A pathwise construction of Birth-Death-Swap systems leading to an averaging result in the presence of two timescales

Sarah Kaakai ∗ Nicole El Karoui †

July 31, 2018

Abstract
This paper deals with the stochastic modeling of a general class of heterogeneous population dynamics structured by discrete subgroups. Such processes generalize classical multitype Birth-Death processes by allowing swap events, i.e. transfers from one subgroup to another. The variability of the environment is also included and the population evolution is not Markovian. We propose a new representation of the population based on its jump measure, characterized as a multivariate counting process with specific support conditions, and which together with the population defines a Birth-Death-Swap (BDS) system. We first prove a general result, on the construction by strong domination of multivariate counting processes solutions of stochastic differential equations driven by extended Poisson measures. Under weaker assumptions than usual, the existence of BDS systems strongly dominated by a Cox-Birth process is obtained. This pathwise comparison is the main tool to obtain tightness results in the second part of the paper. The BDS system in the presence of two timescales is then studied, when swap events occur at a faster timescale than demographic events. A general averaging result for the demographic counting process is proven. Classical averaging results obtained in the Markov case cannot be applied here, and in order to overcome this difficulty, we rely on the stable convergence of processes involved. This mode of convergence is particularly well-suited to our general framework. At the limit, the aggregated population is reduced to a one dimensional Birth-Death process with averaged intensities, resulting from a non-trivial aggregation of the subgroups birth and death intensities.

Key-words: population dynamics, birth-death-swap, point processes, stochastic intensity, SDE driven by Poisson measures, strong comparison, averaging, two timescales, stable convergence

Introduction
This paper was originally motivated by recent developments in the evolution of human populations. Indeed, the past few years have been marked by a renewed demand for more efficient models, based on recent observations of demographic trends that seem to indicate a shift of

∗CMAP, École Polytechnique, route de Saclay, 91128 Palaiseau, France; E-mail: sarah.kaakai@polytechnique.edu
†LPSM, Université Pierre et Marie Curie, 4 Place Jussieu, 75005 Paris, France; Email: nicole.el_karoui@upmc.fr
paradigm over the past decades, toward a more complex and individualized world. In particular, a growing number of studies has reported diverging trends in longevity at several levels, marked by a widening of geographical and socioeconomic inequalities in mortality (Elo (2009), National Research Council (2011), Olshansky et al. (2012), El Karoui et al. (2018)).

Within this new heterogeneous framework, interpretation of empirical measurements made on different subgroups and at different points in time is much more complex, due in part to non-observable changes to populations composition over time (Dowd and Hamoudi (2014)). Furthermore, the behavior of individuals is influenced by interactions with others, resulting in the introduction of *non-linearity* or *density dependence* in the population dynamics. Thus, changes to the population composition add significant complexity to the population dynamics, in a context where demographic rates are already strongly time-inhomogeneous (reduction of mortality, fertility rates) and subject to random changes due to the variability of the global environment (public health, economic growth, medical breakthrough, pollution...).

One of the great challenges posed by this rising heterogeneity is to understand how these changes affect the population and its longevity on an aggregated or macroscopic level. Indeed, the dynamic modeling of the population heterogeneity can produce unexpected or counter-intuitive effects at the aggregated level, that cannot be directly modeled by traditional “macro” models.

At the same time, the recent years have seen a renewed interest in the modeling of heterogeneous population dynamics, motivated in part by the study ecological systems. In the deterministic literature, Auger et al. (2012) deals with population dynamics of communities living on a patchy-environment with migrations between patches, and study the aggregated dynamics produced by such models. Auger et al. (2000) considers a population of fishes divided in two spatial patches. The first one contains food and resources, but is also exposed to parasites attacks, while the second one is safe from parasitism but does not provide enough resources. Other examples are given in Auger et al. (2008). In Fournier and Méléard (2004), the authors introduce a stochastic spatial ecological system, in which the demographic behavior of individuals (birth and death rates) is determined by their location. A pathwise representation of the population dynamics is given using stochastic differential equations (SDEs) driven by Poisson measures, but individuals cannot change of location. In a recent paper focusing of horizontal transfers of biological information, Billiard et al. (2016) included changes of characteristics in their two populations model.

Usually, the temporal evolution of such processes is assumed to be Markovian. In the classical framework of Markov multi-type *Birth-Death* processes, heterogeneous populations are described by demographic events, that is by births (or entry) and deaths occurring in the subgroups. In this paper, moves of individuals from one subgroup to another, called *swap events*, are also taken into account. Furthermore, the population evolution is not assumed to be Markovian, in order to take into account additional randomness expressing the variability of the environment, as well as time-inhomogeneity (for instance the reduction of the mortality intensity over time). This general class of heterogeneous population dynamics is thus called a *Birth-Death-Swap (BDS)* system. To the best of our knowledge, the terminology Birth-Death-Swap has been introduced by Huber (2012) for the purpose of simulating stationary distribution of BDS systems. Due to the complexity already introduced by swap events, the age-structure of the population is not taken into account in the modeling. In this sense, the model does not focus particularly on
human populations. These features have led us to adopt a point of view different than usual. The population is represented by its jumps measure, called jumps counting process, defined as a multivariate counting process on a larger state space and verifying specific support conditions. This representation allows us to obtain an existence result for BDS systems under weak assumptions, by realizing jumps counting processes by strong domination with Cox-Birth processes. This result derives from a more general result for multivariate counting processes, on the realization by strong domination of solutions of SDEs driven by extended Poisson measures. This viewpoint has also proven to be very effective to obtain tightness results in the second part, and allows us to relax usual uniform integrability assumptions.

The complexity generated by the presence of swap events makes it difficult to apprehend the population dynamic directly on an aggregated level. Aggregation methods, which are important in various fields, provide a better understanding of the link between finer-grained dynamics and the aggregated variables they produce. Moreover, the complexity of the studied systems is often reduced in the presence of a timescale separation. For instance, in the field of operation research Yin and Zhang have obtained approximation results for continuous time Markov chains with two timescales, using asymptotic expansions based on the singular perturbation theory (Yin and Zhang (2004), Yin and Zhang (2012)). In Taylor and Véber (2009), the genealogy of a subdivided population that experiences sporadic mass extinction events is studied. In other works, the population is studied in the presence of two timescales after being renormalized (Méléard and Tran (2012), Billiard et al. (2016)).

In the second part of this paper, the BDS system is studied in the presence of a separation of timescale between swap and demographic events: changes to the population composition generated by swap events are fast in comparison with the demographic timescale. For instance, changes in the social and/or geographical structure of the population can be fast in comparison with the demographic timescale. In Auger et al. (2012), Marvá et al. (2013), migrations between different patches or changes of strategies occur at a much faster timescale than demographic events. An averaging result is proven for the multivariate process counting the number of demographic events. As a consequence, the aggregated population dynamics can be reduced to a one dimensional Birth-Death process with averaged intensities: due to swap events, nonlinearities emerge at the aggregated level.

Classical averaging results such as Kurtz (1992) or Yin and Zhang (2012) cannot be applied, since time-inhomogeneity and additional randomness are taken into account. In order to overcome this difficulty, we strongly rely on the stable convergence of concerned processes. This mode of convergence, which extends the convergence in distribution, is very efficient in order to obtain identification results and particularly well-suited to our general framework, by allowing for the randomness of the initial probabilistic structure to be taken into account.

The general model is presented in Section 1. An algebraic decomposition of the population is presented, based on the jumps counting process. BDS systems are then defined by introducing some constraints on the jumps counting processes intensity functional.

Sections 2 and 3 are dedicated to the pathwise representation of the jumps counting process, as the solution of a multivariate SDE driven by an extended Poisson measure. In Section 2 a
result is proven on the construction of multivariate counting processes by strong domination. The existence of BDS systems is then obtained as a corollary in Section 3. At the end of Section 3 an alternative construction of BDS system is presented, more adapted to the simulation, and called the Birth Death Swap decomposition algorithm.

In Section 4 the BDS system is studied when swap events occur at a faster timescale than demographic events. The main result of the section is a general identification result for the demographic counting processes. At the limit, intensities of demographic events are averaged against stable limits of the population, which is seen on a well-adapted space.

In Section 5 the averaging result is applied in in the particular case of deterministic swap intensities, but with general birth and death intensities. In particular, we show that the aggregated population converges to a one dimensional Birth-Death process, whose birth and death intensities have been averaged against stationary distribution of pure Swap processes.

1 Birth Death Swap systems

We describe the evolution of a general class of stochastic heterogeneous population dynamics, in which individuals differ by a finite number of characteristics. The population is structured by discrete subgroups indexed by \(1 \leq p \leq \), with individuals in the same subgroup sharing the same set of characteristics. For instance, subgroups can describe individuals living in the same patch or neighborhood, with the same level of income, similar behavior (level of aggressiveness for non-human populations, smoking status, eating habits...) or all the above.

At a given time, the quantity of interest is the vector of the number of individuals in each subgroup, called the state of the population. Random changes in the population occur at random dates and are described by the move of a single individual at a time. Different types of events can modify the composition of the population. A demographic event is a birth (or entry) or death in a given subgroup. A swap event happens when an individual changes of characteristics, resulting in a move from one subgroup to another. The description of the model differ from Individual-Based Models which focus on describing what happens to each individual in the population, and can be seen as an intermediary scale between the individual and macroscopic level. Moreover, this formal description, only based on changes in the population composition allow us to include a large class of dynamics in the model.

The assumption that individuals in the population differ only by a finite number of characteristics is introduced for ease of exposition, and could be extended to a countable set of characteristics.

1.1 Setup

**Population state space** The \(p\) subgroups composing the population are indexed by the set \(I_p = \{1, 2, \ldots, p\}\). The state of the population is described by the number of individuals in the each subgroup, that is by a vector in \(\mathbb{N}^p\).

**Elementary events representation** In this description, the emphasis is on events that changes the composition of the population, rather than on mechanisms or individuals responsible for these events, which are not specified here. We also give a unified description of swap and demographic
events, based on the introduction of an additional “infinite” population.

a) A swap event from \( i \) to \( j \) (\( j \neq i \)) is a move of an individual from subgroup \( i \) to the subgroup \( j \), and can only happen if the subgroup \( i \) is not empty. Observe that since individuals in the same subgroup are indistinguishable, a swap move can also be interpreted as the simultaneous removal of an individual taken randomly in subgroup \( i \) and addition of an individual in subgroup \( j \). A swap event from \( i \) and \( j \) is indexed by the ordered pair \( \kappa = (i, j) \), and the family of the swap events is indexed by the set \( J^s = \{ \kappa = (i, j) | i, j \in I_p, i \neq j \} \).

b) A demographic event is a birth or death in a subgroup. An event of type “birth in subgroup \( j \)” means that an individual is added to the subgroup \( j \). There is some ambiguity in using the word “birth” in this context since no information is given on the “parents” of the newborn, and a birth event is not a priori endogenous to the population. An event of type “death in subgroup \( i \)” can only occur if the \( i \)th subgroup is not empty, and means that an individual is removed from subgroup \( i \).

In order to have a unified description of swap and demographic events, it is convenient to add an “infinite subpopulation” denoted by \( \{ \infty \} \), composed of all individuals not born yet or already dead. Then, the event “birth in subgroup \( j \)” may be viewed as a swap event \( (\infty, j) = (b, j) \) from \( \infty \) to \( j \), and the event “death in the subgroup \( i \)” may be viewed as a swap event \( (i, \infty) = (d, i) \) from \( i \) to \( \infty \). The indexing sets are modified as follows to include this additional subgroup:

- Characteristics set: \( I = I_p \cup \{ \infty \}, I^{(i)} = I \setminus \{ i \} \).
- Set of all events: \( J = \{ \gamma = (\alpha, \beta); \alpha \in I, \beta \in I^{(\alpha)} \} \).
- Set of swap events: \( J^s = \{ \kappa = (i, j); i \in I_p, i \in I^{(i)} \} \).
- Set of demographic events: \( J^{\text{dem}} = \{ (\infty, j); j \in I_p \} \cup \{ (i, \infty); i \in I_p \} \).
- \( \text{card}(I) = p + 1 \), \( \text{card}(J) = p(p + 1) \), \( \text{card}(J^s) = p(p - 1) \), \( \text{card}(J^{\text{dem}}) = 2p \).

For any vector \( x \), we also denote by \( x^2 = <1, x> \).

Jumps The vector space \( \mathbb{Z}^p \) is equipped with the canonical basis \( (e_i)_{1 \leq i \leq p} \), with \( e_i = \delta(0, \ldots, 0, 1_i, 0) \), and the vector \( e_\infty \) is the null vector, \( e_\infty = \delta(0, \ldots, 0) \). For any \( \gamma = (\alpha, \beta) \in J \), the jump in the population associated with the event of type \( \gamma = (\alpha, \beta) \) is the \( \mathbb{Z}^p \)-valued vector:

\[ \phi(\gamma) = e_\beta - e_\alpha. \]

The state of the population after event \( \gamma \) is \( z + \phi(\gamma) = z + e_\beta - e_\alpha \). In particular, \( \phi(b, j) = \phi(\infty, j) = e_j, \phi(d, i) = \phi(i, \infty) = -e_i \).

\( \phi \) can also be seen as the \((p, p(p + 1))\) matrix with columns \( \phi(\gamma) \). It can be interesting to isolated swap from demographic jumps: the \((p, p(p - 1))\) matrix of swap jumps is denoted by \( \phi^s \), and \( \phi^{\text{dem}} = (\phi^b, \phi^d) \) is the \((p, 2p)\) matrix of demographic events, with \( \phi^b \) the \( p \)-identity matrix and \( \phi^d = -\phi^b \). By abuse of notation, we sometimes write \( \phi = \phi^s + \phi^{\text{dem}} \).

Space of counting vectors We close this section by introducing the space of counting vectors, defined as \( \mathbb{N}^{p(p+1)} \) vectors indexed by the set of all events \( J \), and denoted by \( \nu = (\nu_\gamma)_{\gamma \in J} \).

For \( \gamma \in J \), \( \nu^\gamma \) should be interpreted as the number of events of type \( \gamma \) which occurred in the population. We also use the notation \( \nu = (\nu^s, \nu^{\text{dem}}) \) or \( \nu = \nu^s + \nu^{\text{dem}} \), where \( \nu^s = (\nu^s_\gamma)_{\gamma \in J^s} \) is the swap part of \( \nu \), and \( \nu^{\text{dem}} = (\nu^b, \nu^d) \) is the demographic part, with \( \nu^b = (\nu^b_i)_{i \in I_p} \) and \( \nu^d = (\nu^d_{i})_{i \in I_p} \).
The following calculations will be particularly useful in following, in which the population is represented using counting vectors. \( \phi \circ \nu \in \mathbb{Z}^p \) denotes the matrix product of \( \phi \) with \( \nu \),
\[
\phi \circ \nu = \sum_{\gamma \in J} \phi(\gamma) \nu^\gamma = \phi^s \circ \nu^s + \nu^b - \nu^d,
\]
(1.1)
For each type of event \( \gamma \in \mathcal{J} \), the counting process \( N_\gamma \) defined by \( N_\gamma = \sum_{0<s\leq t} \mathbb{1}_{\{\Delta Z_s = \phi(\gamma)\}} \) is the process which counts the number of events of type \( \gamma \). By assumptions, the processes \( N_\gamma \) have no common jumps. Thus, the process \( \mathbf{N} = (N_\gamma)_{\gamma \in \mathcal{J}} \) indexed by \( \mathcal{J} \) is a well defined multivariate counting process, i.e., a vector of counting processes with no common jumps, called the jumps counting process of \( Z \).

### 1.2 Population process and jumps counting process

From now on, all processes are assumed to be defined on a given probability space \((\Omega, \mathcal{G}, P)\), equipped with a filtration \((\mathcal{F}_t)\) verifying the usual assumptions of right-continuity and completeness. The predictable \(\sigma\)-field generated by adapted càdlàg processes is denoted \(\mathcal{P}(\mathcal{G})\). The evolution of the population is described by an adapted càdlàg process, i.e., a vector of counting processes with no common jumps, called the population process. We also assume that two events cannot happen simultaneously. At time \( t \) the state of the population is \( Z_t = (Z^1_t, ..., Z^n_t) \), with \( Z^i_t \) the number of individuals in subgroup \( i \).

#### 1.2.1 Jumps counting process

**Jumps counting process** As a multivariate pure jump process, the population process \( Z \) has piecewise constant and càdlàg paths, and can be written as the sum of its jumps, denoted by \( \Delta Z_s = Z_s - Z_{s-} \). With the unified notations introduced above, the jump corresponding to the event of type \( \gamma = (\alpha, \beta) \) is \( \phi(\gamma) \), so that,
\[
Z_t = Z_0 + \sum_{0<s\leq t} \Delta Z_s = Z_0 + \sum_{0<s\leq t} \sum_{\gamma \in \mathcal{J}} \mathbb{1}_{\{\Delta Z_s = \phi(\gamma)\}} \phi(\gamma) \tag{1.3}
\]
For each type of event \( \gamma \in \mathcal{J} \), the counting process \( N_\gamma \) defined by \( N_\gamma = \sum_{0<s\leq t} \mathbb{1}_{\{\Delta Z_s = \phi(\gamma)\}} \) is the process which counts the number of events of type \( \gamma \). By assumptions, the processes \( N_\gamma \) have no common jumps. Thus, the process \( \mathbf{N} = (N_\gamma)_{\gamma \in \mathcal{J}} \) indexed by \( \mathcal{J} \) is a well defined multivariate counting process, i.e., a vector of counting processes with no common jumps, called the jumps counting process of \( Z \).

By interchanging the sums in Equation (1.3), the population process can be written as an affine function of the jumps counting process \( \mathbf{N} \),
\[
Affine \ relation \quad Z_t = Z_0 + \sum_{\gamma \in \mathcal{J}} \phi(\gamma) N^\gamma_t = Z_0 + \phi \circ \mathbf{N}_t \tag{1.4}
\]
Observe that \( Z_t \) is an affine function of \((Z_0, \mathbf{N}_t)\), but \( \mathbf{N}_t \) is depending on all the history \([Z]_t = (Z_s)_{0\leq s\leq t}\) of the population process \( Z \). Using the notations introduced in [1.1] the affine relation can also be written as \( Z_t = Z_0 + \phi^s \circ \mathbf{N}^s_t + \mathbf{N}^b_t - \mathbf{N}^d_t \).

An equivalent point of view is to consider the jump measure of the population process, i.e., the random measures on \( \mathbb{R}^+ \times \mathcal{J} \),
\[
J(dt, d\gamma) = \sum_{s>0} \sum_{\gamma \in \mathcal{J}} \mathbb{1}_{\{\Delta Z_s = \phi(\gamma)\}} \delta_{\gamma}(d\gamma)\delta_s(dt) \tag{1.5}
\]
Thus, this representation can be applied to any pure jump process generated by a finite or countable number of jump types. In particular, a similar representation is used in Anderson and Kurtz (2015) in the case of Continuous Time Markov Chains for the modeling of chemical reaction network. To the best of our knowledge, this representation is less usual for population dynamics.

