Conditional limit theorems for multitype branching processes and illustration in epidemiological risk analysis

Sophie Pénisson

Institut für Mathematik der Universität Potsdam
Lehrstuhl für Wahrscheinlichkeitstheorie
Ecole Doctorale Santé Publique Paris-Sud 11 Paris Descartes

CONDITIONAL LIMIT THEOREMS FOR MULTITYPE BRANCHING PROCESSES AND ILLUSTRATION IN EPIDEMIOLOGICAL RISK ANALYSIS

Dissertation / Thèse de Doctorat
eingereicht von / présentée par
Sophie Pénisson

zur Erlangung des akademischen Grades
Doctor rerum naturalium
- Dr. rer. nat. -
in der Wissenschaftsdisziplin Stochastik[1]

pour obtenir le grade de
Docteur
dans le domaine Mathématiques Appliquées[2]

am 16. Juli 2010
an der Mathematisch-Naturwissenschaftlichen Fakultät
der Universität Potsdam.

Prüfungskommission / Membres du jury

Prof. Dr. Gerold Alsmeyer	*Gutachter / Rapporteur*
Dr. Habil. Michel Chavance	
Prof. Dr. Jana Eccard	
Prof. Dr. Elisabeth Gassiat	
Dr. Habil. Christine Jacob (Betreuerin / Directrice de thèse)	
Prof. Dr. Peter Jagers	*Gutachter / Rapporteur*
Prof. Dr. Henning Läuter	*Vorsitz / Président*
Prof. Dr. Sylvie Roelly (Betreuerin / Directrice de thèse)	*Gutachter / Rapporteur*

1 Probabilités et Statistique
2 Angewandte Mathematik

Acknowledgments

I would like to express my gratitude to my supervisors, Prof. Sylvie Rœlly and Dr. Christine Jacob, who were abundantly helpful, patient and devoted. They both offered me invaluable support and guidance throughout these three years.

I must also acknowledge Prof. Jagers and Prof. Alsmeyer who kindly accepted to serve as my external reviewers, as well as all the members of the committee.

I am very grateful to Prof. Rœlly who encouraged me to complete my thesis under a joint supervision, which has been made possible thanks to the Franco-German University. This thesis benefited in addition from the financial assistance of the graduate program IRTG Stochastic Models of Complex Processes, to which I am very thankful for the high-quality scientific training that it offered to me. A very special thanks goes to the Unité MIA from the French National Institute for Agricultural Research in Jouy-en-Josas, who kindly hosted me for several months. I would like to acknowledge more generally the different structures or institutions involved in the development of my thesis: Institut für Mathematik der Universität Potsdam, Institut für Mathematik der Technische Universität Berlin, École Doctorale Santé Publique Paris-Sud 11, and Collège Doctoral Franco-Allemand Paris Ouest Nanterre / Potsdam.

Lastly, I would like to thank my friends and family for their support, and express my gratitude to those who patiently proofread this thesis.

Summary

This thesis is concerned with the issue of extinction of populations composed of different types of individuals, and their behavior before extinction and in case of a very late extinction. We approach this question firstly from a strictly probabilistic viewpoint, and secondly from the standpoint of risk analysis related to the extinction of a particular model of population dynamics. In this context we propose several statistical tools.

The population size is modeled by a branching process, which is either a continuous-time multitype Bienaymé-Galton-Watson process (BGWc), or its continuous-state counterpart, the multitype Feller diffusion process. We are interested in different kinds of conditioning on non-extinction, and in the associated equilibrium states. These ways of conditioning have been widely studied in the monotype case. However the literature on multitype processes is much less extensive, and there is no systematic work establishing connections between the results for BGWc processes and those for Feller diffusion processes.

In the first part of this thesis, we investigate the behavior of the population before its extinction by conditioning the associated branching process X_t on non-extinction ($X_t \neq 0$), or more generally on non-extinction in a near future $0 \leqslant \theta < \infty$ ($X_{t+\theta} \neq 0$), and by letting t tend to infinity. We prove the result, new in the multitype framework and for $\theta > 0$, that this limit exists and is non-degenerate. This reflects a stationary behavior for the dynamics of the population conditioned on non-extinction, and provides a generalization of the so-called Yaglom limit, corresponding to the case $\theta = 0$. In a second step we study the behavior of the population in case of a very late extinction, obtained as the limit when θ tends to infinity of the process conditioned by $X_{t+\theta} \neq 0$. The resulting conditioned process is a known object in the monotype case (sometimes referred to as Q-process), and has also been studied when X_t is a multitype Feller diffusion process. We investigate the not yet considered case where X_t is a multitype BGWc process and prove the existence of the associated Q-process. In addition, we examine its properties, including the asymptotic ones, and propose several interpretations of the process. Finally, we are interested in interchanging the limits in t and θ, as well as in the not yet studied commutativity of these limits with respect to the high-density-type relationship between BGWc processes and Feller processes. We prove an original and exhaustive list of all possible exchanges of limit (long-time limit in t, increasing delay of extinction θ, diffusion limit).

The second part of this work is devoted to the risk analysis related both to the extinction of a population and to its very late extinction. We consider a branching population model (arising notably in the epidemiological context) for which a parameter related to the first moments of the offspring distribution is unknown. We build several estimators adapted to different stages of evolution of the population (phase growth, decay phase, and decay phase when extinction is expected very late), and prove moreover their asymptotic properties (consistency, normality). In particular, we build a least squares estimator adapted to the Q-process, allowing a prediction of the population development in the case of a very late extinction. This would correspond to the best or to the worst-case scenario, depending on whether the population is threatened or invasive. These tools enable us to study the extinction phase of the Bovine Spongiform Encephalopathy epidemic in Great Britain, for which we estimate the infection parameter corresponding to a possible source of horizontal infection persisting after the removal in 1988 of the major route of infection (meat and bone meal). This allows us to predict the evolution of the spread of the disease, including the year of extinction, the number of future cases and the number of infected animals. In particular, we produce a very fine analysis of the evolution of the epidemic in the unlikely event of a very late extinction.

Zusammenfassung

Diese Arbeit befasst sich mit der Frage des Aussterbens von Populationen verschiedener Typen von Individuen. Uns interessiert das Verhalten vor dem Aussterben sowie insbesondere im Falle eines sehr späten Aussterbens. Wir untersuchen diese Fragestellung zum einen von einer rein wahrscheinlichkeitstheoretischen Sicht und zum anderen vom Standpunkt der Risikoanalyse aus, welche im Zusammenhang mit dem Aussterben eines bestimmten Modells der Populationsdynamik steht. In diesem Kontext schlagen wir mehrere statistische Werkzeuge vor.

Die Populationsgröße wird entweder durch einen zeitkontinuierlichen mehrtyp-Bienaymé-Galton-Watson Verzweigungsprozess (BGWc) oder durch sein Analogon mit kontinuierlichem Zustandsraum, den Feller Diffusionsprozess, modelliert. Wir interessieren uns für die unterschiedlichen Arten auf Überleben zu bedingen sowie für die hierbei auftretenden Gleichgewichtszustände. Diese Bedingungen wurden bereits weitreichend im Falle eines einzelnen Typen studiert. Im Kontext von mehrtyp-Verzweigungsprozessen hingegen ist die Literatur weniger umfangreich und es gibt keine systematischen Arbeiten, welche die Ergebnisse von BGWc Prozessen mit denen der Feller Diffusionsprozesse verbinden. Wir versuchen hiermit diese Lücke zu schliessen.

Im ersten Teil dieser Arbeit untersuchen wir das Verhalten von Populationen vor ihrem Aussterben, indem wir das zeitasymptotysche Verhalten des auf Überleben bedingten zugehörigen Verzweigungsprozesses $(X_t|X_t \neq 0)_t$ betrachten (oder allgemeiner auf Überleben in naher Zukunft $0 \leqslant \theta < \infty$, $(X_t|X_{t+\theta} \neq 0)_t$). Wir beweisen das Ergebnis, neuartig im mehrtypen Rahmen und für $\theta > 0$, dass dieser Grenzwert existiert und nicht-degeneriert ist. Dies spiegelt ein stationäres Verhalten für auf Überleben bedingte Bevölkerungsdynamiken wider und liefert eine Verallgemeinerung des sogenannten Yaglom Grenzwertes (welcher dem Fall $\theta = 0$ entspricht). In einem zweiten Schritt studieren wir das Verhalten der Populationen im Falle eines sehr späten Aussterbens, welches wir durch den Grenzübergang auf $\theta \to \infty$ erhalten. Der resultierende Grenzwertprozess ist ein bekanntes Objekt im eintypen Fall (oftmals als Q-Prozess bezeichnet) und wurde ebenfalls im Fall von mehrtyp-Feller-Diffusionsprozessen studiert. Wir untersuchen den bisher nicht betrachteten Fall, in dem X_t ein mehrtyp-BGWc Prozess ist und beweisen die Existenz des zugehörigen Q-Prozesses. Darüber hinaus untersuchen wir seine Eigenschaften einschließlich der asymptotischen und weisen auf mehrere Auslegungen hin. Schließlich interessieren wir uns für die Austauschbarkeit der Grenzwerte in t und θ, und die Vertauschbarkeit dieser Grenzwerte in Bezug auf die Beziehung zwischen BGWc und Feller Prozessen. Wir beweisen die Durchführbarkeit aller möglichen Grenzwertvertauschungen (Langzeitverhalten, wachsende Aussterbeverzögerung, Diffusionslimit).

Der zweite Teil dieser Arbeit ist der Risikoanalyse in Bezug auf das Aussterben und das sehr späte Aussterben von Populationen gewidmet. Wir untersuchen ein Modell einer verzweigten Bevölkerung (welches vor allem im epidemiologischen Rahmen erscheint), für welche ein Parameter der Reproduktionsverteilung unbekannt ist. Wir konstruieren Schätzer, die an die jeweiligen Stufen der Evolution adaptiert sind (Wachstumsphase, Verfallphase sowie die Verfallphase, wenn das Aussterben sehr spät erwartet wird), und beweisen zudem deren asymptotische Eigenschaften (Konsistenz, Normalverteiltheit). Im Besonderen bauen wir einen für Q-Prozesse adaptierten kleinste-Quadrate-Schätzer, der eine Vorhersage der Bevölkerungsentwicklung im Fall eines sehr späten Aussterbens erlaubt. Dies entspricht dem Best- oder Worst-Case-Szenario, abhängig davon, ob die Bevölkerung bedroht oder invasiv ist. Diese Instrumente ermöglichen uns die Betrachtung der Aussterbensphase der Bovinen spongiformen Enzephalopathie Epidemie in Großbritannien. Wir schätzen den Infektionsparameter in Bezug auf mögliche bestehende Quellen der horizontalen Infektion nach der Beseitigung des primären Infektionsweges (Tiermehl) im Jahr 1988. Dies ermöglicht uns eine Vorhersage des Verlaufes der Krankheit inklusive des Jahres des Aussterbens, der Anzahl von zukünftigen Fällen sowie der Anzahl infizierter Tiere. Insbesondere ermöglicht es uns die Erstellung einer sehr detaillierten Analyse des Epidemieverlaufs im unwahrscheinlichen Fall eines sehr späten Aussterbens.

Résumé

Cette thèse s'articule autour de la problématique de l'extinction de populations comportant différents types d'individus, et plus particulièrement de leur comportement avant extinction et/ou en cas d'une extinction très tardive. Nous étudions cette question d'un point de vue strictement probabiliste, puis du point de vue de l'analyse des risques liés à l'extinction pour un modèle particulier de dynamique de population, et proposons plusieurs outils statistiques.

La taille de la population est modélisée soit par un processus de branchement de type Bienaymé-Galton-Watson à temps continu multitype (BGWc), soit par son équivalent dans un espace de valeurs continu, le processus de diffusion de Feller multitype. Nous nous intéressons à différents types de conditionnement à la non-extinction, et aux états d'équilibre associés. Ces conditionnements ont déjà été largement étudiés dans le cas monotype. Cependant la littérature relative aux processus multitypes est beaucoup moins riche, et il n'existe pas de travail systématique établissant des connexions entre les résultats concernant les processus BGWc et ceux concernant les processus de diffusion de Feller. Nous nous y sommes attelés.

Dans la première partie de cette thèse, nous nous intéressons au comportement de la population avant son extinction, en conditionnant le processus de branchement X_t à la non-extinction ($X_t \neq 0$), ou plus généralement à la non-extinction dans un futur proche $0 \leqslant \theta < \infty$ ($X_{t+\theta} \neq 0$), et en faisant tendre t vers l'infini. Nous prouvons le résultat, nouveau dans le cadre multitype et pour $\theta > 0$, que cette limite existe et est non-dégénérée, traduisant ainsi un comportement stationnaire pour la dynamique de la population conditionnée à la non-extinction, et offrant une généralisation de la limite dite de Yaglom (correspondant au cas $\theta = 0$). Nous étudions dans un second temps le comportement de la population en cas d'une extinction très tardive, obtenu comme limite lorsque θ tends vers l'infini du processus X_t conditionné par $X_{t+\theta} \neq 0$. Le processus conditionné ainsi obtenu est un objet connu dans le cadre monotype (parfois dénommé Q-processus), et a également été étudié lorsque le processus X_t est un processus de diffusion de Feller multitype. Nous examinons le cas encore non considéré où X_t est un BGWc multitype, prouvons l'existence du Q-processus associé, examinons ses propriétés, notamment asymptotiques, et en proposons plusieurs interprétations. Enfin, nous nous intéressons aux échanges de limites en t et en θ, ainsi qu'à la commutativité encore non étudiée de ces limites vis-à-vis de la relation de type grande densité reliant processus BGWc et processus de Feller. Nous prouvons ainsi une liste exhaustive et originale de tous les échanges de limites possibles (limite en temps t, retard de l'extinction θ, limite de diffusion).

La deuxième partie de ce travail est consacrée à l'analyse des risques liés à l'extinction d'une population et à son extinction tardive. Nous considérons un certain modèle de population branchante (apparaissant notamment dans un contexte épidémiologique) pour lequel un paramètre lié aux premiers moments de la loi de reproduction est inconnu, et construisons plusieurs estimateurs adaptés à différentes phases de l'évolution de la population (phase de croissance, phase de décroissance, phase de décroissance lorsque l'extinction est supposée tardive), prouvant de plus leurs propriétés asymptotiques (consistence, normalité). En particulier, nous contruisons un estimateur des moindres carrés adapté au Q-processus, permettant ainsi une prédiction de l'évolution de la population dans le meilleur ou le pire des cas (selon que la population est menacée ou au contraire invasive), à savoir celui d'une extinction tardive. Ces outils nous permettent d'étudier la phase d'extinction de l'épidémie d'Encéphalopathie Spongiforme Bovine en Grande-Bretagne, pour laquelle nous estimons le paramètre d'infection correspondant à une possible source d'infection horizontale persistant après la suppression en 1988 de la voie principale d'infection (farines animales). Cela nous permet de prédire l'évolution de la propagation de la maladie, notamment l'année d'extinction, le nombre de cas à venir et le nombre d'animaux infectés, et en particulier de produire une analyse très fine de l'évolution de l'épidémie dans le cas peu probable d'une extinction très tardive.

Contents

Introduction

The evolution of the branching theory is not only imputable to pure mathematical development, it is also strongly related to the desire of resolving a growing number of biological problems, leading to a progressive complexification of the original branching process and to new exciting mathematical inquiries. The history of branching processes finds its origin in the second half of the XIXth century in a demographic context, from the family extinction problem studied notably by Bienaymé, de Candolle, Galton and Watson (see historical overviews such as [Ken66, Jag75, Gut95] or more recently [Jag09]), and famously formulated as follows by Galton in 1873 ([Gal1873]):

"*A large nation, of whom we will only concern ourselves with adult males, N in number, and who each bear separate surnames colonize a district. Their law of population is such that, in each generation, a_0 per cent of the adult males have no male children who reach adult life; a_1 have one such male child; a_2 have two; and so on up to a_5 who have five. Find (1) what proportion of their surnames will have become extinct after r generations; and (2) how many instances there will be of the surname being held by m persons.*"

Branching processes certainly remain the most natural means to model the dynamic of a population consisting of individuals (or particles, or cells...) living and giving birth independently of one another, and as such have encountered various applications in biology ([Hac05]), demography, genealogy, computer science, physics, chemistry *etc.* ([Pak03]). The branching theory gradually left the classical Bienaymé-Galton-Watson process framework to investigate more complex and more realistic processes, considering, for example, an immigration component ([Hea65]), a varying or random environment ([AthKar71, SmiWi69]), a time-structure allowing individuals to have random life spans and give birth during their lives ([CruMod68]), branching populations evolving according to a random walk ([AthNey73]) *etc.* Among the various possible directions of research, we choose to work with branching processes describing populations with several types of individuals, and to consider their diffusion limit as well, known as the Feller diffusion process.

This PhD thesis deals with the following problematic: the extinction of branching populations, and more precisely their behavior before extinction and/or in case of a very late extinction. Extinction obviously plays a major role in the evolution, at different scales, from the global extinction of species to the local extinction of subpopulations. Whether one strives for the extinction of an invasive population or for the survival of an endangered species, a first key tool to estimate the vulnerability of a population is to ascertain its extinction probability. A next natural step, due to Kolmogorov ([Kolm38]), is to study the probability that the population still exists after a large but finite number of generations. Kolmogorov's estimate of the non-extinction probability gradually led to the study of the *Yaglom limit* ([Yag47]), which is the limiting conditional distribution of the branching process under the condition that it is not yet extinct. The Yaglom limit is a *quasi-stationary distribution*, in the sense that it is an equilibrium distribution for the dynamics conditioned on non-extinction. The mathematical and biological interest of such an object, and more generally of stationary behavior for Markov processes with an absorbing state, conditionally on non-absorption, was pointed out by Bartlett in his famous paper on competitive and predatory biological systems: "*the time to extinction may be so long that it is still of more relevance to consider the effectively ultimate distribution*" ([Bar57, Bar60]). Although a lot of papers have been devoted to branching processes conditioned on non-extinction, only a few have dealt with the multitype case and with processes conditioned on not being extinct in the remote future. It appears that some gaps exist in the literature, especially concerning potential commutativity results between different limits of interest for the conditioned processes (increasing delay of extinction, long-time behavior, scaling limit). The first part of our work is dedicated to the statement of new results for conditioned multitype continuous-time Bienaymé-Galton-Watson processes (BGWc) and conditioned multitype Feller diffusions, and to the connection with well-known results. The second part is devoted to the study of vanishing populations from an epidemiological point of view. We present stochastic and statistical tools to predict the evolution of a population, its extinction time, its total size, and its behavior in case of a very late extinction. This enables us to provide a fine epidemiological risk analysis for the Bovine Spongiform Encephalopathy in Great-Britain which is in its extinction phase, studying in particular the worst-case scenario corresponding to a very late extinction.

Chapter 1 first introduces the notion of multitype BGWc process. The classification of the individuals into several types is of immediate interest for the modeling of populations, since these types can correspond, for example, to health states, expressions of a gene, locations *etc.* It adds to the original branching mechanism a Markovian dynamics among the different types of individuals, which brings the original single-type Bienaymé-Galton-Watson process to a higher level of complexity. We next focus on the multitype Feller diffusion process, which can be seen as a limit to BGWc processes with a large number of individuals with small weights: if the mass and time are rescaled appropriately alongside the offspring distribution and number of initial individuals, one indeed obtains as a nontrivial limit a Feller diffusion process ([Fel51]) with continuous-state space $\mathbb{R}_+^d$ (d being the number of types), belonging to the broader category of continuous-state branching processes ([Jir58]). This approximation offers very useful applications, since it drastically reduces the degree of complexity of the original BGWc process from the knowledge of the whole offspring distribution, to the simple knowledge of its first and second-order moments, if these exist. In this chapter we review some basic results present in the literature for multitype BGWc processes with finite first and second-order moments (which is the setting of this work) and for multitype Feller diffusion processes, focusing especially on properties related to the extinction of the processes.

Chapter 2 is dedicated to the conditioning on non-extinction. Conditioning on non-extinction leads to interesting nondegenerate limits and provides a stationarity property to the process. As we will see, this last property holds more generally for Markov processes with an absorbing state conditioned on non-absorption. Moreover, conditioning provides information, crucial from a theoretical point of view as well as for biological applications, about the behavior of the population before extinction and its behavior in case of a late extinction. Indeed, even for populations which are doomed to become extinct, the duration of the extinction phase can be lengthy compared to the observation time scale, and it is often observed that population sizes fluctuate for a large amount of time before extinction actually occurs. After presenting different kinds of conditioning existing in the literature, we focus on two specific conditional limits for multitype BGWc and Feller diffusion processes, for which we provide a new and systematic study. The first limit consists in conditioning the process X_t on the event that it is not extinct at time $t+\theta$, but does eventually die out; the extinction is thus delayed by at least θ. We then consider the limiting behavior of this process as the time t tends to infinity, which is consequently a generalization of the Yaglom limit mentioned previously. The second kind of limit investigated in this work corresponds to what is sometimes referred to in the literature as the Q*-process*, obtained by letting the delay of extinction θ tend to infinity. We define this limit process for multitype BGWc processes, and prove that it is a Doob h-transform of the unconditioned BGWc process. We give an interpretation of this process as a BGWc with immigration (in the monotype case) and as a process with an "immortal" individual. Moreover, we show that it has a stationary behavior, and recall similar results already obtained for multitype Feller diffusion processes ([ChaRoe08]).

Chapter 3 is devoted to the investigation of interchangeability results between the different limits coming into play. A first natural line of inquiry concerns whether a connection exists between the generalized Yaglom limit associated with the BGWc process, and the one associated with the scaling limit of the BGWc process (*i.e.* the Feller diffusion process); the same question also holds for the associated Q-processes. We answer in the affirmative by showing that according to the intuition, the Yaglom-type distribution and the Q-process associated with the scaling limit process, correspond to the scaling limit of the original Yaglom-type distribution and Q-process.

We show moreover that, as expected from the classical single-type result, the Q-process associated with the multitype branching processes admits as a unique stationary probability measure the size-biased Yaglom distribution. Another question concerns the commutativity between the long-time limits of the conditioned processes, that is to say between the limits as the time t or the delay of extinction θ tends to infinity. Our goal is to provide and prove an exhaustive list of commutativity results between the six possible combinations of limit (time limit, large delay of extinction, and high-density limit).

Chapter 4 is dedicated to the risk analysis related to the extinction of a (threatened or harmful) population. The framework is a branching population with Poissonian transitions, which can be seen either as a Markovian process of order $d \geqslant 1$, or as a discrete-time Bienaymé-Galton-Watson process (BGW), with d types corresponding to the memory of the process. After justifying the choice of the model, we provide the distribution of the time of extinction, the total size of the population until extinction, and the behavior in case of a late extinction, making use of the associated Q-process. We next consider that a parameter in the Poissonian transition might be unknown, and provide several estimators for this parameter, corresponding to different phases of the evolution of the population: growth phase, decay phase, and decay phase in case of a very late extinction. One of these estimators stems from the literature on the estimation of the Perron's root for multitype branching processes, while two others are new conditional least squares estimators (CLSE) based either on the chosen process or on the process conditioned on non-extinction at each time step. We prove their strong consistency and asymptotic normality as the initial size of the population tends to infinity or, alternatively, as time tends to infinity. We compare the precision and accuracy of these three estimators on the basis of simulations. The last estimator introduced in this thesis is also original and offers an innovative method for a fine risk analysis of the very late extinction case. It is a CLSE associated with the process conditioned on very late extinction, and we prove that is is strongly consistent and asymptotically normal as time tends to infinity. This estimator enables predictions of the evolution of the population in the *best-case scenario* (if one considers an endangered population for which extinction is feared) or *worst-case scenario* (if one considers on the contrary a harmful population).

Finally, **Chapter 5** concerns the epidemiological study of the Bovine Spongiform Encephalopathy (BSE) in Great-Britain. The previous model provides an adequate epidemic model based on the yearly incidences of cases, which correspond to the available observations. In this model one parameter is unknown, quantifying the remaining infection transmission after the first feed ban law in 1988, which removed the main route of transmission of the disease *via* meat and bone meal. We estimate this parameter thanks to a CLSE presented in the previous chapter, and make use of the obtained estimation to predict the future spread of the disease, including the year of extinction, the number of cases to come and the evolution of the number of infected cattle. As a final object of study, we estimate the infection parameter for the associated Q-process, which leads to a fine analysis of the future behavior of the BSE epidemic in the unlikely case of a very late extinction.

Chapter 1

Multitype branching processes

Let d be the number of types. In this work we use the following notation.

$$\mathbb{N} := \{0, 1, 2, \ldots\}, \quad \mathbb{N}^* := \{1, 2, \ldots\},$$

$$\mathbb{R}_+ := [0, \infty[, \quad \mathbb{R}_+^d := [0, \infty[^d, \quad \overline{\mathbb{R}}_+^d := [0, \infty]^d.$$

If no further indication, any d-dimensional vector $\boldsymbol{x} \in \mathbb{R}^d$ is considered as a row vector $(x_1, \ldots, x_d)$. Its transpose is denoted $\boldsymbol{x}^T$, but in order to avoid heavy notation we omit this subscript when no confusion is possible. $\mathbf{1}$ and $\mathbf{0}$ denote the vectors $(1, \ldots, 1)$ and $(0, \ldots, 0) \in \mathbb{R}^d$, and for all $i = 1 \ldots d$, $\mathbf{e}_i = (0, \ldots, 1, \ldots, 0)$ the basis vector of $\mathbb{R}^d$. The vector $\boldsymbol{\infty}$ denotes the element in $\overline{\mathbb{R}}_+^d$ having all its coordinates equal to ∞. $\boldsymbol{x} \cdot \boldsymbol{y}$ denotes the usual scalar product between $\boldsymbol{x}$ and $\boldsymbol{y}$ in $\mathbb{R}^d$,

$$\boldsymbol{x} \cdot \boldsymbol{y} := x_1 y_1 + \ldots + x_d y_d,$$

$\|\boldsymbol{x}\|$ the Euclidean norm, and $|\boldsymbol{x}|$ the $\mathcal{L}^1$-norm:

$$\|\boldsymbol{x}\| := \sqrt{x_1^2 + \ldots + x_d^2},$$
$$|\boldsymbol{x}| := |x_1| + \ldots + |x_d|.$$

We define moreover

$$\boldsymbol{x}\boldsymbol{y} := (x_1 y_1, \ldots, x_d y_d),$$

and

$$\boldsymbol{x}^{\boldsymbol{y}} := x_1^{y_1} \ldots x_d^{y_d}.$$

We introduce the following partial order on $\mathbb{R}^d$:

$$\boldsymbol{x} \leqslant \boldsymbol{y} \text{ (resp. } \boldsymbol{x} < \boldsymbol{y}\text{) means that for all } i = 1 \ldots d,\ x_i \leqslant y_i \text{ (resp. } x_i < y_i\text{)}.$$

We call a matrix *positive* (resp. *non-negative*) if all its coefficients are > 0 (resp. $\geqslant 0$). In general, any d-dimensional vector or $d \times d$ matrix is denoted by a bold character. The set of the $d \times d$ real matrices is denoted $\mathcal{M}_d(\mathbb{R})$.

Throughout **Chapter 1**, **Chapter 2** and **Chapter 3** we work on the probability space $(\Omega, (\mathbf{X}_t)_{t \geqslant 0}, (\mathcal{F}_t)_{t \geqslant 0})$, where $\Omega := \mathcal{D}\left(\mathbb{R}_+, \mathbb{R}_+^d\right)$ is the canonical space of càdlàg functions from $\mathbb{R}_+$ to $\mathbb{R}_+^d$. For every $t \geqslant 0$, $\mathbf{X}_t$ denotes the canonical projection at time t, and $(\mathcal{F}_t)_{t \geqslant 0}$ the right-continuous filtration generated by the canonical process $(\mathbf{X}_t)_{t \geqslant 0}$. In these chapters, we denote by a subscript on $\mathbb{P}$ or $\mathbb{E}$ the initial distribution of a process with law $\mathbb{P}$. If this subscript is an element of $\mathbb{R}_+^d$, then the initial distribution corresponds to the Dirac measure at this point.

Moreover, for a given infinitesimal generator G with domain $D(G)$ and a given subset $D_0(G) \subseteq D(G)$, we say that a law $\mathbf{P}$ on the probability space $(\Omega, (\mathbf{X}_t)_{t \geqslant 0}, (\mathcal{F}_t)_{t \geqslant 0})$ is a solution to the martingale problem $\mathcal{MP}\left(G, D_0\left(G\right)\right)$ (or $\mathcal{MP}(G)$ to avoid heavy notation) if for all function $f \in D_0(G)$,

$$\left(f(\mathbf{X}_t) - f(\mathbf{X}_0) - \int_0^t (Gf)(\mathbf{X}_{s-})ds\right)_{t\geqslant 0} \quad \text{is a } \left(\mathbf{P}, (\mathcal{F}_t)_{t\geqslant 0}\right)\text{-martingale.}$$

When imposing the initial condition $\boldsymbol{x}_0$ we write $\mathcal{MP}\left(G, D_0\left(G\right), \boldsymbol{x}_0\right)$.

1.1 Continuous-time multitype Bienaymé-Galton-Watson process

Continuous-time Bienaymé-Galton-Watson processes (BGWc) are Markovian processes describing the size evolution of a *splitting* population, where individuals have exponentially distributed life spans and produce at the end of their life-time a random number of offspring. The number of offspring is independent of the parent's life span, and individuals reproduce independently of one another. In the multitype case considered here, where the individuals are classified into different types, the reproduction law as well as the life span might depend on the type of the individual. However, an individual of a given type can possibly give birth to individuals of different types, according to its reproduction law (see Figure 1.1).

The randomness of the life spans is a real amelioration with respect to the classical discrete-time Bienaymé-Galton-Watson process (BGW), where individuals only live one deterministic time-unit. BGWc processes have thus been used a lot in mathematical biology. However one must keep in mind that these processes are of limited biological relevance: first, the splitting mechanism means that children are only born at their parent's death, and second, exponential life spans imply the absence of aging for the individuals.

Much of the early work on the continuous-time branching processes was initiated by the Russian school in the middle of the XXth century ([KolDmi47, Sew51]), and later expanded to the multitype case. On this topic we mostly refer to the monographs [Sew75] and [AthNey72], and might also quote [Har63] and [Mod71].

In this section, we consider a d-type BGWc process with sample paths in $\mathcal{D}\left(\mathbb{R}^+, \mathbb{N}^d\right)$, and we denote by $\mathbb{P}$ its law on the canonical probability space. After giving some definitions and preliminary results in **Subsection 1.1.1**, we recall in **Subsection 1.1.2** the usual basic assumptions on the first-order moments and quote the Perron-Frobenius Theorem, and finally present fundamental known results on the extinction probability of the process in **Subsection 1.1.3**.

1.1.1 Preliminaries: generating functions and infinitesimal generator

We denote by $\left(\mathbf{p}\left(\mathbf{j}\right)\right)_{\mathbf{j}\in\mathbb{N}^d}$ the offspring distribution (or reproduction law) of the branching process, where for all $\mathbf{j} = (j_1, \ldots, j_d) \in \mathbb{N}^d$,

$$\mathbf{p}(\mathbf{j}) = (p_1(\mathbf{j}), \ldots, p_d(\mathbf{j})) \in [0, 1]^d.$$

For every $i = 1 \ldots d$, $p_i(\mathbf{j}) = p_i(j_1, \ldots, j_d) \in [0, 1]$ denotes the probability that a type i individual produces j_1 individuals of type 1, j_2 individuals of type 2 etc. It satisfies for all $i = 1 \ldots d$,

$$\sum_{\mathbf{j}\in\mathbb{N}^d} p_i(\mathbf{j}) = 1.$$

Let $\boldsymbol{f}(\mathbf{r}) = (f_1(\mathbf{r}), \ldots, f_d(\mathbf{r})) \in [0, 1]^d$ be the generating function of the offspring distribution, defined as follows. For all $\mathbf{r} = (r_1, \ldots, r_d) \in [0, 1]^d$,

$$\boldsymbol{f}(\mathbf{r}) := \sum_{\mathbf{j}\in\mathbb{N}^d} \mathbf{p}(\mathbf{j})\mathbf{r}^{\mathbf{j}} = \sum_{\mathbf{j}\in\mathbb{N}^d} \mathbf{p}\left(j_1, \ldots, j_d\right) r_1^{j_1} \ldots r_d^{j_d}, \qquad (1.1.1)$$

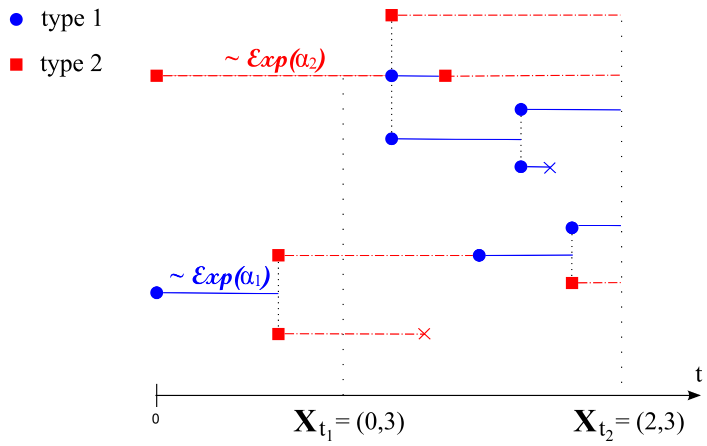

Figure 1.1: A 2-type BGWc process.

or component-wise, for all $i = 1 \dots d$,

$$f_i(\mathbf{r}) = \sum_{\mathbf{j}\in\mathbb{N}^d} p_i(\mathbf{j})\mathbf{r}^{\mathbf{j}} = \sum_{\mathbf{j}\in\mathbb{N}^d} p_i(j_1, \dots, j_d) r_1^{j_1} \dots r_d^{j_d}.$$

We denote by $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_d)$ the vector of the branching rates, meaning that every individual of type i lives an exponentially distributed lifetime of parameter $\alpha_i > 0$ (Figure 1.1), and we introduce the diagonal matrix

$$\mathbf{A} := diag(\boldsymbol{\alpha}) = \begin{pmatrix} \alpha_1 & 0 & \dots & 0 \\ 0 & \alpha_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \alpha_d \end{pmatrix}. \tag{1.1.2}$$

The generating function of the process at time t is denoted $\mathbf{F}_t(\mathbf{r}) = (F_{t,1}(\mathbf{r}), \dots, F_{t,d}(\mathbf{r}))$, $\mathbf{r} \in [0,1]^d$, where for all $i = 1 \dots d$,

$$F_{t,i}(\mathbf{r}) := \mathbb{E}_{\mathbf{e}_i}\left(\mathbf{r}^{\mathbf{X}_t}\right) = \mathbb{E}\left(\mathbf{r}^{\mathbf{X}_t} | \mathbf{X}_0 = \mathbf{e}_i\right) = \sum_{\mathbf{j}\in\mathbb{N}^d} \mathbb{P}_{\mathbf{e}_i}(\mathbf{X}_t = \mathbf{j})\mathbf{r}^{\mathbf{j}} \in [0,1]. \tag{1.1.3}$$

The transition probabilities of the process $P_t(\mathbf{i},\mathbf{j}) := \mathbb{P}(\mathbf{X}_t = \mathbf{j}|\mathbf{X}_0 = \mathbf{i})$, $\mathbf{i},\mathbf{j} \in \mathbb{N}^d$, satisfy the so-called *branching property*

$$P_t(\mathbf{i},\mathbf{j}) = P_t(\mathbf{e}_1,\mathbf{j})^{*i_1} * \dots * P_t(\mathbf{e}_d,\mathbf{j})^{*i_d}, \tag{1.1.4}$$

where $*$ denotes the convolution product. For example, in the monotype case $d = 1$, this formula becomes

$$P_t(i,j) = P_t(1,j)^{*i} = \sum_{j^{(1)}+\dots+j^{(i)}=j} P_t(1,j^{(1)}) \dots P_t(1,j^{(i)}). \tag{1.1.5}$$

The branching property is a fundamental property which is satisfied by every branching process (see (1.2.3) for the continuous-state branching processes). It reflects the *additivity* of the transition probabilities of these processes with respect to the initial condition. In the discrete-state setting, notably for BGWc processes, this means that a BGWc branching process with $\mathbf{i}$ initial individuals is the sum of i_1 independent copies of a BGWc with one type 1 initial individual, i_2 independent copies of a BGWc with one type 2 initial individual *etc.*

This implies notably that for all $\boldsymbol{x} \in \mathbb{N}^d$

$$\mathbb{E}_{\boldsymbol{x}}\left(\mathbf{r}^{\mathbf{X}_t}\right) = \mathbb{E}\left(\mathbf{r}^{\mathbf{X}_t} | \mathbf{X}_0 = \boldsymbol{x}\right) = \left[\mathbf{F}_t(\mathbf{r})\right]^{\boldsymbol{x}} = \prod_{k=1}^{d} \Big[\sum_{\mathbf{j}\in\mathbb{N}^d} F_{t,k}\left(\mathbf{r}\right)\Big]^{x_k}, \tag{1.1.6}$$

hence the additivity turns into a *multiplicativity* dependence on the initial condition for the generating function.

We deduce from (1.1.6) together with the Chapman-Kolmogorov equation the following semigroup property. For all $s, t \geqslant 0$ and $\mathbf{r} \in [0,1]^d$,

$$\mathbf{F}_{t+s}(\mathbf{r}) = \mathbf{F}_t\left(\mathbf{F}_s\left(\mathbf{r}\right)\right). \tag{1.1.7}$$

Finally, the Kolmogorov forward and backward equations applied on $\boldsymbol{x} \mapsto \mathbf{r}^{\boldsymbol{x}}$ lead to ([AthNey72] Section 5.7.1): for $i = 1 \dots d$,

$$\frac{\partial}{\partial t} F_{t,i}(\mathbf{r}) = \sum_{j=1}^{d} \alpha_j \left[f_j\left(\mathbf{r}\right) - r_j\right] \frac{\partial}{\partial r_j} F_{t,i}(\mathbf{r}), \tag{1.1.8}$$

and

$$\frac{\partial}{\partial t} F_{t,i}(\mathbf{r}) = \alpha_i \left[f_i\left(\mathbf{F}_t(\mathbf{r})\right) - F_{t,i}(\mathbf{r})\right]. \tag{1.1.9}$$

Written d-dimensionally, (1.1.8) and (1.1.9) become

$$\begin{aligned} \frac{\partial}{\partial t}\ \mathbf{F}_t(\mathbf{r}) &= \sum_{j=1}^{d} \alpha_j \left[f_j\left(\mathbf{r}\right) - r_j\right] \frac{\partial}{\partial r_j} \mathbf{F}_t(\mathbf{r}), \\ \frac{\partial}{\partial t} \mathbf{F}_t(\mathbf{r}) &= \left[\boldsymbol{f}\left(\mathbf{F}_t(\mathbf{r})\right) - \mathbf{F}_t(\mathbf{r})\right] \mathbf{A}. \end{aligned} \tag{1.1.10}$$

If one assumes that for all $i, j = 1 \dots d$, $\frac{\partial f_i}{\partial r_j}(\mathbf{1}) < \infty$, then as mentioned later (Proposition 1.1.3) this ensures that there cannot be infinitely many individuals produced in a finite time, and also guarantees that (1.1.8) and (1.1.9) subject to the boundary condition $\mathbf{F}_0(\mathbf{r}) = \mathbf{r}$ admits as a unique solution the generating function $\mathbf{F}_t(\mathbf{r})$ of the BGWc process with law $\mathbb{P}$.

