Some stochastic models in Eco-evolution

École de Recherche CIMPA — Mathématiques pour la Biologie

Sylvie Méléard* & Nicolas Champagnat†.

15 septembre 2016

Contents

1 Introduction **2**

2 Birth and Death Processes **3**
2.1 Definition and non-explosion criterion 3
2.2 Kolmogorov equations and invariant measure 6
2.3 Two trajectorial representations of birth and death processes 7
2.4 Extinction criterion 10
2.5 Extinction time 11
2.6 Coming down from infinity 13
2.7 Quasi-stationary distributions 16
2.7.1 Some coupling properties of birth and death processes 16
2.7.2 First properties of quasi-stationary distributions 18
2.7.3 Exponential convergence in total variation to the quasi-stationary distribution 20

3 Scaling Limits for Birth and Death Processes **25**
3.1 Deterministic approximation - Malthusian and logistic Equations 25
3.2 Lotka Volterra models 30

4 Population Point Measure Processes **33**
4.1 Pathwise construction 34
4.2 Examples and simulations 37
4.3 Martingale Properties 40

5 Scaling limits for the individual-based process **41**

*CMAP, Ecole Polytechnique, CNRS, route de Saclay, 91128 Palaiseau Cedex-France; E-mail: *sylvie.meleard@polytechnique.edu*

†IECL (Institut Elie Cartan de Lorraine) Université de Lorraine, Site de Nancy, B.P. 70239, F-54506 Vandœuvre-lès-Nancy Cedex France; E-mail: *Nicolas.Champagnat@inria.fr*

6 Limit of rare mutations - Convergence to the trait substation sequence 47

7 Canonical equation of the adaptive dynamics 49

8 Appendix : Poisson point measures 50

1 Introduction

This course concerns the stochastic modeling of population dynamics. In the first part, we focus on monotypic populations described by one dimensional stochastic differential equations with jumps. We consider their scaling limits for large populations and study the long time behavior of the limiting processes. It is achieved thanks to martingale properties, Poisson measure representations and stochastic calculus. These tools and results will be used and extended to measure-valued processes in the second part. The latter is dedicated to structured populations, where individuals are characterized by a trait belonging to a continuum.

In Section 2, we define birth and death processes with rates depending on the state of the population and recall some long time properties based on recursion equations. Two pathwise representations of the processes using Poisson point measures and Time-changed Poisson processes are introduced, from which we deduce some martingale properties. We carefully study the probability of extinction and in the case it is one, the law and moments of the extinction time. We also give a characterization of the property of coming down from infinity, which means that the birth and death process can be constructed starting from infinity, and immediately reaches finite value at all positive times. We finally study quasi-stationary distributions for birth and death processes, i.e. distributions which are stable conditionally on non-extinction of the population. We prove in particular that, when the process comes down from infinity, there is a unique stationary distribution which uniformly attracts all initial distributions. This result is based on coupling techniques.

In Section 3, we represent the carrying capacity of the underlying environment through a scaling parameter $K \in \mathbb{N}$ and state convergence results in the limit of large K. Depending on the demographic rates, the population size renormalized by K is approximated by the solution of an ordinary differential equation. We give two proofs, one based on time-changed Poisson processes and Gronwall Lemma, and the other on martingale properties and tightness-uniqueness arguments. When the per individual death rate is an affine function of the population size, in the limit we obtain a so called logistic equation. This approach can be generalized to a two-type birth and death process and leads in the large size approximation to a competitive 2d Lotka-Volterra system, whose long-time behavior is analyzed.

The second part of the document concerns structured populations whose individuals are characterized by a type taking values in a continuum. In the mathematical modeling of Darwinian evolution, this type is a heritable trait subject to selection and mutation. In Section 4, the population size process is constructred as a measure-valued Markov process with jumps. The population model includes mutations which may occur during each birth event with some positive probability. The mutant inherits a random perturbation of the ancestor's trait. The individuals compete for resources and the individual death rate depends on the whole population trait distribution, leading to nonlinearities in the limit. We develop some stochastic tools for such processes and use a pathwise representation driven by Poisson point measures to obtain martingale properties.

In the limit of large population size (scaled by the resource parameter K), we derive a nonlinear integro-differential equation in Section 5. The limiting theorem is proved using compactness-uniqueness arguments and the semimartingale decomposition of the measure-valued process. Simulations illustrate the convergence.

Section 6 focuses on the particular case of large population and rare mutations. The time scale at which the process is considered is now much longer (to see the impact of mutations) and the derivation in this case yields an evolutive jump process (for a suitable mutation probability) describing the successive invasions of successful mutants. When the initial population is monomorphic and as long as invasion implies fixation (this assumption can be checked on the parameters), the process jumps from an equilibrium state of the population to another one. This process is known as the Trait Substitution Sequence (TSS) and was first heuristically introduced by Metz et al. [38]. Section 7 is devoted to the study of the TSS in the limit of small mutational jumps. In this case, the TSS converges to the canonical equation of adaptive dynamics, which describes evolution as driven by a fitness gradient. **Notation**

For a Polish space E, $\mathcal{P}(E)$ denotes the space of probability measures on E.

The spaces $C_b^2(\mathbb{R})$, $C_b^2(\mathbb{R}_+)$, $C_b^2(\mathbb{R}^d)$ are the spaces of bounded continuous functions whose first and second derivatives are bounded and continuous, resp. on $\mathbb{R}$, $\mathbb{R}_+$, $\mathbb{R}^d$.

In all what follows, C denotes a constant real number whose value can change from one line to the other.

2 Birth and Death Processes

In this part, we concentrate on one-dimensional models for population dynamics. We recall the main properties of the birth and death processes.

2.1 Definition and non-explosion criterion

Definition 2.1. *A* ***birth and death process*** *is a pure jump Markov process whose jumps steps are equal to* ± 1. *The transition rates are as follows:*

$$\begin{cases} i \to i+1 & \textit{at rate} \quad \lambda_i \\ i \to i-1 & \textit{at rate} \quad \mu_i, \end{cases}$$

$(\lambda_i)_{i\in\mathbb{N}^*}$ *and* $(\mu_i)_{i\in\mathbb{N}^*}$ *being two sequences of positive real numbers and* $\lambda_0 = \mu_0 = 0$.

In this case, the infinitesimal generator is the matrix $(Q_{i,j})$ defined on $\mathbb{N}\times\mathbb{N}$ by

$$Q_{i,i+1} = \lambda_i \ , \ Q_{i,i-1} = \mu_i \ , \ Q_{i,i} = -(\lambda_i + \mu_i) \ , \quad Q_{i,j} = 0 \text{ otherwise.}$$

The global jump rate for a population with size $i \geq 1$ is $\lambda_i + \mu_i$. After a random time distributed according an exponential law with parameter $\lambda_i + \mu_i$, the process increases by 1 with probability $\frac{\lambda_i}{\lambda_i+\mu_i}$ and decreases by -1 with probability $\frac{\mu_i}{\lambda_i+\mu_i}$. If $\lambda_i + \mu_i = 0$, the process is absorbed at i. This construction gives a sequence of jump times $(T_n)_{n\geq 1}$ and a jump chain $(Y_n)_{n\geq 0}$. In the case where $T_\infty = \sup_n T_n < \infty$, the process is taken constant equal to an arbitrary value, say $+\infty$, after time T_∞. We thus obtain a process

$(X_t)_{t\geq 0}$ which is a right-continuous process taking values in $\mathbb{N}\cup\{+\infty\}$ such that, for all $n\geq 0$,

$$X_t = Y_n, \qquad \forall t\in[T_n, T_{n+1})$$

and $X_t=+\infty$ if $t\geq T_\infty=\sup_n T_n$, where $T_0=0$,

- the jump chain $(Y_n)_{n\geq 0}$ is a discrete-time Markov chain on $\mathbb{N}$ such that $Y_0=X_0$ and with transition probability $\frac{\lambda_i}{\lambda_i+\mu_i}$ from i to $i+1$ and $\frac{\mu_i}{\lambda_i+\mu_i}$ from i to $i-1$,
- the inter-jump times defined for all $n\geq 1$ as $S_n=T_n-T_{n-1}$ satisfy that, for all $n\geq 1$, conditional on $Y_0,\dots,Y_{n-1}$, the random variables $S_1,\dots,S_n$ are independent exponential with parameters $\lambda(Y_0)+\mu(Y_0),\dots,\lambda(Y_{n-1})+\mu(Y_{n-1})$ respectively, where we write $\lambda(i)=\lambda_i$ and $\mu(i)=\mu_i$ for all $i\geq 0$.

Recall that if $P(t)=(P_{i,j}(t)), t\in\mathbb{R}_+$ denotes the transition semigroup of the process, i.e. $P_{i,j}(t)=\mathbb{P}(X_t=j\mid X_0=i)$, then

$$P_{i,i+1}(h)=\lambda_i\,h+o(h)\ ;\ P_{i,i-1}(h)=\mu_i\,h+o(h)\ ;\ P_{i,i}(h)=1-(\lambda_i+\mu_i)\,h+o(h).$$

Examples: The constant numbers λ, μ, ρ, c are positive.
1) The Yule process corresponds to the case $\lambda_i=i\lambda,\ \ \mu_i=0$.
2) The branching process or linear birth and death process : $\lambda_i=i\lambda,\ \mu_i=i\mu$.
3) The birth and death process with immigration : $\lambda_i=i\lambda+\rho,\ \mu_i=i\mu$.
4) The logistic birth and death process : $\lambda_i=i\lambda,\ \mu_i=i\mu+c\,i(i-1)$.

The following theorem characterizes the non-explosion in finite time of the process. In this case, the process will have a.s. finite value at any time $t\in\mathbb{R}_+$.

Theorem 2.2. *Suppose that $\lambda_i>0$ for all $i\geq 1$. Then the birth and death process has almost surely an infinite life time if and only if the following series diverges:*

$$\sum_{i\geq 1}\left(\frac{1}{\lambda_i}+\frac{\mu_i}{\lambda_i\lambda_{i-1}}+\cdots+\frac{\mu_i\cdots\mu_2}{\lambda_i\cdots\lambda_2\lambda_1}\right)=+\infty. \tag{2.1}$$

Corollary 2.3. *If for any i, $\lambda_i\leq\lambda\,i$, with $\lambda>0$, the process is well defined on $\mathbb{R}_+$.*

Remark 2.4. One can check that the birth and death processes mentioned in the examples above satisfy this property and are well defined on $\mathbb{R}_+$.

Proof of Theorem 2.2. Let $(T_n)_n$ be the sequence of jump times of the process and $(S_n)_n$ the sequence of the inter-jump times,

$$S_n=T_n-T_{n-1},\quad\forall n\geq 1;\quad T_0=0,\quad S_0=0.$$

We define $T_\infty=\lim_n T_n$. The process doesn't explode almost surely and is well defined on $\mathbb{R}_+$ if and only if for any $i\geq 1$, $\mathbb{P}_i(T_\infty<+\infty)=0$.
The proof consists in showing that the process doesn't explode almost surely if and only if the unique non-negative and bounded solution $x=(x_i)_{i\in\mathbb{N}}$ of $Q\,x=x$ is the null solution. This proof is actually achieved for any integer valued pure jump Markov process. We will then see that it is equivalent to (2.1) for birth and death processes.

For any $i \geq 1$, we set $h_i^{(0)} = 1$ and for $n \geq 1$,

$$h_i^{(n)} = \mathbb{E}_i(\exp(-T_n)) = \mathbb{E}_i\left(\exp(-\sum_{k=1}^{n} S_k)\right).$$

We have

$$\mathbb{E}_i\left(\exp\left(-\sum_{k=1}^{n+1} S_k\right)\big|S_1\right) = \exp(-S_1)\ \mathbb{E}_i\left(\mathbb{E}_{X_{S_1}}\left(\exp(-\sum_{k=1}^{n} S_k)\right)\right),$$

by the Markov property, the independence of S_1 and X_{S_1} and since the jump times of the shifted process are $T_n - S_1$. Moreover,

$$\mathbb{E}_i\left(\mathbb{E}_{X_{S_1}}\left(\exp(-\sum_{k=1}^{n} S_k)\right)\right) = \sum_{j\neq i} \mathbb{P}_i(X_{S_1} = j)\ \mathbb{E}_j\left(\exp(-\sum_{k=1}^{n} S_k)\right) = \sum_{j\neq i} \frac{Q_{i,j}}{q_i}\ h_j^{(n)},$$

where $q_i = \sum_{j\neq i} Q_{i,j}$. Therefore, for all $n \geq 0$,

$$h_i^{(n+1)} = \mathbb{E}_i\left(\mathbb{E}_i\left(\exp(-\sum_{k=1}^{n+1} S_k)\big|S_1\right)\right) = \sum_{j\neq i} \frac{Q_{i,j}}{q_i}\ h_j^{(n)}\ \mathbb{E}_i(\exp(-S_1)).$$

Since $\mathbb{E}_i(\exp(-S_1)) = \int_0^\infty q_i e^{-q_i s} e^{-s} ds = \frac{q_i}{1+q_i}$, we finally obtain that

$$h_i^{(n+1)} = \sum_{j\neq i} \frac{Q_{i,j}}{1+q_i}\ h_j^{(n)}. \tag{2.2}$$

Let $(x_i)_i$ be a non-negative solution of $Qx = x$ bounded by 1. We get $h_i^{(0)} = 1 \geq x_i$ and thanks to the previous formula, we deduce by induction that for all $i \geq 1$ and for all $n \in \mathbb{N}$, $h_i^{(n)} \geq x_i \geq 0$. Indeed if $h_j^{(n)} \geq x_j$, we get $h_i^{(n+1)} \geq \sum_{j\neq i} \frac{Q_{i,j}}{1+q_i} x_j$. As x is solution of $Qx = x$, it satisfies $x_i = \sum_j Q_{i,j}\ x_j = Q_{i,i}x_i + \sum_{j\neq i} Q_{i,j}x_j = -q_i x_i + \sum_{j\neq i} Q_{i,j}x_j$, thus $\sum_{j\neq i} \frac{Q_{i,j}}{1+q_i} x_j = x_i$ and $h_i^{(n+1)} \geq x_i$.
If the process doesn't explode almost surely, we have $T_\infty = +\infty$ a.s. and $\lim_n h_i^{(n)} = 0$. Making n tend to infinity in the previous inequality, we deduce that $x_i = 0$. Thus, in this case, the unique non-negative and bounded solution of $Qx = x$ is the null solution.
Let us now assume that the process explodes with a positive probability. Let $z_i = \mathbb{E}_i(e^{-T_\infty})$. There exists i such that $\mathbb{P}_i(T_\infty < +\infty) > 0$ and for this integer i, $z_i > 0$. Going to the limit with $T_\infty = \lim_n T_n$ and $T_n = \sum_{k=1}^n S_k$ yields $z_j = \lim_n h_j^{(n)}$. Making n tend to infinity proves that z is a non-negative and bounded solution of $Qz = z$, with $z_i > 0$. It ensures that the process doesn't explode almost surely if and only if the unique non-negative and bounded solution $x = (x_i)_{i\in\mathbb{N}}$ of $Q\,x = x$ is $x = 0$.

We apply this result to the birth and death process. We assume that $\lambda_i > 0$ for $i \geq 1$ and $\lambda_0 = \mu_0 = 0$. Let $(x_i)_{i\in\mathbb{N}}$ be a non-negative solution of the equation $Qx = x$. For $n \geq 1$, introduce $\Delta_n = x_n - x_{n-1}$. Equation $Qx = x$ can be written $x_0 = 0$ and

$$\lambda_n x_{n+1} - (\lambda_n + \mu_n)x_n + \mu_n x_{n-1} = x_n\ ,\ \forall n \geq 1.$$

Setting $f_n = \dfrac{1}{\lambda_n}$ and $g_n = \dfrac{\mu_n}{\lambda_n}$, we get

$$\Delta_1 = x_1\,;\;\Delta_2 = \Delta_1\,g_1 + f_1\,x_1\,;\;\ldots\,;\;\Delta_{n+1} = \Delta_n\,g_n + f_n\,x_n.$$

Remark that for all n, $\Delta_n \geq 0$ and the sequence $(x_n)_n$ is non-decreasing. If $x_1 = 0$, the solution is zero. Otherwise we deduce that

$$\Delta_{n+1} = f_n x_n + \sum_{k=1}^{n-1} f_k\,g_{k+1}\cdots g_n\,x_k \;+\; g_1\cdots g_n\,x_1.$$

Since $(x_k)_k$ is non-decreasing and defining $\quad r_n = \dfrac{1}{\lambda_n} + \displaystyle\sum_{k=1}^{n-1} \frac{\mu_{k+1}\cdots\mu_n}{\lambda_k\lambda_{k+1}\cdots\lambda_n} + \frac{\mu_1\cdots\mu_n}{\lambda_1\cdots\lambda_n}$, it follows that $\;r_n\;x_1 \leq \Delta_{n+1} \leq r_n\;x_n$, and by iteration

$$x_1(1 + r_1 + \cdots + r_n) \leq x_{n+1} \leq x_1\;\prod_{k=1}^{n}(1 + r_k).$$

Therefore we have proved that the boundedness of the sequence $(x_n)_n$ is equivalent to the convergence of $\;\sum_k r_k$ and Theorem 2.2 is proved. □

2.2 Kolmogorov equations and invariant measure

Let us recall the Kolmogorov equations, (see for example Karlin-Taylor [29]).

Forward Kolmogorov equation: for all $i, j \in \mathbb{N}$,

$$\begin{aligned}
\frac{dP_{i,j}}{dt}(t) &= \sum_k P_{i,k}(t)\,Q_{k,j} = P_{i,j+1}(t)Q_{j+1,j} + P_{i,j-1}(t)Q_{j-1,j} + P_{i,j}(t)Q_{j,j} \\
&= \mu_{j+1}P_{i,j+1}(t) + \lambda_{j-1}P_{i,j-1}(t) - (\lambda_j + \mu_j)P_{i,j}(t). \qquad (2.3)
\end{aligned}$$

Backward Kolmogorov equation: for all $i, j \in \mathbb{N}$,

$$\begin{aligned}
\frac{dP_{i,j}}{dt}(t) &= \sum_k Q_{i,k}\,P_{k,j}(t) = Q_{i,i-1}P_{i-1,j}(t) + Q_{i,i+1}P_{i+1,j}(t) + Q_{i,i}P_{i,j}(t) \\
&= \mu_i P_{i-1,j}(t) + \lambda_i P_{i+1,j}(t) - (\lambda_i + \mu_i)P_{i,j}(t). \qquad (2.4)
\end{aligned}$$

Let us define for all $j \in \mathbb{N}$ the probability measure

$$p_j(t) = \mathbb{P}(X(t) = j) = \sum_i \mathbb{P}(X(t) = j|\,X_0 = i)\mathbb{P}(X(0) = i) = \sum_i \mathbb{P}(X(0) = i)P_{i,j}(t).$$

A straightforward computation shows that the forward Kolmogorov equation (2.3) reads

$$\frac{d\,p_j}{dt}(t) = \lambda_{j-1}\,p_{j-1}(t) + \mu_{j+1}\,p_{j+1}(t) - (\lambda_j + \mu_j)\,p_j(t). \qquad (2.5)$$

This equation is useful to find an invariant measure, that is a sequence $(q_j)_j$ of nonnegative real numbers with $\;\sum_j q_j < +\infty$ and satisfying for all j,

$$\lambda_{j-1}\,q_{j-1} + \mu_{j+1}\,q_{j+1} - (\lambda_j + \mu_j)\,q_j = 0.$$

2.3 Two trajectorial representations of birth and death processes

We consider as previously a birth and death process with birth rates $(\lambda_n)_n$ and death rates $(\mu_n)_n$. We write $\lambda_n = \lambda(n)$ and $\mu_n = \mu(n)$, where $\lambda(.)$ and $\mu(.)$ are two functions defined on $\mathbb{R}_+$. We assume further that there exist $\bar{\lambda} > 0$ and $\bar{\mu} > 0$ such that for any $x \geq 0$,

$$\lambda(x) \leq \bar{\lambda}\, x \quad ; \quad \mu(x) \leq \bar{\mu}(1 + x^2). \tag{2.6}$$

This assumption is satisfied for the logistic case where $\lambda(x) = \lambda\, x$ and $\mu(x) = cx(x-1) + \mu\, x$.

Assumption (2.6) is a sufficient condition ensuring the existence of the process on $\mathbb{R}_+$, as observed in Corollary 2.3.

Proposition 2.5. *On the same probability space, we consider a Poisson point measure $N(ds, du)$ with intensity $dsdu$ on $\mathbb{R}_+ \times \mathbb{R}_+$ (see Appendix). We also consider a random variable Z_0 independent of N and introduce the filtration $(\mathcal{F}_t)_t$ given by $\mathcal{F}_t = \sigma(Z_0, N((0, s] \times A), s \leq t, A \in \mathcal{B}(\mathbb{R}_+))$.*

The left-continuous and right-limited non-negative Markov process $(Z_t)_{t\geq 0}$ defined by

$$Z_t = Z_0 + \int_0^t \int_{\mathbb{R}_+} \big(\mathbf{1}_{\{u \leq \lambda(Z_{s-})\}} - \mathbf{1}_{\{\lambda(Z_{s-}) \leq u \leq \lambda(Z_{s-}) + \mu(Z_{s-})\}}\big)\, N(ds, du) \tag{2.7}$$

is a birth and death process with birth (resp. death) rates $(\lambda_n)_n$ (resp. $(\mu_n)_n$).

If for $p \geq 1$, $\mathbb{E}((Z_0)^p) < +\infty$, then for any $T > 0$,

$$\mathbb{E}\big(\sup_{t \leq T} (Z_t)^p\big) < +\infty. \tag{2.8}$$

Proof. For $n \in \mathbb{N}$, let us introduce the stopping times

$$T_n = \inf\{t > 0, Z_t \geq n\}.$$

For $s \leq t$, we have

$$\begin{aligned} Z^p_{s\wedge T_n} = Z_0^p \quad &+ \int_0^{s\wedge T_n} ((Z_{s-} + 1)^p - Z^p_{s-}) \mathbf{1}_{\{u \leq \lambda(Z_{s-})\}}\, N(ds, du) \\ &+ \int_0^{s\wedge T_n} ((Z_{s-} - 1)^p - Z^p_{s-}) \mathbf{1}_{\{\lambda(Z_{s-}) \leq u \leq \lambda(Z_{s-}) + \mu(Z_{s-})\}}\, N(ds, du). \end{aligned}$$

The second part of the r.h.s. is non-positive and the first part is increasing in time, yielding the upper bound

$$\sup_{s \leq t} Z^p_{s\wedge T_n} \leq Z_0^p + \int_0^{t\wedge T_n} ((Z_{s-} + 1)^p - Z^p_{s-}) \mathbf{1}_{\{u \leq \lambda(Z_{s-})\}}\, N(ds, du).$$

Since there exists $C > 0$ such that $(1 + x)^p - x^p \leq C(1 + x^{p-1})$ for any $x \geq 0$ and by (2.6), we get

$$\mathbb{E}(\sup_{s \leq t} Z^p_{s\wedge T_n}) \leq \mathbb{E}(Z_0^p) + C\, \bar{\lambda}\, \mathbb{E}\left(\int_0^{t\wedge T_n} Z_s\, (1 + Z_s^{p-1})\, ds\right) \leq \bar{C}\left(1 + \int_0^t \mathbb{E}\big(\sup_{u \leq s\wedge T_n} Z_u^p\big)\, ds\right),$$

where $\bar{C}$ is a positive number independent of n. Since the process is bounded by n before T_n, Gronwall's Lemma implies the existence (for any $T > 0$) of a constant number $C_{T,p}$ independent of n such that

$$\mathbb{E}\big(\sup_{t \le T \wedge T_n} Z_t^p\big) \le C_{T,p}. \tag{2.9}$$

In particular, the sequence $(T_n)_n$ tends to infinity almost surely. Indeed, otherwise there would exist $T_0 > 0$ such that $\mathbb{P}(\sup_n T_n < T_0) > 0$. Hence $\mathbb{E}\big(\sup_{t \le T_0 \wedge T_n} Z_t^p\big) \ge n^p\, \mathbb{P}(\sup_n T_n < T_0)$, which contradicts (2.9). Making n tend to infinity in (2.9) and using Fatou's Lemma yield (2.8).

To prove that $(Z_t)_{t\ge0}$ is a birth and death Markov chain, we first notice that for all $n \ge 1$, since λ and μ are bounded on $\{0, 1, \dots, n\}$, there is no accumulation of jump times in (2.7) up to time T_n. Since $T_\infty = \lim_n T_n = +\infty$, there is no accumulation of jump times in (2.7). We can then define the sequences of inter-jump times $(S_n)_{n\ge1}$ and the jump chain $(Y_n)_{n\ge0}$ associated to $(Z_t)_{t\ge0}$. We prove that these two sequences have the appropriate law by induction: for all $n \ge 1$, we define $J_n = S_0 + \ldots + S_n$ the n-th jump time. We have from (2.7) that S_{n+1} is $t - J_n$ where t is the first time $t > J_n$ such that $N((J_n, t] \times [0, \lambda(Y_{n-1}) + \mu(Y_{n-1})]) = 1$, and $Y_n = Y_{n-1} + 1$ if $N((J_n, t] \times [0, \lambda(Y_{n-1})]) = 1$ and $Y_n = Y_{n-1} - 1$ if $N((J_n, t] \times [\lambda(Y_{n-1}), \lambda(Y_{n-1}) + \mu(Y_{n-1})]) = 1$. Hence, given $(Y_0, \dots, Y_{n-1})$ and $(S_1, \dots, S_n)$, S_{n+1} and Y_n are independent, S_{n+1} is exponential with parameter $\lambda(Y_{n-1}) + \mu(Y_{n-1})$ and

$$\mathbb{P}(Y_n = k+1 \mid Y_{n-1} = k) = 1 - \mathbb{P}(Y_n = k-1 \mid Y_{n-1} = k) = \frac{\lambda(k)}{\lambda(k) + \mu(k)}.$$

The conclusion then follows by induction on n. □

Remark that given Z_0 and N, the process defined by (2.7) is unique. Indeed it can be inductively constructed. It is thus unique in law. Let us now recall its infinitesimal generator and give some martingale properties.

