Ancestral lineages and sampling in populations with density-dependent interactions

Madeleine Kubasch[1,2]

[1]Centre de Mathématiques Appliquées (CMAP), Ecole Polytechnique, Palaiseau, France
[2]Institute of Ecology and Environmental Sciences (iEES), Sorbonne Université, Paris, France

March 2, 2025

Abstract

We study a density-dependent Markov jump process describing a population where each individual is characterized by a type, and reproduces at rates depending both on its type and on the population type distribution. First, using an appropriate change in probability, we exhibit a time-inhomogeneous Markov process, the *auxiliary process*, which allows to capture the behavior of a sampled lineage in the population process. This is achieved through a many-to-one formula, which relates the average of a function over ancestral lineages sampled in the population processes to its average over the auxiliary process, yielding a direct interpretation of the underlying survivorship bias. In addition, this construction allows for more general sampling procedures than what was previously obtained in the literature, such as sampling restricted to subpopulations. Second, we consider the large population regime, when the population size grows to infinity. Under classical assumptions, the population type distribution can then be approached by a diffusion approximation, which captures the fluctuations of the population process around its deterministic large population limit. We establish a many-to-one formula allowing to sample in the diffusion approximation, and quantify the associated approximation error.

Keywords. Interactions; Markov jump process; population process; many-to-one formula; diffusion approximation.

1 Introduction

When considering population processes arising in various fields such as population genetics or epidemiology, the study of ancestral lineages may provide crucial information. For instance, such lineages yield insight on epidemic spread through contamination chains [14], or on the evolution of a trait of interest under selection [11]. As a consequence, several methods have been developed to finely characterize those lineages. On the one hand, a classical approach is to consider a backward-in-time process which reconstructs the genealogy by moving back from time t to time 0, and which is related to the initial population process by duality [8, 1, 22, 12, 14]. On the other hand, there also exists a forward-in-time approach which relies on a second population process, with one distinguished individual (the *spine*) whose lineage behaves as the lineage of a sampled individual in the original process.

More precisely, these *spinal constructions* have originally been introduced for branching processes, using an appropriate change in probability. The key to the construction of the spinal process is that the reproduction rates of the spine are biased towards leaving more numerous descendants than other individuals, making emerge size-biased distributions. Generally speaking, the obtained spinal construction has several strengths. Notably, it allows to establish *many-to-one* formulas (*e.g.* [21, 16, 17, 18]), which are closely related to Feynman-Kac path equations [13, Sections 1.3 and 1.4.4]. Many-to-one formulas translate the average value of a functional evaluated over the lineages in the branching process, into the average of the functional evaluated along the spine, whose trajectories are exponentially weighted to capture the growth of the population. If the exponential weight is deterministic, this immediately implies a numerical advantage for computing such averages through Monte-Carlo simulations. Indeed, simulations of the spine are numerically affordable, whereas simulations of the whole genealogical tree in the original branching process can be numerically challenging [25]. Also, spinal constructions have proven an effective way of establishing classical key results on branching processes, such as the Kesten Stigum theorem [21, 16]. More recently, the semi-group associated to the spinal construction has proven a successful tool in the analysis of non-conservative semi-groups, extending its applications beyond branching processes [4, 5].

While many models for population dynamics arising, for instance, in biology and epidemiology do not satisfy *per se* the branching approximation, a classical approach is to consider regimes in which the population process can be well approached by a branching process, using coupling arguments. For example, in epidemiology, it is well-known that at the beginning of an epidemic, the tree of infections can be captured by a branching process which neglects the depletion in susceptible individuals [2]. Similarly, in order to analyze the lineage of a uniformly sampled individual in a population which is subject to evolution under a changing environment, [11] consider the stationary regime. However, such branching approximations are restricted to specific parts of the dynamics of interest only; see for instance [7] and [6] for details in the case of epidemic models and invasion processes.

In order to address this limitation, there have recently been developments towards capturing the ancestral lineage of a sampled individual, as well as the whole genealogical tree, in populations with interactions. In [3] a spinal construction is developed for this setting, focusing on multi-type processes with discrete type space. The general idea consists in biasing the reproduction rates of the process, both along the spine and outside of it, according to a positive function ψ of the reproducing particle's trait x and the population's type composition $\mathbf{z}$. Intuitively, $\psi(x, \mathbf{z})$ can be regarded as the individual's reproductive value or long-term fertility. Hence, when the spine reproduces, descendances with higher values of ψ given the population state are favored, while the descendances of individuals outside of the spine are biased towards rendering the population more favorable for the spine. Further, this spinal construction has been extended to include more general type spaces [24].

Nevertheless, the ψ-spine construction has two possible drawbacks:

(i) Assuming $\psi > 0$ can be restrictive, as it prohibits for instance sampling restricted to subpopulations. Indeed, this is a pertinent setting arising in applications. For instance, in epidemiology, chains of infection leading to the contamination of vulnerable individuals, such as the elderly or immunocompromised, are naturally of particular interest. In this context, the motivation for a spinal construction designed for targeted sampling is two-fold. First, the dynamics of the spinal lineage may open the door to an intuitive understanding of such contamination chains. Second, when considering the ψ-spine under the condition that $\psi > 0$, sampling at time t restricted to a given subpopulation would amount to sampling within the whole population and subsequently discarding all the trajectories which do not reach the target subpopulation at time t. If the many-to-one formula is to be evaluated by Monte-Carlo methods, this can be computationally expensive if the subpopulation is rare or rarely visited by the spine.

(ii) In addition, in the many-to-one formula, spinal trajectories are penalized by an exponential weight, which depends on a functional of the spine's type and population state (*cf.* Section

2.2). In particular, this exponential weight is not necessarily deterministic. In this case, the interpretation of the spinal process is less straight-forward, as this penalization needs to be taken into account. Further, Monte-Carlo estimations of the many-to-one formula become delicate as rare trajectories may have a tremendous impact.

As a consequence, we aim at proposing an alternative spinal construction which relaxes the positivity assumption (i), and an associated many-to-one formula which does not require exponential weighting of trajectories (ii). This will be achieved by making use of a different change in probability inspired by [23].

Further, a natural regime to consider is the large population limit $K \to +\infty$. Indeed, under classical assumptions, the population type distribution then converges to the unique solution of a dynamical system [15, Chapter 11]. In this setting, the ψ-spine gives rise to a time-inhomogeneous branching process, which describes the lineage of a uniformly sampled individual in the deterministic regime [3]. However, while real populations may frequently be large, they are never of infinite size. It thus is relevant to capture fluctuations of the population process for large but finite population sizes K, around their deterministic limit. Under classical assumptions, this can be achieved using diffusion approximations and the associated approximation error can be quantified using strong approximations of Poisson processes by Brownian motions [20]. In this paper, we derive a many-to-one formula for sampling in the diffusion approximation, and quantify the approximation error.

This paper is structured as follows. The population process of interest is defined in Section 2, and we recall the associated ψ-spine construction introduced by [3]. In Section 3, we introduce a new spinal construction which addresses issues (i) and (ii) mentioned above, whereas Section 4 focuses on sampling in the diffusion approximation. Finally, Section 5 presents a discussion on our results.

2 The population process and its ψ-spine construction

2.1 The population process

We consider a structured population, where each individual has a type $x \in \mathcal{X}$, and we assume that the type space $\mathcal{X}$ is finite. The number of individuals of type x in the population is referred to as $\mathbf{z}_x$, and the corresponding vector $\mathbf{z}$ describes the composition of the population. Here, we will assume that the population size cannot exceed K individuals (carrying capacity, absence of demographic births and deaths, etc.). Thus

$$\mathbf{z} \in \mathcal{Z}_K = \{\mathbf{z} \in (\mathbb{N} \cup \{0\})^{\mathcal{X}} : \|\mathbf{z}\|_1 \leqslant K\}.$$

Further, individuals will reproduce at rates depending on their type and the current population state. More precisely, an individual of type x may be replaced by an offspring $\mathbf{k} = (\mathbf{k}_y, y \in \mathcal{X}) \in \mathcal{Z}_K$, meaning that the individual dies and for any $y \in \mathcal{X}$, $\mathbf{k}_y$ individuals of type y are born. This occurs at rate $\tau_{\mathbf{k}}(x, \mathbf{z})$. We suppose $\tau(x, \mathbf{z}) = \sum_{\mathbf{k}} \tau_{\mathbf{k}}(x, \mathbf{z}) < \infty$ for all $x \in \mathcal{X}$ and $\mathbf{z} \in \mathcal{Z}_K$. Further, let $(\mathbf{e}(x) : x \in \mathcal{X})$ be the canonical base of $\mathcal{Z}_K$, in the sense that for $x \in \mathcal{X}$, the only non-zero component of $\mathbf{e}(x)$ is its x component which equals one. Then, as the population size is bounded by K,

$$\tau_{\mathbf{k}}(x, \mathbf{z}) = 0 \quad \text{if} \quad \|\mathbf{z} + \mathbf{k} - \mathbf{e}(x)\|_1 > K.$$

In order to keep track of the genealogy, we will make use of the Ulam-Harris-Neveu notations. Let $\mathcal{U} = \bigcup_{k \geqslant 1} \mathbb{N}^k$, then $u = (u_1, \ldots, u_k) \in \mathcal{U}$ represents the u_k-th descendent of $(u_1, \ldots, u_{k-1})$ and for $u, v \in U$ we write $u \succeq v$ if v is an ancestor of u . The type of $u \in \mathcal{U}$ will be called x_u. Hence when an individual u is replaced by its offspring $\mathbf{k}$, the new individuals are $(u, 1), \ldots, (u, \|\mathbf{k}\|_1)$

and we need to decide the type of each descendent. We thus consider a probability distribution $\mathcal{Q}_{\mathbf{k}}$ on

$$\mathcal{X}_{\mathbf{k}} = \{\mathbf{x} \in \mathcal{X}^{\|k\|_1} : \forall x \in \mathcal{X}, \#\{i : \mathbf{x}_i = x\} = \mathbf{k}_x\},$$

and $(x_{(u,i)} : i \in [\![1, \|k\|_1]\!])$ is distributed as $\mathcal{Q}_{\mathbf{k}}$.

Let us now introduce the stochastic process of interest. Intuitively, it corresponds to describing the set of individuals alive and their types, at each time $t \geqslant 0$. In order to give its definition, we need the following notations. Consider the set $E = \{(u, x_u) : u \in \mathcal{U}, x_u \in \mathcal{X}\}$, then $\mathcal{M}_{P,K}(E)$ is defined as the set of positive point measures μ on E such that $\mu(E) \leqslant K$. The set of atoms of μ is given by $A(\mu) = \{(u, x_u) : \mu(\{(u, x_u)\}) > 0\}$ and we define $\mathfrak{g}(\mu) \subset \mathcal{U}$ as the set of labels of its atoms, and $\mathbf{z}(\mu) \in \mathcal{Z}_K$ as the corresponding type composition.

We start from an initial set of individuals $\mathbb{G}(0) = \mathfrak{g} \subset \mathbb{N}$, and the population will evolve as explained above. At each time t, let $\mathbb{G}(t) \subset \mathcal{U}$ be the set of individuals alive. The process of interest $(X(t), t \geqslant 0)$ is a Markov jump process with càdlàg trajectories, which can be defined informally by

$$\forall t \geqslant 0, \quad X(t) = \sum_{u \in \mathbb{G}(t)} \delta_{(u,x_u)} \in \mathcal{M}_{P,K}(E).$$

In particular, notice that there cannot be explosion, since there are at most K individuals reproducing at rate less than $\max_{x \in \mathcal{X}, \mathbf{z} \in \mathcal{Z}_K} \tau(x, \mathbf{z})$, which is a finite bound as $\mathcal{X}$ and $\mathcal{Z}_K$ are finite sets. Finally, $Z(t) = \mathbf{z}(X(t))$ yields the composition of the population at time t, and for $u \in \mathbb{G}(t)$ and $s \leqslant t$, $x_u(s)$ stands for the type of the unique ancestor of u alive at time s.

2.2 Reminders on the ψ-spine construction for interacting populations

Now that the population process is properly introduced, let us briefly summarize some of the results obtained in [3] which will be stated in our setting. In particular, we will restrict ourselves to understanding the evolution of the types of the ancestors of a typical individual, and of the population composition over time, instead of dealing with the entire genealogical tree.

Define the set

$$\mathcal{S}_K = \{(x, \mathbf{z}) \in \mathcal{X} \times \mathcal{Z}_K : \mathbf{z}_x \geqslant 1\}. \tag{1}$$

For any $(x, \mathbf{z}) \in \mathcal{S}_K$, we consider an (arbitrary) labeling $\mathfrak{g}(\mathbf{z}) \subset \mathbb{N}$ of individuals and fix $u_x \in \mathfrak{g}(\mathbf{z})$ such that u_x is of type x. We define the corresponding population state

$$\mathfrak{X}(x, \mathbf{z}) = \sum_{u \in \mathfrak{g}(\mathbf{z})} \delta_{(u,x_u)}.$$

We let $\mathbb{E}_{\mathfrak{X}(x,\mathbf{z})}$ and $\mathbb{P}_{\mathfrak{X}(x,\mathbf{z})}$ designate the expectation and probability conditionally on $X(0) = \mathfrak{X}(x, \mathbf{z})$. With these notations,

$$M_t f(x, \mathbf{z}) = \mathbb{E}_{\mathfrak{X}(x,\mathbf{z})}[\sum_{u \in \mathbb{G}(t), u \succeq u_x} f(x_u(t), Z(t))]$$

is a semi-group whose generator $\mathcal{G}$ is defined by its action on functions $f : \mathcal{S}_K \to \mathbb{R}_+$ as follows. For any $(x, \mathbf{z}) \in \mathcal{S}_K$,

$$\begin{aligned} \mathcal{G}f(x, \mathbf{z}) = &\sum_{\mathbf{k} \in \mathcal{Z}_K} \tau_{\mathbf{k}}(x, \mathbf{z}) \left(\sum_{y \in \mathcal{X}} \mathbf{k}_y f(y, \mathbf{z} + \mathbf{k} - \mathbf{e}(x)) - f(x, \mathbf{z}) \right) \\ &+ \sum_{\substack{y \in \mathcal{X} \\ \mathbf{k} \in \mathcal{Z}_K}} (\mathbf{z}_y - \mathbf{1}_{\{x=y\}}) \tau_{\mathbf{k}}(y, \mathbf{z})(f(x, \mathbf{z} + \mathbf{k} - \mathbf{e}(y)) - f(x, \mathbf{z})). \end{aligned} \tag{2}$$