We can explicitly compute its infinitesimal generator L, defined for all smooth function $f : \mathbb{N}^d \to \mathbb{R}$ and all $\boldsymbol{x} \in \mathbb{N}^d$ by

$$\left(Lf\right)\left(\boldsymbol{x}\right) := \lim_{h\to 0} \frac{1}{h} \mathbb{E}_{\boldsymbol{x}}\left[f\left(\mathbf{X}_h\right) - f\left(\boldsymbol{x}\right)\right].$$

For every $h \geqslant 0$, $i = 1 \dots d$ and $\mathbf{k} \in \mathbb{N}^d$ we define the events

$$A^0(h) = \{\text{no branching event in } [0,h]\},$$
$$A_{\mathbf{k}}^{(i)}(h) = \{\text{exactly one branching event in } [0,h]\text{: one } i \text{ individual splits into } \mathbf{k} \text{ offsprings}\},$$

and the disjoint union of events

$$A(h) = A^0(h) \dot{\bigcup_{\substack{1\leqslant i\leqslant d \\ \mathbf{k}\in\mathbb{N}^d}}} A_{\mathbf{k}}^{(i)}(h).$$

Let f be a real-valued function defined on $\mathbb{N}^d$ with compact support. For all $h \geqslant 0$ and $\boldsymbol{x} \in \mathbb{N}^d$ we denote

$$R_h f(\boldsymbol{x}) := \frac{1}{h} \mathbb{E}_{\boldsymbol{x}}\left[\left(f\left(\mathbf{X}_h\right) - f\left(\boldsymbol{x}\right)\right) \mathbf{1}_{A(h)^C}\right].$$

Then for all $\boldsymbol{x} \in \mathbb{N}^d$ we have

$$
\begin{aligned}
&\frac{1}{h}\mathbb{E}_{\boldsymbol{x}}\left[f\left(\mathbf{X}_h\right)-f\left(\boldsymbol{x}\right)\right]\\
&\quad=\frac{1}{h}\left[f\left(\boldsymbol{x}\right)-f\left(\boldsymbol{x}\right)\right]+\frac{1}{h}\mathbb{E}_{\boldsymbol{x}}\Big[\left[f\left(\mathbf{X}_h\right)-f\left(\boldsymbol{x}\right)\right]\sum_{\substack{1\leqslant i\leqslant d\\ \mathbf{k}\in\mathbb{N}^d}}\mathbf{1}_{A_{\mathbf{k}}^{(i)}(h)}\Big]+R_hf(\boldsymbol{x})\\
&\quad=\frac{1}{h}\left[f\left(\boldsymbol{x}\right)-f\left(\boldsymbol{x}\right)\right]+\sum_{i=1}^{d}\sum_{\mathbf{k}\in\mathbb{N}^d}\left[f\left(\boldsymbol{x}+\mathbf{k}-\mathbf{e}_i\right)-f\left(\boldsymbol{x}\right)\right]\frac{1}{h}\mathbb{P}_{\boldsymbol{x}}\left[A_{\mathbf{k}}^{(i)}(h)\right]+R_hf(\boldsymbol{x}). \qquad (1.1.11)
\end{aligned}
$$

But

$$
\begin{aligned}
\frac{1}{h}\mathbb{P}_{\boldsymbol{x}}\left[A_{\mathbf{k}}^{(i)}(h)\right]&=\frac{1}{h}p_i(\mathbf{k})\int_0^h\alpha_ix_ie^{-\alpha_ix_is}e^{-\alpha_i(x_i-1+k_i)(h-s)}\prod_{\substack{j=1..d\\ j\neq i}}e^{-\alpha_jx_js}e^{-\alpha_j(x_j+k_j)(h-s)}ds\\
&=\alpha_ix_ip_i(\mathbf{k})e^{-h\sum_{j=1}^d\alpha_j(x_j+k_j-\delta_{ij})}\frac{1}{h}\int_0^he^{s\sum_{j=1}^d\alpha_j(k_j-\delta_{ij})}ds\\
&=\alpha_ix_ip_i(\mathbf{k})e^{-h\sum_{j=1}^d\alpha_j(x_j+k_j-\delta_{ij})}\frac{e^{h\sum_{j=1}^d\alpha_j(k_j-\delta_{ij})}-1}{h\sum_{j=1}^d\alpha_j(k_j-\delta_{ij})}.
\end{aligned}
$$

Using the fact that $e^{h\sum_{j=1\dots d}\alpha_j(k_j-\delta_{ij})}-1\sim_{h\to0}h\sum_{j=1\dots d}\alpha_j(k_j-\delta_{ij})$ we hence obtain that

$$
\lim_{h\to0}\frac{1}{h}\mathbb{P}_{\boldsymbol{x}}\left[A_{\mathbf{k}}^{(i)}(h)\right]=\alpha_ix_ip_i(\mathbf{k}). \qquad (1.1.12)
$$

It now remains to estimate

$$
R_hf(\boldsymbol{x})=\frac{1}{h}\mathbb{E}_{\boldsymbol{x}}\left[\left(f\left(\mathbf{X}_h\right)-f\left(\boldsymbol{x}\right)\right)\mathbf{1}_{A(h)^C}\right].
$$

For this we note that

$$
\begin{aligned}
A(h)^C&=\{\text{at least two branching events in }[0,h]\}\\
&=\bigcup_{\substack{1\leqslant i,j\leqslant d\\ \mathbf{k}\in\mathbb{N}^d}}\left\{\begin{matrix}\text{at least two branching events in }[0,h]:\\ \text{one } i \text{ individual (splitting into } \mathbf{k} \text{ offsprings), and one } j \text{ individual}\end{matrix}\right\}\\
&=:\bigcup_{\substack{1\leqslant i,j\leqslant d\\ \mathbf{k}\in\mathbb{N}^d}}B_{\mathbf{k}}^{(i,j)}(h).
\end{aligned}
$$

Hence we obtain the following estimate of $R_hf(t,\boldsymbol{x})$. For all $t,h\geqslant0$, $\boldsymbol{x}\in\mathbb{N}^d$,

$$
\begin{aligned}
R_hf(\boldsymbol{x})&\leqslant\frac{1}{h}\mathbb{E}_{\boldsymbol{x}}\Big[\left(f\left(\mathbf{X}_h\right)-f\left(\boldsymbol{x}\right)\right)\sum_{\substack{1\leqslant i,j\leqslant d\\ \mathbf{k}\in\mathbb{N}^d}}\mathbf{1}_{B_{\mathbf{k}}^{(i,j)}(h)}\Big]\\
&\leqslant 2\sup_{\boldsymbol{x}\in\mathbb{N}^d}|f(\boldsymbol{x})|\sum_{\substack{1\leqslant i,j\leqslant d\\ \mathbf{k}\in\mathbb{N}^d}}\frac{1}{h}\mathbb{P}_{\boldsymbol{x}}\left[B_{\mathbf{k}}^{(i,j)}(h)\right],
\end{aligned}
$$

with $\sup_{\boldsymbol{x}\in\mathbb{N}^d}|f(\boldsymbol{x})|<\infty$ by assumption. Moreover,

$$
\begin{aligned}
\frac{1}{h}\mathbb{P}_{\boldsymbol{x}}\left[B_{\mathbf{k}}^{(i,j)}(h)\right]&=p_i(\mathbf{k})\frac{1}{h}\int_0^h\alpha_ix_ie^{-\alpha_ix_is}\left(1-e^{-\alpha_j(x_j+k_j-\delta_{ij})(h-s)}\right)ds\\
&=\alpha_ix_ip_i(\mathbf{k})\left[\frac{1}{h}\int_0^he^{-\alpha_ix_is}ds-e^{-\alpha_j(x_j+k_j-\delta_{ij})h}\frac{1}{h}\int_0^he^{[\alpha_j(x_j+k_j-\delta_{ij})-\alpha_ix_i]s}ds\right]\\
&=\alpha_ix_ip_i(\mathbf{k})\left[\frac{1-e^{-\alpha_ix_ih}}{\alpha_ix_ih}-e^{-\alpha_j(x_j+k_j-\delta_{ij})h}\frac{e^{[\alpha_j(x_j+k_j-\delta_{ij})-\alpha_ix_i]h}-1}{[\alpha_j(x_j+k_j-\delta_{ij})-\alpha_ix_i]\,h}\right],
\end{aligned}
$$

which tends to 0 as $h \to 0$. Hence $\lim_{h\to 0} R_h f(\boldsymbol{x}) = 0$, which combined with (1.1.12) in (1.1.11) implies that for all $\boldsymbol{x} \in \mathbb{N}^d$

$$(Lf)(\boldsymbol{x}) = \sum_{i=1}^{d} \alpha_i x_i \sum_{\mathbf{k}\in\mathbb{N}^d} p_i(\mathbf{k}) \left[f(\boldsymbol{x} + \mathbf{k} - \mathbf{e}_i) - f(\boldsymbol{x})\right]. \tag{1.1.13}$$

Remark 1.1.1. In the monotype case $d = 1$ we obtain the classical result that for all $(t, x) \in \mathbb{R}_+ \times \mathbb{N}$ ([EthKur86]),

$$(Lf)(x) = \alpha x \sum_{k=0}^{\infty} p_k \left[f(x + k - 1) - f(x)\right].$$

1.1.2 Moments and the Perron-Frobenius Theorem

We denote by $\mathbf{M} = [m_{ij}]_{1\leqslant i,j\leqslant d}$ the mean matrix of the offspring distribution, where m_{ij} stands for the expected number of type j offsprings produced by an individual of type i, assumed in all what follows to be finite:

$$m_{ij} := \sum_{\mathbf{k}\in\mathbb{N}^d} k_j p_i(\mathbf{k}) = \frac{\partial f_i}{\partial r_j}(\mathbf{1}) < \infty. \tag{1.1.14}$$

We next introduce $\mathbf{M}(t) = [m_{ij}(t)]_{1\leqslant i,j\leqslant d}$, the mean matrix of the process at time t. For every $i, j = 1 \dots d$ and every $t \geqslant 0$, $m_{ij}(t) := \mathbb{E}_{\mathbf{e}_i}[X_{t,j}]$ denotes the expected number of type j individuals at time t of the process initiated with a single individual of type i. The branching property implies the linearity of $\mathbf{M}(t)$ with respect to the initial number of individuals:

$$\mathbb{E}_{\mathbf{X}_0}[\mathbf{X}_t] = \mathbf{X}_0 \mathbf{M}(t). \tag{1.1.15}$$

The mean matrix satisfies the semigroup property $\mathbf{M}(t+s) = \mathbf{M}(t)\mathbf{M}(s)$ $(s, t \geqslant 0)$ and the continuity condition $\lim_{t\to 0} \mathbf{M}(t) = \mathbf{I}$, hence there exists a matrix $\mathbf{C}$ such that

$$\mathbf{M}(t) = e^{\mathbf{C}t} = \sum_{p=0}^{\infty} \frac{t^p \mathbf{C}^p}{p!}. \tag{1.1.16}$$

We can identify the matrix $\mathbf{C}$ as ([AthNey72] Section 5.7.2)

$$\mathbf{C} := \mathbf{A}(\mathbf{M} - \mathbf{I}). \tag{1.1.17}$$

Remark 1.1.2. Since $\mathbf{A}$ and $\mathbf{M}$ are non-negative matrices, all the non-diagonal terms of the matrix $\mathbf{C}$ are non-negative.

The finiteness of the first-order moments of the offspring distribution guarantees the nonexplosion in finite time. This is however a sufficient but not necessary condition.

Proposition 1.1.3 ([Sew75] Satz 4.4.1)**.** *Assuming that all the entries of $\boldsymbol{M}$ are finite, then for all $i, j = 1 \dots d$ and all $t \geqslant 0$, $m_{ij}(t) < \infty$.*

Let us introduce the following definitions and assumptions.

Definition 1.1.4. A branching process is called *simple* if its generating function $\boldsymbol{f}$ is such that for all $i = 1 \dots d$, $f_i(\mathbf{r})$ is linear in $r_1, \dots, r_d$, with no constant term.

Remark 1.1.5. It means that for all $i = 1 \dots d$, f_i is of the form

$$f_i(\mathbf{r}) = p_i(\mathbf{e}_1) r_1 + \ldots + p_i(\mathbf{e}_d) r_d,$$

which implies that $\sum_{j=1}^{d} p_i(\mathbf{e}_j) = 1$. In this case each individual has exactly one offspring (possibly of different type) and the process has a constant number of individuals.

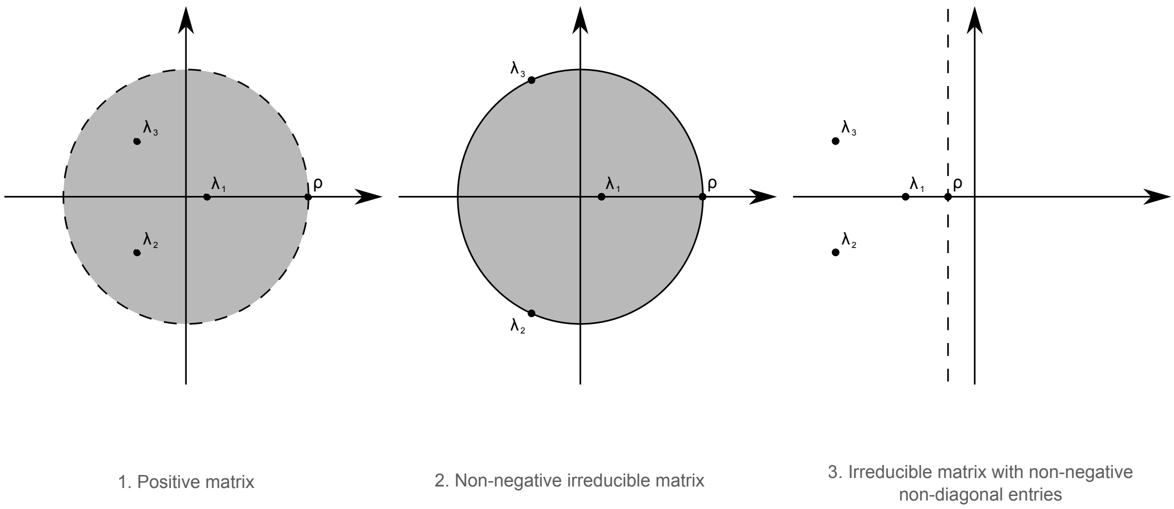

Figure 1.2: Graphical representation of the spectrum and Perron's root ρ for categories of matrices (positive matrix, non-negative irreducible matrix, irreducible matrix with non-negative non-diagonal entries).

To avoid this trivial case, we work under the following assumption:

(B0) The process is *not simple.*

Definition 1.1.6. A matrix $\mathbf{D}$ is called *irreducible* if there does not exist any permutation matrix $\mathbf{S}$ (each row and each column has exactly one 1 entry and all others 0) such that $\mathbf{S}^{-1}\mathbf{D}\mathbf{S}$ is block triangular.

Let us recall a fundamental theorem about non-negative irreducible matrices due to 0. Perron and G. Frobenius ([Per1907, Frob1908]). This theorem can be found *e.g.* in [Sen73], Theorem 1.1, and is illustrated by Figure 1.2.2.

Theorem 1.1.7 (Perron-Frobenius)**.** *An irreducible non-negative matrix always has a real positive eigenvalue ρ, called the Perron's root, such that the moduli of all the other eigenvalues are smaller than are equal to ρ. The "maximal" eigenvalue ρ is simple, and its related eigenspace is one-dimensional. There corresponds a right (resp. left) eigenvector with positive coordinates.*

Remark 1.1.8. If the matrix is positive, then the moduli of all the other eigenvalues are smaller than ρ (see Figure 1.2.1).

In this work we use the following extension of the Perron-Frobenius structure, which concerns irreducible matrices with non-negative non-diagonal entries (see Theorem 2.5 in [Sen73]), illustrated by Figure 1.2.3.

Theorem 1.1.9. *An irreducible matrix with non-negative non-diagonal entries always has a real eigenvalue ρ, called the Perron's root, such that the real part of any other eigenvalue is smaller than ρ. The "maximal" eigenvalue ρ is simple, and its related eigenspace is one-dimensional. There corresponds a right (resp. left) eigenvector with positive coordinates.*

Definition 1.1.10. A process is called *irreducible* if the mean matrix $\mathbf{M}$ is irreducible.

Throughout this work we assume the following.

(B1) The mean matrix $\mathbf{M}$ is *finite* and *irreducible.*

Remark 1.1.11. The irreducibility assumption entails that all the types *communicate* with each other ([Sew75] Satz 4.6.2). We say that a type i and a type j ($i, j = 1 \dots d$) *communicate* (we write $i \leftrightarrow j$) if there exist $s, t > 0$ such that

$$\mathbb{P}_{\mathbf{e}_i}(X_{s,j} > 0) > 0 \quad \text{and} \quad \mathbb{P}_{\mathbf{e}_j}(X_{t,i} > 0) > 0.$$

The binary relation $\leftrightarrow$ is an equivalence relation.

Remark 1.1.12. Obviously, the matrix $\mathbf{M}$ is reducible if and only if $\mathbf{M} - \mathbf{I}$ is reducible. Moreover, since the diagonal matrix $\mathbf{A}$ only multiplies the entries of $\mathbf{M} - \mathbf{I}$ by some positive scalars, the reducibility of $\mathbf{M} - \mathbf{I}$ is equivalent to the one of $\mathbf{C} = \mathbf{A}(\mathbf{M} - \mathbf{I})$. We have in addition the following equivalences ([Sew75] Kapitel IV §6).

$$\begin{aligned} \mathbf{M} \text{ is irreducible } &\iff \mathbf{C} \text{ is irreducible} \\ &\iff \mathbf{M}(t) \text{ is irreducible for all } t > 0 \iff \mathbf{M}(t) > 0 \text{ for all } t > 0. \end{aligned}$$

Note that the last condition holds for example if there exists one $p \in \mathbb{N}^*$ such that $\mathbf{C}^p > 0$.

Let us assume **(B1)**, and apply Theorem 1.1.9 to the irreducible matrix $\mathbf{C}$ (with non-negative non-diagonal entries). We denote by ρ its maximal eigenvalue, and $\boldsymbol{\xi}$ (resp. $\boldsymbol{\eta}$) the associated right (resp. left) eigenvector, with the following normalization convention:

$$\mathbf{C}\boldsymbol{\xi}^T = \rho\boldsymbol{\xi}^T, \quad \boldsymbol{\eta}\mathbf{C} = \rho\boldsymbol{\eta}, \quad \boldsymbol{\eta} \cdot \boldsymbol{\xi} = 1, \quad \boldsymbol{\xi} \cdot \mathbf{1} = 1. \tag{1.1.18}$$

Definition 1.1.13. The process is called *supercritical*, *critical* or *subcritical* according as $\rho > 0$, $\rho = 0$ or $\rho < 0$.

The following result describing the asymptotic behavior of the mean matrix $\mathbf{M}(t)$ as $t \to \infty$ ([Sew75] Satz 4.7.5), provides a justification to the classification into three types of criticality depending on the sign of ρ.

Proposition 1.1.14. *Assume* ***(B1)****. Then the mean matrix* $\boldsymbol{M}(t)$ *satisfies the following asymptotic behavior. For all* $i, j = 1 \dots d$

$$m_{ij}(t) = \xi_i \eta_j e^{\rho t} + o(e^{\tilde{\rho} t}), \quad t \to \infty,$$

with $\tilde{\rho} < \rho$.

Hence the mean value $m_{ij}(t)$ of an irreducible process exponentially decreases (resp. increases) as $t \to \infty$ in the subcritical (resp. supercritical) case, and tends to a finite positive limit in the critical case.

1.1.3 Extinction probability

A key tool for the study of the extinction of a BGWc process, and later for the study of the process conditioned on non-extinction, is its *extinction probability vector* $\mathbf{q} := \lim_{t\to\infty} \mathbf{q}(t)$, where

$$\begin{aligned} \mathbf{q}(t) &:= (q_1(t), \dots, q_d(t)), \\ q_i(t) &:= \mathbb{P}_{\mathbf{e}_i}(\mathbf{X}_t = \mathbf{0}), \quad i = 1 \dots d. \end{aligned} \tag{1.1.19}$$

Note that

$$\begin{aligned} q_i &= \lim_{t\to\infty} q_i(t) = \lim_{t\to\infty} \mathbb{P}_{\mathbf{e}_i}(\mathbf{X}_s = \mathbf{0} \text{ for some } s \leqslant t) \\ &= \mathbb{P}_{\mathbf{e}_i}(\mathbf{X}_s = \mathbf{0} \text{ for some } s \geqslant 0) = \mathbb{P}_{\mathbf{e}_i}(\lim_{t\to\infty} \mathbf{X}_t = \mathbf{0}), \end{aligned}$$

hence the appellation *extinction probability*. By definition, $\mathbf{q}(t) = \mathbf{F}_t(\mathbf{0})$, which by the branching property implies that for all $\boldsymbol{x} \in \mathbb{N}^d$, $\mathbb{P}_{\boldsymbol{x}}(\mathbf{X}_t = \mathbf{0}) = \mathbf{q}(t)^{\boldsymbol{x}}$, and thus

$$\mathbb{P}_{\boldsymbol{x}}\left(\lim_{t\to\infty} \mathbf{X}_t = \mathbf{0}\right) = \mathbf{q}^{\boldsymbol{x}}. \tag{1.1.20}$$

As in the monotype case, there is a dichotomy between explosion of the process and its absorption in the only absorbing point $\mathbf{0}$ (*i.e.* extinction of the process).

Proposition 1.1.15. *Assume **(B0)** and **(B1)**. Then, for all $\boldsymbol{x} \in \mathbb{N}^d$,*

$$\mathbb{P}_{\boldsymbol{x}}\left(\lim_{t\to\infty} \boldsymbol{X}_t = \boldsymbol{0}\right) = 1 - \mathbb{P}_{\boldsymbol{x}}\left(\lim_{t\to\infty} \boldsymbol{X}_t = \infty\right) = \boldsymbol{q}^{\boldsymbol{x}}.$$

Definition 1.1.16. We say that there is almost sure extinction of the process, or that the process almost surely dies out, if $\mathbf{q} = \mathbf{1}$.

Proposition 1.1.17. *The extinction probability vector $\boldsymbol{q}$ satisfies the following equation*

$$\boldsymbol{f}(\boldsymbol{r}) = \boldsymbol{r}. \tag{1.1.21}$$

Proof. Applying (1.1.7) with $\mathbf{r} = \mathbf{0}$, we obtain that $\mathbf{F}_t(\mathbf{F}_s(\mathbf{0})) = \mathbf{F}_{t+s}(\mathbf{0})$, *i.e.* $\mathbf{F}_t(\mathbf{q}(s)) = \mathbf{q}(t+s)$. Letting $s \to \infty$ and using the continuity of $\mathbf{r} \mapsto \mathbf{F}_t(\mathbf{r})$, we obtain $\mathbf{F}_t(\mathbf{q}) = \mathbf{q}$. Hence for all $i = 1 \ldots d$, $t \mapsto \mathbf{F}_t(\mathbf{q})$ is constant, which by (1.1.9) together with the fact that $\alpha_i > 0$ leads to $\boldsymbol{f}(\mathbf{F}_t(\mathbf{q})) = \mathbf{F}_t(\mathbf{q})$, and thus $\boldsymbol{f}(\mathbf{q}) = \mathbf{q}$. ☐

By definition, $\boldsymbol{f}(\mathbf{1}) = \sum_{\mathbf{j}\in\mathbb{N}^d} \mathbf{p}(\mathbf{j}) = \mathbf{1}$, hence $\mathbf{1}$ is also a fixed point of $\boldsymbol{f}$. The following result enables to recognize the extinction probability vector $\mathbf{q}$ among the fixed points of $\boldsymbol{f}$ in $[0,1]^d$ ([Har63] or [Sew75]).

Proposition 1.1.18. *Let us assume **(B0)** and **(B1)**. The function $\boldsymbol{f}$ admits at most one fixed point $\boldsymbol{s}_0$ in $[0,1]^d$ other than $\boldsymbol{1}$. If it exists, then $\boldsymbol{q} = \boldsymbol{s}_0$, and for all $i = 1 \ldots d$, $q_i < 1$. Otherwise, $\boldsymbol{q} = \boldsymbol{1}$.*

The next fundamental result provides a necessary and sufficient condition for the almost sure extinction of the process, under the assumption of irreducibility.

Proposition 1.1.19. *Let us assume **(B0)** and **(B1)** . Then the process almost surely dies out if and only if $\rho \leqslant 0$.*

This result is actually stated in [Sew75] in a more general context than irreducible processes. We briefly recall his result here. Working with reducible processes requires the introduction of the notion of *subprocesses* and *final classes* ([Sew75] Kapitel IV §6).

Definition 1.1.20. A *final class* $C = \{c_1, \ldots, c_p\}$ is a class for the equivalence relation $\leftrightarrow$ defined in Remark 1.1.11, non empty, having the property that there exists one $t > 0$ such that for all $c_i \in C$, the generating function $F_{t,c_i}(\mathbf{r})$ is a linear form with respect to the variables $r_{c_1}, \ldots, r_{c_p}$. This means that we can write $F_{t,c_i}(\mathbf{r})$ in the form

$$F_{t,c_i}(\mathbf{r}) = g_t^{(i,1)}(\mathbf{r}) r_{c_1} + \ldots + g_t^{(i,p)}(\mathbf{r}) r_{c_p}, \tag{1.1.22}$$

where for all $j = 1 \ldots p$, $g_t^{(i,j)}(\mathbf{r})$ depends only on $r_n,\ n \notin C$. If (1.1.22) holds for one $t_0 > 0$, then it holds for all $t > 0$.

The definition of a final class in the discrete-time case might be more intuitive ([Har63] Section 2.10): a final class C has the property that any individual whose type is in C has probability 1 of producing in the next generation exactly one individual whose type is in C (individuals whose types are not in C may also be produced).

Definition 1.1.21. Let $C = \{c_1, \dots, c_p\}$ be a class for the equivalence relation $\leftrightarrow$. We call $C - subprocess$ the process defined by

$$\tilde{\mathbf{X}}_t := (X_{t,c_1}, \dots, X_{t,c_p}).$$

Then $\tilde{\mathbf{X}}$ is still a branching process ([Sew75] Satz 4.6.3), and is by definition irreducible.

We can now quote the more general result of Sewastjanow ([Sew75] Satz 5.2.7).

Proposition 1.1.22. *Let $\rho = \max_C \rho_C$ be the maximal value of the Perron's roots of all the possible C-subprocesses. Then the process almost surely dies out if and only if there are no final classes and $\rho \leqslant 0$.*

We can show that an irreducible process admits a final class only in the trivial case of a simple process, which thus immediately leads to Proposition 1.1.19. Indeed,

Lemma 1.1.23. *Let us assume **(B1)**. Then the process has a final class if and only if the process is simple.*

Proof. If the process is simple, then $\boldsymbol{f}(\mathbf{r})^T = \mathbf{M}\mathbf{r}^T$, and the differential system solved by $\mathbf{F}_t(\mathbf{r})$ becomes the following linear system. For all $\mathbf{r} \in [0,1]^d$,

$$\begin{cases} \dfrac{\partial}{\partial t}\mathbf{F}_t(\mathbf{r})^T = \mathbf{C}\mathbf{F}_t(\mathbf{r})^T, \\ \quad \mathbf{F}_0(\mathbf{r}) = \mathbf{r}. \end{cases} \tag{1.1.23}$$

Hence $\mathbf{F}_t(\mathbf{r})^T = e^{\mathbf{C}t}\mathbf{r}^T$, which implies that for all $t \geqslant 0$ and $\mathbf{r} \in [0,1]^d$, $\mathbf{F}_t(\mathbf{r})$ is a linear form with respect to the variables $r_1, \dots, r_d$. The set of all the types $\{1, \dots, d\}$ consequently builds a final class.

Conversely, if the irreducible process has a final class, then this class must be the set of all the types since they all communicate with each other. Hence the generating function of the process at time t is linear in $\mathbf{r}$ and can be written in the form $\mathbf{F}_t(\mathbf{r})^T = \mathbf{G}(t)\mathbf{r}^T$, for some matrix $\mathbf{G}(t) \in \mathcal{M}_d(\mathbb{R})$. As a consequence of (1.1.7), $\mathbf{G}(t+s) = \mathbf{G}(t)\mathbf{G}(s)$ for all $s, t \geqslant 0$, which, together with the fact that $\mathbf{G}(0)\mathbf{r} = \mathbf{r}$ and thus $\mathbf{G}(0) = \mathbf{I}$, implies there exists some matrix $\mathbf{C} \in \mathcal{M}_d(\mathbb{R})$ such that for all $t \geqslant 0$, $\mathbf{G}(t) = e^{\mathbf{C}t}$. Hence $\mathbf{F}_t(\mathbf{r})^T = e^{\mathbf{C}t}\mathbf{r}^T$, and knowing on the other hand that $\mathbf{F}_t(\mathbf{r})$ is solution of

$$\begin{cases} \dfrac{\partial}{\partial t}F_{t,i}(\mathbf{r}) = \alpha_i \left[f_i\left(\mathbf{F}_t(\mathbf{r})\right) - F_{t,i}(\mathbf{r})\right], \\ \quad F_{0,i}(\mathbf{r}) = r_i, \end{cases} \tag{1.1.24}$$

this imposes that $\boldsymbol{f}$ is linear in $\mathbf{r}$ (namely $\boldsymbol{f}(\mathbf{r}) = \mathbf{r}(\mathbf{A}^{-1}\mathbf{C} + \mathbf{I})$ with $\mathbf{A} := diag(\boldsymbol{\alpha})$). The process is thus simple. $\square$

We now focus on the asymptotic behavior of the extinction probability vector $\mathbf{q}(t)$ in the subcritical and critical cases. Under assumptions **(B1)** and **(B0)** we simply know until now that $\lim_{t\to\infty} \mathbf{q}(t) = \mathbf{1}$. The following proposition gives us the asymptotic behavior of $\mathbf{q}(t)$ as $t \to \infty$ in the subcritical case ([Sew75] Satz 6.2.7). We introduce the following assumption

(XlogX) $\qquad \forall i, j = 1 \dots d, \ \sum_{\mathbf{k}\in\mathbb{N}^d} k_j \ln(k_j) p_i(\mathbf{k}) < \infty.$

According to [Ath68], this condition holds if and only if

$$\forall t > 0, \ \forall i, j = 1 \dots d, \ \mathbb{E}_{\mathbf{e}_i}\left[X_{t,j} \ln X_{t,j}\right] < \infty. \tag{1.1.25}$$

Proposition 1.1.24. *Let us assume **(B1)** and $\rho < 0$. Then the extinction probability vector has the following asymptotic behavior. For all $i = 1 \dots d$,*

- *If **(XlogX)** is satisfied, then there exists $K > 0$ such that*

$$q_i(t) \sim_{t\to\infty} 1 - K\xi_i e^{\rho t}, \tag{1.1.26}$$

- *Otherwise,*

$$q_i(t) =_{t\to\infty} 1 + o(e^{\rho t}). \tag{1.1.27}$$

The next proposition gives us the asymptotic behavior of $\mathbf{q}(t)$ in the critical case ([Sew75] Satz 6.4.4). For this purpose we introduce the following notation.

$$\zeta := \sum_{i,j,k=1}^{d} \alpha_i \frac{\partial^2 f_i}{\partial r_j \partial r_k}(\mathbf{1})\eta_i \xi_j \xi_k. \tag{1.1.28}$$

Proposition 1.1.25. *Let us assume **(B0)**, **(B1)**, $\rho = 0$ and that all the second-order moments of the offspring distribution are finite. Then the extinction probability vector has the following asymptotic behavior. For all $i = 1 \dots d$,*

$$q_i(t) \sim_{t\to\infty} 1 - \frac{2\xi_i}{\zeta t}. \tag{1.1.29}$$

Remark 1.1.26. In the monotype case $d = 1$, if we denote by m and σ^2 the mean and the variance of the offspring distribution, then the matrix $\mathbf{C}$ is equal to the scalar $\alpha(m-1)$. Hence its Perron's root is $\rho = -\alpha(1-m)$, and the related right and left normalized eingenvectors are $\xi = \eta = 1$. Moreover, we have $\zeta = \alpha f''(1) = \alpha\left[\sigma^2 - m(1-m)\right]$. Hence in the critical case $\zeta = \alpha\sigma^2$, and we obtain the following asymptotic behavior,

$$\begin{aligned} &\text{subcritical case} \qquad q(t) \sim_{t\to\infty} 1 - Ke^{-\alpha(1-m)t}, \\ &\text{critical case} \qquad q(t) \sim_{t\to\infty} 1 - \frac{2}{\alpha\sigma^2 t} \quad (\text{if } \sigma^2 > 0). \end{aligned}$$

Remark 1.1.27. The non-simplicity of the critical process in Proposition 1.1.25 implies that $\zeta > 0$. Indeed, if a branching process is such that all the second-order moments of the offspring distribution are null ($\zeta = 0$), then its generating function $\boldsymbol{f}$ is $f_i(\mathbf{r}) = p_i(\mathbf{0}) + \sum_j m_{ij} r_j$, which implies that for all $i = 1 \dots d$, $\sum_j m_{ij} \leqslant 1$. Assuming the existence of a i_0 such that $p_{i_0}(\mathbf{0}) > 0$ would imply moreover that $\sum_j m_{i_0 j} < 1$, and the process would be subcritical. Hence $\zeta = 0$ if and only if the critical process is simple.

1.1.4 Some basic examples

We present here two basic examples in the case $d = 2$, in order to illustrate the previous results about the asymptotic behavior of the extinction probability vector. In Example 1.1.28 we explicitly compute the extinction probability vector at each time t for a 2-dimensional irreducible subcritical process with a very simple offspring generating function. Example 1.1.29 deals with a reducible process, stressing the necessity of the irreducibility assumption in order to obtain an asymptotic behavior as in Proposition 1.1.24 or Proposition 1.1.25.

Example 1.1.28. We consider a process with branching rates $\alpha_1 = \alpha_2 = 1$ and generating function

$$f_1(\mathbf{r}) = f_2(\mathbf{r}) = \frac{1}{3}(1 + r_1 + r_2), \quad \mathbf{r} \in [0,1]^2.$$

By (1.1.9), the extinction probability vector $\mathbf{q}(t)$ is solution of the following linear system of differential equations

$$\begin{cases} \dfrac{d}{dt} q_1(t) = \dfrac{1}{3}(1 - 2q_1(t) + q_2(t)), \\ \dfrac{d}{dt} q_2(t) = \dfrac{1}{3}(1 + q_1(t) - 2q_2(t)), \end{cases} \tag{1.1.30}$$

with initial condition $\mathbf{q}(0) = \mathbf{0}$. Denoting $\mathbf{Q}(t) := \mathbf{q}(t)^T$ we write the system in the canonical form

$$\frac{d}{dt}\mathbf{Q}(t) = \mathbf{C}\mathbf{Q}(t) + \begin{pmatrix} \frac{1}{3} \\ \frac{1}{3} \end{pmatrix}, \quad \mathbf{Q}(0) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{1.1.31}$$

The general solution of the homogeneous system is $\mathbf{Q}_0(t) = e^{t\mathbf{C}}\mathbf{Q}_0(0)$. Denoting $\mathbf{P} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ we get $\mathbf{P}^{-1}\mathbf{B}\mathbf{P} = \begin{pmatrix} -\frac{1}{3} & 0 \\ 0 & -1 \end{pmatrix}$. Then $\mathbf{P}^{-1}e^{t\mathbf{C}}\mathbf{P} = \begin{pmatrix} e^{-\frac{t}{3}} & 0 \\ 0 & e^{-t} \end{pmatrix}$, and

$$\mathbf{Q}_0(t) = \frac{1}{2}\begin{pmatrix} e^{-\frac{t}{3}} + e^{-t} & e^{-\frac{t}{3}} - e^{-t} \\ e^{-\frac{t}{3}} - e^{-t} & e^{-\frac{t}{3}} + e^{-t} \end{pmatrix}\mathbf{Q}_0(0).$$

A particular solution of $\frac{d}{dt}\mathbf{Q}(t) = \mathbf{C}\mathbf{Q}(t) + \begin{pmatrix} \frac{1}{3} \\ \frac{1}{3} \end{pmatrix}$ is $\mathbf{Q}_p(t) = \mathbf{1}^T$. Hence the general solution is

$$\mathbf{Q}(t) = \begin{pmatrix} 1 + (\mu+\nu)e^{-\frac{t}{3}} + (\mu-\nu)e^{-t} \\ 1 + (\mu+\nu)e^{-\frac{t}{3}} + (-\mu+\nu)e^{-t} \end{pmatrix}, \quad \mu, \nu \in \mathbb{R}.$$

The initial condition $\mathbf{Q}(0) = \mathbf{0}^T$ then implies $\mu = \nu = -\frac{1}{2}$, and the solution of (1.1.31) is

$$q_1(t) = q_2(t) = 1 - e^{-\frac{t}{3}}. \tag{1.1.32}$$

Note that in this example, $\mathbf{C} = \begin{pmatrix} -\frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & -\frac{2}{3} \end{pmatrix}$ has eigenvalues $-\frac{1}{3}$ and -1 and Perron's root $\rho = -\frac{1}{3}$. The process is thus irreducible and subcritical. The related right and left eigenvectors with the normalization $\boldsymbol{\xi}\cdot\mathbf{1} = 1$ and $\boldsymbol{\xi}\cdot\boldsymbol{\eta} = 1$ are $\boldsymbol{\xi} = (\frac{1}{2}, \frac{1}{2})$ and $\boldsymbol{\eta} = (1,1)$, and the expression of the extinction probability vector (1.1.32) obviously corresponds at each time t to the asymptotic behavior (1.1.26) obtained in Proposition 1.1.24, as t tends to infinity.

Example 1.1.29. We consider a process with branching rates $\alpha_1 = \alpha_2 = 1$ and generating function

$$\begin{cases} f_1(\mathbf{r}) = r_1 r_2 \\ f_2(\mathbf{r}) = \dfrac{1}{2}(1 + r_2^2) \end{cases}, \qquad \mathbf{r} \in [0,1]^2.$$

The extinction probability vector $\mathbf{q}(t)$ is then solution of the following system of differential equations

$$\begin{cases} \dfrac{d}{dt}q_1(t) = q_1(t)(q_2(t) - 1), \\ \dfrac{d}{dt}q_2(t) = \dfrac{1}{2} - q_2(t) + \dfrac{1}{2}q_2(t)^2, \end{cases} \tag{1.1.33}$$

with initial condition $\mathbf{q}(0) = \mathbf{0}$. A particular solution of $\frac{d}{dt}q_2(t) = \frac{1}{2} - q_2(t) + \frac{1}{2}q_2(t)^2$ is $q_{2,p}(t) = 1$. Denoting $q_2(t) = 1 + y(t)$, then $y(t)$ solves $\frac{dy(t)}{dt} = \frac{1}{2}y(t)^2$, hence $\frac{d}{dt}[\frac{1}{y(t)}] = -\frac{dy(t)}{dt}\frac{1}{y(t)^2} = -\frac{1}{2}$ and $y(t) = \frac{1}{k - \frac{1}{2}t}$, $k \in \mathbb{R}$. Then $q_2(t) = 1 - \frac{2}{t-2k}$, and the initial condition $q_2(0) = 0$ leads to

$$q_2(t) = 1 - \frac{2}{t+2}. \tag{1.1.34}$$

We can now solve $\frac{d}{dt}q_1(t) = q_1(t)(q_2(t) - 1)$, which becomes $\frac{d}{dt}q_1(t) = -q_1(t)\frac{2}{t+2}$. Hence $q_1(t) = ke^{-2\ln(t+2)} = \frac{k}{(t+2)^2}$, $k \in \mathbb{R}$, and the initial condition imposes

$$q_1(t) = 0. \tag{1.1.35}$$

In this example, the process almost surely dies out if initiated by one individual of type 2, but almost surely survives if initiated by one individual of type 1. Obviously, we are not in the framework of Proposition 1.1.24 or Proposition 1.1.25, since $\mathbf{C} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ is a reducible matrix. Note however that initiating the process with one individual of type 2 leads to a single-type binary critical process (with individuals of type 2), and (1.1.34) leads to the right asymptotic behavior in the single-type critical case $q_2(t) \sim_{t\to\infty} 1 - \frac{2}{t}$.

1.2 Multitype Feller diffusion process

In this section we consider an other model of population dynamics which is strongly related with the BGWc process, namely the Feller diffusion process. Quoting Feller from its pioneering paper [Fel51], "*relatively small populations require discrete models, but for large populations it is possible to apply a continuous approximation, and this leads to processes of the diffusion type*". Feller showed indeed in this article that when the population under consideration is large and the time scale is fast, a supercritical monotype BGW process can be approximated by a density $v(t,x)$ satisfying the diffusion equation

$$\frac{\partial}{\partial t}\left[v(t,x)\right] = \frac{\sigma^2}{2}\frac{\partial^2}{\partial x^2}\left[xv(t,x)\right] - \frac{\partial}{\partial x}\left[xv(t,x)\right], \tag{1.2.1}$$

for some $\sigma^2 > 0$, with the boundary condition $v(t,0) = 0$. The corresponding stochastic differential equation is then

$$dX_t = \sigma\sqrt{X_t}dB_t + X_t dt. \tag{1.2.2}$$

This approximation was later made more rigorous and complete notably by Jiřina ([Jir69]) and Lindvall ([Lin72]). The diffusion approximation is now also known for processes with several types of individuals and of any class of criticality, and leads to the multitype Feller diffusion process. We do not give here the details of the approximation, which is the subject of **Subsection 3.2.1**, but focus here on basic properties of the diffusion process, mostly related to its extinction.

We consider a d-type Feller diffusion process with sample paths in $C(\mathbb{R}^+, \mathbb{R}_+^d)$, the space of continuous functions from $\mathbb{R}_+$ to $\mathbb{R}_+^d$. Similarly as for the BGWc process studied in Section 1.1, we denote by $\mathbb{P}$ its law on $(\Omega, \mathbf{X}_t, \mathcal{F}_t)$. Since we never work with the two processes simultaneously, this should not bring any confusion and rather simplify the notation. After introducing in **Subsection 1.2.1** some definitions related to general continuous-state branching processes, and more precisely Feller diffusion processes, we focus in **Subsection 1.2.2** on a fundamental tool for our study which is the martingale problem, and finally discuss in **Subsection 1.2.3** properties related to the extinction of Feller diffusion processes.

1.2.1 Definitions and preliminaries

Multitype continuous-state branching processes

Multitype Feller diffusion processes belong to the broader class of continuous-state branching processes (CB) introduced by Jiřina ([Jir58]). These processes are by definition Markov processes with right-continuous paths whose transition probabilities satisfy the *branching property*, which in the time-homogeneous setting means that

$$P_t(\boldsymbol{x}+\boldsymbol{y}, .) = P_t(\boldsymbol{x}, .) * P_t(\boldsymbol{y}, .), \tag{1.2.3}$$

where $P_t(\boldsymbol{x}, A)$ denotes the transition probability from state $\boldsymbol{x} \in \mathbb{R}_+^d$ at time 0 to the set A at time t. A process starting in $\boldsymbol{x}+\boldsymbol{y}$ has such the same law as the sum of two independent processes starting respectively in $\boldsymbol{x}$ and $\boldsymbol{y}$. This *additive property* for the transition probabilities translates into a *multiplicative property* for the Laplace transform. Indeed, as a consequence of (1.2.3), the Laplace transform of a (time-homogeneous) multitype CB process is multiplicative with respect to the initial condition and can be written in the form, for all $\boldsymbol{\lambda}, \boldsymbol{x} \in \mathbb{R}_+^d$,

$$\int_{\mathbb{R}_+^d} e^{-\boldsymbol{\lambda}\cdot\boldsymbol{y}} P_t(\boldsymbol{x}, d\boldsymbol{y}) = e^{-\boldsymbol{x}\cdot\mathbf{u}_t(\boldsymbol{\lambda})}. \tag{1.2.4}$$

By (1.2.3) together with the Chapman-Kolmogorov equation, the so-called *cumulant* $\mathbf{u}_t(\boldsymbol{\lambda}) = (u_{t,1}(\boldsymbol{\lambda}), \dots, u_{t,d}(\boldsymbol{\lambda}))$ of a CB satisfies for all $s, t \geqslant 0$,

$$\mathbf{u}_{t+s}(\boldsymbol{\lambda}) = \mathbf{u}_t(\mathbf{u}_s(\boldsymbol{\lambda})). \tag{1.2.5}$$

The law of a d-type CB process is characterized by a function $\boldsymbol{\psi}(\boldsymbol{\lambda}) := (\psi_1(\boldsymbol{\lambda}), \dots, \psi_d(\boldsymbol{\lambda}))$ called *branching mechanism function*, which is such that the cumulant is the unique solution of the differential equation

$$\begin{cases} \dfrac{\partial}{\partial t}\mathbf{u}_t(\boldsymbol{\lambda}) = \boldsymbol{\psi}(\mathbf{u}_t(\boldsymbol{\lambda})), \\ \mathbf{u}_0(\boldsymbol{\lambda}) = \boldsymbol{\lambda}. \end{cases} \tag{1.2.6}$$

Before describing the function $\boldsymbol{\psi}$ corresponding to Feller diffusions, let us briefly mention a few facts about the interpretation of the branching mechanism function of general CB processes. Lamperti first proved a random-time change connection between one-dimensional CB processes and spectrally positive Lévy processes, *i.e.* processes with stationary independent increments whose Lévy measure is concentrated on $[0, \infty[$ ([Lamp67], see also [Bin76]). It was then later observed in [LeGaJan98] that there is an explicit formula expressing the *height process* of a subcritical CB process as a functional of a spectrally positive Lévy process whose Laplace exponent is precisely the branching mechanism ψ. The height process of a CB process can be seen as the continuous analogue of the *contour process* of a BGW process (see *e.g.* [DuLeGa02]).

In the single-type case, the branching mechanism function ψ of a general CB process is specified by the Lévy-Khinchin formula

$$\psi(\lambda) = a\lambda - \frac{1}{2}b\lambda^2 + \int_{]0,\infty[} \left(1 - e^{-\lambda r} - \lambda r\right) \Lambda(dr), \quad \lambda \in \mathbb{R}_+, \tag{1.2.7}$$

where $a \in \mathbb{R}$ is the deterministic linear drift, $b \geqslant 0$ is the variance rate of the Brownian component, and the Lévy measure Λ is a Radon measure on $]0, \infty[$ such that $\int_{]0,\infty[}(r \wedge r^2)\Lambda(dr) < \infty$. Denoting $\rho := \psi'(0+)$, the CB process is called *subcritical*, *critical* or *supercritical* according as $\rho < 0$, $\rho = 0$ or $\rho > 0$.