Support Condition An individual can be removed from subgroup \( i \) in several cases: a death, definitive exit, or a swap move to another subgroup \( j \). These events are indexed by couples \((i, \beta)\), \(\beta \in \mathcal{I}(i)\), and cannot happen if the subpopulation \( i \) is empty. This means that for any such \((i, \beta)\), the counting process \( N_{i,\beta} \) satisfies the following support condition

\[
\forall \beta \in \mathcal{I}(i), \quad \int_0^t \mathbf{1}_{\{Z_{s-}=0\}} dN_{i,\beta} = 0, \quad \mathbb{P}\text{-a.s. (1.6)}
\]

Therefore, the processes \( N_{i,\beta} \) cannot be defined independently of the population process in order to respect the constraint that components of \( Z \) are non negative.

The situation is different for events leading to the addition of an individual in subgroup \( j \). Since no reference is made on the origin of the new individual, such events include swap moves, but also birth (from parents in the population) or external entry (of immigrants). In a closed population (without immigration), it is assumed that no event can happen in an empty population, so that

\[
\int_0^t \mathbf{1}_{\{Z_{s-}=0\}} dN_s = 0, \quad \text{or equivalently } \int_0^t \mathbf{1}_{\{Z_{s-}=0\}} dN_s^2 = 0, \quad \mathbb{P}\text{-a.s.}
\]

Initialization procedure In the previous description of the population process, little attention has been paid to the initial condition \( Z_0 \). In many problems, the date 0 for the initial condition is very arbitrary, and the state \( Z_0 \) provides poor information on the population past. Massoulié (1998) suggests to define the population process on \( \mathbb{R} \) in place of \( \mathbb{R}^+ \), and to model the initial condition as the state at time 0 of some \( N^p \)-valued process \((\xi_t)_{t \leq 0}\). It is thus natural to consider a generalized initial condition starting at a random date \( \tau \geq 0 \) (generally a stopping time) from state \( \zeta_\tau \in N^p \). This information is summarized by the so-called entry process \( \xi_t(\zeta) = \zeta_\tau \mathbf{1}_{\{\tau \leq t\}} \) (often denoted \( \xi_\tau \) or \( \xi_{\tau+} \) for simplicity). In this generalized setting,

\[
Z_t = \xi_t + \phi \odot N_t, \quad N_t^\gamma = \sum_{\tau<s \leq t} \mathbf{1}_{\{\Delta Z_s=\phi(\gamma)\}}.
\]

In particular, the jumps counting process is null on the set \( \{ t; \xi_{t-} = 0 \} \).

1.2.2 Population system with entry process

The representation of the population using a larger multivariate counting process is very advantageous, due to the many tools available for the study of point processes. We are thus interested in the inverse modeling approach, that is in the construction of a population process from an entry process and a jumps counting process. Given a couple \((\xi, N)\), the population is obviously characterized by the affine relation (1.4). However, such a process is not necessarily a well defined population since its components can take negative values. A necessary and sufficient condition for the population process to be well-defined is actually the support condition (1.6).

Definition 1.1 (Population system with random departure).

a) Let \((\xi_t = \zeta_\tau \mathbf{1}_{\{\tau \leq t\}})\) be an \( N^p \)-valued entry process. A \( p(p+1) \)-multivariate counting process

\[
7
\]
\( \mathbf{N} \) indexed by \( \mathcal{J} \) is called a jumps counting process starting from \( \xi \) iff

\[
\begin{align*}
\text{Starting condition} & \quad \mathbb{1}_{\{\xi_t=0\}}dN_t = 0 \\
\text{Support condition} & \quad \mathbb{1}_{\{\xi_t+(\phi \circ \mathbf{N})_t=0\}}dN_t^{i,\beta} = 0 \quad \forall i \in \mathcal{T}_p, \quad \forall \beta \in \mathcal{I}^{(i)} 
\end{align*}
\] (1.7)

b) The companion population process of \((\xi, \mathbf{N})\) is defined by \( Z_t = \xi_t + \phi \circ \mathbf{N}_t \). In particular, \( Z \) is a well-defined population process with jumps counting process \( \mathbf{N} \) and entry process \( \xi \). The triplet \((\xi, \mathbf{N}, Z)\) is called a population system.

**Aggregated process** A demographic event in the population corresponds to a jump of one of the processes \( \mathbf{N}^b = (N^{\infty,j}) \) or \( \mathbf{N}^d = (N^{i,\infty}) \). The size of the companion population process, also called the aggregated process, is:

\[
Z^z = \xi^z + (\phi \circ \mathbf{N})^z = \xi^z + N^{b,z} - N^{d,z}. 
\]

This can simply be interpreted as the initial size of the population to which is added the total number of births minus the number of deaths, since swap events don’t change the size of the population.

**Swap processes** Populations in which no demographic event occur are called *Swap processes*. As for population systems, a swap system is defined by a triplet \((\xi, \mathbf{N}^{sw}, X)\), where the swap jumps counting process \( \mathbf{N}^{sw} \) is now indexed by \( \mathcal{J}^{sw} \). The swap process is defined by \( X_t = \xi_t + \phi^s \circ \mathbf{N}^{sw}_t \).

For each \( n \in \mathbb{N} \), the finite space of populations of size \( n \) is denoted by:

\[
U_n = \{ z \in \mathbb{N}^p; \ z^z = n \}. 
\]

Since swap events don’t change the size of the population, the swap size \( X^z \) after \( \tau \) is constant equal to \( \xi^z \), or equivalently \( X_t \in U_{\xi^z} \), for any \( t \geq \tau \).

### 1.2.3 Temporal transformation of population system

Several transformations on population systems can be directly obtained from the previous algebraic definition.

**Population decomposition** Let \((Z_0, \mathbf{N}, Z)\) be a population system starting from 0, and \( \tau \) a random time. The system stopped at time \( \tau \) is \((Z_0, \mathbf{N}^\tau = \mathbf{N}_{t<\tau}, Z^\tau = Z_{t<\tau})\), with jumps counting process \( \mathbf{N}^\tau_t = \int_0^t \mathbb{1}_{[0,\tau]}(s)d\mathbf{N}_s \). A population system starting from \( \tau \) in state \( Z_\tau \) can also be defined, from the entry process \( \xi^\tau_t(Z) = Z_\tau \mathbb{1}_{\{\tau \leq t\}} \) and the jumps counting process \( \mathbf{N}^{\tau+} \):

\[
\mathbf{N}^{\tau+}_t = \mathbb{1}_{[\tau,\infty)} \ast \mathbf{N}_t 
\]

with \( \mathbb{1}_{[\tau,\infty)} \ast \mathbf{N}_t = \int_0^t \mathbb{1}_{[\tau,\infty)}(s)d\mathbf{N}_s = \mathbf{N}_t - \mathbf{N}_t^{\tau} \).

Then, \( \mathbf{N}^{\tau+} \) is a jumps counting process starting from \( \xi^\tau(Z) \), associated with the companion population process \( Z^{\tau+} = \xi^\tau(Z) + \phi \circ \mathbf{N}^{\tau+} \). By definition,

\[
Z = Z_0 + \phi \circ (\mathbf{N} - \mathbf{N}^\tau + \mathbf{N}^{\tau+}) = Z^\tau + \phi \circ \mathbf{N}^{\tau+}. 
\]

**Population process with only one demographic event** Let us now assume that only one demographic event (birth or death) occurs in the population, at time \( T^{dem} = \tau \), which is the first jump of the counting process \( \mathbf{N}^{dem} = N^{k,b} + N^{\varpi,d} \). On \([0,\tau]\), \( \mathbf{N}^{dem} \) is the null process and \( Z_t = Z_0 + \phi^s \circ \mathbf{N}^s_t \) behaves as a “pure swap process”. Since two events cannot occur at the same time, the process \( Z_0 + \phi^s \circ \mathbf{N}^s_{t<\tau} \) is continuous in \( \tau \), and can be considered on \([0,\tau]\) as a
stopped pure Swap process $X^\tau$. By definition, the population process has a jump at time $\tau$ equal to $N^b_\tau - N^d_\tau$, where $N^{\text{dem}}_\tau = (N^b_\tau, N^d_\tau)$ has only one non-zero component. Then, the population process stopped at time $\tau$ can be written as $Z^\tau = X^\tau + N^b_\tau - N^d_\tau$. Since only one demographic event occurs, the process $Z^\tau$ starting at time $\tau$ is a pure Swap process.

Observe that this description actually defines a continuous pasting procedure of two Swap processes evolving on different state spaces. The demographic event is used as a switch process, from state space $UZ^0$ to $UZ^\tau + N^{b,\tau} - N^{d,\tau}$, and also defines the starting state of the second Swap. This decomposition will be the building block of the decomposition algorithm of Section 3.2.

1.3 Birth Death Swap Intensity

By describing events changing the population composition rather than the behavior of individuals, we have obtained a very flexible algebraic description of the population. We can now define the so-called Birth Death Swap (BDS) system. Thanks to the theory on point processes and their pathwise representations, the BDS system is characterized by properties of the multivariate intensity of the jumps counting process. Let us first give a brief recall on intensity processes.

1.3.1 Brief overview on intensity processes

**Intensity process** A $(\mathcal{G}_t)$-adapted counting process $N$ is said to have the $(\mathcal{G}_t)$-intensity process $\lambda$ iff $\lambda$ is a non-negative, $(\mathcal{G}_t)$-predictable process and

$$M^\lambda_t = N_t - \int_0^t \lambda_s \, ds \quad \text{is a } (\mathcal{G}_t)\text{-local martingale.} \quad (1.8)$$

Informally, $\lambda_t \, dt$ is the linear estimate of $N$ between $[t, t+dt]$, conditionally to the strict information given by $\mathcal{G}_t^- : \mathbb{E}[N_{t+dt} - N_t | \mathcal{G}_t^-] \simeq \lambda_t \, dt$.

The predictable intensity process $\lambda$ reflects any predictable support condition on the process $N$, since for any predictable subset $\Delta$,

$$\int_0^t 1_{\Delta}(s) \, dN_s = 0 \iff 1_{\Delta}(s) \lambda_s = 0, \, ds \times dP \text{ a.s..} \quad (1.9)$$

In the standard theory motivated by statistical estimation issues, the intensity is often defined in reference to the minimal filtration $\mathcal{F}_t = \sigma(N_s; s \leq t)$ generated by the past history of the counting process, also called canonical filtration. Then, the intensity may only be a function of the past of the counting process. Working with the larger filtration $(\mathcal{G}_t)$ facilitates the analysis of the jumps counting process, in particular by adding flexibility to the setting.

For a multivariate counting process $N = (N^\gamma)_{\gamma \in \mathcal{J}}$ (we recall that by definition, the components of a multivariate counting process have no common jumps), the $(\mathcal{G}_t)$-multivariate intensity process is the vector $\lambda = (\lambda^\gamma)_{\gamma \in \mathcal{J}}$, where $\lambda^\gamma$ is the $(\mathcal{G}_t)$-intensity of process of $N^\gamma$. The concept of spatial counting measure $N(dt, d\gamma)$ is sometimes preferred to the vector representation. The correspondence is given by: $N([0,t] \times \{\gamma\}) = N^\gamma_t$. The associated intensity measure is $\lambda(dt, d\gamma) = \lambda^\gamma_t dt \, d\gamma$, where $d\gamma$ is the counting measure on $\mathcal{J}$.
1.3.2 Birth Death Swap systems

The jumps counting process of a population system cannot have a deterministic intensity, due to the support conditions (1.7) which are transferred onto the intensity process of the jumps counting process by (1.9). In order for the population system \((\xi, N, Z)\) to become a Birth Death Swap system, additional assumptions are made on the multivariate intensity of \(N\). Only the companion population process \(Z\) is usually observed, and a natural assumption is that the intensity process depends on the population process rather than on the jumps counting process itself. This assumption is implicit in a Markov framework. To go further and take into account some additional time-dependent uncertainty, such as a random environment, the multivariate intensity process is assumed to depend in a predictable way on additional randomness, not explicitly modeled.

**Definition 1.2** (BDS intensity functional and BDS system).

a) A BDS intensity functional \(\mu(\omega, t, z) = \mu(t, z) = (\mu^\gamma(t, z))_{\gamma \in \mathcal{J}}\) is a multivariate \((\mathcal{G}_t)\)-predictable non-negative functional depending on \(z \in \mathbb{N}^p\), satisfying

\[
\mu(t, 0) \equiv 0 \quad \text{and} \quad \sum_{i \in I_p} \sum_{\beta \in I^{(i)}} \mu^{i, \beta}(t, z)1_{\{z^i = 0\}} \equiv 0, \quad dt \otimes dP \text{ a.s.} \tag{1.10}
\]

b) A Birth Death Swap (BDS) system of intensity functional \(\mu\) is a population system \((\xi, N, Z)\) such that \(N\) is a multivariate counting process of \((\mathcal{G}_t)\)-intensity \(\mu(t, \xi^- + \phi \odot N_t^-) = \mu(t, Z_t^-)\).

The second part of the condition (1.10) ensures that the BDS system \((\xi, N, Z)\) verifies the second support condition in (1.7). The condition \(\mu(t, 0) \equiv 0\) ensures that the starting condition is also verified. Indeed, we recall that the starting condition means that \(N\) does not jumps on \(\{\xi_t^- = 0\} = [0, \tau]\), or equivalently that its first jump \(T_1\) verifies \(T_1 > \tau\): \(\int \mathbb{1}_{\{\xi_t^- = 0\} \cap \{t \leq T_1\}} dN_t = 0\). By (1.9), this is equivalent to \(\mathbb{1}_{\{\xi_t^- = 0\} \cap \{t \leq T_1\}} \mu(t, Z_t^-) = 0, dt \otimes dP \text{ p.s.}\) By definition, \(Z_t^- = 0\) on \(\{\xi_t^- = 0\} \cap \{t \leq T_1\}\), and since \(\mu(t, 0) = 0\), this means that \(N\) verifies the starting condition. This condition is actually not necessary, and could have been replaced by more general condition \(\mu(t, 0)1_{\{\xi_t^- = 0\}}\). However, we prefer the former condition which leads to simpler notations.

Finally, for any \(f \in C_b(\mathbb{N}^p)\), the population process verifies:

\[
f(Z_t) = f(\xi_t) + \int_0^t (\sum_{\gamma \in \mathcal{J}} (f(Z_s + \phi(\gamma)) - f(Z_s))dN_t^\gamma, \quad \forall t \geq 0.
\]

By definition, \((N_t^\gamma - \int_0^t \mu^\gamma(s, Z_s)ds)_{\gamma \in \mathcal{J}}\) is a \((\mathcal{G}_t)\)-local martingale, and thus

\[
f(Z_t) - f(\xi_t) - \int_0^t (\sum_{\gamma \in \mathcal{J}} (f(Z_s + \phi(\gamma)) - f(Z_s))\mu^\gamma(s, Z_s))ds\text{ is a }\mathcal{G}_t\text{-local martingale.} \tag{1.11}
\]

In particular, if \(\mu\) is an homogeneous deterministic function \(\mu(z)\), we can see that \(Z\) is solution of a classical martingale problem associated with a Continuous Time Markov Chain.

**Examples of BDS intensities** For an event \(\gamma \in \mathcal{J}\), \(\mu^\gamma(t, Z_t^-)\) is the intensity corresponding to the occurrence of the event of type \(\gamma\) in all the population. This should not be confused with the rate at which an event \(\gamma\) can occur to one individual, which is used in order to describe the
evolution of the population in individual based models.

a) Linear intensities and individual rates: The intensity \( \mu^\gamma \) is called linear if the functional depends linearly on the number of individuals in each subgroup, and a direct interpretation of intensities in term of individual rates can be given in this case:

A classical linear intensity functional for an event \((d, i)\) is \( \mu^{(d,i)}(t, z) = d_i z^i \), which can be interpreted as follow: all individuals in subgroup \( i \) die at rate \( d_i \). A similar interpretation can be given for linear swap intensities \( \mu^{(i,j)}(t, z) = k_i^j z^j \).

When the intensity of an event \((b, j)\) is \( \mu^{(b,j)}(t, z) = b_i^j z^j \), all individuals in subgroup \( j \) give birth to an individual of same characteristics at rate \( b_i^j \). Mutations can be included, by taking \( \mu^{(b,j)}(t, z) = \sum_{i=1}^p b_i^j m_i(i, j) z^i \), where \( m_i(i, j) \) is the random probability for an individual born at time \( t \) from a parent in subgroup \( i \) to be in subgroup \( j \). A stochastic intensity \( \lambda_t \) can also be added to the birth intensity, in order to model the entry of immigrants at rate \( \lambda_t \).

b) Markov BDS system: When the BDS intensity functional is an homogeneous deterministic function \( \mu(z) \), the BDS system is a Continuous Time Markov Chain (CTMC). In this case, the process can be described using classical tools of CTMC. If the initial population is \( Z_0 \), the first jump time is distributed as an exponential of parameter \( \mu^d(Z_0) \), and the event of type \( \gamma \) is chosen independently as the first event with probability \( \mu^\gamma(Z_0)/\mu^d(Z_0) \). The BDS system can be built by iterating the last two steps.

c) Nonlinear Swap intensity: Let us introduce an example of a BDS system with general birth and death intensities and deterministic nonlinear swap intensity functional, to which we come back at the end of the paper. We consider here the case of two subgroups or patches \((p = 2)\), where subgroup \( 2 \) is a favorable subgroup with lower death intensity: \( \mu^{(d,1)}(t, z) \geq \mu^{(d,2)}(t, z) \), a.s. Two regimes can be distinguished for the swap events, depending on the size of the population: when the population is small, \( z^2 \leq M \), individuals can swap more easily to the favorable subgroup \( 2 \), at a rate \( k_{12}(z^2)\alpha \), with \( \alpha > 0 \). When the population size is large, \( z^2 > M \), access to the subgroup \( 2 \) is limited and individuals swap from 1 to 2 at a constant rate \( k_{12}^M \). In both cases, individuals swap from the favorable subgroup \( 2 \) to subgroup \( 1 \) at constant rate \( k_{21} \). To summarize, the swap intensity is defined as follow:

\[
\mu^{(1,2)}(z) = k_{12}(z^2)\alpha z^1 \mathbf{1}_{\{z^2 \leq M\}} + k_{12}^M z^1 \mathbf{1}_{\{z^2 > M\}}; \quad \mu^{(2,1)}(z) = k_{21} z^2. \tag{1.12}
\]

In particular, pure Swap processes with intensity defined as above are CTMC. Since pure Swap processes have a constant size determined by their initial condition, the intensity regime is determined by the size \( n \) of the Swap process, and individuals swap independently from 1 to 2 at a constant rate equal to \( k_{12} n^\alpha \) or \( k_{12}^M \) according to the size of the Swap.