Consider a positive function $\psi : \mathcal{S}_K \to (0, +\infty)$ and let $\lambda = \mathcal{G}\psi/\psi$. The ψ-spine construction corresponds to the stochastic process $(Y(s), \zeta(s))_{s \geqslant 0}$ whose generator $\mathcal{L}$ is defined as follows. For any $f : \mathcal{S}_K \to \mathbb{R}_+$,

$$\mathcal{L}f = \frac{\mathcal{G}(\psi f)}{\psi} - \lambda f.$$

Generally speaking, the idea is that Y yields the evolution of the type along the lineage of a distinguished individual, the spine, whose reproduction rates are biased by $\psi(\cdot, \zeta)$, where ζ describes the spinal population's composition. Similarly, reproduction rates outside of the spine are also biased, favoring descendances which create a population state in which the spine is more fertile. The interest in this ψ-spinal process is illustrated by the following many-to-one formula. Let $(x, \mathbf{z}) \in \mathcal{S}_K$ and consider $\mathbb{E}_{(x,\mathbf{z})}$ the expectation conditionally on the event $(Y(0), \zeta(0)) = (x, \mathbf{z})$. Also, for $t \geqslant 0$, define

$$W(t) = \exp\Big(\int_0^t \lambda(Y(s), \zeta(s))ds\Big). \tag{3}$$

Then for any $t \geqslant 0$ and any measurable function $F : \mathbb{D}([0,t], \mathcal{S}_K) \to \mathbb{R}_+$, for any $(x, \mathbf{z}) \in \mathcal{S}_K$,

$$\mathbb{E}_{\mathfrak{X}(x,\mathbf{z})}\Big[\sum_{\substack{u \in \mathbb{G}(t)\\ u \succeq u_x}} \psi(x_u(t), Z(t)) F((x_u(s), Z(s))_{s \leqslant t})\Big] = \psi(x, \mathbf{z}) \mathbb{E}_{(x,\mathbf{z})}[W(t) F((Y(s), \zeta(s))_{s \leqslant t})]. \tag{4}$$

As mentioned in the introduction, the exponential weight W implies that rare trajectories of the ψ-spine may have a large impact on the right-hand side of Equation (4). This leads to a potential numerical challenge despite the ψ-spine being easier to simulate than a typical lineage of the original population process by itself. The issue vanishes if ψ is an eigenfunction of $\mathcal{G}$, as λ is constant in this case. However, the existence and uniqueness of such a positive eigenfunction is not always satisfied, even for classical models such as the epidemic *SIR* model. This motivates our interest in obtaining a many-to-one formula without exponential weighting the spinal trajectories. Additionally, one may be interested for instance in the typical lineage of individuals who at time t have a specific type x. The positivity assumption $\psi > 0$ does not allow choices such as $\psi = \mathbf{1}_{\{x\}}$, hence this constraint needs to be incorporated in the function F. Again, this emphasizes some potential difficulties for numerically computing the right-hand side of Equation (4), as the event $\{Y(t) = x\}$ may be rare itself, particularly if the type x is rare and/or poorly fertile. This will be addressed in Section 3.

3 The ψ-sampled lineage

In this section, our aim is to propose a many-to-one formula in the line of Equation (4), which relaxes the positivity assumption on ψ and yields a many-to-one formula without exponential penalization of spinal trajectories. In order to do so, we will introduce a change in probability, which yields a time inhomogeneous spinal construction. This approach is inspired by [23], in continuity with which we refer to the obtained time-changed process as the *ψ-auxiliary process.*

3.1 A many-to-one formula

Consider a non-negative function ψ on $\mathcal{X} \times \mathcal{Z}_K$ such that the following application m_ψ is positive on $\mathcal{S}_K \times [0, t]$:

$$m_\psi(x, \mathbf{z}, s) = \mathbb{E}_{\mathfrak{X}(x,\mathbf{z})}\Big[\sum_{\substack{u \in \mathbb{G}(s)\\ u \succeq u_x}} \psi(x_u(t), Z(t))\Big]. \tag{5}$$

Notice that, by the Markov property, for $s \in [0, t]$,

$$m_\psi(x, \mathbf{z}, t-s) = \mathbb{E}\left[\sum_{\substack{u \in \mathbb{G}(t) \\ u \succeq u_x}} \psi(x_u(t), Z(t)) \middle| X(s) = \mathfrak{X}(x, \mathbf{z}) \right].$$

In words, $m_\psi(x, \mathbf{z}, t-s)$ corresponds to the ψ-weighted average of the types of individuals alive at time t who descend from a given individual of type x at time s, given that at time s, the population was in state $\mathbf{z}$. For instance, if $\psi = 1$, this yields the average number of individuals alive at time t, who descend from an individual of type x at time s when the population was in state $\mathbf{z}$. Similarly, if $\psi = \mathbf{1}_{\{y\}}$ for a given $y \in \mathcal{X}$, $m_\psi(x, \mathbf{z}, t-s)$ yields the average number of descendants of type y at time t, starting at time s from an individual of type x in a population of composition $\mathbf{z}$.

Let us now introduce the ψ-auxiliary process, which will allow to capture the behavior of the ancestral lineage of a ψ-weighted sample of the population process. For $t \geqslant 0$ fixed, we will consider the time-inhomogeneous Markov process $(Y^{(t)}(s), \zeta^{(t)}(s))_{s \leqslant t}$ defined as follows. The main idea is to follow the type $Y^{(t)}$ of a distinguished individual, which will be referred to as *spine* in analogy to classical spinal constructions. At time $s \leqslant t$, when of type x in a population of state $\mathbf{z}$, the spine divides to leave descendance $\mathbf{k}$ and switch to type y with rate

$$\rho_{y,\mathbf{k}}^{(t)}(s, x, \mathbf{z}) = \tau_{\mathbf{k}}(x, \mathbf{z}) \mathbf{k}_y \frac{m_\psi(y, \mathbf{z} + \mathbf{k} - \mathbf{e}(x), t-s)}{m_\psi(x, \mathbf{z}, t-s)}.$$

In other words, compared to the original process, at any time $s \leqslant t$, transitions along the distinguished lineage are biased in favor of those which lead to a larger ψ-average descendance at the final time t. However, due to the density-dependence of division rates, it is necessary to keep track of the population state $\zeta^{(t)}$. Again, transitions need to be biased, in order to account for the modified behavior of the distinguished individual when compared to the original process. As a consequence, when the population is in state $\mathbf{z}$ and the spine of type x at time $s \leqslant t$, individuals of type y other than the spine divide to leave descendance $\mathbf{k}$ at rate

$$\widehat{\rho}_{\mathbf{k}}^{(t)}(s, y, x, \mathbf{z}) = \tau_{\mathbf{k}}(y, \mathbf{z}) \frac{m_\psi(x, \mathbf{z} + \mathbf{k} - \mathbf{e}(y), t-s)}{m_\psi(x, \mathbf{z}, t-s)}.$$

Here, the bias favors those transitions which lead to a more favorable environment for the spine, *i.e.* a population composition in which the ψ-average of the spine's descendance is high.

We will now characterize $(Y^{(t)}(s), \zeta^{(t)}(s))_{s \leqslant t}$ as the unique solution of a stochastic differential equation. In order to do so, we let $Y^{(t)}(s) \in \{\mathbf{e}(x) : x \in \mathcal{X}\}$ for any $s \in [0, t]$, where $Y^{(t)}(s) = \mathbf{e}(x)$ means that the spine is of type x. Define $E = \mathbb{R}_+ \times \mathcal{S}_K$, and consider two independent Poisson point processes Q and $\widehat{Q}$ on $\mathbb{R}_+ \times E$, of density $dr \otimes d\theta \otimes n(dy, d\mathbf{k})$ where $dr, d\theta$ designate the Lebesgue measure and n the counting measure on $\mathcal{S}_K$. Here, we assume that Q and $\widehat{Q}$ are defined on the same probability space as and independently from $(Y^{(t)}(0), \zeta^{(t)}(0))$, whose law is supposed to be given. Then, for any $s \in [0, t]$,

$$\begin{aligned}
Y^{(t)}(s) &= Y^{(t)}(0) + \int_0^s \int_E \mathbf{1}_{\left\{\theta \leqslant \rho_{y,\mathbf{k}}^{(t)}(r, Y^{(t)}(r-), \zeta^{(t)}(r-))\right\}} (\mathbf{e}(y) - Y^{(t)}(r-)) Q(dr, d\theta, n(dy, d\mathbf{k})), \\
\zeta^{(t)}(s) &= \zeta^{(t)}(0) + \int_0^s \int_E \mathbf{1}_{\left\{\theta \leqslant \rho_{y,\mathbf{k}}^{(t)}(r, Y^{(t)}(r-), \zeta^{(t)}(r-))\right\}} (\mathbf{k} - Y^{(t)}(r-)) Q(dr, d\theta, n(dy, d\mathbf{k})) \\
&\quad + \int_0^s \int_E \mathbf{1}_{\left\{\theta \leqslant (\zeta_y^{(t)}(r-) - \mathbf{1}_{\left\{Y^{(t)}(r-) = y\right\}}) \widehat{\rho}_{\mathbf{k}}^{(t)}(r, y, Y^{(t)}(r-), \zeta^{(t)}(r-))\right\}} (\mathbf{k} - \mathbf{e}(y)) \widehat{Q}(dr, d\theta, n(dy, d\mathbf{k})).
\end{aligned} \tag{6}$$

Remark 3.1. *Throughout the following, in order to simplify notations, we will make no distinction between the sets* $\mathcal{X}$ *and* $\{\mathbf{e}(x) : x \in \mathcal{X}\}$*, based on the natural bijection between the two sets. For example,* $Y^{(t)}(s) = x$ *is equivalent to* $Y^{(t)}(s) = \mathbf{e}(x)$*. Similarly, to every real-valued function* f *on* S_K*, we assign a function* $\hat{f}$ *on* $\{\mathbf{e}(x) : x \in \mathcal{X}\}$ *by* $\hat{f}(\mathbf{e}(x)) = f(x)$*, the application* $f \mapsto \hat{f}$ *being a bijection between the sets of real-valued functions on* $\mathcal{S}_K$ *and on* $\{\mathbf{e}(x) : x \in \mathcal{X}\}$*. Thus, we will always consider* $Y^{(t)}$ *to take values in the Skorokhod space* $\mathbb{D}([0,t], \mathcal{X})$*, unless mentioned otherwise.*

Our first result shows that the process $(Y^{(t)}(s), \zeta^{(t)}(s))_{s \leqslant t}$ is now well defined, and additionally provides its semi-group $\mathcal{R}^{(t)} = (\mathcal{R}^{(t)}_{r,s}, r \leqslant s \leqslant t)$. We recall that the latter is characterized by its action on non-negative functions f on $\mathcal{S}_K$: for $r \leqslant s \leqslant t$ and $(x, \mathbf{z}) \in \mathcal{S}_K$,

$$\mathcal{R}^{(t)}_{r,s}(x, \mathbf{z}) = \mathbb{E}[f(Y^{(t)}(s), \zeta^{(t)}(s)) | (Y^{(t)}(r), \zeta^{(t)}(r)) = (x, \mathbf{z})].$$

Proposition 3.2. *Equation* (6) *admits a unique strong solution* $(Y^{(t)}, \zeta^{(t)})$ *in the Skorokhod space* $\mathbb{D}([0,t], \mathcal{S}_K)$*. Its semi-group* $\mathcal{R}^{(t)}$ *is defined by:*

$$\forall 0 \leqslant r \leqslant s \leqslant t, \quad \mathcal{R}^{(t)}_{r,s} = \mathrm{e}^{\int_r^s A^{(t)}_\tau d\tau}, \tag{7}$$

where the operator $A^{(t)}$ *is characterized by its action on non-negative functions* f *on* $\mathcal{S}_K$*. For any* $s \in [0, t)$ *and* $(x, \mathbf{z}) \in \mathcal{S}_K$*,*

$$A^{(t)}_s f(x, \mathbf{z}) = m_\psi(x, \mathbf{z}, t-s)^{-1} \left(\mathcal{G}(m_\psi(\cdot, t-s) f(\cdot))(x, \mathbf{z}) - \mathcal{G}(m_\psi(\cdot, t-s))(x, \mathbf{z}) f(x, \mathbf{z}) \right). \tag{8}$$

Notice that the operator $A^{(t)}$ corresponds to the generator of the semi-group $\mathcal{R}^{(t)}$. As our state space $\mathcal{S}_K$ is finite, $A^{(t)}_\tau$ can be represented as a matrix, whose elements correspond to the instantaneous transition rates at time τ which can be recovered by taking $f = \mathbf{1}_{\{(y,\mathbf{v})\}}$ for $(y, \mathbf{v}) \in \mathcal{S}_K$. In particular, this ensures that the generator $A^{(t)}$ uniquely characterizes the semi-group $\mathcal{R}^{(t)}$, and thus the Markov process $(Y^{(t)}, \zeta^{(t)})$. The proof of the proposition is postponed to Section 3.2.

We are now ready to state our main result. With slight abuse of notation, $\mathbb{E}_{(x,\mathbf{z})}$ will designate the expectation conditionally on the event $(Y^{(t)}(0), \zeta^{(t)}(0)) = (x, \mathbf{z})$.

Theorem 3.3. *For any* $t \geqslant 0$ *and any measurable function* $F : \mathbb{D}([0,t], \mathcal{S}_K) \to \mathbb{R}_+$*, for any* $(x, \mathbf{z}) \in \mathcal{S}_K$*,*

$$\mathbb{E}_{\mathfrak{X}(x,\mathbf{z})}[\sum_{\substack{u \in \mathbb{G}(t) \\ u \succeq u_x}} \psi(x_u(t), Z(t)) F((x_u(s), Z(s))_{s \leqslant t})] = m_\psi(x, \mathbf{z}, t) \mathbb{E}_{(x,\mathbf{z})}[F((Y^{(t)}(s), \zeta^{(t)}(s))_{s \leqslant t})]. \tag{9}$$

Before proceeding to the proof of Theorem 3.3, let us compare the obtained ψ-auxiliary process with the ψ-spine from [3]. First, we may notice that both spinal constructions remain similar in spirit, as ours may be regarded as a *time-inhomogeneous* m_ψ*-transform*, instead of the classical ψ-transform. Second, in the special case where ψ is an eigenfunction of the generator $\mathcal{G}$ introduced above, a brief computation shows that Equation (9) amounts to the Feynman-Kac formula of Equation (4), as one would expect.