The first important example is when $\psi(\lambda) = -\frac{1}{2}\lambda^2$. This corresponds to the single-type critical Feller diffusion process. In this case, the height process is known to be a reflected linear Brownian motion. More generally, if the branching mechanism is *quadratic*, $\psi(\lambda) = a\lambda - \frac{1}{2}b\lambda^2$, the corresponding CB process is a Feller diffusion process which is *subcritical*, *critical* or *supercritical* according as $a < 0$, $a = 0$ or $a > 0$. By (1.2.6), the cumulant $u_t(\lambda)$ is solution of

$$\begin{cases} \dfrac{\partial}{\partial t}u_t(\lambda) = au_t(\lambda) - \dfrac{1}{2}b\left(u_t\left(\lambda\right)\right)^2, \\ u_0(\lambda) = \lambda. \end{cases} \tag{1.2.8}$$

This yields in the subcritical and supercritical case $a \neq 0$ the following explicit formula, for every $\lambda \geqslant 0$,

$$u_t(\lambda) = \frac{\lambda e^{at}}{1 - \frac{b}{2a}\lambda(1 - e^{at})}, \tag{1.2.9}$$

while in the critical case $a = 0$,

$$u_t(\lambda) = \frac{\lambda}{1 + \frac{1}{2}b\lambda t}. \tag{1.2.10}$$

An other important class of CB processes corresponds to *α-stable* branching mechanism functions of the form $\psi(\lambda) = c\lambda^\alpha$, $\alpha \in]1, 2[$, $c \in \mathbb{R}$. If $\alpha < 2$ the corresponding Lévy process has (non-negative) jumps, and the sample paths of the CB process are not in $C(\mathbb{R}^+, \mathbb{R}_+)$ as in the diffusion case, but simply in $\mathcal{D}(\mathbb{R}^+, \mathbb{R}_+)$.

Similarly, in the multitype case, the branching mechanism function $\boldsymbol{\psi}$ of a general CB process is of the form (see [RySko70])

$$\psi_i(\boldsymbol{\lambda}) = \sum_{j=1}^d a_{ij}\lambda_j - \frac{1}{2}b_i\lambda_i^2 + \int_{\mathbb{R}_+^d} \left(1 - e^{-\boldsymbol{\lambda}\cdot\mathbf{r}} - \frac{\lambda_i r_i}{1 + \mathbf{r}\cdot\mathbf{r}}\right) \Lambda_i(d\mathbf{r}), \quad \boldsymbol{\lambda} \in \mathbb{R}_+^d, \tag{1.2.11}$$

where $\inf_i b_i > 0$, and the Lévy measure Λ_i is such that $\int_{|\mathbf{r}|<1}(|\mathbf{r}| - r_i + r_i^2)\Lambda_i(d\mathbf{r}) < \infty$.

Multitype Feller diffusion process

Let us consider a d-type Feller diffusion process with law $\mathbb{P}$ on the probability space $(\Omega, \mathbf{X}_t, \mathcal{F}_t)$. Its branching mechanism function $\boldsymbol{\psi}(\boldsymbol{\lambda}) := (\psi_1(\boldsymbol{\lambda}), \dots, \psi_d(\boldsymbol{\lambda}))$ is of the form, for all $\boldsymbol{\lambda} \in \mathbb{R}_+^d$,

$$\psi_i(\boldsymbol{\lambda}) := \sum_{j=1}^d c_{ij}\lambda_j - \frac{1}{2}\sigma_i^2\lambda_i^2, \tag{1.2.12}$$

where for all $i \neq j$, $c_{ij} \geqslant 0$, and $\inf_i \sigma_i^2 > 0$. The matrix $\mathbf{C} \in \mathcal{M}_d(\mathbb{R})$ with entries c_{ij} is called the *mutation matrix* of the process. As shown in **Section 3.2.1**, this matrix as well as the *variance parameters* $\sigma_1^2, \dots, \sigma_d^2$ are related to the first and second-order moments of the approximating BGWc processes. In particular, the mutation matrix $\mathbf{C}$ represents the interaction between the different types.

From what precedes ((1.2.4) and (1.2.6)), the Laplace transform of the Feller diffusion process is given for all $t \geqslant 0$, $\boldsymbol{\lambda} \in \mathbb{R}_+^d$ and $\boldsymbol{x} \in \mathbb{R}_+^d$ by

$$\mathbb{E}_{\boldsymbol{x}}(e^{-\boldsymbol{\lambda}\cdot\mathbf{X}_t}) = e^{-\boldsymbol{x}\cdot\mathbf{u}_t(\boldsymbol{\lambda})},$$

where, for all $i = 1 \dots d$,

$$\begin{cases} \dfrac{\partial}{\partial t}u_{t,i}(\boldsymbol{\lambda}) = \displaystyle\sum_{j=1}^d c_{ij}u_{t,j}(\boldsymbol{\lambda}) - \frac{1}{2}\sigma_i^2 u_{t,i}(\boldsymbol{\lambda})^2, \\ \quad u_{0,i}(\boldsymbol{\lambda}) = \lambda_i. \end{cases} \tag{1.2.13}$$

1.2.2 Associated martingale problem and SDE

The infinitesimal generator of the diffusion process is given on $D(L) := C^2\left(\mathbb{R}_+^d, \mathbb{R}\right)$ by ([RySko70]),

$$(Lf)(\boldsymbol{x}) := \frac{1}{2}\sum_{i=1}^d \sigma_i^2 x_i \frac{\partial^2 f}{\partial x_i^2}(\boldsymbol{x}) + \sum_{i=1}^d x_i \sum_{j=1}^d c_{ij}\frac{\partial f}{\partial x_j}(\boldsymbol{x}). \tag{1.2.14}$$

We denote by $D_0(L) := C_b^2\left(\mathbb{R}_+^d, \mathbb{R}\right)$ the set of bounded C^2-functions on $\mathbb{R}_+^d$. Then $\mathbb{P}$ is the unique solution to the martingale problem $\mathcal{MP}(L, D_0(L))$ (see *e.g.* [EthKur86], Section 8.1, Theorem 1.7).

In particular, considering for a fixed $\boldsymbol{\lambda} \in \mathbb{R}_+^d$ the C^2 bounded function $f(\boldsymbol{x}) := e^{-\boldsymbol{\lambda}\cdot\boldsymbol{x}}$, we have for all $\boldsymbol{x} \in \mathbb{R}_+^d$,

$$(Lf)(\boldsymbol{x}) = \sum_{i=1}^d \Big(\frac{1}{2}\sigma_i^2\lambda_i^2 - \sum_{j=1}^d c_{ij}\lambda_j\Big) x_i e^{-\boldsymbol{\lambda}\cdot\boldsymbol{x}} = -\boldsymbol{\psi}(\boldsymbol{\lambda})\cdot\boldsymbol{x}\ e^{-\boldsymbol{\lambda}\cdot\boldsymbol{x}}.$$

Writing the martingale problem $\mathcal{MP}(L, D_0(L))$ for f, we obtain that for every $\boldsymbol{\lambda} \in \mathbb{R}_+^d$,

$$e^{-\boldsymbol{\lambda}\cdot\mathbf{X}_t} - e^{-\boldsymbol{\lambda}\cdot\mathbf{X}_0} + \int_0^t \mathbf{X}_s\cdot\boldsymbol{\psi}(\boldsymbol{\lambda})\, e^{-\boldsymbol{\lambda}\cdot\mathbf{X}_s} ds \quad \text{is a } (\mathbb{P}, \mathcal{F}_t)\text{-martingale.} \tag{1.2.15}$$

It might also be of interest to mention that the process is the unique weak solution of the stochastic differential equation (written here for column vectors)

$$d\mathbf{X}_t = \boldsymbol{\Sigma}(\mathbf{X}_t)\, d\mathbf{B}_t + \mathbf{C}^T\mathbf{X}_t\, dt, \tag{1.2.16}$$

where $\mathbf{B}_t$ is a standard d-dimensional Brownian motion, and for all $\boldsymbol{x} \in \mathbb{R}_+^d$,

$$\boldsymbol{\Sigma}(\boldsymbol{x}) = \begin{pmatrix} \sigma_1\sqrt{x_1} & & 0 \\ & \ddots & \\ 0 & & \sigma_d\sqrt{x_d} \end{pmatrix}. \tag{1.2.17}$$

Componentwise, (1.2.16) becomes, for all $i = 1 \dots d$,

$$dX_{t,i} = \sigma_i \sqrt{X_{t,i}} dB_{t,i} + \sum_{j=1}^{d} c_{ji} X_{t,j} dt. \tag{1.2.18}$$

The uniqueness of weak solutions for (1.2.16) can be obtained by applying Yamada-Watanabe's criteria ([YaWat71] p.164).

We recognize in the monotype case the classical "Feller" SDE

$$dX_t = \sigma \sqrt{X_t} dB_t + cX_t dt. \tag{1.2.19}$$

1.2.3 Extinction of the process

Let us now consider the problem of extinction of a multitype Feller diffusion process. As for all the CB processes, $\mathbf{0}$ is an absorbing state. Thanks to the branching property, the information about the extinction/absorption strongly depends on the cumulant of the process and does not depend on the initial distribution. For the special case of a Feller diffusion process, this information is related to the mutation matrix.

Let us introduce the following assumption.

(F1) The mutation matrix $\mathbf{C}$ is *irreducible*.

Since by definition all the non-diagonal entries of $\mathbf{C}$ are non-negative, the Perron-Frobenius Theorem (Theorem 1.1.9) can be applied under **(F1)**. Denoting by ρ the Perron's root of $\mathbf{C}$, the Feller diffusion process is called *subcritical*, *critical* or *supercritical* according as $\rho < 0$, $\rho = 0$ or $\rho > 0$. In the following, we denote by $\boldsymbol{\xi}$ (resp. $\boldsymbol{\eta}$) the right (resp. left) eigenvector for ρ with normalization $\boldsymbol{\xi} \cdot \mathbf{1} = 1$, $\boldsymbol{\eta} \cdot \boldsymbol{\xi} = 1$. Again, there should not be any confusion with the eigenvalue and eigenvectors associated with the BGWc studied in Section 1.1 since we never work with the two processes in the same section.

The probability of extinction at time t of the process is given by, for all $\boldsymbol{x} \in \mathbb{R}_+^d$,

$$\mathbb{P}_{\boldsymbol{x}}(\mathbf{X}_t = \mathbf{0}) = \lim_{\boldsymbol{\lambda} \to \infty} \mathbb{E}_{\boldsymbol{x}}\left(e^{-\boldsymbol{\lambda} \cdot \mathbf{X}_t}\right) = e^{-\boldsymbol{x} \cdot \lim_{\boldsymbol{\lambda} \to \infty} \mathbf{u}_t(\boldsymbol{\lambda})}, \tag{1.2.20}$$

and the probability of extinction by

$$\mathbb{P}_{\boldsymbol{x}}\left(\lim_{t \to \infty} \mathbf{X}_t = \mathbf{0}\right) = \lim_{t \to \infty} \lim_{\boldsymbol{\lambda} \to \infty} \mathbb{E}_{\boldsymbol{x}}\left(e^{-\boldsymbol{\lambda} \cdot \mathbf{X}_t}\right) = e^{-\boldsymbol{x} \cdot \lim_{t \to \infty} \lim_{\boldsymbol{\lambda} \to \infty} \mathbf{u}_t(\boldsymbol{\lambda})}. \tag{1.2.21}$$

We thus define the vectors

$$\begin{aligned} \mathbf{u}_t &:= \lim_{\boldsymbol{\lambda} \to \infty} \mathbf{u}_t(\boldsymbol{\lambda}), \\ \mathbf{u} &:= \lim_{t \to \infty} \mathbf{u}_t \geqslant 0, \end{aligned} \tag{1.2.22}$$

and obtain

$$\begin{aligned} \mathbb{P}_{\boldsymbol{x}}(\mathbf{X}_t = \mathbf{0}) &= e^{-\boldsymbol{x} \cdot \mathbf{u}_t}, \\ \mathbb{P}_{\boldsymbol{x}}\left(\lim_{t \to \infty} \mathbf{X}_t = \mathbf{0}\right) &= e^{-\boldsymbol{x} \cdot \mathbf{u}}. \end{aligned}$$

The following result (see *e.g.* [Jir64]) states that, just as for the BGWc process, extinction occurs almost surely in the (sub)critical case.

Proposition 1.2.1. *Let us assume **(F1)**. Then $\boldsymbol{u} = \mathbf{0}$ if and only if $\rho \leqslant 0$.*

Chapter 2

Multitype branching processes conditioned on non-extinction

We mentioned in the introduction how much the study of the extinction of populations is of a great interest in biology. Conditioning on non-extinction can notably lead to a stationary behavior of the process. We point out that branching processes can also result in populations with stable sizes if one allows an immigration component or assume an increasing number of ancestors. But conditioning on extinction or non-extinction provides in addition a lot of information about the evolution of the population before extinction, right before extinction, or in case of a very late extinction. We review in **Section 2.1** some of the historical results related to this topic, and from **Section 2.3** we focus on two kinds of conditioning for multitype BGWc and Feller diffusion processes, for which we provide a systematic study (which is new as far as multitype BGWc processes are concerned). The first conditioning consists in studying the asymptotic behavior of the process $\mathbf{X}_t$ under the condition that it is not extinct at time $t+\theta$, for some $\theta \geqslant 0$, but does eventually die out. This conditional limit distribution is a generalization of the Yaglom limit, obtained for $\theta = 0$, and is the subject of **Section 2.3**. The second object of interest, studied in **Section 2.4**, is the so-called Q-process associated with the multitype branching processes (BGWc or Feller diffusion), *i.e.* the process "conditioned on not being extinct in the distant future and on being extinct in the even more distant future", as described in [AthNey72]. We thus always consider population which are doomed to become extinct. We know from Subsection 1.1.3 and Subsection 1.2.3 that this is the case for subcritical and critical processes, but for the sake of completeness we chose to extend the study to supercritical processes with a positive risk of extinction, which is done thanks to **Section 2.2**.

2.1 Historical introduction: various ways of conditioning

The pioneer historical results on survival chances and conditional population size given non-extinction are due to Kolmogorov and Yaglom. In 1938 Kolmogorov provided the asymptotical survival probability of a single-type BGW process. Denoting by m and σ^2 the mean value and variance of the offspring distribution f, and assuming the existence of a third-order moment (later proved to be unnecessary), he showed that in the subcritical case $m < 1$, for all $x \in \mathbb{N}$ ([Kolm38]),

$$\mathbb{P}_x (X_n > 0) \sim_{n\to\infty} Cxm^n, \tag{2.1.1}$$

where C is some positive constant, while in the critical case $m = 1$, for $0 < \sigma^2 < \infty$,

$$\mathbb{P}_x (X_n > 0) \sim_{n\to\infty} \frac{2x}{\sigma^2 n}. \tag{2.1.2}$$

In both cases the probability of survival thus decreases to 0 as n tends to infinity, but much slower in the critical case that in the subcritical case.

Yaglom limit and quasi-stationary distributions

About ten years later, Yaglom proved under moment restriction the fundamental result that in the subcritical case, the distribution of $(X_n|X_n > 0)$ converges to a proper distribution ([Yag47]). His proof was later simplified and the moment restriction removed in [SenVer66] and [Jof67]. The current usual formulation is the following: if $m < 1$, then for each $x \in \mathbb{N}^*$,

$$\lim_{n\to\infty} \mathbb{P}_x\left(X_n = y \,|\, X_n > 0\right) = \nu(y), \tag{2.1.3}$$

where ν is a probability measure on $\mathbb{N}^*$ which is independent of x, now referred to as the *Yaglom distribution* associated with the process $(X_n)_{n\geqslant 0}$. The generating function $H(r) = \sum_{x=1}^{\infty} \nu(x) r^x$, $r \in [0,1]$, of this distribution satisfies the non-linear implicit equation

$$H \circ f = 1 - m + mH. \tag{2.1.4}$$

In the critical case, Yaglom proved that (again under a third moment assumption, removed later) for every $x \in \mathbb{N}^*$, assuming that $0 < \sigma^2 < \infty$,

$$\lim_{n\to\infty} \mathbb{P}_x\left(\frac{X_n}{n} > y \,|\, X_n > 0\right) = e^{-\frac{2y}{\sigma^2}}. \tag{2.1.5}$$

Hence the rescaled process $\frac{X_n}{n}$ converges in distribution conditionally on non-extinction to an exponential law with parameter $\frac{2}{\sigma^2}$. Corresponding results of (2.1.3) and (2.1.5) for single-type BGWc processes then followed, first in [Sew51] and later in [Con67], the latter using embedding arguments. It then extended to the field of multitype branching processes, first for BGW processes ([JofSpit67]) and later on for BGWc processes ([Sew75], see Proposition 2.3.3 in this work).

It can be proved that the Yaglom distribution as defined by (2.1.3) is a *quasi-stationary distribution* ([SenVer66]), in the sense that it is a stationary distribution for the dynamics conditioned on non-extinction, *i.e.*

$$\mathbb{P}_\nu\left(X_1 = y \,|\, X_1 > 0\right) = \nu(y). \tag{2.1.6}$$

Hence

$$\frac{\sum_{x=1}^{\infty} \nu(x) P(x,y)}{1 - \sum_{x=1}^{\infty} \nu(x) P(x,0)} = \nu(y),$$

or equivalently,

$$\sum_{x=1}^{\infty} \nu(x) P(x,y) = m\nu(y), \tag{2.1.7}$$

since by (2.1.4) applied in 0 together with the branching property, one obtains that

$$1 - m = H\left(f\left(0\right)\right) = H\left(P\left(1,0\right)\right) = \sum_{x=1}^{\infty} \nu(x) P(1,0)^x = \sum_{x=1}^{\infty} \nu(x) P(x,0).$$

It thus appears that ν is a left eigenvector of the transition matrix for the eingenvalue m. Although the Yaglom limit is uniquely defined by property (2.1.3), it is not the only quasi-stationary distribution, and consequently not the only conditional limit distribution. It was proved indeed in [SenVer66] that there exists a continuous range of quasi-stationary distributions $\{\nu_\alpha, 0 < \alpha \leqslant 1\}$ such that for each $0 < \alpha \leqslant 1$, ν_α is a left eigenvector for the eigenvalue m^α (the superscript α denotes here the usual exponentiation):

$$\sum_{x=1}^{\infty} \nu_\alpha(x) P(x,y) = m^\alpha \nu_\alpha(y). \tag{2.1.8}$$

Denoting by H_α the generating function of ν_α, it satisfies the nonlinear implicit equation

$$H_\alpha \circ f = 1 - m^\alpha + m^\alpha H_\alpha. \tag{2.1.9}$$

This paper of Seneta and Vere-Jones actually dealt with the problem of quasi-stationarity for more general absorbing Markov chains, and initiated many other works (see *e.g.* [Fer95], and [Gos01] for an application on state-dependent branching processes). The equivalent theory for diffusion processes started with [Man61] and was then developed by many authors. In our direction of research, we shall quote the results of Lambert who dedicated his paper [Lamb07] on CB processes, and proved that in the subcritical case $\rho := \psi'(0+) < 0$ (see (1.2.7) for a definition of the branching mechanism ψ), there exists a family of quasi-stationary distributions $\{\nu_\gamma,\, 0 < \gamma \leqslant -\rho\}$, *i.e.* satisfying

$$\mathbb{P}_{\nu_\gamma}\left(X_t \in A \,|\, X_t > 0\right) = \nu_\gamma(A). \tag{2.1.10}$$

The Yaglom distribution then corresponds to $\nu_{-\rho}$. For a recent work on this topic, we refer *e.g.* to [Cat09], dealing with the existence, uniqueness and domain of attraction of quasi-stationary distributions for a large class of diffusion models arising from population dynamics.

Quasiextinction probabilities

There are many possible ways to generalize the study of the Yaglom limit. A first idea is to condition the population on *not being too small*, instead of conditioning on non-extinction. Seneta and Vere-Jones proved for example in [SenVer68] that for a subcritical Jiřina process (*i.e.* a discrete-time continuous-state branching process), for each $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left(X_n \leqslant x \,|\, X_n > \varepsilon\right) \tag{2.1.11}$$

exists and is a nondegenerate law on $[\varepsilon, +\infty[$. Similar inquiries were done for CB processes, with a first ecological application in [Gin82], where this distributions are called *quasiextinction probabilities*.

Extinction in a close future

Alternatively, instead of conditioning the process on non-extinction at the present time as in (2.1.3), it might be interesting to condition in addition this process on *extinction in a close future*, and look at the asymptotic distribution. It was proved in [Sen67] that for single-type BGW processes one obtains in the critical case a nondegenerate result: if $p_1 > 0$, for every fixed integer $k > 0$,

$$\lim_{n\to\infty} \mathbb{P}_1\left(X_n = y \,|\, X_n > 0,\, X_{n+k} = 0\right) = \mu_k(y), \tag{2.1.12}$$

where $\sum_{y=1}^{\infty} \mu_k(y) = 1$. Similarly, one can find in [Pak09] that for a critical CB process, for any fixed $\theta > 0$,

$$\lim_{t\to\infty} \mathbb{P}_x(X_t \leqslant y \,|\, X_t > 0,\, X_{t+\theta} = 0) \tag{2.1.13}$$

exists and is the distribution function of a gamma law, depending on θ.

Yaglom-type limits

An other natural investigation is to consider as generalization of (2.1.3) and (2.1.5) that the extinction is delayed by at least k time-units, for some $k \in \mathbb{N}^*$. For example (see [AthNey72] Theorem 1.14.1), for a subcritical BGW process with $p_1 > 0$, one can prove the existence of the following limit,

$$\lim_{n\to\infty} \mathbb{P}_1\left(X_n = y \,|\, X_{n+k} > 0\right) = b_k(y), \tag{2.1.14}$$

where $\sum_{y=1}^{\infty} b_k(y) = 1$. We call in this work this kind of limits the *Yaglom-type limits*. Indeed we investigate in **Section 2.3**, for multitype BGWc and Feller diffusion processes, conditional distributions of the form

$$\lim_{t\to\infty} \mathbb{P}_{\boldsymbol{x}}\left(\mathbf{X}_t \in \,.\, \,|\, \mathbf{X}_{t+\theta} \neq \mathbf{0}\right), \tag{2.1.15}$$

for a given $\theta \geqslant 0$.

Q-processes

Instead of looking at the limit as n tends to infinity of (2.1.14), and obtain a limit distribution depending on k, it might also be interesting to hold n fixed and let k tend to infinity:

$$\lim_{k\to\infty} \mathbb{P}_x \left(X_n = y \,|\, X_{n+k} > 0\right). \tag{2.1.16}$$

It was first pointed out by Harris in [Har51] that such a limit always exists for BGW processes, and was later proved in [LaNey68] by Lamperti and Ney that in the (sub)critical case this defines a Markov process, referred to as *Q-process*. Denoting by $P(x,y)$ the transition probabilities of the original BGW process, the associated *Q-process* has transition probabilities $P^*(x,y)$ satisfying

$$P^*(x,y) = \frac{1}{m}\frac{y}{x}P(x,y), \quad x,y > 0. \tag{2.1.17}$$

In the multitype case, this becomes ([DalJof08])

$$P^*(\boldsymbol{x},\boldsymbol{y}) = \frac{1}{\rho}\frac{\boldsymbol{y}\cdot\boldsymbol{\xi}}{\boldsymbol{x}\cdot\boldsymbol{\xi}}P(\boldsymbol{x},\boldsymbol{y}), \quad \boldsymbol{x},\boldsymbol{y} \neq \boldsymbol{0}, \tag{2.1.18}$$

where $\boldsymbol{\xi}$ denotes the right normalized eigenvector (with positive coordinates) of the mean matrix for its Perron's root ρ. By definition, the Q-process associated with a branching process can be seen as the process conditioned on non-extinction in the remote future, and provides information about the behavior of the population in case of a very late extinction. The study of such objects has now extended to the Dawson-Watanabe process (which is a measure-valued branching process). It was initiated for the single-type case by Rœlly and Rouault ([RoeRou89]), and now covers the multitype case ([ChaRoe08]). In this thesis (see **Section 2.4**), we shall focus on the Q-processes associated with multitype BGWc and Feller diffusion processes.

Limiting diffusion

We have looked at the process $(X_n \,|\, X_{n+k} > 0)$ and considered first its limit as $n \to \infty$ and then as $k \to \infty$. As a complementary result in this direction, we shall mention [LaNey68], where n and k grow simultaneously, and which is also a generalization of (2.1.5). Consider a critical BGW process with finite variance. Then, for each fixed $t < 1$,

$$\lim_{n\to\infty} \mathbb{P}_1 \left(\frac{X_{[nt]}}{n} = .\,|\, X_n > 0\right) = \mathcal{E}xp\left(\frac{2}{t\sigma^2}\right) * \mathcal{E}xp\left(\frac{2}{t(1-t)\sigma^2}\right). \tag{2.1.19}$$

Moreover, looking at $(\frac{X_{[nt]}}{n} \,|\, X_n > 0)$ for fixed n and variable $t \in [0,1]$ as a stochastic process with parameter t, say $Y_n(t)$, one obtains a sequence of processes $Y_1(t), Y_2(t), \ldots$ which converges (in the sense of finite dimensional distributions) as $n \to \infty$ to a limiting diffusion process (which is not time-homogeneous).

Reduced branching processes

An other natural alternative to the Yaglom limit IS to condition the population to be still extant at some fixed time T, but this yields time-inhomogeneous kernels. However, considering only the individuals at each time $t \in [0,T]$ having descendants at time T, one can obtain asymptotical results as T tends to infinity. For this quite different topic we refer to the literature on *reduced branching processes*, for example the seminal work of Fleischmann and Prehn in [FlePre74], or [FleiSieg77].

"On the eve of extinction"

Finally, a last conditional limit distributions of a different kind can be found in a recent paper from Jagers, Klebaner and Sagitov ([Jag07]), which enables to study a process right before extinction. If one considers a single-type subcritical BGWc process X_t^x starting from x individuals and satisfying the **(XlogX)** condition, and denotes by $T_x := \inf\{t \geqslant 0; X_t^x = 0\}$ the time to extinction, then, as $x \to \infty$,

$$X^x_{T_x - u} \longrightarrow Y_u, \tag{2.1.20}$$

in distribution for fixed $u > 0$, and finite-dimensionally. The limit process $(Y_u)_{u\geqslant 0}$, displaying the properties of extinct populations on the eve of their disappearance, is Markovian and precisely described in [Jag07].

As announced, we will from now on focus on the Yaglom-type limits (**Section 2.3**) and Q-processes (**Section 2.4**) associated with multitype BGWc and Feller diffusion processes, after showing in **Section 2.2** how to extend the study to supercritical processes as well.

2.2 Multitype branching processes forced to extinction

As mentioned in the introduction of this chapter, we essentially consider in this work populations which are doomed to become extinct. This is the case for (sub)critical populations, but it also makes sense to work with supercritical populations *conditioned on extinction.* It has indeed been proved by Jagers and Lagerås ([JagLag08]) that general multitype discrete-state branching processes conditioned on extinction remain branching processes, and more specifically that supercritical general branching processes conditioned on extinction are subcritical. It thus means that, contrary to what might be intuitively believed, supercritical populations conditioned on extinction do not first grow exponentially at rate $e^{\rho t}$, $\rho > 0$, and then drop drastically; they instead stabilize as long-lasting subcritical populations. This means moreover that conditioning on extinction in the distant future influences the life careers of the individuals, and more precisely modifies the offspring distribution, but preserves the branching property.

In **Subsection 2.2.1** we consider supercritical multitype BGWc processes conditioned on extinction and provide their explicit parameters as subcritical branching processes, and do the same for Feller diffusion processes in **Subsection 2.2.2**.

2.2.1 BGWc process forced to extinction

As in Section 1.1, let $\mathbb{P}$ be the law of a multitype BGWc process with branching rates $(\alpha_1, \dots, \alpha_d)$ and offspring generating function $\boldsymbol{f}$. We introduce the law $\widetilde{\mathbb{P}}$ of the process *conditioned on extinction,*

$$\widetilde{\mathbb{P}}\,(\,.\,) := \mathbb{P}\left(\,.\mid \lim_{s\to\infty} \mathbf{X}_s = \mathbf{0}\right). \tag{2.2.1}$$

This definition makes sense if for any $\boldsymbol{x} \in \mathbb{N}^d$, $\mathbb{P}_{\boldsymbol{x}}\left(\lim_{s\to\infty} \mathbf{X}_s = \mathbf{0}\right) = \mathbf{q}^{\boldsymbol{x}} > 0$. We thus work under the assumption

(B2) The BGWc process has a *positive risk of extinction* $\mathbf{q} > \mathbf{0}$.

Then $\widetilde{\mathbb{P}}$ is a well-defined probability measure on $(\Omega, (\mathbf{X}_t)_{t\geqslant 0}, (\mathcal{F}_t)_{t\geqslant 0})$. If the branching process with law $\mathbb{P}$ is supercritical, assumption **(B2)** excludes the degenerate case for which the process explodes almost surely (see Proposition 1.1.15). If the process is critical, assumption **(B2)** avoids the trivial case of a *simple* process (see Definition 1.1.4). Note that under **(B2)**, a (sub)critical process almost surely dies out (Proposition 1.1.19), hence conditioning on extinction in (2.2.1) does not change the measure and we have $\widetilde{\mathbb{P}} = \mathbb{P}$.

We already know from [JagLag08] that the conditioned process with law $\widetilde{\mathbb{P}}$ is a subcritical branching process if $\rho > 0$ (and obviously if $\rho < 0$). We prove again this result in our specific case, and obtain the explicit parameters of the process with law $\widetilde{\mathbb{P}}$.

Proposition 2.2.1. *Let us assume **(B1)** and **(B2)**. Then $\widetilde{\mathbb{P}}$ is a Doob h-transform of $\mathbb{P}$ satisfying for all $t \geqslant 0$ and all $\boldsymbol{x} \in \mathbb{N}^d$,*

$$d\widetilde{\mathbb{P}}_{\boldsymbol{x}}|_{\mathcal{F}_t} = \frac{\boldsymbol{q}^{\mathbf{X}_t}}{\boldsymbol{q}^{\boldsymbol{x}}} d\mathbb{P}_{\boldsymbol{x}}|_{\mathcal{F}_t}. \tag{2.2.2}$$

Moreover, if $\rho \neq 0$, $\widetilde{\mathbb{P}}$ is the law of an irreducible subcritical branching process with branching rates $(\alpha_1, \ldots, \alpha_d)$ and offspring generating function $\widetilde{\boldsymbol{f}}$ defined by

$$\widetilde{f}_i(\boldsymbol{r}) := \frac{1}{q_i} f_i(\boldsymbol{q}\boldsymbol{r}). \tag{2.2.3}$$

Proof. Let $t \geqslant 0$, $B \in \mathcal{F}_t$ and $\boldsymbol{x} \in \mathbb{N}^d$. By definition,

$$\widetilde{\mathbb{E}}_{\boldsymbol{x}}\left(\mathbf{1}_B\right) = \frac{\mathbb{E}_{\boldsymbol{x}}\left[\mathbf{1}_B \mathbb{P}_{\boldsymbol{x}}\left(\lim_{s\to\infty} \mathbf{X}_s = \mathbf{0} \,|\, \mathcal{F}_t\right)\right]}{\mathbb{P}_{\boldsymbol{x}}\left(\lim_{s\to\infty} \mathbf{X}_s = \mathbf{0}\right)}.$$

Using (1.1.20) together with the Markov property, we obtain

$$\widetilde{\mathbb{E}}_{\boldsymbol{x}}\left(\mathbf{1}_B\right) = \mathbb{E}_{\boldsymbol{x}}\left[\mathbf{1}_B \frac{\mathbf{q}^{\mathbf{X}_t}}{\mathbf{q}^{\boldsymbol{x}}}\right].$$

It ensues (2.2.2), and that $\left(\mathbf{q}^{\mathbf{X}_t}\right)_{t\geqslant 0}$ is a $(\mathbb{P}, \mathcal{F}_t)$-martingale. Defining for all $\boldsymbol{x} \in \mathbb{N}^d$, $\widetilde{h}(\boldsymbol{x}) := \mathbf{q}^{\boldsymbol{x}}$, the infinitesimal generator $\widetilde{L}$ of the conditioned process with law $\widetilde{\mathbb{P}}$ is then given for all smooth function $f : \mathbb{N}^d \to \mathbb{R}$ by

$$\widetilde{L}f := \frac{1}{\widetilde{h}} L(\widetilde{h}f).$$

Hence, for all $\boldsymbol{x} \in \mathbb{N}^d$,

$$\left(\widetilde{L}f\right)(\boldsymbol{x}) = \sum_{i=1}^{d} \alpha_i x_i \sum_{\mathbf{k}\in\mathbb{N}^d} p_i(\mathbf{k}) \left[\frac{1}{q_i}\mathbf{q}^{\mathbf{k}} f\left(\boldsymbol{x} + \mathbf{k} - \mathbf{e}_i\right) - f\left(\boldsymbol{x}\right)\right].$$

As a fixed point of the generating function $\boldsymbol{f}$ (Proposition 1.1.17), the extinction probability vector $\mathbf{q}$ satisfies for all $i = 1\ldots d$, $\sum_{\mathbf{k}\in\mathbb{N}^d} p_i(\mathbf{k})\mathbf{q}^{\mathbf{k}} = q_i$, and it follows

$$\left(\widetilde{L}f\right)(\boldsymbol{x}) = \sum_{i=1}^{d} \alpha_i x_i \sum_{\mathbf{k}\in\mathbb{N}^d} \widetilde{p}_i(\mathbf{k}) \left[f\left(\boldsymbol{x} + \mathbf{k} - \mathbf{e}_i\right) - f\left(\boldsymbol{x}\right)\right], \tag{2.2.4}$$

where

$$\widetilde{p}_i(\mathbf{k}) := \frac{1}{q_i}\mathbf{q}^{\mathbf{k}} p_i(\mathbf{k}), \quad i = 1\ldots d, \ \mathbf{k} \in \mathbb{N}^d. \tag{2.2.5}$$

$\widetilde{\mathbb{P}}$ is thus the law of a branching process with branching rates $(\alpha_1, \ldots, \alpha_d)$ and offspring generating function $\widetilde{\boldsymbol{f}}$ given by (2.2.3).

It remains to show that this process is irreducible and subcritical. The irreducibility of the mean matrix (denoted $\widetilde{\mathbf{M}}$) comes from the fact that

$$\widetilde{m}_{ij} = \frac{\partial \widetilde{f}_i}{\partial r_j}(\mathbf{1}) = \frac{q_j}{q_i}\frac{\partial f_i}{\partial r_j}(\mathbf{q}). \tag{2.2.6}$$

Hence if $\widetilde{m}_{ij} = 0$, then $\frac{\partial f_i}{\partial r_j}(\mathbf{q}) = 0$, which can be the case if and only if each coefficient in $\frac{\partial f_i}{\partial r_j}(\mathbf{r})$ is null, implying *a fortiori* that $m_{ij} = \frac{\partial f_i}{\partial r_j}(\mathbf{1}) = 0$. Conversely, if $m_{ij} = 0$ then for the same reason we have $\widetilde{m}_{ij} = 0$. The positive entries of $\widetilde{\mathbf{M}}$ thus coincide with those of $\mathbf{M}$, and $\widetilde{\mathbf{M}}$ is irreducible as well. Moreover, for all $i, j = 1\ldots d$,

$$\widetilde{\mathbb{E}}_{\mathbf{e}_i}\left(X_{t,j}\right) = \frac{1}{q_i}\mathbb{E}_{\mathbf{e}_i}\left(\mathbf{q}^{\mathbf{X}_t} X_{t,j}\right),$$

hence in the supercritical case $\rho > 0$, for which $\mathbf{q} < \mathbf{1}$, we obtain by dominated convergence that $\lim_{t\to\infty} \widetilde{\mathbb{E}}_{\mathbf{e}_i}\left(X_{t,j}\right) = 0$, which proves that the irreducible process with law $\widetilde{\mathbb{P}}$ is subcritical. $\square$

Remark 2.2.2. Due to the equivalence between extinction of the continuous-time process and extinction of its embedded generation counting process, we obtain unsurprisingly the same offspring generating function $\widetilde{\boldsymbol{f}}$ as the one computed in [DalJof08] for supercritical BGW processes conditioned on extinction.

Definition 2.2.3. We denote by $\widetilde{\mathbf{q}} := \lim_{t\to\infty} \widetilde{\mathbf{q}}(t)$ the extinction probability vector of the process with law $\widetilde{\mathbb{P}}$, where

$$\widetilde{q}_i(t) := \widetilde{\mathbb{P}}_{\mathbf{e}_i}(\mathbf{X}_t = \mathbf{0}), \quad i = 1 \dots d. \tag{2.2.7}$$

Moreover, we denote by $\widetilde{\rho}$, $\widetilde{\boldsymbol{\xi}}$ and $\widetilde{\boldsymbol{\eta}}$ the Perron's root and associated right and left eigenvectors of the matrix $\widetilde{\mathbf{C}} := \mathbf{A}(\widetilde{\mathbf{M}} - \mathbf{I})$, with the usual normalization convention.

2.2.2 Feller diffusion process forced to extinction

As in Section 1.2, let $\mathbb{P}$ be the law of a multitype Feller diffusion process with mutation matrix $\mathbf{C}$ and variance parameters $(\sigma_1^2, \dots, \sigma_d^2)$. We introduce the law $\widetilde{\mathbb{P}}$ of the process conditioned on *extinction*,

$$\widetilde{\mathbb{P}}\,(\,.\,) := \mathbb{P}\left(\,.\mid \lim_{s\to\infty} \mathbf{X}_s = \mathbf{0}\right). \tag{2.2.8}$$

Again, this definition makes sense if for any $\boldsymbol{x} \in \mathbb{N}^d$, $\mathbb{P}_{\boldsymbol{x}}\left(\lim_{s\to\infty} \mathbf{X}_s = \mathbf{0}\right) = e^{-\boldsymbol{x}\cdot\mathbf{u}} > 0$. We thus introduce the following assumption.

(F2) The Feller diffusion process has a *positive risk of extinction*, *i.e.* $\mathbf{u} < \infty$.

We next prove that the conditioned Feller diffusion process with law $\widetilde{\mathbb{P}}$ is a subcritical diffusion process, if $\rho > 0$, while we obviously have $\widetilde{\mathbb{P}} = \mathbb{P}$ if $\rho \leqslant 0$.

Proposition 2.2.4. *Let us assume **(F1)** and **(F2)**. Then $\widetilde{\mathbb{P}}$ is a Doob h-transform of $\mathbb{P}$ satisfying for all $t \geqslant 0$ and all $\boldsymbol{x} \in \mathbb{R}_+^d$,*

$$d\widetilde{\mathbb{P}}_{\boldsymbol{x}}|_{\mathcal{F}_t} = \frac{e^{-\mathbf{X}_t\cdot\boldsymbol{u}}}{e^{-\boldsymbol{x}\cdot\boldsymbol{u}}} d\mathbb{P}_{\boldsymbol{x}}|_{\mathcal{F}_t}. \tag{2.2.9}$$

Moreover, if $\rho \neq 0$, $\widetilde{\mathbb{P}}$ is the law of a Feller diffusion process with variance parameters $\sigma_1^2, \dots, \sigma_d^2$ and subcritical irreducible mutation matrix $\widetilde{\boldsymbol{C}}$ with entries

$$\widetilde{c}_{ij} := c_{ij} - \sigma_i^2 u_i \delta_{ij}. \tag{2.2.10}$$

Proof. Let $t \geqslant 0$, $B \in \mathcal{F}_t$ and $\boldsymbol{x} \in \mathbb{R}_+^d$. The branching and Markov properties imply that

$$\widetilde{\mathbb{E}}_{\boldsymbol{x}}\left(\mathbf{1}_B\right) = \frac{\mathbb{E}_{\boldsymbol{x}}\left[\mathbf{1}_B \mathbb{P}_{\boldsymbol{x}}\left(\lim_{s\to\infty} \mathbf{X}_s = \mathbf{0}\,|\,\mathcal{F}_t\right)\right]}{\mathbb{P}_{\boldsymbol{x}}\left(\lim_{s\to\infty} \mathbf{X}_s = \mathbf{0}\right)} = \mathbb{E}_{\boldsymbol{x}}\left[\mathbf{1}_B \frac{e^{-\mathbf{X}_t\cdot\mathbf{u}}}{e^{-\boldsymbol{x}\cdot\mathbf{u}}}\right].$$

It ensues (2.2.9), and that $\left(e^{-\mathbf{X}_t\cdot\mathbf{u}}\right)_{t\geqslant 0}$ is a $(\mathbb{P}, \mathcal{F}_t)$-martingale. Defining for all $\boldsymbol{x} \in \mathbb{N}^d$, $\widetilde{h}(\boldsymbol{x}) := e^{-\boldsymbol{x}\cdot\mathbf{u}}$, the infinitesimal generator $\widetilde{L}$ of the conditioned process with law $\widetilde{\mathbb{P}}$ is then given for all smooth function $f : \mathbb{R}_+^d \to \mathbb{R}$ by $\widetilde{L}f := \frac{1}{\widetilde{h}} L(\widetilde{h}f)$. Hence, for all $\boldsymbol{x} \in \mathbb{R}_+^d$,

$$\begin{aligned}
\left(\widetilde{L}f\right)(\boldsymbol{x}) &= \frac{1}{2}\sum_{i=1}^d \sigma_i^2 x_i \left[u_i^2 f(\boldsymbol{x}) - 2u_i \frac{\partial f}{\partial x_i}(\boldsymbol{x}) + \frac{\partial^2 f}{\partial x_i^2}(\boldsymbol{x})\right] + \sum_{i=1}^d x_i \sum_{j=1}^d c_{ij}\left[-u_j f(\boldsymbol{x}) + \frac{\partial f}{\partial x_j}(\boldsymbol{x})\right] \\
&= f(\boldsymbol{x})\Big[\frac{1}{2}\sum_{i=1}^d \sigma_i^2 x_i u_i^2 - \sum_{i=1}^d x_i \sum_{j=1}^d c_{ij} u_j\Big] + \frac{1}{2}\sum_{i=1}^d \sigma_i^2 x_i \frac{\partial^2 f}{\partial x_i^2}(\boldsymbol{x}) + \sum_{i=1}^d x_i \sum_{j=1}^d \widetilde{c}_{ij} \frac{\partial f}{\partial x_j}(\boldsymbol{x}) \\
&= f(\boldsymbol{x}) e^{\boldsymbol{x}\cdot\mathbf{u}} \left(L\widetilde{h}\right)(\boldsymbol{x}) + \frac{1}{2}\sum_{i=1}^d \sigma_i^2 x_i \frac{\partial^2 f}{\partial x_i^2}(\boldsymbol{x}) + \sum_{i=1}^d x_i \sum_{j=1}^d \widetilde{c}_{ij} \frac{\partial f}{\partial x_j}(\boldsymbol{x}) \\
&= \frac{1}{2}\sum_{i=1}^d \sigma_i^2 x_i \frac{\partial^2 f}{\partial x_i^2}(\boldsymbol{x}) + \sum_{i=1}^d x_i \sum_{j=1}^d \widetilde{c}_{ij} \frac{\partial f}{\partial x_j}(\boldsymbol{x}).
\end{aligned} \tag{2.2.11}$$

We have indeed $(L\widetilde{h})(\boldsymbol{x}) = 0$ since $(\widetilde{h}(\mathbf{X}_t))_{t\geqslant 0}$ is a $(\mathbb{P}, \mathcal{F}_t)$-martingale. By (2.2.4) it appears that the process with law $\widetilde{\mathbb{P}}$ is a Feller diffusion process with a different mutation matrix $\widetilde{\mathbf{C}}$ and unchanged variance parameters. Since $\widetilde{\mathbf{C}}$ only differs from $\mathbf{C}$ on the diagonal, it is irreducible as well. Moreover, by definition,

$$\widetilde{\mathbb{E}}_{\boldsymbol{x}}\left(X_{t,j}\right) = \frac{1}{e^{-\boldsymbol{x}\cdot\mathbf{u}}}\mathbb{E}_{\boldsymbol{x}}\left(e^{-\mathbf{X}_t\cdot\mathbf{u}}X_{t,j}\right),$$

hence in the supercritical case $\rho > 0$, for which $\mathbf{u} > \mathbf{0}$, we obtain by dominated convergence that $\lim_{t\to\infty}\widetilde{\mathbb{E}}_{\boldsymbol{x}}\left(X_{t,j}\right) = 0$, which proves that the irreducible process with law $\widetilde{\mathbb{P}}$ is subcritical. □

Definition 2.2.5. We denote by $\widetilde{\rho}$, $\widetilde{\boldsymbol{\xi}}$ and $\widetilde{\boldsymbol{\eta}}$ the Perron's root and associated right and left eigenvectors of the mutation matrix $\widetilde{\mathbf{C}}$, with the usual normalization convention.