2 Pathwise comparison of multivariate counting processes

The question of the existence of BDS systems can be complex considering the feedback effect induced by the intensity functional. Markovian assumptions are generally made for Birth Death processes or for Swap processes, yielding to a distributional point of view on the existence of such processes. Since the 1990s, a pathwise point of view, allowing more flexibility on intensity processes, has been considered by many authors (see e.g. Brémaud and Massoulié (1996), Massoulié...
This point of view is based on the pathwise realization of point processes as solutions of Stochastic Differential Equations (SDE) driven by Poisson measures, obtained from the thinning of “augmented Poisson measures”.

In this section, we rely on these representations in order to give a pathwise representation of the BDS system, based on the realization of the jumps counting process as the solution of a multivariate SDE. Focusing on the jumps counting process allows us to adopt the point of view of point processes, within a general framework similar to that of Massoulie (1998). Thanks to this shift in focus, the existence of BDS systems is derived from a more general result obtained in this section, on the construction of multivariate counting processes by strong domination with a non-explosive process. In particular, this construction allows us to relax some of the usual assumptions on the intensity functional that are Lipschitz or sublinear growth conditions. Furthermore, the strong domination construction will be very useful in Section 4 by providing “free” tightness properties.

In this section, multivariate objects are denoted by the symbol − to reflect the generality of the obtained results.

2.1 Thinning of Poisson measure

By convention, all Poisson measures are assumed to be defined on the given probability space (Ω, G, P).

\( \mathcal{G}_t \)-Poisson measure. We recall that a random measure \( Q(dt, d\theta) \) on \( \mathbb{R}^+ \times \mathbb{R}^+ \) is a \( \mathcal{G}_t \)-Poisson measure of intensity measure \( dt \times d\theta \), if for disjoints sets \( A_1, \ldots, A_n \subset \mathbb{R}^+ \) such that \( \int_{A_i} d\theta < \infty \), the processes \( \{Q_t(A_i) = Q([0, t] \times A_i), i = 1 \ldots n, \} \) are independent Poisson processes (they have no common jumps) of \( \mathcal{G}_t \)-intensity \( \int_{A_i} d\theta \). In particular, all properties of Poisson measures are defined in reference to the large filtration \( \mathcal{G}_t \).

\( Q \) can also be characterized by a set \( \{ (T_i, \Theta_i) \} \) of random times \( T_i \) associated with marks \( \Theta_i \), 
\( Q(dt, d\theta) = \sum_i \delta_{(T_i, \Theta_i)}(dt, d\theta) \). However, since the Lebesgue measure is only \( \sigma \)-finite, there is no ordered enumeration of the jump times \( T_i \).

Thinning procedure. For any random set \( \Delta \in \mathcal{P}(\mathcal{G}) \otimes \mathcal{B}(\mathbb{R}^+) \), we can defined a point process \( Q^\Delta \) of intensity \( q^\Delta(dt, d\theta) = 1_{\Delta}(t, \theta)dt d\theta \), as the restriction of the Poisson measure \( Q \) to \( \Delta \), 
\[ Q^\Delta(dt, d\theta) = 1_{\Delta}(t, \theta)Q(dt, d\theta). \]

A fundamental example is to take \( \Delta = \{ (\omega, t, \theta); 0 < \theta \leq \lambda_i(\omega) \} \), with \( \lambda \) a non negative predictable process, \( dt \)-integrable over compact intervals. Then, the marginal of \( Q^\Delta \) on the first coordinate is the Cox process of \( \mathcal{G}_t \)-intensity \( \int_{\mathbb{R}^+} 1_{\{0<\theta\leq \lambda_i\} } d\theta = \lambda_i \):
\[ Q^\lambda_1 = \int_0^t \int_{\mathbb{R}^+} Q^\Delta(dt, d\theta) = \int_0^t \int_{\mathbb{R}^+} 1_{\{0<\theta\leq \lambda_i\} } Q(ds, d\theta) = \int_0^t Q(ds, 0, \lambda_i), \]

obtained by thinning and projection of the Poisson measure \( Q \). By definition, \( Q^\lambda \) is characterized by an increasing sequence of jump times \( \{ T^\lambda_i \} \) and thus \( Q^\Delta \) is characterized by the sequence \( \{ T^\lambda_i, \Theta^\lambda_i \} \), with \( \Theta^\lambda_i \) the mark of \( Q \) associated with \( T^\lambda_i \). This increasing enumeration of the jump times of \( Q^\Delta \) is particularly useful since it allows us to define a construction algorithm for processes obtained by thinning of \( Q^\Delta \), based on acceptance/reject of jump times of \( Q^\Delta \), while this is not
possible for processes obtained by thinning of the Poisson measure \(Q\).

The thinning procedure can be extended to more general Poisson measures \(Q(dt, dx, d\theta)\), with marks \((x, \theta)\) defined on \((E \times \mathbb{R}^+, \mathcal{E} \times \mathcal{B}(\mathbb{R}^+))\) and intensity measure \(dt \times \nu(dx) \times d\theta\), where \(\nu(dx)\) is a \(\sigma\)-finite measure on \(E\). In many applications, \(E\) is a finite space of cardinal \(\rho\), with elements denoted by \(\imath\) and equipped with the uniform counting measure \(\im\). In the case of BDS systems for instance, \(E\) is the set \(\mathcal{F}\) of all events types \((\rho = p(p+1))\). It is thus interesting to use a vector version of the foregoing, in which the Poisson measure \(Q(dt, \im, d\theta)\) is reinterpreted as a family of independent Poisson measures \(Q(dt, d\theta) = (Q^i(dt, d\theta); \im \in E)\), with no common jumps. Given a multivariate predictable intensity process \(\bar{\lambda} = (\lambda^i)_{i \in E}\), a multivariate Cox process can be defined as \(\bar{Q}^\lambda_t = \int_0^t \mathbf{1}_{[0, \bar{\lambda}^i]}(\theta)Q(ds, d\theta) = \left( \int_0^t Q(ds, d\im, [0, \lambda^i(s, \im)]) \right)_{\im \in E}\), associated with the multivariate random measure \(\bar{Q}^\Delta(dt, d\theta) = \mathbf{1}_{[0, \bar{\lambda}]}(\theta)\bar{Q}(dt, d\theta)\). Due to the independence of the Poisson measures \(Q^i\), the components of \(\bar{Q}^\lambda\) and \(\bar{Q}^\Delta\) also have no common jumps.

**SDE driven by Poisson measures** When the intensity \(\bar{\lambda}\) of the multivariate counting process \(\bar{Q}^\lambda\) is a predictable functional of \(Q^\lambda\) itself, \(\bar{\lambda}_t = \bar{\alpha}(\omega, t, \bar{Q}^\lambda_{t-})\), the thinning equation becomes a stochastic differential equation (SDE) driven by the multivariate Poisson measure \(Q\), and \(\bar{Q}^\lambda\) is now solution of the multivariate SDE:

\[
d\bar{Y}^\alpha_t = \bar{Q}(dt, [0, \alpha(t, \bar{Y}^\alpha_{t-}))].
\]

In their paper on the modeling of recurrent events (another terminology used for counting processes), Giessing et al. (2010) highlight the importance of non-explosion properties in order to avoid fitting errors due to the uncontrollable nature of exploding systems. In the remainder of this paper, all solutions of stochastic differential equations are considered to be well-defined if they stay finite in finite time with probability one.

2.2 Existence and strong domination of multivariate counting processes as solutions of SDEs driven by Poisson measures

**Strong order** Let us introduce the different orders that we shall need in the sequel. A \(\rho\)-multivariate counting process \(\bar{Y}\) is said to be strongly dominated by \(\bar{X}\), \(\bar{Y} \prec \bar{X}\) iff

\[
\bar{X} - \bar{Y} \text{ is a multivariate counting process,}
\]

or equivalently iff all jumps of \(\bar{Y}\) are jumps of \(\bar{X}\).

We also consider a strong order on stochastic intensity functionals: let \(\bar{\alpha}(t, \bar{y}) = (\alpha^i(t, \bar{y}))_{i \in \rho}\) and \(\bar{\beta}(t, \bar{y}) = (\beta^i(t, \bar{y}))_{i \in \rho}\) be two intensity functionals. \(\bar{\alpha}\) and \(\bar{\beta}\) are strongly ordered, \(\bar{\alpha} \leq_s \bar{\beta}\) iff:

\[
a.s. \forall 1 \leq i \leq \rho, t \geq 0 \text{ and } \bar{y} \leq \bar{x} \in \mathbf{N}^\rho \quad \alpha^i(t, \bar{y}) \leq \beta^i(t, \bar{x}).
\]

Observe that \(\bar{\beta}\) is dominated by itself iff \(\bar{\beta}\) is non-decreasing in \(y\).

**Pathwise comparison of Cox processes** The comparison of point processes with ordered (stochastic) intensities has been the subject of many papers (see e.g. Preston (1973), Rolski and Szekli (1991)). The thinning procedure is well-adapted to solve this problem, and for Cox processes, the answer is immediate. If two Cox processes \(\bar{Q}^\lambda_1 = \int_0^t Q(ds, [0, \lambda^i_1])\), \(i = 1, 2\) have ordered intensities \(\lambda^1_1 \leq \lambda^2_1\), then the thinning construction described above using the same Poisson measure for both processes yields that \(\bar{Q}^\lambda_1 \leq \bar{Q}^\lambda_2\). Actually, a stronger property is verified. Since,
\[ \lambda_1 \leq \lambda_2^2. \] 

\[ Q^{\lambda^3} \text{ can be rewritten as } Q^{\lambda^3}_t = \int_0^t Q(ds, [0, \lambda^1_t \wedge \lambda^2_t]) = \int_0^t Q^{\lambda^2}(ds, [0, \lambda^1_t]). \] 

This means that \( Q^{\lambda^3} \) can be obtained by thinning of \( Q^{\lambda^2} \) instead of \( Q \). In particular, all jump times of \( Q^{\lambda^3} \) are jump times of \( Q^{\lambda^2} \) and \( Q^{\lambda^3} \) is strongly dominated by \( Q^{\lambda^2} \).

**Strong domination of multivariate counting processes** The direct application to general multivariate counting processes is not so easy, since the natural order of intensity functionals does not necessarily imply a natural order on the stochastic intensities. Nevertheless, for online Markov Birth processes, Bhaskaran (1986) (see also Bezborodov (2015)) showed that for non exploding processes, an intensity functional inequality implied a strong domination result.

Theorem 2.1 sets a general framework for the pathwise comparison of multivariate counting processes with ordered intensity functionals. The result allows us to build the solution \( \hat{Y}^\alpha \) of the multivariate SDE (2.1) by strong comparison with a dominating process \( \hat{Y}^\beta \) of multivariate stochastic intensity functional \( \beta(t, \bar{y}) \), assuming that the intensity functional \( \tilde{\alpha}(t, \bar{y}) \) of \( Y^\alpha \) is strongly dominated by \( \beta(t, \bar{y}) \).

**Theorem 2.1** (Strong comparison of \( \rho \)-multivariate counting processes). Let \( \tilde{Q}(dt, d\theta) = (Q^\rho(dt, d\theta))_{s \in E} \) be a multivariate Poisson measure, and \( \beta \) a \( \rho \)-intensity functionals such that there exists a unique well-defined solution \( \hat{Y}^\beta \in \mathbb{N}^\rho \) of the multivariate SDE:

\[ d\hat{Y}^\beta_t = \tilde{Q}(dt, [0, \beta(t, \hat{Y}^\beta_{t-})]). \]  \hspace{1cm} (2.2)

Then, for all intensity functional \( \tilde{\alpha} \) strongly dominated by \( \beta \) (\( \tilde{\alpha} \leq_s \beta \)), there exists a unique solution of the equation

\[ d\hat{Y}^\alpha_t = \tilde{Q}(dt, [0, \tilde{\alpha}(t, \hat{Y}^\alpha_{t-})]). \]  \hspace{1cm} (2.3)

Furthermore, \( \hat{Y}^\alpha \) is strongly dominated by \( \hat{Y}^\beta \); \( \hat{Y}^\alpha \prec \hat{Y}^\beta \).

**Proof.** The main argument of the proof is similar to the case of Cox processes, and relies on replacing the driving Poisson measure \( \tilde{Q} \) by the random measure \( \hat{Q}^{\Delta_\beta} \) associated with the dominating process \( \hat{Y}^\beta \).

a) Let us introduce of a slightly different version of Equation (2.3), based on the multivariate random measure \( \hat{Q}^{\Delta_\beta} \) defined by,

\[ \hat{Q}^{\Delta_\beta}(dt, d\theta) = \mathbb{1}_{\Delta_\beta}(t, \theta)\tilde{Q}(dt, d\theta), \text{ with } \Delta_\beta = \{(t, \theta); 0 < \theta \leq \beta(t, \hat{Y}^\beta_{t-})\}. \]

By assumption, the solution \( \hat{Y}^\beta \) of Equation (2.2) is a well-defined multivariate counting process. Thus, its jumps can be enumerated by a sequence \( (T_j, \bar{y}_j)_{j \geq 1} \), where \( (T_j) \) is the increasing sequence of the jump times of \( \hat{Y}^\beta \), \( \lim T_j = +\infty \), and \( \bar{y}_j \) is the index of the component of \( \hat{Y}^\beta \) jumping at time \( T_j \). As a consequence, \( \hat{Q}^{\Delta_\beta} \) is a marked multivariate counting process, and can be characterized by the sequence \( (T_j, t_j, \Theta_j)_{j \geq 1} \), where \( \Theta_j \) is the mark of the Poisson measure \( \hat{Q}^{t_j} \) associated with \( T_j \).

b) **Existence for Equation (2.3):** The modified equation driven by \( \hat{Q}^{\Delta_\beta} \) is defined by

\[ d\hat{Y}^\alpha_t = \hat{Q}^{\Delta_\beta}(dt, [0, \tilde{\alpha}(t, \hat{Y}^\alpha_{t-})]) = \hat{Q}(dt, [0, \tilde{\alpha}(t, \hat{Y}^\alpha_{t-}) \wedge \beta(t, \hat{Y}^\beta_{t-})]), \]  \hspace{1cm} (2.4)

Thanks to the increasing enumeration of jump times of \( \hat{Q}^{\Delta_\beta} \), the unique solution of (2.4) can be built recursively as the the solution of:

14
\[ \tilde{Y}_t^{\alpha,\beta} = \sum_{j=1}^{\infty} \mathbf{1}_{\{T_j \leq t\}} \mathbf{1}_{\{\tau_j \leq \alpha(T_j, \tilde{Y}_{T_{j-1}}^{\alpha})\}}, \quad \forall t \in E. \]

Since \( \tilde{Y}^{\alpha} \) is obtained by thinning of \( \tilde{Q}^{\alpha,\beta} \), the multivariate counting process is strongly dominated by \( \tilde{Y}^{\beta} \) by definition. In particular, \( \tilde{Y}_t^{\alpha} \leq \tilde{Y}_t^{\beta} \) for all \( t \geq 0 \). Then, since \( \bar{\alpha} \leq_s \bar{\beta} \), \( \bar{\alpha}(t, \tilde{Y}_t^{\alpha}) \leq \bar{\beta}(t, \tilde{Y}_t^{\beta}) \), and thus \( \tilde{Y}^{\alpha} \) is solution of (2.3), which achieves to prove existence.

c) **Uniqueness**: It remains to prove that any solution of (2.3) is solution of (2.4). Let \( \tilde{Y}^{\alpha} \) be a solution of (2.3) and \( T_1^{\alpha} \) its first jump time, associated with the jumping component \( \Delta_1^{\alpha} \) and the mark \( \Theta_1^{\alpha} \) of \( Q^{\alpha,\beta} \). By definition of the thinning procedure, \( \Theta^{\alpha} \leq \alpha(T_1^{\alpha}, 0) \). By assumption, \( \alpha(T_1^{\alpha}, 0) \leq \beta(T_1^{\alpha}, 0) \) and thus \( (T_1^{\alpha}, \Delta_1^{\alpha}) \) is also a jump of \( \tilde{Y}^{\beta} \). By iterating this argument, we obtain that all jump times of \( \tilde{Y}^{\alpha} \) are jump times of \( \tilde{Y}^{\beta} \), or equivalently that \( \tilde{Y}^{\alpha} \prec \tilde{Y}^{\beta} \). In particular, \( \tilde{Y}^{\alpha} \) is the unique solution of (2.4), which achieves the proof of the theorem.

Let \( S_\beta \) be the class of processes \( \tilde{Y}^{\alpha} \) solution of (2.3), with \( \bar{\alpha} \leq_s \bar{\beta} \). Then, by the previous theorem, all processes in \( S_\beta \) are strongly dominated by the solution \( \tilde{Y}^{\beta} \) of (2.2). This yields interesting properties shared by elements of \( S_\beta \), and the following corollary is particularly useful.

**Corollary 2.2.** Under the assumptions and notations of Theorem [2.1], the sequence \( (S_{\beta}^p) \) of jump times of \( \tilde{Y}^{\beta} \) is a universal localizing sequence of \( (G_t) \)-local martingales \( \hat{M}_t^{\alpha} = Y_t^{\alpha} - \int_0^t \bar{\alpha}(s, Y_s^{\alpha})ds \), \( \forall \tilde{Y}^{\alpha} \in S_\beta \).

**Proof.** Let us consider the sup norm on \( \mathbb{R}^\alpha \), \( |y| = \sup_{i=1,\ldots,\alpha} y^i \), and let \( \tilde{Y}^{\alpha} \in S_\beta \). For all \( t, p \geq 0 \),
\[
E[\sup_{0 \leq s \leq t} |\hat{M}_t^{\alpha}_{t \wedge S_p}|] \leq 2E[|\tilde{Y}_{t \wedge S_p}^{\alpha}|] \leq 2E[|\tilde{Y}_{t \wedge S_p}^{\beta}|] \leq 2p.
\]
Thus, \( \hat{M}_t^{\alpha}_{t \wedge S_p} \) is a martingale, and since \( S_p^\beta \to \infty \), \( (S_{\beta}^p) \) is a localizing sequence of \( \hat{M}^{\alpha} \).

**2.3 Class of dominating processes**

In order to apply Theorem 2.1, it remains to define a class of dominating intensities \( \beta \) associated with non-exploding processes, large enough to enable us to obtain the existence of multivariate counting processes under reasonable assumptions. In particular, since the existence or uniqueness is not obtain by controlling moments of the counting processes, the usual assumption of sublinear growth of the intensity is not determinant here.

**Multivariate Markov Birth process** The most simple example of such processes are probably one dimensional Markov pure Birth processes, which have been extensively studied since the 60s. The one dimensional Markov Birth process is characterized by its intensity function, here denoted by \( Kg(y) \) (with \( g(0) = 0 \)). The well-known Feller criterion \( \sum_{j=1}^{\infty} \frac{1}{y_j} = \infty \) (Feller 1968) guarantees the non-explosion of the online Markov Birth process, with the most classical example being the Yule process in the linear case when \( g(y) = by \).