Theorem 2.6. *Let us assume that $\mathbb{E}(Z_0^p) < \infty$, for $p \ge 2$.*

(i) The infinitesimal generator of the Markov process Z is defined for any bounded measurable function ϕ from $\mathbb{R}_+$ into $\mathbb{R}$ by

$$L\phi(z) = \lambda(z)(\phi(z+1) - \phi(z)) + \mu(z)(\phi(z-1) - \phi(z)).$$

(ii) For any measurable function ϕ such that $|\phi(x)| + |L\phi(x)| \le C\,(1 + x^p)$, the process M^ϕ defined by

$$M_t^\phi = \phi(Z_t) - \phi(Z_0) - \int_0^t L\phi(Z_s)ds \tag{2.10}$$

is a left-limited and right-continous (càdlàg) $(\mathcal{F}_t)_t$-martingale.

(iii) The process M defined by

$$M_t = Z_t - Z_0 - \int_0^t (\lambda(Z_s) - \mu(Z_s))ds \tag{2.11}$$

is a square-integrable martingale with quadratic variation

$$\langle M \rangle_t = \int_0^t (\lambda(Z_s) + \mu(Z_s))ds. \tag{2.12}$$

Remark that the drift term of (2.11) involves the difference between the birth and death rates (i.e. the growth rate), while (2.12) involves the sum of both rates. Indeed the drift term describes the mean behavior whereas the quadratic variation reports the random fluctuations.

Proof. (i) is well known.

(ii) Dynkin's theorem implies that M^ϕ is a local martingale. By the assumption on ϕ, all the terms of the r.h.s. of (2.10) are integrable. Therefore M^ϕ is a martingale.

(iii) We first assume that $\mathbb{E}(Z_0^3) < +\infty$. By (2.6), we may apply (ii) to both functions $\phi_1(x) = x$ and $\phi_2(x) = x^2$. Hence $M_t = Z_t - Z_0 - \int_0^t (\lambda(Z_s) - \mu(Z_s))ds$ and $Z_t^2 - Z_0^2 - \int_0^t \big(\lambda(Z_s)(2Z_s + 1) - \mu(Z_s)(1 - 2Z_s)\big)ds$ are martingales. The process Z is a semi-martingale and Itô's formula applied to Z^2 gives that $Z_t^2 - Z_0^2 - \int_0^t 2Z_s\big(\lambda(Z_s) - \mu(Z_s)\big)ds - \langle M\rangle_t$ is a martingale. The uniqueness of the Doob-Meyer decomposition leads to (2.12). The general case $\mathbb{E}(Z_0^2) < +\infty$ follows by a standard localization argument. $\square$

We end this section with another trajectorial construction of birth and death processes as time changed Poisson processes.

Proposition 2.7. *On the same probability space, we consider two independent (standard) Poisson processes $(P_1(t))_{t\geq 0}$ and $(P_2(t))_{t\geq 0}$, independent of the random variable X_0. We make no assumption on the infinitesimal generator Q we consider here. In particular, it may be explosive. Then the equation*

$$X_t = \begin{cases} X_0 + P_1\left(\int_0^t \lambda(X_s)ds\right) - P_2\left(\int_0^t \mu(X_s)ds\right), & \forall t \geq 0 \text{ s.t. both integrals are finite,} \\ +\infty & \text{otherwise.} \end{cases} \tag{2.13}$$

admits a.s. a unique solution which is a birth and death process with generator Q.

Proof. Since the jump times of Poisson processes are isolated, the construction of $(X_t)_{t\geq 0}$ can be done pathwise inductively along the successive jump times of X_t. Given the n first values of the jump chain $Y_0, \dots, Y_{n-1}$ and the n first holding times $S_1, \dots, S_n$, we set $T_n = S_1 + \ldots + S_n$ the n-th jump time. Then the next jump time $T_n + S_{n+1}$ is the first time $t > T_n$ such that

$$P_1\left(\int_0^{T_n} \lambda(X_s)ds + \lambda(Y_{n-1})(t - T_n)\right) - P_1\left(\int_0^{T_n} \lambda(X_s)ds\right) \neq 0, \tag{2.14}$$

$$\text{or } P_1\left(\int_0^{T_n} \mu(X_s)ds + \mu(Y_{n-1})(t - T_n)\right) - P_1\left(\int_0^{T_n} \mu(X_s)ds\right) \neq 0, \tag{2.15}$$

and the next value Y_n of the jump chain is $Y_{n-1}+1$ (resp. $Y_{n-1}-1$) if (2.14) (resp. (2.15)) is satisfied first. By Markov's property for Poisson processes, the process in (2.14) is a Poisson process with rate $\lambda(Y_{n-1})$ and the one in (2.15) is a Poisson process with rate $\mu(Y_{n-1})$ independent of the first one. Hence, we deduce that, conditional on $(Y_0, \dots, Y_{n-1})$ and $(S_1, \dots, S_n)$, S_n is the infimum of two exponential variables with parameters $\lambda(Y_{n-1})$ and $\mu(Y_{n-1})$, hence is exponential of parameter $\lambda(Y_{n-1})+\mu(Y_{n-1})$, and Y_n is independent of S_{n+1}, with value $Y_{n-1} + 1$ with probability $\frac{\lambda(Y_{n-1})}{\lambda(Y_{n-1})+\mu(Y_{n-1})}$ and $Y_{n-1} - 1$ otherwise.

This proves that the sequences $(Y_n)_n$ and $(S_n)_n$ have the distribution corresponding to that of a birth and death process with infinitesimal generator Q.

Since the jump times of Poisson processes are isolated, it is also clear that the first accumulation point of the sequence of jump times $(T_n)_n$ is exactly the first time where either $\int_0^t \lambda(X_s)ds = +\infty$ or $\int_0^t \mu(X_s)ds = +\infty$. □

2.4 Extinction criterion

Let us come back to the general case.

Some of the following computation can be found in [29] or in [2], but they are finely developed in [4].

Let T_0 denote the extinction time and $u_i = \mathbb{P}_i(T_0 < \infty)$ the probability to see extinction in finite time starting from state i.
Conditioning by the first jump $X_{T_1} \in \{-1, +1\}$, we get the following recurrence property: for all $i \geq 1$,

$$\lambda_i u_{i+1} - (\lambda_i + \mu_i)u_i + \mu_i u_{i-1} = 0 \tag{2.16}$$

This equation can also be easily obtained from the backward Kolmogorov equation (2.4). Indeed

$$u_i = \mathbb{P}_i(\exists t > 0, X_t = 0) = \mathbb{P}_i(\cup_t \{X_t = 0\}) = \lim_{t\to\infty} P_{i,0}(t),$$

and

$$\frac{dP_{i,0}}{dt}(t) = \mu_i P_{i-1,0}(t) + \lambda_i P_{i+1,0}(t) - (\lambda_i + \mu_i)P_{i,0}(t).$$

Let us solve (2.16). We know that $u_0 = 1$. Let us first assume that for a state N, $\lambda_N = 0$ and $\lambda_i > 0$ for $i < N$. Define $u_i^{(N)} = \mathbb{P}_i(T_0 < T_N)$, where T_N is the hitting time of N. Thus $u_0^N = 1$ et $u_N^N = 0$. Setting

$$U_N = \sum_{k=1}^{N-1} \frac{\mu_1 \cdots \mu_k}{\lambda_1 \cdots \lambda_k},$$

straightforward computations using (2.16) yield that for $i \in \{1, \cdots, N-1\}$

$$u_i^{(N)} = (1 + U_N)^{-1} \sum_{k=i}^{N-1} \frac{\mu_1 \cdots \mu_k}{\lambda_1 \cdots \lambda_k} \quad \textit{and in particular} \quad u_1^{(N)} = \frac{U_N}{1 + U_N}.$$

For the general case, let N tend to infinity. We observe that extinction will happen (or not) almost surely in finite time depending on the convergence of the series $\displaystyle\sum_{k=1}^{\infty} \frac{\mu_1 \cdots \mu_k}{\lambda_1 \cdots \lambda_k}$.

Theorem 2.8. *(i) If* $\displaystyle\sum_{k=1}^{\infty} \frac{\mu_1 \cdots \mu_k}{\lambda_1 \cdots \lambda_k} = +\infty$, *then the extinction probabilities* u_i *are equal to* 1. *Hence we have almost-sure extinction of the birth and death process for any finite initial condition.*

(ii) If $\displaystyle\sum_{k=1}^{\infty} \frac{\mu_1 \cdots \mu_k}{\lambda_1 \cdots \lambda_k} = U_\infty < \infty$, *then for* $i \geq 1$,

$$u_i = (1 + U_\infty)^{-1} \sum_{k=i}^{\infty} \frac{\mu_1 \cdots \mu_k}{\lambda_1 \cdots \lambda_k}.$$

There is a positive probability for the process to survive for any positive inital condition.

Application of Theorem 2.8 to the binary branching process (linear birth and death process): any individual gives birth at rate λ and dies at rate μ. The population process is a binary branching process and individual life times are exponential variables with parameter $\lambda + \mu$. An individual either gives birth to 2 individuals with probability $\frac{\lambda}{\lambda+\mu}$ or dies with probability $\frac{\mu}{\lambda+\mu}$.

Applying the previous results, one gets that when $\lambda \leq \mu$, i.e. when the process is subcritical or critical, the sequence $(U_N)_N$ tends to infinity with N and there is extinction with probability 1. Conversely, if $\lambda > \mu$, the sequence $(U_N)_N$ converges to $\frac{\mu}{\lambda-\mu}$ and straightforward computations yield $u_i = (\mu/\lambda)^i$.

Application of Theorem 2.8 to the logistic birth and death process. Let us assume that the birth and death rates are given by

$$\lambda_i = \lambda\, i \; ; \; \mu_i = \mu\, i + c\, i(i-1). \tag{2.17}$$

The parameter c models the competition pressure between two individuals. It's easy to show that in this case, the series $\displaystyle\sum_{k=1}^{\infty} \frac{\mu_1 \cdots \mu_k}{\lambda_1 \cdots \lambda_k}$ diverges, leading to the almost sure extinction of the process. Hence the competition between individuals makes the extinction inevitable.

2.5 Extinction time

Let us now come back to the general case and assume that the series $\displaystyle\sum_{k=1}^{\infty} \frac{\mu_1 \cdots \mu_k}{\lambda_1 \cdots \lambda_k}$ diverges. The extinction time T_0 is well defined and we wish to compute its moments.

We use the standard notation

$$\pi_1 = \frac{1}{\mu_1} \; ; \; \pi_n = \frac{\lambda_1 \ldots \lambda_{n-1}}{\mu_1 \ldots \mu_n} \qquad \forall n \geq 2.$$

We now focus on the time spent by the process $(X(t), t \geq 0)$ to go from level $n+1$ to level n. For $n \geq 0$, we introduce the function

$$G_n(a) := \mathbb{E}_{n+1}(\exp(-aT_n)), \; a > 0,$$

where T_n is the first hitting time of the level n.

Proposition 2.9. *Let us assume that*

$$\sum_{k=1}^{\infty} \frac{\mu_1 \cdots \mu_k}{\lambda_1 \cdots \lambda_k} = \sum_n \frac{1}{\lambda_n \pi_n} = +\infty. \tag{2.18}$$

Then

(i) For any $a > 0$ and $n \geq 1$,

$$G_n(a) = \mathbb{E}_{n+1}(\exp(-aT_n)) = 1 + \frac{\mu_n + a}{\lambda_n} - \frac{\mu_n}{\lambda_n} \frac{1}{G_{n-1}(a)}. \tag{2.19}$$

(ii) $\mathbb{E}_1(T_0) = \sum_{k\geq 1} \pi_k$ and for every $n \geq 2$,

$$\mathbb{E}_{n+1}(T_n) \quad = \quad \frac{1}{\lambda_n \pi_n} \sum_{i \geq n+1} \pi_i \,, \tag{2.20}$$

$$\mathbb{E}_n(T_0) = \sum_{k\geq 1} \pi_k + \sum_{k=1}^{n-1} \frac{1}{\lambda_k \pi_k} \sum_{i\geq k+1} \pi_i = \sum_{k=1}^{n-1} \left(\sum_{i \geq k+1} \frac{\lambda_{k+1} \ldots \lambda_{i-1}}{\mu_{k+1} \ldots \mu_i} \right). \tag{2.21}$$

Proof. (i) Let τ_n be a random variable distributed as T_n under $\mathbb{P}_{n+1}$ and consider the Laplace transform of τ_n. Following [3, p. 264] and by the Markov property, we have

$$\tau_{n-1} \overset{(d)}{=} \mathbf{1}_{\{Y_n=-1\}} \mathcal{E}_n + \mathbf{1}_{\{Y_n=1\}} \left(\mathcal{E}_n + \tau_n + \tau'_{n-1} \right)$$

where Y_n, $\mathcal{E}_n$, τ'_{n-1} and τ_n are independent random variables, $\mathcal{E}_n$ is an exponential random variable with parameter $\lambda_n + \mu_n$ and τ'_{n-1} is distributed as τ_{n-1} and $\mathbb{P}(Y_n = 1) = 1 - \mathbb{P}(Y_n = -1) = \lambda_n/(\lambda_n + \mu_n)$. Hence, we get

$$G_{n-1}(a) = \frac{\lambda_n + \mu_n}{a + \lambda_n + \mu_n} \left(G_n(a) G_{n-1}(a) \frac{\lambda_n}{\lambda_n + \mu_n} + \frac{\mu_n}{\lambda_n + \mu_n} \right)$$

and (2.19) follows.

(ii) Differentiating (2.19) at $a = 0$, we get

$$\mathbb{E}_n(T_{n-1}) = \frac{\lambda_n}{\mu_n} \, \mathbb{E}_{n+1}(T_n) + \frac{1}{\mu_n}, \quad n \geq 1.$$

Following the proof of Theorem 2.8, we first deal with the particular case when $\lambda_N = 0$ for some $N > n$, $\mathbb{E}_N(T_{N-1}) = \frac{1}{\mu_N}$ and a simple induction gives

$$\mathbb{E}_n(T_{n-1}) = \frac{1}{\mu_n} + \sum_{i=n+1}^{N} \frac{\lambda_n \ldots \lambda_{i-1}}{\mu_n \ldots \mu_i}.$$

We get $\mathbb{E}_1(T_0) = \sum_{k=1}^N \pi_k$ and writing $\mathbb{E}_n(T_0) = \sum_{k=1}^n \mathbb{E}_k(T_{k-1})$, we deduce that

$$\mathbb{E}_n(T_0) = \sum_{k=1}^{N} \pi_k + \sum_{k=1}^{n-1} \frac{1}{\lambda_k \pi_k} \sum_{i=k+1}^{N} \pi_i.$$

In the general case, let $N > n$. Thanks to (2.18), T_0 is finite and the process a.s. does not explode in finite time for any initial condition. Then $T_N \to \infty$ $\mathbb{P}_n$-a.s., where we use the convention $\{T_N = +\infty\}$ on the event where the process does not attain N. The monotone convergence theorem yields

$$\mathbb{E}_n(T_0; T_0 \le T_N) \underset{N\to+\infty}{\longrightarrow} \mathbb{E}_n(T_0).$$

Let us consider a birth and death process X^N with birth and death rates $(\lambda_k^N, \mu_k^N : k \ge 0)$ such that $(\lambda_k^N, \mu_k^N) = (\lambda_k, \mu_k)$ for $k \ne N$ and $\lambda_N^N = 0, \mu_N^N = \mu_N$.
Since $(X_t : t \le T_N)$ and $(X_t^N : t \le T_N^N)$ have the same distribution under $\mathbb{P}_n$, we get

$$\mathbb{E}_n\left(T_0; T_0 \le T_N\right) = \mathbb{E}_n\left(T_0^N; T_0^N \le T_N^N\right),$$

which yields

$$\mathbb{E}_n(T_0) = \lim_{N\to\infty} \mathbb{E}_n\left(T_0^N; T_0^N \le T_N^N\right) \le \lim_{N\to\infty} \mathbb{E}_n\left(T_0^N\right),$$

where the convergence of the last term is due to the stochastic monotonicity of T_0^N with respect to N under $\mathbb{P}_n$. Using now that T_0^N is stochastically smaller than T_0 under $\mathbb{P}_n$, we have also

$$\mathbb{E}_n(T_0) \ge \mathbb{E}_n(T_0^N).$$

We deduce that

$$\mathbb{E}_n(T_0) = \lim_{N\to\infty} \mathbb{E}_n(T_0^N) = \lim_{N\to\infty} \sum_{k=1}^{N} \pi_k + \sum_{k=1}^{n-1} \frac{1}{\lambda_k \pi_k} \sum_{i=k+1}^{N} \pi_i,$$

which ends up the proof. □

The proof of the next proposition is left to the reader.

Proposition 2.10. *Assume* (2.18). *Show that for every $n \ge 0$,*

$$\begin{aligned}
\mathbb{E}_{n+1}(T_n^2) &= \frac{2}{\lambda_n \pi_n} \sum_{i\ge n} \lambda_i \pi_i \, \mathbb{E}_{i+1}(T_i)^2; \\
\mathbb{E}_{n+1}(T_n^3) &= \frac{6}{\lambda_n \pi_n} \sum_{i\ge n} \lambda_i \pi_i \, \mathbb{E}_{i+1}(T_i) \, \mathrm{Var}_{i+1}(T_i).
\end{aligned}$$

2.6 Coming down from infinity

The first lemma allows us to define the law of the process starting from infinity. As in Donnelly [15], our main tool is a monotonicity argument. We set $\overline{\mathbb{N}} := \{0, 1, \ldots\} \cup \{\infty\}$ and for any $T > 0$, we denote by $\mathbb{D}_{\overline{\mathbb{N}}}([0,T])$ the Skorohod space of càdlàg functions on $[0,T]$ with values in $\overline{\mathbb{N}}$.

Lemma 2.11. *Under* (2.18), *the sequence $(\mathbb{P}_n)_n$ converges weakly in the space of probability measures on $\mathbb{D}_{\overline{\mathbb{N}}}([0,T])$ to a probability measure $\mathbb{P}_\infty$.*

At this point, the limiting process is not assumed to be finite for positive times.

Proof. We follow the tightness argument given in the first part of the proof of Theorem 1 by Donnelly in [15]. Indeed, no integer is an instantaneous state for the process ($\lambda_n, \mu_n < \infty$ for each $n \geq 0$) and the process is stochastically monotone with respect to the initial condition. It ensures that Assumption (A1) of [15] holds. In addition, Assumption (2.18) ensures that the process almost surely does not explode and (A2) of [15, Thm. 1] is also satisfied by denoting B_n^N the birth and death process X issued from n and stopped in N. Then the tightness holds and we identify the finite marginal distributions by noticing that for $k \geq 1$, for $t_1, \ldots, t_k \geq 0$ and for $a_1, \ldots, a_k \in \mathbb{N}$, the quantities $\mathbb{P}_n(X(t_1) \leq a_1, \cdots, X(t_k) \leq a_k)$ are non-increasing with respect to $n \in \mathbb{N}$ (and thus converge). $\square$

When the process starting from infinity is non-degenerate, it hits finite values in finite time with positive probability. More precisely, we say that the process *comes down from infinity* if there exist $t > 0$ and $y \in \mathbb{R}_+$, such that $\mathbb{P}_\infty(T_y < t) > 0$.

Characterizations of the coming down from infinity have been given in [3, 8]. They rely on the convergence of the mean time of absorption when the initial condition goes to infinity or equivalently to the convergence of the non-decreasing sequence $\mathbb{E}_n(T_0)$ as $n \to \infty$:

$$S = \mathbb{E}_\infty(T_0) = \sum_{i\geq 1} \pi_i + \sum_{n\geq 1} \frac{1}{\lambda_n \pi_n} \sum_{i \geq n+1} \pi_i = \sum_{n\geq 0} \left(\frac{1}{\mu_{n+1}} + \sum_{i\geq n+2} \frac{\lambda_{n+1} \cdots \lambda_{i-1}}{\mu_{n+1} \cdots \mu_i} \right) < +\infty. \tag{2.22}$$

This is equivalent to the existence and uniqueness of the quasi-stationary distribution related to the absorbing point zero (see [40], [8]) and to the finiteness of some exponential moments of T_0.

In the next lines, we show, using monotonicity properties, that it is also equivalent to instantaneous almost-sure coming down from infinity (Proposition 2.13). The latter is a stronger notion of coming down from infinity corresponding to the behavior of birth and death processes under (2.18) and (2.22).

Definition 2.12. *The process* $(X(t), t \geq 0)$ *instantaneously comes down from infinity if*

$$\mathbb{P}_\infty(\forall t > 0, X(t) < +\infty) = 1. \tag{2.23}$$

Using Lemma 2.11 and that $X \in \mathbb{D}_{\overline{\mathbb{N}}}([0,T])$, is quasi-left continuous and $+\infty$ is not accessible from $\mathbb{N}$, we have for any $0 < t_0 < T$,

$$\mathbb{P}_\infty(\forall t \in [t_0, T), X(t) < +\infty) = \lim_{m\to\infty} \lim_{k\to\infty} \mathbb{P}_k(\forall t \in [t_0, T), X(t) \leq m),$$

we get the equivalence between (2.23) and

$$\forall t \in [0, T], \qquad \lim_{m\to\infty} \lim_{k\to+\infty} \mathbb{P}_k(T_m < t) = 1.$$

In particular, a process satisfying (2.23) comes down from infinity.

Let us now show that (2.23) is satisfied under (2.18) and (2.22). In fact we give several necessary and sufficient conditions for $(X(t), t \geq 0)$ to come down from infinity. The first two ones are directly taken from [8]. We add here an exponential moment criterion. We also mention that it is equivalent to the existence (cf. [40]) and uniqueness (cf. [8]) of a quasi-stationary distribution for the process X.

Proposition 2.13. *Under condition* (2.18)*, the following assertions are equivalent:*

(i) The process $(X(t), t \geq 0)$ *comes down from infinity.*

(ii) The process $(X(t), t \geq 0)$ *instantaneously comes down from infinity.*

(iii) Assumption (2.22) *is satisfied:* $S < +\infty$.

(iv) $\sup_{k\geq 0} \mathbb{E}_k[T_0] < +\infty$.

(v) For all $a > 0$*, there exists* $k_a \in \mathbb{N}$ *such that* $\mathbb{E}_\infty\left(\exp(aT_{k_a})\right) < +\infty$.

Proof. (iii) and (iv) are clearly equivalent, using (2.21). As already mentioned, (ii) implies (i). From [3], Section 8.1, we have that (iii) and (i) are equivalent. We now prove that (v) implies (i) and that (iii) implies (v) and that (iii) implies (ii) to complete the proof.

First, we check that (v) implies that X comes down from infinity. Indeed, taking $a = 1$ in (v), we have $M := \mathbb{E}_\infty\left(\exp(T_{k_1})\right) < +\infty$. Then, Markov inequality ensures that for all $k \geq k_1$ and $t \geq 0$, $\mathbb{P}_k(T_{k_1} < t) \geq 1 - \exp(-t)M$. Choosing t large enough ensures $\mathbb{P}_\infty(T_{k_1} < t) > 0$ and (i) holds.

We then prove that (iii) implies (v). We fix $a > 0$ and using $S < +\infty$, there exists $k_a > 1$ such that

$$\sum_{n\geq k_a-1} \frac{1}{\lambda_n \pi_n} \sum_{i\geq n+1} \pi_i \leq \frac{1}{a}.$$

We now define the Lyapounov function J_a as

$$J_a(m) := \begin{cases} \displaystyle\sum_{n=k_a-1}^{m-1} \frac{1}{\lambda_n \pi_n} \sum_{i\geq n+1} \pi_i & \text{if } m \geq k_a\,, \\ 0 & \text{if } m < k_a\,. \end{cases}$$

We notice that J_a is non-decreasing and bounded and we introduce the infinitesimal generator L of X, defined by

$$L(f)(n) = \left(f(n+1) - f(n)\right)\lambda_n + \left(f(n-1) - f(n)\right)\mu_n,$$

for any bounded function f and any $n \geq 1$. Then, the process

$$M_t := e^{at} J_a(X(t)) - \int_0^t e^{au}\left(aJ_a(X(u)) + LJ_a(X(u))\right) \mathrm{d}u, \qquad (t \geq 0)$$

is a martingale with respect to the natural filtration of X. Adding that $LJ_a(m) = -1$ for any $m \geq k_a$ and that $J_a(X(u)) \leq J_a(\infty) \leq 1/a$, we get for all $k \geq k_a$ and $t \geq 0$,

$$\begin{aligned} \mathbb{E}_k\left(e^{at\wedge T_{k_a}} J_a(X(t\wedge T_{k_a}))\right) &= \mathbb{E}_k\left(\int_0^{t\wedge T_{k_a}} e^{au}\left(aJ_a(X(u)) + LJ_a(X(u))\right)\mathrm{d}u\right) + J_a(k) \\ &= \mathbb{E}_k\left(\int_0^{t\wedge T_{k_a}} e^{au}\left(aJ_a(X(u)) - 1\right)\mathrm{d}u\right) + J_a(k) \\ &\leq J_a(k). \end{aligned}$$

Adding that for any $k \geq k_a$, $\mathbf{p}_k$-a.s. $J_a(X(t \wedge T_{k_a})) \geq J_a(k_a)$, we get $\mathbb{E}_k\left(e^{at\wedge T_{k_a}}\right) \leq \frac{J_a(k)}{J_a(k_a)}$. Then (v) follows from the monotone convergence theorem and Assumption (iii).