Moreover, we denote by $\widetilde{\mathbf{u}}_t(\boldsymbol{\lambda})$ the cumulant at time t of the process with law $\widetilde{\mathbb{P}}$, and by $\widetilde{\mathbf{u}}_t$ its limit as $\boldsymbol{\lambda} \to \boldsymbol{\infty}$.

2.3 Yaglom-type limits

In the following **Subsection 2.3.1** (resp. **Subsection 2.3.2**), we consider the conditional limit distribution

$$\lim_{t\to\infty}\widetilde{\mathbb{P}}\left(\mathbf{X}_t \in\ .\,|\mathbf{X}_{t+\theta} \neq \mathbf{0}\right), \tag{2.3.1}$$

for some fixed $\theta \geqslant 0$, where $\widetilde{\mathbb{P}}$ is the law of the multitype BGWc process (resp. Feller diffusion process) conditioned on extinction, introduced in Subsection 2.2.1 (resp. Subsection 2.2.2).

2.3.1 Yaglom-type limits for the multitype BGWc process

As we have seen in the historical introduction, the asymptotic behavior of a single-type BGW process differs drastically whether the process is critical or subcritical: in the subcritical case (see (2.1.3)), the limit (2.3.1) defines a probability distribution when $\theta = 0$, while this limit is degenerate in the critical case (see (2.1.5)). Indeed, the process X_n explodes conditionally on $X_n > 0$ when $n \to \infty$, and the suitable normalization in order to obtain a probability distribution is of the form $\frac{X_n}{n}$.

In this subsection we generalize these results for multitype BGWc processes, and for any $\theta \geqslant 0$.

Noncritical case

Let us first assume that $\rho \neq 0$. By Proposition 2.2.1, the process with law $\widetilde{\mathbb{P}}$ is then a subcritical process. We know from [Sew75] (Satz 6.2.8) that for $\theta = 0$, the limit (2.3.1) defines a distribution on $\mathbb{N}^d\backslash\{\mathbf{0}\}$, the so-called Yaglom distribution. We denote by F^0 its generating function. We thus have, for all $\mathbf{r} \in [0,1]^d$ and all $\boldsymbol{x} \in \mathbb{N}^d$, $\boldsymbol{x} \neq \mathbf{0}$,

$$F^0(\mathbf{r}) = \lim_{t\to\infty}\widetilde{\mathbb{E}}_{\boldsymbol{x}}\left[\mathbf{r}^{\mathbf{X}_t}\,|\,\mathbf{X}_t \neq \mathbf{0}\right]. \tag{2.3.2}$$

Let us generalize this result to $\theta \geqslant 0$.

Proposition 2.3.1. *Let us assume **(B1)**, **(B2)**, $\rho \neq 0$. If $\rho < 0$, we assume moreover that for all $i, j = 1 \dots d$,*

$$\sum_{\boldsymbol{k}\in\mathbb{N}^d} k_j\ln(k_j)p_i(\boldsymbol{k}) < \infty. \tag{2.3.3}$$

Then for any $\boldsymbol{x} \in \mathbb{N}^d\backslash\{\boldsymbol{0}\}$ and $\theta \geqslant 0$, the limit

$$\lim_{t\to\infty}\widetilde{\mathbb{P}}_{\boldsymbol{x}}\left(\boldsymbol{X}_t \in\ .\,|\boldsymbol{X}_{t+\theta} \neq \boldsymbol{0}\right) \tag{2.3.4}$$

defines a probability distribution independent of $\boldsymbol{x}$, whose generating function $F^\theta(\boldsymbol{r})$ satisfies

$$F^\theta(\boldsymbol{r}) = e^{-\widetilde{\rho}\theta}\left[F^0(\boldsymbol{r}) - F^0\left(\boldsymbol{r}\widetilde{\mathbf{q}}(\theta)\right)\right], \qquad \boldsymbol{r} \in [0,1]^d, \tag{2.3.5}$$

where $\widetilde{\rho}$ and $\widetilde{\mathbf{q}}$ are given in Definition 2.2.3. In the following we refer to this distribution as the θ-Yaglom distribution.

Proof. By means of Theorem 2 in [JofSpit67] and its extension to the continuous-time case *via* the embedded process in Theorem 6.1 in [Ogu75], we obtain that there exists a non-negative real function γ on $[0,1]^d$ such that

$$\lim_{t\to\infty} e^{-\widetilde{\rho}t}\left(\mathbf{1} - \widetilde{\mathbf{F}}_t(\mathbf{r})\right) = \gamma(\mathbf{r})\widetilde{\boldsymbol{\xi}}, \tag{2.3.6}$$

where $\widetilde{\mathbf{F}}_t$ is the generating function at time t of the subcritical process with law $\widetilde{\mathbb{P}}$, $\widetilde{F}_{t,i}(\mathbf{r}) := \widetilde{\mathbb{E}}_{\mathbf{e}_i}(\mathbf{r}^{\mathbf{X}_t})$. In particular, $\widetilde{\mathbf{F}}_t(\mathbf{0}) = \widetilde{\mathbf{q}}(t)$, and thus

$$\lim_{t\to\infty} e^{-\widetilde{\rho}t}\left(\mathbf{1} - \widetilde{\mathbf{q}}(t)\right) = \gamma(\mathbf{0})\widetilde{\boldsymbol{\xi}}. \tag{2.3.7}$$

If $\rho > 0$, then $\mathbf{q} < \mathbf{1}$ which implies that

$$\widetilde{\mathbb{E}}_{\mathbf{e}_i}\left[X_{t,j}\ln X_{t,j}\right] = \frac{1}{q_i}\mathbb{E}_{\mathbf{e}_i}\left[\mathbf{q}^{\mathbf{X}_t}X_{t,j}\ln X_{t,j}\right] < \infty.$$

Thanks to (1.1.25), this means that if $\rho > 0$, the subcritical process with law $\widetilde{\mathbb{P}}$ satisfies the **(XlogX)** assumption. This is obviously also the case if $\rho < 0$, by assumption. Considering (2.3.6), we then immediately see as a consequence of Proposition 1.1.24 that $\gamma(\mathbf{0}) > 0$.

From the Markov and branching properties of $\widetilde{\mathbb{P}}$, for all $\theta \geqslant 0$, $\mathbf{r} \in [0,1]^d$ and $\boldsymbol{x} \in \mathbb{N}^d\backslash\{\mathbf{0}\}$,

$$\begin{aligned}
\widetilde{\mathbb{E}}_{\boldsymbol{x}}\left[\mathbf{r}^{\mathbf{X}_t} \mid \mathbf{X}_{t+\theta} \neq \mathbf{0}\right] &= \frac{\widetilde{\mathbb{E}}_{\boldsymbol{x}}\left[\mathbf{r}^{\mathbf{X}_t}\right] - \widetilde{\mathbb{E}}_{\boldsymbol{x}}\left[\mathbf{r}^{\mathbf{X}_t}\widetilde{\mathbb{P}}_{\mathbf{X}_t}\left[\mathbf{X}_\theta = \mathbf{0}\right]\right]}{1 - \widetilde{\mathbb{P}}_{\boldsymbol{x}}\left[\mathbf{X}_{t+\theta} = \mathbf{0}\right]} \\
&= \frac{\widetilde{\mathbb{E}}_{\boldsymbol{x}}\left[\mathbf{r}^{\mathbf{X}_t}\right] - \widetilde{\mathbb{E}}_{\boldsymbol{x}}\left[\mathbf{r}^{\mathbf{X}_t}\widetilde{\mathbf{F}}_\theta(\mathbf{0})^{\mathbf{X}_t}\right]}{1 - \widetilde{\mathbf{F}}_{t+\theta}(\mathbf{0})^{\boldsymbol{x}}} \\
&= \frac{\widetilde{\mathbf{F}}_t(\mathbf{r})^{\boldsymbol{x}} - \widetilde{\mathbf{F}}_t\left(\mathbf{r}\widetilde{\mathbf{q}}(\theta)\right)^{\boldsymbol{x}}}{1 - \widetilde{\mathbf{F}}_{t+\theta}\left(\mathbf{0}\right)^{\boldsymbol{x}}},
\end{aligned}$$

from which it ensues together with (2.3.6) that

$$\lim_{t\to\infty} \widetilde{\mathbb{E}}_{\boldsymbol{x}}\left[\mathbf{r}^{\mathbf{X}_t} \mid \mathbf{X}_{t+\theta} \neq \mathbf{0}\right] = \frac{\gamma\left(\mathbf{r}\widetilde{\mathbf{q}}(\theta)\right) - \gamma(\mathbf{r})}{e^{\widetilde{\rho}\theta}\gamma(\mathbf{0})} =: F^\theta(\mathbf{r}). \tag{2.3.8}$$

Now for $\theta = 0$ this relation becomes

$$F^0(\mathbf{r}) = 1 - \frac{\gamma(\mathbf{r})}{\gamma(\mathbf{0})}. \tag{2.3.9}$$

Consequently,

$$\begin{aligned}
F^\theta(\mathbf{r}) &= e^{-\widetilde{\rho}\theta}\left[\frac{\gamma\left(\mathbf{r}\widetilde{\mathbf{q}}(\theta)\right)}{\gamma(\mathbf{0})} - \frac{\gamma(\mathbf{r})}{\gamma(\mathbf{0})}\right] \\
&= e^{-\widetilde{\rho}\theta}\left[F^0(\mathbf{r}) - F^0\left(\mathbf{r}\widetilde{\mathbf{q}}(\theta)\right)\right].
\end{aligned}$$

In addition, F^0 is continuous in $\mathbf{r} = \mathbf{1}$, with $F^0(\mathbf{1}) = 1$, hence for each $\theta \geqslant 0$, F^θ is continuous in $\mathbf{1}$ as well, with

$$F^\theta(\mathbf{1}) = \frac{e^{-\widetilde{\rho}\theta}\gamma(\widetilde{\mathbf{F}}_\theta(\mathbf{0}))}{\gamma(\mathbf{0})}.$$

Moreover, thanks to the semi-group property of the generating function we have,

$$e^{-\widetilde{\rho}(t+\theta)}(\mathbf{1} - \widetilde{\mathbf{F}}_{t+\theta}(\mathbf{0})) = e^{-\widetilde{\rho}\theta}e^{-\widetilde{\rho}t}(\mathbf{1} - \mathbf{F}_t(\mathbf{F}_\theta(\mathbf{0}))),$$

which by (2.3.6) becomes as t tends to infinity

$$\gamma(\mathbf{0})\widetilde{\boldsymbol{\xi}} = e^{-\widetilde{\rho}\theta}\gamma\left(\mathbf{F}_\theta(\mathbf{0})\right)\widetilde{\boldsymbol{\xi}}.$$

Consequently, $F^\theta(\mathbf{1}) = 1$, and F^θ defines a probability generating function as well. □

We thus have obtained an expression of the θ-Yaglom distribution with respect to the original Yaglom limit (see (2.3.5)). The latter is almost never explicitly known, but has the following implicit characterization.

Proposition 2.3.2 ([Sew75] Satz 6.2.8)**.** *The generating function $F^0(\boldsymbol{r})$ of the Yaglom distribution is solution of the following partial differential equation. For all $\boldsymbol{r} \in [0,1]^d$,*

$$\begin{cases} \displaystyle\sum_{i=1}^d \alpha_i\left(\widetilde{f}_i(\boldsymbol{r}) - r_i\right)\frac{\partial F^0(\boldsymbol{r})}{\partial r_i} = -\widetilde{\rho}\left(1 - F^0(\boldsymbol{r})\right), \\ F^0(\boldsymbol{0}) = 0. \end{cases} \tag{2.3.10}$$

This proposition is a consequence of the following intermediate result. The generating function $F^0(\mathbf{r})$ of the Yaglom distribution satisfies the following implicit equation: for all $\mathbf{r} \in [0,1]^d$ and all $t \geqslant 0$,

$$F^0\left(\widetilde{\mathbf{F}}_t\left(\mathbf{r}\right)\right) = 1 - e^{\widetilde{\rho}t}\left(1 - F^0(\mathbf{r})\right). \tag{2.3.11}$$

Critical case

We now consider the critical case $\rho = 0$. We recall that in this case we have $\mathbb{P} = \widetilde{\mathbb{P}}$. The equivalent of the Kolmogorov-Yaglom exponential law (2.1.5) for multitype BGWc processes is given by the following proposition.

Proposition 2.3.3 ([Sew75] Satz 6.3.5)**.** *Let us assume **(B1)**, **(B2)**, $\rho = 0$ and that all the second-order moments of the offspring distribution are finite. Then, for all $\boldsymbol{x} \in \mathbb{N}^d\backslash\{\boldsymbol{0}\}$, and all $\boldsymbol{u} \geqslant \boldsymbol{0}$,*

$$\lim_{t\to\infty}\mathbb{P}_{\boldsymbol{x}}\left[\frac{2X_{t,1}}{\zeta\eta_1 t} > u_1, \ldots, \frac{2X_{t,d}}{\zeta\eta_d t} > u_d \,|\boldsymbol{X}_t \neq \boldsymbol{0}\right] = e^{-\max_i u_i}.$$

In other words, the random vector $\left(\frac{2X_{t,1}}{\zeta\eta_1 t}, \ldots, \frac{2X_{t,d}}{\zeta\eta_d t}\right)$ converges conditionally on $\mathbf{X}_t \neq \mathbf{0}$ to a random vector $\mathbf{Y}$ independent of $\boldsymbol{x}$, with coordinates $Y_1 = \ldots = Y_d$ almost surely, and such that each Y_i is exponentially distributed with parameter 1.

We provide in the following proposition a generalization of this result to the case $\theta \geqslant 0$. We show that the limit law for $t \to \infty$ of the random vector $(\frac{2X_{t,1}}{\zeta\eta_1 t}, \ldots, \frac{2X_{t,d}}{\zeta\eta_d t})$ conditionally on $\mathbf{X}_{t+\theta} \neq \mathbf{0}$ is the same for every $\theta \geqslant 0$. The limiting vector $\mathbf{Y}$ thus depends neither on $\boldsymbol{x}$ nor on θ, which comes intuitively from the fact that rescaling the process by t or by $t+\theta$ does not make a difference any more once t tends to ∞.

Proposition 2.3.4. *Let us assume **(B1)**, **(B2)**, $\rho = 0$ and that all the second-order moments of the offspring distribution are finite. Then, for all $\theta \geqslant 0$, all $\boldsymbol{x} \in \mathbb{N}^d\backslash\{\boldsymbol{0}\}$, and all $\boldsymbol{u} \geqslant \boldsymbol{0}$,*

$$\lim_{t\to\infty}\mathbb{P}_{\boldsymbol{x}}\left[\frac{2X_{t,1}}{\zeta\eta_1 t} > u_1, \ldots, \frac{2X_{t,d}}{\zeta\eta_d t} > u_d \,|\boldsymbol{X}_{t+\theta} \neq \boldsymbol{0}\right] = e^{-\max_i u_i}.$$

In other words, the random vector $\left(\frac{2X_{t,1}}{\zeta\eta_1 t}, \ldots, \frac{2X_{t,d}}{\zeta\eta_d t}\right)$ converges conditionally on $\mathbf{X}_{t+\theta} \neq \mathbf{0}$ to a random vector $\mathbf{Y}$ independent of $\boldsymbol{x}$ and of θ, with coordinates $Y_1 = \ldots = Y_d$ almost surely, and such that each Y_i is exponentially distributed with parameter 1.

Proof. From [Sew75], Satz 6.3.1, we know that under these assumptions, the generating function of the process satisfies the following asymptotic behavior. For all $i = 1 \ldots d$, uniformly in $\mathbf{r} \in [0,1]^d$, $\mathbf{r} \neq \mathbf{1}$, as $t \to \infty$,

$$F_{t,i}(\mathbf{r}) \sim 1 - \frac{\xi_i\, \boldsymbol{\eta} \cdot (\mathbf{1} - \mathbf{r})}{1 + \frac{\zeta t}{2}\, \boldsymbol{\eta} \cdot (\mathbf{1} - \mathbf{r})}. \tag{2.3.12}$$

For all $\boldsymbol{\lambda} \geqslant \mathbf{0}$ and all $\boldsymbol{x} \in \mathbb{N}^d \setminus \{\mathbf{0}\}$, denoting by $\mathbf{r}_{\boldsymbol{\lambda},t}$ the vector with coordinates $(e^{-\frac{2\lambda_1}{\zeta\eta_1 t}}, \ldots, e^{-\frac{2\lambda_d}{\zeta\eta_d t}})$,

$$\begin{aligned}\mathbb{E}_{\boldsymbol{x}}\left[e^{-\left(\lambda_1 \frac{2X_{t,1}}{\zeta\eta_1 t} + \ldots + \lambda_d \frac{2X_{t,d}}{\zeta\eta_d t}\right)} \mid \mathbf{X}_{t+\theta} \neq \mathbf{0}\right] &= \mathbb{E}_{\boldsymbol{x}}\left[(\mathbf{r}_{\boldsymbol{\lambda},t})^{\mathbf{X}_t} \mid \mathbf{X}_{t+\theta} \neq \mathbf{0}\right] \\ &= \frac{\mathbf{F}_t(\mathbf{r}_{\boldsymbol{\lambda},t})^{\boldsymbol{x}} - \mathbf{F}_t(\mathbf{r}_{\boldsymbol{\lambda},t}\mathbf{F}_\theta(\mathbf{0}))^{\boldsymbol{x}}}{1 - \mathbf{F}_{t+\theta}(\mathbf{0})^{\boldsymbol{x}}}. \end{aligned} \tag{2.3.13}$$

We have

$$\boldsymbol{\eta} \cdot (\mathbf{1} - \mathbf{r}_{\boldsymbol{\lambda},t}) = \sum_{i=1}^{d} \eta_i (1 - e^{-\frac{2\lambda_i}{\zeta\eta_i t}}) \sim_{t\to\infty} \frac{2}{\zeta t} \boldsymbol{\lambda} \cdot \mathbf{1},$$

hence, by (2.3.12),

$$\mathbf{F}_t(\mathbf{r}_{\boldsymbol{\lambda},t})^{\boldsymbol{x}} \sim_{t\to\infty} 1 - \frac{2\boldsymbol{x}\cdot\boldsymbol{\xi}}{\zeta t} \frac{\boldsymbol{\lambda}\cdot\mathbf{1}}{1 + \boldsymbol{\lambda}\cdot\mathbf{1}}. \tag{2.3.14}$$

Similarly,

$$\boldsymbol{\eta} \cdot (\mathbf{1} - \mathbf{r}_{\boldsymbol{\lambda},t}\mathbf{F}_\theta(\mathbf{0})) \sim_{t\to\infty} \boldsymbol{\eta} \cdot (\mathbf{1} - \mathbf{F}_\theta(\mathbf{0})) + \frac{2}{\zeta t} \boldsymbol{\lambda} \cdot \mathbf{F}_\theta(\mathbf{0}),$$

and thus

$$\mathbf{F}_t\left(\mathbf{r}_{\boldsymbol{\lambda},t}\mathbf{F}_\theta\left(\mathbf{0}\right)\right)^{\boldsymbol{x}} \sim_{t\to\infty} 1 - \frac{2\boldsymbol{x}\cdot\boldsymbol{\xi}}{\zeta t} \frac{\boldsymbol{\lambda}\cdot\mathbf{F}_\theta(\mathbf{0}) + \frac{\zeta t}{2}\boldsymbol{\eta}\cdot(\mathbf{1} - \mathbf{F}_\theta(\mathbf{0}))}{1 + \boldsymbol{\lambda}\cdot\mathbf{F}_\theta(\mathbf{0}) + \frac{\zeta t}{2}\boldsymbol{\eta}\cdot(\mathbf{1} - \mathbf{F}_\theta(\mathbf{0}))}. \tag{2.3.15}$$

Finally, by Proposition 1.1.25,

$$\mathbf{F}_{t+\theta}(\mathbf{0})^{\boldsymbol{x}} \sim_{t\to\infty} 1 - \frac{2\boldsymbol{x}\cdot\boldsymbol{\xi}}{\zeta(t+\theta)}, \tag{2.3.16}$$

which, together with (2.3.14) and (2.3.15) used in (2.3.13) leads to

$$\begin{aligned} &\lim_{t\to\infty} \mathbb{E}_{\boldsymbol{x}}\left[e^{-\left(\lambda_1 \frac{2X_{t,1}}{\zeta\eta_1 t} + \ldots + \lambda_d \frac{2X_{t,d}}{\zeta\eta_d t}\right)} \mid \mathbf{X}_{t+\theta} \neq \mathbf{0}\right] \\ &\quad = \lim_{t\to\infty} \frac{t+\theta}{t}\left[\frac{\boldsymbol{\lambda}\cdot\mathbf{F}_\theta(\mathbf{0}) + \frac{\zeta t}{2}\boldsymbol{\eta}\cdot(\mathbf{1} - \mathbf{F}_\theta(\mathbf{0}))}{1 + \boldsymbol{\lambda}\cdot\mathbf{F}_\theta(\mathbf{0}) + \frac{\zeta t}{2}\boldsymbol{\eta}\cdot(\mathbf{1} - \mathbf{F}_\theta(\mathbf{0}))} - \frac{\boldsymbol{\lambda}\cdot\mathbf{1}}{1 + \boldsymbol{\lambda}\cdot\mathbf{1}}\right] \\ &\quad = \frac{1}{1 + \boldsymbol{\lambda}\cdot\mathbf{1}}, \end{aligned} \tag{2.3.17}$$

which is the Laplace transform of a random vector with almost surely equal coordinates, each of them exponentially distributed with parameter 1. □

2.3.2 Yaglom-type limits for the multitype Feller diffusion process

Again, we must here separate the noncritical case from the critical case, for which limit (2.3.1) is degenerate.

Noncritical case

Proposition 2.3.5. *Let us assume **(F1)**, **(F2)** and $\rho \neq 0$. Then for any $\boldsymbol{x} \in \mathbb{N}^d \backslash \{\boldsymbol{0}\}$ and $\theta \geqslant 0$, the limit*

$$\lim_{t\to\infty} \widetilde{\mathbb{P}}_{\boldsymbol{x}} \left(\boldsymbol{X}_t \in \, . \, | \boldsymbol{X}_{t+\theta} \neq \boldsymbol{0} \right), \tag{2.3.18}$$

defines a probability distribution independent of $\boldsymbol{x}$, whose Laplace transform $\Phi^\theta(\boldsymbol{\lambda})$ satisfies

$$\Phi^\theta(\boldsymbol{\lambda}) = e^{-\widetilde{\rho}\theta} \left[\Phi^0(\boldsymbol{\lambda}) - \Phi^0\left(\boldsymbol{\lambda} + \widetilde{\boldsymbol{u}}_\theta\right) \right], \quad \boldsymbol{\lambda} \in \mathbb{R}_+^d, \tag{2.3.19}$$

where $\widetilde{\rho}$ and $\widetilde{\boldsymbol{u}}_\theta$ are given in Definition 2.2.5. We refer to this distribution as the θ-Yaglom distribution.

Proof. From the Markov and branching properties of $\widetilde{\mathbb{P}}$, for all $\theta \geqslant 0$, $\boldsymbol{\lambda} \in \mathbb{R}_+^d$ and $\boldsymbol{x} \in \mathbb{R}_+^d \backslash \{\mathbf{0}\}$,

$$\begin{aligned} \widetilde{\mathbb{E}}_{\boldsymbol{x}} \left[e^{-\boldsymbol{\lambda}\cdot\mathbf{X}_t} \,|\, \mathbf{X}_{t+\theta} \neq \mathbf{0} \right] &= \frac{\widetilde{\mathbb{E}}_{\boldsymbol{x}} \left[e^{-\boldsymbol{\lambda}\cdot\mathbf{X}_t} \right] - \widetilde{\mathbb{E}}_{\boldsymbol{x}} \left[e^{-\boldsymbol{\lambda}\cdot\mathbf{X}_t} e^{-\widetilde{\mathbf{u}}_\theta\cdot\mathbf{X}_t} \right]}{1 - \widetilde{\mathbf{F}}_{t+\theta}(\mathbf{0})^{\boldsymbol{x}}} \\ &= \frac{e^{-\boldsymbol{x}\cdot\widetilde{\mathbf{u}}_t(\boldsymbol{\lambda})} - e^{-\boldsymbol{x}\cdot\widetilde{\mathbf{u}}_t(\boldsymbol{\lambda}+\widetilde{\mathbf{u}}_\theta)}}{1 - e^{-\boldsymbol{x}\cdot\widetilde{\mathbf{u}}_{t+\theta}}}. \end{aligned}$$

Moreover, from the proof of Theorem 3.7 in [ChaRoe08], we know that there exists a positive function $\kappa(\boldsymbol{\lambda})$ with $\kappa := \lim_{\boldsymbol{\lambda}\to\boldsymbol{\infty}} \kappa(\boldsymbol{\lambda}) > 0$ such that

$$\lim_{t\to\infty} e^{-\widetilde{\rho}t} \widetilde{\mathbf{u}}_t(\boldsymbol{\lambda}) = \kappa(\boldsymbol{\lambda}) \widetilde{\boldsymbol{\xi}}, \tag{2.3.20}$$

$$\lim_{t\to\infty} e^{-\widetilde{\rho}t} \widetilde{\mathbf{u}}_t = \kappa \widetilde{\boldsymbol{\xi}}. \tag{2.3.21}$$

This means that both vectors $e^{-\widetilde{\rho}t}\widetilde{\mathbf{u}}_t(\boldsymbol{\lambda})$ and $e^{-\widetilde{\rho}t}\widetilde{\mathbf{u}}_t$ converge as $t \to \infty$ to a positive limit which is proportional to the eigenvector $\widetilde{\boldsymbol{\xi}}$. It ensues that

$$\lim_{t\to\infty} \widetilde{\mathbb{E}}_{\boldsymbol{x}} \left[e^{-\boldsymbol{\lambda}\cdot\mathbf{X}_t} \,|\, \mathbf{X}_{t+\theta} \neq \mathbf{0} \right] = \frac{\kappa\left(\boldsymbol{\lambda} + \widetilde{\mathbf{u}}_\theta\right) - \kappa\left(\boldsymbol{\lambda}\right)}{e^{\widetilde{\rho}\theta}\kappa} =: \Phi^\theta(\boldsymbol{\lambda}). \tag{2.3.22}$$

Now for $\theta = 0$ this relation becomes

$$\lim_{t\to\infty} \widetilde{\mathbb{E}}_{\boldsymbol{x}} \left[e^{-\boldsymbol{\lambda}\cdot\mathbf{X}_t} \,|\, \mathbf{X}_t \neq \mathbf{0} \right] = \Phi^0(\boldsymbol{\lambda}) = 1 - \frac{\kappa(\boldsymbol{\lambda})}{\kappa}, \tag{2.3.23}$$

and thus

$$\begin{aligned} \Phi^\theta(\boldsymbol{\lambda}) &= e^{-\widetilde{\rho}\theta} \left[\frac{\kappa\left(\boldsymbol{\lambda} + \widetilde{\mathbf{u}}_\theta\right)}{\kappa} - \frac{\kappa\left(\boldsymbol{\lambda}\right)}{\kappa} \right] \\ &= e^{-\widetilde{\rho}\theta} \left[\Phi^0(\boldsymbol{\lambda}) - \Phi^0\left(\boldsymbol{\lambda} + \widetilde{\mathbf{u}}_\theta\right) \right]. \end{aligned}$$

Moreover, $\Phi^0(\boldsymbol{\lambda})$ is continuous in $\mathbf{0}$, with $\Phi^0(\mathbf{0}) = 1$, hence for each $\theta \geqslant 0$, Φ^θ is continuous in $\mathbf{0}$ as well, with

$$\Phi^\theta(\mathbf{0}) = e^{-\widetilde{\rho}\theta} \frac{\kappa\left(\widetilde{\mathbf{u}}_\theta\right)}{\kappa}.$$

Thanks to the semi-group property of the cumulant,

$$e^{-\widetilde{\rho}(t+\theta)} \widetilde{\mathbf{u}}_{t+\theta} = e^{-\widetilde{\rho}\theta} e^{-\widetilde{\rho}t} \widetilde{\mathbf{u}}_t(\widetilde{\mathbf{u}}_\theta),$$

hence by (2.3.20) and (2.3.21) one obtains as $t \to \infty$ that

$$\kappa \widetilde{\boldsymbol{\xi}} = e^{-\widetilde{\rho}\theta} \kappa(\widetilde{\mathbf{u}}_\theta) \widetilde{\boldsymbol{\xi}},$$

and thus $\Phi^\theta(\mathbf{0}) = 1$. This ensures that Φ^θ is the Laplace transform of a probability distribution. $\square$

Remark 2.3.6. One can find in the monotype case for a general CB process a similar result in [Li00], Theorem 4.3.

We provide in the following proposition a characterization of the Laplace transform $\Phi^0(\boldsymbol{\lambda})$ as solution of a partial differential equation.

Proposition 2.3.7. *The Laplace transform $\Phi^0(\boldsymbol{\lambda})$ of the Yaglom distribution satisfies the following partial differential equation. For all $\boldsymbol{\lambda} \in \mathbb{R}_+^d$,*

$$\begin{cases} \displaystyle\sum_{i=1}^{d}\sum_{j=1}^{d}\left(\widetilde{c}_{ij}\lambda_j - \frac{1}{2}\sigma_i^2\lambda_i^2\right)\frac{\partial\Phi^0(\boldsymbol{\lambda})}{\partial\lambda_i} = -\widetilde{\rho}\left(1-\Phi^0(\boldsymbol{\lambda})\right), \\ \Phi^0(\boldsymbol{0}) = 1. \end{cases} \tag{2.3.24}$$

Proof. The semi-group property of the cumulant $\widetilde{\mathbf{u}}_\theta(\boldsymbol{\lambda})$ implies that for all $s,t \geqslant 0$ and all $\boldsymbol{\lambda} \in \mathbb{R}_+^d$,

$$e^{-\widetilde{\rho}(s+t)}\widetilde{\mathbf{u}}_{s+t}(\boldsymbol{\lambda}) = e^{-\widetilde{\rho}t}e^{-\widetilde{\rho}s}\widetilde{\mathbf{u}}_s(\widetilde{\mathbf{u}}_\theta(\boldsymbol{\lambda})) \tag{2.3.25}$$

Letting s tends to infinity one deduces from (2.3.20)-(2.3.21) that

$$\kappa(\boldsymbol{\lambda}) = e^{-\widetilde{\rho}t}\kappa(\widetilde{\mathbf{u}}_t(\boldsymbol{\lambda})).$$

We thus have by (2.3.23),

$$\Phi^0(\widetilde{\mathbf{u}}_t(\boldsymbol{\lambda})) = 1 - \frac{\kappa(\widetilde{\mathbf{u}}_t(\boldsymbol{\lambda}))}{\kappa} = 1 - e^{\widetilde{\rho}t}\frac{\kappa(\boldsymbol{\lambda})}{\kappa}.$$

Together with (2.3.23) this yields the following implicit characterization of the Laplace transform $\Phi^0(\boldsymbol{\lambda})$:

$$\Phi^0(\widetilde{\mathbf{u}}_t(\boldsymbol{\lambda})) = 1 - e^{\widetilde{\rho}t}\left(1-\Phi^0(\boldsymbol{\lambda})\right).$$

We will now deduce from this relation the differential equation (2.3.24). In the neighborhood of $t = 0$ we have on the one hand,

$$\begin{aligned}\Phi^0(\widetilde{\mathbf{u}}_t(\boldsymbol{\lambda})) &= \Phi^0(\boldsymbol{\lambda}) + \sum_{i=1}^{d}\frac{\partial\Phi^0(\boldsymbol{\lambda})}{\partial\lambda_i}\frac{\partial\widetilde{u}_{t,i}(\boldsymbol{\lambda})}{\partial t}\Big|_{t=0}\, t + o(t) \\ &= \Phi^0(\boldsymbol{\lambda}) + \sum_{i=1}^{d}\frac{\partial\Phi^0(\boldsymbol{\lambda})}{\partial\lambda_i}\sum_{j=1}^{d}\left(\widetilde{c}_{ij}\lambda_j - \frac{1}{2}\sigma_i^2\lambda_i^2\right)t + o(t),\end{aligned}$$

and on the other hand

$$1 - e^{\widetilde{\rho}t}\left(1-\Phi^0(\boldsymbol{\lambda})\right) = 1 - (1+\widetilde{\rho}t + o(t))\left(1-\Phi^0(\boldsymbol{\lambda})\right).$$

We then obtain the result by letting t tend to 0. □

Remark 2.3.8. In the monotype case $d = 1$, (2.3.24) becomes

$$\begin{cases} \left(\widetilde{\rho}\lambda - \frac{1}{2}\sigma^2\lambda^2\right)\dfrac{d\Phi^0(\lambda)}{d\lambda} = -\widetilde{\rho}\left(1-\Phi^0(\lambda)\right), \quad \lambda > 0, \\ \Phi^0(0) = 1. \end{cases} \tag{2.3.26}$$

Hence

$$\Phi^0(\lambda) = 1 - \exp\left[\int_\lambda^\infty \frac{-\widetilde{\rho}}{\widetilde{\rho}u - \frac{1}{2}\sigma^2u^2}du\right] = \frac{1}{1+\frac{\sigma^2}{2|\widetilde{\rho}|}\lambda}, \tag{2.3.27}$$

and we obtain the well-known result that the Yaglom distribution is the exponential law with parameter $\frac{2|\widetilde{\rho}|}{\sigma^2}$ (see for example Theorem 3.1 in [Lamb07]).

We deduce from (1.2.9) that for all $\theta \geqslant 0$

$$\widetilde{u}_\theta = \lim_{\lambda\to\infty} \widetilde{u}_\theta(\lambda) = \frac{2|\widetilde{\rho}|e^{\widetilde{\rho}\theta}}{\sigma^2(1-e^{\widetilde{\rho}\theta})}. \tag{2.3.28}$$

Thanks to (2.3.19) we thus obtain that for every $\theta \geqslant 0$, the generating function of the θ-Yaglom distribution is given by

$$\begin{aligned}
\Phi^\theta(\lambda) &= e^{-\widetilde{\rho}\theta}\left[\frac{1}{1+\frac{\sigma^2}{2|\widetilde{\rho}|}\lambda} - \frac{1}{1+\frac{\sigma^2}{2|\widetilde{\rho}|}\left(\lambda + \frac{2|\widetilde{\rho}|e^{\widetilde{\rho}\theta}}{\sigma^2(1-e^{\widetilde{\rho}\theta})}\right)}\right] \\
&= e^{-\widetilde{\rho}\theta}\left[\frac{2|\widetilde{\rho}|}{2|\widetilde{\rho}|+\sigma^2\lambda} - \frac{2|\widetilde{\rho}|(1-e^{\widetilde{\rho}\theta})}{2|\widetilde{\rho}|+\sigma^2(1-e^{\widetilde{\rho}\theta})\lambda}\right] \\
&= e^{-\widetilde{\rho}\theta}\left[\frac{2|\widetilde{\rho}|\left(2|\widetilde{\rho}|+\sigma^2(1-e^{\widetilde{\rho}\theta})\lambda\right)}{2|\widetilde{\rho}|+\sigma^2\lambda} - \frac{2|\widetilde{\rho}|(1-e^{\widetilde{\rho}\theta})\left(2|\widetilde{\rho}|+\sigma^2\lambda\right)}{2|\widetilde{\rho}|+\sigma^2(1-e^{\widetilde{\rho}\theta})\lambda}\right] \\
&= \frac{2|\widetilde{\rho}|}{2|\widetilde{\rho}|+\sigma^2\lambda} \times \frac{2|\widetilde{\rho}|}{2|\widetilde{\rho}|+\sigma^2(1-e^{\widetilde{\rho}\theta})\lambda} \\
&= \frac{1}{1+\frac{\sigma^2}{2|\widetilde{\rho}|}\lambda}.\frac{1}{1+\frac{\sigma^2}{2|\widetilde{\rho}|}(1-e^{\widetilde{\rho}\theta})\lambda}.
\end{aligned}$$

We thus obtain the same result as in [ChaRoe08], Proposition 3.3, which is that the θ-Yaglom distribution is the sum of two independent exponential random variables with respective parameters $\frac{2|\widetilde{\rho}|}{\sigma^2}$ and $\frac{2|\widetilde{\rho}|}{\sigma^2}(1-e^{\widetilde{\rho}\theta})$.

We point out that the monotype case is the only case where one can obtain such an explicit expression of the Yaglom or θ-Yaglom distributions.

Critical case

It is known that in the monotype case, the process, conditionally on $X_t \neq 0$, grows linearly in t ([EvPer90] Lemma 2.1). In fact the same result holds by conditioning the process on any event $X_{t+\theta} \neq 0$, for $\theta \geqslant 0$. We thus obtain that the conditioned monotype process, once normalized by t, converges as $t \to \infty$ to an exponential distribution independent of the initial state and of θ.

Proposition 2.3.9. *Assume $\rho = 0$ and $\sigma^2 > 0$. Then for all $x > 0$, $\theta \geqslant 0$ and $u \geqslant 0$,*

$$\lim_{t\to\infty} \mathbb{P}_x\left[\frac{2X_t}{\alpha\sigma^2 t} > u \,|X_{t+\theta} > 0\right] = e^{-u}.$$

In other words, the random variable $\frac{2X_t}{\alpha\sigma^2 t}$ converges conditionally on $X_{t+\theta} > 0$ to a nondegenerate limit, independent of x and of θ, which is exponentially distributed with parameter 1.

Proof. We deduce from the explicit form of the cumulant (1.2.10) that

$$u_t = \lim_{\lambda\to\infty} u_t(\lambda) = \frac{2}{\alpha\sigma^2 t}. \tag{2.3.29}$$

Then, for all $\theta \geqslant 0$ and all $\lambda \geqslant 0$,

$$\begin{aligned}
\lim_{t\to\infty} \mathbb{E}_x\left[e^{-\lambda\frac{2X_t}{\alpha\sigma^2 t}} \,|\, X_{t+\theta} > 0\right] &= \lim_{t\to\infty} \frac{e^{-xu_t(\frac{2\lambda}{\alpha\sigma^2 t})} - e^{-xu_t\left(\frac{2\lambda}{\alpha\sigma^2 t}+u_\theta\right)}}{1-e^{-xu_{t+\theta}}} \\
&= \lim_{t\to\infty} \frac{u_t\left(\frac{2\lambda}{\alpha\sigma^2 t}+\frac{2}{\alpha\sigma^2\theta}\right) - u_t(\frac{2\lambda}{\alpha\sigma^2 t})}{\frac{2}{\alpha\sigma^2(t+\theta)}} \\
&= \lim_{t\to\infty} \frac{t+\theta}{t}\left[\frac{\lambda+\frac{t}{\theta}}{1+\lambda+\frac{t}{\theta}} - \frac{\lambda}{1+\lambda}\right] \\
&= \frac{1}{1+\lambda},
\end{aligned}$$

which is the Laplace transform of an exponential distribution with parameter 1. □

Remark 2.3.10. In order to obtain a similar result in the multitype case as Proposition 2.3.9, we would need an expression of the asymptotic of the cumulant $\mathbf{u}_t(\boldsymbol{\lambda})$ as t tends to infinity. Since this is not explicitly known, we make the following conjecture. Denoting $\zeta := \sum_{i=1}^d \alpha_i \sigma_i^2 \eta_i \xi_i^2$,

$$u_{t,i}(\boldsymbol{\lambda}) \sim_{t\to\infty} \frac{\xi_i \boldsymbol{\eta}\cdot\boldsymbol{\lambda}}{1+\frac{\zeta t}{2}\boldsymbol{\eta}\cdot\boldsymbol{\lambda}}, \quad \text{and thus} \quad u_{t,i} \sim_{t\to\infty} \frac{2\xi_i}{\zeta t}. \tag{2.3.30}$$

Then, under this conjecture, we show that the random vector $\left(\frac{2X_{t,1}}{\zeta\eta_1 t}, \ldots, \frac{2X_{t,d}}{\zeta\eta_d t}\right)$ converges conditionally on $\mathbf{X}_{t+\theta} \neq \mathbf{0}$ to a random vector $\mathbf{Y}$ independent of $\boldsymbol{x}$ and of θ, with coordinates $Y_1 = \ldots = Y_d$ almost surely, and such that each Y_i is exponentially distributed with parameter 1.