The multivariate case is an easy extension under the following assumption: the \( \rho \)-multivariate Markov birth process has an intensity functional which is a deterministic function of the size \( \tilde{g}^\alpha = \sum_i^\rho y^i \) of the birth process, \( \bar{g}(\tilde{g}) = (g^i(\tilde{g}^i)) \), with \( \sum_{i=0}^{\rho} g^i \) satisfying the Feller criterion. A realization \( \tilde{B} \) of such a process can be obtained as the solution of the following multivariate SDE
driven by a $\rho$-Poisson measure $\bar{Q}$:

\[
\begin{aligned}
B_t(\bar{y}) &= \bar{y} + \bar{N}^B_t \\
\text{Multivariate Feller criterion} \quad \sum_{n=1}^{\infty} \frac{1}{\sum_{t=0}^{n} \sigma(n)} = \infty
\end{aligned}
\tag{2.5}
\]

Observe that the multivariate Feller criterion is equivalent to the following property: for each $i = 1..\rho$, the function $g^i$ verifies the Feller criterion.

**Sketch of the proof:** Let us give some elements of the proof of the (strong) existence and uniqueness of the previous equation. The problem can be reduced to the study of the one-dimensional SDE:

\[
dN^B_t = \sum_{i=1,..\rho} Q^i(dt, [0, K\bar{g}^i(\bar{y}^i + N^B_t)])
\]

obtained with thinning of the Poisson measure $\bar{Q}(dt, d\theta, dt)$. Before its first jump time $T^{1.2}$, $N^{B,2}$ coincides with $\sum_{i=1,..\rho} Q^i([0, K\bar{g}^i(\bar{y}^i)])$, which is a Poisson process of intensity $K \sum_{i=0}^{\rho} g^i(\bar{y}^i)$. Thus, $T^{1.2}$ is defined uniquely and the step can be iterated, by starting from $\bar{y}^i + 1$ at time $T^{1.2}$. $N^{B,2}$ is an one dimensional Markov birth process of intensity function verifying the Feller criterion. Thus, $N^{B,2}$ is non-explosive, and is by construction the unique solution of the above equation. Then, the solution $\bar{N}^B$ of (2.3) is a simple Cox process obtained by thinning of $\bar{Q}$.

**Cox Birth process:** As a direct application of Theorem 2.1 we can obtain the existence and uniqueness of the solution of (2.3) for all predictable intensity functionals $\alpha \leq \beta(t, \bar{y}) = K\bar{g}(\bar{y}^i)$.

In the one dimensional case, this condition is sometimes known as the Jacobsen condition (Jacobsen (1982)). However, the domination by a Markov birth process is often not satisfactory. The assumption can be relaxed by using Cox Birth processes as dominating processes. Cox Birth processes are defined as multivariate counting processes with product intensity $\alpha(t, \bar{y}) = k_t\bar{g}(\bar{y}^i)$, with $(k_t)$ a locally bounded predictable process and $\bar{g}$ a function verifying the multivariate Feller condition. The existence and non-explosion of Cox Birth processes is a corollary of Theorem 2.1.

**Corollary 2.3.** Let $(k_t)$ a locally bounded predictable process and $\bar{g}$ verifying the multivariate Feller condition and $n \in \mathbb{N}$. Then, there exists a unique well-defined solution of the stochastic differential system:

\[
d\bar{Y}_t = \bar{Q}(dt, [0, k_t\bar{g}(n + \bar{Y}^2_t)])
\tag{2.6}
\]

Such a process is called a $\rho$ Cox-Birth process of intensity functional $k_t\bar{g}$.

**Proof.** $(k_t)$ is bounded by a sequence $(K_p)$ along a non-decreasing sequence of stopping times $(S_p)$ going to $\infty$. By Theorem 2.1, there exists for each $p \geq 0$ a unique solution $\bar{Y}^p$ of Equation (2.3) with intensity functional $\alpha^p(t, \bar{y}) = (k_t \wedge K_p)\bar{g}(\bar{y}^i)$ which does not depend on $p$ on $[0, S_p]$. In particular, there exists a unique solution well-defined solution of (2.6) on each interval $[0, S_p]$ and by letting $p$ go to 0, we obtain the unique solution of (2.6).

\[\square\]

### 3 BDS multivariate SDE dominated by a Cox Birth process

Due to the representation of BDS systems introduced in Section 1, the existence of BDS systems boils down to the existence of BDS jumps counting processes, and can be seen as particular case of the foregoing. In the following, Theorem 2.1 is applied in order to realize jumps counting processes of BDS systems, by strong domination with a multivariate Cox Birth process.
3.1 Existence and uniqueness of the BDS multivariate SDE by strong domination

Recall that a BDS system \((\xi, N, Z)\) of intensity functional \(\mu\) is defined by an entry process \(\xi\) and a jumps counting process \(N\) of intensity \(\mu(t, Z_{\cdot -})\), with \(Z_t = \xi_t + \phi \odot N_t\) the companion population process. As a multivariate counting process, \(N\) can thus be represented as a solution of a multivariate SDE driven by a Poisson measure \(Q = (Q^\gamma)_{\gamma \in \mathcal{J}}\) indexed by \(\mathcal{J}\).

**Definition 3.1 (BDS multivariate SDE).** Let \(Q = (Q^\gamma)_{\gamma \in \mathcal{J}}\) be a multivariate Poisson measure, \(\mu\) a BDS intensity functional and \((\xi_t)\) an entry process. The Birth Death Swap multivariate SDE associated with the entry process \(\xi\) and intensity functional \(\mu\) is defined by

\[
dN_t = Q(dt, 0, \mu(t, \xi_{t-} + \phi \odot N_{t-})), \quad Z_t = \xi_t + \phi \odot N_t.
\]

If \(N\) is a solution of (3.1), then \((\xi, N, Z)\) is a BDS system of entry process \(\xi\) and intensity functional \(\mu\).

The intensity functional associated with \(N\) can be written as a functional \(\lambda(t, \nu) = \mu(t, \xi_t + \phi \odot \nu)\), defined on the space of counting vectors \(\nu\). In order to apply Theorem 2.1, the problem is thus reduced to the existence of a dominating non-exploding multivariate counting process, whose intensity functional dominates the \(\nu\)-intensity functional \(\lambda(t, \nu)\).

**Cox Birth Dominating Assumption** We recall that a counting vector \(\nu = (\nu^\gamma)_{\gamma \in \mathcal{J}}\) can be decomposed into a demographic component \(\nu^{\text{dem}} = (\nu^b, \nu^d) = (\nu^\gamma)_{\gamma \in \mathcal{J}^{\text{dem}}}\) and a swap component \(\nu^s = (\nu^\gamma)_{\gamma \in \mathcal{J}^s}\). In particular, \((\phi \odot \nu)^2 = (\phi \odot \nu^{\text{dem}})^2 = \nu^b + \nu^d \leq \nu^s\).

The main problem is the control of the size \(Z_t\) of the population via the control of the size of the counting birth process \(N^{b, 2}\). In the following, the birth part \(\lambda^b\) of the intensity functional is now assumed to be dominated by a Cox Birth intensity functional:

**Assumption 1** (Cox Birth domination assumption). A BDS intensity functional \(\lambda(t, \nu) = \mu(t, \xi_t + \phi \odot \nu)\) is said to verify the Cox Birth domination assumption iff

\[
\text{Cox Birth Hyp} \quad \forall i \in \mathcal{I}_p, \quad \lambda^{(b, i)}(t, \nu) = \mu^{(b, i)}(t, \xi_t + \phi \odot \nu) \leq k_t g^{(b, i)}(\xi_t^b + \nu^{b, 2}) \quad \text{a.s.,} \quad (3.2)
\]

where \(g^b = (g^{(b, i)}_{i=1, p})\) satisfies the multivariate Feller criterion and \((k_t)\) is a predictable locally bounded process.

Thanks to specific properties of BDS systems, no particular assumptions are made on the swap and death intensities. The swap and death part of the intensity functional are naturally dominated by an increasing multivariate functional depending only on \(\nu^b\). Indeed, when expressed in terms of population \(z\), swap and death intensity functionals are smaller than the maximum intensity over the finite space of population of size smaller than \(n = 2^3\), defined by

\[
\hat{\lambda}^{(i, \beta)}(t, n) = \sup_{|\zeta| \leq n} \mu^{(i, \beta)}(t, \zeta), \quad \forall (i, \beta) \in \mathcal{J}^s \cup \mathcal{J}^d.
\]

The functionals \(\hat{\lambda}^{(i, \beta)}(t, n)\) are non-decreasing in \(n\), and thus:

\[
\lambda^{(i, \beta)}(t, \nu) = \mu^{(i, \beta)}(t, \xi_t + \phi \odot \nu) \leq \hat{\lambda}^{(i, \beta)}(t, \xi_t^b + \nu^{b, 2}), \quad \forall (i, \beta) \in \mathcal{J}^s \cup \mathcal{J}^d.
\]

(3.3)
Remark 3.1. A usual assumption made for birth intensities is the assumption of sublinear growth,
\[ \lambda^{(b,i)}(t, \nu) \leq K(\xi_t + \phi \odot \nu), \]
Thanks to the construction by strong domination, the Cox Birth domination assumption allows us to weaken this hypothesis, by including randomness and non linearity in the dominating intensity.

Dominating multivariate counting process The \((p+1)\) dominating Cox-Birth process \(G = (G^b, G^d, G^s)\) can be directly defined by Corollary 2.3. Due to the particular structure of the domination assumption, we can also realize \(G\) in the two following steps:

(i) The first step is to introduce the \(p\)-Cox Birth process \(G^b\), solution of the multivariate SDE:
\[ dG^b_t = Q^b(dt, [0, k_t g^b(\xi^b_t - + G^b_t - )]). \]
(3.4)

By Corollary 2.3, \(G^b\) is well-defined.

(ii) The second step is to add “swap and death coordinates” to \(G^b\), by defining the \(p\) and \((p-1)\) multivariate Cox processes:
\[ G^d_t = \int_{[0,t] \times \mathbb{R}^+} 1_{\{\theta \leq \hat{\mu}^d(s, \xi^d_s - + G^d_s - )\}} Q^d(dt, d\theta), \quad G^s_t = \int_{[0,t] \times \mathbb{R}^+} 1_{\{\theta \leq \hat{\mu}^s(s, \xi^s_s - + G^s_s - )\}} Q^s(dt, d\theta). \]
(3.5)

Observe that the previous thinning equation is not a differential equation, since intensities of the processes \(G^d\) and \(G^s\) do not depend on the counting processes themselves. Furthermore, the dominating process \(G\) does not define a population process \(\xi + \phi \odot G\), since \(G\) does not verify the support conditions (1.7).

Existence and uniqueness of BDS SDE By construction, the BDS intensity functional \(\mu\) is dominated by the intensity functional of \(G\), an Theorem 2.1 allows us to conclude:

Theorem 3.1. Assume that the Cox Birth assumption (1) is verified. Then, there exists a unique well-defined solution of Equation (3.1):
\[ dN_t = Q(dt, [0, \mu(t, \xi_t - + \phi \odot N_t - )]), \] with \(Z_t = \xi_t + \phi \odot N_t\), defining a BDS system \((\xi, N, Z)\) of BDS intensity functional \(\mu\) and entry process \(\xi\).

Furthermore, \(N\) is strongly dominated by the non-exploding dominating process \(G\) solution of (3.4)-(3.5): \(N \prec G\).

Since \(N^b \prec G^b\), all jumps of \(N^b\) are jumps of \(G^b\). This can interpreted as follow: “all individuals born in the population \(Z\) are also born in the Cox Birth population \(\xi + G^b\)”.

3.2 BDS decomposition algorithm

There are many advantages in the construction by strong domination, since jump times of the jumps counting process are all jump times of the dominating process \(G\) which does not depend on \(N\). However, the same property can be a drawback when simulating the BDS system by strong domination. Indeed, the dominating process \(G\) can have much more jumps that \(N\), making the simulation inefficient.

An alternative construction of the solution of (3.1) is presented in the following, called the Birth Death Swap decomposition algorithm and based on the disentanglement of swap and demographic...
events. The construction is better suited to the simulation of BDS systems when swap and demographic intensities are of different nature, for instance when they have their own timescale. Let us consider a BDS system solution of (3.1). Swap events can be distinguished from demographic events. The construction is better suited to the simulation of BDS systems when swap and demographic events. The main idea to disentangle swap and demographic events is to generalize the decomposition of population with only one demographic event given in [12,3]

**Decomposition of the BDS system at the first demographic event**

a) Let $T_1$ be the first demographic event. On the first demographic interval $[0, T_1]$, the demographic counting process is null and the population process behaves as a pure swap process since $Z_t = \xi^\circ + \phi^\circ \circ N_0^s + N_1^b - N_1^d$, with $dN^s = Q^s(dt, 0, \mu^s(t, \xi^\circ + \phi^\circ \circ N_0^s))$. Since $Q^s$ doesn’t jump at $T_1$, the process $X^{0,1}_t = \xi^\circ + \phi^\circ \circ N_{t \land T_1}$ is continuous at $T_1$, and can be seen as a pure swap process $X^0$ stopped at $T_1$. The definition of $X^0$ is not unique. The most straightforward construction might be to consider a swap of intensity functional $\mu^s(t, z)I_{(t \leq T_1)}$, null on $[T_1, \infty] = \{N_{\infty}^\infty > 0\}$. However, our goal is to disentangle swap and demographic events, and we prefer to define $X^0$ as the solution of the following Swap multivariate SDE:

$$X^0_t = \xi^\circ + \phi^\circ \circ N^{sw,0}_t, \quad dN^{sw,0}_t = Q^s(dt, 0, \mu^s(t, X^0_0)).$$

By unicity of the previous multivariate SDE, $X^{0,1}_t = X^{0,1}_{0 \land T_1}$ and the population process coincides with the pure Swap process $X^0$ on the first demographic interval.

b) At the first demographic event time $T_1$, the population process jump of $N_{T_1}^b - N_{T_1}^d$. By definition, one and only one component of the vector $N^{dem}$ jumps at $T_1$ and an individual is either added (birth) or removed (death) to one of the subgroups. The important point is that the first demographic event is completely characterized from the Poisson measure $Q^{dem}$ and the Swap process $X^0$, since $dN^{dem}_t = Q^{dem}(dt, 0, \mu^{dem}(t, Z\infty_\infty))$ and $Z\infty_\infty = X^{0,\infty}_\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\in\i
system by pasting Swap processes. Swap trajectories are stopped using the Switch processes, which define demographic events, as well as the starting state of a new Swap process. The construction of the process \( N_{\text{dem}} \) counting demographic events can be seen as a generalization of the construction of Markov Birth processes, where Poisson processes are replaced by Switch processes. Furthermore, a parallel can be drawn with the classical construction of Piecewise Deterministic Markov Processes (PDMP) (Davis (1984)), where Swap processes play the role of the deterministic evolution between two jump times (demographic events times). In particular, the thinning construction allows us to give a pathwise version of the usual construction.

4 Averaging result in presence of two timescales

The study aggregated population process \( Z^\natural = \sum_{i=1}^{p} Z^i \) is of particular interest, since its birth and death intensity give information of the average birth/death rate of an individual taken at random in the population. However, its evolution may be quite complex, since \( Z^\natural \) is not a “true” Birth Death process: its birth and death intensities functionals \( \mu^{b,\natural} \) and \( \mu^{d,\natural} \) depend on the whole structure of the population, and the population composition is not constant between two demographic events, due to swap events. Nevertheless, as it is often the case, the complexity of the aggregated dynamic can be reduced in the presence of two timescales.

We recall that the aggregated process can be written as a simple function of the demographic counting process \( N_{\text{dem}} = (N^b, N^d) \) counting the number of demographic events in the population:

\[
Z^\natural_t = \xi^\natural_t + N^b_t - N^d_t, \quad \forall t \geq 0.
\]

This richer \( 2p \)-multivariate counting process is the process of interest in the following.

4.1 Two timescales BDS system

From now on, swap events are assumed to occur at a much faster timescale than demographic events. Intuitively, this means that demographic events happen with low intensities of order “\( O(1) \)”, in comparison with swap events occurring with intensities of order “\( O(\frac{1}{\epsilon}) \)”, depending on a small parameter \( \epsilon \). Figure 1 gives an example of the distribution of the different types of events. In this case, swap events can have an averaging effect on demographic intensities, allowing for the aggregated population to be approximated by a simpler (although non-linear) dynamic.

![Figure 1: Example of distribution of swap events and demographic events](image)

More formally, the BDS system is now depending on a small parameter \( \epsilon \) and is denoted by
$\left(\xi_t, N^\epsilon, Z^\epsilon\right)$. The BDS multivariate SDE (3.1) becomes:

$$Z_t^\epsilon = \xi_t + \phi^\epsilon \odot N_t^{s,\epsilon} + N_t^{b,\epsilon} - N_t^{d,\epsilon}, \quad N_t^{\text{dem},\epsilon} = (N_t^{b,\epsilon}, N_t^{s,\epsilon})$$

$$\text{d}N_t^{b,\epsilon} = Q^\epsilon(\text{d}t, [0, 1 - \mu^b(t, Z_t^\epsilon)]), \quad \text{d}N_t^{s,\epsilon} = Q^\epsilon(\text{d}t, [0, 1 - \mu^s(t, Z_t^\epsilon)]) \quad (4.1)$$

The construction of the two timescales BDS differs from constructions proposed in the literature. Indeed, the fast evolution is often modeled by a change of time and adapted to a filtration indexed by $\epsilon$. Here, the thinning construction (4.1) of the two timescales BDS system uses the space component $\theta$ of the Poisson measure $Q$, and all processes are adapted to the given filtration $(\mathcal{G}_t)$. 

**DOMINATING COUNTING PROCESS** The solution of (4.1) is still obtained by strong domination with a dominating Cox-Birth process, now denoted by $G$ and functional of $G$ still defined by Equations (3.4) and (3.5), which are not modified. In the following, the intensity product probability measure $P$ is uniformly strongly dominated by the multivariate counting process $N^{\text{dem}}$ of processes strongly dominated by $G$. The critical property of the dominating process $G^\epsilon$ is that its demographic part $G^{\text{dem}} = (G^b, G^d)$ does not depend on $\epsilon$: $G^b$ and $G^d$ are still defined by Equations (3.4) and (3.5), which are not modified. In the following, the intensity functional of $G^{\text{dem}}$ is denoted by $\hat{\lambda}^{\text{dem}}$. As a consequence, all processes $N^{\text{dem},\epsilon}$ are in the class of processes strongly dominated by $G^{\text{dem}}$ and by application of Theorem 2.1 and Corollary 2.2:

**Lemma 4.1** (Uniform strong domination). The family of demographic counting processes $(N^{\text{dem},\epsilon})_{\epsilon > 0}$ is uniformly strongly dominated by the multivariate counting process $G^{\text{dem}}$ solution of (3.4)-(3.5).