It remains to show that (iii) implies (ii). On the one hand, according to (2.20), $\mathbb{E}_\infty(T_n) = \sum_{i\geq n}\mathbb{E}_{i+1}(T_i)$ and Assumption (iii) entails that $\mathbb{E}_\infty(T_n)$ vanishes as $n \to \infty$ as the rest of the finite series S. On the other hand, under $\mathbb{P}_\infty$, the sequence $(T_n)_{n\geq 0}$ decreases to the random variable $T_\mathbb{N}$. Then, from the monotone convergence theorem, $\mathbb{E}_\infty(T_n)$ decreases to $\mathbb{E}_\infty(T_\mathbb{N})$ and $\mathbb{E}_\infty(T_\mathbb{N}) = 0$. It ensures that $T_\mathbb{N} = 0$ $\mathbb{P}_\infty$ a.s. and X instantaneously comes down from infinity. The proof is then complete. □

2.7 Quasi-stationary distributions

2.7.1 Some coupling properties of birth and death processes

Given a pair of random variables $(\bar{X}_0, \bar{Y}_0)$ and the infinitesimal generator Q of a non-explosive birth and death process, we can construct on the same probability space two birth and death processes $(X_t)_{t\geq 0}$ and $(Y_t)_{t\geq 0}$ with the same infinitesimal generator Q such that (X_0, Y_0) has the same law as $(\bar{X}_0, \bar{Y}_0)$ and such that, for all $i, j \in \mathbb{N}$, given $(X_0, Y_0) = (i, j)$, the processes $(X_t)_{t\geq 0}$ and $(Y_t)_{t\geq 0}$ are independent. We define their coupling time τ as

$$\tau = \inf\{t \geq 0, X_t = Y_t\}.$$

We also define the process $(\widetilde{X}_t)_{t\geq 0}$ as

$$\widetilde{X}_t = \begin{cases} X_t & \text{if } t < \tau, \\ Y_t & \text{if } t \geq \tau. \end{cases} \tag{2.24}$$

Proposition 2.14. *The process $(\widetilde{X}_t)_{t\geq 0}$ is a birth and death process with initial value X_0 and infinitesimal generator Q.*

Proof. The process $(X_t, Y_t)_{t\geq 0}$ is a strong Markov process in $\mathbb{N}^2$ and τ is a stopping time for this process. Hence, defining $\mathcal{F}_t = \sigma(X_s, Y_s, s \leq t)$ and the stopped σ-field $\mathcal{F}_\tau$ as usual, for fixed $n \geq 1$, $0 = t_0 \leq \ldots \leq t_n$ and $i_0, \ldots, i_n \in \mathbb{N}$,

$$\begin{aligned}&\mathbb{P}(\widetilde{X}_{t_0} = i_0, \ldots, \widetilde{X}_{t_n} = i_n \mid \mathcal{F}_\tau) \\ &\qquad = \sum_{k=0}^{n} \mathbb{1}_{t_k \leq \tau < t_{k+1}} \mathbb{P}(X_{t_0} = i_0, \ldots, X_{t_k} = i_k, Y_{t_{k+1}} = i_{k+1}, \ldots, Y_{t_n} = i_n \mid \mathcal{F}_\tau),\end{aligned}$$

with $t_{n+1} = +\infty$. Hence, denoting $\mathbb{P}_{(i,j)} = \mathbb{P}(\cdot \mid (X_0, Y_0) = (i, j))$ and $\mathbb{P}_i = \mathbb{P}(\cdot \mid X_0 = i)$, it follows from the Markov property at time τ that

$$\begin{aligned}
&\mathbb{P}(\widetilde{X}_{t_0}=i_0,\dots,\widetilde{X}_{t_n}=i_n \mid \mathcal{F}_\tau)\\
&\quad=\sum_{k=0}^{n}\mathbb{1}_{t_k\le\tau<t_{k+1}}\,\mathbb{1}_{X_{t_0}=i_0,\dots,X_{t_k}=i_k}\,\mathbb{P}_{(X_\tau,Y_\tau)}(Y_{t_{k+1}-\tau}=i_{k+1},\dots,Y_{t_n-\tau}=i_n)\\
&\quad=\sum_{k=0}^{n}\mathbb{1}_{t_k\le\tau<t_{k+1}}\,\mathbb{1}_{X_{t_0}=i_0,\dots,X_{t_k}=i_k}\,\mathbb{P}_{X_\tau}(X_{t_{k+1}-\tau}=i_{k+1},\dots,X_{t_n-\tau}=i_n)\\
&\quad=\sum_{k=0}^{n}\mathbb{1}_{t_k\le\tau<t_{k+1}}\,\mathbb{1}_{X_{t_0}=i_0,\dots,X_{t_k}=i_k}\,\mathbb{P}(X_{t_{k+1}}=i_{k+1},\dots,X_{t_n}=i_n \mid \mathcal{F}_\tau)\\
&\quad=\mathbb{P}(X_{t_0}=i_0,\dots,X_{t_n}=i_n \mid \mathcal{F}_\tau).
\end{aligned}$$

Taking the expectation of both sides, we deduce that the processes $(\widetilde{X}_t)_{t\ge0}$ and $(X_t)_{t\ge0}$ have the same law. □

This property allows to give bounds on the total variation distance between the distributions of two birth and death processes. Let us first recall the definition of the total variation distance between probability measures.

Definition 2.15. *Given π,ν two probability measures on $\mathbb{N}$, their total variation distance is given by one the following equivalent formulas*

$$\|\pi-\nu\|_{TV}=2\sup_{A\subset\mathbb{N}}|\pi(A)-\nu(A)|=\sum_{i\in\mathbb{N}}|\pi(i)-\nu(i)|=\sup_{f:\mathbb{N}\to\mathbb{R},\ \|f\|_\infty\le1}|\pi(f)-\nu(f)|, \quad (2.25)$$

where $\pi(f)=\sum_{i\in\mathbb{N}}f(i)\pi(i)$.

For the next result, we use standard semigroup notations: for all $x\in\mathbb{N}$ and $f:\mathbb{N}\to\mathbb{R}$ bounded, we denote $\delta_xP_tf=P_tf(x)=\mathbb{E}_xf(X_t)$ and for all probability measure π on $\mathbb{N}$, $\pi P_tf=\sum_{i\ge0}\pi(i)\delta_iP_tf=\mathbb{E}_\pi f(X_t)$. Hence δ_xP_t is the distribution of the random variable X_t given $X_0=x$.

Corollary 2.16. *For all $x,y\in\mathbb{N}$ and all $t\ge0$, there exists a probability measure ν on $\mathbb{N}^*$ such that*

$$\delta_xP_t(\mathbb{1}_A)\ge\nu(A)\mathbb{P}(\tau\le t)\ \text{and}\ \delta_yP_t(\mathbb{1}_A)\ge\nu(A)\mathbb{P}(\tau\le t),\quad\forall A\subset\mathbb{N}^*, \quad (2.26)$$

where τ is constructed as above with $\bar{X}_0=x$ and $\bar{Y}_0=y$. In particular,

$$\|\delta_xP_t-\delta_yP_t\|_{TV}\le2\mathbb{P}(t<\tau),\quad\forall t\ge0. \quad (2.27)$$

Proof. From the above construction we see that, for all $A\subset\mathbb{N}$,

$$\mathbb{P}(X_t\in A)=\mathbb{P}(\widetilde{X}_t\in A)=\mathbb{P}(Y_t\in A,\tau\le t)+\mathbb{P}(X_t\in A,t<\tau)\ge\mathbb{P}(Y_t\in A,\tau\le t).$$

Since the same inequality is trivial for $(Y_t)_{t\ge0}$, we have proved (2.26) with

$$\nu(A)=\frac{\mathbb{P}(Y_t\in A,\tau\le t)}{\mathbb{P}(\tau\le t)}.$$

We also deduce that

$$|\mathbb{P}(X_t\in A)-\mathbb{P}(Y_t\in A)|=|\mathbb{P}(X_t\in A,t<\tau)-\mathbb{P}(Y_t\in A,t<\tau)|\le\mathbb{P}(t<\tau).$$

Taking the supremum over $A\subset\mathbb{N}$ entails (2.27). □

Since birth and death processes only make jumps of size ± 1, it is also clear that $X_t - Y_t$ has constant sign before time τ. The next result is then clear.

Corollary 2.17. *If $\bar{X}_0 \leq \bar{Y}_0$ a.s., then $X_t \leq Y_t$ a.s. for all $t \geq 0$. In particular, $x \mapsto \mathbb{P}_x(X_t \geq z)$ is non-decreasing for all $z \in \mathbb{N}$, and if T_0 denotes the first hitting time of 0 by the birth and death process, $x \mapsto \mathbb{P}_x(t < T_0)$ is non-decreasing for all $t \geq 0$.*

2.7.2 First properties of quasi-stationary distributions

We will assume in all this section that the birth and death process $(X_t)_{t\geq 0}$ gets almost surely extinct after a finite time T_0, i.e. that its infinitesimal generator satisfies the condition of Theorem 2.8(i). In this case the stationary behavior of the process is trivial and δ_0 is the only stationary distribution. However, it may happen that extinction only occurs after a long time and it is then interesting to characterize a *stationary behavior of the process before extinction.* This can be done using the notions of quasi-stationary distribution and quasi-limiting distribution as defined below.

Definition 2.18. (a) *A probability measure ν on $\mathbb{N}^*$ is a quasi-stationary distribution for the birth and death process $(X_t)_{t\geq 0}$ if, for all $t \geq 0$ and all $A \subset \mathbb{N}$,*

$$\mathbb{P}_\nu(X_t \in A \mid t < T_0) = \nu(A).$$

(b) *A probability measure ν on $\mathbb{N}^*$ is a quasi-limiting distribution for the birth and death process $(X_t)_{t\geq 0}$ if there exists a probability measure π on $\mathbb{N}$ such that*

$$\lim_{t\to+\infty} \mathbb{P}_\pi(X_t \in \cdot \mid t < T_0) = \nu.$$

where the convergence holds in total variation.

The next results shows that the two notions are the same.

Proposition 2.19. *ν is a quasi-stationary distribution if and only if ν is a quasi-limiting distribution.*

Proof. A quasi-stationary distribution ν is a quasi-limiting distribution since Definition 2.18(b) holds for $\pi = \nu$.

Assume that ν is a quasi-limiting distribution and take π as in Definition 2.18(b). Then, for all $f : \mathbb{N} \to \mathbb{R}$ bounded,

$$\nu(f) = \sum_{i\geq 0} \nu(i) f(i) = \lim_{t\to+\infty} \mathbb{E}_\pi(f(X_t) \mid t < T_0).$$

Now, for fixed $s > 0$, and $A \subset \mathbb{N}^*$,

$$\mathbb{P}_\nu(X_s \in A \mid s < T_0) = \frac{\mathbb{P}_\nu(X_s \in A,\, s < T_0)}{\mathbb{P}_\nu(s < T_0)} = \frac{\mathbb{P}_\nu(X_s \in A)}{\mathbb{P}_\nu(s < T_0)} = \frac{\nu(f_s)}{\nu(g_s)},$$

where $f_s(x) = \mathbb{P}_x(X_s \in A)$ and $g_s(x) = \mathbb{P}_x(s < T_0) > 0$ for all $x \in \mathbb{N}$. Then

$$\mathbb{P}_\nu(X_s \in A \mid s < T_0) = \lim_{t\to\infty} \frac{\mathbb{E}_\pi(f_s(X_t) \mid t < T_0)}{\mathbb{E}_\pi(g_s(X_t) \mid t < T_0)} = \lim_{t\to\infty} \frac{\mathbb{E}_\pi(f_s(X_t))}{\mathbb{E}_\pi(g_s(X_t))} = \lim_{t\to\infty} \frac{\mathbb{P}_\pi(X_{t+s} \in A)}{\mathbb{P}_\pi(t+s < T_0)},$$

where we used Markov's property in the last equality. Hence

$$\mathbb{P}_\nu(X_s \in A \mid s < T_0) = \lim_{t\to\infty} \mathbb{P}_\pi(X_{t+s} \in A \mid t+s < T_0) = \nu(A).$$

Therefore ν is a quasi-stationary distribution. $\square$

A birth and death process started from its quasi-stationary distribution has remarkable properties.

Proposition 2.20. *Let ν be a quasi-stationary distribution for the birth and death process $(X_t)_{t\geq 0}$. Then*

(i) *there exists $\theta > 0$ such that*

$$\mathbb{P}_\nu(t < T_0) = e^{-\theta t}, \quad \forall t \geq 0,$$

i.e. T_0 has exponential distribution with parameter θ under $\mathbb{P}_\nu$.

(ii) *under $\mathbb{P}_\nu$, X_{T_0} is independent of T_0.*

Note that (i) implies that $\mathbb{E}_\nu(e^{\alpha T_0}) < \infty$ for $0 < \alpha < \theta$ and hence that $\mathbb{E}_x(e^{\alpha T_0}) < \infty$ for some $x \in \mathbb{N}^*$. Proposition 2.13 then suggests that there may exist a quasi-stationary distribution when the birth and death process comes down from infinity. We will see that this condition is actually necessary and sufficient for the existence and uniqueness of a quasi-stationary distribution.

Proof. To prove (i), we use Markov's property:

$$\begin{aligned}
\mathbb{P}_\nu(t + s < T_0) &= \mathbb{E}_\nu[\mathbb{1}_{t<T_0}\mathbb{P}_{X_t}(s < T_0)] \\
&= \mathbb{E}_\nu[\mathbb{P}_{X_t}(s < T_0) \mid t < T_0]\mathbb{P}_\nu(t < T_0) \\
&= \mathbb{P}_\nu(s < T_0)\mathbb{P}_\nu(t < T_0).
\end{aligned}$$

This is the standard property of lack of memory characterizing exponential distributions with some parameter $\theta \in [0, +\infty]$. The case $\theta = 0$ corresponds to $T_0 = +\infty$ a.s., which is excluded by assumption, and the case $\theta = +\infty$ corresponds to $T_0 = 0$ a.s., which never holds for birth and death processes.

To prove (ii), we use a similar computation: given $f : \mathbb{N} \to \mathbb{R}$ bounded and $t \geq 0$,

$$\begin{aligned}
\mathbb{E}_\nu[f(X_{T_0})\mathbb{1}_{t<T-0}] &= \mathbb{E}_\nu[\mathbb{1}_{t<T_0}\mathbb{E}_{X_t}(f(X_{T_0}))] \\
&= \mathbb{E}_\nu[\mathbb{E}_{X_t}(f(X_{T_0})) \mid t < T_0]\mathbb{P}_\nu(t < T_0) \\
&= \mathbb{E}_{X_t}(f(X_{T_0}))\mathbb{P}_\nu(t < T_0).
\end{aligned}$$

Hence X_{T_0} and T_0 are independent under $\mathbb{P}_\nu$. $\square$

To conclude these first properties, we give the characterization of quasi-stationary distributions as eigenfunctions of the adjoint generator.

Proposition 2.21. *Let ν be a probability measure on $\mathbb{N}^*$. Then ν is a quasi-stationary distribution if and only if there exists $\theta > 0$ such that*

$$\lambda_{i-1}\nu(i-1) - (\lambda_i + \mu_i)\nu(i) + \mu_{i+1}\nu(i+1) = -\theta\nu(i), \quad \forall i \geq 1. \tag{2.28}$$

Proof. If ν is a quasi-stationary distribution, we deduce from Proposition 2.20 that, for all $i \in \mathbb{N}^*$,

$$\nu(i) = \frac{\nu P_t \mathbb{1}_i}{\nu P_t \mathbb{1}_{N^*}} = e^{\theta t}\nu P_t \mathbb{1}_i.$$

Now, Kolmogorov's forward equation entails that

$$\left|\frac{\partial P_t \mathbb{1}_i}{\partial t}(x)\right| = |P_t(Q\mathbb{1}_i)(x)| \leq \|Q\mathbb{1}_i\|_\infty < \infty.$$

Hence one can differentiate $\nu P_t \mathbb{1}_i = \sum_{j\geq 1} \nu(j) P_t \mathbb{1}_i(j)$ under the sum, which imples that $\nu(Q\mathbb{1}_i) = -\theta\nu(i)$. Hence (2.28) is proved.

Conversely, assume that (2.28) holds true for some $\theta > 0$. Since $f = \mathbb{1}_j$ belongs to the domain of the infinitesimal generator Q, for all $x \in \mathbb{N}^*$,

$$\frac{\partial P_t f}{\partial t}(x) = P_t Q f(x) = Q P_t f(x).$$

As above, this is a bounded function of x and we can differentiate $\nu(P_t f)$ as

$$\frac{d\nu(P_t f)}{dt} = \nu(L P_t f) = \sum_{i\geq 1} P_t f(i)[\lambda_{i-1}\nu(i-1) - (\lambda_i + \mu_i)\nu(i) + \mu_{i+1}\nu(i+1)] = -\theta\nu(P_t f).$$

Solving this ODE gives

$$\nu P_t \mathbb{1}_j = e^{-\theta t}\nu(j).$$

By monotone convergence, we deduce that, for all $A \subset \mathbb{N}^*$, $\nu P_t \mathbb{1}_A = e^{-\theta t}\nu(A)$ and, in particular, $\nu P_t \mathbb{1}_{\mathbb{N}^*} = e^{-\theta t}$. This implies that ν is a quasi-stationary distribution. □

2.7.3 Exponential convergence in total variation to the quasi-stationary distribution

The goal of this section is to prove the next result.

Theorem 2.22 (Martinez, San Martin, Villemonais [35])**.** *If the birth and death process $(X_t)_{t\geq 0}$ comes down from infinity, then $(X_t)_{t\geq 0}$ admits a unique quasi-stationary distribution ν and there exist constants $C, \gamma > 0$ such that, for all probability measure π on $\mathbb{N}^*$,*

$$\|\mathbb{P}_\pi(X_t \in \cdot \mid t < T_0) - \nu\|_{TV} \leq C e^{-\gamma t}, \quad \forall t \geq 0. \tag{2.29}$$

The proof given here is adapted from [12]. We start with some Lemmas.

Lemma 2.23. *Let $(X_t)_{t\geq 0}$ be a birth and death process coming down from infinity. Then, there exists a constant $c > 0$ such that*

$$\inf_{x\in\mathbb{N}^*} \mathbb{P}_x(t < T_0) \geq c \sup_{x\in\mathbb{N}^*} \mathbb{P}_x(t < T_0), \quad \forall t \geq 0.$$

Proof. Let us recall from Corollary 2.17 that $\inf_{x\in\mathbb{N}^*} \mathbb{P}_x(t < T_0) = \mathbb{P}_1(t < T_0)$ and from the property of coming down from infinity that $\sup_{x\in\mathbb{N}^*} \mathbb{P}_x(t < T_0) = \mathbb{P}_\infty(t < T_0)$.

In view of Proposition 2.13(v), setting $a = 1 + \mu_1 + \lambda_1$, we can find $z \geq 1$ such that, defining the finite set $K = \{1, 2, \dots, z\}$ and $T_K = \inf\{t \geq 0,\ X_t \in K\}$,

$$A := \sup_{x\geq 1} \mathbb{E}_x(e^{a(T_K \wedge T_0)}) < \infty. \tag{2.30}$$

Let us first observe that for all $y, z \in K$, $\mathbb{P}_y(X_1 = z)\mathbb{P}_z(t < T_0) \leq \mathbb{P}_y(t + 1 < T_0) \leq \mathbb{P}_y(t < T_0)$. Therefore, the constant $C^{-1} := \inf_{y,z\in K} \mathbb{P}_y(X_1 = z) > 0$ satisfies the following inequality:

$$\sup_{x\in K} \mathbb{P}_x(t < T_0) \leq C \inf_{x\in K} \mathbb{P}_x(t < T_0), \quad \forall t \geq 0. \tag{2.31}$$

Moreover, since a is larger than $\lambda_1+\mu_1$, the jump rate of X from 1,

$$e^{-as}\mathbb{P}_1(t-s<T_0)\leq \mathbb{P}_1(X_s=1)\mathbb{P}_1(t-s<T_0)\leq \mathbb{P}_1(t<T_0).$$

For all $x\geq 1$, we deduce from Chebyshev's inequality and (2.30) that

$$\mathbb{P}_x(t<T_K\wedge T_0)\leq Ae^{-at}.$$

Using the last three inequalities and the strong Markov property, we have

$$\begin{aligned}\mathbb{P}_x(t<T_0)&=\mathbb{P}_x(t<T_K\wedge T_0)+\mathbb{P}_x(T_K\wedge T_0\leq t<T_0)\\&\leq Ae^{-at}+\int_0^t \sup_{y\in K\cup\{0\}}\mathbb{P}_y(t-s<T_0)\mathbb{P}_x(T_K\wedge T_0\in ds)\\&\leq A\mathbb{P}_1(t<T_0)+C\int_0^t \mathbb{P}_1(t-s<T_0)\mathbb{P}_x(T_K\wedge T_0\in ds)\\&\leq A\mathbb{P}_1(t<T_0)+C\,\mathbb{P}_1(t<T_0)\int_0^t e^{as}\,\mathbb{P}_x(T_K\wedge T_0\in ds)\\&\leq A(1+C)\mathbb{P}_1(t<T_0).\end{aligned}$$

Combining this with (2.31) ends the proof of the lemma. □

Lemma 2.24. *Let us define, for all $0\leq s\leq t\leq T$ the linear operator $R^T_{s,t}$ by*

$$\begin{aligned}R^T_{s,t}f(x)&=\mathbb{E}_x(f(X_{t-s})\mid T-s<\tau_\partial)\\&=\mathbb{E}(f(X_t)\mid X_s=x,\ T<\tau_\partial),\end{aligned}$$

by the Markov property. This family of operators forms a time-inhomogeneous semigroup, in the sense that, for all $0\leq u\leq s\leq t\leq T$, all $x\geq 1$ and all $f:\mathbb{N}^*\to\mathbb{R}$ bounded,

$$R^T_{u,s}(R^T_{s,t}f)(x)=R^T_{u,t}f(x).$$

Proof. We have, for all $0\leq u\leq s\leq t\leq T$,

$$R^T_{u,s}(R^T_{s,t}f)(x)=\mathbb{E}_x(\mathbb{E}_{X_{s-u}}(f(X_{t-s})\mid T-s<T_0)\mid T-u<T_0).$$

For any bounded measurable function g, the Markov property implies that

$$\begin{aligned}\mathbb{E}_x\left(g(X_{s-u})\mathbb{1}_{T-u<T_0}\right)&=\mathbb{E}_x\left(g(X_{s-u})\mathbb{P}_{X_{s-u}}(T-u-(s-u)<T_0)\right)\\&=\mathbb{E}_x\left(g(X_{s-u})\mathbb{P}_{X_{s-u}}(T-s<T_0)\right)\end{aligned}$$

Applying this equality to $g:y\mapsto \mathbb{E}_x(f(X_{t-s})\mid T-s<T_0)$, we deduce that

$$\begin{aligned}R^T_{u,s}(R^T_{s,t}f)(x)&=\frac{\mathbb{E}_x(\mathbb{E}_{X_{s-u}}(f(X_{t-s})\mathbb{1}_{T-s<T_0}))}{\mathbb{P}_x(T-u<T_0)}\\&=\frac{\mathbb{E}_x(f(X_{t-s+(s-u)})\mathbb{1}_{T-s+(s-u)<T_0}))}{\mathbb{P}_x(T-u<T_0)}\\&=R^T_{u,t}f(x),\end{aligned}$$

where we have used the Markov property a second time. □

Lemma 2.25. *Assume that there exist constants $C, \gamma > 0$ such that, for all $x, y \in \mathbb{N}^*$,*

$$\|\mathbb{P}_x(X_t \in \cdot \mid t < T_0) - \mathbb{P}_y(X_t \in \cdot \mid t < T_0)\|_{TV} \leq Ce^{-\gamma t}, \quad \forall t \geq 0.$$

Then, for all probability measures π_1, π_2 on $\mathbb{N}^*$,

$$\|\mathbb{P}_{\pi_1}(X_t \in \cdot \mid t < T_0) - \mathbb{P}_{\pi_2}(X_t \in \cdot \mid t < T_0)\|_{TV} \leq Ce^{-\gamma t}, \quad \forall t \geq 0.$$