Indeed, for all $\boldsymbol{\lambda} \geqslant \mathbf{0}$ and $\boldsymbol{x} \in \mathbb{R}_+^d \backslash\{\mathbf{0}\}$, denoting by $\frac{\boldsymbol{\lambda}}{\boldsymbol{\eta}}$ the vector with coordinates $(\frac{\lambda_1}{\eta_1}, \ldots, \frac{\lambda_d}{\eta_d})$, we have, assuming (2.3.30),

$$\begin{aligned}
\lim_{t\to\infty} \mathbb{E}_{\boldsymbol{x}}\left[e^{-\left(\lambda_1 \frac{2X_{t,1}}{\zeta\eta_1 t}+\ldots+\lambda_d\frac{2X_{t,d}}{\zeta\eta_d t}\right)} \,|\, \mathbf{X}_{t+\theta}\neq\mathbf{0}\right] &= \lim_{t\to\infty} \frac{e^{-\boldsymbol{x}\cdot\mathbf{u}_t(\frac{2}{\zeta t}\frac{\boldsymbol{\lambda}}{\boldsymbol{\eta}})} - e^{-\boldsymbol{x}\cdot\mathbf{u}_t\left(\frac{2}{\zeta t}\frac{\boldsymbol{\lambda}}{\boldsymbol{\eta}}+\mathbf{u}_\theta\right)}}{1-e^{-\boldsymbol{x}\cdot\mathbf{u}_{t+\theta}}} \\
&= \lim_{t\to\infty} \frac{\boldsymbol{x}\cdot\mathbf{u}_t\left(\frac{2}{\zeta t}\frac{\boldsymbol{\lambda}}{\boldsymbol{\eta}}+\frac{2\boldsymbol{\xi}}{\zeta\theta}\right) - \boldsymbol{x}\cdot\mathbf{u}_t(\frac{2}{\zeta t}\frac{\boldsymbol{\lambda}}{\boldsymbol{\eta}})}{\frac{2\boldsymbol{x}\cdot\boldsymbol{\xi}}{\zeta(t+\theta)}} \\
&= \lim_{t\to\infty} \frac{t+\theta}{t}\left[\frac{\boldsymbol{\lambda}\cdot\mathbf{1}+\frac{t}{\theta}\boldsymbol{\eta}\cdot\mathbf{1}}{1+\boldsymbol{\lambda}\cdot\mathbf{1}+\frac{t}{\theta}\boldsymbol{\eta}\cdot\mathbf{1}} - \frac{\boldsymbol{\lambda}\cdot\mathbf{1}}{1+\boldsymbol{\lambda}\cdot\mathbf{1}}\right] \\
&= \frac{1}{1+\boldsymbol{\lambda}\cdot\mathbf{1}},
\end{aligned}$$

which is the Laplace transform of a random vector with almost surely equal coordinates, each of them exponentially distributed with parameter 1. Hence for all $\theta \geqslant 0$, $\boldsymbol{x} \in \mathbb{R}_+^d\backslash\{\mathbf{0}\}$, and $\mathbf{u} \geqslant \mathbf{0}$,

$$\lim_{t\to\infty} \mathbb{P}_{\boldsymbol{x}}\left[\frac{2X_{t,1}}{\zeta\eta_1 t} > u_1, \ldots, \frac{2X_{t,d}}{\zeta\eta_d t} > u_d \,|\mathbf{X}_{t+\theta}\neq\mathbf{0}\right] = e^{-\max_i u_i}.$$

2.4 Q-process

In the following **Subsection 2.4.1** (resp. **Subsection 2.4.2**), we consider the following conditional limit

$$\lim_{\theta\to\infty} \widetilde{\mathbb{P}}\left(\mathbf{X}_t \in \,.\, |\mathbf{X}_{t+\theta}\neq\mathbf{0}\right), \tag{2.4.1}$$

where $\widetilde{\mathbb{P}}$ is the law of the multitype BGWc process (resp. Feller diffusion process) conditioned on extinction, introduced in Subsection 2.2.1 (resp. Subsection 2.2.2). This means that we first *delay the extinction* of at least θ, and then let θ tend to infinity. As shown in the following, this limit defines a Markov process, called the associated *Q-process*, which takes its values in $\mathbb{N}^d\backslash\{\mathbf{0}\}$ (resp. $\mathbb{R}_+^d\backslash\{\mathbf{0}\}$). To draw a parallel with the Q-process defined in [DalJof08] for a multitype Bienaymé-Galton-Watson process, we can also write (2.4.1) as follows,

$$\lim_{\theta\to\infty} \mathbb{P}\left(\mathbf{X}_t \in \,.\,|\,\mathbf{X}_{t+\theta}\neq\mathbf{0}, \lim_{s\to\infty}\mathbf{X}_s=\mathbf{0}\right). \tag{2.4.2}$$

It appears in this way that the law of the Q-process can be roughly thought as the law of the process conditioned on not being extinct in the distant future and on being extinct in the even more distant future. We will thus sometimes refer to this process as the process *conditioned on very late extinction*, or *conditioned on extinction in the remote future*.

2.4.1 Q-process associated with the multitype BGWc process

For all $t \geqslant 0$ and all $B \in \mathcal{F}_t$, we define

$$\mathbb{P}^*(B) := \lim_{\theta \to \infty} \widetilde{\mathbb{P}}\left(B \,|\, \mathbf{X}_{t+\theta} \neq \mathbf{0}\right), \tag{2.4.3}$$

if this limit exists.

The conditioned BGWc process as h-process

We prove in the following theorem that the limit $\mathbb{P}^*$ written in (2.4.3) does exist and is a well-defined probability measure on $\mathcal{F}_t$ which is absolutely continuous with respect to $\mathbb{P}|_{\mathcal{F}_t}$. We show moreover that $\mathbb{P}^*$ is actually a Doob h-transform of $\mathbb{P}$: the local definitions of $\mathbb{P}^*$ on several $\mathcal{F}_t$ are compatible and thus define a unique probability measure on $\vee_{t\geqslant 0}\mathcal{F}_t$.

Theorem 2.4.1. *Let us assume **(B1)** and **(B2)**. We assume moreover that*

(i) if $\rho = 0$, all the second order moments of the offspring distribution are finite,

(ii) if $\rho < 0$, for all $i, j = 1 \ldots d$, $\sum_{\boldsymbol{k} \in \mathbb{N}^d} k_j \ln(k_j) p_i(\boldsymbol{k}) < \infty$.

Then $\mathbb{P}^*$ is a Doob h-transform of $\mathbb{P}$ satisfying for all $t \geqslant 0$ and all $\boldsymbol{x} \in \mathbb{N}^d$, $\boldsymbol{x} \neq \boldsymbol{0}$,

$$d\mathbb{P}^*_{\boldsymbol{x}}|_{\mathcal{F}_t} = e^{-\widetilde{\rho} t} \frac{\boldsymbol{q}^{\boldsymbol{X}_t}}{\boldsymbol{q}^{\boldsymbol{x}}} \frac{\boldsymbol{X}_t \cdot \widetilde{\boldsymbol{\xi}}}{\boldsymbol{x} \cdot \widetilde{\boldsymbol{\xi}}}\, d\mathbb{P}_{\boldsymbol{x}}|_{\mathcal{F}_t}. \tag{2.4.4}$$

In particular, if $\rho \leqslant 0$, $\mathbb{P}^*$ satisfies for all $t \geqslant 0$ and all $\boldsymbol{x} \in \mathbb{N}^d$, $\boldsymbol{x} \neq \boldsymbol{0}$,

$$d\mathbb{P}^*_{\boldsymbol{x}}|_{\mathcal{F}_t} = e^{-\rho t} \frac{\boldsymbol{X}_t \cdot \boldsymbol{\xi}}{\boldsymbol{x} \cdot \boldsymbol{\xi}}\, d\mathbb{P}_{\boldsymbol{x}}|_{\mathcal{F}_t}. \tag{2.4.5}$$

Proof. The proof relies mostly on the asymptotical properties of the extinction probability vector $\mathbf{q}(t)$ as t tends to infinity, in both critical and subcritical cases. Thanks to Proposition 2.2.1, the supercritical case is then simply reduced to the subcritical case.

Let $t \geqslant 0$ and $B \in \mathcal{F}_t$. By definition, for all $\theta \geqslant 0$ and all $\boldsymbol{x} \in \mathbb{N}^d$, $\boldsymbol{x} \neq \mathbf{0}$,

$$\begin{aligned}\widetilde{\mathbb{E}}_{\boldsymbol{x}}\left[\mathbf{1}_B \mid \mathbf{X}_{t+\theta} \neq \mathbf{0}\right] &= \frac{\widetilde{\mathbb{E}}_{\boldsymbol{x}}\left[\widetilde{\mathbb{E}}_{\boldsymbol{x}}\left(\mathbf{1}_B \mathbf{1}_{\{\mathbf{X}_{t+\theta} \neq \mathbf{0}\}} | \mathcal{F}_t\right)\right]}{\widetilde{\mathbb{P}}_{\boldsymbol{x}}\left(\mathbf{X}_{t+\theta} \neq \mathbf{0}\right)} \\ &= \frac{\widetilde{\mathbb{E}}_{\boldsymbol{x}}\left[\mathbf{1}_B \widetilde{\mathbb{P}}_{\boldsymbol{x}}\left(\mathbf{X}_{t+\theta} \neq \mathbf{0} | \mathcal{F}_t\right)\right]}{\widetilde{\mathbb{P}}_{\boldsymbol{x}}\left(\mathbf{X}_{t+\theta} \neq \mathbf{0}\right)}.\end{aligned}$$

By virtue of the Markov and branching properties we obtain

$$\begin{aligned}\widetilde{\mathbb{E}}_{\boldsymbol{x}}\left[\mathbf{1}_B \mid \mathbf{X}_{t+\theta} \neq \mathbf{0}\right] &= \frac{\widetilde{\mathbb{E}}_{\boldsymbol{x}}\left[\mathbf{1}_B \left(1 - \widetilde{\mathbb{P}}_{\mathbf{X}_t}\left(\mathbf{X}_\theta = \mathbf{0}\right)\right)\right]}{1 - \widetilde{\mathbb{P}}_{\boldsymbol{x}}\left(\mathbf{X}_{t+\theta} = \mathbf{0}\right)} \\ &= \frac{\widetilde{\mathbb{E}}_{\boldsymbol{x}}\left[\mathbf{1}_B \left(1 - \widetilde{\mathbf{q}}(\theta)^{\mathbf{X}_t}\right)\right]}{1 - \widetilde{\mathbf{q}}(t+\theta)^{\boldsymbol{x}}}.\end{aligned} \tag{2.4.6}$$

- **Critical case $\rho = 0$.** We have $\widetilde{\mathbb{P}} = \mathbb{P}$, and (2.4.6) becomes

$$\mathbb{E}_{\boldsymbol{x}}\left[\mathbf{1}_B \mid \mathbf{X}_{t+\theta} \neq \mathbf{0}\right] = \mathbb{E}_{\boldsymbol{x}}\left[\mathbf{1}_B \frac{1 - \mathbf{q}(\theta)^{\mathbf{X}_t}}{1 - \mathbf{q}(t+\theta)^{\boldsymbol{x}}}\right]. \tag{2.4.7}$$

Using the asymptotic behavior of $\mathbf{q}(t)$ given by Proposition 1.1.25 (holding under assumption *(i)*), we obtain that for all $t \geqslant 0$ and all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{N}^d$, $\boldsymbol{x} \neq \mathbf{0}$,

$$\begin{aligned}
\lim_{\theta\to\infty} \frac{1-\mathbf{q}(\theta)^{\boldsymbol{y}}}{1-\mathbf{q}(t+\theta)^{\boldsymbol{x}}} &= \lim_{\theta\to\infty} \frac{1-\prod_{i=1}^d \left[1-\frac{2\xi_i}{\zeta\theta}\right]^{y_i}}{1-\prod_{i=1}^d \left[1-\frac{2\xi_i}{\zeta(t+\theta)}\right]^{x_i}} \\
&= \lim_{\theta\to\infty} \frac{1-\prod_{i=1}^d \left[1-\frac{2y_i\xi_i}{\zeta\theta}+o(\frac{1}{\theta})\right]}{1-\prod_{i=1}^d \left[1-\frac{2x_i\xi_i}{\zeta(t+\theta)}+o(\frac{1}{t+\theta})\right]} \\
&= \lim_{\theta\to\infty} \frac{\sum_{i=1}^d \frac{2y_i\xi_i}{\zeta\theta}+o(\frac{1}{\theta})}{\sum_{i=1}^d \frac{2x_i\xi_i}{\zeta(t+\theta)}+o(\frac{1}{t+\theta})} \\
&= \lim_{\theta\to\infty} \frac{t+\theta}{\theta}\frac{\boldsymbol{y}\cdot\boldsymbol{\xi}+o(1)}{\boldsymbol{x}\cdot\boldsymbol{\xi}+o(1)} \\
&= \frac{\boldsymbol{y}\cdot\boldsymbol{\xi}}{\boldsymbol{x}\cdot\boldsymbol{\xi}}.
\end{aligned}$$

Moreover, for θ large enough,

$$\left|\frac{1-\mathbf{q}(\theta)^{\boldsymbol{y}}}{1-\mathbf{q}(t+\theta)^{\boldsymbol{x}}}\right| \leqslant 2\frac{\boldsymbol{y}\cdot\boldsymbol{\xi}}{\boldsymbol{x}\cdot\boldsymbol{\xi}}.$$

Since by (1.1.15) and Proposition 1.1.3, $\mathbb{E}_{\boldsymbol{x}}(\mathbf{X}_t\cdot\boldsymbol{\xi}) = \sum_{j=1}^d \xi_j \sum_{i=1}^d x_i m_{ij}(t) < \infty$, we obtain by dominated convergence in (2.4.7) that

$$\lim_{\theta\to\infty} \mathbb{E}_{\boldsymbol{x}}\left[\mathbf{1}_B \mid \mathbf{X}_{t+\theta}\neq\mathbf{0}\right] = \mathbb{E}_{\boldsymbol{x}}\left[\mathbf{1}_B\frac{\mathbf{X}_t\cdot\boldsymbol{\xi}}{\boldsymbol{x}\cdot\boldsymbol{\xi}}\right],$$

which leads to (2.4.5).

- **Subcritical case $\rho < 0$.** We similarly use the known asymptotic behavior of $\mathbf{q}(t)$ given by Proposition 1.1.24, which holds under assumption *(ii)*. For all $t \geqslant 0$ and for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{N}^d$, $\boldsymbol{x} \neq \mathbf{0}$,

$$\begin{aligned}
\lim_{\theta\to\infty} \frac{1-\mathbf{q}(\theta)^{\boldsymbol{y}}}{1-\mathbf{q}(t+\theta)^{\boldsymbol{x}}} &= \lim_{\theta\to\infty} \frac{1-\prod_{i=1}^d \left[1-K\xi_i e^{\rho\theta}\right]^{y_i}}{1-\prod_{i=1}^d \left[1-K\xi_i e^{\rho(t+\theta)}\right]^{x_i}} \\
&= \lim_{\theta\to\infty} \frac{1-\prod_{i=1}^d \left[1-Ky_i\xi_i e^{\rho\theta}+o(e^{\rho\theta})\right]}{1-\prod_{i=1}^d \left[1-Kx_i\xi_i e^{\rho(t+\theta)}+o(e^{\rho(t+\theta)})\right]} \\
&= \lim_{\theta\to\infty} \frac{Ke^{\rho\theta}\sum_{i=1}^d y_i\xi_i+o(e^{\rho\theta})}{Ke^{\rho(t+\theta)}\sum_{i=1}^d x_i\xi_i+o(e^{\rho(t+\theta)})} \\
&= \lim_{\theta\to\infty} \frac{e^{\rho\theta}}{e^{\rho(t+\theta)}}\frac{\boldsymbol{y}\cdot\boldsymbol{\xi}+o(1)}{\boldsymbol{x}\cdot\boldsymbol{\xi}+o(1)} \\
&= e^{-\rho t}\frac{\boldsymbol{y}\cdot\boldsymbol{\xi}}{\boldsymbol{x}\cdot\boldsymbol{\xi}}.
\end{aligned}$$

which by dominated convergence in (2.4.7) leads to (2.4.5).

- **Supercritical case $\rho > 0$.** We apply the previous result to the subcritical process with law $\widetilde{\mathbb{P}}$, which is allowed since, as detailed in the proof of Propostion (2.3.1), it automatically satisfies the required **(XlogX)** condition. Hence, for all $t \geqslant 0$ and all $\boldsymbol{x} \in \mathbb{N}^d$, $\boldsymbol{x} \neq \mathbf{0}$,

$$d\mathbb{P}^*_{\boldsymbol{x}}|_{\mathcal{F}_t} = e^{-\widetilde{\rho}t}\frac{\mathbf{X}_t\cdot\widetilde{\boldsymbol{\xi}}}{\boldsymbol{x}\cdot\widetilde{\boldsymbol{\xi}}}\, d\widetilde{\mathbb{P}}_{\boldsymbol{x}}|_{\mathcal{F}_t}, \tag{2.4.8}$$

which combined with Proposition 2.2.1 leads to (2.4.4).

We finally check that for any class of criticality $(e^{-\widetilde{\rho}t}\mathbf{q}^{\mathbf{X}_t}\mathbf{X}_t\cdot\widetilde{\boldsymbol{\xi}})_{t\geqslant 0}$ is a $(\mathbb{P},\mathcal{F}_t)$-martingale. We have for all $t\geqslant s\geqslant 0$, using (1.1.15) and (1.1.16),

$$\begin{aligned}\widetilde{\mathbb{E}}\left[e^{-\widetilde{\rho}t}\mathbf{X}_t\cdot\widetilde{\boldsymbol{\xi}}|\mathcal{F}_s\right] &= e^{-\widetilde{\rho}t}\widetilde{\mathbb{E}}\left[\mathbf{X}_t|\mathcal{F}_s\right]\cdot\widetilde{\boldsymbol{\xi}}\\ &= e^{-\widetilde{\rho}t}\left(\mathbf{X}_s e^{\widetilde{\mathbf{C}}(t-s)}\right)\cdot\widetilde{\boldsymbol{\xi}} = e^{-\widetilde{\rho}t}\mathbf{X}_s\cdot\left(e^{\widetilde{\mathbf{C}}(t-s)}\widetilde{\boldsymbol{\xi}}^T\right)\\ &= e^{-\widetilde{\rho}t}\mathbf{X}_s\cdot\left(e^{\widetilde{\rho}(t-s)}\widetilde{\boldsymbol{\xi}}^T\right) = e^{-\widetilde{\rho}s}\mathbf{X}_s\cdot\widetilde{\boldsymbol{\xi}}.\end{aligned}$$

Hence $(e^{-\widetilde{\rho}t}\mathbf{X}_t\cdot\widetilde{\boldsymbol{\xi}})_{t\geqslant 0}$, is a $(\widetilde{\mathbb{P}},\mathcal{F}_t)$- martingale, which together with the fact that $(\mathbf{q}^{\mathbf{X}_t})_{t\geqslant 0}$ is a $(\mathbb{P},\mathcal{F}_t)$-martingale leads to the result, and implies that $\mathbb{P}^*$ is a Doob h-transform of $\mathbb{P}$. □

Let us compute the infinitesimal generator L^* of the process with law $\mathbb{P}^*$. Applying Theorem 2.4.1, we have for all smooth function $f:\mathbb{N}^d\backslash\{\mathbf{0}\}\to\mathbb{R}$

$$L^*f := \frac{1}{h}\widetilde{L}(hf),$$

and thus

$$\begin{aligned}(L^*f)(\boldsymbol{x}) &= -\widetilde{\rho}f(\boldsymbol{x})\\ &\quad+\frac{1}{\boldsymbol{x}\cdot\widetilde{\boldsymbol{\xi}}}\sum_{i=1}^d\alpha_i x_i\sum_{\mathbf{k}\in\mathbb{N}^d}\widetilde{p}_i(\mathbf{k})\left[(\boldsymbol{x}+\mathbf{k}-\mathbf{e}_i)\cdot\widetilde{\boldsymbol{\xi}}\ f(\boldsymbol{x}+\mathbf{k}-\mathbf{e}_i)-\boldsymbol{x}\cdot\widetilde{\boldsymbol{\xi}}\ f(\boldsymbol{x})\right].\end{aligned}\tag{2.4.9}$$

Using the definition of $\widetilde{\rho}$ and $\widetilde{\boldsymbol{\xi}}$ given in Definition 2.2.3, we obtain that

$$\widetilde{\rho} = \frac{1}{\boldsymbol{x}\cdot\widetilde{\boldsymbol{\xi}}}\sum_{i=1}^d\alpha_i x_i\sum_{\mathbf{k}\in\mathbb{N}^d}\widetilde{p}_i(\mathbf{k})\ (\mathbf{k}-\mathbf{e}_i)\cdot\widetilde{\boldsymbol{\xi}},$$

hence (2.4.9) becomes

$$(L^*f)(\boldsymbol{x}) = \sum_{i=1}^d\alpha_i x_i\sum_{\mathbf{k}\in\mathbb{N}^d}\widetilde{p}_i(\mathbf{k})\frac{(\boldsymbol{x}+\mathbf{k}-\mathbf{e}_i)\cdot\widetilde{\boldsymbol{\xi}}}{\boldsymbol{x}\cdot\widetilde{\boldsymbol{\xi}}}\left[f(\boldsymbol{x}+\mathbf{k}-\mathbf{e}_i)-f(\boldsymbol{x})\right].\tag{2.4.10}$$

The conditioned paths and the immortal particle

As already mentioned, Jagers and Lagerås proved in [JagLag08] that a branching process conditioned to die out remains a branching process, and that in the supercritical case the conditioned process becomes subcritical. Conditioning on extinction in the remote future thus influences the life careers of the individuals but preserves the branching property. We will see that the same does not occur when conditioning on very late extinction, and the purpose of this section is to describe the structure of the conditioned process obtained in Theorem 2.4.1.

One can easily verify that the branching property is not preserved for the process with law $\mathbb{P}^*$. For $d=1$ and $\rho=0$ we have for instance, using the fact that $\mathbb{E}_1(X_t^2)=\alpha\sigma^2 t+1$ (proved *e.g.* in [AthNey72] Section 3.4),

$$\begin{aligned}\mathbb{E}^*_x(X_t) &= \frac{1}{x}\mathbb{E}_x(X_t^2) = \frac{1}{x}\left[x\mathbb{V}ar_1(X_t)+(x\mathbb{E}_1(X_t))^2\right]\\ &= \frac{1}{x}\left[x\alpha\sigma^2 t+x^2\right] = x+\alpha\sigma^2 t.\end{aligned}$$

So the linearity with respect to the initial condition is no more satisfied, since

$$\mathbb{E}^*_x(X_t)\neq x\mathbb{E}^*_1(X_t)$$

for $x\neq 1$.

Nevertheless we will see that the branching structure is somehow preserved: the conditioned process with law $\mathbb{P}^*$ behaves like an unconditioned (sub)critical branching process to which an external force is added, compelling the process to die out very late.

Monotype case. In the monotype case $d = 1$, this external input is a standard immigration. We show here that a single-type BGWc process of any class of criticality conditioned on very late extinction, and from which one removes one individual, has the same law as a BGWc process with immigration. This result is already known for critical BGW processes ([KaWat71]).

First, note that the infinitesimal generator L^* can also be written as a perturbation of $\widetilde{L}$:

$$(L^* f)(\boldsymbol{x}) = \left(\widetilde{L} f\right)(\boldsymbol{x}) + \frac{1}{\boldsymbol{x} \cdot \boldsymbol{\xi}} \sum_{i=1}^{d} \alpha_i x_i \sum_{\mathbf{k} \in \mathbb{N}^d} \widetilde{p}_i(\mathbf{k})(\mathbf{k} - \mathbf{e}_i) \cdot \boldsymbol{\xi} \left[f(\boldsymbol{x} + \mathbf{k} - \mathbf{e}_i) - f(\boldsymbol{x}) \right].$$

We point out that for $\boldsymbol{x} = \mathbf{e}_i$ and $\mathbf{k} = \mathbf{0}$ the term $f(\boldsymbol{x} + \mathbf{k} - \mathbf{e}_i)$ is not defined in the last summand, but is canceled by the same term in $\left(\widetilde{L} f\right)(\boldsymbol{x})$.

In the monotype case, this relation reduces to

$$(L^* f)(x) = \left(\widetilde{L} f\right)(x) + \alpha \sum_{k=0}^{\infty} (k-1) \widetilde{p}_k \left[f(x+k-1) - f(x) \right], \tag{2.4.11}$$

which again only makes sense for the terms $k > 0$. So we can interpret the additional term as some constant immigration ("constant" in the sense that it does not depend on x). We define $\widehat{\mathbb{P}}$, the law of the Q-process shifted downwards by 1, as follows: for all $x \in \mathbb{N}$,

$$\widehat{\mathbb{P}}_x (X_t = .) := \mathbb{P}^*_{x+1} (X_t - 1 = .).$$

Let us show that $\widehat{\mathbb{P}}$ is the law of a BGWc process with constant immigration, *i.e.* that for all $x \in \mathbb{N}$ and $r \in [0, 1]$,

$$\widehat{\mathbb{E}}_x(r^{X_t}) = [\widetilde{F}_t(r)]^x \exp[\int_0^t h(\widetilde{F}_u(r)) du],$$

where the immigration function h is of the form (see *e.g.* [Sew75] Kapitel 7.1)

$$\sum_{k=0}^{\infty} v_k r^k, \text{ with } \sum_{k=0}^{\infty} v_k = 0,\ v_0 < 0 \text{ and for all } k \geqslant 1,\ v_k \geqslant 0.$$

If so, the probability that no individual immigrates (resp. $k \geqslant 1$ individuals immigrate) during a time-interval Δt with $\Delta t \to 0$ is given by $1 + v_0 \Delta t + o(\Delta t)$ (resp. $v_k \Delta t + o(\Delta t)$).

Note that a single-type BGW process with constant immigration would have the following generating function at time n (c.f. [KaWat71] p.45),

$$\mathbb{E}_i(r^{X_n}) = f_n(r)^i \prod_{k=0}^{n-1} g\left(f_k(r)\right),$$

where f (resp. g) stands for the generating function of the offspring distribution (resp. of the immigration distribution). A comparison between the continuous and discrete-time cases then enables us to draw an analogy between the immigration function h and $\log g$. Indeed, writing $g(r) = \sum_{k=0}^{\infty} u_k r^k$ with $\sum_{k=0}^{\infty} u_k = 1$, the analogy $h \simeq \log g$ then explains why $h(1) \simeq \log g(1) = \log 1 = 0$, and why $v_0 = h(0) \simeq \log u_0 < 0$.

Applying Theorem 2.4.1, we have for all $x \in \mathbb{N}$ and $r \in [0, 1]$,

$$\widehat{\mathbb{E}}_x(r^{X_t}) = \mathbb{E}^*_{x+1}(r^{X_t - 1}) = e^{-\alpha(\widetilde{m}-1)t} \frac{1}{x+1} \widetilde{\mathbb{E}}_{x+1}\left(X_t r^{X_t - 1}\right).$$

But

$$\widetilde{\mathbb{E}}_{x+1}\left(X_t r^{X_t - 1}\right) = \frac{\partial}{\partial r}\left(\widetilde{\mathbb{E}}_{x+1}(r^{X_t})\right) = \frac{\partial}{\partial r}\left(\left[\widetilde{F}_t(r)\right]^{x+1}\right) = (x+1)\left[\widetilde{F}_t(r)\right]^x \frac{\partial \widetilde{F}_t(r)}{\partial r},$$

and we know from (1.1.8) and (1.1.9) that $\widetilde{F}_t(r)$ satisfies

$$\frac{\partial \widetilde{F}_t(r)}{\partial t} = \alpha\left[\widetilde{f}\left(\widetilde{F}_t(r)\right) - \widetilde{F}_t(r)\right] = \alpha\left[\widetilde{f}(r) - r\right]\frac{\partial \widetilde{F}_t(r)}{\partial r},$$

which implies

$$\frac{\partial \widetilde{F}_t(r)}{\partial r} = \frac{\widetilde{f}\left(\widetilde{F}_t(r)\right) - \widetilde{F}_t(r)}{\widetilde{f}(r) - r} = \frac{\widetilde{f}\left(\widetilde{F}_t(r)\right) - \widetilde{F}_t(r)}{\widetilde{f}(\widetilde{F}_0(r)) - \widetilde{F}_0(r)}.$$

Hence

$$\begin{aligned}
\widehat{\mathbb{E}}_x(r^{X_t}) &= \left[\widetilde{F}_t(r)\right]^x \exp\Big(-\alpha(\widetilde{m}-1)t + \ln\left[\widetilde{f}(\widetilde{F}_t(r)) - \widetilde{F}_t(r)\right] - \ln\left[\widetilde{f}(\widetilde{F}_0(r)) - \widetilde{F}_0(r)\right]\Big) \\
&= \left[\widetilde{F}_t(r)\right]^x \exp\left(\int_0^t \left(-\alpha(\widetilde{m}-1) + \frac{\frac{\partial}{\partial u}\left[\widetilde{f}(\widetilde{F}_u(r)) - \widetilde{F}_u(r)\right]}{\widetilde{f}(\widetilde{F}_u(r)) - \widetilde{F}_u(r)}\right) du\right) \\
&= \left[\widetilde{F}_t(r)\right]^x \exp\left(\int_0^t \left(-\alpha(\widetilde{m}-1) + \frac{\partial \widetilde{F}_u(r)}{\partial u}\frac{\widetilde{f}'(\widetilde{F}_u(r)) - 1}{\widetilde{f}(\widetilde{F}_u(r)) - \widetilde{F}_u(r)}\right) du\right) \\
&= \left[\widetilde{F}_t(r)\right]^x \exp\left(\int_0^t \left(-\alpha(\widetilde{m}-1) + \alpha\left(\widetilde{f}'(\widetilde{F}_u(r)) - 1\right)\right) du\right) \\
&= \left[\widetilde{F}_t(r)\right]^x \exp\left(\int_0^t \left(\alpha\left(\widetilde{f}'(\widetilde{F}_u(r)) - \widetilde{m}\right)\right) du\right).
\end{aligned}$$

We finally check that $h := \alpha(\widetilde{f}' - \widetilde{m})$ defines an immigration function as described above, which leads to the following proposition.

Proposition 2.4.2. *Let us assume $d = 1$, $m < \infty$ and **(B2)**. We assume moreover that*

(i) if $\rho = 0$, $\sigma^2 < \infty$,

(ii) if $\rho < 0$, $\sum_{k\in\mathbb{N}} k \ln k\, p(k) < \infty$.

Then the Q-process shifted downwards by 1 is a BGWc process with constant immigration, where the branching process has branching rate α and offspring generating function $\widetilde{f}$. The immigration is described in the following sense by $\alpha(\widetilde{f}' - \widetilde{m})$: the probability that during a time-interval Δt with $\Delta t \to 0$

- *$k \geqslant 1$ individuals immigrate, is*

$$\alpha(k+1)\widetilde{p}_{k+1}\Delta t + o(\Delta t),$$

- *no individual immigrates, is*

$$1 - \alpha(\widetilde{m} - \widetilde{p}_1)\Delta t + o(\Delta t).$$

Another possible interpretation in the monotype case is to consider this constant 1 as an *immortal individual*, and to consider the immigrants as the offspring of this individual, counting in addition the individual himself as its own descendant (see Figure 2.1). The probability that the immortal individual produces $k > 1$ offsprings is thus related to the probability that $k-1$ individuals immigrate, *i.e.* $\alpha k \widetilde{p}_k$. Since $\sum_{k\geqslant 1} \alpha k \widetilde{p}_k = \alpha \widetilde{m}$, we set

$$s(k) := \frac{1}{m} k \widetilde{p}_k, \quad k \geqslant 1. \tag{2.4.12}$$

Then the Q-process can be described as the independent sum of a non-conditioned BGWc process and an immortal individual, producing offsprings (including himself) at rate αm, according to the *size-biased* distribution $(s(k))_{k\geqslant 1}$. Note that since the distribution s is concentrated on $\mathbb{N}^*$, the immortal individual always produces at least one offspring, hence the appellation "immortal".

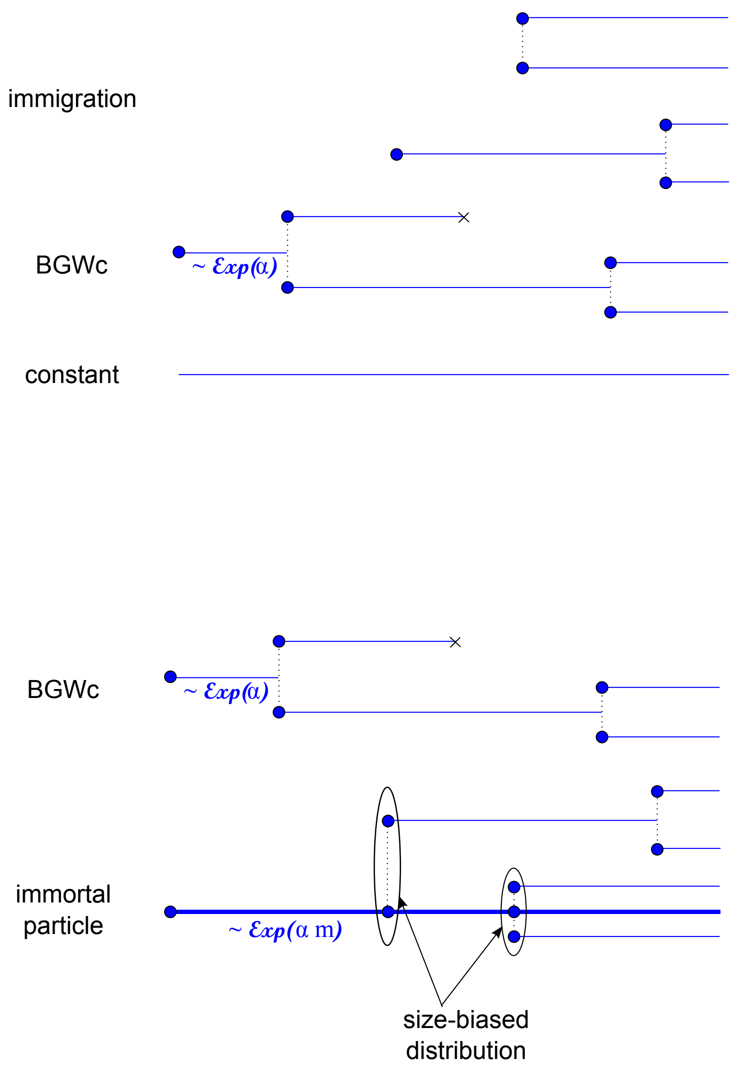

Figure 2.1: Two interpretations of the Q-process in the monotype case.

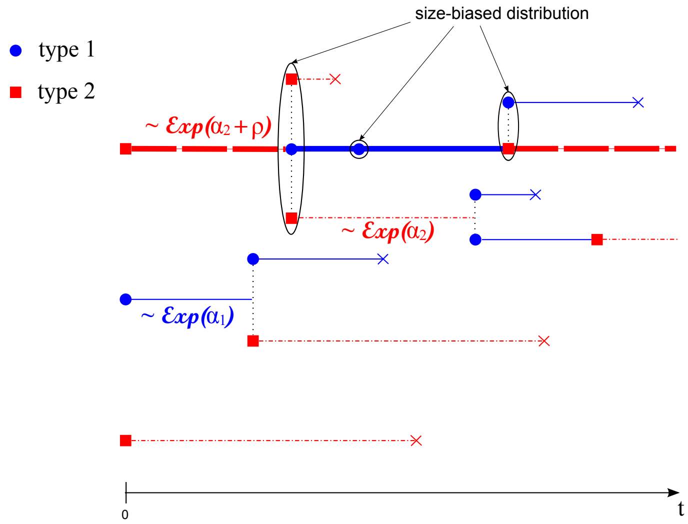

Figure 2.2: The Q-process and the immortal particle.

Multitype case. In the multitype case, an interpretation of the Q-process with a classical branching process with immigration as in Proposition 2.4.2 is not possible, since the removed constant should be allowed to change its type.

The second interpretation in the monotype case which involves an immortal individual can however be extended to the multitype case $d \geqslant 1$, and differs from the monotype case by the fact that the immortal individual can mutate from one type to another. We prove indeed in Proposition 2.4.3 that the external input which prevents the process from dying out comes from an *immortal individual* or *immortal particle* (so called in reference to [Ev93]). More precisely, the behavior of the immortal individual is the one of the trunk of a size-biased multitype Galton-Watson tree in its continuous-time version, introduced in [GeoBa03].

Let us introduce the following *size-biased* offspring distribution $(s_i(\mathbf{k}))_{\mathbf{k}\in\mathbb{N}^d}$ with respect to the offspring distribution $(\widetilde{p}_i(\mathbf{k}))_{\mathbf{k}\in\mathbb{N}^d}$ defined in (2.2.5). For all $i = 1 \ldots d$,

$$s_i(\mathbf{k}) := \frac{\alpha_i}{(\alpha_i+\widetilde{\rho})\widetilde{\xi}_i}\mathbf{k}\cdot\widetilde{\boldsymbol{\xi}}\,\widetilde{p}_i(\mathbf{k}), \quad \mathbf{k}\in\mathbb{N}^d. \tag{2.4.13}$$

We easily check that for all $i = 1 \ldots d$, $(s_i(\mathbf{k}))_{\mathbf{k}\in\mathbb{N}^d}$ is a probability distribution:

$$\sum_{\mathbf{k}\in\mathbb{N}^d} s_i(\mathbf{k}) = \frac{1}{(\alpha_i+\widetilde{\rho})\xi_i}\sum_{j=1}^{d}\alpha_i\widetilde{m}_{ij}\widetilde{\xi}_j = \frac{1}{(\alpha_i+\widetilde{\rho})\widetilde{\xi}_i}\Big[\sum_{j=1}^{d}\widetilde{c}_{ij}\widetilde{\xi}_j+\alpha_i\widetilde{\xi}_i\Big] = 1,$$

and that it is concentrated on $\mathbb{N}^d\backslash\{\mathbf{0}\}$.

Let us now describe in detail the structure of the conditioned process with law $\mathbb{P}^*$ (see Figure 2.2).

Proposition 2.4.3. *Let us assume **(B1)** and **(B2)**. We assume moreover that*

(i) if $\rho = 0$, all the second order moments of the offspring distribution are finite,

(ii) if $\rho < 0$, for all $i, j = 1 \ldots d$, $\sum_{\boldsymbol{k}\in\mathbb{N}^d} k_j \ln(k_j)p_i(\boldsymbol{k}) < \infty$.

Then $\mathbb{P}^*$ is the law of the independent sum of a (sub)critical branching process and of an "immortal particle" or immortal individual.

The branching process has branching rates $\alpha_1, \ldots, \alpha_d$ and offspring generating function $\widetilde{\boldsymbol{f}}$.

Given that the immortal individual is of type i, this individual has an exponential life-time of parameter $\alpha_i + \widetilde{\rho}$ and an offspring distribution $(s_i(\boldsymbol{k}))_{\boldsymbol{k}\in\mathbb{N}^d}$ given by (2.4.13). Its initial type is i

with probability $x_i\widetilde{\xi}_i/\boldsymbol{x}\cdot\widetilde{\boldsymbol{\xi}}$, where $\boldsymbol{x}\in\mathbb{N}^d$ is the initial number of individuals, and if it produces $\boldsymbol{k}$ offspring it mutates to type j with probability $k_j\widetilde{\xi}_j/\boldsymbol{k}\cdot\widetilde{\boldsymbol{\xi}}$.

Proof. Denoting by $\overline{\mathbb{P}}$ the law of the time-homogeneous Markov process described in Proposition 2.4.3, its infinitesimal generator $\overline{L}$ is by definition

$$(\overline{L}f)(\boldsymbol{x}) := \lim_{h\to 0}\frac{1}{h}\overline{\mathbb{E}}_{\boldsymbol{x}}\left[f\left(\mathbf{X}_h\right)-f\left(\boldsymbol{x}\right)\right].$$

Let $h>0$. It appears that during the time-interval $[0,h]$, the only nontrivial events whose probabilities are of order h as $h\to 0$ are the ones consisting of exactly one branching event, either of the unconditioned branching process or of the immortal individual. The first possibility is that during $[0,h]$, the immortal individual of type i splits into $\mathbf{k}$ offspring, with no other event occurring. The probability of this event is then

$$\frac{x_i\widetilde{\xi}_i}{\boldsymbol{x}\cdot\widetilde{\boldsymbol{\xi}}}(\alpha_i+\widetilde{\rho})s_i(\mathbf{k})h=\alpha_i x_i\frac{\mathbf{k}\cdot\widetilde{\boldsymbol{\xi}}}{\boldsymbol{x}\cdot\widetilde{\boldsymbol{\xi}}}\widetilde{p}_i(\mathbf{k})h.$$

The second possible type of event appearing in the computation of $\overline{L}$ is that one individual of type i in the unconditioned process splits into $\mathbf{k}$ offspring, while the immortal individual is of type j and no other event occurs. The probability of this event is then

$$\frac{x_j\widetilde{\xi}_j}{\boldsymbol{x}\cdot\widetilde{\boldsymbol{\xi}}}\alpha_i(x_i-\delta_{ij})\widetilde{p}_i(\mathbf{k}).$$

The infinitesimal generator $\overline{L}$ is thus given by

$$\begin{aligned}\left(\overline{L}f\right)(\boldsymbol{x}) = \sum_{i=1}^{d}\sum_{\mathbf{k}\in\mathbb{N}^d}\alpha_i x_i\frac{\mathbf{k}\cdot\widetilde{\boldsymbol{\xi}}}{\boldsymbol{x}\cdot\widetilde{\boldsymbol{\xi}}}\widetilde{p}_i(\mathbf{k})\left[f\left(\boldsymbol{x}+\mathbf{k}-\mathbf{e}_i\right)-f\left(\boldsymbol{x}\right)\right] \\ +\sum_{i=1}^{d}\sum_{j=1}^{d}\sum_{\mathbf{k}\in\mathbb{N}^d}\frac{x_j\widetilde{\xi}_j}{\boldsymbol{x}\cdot\widetilde{\boldsymbol{\xi}}}\alpha_i(x_i-\delta_{ij})\widetilde{p}_i(\mathbf{k})\left[f\left(\boldsymbol{x}+\mathbf{k}-\mathbf{e}_i\right)-f\left(\boldsymbol{x}\right)\right],\end{aligned}$$

and we observe that $\overline{L}=L^*$. □

Remark 2.4.4. We see thanks to this proposition in an obvious way that the process conditioned on very late extinction never reaches the absorbing state $\mathbf{0}$. Indeed, the so-called immortal particle is really "immortal" since it always produces at least one descendant according to the size-biased distribution, and thus indefinitely "regenerates" itself.

Remark 2.4.5. The same result can be proved for a Q-process associated with a BGW process. We just sketch the proof: assume for simplicity that the considered BGW process is (sub)critical, with offspring distribution given by $p_j(\mathbf{k})$. Then the Q-process has the same law as the sum of a BGW process and an independent immortal individual which, given that it is of type j, produces $\mathbf{k}$ offsprings according to the size-biased distribution

$$\frac{1}{\rho\xi_j}\mathbf{k}\cdot\boldsymbol{\xi}\,p_j(\mathbf{k}).$$

The type of the immortal individual is determined as follows: if the initial population is $\mathbf{i}$, the initial type of the individual will be j with probability $\frac{i_j\xi_j}{\mathbf{i}\cdot\boldsymbol{\xi}}$. If the immortal individual produces $\mathbf{k}$ offspring, it then mutate to type l with probability $\frac{k_l\xi_l}{\mathbf{k}\cdot\boldsymbol{\xi}}$.

An numerical example. We consider a 2-type BGWc process with branching rates $\alpha_1 = \alpha_2 = 1$ and generating function

$$\begin{cases} f_1(\mathbf{r}) = \dfrac{7}{12} + \dfrac{1}{6}r_1 + \dfrac{1}{12}r_2 + \dfrac{1}{12}r_2^2 + \dfrac{1}{12}r_1^2 r_2 \\ f_2(\mathbf{r}) = \dfrac{7}{8} + \dfrac{1}{8}r_1 r_2^2 \end{cases}, \qquad \forall \mathbf{r} \in [0,1]^2.$$

The mean matrix is $\mathbf{M} = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} \\ \frac{1}{8} & \frac{1}{4} \end{pmatrix}$, and thus $\mathbf{B} = \begin{pmatrix} -\frac{2}{3} & \frac{1}{3} \\ \frac{1}{8} & -\frac{3}{4} \end{pmatrix}$, which is irreducible. The eigenvalues of $\mathbf{B}$ are $-\frac{1}{2}$ and $-\frac{11}{12}$, hence $\rho = -\frac{1}{2}$ and we are in the subcritical case. The related right and left eigenvectors with the normalization $\boldsymbol{\xi} \cdot \mathbf{1} = 1$ and $\boldsymbol{\xi} \cdot \boldsymbol{\eta} = 1$ are $\boldsymbol{\xi} = (\frac{2}{3}, \frac{1}{3})$, $\boldsymbol{\eta} = (\frac{9}{10}, \frac{6}{5})$. Then, if the process is initiated by one individual of type 1 and one individual of type 2, the immortal individual will be initially of type 1 with probability $\frac{2}{3}$, and of type 2 with probability $\frac{1}{3}$. Applying (2.4.13), we obtain the following expression of the size-biased offspring distribution of the immortal individual. For all $\mathbf{k} \in \mathbb{N}^d$,

$$s_1(\mathbf{k}) = (2k_1 + k_2)p_1(\mathbf{k}),$$
$$s_2(\mathbf{k}) = (4k_1 + 2k_2)p_2(\mathbf{k}).$$

The following table compares the unbiased and size-biased distributions.