3.2 Proofs

The general idea is to proceed as follows. We start by showing that the ψ-auxiliary process is well defined, by Proposition 3.2, and compute its generator $A^{(t)}$. Next, we introduce a time-inhomogeneous semi-group corresponding essentially to the left-hand side of the many-to-one formula given by Equation (9), and show that its generator is equal to $A^{(t)}$. As mentioned previously, the considered state space being finite, the generator uniquely characterizes the time-inhomogeneous semi-group. This finally allows to establish Theorem 3.3.

3.2.1 Existence and uniqueness of the ψ-auxiliary process

We first establish Proposition 3.2, ensuring that the ψ-auxiliary process is well defined. We start with a technical lemma.

Lemma 3.4. *For any $t \geqslant 0$, for any $(x, \mathbf{z}) \in \mathcal{S}_K$, the function $s \mapsto m_\psi(x, \mathbf{z}, t-s)$ is differentiable on $(0, t)$, and we have:*

$$\partial_s m_\psi(x, \mathbf{z}, t-s) = -\mathcal{G}(m_\psi(\cdot, t-s))(x, \mathbf{z}).$$

Proof. Let $(x, \mathbf{z}) \in \mathcal{S}_K$. Showing that $t \mapsto m_\psi(x, \mathbf{z}, t)$ is differentiable on $\mathbb{R}_+$ and computing its derivative is sufficient, as the desired result follows by composition. Let $t \geqslant 0$ and $h > 0$. The Markov property ensures that

$$m_\psi(x, \mathbf{z}, t+h) = \mathbb{E}_{\mathfrak{X}(x,\mathbf{z})}[\sum_{\substack{u \in \mathbb{G}(h) \\ u \succeq u_x(0)}} m_\psi(x_u(h), Z(h), t)].$$

For $i \geqslant 1$, Let T_i be the time of the i-th jump of the population process. Then on the one hand, if $T_1 > h$, then the population at time h is identical to the population at time 0, and thus:

$$m_\psi(x, \mathbf{z}, t+h) = \mathbb{E}_{\mathfrak{X}(x,\mathbf{z})}[\sum_{\substack{u \in \mathbb{G}(h) \\ u \succeq u_x(0)}} m_\psi(x_u(h), Z(h), t)\mathbf{1}_{\{T_1 < h\}}] + m_\psi(x, \mathbf{z}, t)\mathbb{P}_{\mathbf{z}}(T_1 < h).$$

Similarly, on the event $\{T_1 < h < T_2\}$, $Z(h) = Z(T_1)$ whence

$$\mathbb{E}_{\mathfrak{X}(x,\mathbf{z})}[\sum_{\substack{u \in \mathbb{G}(h) \\ u \succeq u_x(0)}} m_\psi(x_u(h), Z(h), t)\mathbf{1}_{\{T_1 < h\}}] = a(h) + b(h),$$

where

$$a(h) = \mathbb{E}_{\mathfrak{X}(x,\mathbf{z})}[\sum_{\substack{u \in \mathbb{G}(h) \\ u \succeq u_x(0)}} m_\psi(x_u(T_1), Z(T_1), t)\mathbf{1}_{\{T_1 < h < T_2\}}],$$

$$b(h) = \mathbb{E}_{\mathfrak{X}(x,\mathbf{z})}[\sum_{\substack{u \in \mathbb{G}(h) \\ u \succeq u_x(0)}} m_\psi(x_u(h), Z(h), t)\mathbf{1}_{\{T_2 < h\}}].$$

As a consequence, we obtain that

$$m_\psi(x, \mathbf{z}, t+h) - m_\psi(x, \mathbf{z}, t) = A(h) + B(h), \tag{10}$$

with $A(h) = a(h) - m_\psi(x, \mathbf{z}, t)\mathbb{P}_{\mathbf{z}}(T_1 < h < T_2)$ and $B(h) = b(h) - m_\psi(x, \mathbf{z}, t)\mathbb{P}_{\mathbf{z}}(T_2 < h)$.

Let us first focus on $B(h)$. For any $t \geqslant 0$ and $(y, \mathbf{v}) \in \mathcal{S}_K$, it holds that $m_\psi(y, \mathbf{v}, t) \leqslant K \|\psi\|_\infty$. As $\mathcal{S}_K$ is a finite set, it follows that there exists a constant $c > 0$ such that

$$B(h) \leqslant c\mathbb{P}_{\mathbf{z}}(T_2 < h).$$

For $\mathbf{v} \in \mathcal{Z}_K$, let us write $\Lambda(\mathbf{v}) = \sum_{y \in \mathcal{X}} \sum_{\mathbf{k} \in \mathcal{Z}_K} \mathbf{v}_y \tau_{\mathbf{k}}(y, \mathbf{v})$ for the total jump rate in a population whose type distribution is given by $\mathbf{v}$. In particular, Λ is bounded on $\mathcal{Z}_K$. Using the law of T_1 given $Z(0) = \mathbf{z}$ and the law of $T_2 - T_1$ given T_1 and $Z(T_1)$, we then obtain:

$$\mathbb{P}_{\mathbf{z}}(T_2 < h) = \int_0^h e^{-\Lambda(\mathbf{z})t_1} \sum_{\substack{y \in \mathcal{X} \\ \mathbf{k} \in \mathcal{Z}_K}} \mathbf{z}_y \tau_{\mathbf{k}}(y, \mathbf{z}) \int_0^{h-t_1} \Lambda(\mathbf{z} + \mathbf{k} - \mathbf{e}(y))\mathrm{e}^{-\Lambda(\mathbf{z}+\mathbf{k}-\mathbf{e}(y))t_2} dt_2 dt_1 \leqslant \frac{\|\Lambda\|_\infty^2}{2} h^2.$$

We deduce that

$$\frac{B(h)}{h} \xrightarrow[h\to 0^+]{} 0. \tag{11}$$

Let us now focus on $A(h)$. Proceeding in the same way, we have

$$A(h) = \int_0^h e^{-\Lambda(\mathbf{z})t_1} e^{-\Lambda(\mathbf{z}+\mathbf{k}-\mathbf{e}(x))(h-t_1)} dt_1 \sum_{\mathbf{k}\in\mathcal{Z}_K} \tau_{\mathbf{k}}(x,\mathbf{z}) \sum_{y\in\mathcal{X}} \mathbf{k}_y (m_\psi(y,\mathbf{z}+\mathbf{k}-\mathbf{e}(x),t) - m_\psi(x,\mathbf{z},t))$$
$$+ \sum_{\substack{y\in\mathcal{X}\\ \mathbf{k}\in\mathcal{Z}_K}} \int_0^h e^{-\Lambda(\mathbf{z})t_1} e^{-\Lambda(\mathbf{z}+\mathbf{k}-\mathbf{e}(y))(h-t_1)} dt_1 \, \tau_{\mathbf{k}}(y,\mathbf{z})(m_\psi(x,\mathbf{z}+\mathbf{k}-\mathbf{e}(y),t) - m_\psi(x,\mathbf{z},t)),$$

from which it follows that

$$\begin{aligned}\frac{A(h)}{h} \xrightarrow[h\to 0^+]{} & \sum_{\mathbf{k}\in\mathcal{Z}_K} \tau_{\mathbf{k}}(x,\mathbf{z}) \sum_{y\in\mathcal{X}} \mathbf{k}_y (m_\psi(y,\mathbf{z}+\mathbf{k}-\mathbf{e}(x),t) - m_\psi(x,\mathbf{z},t)) \\ & + \sum_{\substack{y\in\mathcal{X}\\ \mathbf{k}\in\mathcal{Z}_K}} \tau_{\mathbf{k}}(y,\mathbf{z})(m_\psi(x,\mathbf{z}+\mathbf{k}-\mathbf{e}(y),t) - m_\psi(x,\mathbf{z},t)).\end{aligned} \tag{12}$$

As a consequence, right differentiability of $t \mapsto m_\psi(x,\mathbf{z},t)$ is established by Equations (10), (11) and (12), and its right derivative is given by the right-hand side of Equation (12). As this corresponds to a continuous function on $\mathbb{R}_+$, we deduce that $t \mapsto m_\psi(x,\mathbf{z},t)$ is differentiable on $\mathbb{R}_+$ (see *e.g.* Corollary 1.2 of Chapter 2 in [27]) and

$$\begin{aligned}\frac{d}{dt} m_\psi(x,\mathbf{z},t) = & \sum_{\mathbf{k}\in\mathcal{Z}_K} \tau_{\mathbf{k}}(x,\mathbf{z}) \sum_{y\in\mathcal{X}} \mathbf{k}_y (m_\psi(y,\mathbf{z}+\mathbf{k}-\mathbf{e}(x),t) - m_\psi(x,\mathbf{z},t)) \\ & + \sum_{\substack{y\in\mathcal{X}\\ \mathbf{k}\in\mathcal{Z}_K}} \tau_{\mathbf{k}}(y,\mathbf{z})(m_\psi(x,\mathbf{z}+\mathbf{k}-\mathbf{e}(y),t) - m_\psi(x,\mathbf{z},t)).\end{aligned}$$

This concludes the proof. □

We are now ready to establish the desired result.

Proof of Proposition 3.2. The proof is decomposed in three steps, establishing (i) existence and (ii) uniqueness of the solution to Equation (6) by classical arguments, before (iii) characterizing the associated semi-group $\mathcal{R}^{(t)}$.

For ease of notation, throughout the proof, for $0 \leqslant s \leqslant t$ we let $\mathbb{Y}^{(t)}(s) = (Y^{(t)}(s), \zeta^{(t)}(s))$.

(i) *Existence.* First, notice that by assumption on $\tau(x,\mathbf{z})$ for $(x,\mathbf{z}) \in \mathcal{S}_K$ and continuity of m_ψ, both applications $\rho^{(t)}_{y,\mathbf{k}}$ and $\hat{\rho}^{(t)}_{\mathbf{k}}$ are bounded for any $y \in \mathcal{X}$ and $\mathbf{k} \in \mathcal{Z}_K$. As a consequence, existence of at least one solution to Equation (6) is ensured, as the associated sequence of jump times $(T_k)_{k\geqslant 0}$ cannot admit an accumulation point on $\mathbb{R}_+$.

(ii) *Uniqueness.* Subsequently, in order to establish uniqueness, let us show by induction that for any $k \geqslant 0$ such that $T_k \leqslant t$, $(T_k, \mathbb{Y}^{(t)}(T_k))$ is entirely determined by $(\mathbb{Y}^{(t)}(0), Q, \widehat{Q})$. As $T_0 = 0$, initialization of the induction argument is immediate. If the property holds for $k \geqslant 1$, then by construction, T_{k+1} only depends on $(T_k, \mathbb{Y}^{(t)}(T_k), Q, \widehat{Q})$. Similarly, given T_{k+1} and the corresponding atoms A_{k+1} and $\widehat{A}_{k+1}$ of Q and $\widehat{Q}$, it is clear that $\mathbb{Y}^{(t)}(T_{k+1})$ is fixed by $(T_{k+1}, A_{k+1}, \widehat{A}_{k+1}, \mathbb{Y}^{(t)}(T_k))$. The desired conclusion thus is a consequence of the induction hypothesis.

(ii) *Characterization of* $\mathcal{R}^{(t)}$. In order to establish Equation (7), it is sufficient to show that for any non-negative function f on $\mathcal{S}_K$ and $(x,\mathbf{z}) \in \mathcal{S}_K$, the function

$$\tau \mapsto \mathcal{R}^{(t)}_{s,\tau} f(x,\mathbf{z}) = \mathbb{E}[f(\mathbb{Y}^{(t)}(\tau)) | \mathbb{Y}^{(t)}(s) = (x,\mathbf{z})]$$

is right differentiable at $\tau = s$. Indeed, it then follows that Equation (7) holds with the operator $\mathcal{A}^{(t)}$ defined by

$$\forall f : \mathcal{S}_K \to \mathbb{R}_+ \ \forall (x,\mathbf{z}) \in \mathcal{S}_K, \quad \mathcal{A}^{(t)}_s f(x,\mathbf{z}) = \lim_{h \to 0+} \frac{1}{h} \left(\mathcal{R}^{(t)}_{s,s+h} f(x,\mathbf{z}) - f(x,\mathbf{z}) \right). \tag{13}$$

As we will see, computing the right-hand side of Equation (13) leads to Equation (8).

Let $f : \mathcal{S}_K \to \mathbb{R}_+$ and $(x,\mathbf{z}) \in \mathcal{S}_K$. We introduce the following notations. For any $(y,\mathbf{k}) \in \mathcal{S}_K$ such that $\tau_{\mathbf{k}}(x,\mathbf{z}) > 0$,

$$\mathfrak{d}_{y,\mathbf{k}} f(x,\mathbf{z}) = f(y, \mathbf{z} + \mathbf{k} - \mathbf{e}(x)) - f(x,\mathbf{z}).$$

Further, for any $y \in \mathcal{X}$ such that $\mathbf{z}_y > 0$ and $\mathbf{k} \in \mathcal{Z}_K$ such that $\tau_{\mathbf{k}}(y,\mathbf{z}) > 0$, let

$$\widehat{\mathfrak{d}}_{y,\mathbf{k}} f(x,\mathbf{z}) = f(x, \mathbf{z} + \mathbf{k} - \mathbf{e}(y)) - f(x,\mathbf{z}).$$

Equation (6) then ensures that, on the event $\mathbb{Y}^{(t)}(s) = (x,\mathbf{z})$, we have for any $h \in [0, t-s]$:

$$\begin{aligned} f(\mathbb{Y}^{(t)}(s+h)) - f(x,\mathbf{z}) &= \int_s^{s+h} \int_E \mathbf{1}_{\left\{\theta \leqslant \rho^{(t)}_{y,\mathbf{k}}(r, \mathbb{Y}^{(t)}(r-))\right\}} \mathfrak{d}_{y,\mathbf{k}} f(\mathbb{Y}^{(t)}(r-)) Q(dr, d\theta, n(dy, d\mathbf{k})) \\ &+ \int_s^{s+h} \int_E \mathbf{1}_{\left\{\theta \leqslant (\zeta^{(t)}_y(r-) - \mathbf{1}_{\{Y^{(t)}(r-) = y\}}) \widehat{\rho}^{(t)}_{\mathbf{k}}(r, y, \mathbb{Y}^{(t)}(r-))\right\}} \widehat{\mathfrak{d}}_{y,\mathbf{k}} f(\mathbb{Y}^{(t)}(r-)) \widehat{Q}(dr, d\theta, n(dy, d\mathbf{k})). \end{aligned}$$