Proof. Let π_1 be a probability measure on E and $x \in E$. We have

$$\begin{aligned}
&\|\mathbb{P}_{\pi_1}(X_t \in \cdot \mid t < T_0) - \mathbb{P}_x(X_t \in \cdot \mid t < T_0)\|_{TV} \\
&= \frac{1}{\mathbb{P}_{\pi_1}(t < T_0)} \|\mathbb{P}_{\pi_1}(X_t \in \cdot) - \mathbb{P}_{\pi_1}(t < T_0)\mathbb{P}_x(X_t \in \cdot \mid t < T_0)\|_{TV} \\
&\leq \frac{1}{\mathbb{P}_{\pi_1}(t < T_0)} \sum_{y \geq 1} \|\mathbb{P}_y(X_t \in \cdot) - \mathbb{P}_y(t < T_0)\mathbb{P}_x(X_t \in \cdot \mid t < T_0)\|_{TV}\pi_1(y) \\
&\leq \frac{1}{\mathbb{P}_{\pi_1}(t < T_0)} \sum_{y \geq 1} \mathbb{P}_y(t < T_0)\|\mathbb{P}_y(X_t \in \cdot \mid t < T_0) - \mathbb{P}_x(X_t \in \cdot \mid t < T_0)\|_{TV}\pi_1(y) \\
&\leq \frac{1}{\mathbb{P}_{\pi_1}(t < T_0)} \int_{y \in E} \mathbb{P}_y(t < T_0)Ce^{-\gamma t} d\pi_1(y) \leq Ce^{-\gamma t}.
\end{aligned}$$

The same computation, replacing δ_x by any probability measure, concludes the proof of Lemma 2.25. □

Proof of Theorem 2.22. We use the coupling technique of Section 2.7.1: we construct on the same probability space two independent birth and death processes with generator Q, $(X_t^1)_{t \geq 0}$ and $(X_t^\infty)_{t \geq 0}$, one starting from $X_0^1 = 1$ and the other one from $X_0^\infty = \infty$, and we call $\tau^{1,\infty}$ their coupling time. Since X^∞ gets a.s. extinct in finite time and the two processes are independent, there exist $t_0 > 0$ such that

$$c_0 := \mathbb{P}(\tau^{1,\infty} < t_0,\, X_{t_0}^1 > 0) > 0.$$

As in Corollary 2.16, we deduce that there exists a probability measure $\nu^{1,\infty}$ such that

$$\mathbb{P}_1(X_{t_0} \in A) \geq c_0 \nu^{1,\infty}(A), \quad \mathbb{P}_\infty(X_{t_0} \in A) \geq c_0 \nu^{1,\infty}(A), \quad \forall A \subset \mathbb{N}^*.$$

Since we can similarly couple on the same probability the three processes X^1, X^∞ and X^x (a birth and death process started from $X_0 = x \in \mathbb{N}^*$) and since the coupling times $\tau^{1,x}$ between X^1 and X^∞ and $\tau^{x,\infty}$ between X^x and X^∞ are clearly smaller than $\tau^{1,\infty}$, we also deduce that

$$\mathbb{P}_x(X_{t_0} \in A) \geq c_0 \nu^{1,\infty}(A), \quad \forall A \subset \mathbb{N}^*,\ \forall x \in \mathbb{N}^*.$$

Then, for all $x \in \mathbb{N}^*$ and $t \geq t_0$,

$$\mathbb{P}_x(X_{t_0} \in A,\, t < T_0) = \mathbb{E}_x[\mathbb{1}_{X_{t_0} \in A} \mathbb{P}_{X_{t_0}}(t - t_0 < T_0)] \geq c_0 \nu^{1,\infty}[\mathbb{1}_A \mathbb{P}_\cdot(t - t_0 < T_0)].$$

Dividing by $\mathbb{P}_x(t < T_0)$ and using the inequality $\mathbb{P}_x(t < T_0) \leq \mathbb{P}_x(t - t_0 < T_0)$, we obtain

$$\mathbb{P}_x(X_{t_0} \in A \mid t < T_0) \geq c_0 \frac{\nu^{1,\infty}[\mathbb{1}_A \mathbb{P}_\cdot(t - t_0 < T_0)]}{\sup_{y \geq 1} \mathbb{P}_y(t - t_0 < T_0)}.$$

Now Lemma 2.23 entails that the measure $A \mapsto \frac{\nu^{1,\infty}[\mathbb{1}_A \mathbb{P}_\cdot(t-t_0<T_0)]}{\sup_{y\geq 1}\mathbb{P}_y(t-t_0<T_0)}$ has a mass greater than c_0. Since it does not depend on x, we have proved that there exists a probability measure ν_t on $\mathbb{N}^*$ such that, for all $t \geq t_0$,

$$\mathbb{P}_x(X_{t_0} \in A \mid t < T_0) \geq cc_0 \nu_t(A), \quad \forall x \geq 1,\ \forall A \subset \mathbb{N}^*.$$

In other words, using the notations of Lemma 2.24, we have proved that, for all $0 \leq s \leq s + t_0 \leq T$,

$$\delta_x R^T_{s,s+t_0} \geq cc_0 \nu_{T-s}, \quad \forall x \geq 1.$$

Therefore, for all $x \neq y$ in $\mathbb{N}^*$,

$$\|\delta_x R^T_{s,s+t_0} - \delta_y R^T_{s,s+t_0}\|_{TV} \leq 2(1 - cc_0).$$

Given two mutually singular probability measures π_1, π_2 on E, we have

$$\begin{aligned}\left\|\pi_1 R^T_{s,s+t_0} - \pi_2 R^T_{s,s+t_0}\right\|_{TV} &\leq \sum_{x\geq 1,\, y \geq 1} \left\|\delta_x R^T_{s,s+t_0} - \delta_y R^T_{s,s+t_0}\right\|_{TV} \pi_1(x)\pi_2(y) \\ &\leq 2(1-cc_0) = (1-cc_0)\|\pi_1 - \pi_2\|_{TV}.\end{aligned}$$

This inequality extends to probability measures which are non-singular since one can apply the last inequality to the mutually singular probability measures $\bar\pi_+ := \frac{(\pi_1-\pi_2)_+}{(\pi_1-\pi_2)_+(\mathbb{N}^*)}$ and $\bar\pi_- := \frac{(\pi_1-\pi_2)_-}{(\pi_1-\pi_2)_-(N^*)}$. Then

$$\left\|\bar\pi_+ R^T_{s,s+t_0} - \bar\pi_- R^T_{s,s+t_0}\right\|_{TV} \leq 2(1-cc_0).$$

Since $\pi_1(\mathbb{N}^*) = \pi_2(\mathbb{N}^*) = 1$, we have $(\pi_1 - \pi_2)_+(\mathbb{N}^*) = (\pi_1 - \pi_2)_-(\mathbb{N}^*)$. So multiplying the last inequality by $(\pi_1 - \pi_2)_+(\mathbb{N}^*)$, we deduce that

$$\begin{aligned}&\|(\pi_1 - \pi_2)_+ R^T_{s,s+t_0} - (\pi_1 - \pi_2)_- R^T_{s,s+t_0}\|_{TV} \\ &\qquad\qquad \leq 2(1-cc_0)(\pi_1-\pi_2)_+(\mathbb{N}^*) = (1-cc_0)\|\pi_1-\pi_2\|_{TV}.\end{aligned}$$

Since $(\pi_1 - \pi_2)_+ - (\pi_1 - \pi_2)_- = \pi_1 - \pi_2$, we obtain

$$\|\pi_1 R^T_{s,s+t_0} - \pi_2 R^T_{s,s+t_0}\|_{TV} \leq (1 - c_1 c_2)\|\pi_1 - \pi_2\|_{TV}.$$

We can now use the semigroup property of Lemma 2.24: for any $x, y \in E$,

$$\begin{aligned}\|\delta_x R^T_{0,T} - \delta_y R^T_{0,T}\|_{TV} &= \|\delta_x R^T_{0,T-t_0} R^T_{T-t_0,T} - \delta_y R^T_{0,T-t_0} R^T_{T-t_0,T}\|_{TV} \\ &\leq (1-cc_0)\,\|\delta_x R^T_{0,T-t_0} - \delta_y R^T_{0,T-t_0}\|_{TV} \leq \dots \\ &\leq (1-cc_0)^{\lfloor T/t_0 \rfloor}\,\|\delta_x R^T_{0,T-t_0\lfloor T/t_0\rfloor} - \delta_y R^T_{0,T-t_0\lfloor T/t_0\rfloor}\|_{TV} \\ &\leq 2\,(1-cc_0)^{\lfloor T/t_0\rfloor}.\end{aligned}$$

Therefore, by Lemma 2.25, we have proved that there exist constants $C, \gamma > 0$ such that, for all probability measures π_1, π_2 on $\mathbb{N}^*$,

$$\|\mathbb{P}_{\pi_1}(X_t \in \cdot \mid t < T_0) - \mathbb{P}_{\pi_2}(X_t \in \cdot \mid t < T_0)\|_{TV} \leq Ce^{-\gamma t}, \quad \forall t \geq 0. \tag{2.32}$$

Given two quasi-stationary distributions ν and ν', the last inequality applied to $\pi_1 = \nu$ and $\pi_2 = \nu'$ implies the uniqueness of the quasi-stationary distribution. Let us now prove the existence of a quasi-stationary distribution. By Proposition 2.19, this is equivalent to prove the existence of a quasi-limiting distribution for X, so we only need to prove that $\mathbb{P}_x(X_t \in \cdot \mid t < T_0)$ converges when t goes to infinity, for some $x \geq 1$. We have, for all $s, t \geq 0$ and $x \geq 1$,

$$\begin{aligned}\mathbb{P}_x\left(X_{t+s} \in \cdot \mid t+s < T_0\right) &= \frac{\delta_x P_{t+s}}{\delta_x P_{t+s}\mathbb{1}_{\mathbb{N}^*}} = \frac{\delta_x P_t P_s}{\delta_x P_t P_s \mathbb{1}_{\mathbb{N}^*}} = \frac{\delta_x R^s_{0,s} P_t}{\delta_x R^s_{0,s} P_t \mathbb{1}_{\mathbb{N}^*}} \\ &= \mathbb{P}_{\delta_x R^s_{0,s}}\left(X_t \in \cdot \mid t < T_0\right).\end{aligned}$$

Hence,

$$\begin{aligned}&\|\mathbb{P}_x(X_t \in \cdot \mid t < T_0) - \mathbb{P}_x(X_{t+s} \in \cdot \mid t+s < T_0)\|_{TV} \\ &\qquad = \|\mathbb{P}_x(X_t \in \cdot \mid t < T_0) - \mathbb{P}_{\delta_x R^s_{0,s}}(X_t \in \cdot \mid t < T_0)\|_{TV} \\ &\qquad \leq 2\left(1 - cc_0\right)^{\lfloor t/t_0 \rfloor} \xrightarrow[s,t\to+\infty]{} 0.\end{aligned}$$

Therefore, the sequence $(\mathbb{P}_x(X_t \in \cdot \mid t < T_0))_{t\geq 0}$ is a Cauchy sequence for the total variation norm, hence converges when t goes to infinity to some probability measure ν on E, which is a quasi-limiting distribution, hence a quasi-stationary distribution.

Finally (2.29) follows from (2.32) with $\pi_1 = \pi$ and $\pi_2 = \nu$. □

Theorem 2.22 has the following converse.

Theorem 2.26 (van Doorn [40])**.** *A non-explosive birth and death process with almost sure extinction $(X_t)_{t\geq 0}$ admits a unique quasi-stationary distribution if and only if it comes down from infinity.*

We do not give the proof of this result here. Instead, we prove a weaker converse statement.

Theorem 2.27. *A non-explosive birth and death process with almost sure extinction $(X_t)_{t\geq 0}$ admits a unique quasi-stationary distribution ν such that, for all probability measure π on $\mathbb{N}^*$,*

$$\|\mathbb{P}_\pi(X_t \in \cdot \mid t < T_0) - \nu\|_{TV} \leq Ce^{-\gamma t}, \quad \forall t \geq 0 \tag{2.33}$$

for some constants $C, \gamma > 0$, if and only if it comes down from infinity.

Proof. One implication is given by Theorem 2.22, so let us assume (2.33) and that $(X_t)_{t\geq 0}$ does not come back from infinity, and try to reach a contradiction. By definition of the property of coming down from infinity, we have that, for all $t > 0$ and $y \geq 1$, $\mathbb{P}_\infty(T_y > t) = 1$. Let us choose $t > 0$ such that $Ce^{-\gamma t} < 1/3$ and y_0 such that $\nu(\{1, \dots, y_0\}) \geq 2/3$. It then follows from (2.33) that, for all $y \geq 1$,

$$\mathbb{P}_y(X_t \leq y_0) \geq \mathbb{P}_y(X_t \leq y_0 \mid t < T_0) \geq 1/3.$$

This is impossible since $\lim_{y\to+\infty} \mathbb{P}_y(X_t \leq y_0) = \mathbb{P}_\infty(X_t \leq y_0) = 0$. □

3 Scaling Limits for Birth and Death Processes

If the population is large, so many birth and death events occur that the dynamics becomes difficult to describe individual per individual. Living systems need resources in order to survive and reproduce and the biomass per capita depends on the order of magnitude of these resources. We introduce a parameter $K \in \mathbb{N}^* = \{1, 2, \ldots\}$ scaling either the size of the population or the total amount of resources. We assume that the individuals are weighted by $\frac{1}{K}$.

In this section, we show that depending on the scaling relations between the population size and the demographic parameters, the population size process will be approximated either by a deterministic process or by a stochastic process. These approximations will lead to different long time behaviors.

In the rest of this section, we consider a sequence of birth and death processes Z^K parameterized by K, where the birth and death rates for the population state $n \in \mathbb{N}$ are given by $\lambda_K(n)$ and $\mu_K(n)$. Since the individuals are weighted by $\frac{1}{K}$, the population dynamics is modeled by the process $(X^K_t, t \geq 0) \in \mathbb{D}(\mathbb{R}_+, \mathbb{R}_+)$ with jump amplitudes $\pm\frac{1}{K}$ and defined for $t \geq 0$ by

$$X^K_t = \frac{Z^K_t}{K}. \tag{3.1}$$

This process is a Markov process with generator

$$L_K\phi(x) = \lambda_K(Kx)\big(\phi(x + \frac{1}{K}) - \phi(x)\big) + \mu_K(Kx)\big(\phi(x - \frac{1}{K}) - \phi(x)\big). \tag{3.2}$$

Therefore, adapting Proposition 2.5 and Theorem 2.6, one can easily show that if $\lambda_K(n) \leq \bar{\lambda} n$ (uniformly in K) and if

$$\sup_K \mathbb{E}((X^K_0)^3) < +\infty, \tag{3.3}$$

then

$$\sup_K \mathbb{E}(\sup_{t\leq T}(X^K_t)^3) < +\infty, \tag{3.4}$$

and for any $K \in \mathbb{N}^*$, the process

$$M^K_t = X^K_t - X^K_0 - \frac{1}{K}\int_0^t (\lambda_K(Z^K_s) - \mu_K(Z^K_s))ds \tag{3.5}$$

is a square integrable martingale with quadratic variation

$$\langle M^K\rangle_t = \frac{1}{K^2}\int_0^t (\lambda_K(Z^K_s) + \mu_K(Z^K_s))ds. \tag{3.6}$$

3.1 Deterministic approximation - Malthusian and logistic Equations

Let us now assume that the birth and death rates satisfy the following assumption:

$$\begin{aligned}
&\lambda_K(n) = n\lambda\left(\frac{n}{K}\right);\ \mu_K(n) = n\mu\left(\frac{n}{K}\right),\ \text{where the functions}\\
&\lambda \text{ and } \mu \text{ are non negative and Lipschitz continuous on } \mathbb{R}_+,\\
&\lambda(x) \leq \bar{\lambda} \quad ; \quad \mu(x) \leq \bar{\mu}(1 + x).
\end{aligned} \tag{3.7}$$

We will focus on two particular cases:

The linear case: $\lambda_K(n) = n\lambda$ and $\mu_K(n) = n\mu$, with $\lambda, \mu > 0$.

The logistic case: $\lambda_K(n) = n\lambda$ and $\mu_K(n) = n(\mu + \frac{c}{K}n)$ with $\lambda, \mu, c > 0$.

By (3.3), the population size is of the order of magnitude of K and the biomass per capita is of order $\frac{1}{K}$. This explains that the competition pressure from one individual to another one in the logistic case is proportional to $\frac{1}{K}$.

We are interested in the limiting behavior of the process $(X^K_t, t \geq 0)$ when $K \to \infty$. We are actually going to prove two versions of the next result.

Theorem 3.1 (Ethier and Kurtz [17]). *Let us assume* (3.7)*, that* $\lambda(x) \leq \mu(x)$ *for all* x *large enough and that the sequence* $(X^K_0)_K$ *converges a.s. to a real number* x_0*. Then for any* $T > 0$*, the sequence of processes* $(X^K_t, t \in [0, T])$ *constructed as in Proposition 2.7 from two given Poisson processes* $(P_1(t))_{t\geq 0}$ *and* $(P_2(t))_{t\geq 0}$*, converges in probability for the* $L^\infty([0, T])$ *norm to the continuous deterministic function* $(x(t), t \in [0, T])$ *solution of the ordinary differential equation*

$$x'(t) = x(t)(\lambda(x(t)) - \mu(x(t))) \; ; x(0) = x_0. \tag{3.8}$$

In the linear case, the limiting equation is the Malthusian equation

$$x'(t) = x(t)(\lambda - \mu).$$

In the logistic case, one obtains the logistic equation

$$x'(t) = x(t)(\lambda - \mu - c\,x(t)). \tag{3.9}$$

These two equations have different long time behaviors. In the Malthusian case, depending on the sign of $\lambda - \mu$, the solution of the equation tends to $+\infty$ or to 0 as time goes to infinity, modeling the explosion or extinction of the population. In the logistic case and if the growth rate $\lambda - \mu$ is positive, the solution converges to the carrying capacity $\dfrac{\lambda - \mu}{c} > 0$. The competition between individuals yields a regulation of the population size.

Proof. We first make the proof assuming that $x \mapsto x\lambda(x)$ and $x \mapsto x\mu(x)$ are bounded and globally Lipschitz functions on $\mathbb{R}_+$. The extension to the general case will be done afterwards. Using the construction of Proposition 2.7, for all $K \geq 1$,

$$X^K_t = X^K_0 + \frac{1}{K} P_1\left(K\int_0^t X^K_s \lambda(X^K_s)ds\right) - \frac{1}{K} P_2\left(K\int_0^t X^K_s \mu(X^K_s)ds\right), \quad \forall t \geq 0.$$

Introducing $\widetilde{P}_1(t) = P_1(t) - t$ and $\widetilde{P}_2(t) = P_2(t) - t$ the compensated Poisson processes, we obtain

$$X^K_t = X^K_0 + \int_0^t F(X^K_s)ds + \frac{1}{K}\widetilde{P}_1\left(K\int_0^t X^K_s \lambda(X^K_s)ds\right) - \frac{1}{K}\widetilde{P}_2\left(K\int_0^t X^K_s \mu(X^K_s)ds\right),$$

where $F(x) = x(\lambda(x) - \mu(x))$. Therefore, introducing M such that $x \mapsto x\lambda(x)$ and $x \mapsto x\mu(x)$ are bounded by M and M-Lipschitz, for all $t \leq T$,

$$
\begin{aligned}
|X_t^K - x(t)| &\leq |X_0^K - x(0)| + \int_0^t |F(X_s^K) - F(x(s))|ds + \frac{1}{K} \sup_{0\leq s\leq KMT} \left\{|\widetilde{P}_1(s)| + |\widetilde{P}_2(s)|\right\} \\
&\leq |X_0^K - x(0)| + M \int_0^t |X_s^K - x(s)|ds + \frac{1}{K} \sup_{0\leq s\leq KMT} \left\{|\widetilde{P}_1(s)| + |\widetilde{P}_2(s)|\right\}.
\end{aligned}
$$

Therefore, the result follows from Gronwall's lemma and the next lemma.

Lemma 3.2. *For any Poisson process $(P(t))_{t\geq 0}$ and for all $\alpha > 1/2$,*

$$\frac{1}{n^\alpha} \sup_{t\in[0,n]} |P(t) - t| \xrightarrow[n\to+\infty]{} 0 \quad a.s.$$

This result will be proved after the current proof.

We now consider the general case, where the functions $x \mapsto x\lambda(x)$ and $x \mapsto x\mu(x)$ are only locally bounded and locally Lipschitz on $\mathbb{R}_+$. Since we assume that $\lambda(x) \leq \mu(x)$ for all $x \geq x_1$, the solution $x(t)$ to (3.10) remains smaller that $\max\{x(0), x_1\}$. So let us define $\bar{x} = 1 + \max\{x(0), x_1\}$ and

$$\bar{\lambda}(x) = \begin{cases} \lambda(x) & \text{if } x \leq \bar{x}, \\ \frac{\bar{x}}{x}\lambda(\bar{x}) & \text{otherwise}, \end{cases}, \quad \bar{\mu}(x) = \begin{cases} \mu(x) & \text{if } x \leq \bar{x}, \\ \frac{\bar{x}}{x}\mu(\bar{x}) & \text{otherwise}. \end{cases}$$

The first part of the proof applies to the processes $(\bar{X}_t^K)_{t\geq 0}$ constructed from the functions $\bar{\lambda}$ and $\bar{\mu}$ and the Poisson processes $(P_1(t))_{t\geq 0}$ and $(P_2(t))_{t\geq 0}$. Hence, for any $T > 0$, $\sup_{t\in[0,T]} |\bar{X}_t^K - \bar{x}(t)| \to 0$ when $K \to +\infty$ and

$$\bar{x}'(t) = \bar{x}(t)(\bar{\lambda}(\bar{x}(t)) - \bar{\mu}(\bar{x}(t))), \quad \bar{x}(0) = x(0).$$

Since $\lambda(x) \leq \mu(x)$ for $x \geq x_1$, we have $\bar{x}(t) = x(t) \leq \bar{x} - 1$ for all $t \geq 0$. In addition, for K large enough so that $\sup_{t\in[0,T]} |\bar{X}_t^K - \bar{x}(t)| \leq 1$, we have $\sup_{t\in[0,T]} \bar{X}_t^K \leq \bar{x}$. Since $\bar{\lambda}(x) = \lambda(x)$ and $\bar{\mu}(x) = \mu(x)$ for all $x \leq \bar{x}$, we deduce that $\bar{X}_t^K = X_t^K$ for all $t \in [0, T]$. Combining these two facts, we deduce that

$$\sup_{t\in[0,T]} |X_t^K - x(t)| \xrightarrow[K\to+\infty]{} 0 \quad a.s.,$$

which ends the proof of Theorem 3.1. □

Proof of Lemma 3.2. Using the Laplace transform of $P(t) - t$ and Chebychev's exponential inequality, we obtain that, for all $\gamma > 0$ and $\varepsilon > 0$,

$$\mathbb{P}(P(t) - t > \varepsilon) \leq e^{-\gamma\varepsilon}\mathbb{E}[\exp(\gamma(P(t) - t))] = \exp[t(e^\gamma - 1 - \gamma) - \gamma\varepsilon].$$

Taking the infimum of the right-hand side with respect to $\gamma > 0$, we obtain

$$\mathbb{P}(P(t) - t > \varepsilon) \leq \frac{e^\varepsilon}{(1 + \varepsilon/t)^{t+\varepsilon}}.$$

Similarly,

$$\mathbb{P}(P(t) - t > -\varepsilon) \leq \frac{e^{-\varepsilon}}{(1 - \varepsilon/t)^{t-\varepsilon}}.$$

Now, we fix $1/2 < \alpha < 1$ and take $\varepsilon = t^\alpha$. We deduce that

$$\mathbb{P}(|P(t) - t| > t^\alpha) \leq \frac{e^{t^\alpha}}{(1 + 1/t^{1-\alpha})^{t+t^\alpha}} + \frac{e^{-t^\alpha}}{(1 - 1/t^{1-\alpha})^{t-t^\alpha}}$$

and one then checks that, when $t \to +\infty$,

$$\mathbb{P}(|P(t) - t| > t^\alpha) \leq 2 \exp\left(-\frac{t^{2\alpha-1}}{2} + O(t^{3\alpha-2}) \right).$$

Since the right-hand side is sommable w.r.t. $t \in \mathbb{N}^*$, Borel-Cantelli Lemma implies that

$$\sup_{n \in \mathbb{N}^*} \frac{|P(n) - n|}{n^\alpha} < \infty \quad \text{a.s.}$$

Since $P(t)$ is non-decreasing, $P(\lfloor t \rfloor) - \lfloor t \rfloor - 1 \leq P(t) - t \leq P(\lceil t \rceil) - \lceil t \rceil + 1$ for all $t \geq 1$. Hence

$$\sup_{t \in \mathbb{R}_+} \frac{|P(t) - t|}{(t \vee 1)^\alpha} < \infty \quad \text{a.s.}$$

Therefore, for all $\eta > 0$,

$$\frac{1}{n^{\alpha+\eta}} \sup_{t \in [0,n]} |P(t) - t| \leq \frac{1}{n^\eta} \sup_{t \in [0,n]} \frac{|P(t) - t|}{(t \vee 1)^\alpha} \xrightarrow[n \to +\infty]{} 0 \quad \text{a.s.} \quad \square$$

Theorem 3.1 gives strong convergence (almost sure convergence for the L^∞ norm) of the birth and death process to a deterministic limit. We also state and prove another version giving convergence in law because it will allow us to explain in detail a general method applying to much more general situations to which the method of Theorem 3.1 does not apply (see for example Section 5).