In particular:

(i) The sequence $(S_p)_{p \geq 0}$ of jump times of $G^{\text{dem}}$ is a universal localizing sequence of the local martingales $(N_{t}^{\text{dem},\epsilon} - \int_0^t \mu^{\text{dem}}(s, Z_s^\epsilon) \text{d}s)$.

(ii) Furthermore, the aggregated processes $(Z^{\epsilon,\hat{\lambda}})$ are bounded by $\xi^\epsilon + G^{b,\epsilon}$, and

$$\mu^{\text{dem}}(t, Z_t^\epsilon) \leq \hat{\lambda}^{\text{dem}}(t, G_t^{b,\epsilon}), \quad \forall t \geq 0, \epsilon > 0. \quad (4.2)$$

Thus, the swap counting process $N^{s,\epsilon}$ and the demographic counting process $N^{\text{dem},\epsilon}$, which are entangled due to their dependence on $Z_t^\epsilon = F(N_t^{\text{dem},\epsilon}, N_t^{s,\epsilon})$, have very different behavior:

- $N^{s,\epsilon}$ evolves on a fast timescale. Its intensity is proportional to $\frac{1}{\epsilon}$ and the process explodes when $\epsilon \to 0$.
- $N^{\text{dem},\epsilon}$ only depends on $\epsilon$ through $Z^\epsilon$ and is controlled by $G^{\text{dem}}$.

The aim of this section is the study of the convergence of the BDS system (4.1) when swap events become instantaneous with respect to demographic events ($\epsilon \to 0$). We consider the convergence of the demographic counting processes $N^{\text{dem},\epsilon}$ as dynamic processes, i.e. viewed as random variables taking values in the space of 2p-multivariate counting processes.

The situation is different for the family of population processes $(Z^\epsilon)_{\epsilon > 0}$, which is not tight in the space of càdlàg pure jump processes on $\mathbb{N}^p$, due to the explosion of swap events. However, the tightness of $(Z^\epsilon)$ is obtained in a weaker framework, by considering population processes as $\mathbb{N}^p$-valued random variable $Z^\epsilon(\omega, s)$ defined on the product space $\Omega \times \mathbb{R}^+$ (equipped with the product probability measure $P \otimes \lambda^\epsilon$, where $\lambda^\epsilon(\text{d}t) = e^{-\epsilon t} \text{d}t$), rather than dynamic processes.

In the non-Markov framework of BDS systems, intensities are random functionals of the population, and classical averaging results based on the convergence in distribution, such as Kurtz
cannot be applied here. Nevertheless, the notion of stable convergence allows us to overcome these difficulties. For ease of reading, we start in Subsection 4.2 by giving an overview of this mode of convergence, whose advantage is that no regularity assumptions are needed on the dependence on \((\omega, t)\) of the intensity functionals.

Combined with the construction by strong domination, the use of the stable convergence is the cornerstone of the main result of this section. The stable convergence is first applied to the population and the demographic counting processes with different point of views, in order to realize stable limits of the demographic counting processes and identify their intensity. At the end of the section, a general identification result is proven, identifying stable limits of the demographic counting processes of \((N_{\text{dem}}, \epsilon)\) with multivariate counting processes of average intensity, corresponding to the demographic intensity functional \(\mu_{\text{dem}}\) averaged against stable limits of the population process \(Z^\epsilon(\omega, s)\).

### 4.2 Overview on the stable convergence and space of rules

Originated by Alfred Rényi, the notion of stable convergence, which is stronger than the convergence of in distribution, is used in many limit theorems in probability and statistics in random environment, and has also been used to study the convergence of counting processes (see e.g. Brown (1981) or Jacod (1987)). Some useful characterizations and properties of the stable convergence may be found in Jacod and Mémin (1981), Jacod and Shiryaev (2003) or Aldous et al. (1978). A very detailed presentation is also given in the recent book of Häusler and Luschgy (2015). The introduction of this mode of convergence can be motivated in different ways. The point of view best suited to our purpose defines the stable convergence as a convergence of probability measures defined on an enlarged space, which allows for the randomness of the initial structure to be taken into account.

#### 4.2.1 Enlarged space and space of rules

The initial probabilistic structure is a given probability space \((\Omega, \mathcal{G}, P)\), with \(b\mathcal{G}\) the space of bounded \(\mathcal{G}\)-measurable variables. We add to this structure a Polish space \(\mathcal{X}\), equipped with its Borel \(\sigma\)-field \(\mathcal{B}(\mathcal{X})\), assumed to be the state space of the random variables of interest.

A natural extension preserving the initial structure is to consider the enlarged product space \((\bar{\Omega}, \bar{\mathcal{G}}) = (\Omega \times \mathcal{X}, \mathcal{G} \otimes \mathcal{B}(\mathcal{X}))\), where the “canonical” variable on \(\mathcal{X}\) is denoted by \(\Upsilon(\omega, \chi) = \chi\). Admissible probability measures on \((\bar{\Omega}, \bar{\mathcal{G}})\), called rules, are probability measures whose marginal on \((\Omega, \mathcal{G})\) is the given probability measure \(P\). The set of rules on \((\bar{\Omega}, \bar{\mathcal{G}})\) is denoted by \(\mathcal{R}(P, \mathcal{X})\).

An example is given by the rule \(R^Y\) associated with an \(\mathcal{X}\)-valued \(\mathcal{G}\)-random variable \(Y\), defined by

\[
R^Y(H) = \int_{\Omega \times \mathcal{X}} R^Y(d\omega, d\chi) H(\omega, \chi) = \int_\Omega P(d\omega) H(\omega, Y(\omega)) = E[H(., Y)],
\]

for any \(\mathcal{G}\)-r.v. \(H(\omega, x)\). In particular, the \(R^Y\)-probability distribution of the canonical variable \(\Upsilon\) is the \(P\)-probability distribution \(\mu^Y\) of \(Y\).

**Lemma 4.2** (Rules disintegration and immersion). Recall that \(\mathcal{X}\) is a Polish space.

(i) Any rule \(R \in \mathcal{R}(P, \mathcal{X})\) can be disintegrated into a transition probability kernel \(\Gamma(\omega, d\chi)\) from
(Ω, ℓ) to (X, ℬ(X)), which is the ℛ-conditional distribution of the canonical variable Υ given ℓ:

\[ R(\omega, d\chi) = P(\omega)\Gamma(\omega, d\chi), \quad R(h(\Upsilon)|\ell) = \Gamma(\cdot, h), \quad P - a.s. \quad (4.4) \]

For example, if R^ℓ is a rule associated with a random variable Y, Γ^ℓ(ω, dχ) = δ_\Upsilon^\ell(\omega)(dχ).

(ii) Immersion property Let \mathcal{F} \subset \mathcal{G} be a σ-field. A rule R \in \mathcal{R}(P, X) is said to have the \mathcal{F}-immersion property if

\[ \forall U \in b\mathcal{G}, \ h \in C_b(X), \ R(U h(Y)) = R(E(U|\mathcal{F})h(Y)) \quad P-a.s. \quad (4.5) \]

An equivalent formulation in term of \mathcal{G}-transition kernel \Gamma is that Γ(\cdot, d\chi) is \mathcal{F}-measurable.

Proof. (i) For any \mathcal{G}-variable H(\omega, x) = U(\omega)h(\chi),

\[ \int_{\Omega \times X} U(\omega)h(\chi)R(d\omega, d\chi) = \int_{\Omega} U(\omega)dP(h(Y)|\mathcal{G}). \]

Since X is a Polish space, there exists a regular version of the conditional expectations R(h(\Upsilon)|\mathcal{G}), i.e. a \mathcal{G}-kernel Γ, such that R(h(\Upsilon)|\mathcal{G}) = Γ(\cdot, h), P a.s.. In particular, R(U h(Y)) = E[U\Gamma(h)].

(ii) Let R be a rule verifying the immersion property, E[U\Gamma(h)] = E[E[U|\mathcal{F}]\Gamma(h)]. Then,

\[ E[E[U|\mathcal{F}]\Gamma(h)] = E[E[U|\mathcal{F}]E[\Gamma(h)|\mathcal{F}]] = E[U E[\Gamma(h)|\mathcal{F}]]. \]

Since this equality holds true for any \mathcal{G}-variable U, then E[\Gamma(h)|\mathcal{F}] = Γ(h), P.a.s.. Since X is a Polish space, this property implies that Γ is \mathcal{F}-measurable. \qed

4.2.2 Stable convergence of rules

The different modes of convergence of probability measures are usually characterized by their family of tests functions. For example, test functions characterizing the convergence in distribution are bounded continuous functions on X, here denoted by C_bc(X), (C_b(X) for bounded functions). The stable convergence can be interpreted as an extension of the weak convergence to the space of rules. The class of test functions is extended to the family C_{bmc}(Ω × X), of b\mathcal{G} functions \mathcal{H}(\omega, \chi), continuous in \chi for any \omega, but without any regularity in \omega. In particular, no topological structure is required on the space Ω. In the sequel, we sometimes represent a rule R by its kernel Γ and write R = (P, Γ).

Definition 4.1 (Stable convergence). (i) A sequence of rules (R^n) converges stably to R \in \mathcal{R}(P, X) iff

\[ \forall H \in C_{bmc}(\Omega \times X), \ R^n(H) = E[\Gamma^n(H)] \text{ converges to } R(H) = E[\Gamma(H)]. \]

(ii) Let \mathcal{F} \subset \mathcal{G} be a σ-field. The stable limit R of a sequence of rules (R^n) satisfying the \mathcal{F}-immersion property also satisfies this property and its kernel Γ is \mathcal{F}-measurable.

Remark 4.1. We recall that a sequence of random variables \( (\xi_n) \) is said to converges to \( \xi \) weakly in \( L^1(\Omega, \mathcal{G}, P) \), or for the \( \sigma(L^1, L^\infty) \)-topology, iff for any \( U \in b\mathcal{G} \), \( E[U \xi_n] \to E[U \xi] \). Thus, \( (R^n) \) converges stably to R iff \( (\Gamma^n(h)) \) converges to \( \Gamma(h) \) weak-L^1(\Omega, \mathcal{G}, P), \( \forall h \in C_{bmc}(X) \). However, this does not mean that \( (\Gamma^n) \) converges in distribution (as measured-valued variables) to \( \Gamma \), since the weak-L^1 convergence of \( (\Gamma^n(h)) \) does not imply that \( (\Gamma^n(h)) \) converges in distribution to \( \Gamma(h) \). Thus, the interpretation of the stable convergence as a convergence of random measures is different from the usual weak convergence of random measures. \cite{Kallenberg(1973)}.
Some extensions of the stable convergence Many useful properties true for the weak convergence can be extended to the stable convergence:

(i) “Porte-manteau” theorem: let \( F \in \mathcal{G} \), with closed sections \( F_\omega = \{ y; (\omega, y) \in F \} \), \( \forall \omega \in \Omega \). Then, if \( (R^n) \) converges to \( R \) in \( \mathcal{R}(P, \mathcal{X}) \), \( \limsup_n R^n(F) \leq R(F) \).

In particular, if the rules \( R^n \) have supports in a space \( F \) verifying the previous condition, \( (R^n(F) = 1, \forall n \geq 0) \) then the same is true for the limit rule \( (R(F) = 1) \).

(ii) The stable convergence can be extended to a larger class of test functions:

- The boundedness requirement can be replaced by
  \[
  \lim_{n \to \infty} \sup_n R^n(H(1_{|H| > a})) = 0,
  \]

- The continuity of \( H(\omega, \cdot) \) can be relaxed if there exists a subspace \( A \) with \( R^n(A) \to 1 \) and \( R(A) = 1 \), and such that \( P \)-a.s., the restrictions of \( H(\omega, \cdot) \) to \( A_\omega \) are continuous, \( \Gamma(\omega, \cdot) \)-a.s.

### 4.2.3 Stable convergence of random variables and tightness

When the rules \( (R^n) \) are generated by a sequence of random variables \( (Y_n) \) defined on the same probability space, \( R^n(\mathcal{H}) = E[H(\cdot, Y_n)] \) with \( \Gamma^n(\cdot, d\chi) = \delta_{Y_n}(d\chi) \), the stable convergence of \( (R^n) \) to a rule \( R \) associated with a transition kernel \( \Gamma \) can be reinterpreted as the stable convergence of the random variables \( (Y^n) \), with two viewpoints:

- In the first viewpoint, the stable convergence is considered on the given probability space only, and defined by: for any \( U \in b\mathcal{G} \) and \( f \in C_b(\mathcal{X}) \), \( E[U f(Y_n)] \to E[U \Gamma(f)] \).

- In the second viewpoint, the reference probability space is the enlarged space \( (\bar{\Omega}, \bar{\mathcal{G}}, \bar{R}) \), equipped with the limit rule \( R \). The stable convergence of the sequence \( (Y_n) \), which is naturally extended on \( (\bar{\Omega}, \bar{\mathcal{G}}, \bar{R}) \), is now the stable convergence of \( (Y^n) \) to the canonical variable \( \Upsilon \): \( R[U f(Y_n)] \to R[U f(\Upsilon)] \).

According to the chosen viewpoint, we will alternatively say that \( (Y^n) \) converges stably to the rule \( R \), to the random kernel \( \Gamma \), or to \( \Upsilon \) in \( (\bar{\Omega}, \bar{\mathcal{G}}, \bar{R}) \).

Observe that the extension \((\text{LG})\) of the stable convergence to non-bounded functions can be reinterpreted as a uniform integrability condition for the family of random variables \( (H(\omega, Y_n)) \).

Obviously, the stable convergence of the sequence \( (Y_n) \) implies the convergence in distribution to the marginal \( R_X \) of \( R \) on \( \mathcal{X} \).

Stable relative compactness of random variables One of the most interesting property of the stable convergence is that relative compactness for the stable convergence is “free” (Proposition 2.4, Jacod and Mémin [1981]). In terms of random variables, this means that from any sequence \( (Y_n) \) with distribution \( (\mu_n) \) converging weakly to \( \mu \), there exists a subsequence of \( (Y_n) \) converging stably to a rule \( R \) such that \( \mu = R_X \). Equivalently, \( (R^n) \) is stably relatively compact in \( \mathcal{R}(P, \mathcal{X}) \) iff \( (\mu_n) \) is tight in \( \mathcal{X} \). For simplicity, we also say “\( (Y_n) \) or \( (\Gamma^n) \) is stably relatively compact” and use the abuse of language “\( (Y^n) \) is tight in \( \mathcal{X} \)”.

### 4.3 Stable limits of the random variables \( Z^c \) on \( \Omega \times \mathbb{R}^+ \)

Due to the explosion of swap events, the family of population processes \( (Z^c) \) is not tight in the space of \( \mathbb{N}^p \)-valued càdlàg processes. However, due to the construction by strong domination
(Lemma [1.1], the aggregated processes \(Z_{t}^{\tilde{\xi}} = \xi_{t}^{\tilde{\varepsilon}} + N_{t}^{p,b,c} - N_{t}^{d,\varepsilon}\) are bounded by \(\xi_{t}^{\tilde{\varepsilon}} + \tilde{\varepsilon}_{t}\). Our aim is to define the right probability space on which this property can be used to obtain stable relative compactness properties. The idea is to see the processes \(Z^{\varepsilon}\) not as dynamic processes defined on \(\Omega\) anymore, but rather as \(N^{p}\)-valued random variables \(\tilde{Z}^{\varepsilon}\) depending on \(\tilde{\omega} = (\omega, s)\), defined by \(\tilde{Z}^{\varepsilon}(\omega, s) = Z_{s}^{\varepsilon}(\omega)\).

THE NEW “GIVEN” PROBABILITY SPACE The new reference space is \(\tilde{\Omega} = \Omega \times \mathbb{R}^{+}\), which can be equipped with different \(\sigma\)-fields: the product \(\sigma\)-field \(\tilde{\mathcal{G}} = \mathcal{G} \otimes \mathcal{B}(\mathbb{R}^{+})\), but also with some sub \(\sigma\)-field of \(\tilde{\mathcal{G}}\) such as the optional \(\sigma\)-field \(\mathcal{O}\) (generated by adapted càdlàg processes), or the predictable one \(\mathcal{P}\) (generated by adapted left continuous processes). \((\tilde{\Omega}, \tilde{\mathcal{G}})\) is equipped with the probability measure \(\tilde{\mathbb{P}} = \mathbb{P} \otimes \lambda^{e}\), with \(\lambda^{e}(ds) = e^{-s}ds\).

In this new given space, for any \(\tilde{\mathcal{G}}\)-measurable bounded variable \(\tilde{H}(\tilde{\omega}) = \tilde{H}(\omega, s) = H_{s}(\omega)\),

\[
\mathbb{E}[\tilde{H}(\tilde{\omega})] = \mathbb{E}[^{\infty}_{0} \tilde{H}(\omega, s)\lambda^{e}(ds)] = \mathbb{E}[^{\infty}_{0} e^{-s}\tilde{H}(\omega, s)ds] = \mathbb{E}[^{\infty}_{0} e^{-s}H_{s}(\omega)ds].
\]

4.3.1 Stable relative compactness of the family \(\tilde{Z}^{\varepsilon}\) and stable limits

The tightness of the \(\mathcal{O}\)-measurable variables \(\tilde{Z}^{\varepsilon}\) can be obtained on this new given space. The first application of the stable convergence is based on the first viewpoint, which defines stable limits of the population variables as random kernels \(\tilde{\Gamma}\) from \(\tilde{\Omega}\) to \(N^{p}\). The rules generated by these variables satisfy the immersion property between \(\mathcal{O}\) and \(\tilde{\mathcal{G}}\), and therefore their random kernel \(\tilde{\Gamma}\) are optional as well.

**Proposition 4.3.** The family of \(N^{p}\)-valued random variables \(\tilde{Z}^{\varepsilon}\) defined on \((\tilde{\Omega}, \tilde{\mathcal{G}}, \tilde{\mathbb{P}})\) is stably relatively compact in \(\mathcal{R}(\tilde{\mathbb{P}}, N^{p})\).

(i) The random kernel \(\tilde{\Gamma}(\tilde{\omega}, dz)\) associated with any limit rule is optional.