	Unbiased distribution	*Size-biased distribution*
$\mathbf{k}$	$p_1(\mathbf{k})$	$s_1(\mathbf{k})$
$(0,0)$	7/12	0
$(1,0)$	1/6	1/3
$(0,1)$	1/12	1/12
$(0,2)$	1/12	1/6
$(2,1)$	1/12	5/12
	$p_2(\mathbf{k})$	$s_2(\mathbf{k})$
$(0,0)$	7/8	0
$(1,2)$	1/8	1

This numerical example illustrates the fact that the size-biased distribution gives a larger weight to the non-zero values than the original one, and gives no weight to the $\mathbf{0}$ value. In particular, any distribution with a support composed of $\mathbf{0}$ and of a non-zero element (such as p_2, with support $\{(0,0)\ (1,2)\}$) ends up in a constant distribution concentrated on the non-zero value (in our example, $s_2(1,2) = 1$).

Analogy with the Q-process associated with a BGW process

In the critical case we can draw an analogy between the (state-dependent) offspring distribution of the Q-process with law $\mathbb{P}^*$, and the transition probabilities of its discrete-time analog studied in [DalJof08]. In the critical case the infinitesimal generator L^* of the Q-process becomes indeed

$$(L^* f)(\boldsymbol{x}) = \sum_{i=1}^{d} \alpha_i x_i \sum_{\mathbf{k} \in \mathbb{N}^d} p_i(\mathbf{k}) \frac{(\boldsymbol{x} + \mathbf{k} - \mathbf{e}_i) \cdot \boldsymbol{\xi}}{\boldsymbol{x} \cdot \boldsymbol{\xi}} \left[f(\boldsymbol{x} + \mathbf{k} - \mathbf{e}_i) - f(\boldsymbol{x}) \right],$$

Let us define for every $i = 1 \dots d$ and $\boldsymbol{x} \in \mathbb{N}^d$, $\boldsymbol{x} \neq \mathbf{0}$,

$$p_i^*(\boldsymbol{x}, \mathbf{k}) := \frac{(\boldsymbol{x} + \mathbf{k} - \mathbf{e}_i) \cdot \boldsymbol{\xi}}{\boldsymbol{x} \cdot \boldsymbol{\xi}} p_i(\mathbf{k}). \tag{2.4.14}$$

Then

$$\sum_{\mathbf{k} \in \mathbb{N}^d} p_i^*(\boldsymbol{x}, \mathbf{k}) = 1 + \frac{1}{\boldsymbol{x} \cdot \boldsymbol{\xi}} \sum_{\mathbf{k} \in \mathbb{N}^d} (\mathbf{k} - \mathbf{e}_i) \cdot \boldsymbol{\xi}\, p_i(\mathbf{k}) = 1 + \frac{1}{\boldsymbol{x} \cdot \boldsymbol{\xi}} \sum_{j=1}^{d} c_{ij} \xi_j = 1.$$

Hence for every $i = 1 \dots d$ and $\boldsymbol{x} \in \mathbb{N}^d \backslash \{\mathbf{0}\}$, $(p_i^*(\boldsymbol{x}, \mathbf{k}))_{\mathbf{k} \in \mathbb{N}^d}$ is a probability distribution, and the analogy between (2.4.14) and formula (2.1.18), providing the relation between the transition probabilities of the BGW process and of the associated Q-process, is then obvious.

2.4.2 Q-process associated with the multitype Feller diffusion process

For all $t \geqslant 0$ and all $B \in \mathcal{F}_t$, we define

$$\mathbb{P}^*(B) := \lim_{\theta \to \infty} \widetilde{\mathbb{P}}(B \,|\, \mathbf{X}_{t+\theta} \neq \mathbf{0}), \tag{2.4.15}$$

if this limit exists.

The conditioned Feller diffusion process as h-process

We prove in the following theorem that the limit $\mathbb{P}^*$ given by (2.4.15) is a well-defined probability measure on $\mathcal{F}_t$ which is absolutely continuous with respect to $\mathbb{P}|_{\mathcal{F}_t}$, and that $\mathbb{P}^*$ as a Doob h-transform of $\mathbb{P}$ defines a probability measure on $\vee_{t \geqslant 0} \mathcal{F}_t$. Our result is a generalization of a result from N. Champagnat and S. Rœlly (Theorem 2.2 in [ChaRoe08]), dealing with critical or subcritical multitype Dawson-Watanabe processes, to Feller diffusion processes of any class of criticality. It is straightforward by using the fact that by conditioning a supercritical processes on extinction, one recovers a subcritical process (see Proposition 2.2.4). For this reason we omit the proof of the following statement.

Theorem 2.4.6. *Let us assume **(F1)** and **(F2)**. Then $\mathbb{P}^*$ is a Doob h-transform of $\mathbb{P}$ satisfying for all $t \geqslant 0$ and all $\boldsymbol{x} \in \mathbb{R}_+^d$, $\boldsymbol{x} \neq \boldsymbol{0}$,*

$$d\mathbb{P}^*_{\boldsymbol{x}}|_{\mathcal{F}_t} = e^{-\widetilde{\rho} t} \frac{e^{-\boldsymbol{X}_t \cdot \boldsymbol{u}}}{e^{-\boldsymbol{x} \cdot \boldsymbol{u}}} \frac{\boldsymbol{X}_t \cdot \widetilde{\boldsymbol{\xi}}}{\boldsymbol{x} \cdot \widetilde{\boldsymbol{\xi}}} \, d\mathbb{P}_{\boldsymbol{x}}|_{\mathcal{F}_t}. \tag{2.4.16}$$

In particular, if $\rho \leqslant 0$, $\mathbb{P}^*$ satisfies for all $t \geqslant 0$ and all $\boldsymbol{x} \in \mathbb{N}^d$, $\boldsymbol{x} \neq \boldsymbol{0}$,

$$d\mathbb{P}^*_{\boldsymbol{x}}|_{\mathcal{F}_t} = e^{-\rho t} \frac{\boldsymbol{X}_t \cdot \boldsymbol{\xi}}{\boldsymbol{x} \cdot \boldsymbol{\xi}} \, d\mathbb{P}_{\boldsymbol{x}}|_{\mathcal{F}_t}. \tag{2.4.17}$$

Denoting

$$h(t, \boldsymbol{x}) := e^{-\widetilde{\rho} t} \boldsymbol{x} \cdot \widetilde{\boldsymbol{\xi}}, \tag{2.4.18}$$

the infinitesimal generator L^* of the Q-process is then given on $D(L^*) := C^2\left(\mathbb{R}_+^d \backslash \{\mathbf{0}\}, \mathbb{R}\right)$ by

$$L^* f := \frac{1}{h} \widetilde{L}(hf).$$

Hence, for all $\boldsymbol{x} \in \mathbb{R}_+^d \backslash \{\mathbf{0}\}$,

$$(L^* f)(\boldsymbol{x}) = -\rho f(\boldsymbol{x}) + \frac{1}{2} \sum_{i=1}^d \sigma_i^2 x_i \left[\frac{2\widetilde{\xi}_i}{\boldsymbol{x} \cdot \widetilde{\boldsymbol{\xi}}} \frac{\partial f}{\partial x_i}(\boldsymbol{x}) + \frac{\partial^2 f}{\partial x_i^2}(\boldsymbol{x}) \right] + \sum_{i=1}^d x_i \sum_{j=1}^d \widetilde{c}_{ij} \left[\frac{\widetilde{\xi}_j}{\boldsymbol{x} \cdot \widetilde{\boldsymbol{\xi}}} f(\boldsymbol{x}) + \frac{\partial f}{\partial x_j}(\boldsymbol{x}) \right].$$

But $\sum_{i=1}^d x_i \sum_{j=1}^d \widetilde{c}_{ij} \widetilde{\xi}_j = \widetilde{\rho} \boldsymbol{x} \cdot \widetilde{\boldsymbol{\xi}}$, hence

$$(L^* f)(\boldsymbol{x}) = \frac{1}{2} \sum_{i=1}^d \sigma_i^2 x_i \frac{\partial^2 f}{\partial x_i^2}(\boldsymbol{x}) + \sum_{i=1}^d x_i \sum_{j=1}^d \left(\widetilde{c}_{ij} + \frac{\sigma_i^2 \widetilde{\xi}_i}{\boldsymbol{x} \cdot \widetilde{\boldsymbol{\xi}}} \delta_{ij} \right) \frac{\partial f}{\partial x_j}(\boldsymbol{x}). \tag{2.4.19}$$

The conditioned Feller diffusion process can thus be considered as a Feller diffusion with variance parameters $\sigma_1^2, \dots, \sigma_d^2$ and *state-dependent* mutation matrix $\mathbf{C}(\boldsymbol{x})$ defined for $\boldsymbol{x} \neq \mathbf{0}$ by

$$c_{ij}(\boldsymbol{x}) := \widetilde{c}_{ij} + \frac{\sigma_i^2 \widetilde{\xi}_i}{\boldsymbol{x} \cdot \widetilde{\boldsymbol{\xi}}} \delta_{ij}. \tag{2.4.20}$$

Associated martingale problem and SDE

Let us prove that the martingale problem $\mathcal{MP}\left(L^*, C_b^2\left(\mathbb{R}_+^d\backslash\{\mathbf{0}\}, \mathbb{R}\right), \boldsymbol{x}\right)$ admits as a unique solution the conditioned law $\mathbb{P}^*_{\boldsymbol{x}}$.

For this purpose we define the following subset of bounded C^2-functions on $\mathbb{R}_+^d\backslash\{\mathbf{0}\}$,

$$D_0\left(L^*\right) := \left\{\boldsymbol{x} \mapsto e^{-\boldsymbol{\lambda}\cdot\boldsymbol{x}},\ \boldsymbol{\lambda} \in \mathbb{R}_+^d\right\}.$$

Let $f(\boldsymbol{x}) := e^{-\boldsymbol{\lambda}\cdot\boldsymbol{x}}$, $\boldsymbol{\lambda} \in \mathbb{R}_+^d$. Then

$$\begin{aligned}
\left(L^* f\right)(\boldsymbol{x}) &= \left[\frac{1}{2}\sum_{i=1}^d \sigma_i^2 x_i \lambda_i^2 - \sum_{i=1}^d x_i \sum_{j=1}^d \left(\widetilde{c}_{ij} + \frac{\sigma_i^2 \widetilde{\xi}_i}{\boldsymbol{x}\cdot\widetilde{\boldsymbol{\xi}}}\delta_{ij}\right)\lambda_j\right] e^{-\boldsymbol{\lambda}\cdot\boldsymbol{x}} \\
&= -\left[\sum_{i=1}^d x_i \left(\sum_{j=1}^d \widetilde{c}_{ij}\lambda_j - \frac{1}{2}\sigma_i^2\lambda_i^2\right) + \frac{1}{\boldsymbol{x}\cdot\widetilde{\boldsymbol{\xi}}}\sum_{i=1}^d x_i \sigma_i^2 \widetilde{\xi}_i \lambda_i\right] e^{-\boldsymbol{\lambda}\cdot\boldsymbol{x}} \\
&= -\left[\boldsymbol{x}\cdot\widetilde{\boldsymbol{\psi}}(\boldsymbol{\lambda}) + \frac{1}{\boldsymbol{x}\cdot\widetilde{\boldsymbol{\xi}}}\left(\boldsymbol{\sigma}^2\boldsymbol{\lambda}\boldsymbol{x}\right)\cdot\widetilde{\boldsymbol{\xi}}\right] e^{-\boldsymbol{\lambda}\cdot\boldsymbol{x}},
\end{aligned}$$

where $\boldsymbol{\sigma}^2 := \left(\sigma_1^2, \ldots, \sigma_1^d\right)$, and $\widetilde{\boldsymbol{\psi}}$ is the branching mechanism of the (sub)critical Feller diffusion process with law $\widetilde{\mathbb{P}}$:

$$\widetilde{\psi}_i(\boldsymbol{\lambda}) := \sum_{j=1}^d \widetilde{c}_{ij}\lambda_j - \frac{1}{2}\sigma_i^2\lambda_i^2, \quad i = 1\ldots d. \tag{2.4.21}$$

Hence L^* maps $D_0\left(L^*\right)$ into bounded continuous functions on $\mathbb{R}_+^d\backslash\{\mathbf{0}\}$, and thus $D_0\left(L^*\right)$ is a core for L^*. As a consequence, in order to obtain the uniqueness of solution to the martingale problem $\mathcal{MP}\left(L^*, C_b^2\left(\mathbb{R}_+^d\backslash\{\mathbf{0}\}, \mathbb{R}\right), \boldsymbol{x}\right)$, it is enough to prove that $\mathcal{MP}\left(L^*, D_0\left(L^*\right), \boldsymbol{x}\right)$ admits a unique solution, *i.e.* that there exists one and only one $\mathbf{P}$ such that for all $\boldsymbol{\lambda} \in \mathbb{R}_+^d$,

$$e^{-\boldsymbol{\lambda}\cdot\mathbf{X}_t} - e^{-\boldsymbol{\lambda}\cdot\boldsymbol{x}} - \int_0^t \left(L^* f\right)\left(\mathbf{X}_s\right) ds \tag{2.4.22}$$

is a $(\mathbf{P}, \mathcal{F}_t)$-martingale, or equivalently, such that

$$e^{-\boldsymbol{\lambda}\cdot\mathbf{X}_t} - e^{-\boldsymbol{\lambda}\cdot\boldsymbol{x}} + \int_0^t \left(\mathbf{X}_s\cdot\widetilde{\boldsymbol{\psi}}(\boldsymbol{\lambda}) + \frac{\left(\mathbf{X}_s\boldsymbol{\sigma}^2\boldsymbol{\lambda}\right)\cdot\widetilde{\boldsymbol{\xi}}}{\mathbf{X}_s\cdot\boldsymbol{\xi}}\right) e^{-\boldsymbol{\lambda}\cdot\mathbf{X}_s} ds \tag{2.4.23}$$

is a $(\mathbf{P}, \mathcal{F}_t)$-martingale.

Applying the martingale problem $\mathcal{MP}\left(L, D_0\left(L\right), \boldsymbol{x}\right)$ to the function $f(t, \boldsymbol{x}) := e^{-\widetilde{\rho}t}\boldsymbol{x}\cdot\widetilde{\boldsymbol{\xi}}e^{-\boldsymbol{\lambda}\cdot\boldsymbol{x}}$, $\boldsymbol{\lambda} \in \mathbb{R}_+^d$, we know that

$$e^{-\widetilde{\rho}t}\mathbf{X}_t\cdot\widetilde{\boldsymbol{\xi}}e^{-\boldsymbol{\lambda}\cdot\mathbf{X}_t} - \boldsymbol{x}\cdot\widetilde{\boldsymbol{\xi}}e^{-\boldsymbol{\lambda}\cdot\boldsymbol{x}} + \int_0^t e^{-\widetilde{\rho}s}\mathbf{X}_s\cdot\widetilde{\boldsymbol{\xi}}\left(\mathbf{X}_s\cdot\widetilde{\boldsymbol{\psi}}(\boldsymbol{\lambda}) + \frac{\left(\mathbf{X}_s\boldsymbol{\sigma}^2\boldsymbol{\lambda}\right)\cdot\widetilde{\boldsymbol{\xi}}}{\mathbf{X}_s\cdot\widetilde{\boldsymbol{\xi}}}\right) e^{-\boldsymbol{\lambda}\cdot\mathbf{X}_s} ds \tag{2.4.24}$$

is a $(\mathbb{P}_{\boldsymbol{x}}, \mathcal{F}_t)$-martingale. Denoting $h(t, \boldsymbol{x}) := e^{-\widetilde{\rho}t}\boldsymbol{x}\cdot\widetilde{\boldsymbol{\xi}}$ as previously (see (2.4.18)), this becomes: for all $\boldsymbol{\lambda} \in \mathbb{R}_+^d$,

$$h(t, \mathbf{X}_t)e^{-\boldsymbol{\lambda}\cdot\mathbf{X}_t} - h(0, \boldsymbol{x})e^{-\boldsymbol{\lambda}\cdot\boldsymbol{x}} - \int_0^t h(s, \mathbf{X}_s)\left(\mathbf{X}_s\cdot\widetilde{\boldsymbol{\psi}}(\boldsymbol{\lambda}) + \frac{\left(\mathbf{X}_s\boldsymbol{\sigma}^2\boldsymbol{\lambda}\right)\cdot\widetilde{\boldsymbol{\xi}}}{\mathbf{X}_s\cdot\widetilde{\boldsymbol{\xi}}}\right) e^{-\boldsymbol{\lambda}\cdot\mathbf{X}_s} ds$$

is a $(\mathbb{P}_{\boldsymbol{x}}, \mathcal{F}_t)$-martingale, hence for all $s \leqslant t$,

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{x}}\left[h(t, \mathbf{X}_t)e^{-\boldsymbol{\lambda}\cdot\mathbf{X}_t}\middle|\mathcal{F}_s\right] - h(s, \mathbf{X}_s)e^{-\boldsymbol{\lambda}\cdot\mathbf{X}_s} \\
&\qquad - \int_s^t \mathbb{E}_{\boldsymbol{x}}\left[h(u, \mathbf{X}_u)\left(\mathbf{X}_u\cdot\widetilde{\boldsymbol{\psi}}(\boldsymbol{\lambda}) + \frac{\left(\mathbf{X}_u\boldsymbol{\sigma}^2\boldsymbol{\lambda}\right)\cdot\widetilde{\boldsymbol{\xi}}}{\mathbf{X}_u\cdot\widetilde{\boldsymbol{\xi}}}\right) e^{-\boldsymbol{\lambda}\cdot\mathbf{X}_u}\middle|\mathcal{F}_s\right] du = 0.
\end{aligned} \tag{2.4.25}$$

Furthermore, according to Theorem 2.4.1, $\mathbb{P}^*_{\boldsymbol{x}}$ is a Doob h-transform of $\mathbb{P}_{\boldsymbol{x}}$ *via* the space-time harmonic function h. As a consequence, for all $0 \leqslant s \leqslant t$ and any $\mathcal{F}_t$-measurable random variable Y_t, the following relation holds:

$$\mathbb{E}^*_{\boldsymbol{x}}\left[Y_t|\mathcal{F}_s\right] = \frac{1}{h(s,\mathbf{X}_s)}\mathbb{E}_{\boldsymbol{x}}\left[h(t,\mathbf{X}_t)Y_t|\mathcal{F}_s\right]. \tag{2.4.26}$$

Indeed, for all $B \in \mathcal{F}_s$,

$$\begin{aligned}
\mathbb{E}^*_{\boldsymbol{x}}\left[Y_t\mathbf{1}_B\right] &= \frac{1}{h(0,\boldsymbol{x})}\mathbb{E}_{\boldsymbol{x}}\left[h(t,\mathbf{X}_t)Y_t\mathbf{1}_B\right] \\
&= \frac{1}{h(0,\boldsymbol{x})}\mathbb{E}_{\boldsymbol{x}}\left[\mathbb{E}_{\boldsymbol{x}}\left[h(t,\mathbf{X}_t)Y_t|\mathcal{F}_s\right]\mathbf{1}_B\right] \\
&= \frac{1}{h(0,\boldsymbol{x})}\mathbb{E}_{\boldsymbol{x}}\left[h(s,\mathbf{X}_s)\frac{\mathbb{E}_{\boldsymbol{x}}\left[h(t,\mathbf{X}_t)Y_t|\mathcal{F}_s\right]}{h(s,\mathbf{X}_s)}\mathbf{1}_B\right] \\
&= \mathbb{E}^*_{\boldsymbol{x}}\left[\frac{\mathbb{E}_{\boldsymbol{x}}\left[h(t,\mathbf{X}_t)Y_t|\mathcal{F}_s\right]}{h(s,\mathbf{X}_s)}\mathbf{1}_B\right].
\end{aligned}$$

Applying relations (2.4.25) and (2.4.26) it comes that for all $0 \leqslant s \leqslant t$,

$$\begin{aligned}
&\mathbb{E}^*_{\boldsymbol{x}}\left[e^{-\boldsymbol{\lambda}\cdot\mathbf{X}_t} - e^{-\boldsymbol{\lambda}\cdot\mathbf{X}_s} - \int_s^t\left(\mathbf{X}_u\cdot\widetilde{\boldsymbol{\psi}}(\boldsymbol{\lambda}) + \frac{(\mathbf{X}_u\boldsymbol{\sigma}^2\boldsymbol{\lambda})\cdot\widetilde{\boldsymbol{\xi}}}{\mathbf{X}_u\cdot\widetilde{\boldsymbol{\xi}}}\right)e^{-\boldsymbol{\lambda}\cdot\mathbf{X}_u}du\Big|\mathcal{F}_s\right] \\
&= \mathbb{E}^*_{\boldsymbol{x}}\left[e^{-\boldsymbol{\lambda}\cdot\mathbf{X}_t}\Big|\mathcal{F}_s\right] - e^{-\boldsymbol{\lambda}\cdot\mathbf{X}_s} - \int_s^t\mathbb{E}^*\left[\left(\mathbf{X}_u\cdot\widetilde{\boldsymbol{\psi}}(\boldsymbol{\lambda}) + \frac{(\mathbf{X}_u\boldsymbol{\sigma}^2\boldsymbol{\lambda})\cdot\widetilde{\boldsymbol{\xi}}}{\mathbf{X}_u\cdot\widetilde{\boldsymbol{\xi}}}\right)e^{-\boldsymbol{\lambda}\cdot\mathbf{X}_u}\Big|\mathcal{F}_s\right]du \\
&= \frac{1}{h(s,\mathbf{X}_s)}\Bigg[\mathbb{E}_{\boldsymbol{x}}\left[h(t,\mathbf{X}_t)e^{-\boldsymbol{\lambda}\cdot\mathbf{X}_t}\Big|\mathcal{F}_s\right] - h(s,\mathbf{X}_s)e^{-\boldsymbol{\lambda}\cdot\mathbf{X}_s} \\
&\qquad - \int_s^t\mathbb{E}_{\boldsymbol{x}}\left[h(u,\mathbf{X}_u)\left(\mathbf{X}_u\cdot\widetilde{\boldsymbol{\psi}}(\boldsymbol{\lambda}) + \frac{(\mathbf{X}_u\boldsymbol{\sigma}^2\boldsymbol{\lambda})\cdot\widetilde{\boldsymbol{\xi}}}{\mathbf{X}_u\cdot\widetilde{\boldsymbol{\xi}}}\right)e^{-\boldsymbol{\lambda}\cdot\mathbf{X}_u}\Big|\mathcal{F}_s\right]du\Bigg] \\
&= 0,
\end{aligned}$$

hence for all $\boldsymbol{\lambda} \in \mathbb{R}^d_+$, (2.4.23) is a $(\mathbb{P}^*_{\boldsymbol{x}}, \mathcal{F}_t)$-martingale for all $\boldsymbol{\lambda} \in \mathbb{R}^d_+$. The law $\mathbb{P}^*_{\boldsymbol{x}}$ is thus a solution to $\mathcal{MP}\left(L^*, D_0\left(L^*\right), \boldsymbol{x}\right)$.

Let us now assume that there exists an other distribution $\widehat{\mathbb{P}}^*_{\boldsymbol{x}}$ such that for all $\boldsymbol{\lambda} \in \mathbb{R}^d_+$, (2.4.23) is a $(\widehat{\mathbb{P}}^*_{\boldsymbol{x}}, \mathcal{F}_t)$-martingale, and let us define $\widehat{\mathbb{P}}_{\boldsymbol{x}}$ such that

$$d\widehat{\mathbb{P}}_{\boldsymbol{x}}|_{\mathcal{F}_t} := \frac{h(0,\boldsymbol{x})}{h(t,\mathbf{X}_t)}d\widehat{\mathbb{P}}^*_{\boldsymbol{x}}|_{\mathcal{F}_t}.$$

Then, according to the previous computation, for all $\boldsymbol{\lambda} \in \mathbb{R}^d_+$, (2.4.24) is a $(\widehat{\mathbb{P}}_{\boldsymbol{x}}, \mathcal{F}_t)$-martingale. By uniqueness of the solution to the martingale problem $\mathcal{MP}(L)$, this implies $\widehat{\mathbb{P}}_{\boldsymbol{x}} = \mathbb{P}_{\boldsymbol{x}}$, and thus $\widehat{\mathbb{P}}^*_{\boldsymbol{x}} = \mathbb{P}^*_{\boldsymbol{x}}$.

We just proved that $\mathbb{P}^*$ is the unique solution to the martingale problem

$$\mathcal{MP}\left(L^*, C^2_b(\mathbb{R}^d_+\backslash\{\mathbf{0}\}, \mathbb{R}), \boldsymbol{x}\right).$$

As a consequence, the Q-process with law $\mathbb{P}^*$ is the unique weak solution of the stochastic differential equation

$$d\mathbf{X}_t = \boldsymbol{\Sigma}(\mathbf{X}_t)\,d\mathbf{B}_t + \mathbf{C}^T\mathbf{X}_t\,dt + \frac{1}{\mathbf{X}_t\cdot\widetilde{\boldsymbol{\xi}}}\boldsymbol{\sigma}^2\mathbf{X}_t\widetilde{\boldsymbol{\xi}}\,dt, \tag{2.4.27}$$

where $\boldsymbol{\Sigma}$ is defined in (1.2.17). Componentwise, (2.4.27) becomes, for all $i = 1 \dots d$,

$$dX_{t,i} = \sigma_i \sqrt{X_{t,i}} dB_{t,i} + \sum_{j=1}^{d} \widetilde{c}_{ji} X_{t,j} dt + \frac{1}{\mathbf{X}_t \cdot \widetilde{\boldsymbol{\xi}}} \sigma_i^2 \widetilde{\xi}_i X_{t,i} dt. \tag{2.4.28}$$

The equation solved by the conditioned process thus has an additional non-linear term compared with the equation solved by the non-conditioned process. We show in the following paragraph that this term corresponds to a state-dependent immigration.

The conditioned Feller diffusion process as process with state-dependent immigration

Monotype case. If $d = 1$, the previous SDE becomes

$$dX_t = \sigma \sqrt{X_t} dB_t + \widetilde{c} X_t dt + \sigma^2 dt. \tag{2.4.29}$$

The additional deterministic term $\sigma^2 dt$ suggests that the conditioned process can be interpreted as a process with immigration. Let us prove this result. Applying Theorem 2.4.6, we obtain that

$$\mathbb{E}_x^* \left(e^{-\lambda X_t}\right) = e^{-\widetilde{c}t} \frac{1}{x} \widetilde{\mathbb{E}}_x \left(X_t e^{-\lambda X_t}\right) = e^{-\widetilde{c}t} \frac{1}{x} \frac{\partial}{\partial \lambda} \left[-e^{-x\widetilde{u}_t(\lambda)}\right] = e^{-\widetilde{c}t} \frac{\partial \widetilde{u}_t(\lambda)}{\partial \lambda} e^{-x\widetilde{u}_t(\lambda)}.$$

Denoting $v_t(\lambda) := \frac{\partial \widetilde{u}_t(\lambda)}{\partial \lambda}$, we thus have

$$\mathbb{E}_x^* \left(e^{-\lambda X_t}\right) = e^{-\widetilde{c}t} v_t(\lambda) \widetilde{\mathbb{E}}_x \left(e^{-\lambda X_t}\right), \tag{2.4.30}$$

where $v_t(\lambda)$ is the unique solution of (we use here (1.2.6)),

$$\begin{cases} \dfrac{\partial}{\partial t} v_t(\lambda) = \widetilde{c} v_t(\lambda) - \sigma^2 v_t(\lambda) \widetilde{u}_t(\lambda), \\ \quad v_0(\lambda) = 1. \end{cases} \tag{2.4.31}$$

Then (2.4.30) becomes

$$\begin{aligned} \mathbb{E}_x^* \left(e^{-\lambda X_t}\right) &= \exp\left[-\widetilde{c}t + \ln\left(v_t(\lambda)\right)\right] \widetilde{\mathbb{E}}_x \left(e^{-\lambda X_t}\right) \\ &= \exp\left[\int_0^t \left(-\widetilde{c} + \frac{\frac{\partial}{\partial s} v_s(\lambda)}{v_s(\lambda)}\right) du\right] \widetilde{\mathbb{E}}_x \left(e^{-\lambda X_t}\right) \\ &= \exp\left[-\int_0^t \sigma^2 \widetilde{u}_s(\lambda) ds\right] \widetilde{\mathbb{E}}_x \left(e^{-\lambda X_t}\right). \end{aligned} \tag{2.4.32}$$

We recognize here that the Laplace transform of the Q-process associated with the monotype Feller diffusion process is the Laplace transform of a Feller diffusion process with immigration.

Indeed, as described by Lambert in [Lamb07], a Feller diffusion process (and more generally a CB process) with immigration is a strong Markov process with Laplace transform

$$\exp\left[-x u_t(\lambda) - \int_0^t \chi\left(u_s(\lambda)\right) ds\right], \ x, \lambda \in \mathbb{R}_+ \tag{2.4.33}$$

where $u_t(\lambda)$ is the cumulant of the original CB process, and χ is the Laplace exponent of the subordinator describing the immigration. The equivalent of this subordinator for BGW processes would be the number of immigrants until generation n. The author provides a general result for monotype CB processes (Theorem 4.1 in [Lamb07]), stating that the Q-process associated with a (sub)critical CB process with branching mechanism $\widetilde{\psi}$ is a CB process with immigration, the latter being given by $\chi(\lambda) = \widetilde{\psi}'(0) - \widetilde{\psi}'(\lambda)$. Applying this to $\widetilde{\psi}(\lambda) = \widetilde{c}\lambda - \frac{1}{2}\sigma^2\lambda^2$ we obtain $\chi(\lambda) = \sigma^2\lambda$, and thus our result (2.4.32) coincides with (2.4.33).

Multitype case. We can find a generalization of (2.4.30) to the multitype case in [ChaRoe08]. It is there proved that the Laplace transform of the Q-process is given by

$$\mathbb{E}^*_{\boldsymbol{x}}\left[e^{-\boldsymbol{\lambda}\cdot\mathbf{X}_t}\right] = e^{-\rho t}\frac{\boldsymbol{x}\cdot\boldsymbol{v}_t(\boldsymbol{\lambda})}{\boldsymbol{x}\cdot\boldsymbol{\xi}}\mathbb{E}_{\boldsymbol{x}}\left[e^{-\boldsymbol{\lambda}\cdot\mathbf{X}_t}\right],$$

where $\boldsymbol{v}_t(\boldsymbol{\lambda})$ is the unique solution of the differential system

$$\begin{cases} \dfrac{\partial}{\partial t}v_{t,i}(\boldsymbol{\lambda}) = \displaystyle\sum_{j=1}^{d}\widetilde{c}_{ij}v_{t,j}(\boldsymbol{\lambda}) - \sigma_i^2 u_{t,i}(\boldsymbol{\lambda})v_{t,i}(\boldsymbol{\lambda}), \quad i = 1\dots d, \\ v_{0,i}(\boldsymbol{\lambda}) = \xi_i. \end{cases} \tag{2.4.34}$$

The multiplicative term $e^{-\rho t}\frac{\boldsymbol{x}\cdot\boldsymbol{v}_t(\boldsymbol{\lambda})}{\boldsymbol{x}\cdot\boldsymbol{\xi}}$ appearing in the Laplace transform corresponds to the immigration in the Q-process, which is confirmed by the form of the martingale problem (2.4.23) solved by $\mathbb{P}^*$: the Q-process is a multitype Feller diffusion process with immigration whose rate at time t, if conditioned by $\mathbf{X}_t$, is a random variable with Laplace transform

$$\exp\left[-\frac{\left(\mathbf{X}_t\boldsymbol{\sigma}^2\boldsymbol{\lambda}\right)\cdot\widetilde{\boldsymbol{\xi}}}{\mathbf{X}_t\cdot\widetilde{\boldsymbol{\xi}}}\right].$$

Chapter 3

Commutativity results for the conditioned processes

The whole previous chapter has been dedicated to multitype BGWc and Feller diffusion processes conditioned on a delayed extinction. We investigated both limits as the time t tends to infinity (Yaglom-type limits, Section 2.3) and as the delay of extinction θ tends to infinity (Q-process, Section 2.4). It is very natural to wonder whether one would obtain an interesting mathematical object by taking first one limit and then the other one, and if these two limit procedures could be interchanged.

Moreover, in the last chapter we observed a parallel behavior for BGWc processes and their continuous counterparts, the Feller diffusion processes: compare for example Proposition 2.3.1 with Proposition 2.3.5, or Theorem 2.4.1 with Theorem 2.4.6. Since a Feller diffusion process can be obtained as a high-density limit of a BGWc process (once rescaled appropriately in time and space with some scaling parameter n, as detailed in **Subsection 3.2.1**), we will rigorously justify these similarities by showing that the random objects associated with the Feller diffusion process (Yaglom-type limits and Q-process) can also be obtained as a limit of the same objects associated with the BGWc process.

More generally, the purpose of this section is to show the commutativity between the three possible limits in n, t and θ of

$$P(n,t,\theta) := \mathbb{P}^n_{\frac{1}{n}\boldsymbol{x}^n}\left(\mathbf{X}_t \in \,.\,|\, \mathbf{X}_{t+\theta} \neq \mathbf{0}, \lim_{s\to\infty} \mathbf{X}_s = \mathbf{0}\right), \tag{3.0.1}$$

where $\mathbb{P}^n$ denotes the law of the BGWc process rescaled with the scaling parameter n defined in **Subsection 3.2.1**, and $\boldsymbol{x}^n \in \mathbb{N}^d$ is such that $\lim_{n\to\infty} \frac{1}{n}\boldsymbol{x}^n = \boldsymbol{x} \in \mathbb{R}^d_+\backslash\{\mathbf{0}\}$.

In **Section 3.1** we first focus on the interchangeability of the limits in t and θ, for both the BGWc and Feller diffusion process, in order to obtain an equality of the form (BGWc process)

$$\lim_n \boldsymbol{\lim_\theta} \boldsymbol{\lim_t} P(n,t,\theta) = \lim_n \boldsymbol{\lim_t} \boldsymbol{\lim_\theta} P(n,t,\theta), \tag{3.0.2}$$

and (Feller diffusion process)

$$\boldsymbol{\lim_\theta} \boldsymbol{\lim_t} \lim_n P(n,t,\theta) = \boldsymbol{\lim_t} \boldsymbol{\lim_\theta} \lim_n P(n,t,\theta). \tag{3.0.3}$$

In **Subsection 3.2.3** we prove the following intuitive relation between Yaglom-type limits of a BGWc and a Feller diffusion process,

$$\lim_\theta \boldsymbol{\lim_n} \boldsymbol{\lim_t} P(n,t,\theta) = \lim_\theta \boldsymbol{\lim_t} \boldsymbol{\lim_n} P(n,t,\theta). \tag{3.0.4}$$

Next, in **Subsection 3.2.4**, we prove that the diffusion limit of the Q-process is the Q-process of the diffusion limit, which means that

$$\lim_t \boldsymbol{\lim_\theta} \boldsymbol{\lim_n} P(n,t,\theta) = \lim_t \boldsymbol{\lim_n} \boldsymbol{\lim_\theta} P(n,t,\theta). \tag{3.0.5}$$

Finally we prove in **Subsection 3.2.5** that

$$\lim_{\boldsymbol{n}} \lim_{\boldsymbol{t}} \lim_{\theta} P(n,t,\theta) = \lim_{\boldsymbol{t}} \lim_{\boldsymbol{n}} \lim_{\theta} P(n,t,\theta), \tag{3.0.6}$$

or in other words that the asymptotic behavior of the Q-process associated with the n-rescaled BGWc process converges as n tends to infinity to the one of the Q-process associated with the Feller diffusion process.

3.1 Commutativity of the long-time limits

In this section we are interested in the long-time behavior of the Q-process associated with the BGWc (resp. Feller diffusion process). We show in Proposition 3.1.1 (resp. Proposition 3.1.4) that the long-time limit is nondegenerate in the subcritical and supercritical cases, and that it is a probability distribution independent of the initial condition. By definition of the Q-process, this limit is obtained by letting first θ and then t tend to infinity in the law of $\mathbf{X}_t$ conditioned on extinction delayed by at least θ. We have asked ourselves whether the order of those two limits in t and θ can be exchanged. Proposition 3.1.1 (resp. Proposition 3.1.4) provides an affirmative answer: the nondegenerate limit mentioned above can also be obtained by letting first t and then θ tend to infinity, *i.e.* as the limit as θ tends to infinity of the so-called θ-Yaglom limit introduced in Subsection 2.3.1 (resp. Subsection 2.3.2).

3.1.1 Long-time limits of the conditioned BGWc process

Proposition 3.1.1. *Let us assume **(B1)** and **(B2)**. We assume moreover that*

(i) if $\rho = 0$, all the second order moments of the offspring distribution are finite,

(ii) if $\rho < 0$, for all $i, j = 1 \dots d$, $\sum_{\boldsymbol{k} \in \mathbb{N}^d} k_j \ln(k_j) p_i(\boldsymbol{k}) < \infty$.

Then the following holds.

- *In the critical case, the Q-process associated with the BGWc process explodes as t tends to infinity, i.e. for all $\boldsymbol{x} \in \mathbb{N}^d \backslash \{\boldsymbol{0}\}$ and $\boldsymbol{u} \geqslant \boldsymbol{0}$,*

$$\lim_{t\to\infty} \lim_{\theta\to\infty} \mathbb{P}_{\boldsymbol{x}} \left(\boldsymbol{X}_t \leqslant \boldsymbol{u} \,|\, \boldsymbol{X}_{t+\theta} \neq \boldsymbol{0} \right) = 0. \tag{3.1.1}$$

Furthermore, one can interchange both limits in t and θ:

$$\lim_{t\to\infty} \lim_{\theta\to\infty} \mathbb{P}_{\boldsymbol{x}} \left(\boldsymbol{X}_t \leqslant \boldsymbol{u} \,|\, \boldsymbol{X}_{t+\theta} \neq \boldsymbol{0} \right) = \lim_{\theta\to\infty} \lim_{t\to\infty} \mathbb{P}_{\boldsymbol{x}} \left(\boldsymbol{X}_t \leqslant \boldsymbol{u} \,|\, \boldsymbol{X}_{t+\theta} \neq \boldsymbol{0} \right) = 0. \tag{3.1.2}$$

- *In the noncritical case, the Q-process converges as t tends to infinity to a nontrivial limit which does not depend on the initial condition $\boldsymbol{x} \in \mathbb{N}^d \backslash \{\boldsymbol{0}\}$, and which corresponds to the size-biased Yaglom distribution (see (3.1.7)). Furthermore, one can interchange both limits in t and θ:*

$$\begin{aligned} &\lim_{t\to\infty} \lim_{\theta\to\infty} \mathbb{P}_{\boldsymbol{x}} \left(\boldsymbol{X}_t \in . | \boldsymbol{X}_{t+\theta} \neq \boldsymbol{0}, \lim_{s\to\infty} \boldsymbol{X}_s = \boldsymbol{0} \right) \\ &\qquad = \lim_{\theta\to\infty} \lim_{t\to\infty} \mathbb{P}_{\boldsymbol{x}} \left(\boldsymbol{X}_t \in . | \boldsymbol{X}_{t+\theta} \neq \boldsymbol{0}, \lim_{s\to\infty} \boldsymbol{X}_s = \boldsymbol{0} \right). \end{aligned} \tag{3.1.3}$$

Proof. We first assume that $\rho \neq 0$ and focus on the right term of (3.1.3). By (2.3.6) and (2.3.8) we have, for all $\mathbf{r} \in [0,1]^d$ and all $\boldsymbol{x} \in \mathbb{N}^d$, $\boldsymbol{x} \neq \mathbf{0}$,

$$\begin{aligned} \lim_{\theta\to\infty} \lim_{t\to\infty} \widetilde{\mathbb{E}}_{\boldsymbol{x}} \left[\mathbf{r}^{\mathbf{X}_t} \,|\, \mathbf{X}_{t+\theta} \neq \mathbf{0} \right] &= \lim_{\theta\to\infty} \frac{\gamma\left(\mathbf{r} - e^{\widetilde{\rho}\theta} \gamma(\mathbf{0}) \widetilde{\boldsymbol{\xi}} \right) - \gamma(\mathbf{r})}{e^{\widetilde{\rho}\theta} \gamma(\mathbf{0})} \\ &= -\sum_{i=1}^{d} r_i \widetilde{\xi}_i \frac{\partial \gamma(\mathbf{r})}{\partial r_i}, \end{aligned} \tag{3.1.4}$$

the differentiability of γ stemming from (2.3.9).

Let us focus on the left term of (3.1.3). Using (2.4.8) we obtain

$$\begin{aligned}
\lim_{\theta\to\infty} \widetilde{\mathbb{E}}_{\boldsymbol{x}}\left[\mathbf{r}^{\mathbf{X}_t} \,|\, \mathbf{X}_{t+\theta}\neq \mathbf{0}\right] &= e^{-\widetilde{\rho}t}\frac{1}{\boldsymbol{x}\cdot\widetilde{\boldsymbol{\xi}}}\widetilde{\mathbb{E}}_{\boldsymbol{x}}\left[\mathbf{X}_t\cdot\widetilde{\boldsymbol{\xi}}\mathbf{r}^{\mathbf{X}_t}\right] \\
&= e^{-\widetilde{\rho}t}\frac{1}{\boldsymbol{x}\cdot\widetilde{\boldsymbol{\xi}}}\sum_{i=1}^{d} r_i\widetilde{\xi}_i\frac{\partial}{\partial r_i}\left[\widetilde{\mathbf{F}}_t(\mathbf{r})^{\boldsymbol{x}}\right] \\
&= e^{-\widetilde{\rho}t}\frac{1}{\boldsymbol{x}\cdot\widetilde{\boldsymbol{\xi}}}\sum_{i=1}^{d} r_i\widetilde{\xi}_i\sum_{j=1}^{d} x_j\frac{\partial}{\partial r_i}\left[\ln\widetilde{F}_{t,j}(\mathbf{r})\right]\widetilde{\mathbf{F}}_t(\mathbf{r})^{\boldsymbol{x}}. \qquad (3.1.5)
\end{aligned}$$

But for all $i,j=1\dots d$ and all $\mathbf{r}\in[0,1]^d$ such that $r_i>0$ we have (see Lemma 3.1.3 below)

$$\lim_{t\to\infty} e^{-\widetilde{\rho}t}\frac{\partial}{\partial r_i}\left[\ln\widetilde{F}_{t,j}(\mathbf{r})\right] = -\frac{\partial\gamma(\mathbf{r})}{\partial r_i}\widetilde{\xi}_j,$$

which together with (3.1.4) and (3.1.5) leads to the equality

$$\lim_{\theta\to\infty}\lim_{t\to\infty}\widetilde{\mathbb{E}}_{\boldsymbol{x}}\left[\mathbf{r}^{\mathbf{X}_t} \,|\, \mathbf{X}_{t+\theta}\neq\mathbf{0}\right] = \lim_{t\to\infty}\lim_{\theta\to\infty}\widetilde{\mathbb{E}}_{\boldsymbol{x}}\left[\mathbf{r}^{\mathbf{X}_t} \,|\, \mathbf{X}_{t+\theta}\neq\mathbf{0}\right],$$

for all $\mathbf{r}\in[0,1]^d$.