Theorem 3.3. *Let us assume* (3.7), (3.3) *and that the sequence* $(X_0^K)_K$ *converges in law (and in probability) to a real number* x_0. *Then for any* $T > 0$, *the sequence of processes* $(X_t^K, t \in [0,T])$ *converges in law (and hence in probability), in* $\mathbb{D}([0,T], \mathbb{R}_+)$, *to the continuous deterministic function* $(x(t), t \in [0,T])$ *solution of the ordinary differential equation*

$$x'(t) = x(t)(\lambda(x(t)) - \mu(x(t))) \; ; x(0) = x_0. \tag{3.10}$$

Proof. The proof is based on a compactness-uniqueness argument. More precisely, the scheme of the proof is the following:

1) Uniqueness of the limit.

2) Uniform estimates on the moments (which are given by Proposition 2.5).

3) Tightness of the sequence of laws of $(X_t^K, t \in [0,T])$ in the Skorohod space. We will use the Aldous and Rebolledo criterion.

4) Identification of the limit.

Thanks to Assumption (3.7), the uniqueness of the solution of equation (3.10) is obvious. We also have (3.4). Therefore it remains to prove the tightness of the sequence of laws and to identify the limit. Recall (see for example [17] or [25]) that since the processes

$(X^K_t = X^K_0 + M^K_t + A^K_t)_t$ are semimartingales, tightness will be proved as soon as we have

(i) The sequence of laws of $(\sup_{t\leq T}|X^K_t|)$ is tight,

(ii) The finite variation processes $\langle M^K\rangle$ and A^K satisfy the Aldous conditions.

Let us recall the Aldous condition (see [1]): let $(Y^K)_K$ be a sequence of $\mathcal{F}_t$-adapted processes and τ the set of stopping times for the filtration $(\mathcal{F}_t)_t$. The Aldous condition can be written: $\forall\varepsilon>0$, $\forall\eta>0$, $\exists\delta>0$, K_0 such that

$$\sup_{K\geq K_0}\ \sup_{S,S'\in\tau; S\leq S'\leq (S+\delta)\wedge T} \mathbb{P}(|Y^K_{S'}-Y^K_S|>\varepsilon)\leq\eta.$$

Let us show this property for the sequence $(A^K)_K$. We have

$$\begin{aligned}
\mathbb{E}(|A^K_{S'}-A^K_S|) &\leq \mathbb{E}\left(\int_S^{S'} X^K_s|\lambda(X^K_s)-\mu(X^K_s)|ds\right)\\
&\leq C\mathbb{E}\left(\int_S^{S'}(1+(X^K_s)^2)ds\right) \quad \text{by } (3.7)\\
&\leq C\,\delta\,\mathbb{E}\left(\sup_{s\leq T}(1+(X^K_s)^2)\right)
\end{aligned}$$

which tends to 0 uniformly in K as δ tends to 0. We use a similar argument for $(\langle M^K\rangle)_K$ to conclude for the tightness of the laws of $(X^K)_K$. Prokhorov's Theorem implies the relative compactness of this family of laws in the set of probability measures on $\mathbb{D}([0,T],\mathbb{R})$, leading to the existence of a limiting value Q.

Let us now identify the limit. The jumps of X^K have the amplitude $\frac{1}{K}$. Since the mapping $x\to\sup_{t\leq T}|\Delta x(t)|$ is continuous from $\mathbb{D}([0,T],\mathbb{R})$ into $\mathbb{R}_+$, then the probability measure Q only charges the subset of continuous functions. For any $t>0$, we define on $\mathbb{D}([0,T],\mathbb{R})$ the function

$$\psi_t(x) = x_t - x_0 - \int_0^t (\lambda(x_s)-\mu(x_s))\ x_s ds.$$

The assumptions yield

$$|\psi_t(x)|\leq C\sup_{t\leq T}(1+(x_t)^2)$$

and we deduce the uniform integrability of the sequence $(\psi_t(X^K))_K$ from (3.4). The projection mapping $x\to x_t$ isn't continuous on $\mathbb{D}([0,T],\mathbb{R})$ but since Q only charges the continuous paths, we deduce that $X\to\psi_t(X)$ is Q-a.s. continuous, if X denotes the canonical process. Therefore, since Q is the weak limit of a subsequence of $(\mathcal{L}(X^K))_K$ (that for simplicity we still denote $\mathcal{L}(X^K)$) and using the uniform integrability of $(\psi_t(X^K))_K$, we get

$$\mathbb{E}_Q(|\psi_t(X)|) = \lim_K \mathbb{E}(|\psi_t(X^K)|) = \lim_K \mathbb{E}(|M^K_t|).$$

But

$$\mathbb{E}(|M^K_t|)\leq \left(\mathbb{E}(|M^K_t|^2)\right)^{1/2}$$

tends to 0 by (3.6), (3.7) and (3.4). Hence the limiting process X is the deterministic solution of the equation

$$x(t) = x_0 + \int_0^t x_s(\lambda(x_s)-\mu(x_s))ds.$$

That ends the proof. $\square$

3.2 Lotka Volterra models

In the previous section, we have considered the large approximation of an homogeneous population, where demographic rates are similar for all individuals. We could generalize our approach to a set of different subpopulations interacting together by considering a multitype birth and death process ad its large population size approximation.

Let us focus here on the case of two sub-populations characterized by two different types 1 and 2. For $i = 1, 2$, the growth rates of these populations are r_1 and r_2. Individuals compete for resources either inside the same species (intra-specific competition) or with individuals of the other species (inter-specific competition). As before, let K be the scaling parameter describing the capacity of the environment. The competition pressure exerted by an individual of type 1 on an individual of type 1 (resp. type 2) is given by $\dfrac{c_{11}}{K}$ (resp. $\dfrac{c_{21}}{K}$). The competition pressure exerted by an individual of type 2 is respectively given by $\dfrac{c_{12}}{K}$ and $\dfrac{c_{22}}{K}$. The parameters c_{ij} are assumed to be positive.

By similar arguments as in Subsection 3.1, one can prove that the large K-approximation of the population dynamics is described by the well known competitive Lotka-Volterra dynamical system. Let $x_1(t)$ (resp. $x_2(t)$) be the limiting renormalized 1-population size (resp. 2-population size). We get

$$\begin{cases} x_1'(t) = x_1(t)\,(r_1 - c_{11}\,x_1(t) - c_{12}\,x_2(t)); \\ x_2'(t) = x_2(t)\,(r_2 - c_{21}\,x_1(t) - c_{22}\,x_2(t)). \end{cases} \tag{3.11}$$

This system has been extensively studied and its long time behavior is well known. Let us assume that $c_{11}c_{22} - c_{12}c_{21} \neq 0$. Then there are 4 possible equilibria: the unstable equilibrium $(0, 0)$, two trivial equilibria $(\bar{x}_1, 0) = (\frac{r_1}{c_{11}}, 0)$, $(0, \bar{x}_2) = (0, \frac{r_2}{c_{22}})$ and a non-trivial equilibrium (x_1^*, x_2^*) given by

$$x_1^* = \frac{r_1c_{22} - r_2c_{12}}{c_{11}c_{22} - c_{12}c_{21}} \;;\; x_2^* = \frac{r_2c_{11} - r_1c_{21}}{c_{11}c_{22} - c_{12}c_{21}}.$$

Of course, the latter is possible if the two coordinates are positive. The asymptotic behavior of (3.11) is given by the next result.

Proposition 3.4. (i) *Any solution to* (3.11) *with initial condition in* $\mathbb{R}_+^2$ *converges to a finite equilibirum of* (3.11) *in* $\mathbb{R}_+^2$ *when* $t \to +\infty$.

(ii) *The equilibrium* $(\bar{x}_1, 0)$ *is locally asymptotically stable if*

$$r_2c_{11} - r_1c_{21} < 0,$$

It is globally asymptotically stable and attracts all the initial conditions in $\mathbb{R}_+ \times \mathbb{R}_+^*$ *if*

$$r_2c_{11} - r_1c_{21} < 0 \quad \textit{and} \quad r_1c_{22} - r_2c_{12} > 0. \tag{3.12}$$

(iii) *The system* (3.11) *admits a unique non-trivial equilibrium in* $(\mathbb{R}_+^*)^2$ *if and only if*

$$(r_2c_{11} - r_1c_{21})(r_1c_{22} - r_2c_{12}) > 0. \tag{3.13}$$

It is unstable if

$$r_2c_{11} - r_1c_{21} < 0 \quad \textit{and} \quad r_1c_{22} - r_2c_{12} < 0. \tag{3.14}$$

It is globally asymptotically stable and attracts all the initial conditions in $(\mathbb{R}_+^*)^2$ *if*

$$r_2c_{11} - r_1c_{21} > 0 \quad \textit{and} \quad r_1c_{22} - r_2c_{12} > 0. \tag{3.15}$$

Proof. To prove (i), we divide $\mathbb{R}_+^2$ according to the sign of $\dot{x}_1$ and $\dot{x}_2$: $\dot{x}_1$ is positive under the line $r_1 - c_{11}x_1 - c_{12}x_2$ and $\dot{x}_2$ under the line $r_2 - c_{21}x_1 - c_{22}x_2$. Each of these two lines cut the coordinate axes at nonnegative coordinates. In particular, there exists a non-trivial equilibrium in $(\mathbb{R}_+^*)^2$ iff the two lines cut at a point with coordinates having the same sign, hence if and only if (3.13) is satisfied.

We obtain four possible configurations shown in Fig. 1, where the small arrows represent the direction of the flow. Fig. 1 (a) corresponds to the case $r_2c_{11} - r_1c_{21} < 0$ and

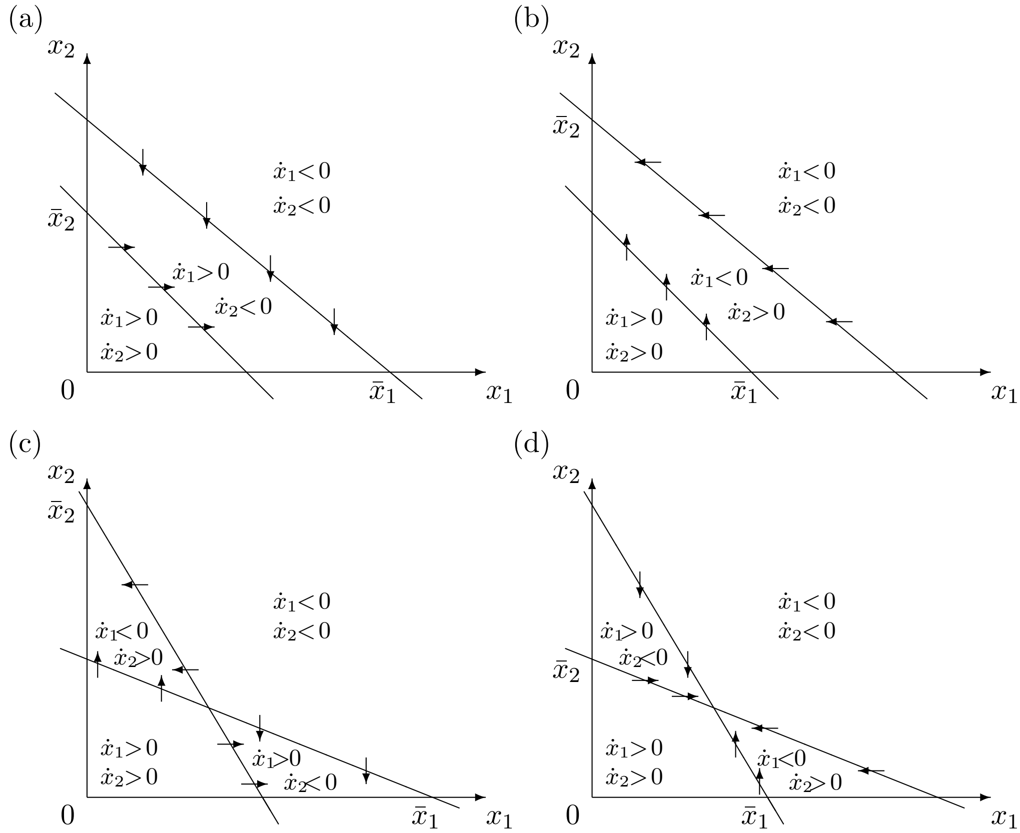


Figure 1: Possible configurations for the signs of $\dot{x}_1$ and $\dot{x}_2$ for the system (3.11).

$r_1c_{22} - r_2c_{12} > 0$, Fig. (b) to the case $r_2c_{11} - r_1c_{21} > 0$ et $r_1c_{22} - r_2c_{12} < 0$, Fig. (c) to the case $r_2c_{11} - r_1c_{21} < 0$ and $r_1c_{22} - r_2c_{12} < 0$, and Fig. (d) to the case $r_2c_{11} - r_1c_{21} > 0$ and $r_1c_{22} - r_2c_{12} > 0$. Cases where one of these quantities is zero correspond to cases where the two lines intersect on one of the coordinate axes, and the two lines are equal if and only if $r_2c_{11} - r_1c_{21} = r_1c_{22} - r_2c_{12} = 0$.

Consider now any solution to (3.11) in each of the cases (a) to (d). If this solution starts in the domain where $\dot{x}_1 \leq 0$ and $\dot{x}_2 \geq 0$, looking at the signs of the derivatives of

x_1 and x_2 at the boundary, we see that the solution cannot exit from this domain. Hence its two coordinates are monotonous and converge to some limit. A similar result holds if the solution starts from the domain where $\dot{x}_1 \geq 0$ and $\dot{x}_2 \leq 0$. When the initial condition satisfies $\dot{x}_1(0) > 0$ and $\dot{x}_2(0) > 0$, the solution can either stay in this domain forever (and then converge as a monotonous function time), or leave it after a finite time and reach one of the first two domains considered above, where the solution will remain forever and converge to some limit. A similar situation holds if the initial condition satisfies $\dot{x}_1(0) < 0$ and $\dot{x}_2(0) < 0$. This ends the proof of (i).

To prove (ii), we first observe that (i) implies that any solution to (3.11) is bounded. Since

$$x_1(t) = x_1(0) \exp\left(\int_0^t (r_1 - c_{11}x_1(s) - c_{12}x_2(s))ds\right)$$
$$x_2(t) = x_2(0) \exp\left(\int_0^t (r_2 - c_{21}x_1(s) - c_{22}x_2(s))ds\right),$$

we deduce that $x_1(t)$ and $x_2(t)$ never reach zero if $x_1(0) > 0$ and $x_2(0) > 0$.

Computing the Jacobian matrix of the system, it is easy to see that the linear criterion of local stability for the equilibrium $(\bar{x}_1, 0)$ is $r_2c_{11} - r_1c_{21} < 0$, and the equilibrium is unstable if $r_2c_{11} - r_1c_{21} > 0$. This gives the first part of (ii).

To prove the global asymptotic stability of $(\bar{x}_1, 0)$ under condition (3.12), we first notice that, because of criterion (3.13), the non-trivial equilibrium does not belong to $(\mathbb{R}_+^*)^2$ in this case. Hence any solution to (3.11) must converge to either $(\bar{x}_1, 0)$ or $(0, \bar{x}_2)$ (except when the initial condition is 0). To end the proof of (ii), It therefore suffices to prove that no solution started from $(\mathbb{R}_+^*)^2$ can converge to $(0, \bar{x}_2)$. Since $r_1c_{22} - r_2c_{12} > 0$, $\dot{x}_1 > 0$ at all point close enough to $(0, \bar{x}_2)$ with $x_1 > 0$. Since we have proved that no solution started from $(\mathbb{R}_+^*)^2$ can hit the axis $\{x_1 = 0\}$, the proof of (ii) is completed.

We now come to the proof of (iii). We already proved that (3.13) is equivalent to the existence of the non-trivial equilibrium (x_1^*, x_2^*), so let us assume (3.13). The Jacobian matrix of the system at (x_1^*, x_2^*) is

$$\begin{pmatrix} -c_{11}x_1^* & -c_{12}x_1^* \\ -c_{21}x_2^* & -c_{22}x_2^* \end{pmatrix}. \tag{3.16}$$

Its determinant $(c_{11}c_{22} - c_{12}c_{21})x_1^*x_2^*$ is striclty negative if $c_{11}c_{22} - c_{12}c_{21} < 0$. So in this case, the Jacobian matrix has a positive eigenvalue and the equilibrium is unstable. Because of the expression of (x_1^*, x_2^*), this case is equivalent to Condition (3.14).

If (3.15) holds true, as was proved in (ii), a solution to (3.11) started from $(\mathbb{R}_+^*)^2$ cannot converge to $(0, 0)$, $(\bar{x}_1, 0)$ or $(0, \bar{x}_2)$. By (i), this solution must converge to an equilibrium so it must be (x_1^*, x_2^*), which ends the proof of (iii). $\square$

One could extend (3.11) to negative coefficients c_{ij}, describing a cooperation effect of species j on the growth of species i. The long time behavior can be totally different. For example, the prey-predator models have been extensively studied in ecology (see [22], Part 1). The simplest prey-predator system

$$\begin{cases} x_1'(t) = x_1(t)\,(r_1 - c_{12}\,x_2(t)); \\ x_2'(t) = x_2(t)\,(c_{21}\,x_1(t) - r_2), \end{cases} \tag{3.17}$$

with $r_1, r_2, c_{12}, c_{21} > 0$, has periodic solutions.

Of course, we could also study multi-dimensional systems corresponding to multi-type population models. In what follows we are more interested in modeling the case where the types of the individuals belong to a continuum. That will allow us to add mutation events where the offspring of an individual may randomly mutate and create a new type.

4 Population Point Measure Processes

We are now interested in the mathematical modeling of Darwinian evolution. Even if the evolution appears as a global change in the state of a population, its basic mechanisms, mutation and selection, operate at the level of individuals. Consequently, the evolving population is modeled as a stochastic system of competing individuals (sharing limited resources). Each individual is characterized by a vector of phenotypic trait values, heritable except when mutation occurs. The trait space $\mathcal{X}$ is assumed to be a compact subset of $\mathbb{R}^d$, for some $d \geq 1$. The population is described by a random point measure with support on the trait space.

We will denote by $M_F(\mathcal{X})$ the set of all finite non-negative measures on $\mathcal{X}$. Let $\mathcal{M}$ be the subset of $M_F(\mathcal{X})$ consisting of all finite point measures:

$$\mathcal{M} = \left\{ \sum_{i=1}^{n} \delta_{x_i},\ n \geq 0, x_1, ..., x_n \in \mathcal{X} \right\}.$$

Here and below, δ_x denotes the Dirac mass at x. For any $\mu \in M_F(\mathcal{X})$ and any measurable function f on $\mathcal{X}$, we set $\langle \mu, f \rangle = \int_{\mathcal{X}} f d\mu$.

We wish to study the stochastic process $(Y_t, t \geq 0)$, taking its values in $\mathcal{M}$, and describing the distribution of individuals and traits at time t. We define

$$Y_t = \sum_{i=1}^{N_t} \delta_{X_t^i}, \tag{4.1}$$

$N_t = \langle Y_t, 1 \rangle \in \mathbb{N}$ standing for the number of individuals alive at time t, and $X_t^1, ..., X_t^{N_t}$ describing the individuals' traits (in $\mathcal{X}$).

We assume that the birth rate of an individual with trait x is $b(x)$ and that for a population $\nu = \sum_{i=1}^{N} \delta_{x^i}$, its death rate is given by $d(x, C * \nu(x)) = d(x, \sum_{i=1}^{N} C(x - x^i))$. This death rate takes into account the intrinsic death rate of the individual, depending on its phenotypic trait x but also on the competition pressure exerted by the other individuals alive, modeled by the competition kernel C. Let $p(x)$ and $m(x,z)dz$ be respectively the probability that an offspring produced by an individual with trait x carries a mutated trait and the law of this mutant trait.
Thus, the population dynamics can be roughly summarized as follows. The initial population is characterized by a (possibly random) counting measure $\nu_0 \in \mathcal{M}$ at time 0, and any individual with trait x at time t has two independent random exponentially distributed "clocks": a birth clock with parameter $b(x)$, and a death clock with parameter $d(x, C * Y_t(x))$. If the death clock of an individual rings, this individual dies and disappears. If the birth clock of an individual with trait x rings, this individual produces an offspring. With probability $1-p(x)$ the offspring carries the same trait x; with probability

$p(x)$ the trait is mutated. If a mutation occurs, the mutated offspring instantly acquires a new trait z, picked randomly according to the mutation step measure $m(x,z)dz$. When one of these events occurs, all individual's clocks are reset to 0.

We are looking for a $\mathcal{M}$-valued Markov process $(Y_t)_{t\geq 0}$ with infinitesimal generator L, defined for all real bounded functions ϕ and $\nu \in \mathcal{M}$ by

$$\begin{aligned} L\phi(\nu) &= \sum_{i=1}^{N} b(x^i)(1-p(x^i))(\phi(\nu+\delta_{x^i})-\phi(\nu)) \\ &+ \sum_{i=1}^{N} b(x^i)p(x^i)\int_{\mathcal{X}}(\phi(\nu+\delta_z)-\phi(\nu))m(x^i,z)dz \\ &+ \sum_{i=1}^{N} d(x^i, C*\nu(x^i))(\phi(\nu-\delta_{x^i})-\phi(\nu)). \end{aligned} \tag{4.2}$$

The first term in (4.2) captures the effect of births without mutation, the second term the effect of births with mutation and the last term the effect of deaths. The density-dependence makes the third term nonlinear.

4.1 Pathwise construction

Let us justify the existence of a Markov process with infinitesimal generator L. The explicit construction of $(Y_t)_{t\geq 0}$ also yields two side benefits: providing a rigorous and efficient algorithm for numerical simulations (given hereafter) and establishing a general method that will be used to derive some large population limits (Section 5).

We make the biologically natural assumption that the trait dependency of birth parameters is "bounded", and at most linear for the death rate. Specifically, we assume

Assumption 4.1. *There exist constants $\bar{b}$, $\bar{d}$, $\bar{C}$, and α and a probability density function $\bar{m}$ on $\mathbb{R}^d$ such that for each $\nu = \sum_{i=1}^{N} \delta_{x^i}$ and for $x, z \in \mathcal{X}$, $\zeta \in \mathbb{R}$,*

$$\begin{gathered} b(x) \leq \bar{b}, \quad d(x,\zeta) \leq \bar{d}(1+|\zeta|), \\ 0 < C^* \leq C(x) \leq \bar{C}, \\ m(x,z) \leq \alpha\,\bar{m}(z-x). \end{gathered}$$

These assumptions ensure that there exists a constant $\widehat{C}$, such that for a population measure $\nu = \sum_{i=1}^{N} \delta_{x^i}$, the total event rate, obtained as the sum of all event rates, is bounded by $\widehat{C}N(1+N)$.

Let us now give a pathwise description of the population process $(Y_t)_{t\geq 0}$. We introduce the following notation.

Notation 1. Let $\mathbb{N}^* = \mathbb{N}\backslash\{0\}$. Let $H = (H^1, ..., H^k, ...) : \mathcal{M} \mapsto (\mathbb{R}^d)^{\mathbb{N}^*}$ be defined by $H\left(\sum_{i=1}^{n} \delta_{x_i}\right) = (x_{\sigma(1)}, ..., x_{\sigma(n)}, 0, ..., 0, ...)$, where σ is a permutation such that $x_{\sigma(1)} \preccurlyeq ... \preccurlyeq x_{\sigma(n)}$, for some arbitrary order $\preccurlyeq$ on $\mathbb{R}^d$ (for example the lexicographic order).

This function H allows us to overcome the following (purely notational) problem. Choosing a trait uniformly among all traits in a population $\nu \in \mathcal{M}$ consists in choosing i uniformly in $\{1, ..., \langle \nu, 1\rangle\}$, and then in choosing the individual *number* i (from the arbitrary order point of view). The trait value of such an individual is thus $H^i(\nu)$.

We now introduce the probabilistic objects we will need.

Definition 4.1. *Let $(\Omega, \mathcal{F}, P)$ be a (sufficiently large) probability space. On this space, we consider the following four independent random elements:*

(i) *a $\mathcal{M}$-valued random variable Y_0 (the initial distribution),*

(ii) *independent Poisson point measures $N_1(ds, di, d\theta)$, and $N_3(ds, di, d\theta)$ on $\mathbb{R}_+ \times \mathbb{N}^* \times \mathbb{R}^+$, with the same intensity measure $ds \left(\sum_{k\geq 1} \delta_k(di)\right) d\theta$ (the "clonal" birth and the death Poisson measures),*

(iii) *a Poisson point measure $N_2(ds, di, dz, d\theta)$ on $\mathbb{R}_+ \times \mathbb{N}^* \times \mathcal{X} \times \mathbb{R}^+$, with intensity measure $ds \left(\sum_{k\geq 1} \delta_k(di)\right) dzd\theta$ (the mutation Poisson point measure).*

Let us denote by $(\mathcal{F}_t)_{t\geq 0}$ the canonical filtration generated by these processes.

We finally define the population process in terms of these stochastic objects.