(ii) Up to a subsequence, the family \(\tilde{Z}^{\varepsilon}(\tilde{\omega})\) converges stably to \(\tilde{\Gamma}\) and for all \(\tilde{H} \in C_{mc}(\tilde{\Omega} \times N^{p})\) verifying the uniform integrability condition (1.0):

\[
\mathbb{E}[\tilde{H}(\tilde{\omega}, \tilde{Z}^{\varepsilon}(\tilde{\omega}))] \to_{\varepsilon \to 0} \mathbb{E}[\int_{N^{p}} \tilde{H}(\tilde{\omega}, z)\tilde{\Gamma}(\tilde{\omega}, dz)] = \mathbb{E}[\int_{0}^{\infty} \tilde{\Gamma}(\omega, s, \tilde{H}(\omega, s, \cdot))\lambda^{e}(ds)].
\]  

(4.7)

**Proof.** Thanks to the uniform domination of \(\tilde{Z}^{\varepsilon}(\tilde{\omega})\) by \(\tilde{\xi}(\tilde{\omega}) + \tilde{\varepsilon}_{t}(\tilde{\omega})\), it directly follows that the family \((\tilde{Z}^{\varepsilon})\) is tight in \(N^{p}\), and is thus stably relatively compact in \(\mathcal{R}(\tilde{\mathbb{P}}, N^{p})\). Since \(Z^{\varepsilon}\) is \(\mathcal{O}\)-measurable for all \(\varepsilon > 0\), any limit kernel \(\tilde{\Gamma}\) is also \(\mathcal{O}\)-measurable by the immersion property. □

4.3.2 Application to the demographic intensity

The stable convergence of \(\tilde{Z}^{\varepsilon}\) in \(\mathcal{R}(\tilde{\mathbb{P}}, N^{p})\) can be reinterpreted as the weak- \(L^{1}(\Omega, \mathcal{G}, \mathbb{P})\) convergence of integrals of functionals of the population processes (with respect to the Lebesgue measure). This point of view is well-suited to the study of the compensators of the demographic counting processes, i.e. the increasing processes \(A^{p}_{t} = \int_{0}^{t} \mu^{dem}(s, Z^{p}_{s})ds\). In particular, the universal localizing sequence \((S_{p})\) is particularly useful, since it allows us to obtain results without the usual making uniform integrability assumptions on the compensators:
Corollary 4.4. Let $\tilde{\Gamma}(\omega,s,dz)(=\Gamma_s(\omega,dz))$ be a stable limit point of $(\tilde{Z}^t)$ in $\mathcal{R}(\mathcal{P},\mathbb{N}^p)$. Then, up to a subsequence and $\forall \, p \geq 0$ and $(\mathcal{G}_t)$-stopping time $\tau$,

$$\mathcal{A}_{\tau}^{\tau_{\wedge}S_p} = \int_{0}^{\tau_{\wedge}S_p} \mu_{\text{dem}}(s,Z_s^t)ds \text{ cv weakly in } L^1(\Omega,\mathcal{G},\mathcal{P}) \text{ to } \int_{0}^{\tau_{\wedge}S_p} \tilde{\Gamma}(s,\mu_{\text{dem}}(s,\cdot))ds. \quad (4.8)$$

At the limit, the demographic intensity is $\mu_{\text{dem}}(s,z)$ is averaged against the optional kernel $\Gamma_s(\cdot,dz)$, that is $\tilde{\Gamma}(s,\mu_{\text{dem}}(s,\cdot)) = \int_{\mathbb{N}^p} \mu_{\text{dem}}(s,z)\Gamma_s(\cdot,dz)$.

Proof. Let $B \in \mathcal{G}$ and $\tau$ a $(\mathcal{G}_t)$-stopping time. Then,

$$E[I_B \int_{0}^{\tau_{\wedge}S_p} \mu_{\text{dem}}(s,Z_s^t)ds] = E[I_B \int_{0}^{\tau_{\wedge}S_p} \mu_{\text{dem}}(\tilde{\omega},Z_s^t(\tilde{\omega}))].$$

Hence, in order to apply Proposition 4.3, we need to prove that the family of random variables $(I_{B \times [0,\tau_{\wedge}S_p]}e^{s\mu_{\text{dem}}(\tilde{\omega},Z_s^t(\tilde{\omega}))})_\epsilon$ satisfies the uniform integrability condition (4.6).

By (ii) of Lemma 4.1, $\mu_{\text{dem}}(s,Z_s^t) \leq \lambda_{\text{dem}}(s,G_s^{\text{dem}})$ and

$$E[I_B \int_{0}^{\tau_{\wedge}S_p} \epsilon \lambda_{\text{dem}}(\cdot,G_s^{\text{dem}})] = E[I_B \int_{0}^{\tau_{\wedge}S_p} \epsilon \lambda_{\text{dem}}(s,G_s^{\text{dem}})ds] = E[G_{\text{dem}}^{\tau_{\wedge}S_p}] \leq p.$$ 

Hence, the family $(I_{B \times [0,\tau_{\wedge}S_p]}e^{s\mu_{\text{dem}}(\tilde{\omega},Z_s^t(\tilde{\omega}))})_\epsilon$ is dominated by a random variable in $L^1(\Omega,\mathcal{G},\mathcal{P})$, and is therefore uniformly integrable.

Remark 4.2. A number of classical results such as [Kurtz 1992] rely on identifying weak limits in distribution of the compensators. However, the stable convergence of $Z^t$ in $\mathcal{R}(\mathcal{P},\mathbb{N}^p)$ does not imply the convergence in distribution of $A^t_\epsilon$ to the the averaged process. Indeed, the weak-$L^1$ convergence does not imply the convergence in distribution, and the reverse is not true either. Moreover, if both modes of convergence are true, the weak-$L^1$ limit does not necessarily have the same distribution than the limit distribution. In fact, this is true only if convergence in probability holds ([Jajte and Paszkiewicz 1999]). Nevertheless, the weak-$L^1$ convergence preserves martingale properties, and this property is enough to obtain results.

4.4 Stable limits of the demographic counting processes

We now turn to the study of the demographic counting processes $(N_{\text{dem},\epsilon}^t)$. Contrary to the population processes $Z^t$, the demographic counting processes only depend on $\epsilon$ through $Z^t$, and can be considered as dynamic processes, that is as random variables on $(\Omega,\mathcal{G},\mathcal{P})$, taking values in the subspace $\mathcal{A}^{2p}$ of the Skorohod space $\mathcal{D}(\mathbb{R}^+,\mathbb{N}^{2p})$, composed of $\mathbb{N}^{2p}$-valued functions whose components only have unit jump and no common jump ([Jacod and Shiryaev 2003] for more details on the space $\mathcal{A}^{2p}$). The key to the tightness of the demographic counting processes is again the strong domination by $G_{\text{dem}}$.

Lemma 4.5 (Tightness of the demographic counting processes). The family of demographic counting processes $(N_{\text{dem},\epsilon}^t)$ is tight in $\mathcal{A}^{2p}$ (endowed with the induced Skorohod topology). Equivalently, $(N_{\text{dem},\epsilon}^t)$ is also stably relatively compact in $\mathcal{R}(\mathcal{P},\mathcal{A}^{2p})$.

Proof. By Lemma 4.1, $N_{\text{dem},\epsilon}^t \prec G_{\text{dem},\epsilon}^t$, for all $\epsilon > 0$, and the tightness of $N_{\text{dem},\epsilon}^t$ can obtained by application of Proposition VI.3.35 in [Jacod and Shiryaev 2003] on the tightness of strongly dominated increasing processes. The tightness in $\mathcal{A}^{2p}$ of $(N_{\text{dem},\epsilon}^t)$ is a consequence of Proposition VI.3.36 in [Jacod and Shiryaev 2003].
In this second application of the stable convergence, we are now interested in realizing stable limits of demographic counting processes on an enlarged space, in order to preserve the structure of the problem, in particular the strong domination by $G^{\text{dem}}$.

Set of rules Using notations of Section 4.2, the given space is $(\Omega, \mathcal{G}, P)$, and $\mathcal{X}$ is the space of functions $\mathcal{A}^2_p$, equipped with the Skorohod topology. The canonical filtration of $\mathcal{A}^2_p$ is denoted by $\mathcal{F}_t^A = (\sigma(\alpha(s); s \leq t, \alpha \in \mathcal{A}^2_p))$. The canonical variable on $\mathcal{A}^2_p$ is denoted by $\mathcal{N}^{\text{dem}}(\omega, \alpha) = \alpha$.

On the enlarged filtered space $(\bar{\Omega}, (\bar{\mathcal{G}}_t)) = (\Omega \times \mathcal{A}^2_p, (\mathcal{G}_t \otimes \mathcal{F}_t^A))$, the set of admissible probability measures are the rules $R$ in $\mathcal{R}(P, \mathcal{A}^2_p)$, and tests functions are the product functions $1_B h(N^{\text{dem}})$ with $B \in b\mathcal{G}$ and $h \in C_{bc}(\mathcal{A}^2_p)$.

By the support property deduced from the “portemanteau” inequality, the subset $B \subset \bar{\Omega}$ with $\mathcal{P}(B) \in \mathcal{R}(\mathcal{A}^2_p)$ measures are the rules $\mathcal{G}$ in $\mathcal{R}(P, \mathcal{A}^2_p)$ of probability measures with support included in the domain $F = \{ (\omega, \alpha) \in \bar{\Omega}; \alpha \prec G^{\text{dem}}(\omega) \}$ (whose sections in $\omega$ are closed in of $\mathcal{A}^2_p$) is closed for the stable convergence. Thus, for any rule $R \in \mathcal{R}^G(P, \mathcal{A}^2_p)$, $\mathcal{N}^{\text{dem}}$ is $R$ a.s. strongly dominated by $G^{\text{dem}}$.

Realization of stable limits of $N^{\text{dem}, \epsilon}$. For all $\epsilon > 0$, $N^{\text{dem}, \epsilon}$ is strongly dominated by $G^{\text{dem}}$, i.e. $R^\epsilon = R^{N^{\text{dem}, \epsilon}} \in \mathcal{R}^G(P, \mathcal{A}^2_p)$. For each sequence $(N^{\text{dem}, \epsilon_k})$ converging in distribution there exists a subsequence (still denoted with the same notations) converging stably to a rule $R \in \mathcal{R}^G(P, \mathcal{A}^2_p)$, whose marginal on $\mathcal{A}^2_p$ is the limit distribution of $(N^{\text{dem}, \epsilon_k})$. Thus, $(N^{\text{dem}, \epsilon_k})$ converges stably to $N^{\text{dem}}$ in $\mathcal{G}$ and $N^{\text{dem}} \prec G^{\text{dem}}$, $R$-a.s.

This representation of limit distributions is particularly interesting since it allows the structure of the primary problem to be maintained. Jump times of $N^{\text{dem}}$ are also jump times of $G^{\text{dem}}$ and in particular $N^{\text{dem}}$ has an intensity.

The following technical result is used to prove the main result of this section.

Lemma 4.6. Let $x \in \mathcal{A}^2_p$ and let $(t^x_p)_{p \geq 1}$ be the increasing sequence of jump times of $x$, with $t^x_p \to +\infty$. Let $A^x = \{ \alpha \in \mathcal{A}^2_p; \alpha \prec x \}$. Then, for all $p \geq 1$, the function $\pi_{t^x_p} : \alpha \in A^x \to \alpha(t^x_p) \in \mathcal{N}^p$ is a continuous function on $A^x$.

Proof. Let $\alpha \in A^x$ and $p \geq 1$. we recall that $\alpha \to \alpha(t)$ is a continuous function on $\mathcal{A}^2_p$ (embedded with the Skorohod topology) iff $t$ is not a jump time of $\alpha$. Hence, if $t^x_p$ is not a jump time of $\alpha$, $\pi_{t^x_p}$ is continuous in $\alpha$.

Let us now consider the case when $\Delta \alpha(t^x_p) \neq 0 \Leftrightarrow \Delta \alpha(t^x_p) = \Delta x(t^x_p)$ since $\alpha \prec x$. By observing that $t^x_p$ is the first jump time of $\alpha(t^x_p) = \alpha(t^x_p)$, we can only consider the case when $p = 1$. In this case, this means that $t^x_1$ is the first jump time of $\alpha$ and $\alpha(t^x_1) = x(t^x_1)$. Let $(\alpha_n)$ be a sequence in $A^x$ converging to $\alpha$, and let $T^n_1$ be the first jump time of $\alpha_n$. Since $\alpha_n \to \alpha$, $T^n_1 \to_n t^x_1$. Furthermore, $\alpha_n$ is strongly dominated by $x$ for all $n \geq 0$, and hence the sequence $(T^n_1)$ is a subset of the discrete space $\{ t^x_p; p \geq 0 \}$. Thus, $(T^n_1)$ is necessarily constant equal to $t^x_1$ above a given rank $N$. This means that for all $n \geq N$, $t^x_1$ is also the first jump time of $\alpha_n$, so that $\pi_{t^x_1}(\alpha_n) = \alpha_n(t^x_1) = x(t^x_1) = \alpha(t^x_1)$. Therefore $\pi_{t^x_1}$ is continuous in $\alpha$. \qed
4.5 Joint convergence and identification result

Let us now state the main result of this section, which characterizes stable limits of the demographic counting processes. To that matter, we shall need to consider the joint stable convergence of the processes \((\mathbf{N}^{\text{dem}})\) and the random variables \((\bar{Z}^t)\). Using the different point of views of the stable convergence, the intensity of stable limits are identified with the demographic intensity functional \(\mu^\text{dem}\), averaged against the conditional kernel of stable limits of \((\bar{Z}^t)\).

For the sake of clarity, functions in \(\mathcal{A}^{2p}\) are denoted by \([\alpha]\), with \([\alpha]_s = \alpha(\cdot \wedge t)\). For any function \(\tilde{h}\) on \(\Omega \times \mathbb{R}^+ \times \mathcal{A}^{2p}\), we also denote by \(\tilde{h}_s\) the function defined by:

\[\tilde{h}_s(\omega, s, [\alpha]) = \tilde{h}(\omega, s, [\alpha]_s), \quad \forall (\omega, s, [\alpha]) \in \Omega \times \mathbb{R}^+ \times \mathcal{A}^{2p}.\]

4.5.1 Extension of \((\mathbf{N}^{\text{dem}, \epsilon})\) to \((\bar{\Omega}, \bar{\mathcal{G}}, \bar{\mathcal{P}})\)

The first step is to define \((\mathbf{N}^{\text{dem}, \epsilon})\) and \((\bar{Z}^t)\) on the same probability space. In Subsection 4.3 the population processes have been considered as \(\mathbb{N}^p\)-valued random variables \(\bar{Z}^t\) on the product space \(\bar{\Omega} = \Omega \times \mathbb{R}^+\), while the demographic counting processes \(\mathbf{N}^{\text{dem}, \epsilon}\) are considered as \(\mathcal{A}^{2p}\)-valued random variables on \(\Omega\). However, the demographic counting processes \(\mathbf{N}^{\text{dem}, \epsilon}\) can also be seen as \(\mathcal{A}^{2p}\)-valued random variables on \(\bar{\Omega}\), by setting

\[\mathbf{N}^{\text{dem}, \epsilon}(\bar{\omega}) = [\mathbf{N}^{\text{dem}, \epsilon}(\omega)]_s, \quad \forall \bar{\omega} = (\omega, s) \in \bar{\Omega}.\]  (4.9)

By definition, \((\mathbf{N}^{\text{dem}, \epsilon})\) is a family of \(\mathcal{O}\)-measurable \(\mathcal{A}^{2p}\)-valued variables, and \(\bar{\mathbb{E}}[\bar{h}(\mathbf{N}^{\text{dem}, \epsilon})] = \bar{\mathbb{E}}[\bar{h}_s(\mathbf{N}^{\text{dem}, \epsilon})]\) for any measurable function \(\bar{h}\). Obviously, \((\mathbf{N}^{\text{dem}, \epsilon})\) is stably relatively compact in \(\mathcal{R}(\mathbb{P} \otimes \lambda^\epsilon, \mathcal{A}^{2p})\) since \(\mathbf{N}^{\text{dem}, \epsilon}(\omega, s) \prec [\mathbf{G}^{\text{dem}}(\omega)]_s\). Any limit rule \(\tilde{\mathbb{R}}^\mathbb{N}\) inherit the immersion property (13) with \(\mathcal{F} = \mathcal{O}\), so that

\[\tilde{\mathbb{R}}^\mathbb{N}(\bar{h}) = \tilde{\mathbb{R}}^\mathbb{N}(\bar{h}_s), \quad \forall \bar{\mathcal{G}} \otimes \mathcal{F}_\infty^\mathbb{A}\)-measurable \(\bar{h}\).

Limit rules \(\tilde{\mathbb{R}}^\mathbb{N}(d\bar{\omega}, d[\alpha]) \in \mathcal{R}(\bar{\mathbb{P}}, \bar{\mathcal{A}}^{2p})\) of \((\mathbf{N}^{\text{dem}, \epsilon}(\bar{\omega}))\) can be linked to stable limits \(\mathbb{R}^\mathbb{N}(d\omega, d[\alpha]) \in \mathcal{R}(\mathbb{P}, \mathcal{A}^{2p})\) of \((\mathbf{N}^{\text{dem}, \epsilon}(\omega))\) that have been defined in the previous subsection. We recall that the canonical variable on \(\mathcal{A}^{2p}\) defined on \((\bar{\Omega}, (\bar{\mathcal{G}}_t)) = (\Omega \times \mathcal{A}^{2p}, (\bar{\mathcal{G}}_t \otimes \mathcal{F}_t^\mathbb{A}))\) is \(\mathbf{N}^{\text{dem}}(\omega, \alpha) = \alpha\).