Now from [Sew75] Satz 6.2.8 we know that under assumption *(ii)* of Theorem 2.4.1, the generating function F^0 is differentiable in $\mathbf{r}=\mathbf{1}$, and its derivative satisfies, for all $i=1\dots d$,

$$\frac{\partial F^0(\mathbf{1})}{\partial r_i} = \frac{\widetilde{\eta}_i}{\gamma(\mathbf{0})}. \qquad (3.1.6)$$

Using Lebesgue's dominated convergence theorem together with the fact that (3.1.6) is finite, it comes that $\frac{\partial F^0(\mathbf{r})}{\partial r_i}$ is continuous in $\mathbf{1}$, which thanks to (2.3.9) implies the continuity in $\mathbf{r}=\mathbf{1}$ of the right term of (3.1.4). Lévy's continuity theorem thus ensures the convergence law.

Let us finally show that the probability distribution π obtained in (3.1.3) is not trivially reduced to the Dirac measure in $\mathbf{0}$, and corresponds to the size-biased Yaglom distribution. By this we mean that, denoting by Υ the Yaglom distribution, we have for all $\mathbf{k}\in\mathbb{N}^d$,

$$\pi(\mathbf{k}) = \frac{1}{\sum_{\mathbf{i}\in\mathbb{N}^d}\mathbf{i}\cdot\widetilde{\boldsymbol{\xi}}\Upsilon(\mathbf{i})}\mathbf{k}\cdot\widetilde{\boldsymbol{\xi}}\,\Upsilon(\mathbf{k}). \qquad (3.1.7)$$

Indeed, using (2.3.9) and (3.1.4), the generating function of π is given for all $\mathbf{r}\in[0,1]^d$ by

$$-\sum_{i=1}^{d} r_i\widetilde{\xi}_i\frac{\partial\gamma(\mathbf{r})}{\partial r_i} = \gamma(\mathbf{0})\sum_{i=1}^{d} r_i\widetilde{\xi}_i\frac{\partial F_0(\mathbf{r})}{\partial r_i} = \gamma(\mathbf{0})\sum_{\mathbf{k}\in\mathbb{N}^d\setminus\{\mathbf{0}\}}\mathbf{k}\cdot\widetilde{\boldsymbol{\xi}}\,\Upsilon(\mathbf{k})\,\mathbf{r}^{\mathbf{k}}. \qquad (3.1.8)$$

and we know by (3.1.6) that

$$\sum_{\mathbf{i}\in\mathbb{N}^d}\mathbf{i}\cdot\widetilde{\boldsymbol{\xi}}\Upsilon(\mathbf{i}) = \sum_{j=1}^{d}\widetilde{\xi}_j\frac{\partial F^0(\mathbf{1})}{\partial r_j} = \frac{\widetilde{\boldsymbol{\eta}}\cdot\widetilde{\boldsymbol{\xi}}}{\gamma(\mathbf{0})} = \frac{1}{\gamma(\mathbf{0})}.$$

We now consider the critical case $\rho=0$. We deduce from Proposition 2.3.4 that for all $\theta\geqslant 0$, $\mathbf{r}\in C_d$, $\mathbf{r}\neq\mathbf{1}$, $\lim_{t\to\infty}\mathbb{E}_{\boldsymbol{x}}\left[\mathbf{r}^{\mathbf{X}_t} \,|\, \mathbf{X}_{t+\theta}\neq\mathbf{0}\right]=0$. On the other hand,

$$\begin{aligned}
\lim_{\theta\to\infty}\mathbb{E}_{\boldsymbol{x}}\left[\mathbf{r}^{\mathbf{X}_t} \,|\, \mathbf{X}_{t+\theta}\neq\mathbf{0}\right] &= \frac{1}{\boldsymbol{x}\cdot\boldsymbol{\xi}}\mathbb{E}_{\boldsymbol{x}}\left[\mathbf{X}_t\cdot\boldsymbol{\xi}\mathbf{r}^{\mathbf{X}_t}\right] \\
&= \frac{1}{\boldsymbol{x}\cdot\boldsymbol{\xi}}\sum_{i=1}^{d} r_i\xi_i\frac{\partial}{\partial r_i}\left[\mathbf{F}_t(\mathbf{r})^{\boldsymbol{x}}\right] \\
&= \frac{1}{\boldsymbol{x}\cdot\boldsymbol{\xi}}\sum_{i=1}^{d} r_i\xi_i\sum_{j=1}^{d} x_j\frac{1}{F_{t,j}(\mathbf{r})}\frac{\partial}{\partial r_i}\left[F_{t,j}(\mathbf{r})\right]\mathbf{F}_t(\mathbf{r})^{\boldsymbol{x}}. \qquad (3.1.9)
\end{aligned}$$

In the critical case, $\lim_{t\to\infty} F_{t,j}(\mathbf{r}) = 1$, and we deduce from (1.1.8) and (1.1.9) that

$$\lim_{t\to\infty}\left[\sum_{j=1}^{d}\alpha_j\left[f_j(\mathbf{r})-r_j\right]\frac{\partial}{\partial r_j}F_{t,i}(\mathbf{r})\right] = \lim_{t\to\infty}\alpha_i\left[f_i\left(\mathbf{F}_t(\mathbf{r})\right)-F_{t,i}(\mathbf{r})\right] = \alpha_i\left[f_i(\mathbf{1})-1\right]=0.$$

Since all the terms of the left member are nonnegative, this implies that

$$\lim_{t\to\infty}\frac{\partial}{\partial r_j}F_{t,i}(\mathbf{r}) = 0.$$

Using this in (3.1.9), we finally obtain

$$\lim_{t\to\infty}\lim_{\theta\to\infty}\mathbb{E}_{\boldsymbol{x}}\left[\mathbf{r}^{\mathbf{X}_t}\,|\,\mathbf{X}_{t+\theta}\neq\mathbf{0}\right]=0.$$

□

Remark 3.1.2. It appears thanks to (3.1.8) that the probability distribution defined by (3.1.3) is reduced to a Dirac measure on $\boldsymbol{x}$ for some $\boldsymbol{x}\in\mathbb{N}^d$, if and only if the Yaglom measure is reduced to the same Dirac measure. Indeed, if $\Upsilon(\boldsymbol{x})=1$, then, by (3.1.6)

$$\frac{\widetilde{\eta}_i}{\gamma(\mathbf{0})} = \frac{\partial F^0(\mathbf{1})}{\partial r_i} = x_i, \tag{3.1.10}$$

which implies that $\gamma(\mathbf{0})\,\boldsymbol{x}\cdot\widetilde{\boldsymbol{\xi}}=1$, and thus the probability distribution has generating function

$$\gamma(\mathbf{0})\sum_{i=1}^{d}\widetilde{\xi}_i\sum_{\mathbf{k}\in\mathbb{N}^d\setminus\{\mathbf{0}\}}\Upsilon(\mathbf{k})k_i\mathbf{r}^{\mathbf{k}} = \gamma(\mathbf{0})\,\boldsymbol{x}\cdot\widetilde{\boldsymbol{\xi}}\,\mathbf{r}^{\boldsymbol{x}} = \mathbf{r}^{\boldsymbol{x}}. \tag{3.1.11}$$

Conversely, if $\gamma(\mathbf{0})\sum_{i=1}^{d}\widetilde{\xi}_i\sum_{\mathbf{k}\in\mathbb{N}^d\setminus\{\mathbf{0}\}}\Upsilon(\mathbf{k})k_i\mathbf{r}^{\mathbf{k}}=\mathbf{r}^{\boldsymbol{x}}$, then we necessarily have $\Upsilon(\boldsymbol{x})=1$.

Let us prove the following technical lemma, needed in the proof of Proposition 3.1.1.

Lemma 3.1.3. *For all $i,j=1\dots d$ and all $\boldsymbol{r}\in[0,1]^d$, $\boldsymbol{r}\neq\boldsymbol{1}$, such that $r_i>0$,*

$$\lim_{t\to\infty}e^{-\widetilde{\rho}t}\frac{\partial}{\partial r_i}\left[\ln\widetilde{F}_{t,j}(\boldsymbol{r})\right] = -\frac{\partial\gamma(\boldsymbol{r})}{\partial r_i}\widetilde{\xi}_j. \tag{3.1.12}$$

Proof. Deducing from (2.3.6) that $\lim_{t\to\infty}e^{-\widetilde{\rho}t}\ln\widetilde{F}_{t,j}(\mathbf{r})=-\gamma(\mathbf{r})\widetilde{\xi}_j$, we obtain that for all $h\geqslant 0$,

$$\gamma(\mathbf{r})\widetilde{\xi}_j-\gamma(\mathbf{r}+h\mathbf{e}_i)\widetilde{\xi}_j = \lim_{t\to\infty}\int_0^h e^{-\widetilde{\rho}t}\frac{\partial}{\partial r_i}\left[\ln\widetilde{F}_{t,j}(\mathbf{r}+u\mathbf{e}_i)\right]du. \tag{3.1.13}$$

Moreover,

$$0\leqslant e^{-\widetilde{\rho}t}\frac{\partial}{\partial r_i}\left[\ln\widetilde{F}_{t,j}(\mathbf{r})\right]\leqslant e^{-\widetilde{\rho}t}\frac{1}{\widetilde{F}_{t,j}(\mathbf{r})}\frac{1}{r_i}\widetilde{\mathbb{E}}_{\mathbf{e}_j}\left[X_{t,i}\right],$$

and by Proposition 1.1.14, $\lim_{t\to\infty}e^{-\widetilde{\rho}t}\widetilde{\mathbb{E}}_{\mathbf{e}_j}\left[X_{t,i}\right]=\widetilde{\xi}_j\widetilde{\eta}_i$. We then have by the continuity of $\widetilde{F}_{t,j}$ in $\mathbf{r}$ the existence of a constant $C>0$ such that for all $t\geqslant 0$ and all $u\in[0,h]$,

$$\left|e^{-\rho t}\frac{\partial}{\partial r_i}\left[\ln\widetilde{F}_{t,j}(\mathbf{r}+u\mathbf{e}_i)\right]\right|\leqslant\frac{C}{r_i+u}.$$

Using this upper bound integrable on $u\in[0,h]$ together with Lebesgue's dominated convergence theorem, (3.1.13) leads to

$$\gamma(\mathbf{r})\widetilde{\xi}_j-\gamma(\mathbf{r}+h\mathbf{e}_i)\widetilde{\xi}_j=\int_0^h\limsup_{t\to\infty}\left[e^{-\widetilde{\rho}t}\frac{\partial}{\partial r_i}\left(\ln\widetilde{F}_{t,j}(\mathbf{r}+u\mathbf{e}_i)\right)\right]du,$$

and thus $\limsup_{t\to\infty}e^{-\widetilde{\rho}t}\frac{\partial}{\partial r_i}\left[\ln\widetilde{F}_{t,j}(\mathbf{r})\right]=-\frac{\partial\gamma(\mathbf{r})}{\partial r_i}\widetilde{\xi}_j$. Proving the same way the result for the limit inferior we finally obtain (3.1.12). □

3.1.2 Long-time limits of the Feller diffusion process

In this section, we generalize a result from Champagnat and Rœlly (Theorem 3.6 and Theorem 3.7, [ChaRoe08]) which is the analog of Proposition 3.1.1 for multitype Feller diffusion processes. It proves the interchangeability of the long-time limits in t and θ for (sub)critical processes, and we generalize this result by including the supercritical case as well. The generalization is straight forward thanks to Proposition 2.2.4, and we thus omit the proof.

Proposition 3.1.4. *Let us assume **(F1)** and **(F2)**. Then the following holds.*

- *In the critical case, the Q-process associated with the Feller process explodes as t tends to infinity, i.e. for all $\boldsymbol{x} \in \mathbb{R}_+^d \backslash \{\boldsymbol{0}\}$ and $\boldsymbol{u} \geqslant \boldsymbol{0}$,*

$$\lim_{t\to\infty}\lim_{\theta\to\infty} \mathbb{P}_{\boldsymbol{x}}\left(\boldsymbol{X}_t \leqslant \boldsymbol{u} \,|\, \boldsymbol{X}_{t+\theta} \neq \boldsymbol{0}\right) = 0. \tag{3.1.14}$$

Furthermore, one can interchange both limits in t and θ:

$$\lim_{t\to\infty}\lim_{\theta\to\infty} \mathbb{P}_{\boldsymbol{x}}\left(\boldsymbol{X}_t \leqslant \boldsymbol{u} \,|\, \boldsymbol{X}_{t+\theta} \neq \boldsymbol{0}\right) = \lim_{\theta\to\infty}\lim_{t\to\infty} \mathbb{P}_{\boldsymbol{x}}\left(\boldsymbol{X}_t \leqslant \boldsymbol{u} \,|\, \boldsymbol{X}_{t+\theta} \neq \boldsymbol{0}\right) = 0. \tag{3.1.15}$$

- *In the noncritical case, the Q-process converges as t tends to infinity to a nontrivial limit which does not depend on the initial condition $\boldsymbol{x} \in \mathbb{R}_+^d \backslash \{\boldsymbol{0}\}$. Furthermore, one can interchange both limits in t and θ:*

$$\begin{aligned}\lim_{t\to\infty}\lim_{\theta\to\infty} \mathbb{P}_{\boldsymbol{x}}&\left(\boldsymbol{X}_t \in .|\boldsymbol{X}_{t+\theta} \neq \boldsymbol{0}, \lim_{s\to\infty}\boldsymbol{X}_s = \boldsymbol{0}\right)\\ &= \lim_{\theta\to\infty}\lim_{t\to\infty} \mathbb{P}_{\boldsymbol{x}}\left(\boldsymbol{X}_t \in .|\boldsymbol{X}_{t+\theta} \neq \boldsymbol{0}, \lim_{s\to\infty}\boldsymbol{X}_s = \boldsymbol{0}\right).\end{aligned} \tag{3.1.16}$$

Remark 3.1.5. In the monotype case, this nontrivial limit is known to be a Gamma distribution (see *e.g.* [ChaRoe08] Proposition 3.3), with parameter 2 and $\frac{2|\tilde{\rho}|}{\sigma^2}$.

3.2 Commutativity between rescaling and conditioning

The critical single-type Feller diffusion process was originally introduced in [Fel51] as a continuous approximation for large branching populations. It is also well-known that a similar approximation holds for multitype BGWc processes of any class of criticality, as shown for example in [JofMet86], Theorem 4.4.2. In this section, we shall relate the conditional limit theorems obtained for BGWc processes with the ones obtained for Feller diffusion processes by using this same approximation, and deduce from this that *rescaling and conditioning commute.* After presenting in **Subsection 3.2.1** sufficient assumptions on the rescaled BGWc process to obtain a Feller diffusion limit, we prove the "interchangeability" between rescaling and conditioning on extinction (**Subsection 3.2.2**), rescaling and the Yaglom-type limits (**Subsection 3.2.3**), rescaling and the Q-process (**Subsection 3.2.4**), and finally between rescaling and the long-time behavior of the Q-process (**Subsection 3.2.5**).

In the following **Subsections 3.2.1-3.2.5**, $\mathbb{P}$ denotes the law of the multitype Feller diffusion process introduced in Section 1.2, with mutation matrix $\mathbf{C}$ and variance parameters $\sigma_1^2, \ldots, \sigma_d^2$.

3.2.1 Rescaled BGWc process

For every $n \in \mathbb{N}^*$, we consider $\mathbb{P}^n$ the law of a BGWc process with offspring distribution $(\mathbf{p}^n(\mathbf{k}))_{\mathbf{k}\in\mathbb{N}^d}$, branching rates $\alpha_i := n$, $i = 1 \ldots d$, and rescaled by $\frac{1}{n}$. On the one hand, the relation $\alpha_i = n$ implies that for n large, the branching dynamics is accelerated (the life expectancy of the individuals are shorter, hence the branching events occur more often). On the other hand, the rescaling in space means that each individual is given a small weight $\frac{1}{n}$. The process with law $\mathbb{P}^n$ consequently takes its values in $\frac{1}{n}\mathbb{N}^d$. We denote by m_{ij}^n the first-order moments of the offspring distribution $\mathbf{p}^n(\mathbf{k})$,

$$m_{ij}^n := \sum_{\mathbf{k}\in\mathbb{N}^d} k_j p_i^n(\mathbf{k}), \tag{3.2.1}$$

and introduce the second-order moments: for all $i, j, k = 1 \dots d$,

$$\sigma_{ij}^n(l) := \sum_{\mathbf{k}\in\mathbb{N}^d} (k_i - \delta_{li})(k_j - \delta_{lj}) p_l^n(\mathbf{k}). \tag{3.2.2}$$

We denote by $\mathbf{M}^n$ the mean matrix and define $\mathbf{C}^n := n(\mathbf{M}^n - \mathbf{I})$.

We point out that here and in was follows the superscript n stands for the rescaling parameter and not for an exponentiation, except in formula (3.2.5) and more generally for expressions involving $\mathbf{q}_n$.

The infinitesimal generator of the rescaled BGWc with law $\mathbb{P}^n$ is then, for all smooth function $f : \frac{1}{n}\mathbb{N}^d \to \mathbb{R}$ and all $\boldsymbol{x} \in \frac{1}{n}\mathbb{N}^d$,

$$(L^n f)(\boldsymbol{x}) := n^2 \sum_{i=1}^d x_i \sum_{\mathbf{k}\in\mathbb{N}^d} p_i^n(\mathbf{k}) \left[f\left(\boldsymbol{x} + \frac{\mathbf{k} - \mathbf{e}_i}{n}\right) - f(\boldsymbol{x}) \right]. \tag{3.2.3}$$

One factor n stems from the branching rates, and the other scalings are a result of describing $n\boldsymbol{x}$ individuals of mass $\frac{1}{n}$.

Under appropriate assumptions on the initial distribution and on the first and second-order moments, the sequence of BGWc processes with law $\mathbb{P}^n$ is a nice approximation of the Feller diffusion process with law $\mathbb{P}$. We quote here the result of [JofMet86], Theorem 4.4.2.

Proposition 3.2.1. *Let us assume that for all $i, j = 1 \dots d$, as n tends to infinity,*

$$\begin{aligned}
&\textbf{\textit{(A1)}} \qquad m_{ij}^n = \delta_{ij} + \frac{1}{n} c_{ij} + o\left(\frac{1}{n}\right),\\
&\textbf{\textit{(A2)}} \qquad \sigma_{ii}^n(i) = \sigma_i^2 + o(1),\\
&\textbf{\textit{(A3)}} \qquad \lim_{N\to\infty} \sup_{n\in\mathbb{N}^*} \sum_{\boldsymbol{k}/\|\boldsymbol{k}\|>N} \|\boldsymbol{k}\|^2 p_i^n(\boldsymbol{k}) = 0.
\end{aligned}$$

Then, for any $\boldsymbol{x} \in \mathbb{R}_+^d$, $\boldsymbol{x} \neq \boldsymbol{0}$, and any $\mathbb{N}^d$-valued sequence $\boldsymbol{x}^n$ such that $\lim_{n\to\infty} \frac{1}{n}\boldsymbol{x}^n = \boldsymbol{x}$, the sequence of probability measures $\mathbb{P}^n_{\frac{1}{n}\boldsymbol{x}^n}$ converges weakly to $\mathbb{P}_{\boldsymbol{x}}$ as n tends to infinity.

Remark 3.2.2. Assumption **(A1)** means that $\lim_{n\to\infty} n(\mathbf{M}^n - \mathbf{I}) = \mathbf{C}$, hence the mutation matrix $\mathbf{C}$ of the Feller diffusion process measures the rescaled discrepancy between the mean matrix $\mathbf{M}^n$ and the identity matrix $\mathbf{I}$, which represents the case of independent types.

Remark 3.2.3. Assumptions **(A1)**-**(A3)** imply

$$\lim_{n\to\infty} \sigma_{ij}^n(l) = \begin{cases} \sigma_i^2 & \text{if } (i,j) = (l,l),\\ 0 & \text{otherwise.} \end{cases} \tag{3.2.4}$$

Indeed if $(i, j) \neq (l, l)$, say $i \neq l$, we have for any $N > 0$,

$$\begin{aligned}
\sigma_{ij}^n(l) &= \sum_{\mathbf{k}\in\mathbb{N}^d} k_i k_j p_l^n(\mathbf{k}) - \delta_{lj} m_{li}^n\\
&\leqslant \sum_{\|\mathbf{k}\|\leqslant N} k_i k_j p_l^n(\mathbf{k}) + \sum_{\|\mathbf{k}\|>N} k_i k_j p_l^n(\mathbf{k})\\
&\leqslant N \sum_{\|\mathbf{k}\|\leqslant N} k_i p_l^n(\mathbf{k}) + \frac{1}{2}\left[\sum_{\|\mathbf{k}\|>N} k_i^2 p_l^n(\mathbf{k}) + \sum_{\|\mathbf{k}\|>N} k_j^2 p_l^n(\mathbf{k}) - \sum_{\|\mathbf{k}\|>N} (k_i - k_j)^2 p_l^n(\mathbf{k})\right]\\
&\leqslant N m_{li}^n + \frac{1}{2} \sup_{n\in\mathbb{N}^*} \sum_{\|\mathbf{k}\|>N} \|\mathbf{k}\|^2 p_l^n(\mathbf{k}).
\end{aligned}$$

Let $\varepsilon > 0$. **(A3)** implies that there exists N such that $\sup_{n\in\mathbb{N}^*} \sum_{\|\mathbf{k}\|>N} \|\mathbf{k}\|^2 p_l^n(\mathbf{k}) < \varepsilon$, and from **(A1)** there exists $N_0 \in \mathbb{N}^*$ such that for all $n \geqslant N_0$, $m_{li}^n < \frac{\varepsilon}{2N}$. Then for all $n \geqslant N_0$, $\sigma_{ij}^n(l) < \varepsilon$, which completed with **(A2)** leads to (3.2.4).

Whenever $\mathbf{M}^n$ is finite and irreducible, we shall denote by ρ^n the Perron's root of $\mathbf{C}^n$, and by $\boldsymbol{\xi}^n$ and $\boldsymbol{\eta}^n$ the associated right and left eigenvectors with the usual normalization convention $\boldsymbol{\xi}^n \cdot \mathbf{1} = 1$ and $\boldsymbol{\eta}^n \cdot \boldsymbol{\xi}^n = 1$.

3.2.2 Scaling limit of the BGWc process conditioned on extinction

The aim of this section is to prove that the sequence of rescaled BGWc processes introduced in Subsection 3.2.1, once *conditioned on extinction*, converges to the multitype Feller diffusion process with law $\mathbb{P}$ *conditioned on extinction* as well. Since the convergence of the unconditioned processes is known, this result means that the procedures of conditioning on extinction and taking the scaling limit are interchangeable.

Whenever it is defined, we denote by $\widetilde{\mathbb{P}}^n$ the law of the BGWc process with law $\mathbb{P}^n$, conditioned on extinction (see Proposition 2.2.1). Similar to (2.2.5) we denote by $(\widetilde{\mathbf{p}}^n(\mathbf{k}))_{\mathbf{k}\in\mathbb{N}^d}$ the probability distribution

$$\widetilde{p}_i^n(\mathbf{k}) := q_{n,i}^{-\frac{1}{n}} (\mathbf{q}_n)^{\frac{\mathbf{k}}{n}} p_i^n(\mathbf{k}), \tag{3.2.5}$$

where $\mathbf{q}_n$ stands for the extinction probability vector of the rescaled BGWc process:

$$q_{n,i} := \lim_{t\to\infty} \mathbb{P}^n_{\mathbf{e}_i}(\mathbf{X}_t = \mathbf{0}). \tag{3.2.6}$$

We introduce the associated first and second-order moments

$$\widetilde{m}_{ij}^n := \sum_{\mathbf{k}\in\mathbb{N}^d} k_j \widetilde{p}_i^n(\mathbf{k}), \tag{3.2.7}$$

$$\widetilde{\sigma}_{ij}^n(l) := \sum_{\mathbf{k}\in\mathbb{N}^d} (k_i - \delta_{li})(k_j - \delta_{lj}) \widetilde{p}_l^n(\mathbf{k}), \tag{3.2.8}$$

and define the matrices $\widetilde{\mathbf{M}}^n := \left(\widetilde{m}_{ij}^n\right)_{i,j=1\ldots d}$ and $\widetilde{\mathbf{C}}^n := n(\widetilde{\mathbf{M}}^n - \mathbf{I})$.

Proposition 3.2.4. *Let us assume **(A1)-(A3)** and **(F1)-(F2)**. Then for any $\boldsymbol{x} \in \mathbb{R}_+^d$, $\boldsymbol{x} \neq \boldsymbol{0}$, and any $\mathbb{N}^d$-valued sequence $\boldsymbol{x}^n$ such that $\lim_{n\to\infty} \frac{1}{n}\boldsymbol{x}^n = \boldsymbol{x}$, the following diagram is commutative,*

$$\begin{array}{ccc} \mathbb{P}^n_{\frac{1}{n}\boldsymbol{x}^n} & \rightsquigarrow & \widetilde{\mathbb{P}}^n_{\frac{1}{n}\boldsymbol{x}^n} \\ \Downarrow & & \Downarrow \\ \mathbb{P}_{\boldsymbol{x}} & \rightsquigarrow & \widetilde{\mathbb{P}}_{\boldsymbol{x}} \end{array}, \tag{3.2.9}$$

where $\rightsquigarrow$ stands for the transform by conditioning on extinction, and $\Rightarrow$ for the weak convergence of probability measures as n tends to infinity.

Proof. We already know from Proposition 3.2.1 that the weak convergence

$$\mathbb{P}^n_{\frac{1}{n}\boldsymbol{x}^n} \Longrightarrow \mathbb{P}_{\boldsymbol{x}} \tag{3.2.10}$$

holds as n tends to infinity.

On the other hand, we deduce from Proposition 3.2.1 that for all $i = 1 \ldots d$,

$$\lim_{n\to\infty} q_{n,i} = e^{-u_i}, \tag{3.2.11}$$

which by **(F2)** implies that for n large enough the rescaled branching process with law $\mathbb{P}^n$ has a positive risk of extinction $\mathbf{q}_n > \mathbf{0}$. For n large enough the conditioned law $\widetilde{\mathbb{P}}^n$ is thus well defined, and by Proposition 2.2.1 is the law of a (sub)critical BGWc process. The weak convergence to $\widetilde{\mathbb{P}}$

can thus be obtained by showing that assumptions **(A1)-(A3)** hold for the conditioned BGWc process as well, this time for the parameters $(\widetilde{c}_{ij})_{i,j}$ and $(\sigma_i^2)_i$.

First, we deduce from (3.2.11) that for all $i = 1 \dots d$,

$$\lim_{n\to\infty} q_{n,i}^{\frac{1}{n}} = 1, \tag{3.2.12}$$

and that for all $\mathbf{k} \in \mathbb{N}^d$,

$$\lim_{n\to\infty} n\left(q_{n,i}^{-\frac{1}{n}} (\mathbf{q}_n)^{\frac{\mathbf{k}}{n}} - 1\right) = -(k_i - 1)u_i.$$

Since we have on the one hand $\widetilde{c}_{ij} = c_{ij} - \sigma_i^2 u_i \delta_{ij}$, and on the other hand

$$\begin{aligned}\widetilde{c}_{ij}^n &= n\left(\frac{1}{q_{n,i}} \sum_{\mathbf{k}\in\mathbb{N}^d} (\mathbf{q}_n)^{\mathbf{k}} k_j p_i^n(\mathbf{k}) - \delta_{ij}\right) \\ &= c_{ij}^n + \sum_{\mathbf{k}\in\mathbb{N}^d} n\left(q_{n,i}^{-1} (\mathbf{q}_n)^{\mathbf{k}} - 1\right) k_j p_i^n(\mathbf{k}),\end{aligned}$$

it ensues that

$$\lim_{n\to\infty} \widetilde{\mathbf{C}}^n = \widetilde{\mathbf{C}}. \tag{3.2.13}$$

Second, we have

$$\lim_{n\to\infty} \widetilde{\sigma}_{ii}^n(i) = \lim_{n\to\infty} q_{n,i}^{-\frac{1}{n}} \sum_{\mathbf{k}\in\mathbb{N}^d} (k_i - 1)^2 (\mathbf{q}_n)^{\frac{\mathbf{k}}{n}} p_i^n(\mathbf{k}) = \sigma_i^2. \tag{3.2.14}$$

Finally,

$$\lim_{N\to\infty} \sup_{n\in\mathbb{N}^*} \sum_{\mathbf{k}/\|\mathbf{k}\|>N} \|\mathbf{k}\|^2 \widetilde{p}_i^n(\mathbf{k}) = \lim_{N\to\infty} \sup_{n\in\mathbb{N}^*} q_{n,i}^{-\frac{1}{n}} \sum_{\mathbf{k}/\|\mathbf{k}\|>N} \|\mathbf{k}\|^2 (\mathbf{q}_n)^{\frac{\mathbf{k}}{n}} p_i^n(\mathbf{k}) = 0. \tag{3.2.15}$$

$\square$

Remark 3.2.5. From the definition of $\widetilde{\sigma}_{ij}^n(l)$ we easily obtain thanks to (3.2.4) and (3.2.12) that

$$\lim_{n\to\infty} \widetilde{\sigma}_{ij}^n(l) = \begin{cases} \sigma_i^2 & \text{if } (i,j) = (l,l), \\ 0 & \text{otherwise.} \end{cases} \tag{3.2.16}$$

Assumption **(A1)** means that $\lim_{n\to\infty} \mathbf{C}^n = \mathbf{C}$, which implies that for n large enough, every matrix $\mathbf{C}^n$ has at least as many positive non-diagonal entries as $\mathbf{C}$ and is thus irreducible. By Proposition 2.2.1, this implies that the matrices $\widetilde{\mathbf{C}}^n$ are irreducible as well, and we denote in the following by $\widetilde{\rho}^n$, $\widetilde{\boldsymbol{\xi}}^n$ and $\widetilde{\boldsymbol{\eta}}^n$ their Perron's roots and right and left eigenvectors.

3.2.3 Scaling limit of the Yaglom-type distributions

The purpose of this section is to show that, for any fixed $\theta \geqslant 0$, the scaling limit of the θ-Yaglom distribution associated with the sequence of rescaled BGWc processes, is equal to the θ-Yaglom limit of the Feller diffusion process. We thus prove that

$$\lim_n \lim_t P(n,t,\theta) = \lim_t \lim_n P(n,t,\theta), \tag{3.2.17}$$

where $P(n,t,\theta)$ is the probability distribution introduced in (3.0.1).

Proposition 3.2.6. *Let us assume **(A1)-(A3)**, **(F1)-(F2)** and $\rho \neq 0$. Then, for every $\theta \geqslant 0$, $\boldsymbol{x} \in \mathbb{R}_+^d \backslash \{\boldsymbol{0}\}$ and $\mathbb{N}^d$-valued sequence $\boldsymbol{x}^n$ such that $\lim_{n\to\infty} \frac{1}{n}\boldsymbol{x}^n = \boldsymbol{x}$, the following nondegenerate probability distributions are equal:*

$$\lim_{n\to\infty}\lim_{t\to\infty} \mathbb{P}^n_{\frac{1}{n}\boldsymbol{x}^n}\left(\boldsymbol{X}_t \in \,.\,|\, \boldsymbol{X}_{t+\theta} \neq \boldsymbol{0}, \lim_{s\to\infty} \boldsymbol{X}_s = \boldsymbol{0}\right)$$
$$= \lim_{t\to\infty}\lim_{n\to\infty} \mathbb{P}^n_{\frac{1}{n}\boldsymbol{x}^n}\left(\boldsymbol{X}_t \in \,.\,|\, \boldsymbol{X}_{t+\theta} \neq \boldsymbol{0}, \lim_{s\to\infty} \boldsymbol{X}_s = \boldsymbol{0}\right). \quad (3.2.18)$$

Proof. We have shown in the proof of Proposition 3.2.4 that, under assumptions **(A1)-(A3)** and **(F1)-(F2)**, for n large enough, $\mathbf{C}^n$ is irreducible and $\mathbf{q}_n > \mathbf{0}$. Moreover, Remark 3.2.3 ensures finiteness of the second-order moments for the distribution $\mathbf{p}^n(\mathbf{k})$ from a certain rank, and finally we deduce from **(A1)** that $\lim_{n\to\infty}\rho^n = \rho \neq 0$, hence that $\rho_n \neq 0$ from a certain rank. We can consequently apply Proposition 2.3.1 for any fixed n large enough satisfying these conditions. We obtain that the long-time limit

$$\lim_{t\to\infty} \mathbb{P}^n_{\frac{1}{n}\boldsymbol{x}^n}\left(\mathbf{X}_t \in \,.\,|\, \mathbf{X}_{t+\theta} \neq \mathbf{0}, \lim_{s\to\infty} \mathbf{X}_s = \mathbf{0}\right)$$

defines a probability distribution on $\frac{1}{n}\mathbb{N}^d$, with generating function $F^{\theta,n}$ given by

$$F^{\theta,n}(\mathbf{r}) = e^{-\widetilde{\rho}^n\theta}\left[F^{0,n}(\mathbf{r}) - F^{0,n}\left(\mathbf{r}\widetilde{\mathbf{q}}_n(\theta)^n\right)\right]. \quad (3.2.19)$$

Here $\widetilde{\mathbf{q}}_n(\theta)$ denotes the extinction probability vector at time θ for the subcritical process with law $\widetilde{\mathbb{P}}^n$, and $F^{0,n}$ denotes the generating function of the Yaglom distribution associated with the subcritical process with law $\widetilde{\mathbb{P}}^n$, $F^{0,n}(\mathbf{r}) := \lim_{t\to\infty}\widetilde{\mathbb{E}}^n\left[\mathbf{r}^{\mathbf{X}_t} \,|\, \mathbf{X}_t \neq \mathbf{0}\right]$.

On the other hand, by Proposition 3.2.1, the right side of (3.2.18) equals

$$\lim_{t\to\infty} \widetilde{\mathbb{P}}_{\boldsymbol{x}}\left(\mathbf{X}_t \in \,.\,|\, \mathbf{X}_{t+\theta} \neq \mathbf{0}\right),$$

which by Proposition 2.3.5 defines a probability distribution on $\mathbb{R}_+^d$ with Laplace transform Φ^θ given by

$$\Phi^\theta(\boldsymbol{\lambda}) = e^{-\widetilde{\rho}\theta}\left[\Phi^0(\boldsymbol{\lambda}) - \Phi^0\left(\boldsymbol{\lambda} + \widetilde{\mathbf{u}}(\theta)\right)\right]. \quad (3.2.20)$$

Hence we need to prove that for any $\boldsymbol{\lambda} \geqslant \mathbf{0}$, denoting $\mathbf{e}^{-\boldsymbol{\lambda}} := (e^{-\lambda_1}, \dots, e^{-\lambda_d})$,

$$\lim_{n\to\infty} F^{\theta,n}(\mathbf{e}^{-\boldsymbol{\lambda}}) = \Phi^\theta(\boldsymbol{\lambda}). \quad (3.2.21)$$

From Proposition 3.2.4 we have $\lim_{n\to\infty}\widetilde{q}_{n,i}(\theta) = e^{-\widetilde{u}_i(\theta)}$, and since $\lim_{n\to\infty}\widetilde{\mathbf{C}}^n = \widetilde{\mathbf{C}}$ we can show easily that

$$\lim_{n\to\infty}\widetilde{\rho}^n = \widetilde{\rho}, \quad \lim_{n\to\infty}\widetilde{\boldsymbol{\xi}}^n = \widetilde{\boldsymbol{\xi}}. \quad (3.2.22)$$

As a consequence we see by (3.2.19) and (3.2.20) that the convergence (3.2.21) holds as soon as it is true for $\theta = 0$, *i.e.*

$$\lim_{n\to\infty} F^{0,n}(\mathbf{e}^{-\boldsymbol{\lambda}}) = \Phi^0(\boldsymbol{\lambda}). \quad (3.2.23)$$

On the one hand we have, as seen in (2.3.9),

$$F^{0,n}(\mathbf{r}) = 1 - \frac{\gamma^n(\mathbf{r})}{\gamma^n(\mathbf{0})},$$

where $\gamma^n(\mathbf{r})$ satisfies (see (2.3.6))

$$\lim_{t\to\infty} e^{-\widetilde{\rho}^n t}\widetilde{\boldsymbol{\eta}}^n \cdot \left(\mathbf{1} - \widetilde{\mathbf{F}}_t^n(\mathbf{r})\right) = \gamma^n(\mathbf{r}). \quad (3.2.24)$$

Here $\widetilde{\mathbf{F}}_t^n$ denotes the generating function at time t of the rescaled process with law $\widetilde{\mathbb{P}}^n$, defined by $\widetilde{F}^n_{t,i}(\mathbf{r}) := \widetilde{\mathbb{E}}^n_{\mathbf{e_i}}(\mathbf{r}^{\mathbf{X}_t})$.

On the other hand, as seen in Subsection 2.3.2,

$$\Phi^0(\boldsymbol{\lambda}) = 1 - \frac{\kappa(\boldsymbol{\lambda})}{\kappa},$$

where $\kappa(\boldsymbol{\lambda})$ and $\kappa > 0$ satisfy

$$\lim_{t\to\infty} e^{-\widetilde{\rho}t}\widetilde{\boldsymbol{\eta}} \cdot \widetilde{\mathbf{u}}_t(\boldsymbol{\lambda}) = \kappa(\boldsymbol{\lambda}),$$
$$\lim_{t\to\infty} e^{-\widetilde{\rho}t}\widetilde{\boldsymbol{\eta}} \cdot \left(\lim_{\boldsymbol{\lambda}\to\boldsymbol{\infty}} \widetilde{\mathbf{u}}_t(\boldsymbol{\lambda})\right) = \kappa.$$

In order to obtain (3.2.23) we first prove that the convergence (3.2.24) is uniform in n. As already mentioned, (2.3.6) (and thus (3.2.24)) is obtained as an extension *via* the embedded process of a convergence result for BGW processes (Theorem 2 in [JofSpit67]), *via* a method detailed in the proof of Theorem 6.1 in [Ogu75]. Analyzing this proof, it appears that the uniform convergence in n for (3.2.24) would stem from the uniform convergence in n of

$$\lim_{\substack{k\to\infty\\ k\in\mathbb{N}}} e^{-\widetilde{\rho}^n k}\widetilde{\boldsymbol{\eta}}^n \cdot \left(\mathbf{1} - \widetilde{\mathbf{F}}^n_k(\mathbf{r})\right) = \gamma^n(\mathbf{r}). \tag{3.2.25}$$

Let us prove the uniform convergence of (3.2.25). For this purpose, we consider the embedded subcritical BGW process with offspring generating function $\widetilde{\mathbf{F}}^n_1$, mean matrix $\exp(\widetilde{\mathbf{C}}^n)$, maximal eigenvalue $e^{\widetilde{\rho}^n}$, associated eigenvectors $\widetilde{\boldsymbol{\xi}}^n$ and $\widetilde{\boldsymbol{\eta}}^n$. By the integral form of the remainder term in the Taylor expansion of $\widetilde{\mathbf{F}}^n_1$, there exists a non-negative matrix-valued rest $\mathbf{A}^n(\mathbf{r})$ such that for all $\mathbf{r} \in [0,1]^d$,

$$\mathbf{1} - \widetilde{\mathbf{F}}^n_1(\mathbf{r}) = \left(\exp(\widetilde{\mathbf{C}}^n) - \mathbf{A}^n(\mathbf{r})\right)(\mathbf{1} - \mathbf{r}), \tag{3.2.26}$$

satisfying $\mathbf{A}^n(\mathbf{r}) = O(\|\mathbf{1}-\mathbf{r}\|)$ as $\mathbf{r} \to \mathbf{1}$. Moreover, the second-order derivatives of $\widetilde{\mathbf{F}}^n_1$ being finite and bounded in n thanks to (3.2.16), we have in the neighborhood of 1,

$$\mathbf{A}^n(\mathbf{r}) = O(\|\mathbf{1}-\mathbf{r}\|) \quad \text{uniformly in } n. \tag{3.2.27}$$

Let us denote

$$\Delta^n_k(\mathbf{r}) := e^{-\widetilde{\rho}^n k}\widetilde{\boldsymbol{\eta}}^n \cdot \left(\mathbf{1} - \widetilde{\mathbf{F}}^n_k(\mathbf{r})\right).$$

By (3.2.26) we have

$$\Delta^n_{k+1}(\mathbf{r}) - \Delta^n_k(\mathbf{r}) = -e^{-(k+1)\widetilde{\rho}^n}\widetilde{\boldsymbol{\eta}}^n \cdot \mathbf{A}^n\left(\widetilde{\mathbf{F}}^n_k(\mathbf{r})\right)\left(\mathbf{1} - \widetilde{\mathbf{F}}^n_k(\mathbf{r})\right),$$

hence for every n and $\mathbf{r}$, $\Delta^n_k(\mathbf{r})$ is decreasing in k. It follows that for every $n, k \in \mathbb{N}$ and $\mathbf{r} \in [0,1]^d$,

$$\Delta^n_k(\mathbf{r}) \leqslant \Delta^n_0(\mathbf{r}) = \widetilde{\boldsymbol{\eta}}^n \cdot (\mathbf{1} - \mathbf{r}) \leqslant \sup_{n\in\mathbb{N}} \widetilde{\boldsymbol{\eta}}^n \cdot \mathbf{1},$$

the right term being finite by means of (3.2.13). This ensures that $\sup_{n\in\mathbb{N}} \gamma^n(\mathbf{r}) < \infty$.

Let $N \in \mathbb{N}$ such that

$$\sup_{n\geqslant N} \widetilde{\rho}^n < 0.$$

From (2.3.6) we obtain that for every n,

$$\mathbf{1} - \widetilde{\mathbf{F}}^n_k(\mathbf{r}) \sim_{k\to\infty} e^{\widetilde{\rho}^n k}\gamma^n(\mathbf{r})\widetilde{\boldsymbol{\xi}}^n,$$

from which we deduce thanks to (3.2.22) that

$$\lim_{k\to\infty} \widetilde{\mathbf{F}}^n_k(\mathbf{r}) = \mathbf{1} \quad \text{uniformly in } n \text{ up to } N.$$

(Note that the convergence is also uniform in $\mathbf{r} \in [0,1]^d$). Together with (3.2.27) this implies the existence of $C_1 > 0$ and $K > 0$ such that for all $k \geqslant K$ and all $n \geqslant N$,

$$\mathbf{A}^n \left(\widetilde{\mathbf{F}}_k^n (\mathbf{r}) \right) \leqslant C_1 \| \mathbf{1} - \widetilde{\mathbf{F}}_k^n (\mathbf{r}) \| \mathbf{I}. \tag{3.2.28}$$

Since

$$e^{-\widetilde{\rho}^n k} \left(1 - \widetilde{F}_{k,i}^n (\mathbf{r}) \right) \leqslant \frac{1}{\inf_{n,i} \widetilde{\eta}_i^n} e^{-\widetilde{\rho}^n k} \widetilde{\boldsymbol{\eta}}^n \cdot \left(\mathbf{1} - \widetilde{\mathbf{F}}_k^n (\mathbf{r}) \right),$$

there exists $C_2 > 0$ such that for all k and all $n \geqslant N$,

$$e^{-\widetilde{\rho}^n k} \| \mathbf{1} - \widetilde{\mathbf{F}}_k^n (\mathbf{r}) \| \leqslant C_2. \tag{3.2.29}$$

Now, for every $k \geqslant K$, $p \geqslant 0$ and every $n \geqslant N$,

$$\begin{aligned} \Delta_k^n(\mathbf{r}) - \Delta_{k+p}^n(\mathbf{r}) &= \sum_{i=0}^{p-1} e^{-\widetilde{\rho}^n (k+1+i)} \widetilde{\boldsymbol{\eta}}^n \cdot \mathbf{A}^n \left(\widetilde{\mathbf{F}}_{k+i}^n(\mathbf{r}) \right) \left(\mathbf{1} - \widetilde{\mathbf{F}}_{k+i}^n(\mathbf{r}) \right) \\ &\leqslant e^{(k-1) \sup_n \widetilde{\rho}^n} C_1 (C_2)^2 \sup_n \left(\widetilde{\boldsymbol{\eta}}^n \cdot \mathbf{1} \right) \frac{1}{1 - e^{\sup_n \widetilde{\rho}^n}}. \end{aligned}$$

We thus obtain by virtue of Cauchy criterion that the convergence (3.2.25) is uniform in n up to N.