Notice that, for instance,

$$\begin{aligned} &\mathbb{E}\Big[\int_s^{s+h} \int_E \mathbf{1}_{\left\{\theta \leqslant \rho^{(t)}_{y,\mathbf{k}}(r, \mathbb{Y}^{(t)}(r-))\right\}} \mathfrak{d}_{y,\mathbf{k}} f(\mathbb{Y}^{(t)}(r-)) Q(dr, d\theta, n(dy, d\mathbf{k})) | \mathbb{Y}^{(t)}(s) = (x,\mathbf{z})\Big] \\ &\qquad = \mathbb{E}\Big[\int_s^{s+h} \sum_{(y,\mathbf{k}) \in \mathcal{S}_K} \rho^{(t)}_{y,\mathbf{k}}(r, \mathbb{Y}^{(t)}(r)) \mathfrak{d}_{y,\mathbf{k}} f(\mathbb{Y}^{(t)}(r)) dr | \mathbb{Y}^{(t)}(s) = (x,\mathbf{z})\Big]. \end{aligned}$$

On the one hand, almost surely,

$$\lim_{h \to 0+} \frac{1}{h} \int_s^{s+h} \sum_{(y,\mathbf{k}) \in \mathcal{S}_K} \rho^{(t)}_{y,\mathbf{k}}(r, \mathbb{Y}^{(t)}(r)) \mathfrak{d}_{y,\mathbf{k}} f(\mathbb{Y}^{(t)}(r)) dr = \sum_{(y,\mathbf{k}) \in \mathcal{S}_K} \rho^{(t)}_{y,\mathbf{k}}(s, \mathbb{Y}^{(t)}(s)) \mathfrak{d}_{y,\mathbf{k}} f(\mathbb{Y}^{(t)}(s)).$$

On the other hand, as mentioned at the beginning of the proof,

$$\|\rho^{(t)}\|_\infty = \max_{s \in [0,t], (y,\mathbf{k}) \in \mathcal{S}, (x,\mathbf{z}) \in \mathcal{S}_K} \rho^{(t)}_{y,\mathbf{k}}(s, x, \mathbf{z}) < \infty.$$

Further, as $\mathcal{S}_K$ is a finite set, $\|f\|_\infty = \max_{(x,\mathbf{z}) \in \mathcal{S}_K} f(x,\mathbf{z}) < \infty$. Thus, for any $h \in [0, t-s]$,

$$\frac{1}{h} \int_s^{s+h} \left| \sum_{(y,\mathbf{k}) \in \mathcal{S}_K} \rho^{(t)}_{y,\mathbf{k}}(r, \mathbb{Y}^{(t)}(r)) \mathfrak{d}_{y,\mathbf{k}} f(\mathbb{Y}^{(t)}(r)) \right| dr \leqslant 2 \mathrm{Card}(\mathcal{S}_K) \|\rho^{(t)}\|_\infty \|f\|_\infty < \infty.$$

Taken together, we obtain by dominated convergence:

$$\begin{aligned} &\frac{1}{h} \mathbb{E}\Big[\int_s^{s+h} \int_E \mathbf{1}_{\left\{\theta \leqslant \rho^{(t)}_{y,\mathbf{k}}(r, \mathbb{Y}^{(t)}(r-))\right\}} \mathfrak{d}_{y,\mathbf{k}} f(\mathbb{Y}^{(t)}(r-)) Q(dr, d\theta, n(dy, d\mathbf{k})) | \mathbb{Y}^{(t)}(s) = (x,\mathbf{z})\Big] \\ &\qquad \xrightarrow[h \to 0+]{} \sum_{(y,\mathbf{k}) \in \mathcal{S}_K} \rho^{(t)}_{y,\mathbf{k}}(s, x, \mathbf{z}) \mathfrak{d}_{y,\mathbf{k}} f(x,\mathbf{z}). \end{aligned}$$

The other terms arising on the right-hand side of Equation (13) can be treated analogously. This leads to the desired result. $\square$

3.2.2 Proof of the many-to-one formula

We are now ready to turn to the proof of Theorem 3.3, which comprises several steps. Let $t \geqslant 0$, and start by introducing the time-inhomogeneous semi-group of interest $\mathcal{P}^{(t)} = (\mathcal{P}^{(t)}_{r,s}, r \leqslant s \leqslant t)$ through its action on applications $f : \mathcal{S}_K \to \mathbb{R}_+$. For $s \leqslant t$, $u_x(s)$ will designate a chosen individual of type x in $\mathbb{G}(s)$, if it exists. For any $(x, \mathbf{z}) \in \mathcal{S}_K$ and $0 \leqslant r \leqslant s \leqslant t$,

$$\mathcal{P}^{(t)}_{r,s} f(x,\mathbf{z}) = m_\psi(x,\mathbf{z},t-r)^{-1}\mathbb{E}[\sum_{\substack{u\in\mathbb{G}(t)\\ u\succeq u_x(r)}} \psi(x_u(t),Z(t))f(x_u(s),Z(s))|X(r)=\mathfrak{X}(x,\mathbf{z})]. \tag{14}$$

Lemma 3.5. *$(\mathcal{P}^{(t)}_{r,s}, r \leqslant s \leqslant t)$ defines a conservative, time-inhomogeneous semi-group acting on the set of functions $\{f : \mathcal{S}_K \to \mathbb{R}_+\}$.*

Proof. The conservativity of $\mathcal{P}^{(t)}$ follows directly from Equation (14) applied to $f \equiv 1$, which shows that $\mathcal{P}^{(t)}1 \equiv 1$.

Let us now turn to the inhomogeneous semi-group property. Let $r \leqslant \tau \leqslant t$, and consider $f : \mathcal{S}_K \to \mathbb{R}_+$ and $(x, \mathbf{z}) \in \mathcal{S}_K$. Throughout the proof, we let $\mathfrak{X}_0 = \mathfrak{X}(x, \mathbf{z})$. By definition of the semi-group,

$$\begin{aligned}
\mathcal{P}^{(t)}_{r,s} f(x,\mathbf{z}) &= m_\psi(x,\mathbf{z},t-r)^{-1}\mathbb{E}[\sum_{\substack{u\in\mathbb{G}(t)\\ u\succeq u_x(r)}} \psi(x_u(t),Z(t))f(x_u(s),Z(s))|X(r)=\mathfrak{X}_0]\\
&= m_\psi(x,\mathbf{z},t-r)^{-1}\mathbb{E}[\sum_{\substack{v\in\mathbb{G}(\tau)\\ v\succeq u_x(r)}}\sum_{\substack{u\in\mathbb{G}(t)\\ u\succeq v}} \psi(x_u(t),Z(t))f(x_u(s),Z(s))|X(r)=\mathfrak{X}_0]\\
\mathcal{P}^{(t)}_{r,s} f(x,\mathbf{z}) &= m_\psi(x,\mathbf{z},t-r)^{-1}\mathbb{E}[\sum_{\substack{v\in\mathbb{G}(\tau)\\ v\succeq u_x(r)}} g(x_v(\tau),Z(\tau))|X(r)=\mathfrak{X}_0],
\end{aligned} \tag{15}$$

where we define the function $g : \mathcal{S}_K \to \mathbb{R}_+$ by

$$g(x,\mathbf{z}) = \mathbb{E}[\sum_{\substack{u\in\mathbb{G}(t)\\ u\succeq u_x(\tau)}} \psi(x_u(t),Z(t))f(x_u(s),Z(s))|X(\tau)=\mathfrak{X}(x,\mathbf{z})].$$

Notice that, for any measurable function $G : \mathbb{D}([0,\tau],\mathcal{S}_K) \to \mathbb{R}_+$,

$$\begin{aligned}
&\mathbb{E}[\sum_{\substack{u\in\mathbb{G}(t)\\ u\succeq u_x(r)}} \psi(x_u(t),Z(t))m_\psi(x_u(\tau),Z(\tau),t-\tau)^{-1}G((x_u(s),Z(s))_{s\leqslant\tau})|X(r)=\mathfrak{X}_0]\\
&= \mathbb{E}[\sum_{\substack{v\in\mathbb{G}(\tau)\\ v\succeq u_x(r)}} \mathbb{E}[\sum_{\substack{u\in\mathbb{G}(t)\\ u\succeq v}} \psi(x_u(t),Z(t))|X(\tau)]m_\psi(x_v(\tau),Z(\tau),t-\tau)^{-1}G((x_v(s),Z(s))_{s\leqslant\tau})|X(r)=\mathfrak{X}_0]\\
&= \mathbb{E}[\sum_{\substack{v\in\mathbb{G}(\tau)\\ v\succeq u_x(r)}} G((x_v(s),Z(s))_{s\leqslant\tau})|X(r)=\mathfrak{X}_0].
\end{aligned} \tag{16}$$

Applying this equality to $G((x_v(s), Z(s)_{s\leqslant\tau}) = g(x(\tau), Z(\tau))$ finally yields the desired semi-group property:

$$\mathcal{P}^{(t)}_{r,s} f(x,\mathbf{z}) = \mathcal{P}^{(t)}_{r,\tau}\mathcal{P}^{(t)}_{\tau,s} f(x,\mathbf{z}).$$

This concludes the proof. □

Let us now compute the generator of $(\mathcal{P}^{(t)}_{r,s}, r \leqslant s \leqslant t)$.

Lemma 3.6. *Let $t \geqslant 0$. The generator of the semi-group $(\mathcal{P}^{(t)}_{r,s}, r \leqslant s \leqslant t)$ is $(\mathcal{A}^{(t)}_s, s \leqslant t)$.*

Proof. Consider $f : \mathcal{S}_K \to \mathbb{R}_+$. Let $(x, \mathbf{z}) \in \mathcal{S}_K$ and $t \geqslant 0$. For any $0 \leqslant s \leqslant t$ and $h > 0$ such that $s + h \leqslant t$, it follows from Equation (16) and the Markov property that

$$\mathcal{P}^{(t)}_{s,s+h} f(x, \mathbf{z}) = m_\psi(x, \mathbf{z}, t - s)^{-1} \mathbb{E}_{\mathfrak{X}(x,\mathbf{z})}[\sum_{\substack{u \in \mathbb{G}(h) \\ u \succeq u_x(0)}} m_\psi(x_u(h), Z(h), t - (s + h)) f(x_u(h), Z(h))].$$

Using Lemma 3.4 as well as the fact that $\mathcal{S}_K$ is a finite set, we obtain the following Taylor expansion:

$$\begin{aligned} m_\psi(x, \mathbf{z}, t - s)\mathcal{P}^{(t)}_{s,s+h} f(x, \mathbf{z}) = \mathbb{E}_{\mathfrak{X}(x,\mathbf{z})}[&\sum_{\substack{u \in \mathbb{G}(h) \\ u \succeq u_x(0)}} m_\psi(x_u(h), Z(h), t - s) f(x_u(h), Z(h))] \\ &+ h\mathbb{E}_{\mathfrak{X}(x,\mathbf{z})}[\sum_{\substack{u \in \mathbb{G}(h) \\ u \succeq u_x(0)}} \partial_s m_\psi(x_u(h), Z(h), t - s) f(x_u(h), Z(h))] + o(h). \end{aligned}$$

As a consequence,

$$\begin{aligned} m_\psi(x, \mathbf{z}, t - s) \frac{\mathcal{P}^{(t)}_{s,s+h} f(x, \mathbf{z}) - f(x, \mathbf{z})}{h} &= \mathbb{E}_{\mathfrak{X}(x,\mathbf{z})}[\sum_{\substack{u \in \mathbb{G}(h) \\ u \succeq u_x(0)}} \partial_s m_\psi(x_u(h), Z(h), t - s) f(x_u(h), Z(h))] \\ + h^{-1}\Big(\mathbb{E}_{\mathfrak{X}(x,\mathbf{z})}[\sum_{\substack{u \in \mathbb{G}(h) \\ u \succeq u_x(0)}} m_\psi(x_u(h), Z(h), t - s) f(x_u(h), Z(h))] &- m_\psi(x, \mathbf{z}, t - s) f(x, \mathbf{z})\Big) + \epsilon(h), \end{aligned}$$

where $\epsilon(h)$ is such that $\lim_{h \to 0+} \epsilon(h) = 0$. We thus obtain that

$$\lim_{h \to 0+} \frac{\mathcal{P}^{(t)}_{s,s+h} f(x, \mathbf{z}) - f(x, \mathbf{z})}{h} = m_\psi(x, \mathbf{z}, t - s)^{-1} \left(\mathcal{G}(m_\psi(\cdot, t - s) f(\cdot))(x, \mathbf{z}) + \partial_s m_\psi(x, \mathbf{z}, t - s) f(x, \mathbf{z})\right),$$

where we recall that $\mathcal{G}$ is defined by (2). Lemma 3.4 yields the desired result. ☐

We finally are ready to establish Theorem 3.3. The proof follows the lines of [23], and is detailed here for the sake of completeness.

Proof of Theorem 3.3. Throughout this proof, for readability, we will make use of the following notations. On the one hand, for $t \geqslant 0$ and u such that there exists $v \in \mathbb{G}(t)$ satisfying $u \succeq v$, let

$$\mathbb{X}_u(t) = (x_u(t), Z(t)).$$

Similarly, for $0 \leqslant s \leqslant t$, we let

$$\mathbb{Y}^{(t)}(s) = (Y^{(t)}(s), \zeta^{(t)}(s)).$$

Let us start by showing that Equation (9) holds for $F((x(s), \mathbf{z}(s))_{s \leqslant t}) = \prod_{j=1}^{k} f_j(x(s_j), \mathbf{z}(s_j))$ where $k \geqslant 1$, $0 \leqslant s_1 \leqslant \cdots \leqslant s_k \leqslant t$ and $f_1, \ldots, f_k : \mathcal{S}_K \to \mathbb{R}_+$.