**Lemma 4.7.** \((\mathbf{N}^{\text{dem}, \epsilon}(\bar{\omega}))\) converges stably to \(\tilde{\mathbb{R}}^\mathbb{N}\) in \(\mathcal{R}(\mathbb{P} \otimes \lambda^\epsilon, \mathcal{A}^{2p})\) iff \((\mathbf{N}^{\text{dem}, \epsilon}(\omega))\) converges stably to a rule \(\mathbb{R}^\mathbb{N}\) in \(\mathcal{R}(\mathbb{P}, \mathcal{A}^{2p})\). Furthermore, for all \(\bar{\mathcal{G}} \otimes \mathcal{F}_\infty^\mathbb{A}\)-measurable function \(\tilde{h}\),

\[\tilde{\mathbb{R}}^\mathbb{N}(\tilde{h}) = \mathbb{R}^\mathbb{N}[\int_0^\infty \bar{h}_s(s, [\mathbf{N}^{\text{dem}}]) \lambda^\epsilon(ds)].\]  (4.10)

**Proof.** Let us first assume that \((\mathbf{N}^{\text{dem}, \epsilon}(\omega))\) converges stably to a rule \(\mathbb{R}^\mathbb{N}\) in \(\mathcal{R}(\mathbb{P}, \mathcal{A}^{2p})\). Let \(\bar{h} \in \mathcal{C}_{bmc}(\bar{\Omega} \times \mathcal{A}^{2p})\), and let \(H(\omega, [\alpha]) = \int_0^\infty \bar{h}_s(\omega, s, [\alpha]) \lambda^\epsilon(ds)\), so that:

\[\bar{\mathbb{E}}[\bar{h}^\epsilon(\cdot, [\mathbf{N}^{\text{dem}}])] = \mathbb{E}[H^\epsilon(\cdot, [\mathbf{N}^{\text{dem}}])].\]

The function \(\alpha \in \mathcal{A}^{2p} \rightarrow \tilde{h}_s(\omega, s, [\alpha])\) is continuous for all \(s > 0\) such that \(\Delta \alpha(s) = 0\), and since \(\alpha\) only have a countable number of discontinuities, \(H(\omega, \cdot)\) is continuous, i.e. \(H \in \mathcal{C}_{bmc}(\Omega \times \mathcal{A}^{2p})\). By taking stable limits in both sides of the previous equation, we obtain that \((\mathbf{N}^{\text{dem}, \epsilon})\) converges stably in \(\mathcal{R}(\mathbb{P} \otimes \lambda^\epsilon, \mathcal{A}^{2p})\) to the rule \(\tilde{\mathbb{R}}^\mathbb{N}\) defined by Equation (4.10).
Reciprocally, assume that \((\tilde{N}^{\text{dem},\epsilon}(\tilde{\omega}))\) converges stably to a rule \(\tilde{R}^N \in \mathcal{R}(\mathbb{P} \otimes \lambda', \mathbb{A}^{2p})\). Let \(H \in \mathcal{C}_{bmc}(\Omega \times \mathbb{A}^{2p})\), and \(\tilde{h}(\omega, s, [\alpha]) = \frac{1}{\lambda'(t, s)} 1_{\mathbb{A}^{2p}}(s)H(\omega, [\alpha]) \in \mathcal{C}_{bmc}(\Omega \times \mathbb{A}^{2p})\), for \(t \geq 0\). Then,

\[
\mathbb{E}[\tilde{h}(\cdot, [\tilde{N}^{\text{dem},\epsilon}])] = \mathbb{E}\left[\frac{1}{\lambda'(t, s)} \int_t^\infty H(\cdot, [N^{\text{dem},\epsilon}]_{t, s}) \lambda'(ds)\right] = \mathbb{E}[H(\cdot, [N^{\text{dem},\epsilon}])] \quad \forall \epsilon > 0.
\]

By passing to the limit, we obtain that \(\mathbb{E}[H(t, [N^{\text{dem},\epsilon}])]\) converges for any \(\mathbb{G} \otimes \mathcal{F}_t^{A}\)-measurable function \(H_t \in \mathcal{C}_{bmc}(\Omega \times \mathbb{A}^{2p})\), and hence \((N^{\text{dem},\epsilon})\) converges stably in \(\mathcal{R}(\mathbb{P}, \mathbb{A}^{2p})\) (see e.g. Proposition 3.12 in [Häusler and Luschgy (2015)]).

4.5.2 Joint stable convergence of \((\tilde{N}^{\text{dem},\epsilon}, \tilde{Z}^\epsilon)\) and averaging result

By tightness of \((\tilde{N}^{\text{dem},\epsilon})\) and \((\tilde{Z}^\epsilon)\), the family of random vectors \(((\tilde{N}^{\text{dem},\epsilon}, \tilde{Z}^\epsilon))\) is stably relatively compact in \(\mathcal{R}(\tilde{\mathbb{P}}, \mathbb{A}^{2p} \times \mathbb{N}^p)\). Let \(\tilde{R}^{N,Z} \in \mathcal{R}(\tilde{\mathbb{P}}, \mathbb{A}^{2p} \times \mathbb{N}^p)\) be a limit rule. Up to a subsequence also denoted by \(\epsilon\), \(((\tilde{N}^{\text{dem},\epsilon}, \tilde{Z}^\epsilon))\) converges stably to \(\tilde{R}^{N,Z}\). In particular, \((\tilde{N}^{\text{dem},\epsilon})\) converges stably to the marginal \(\tilde{R}^N\) of \(\tilde{R}^{N,Z}\) on \(\Omega \times \mathbb{A}^{2p}\). \(\tilde{R}^{N,Z}\) can be disintegrated with respect to \(\tilde{R}^N\) as follows:

\[
\tilde{R}^{N,Z}(\tilde{g}) = \tilde{R}^N\left(\int_{\mathbb{N}^p} \tilde{g}(\tilde{\omega}, [\alpha], dz) \Gamma(\tilde{\omega}, [\alpha])\right) = \tilde{R}^N\left(\Gamma(\tilde{\omega}, [\alpha], \tilde{g}(\tilde{\omega}, [\alpha], \cdot))\right),
\]

where \(\Gamma\) is a random kernel from \((\tilde{\Omega} \times \mathbb{A}^{2p}, \mathcal{O} \otimes \mathcal{F}_\alpha^A)\) to \(\mathbb{N}^p\), which can be interpreted as the random kernel associated with the stable convergence of \((\tilde{Z}^\epsilon)\), “conditionally to \(\tilde{N}^{\text{dem},\epsilon}\)\).

By Lemma 4.7, the demographic counting processes \(N^{\text{dem},\epsilon}\) also converge stably to a rule \(R^N \in \mathcal{R}(\mathbb{P}, \mathbb{A}^{2p})\). By applying the relation (4.10) to \(\tilde{h}(\tilde{\omega}, [\alpha]) = \Gamma(\tilde{\omega}, [\alpha], \tilde{g}(\tilde{\omega}, [\alpha], \cdot))\) the limit rule \(\tilde{R}^{N,Z}\) can be rewritten as follows:

\[
\tilde{R}^{N,Z}(\tilde{g}) = \mathbb{R}^N\left[\int_0^\infty \Gamma(s, [N^{\text{dem}}]_s, \tilde{g}(s, [N^{\text{dem}}]_s, \cdot)) \lambda'(ds)\right].
\]

In particular, the random kernel \((\Gamma(t, [N^{\text{dem}}]_t, dz))\), viewed as a dynamic random measure, is \((\mathbb{G}_t \otimes \mathcal{F}_\alpha^A)\)-adapted.

The previous decomposition allows us to define stable joint limits of \((\tilde{N}^{\text{dem},\epsilon}, \tilde{Z}^\epsilon)\) in such way that the point of views on the stable convergence of \((\tilde{Z}^\epsilon)\) (convergence to a random kernel), and of \((N^{\text{dem},\epsilon})\) (realization of the stable limit on the enlarged space), as developed in [Lessay and Luschgy (2014)] are preserved. Based on this representation, stable limits of \((\tilde{N}^{\text{dem},\epsilon}, \tilde{Z}^\epsilon)\) are now denoted by \((R^N, \Gamma)\).

**Theorem 4.8 (Identification result).** Let \(R^N \in \mathcal{R}(\mathbb{P}, \mathbb{A}^{2p})\) be a stable limit of \((N^{\text{dem},\epsilon})\). Then, up to a subsequence, \((\tilde{N}^{\text{dem},\epsilon}, \tilde{Z}^\epsilon)\) converges stably to a limit \((R^N, \Gamma)\) as defined above, with \(R^N\) the stable limit of \((N^{\text{dem},\epsilon})\).

Furthermore, the realization \(\tilde{N}^{\text{dem}}\) of the stable limit of \((N^{\text{dem},\epsilon})\) on the enlarged space \((\tilde{\Omega}, (\mathbb{G}_t), R^N)\) is characterized by the following property:

\[
\tilde{N}^{\text{dem}} - \int_0^t \Gamma(s, [N^{\text{dem}}]_s, \mu^{\text{dem}}) ds \text{ is a } (\mathbb{G}_t)\text{-local martingale.}
\]

Thus, \(\tilde{N}^{\text{dem}}\) has the \((\mathbb{G}_t)\)-averaged intensity \(\Gamma(t, [\tilde{N}^{\text{dem}}]_t, \mu^{\text{dem}}) = \int_{\mathbb{N}^2p} \mu^{\text{dem}}(t, z) \Gamma(t, [N^{\text{dem}}]_t, dz)\).
Proof. Let \( \epsilon_k \to 0 \) be a subsequence along which \((\bar{N}_{\text{dem},\epsilon_k})\) converges stably to \(\mathbb{R}\). Then, by Lemma 4.7, \((\bar{N}_{\text{dem},\epsilon_k})\) converges to the rule \(\bar{N}\) defined by (1.10), and we can take a subsequence of \((\epsilon_k)\) (also denoted by \((\epsilon_k)\)) along which \((\bar{N}_{\text{dem},\epsilon_k},Z_{\epsilon_k})\) converges stably to a rule \(\bar{R}(\mathbb{N},\mathbb{Z}) = (\mathbb{R}\mathbb{N},\Gamma)\), as defined in (1.11).

Let \(0 \leq u \leq t\) and \(h_u \in C_{\text{bmc}}(\Omega \times \mathcal{A}^p)\), an \(G_u \otimes \mathcal{F}_u\text{-measurable function } (h_u(\omega, [\alpha]) = h_u(\omega,[\alpha]_u)).\) The martingale property on \((\bar{N}_{\text{dem},\epsilon_k} - A_{\epsilon_k}\mathbb{N}_\mathbb{P})\) gives for all \(p \geq 0:\)

\[
E[h_u(\bar{N}_{\text{dem},\epsilon_k})(\bar{N}_{\text{dem},\epsilon_k} - \bar{N}_{\text{dem},\epsilon_k}^u)] = E[h_u(\bar{N}_{\text{dem},\epsilon_k}) \int_{t \wedge S_p}^{t \wedge S_p} \mu_{\text{dem}}(s, Z^s_{\epsilon_k}) ds]. \tag{4.13}
\]

Let \(H(\omega, \alpha) \equiv h_u(\alpha)(\alpha(t \wedge S_p(\omega)) - \alpha(u \wedge S_p(\omega))).\) The right hand side of (4.13) is equal to \(E[H(\omega, \bar{N}_{\text{dem},\epsilon_k})].\) \(H(\omega, \cdot)\) is not a continuous and bounded function. However, for all \(\epsilon > 0, H(\cdot, \bar{N}_{\text{dem},\epsilon}) \leq \|H\|_\infty(\bar{G}_{\epsilon,\mathbb{P}} - \bar{G}_{\text{dem},\mathbb{P}})\) which is in \(L^1(\Omega, \mathcal{G}, \mathbb{P})\), and thus \(H\) verifies the uniform integrability condition (1.10).

Furthermore, let \(A = \{(\omega, \alpha) \in \Omega \times \mathcal{A}^p; \alpha < \bar{G}^{\text{dem}}(\omega)\) and \(\Delta_\alpha(t) = \Delta_\alpha(u) = 0\}\) be the space of \((\omega, \alpha)\) with \(\alpha\) strongly dominated by \(G^{\text{dem}}(\omega)\) and with no jumps in \(t\) or \(u\). Since \(\bar{N}_{\text{dem},\epsilon} < G^{\text{dem}}\) and \(\bar{N}^{\text{dem}} < \bar{G}^{\text{dem}}, \mathbb{R}^r(A) = E[I_{A}(\bar{N}^{\text{dem},\epsilon})] = \mathbb{R}(A) = 1.\) Moreover, by Lemma 4.6, the restriction of \(H(\omega, \cdot)\) to \(\bar{A}\) is continuous, \(\mathbb{P}\text{-a.s.}\). Thus, the stable convergence of \((\bar{N}_{\text{dem},\epsilon_k})\) to \(\mathbb{R}\) can be applied to \(H\).

Now, let \(\bar{g}(\omega, s, \alpha, z) = I_{[u \wedge S_p, t \wedge S_p]}(\omega, s) e^z h_u(\omega, [\alpha]_u) \mu_{\text{dem}}(\omega, s, z).\) The left hand side of (4.13) is equal to \(E[\bar{g}(\omega, \bar{N}_{\text{dem},\epsilon_k}, \bar{Z}_{\epsilon_k})].\) Using the same argument of uniform integrability than in the proof of Corollary 4.4, the joint stable convergence (4.11) can be applied to \(\bar{g}\). By taking the stable limits in both sides of Equation (4.13), we obtain that:

\[
\mathbb{R}^N[h_u(\bar{N}_{\text{dem}})(\bar{N}_{\text{dem}} - \bar{N}_{u \wedge S_p})] = \mathbb{R}^N[h_u(\bar{N}_{\text{dem}}) \int_{u \wedge S_p}^{t \wedge S_p} \int_{\mathbb{N}^p} \mu_{\text{dem}}(s, z) \Gamma(s, [\bar{N}_{\text{dem}}]_s, dz) ds],
\]

for all \(G_u \otimes \mathcal{F}_u\text{-measurable } h_u,\) which achieves to prove the theorem. \(\square\)

An aggregated limit population process can be defined on \((\bar{\Omega}, \bar{\mathcal{G}}_t, \mathbb{R}^N):\)

\[
\bar{Z}_{s}^2 = \bar{\xi}_{s}^2 + \bar{N}_{s}^{\bar{\Delta}} - \bar{N}_{s}^{\bar{\Delta}}, \quad s \geq 0.
\]

Theorem 4.8 does not give information on the relationship between the random kernel \(\Gamma\) and the limit multivariate counting process \(\bar{N}_{\text{dem}}.\) Actually, the result may be specified. The equation \(\bar{Z}_{s}^{\bar{\Delta}} = \bar{\xi}_{s}^2 + N_{s}^{b,\bar{\Delta}} - N_{s}^{d,\bar{\Delta}},\) linking the aggregated processes \(\bar{Z}_{s}^{\bar{\Delta}}\) to the demographic counting processes \(\bar{N}_{\text{dem}}\) and the initial condition \(\bar{\xi}_{s}^2 = \xi_{s}^2 I_{\{s \geq \tau\}}\) can be translated into a support property of the random kernel \(\Gamma.\)

**Corollary 4.9.** Let \((\mathbb{R}^N, \Gamma)\) be a limit point of the stably relatively compact family \((\bar{N}_{\text{dem},\epsilon}, \bar{Z}_{\epsilon}).\) Then, the support of \(\Gamma(\omega, s, [\bar{N}_{\text{dem}}]_s, dz)\) is included in the space \(U_{Z_{s}^2}\) of populations of size \(Z_{s}^2:\)

\[
\Gamma(\omega, s, [\bar{N}_{\text{dem}}]_s, U_{Z_{s}^2}) = 1 \quad \mathbb{R}^N \otimes ds \text{ a.s..} \tag{4.14}
\]

In particular, for all \(s \leq \tau, \Gamma(\omega, s, [\bar{N}_{\text{dem}}]_s, \cdot) = 0.\)
Proof. Let \( F = \{(\omega, s, \alpha, z); z^\epsilon = \xi^\epsilon_x(\omega) + \alpha b^\epsilon(s) - \alpha d^\epsilon(s)\} \). We have for all \( \epsilon > 0 \):
\[
\mathbb{E}[\mathbb{I}_F(\cdot, \hat{N}^{\text{dem},\epsilon}, \hat{Z}^\epsilon)] = \mathbb{E}[\int_0^\infty \mathbb{I}_{\{Z^\epsilon_t = \xi^\epsilon_x + N^{b,\epsilon} - N^{d,\epsilon}\}} \lambda^\epsilon(ds)] = 1.
\]
The extension of the porte-manteau theorem cannot be directly applied to \( F \), since the sections \( F(\omega, s) \) are not closed in \( \mathcal{A}^{2p} \times \mathbb{N}^p \). However, the extension of the stable convergence to discontinuous functions can be applied again, with \( \tilde{g} = \mathbb{I}_F \) and \( A = \{(\omega, s, \alpha, z); \Delta \alpha(s) = 0\} \). Thus, the stable convergence along a subsequence of \((\hat{N}^{\text{dem},\epsilon}, \hat{Z}^\epsilon)\) to \((\mathbb{R}^N, \Gamma)\) can be applied to \( \tilde{g} \), and hence:
\[
\mathbb{R}^N[\int_0^\infty \mathbb{I}_F \lambda^\epsilon(ds)] = \mathbb{R}^N[\int_0^\infty \Gamma(\cdot, s, [\hat{N}^{\text{dem}}], \mathcal{U}_{Z^\epsilon}^s)\lambda^\epsilon(ds)] = 1.
\]

5 Application to Markov swaps

In the previous section, we characterized limiting distribution of the demographic counting processes in the presence of fast swaps. Without more information on the swap intensity functional, we cannot hope to have a unique limiting distribution. Most of the papers concerned with similar questions are set in a Markov framework, with often ergodic requirements. In the case where all intensities are deterministic and time homogeneous, an averaging result for the demographic counting processes can be derived from Kurtz (1992) or Yin and Zhang (2012), under stationary assumptions for the swap processes (and using the construction by strong domination).

In our setting, we want to restrict only the behavior of the swap components of the population process: for the remainder of this section the swap intensity functional \( \mu^\epsilon \) is assumed to be deterministic and time-homogeneous, \( \mu^\epsilon : z \in \mathbb{N}^p \rightarrow \mu^\epsilon(z) \in \mathbb{R}^{p(p-1)} \). However, no other assumptions than the Cox-Birth domination assumption \( \| \) are made on the demographic intensity functional \( \mu^{\text{dem}}(\omega, t, z) \). The two timescales BDS \((\xi^\tau, N^\epsilon, Z^\epsilon)\) is now solution of:

\[
Z^\epsilon_t = \xi_t + \phi^\epsilon \odot N_t^{s,\epsilon} + N_t^{b,\epsilon} - N_t^{d,\epsilon},
\]

\[
dN_t^{s,\epsilon} = Q^\epsilon(dt, [0, \frac{1}{\epsilon} \mu^\epsilon(Z^\epsilon_t - )]), \quad dN_t^{\text{dem},\epsilon} = Q^{\text{dem}}(dt, [0, \mu^{\text{dem}}(\omega, t, Z^\epsilon_t - )]).
\]

Under additional stationary assumptions for Swap processes, we prove that the demographic counting processes converge in distribution to a multivariate counting process with averaged intensities. Due to the averaging effect of swap events, the birth and death intensity functionals only depend on the aggregated process at the limit. In particular, the aggregated processes \( Z^\epsilon_t^\tau \) converge in distribution to a true (Non-Markov) one dimensional Birth-Death process \( Z^\tau \). Finally, we show on a toy example how non-linearities in the aggregated birth and death intensities can emerged, resulting from a non-trivial aggregation of the subgroups birth and death intensities.