Definition 4.2. *Assume (H). An $(\mathcal{F}_t)_{t\geq 0}$-adapted stochastic process $\nu = (Y_t)_{t\geq 0}$ is called a population process if a.s., for all $t \geq 0$,*

$$\begin{aligned} Y_t = Y_0 + &\int_{[0,t]\times\mathbb{N}^*\times\mathbb{R}^+} \delta_{H^i(Y_{s-})} \mathbf{1}_{\{i\leq\langle Y_{s-},1\rangle\}}\ \mathbf{1}_{\{\theta\leq b(H^i(Y_{s-}))(1-p(H^i(Y_{s-})))\}} N_1(ds, di, d\theta) \\ &+ \int_{[0,t]\times\mathbb{N}^*\times\mathcal{X}\times\mathbb{R}^+} \delta_z \mathbf{1}_{\{i\leq\langle Y_{s-},1\rangle\}}\ \mathbf{1}_{\{\theta\leq b(H^i(Y_{s-}))p(H^i(Y_{s-}))\, m(H^i(Y_{s-}),z)\}} N_2(ds, di, dz, d\theta) \\ &- \int_{[0,t]\times\mathbb{N}^*\times\mathbb{R}^+} \delta_{H^i(Y_{s-})} \mathbf{1}_{\{i\leq\langle Y_{s-},1\rangle\}} \mathbf{1}_{\{\theta\leq d(H^i(Y_{s-}),C*Y_{s-}(H^i(Y_{s-})))\}} N_3(ds, di, d\theta) \end{aligned} \tag{4.3}$$

Let us now show that if Y solves (4.3), then Y follows the Markovian dynamics we are interested in.

Proposition 4.3. *Assume Assumption 4.1 holds and consider a solution $(Y_t)_{t\geq 0}$ of (4.3) such that $\mathbb{E}(\sup_{t\leq T}\langle Y_t, \mathbf{1}\rangle^2) < +\infty$, $\forall T > 0$. Then $(Y_t)_{t\geq 0}$ is a Markov process. Its infinitesimal generator L is defined by (4.2). In particular, the law of $(Y_t)_{t\geq 0}$ does not depend on the chosen order $\preccurlyeq$.*

Proof. The fact that $(Y_t)_{t\geq 0}$ is a Markov process is classical. Let us now consider a measurable bounded function ϕ. With our notation, $Y_0 = \sum_{i=1}^{\langle Y_0, 1\rangle} \delta_{H^i(Y_0)}$. A simple computation, using the fact that a.s., $\phi(Y_t) = \phi(Y_0) + \sum_{s\leq t}(\phi(Y_{s-} + (Y_s - Y_{s-})) - \phi(Y_{s-}))$, shows that

$$\begin{aligned}\phi(Y_t) = \phi(Y_0) + \int_{[0,t]\times\mathbb{N}^*\times\mathbb{R}^+} & \big(\phi(Y_{s-} + \delta_{H^i(Y_{s-})}) - \phi(Y_{s-})\big)\, \mathbf{1}_{\{i\leq \langle Y_{s-},1\rangle\}} \\ & \mathbf{1}_{\{\theta\leq b(H^i(Y_{s-}))(1-p(H^i(Y_{s-})))\}} N_1(ds,di,d\theta) \\ + \int_{[0,t]\times\mathbb{N}^*\times\mathcal{X}\times\mathbb{R}^+} & (\phi(Y_{s-}+\delta_z) - \phi(Y_{s-}))\, \mathbf{1}_{\{i\leq \langle Y_{s-},1\rangle\}} \\ & \mathbf{1}_{\{\theta\leq b(H^i(Y_{s-}))p(H^i(Y_{s-}))\, m(H^i(Y_{s-}),z)\}} N_2(ds,di,dz,d\theta) \\ + \int_{[0,t]\times\mathbb{N}^*\times\mathbb{R}^+} & \big(\phi(Y_{s-} - \delta_{H^i(Y_{s-})}) - \phi(Y_{s-})\big)\, \mathbf{1}_{\{i\leq \langle Y_{s-},1\rangle\}} \\ & \mathbf{1}_{\{\theta\leq d(H^i(Y_{s-}),C*Y_{s-}(H^i(Y_{s-})))\}} N_3(ds,di,d\theta).\end{aligned}$$

Taking expectations, we obtain

$$\begin{aligned}\mathbb{E}(\phi(Y_t)) &= \mathbb{E}(\phi(Y_0)) \\ &+ \int_0^t \mathbb{E}\Big(\sum_{i=1}^{\langle Y_s,1\rangle} \Big\{ \big(\phi(Y_s+\delta_{H^i(Y_s)}) - \phi(Y_s)\big)\, b(H^i(Y_s))(1-p(H^i(Y_s))) \\ &+ \int_{\mathcal{X}} (\phi(Y_s+\delta_z) - \phi(Y_s))\, b(H^i(Y_s))p(H^i(Y_s))\, m(H^i(Y_s),z)dz \\ &+ \big(\phi(Y_s - \delta_{H^i(Y_s)}) - \phi(Y_s)\big)\, d(H^i(Y_s), C*Y_s(H^i(Y_s)))\Big\}\Big)ds\end{aligned}$$

Differentiating this expression at $t = 0$ leads to (4.2). □

Let us show the existence and some moment properties for the population process.

Theorem 4.4. (i) *Assume Assumption 4.1 holds and that* $\mathbb{E}\left(\langle Y_0,1\rangle\right) < \infty$. *Then the process* $(Y_t)_{t\geq 0}$ *defined in Definition 4.2 is well defined on* $\mathbb{R}_+$.

(ii) *If furthermore for some* $p \geq 1$, $\mathbb{E}\left(\langle Y_0,1\rangle^p\right) < \infty$, *then for any* $T < \infty$,

$$\mathbb{E}(\sup_{t\in[0,T]} \langle Y_t,1\rangle^p) < +\infty. \tag{4.4}$$

Proof. We first prove (ii). Consider the process $(Y_t)_{t\geq 0}$. We introduce for each n the stopping time $\tau_n = \inf\{t\geq 0,\ \langle Y_t,1\rangle \geq n\}$. Then a simple computation using Assumption 4.1 shows that, dropping the non-positive death terms,

$$\begin{aligned}\sup_{s\in[0,t\wedge\tau_n]} \langle Y_s,1\rangle^p \leq \langle Y_0,1\rangle^p + \int_{[0,t\wedge\tau_n]\times\mathbb{N}^*\times\mathbb{R}^+} & ((\langle Y_{s-},1\rangle+1)^p - \langle Y_{s-},1\rangle^p)\, \mathbf{1}_{\{i\leq \langle Y_{s-},1\rangle\}} \\ & \mathbf{1}_{\{\theta\leq b(H^i(Y_{s-}))(1-p(H^i(Y_{s-})))\}} N_1(ds,di,d\theta) \\ + \int_{[0,t]\times\mathbb{N}^*\times\mathcal{X}\times\mathbb{R}^+} & ((\langle Y_{s-},1\rangle+1)^p - \langle Y_{s-},1\rangle^p)\, \mathbf{1}_{\{i\leq \langle Y_{s-},1\rangle\}} \\ & \mathbf{1}_{\{\theta\leq b(H^i(Y_{s-}))p(H^i(Y_{s-}))\, m(H^i(Y_{s-}),z)\}} N_2(ds,di,dz,d\theta).\end{aligned}$$

Using the inequality $(1+x)^p - x^p \leq C_p(1+x^{p-1})$ and taking expectations, we thus obtain, the value of C_p changing from one line to the other,

$$\mathbb{E}(\sup_{s\in[0,t\wedge\tau_n]} \langle Y_s, 1\rangle^p) \leq C_p \left(1 + \mathbb{E}\left(\int_0^{t\wedge\tau_n} \bar{b}\left(\langle Y_{s-}, 1\rangle + \langle Y_{s-}, 1\rangle^p\right) ds\right)\right)$$
$$\leq C_p \left(1 + \mathbb{E}\left(\int_0^t (1 + \langle Y_{s\wedge\tau_n}, 1\rangle^p)\, ds\right)\right).$$

The Gronwall Lemma allows us to conclude that for any $T < \infty$, there exists a constant $C_{p,T}$, not depending on n, such that

$$\mathbb{E}(\sup_{t\in[0,T\wedge\tau_n]} \langle Y_t, 1\rangle^p) \leq C_{p,T}. \tag{4.5}$$

First, we deduce that τ_n tends a.s. to infinity. Indeed, if not, one may find a $T_0 < \infty$ such that $\epsilon_{T_0} = P\left(\sup_n \tau_n < T_0\right) > 0$. This would imply that $\mathbb{E}\left(\sup_{t\in[0,T_0\wedge\tau_n]} \langle Y_t, 1\rangle^p\right) \geq \epsilon_{T_0} n^p$ for all n, which contradicts (4.5). We may let n go to infinity in (4.5) thanks to the Fatou Lemma. This leads to (4.4).

Point (i) is a consequence of point (ii). Indeed, one builds the solution $(Y_t)_{t\geq 0}$ step by step. One only has to check that the sequence of jump instants T_n goes a.s. to infinity as n tends to infinity. But this follows from (4.4) with $p = 1$. □

4.2 Examples and simulations

Let us remark that Assumption 4.1 is satisfied in the case where

$$d(x, C*\nu(x)) = d(x) + \alpha(x)\int_{\mathcal{X}} C(x-y)\nu(dy), \tag{4.6}$$

and b, d and α are bounded functions.

In the case where moreover, $p \equiv 1$, this individual-based model can also be interpreted as a model of "spatially structured population", where the trait is viewed as a spatial location and the mutation at each birth event is viewed as dispersal. This kind of models have been introduced by Bolker and Pacala ([5, 6]) and Law et al. ([34]), and mathematically studied by Fournier and Méléard [19]. The case $C \equiv 1$ corresponds to a density-dependence in the total population size.

Later, we will consider the particular set of parameters leading to the logistic interaction model, taken from Kisdi [30] and corresponding to a model of asymmetric competition:

$$\mathcal{X} = [0,4], \quad d(x) = 0, \quad \alpha(x) = 1, \quad p(x) = p,$$
$$b(x) = 4 - x, \quad C(x-y) = \frac{2}{K}\left(1 - \frac{1}{1 + 1.2\exp(-4(x-y))}\right) \tag{4.7}$$

and $m(x,z)dz$ is a Gaussian law with mean x and variance σ^2 conditioned to stay in $[0,4]$. As we will see in Section 5, the constant K scaling the strength of competition also scales the population size (when the initial population size is proportional to K). In this model, the trait x can be interpreted as body size. Equation (4.7) means that body size influences the birth rate negatively, and creates asymmetrical competition reflected in the sigmoid shape of C (being larger is competitively advantageous).

Let us give now an algorithmic construction of the population process (in the general case), simulating the size N_t of the population and the trait vector $\mathbf{X}_t$ of all individuals alive at time t.

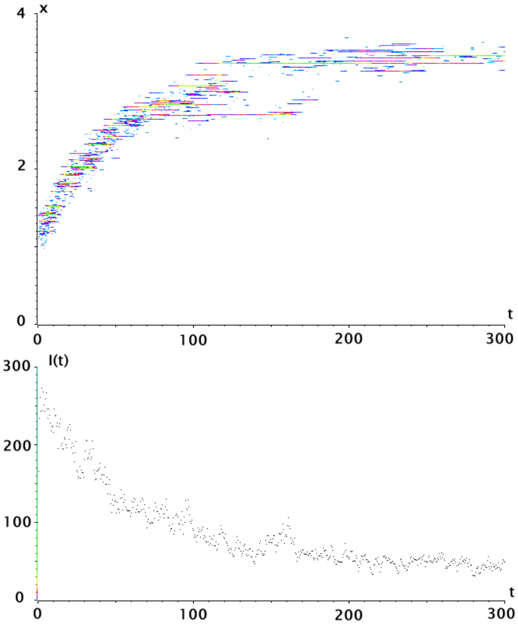
At time $t = 0$, the initial population Y_0 contains N_0 individuals and the corresponding trait vector is $\mathbf{X}_0 = (X_0^i)_{1\leq i\leq N_0}$. We introduce the following sequences of independent random variables, which will drive the algorithm.

- The type of birth or death events will be selected according to the values of a sequence of random variables $(W_k)_{k\in\mathbb{N}^*}$ with uniform law on $[0,1]$.
- The times at which events may be realized will be described using a sequence of random variables $(\tau_k)_{k\in\mathbb{N}}$ with exponential law with parameter $\widehat{C}$.
- The mutation steps will be driven by a sequence of random variables $(Z_k)_{k\in\mathbb{N}}$ with law $\bar{m}(z)dz$.

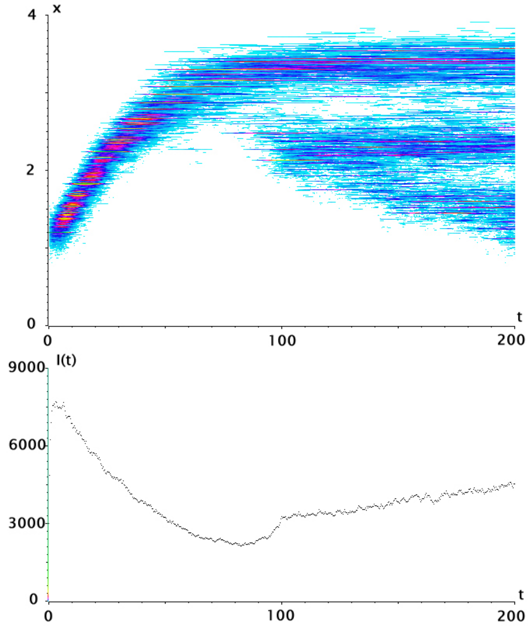
We set $T_0 = 0$ and construct the process inductively for $k \geq 1$ as follows.
At step $k-1$, the number of individuals is N_{k-1}, and the trait vector of these individuals is $\mathbf{X}_{T_{k-1}}$.
Let $T_k = T_{k-1} + \dfrac{\tau_k}{N_{k-1}(N_{k-1}+1)}$. Notice that $\dfrac{\tau_k}{N_{k-1}(N_{k-1}+1)}$ represents the time between jumps for N_{k-1} individuals, and $\widehat{C}(N_{k-1}+1)$ gives an upper bound of the total rate of events affecting each individual.

At time T_k, one chooses an individual $i_k = i$ uniformly at random among the N_{k-1} alive in the time interval $[T_{k-1}, T_k)$; its trait is $X^i_{T_{k-1}}$. (If $N_{k-1} = 0$ then $Y_t = 0$ for all $t \geq T_{k-1}$.)

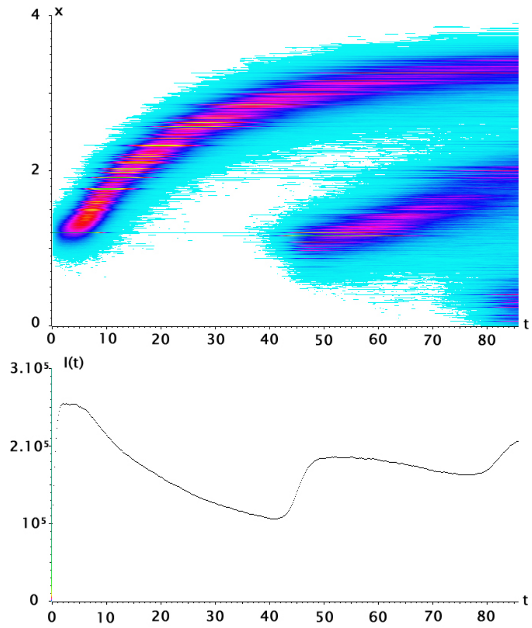
- If $0 \leq W_k \leq \dfrac{d(X^i_{T_{k-1}}, \sum_{j=1}^{I_{k-1}} C(X^i_{T_{k-1}} - X^j_{T_{k-1}}))}{\widehat{C}(N_{k-1}+1)} = W_1^i(\mathbf{X}_{T_{k-1}})$, then the chosen individual dies, and $N_k = N_{k-1} - 1$.
- If $W_1^i(\mathbf{X}_{T_{k-1}}) < W_k \leq W_2^i(\mathbf{X}_{T_{k-1}})$, where
$$W_2^i(\mathbf{X}_{T_{k-1}}) = W_1^i(\mathbf{X}_{T_{k-1}}) + \frac{[1 - p(X^i_{T_{k-1}})]b(X^i_{T_{k-1}})}{\bar{C}(N_{k-1}+1)},$$
then the chosen individual gives birth to an offspring with trait $X^i_{T_{k-1}}$, and $N_k = N_{k-1} + 1$.
- If $W_2^i(\mathbf{X}_{T_{k-1}}) < W_k \leq W_3^i(\mathbf{X}_{T_{k-1}}, Z_k)$, where
$$W_3^i(\mathbf{X}_{T_{k-1}}, Z_k) \quad = \quad W_2^i(\mathbf{X}_{T_{k-1}}) \ + \ \frac{p(X^i_{T_{k-1}})b(X^i_{T_{k-1}})\, m(X^i_{T_{k-1}}, X^i_{T_{k-1}} + Z_k)}{\widehat{C}\bar{m}(Z_k)(N_{k-1}+1)},$$
then the chosen individual gives birth to a mutant offspring with trait $X^i_{T_{k-1}} + Z_k$, and $N_k = N_{k-1} + 1$.
- If $W_k > W_3^i(\mathbf{X}_{T_{k-1}}, Z_k)$, nothing happens, and $N_k = N_{k-1}$.



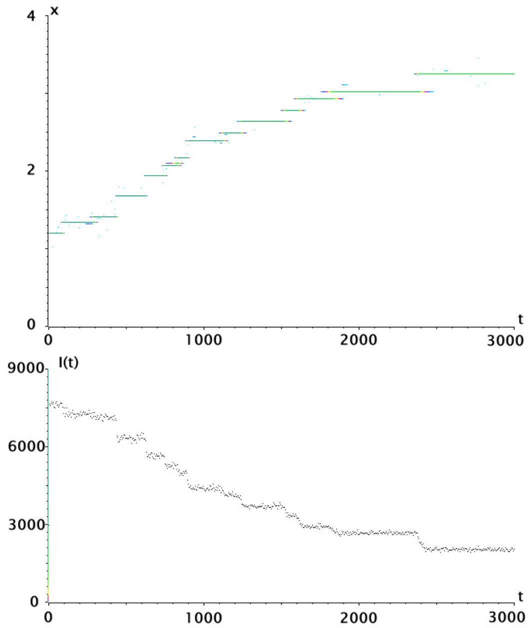
(a) $p = 0.03$, $K = 100$, $\sigma = 0.1$.



(b) $p = 0.03$, $K = 3000$, $\sigma = 0.1$.



(c) $p = 0.03$, $K = 100000$, $\sigma = 0.1$.



(d) $p = 0.00001$, $K = 3000$, $\sigma = 0.1$.

Figure 2: *Numerical simulations of trait distributions (upper panels, darker means higher frequency) and population size (lower panels). The initial population is monomorphic with trait value* 1.2 *and contains* K *individuals. (a–c) Qualitative effect of increasing the system size (measured by the parameter* K*). (d) Large system size with rare mutations*

Then, at any time $t \geq 0$, the number of individuals and the population process are defined by

$$N_t = \sum_{k\geq 0} 1_{\{T_k \leq t < T_{k+1}\}} N_k, \qquad Y_t = \sum_{k\geq 0} 1_{\{T_k \leq t < T_{k+1}\}} \sum_{i=1}^{N_k} \delta_{X^i_{T_k}}.$$

The simulation of Kisdi's example (4.7) can be carried out following this algorithm. We can show a very wide variety of qualitative behaviors depending on the value of the parameters σ, p and K.
In Figure 2, the upper part gives the distribution of the traits in the population at any time, using a grey scale code for the number of individuals holding a given trait. The lower part of the simulation represents the dynamics of the total population size N_t.
These simulations will serve to illustrate the mathematical scalings described in Section 5. In Fig. 2 (a)–(c), we see the qualitative effect of increasing scalings K, from a finite trait support process for small K to a wide population density for large K. The simulation (d) illustrates the case of rare mutations in a longer time scale, studied in Section 6. Although issued from the same individual system, these simulations show very different qualitative behaviors. The end of the notes will be devoted to the mathematical study of these asymptotics.

4.3 Martingale Properties

The martingale properties of the process $(Y_t)_{t\geq 0}$ are the key point of our approach.

Theorem 4.5. *Suppose Assumption 4.1 holds and that for some $p \geq 2$, $E\left(\langle Y_0, 1\rangle^p\right) < \infty$.*

(i) *For all measurable functions ϕ from $\mathcal{M}$ into $\mathbb{R}$ such that for some constant C, for all $\nu \in \mathcal{M}$, $|\phi(\nu)| + |L\phi(\nu)| \leq C(1 + \langle \nu, 1\rangle^p)$, the process*

$$\phi(Y_t) - \phi(Y_0) - \int_0^t L\phi(Y_s)ds \tag{4.8}$$

is a càdlàg $(\mathcal{F}_t)_{t\geq 0}$-martingale starting from 0.

(ii) *Point (i) applies to any function $\phi(\nu) = \langle \nu, f\rangle^q$, with $0 \leq q \leq p-1$ and with f bounded and measurable on $\mathcal{X}$.*

(iii) *For such a function f, the process*

$$\begin{aligned} M_t^f = \langle Y_t, f\rangle - \langle Y_0, f\rangle - \int_0^t \int_{\mathcal{X}} \bigg\{ \Big((1-p(x))b(x) - d(x, C * Y_s(x)) \Big) f(x) \\ + p(x)b(x) \int_{\mathcal{X}} f(z)\, m(x,z)dz \bigg\} Y_s(dx)ds \end{aligned} \tag{4.9}$$

is a càdlàg square integrable martingale starting from 0 with quadratic variation

$$\begin{aligned} \langle M^f\rangle_t = \int_0^t \int_{\mathcal{X}} \bigg\{ \Big((1-p(x))b(x) - d(x, C * Y_s(x)) \Big) f^2(x) \\ + p(x)b(x) \int_{\mathcal{X}} f^2(z)\, m(x,z)dz \bigg\} Y_s(dx)ds. \end{aligned} \tag{4.10}$$

Proof. The proof follows the proof of Theorem 2.6. First of all, note that point (i) is immediate thanks to Proposition 4.3 and (4.4). Point (ii) follows from a straightforward computation using (4.2). To prove (iii), we first assume that $E\left(\langle Y_0, 1\rangle^3\right) < \infty$. We apply (i) with $\phi(\nu) = \langle \nu, f\rangle$. This gives us that M^f is a martingale. To compute its bracket, we first apply (i) with $\phi(\nu) = \langle \nu, f\rangle^2$ and obtain that

$$\langle Y_t, f\rangle^2 - \langle Y_0, f\rangle^2 - \int_0^t \int_{\mathcal{X}} \Bigg\{ \Big((1-p(x))b(x)(f^2(x) + 2f(x)\langle Y_s, f\rangle) \\ + d(x, C * Y_s(x))(f^2(x) - 2f(x)\langle Y_s, f\rangle) \Big) \\ + p(x)b(x) \int_{\mathcal{X}} (f^2(z) + 2f(z)\langle Y_s, f\rangle)\, m(x,z)dz \Bigg\} Y_s(dx)ds \tag{4.11}$$

is a martingale. On the other hand, we apply the Itô formula to compute $\langle Y_t, f\rangle^2$ from (4.9). We deduce that

$$\langle Y_t, f\rangle^2 - \langle Y_0, f\rangle^2 - \int_0^t 2\langle Y_s, f\rangle \int_{\mathcal{X}} \Bigg\{ \Big((1-p(x))b(x) - d(x, C * Y_s(x)) \Big) f(x) \\ + p(x)b(x) \int_{\mathcal{X}} f(z)m(x,z)dz \Bigg\} Y_s(dx)ds - \langle M^f\rangle_t \tag{4.12}$$

is a martingale. Comparing (4.11) and (4.12) leads to (4.10). The extension to the case where only $E\left(\langle Y_0, 1\rangle^2\right) < \infty$ is straightforward by a localization argument, since also in this case, $E(\langle M^f\rangle_t) < \infty$ thanks to (4.4) with $p = 2$. $\square$

5 Scaling limits for the individual-based process

As in Section 3, we consider the case where the system size becomes very large. We scale this size by the integer K and look for approximations of the conveniently renormalized measure-valued population process, when K tends to infinity.

For any $K \in \mathbb{N}^*$, let the set of functions C_K, b, d, m, p satisfy Assumption 4.1. Let Y_t^K be the counting measure of the population at time t. We define the measure-valued Markov process $(X_t^K)_{t\geq 0}$ by

$$X_t^K = \frac{1}{K} Y_t^K.$$

As the system size K goes to infinity, we need to assume the

Assumption 5.1. *The parameters C_K, b, d, m and p are continuous, $\zeta \mapsto d(x,\zeta)$ is Lipschitz continuous for any $x \in \mathcal{X}$ and*

$$C_K(x) = \frac{C(x)}{K}.$$

A biological interpretation of this renormalization is that larger systems are made up of smaller individuals, which may be a consequence of a fixed amount of available resources to be partitioned among individuals. Indeed, the biomass of each interacting individual scales like $1/K$, which may imply that the interaction effect of the global population on a focal individual is of order 1. The parameter K may also be interpreted as scaling the amount of resources available, so that the renormalization of C_K reflects the decrease of competition for resources.

The generator $\tilde{L}^K$ of $(Y_t^K)_{t\geq 0}$ is given by (4.2), with C_K instead of C. The generator L^K of $(X_t^K)_{t\geq 0}$ is obtained by writing, for any measurable function ϕ from $M_F(\mathcal{X})$ into $\mathbb{R}$ and any $\nu \in M_F(\mathcal{X})$,

$$L^K\phi(\nu) = \partial_t \mathbb{E}_\nu(\phi(X_t^K))_{t=0} = \partial_t \mathbb{E}_{K\nu}(\phi(Y_t^K/K))_{t=0} = \tilde{L}^K\phi^K(K\nu)$$

where $\phi^K(\mu) = \phi(\mu/K)$.