This part of the proof proceeds by induction. For $k \geqslant 1$, let H_k be the property that for any $0 \leqslant s_1 \leqslant \cdots \leqslant s_k \leqslant t$ and $f_1, \ldots, f_k : \mathcal{S}_K \to \mathbb{R}_+$,

$$\mathbb{E}_{\mathfrak{X}(x,\mathbf{z})}[\sum_{u \in \mathbb{G}(t),\, u \succeq u_x(0)} \psi(\mathbb{X}_u(t)) \prod_{j=1}^{k} f_j(\mathbb{X}_u(s_j))] = m_\psi(x, \mathbf{z}, t)\mathbb{E}_{(x,\mathbf{z})}[\prod_{j=1}^{k} f_j(\mathbb{Y}^{(t)}(s_j))].$$

Let us turn our attention to the initialization step. As $\mathcal{S}_K$ is a finite set, a semi-group acting on non-negative functions on $\mathcal{S}_K$ is uniquely characterized by its generator. Thus Lemma

3.6 implies that the semi-groups $\mathcal{P}^{(t)}$ and $\mathcal{R}^{(t)}$ are identical. Hence for any $s \in [0,t]$ and $f : \mathcal{S}_K \to \mathbb{R}_+$, Equation (14) becomes

$$\mathbb{E}_{\mathfrak{X}(x,\mathbf{z})}[\sum_{u \in \mathbb{G}(t),\, u \succeq u_x(0)} \psi(\mathbb{X}_u(t)) f(\mathbb{X}_u(s))] = m_\psi(x,\mathbf{z},t)\mathcal{R}^{(t)}_{0,s} f(x,\mathbf{z}).$$

This exactly corresponds to H_1 by definition of $\mathcal{R}^{(t)}$.

Suppose now that H_{k-1} is true for $k > 1$, and let us show that H_k follows. Consider functions $f_1, \dots, f_k : \mathcal{S}_K \to \mathbb{R}_+$ and $0 \leqslant s_1 \leqslant \dots \leqslant s_k \leqslant t$. Notice that

$$\mathbb{E}_{\mathfrak{X}(x,\mathbf{z})}[\sum_{\substack{u \in \mathbb{G}(t) \\ u \succeq u_x(0)}} \psi(\mathbb{X}_u(t)) \prod_{j=1}^{k} f_j(\mathbb{X}_u(s_j))] =$$

$$\mathbb{E}_{\mathfrak{X}(x,\mathbf{z})}[\sum_{\substack{u \in \mathbb{G}(s_{k-1}) \\ u \succeq u_x(0)}} \prod_{j=1}^{k-1} f_j(\mathbb{X}_u(s_j)) \mathbb{E}[\sum_{\substack{v \in \mathbb{G}(t) \\ v \succeq u_{x_u(s_{k-1})}}} \psi(\mathbb{X}_u(t)) f_k(\mathbb{X}_v(s_k)) | X(s_{k-1}) = \mathfrak{X}(\mathbb{X}_u(s_{k-1}))]].$$

Using the Markov property and H_1 leads to:

$$\mathbb{E}_{\mathfrak{X}(x,\mathbf{z})}[\sum_{\substack{u \in \mathbb{G}(t) \\ u \succeq u_x(0)}} \psi(\mathbb{X}_u(t)) \prod_{j=1}^{k} f_j(\mathbb{X}_u(s_j))] =$$

$$\mathbb{E}_{\mathfrak{X}(x,\mathbf{z})}\Big[\sum_{\substack{u \in \mathbb{G}(s_{k-1}) \\ u \succeq u_x(0)}} m_\psi(\mathbb{X}_u(s_{k-1}), t - s_{k-1}) \prod_{j=1}^{k-1} f_j(\mathbb{X}_u(s_j)) \mathbb{E}[f_k(\mathbb{Y}^{(t)}(s_k)) | \mathbb{Y}^{(t)}(s_{k-1}) = \mathbb{X}_u(s_{k-1})]\Big].$$

Equation (16) allows to rewrite this as:

$$\mathbb{E}_{\mathfrak{X}(x,\mathbf{z})}[\sum_{\substack{u \in \mathbb{G}(t) \\ u \succeq u_x(0)}} \psi(\mathbb{X}_u(t)) \prod_{j=1}^{k} f_j(\mathbb{X}_u(s_j))] =$$

$$\mathbb{E}_{\mathfrak{X}(x,\mathbf{z})}\Big[\sum_{\substack{u \in \mathbb{G}(t) \\ u \succeq u_x(0)}} \psi(\mathbb{X}_u(t)) \prod_{j=1}^{k-1} f_j(\mathbb{X}_u(s_j)) \mathbb{E}[f_k(\mathbb{Y}^{(t)}(s_k)) | \mathbb{Y}^{(t)}(s_{k-1}) = \mathbb{X}_u(s_{k-1})]\Big].$$

Finally, H_{k-1} yields:

$$\begin{aligned}
&\mathbb{E}_{\mathfrak{X}(x,\mathbf{z})}[\sum_{\substack{u \in \mathbb{G}(t) \\ u \succeq u_x(0)}} \psi(\mathbb{X}_u(t)) \prod_{j=1}^{k} f_j(\mathbb{X}_u(s_j))] \\
&\qquad = m_\psi(x,\mathbf{z},t)\mathbb{E}_{(x,\mathbf{z})}[\prod_{j=1}^{k-1} f_j(\mathbb{Y}^{(t)}(s_j)) \mathbb{E}[f_k(\mathbb{Y}^{(t)}(s_k)) | \mathbb{Y}^{(t)}(s_{k-1})]] \\
&\qquad = m_\psi(x,\mathbf{z},t)\mathbb{E}_{(x,\mathbf{z})}[\prod_{j=1}^{k} f_j(\mathbb{Y}^{(t)}(s_j))].
\end{aligned}$$

This concludes the induction argument.

In order to obtain the desired result, we will reason using the monotone class theorem. Let us introduce the set

$$I = \left\{ \bigcap_{j=1}^{k} \{x \in \mathbb{D}([0,t], \mathcal{S}_K) : x(s_j) \in B_j\}, k \in \mathbb{N}, s_j \in [0,t], B_j \in \mathcal{P}(\mathcal{S}_K) \right\}$$

where $\mathcal{P}(\mathcal{S}_K)$ is the set of subsets of $\mathcal{S}_K$. The set I is a π-system, which induces the Borel σ-algebra $\mathcal{B}(\mathbb{D}([0,t],\mathcal{S}_K))$ on the Skorokhod space $\mathbb{D}([0,t],\mathcal{S}_K)$ (Theorem 12.5 in [9]). Further, define

$$M = \{\mathbf{B} \in \mathcal{B}(\mathbb{D}([0,t],\mathcal{S}_K)) : \text{ Equation (9) is satisfied for } F = \mathbf{1}_{\mathbf{B}}\}.$$

M is a monotone class which contains I according to our induction argument. It thus follows from the monotone class theorem that $M = \mathcal{B}(\mathbb{D}([0,t],\mathcal{S}_K))$. In other words, for any $\mathbf{B} \in \mathcal{B}(\mathbb{D}([0,t],\mathcal{S}_K))$, Equation (9) is satisfied for $F = \mathbf{1}_{\mathbf{B}}$. As a consequence, Equation (9) holds for any positive measurable function $F : \mathbb{D}([0,t],\mathcal{S}_K) \to \mathbb{R}_+$ as there exists an increasing sequence of simple functions converging pointwise to F, from which the result follows by monotone convergence. ☐

4 Typical lineages and sampling in the diffusion approximation

In this section, we are interested in sampling in large populations. Indeed, when the population size K grows to infinity, the population trait distribution converges under classical assumptions to a deterministic limit z, characterized as the unique solution of a dynamical system. In particular, it is possible to sample in the limit z, thanks to a time-inhomogeneous spinal process emerging from the ψ-spine construction in the large population limit [3].

In order to refine this approximation and capture deviations of the population type distribution from its deterministic limit, we consider in this section the diffusion approximation of the population type distribution [15, Chapter 11]. More precisely, we consider the case of uniform sampling ($\psi = 1$) and establish a many-to-one formula for sampling in the diffusion approximation. In addition, we explicitly quantify the associated approximation error.

4.1 A many-to-one formula for diffusion approximations

Let us start by properly introducing the diffusion approximation of the type distribution in the original population process.

Let $d = \mathrm{Card}(\mathcal{X})$ and $\mathcal{Z} = [0,1]^d$. Let $(\tau_{\mathbf{k}}, \mathbf{k} \in \mathbb{N}^d)$ be a family of continuous bounded functions $\tau_{\mathbf{k}} : \mathcal{X} \times \mathcal{Z} \to \mathbb{R}_+$ such that the set $J = \{(x,\mathbf{k}) \in \mathcal{X} \times \mathbb{N}^d : \tau_{\mathbf{k}}(x,\cdot) \neq 0\}$ is finite.

For $K \geqslant 1$, consider the population process X^K where an individual of type x in a population of composition $\mathbf{z} \in \mathcal{Z}_K$ is replaced by descendance $\mathbf{k}$ at rate $\tau_{\mathbf{k}}(x, \mathbf{z}/K)$. For the population size to be bounded by K, this imposes the following condition:

$$\forall K \geqslant 1, \forall (x,\mathbf{k}) \in J, \forall \mathbf{z} \in \mathcal{Z}_K, \quad \tau_{\mathbf{k}}(x, \mathbf{z}/K) = 0 \text{ if } \|\mathbf{z} + \mathbf{k} - \mathbf{e}(x)\|_1 \geqslant K. \tag{17}$$

Notice that for instance, classical epidemic models satisfy this condition, as the effective population size is often kept constant. Letting $Z^K = Z(X^K)/K$ then ensures that $Z^K(t) \in \mathcal{Z}$ almost surely, for every $t \geqslant 0$.

Consider the set

$$\mathcal{S} = \{(x,\mathbf{z}) \in \mathcal{X} \times [0,1]^d : \mathbf{z}_x > 0\}.$$

We consider a given sequence of initial conditions $(x, \mathbf{z}_0^K)_{K \geqslant 2}$ such that for any K, $\mathbf{z}_0^K \in \mathcal{Z}_K/K$, $(\mathbf{z}_0^K)_x \geqslant 1/K$ and $\lim_{K\to\infty} \mathbf{z}_0^K = \mathbf{z}_0 \in \mathcal{Z}$. In particular, $(x, \mathbf{z}_0^K) \in \mathcal{S}$. Define $A(\mathbf{z}) = (A_{x,y}(\mathbf{z}))_{x,y\in\mathcal{X}}$ for $\mathbf{z} \in \mathcal{Z}$ by $A_{x,y}(\mathbf{z}) = \sum_{\mathbf{k}} (\mathbf{k}_y - 1)\tau_{\mathbf{k}}(x,\mathbf{z})$. Throughout the section, we work under the following regularity assumption:

Assumption 4.1. *For every $(x,\mathbf{k}) \in J$, there exists a positive constant $L_{x,\mathbf{k}}$ such that the application $\tau_{\mathbf{k}}(x,\cdot)$ is $L_{x,\mathbf{k}}$-Lipschitz continuous on $[0,1]^d$.*

In particular, this implies that $\mathbf{z} \mapsto A(\mathbf{z})$ is Lipschitz continuous on $\mathcal{Z}$. Thus, there exists a unique solution z to the differential equation

$$z'(t) = z(t)A(z(t)), \quad z(0) = \mathbf{z}_0.$$

Then Z^K converges uniformly in probability to z on finite time intervals [15, Theorem 3.1, Chapter 11].

However, realistic populations may be large, but never of infinite size. As a consequence, it is natural to aim at capturing fluctuations of Z^K around its large population limit, when the population size K is large. This can be achieved thanks to its diffusion approximation defined as follows. For $(x, \mathbf{k}) \in J$, let $h(x, \mathbf{k}) = \sum_{y \in X} (\mathbf{k}_y - \delta^x_y)\mathbf{e}(y)$. As the set J is finite, and thanks to Assumption 4.1, Theorem 2.3 in [19, Chapter IV] ensures existence of a family of independent Brownian motions $(B_{x,\mathbf{k}}(t), (x, \mathbf{k}) \in J, t \geqslant 0)$ and of a d-dimensional càdlàg process $\widehat{Z}^K = (\widehat{Z}^K_x)_{x \in \mathcal{X}}$ such that almost surely for all $t \geqslant 0$,

$$\widehat{Z}^K(t) = \mathbf{z}^K_0 + \sum_{(x,\mathbf{k}) \in J} h(x, \mathbf{k}) \left(\int_0^t \widehat{Z}^K_x(s)\tau_{\mathbf{k}}(x, \widehat{Z}^K(s))ds + \frac{1}{\sqrt{K}} \int_0^t \sqrt{\widehat{Z}^K_x(s)\tau_{\mathbf{k}}(x, \widehat{Z}^K(s))} dB_{x,\mathbf{k}}(s) \right). \quad (18)$$

In order to construct a spinal process associated to the diffusion approximation, the reasoning goes as follows. Recall that we are interested in uniform sampling, *i.e.* letting $\psi = 1$. As a consequence, for the ψ-spine construction of [3], only the reproduction rates along the ψ-spine are biased; individuals other than the spine behave as in the original population process. In addition, as the population size grows large, the spine itself becomes negligible and its impact on the type distribution in the spinal population is expected to vanish, at least on finite time intervals. We thus expect that $\widehat{Z}^K$ actually approximates not only Z^K, but also the type distribution of the spinal population.

It remains to describe the behavior of the spine itself. Assume that the spine is of type x in a population of type distribution $\mathbf{z} \in \mathcal{Z}$. Let $\mathbf{k}$ be such that $(x, \mathbf{k}) \in J$. Then the spine reproduces to leave descendance $\mathbf{k}$ and become of type $y \in \mathcal{X}$ at rate $\mathbf{k}_y\tau_{\mathbf{k}}(x, \mathbf{z})$. Let $(\widehat{Y}^K(t))_{t \geqslant 0}$ be such that $\widehat{Y}^K(t) = x$ means that at time t, the spine is of type x. Let $J^* = \{(x, \mathbf{k}, y) : (x, \mathbf{k}) \in J, \mathbf{k}_y > 0\}$ and consider a family $(Q_{x,\mathbf{k},y}, (x, \mathbf{k}, y) \in J^*)$ of Poisson Point Processes on $\mathbb{R}^2_+$ with intensity the Lebesgue measure, defined on the same probability space as and independent from $(B_{x,\mathbf{k}}(t), (x, \mathbf{k}) \in J, t \geqslant 0)$. As previously, $\widehat{Y}^K$ can properly be defined by the following SDE:

$$\widehat{Y}^K(t) = y_0 + \sum_{(x,\mathbf{k},y) \in J^*} \int_0^t \int_0^{+\infty} \mathbf{1}_{\left\{\widehat{Y}^K(s-) = x, \theta \leqslant \mathbf{k}_y\tau_{\mathbf{k}}(x, \widehat{Z}^K(s-))\right\}} (\mathbf{e}(y) - \mathbf{e}(x))Q_{x,\mathbf{k},y}(ds, d\theta). \quad (19)$$

The spinal process associated to the diffusion approximation is thus given by $(\widehat{Y}^K, \widehat{Z}^K)$. In order to relate it to lineages sampled in the original population process, we will establish an approximate many-to-one formula in the spirit of Equation (4) which becomes exact in the large population limit.