5.1 Markov Swap processes

Swap CTMC A pure swap process (i.e. a population in which only swap events occur), of intensity function \( \mu^s \) and initial state \( \xi_0 \), can be realized as the solution of the following multivariate SDE driven by the \( p(p-1) \) Poisson measure \( Q^s \):

\[
X_t = \xi_0 + \phi^s \odot N_t^{w,\epsilon}, \quad dN_t^{w,\epsilon} = Q^s(dt, [0, \mu^s(X_t - )]).
\]

31
Since swap events don’t change the size of the population, the size of the swap is constant, \( X^2_t = \xi^2_0 \). Thus, conditionally to the initial population \( \xi_0 \), \( X \) has a finite state space and the existence and uniqueness of \( \pi_f \) is trivial. Furthermore,

\[
f(X_t) - f(\xi_0) - \int_0^t \sum_{(i,j) \in \mathcal{J}} (f(X_s + \phi(i,j)) - f(X_s)) \mu^{(i,j)}(X_s) ds \text{ is a } (\mathcal{G}_t)\text{-local martingale.}
\]

Thus, the Swap process \( X \) solution of \( \mathbf{5.2} \) is a \((\mathcal{G}_t)\)-CTMC, of intensity matrix \( L^{sw} \), defined by:

\[
L^{sw}(z) = \sum_{(i,j) \in \mathcal{J}} (f(z + \phi(i,j)) - f(z)) \mu^{(i,j)}(z), \quad \forall z \in \mathbb{N}^p.
\]

**Swap CTMC stationary measures** Swap CTMC, when viewed as Markov processes on the whole state space \( \mathbb{N}^p \), cannot be ergodic processes since if the Swap starts in the space \( \mathcal{U}_n \) of population of size \( n \), it will stay in \( \mathcal{U}_n \) and not visit the entire state space. Nevertheless, restrictions of Swap processes to the spaces \( \mathcal{U}_n \) are from now on assumed to have a unique stationary measure:

**Assumption 2** (Ergodicity of the Swap process on \( \mathcal{U}_n \)). \( \forall n \geq 0 \), The Swap CTMC restricted to \( \mathcal{U}_n \) is assumed to be irreducible. Since \( \mathcal{U}_n \) is finite, this means that the Swap CTMC restricted to \( \mathcal{U}_n \) admits a unique stationary measure denoted by \((\pi(n,dz))_{z \in \mathcal{U}_n}\).

In particular, \( \pi(n,\cdot) \) is the unique probability measure on \( \mathcal{U}_n \) such that:

\[
\pi(n, L^{sw} f) = \sum_{z \in \mathcal{U}_n} L^{sw}(z) \pi(n, dz) = 0, \quad \forall f : \mathcal{U}_n \to \mathbb{R}. \tag{5.3}
\]

### 5.2 Birth Death limit of demographic counting processes

The aggregated process \( Z^{2,\epsilon} \) is not a one dimensional Birth-Death process, due to swap events and since the aggregated birth and death intensities functionals \( \mu^{b,\epsilon} \) and \( \mu^{d,\epsilon} \) depend on the whole structure of the population and not just on \( Z^{2,\epsilon} \). In the setting of deterministic swap intensities and under the stationary Assumption 2, the following theorem shows that the family of demographic counting processes converges in distribution to a multivariate counting process with averaged intensity. Furthermore, due to the averaging effect of swap events occurring between two successive demographic events, the demographic intensities only depend on the aggregated population at the limit.

**Theorem 5.1** (Convergence of the demographic counting processes). **For all** \( \epsilon > 0 \), **let** \((\xi^\tau, \mathbf{N}^\epsilon, Z^\epsilon)\) **be the BDS system solution of** \( \mathbf{5.1} \), **under the stationary Assumption 2**. **Then:**
(i) The family of demographic counting processes \((\mathbf{N}^{\text{dem,}\epsilon})\) converges in distribution in \( \mathcal{A}^{2p} \).
(ii) For any stable limit rule \( \mathbb{R}^\mathbf{N} \in \mathbb{R}(\mathcal{P}, \mathcal{A}^{2p}) \), the limit process \( \bar{\mathbf{N}}^{\text{dem}} \) defined on the enlarged space \((\bar{\Omega}, (\bar{\mathcal{G}}_t), \mathbb{R}^\mathbf{N})\) has the \((\bar{\mathcal{G}}_t)\)-intensity:

\[
\pi(\bar{Z}^\epsilon_t, \bar{\mu}^{\text{dem}}(t, \cdot)), \quad \text{where} \quad \bar{Z}^\epsilon_t = \xi^\epsilon_t + \bar{N}^{b,\epsilon}_t - \bar{N}^{d,\epsilon}_t.
\]

**Corollary 5.2** (Convergence of the aggregated processes to a Birth-Death process). **The family of aggregated processes** \((Z^{2,\epsilon})\) **converges in distribution to the (non-Markov) one dimensional Birth-Death process** \( Z^2 \), **starting in state** \((\xi)\) **and with birth and death intensity functionals respectively defined by:**

\[
\bar{\mu}^{b}(t, n) = \pi(n, \mu^{b,\epsilon}(t, \cdot)), \quad \bar{\mu}^{d}(t, n) = \pi(n, \mu^{d,\epsilon}(t, \cdot)), \quad \forall n \geq 0. \tag{5.4}
\]
The proof of Theorem 5.1 is similar to classical identification results in the Markov case, based on the convergence of well chosen martingales (see e.g. Kurtz (1992), Fournier and Méléard (2004), or Méléard and Tran (2012)). The ideas of the proof are here adapted to the stable convergence framework.

Proof. Let \( \epsilon_k \to 0 \) be a sequence along which \( (N_{\text{dem},\epsilon}) \) converges in distribution. Then, there exits a rule \( R^N \in \mathcal{R}(P, A^{2p}) \) and a subsequence of \( (\epsilon_k) \) along which \( (N_{\text{dem},\epsilon}) \) converges stably (and thus in distribution) to \( N_{\text{dem}}(\omega, \alpha) = \alpha \in (\Omega, \mathcal{G}, R^N) \).

By Theorem 1.8 there also exists a subsequence along which \( (N_{\text{dem},\epsilon}, \tilde{Z}^\epsilon) \) converges stably to a limit \( (R^N, \Gamma) \) in the sense of (4.11), and such that \( N_{\text{dem}} \) has the \( (\mathcal{G}_t) \)-intensity \( \Gamma(t, [N_{\text{dem}}]^t, \mu_{\text{dem}}) \).

In order to prove Theorem 5.1 it is thus sufficient to prove that:

\[
\Gamma(\omega, s, [N_{\text{dem}}]^s, d\omega) = \pi(\bar{Z}^s, d\omega) \, R^N \otimes ds - \text{a.s.}
\]

Since we only want to identify \( \Gamma \), we may assume for the proof of the fact that \( (\bar{N}_{\text{dem},\epsilon}, \tilde{Z}^\epsilon) \) converges stably to \( (R^N, \Gamma) \). Let \( f \) be a bounded function on \( \mathbb{N}^p \), \( p \geq 0 \), \( 0 \leq u < t \) and \( h_u \) a \( \mathcal{G}_u \otimes \mathcal{F}_u^A \)-measurable function in \( C_{\text{boc}}(\Omega \times A^{2p}) \). Then,

\[
\mathbb{E}[h_u(N_{\text{dem},\epsilon}) \left( f(Z^\epsilon_{t\wedge S^p}) - f(Z^\epsilon_{u\wedge S^p}) \right)] =
\mathbb{E}[h_u(N_{\text{dem},\epsilon}) \int_{u\wedge S^p}^{t\wedge S^p} \sum_{\gamma \in J_{\text{dem}}^s} (f(Z^\epsilon_{s} + \phi(\gamma)) - f(Z^\epsilon_{s})) \mu(\gamma, Z^\epsilon_{s}) d\gamma + \mathbb{E}[h_u(N_{\text{dem},\epsilon})] \frac{1}{\epsilon} \int_{u\wedge S^p}^{t\wedge S^p} L^{sw} f(Z^\epsilon_{s}) d\gamma]
\]

By multiplying the equation above by \( \epsilon \) and letting \( \epsilon \to 0 \) we obtain that:

\[
R^N[h_u(\bar{N}_{\text{dem}})] \int_{u\wedge S^p}^{t\wedge S^p} \Gamma(\cdot, s, [\bar{N}_{\text{dem}}]^s, L^{sw} f) d\gamma = 0.
\]

The foregoing yields that \( \int_0^T \Gamma(\cdot, s, [\bar{N}_{\text{dem}}]^s, L^{sw} f) d\gamma \) is included in the space \( U_{\bar{Z}^s} \), the support of \( \Gamma(\omega, s, [\bar{N}_{\text{dem}}]^s, d\omega) \) is included in the space \( U_{\bar{Z}^s} \) verifying this property is \( \pi(\bar{Z}^s, d\omega) \).

5.3 Application

In this last subsection, we illustrate Theorem 5.1 in the simple example of non-linear swap intensities presented in [1.12]. Let us first recall the model.

A population composed of two subgroups is considered. The second subgroup (subgroup 2) has a lower stochastic death intensity, \( \mu^{(d,1)}(t, z) \geq \mu^{(d,2)}(t, z) \). When the population is smaller than a given size \( M \), individuals swap to the favorable subgroup 2 at rate \( k_{12}(z)^\alpha \), \( \alpha > 0 \). When the
population is larger than $M$ access to the subgroup 2 is restricted and individuals swap from 1 to 2 at a lower constant rate $k_2^{M}$. Individuals swap from the favorable subgroup 2 to subgroup 1 at constant rate $k_2$. The swap intensity is thus defined by:

$$\mu^{(1,2)}(z) = k_{12} z^1 (z^2)^{\alpha} \mathbf{1}_{\{z^2 \leq M\}} + k_{12}^{M} z^1 \mathbf{1}_{\{z^2 > M\}}, \quad \mu^{(2,1)}(z) = k_{21} z^2.$$  

**Swap CTMC stationary measure.** The intensity matrix of the Swap CTMC restricted to the space $\mathcal{U}_n$ (of population of size $n$) is defined for all $f : \mathcal{U}_n \to \mathbb{R}$ by,

$$L_n^{sw} f(z) = \begin{cases} 
    k_{12} n^\alpha z^1 (f(z + \mathbf{e}_2 - \mathbf{e}_1) - f(z)) + k_{21} z^2 (f(z + \mathbf{e}_1 - \mathbf{e}_2) - f(z)), & \text{if } n \leq M \\
    k_{12}^{M} z^1 (f(z + \mathbf{e}_2 - \mathbf{e}_1) - f(z)) + k_{21} z^2 (f(z + \mathbf{e}_1 - \mathbf{e}_2) - f(z)), & \text{if } n > M 
\end{cases}$$

Actually, the Swap CTMC can be reinterpreted as follow: all individuals evolve as “independent CTMC” on the state space $\{1, 2\}$, with constant transition rates depending on the initial number of individuals. In particular, this means that the stationary measure of the Swap CTMC restricted to $\mathcal{U}_n$, $\pi(n, \cdot)$, is defined as the distribution of the sum of $n$ i.i.d random variables of distribution $\nu$ defined by:

$$\nu_2(n) = \begin{cases} 
    \frac{k_{12} n^\alpha}{k_{12} n^\alpha + k_{21}} & \text{if } n \leq M \\
    \frac{k_{12}^{M}}{k_{12}^{M} + k_{21}} & \text{if } n > M 
\end{cases},$$

and with $\nu_1(n) = 1 - \nu_2(n)$. In particular, $\pi(n, z^2) = n \nu_2(n)$.

Thus, there can be two situations at the limit. If the size of the population is small enough individuals are able to move more easily to the favorable subgroup 2 which is more populated on average. On the other hand, when the population becomes crowded, the access to subgroup 2 is restricted, the proportion of individual in subgroup 2 becomes smaller.

**Linear death intensities.** Let us now assume that the death intensity functionals $\mu^{(i,\alpha)}(\omega, t, z)$, $i = 1, 2$, are linear, equal to:

$$\mu^{(i,\alpha)}(\omega, t, z) = d^i(\omega, t) z^i.$$  

By Theorem 5.1 the aggregated death intensity of the limit process is:

$$\lambda^d(\omega, t, n) = \pi(n, \mu^d, \nu),$$

where $\mu^d = \mu^{d,1} + \mu^{d,2}$.

In the case of linear death intensities, the previous equation can be rewritten as:

$$\lambda^d(\omega, t, n) = \sum_{i=1,2} d^i(\omega, t) \pi(n, z^i) = n (d^1(\omega, t) (1 - \nu_2(n)) + d^2(\omega, t) \nu_2(n)). \quad (5.5)$$

This shows how non-linearities in the death intensity can emerge in the limit aggregated population, even when death intensities in each subgroup are linear. This results from a non trivial aggregation of the subgroup specific death intensities, due to the swap events. In this particular example, the death rate of individuals in the limit aggregated population - which approximates the behavior of the aggregated population when swaps occur on a faster timescale - depends non trivially on the size of the population. Due to the two regimes of swap events, the individual
death rate in the limit aggregated population is lower when the population is small. Thus, thanks to the approximation of the aggregated process by a simpler Birth Death process in the two timescale framework, we can better understand how swap events modify the behavior of the population, by creating non-linearities at the aggregated level.

References

Aldous, D. J., Eagleson, G., et al. (1978). On mixing and stability of limit theorems. *The Annals of Probability, 6*(2):325–331. 22

Anderson, D. and Kurtz, T. (2015). *Stochastic Analysis of Biochemical Systems*. Stochastics in Biological Systems. Springer International Publishing. 7

Auger, P., Charles, S., Viala, M., and Poggiale, J.-C. (2000). Aggregation and emergence in ecological modelling: integration of ecological levels. *Ecological Modelling, 127*(1):11–20. 2

Auger, P., De La Parra, R. B., Poggiale, J.-C., Sánchez, E., and Nguyen-Huu, T. (2008). Aggregation of variables and applications to population dynamics. In *Structured population models in biology and epidemiology*, pages 209–263. Springer, Berlin. 2

Auger, P., Poggiale, J., and Sánchez, E. (2012). A review on spatial aggregation methods involving several time scales. *Ecological Complexity, 10*:12–25. 2, 5

Bansaye, V. and Méliard, S. (2015). Stochastic models for structured populations. *Mathematical Biosciences Institute Lecture Series. Springer edition*. 12

Bezborodov, V. (2015). Markov birth-and-death dynamics of populations. *arXiv preprint arXiv:1502.06783*. 14

Bhaskaran, B. (1986). Almost sure comparison of birth and death processes with application to m/m/s queueing systems. *Queueing Systems, 1*(1):103–127. 14

Billiard, S., Collet, P., Ferrière, R., Méliard, S., and Tran, V. C. (2016). The effect of competition and horizontal trait inheritance on invasion, fixation, and polymorphism. *Journal of Theoretical Biology, 411*:48–58. 2, 5

Brémaud, P. and Massoulié, L. (1996). Stability of nonlinear Hawkes processes. *The Annals of Probability, pages 1563–1588*. 11

Brown, T. (1981). Compensators and Cox convergence. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 90, pages 305–319. Cambridge Univ Press. 22

Davis, M. H. (1984). Piecewise-deterministic markov processes: a general class of non-diffusion stochastic models. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 353–388. 20

Dowd, J. B. and Hamoudi, A. (2014). Is life expectancy really falling for groups of low socio-economic status? Lagged selection bias and artefactual trends in mortality. *International Journal of Epidemiology, 43*(4):983–988. 2

El Karoui, N., Hadjì, K., and Kaakai, S. (2018). Inextricable complexity of two centuries of worldwide demographic transition: a fascinating modeling challenge. *Hal-01745901*. 2

Elo, I. T. (2009). Social class differentials in health and mortality: Patterns and explanations in comparative perspective. *Annual Review of Sociology, 35*:553–572. 2

Feller, W. (1968). *An introduction to probability theory and its applications*, volume 1. Wiley, New York. 15

Fournier, N. and Méliard, S. (2004). A microscopic probabilistic description of a locally regulated population and macroscopic approximations. *Ann. Appl. Probab., 14*(4):1880–1919. 2, 12, 13

35
Garcia, N. and Kurtz, T. (2008). Spatial point processes and the projection method. In In and Out of Equilibrium 2, pages 271–298. Springer. 12
Garcia, N. L. and Kurtz, T. G. (2006). Spatial birth and death processes as solutions of stochastic equations. ALEA Lat. Am. J. Probab. Math. Stat, pages 281–303. 12
Gjessing, H., Røysland, K. and Pena, E., and Aalen, O. (2010). Recurrent events and the exploding Cox model. Lifetime Data Analysis, 16(4):525–546. 13
Häusler, E. and Luschgy, H. (2015). Stable Convergence and Stable Limit Theorems, volume 74. Springer. 22, 29
Huber, M. (2012). Spatial birth–death swap chains. Bernoulli, 18(3):1031–1041. 2
Jacobsen, M. (1982). Statistical analysis of counting processes, volume 12. Springer, lectures note in statistics edition. 16
Jacod, J. (1987). Sur la convergence des processus ponctuels. Probability Theory and Related Fields, 76(4):573–586. 22
Jacod, J. and Mémin, J. (1981). Sur un type de convergence intermédiaire entre la convergence en loi et la convergence en probabilité. Séminaire de Probabilités XV 1979/80, pages 529–546. 22, 24
Jacod, J. and Shiryaev, A. (2003). Limit theorems for stochastic processes, volume 288 of Grundlehren der mathematischen Wissenschaften. Springer–Verlag Berlin Heidelberg. 22, 26
Jajte, R. and Paszkiewicz, A. (1999). Conditioning and weak convergence. Probability and Mathematical Statistics, 19(2):453–461. 26
Kallenberg, O. (1975). Random Measures. Schriftenreihe. Akad.-Verlag. 23
Kurtz, T. G. (1992). Averaging for martingale problems and stochastic approximation. In Applied Stochastic Analysis, pages 186–209. Springer Berlin Heidelberg. 3, 21, 26, 33
Marvá, M., Moussaoui, A., de la Parra, R. B., and Auger, P. (2013). A density-dependent model describing age-structured population dynamics using hawk-dove tactics. Journal of Difference Equations and Applications, 19(6):1022–1034. 3
Massoulié, L. (1998). Stability results for a general class of interacting point processes dynamics, and applications. Stochastic Processes and their Applications, 75(1):1–30. 7, 11, 12
Méléard, S. and Tran, V. (2012). Slow and fast scales for superprocess limits of age-structured populations. Stochastic Processes and their Applications, 122(1):250–276. 3, 33
National Research Council (2011). Explaining divergent levels of longevity in high-income countries. National Academies Press. 2
Olshansky, S. J., Antonucci, T., Berkman, L., Binstock, R. H., Boersch-Supan, A., Cacioppo, J. T., Carnes, B. A., Carstensen, L. L., Fried, L. P., Goldman, D. P., Jackson, J., Kohli, M., Rother, J., Zheng, Y., and Rowe, J. (2012). Differences in life expectancy due to race and educational differences are widening, and many may not catch up. Health Affairs, 31(8):1803–1813. 2
Preston, C. (1975). Spatial birth and death processes. Advances in Applied Probability, 7(3):465–466. 13
Rolski, T. and Szekli, R. (1991). Stochastic ordering and thinning of point processes. Stochastic Processes and their Applications, 37(2):299–312. 13
Taylor, J. E. and Véber, A. (2009). Coalescent processes in subdivided populations subject to recurrent mass extinctions. Electron. J. Probab, 14:242–288. 3
Yin, G. and Zhang, H. (2004). Two-time-scale markov chains and applications to quasi-birth-death queues. SIAM Journal on Applied Mathematics, 65(2):567–586. 3
Yin, G. and Zhang, Q. (2012). Continuous-time Markov chains and applications: a two-time-scale approach, volume 37. Springer Science & Business Media. 4, 41