By a similar proof as that carried out in Section 4.3, we may summarize the moment and martingale properties of X^K.

Proposition 5.1. *Assume that for some $p \geq 2$, $\mathbb{E}(\langle X_0^K, 1\rangle^p) < +\infty$.*

(1) *For any $T > 0$, $\mathbb{E}\big(\sup_{t\in[0,T]}\langle X_t^K, 1\rangle^p\big) < +\infty$.*

(2) *For any bounded and measurable function ϕ on M_F such that $|\phi(\nu)| + |L^K\phi(\nu)| \leq C(1 + \langle \nu, 1\rangle^p)$, the process*

$$\phi(X_t^K) - \phi(X_0^K) - \int_0^t L^K\phi(X_s^K)ds \tag{5.1}$$

is a càdlàg martingale.

(3) *For each measurable bounded function f, the process*

$$\begin{aligned} M_t^{K,f} &= \langle X_t^K, f\rangle - \langle X_0^K, f\rangle \\ &- \int_0^t \int_{\mathcal{X}} (b(x) - d(x, C * X_s^K(x)))f(x)X_s^K(dx)ds \\ &- \int_0^t \int_{\mathcal{X}} p(x)b(x)\Big(\int_{\mathcal{X}} f(z)m_K(x,z)dz - f(x)\Big)X_s^K(dx)ds \end{aligned} \tag{5.2}$$

is a square integrable martingale with quadratic variation

$$\begin{aligned} \langle M^{K,f}\rangle_t = \frac{1}{K}\Big\{ &\int_0^t \int_{\mathcal{X}} p(x)b(x)\Big(\int_{\mathcal{X}} f^2(z)\, m(x,z)dz - f^2(x)\Big)X_s^K(dx)ds \\ &+ \int_0^t \int_{\mathcal{X}} (b(x) + d(x, C * X_s^K(x)))f^2(x)X_s^K(dx)ds\Big\}. \end{aligned} \tag{5.3}$$

Let us make K tend to infinity.

Theorem 5.2. *Assume Assumptions 4.1 and 5.1 hold. Assume moreover that $\sup_K E(\langle X_0^K, 1\rangle^3) < +\infty$ and that the initial conditions X_0^K converge in law and for the weak topology on $M_F(\mathcal{X})$ as K increases, to a finite deterministic measure ξ_0.*

Then for any $T > 0$, the process $(X_t^K)_{t\geq 0}$ converges in law, in the Skorohod space $\mathbb{D}([0,T], M_F(\mathcal{X}))$, as K goes to infinity, to the unique deterministic continuous function $\xi \in C([0,T], M_F(\mathcal{X}))$ satisfying for any continuous $f : \mathcal{X} \to \mathbb{R}$

$$\langle \xi_t, f\rangle = \langle \xi_0, f\rangle + \int_0^t \int_{\mathcal{X}} f(x)[(1-p(x))b(x) - d((x, C * \xi_s(x))]\xi_s(dx)ds \\ + \int_0^t \int_{\mathcal{X}} p(x)b(x) \left(\int_{\mathcal{X}} f(z)\, m(x,z)dz \right) \xi_s(dx)ds \tag{5.4}$$

This result is illustrated by the simulations of Fig. 2 (a)–(c).

Proof. We divide the proof in five steps. Let us fix $T > 0$.

Step 1 Let us first show the uniqueness of a solution of the equation (5.4).
Let us consider two solutions $(\xi_t)_{t\geq 0}$ and $(\bar{\xi}_t)_{t\geq 0}$ of (5.4) satisfying $\sup_{t\in[0,T]} \langle \xi_t + \bar{\xi}_t, 1\rangle = A_T < +\infty$. Recall that the total variation norm is given for μ_1 and μ_2 in M_F by

$$||\mu_1 - \mu_2|| = \sup_{f\in L^\infty(\mathbb{R}^d),\, ||f||_\infty \leq 1} |\langle \mu_1 - \mu_2, f\rangle|. \tag{5.5}$$

Let f be a bounded measurable function on $\mathcal{X}$ such that $||f||_\infty \leq 1$. We get

$$|\langle \xi_t - \bar{\xi}_t, f\rangle| \leq \int_0^t \bigg| \int_{\mathbb{R}^d} [\xi_s(dx) - \bar{\xi}_s(dx)] \left((1-p(x)b(x) - d(x, C * \xi_s(x))\right) f(x) \\ + p(x)b(x) \left(\int_{\mathcal{X}} f(z)\, m(x,z)dz \right) \bigg| ds \\ + \int_0^t \left| \int_{\mathbb{R}^d} \bar{\xi}_s(dx)(d(x, C * \xi_s(x)) - d(x, C * \bar{\xi}_s(x)))f(x) \right| ds. \tag{5.6}$$

Since $||f||_\infty \leq 1$, for all $x \in \mathbb{R}^d$,

$$\left| \Big((1-p(x)b(x) - d(x, C * \xi_s(x)) \Big) f(x) + p(x)b(x) \Big(\int_{\mathcal{X}} f(z)\, m(x,z)dz \Big) \right| \leq \bar{b} + \bar{d}(1 + \bar{C}A_T).$$

Moreover, d is Lipschitz continuous in its second variable with Lipschitz constant K_d. Thus we obtain from (5.6) that

$$|\langle \xi_t - \bar{\xi}_t, f\rangle| \leq \left[\bar{b} + \bar{d}(1 + \bar{C}A_T) + K_d A_T \bar{C}\right] \int_0^t ||\xi_s - \bar{\xi}_s|| ds. \tag{5.7}$$

Taking the supremum over all functions f such that $||f||_\infty \leq 1$, and using Gronwall's Lemma, we finally deduce that for all $t \leq T$, $||\xi_t - \bar{\xi}_t|| = 0$. Uniqueness holds.

Step 2 Next, we need some moment estimates on the time interval $[0,T]$, $T > 0$. To this end, we consider (5.2) with $f = 1$. Dropping the non-positive death term, using a localization argument, the assumption $\sup_K E(\langle X_0^K, 1\rangle^3) < +\infty$ and Gronwall's Lemma (as in the proof of (2.8)) lead to

$$\sup_K \mathbb{E}\big(\sup_{t\in[0,T]} \langle X_t^K, 1\rangle^3 \big) < \infty. \tag{5.8}$$

Step 3 Recall that M_F is endowed with the weak topology. To show the tightness of the sequence of laws $Q^K = \mathcal{L}(X^K)$ in $\mathcal{P}(\mathbb{D}([0,T], M_F))$, it suffices, following Roelly [39], to show that for any continuous bounded function f on $\mathbb{R}^d$, the sequence of laws of the processes $\langle X^K, f\rangle$ is tight in $\mathbb{D}([0,T],\mathbb{R})$. To this end, we use the Aldous criterion [1] and the Rebolledo criterion (see [25]). We have to show that

$$\sup_K \mathbb{E}\big(\sup_{t\in[0,T]} |\langle X^K_t, f\rangle|\big) < \infty, \tag{5.9}$$

and the Aldous condition respectively for the predictable quadratic variation of the martingale part and for the drift part of the semimartingales $\langle X^K, f\rangle$.
Since f is bounded, (5.9) is a consequence of (5.8): let us thus consider a couple (S, S') of stopping times satisfying a.s. $0 \le S \le S' \le S+\delta \le T$. By the assumptions on the parameters and (5.8), there exist constants $C, C' > 0$ such that

$$\mathbb{E}\left(\langle M^{K,f}\rangle_{S'} - \langle M^{K,f}\rangle_S\right) \le C\mathbb{E}\left(\int_S^{S+\delta} \left(\langle X^K_s, 1\rangle + \langle X^K_s, 1\rangle^2\right) ds\right) \le C'\delta.$$

In a similar way, the expectation of the finite variation part of $\langle X^K_{S'}, f\rangle - \langle X^K_S, f\rangle$ is bounded by $C'\delta$.
Hence, the sequence $(Q^K = \mathcal{L}(X^K))$ is tight in $\mathcal{P}(\mathbb{D}([0,T], M_F))$.

Step 4 Let us now denote by Q the weak limit in $\mathcal{P}(\mathbb{D}([0,T], M_F))$ of a subsequence of (Q^K), still denoted by (Q^K) for convenience. Let $X = (X_t)_{t\ge 0}$ a process with law Q. We remark that by construction, almost surely,

$$\sup_{t\in[0,T]} \sup_{f\in L^\infty(\mathbb{R}^d), ||f||_\infty \le 1} |\langle X^K_t, f\rangle - \langle X^K_{t-}, f\rangle| \le 1/K.$$

Since, for each f in a countable measure-determining set of continuous functions on $\mathcal{X}$, the mapping $\nu \mapsto \sup_{t\le T} |\langle \nu_t, f\rangle - \langle \nu_{t-}, f\rangle|$ is continuous on $\mathbb{D}([0,T], M_F)$, one deduces that Q only charges the continuous processes from $[0,T]$ into M_F.

Step 5 Let us now check that almost surely, the process X is the unique solution of (5.4). Thanks to (5.8), it satisfies $\sup_{t\in[0,T]}\langle X_t, 1\rangle < +\infty$ a.s.. Let f be continuous on $\mathcal{X}$.
For $\nu \in C([0,T], M_F)$, denote

$$\Psi_t(\nu) = \langle \nu_t, f\rangle - \langle \nu_0, f\rangle - \int_0^t \int_{\mathcal{X}} \Big((b(x) - d(x, C*\nu_s(x))) - p(x)b(x)\Big(\int_{\mathcal{X}} f(z)\, m(x,z)dz\Big)\Big)\nu_s(dx)ds. \tag{5.10}$$

Our aim is to show that

$$\mathbb{E}\left(|\Psi_t(X)|\right) = 0, \tag{5.11}$$

implying that X is solution of (5.4).
By (5.2), we know that for each K, $M^{K,f}_t = \Psi_t(X^K)$. Moreover, (5.8) implies that for each K,

$$\mathbb{E}\left(|M^{K,f}_t|^2\right) = \mathbb{E}\left(\langle M^{K,f}\rangle_t\right) \le \frac{C_f K^\eta}{K}\mathbb{E}\left(\int_0^t \left\{\langle X^K_s, 1\rangle + \langle X^K_s, 1\rangle^2\right\} ds\right) \le \frac{C_{f,T}K^\eta}{K}, \tag{5.12}$$

which goes to 0 as K tends to infinity, since $0 < \eta < 1$. Therefore,

$$\lim_K \mathbb{E}(|\Psi_t(X^K)|) = 0.$$

Since X is a.s. strongly continuous (for the weak topology) and since f is continuous and thanks to the continuity of the parameters, the functions Ψ_t is a.s. continuous at X. Furthermore, for any $\nu \in \mathbb{D}([0,T], M_F)$,

$$|\Psi_t(\nu)| \leq C_{f,T} \sup_{s\in[0,T]} \left(1 + \langle \nu_s, 1\rangle^2\right). \tag{5.13}$$

Hence the sequence $(\Psi_t(X^K))_K$ is uniformly integrable by (5.8) and thus

$$\lim_K \mathbb{E}\left(|\Psi_t(X^K)|\right) = \mathbb{E}\left(|\Psi_t(X)|\right) = 0. \tag{5.14}$$

□

Main Examples:

(1) A density case.

Proposition 5.3. *Suppose that ξ_0 is absolutely continuous with respect to Lebesgue measure. Then for all $t \geq 0$, ξ_t is absolutely continuous where respect to Lebesgue measure and $\xi_t(dx) = \xi_t(x)dx$, where $\xi_t(x)$ is the solution of the functional equation*

$$\partial_t \xi_t(x) = \left[(1-p(x))b(x) - d(x, C * \xi_t(x))\right]\xi_t(x) + \int_{\mathbb{R}^d} m(y,x)p(y)b(y)\xi_t(y)dy \tag{5.15}$$

for all $x \in \mathcal{X}$ and $t \geq 0$.

Some people refer to $\xi_t(.)$ as the population number density.

Proof. Consider a Borel set A of $\mathbb{R}^d$ with Lebesgue measure zero. A simple computation allows us to obtain, for all $t \geq 0$,

$$\langle \xi_t, \mathbf{1}_A\rangle \leq \langle \xi_0, \mathbf{1}_A\rangle) + \bar{b} \int_0^t \int_{\mathcal{X}} \mathbf{1}_A(x)\xi_s(dx)ds + \bar{b} \int_0^t \int_{\mathcal{X}} \int_{\mathcal{X}} \mathbf{1}_A(z)m(x,z)dz\xi_s(dx)ds.$$

By assumption, the first term on the RHS is zero. The third term is also zero, since for any $x \in \mathcal{X}$, $\int_{\mathcal{X}} \mathbf{1}_A(z)m(x,z)dz = 0$. Using Gronwall's Lemma, we conclude that $\langle \xi_t, \mathbf{1}_A\rangle = 0$ and then, the measure $\xi_t(dx)$ is absolutely continuous w.r.t. Lebesgue measure. One can easily prove from (5.4) that the density function $\xi_t(.)$ is solution of the functional equation (5.15). □

(2) The mean field case. The case of structured populations with $d(x, C * \xi(x)) = d + \alpha C * \xi(x)$ with constant rates b, d, α is meaningful, and has been developed in a spatial context where the kernel C describes the resources available (see for example [30]). In this context, (5.4) leads to the following equation on the total mass $n_t = \langle \xi_t, 1\rangle$:

$$\partial_t n_t = (b-d)n_t - \alpha \int_{\mathcal{X}\times\mathcal{X}} C(x-y)\xi_t(dx)\xi_t(dy). \tag{5.16}$$

This equation is not closed in $(n_t)_t$ and presents an unresolved hierarchy of nonlinearities. In the very particular case of uniform competition where $C \equiv 1$ (usually called "mean-field case"), there is a "decorrelative" effect and we recover the classical mean-field logistic equation of population growth:

$$\partial_t n_t = (b-d)n_t - \alpha n_t^2.$$

(3) Monomorphic and dimorphic cases with no mutation. We assume here that the mutation probability is $p = 0$. Without mutation, the trait space is fixed at time 0.

(a) Monomorphic case: All the individuals have the same trait x. Thus, we can write $X_0^K = n_0^K(x)\delta_x$, and then $X_t^K = n_t^K(x)\delta_x$ for any time t. In this case, Theorem 5.2 recasts into $n_t^K(x) \to n_t(x)$ with $\xi_t = n_t(x)\delta_x$, and (5.4) reads

$$\frac{d}{dt}n_t(x) = n_t(x)\big(b(x) - d(x, C(0)n_t(x))\big), \tag{5.17}$$

When $d(x, C * \xi(x)) = d + \alpha\, C * \xi(x)$, we recognize the logistic equation (3.9).

(b) Dimorphic case: The population consists in two subpopulations of individuals with traits x and y, i.e. $X_0^K = n_0^K(x)\delta_x + n_0^K(y)\delta_y$ and when K tends to infinity, the limit of X_t^K is given by $\xi_t = n_t(x)\delta_x + n_t(y)\delta_y$ satisfying (5.4), which recasts into the following system of coupled ordinary differential equations:

$$\begin{aligned} &\frac{d}{dt}n_t(x) = n_t(x)\big(b(x) - d(x, C(0)n_t(x) + C(x-y)n_t(y))\big) \\ &\frac{d}{dt}n_t(y) = n_t(y)\big(b(y) - d(y, C(0)n_t(y) + C(y-x)n_t(x))\big). \end{aligned} \tag{5.18}$$

When $d(x, C * \xi(x)) = d + \alpha\, C * \xi(x)$, we obtain the competitive Lotka-Volterra system defined in Section 3.2. In this case, we recall from Proposition 3.4 that there are in this case 4 equilibria: $(0,0)$, $(\bar{n}(x), 0)$, $(0, \bar{n}(y))$, and a non trivial equilibrium $(\bar{n}^1_{xy}, \bar{n}^2_{xy}) \in (\mathbb{R}_+^*)^2$.

In what follows, we will describe the invasion of a mutant trait y in a resident population with trait x. Immediately after its birth, the population's size issued from the mutant individual is almost zero and we can neglect it. We may define the *invasion fitness of the mutant trait y in the resident population with trait x* . This function approximates the mutant population growth rate at the beginning of its appearance in the resident population at equilibrium. It describes the ability of the mutant trait y to invade the resident population x and is given by

$$f(y; x) = b(y) - d(y) - C(y-x)\bar{n}(x), \tag{5.19}$$

where $\bar{n}(x)$ is the non trivial equilibrium of the logistic equation (5.17), given by

$$\bar{n}(x) = \frac{b(x) - d(x)}{C(0)}$$

.

Let us remark that $f(x; x) = 0$ and that f is actually not symmetric.

Using Proposition 3.4, we can characterize the stability of the equilibria of the system (5.18) thanks to the sign of the fitness function.

- $(\bar{n}(x), 0)$ is unstable if $f(y;x) > 0$ and stable si $f(y;x) < 0$.
- If $f(y;x) > 0$ and $f(x;y) < 0$, $(0, \bar{n}(y))$ is stable (fixation of the trait y).
- If $f(y;x) > 0$ and $f(x;y) > 0$, there is a nontrivial and stable equilibrium (coexistence of traits x and y).

6 Limit of rare mutations - Convergence to the trait substation sequence

In this part, we consider the scales of the adaptive dynamics: large population and small biomass, rare mutations and long mutation time scale.

We will assume that the ecological coefficients impede the coexistence of two traits and that there is a time scale separation between the ecological fast time scale in which the population comes back to equilibrium after competition and the mutation time scale which is much longer. Under these assumptions, the limiting process at the mutation time scale will be a jump process (see Fig.1 (d)), heuristically introduced in [38] and rigorously studied in [10], as Trait Substitution Sequence.

Assumption 6.1. *The initial population is composed of n_0^K individuals with $n_0^K \to \bar{n}(x_0)$, the nontrivial equilibrium of* (5.17) *with trait x_0.*

Assumption 6.2. *Invasion implies Fixation. Given $x \in \mathcal{X}$, a.s. any $z \in \mathcal{X}$ satisfies one of the two following conditions: either $f(z;x) < 0$, or $f(z;x) > 0$ and $f(x;z) > 0$.*

Let p_K be the individual mutation probability, going to 0 when $K \to \infty$). Therefore, the global population mutation rate is of order Kp_K, and $\frac{t}{Kp_K}$ will represent the mutation time scale.

We study the asymptotic behavior of $(\nu^K_{\frac{t}{Kp_K}}, t \geq 0)$, illustrated by Fig. 2(d).

Theorem 6.1. *Assume that Assumptions 4.1, 6.1 and 6.2 are fulfilled and assume moreover the following time scale separation: for all $C > 0$,*

$$\ln K \ll \frac{1}{Kp_K} \ll e^{CK}. \tag{6.1}$$

(where $f \ll g$ means $\frac{f}{g} \to 0$). Then the process $(\nu^K_{\frac{t}{Kp_K}}, t \geq 0)$ converges in the sense of finite marginals to a pure jump process $(\Lambda_t, t \geq 0)$,with values in

$$\mathcal{M}_0 = \left\{ \bar{n}(x)\delta_x \; ; \;\; \bar{n}(x) \text{ equilibrium of } (5.17) \right\},$$

and transitions from $\bar{n}(x)\delta_x$ to $\bar{n}(z)\delta_z$ at rate

$$p(x)b(x)\,\bar{n}(x)\,\frac{[f(z;x)]_+}{b(z)}\,m(x,z)dz. \tag{6.2}$$

The process $(\Lambda_t, t \geq 0)$ writes $\Lambda_t = \bar{n}_{X_t} \delta_{X_t}$. The process $(X_t, t \geq 0)$ (with $X_0 = x_0$) describes the support of $(\Lambda_t, t \geq 0)$. It is a pure jump Markov process with infinitesimal generator given by

$$A\varphi(x) = \int_{\mathcal{X}} (\varphi(z) - \varphi(x)) p(x) b(x) \, \bar{n}(x) \, \frac{[f(z;x)]_+}{b(z)} m(x,z) dz. \tag{6.3}$$

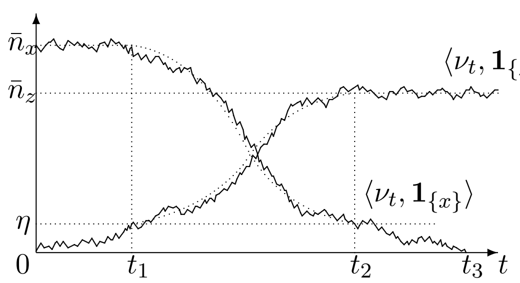
The process $(X_t, t \geq 0)$ is called the Trait Substitution Sequence (TSS).

Let us now give ideas of the proof of Theorem 6.1. For a detailed proof, we refer to [10].

The main idea is as follows. If mutations are rare, the selection has time to eliminate the deleterious traits or to fix advantageous traits before a new mutant arrives. The large population assumption allows us to approximate the birth and death dynamics between mutations by a deterministic Lotka-Volterra system which is more tractable. That will allow us to predict the outcome of the competition after a mutant arrival.

One describes the steps of the invasion of a mutant and the stabilization of the population which follows, with or without fixation of the mutant trait. Let us fix $\eta > 0$. Assume that at $t = 0$, the population is monomorphic with trait x and that Assumption 6.1 is satisfied. For t and K large enough, the density process $\langle \nu_t^K, \mathbf{1}_x \rangle$ belongs to a η-neighborhood of $\bar{n}(x)$ with large probability (cf. Theorem 5.2). We need the process to stay in this neighborhood before the first mutation occurs. We thus use a large deviations result for exit time of neighborhoods of $\bar{n}(x)$, stated in Freidlin-Wentzell [20] and Feng-Kurtz [18]: the time taken by the density process to leave the η-neighborhood of $\bar{n}(x)$ is larger than $\exp(CK)$, for some $C > 0$, with high probability. Hence, the first mutant will appear with large probability before the population exits the η–neighborhood of $\bar{n}(x)$, if one assumes that $\frac{1}{Kp_K} \ll \exp(CK)$.

Let us now study the invasion dynamics of the mutant with trait z. We will divide this event in three steps, as described in the figure below and developed in [10].



First step: Between 0 and t_1, the number of individuals with mutant trait z is small and the resident population's size is close to $\bar{n}(x)$. Thus the mutant dynamics is close to the one of a linear birth and death process with rates $b(z)$ and $d(z)+C(z-x)\bar{n}(x)$. Its growth rate approximatively equals the invasion fitness $f(z;x) = b(z) - d(z) - C(z-x)\bar{n}(x)$. If $f(z;x) > 0$, the birth and death process is supercritical, and therefore, for large K (see Section 2.4),

$$\begin{aligned}
&\mathbb{P}(\text{ the mutant population's size attains } \eta) \\
&\simeq \mathbb{P}(\text{the branching process attains } \ \eta K) \\
&\simeq \frac{[f(z;x)]_+}{b(z)} \quad (\text{ survival probability}).
\end{aligned}$$

Between t_1 and t_2, it is the competition step. When K increases, the density process $(\langle \nu_t^K, \mathbf{1}_x\rangle, \langle \nu_t^K, \mathbf{1}_z\rangle)$ tends to the solution of the Lotka-Volterra system (5.18). Thus the population process will attain with large probability a η-neighborhood of the unique globally asymptotically stable equilibrium n^* of (5.18) in time t_2, for small η.

The third step describes the stabilization of the population. If $f(z;x) > 0$, so that invasion of the mutant in the first step has a positive probability (as in the figure above), then our Assumption 6.2 implies that $f(x;z) < 0$, which means that $n^* = (0, \bar{n}_z)$ (see Proposition 3.4). As in the first step, we can approximate the density process for trait x is approximated by a subcritical birth and death process. Thus it will attain 0 in finite time and only individuals with trait z will remain in the population. Their density will stabilize around $\bar{n}(z) > 0$.

If the initial population is of order K, then the time taken for these three steps is of order $\ln K$, which is the order of the expectation of the extinction time for a birth and death process. Hence, if $\ln K \ll \frac{1}{Kp_K}$, with a large probability these three phases of competition-stabilization will happen before a next mutation occurs and we can reiterate the reasoning after every mutation event.

Thanks to this analysis, we obtain the TSS $(\Lambda_t, t \geq 0)$ which describes the successive stationary states and only keeps in its support the favorable mutations. It takes its values in $\mathcal{M}_0$. Let us assume that at some time t, $\Lambda_t = \bar{n}(x)\delta_x$. If the process belongs to a η-neighborhood of $\bar{n}(x)$, the mutation rate from an individual with trait x is close to $p_K p(x) b(x) K \bar{n}(x)$. Hence, in the time scale $\frac{t}{Kp_K}$, it is approximatively $b(x)p(x)\bar{n}(x)$. When a mutation occurs, the mutant trait z is chosen following $m(x,z)dz$. Its invasion probability is then approximatively the survival probability of the mutant with trait z in the resident population, given by $\frac{[f(z;x)]_+}{b(z)}$. In this case, the process will jump to $\bar{n}(z)\delta_z$. This explains Formula (6.2).

7 Canonical equation of the adaptive dynamics

Fig. 2(d) suggests that, for small mutation jumps, the TSS of Theorem 6.1 could be approximated by some continuous process. We will hence add an assumption of small mutations. In this section, we assume by simplicity that the trait space $\mathcal{X}$ is convex and symmetric. We introduce a mutation law $m(x,h)dh$, which is symmetric with bounded three-order moments and a parameter $\varepsilon > 0$ which will scale the mutation's size. Let H_ε be the function defined by $H_\varepsilon(h) = \varepsilon h$. The distribution of mutant traits from an individual with trait x is given by $m_\varepsilon(x, dh) = (m(x,h)dh) \circ H_\varepsilon^{-1}$.