In particular, similarly to Equation (4), the many-to-one formula requires an exponential weighting of the spinal trajectories. For $(x, \mathbf{z}) \in \mathcal{X} \times \mathcal{Z}$, let

$$\lambda(x, \mathbf{z}) = \sum_{\mathbf{k}:(x,\mathbf{k}) \in J} (\|\mathbf{k}\|_1 - 1)\tau_{\mathbf{k}}(x, \mathbf{z}).$$

For $t \geqslant 0$, define

$$\widehat{W}^K(t) = \exp\left(\int_0^t \lambda(\widehat{Y}^K(s), \widehat{Z}^K(s))ds \right).$$

Finally, we introduce the following family functions:

$$\mathfrak{L} = \Big\{ F : \mathbb{D}([0,t],\mathcal{X}) \times \mathbb{D}([0,t],\mathcal{Z}) \to \mathbb{R} \text{ bounded s.t. } \exists L : \forall x \in \mathbb{D}([0,t],\mathcal{X}),\ \forall \mathbf{z}_1, \mathbf{z}_2 \in \mathbb{D}([0,t],\mathcal{Z}),$$
$$|F((x(s),\mathbf{z}_1(s))_{s\leqslant t}) - F((x(s),\mathbf{z}_2(s))_{s\leqslant t})| \leqslant L_F \sup_{s\in[0,t]} \|\mathbf{z}_1(s) - \mathbf{z}_2(s)\|_1 \Big\}.$$

We then obtain the main result.

Theorem 4.2. *Let $t > 0$ and $F \in \mathfrak{L}$. There exists $K_0 \in \mathbb{N}$ and $C > 0$ such that for every $K \geqslant K_0$,*

$$\left| \mathbb{E}_{\mathfrak{X}(x,\mathbf{z}_0^K)} \left[\sum_{\substack{u\in\mathbb{G}(t),\\ u\succeq u_x}} F((x_u(s), Z^K(s))_{s\leqslant t}) \right] - \mathbb{E}_{(x,\mathbf{z}_0^K)} \left[\widehat{W}^K(t) F((\widehat{Y}^K(s), \widehat{Z}^K(s))_{s\leqslant t}) \right] \right| \leqslant C K^{-1/4}.$$

This theorem establishes that it is possible to approximate the expectation of F over lineages sampled uniformly in the original population, by the ($\widehat{W}^K$-weighted) expectation of F over trajectories of the spine in the diffusion approximation. In particular, this approximation is asymptotically exact in the large population limit, and we control its speed of convergence.

4.2 Proof of Theorem 4.2

The proof of Theorem 4.2 proceeds as follows. We couple the spine of the initial population process (Y^K, ζ^K) with the spine $(\widehat{Y}^K, \widehat{Z}^K)$ associated to its diffusion approximation. We then make use of classical results on diffusion approximations in order to show that, on finite time intervals, (Y^K, ζ^K) and $(\widehat{Y}^K, \widehat{Z}^K)$ grow arbitrarily close with high probability as the population size grow large. This finally allows to obtain the desired result.

4.2.1 Coupling the spine with its diffusion approximation

Let us start by recalling the behavior of the ψ-spine with for the original population process [3], in the case of uniform sampling $\psi = 1$.

As usual, we want to keep track of the type of the spine Y^K and of the type distribution in the spinal population. We will do so while introducing a slight change in the type space, which will allow us to derive an equation for the spinal population which does not depend on Y^K, allowing for classical results on diffusion approximations to apply. More precisely, the type space now becomes $\mathcal{X}^* = \{0,1\} \times \mathcal{X}$. An individual of type $(0,x) \in \mathcal{X}^*$ corresponds to an individual of type x which is not the spine, whereas an individual is of type $(1,x) \in \mathcal{X}^*$ if it is the spine and of type x. We let $\zeta^K = (\zeta^K_{i,x}, (i,x) \in \mathcal{X}^*)$ designate the corresponding type distribution of the spinal population. Finally, we define $\zeta^K|_{\mathcal{X}} = ((\zeta^K|_{\mathcal{X}})_x, x \in \mathcal{X})$ by

$$\left(\zeta^K|_{\mathcal{X}}\right)_x = \zeta^K_{1,x} + \zeta^K_{0,x} \quad \forall x \in \mathcal{X}.$$

As mentioned previously, since $\psi = 1$, individuals other than the spine behave exactly as in the original population process. Whenever the spine is of type x in a population of type distribution $\mathbf{z} = (\mathbf{z}_{i,x}, (i,x) \in \mathcal{X}^*) \in \mathcal{Z}^2$ such that $\mathbf{z}_{1,x} > 0$, it reproduces to leave descendance $\mathbf{k}$ and become of type y at rate $\mathbf{k}_y \tau_{\mathbf{k}}(x, \mathbf{z}|_{\mathcal{X}})$. Throughout the following, with slight abuse of notation, we write $\tau_{\mathbf{k}}(x,\mathbf{z})$ instead of $\tau_{\mathbf{k}}(x, \mathbf{z}|_{\mathcal{X}})$ in order to simplify notations.

For $(x,\mathbf{k}) \in J$ and $(x,\mathbf{k},y) \in J^*$, we let

$$h_0(x,\mathbf{k}) = \sum_{y\in\mathcal{X}} (\mathbf{k}_y - \delta^x_y)\mathbf{e}(0,y) \quad \text{and} \quad h_1(x,\mathbf{k},y) = \sum_{w\in\mathcal{X}} (\mathbf{k}_w - \delta^y_w)\mathbf{e}(0,w) + \mathbf{e}(1,y) - \mathbf{e}(1,x).$$

Finally, our initial condition $(y, \mathbf{z}_0^K)$ becomes $(y, \overline{\mathbf{z}}_0^K)$ with

$$\overline{\mathbf{z}}_0^K = \frac{1}{K}\mathbf{e}(1, y) + \sum_{x \in \mathcal{X}} \left((\mathbf{z}_0^K)_x - \frac{1}{K}\delta_y^x \right) \mathbf{e}(0, x).$$

Given a family of independent Poisson Point Processes $(Q_j, j \in J \cup J^*)$ of intensity the Lebesgue measure on $\mathbb{R}_+^2$, the process (Y^K, ζ^K) can then be defined as follows:

$$\begin{aligned}
Y^K(t) &= y_0 + \sum_{(x,\mathbf{k},y) \in J^*} \int_0^t \int_0^{+\infty} \mathbf{1}_{\{Y^K(s-)=x, \theta \leqslant \mathbf{k}_y \tau_{\mathbf{k}}(x, \zeta^K(s-))\}} (\mathbf{e}(y) - \mathbf{e}(x)) Q_{x,\mathbf{k},y}(ds, d\theta), \\
\zeta^K(t) &= \overline{\mathbf{z}}_0^K + \frac{1}{K} \sum_{(x,\mathbf{k},y) \in J^*} h_1(x, \mathbf{k}, y) \int_0^t \int_0^{+\infty} \mathbf{1}_{\left\{\theta \leqslant K \zeta_{1,x}^K(s-) \mathbf{k}_y \tau_{\mathbf{k}}(x, \zeta^K(s-))\right\}} Q_{x,\mathbf{k},y}(ds, d\theta) \\
&\quad + \frac{1}{K} \sum_{(x,\mathbf{k}) \in J} h_0(x, \mathbf{k}) \int_0^t \int_0^{+\infty} \mathbf{1}_{\left\{\theta \leqslant K \zeta_{0,x}^K(s-) \tau_{\mathbf{k}}(x, \zeta^K(s-))\right\}} Q_{x,\mathbf{k}}(ds, d\theta).
\end{aligned} \tag{20}$$

Let us now turn to the coupling of the processes (Y^K, ζ^K) and $(\hat{Y}^K, \hat{Z}^K)$, which we do by making use of the diffusion approximation $\hat{\zeta}^K$ of ζ^K, which we define below. This will allow in the next section to quantify the approximation error, by controlling both $\|\zeta^K - \hat{\zeta}^K\|_1$ and $\|\widehat{\zeta^K} - \hat{Z}^K\|_1$ on finite time intervals.

As previously, since the set J is finite and using Assumption 4.1, Theorem 2.3 in [19, Chapter IV] ensures existence of a family of independent Brownian motions $(B_j(t), j \in J \cup J^*, t \geqslant 0)$, and of a d-dimensional càdlàg process $\hat{\zeta}^K$ such that almost surely for all $t \geqslant 0$,

$$\begin{aligned}
\hat{\zeta}^K(t) = \overline{\mathbf{z}}_0^K &+ \sum_{(x,\mathbf{k}) \in J} h_0(x, \mathbf{k}) \left(\int_0^t \hat{\zeta}_{(0,x)}^K(s) \tau_{\mathbf{k}}(x, \hat{\zeta}^K(s)) ds + \frac{1}{\sqrt{K}} \int_0^t \sqrt{\hat{\zeta}_{(0,x)}^K(s) \tau_{\mathbf{k}}(x, \hat{\zeta}^K(s))} dB_{x,\mathbf{k}}(s) \right) \\
&+ \sum_{(x,\mathbf{k},y) \in J^*} h_1(x, \mathbf{k}, y) \left(\int_0^t \hat{\zeta}_{(1,x)}^K(s) \mathbf{k}_y \tau_{\mathbf{k}}(x, \hat{\zeta}^K(s)) ds + \frac{1}{\sqrt{K}} \int_0^t \sqrt{\hat{\zeta}_{(1,x)}^K(s) \mathbf{k}_y \tau_{\mathbf{k}}(x, \hat{\zeta}^K(s))} dB_{x,\mathbf{k},y}(s) \right).
\end{aligned} \tag{21}$$

This process $\hat{\zeta}^K$ corresponds to the diffusion approximation of ζ^K. In order to compare it to ζ^K, we make use of strong approximations theorems. Indeed, Theorem 3.1 in [15, Chapter 11] assures that we may assume the existence of a family of independent Poisson Point Processes $(Q_j, j \in J \cup J^*)$ of intensity the Lebesgue measure on $\mathbb{R}_+^2$ such that the process (Y^K, ζ^K) defined by Equation (20) using this family satisfies the following. For all $t \geqslant 0$, there exist constants $C(t)$, $C'(t)$ independent from K such that

$$\mathbb{P}\left(\sup_{s \in [0,t]} \|\zeta^K(s) - \hat{\zeta}^K(s)\|_1 > \frac{C(t) \log(K)}{K} \right) \leqslant \frac{C'(t)}{K^2}. \tag{22}$$

Finally, in this section, we consider $\hat{Z}^K$ to take values in $\mathcal{X}^*$. Notice that, as the families $(B_j, j \in J)$ and $(B_j, j \in J^*)$ are independent, we can further assume the family $(Q_j, j \in J^*)$ to be independent from $(B_j, j \in J)$. Let $\hat{\mathbf{z}}_0^K = \sum_{x \in X} (\mathbf{z}_0^K)_x \mathbf{e}(0, x)$. We assume from now the process $(\hat{Y}^K, \hat{Z}^K)$ to be defined by:

$$\begin{aligned}
\hat{Y}^K(t) &= y_0 + \sum_{(x,\mathbf{k},y) \in J^*} \int_0^t \int_0^{+\infty} \mathbf{1}_{\left\{\hat{Y}^K(s-)=x, \theta \leqslant \mathbf{k}_y \tau_{\mathbf{k}}(x, \hat{Z}^K(s-))\right\}} (\mathbf{e}(y) - \mathbf{e}(x)) Q_{x,\mathbf{k},y}(ds, d\theta), \\
\hat{Z}^K(t) &= \hat{\mathbf{z}}_0^K + \sum_{(x,\mathbf{k}) \in J} h_0(x, \mathbf{k}) \left(\int_0^t \hat{Z}_x^K(s) \tau_{\mathbf{k}}(x, \hat{Z}^K(s)) ds + \frac{1}{\sqrt{K}} \int_0^t \sqrt{\hat{Z}_x^K(s) \tau_{\mathbf{k}}(x, \hat{Z}^K(s))} dB_{x,\mathbf{k}}(s) \right).
\end{aligned}$$

This concludes the coupling of the processes (Y^K, ζ^K), $\hat{\zeta}^K$ and $(\hat{Y}^K, \hat{Z}^K)$.

4.2.2 Quantifying the approximation error

We are now ready to turn to the proof of Theorem 4.2. We start by comparing ζ^K and $\widehat{Z}^K$.

Lemma 4.3. *Let $t \geqslant 0$ and $p \in (0,1/2)$. There exist positive constants $C(t)$ and $M(t)$ depending only on t such that for every $K \geqslant 2$, letting $\varepsilon_K \geqslant 2C(t)K^{-1}\log(K)$ and $\alpha_K = M(t)(K^{-2} + K^{-1/2}\varepsilon_K^{-1})$,*

$$\mathbb{P}\left(\sup_{s\in[0,t]} \|\zeta^K(s) - \widehat{Z}^K(s)\|_1 \geqslant \varepsilon_K\right) \leqslant \alpha_K. \tag{23}$$

Remark 4.4. *The most interesting setting is for both α_K and ε_K to converge to zero as K grows to infinity. This incites to consider $\varepsilon_K = 2\max(K^{-p}, C(t)K^{-1}\log(K))$ for $p \in (0,1/2)$. In particular, for K sufficiently large, we then have $\varepsilon_K = 2K^{-p}$.*

Proof. Start by noticing that

$$\begin{aligned}\mathbb{P}\left(\sup_{s\in[0,t]} \|\zeta^K(s) - \widehat{Z}^K(s)\|_1 \geqslant \varepsilon_K\right) &\leqslant \mathbb{P}\left(\sup_{s\in[0,t]} \|\zeta^K(s) - \widehat{\zeta}^K(s)\|_1 \geqslant \frac{\varepsilon_K}{2}\right)\\ &\quad + \mathbb{P}\left(\sup_{s\in[0,t]} \|\widehat{\zeta}^K(s) - \widehat{Z}^K(s)\|_1 \geqslant \frac{\varepsilon_K}{2}\right).\end{aligned}$$

We consider $C(t)$ to be the constant appearing in Equation (22). On the one hand, it then follows from Equation (22) that there exists a constant $C'(t)$ independent from K such that

$$\mathbb{P}\left(\sup_{s\in[0,t]} \|\zeta^K(s) - \widehat{\zeta}^K(s)\|_1 \geqslant \frac{\varepsilon_K}{2}\right) \leqslant \frac{C'(t)}{K^2}. \tag{24}$$

It remain to focus on controlling $\sup_{s\in[0,t]} \|\widehat{\zeta}^K(s) - \widehat{Z}^K(s)\|_1$. Let $s \in [0,t]$ and start by noticing that

$$\|\widehat{\zeta}^K(s) - \widehat{Z}^K(s)\|_1 = \sum_{x\in\mathcal{X}} |\ \widehat{\zeta}^K_{0,x}(s) - \widehat{Z}^K_{0,x}(s)|\ + \sum_{x\in\mathcal{X}} |\widehat{\zeta}^K_{1,x}(s)|. \tag{25}$$

Equation (21) ensures that, almost surely,

$$\sum_{x\in\mathcal{X}} |\widehat{\zeta}^K_{1,x}(s)| = \sum_{x\in\mathcal{X}} |(\overline{\mathbf{z}}^K_0)_{1,x}| = \frac{1}{K}. \tag{26}$$

Throughout the following, c and $c(t)$ designate positive constants which may change throughout the computation, the latter depending on t. For $(x,\mathbf{k}) \in J$, let $\rho_{x,\mathbf{k}}(\mathbf{z}) = \mathbf{z}_{0,x}\tau_{\mathbf{k}}(x,\mathbf{z})$. Similarly, for $(x,\mathbf{k},y) \in J^*$, let $\rho^*_{x,\mathbf{k},y} = \mathbf{k}_y\mathbf{z}_{1,x}\tau_{\mathbf{k}}(x,\mathbf{z})$. Using the fact that J and J^* are finite sets and Equation (21), we obtain that:

$$\begin{aligned}|\widehat{\zeta}^K_{0,x}(s) - \widehat{Z}^K_{0,x}(s)| \leqslant\ & |(\overline{\mathbf{z}}^K_0)_{0,x} - (\widehat{\mathbf{z}}^K_0)_{0,x}| + c \sum_{(x,\mathbf{k})\in J} \int_0^s |\rho_{x,\mathbf{k}}(\widehat{\zeta}^K(u)) - \rho_{x,\mathbf{k}}(\widehat{Z}^K(u))\ | du\\ &+ c \sum_{(x,\mathbf{k},y)\in J^*} \int_0^s |\rho^*_{x,\mathbf{k},y}(\widehat{\zeta}^K(u))| du + \frac{c}{\sqrt{K}}\left(\sum_{(x,\mathbf{k})\in J} |M_{x,\mathbf{k}}(s)| + \sum_{(x,\mathbf{k},y)\in J^*} |N_{x,\mathbf{k},y}(s)|\right),\end{aligned}$$

where for $(x,\mathbf{k}) \in J$ and $(x,\mathbf{k},y) \in J^*$,

$$\begin{aligned}M_{x,\mathbf{k}}(s) &= \int_0^s \left(\sqrt{\rho_{x,\mathbf{k}}(\widehat{\zeta}^K(s))} - \sqrt{\rho_{x,\mathbf{k}}(\widehat{Z}^K(s))}\right) dB_{x,\mathbf{k}}(ds),\\ N_{x,\mathbf{k},y}(s) &= \int_0^s \sqrt{\rho^*_{x,\mathbf{k},y}(\widehat{\zeta}^K(s))} ds.\end{aligned}$$

By definition, we have $\sum_{x\in\mathcal{X}} |(\overline{\mathbf{z}}_0^K)_{0,x} - (\hat{\mathbf{z}}_0^K)_{0,x}| = 1/K$. It further follows from Equation (17) that almost surely, for every $s \geqslant 0$, $\|\widehat{\zeta}^K(s)\|_1 \leqslant 1$ and $\|\widehat{Z}^K(s)\|_1 \leqslant 1$. In addition to Assumption 4.1, this yields

$$\sum_{(x,\mathbf{k})\in J} \int_0^s |\rho_{x,\mathbf{k}}(\widehat{\zeta}^K(u)) - \rho_{x,\mathbf{k}}(\widehat{Z}^K(u))\ |du \leqslant c\int_0^s \|\widehat{\zeta}^K(u) - \widehat{Z}^K(u)\|_1 du.$$

Since J^* is finite and the reproduction rates are bounded, Equation (26) further implies that

$$\sum_{(x,\mathbf{k},y)\in J^*} \int_0^s |\rho^*_{x,\mathbf{k},y}(\widehat{\zeta}^K(u))|du \leqslant \frac{ct}{K}.$$

We thus obtain that

$$\mathbb{E}\left[\sup_{u\in[0,s]} \|\widehat{\zeta}^K(u) - \widehat{Z}^K(u)\ \|_1\right] \leqslant c\int_0^s \mathbb{E}\left[\sup_{r\in[0,u]} \|\widehat{\zeta}^K(r) - \widehat{Z}^K(r)\|_1\right] du + \frac{c(t)}{K}$$
$$+\frac{c}{\sqrt{K}}\left(\sum_{(x,\mathbf{k})\in J} \mathbb{E}\left[\sup_{u\in[0,s]} |M_{x,\mathbf{k}}(u)|\right] + \sum_{(x,\mathbf{k},y)\in J^*} \mathbb{E}\left[\sup_{u\in[0,s]} |N_{x,\mathbf{k},y}(u)|\right]\right).$$

Let $(x,\mathbf{k},y) \in J^*$. We know that there exists a constant m such that $\rho^*_{x,\mathbf{k},y}$ is bounded by m. It follows that, for any $r > 0$,

$$\mathbb{P}\left(\sup_{u\in[0,s]} |N_{x,\mathbf{k},y}(u)| \geqslant r\right) \leqslant 4\mathbb{P}(B_{x,\mathbf{k},y}(m^2 t) \geqslant r).$$

In particular, this implies existence of a constant $m_1(t)$ depending on t but not on K such that, for every $(x,\mathbf{k},y) \in J^*$ and $s \in [0,t]$,

$$\mathbb{E}\left[\sup_{u\in[0,s]} |N_{x,\mathbf{k},y}(u)|\right] \leqslant m_1(t).$$

Proceeding in the same way, there exists $m_2(t)$ depending on t but not on K such that, for every $(x,\mathbf{k}) \in J$ and $s \in [0,t]$,

$$\mathbb{E}\left[\sup_{u\in[0,s]} |M_{x,\mathbf{k}}(u)|\right] \leqslant m_2(t).$$

Taken together, this leads to the following inequality: for any $s \in [0,t]$,

$$\mathbb{E}\left[\sup_{u\in[0,s]} \|\widehat{\zeta}^K(u) - \widehat{Z}^K(u)\ \|_1\right] \leqslant c\int_0^s \mathbb{E}\left[\sup_{r\in[0,u]} \|\widehat{\zeta}^K(r) - \widehat{Z}^K(r)\|_1\right] du + \frac{c(t)}{\sqrt{K}}.$$

Thus, Gronwall's Lemma yields:

$$\mathbb{E}\left[\sup_{s\in[0,t]} \|\widehat{\zeta}^K(s) - \widehat{Z}^K(s)\ \|_1\right] \leqslant \frac{c(t)}{\sqrt{K}},$$

which finally implies that

$$\mathbb{P}\left(\sup_{s\in[0,t]} \|\widehat{\zeta}^K(s) - \widehat{Z}^K(s)\ \|_1 \geqslant \frac{\varepsilon_K}{2}\right) \leqslant \frac{c(t)}{\sqrt{K}\varepsilon_K}. \tag{27}$$

The conclusion follows by combining Equations (24) and (27). ☐

Next, we focus on quantifying how well $\widehat{Y}^K$ approaches Y^K, yielding the approximation error associated to replacing (Y^K, ζ^K) by $(\widehat{Y}^K, \widehat{\zeta}^K)$ on finite time intervals.

Lemma 4.5. *Let $t > 0$. Under Assumption 4.1, there exists a constant $C(t) > 0$ such that such that for every $K \geqslant 1$,*

$$\mathbb{P}(\forall s \in [0, t], Y^K(s) = \widehat{Y}^K(s) \text{ and } \|\zeta^K(s) - \widehat{Z}^K(s)\|_1 \leqslant \varepsilon_K) \geqslant 1 - C(t)(\alpha_K + \varepsilon_K).$$

Proof. We are interested in the first instant T_K at which $\widehat{Y}^K$ differs from Y^K:

$$T_K = \inf\{t \geqslant 0 : Y^K(t) \neq \widehat{Y}^K(t)\}.$$

Let $K \geqslant 1$. For two sets A and B, we let $A\Delta B$ designate their symmetric difference. For $(x, \mathbf{k}, y) \in J^*$, $(x', \mathbf{z}) \in \mathcal{S}$ and $\theta > 0$, define the event

$$\mathcal{E}_{x,\mathbf{k},y}(\theta, y, \mathbf{z}) = \{x' = x, \theta \leqslant \mathbf{k}_y \tau_{\mathbf{k}}(x, \mathbf{z}).\}$$

Notice that, by the coupling of Y^K and $\widehat{Y}^K$,

$$\begin{aligned}\{T_K \geqslant t\} &\supseteq \left\{ T_K \geqslant t, \sum_{(x,\mathbf{k},y)\in J*} \int_0^t \int_0^{+\infty} \mathbf{1}_{\{\mathcal{E}_{x,\mathbf{k},y}(\theta, Y^K(s-), \zeta^K(s-))\Delta\mathcal{E}_{x,\mathbf{k},y}(\theta, \hat{Y}^K(s-), \hat{Z}^K(s-))\}} Q_{x,\mathbf{k},y}(ds, d\theta) = 0 \right\} \\ &\supseteq \left\{ T_K \geqslant t, \sum_{(x,\mathbf{k},y)\in J*} \int_0^t \int_0^{+\infty} \mathbf{1}_{\{\mathcal{E}_{x,\mathbf{k},y}(\theta, Y^K(s-), \zeta^K(s-))\Delta\mathcal{E}_{x,\mathbf{k},y}(\theta, Y^K(s-), \hat{Z}^K(s-))\}} Q_{x,\mathbf{k},y}(ds, d\theta) = 0 \right\} \\ \{T_K \geqslant t\} &\supseteq \left\{ \sum_{(x,\mathbf{k},y)\in J*} \int_0^t \int_0^{+\infty} \mathbf{1}_{\{\mathcal{E}_{x,\mathbf{k},y}(\theta, Y^K(s-), \zeta^K(s-))\Delta\mathcal{E}_{x,\mathbf{k},y}(\theta, Y^K(s-), \hat{Z}^K(s-))\}} Q_{x,\mathbf{k},y}(ds, d\theta) = 0 \right\}.\end{aligned}$$

Let $\mathcal{A} = \{(y, \mathbf{k}) : \exists x \in \mathcal{X} \text{ s.t. } (x, \mathbf{k}) \in J, \mathbf{k}_y > 0\}$. Recall from Assumption 4.1 that for any $(x, \mathbf{k}) \in J$, $\tau_{\mathbf{k}}(x, \cdot)$ is $L_{x,\mathbf{k}}$-Lipschitz continuous. Let $L = \max_{(x,\mathbf{k})\in J} \|k\|_1 L_{x,\mathbf{k}}$. We introduce the event

$$A_K = \left\{ \max_{(y,\mathbf{k})\in\mathcal{A}} \sup_{s\in[0,t]} \mathbf{k}_y |\tau_{y,\mathbf{k}}(Y(s), \zeta^K(s)) - \tau_{y,\mathbf{k}}(Y^K(s), \widehat{Z}^K(s))| < L\varepsilon_K \right\}.$$

It follows that

$$\begin{aligned}\mathbb{P}(T_K \leqslant t_K) \leqslant \mathbb{P}(A_K, &\sum_{(x,\mathbf{k},y)\in J*} \int_0^t \int_0^{+\infty} \mathbf{1}_{\{\mathcal{E}_{x,\mathbf{k},y}(\theta, Y^K(s-), \zeta^K(s-))\Delta\mathcal{E}_{x,\mathbf{k},y}(\theta, Y^K(s-), \hat{Z}^K(s-))\}} Q_{x,\mathbf{k},y}(ds, d\theta) \geqslant 1) \\ &+ \mathbb{P}(A_K^C).\end{aligned} \tag{28}$$

First, we may notice that Assumption 4.1 ensures that

$$A_K^C \subseteq \{ \sup_{s\in[0,t]} \|\zeta^K(s) - \widehat{Z}^K(s)\|_1 \geqslant \varepsilon_K \},$$

from which we deduce that

$$\mathbb{P}(A_K^C) \leqslant \alpha_K. \tag{29}$$

Second, on the event A_K, it holds that for any $\theta \geqslant 0$ and $(x, \mathbf{k}, y) \in J^*$,

$$\begin{aligned}&\mathcal{E}_{x,\mathbf{k},y}(\theta, Y^K(s-), \zeta^K(s-))\Delta\mathcal{E}_{x,\mathbf{k},y}(\theta, Y^K(s-), \widehat{Z}^K(s-)) \\ &\qquad \subseteq \{\theta \in [\mathbf{k}_y \tau_{\mathbf{k}}(Y^K(s-), \zeta^K(s-)) - L\varepsilon_K, \mathbf{k}_y \tau_{\mathbf{k}}(Y^K(s-), \zeta^K(s-)) + L\varepsilon_K]\}.\end{aligned}$$

As a consequence,

$$\begin{aligned}
&\Big\{A_K, \sum_{(x,\mathbf{k},y)\in J*} \int_0^t \int_0^{+\infty} \mathbf{1}_{\{\mathcal{E}_{x,\mathbf{k},y}(\theta,Y^K(s-),\zeta^K(s-))\Delta\mathcal{E}_{x,\mathbf{k},y}(\theta,Y^K(s-),\widehat{Z}^K(s-))\}} Q_{x,\mathbf{k},y}(ds,d\theta) \geqslant 1\Big\} \\
&\qquad \subseteq \Big\{ \sum_{(x,\mathbf{k},y)\in J*} \int_0^t \int_0^{+\infty} \mathbf{1}_{\{\theta\in[\mathbf{k}_y\tau_{\mathbf{k}}(Y^K(s-),\zeta^K(s-))-L\varepsilon_K, \mathbf{k}_y\tau_{\mathbf{k}}(Y^K(s-),\zeta^K(s-))+L\varepsilon_K]\}} Q_{x,\mathbf{k},y}(ds,d\theta) \geqslant 1\Big\}.
\end{aligned}$$

Hence, Markov's inequality leads to

$$\begin{aligned}
&\mathbb{P}(A_K, \sum_{(x,\mathbf{k},y)\in J*} \int_0^t \int_0^{+\infty} \mathbf{1}_{\{\mathcal{E}_{x,\mathbf{k},y}(\theta,Y^K(s-),\zeta^K(s-))\Delta\mathcal{E}_{x,\mathbf{k},y}(\theta,Y^K(s-),\xi^K(s-))\}} Q_{x,\mathbf{k},y}(ds,d\theta) \geqslant 1) \\
&\leqslant \mathbb{E}[\sum_{(x,\mathbf{k},y)\in J*} \int_0^t \int_0^{+\infty} \mathbf{1}_{\{\theta\in[\mathbf{k}_y\tau_{\mathbf{k}}(Y^K(s-),\zeta^K(s-))-\varepsilon_K, \mathbf{k}_y\tau_{\mathbf{k}}(Y^K(s-),\zeta^K(s-))+\varepsilon_K]\}} Q_{x,\mathbf{k},y}(ds,d\theta)] \\
&\leqslant Ct\varepsilon_K,
\end{aligned} \tag{30}$$

with $C = 2L\text{Card}(J^*)$. By injecting Inequalities (29) and (30) into Equation (28), we thus obtain that

$$\mathbb{P}(\exists t \leqslant t : Y^K(t) \neq \widehat{Y}^K(t)) \leqslant \alpha_K + Ct\varepsilon_K. \tag{31}$$

The conclusion follows by combining Equations (23) and (31). □

In order to simplify notations, for any function $G : \mathbb{D}([0,t],\mathcal{X}) \times \mathbb{D}([0,t],\mathcal{Z}) \to \mathbb{R}$ as well as càdlàg trajectories $(x(s))_{s\leqslant t}$ in $\mathcal{X}$ and $(\mathbf{z}(s))_{s\leqslant t} = (\mathbf{z}_{i,x}(s))_{(i,x)\in\mathcal{X}*,s\leqslant t}$ in $\mathcal{Z}^2$, we let

$$G((x(s),\mathbf{z}(s))_{s\leqslant t}) \coloneqq G((x(s), \mathbf{z}|_{\mathcal{X}}(s))_{s\leqslant t}).$$

We are now ready to establish the main result of this Section.