Under the same assumptions as in the previous section and replacing m by m_ε in (6.3), we obtain the TSS X^ε. To observe a non trivial limit, we need to change the time scale t in $\frac{t}{\varepsilon^2}$.

Theorem 7.1. *The process $(X^\varepsilon_{\frac{t}{\varepsilon^2}}, t \geq 0)$ converges in law as $\varepsilon \to 0$, to the deterministic monomorphic process $t \to \bar{n}(x(t))\delta_{x(t)}$, where $x(.)$ is the unique solution of the ordinary differential equation*

$$\frac{dx}{dt} = \frac{1}{2} p(x)\bar{n}(x)\partial_1 f(x;x) \int_{\mathbb{R}} h^2 m(x,h)dh. \tag{7.1}$$

This equation has been heuristically introduced by Dieckmann and Law [14] and is called *canonical equation of the adaptive dynamics.* When the mutation law m is not symmetric, (7.1) involves the whole measure m, instead of its variance.

Idea of the proof. It is again based on a compactness-uniqueness argument. The process $(X^\varepsilon_{\frac{t}{\varepsilon^2}}, t \geq 0)$ has the generator

$$L^\varepsilon\varphi(x) = \frac{1}{\varepsilon^2}\int_{\mathbb{R}}(\varphi(x+\varepsilon h) - \varphi(x))p(x)b(x)\,\bar{n}(x)\,\frac{[f(x+\varepsilon h;x)]_+}{b(x+\varepsilon h)}m(x,h)dh. \qquad (7.2)$$

Its uniqueness is obtained by a standard theorem (boundedness of the coefficients). As $f(x;x) = 0$, and by an expansion of $\varphi(x+\varepsilon h)$ and $f(x+\varepsilon h;x)$ at order 2 around $\varepsilon = 0$, one can show that

$$L^\varepsilon\varphi(x) \underset{\varepsilon\to 0}{\longrightarrow} \frac{1}{2}p(x)\bar{n}(x)\partial_1 f(x;x)\varphi'(x)\int_{\mathbb{R}} h^2 m(x,h)dh. \qquad (7.3)$$

The process $(X^\varepsilon_{\frac{t}{\varepsilon^2}}, t \geq 0)$ is a semimartingale. Uniform tightness (in ε) of the laws of $(X^\varepsilon_{\frac{\cdot}{\varepsilon^2}})$ comes from uniform estimates for the martingale part and for the total variation part of its decomposition. The characterization of the limiting martingale problem is deduced from (7.3).

8 Appendix : Poisson point measures

In this appendix, we summarize the main definitions and results concerning the Poisson point measures. The reader can consult the two main books by Ikeda-Watanabe [23] and by Jacod-Shiryaev [24] for more details.

Definition 8.1. *Let $(E,\mathcal{E})$ be a measurable space and μ a σ-finite measure on this space. A (homogeneous) Poisson point measure N with intensity $\mu(dh)dt$ on $\mathbb{R}_+ \times E$ is a $(\mathbb{R}_+ \times E, \mathcal{B}(\mathbb{R}_+)\otimes\mathcal{E})$-random measure defined on a probability space $(\Omega,\mathcal{F},\mathbb{P})$ which satisfies the following properties:*

1. *N is a counting measure: $\forall \widehat{A} \in \mathcal{B}(\mathbb{R}_+)\otimes\mathcal{E}$, $\forall\omega\in\Omega$, $N(\omega,\widehat{A}) \in \mathbb{N}\cup\{+\infty\}$.*
2. *$\forall\omega\in\Omega$, $N(\omega,\{0\}\times E) = 0$: no jump at time 0.*
3. *$\forall\widehat{A}\in\mathcal{B}(\mathbb{R}_+)\otimes\mathcal{E}$, $\mathbb{E}(N(\widehat{A})) = \nu(\widehat{A})$, where $\nu(dt,dh) = \mu(dh)dt$.*
4. *If $\widehat{A}$ and $\widehat{B}$ are disjoint in $\mathcal{B}(\mathbb{R}_+)\otimes\mathcal{E}$ and if $\nu(\widehat{A}) < +\infty, \nu(\hat{B}) < +\infty$, then the random variables $N(\widehat{A})$ and $N(\widehat{B})$ are independent.*

The existence of such a Poisson point measure with intensity $\mu(dh)dt$ is proven in [24], for any σ-finite measure μ on $(E,\mathcal{E})$.

Let us remark that for any $A\in\mathcal{E}$ with $\mu(A) < \infty$ the process defined by

$$N_t(A) = N((0,t]\times A)$$

is a Poisson process with intensity $\mu(A)$.

Definition 8.2. *The filtration $(\mathcal{F}_t)_t$ generated by N is given by*

$$\mathcal{F}_t = \sigma(N((0,s] \times A), \forall s \leq t, \forall A \in \mathcal{E}).$$

If $\widehat{A} \in (s,t] \times \mathcal{E}$ and $\nu(\widehat{A}) < \infty$, then $N(\widehat{A})$ is independent of $\mathcal{F}_s$.

Let us first assume that the measure μ is finite on $(E, \mathcal{E})$. Then $(N_t(E), t \geq 0)$ is a Poisson process with intensity $\mu(E)$. The point measure is associated with a compound Poisson process. Indeed, let us write

$$\mu(dh) = \mu(E) \, \frac{\mu(dh)}{\mu(E)},$$

the decomposition of the measure μ as the product of the jump rate $\mu(E)$ and the jump amplitude law $\frac{\mu(dh)}{\mu(E)}$. Let us fix $T > 0$ and introduce $T_1, \ldots, T_\gamma$ the jump times of the process $(N_t(E), t \geq 0)$ between 0 and T. We know that the jump number γ is a Poisson variable with parameter $T\mu(E)$. Moreover, conditionally on γ, $T_1, \ldots, T_\gamma$, the jumps $(U_n)_{n=1,\ldots,\gamma}$ are independent with the same law $\frac{\mu(dh)}{\mu(E)}$. We can write in this case

$$N(dt, dh) = \sum_{n=1}^{\gamma} \delta_{(T_n, U_n)}.$$

Therefore, one can define for any measurable function $G(\omega, s, h)$ defined on $\Omega \times \mathbb{R}_+ \times E$ the random variable

$$\int_0^T \int_E G(\omega, s, h) N(\omega, ds, dh) = \sum_{n=1}^{\gamma} G(\omega, T_n, U_n).$$

In the following, we will forget the ω. Let us remark that $T \longrightarrow \int_0^T \int_E G(s,h) N(ds, dh)$ is a finite variation process which is increasing if G is positive. A main example is the case where $G(\omega, s, h) = h$. Then

$$X_T = \int_0^T \int_E h \, N(ds, dh) = \sum_{n=1}^{\gamma} U_n = \sum_{s \leq T} \Delta X_s$$

is the sum of the jumps between 0 and T.

Our aim now is to generalize the definition of the integral of G with respect to N when $\mu(E) = +\infty$. In this case, one can have an accumulation of jumps during the finite time interval $[0,T]$ and the counting measure N is associated with a countable set of points:

$$N = \sum_{n \geq 1} \delta_{(T_n, U_n)}.$$

We need additional properties on the process G.
Since μ is σ-finite, there exists an increasing sequence $(E_p)_{p \in \mathbb{N}}$ of subsets of E such that $\mu(E_p) < \infty$ for each p and $E = \cup_p E_p$. As before we can define $\int_0^T \int_{E_p} G(s,h) N(ds, dh)$ for any p.

We introduce the predictable σ-field $\mathcal{P}$ on $\Omega \times \mathbb{R}_+$ (generated by all left-continuous adapted processes) and define a predictable process $(G(s,h), s \in \mathbb{R}_+, h \in E)$ as a $\mathcal{P} \otimes \mathcal{E}$ measurable process.

Theorem 8.3. *Let us consider a predictable process G(s, h) and assume that*

$$\mathbb{E}\left(\int_0^T \int_E |G(s,h)|\mu(dh)ds\right) < +\infty. \tag{8.1}$$

1) The sequence of random variables $\left(\int_0^T \int_{E_p} G(s,h)N(ds,dh)\right)_p$ *is Cauchy in* $\mathbb{L}^1$ *and converges to a* $\mathbb{L}^1$*-random variable that we denote by* $\int_0^T \int_E G(s,h)N(ds,dh)$*. It's an increasing process if* G *is non-negative. Moreover, we get*

$$\mathbb{E}\left(\int_0^T \int_E G(s,h)N(ds,dh)\right) = \mathbb{E}\left(\int_0^T \int_E G(s,h)\mu(dh)ds\right)$$

2) The process $M = (\int_0^t \int_E G(s,h)N(ds,dh) - \int_0^t \int_E G(s,h)\mu(dh)ds, t \leq T)$ *is a martingale.*
The random measure

$$\widetilde{N}(ds,dh) = N(ds,dh) - \mu(dh)ds$$

is called the compensated martingale-measure of N*.*

3) If we assume moreover that

$$\mathbb{E}\left(\int_0^T \int_E G^2(s,h)\mu(dh)ds\right) < +\infty, \tag{8.2}$$

then the martingale M *is square-integrable with quadratic variation*

$$\langle M\rangle_t = \int_0^t \int_E G^2(s,h)\mu(dh)ds.$$

Let us remark that when (8.1) holds, the random integral $\int_0^t \int_E G(s,h)N(ds,dh)$ can be defined without the predictability assumption on H but the martingale property of the stochastic integral $\int_0^t \int_E G(s,h)\widetilde{N}(ds,dh)$ is only true under this assumption.

We can improve the condition under which the martingale (M_t) can be defined. The proof of the next theorem is tricky and consists in studying the $\mathbb{L}^2$-limit of the sequence of martingales $\int_0^t \int_{E_p} G(s,h)\widetilde{N}(ds,dh)$ as p tends to infinity. Once again, this sequence is Cauchy in $\mathbb{L}^2$ and converges to a limit which is a square-integrable martingale. Let us recall that the quadratic variation of a square-integrable martingale M is the unique predictable process $\langle M\rangle$ such that $M^2 - \langle M\rangle$ is a martingale.

Theorem 8.4. *Let us consider a predictable process G(s, h) satisfying* (8.2)*. Then the process* $M = (\int_0^t \int_E G(s,h)\widetilde{N}(ds,dh), t \leq T)$ *is a square-integrable martingale with quadratic variation*

$$\langle M\rangle_t = \int_0^t \int_E G^2(s,h)\mu(dh)ds.$$

If (8.2) is satisfied but not (8.1), the definition of M comes from a $\mathbb{L}^2$- limiting argument, as for the usual stochastic integrals. In this case the quantity $\int_0^t \int_E G(s,h)N(ds,dh)$ isn't always well defined and we are obliged to compensate.

Example: Let $\alpha \in (0,2)$. A symmetric α-stable process S can be written

$$S_t = \int_0^t \int_{\mathbb{R}} h \mathbf{1}_{\{0<|h|<1\}} \widetilde{N}(ds, dh) + \int_0^t \int_{\mathbb{R}} h \mathbf{1}_{\{|h|\geq 1\}} N(ds, dh), \tag{8.3}$$

where $N(ds, dh)$ is a Poisson point measure with intensity $\mu(dh)ds = \frac{1}{|h|^{1+\alpha}} dhds$. There is an accumulation of small jumps and the first term in the r.h.s. of (8.3) is defined as a compensated martingale. The second term corresponds to the big jumps, which are in finite number on any finite time interval.
If $\alpha \in (1,2)$, then $\int h \wedge h^2 \mu(dh) < \infty$ and the process is integrable. If $\alpha \in (0,1)$, we only have that $\int 1 \wedge h^2 \mu(dh) < \infty$ and the integrability of the process can fail.

Let us now consider a stochastic differential equation driven both by a Brownian term and a Poisson point measure. We consider a random variable X_0, a Brownian motion B and a Poisson point measure $N(ds, dh)$ on $\mathbb{R}_+ \times \mathbb{R}$ with intensity $\mu(dh)ds$. Let us fix some measurable functions b and σ on $\mathbb{R}$ and $G(x,h)$ and $K(x,h)$ on $\mathbb{R} \times \mathbb{R}$.

We consider a process $X \in \mathbb{D}(\mathbb{R}_+, \mathbb{R})$ such that for any $t > 0$,

$$\begin{aligned} X_t &= X_0 + \int_0^t b(X_s)ds + \int_0^t \sigma(X_s)dB_s \\ &\quad + \int_0^t \int_{\mathbb{R}} G(X_{s-}, h)N(ds, dh) + \int_0^t \int_{\mathbb{R}} K(X_{s-}, h)\widetilde{N}(ds, dh). \end{aligned} \tag{8.4}$$

To give a sense to the equation, one expects that for any $T > 0$,

$$\mathbb{E}\left(\int_0^T \int_{\mathbb{R}} |G(X_s, h)|\mu(dh)ds\right) < +\infty \ ; \ \mathbb{E}\left(\int_0^T \int_{\mathbb{R}} K^2(X_s, h)\mu(dh)ds\right) < +\infty.$$

We refer to [23] Chapter IV-9 for general existence and uniqueness assumptions (generalizing the Lipschitz continuity assumptions asked in the case without jump).

Let us assume that a solution of (8.4) exists. The process X is a left-limited and right-continuous semimartingale. A standard question is to ask when the process $f(X_t)$ is a semimartingale and to know its Doob-Meyer decomposition. For a smooth function f, there is an Itô's formula generalizing the usual one stated for continuous semimartingales.

Recall (cf. Dellacherie-Meyer VIII-25 [13]) that for a function $a(t)$ with bounded variation, the change of variable formula gives that for a C^1-function f,

$$f(a(t)) = f(a(0)) + \int_{(0,t]} f'(a(s))da(s) + \sum_{0<s\leq t} (f(a(s) - f(a(s^-) - \Delta a(s) f'(a(s^-)).$$

We wish to replace a by a semimartingale. We have to add smoothness to f and we will get two additional terms in the formula because of the two martingale terms. As in the continuous case, we assume that the function f is C^2.

Theorem 8.5. *(see [23] Theorem 5.1 in Chapter II). Let f a C^2-function. Then $f(X)$ is a semimartingale and for any t,*

$$
\begin{aligned}
f(X_t) = f(X_0) &+ \int_0^t f'(X_s)b(X_s)ds + \int_0^t f'(X_s)\sigma(X_s)dB_s + \frac{1}{2}\int_0^t f''(X_s)\sigma^2(X_s)ds \\
&+ \int_0^t \int_{\mathbb{R}} (f(X_{s-} + G(X_{s-}, h)) - f(X_{s-}))N(ds, dh) \\
&+ \int_0^t \int_{\mathbb{R}} (f(X_{s-} + K(X_{s-}, h)) - f(X_{s-}))\widetilde{N}(ds, dh) \\
&+ \int_0^t \int_{\mathbb{R}} \big(f(X_s + K(X_s, h)) - f(X_s) - K(X_s, h)f'(X_s)\big)\,\mu(dh)ds. \qquad (8.5)
\end{aligned}
$$

Corollary 8.6. *Under suitable integrability and regularity conditions on b, σ, G, K and μ, the process X is a Markov process with extended generator: for any C^2-function f, for $x \in \mathbb{R}$,*

$$
\begin{aligned}
Lf(x) &= b(x)f'(x) + \frac{1}{2}\sigma^2(x)\mathcal{F}''(x) + \int_{\mathbb{R}} (f(x + G(x, h)) - f(x))\,\mu(dh) \\
&\quad + \int_{\mathbb{R}} \big(f(x + K(x, h)) - f(x) - K(x, h)f'(x)\big)\,\mu(dh). \qquad (8.6)
\end{aligned}
$$

Example: let us study the case where

$$X_t = X_0 + \int_0^t b(X_s)ds + \int_0^t \sigma(X_s)dB_s + S_t,$$

where S is the stable process introduced in (8.3). Let us consider a C^2-function f. Then $f(X)$ is a semimartingale and writes

$$
\begin{aligned}
f(X_t) &= f(X_0) + M_t + \int_0^t f'(X_s)b(X_s)ds + \frac{1}{2}\int_0^t f''(X_s)\sigma^2(X_s)ds \\
&\quad + \int_0^t \int_{\mathbb{R}} (f(X_{s-} + h\mathbf{1}_{\{|h|>1\}}) - f(X_{s-}))\frac{1}{|h|^{1+\alpha}}dhds \\
&\quad + \int_0^t \int_{\mathbb{R}} \big(f(X_{s-} + h\mathbf{1}_{\{|h|\le 1\}}) - f(X_{s-}) - h\mathbf{1}_{\{|h|\le 1\}}f'(X_{s-})\big)\,\frac{1}{|h|^{1+\alpha}}dhds \\
&= f(X_0) + M_t + \int_0^t f'(X_s)b(X_s)ds + \frac{1}{2}\int_0^t f''(X_s)\sigma^2(X_s)ds \\
&\quad + \int_0^t \int_{\mathbb{R}} \big(f(X_{s-} + h) - f(X_{s-}) - h\mathbf{1}_{\{|h|\le 1\}}f'(X_{s-})\big)\,\frac{1}{|h|^{1+\alpha}}dhds,
\end{aligned}
$$

where M is a martingale.

Let us come back to the general case and apply Itô's formula (8.7) to $f(x) = x^2$:

$$\begin{aligned} X_t^2 = X_0^2 &+ \int_0^t 2X_s b(X_s)ds + \int_0^t 2X_{s-}\sigma(X_{s-})dB_s + \int_0^t \sigma^2(X_s)ds \\ &+ \int_0^t \int_{\mathbb{R}} (2X_{s-}G(X_{s-},h) + (G(X_{s-},h))^2)N(ds,dh) \\ &+ \int_0^t \int_{\mathbb{R}} (2X_{s-}K(X_{s-},h) + (K(X_{s-},h))^2)\widetilde{N}(ds,dh) \\ &+ \int_0^t \int_{\mathbb{R}} (K(X_{s-},h))^2 \mu(dh)ds. \end{aligned} \tag{8.7}$$

In the other hand, since

$$X_t = X_0 + M_t + A_t,$$

where M is square-integrable and A has finite variation, then

$$X_t^2 = X_0^2 + N_t + \int_0^t 2X_{s-}dA_s + \langle M \rangle_t.$$

Doob-Meyer's decomposition allows us to identify the martingale parts and the finite variation parts in the two previous decompositions and therefore

$$\langle M \rangle_t \quad = \quad \int_0^t \sigma^2(X_s)ds + \int_0^t \int_{\mathbb{R}} (G^2(X_{s-},h) + K^2(X_{s-},h))\mu(dh)ds.$$

References

[1] D. Aldous. Stopping times and tightness. *Ann. Probab.* **6**, 335–340, 1978.

[2] L. J. S. Allen. *An introduction to stochastic processes with applications to biology.* CRC Press, Boca Raton, FL, second edition, 2011.

[3] W. J. Anderson. *Continuous-time Markov chains.* Springer Series in Statistics: Probability and its Applications. Springer-Verlag, New York, 1991. An applications-oriented approach.

[4] V. Bansaye, S. Méléard, M. Richard. How do birth and death processes come down from infinity? *Preprint avialable via http://arxiv.org/abs/1310.7402*, 2014.

[5] B. Bolker, S.W. Pacala. Using moment equations to understand stochastically driven spatial pattern formation in ecological systems. *Theor. Pop. Biol.* **52**, 179–197, 1997.

[6] B. Bolker, S.W. Pacala. Spatial moment equations for plant competition: Understanding spatial strategies and the advantages of short dispersal. *Am. Nat.* **153**, 575–602, 1999.

[7] P. Cattiaux and S. Méléard. Competitive or weak cooperative stochastic Lotka-Volterra systems conditioned on non-extinction. J. Math. Biology 6, 797–829, 2010.

[8] P. Cattiaux, P. Collet, A. Lambert, S. Martinez, S. Méléard, and J. San Martin. Quasi-stationary distributions and diffusion models in population dynamics. *Ann. Probab.*, 37(5):1926–1969, 2009.

[9] N. Champagnat, R. Ferrière and S. Méléard : Unifying evolutionary dynamics: From individual stochastic processes to macroscopic models. Theor. Pop. Biol. **69**, 297–321, 2006.

[10] Champagnat, N.: A microscopic interpretation for adaptive dynamics trait substitution sequence models. Stochastic Process. Appl. 116 (2006), no. 8, 1127–1160 .

[11] N. Champagnat and S. Méléard. Invasion and adaptive evolution for individual-based spatially structured populations. *Journal of Mathematical Biology* **55**, 147–188, 2007.

[12] N. Champagnat and D. Villemonais Exponential convergence to quasi-stationary distribution and Q-process. *Probability Theory and Related Fields* **164(1)**, 243–283, 2016.

[13] C. Dellacherie and P.A. Meyer. *Probabilités et potentiel - Théorie des martingales.* Hermann, 1985.

[14] Dieckmann, U., Law, R.: The dynamical theory of coevolution: A derivation from stochastic ecological processes. J. Math. Biol. **34**, 579–612 (1996).

[15] P. Donnelly. Weak convergence to a Markov chain with an entrance boundary: ancestral processes in population genetics. *Ann. Probab.*, 19(3):1102–1117, 1991.

[16] A. Etheridge : Survival and extinction in a locally regulated population. *Ann. Appl. Probab.* **14**, 188–214, 2004.

[17] S. N. Ethier and T. G. Kurtz. *Markov processes: characterization and convergence.* Wiley, 1986.

[18] J. Feng and T. G. Kurtz. *Large deviations for stochastic processes.* Mathematical Surveys and Monographs, 131. American Mathematical Society, Providence, 2006.

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

[20] Freidlin, M.I., Wentzel, A.D.: Random Perturbations of Dynamical Systems. Springer-Verlag, Berlin, (1984).

[21] A. Grimvall. On the convergence of sequences of branching processes. *Ann. Probability*, 2:1027–1045, 1974.

[22] J. Hofbauer and K. Sigmund. *Evolutionary Games and Population Dynamics.* Cambridge Univ. Press (2002).

[23] N. Ikeda and S. Watanabe. *Stochastic differential equations and diffusion processes, 2nd ed.* North-Holland, 1989.

[24] J. Jacod and A.N. Shiryaev *Limit Theorems for Stochastic Processes.* Springer, 1987.

[25] A. Joffe and M. Métivier. Weak convergence of sequences of semimartingales with applications to multitype branching processes. *Adv. Appl. Probab.* **18**, 20–65, 2012.

[26] B. Jourdain, S. Méléard ad W. Woyczynsky. Lévy flights in evolutionary ecology. *J. Math. Biol.* **65**, 677–707, 1986.

[27] I. Karatzas and S.E. Shreve. *Brownian Motion and Stochastic Calculus, 2nd ed.* Springer, 1998.

[28] S. Karlin and J. L. McGregor. The differential equations of birth-and-death processes, and the Stieltjes moment problem. *Trans. Amer. Math. Soc.*, 85:489–546, 1957.

[29] S. Karlin and H. M. Taylor. *A first course in stochastic processes.* Academic Press [A subsidiary of Harcourt Brace Jovanovich, Publishers], New York-London, second edition, 1975.

[30] E. Kisdi. Evolutionary branching under asymmetric competition. J. Theor. Biol. **197**, 149–162, 1999.

[31] T. G. Kurtz. Diffusion approximations for branching processes. In *Branching processes (Conf., Saint Hippolyte, Que., 1976)*, volume 5 of *Adv. Probab. Related Topics*, pages 269–292. Dekker, New York, 1978.

[32] A. Lambert. The branching process with logistic growth. *Ann. Appl. Probab.*, 15 no.2, 150–1535, 2005.

[33] A. Lambert. Population dynamics and random genealogies. *Stoch. Models*, 24 (suppl. 1), 45–163, 2008.

[34] R. Law, D. J. Murrell and U. Dieckmann. Population growth in space and time: Spatial logistic equations. *Ecology* **84**, 252–262, 2003.

[35] S. Martinez, J. San Martin, and D. Villemonais. Existence and uniqueness of a quasi-stationary distribution for Markov processes with fast return from infinity. *J. Appl. Probab.*, **51(3)**, 756–768, 2014.

[36] S. Méléard and S. Roelly. Sur les convergences étroite ou vague de processus à valeurs mesures. *C. R. Acad. Sci. Paris Sér. I Math.* **317**, 785–788, 1993.

[37] J.A.J. Metz, R.M. Nisbet and S.A.H. Geritz. How should we define fitness for general ecological scenarios. Trends Ecol. Evol. **7**, 198–202, 1992.

[38] J.A.J. Metz, S.A.H. Geritz, G. Meszeena, F.A.J. Jacobs, J. S. van Heerwaarden. Adaptive Dynamics, a geometrical study of the consequences of nearly faithful reproduction. Pages 183-231 in Stochastic and Spatial Structures of Dynamical Systems (S.J. van Strien, S.M. Verduyn Lunel, editors). North Holland, Amsterdam, 1996.

[39] S. Roelly-Coppoletta. A criterion of convergence of measure-valued processes: application to measure branching processes. *Stoch. Stoch. Rep.* **17**, 43–65, 1986.

[40] E. A. van Doorn. Quasi-stationary distributions and convergence to quasi-stationarity of birth-death processes. *Adv. in Appl. Probab.*, 23(4):683–700, 1991.