Proof of Theorem 4.2. Throughout the proof, we write $\mathbb{Y}^K = (Y^K, \zeta^K)$ for the spinal construction in the initial population process, and $\widehat{\mathbb{Y}}^K = (\widehat{Y}^K, \widehat{\zeta}^K)$ for the spinal process in the diffusion approximation.

Let $t \geqslant 0$ and $F \in \mathfrak{L}$. It follows from Proposition 7 in [3] that

$$\mathbb{E}_{\mathfrak{X}(x,\mathbf{z}_0^K)}\left[\sum_{\substack{u\in\mathbb{G}(t),\\ u\succeq u_x}} F((x_u(s), Z^K(s))_{s\leqslant t})\right] = \mathbb{E}\left[H((\mathbb{Y}^K(s))_{s\leqslant t})\right],$$

where

$$H((\mathbb{Y}^K(s))_{s\leqslant t}) = \mathrm{e}^{\int_0^t \lambda(\mathbb{Y}^K(s))ds} F((\mathbb{Y}^K(s))_{s\leqslant t}).$$

As J is a finite set, Equation (17) and Assumption 4.1 imply that $H \in \mathfrak{L}$. In other words, there exists a constant M depending on t and F such that for any $x \in \mathbb{D}([0,t],\mathcal{X})$,

$$|H((x(s),\zeta^K(s))_{s\leqslant t}) - H((x(s),\widehat{\zeta}^K(s))_{s\leqslant t})| \leqslant M \sup_{s\in[0,t]} \|\zeta^K(s) - \widehat{\zeta}^K(s)\|_1. \tag{32}$$

Consider the event

$$A_K = \left\{\forall s \in [0,t], Y^K(s) = \widehat{Y}^K(s) \text{ and } \|\zeta^K(s) - \widehat{Z}^K(s)\|_1 \leqslant \varepsilon_K\right\}.$$

We have

$$\begin{aligned}
\mathbb{E}[|H((\mathbb{Y}^K(s))_{s\leqslant t}) - H((\widehat{\mathbb{Y}}^K(s))_{s\leqslant t})\ |] &\leqslant \mathbb{E}[|H((\mathbb{Y}^K(s))_{s\leqslant t}) - H((\widehat{\mathbb{Y}}^K(s))_{s\leqslant t})\ |\mathbf{1}_{A_K}] \\
&\quad + \mathbb{E}[|H((\mathbb{Y}^K(s))_{s\leqslant t}) - H((\widehat{\mathbb{Y}}^K(s))_{s\leqslant t})\ |\mathbf{1}_{A_K^C}]
\end{aligned}$$

On the one hand, it follows from the definition of A_K and Equation (32) that

$$\mathbb{E}[|H((\mathbb{Y}^K(s))_{s\leqslant t}) - H((\widehat{\mathbb{Y}}^K(s))_{s\leqslant t})\ |\mathbf{1}_{A_K}] \leqslant M\varepsilon_K.$$

On the other hand, the boundedness of H and Lemma 4.5 imply the existence of a constant c such that

$$\mathbb{E}[|H((\mathbb{Y}^K(s))_{s\leqslant t}) - H((\widehat{\mathbb{Y}}^K(s))_{s\leqslant t})\ |\mathbf{1}_{A_K^C}] \leqslant c(\alpha_K + \varepsilon_K).$$

Taken together, we thus obtain existence of a constant C such that

$$\mathbb{E}[|H((\mathbb{Y}^K(s))_{s\leqslant t}) - H((\widehat{\mathbb{Y}}^K(s))_{s\leqslant t})] \leqslant C(\alpha_K + \varepsilon_K).$$

Pick $\varepsilon_K = 2\max(K^{-1/4}, C(t)K^{-1}\log(K))$ with $C(t)$ as in Lemma 4.3. The conclusion follows by choosing K_0 such that for any $K \geqslant K_0$, $\varepsilon_K = K^{-1/4}$.

□

5 Discussion

Using a change of probability, we first exhibit a time-inhomogeneous Markov process which allows to establish a many-to-one formula, *i.e.* which behaves like a sampled lineage in the population process. When compared to the ψ-spine of [3], the general idea remains familiar. Indeed, reproduction rates are biased both along the spine and outside of it, transitions being most likely if they increase the average fertility of the spine, where fertility is measured by evaluating ψ over the spine's descendants at time t.

However, the ψ-auxiliary process has the advantage of relaxing the initial positivity assumption on ψ, which for instance opens the door to more efficient sampling in subpopulations. Further, the corresponding many-to-one formula does not require exponential weighting of trajectories. In particular, this implies that the bias of reproduction rates relying on the application m_ψ defined in Equation (5) accurately depicts the survivorship bias. In addition, this may facilitate the numerical evaluation of the many-to-one formula. However, it comes at the cost of introducing a time-inhomogeneity, a possible inconvenient being the necessity to evaluate m_ψ. Indeed, it can be delicate to compute explicitly, even for classical models such as the *SIR* model. As a consequence, it may need to be computed numerically, in which case simulating trajectories of the ψ-auxiliary process through standard algorithms such as [28] may require excessive computation time.

In this context, the large population approximation may be a pertinent regime. Heuristically, one would expect a similar result as in the case of the ψ-spine: under appropriate assumptions, the effect of the spine on the population state should vanish, as well as the bias of reproduction rates outside of the spine. Hence, the dynamics of the population's type composition are given by the deterministic limit z arising from the original population process. At time $s \leqslant t$, a spine of type x is thus expected to be replaced by descendance $\mathbf{k}$ and switch to type y at rate

$$\tau_{\mathbf{k}}(x, z(s))\mathbf{k}_y \frac{m_\psi(y, z(s), s, t)}{m_\psi(x, z(s), s, t)}.$$

In particular, the process describing the descendance of the spine then corresponds to a time-inhomogeneous multi-type branching process whose reproduction rates depend on a changing environment, given by z. If similarly to [11], we assume that z admits a stable equilibrium, then starting from this equilibrium, the descendance is described by a classical multi-type branching process.

Second, we focus on sampling in large populations, when the population type distribution can be approximated by a diffusion approximation capturing its fluctuations around the deterministic limit z. We obtain a many-to-one formula allowing for sampling in the diffusion approximation, and quantify the error made by sampling in the diffusion approximation rather than the original population process.

Notably, our means of quantification of the approximation error based on a coupling argument and on controlling the deviation of the spinal population type distribution from its approximation is not restricted to diffusion approximations. In particular, it may allow to quantify the speed of convergence of the ψ-spine to its large population limit [3]. In order to achieve this, we need some control of the form:

$$\mathbb{P}\left(\sup_{s\in[0,t_K]} \|\zeta^K(s) - z(s)\|_1 \geqslant \varepsilon_K\right) \leqslant \alpha_K,$$

with ε_K and α_K converging to zero and ideally t_K growing to infinity, as K grows large. We consider that expecting such control is reasonable. Indeed, it corresponds to understanding and controlling the fluctuations of the finite-population process ζ^K around its deterministic limit, which is a well studied question with several classical regimes: Gaussian fluctuations for $\varepsilon_K = O(K^{1/2})$, which are related to the diffusion approximation, moderate deviations with $\varepsilon_K = O(K^p)$ for $p \in (0, 1/2)$ and large deviations where $\varepsilon_K = O(1)$ (e.g. [26], [10, Chapters 2 and 4] for epidemic models). In particular, moderate and large deviations may be an interesting regime, as they allow to consider longer time scales. Nevertheless, in the case of the spinal population process, they are not immediate, due to a boundary problem arising from the fact that the spinal individual becomes negligible in large populations. As a consequence, these considerations are left for a futur work.

References

[1] Ellen Baake, Fernando Cordero, and Sebastian Hummel. A probabilistic view on the deterministic mutation–selection equation: Dynamics, equilibria, and ancestry via individual lines of descent. *Journal of Mathematical Biology*, 77(3):795–820, September 2018.

[2] Frank G. Ball and Peter Donnelly. Strong approximations for epidemic models. *Stochastic Processes and their Applications*, 55(1):1–21, January 1995.

[3] Vincent Bansaye. Spine for interacting populations and sampling. *Bernoulli*, 30(2):1555–1585, May 2024.

[4] Vincent Bansaye, Bertrand Cloez, and Pierre Gabriel. Ergodic behavior of non-conservative semigroups via generalized Doeblin's conditions. *Acta Applicandae Mathematicae*, 166(1):29–72, April 2020.

[5] Vincent Bansaye, Bertrand Cloez, Pierre Gabriel, and Aline Marguet. A non-conservative Harris ergodic theorem. *Journal of the London Mathematical Society*, 106(3), 2022.

[6] Vincent Bansaye, Xavier Erny, and Sylvie Méléard. Sharp approximation and hitting times for stochastic invasion processes, April 2023.

[7] Andrew D. Barbour and Gesine Reinert. Approximating the epidemic curve. *Electronic Journal of Probability*, 18:1–30, January 2013.

[8] Nathanaël Berestycki. Recent progress in coalescent theory. *Ensaios Matemáticos*, 16(1), 2009.

[9] Patrick Billingsley. *Convergence of Probability Measures.* Wiley Series in Probability and Statistics. Probability and Statistics Section. Wiley, New York, 2nd ed edition, 1999.

[10] Tom Britton and Etienne Pardoux, editors. *Stochastic Epidemics in a Homogeneous Community*, volume 2255. Springer Cham, 2019.

[11] Vincent Calvez, Benoît Henry, Sylvie Méléard, and Viet Chi Tran. Dynamics of lineages in adaptation to a gradual environmental change. *Annales Henri Lebesgue*, 5:729–777, 2022.

[12] Fernando Cordero, Sebastian Hummel, and Emmanuel Schertzer. General selection models: Bernstein duality and minimal ancestral structures. *The Annals of Applied Probability*, 32(3):1499–1556, June 2022.

[13] Pierre Del Moral. *Feynman-Kac Formulae.* Probability and Its Applications. Springer, New York, NY, 2004.

[14] Jean-Jil Duchamps, Félix Foutel-Rodier, and Emmanuel Schertzer. General epidemiological models: Law of large numbers and contact tracing. *Electronic Journal of Probability*, 28(none):1–37, January 2023.

[15] Stewart N. Ethier and Thomas G. Kurtz. *Markov Processes: Characterization and Convergence.* Wiley, New York, NY, April 1986.

[16] Hans-Otto Georgii and Ellen Baake. Supercritical multitype branching processes: The ancestral types of typical individuals. *Advances in Applied Probability*, 35(4):1090–1110, December 2003.

[17] Simon C. Harris, Marion Hesse, and Andreas E. Kyprianou. Branching Brownian motion in a strip: Survival near criticality. *The Annals of Probability*, 44(1):235–275, January 2016.

[18] Simon C. Harris and Matthew I. Roberts. The many-to-few lemma and multiple spines. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 53(1):226–242, February 2017.

[19] N. Ikeda and S. Watanabe. *Stochastic Differential Equations and Diffusion Processes.* Elsevier, June 2014.

[20] Thomas G. Kurtz. Strong approximation theorems for density dependent Markov chains. *Stochastic Processes and their Applications*, 6(3):223–240, February 1978.

[21] Russell Lyons, Robin Pemantle, and Yuval Peres. Conceptual Proofs of LLogL Criteria for Mean Behavior of Branching Processes. *The Annals of Probability*, 23(3):1125–1138, July 1995.

[22] Tibor Mach, Anja Sturm, and Jan M. Swart. Recursive tree processes and the mean-field limit of stochastic flows. *Electronic Journal of Probability*, 25(none):1–63, January 2020.

[23] Aline Marguet. Uniform sampling in a structured branching population. *Bernoulli*, 25(4A):2649–2695, November 2019.

[24] Charles Medous. Spinal constructions for continuous type-space branching processes with interactions, May 2024.

[25] Kai Nagel, Christian Rakow, and Sebastian A. Müller. Realistic agent-based simulation of infection dynamics and percolation. *Physica A: Statistical Mechanics and its Applications*, 584:126322, December 2021.

[26] Etienne Pardoux. Moderate deviations and extinction of an epidemic. *Electronic Journal of Probability*, 25(none):1–27, January 2020.

[27] Amnon Pazy. *Semigroups of Linear Operators and Applications to Partial Differential Equations*. Springer, Place of publication not identified, 2012.

[28] Vo Hong Thanh and Corrado Priami. Simulation of biochemical reactions with time-dependent rates by the rejection-based algorithm. *The Journal of Chemical Physics*, 143(5):054104, August 2015.