As a consequence, the convergence (3.2.24) is uniform as well, and we obtain that for all $\boldsymbol{\lambda} \in \mathbb{R}_+^d$,

$$\begin{aligned} \lim_{n\to\infty} \gamma^n(\mathbf{e}^{-\boldsymbol{\lambda}}) &= \lim_{n\to\infty} \lim_{t\to\infty} e^{-\widetilde{\rho}^n t} \widetilde{\boldsymbol{\eta}}^n \cdot \left(\mathbf{1} - \widetilde{\mathbf{F}}_t^n \left(\mathbf{e}^{-\boldsymbol{\lambda}} \right) \right) \\ &= \lim_{t\to\infty} \lim_{n\to\infty} e^{-\widetilde{\rho}^n t} \widetilde{\boldsymbol{\eta}}^n \cdot \left(\mathbf{1} - \widetilde{\mathbf{F}}_t^n \left(\mathbf{e}^{-\boldsymbol{\lambda}} \right) \right) \\ &= \lim_{t\to\infty} e^{-\widetilde{\rho} t} \widetilde{\boldsymbol{\eta}} \cdot \left(\mathbf{1} - \mathbf{e}^{-\widetilde{\mathbf{u}}_t(\boldsymbol{\lambda})} \right) \\ &= \kappa(\boldsymbol{\lambda}). \end{aligned}$$

Similarly

$$\lim_{n\to\infty} \gamma^n(\mathbf{0}) = \kappa.$$

This finally proves (3.2.23), which ensures (3.2.18). □

3.2.4 Scaling limit of the Q-process

It has been proved in [LaNey68] that the diffusion limit of branching processes conditioned on very late extinction was also a Feller diffusion process conditioned on very late extinction. However, the result only concerns single-type processes and the convergence of the finite-dimensional distributions. We wish to generalize this result to multitype BGWc processes.

In this section we present the Q-process associated with the Feller diffusion process as the solution to a martingale problem. As shown in Subsection 2.4.2, this Q-process is the unique solution to the martingale problem $\mathcal{MP}(\frac{1}{h}L(h.))$, where h is the space-time harmonic function for the infinitesimal generator L of the Feller diffusion process given by (2.4.18). Denoting by L^n the infinitesimal generator of the rescaled BGWc process (as defined in (3.2.3)), we know from Theorem 2.4.1 that the corresponding Q-process is a solution to the martingale problem $\mathcal{MP}(\frac{1}{h^n}L^n(h^n.))$, where h^n is an appropriate space-time harmonic function for L^n. On the other side it is known that any limit of solutions to the martingale problems $\mathcal{MP}(L^n)$ is a solution to the martingale problem $\mathcal{MP}(L)$ (Proposition 3.2.1). Our aim is now to prove that any limit of solutions to $\mathcal{MP}(\frac{1}{h^n}L^n(h^n.))$ is a solution to $\mathcal{MP}(\frac{1}{h}L(h.))$. The result is illustrated in the following commutative diagram, where $\dashrightarrow$ stands for the transform by conditioning on very late extinction, and $\Rightarrow$ for the weak convergence of probability measures when the scaling parameter n tends to infinity:

$$\begin{array}{ccc} \mathcal{MP}(L^n) & \dashrightarrow & \mathcal{MP}(\frac{1}{h^n}L^n(h^n.)) \\ \Downarrow & & \Downarrow \\ \mathcal{MP}(L) & \dashrightarrow & \mathcal{MP}(\frac{1}{h}L(h.)) \end{array} \tag{3.2.30}$$

Let us approximate the conditioned Feller diffusion process with law $\mathbb{P}^*$ by discrete-state processes. According to the intuition, these approximating processes could be BGWc processes conditioned on very late extinction. For every $n \in \mathbb{N}^*$, let us denote by $\mathbb{P}^{n,*}$ the law of the Q-process associated with the BGWc process with law $\mathbb{P}^n$, as defined in (2.4.3). The following theorem then states that under the assumptions of Proposition 3.2.1 and under an additional technical assumption on the third-order moments (assumption **(A4)**), the sequence of conditioned laws $\mathbb{P}^{n,*}$ converges weakly to the conditioned law $\mathbb{P}^*$.

Theorem 3.2.7. *Let us assume **(A1)-(A3)**, **(F1)-(F2)**. We assume moreover*

$$\textbf{(A4)} \qquad \sup_{n\in\mathbb{N}^*} \sum_{\boldsymbol{k}\in\mathbb{N}^d} k_j^2 k_l p_i^n(\boldsymbol{k}) < \infty.$$

Then, for any $\boldsymbol{x} \in \mathbb{R}_+^d$, $\boldsymbol{x} \neq \boldsymbol{0}$, and any $\mathbb{N}^d$-valued sequence $\boldsymbol{x}^n$ such that $\lim_{n\to\infty} \frac{1}{n}\boldsymbol{x}^n = \boldsymbol{x}$, the following diagram is commutative,

$$\begin{array}{ccc} \mathbb{P}^n_{\frac{1}{n}\boldsymbol{x}^n} & \dashrightarrow & \mathbb{P}^{n,*}_{\frac{1}{n}\boldsymbol{x}^n} \\ \Downarrow & & \Downarrow \\ \mathbb{P}_{\boldsymbol{x}} & \dashrightarrow & \mathbb{P}^*_{\boldsymbol{x}} \end{array}, \tag{3.2.31}$$

where $\dashrightarrow$ stands for the transform by conditioning on very late extinction, and $\Rightarrow$ for the weak convergence of probability measures as n tends to infinity.

Proof. Since $\mathbb{P}^n_{n^{-1}\boldsymbol{x}^n} \Rightarrow \mathbb{P}_{\boldsymbol{x}}$ is ensured by Proposition 3.2.1, it only remains to prove $\mathbb{P}^{n,*}_{n^{-1}\boldsymbol{x}^n} \Rightarrow \mathbb{P}^*_{\boldsymbol{x}}$. As already mentioned in the proof of Proposition 3.2.6, the BGWc with law $\mathbb{P}^n$ has for n large enough a positive risk of extinction, and the second-order moments of the distribution $\mathbf{p}^n(\mathbf{k})$ are finite from a certain rank. Hence for n large enough, $\mathbb{P}^n$ satisfies the assumptions of Theorem 2.4.1, which will be assumed from now on. From (2.4.10) the infinitesimal generator $L^{n,*}$ of the Q-process with law $\mathbb{P}^{n,*}$ is, for all smooth function $f : \frac{1}{n}\mathbb{N}^d\backslash\{\mathbf{0}\} \to \mathbb{R}$ and all $\boldsymbol{x} \in \frac{1}{n}\mathbb{N}^d$, $\boldsymbol{x} \neq \mathbf{0}$,

$$\left(L^{n,*}f\right)(\boldsymbol{x}) := n^2 \sum_{i=1}^d x_i \sum_{\mathbf{k}\in\mathbb{N}^d} \widetilde{p}_i^n(\mathbf{k}) \frac{(n\boldsymbol{x} + \mathbf{k} - \mathbf{e}_i)\cdot \widetilde{\boldsymbol{\xi}}^n}{n\boldsymbol{x}\cdot\widetilde{\boldsymbol{\xi}}^n} \left[f\left(\boldsymbol{x} + \frac{\mathbf{k}-\mathbf{e}_i}{n}\right) - f(\boldsymbol{x})\right]. \tag{3.2.32}$$

Introducing the state-dependent probability distribution $(\mathbf{s}^n(\boldsymbol{x},\mathbf{k}))_{\mathbf{k}\in\mathbb{N}^d}$ and branching rates $\alpha_i^n(\boldsymbol{x})$ defined for all $\boldsymbol{x} \in \mathbb{N}^d$, $\boldsymbol{x} \neq \mathbf{0}$, by

$$s_i^n(\boldsymbol{x},\mathbf{k}) := \frac{(\boldsymbol{x} + \mathbf{k} - \mathbf{e}_i)\cdot\widetilde{\boldsymbol{\xi}}^n}{\boldsymbol{x}\cdot\widetilde{\boldsymbol{\xi}}^n + \widetilde{\rho}^n\widetilde{\xi}_i^n}\widetilde{p}_i^n(\mathbf{k}),$$

$$\alpha_i^n(\boldsymbol{x}) := \frac{\boldsymbol{x}\cdot\widetilde{\boldsymbol{\xi}}^n + \widetilde{\rho}^n\widetilde{\xi}_i^n}{\boldsymbol{x}\cdot\widetilde{\boldsymbol{\xi}}^n},$$

the infinitesimal generator $L^{n,*}$ can be written

$$\left(L^{n,*}f\right)(\boldsymbol{x}) = n^2 \sum_{i=1}^d \alpha_i^n(n\boldsymbol{x})\, x_i \sum_{\mathbf{k}\in\mathbb{N}^d} s_i^n(n\boldsymbol{x},\mathbf{k}) \left[f\left(\boldsymbol{x} + \frac{\mathbf{k}-\mathbf{e}_i}{n}\right) - f(\boldsymbol{x})\right]. \tag{3.2.33}$$

The law $\mathbb{P}^{n,*}$ is thus the law of a branching process with *state-dependent* offspring distribution and branching rates.

This is also the case for the Q-process associated with the Feller diffusion process (which has a state-dependant mutation matrix, see (2.4.19)). We can thus apply the convergence criterion provided in [JofMet86] for multitype state-dependent branching processes. For this purpose, we introduce the moments associated with the probability distribution $(\mathbf{s}^n(\boldsymbol{x},\mathbf{k}))_{\mathbf{k}\in\mathbb{N}^d}$. For every $\boldsymbol{x}\in\mathbb{N}^d$, $\boldsymbol{x}\neq\mathbf{0}$,

$$\begin{aligned} m_{ij}^n(\boldsymbol{x}) &:= \sum_{\mathbf{k}\in\mathbb{N}^d} s_i^n(\boldsymbol{x},\mathbf{k})k_j, \\ \sigma_{ij}^n(l)(\boldsymbol{x}) &:= \sum_{\mathbf{k}\in\mathbb{N}^d} s_l^n(\boldsymbol{x},\mathbf{k})(k_i-\delta_{li})(k_j-\delta_{lj}), \end{aligned} \tag{3.2.34}$$

and we define

$$c_{ij}^n(\boldsymbol{x}) := n\alpha_i^n(\boldsymbol{x})\left(m_{ij}^n(\boldsymbol{x})-\delta_{ij}\right). \tag{3.2.35}$$

Then, according to Theorem 4.4.2 in [JofMet86], the weak convergence of $\mathbb{P}^{n,*}_{n^{-1}\boldsymbol{x}^n}$ to $\mathbb{P}^*_{\boldsymbol{x}}$ holds as soon as for all $i,j=1\dots d$ and all $\boldsymbol{x}\in\mathbb{N}^d$, $\boldsymbol{x}\neq\mathbf{0}$,

$$\sup_{n\in\mathbb{N}^*}\sup_{\substack{\boldsymbol{x}\in\mathbb{N}^d\\ \boldsymbol{x}\neq\mathbf{0}}} \alpha_i^n(n\boldsymbol{x})<\infty, \qquad \lim_{n\to\infty}\alpha_i^n(n\boldsymbol{x})=1, \tag{3.2.36}$$

$$\sup_{n\in\mathbb{N}^*}\sup_{\substack{\boldsymbol{x}\in\mathbb{N}^d\\ \boldsymbol{x}\neq\mathbf{0}}} c_{ij}^n(n\boldsymbol{x})<\infty, \qquad \lim_{n\to\infty}c_{ij}^n(n\boldsymbol{x})=c_{ij}(\boldsymbol{x}), \tag{3.2.37}$$

$$\lim_{N\to\infty}\sup_{n\in\mathbb{N}^*}\sup_{\substack{\boldsymbol{x}\in\mathbb{N}^d\\ \boldsymbol{x}\neq\mathbf{0}}}\sum_{\mathbf{k}/\|\mathbf{k}\|>N}\|\mathbf{k}\|^2 s_i^n(\boldsymbol{x},\mathbf{k})=0, \tag{3.2.38}$$

$$\sup_{n\in\mathbb{N}^*}\sup_{\substack{\boldsymbol{x}\in\mathbb{N}^d\\ \boldsymbol{x}\neq\mathbf{0}}} \sigma_{ii}^n(i)(\boldsymbol{x})<\infty, \qquad \lim_{n\to\infty}\alpha_i^n(n\boldsymbol{x})\sigma_{ii}^n(i)(n\boldsymbol{x})=\sigma_i^2, \tag{3.2.39}$$

and if the martingale problem $\mathcal{MP}\left(L^*, C_b^2\left(\mathbb{R}_+^d\backslash\{\mathbf{0}\},\mathbb{R}\right),\boldsymbol{x}\right)$ has a unique solution, which has been shown in Subsection 2.4.2.

Criteria (3.2.36)-(3.2.37) mean that the state-dependant branching rates $\alpha_i^n(n\boldsymbol{x})$ and matrix entries $c_{ij}^n(n\boldsymbol{x})$ converge uniformly in $\boldsymbol{x}\in\mathbb{N}^d\backslash\{\mathbf{0}\}$, while criterion (3.2.38) implies the uniform convergence in n and x of the series of the second-order moments.

Let us prove that (3.2.36)-(3.2.39) are satisfied. (3.2.36) is immediate. Let us show (3.2.37). For all $\boldsymbol{x}\in\mathbb{N}^d$, $\boldsymbol{x}\neq\mathbf{0}$, we have

$$\begin{aligned} c_{ij}^n(n\boldsymbol{x}) &= n\left[\sum_{\mathbf{k}\in\mathbb{N}^d} k_j\frac{(n\boldsymbol{x}+\mathbf{k}-\mathbf{e}_i)\cdot\widetilde{\boldsymbol{\xi}}^n}{n\boldsymbol{x}\cdot\widetilde{\boldsymbol{\xi}}^n}\widetilde{p}_i^n(\mathbf{k})-\frac{n\boldsymbol{x}\cdot\widetilde{\boldsymbol{\xi}}^n+\widetilde{\rho}^n\widetilde{\xi}_i^n}{n\boldsymbol{x}\cdot\widetilde{\boldsymbol{\xi}}^n}\delta_{ij}\right] \\ &= n\left(\widetilde{m}_{ij}^n-\delta_{ij}\right)+\frac{1}{\boldsymbol{x}\cdot\widetilde{\boldsymbol{\xi}}^n}\left[\sum_{\mathbf{k}\in\mathbb{N}^d}k_j\left(\mathbf{k}-\mathbf{e}_i\right)\cdot\widetilde{\boldsymbol{\xi}}^n\widetilde{p}_i^n(\mathbf{k})-\widetilde{\rho}^n\widetilde{\xi}_i^n\delta_{ij}\right] \\ &= \widetilde{c}_{ij}^n+\frac{1}{\boldsymbol{x}\cdot\widetilde{\boldsymbol{\xi}}^n}\sum_{l=1}^d\widetilde{\xi}_l^n\widetilde{\sigma}_{jl}^n(i). \end{aligned} \tag{3.2.40}$$

(3.2.16) together with (3.2.13), (3.2.22) and (3.2.40) implies that $\sup_{n\in\mathbb{N}^*}\sup_{\substack{\boldsymbol{x}\in\mathbb{N}^d\\ \boldsymbol{x}\neq\mathbf{0}}} c_{ij}^n(n\boldsymbol{x})<\infty$, and that

$$\lim_{n\to\infty}c_{ij}^n(n\boldsymbol{x})=\widetilde{c}_{ij}+\frac{\sigma_i^2\widetilde{\xi}_i}{\boldsymbol{x}.\widetilde{\boldsymbol{\xi}}}\delta_{ij}.$$

Let us next prove (3.2.38). For all $N\in\mathbb{N}^*$ and all $\boldsymbol{x}\in\mathbb{N}^d$, $\boldsymbol{x}\neq\mathbf{0}$,

$$\sum_{\|\mathbf{k}\|>N}\|\mathbf{k}\|^2 s_i^n(\boldsymbol{x},\mathbf{k})=\frac{1}{\boldsymbol{x}\cdot\widetilde{\boldsymbol{\xi}}^n+\widetilde{\rho}^n\widetilde{\xi}_i^n}\left[\boldsymbol{x}\cdot\widetilde{\boldsymbol{\xi}}^n\sum_{\|\mathbf{k}\|>N}\|\mathbf{k}\|^2\widetilde{p}_i^n(\mathbf{k})+\sum_{\|\mathbf{k}\|>N}\|\mathbf{k}\|^2\left(\mathbf{k}-\mathbf{e}_i\right)\cdot\widetilde{\boldsymbol{\xi}}^n\widetilde{p}_i^n(\mathbf{k})\right].$$

We have on the one hand

$$\sum_{\|\mathbf{k}\|>N} \|\mathbf{k}\|^2 \widetilde{p}_i^n(\mathbf{k}) \leqslant \frac{1}{q_{n,i}} \sum_{\|\mathbf{k}\|>N} \|\mathbf{k}\|^2 p_i^n(\mathbf{k}),$$

and on the other hand

$$\sum_{\|\mathbf{k}\|>N} \|\mathbf{k}\|^2 \,(\mathbf{k}-\mathbf{e}_i)\cdot \widetilde{\boldsymbol{\xi}}^n \widetilde{p}_i^n(\mathbf{k}) \leqslant \frac{1}{q_{n,i}} \sum_{l=1}^d \widetilde{\xi}_l^n \sum_{j=1}^d \sum_{\mathbf{k}\in\mathbb{N}^d} k_j^2 k_l p_i^n(\mathbf{k}).$$

This together with assumptions **(A3)-(A4)** and the convergence results given by (3.2.12) and (3.2.22) leads to (3.2.38). It remains to prove (3.2.39). For all $\boldsymbol{x}\in\mathbb{N}^d$, $\boldsymbol{x}\neq\mathbf{0}$,

$$\begin{aligned}
\alpha_i^n(n\boldsymbol{x})\sigma_{ii}^n(i)(n\boldsymbol{x}) &= \sum_{\mathbf{k}\in\mathbb{N}^d} \frac{(n\boldsymbol{x}+\mathbf{k}-\mathbf{e}_i)\cdot\widetilde{\boldsymbol{\xi}}^n}{n\boldsymbol{x}\cdot\widetilde{\boldsymbol{\xi}}^n}(k_i-1)^2\widetilde{p}_i^n(\mathbf{k}) \\
&= \widetilde{\sigma}_{ii}^n(i) + \frac{1}{n}\frac{1}{\boldsymbol{x}\cdot\widetilde{\boldsymbol{\xi}}^n}\sum_{l=1}^d \widetilde{\xi}_l^n \sum_{\mathbf{k}\in\mathbb{N}^d}(k_l-\delta_{il})\left(k_i^2-2k_i+1\right)\widetilde{p}_i^n(\mathbf{k}) \\
&= \widetilde{\sigma}_{ii}^n(i) + \frac{1}{n}\frac{1}{\boldsymbol{x}\cdot\widetilde{\boldsymbol{\xi}}^n}\sum_{l=1}^d \widetilde{\xi}_l^n \left(\sum_{\mathbf{k}\in\mathbb{N}^d} k_i^2 k_l \widetilde{p}_i^n(\mathbf{k}) - 2\widetilde{\sigma}_{il}^n(i) + \widetilde{m}_{il}^n - \delta_{il}\right),
\end{aligned}$$

which thanks to (3.2.13), (3.2.22), (3.2.16) and assumption **(A4)** converges to σ_i^2 as n tends to infinity. Writing

$$\sigma_{ii}^n(i)(\boldsymbol{x}) = \frac{\boldsymbol{x}\cdot\widetilde{\boldsymbol{\xi}}^n}{\boldsymbol{x}\cdot\widetilde{\boldsymbol{\xi}}^n+\widetilde{\rho}^n\widetilde{\xi}_i^n}\left[\widetilde{\sigma}_{ii}^n(i) + \frac{1}{\boldsymbol{x}\cdot\widetilde{\boldsymbol{\xi}}^n}\sum_{l=1}^d \widetilde{\xi}_l^n\left(\sum_{\mathbf{k}\in\mathbb{N}^d} k_i^2 k_l \widetilde{p}_i^n(\mathbf{k}) - 2\widetilde{\sigma}_{il}^n(i) + \widetilde{m}_{il}^n - \delta_{il}\right)\right],$$

we obtain (3.2.39) thanks to the same convergence results. □

3.2.5 Scaling limit of the time asymptotic of the Q-process

In this section, we show that the time asymptotic of the Q-process associated with the rescaled BGWc process converges as n tends to infinity to the one of the Q-process associated with the Feller diffusion process. We thus prove that

$$\lim_n \lim_t \lim_\theta P(n,t,\theta) = \lim_t \lim_n \lim_\theta P(n,t,\theta), \tag{3.2.41}$$

where $P(n,t,\theta)$ is defined in (3.0.1).

Proposition 3.2.8. *Let us assume **(A1)-(A3)**, **(F1)-(F2)** and $\rho\neq 0$. Then for every $\boldsymbol{x}\in\mathbb{R}_+^d\backslash\{\boldsymbol{0}\}$ and $\mathbb{N}^d$-valued sequence $\boldsymbol{x}^n$ such that $\lim_{n\to\infty}\frac{1}{n}\boldsymbol{x}^n=\boldsymbol{x}$, the following nondegenerate probability distributions are equal:*

$$\begin{aligned}
&\lim_{n\to\infty}\lim_{t\to\infty}\lim_{\theta\to\infty}\mathbb{P}^n_{\frac{1}{n}\boldsymbol{x}^n}\left(\boldsymbol{X}_t\in\,.\,|\,\boldsymbol{X}_{t+\theta}\neq\boldsymbol{0},\lim_{s\to\infty}\boldsymbol{X}_s=\boldsymbol{0}\right) \\
&\qquad= \lim_{t\to\infty}\lim_{n\to\infty}\lim_{\theta\to\infty}\mathbb{P}^n_{\frac{1}{n}\boldsymbol{x}^n}\left(\boldsymbol{X}_t\in\,.\,|\,\boldsymbol{X}_{t+\theta}\neq\boldsymbol{0},\lim_{s\to\infty}\boldsymbol{X}_s=\boldsymbol{0}\right).
\end{aligned} \tag{3.2.42}$$

Proof. Applying Proposition 3.1.1 to the BGWc process with law $\mathbb{P}^n$, we know that the left term of (3.2.42) corresponds for every n fixed (before considering the limit in n) to a nondegenerate probability distribution with generating function given by

$$\lim_{t\to\infty}\mathbb{E}^{n,*}_{\frac{1}{n}\boldsymbol{x}^n}\left[\mathbf{r}^{\mathbf{X}_t}\right] = -\sum_{i=1}^d r_i\widetilde{\xi}_i^n\frac{\partial\gamma^n(\mathbf{r})}{\partial r_i}. \tag{3.2.43}$$

In order to obtain the interchangeability of limits

$$\lim_{n\to\infty}\lim_{t\to\infty}\mathbb{E}^{n,*}_{\frac{1}{n}\boldsymbol{x}^n}\left[\mathbf{r}^{\mathbf{X}_t}\right]=\lim_{t\to\infty}\lim_{n\to\infty}\mathbb{E}^{n,*}_{\frac{1}{n}\boldsymbol{x}^n}\left[\mathbf{r}^{\mathbf{X}_t}\right],$$

it is enough to prove that the convergence (3.2.43) proved for every n is actually uniform in n. As seen in the proof of Proposition 3.1.1, it is enough to prove that

$$\lim_{t\to\infty} e^{-\widetilde{\rho}^n t}\frac{\partial}{\partial r_i}\left[\ln \widetilde{F}^n_{t,j}(\mathbf{r})\right]=-\frac{\partial\gamma^n(\mathbf{r})}{\partial r_i}\widetilde{\xi}^n_j \quad \text{uniformly in } n. \tag{3.2.44}$$

For this purpose we follow the steps of Lemma 3.1.3. First we prove that

$$\lim_{t\to\infty} e^{-\widetilde{\rho}^n t}\left(\mathbf{1}-\mathbf{F}^n_t(\mathbf{r})\right)=\gamma^n(\mathbf{r})\,\widetilde{\boldsymbol{\xi}}^n \quad \text{uniformly in } n, \tag{3.2.45}$$

which similarly as for (3.2.24) can be deduced from the convergence of the embedded BGW process

$$\lim_{\substack{k\to\infty\\ k\in\mathbb{N}}} e^{-\widetilde{\rho}^n k}\left(\mathbf{1}-\mathbf{F}^n_k(\mathbf{r})\right)=\gamma^n(\mathbf{r})\,\widetilde{\boldsymbol{\xi}}^n \quad \text{uniformly in } n. \tag{3.2.46}$$

(3.2.46) will be itself a consequence from (3.2.25) together with

$$\lim_{k\to\infty}\frac{\mathbf{1}-\widetilde{\mathbf{F}}^n_k(\mathbf{r})}{\widetilde{\boldsymbol{\eta}}^n\cdot\left(\mathbf{1}-\widetilde{\mathbf{F}}^n_k(\mathbf{r})\right)}=\widetilde{\boldsymbol{\xi}}^n \quad \text{uniformly in } n. \tag{3.2.47}$$

Let us prove (3.2.47). We denote by $\{\lambda^n_l,\,l\}$ the eigenvalues of $\widetilde{\mathbf{C}}^n$ different from $\widetilde{\rho}^n$, and define $R^n:=\min_l\{\widetilde{\rho}^n-\Re(\lambda^n_l)\}>0$, where $\Re$ denotes the real part of λ^n_l. We similarly define R for the matrix $\widetilde{\mathbf{C}}$. By (3.2.13) we have $\lim_{n\to\infty}R^n=R$. Moreover, for all $i,j=1\dots d$ and $k\in\mathbb{N}^*$,

$$\left[\exp(k\widetilde{\mathbf{C}}^n)\right]_{ij}=\widetilde{\xi}^n_i\widetilde{\eta}^n_j e^{\widetilde{\rho}^n k}+\sum_l \varphi^n_{ij,l}(k)e^{\lambda^n_l k},$$

where $\varphi^n_{ij,l}$ is a complex-valued polynomial with degree smaller than the algebraic multiplicity of λ^n_l. Since the $\varphi^n_{ij,l}$ converge as n tends to infinity, we have $\sup_n|\varphi^n_{ij,l}(k)|<\infty$ and we can write

$$\begin{aligned}\left|e^{-\widetilde{\rho}^n k}\left[\exp(k\widetilde{\mathbf{C}}^n)\right]_{ij}-\widetilde{\xi}^n_i\widetilde{\eta}^n_j\right| &\leqslant \sum_l\left|\varphi^n_{ij,l}(k)\right|e^{(\Re(\lambda^n_l)-\widetilde{\rho}^n)k}\\ &\leqslant \sum_l\sup_n\left|\varphi^n_{ij,l}(k)\right|e^{-\inf_n R^n k}.\end{aligned}$$

Consequently, denoting by $\mathbf{P}^n$ the matrix with entries $\widetilde{\xi}^n_i\widetilde{\eta}^n_j$,

$$\lim_{k\to\infty} e^{-\widetilde{\rho}^n k}\exp(k\widetilde{\mathbf{C}}^n)=\mathbf{P}^n \quad \text{uniformly in } n, \tag{3.2.48}$$

and we can find a null sequence u_k such that for all $k,n\geqslant 1$,

$$(1-u_k)\,\mathbf{P}^n\leqslant e^{-\widetilde{\rho}^n k}\exp(k\widetilde{\mathbf{C}}^n)\leqslant(1+u_k)\,\mathbf{P}^n.$$

On the other hand, we know by (3.2.28) and (3.2.29) that $\lim_{k\to\infty}e^{-\widetilde{\rho}^n}\mathbf{A}^n(\widetilde{\mathbf{F}}^n_{k-1}(\mathbf{r}))=\mathbf{0}$ uniformly in n up to N, hence we can choose a null sequence v_k such that for all $k\geqslant 1$ and all $n\geqslant N$,

$$0\leqslant e^{-\widetilde{\rho}^n}\mathbf{A}^n\left(\widetilde{\mathbf{F}}^n_{k-1}(\mathbf{r})\right)\leqslant v_k\mathbf{P}^n.$$

Then, as detailed in the proof of Theorem 1 in [JofSpit67], we have for any $k\geqslant l\geqslant 1$ and any $n\geqslant N$,

$$\left\|\frac{\mathbf{1}-\widetilde{\mathbf{F}}^n_k(\mathbf{r})}{\widetilde{\boldsymbol{\eta}}^n\cdot\left(\mathbf{1}-\widetilde{\mathbf{F}}^n_k(\mathbf{r})\right)}-\widetilde{\boldsymbol{\xi}}^n\right\|\leqslant\frac{2u_l+\sum_{m=k-l+1}^k v_m}{1-\sum_{m=k-l+1}^k v_m-u_l}.$$

By letting first k tend to infinity, and then l, we obtain (3.2.47) and thus (3.2.45). Now from (3.2.45) we deduce that the convergence $\lim_{t\to\infty} e^{-\widetilde{\rho}^n t} \ln \widetilde{F}^n_{t,j}(\mathbf{r}) = -\gamma^n(\mathbf{r})\widetilde{\xi}^n_j$ is uniform in n too, and for all $h \geqslant 0$,

$$\gamma(\mathbf{r})^n \widetilde{\xi}^n_j - \gamma^n(\mathbf{r}+h\mathbf{e_i})\widetilde{\xi}^n_j = \lim_{t\to\infty} \int_0^h e^{-\widetilde{\rho}^n t} \frac{\partial}{\partial r_i}\left[\ln \widetilde{F}^n_{t,j}(\mathbf{r}+u\mathbf{e_i})\right] du \ \text{ uniformly in } n.$$

Moreover,

$$0 \leqslant e^{-\widetilde{\rho}^n t}\frac{\partial}{\partial r_i}\left[\ln \widetilde{F}^n_{t,j}(\mathbf{r})\right] \leqslant e^{-\widetilde{\rho}^n t}\frac{1}{\widetilde{F}^n_{t,j}(\mathbf{r})}\frac{1}{r_i}\widetilde{\mathbb{E}}^n_{\mathbf{e_j}}\left[X_{t,i}\right],$$

and by (3.2.48)

$$\lim_{t\to\infty} e^{-\widetilde{\rho}^n t}\widetilde{\mathbb{E}}^n_{\mathbf{e_j}}\left[X_{t,i}\right] = \widetilde{\xi}^n_j\widetilde{\eta}^n_i \ \text{ uniformly in } n.$$

Since $\lim_{t\to\infty} \widetilde{F}^n_{t,j}(\mathbf{r}) = 1$ uniformly in n and $\mathbf{r}$, there exists a constant $C > 0$ such that for all $n \in \mathbb{N}$, $t \geqslant 0$ and all $u \in [0,h]$,

$$\left| e^{-\rho^n t}\frac{\partial}{\partial r_i}\left[\ln \widetilde{F}^n_{t,j}(\mathbf{r}+u\mathbf{e_i})\right]\right| \leqslant \frac{C}{r_i+u},$$

which is integrable in u. By Lebesgue's dominated convergence theorem we thus have

$$\gamma^n(\mathbf{r})\widetilde{\xi}^n_j - \gamma^n(\mathbf{r}+h\mathbf{e_i})\widetilde{\xi}^n_j = \int_0^h \lim_{t\to\infty}\left[e^{-\widetilde{\rho}^n t}\frac{\partial}{\partial r_i}\left(\ln \widetilde{F}^n_{t,j}(\mathbf{r}+u\mathbf{e_i})\right)\right] du \ \text{ uniformly in } n,$$

which leads to (3.2.44). □

3.3 Overview of the commutativity results

Summing up the results of this chapter we consequently obtain the following statement.

Theorem 3.3.1. *Let us assume* ***(A1)-(A4)****,* ***(F1)-(F2)*** *and $\rho \neq 0$. Then the three limits interchange*

$$\lim_{n,\,t,\,\theta} \mathbb{P}^n_{\frac{1}{n}\boldsymbol{x}^n}\left(\boldsymbol{X}_t \in .\ |\ \boldsymbol{X}_{t+\theta} \neq \boldsymbol{0},\ \lim_{s\to\infty}\boldsymbol{X}_s = \boldsymbol{0}\right). \tag{3.3.1}$$

Furthermore, the obtained limit is non-degenerate, and defines a non-trivial probability distribution on $\mathbb{R}^d_+$ which does not depend on $\boldsymbol{x} := \lim_{n\to\infty}\frac{1}{n}\boldsymbol{x}^n$.

We illustrate Proposition 3.3.1 in the commutative diagram presented in Figure 3.1.

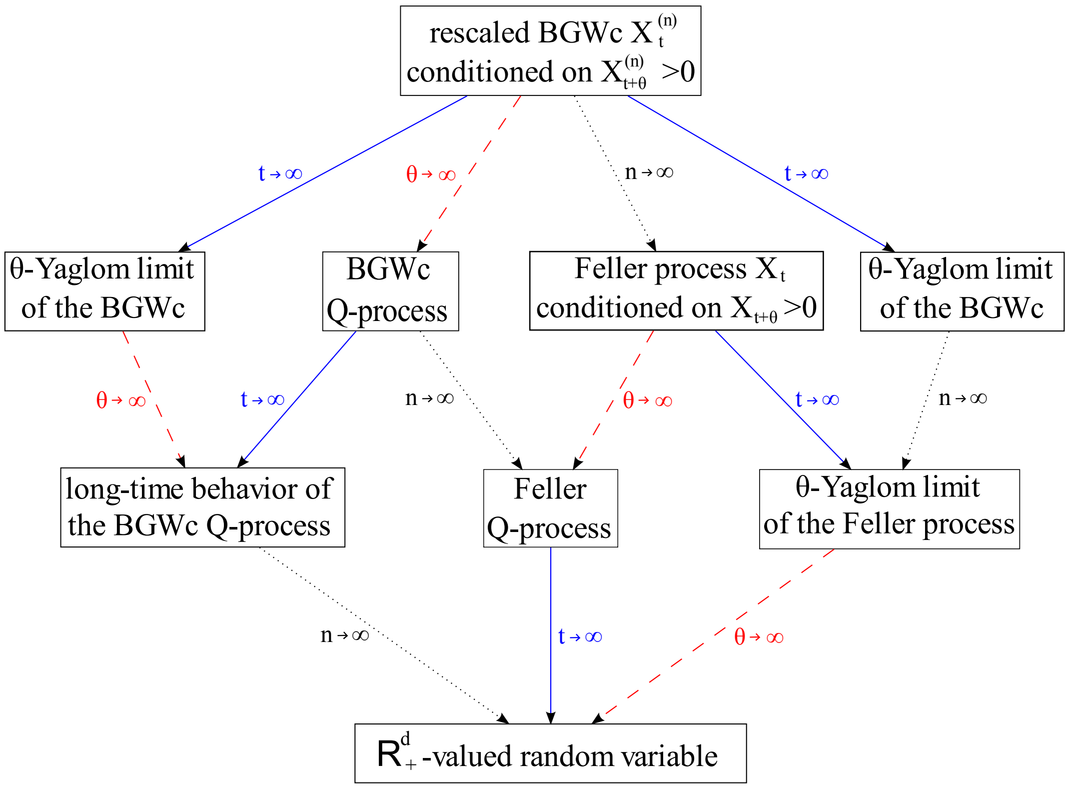

Figure 3.1: Interchangeability of limits.

Chapter 4

Risk analysis for vanishing branching populations

This chapter is devoted to the risk analysis related to the extinction of a population, namely the estimation of the time to extinction, of the total size of the population until extinction (*tree size*), and more generally the prediction of the behavior of the population until extinction. We provide notably an innovative statistical tool which enables to predict the evolution of the population in case of a *very late extinction*, corresponding either to the *best-case scenario* or to the *worst-case scenario*, depending on whether the extinction of the population is desirable or not.

We consider a branching population with Poissonian transitions, which can be seen either as a Markovian process of order $d \geqslant 1$, or as a discrete-time Bienaymé-Galton-Watson process (BGW), with d types corresponding to the memory of the process. As we will see in **Chapter 5**, this kind of model arises notably in the epidemiological context. We assume moreover that a certain parameter in the Poissonian transition, corresponding to the infection parameter in the latter context, is unknown. The estimation of a key parameter such as the Perron's root, which determines whether extinction is certain or not, is of very large interest and has been studied a lot in the literature. Since the unknown parameter is, in our model, an explicit function of the Perron's root, it is very natural either to build a new estimator specifically designed for the model, or to investigate the existing results in the estimation theory. As detailed in **Section 4.2**, estimators of the Perron's root for general multitype branching processes usually require the knowledge of the whole or partial genealogy of the process (for example individual offspring sizes, or parent-offspring type combination counts). Such data are mostly non available, which is what we assume in our work. S. Asmussen and N. Keiding, however, introduced in [AsmKei78] an explicit estimator based only on the total generation sizes, which is of direct practical applicability for our model. We deduce from this estimator a first estimator of the infection parameter. Despite the potentially large order of the Markovian process that we consider, its Poissonian character ensures many properties which make it easy to derive estimators with interesting characteristics. We thus build two conditional least squares estimators (CLSE) based either on the chosen process or on the process conditioned on non-extinction at each time step. In addition, we build an estimator corresponding to the Q-process associated with the model, which enables predictions of the evolution of the population in the best-case or worst-case scenario.

After presenting the model in **Section 4.1**, we provide in **Section 4.2** three estimators of the unknown parameter, which are all only based on the available observations (*i.e.* the size of the generations). We aim at asymptotic results, either as the size of the initial population tends to infinity, or as time tends to infinity. It could be of a great mathematical interest to study the asymptotic behavior when both the number of ancestors and the number of generations of the branching process simultaneously tend to infinity, as it is done in [DioYa97] for the single-type case, but we choose to focus on asymptotic properties of an immediate practical interest. We first build in **Subsection 4.2.1** a CLSE which is consistent and asymptotically normal, as the initial population size grows to infinity. This estimator is thus appropriate either in the growth phase of the population or in its decay phase, provided that the initial time of the model corresponds to a large number of individuals. In **Subsection 4.2.2** we focus on the subcritical case that is particularly designed for the decay phase, and build a CLSE based on the process conditioned on its non-extinction at each time step. We prove its consistency and asymptotic normality, as time tends to infinity. We finally provide in **Subsection 4.2.3** an explicit estimator derived from the estimator of the Perron's root introduced in [AsmKei78], and we deduce the consistency and asymptotic normality of our estimator in the supercritical case, on the set of non-extinction, as time tends to infinity. This last estimator is especially suitable in the growth phase of the population. In **Subsection 4.2.4** we compare these three estimators for several values of initial population size and time, and illustrate by means of simulations their asymptotic distributions. The final **Section 4.3** is dedicated to the very late extinction case. For this purpose, we study in **Subsection 4.3.1** the Q-process associated with the model, *i.e.* the process conditioned on very late extinction, and we prove that this process can be described very easily, with recognizable transition laws. In order to make long-term predictions for this conditioned process, we build in **Subsection 4.3.2** a CLSE of the unknown parameter designed for the Q-process, and prove its consistency and asymptotic normality, as time tends to infinity.

4.1 Stochastic branching model

In this section we introduce a stochastic process, which can arise notably as a model for the propagation of a rare SEIR disease in a large branching population. This will be detailed in **Chapter 5**, while we focus here on the theoretical aspect of the process (statistical properties, and stochastic properties of the associated Q-process).

Throughout this chapter we consider the following Markovian process of order $d \geqslant 1$ (see Figure 4.1),

$$X_n = \sum_{k=1}^{d} \sum_{i=1}^{X_{n-k}} \zeta_{n-k,n,i}, \tag{4.1.1}$$

where the $\{\zeta_{n-k,n,i}\}_i$ are i.i.d. given $\mathcal{F}_{n-1} := \sigma\left(\{X_{n-k}\}_{k\geqslant 1}\right)$, and follow a Poisson distribution with some parameter Ψ_k independent of n (time-homogeneous setting). Moreover, the $\{\zeta_{n-k,n,i}\}_{i,k}$ are assumed to be independent given $\mathcal{F}_{n-1}$.

The quantity k represents the maturation period needed to produce the "mathematical offsprings" $\zeta_{n-k,n,i}$. Note that the $\zeta_{n-k,n,i}$ are not offsprings of individual i in the usual sense, since the latter is only counted in the population at time $n-k$.

We point out that in the simple case $d = 1$, the process is a single-type BGW branching process with a Poisson offspring distribution.

We easily derive from (4.1.1) the conditional law of X_n,

$$X_n|(X_{n-1}, \ldots, X_{n-d}) \sim \mathcal{P}oisson\left(\sum_{k=1}^{d} X_{n-k}\Psi_k\right). \tag{4.1.2}$$

It appears moreover thanks to (4.1.1) that X_n can be written as a multitype BGW process. We do not introduce in this work the definition and properties of multitype BGW processes, since there are many similarities with the multitype BGWc processes introduced in Section 1.1. We refer instead to [AthNey72].

Let us define the d-dimensional process $\mathbf{X}_n := (X_{n,1}, \ldots, X_{n,d})$ such that

$$X_{n,i} := X_{n-i+1}, \quad i = 1 \ldots d. \tag{4.1.3}$$

We thus have

$$\mathbf{X}_n = (X_n, X_{n-1}, \ldots, X_{n-d+1}). \tag{4.1.4}$$

Note that in particular the first coordinate $X_{n,1}$ corresponds to the value of the single-type process at time n. Then $(\mathbf{X}_n)_{n\geqslant 0}$ is a d-type BGW process with offspring generating function

$$\begin{cases} f_i(\mathbf{r}) := \displaystyle\sum_{k=0}^{\infty} \frac{(\Psi_i)^k}{k!} e^{-\Psi_i} r_1^k r_{i+1} = e^{-\Psi_i(1-r_1)} r_{i+1}, \quad i = 1 \dots d-1, \\ f_d(\mathbf{r}) := \displaystyle\sum_{k=0}^{\infty} \frac{(\Psi_d)^k}{k!} e^{-\Psi_d} r_1^k = e^{-\Psi_d(1-r_1)}, \end{cases} \tag{4.1.5}$$

and mean matrix

$$\mathbf{M} := \begin{pmatrix} \Psi_1 & 1 & 0 & \dots & 0 \\ \Psi_2 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ \Psi_{d-1} & 0 & \dots & \dots & 1 \\ \Psi_d & 0 & \dots & \dots & 0 \end{pmatrix}. \tag{4.1.6}$$

The process $(\mathbf{X}_n)_{n\geqslant 0}$ is obviously non simple (the definition is the same as for BGWc processes, see Definition 1.1.4). We assume throughout this paper that $\Psi_d > 0$, and that there exists some $i = 1 \dots d-1$ with $\Psi_i > 0$. Under this assumption, there exists some $p \in \mathbb{N}$ such that $\mathbf{M}^p$ has all its entries positive, and the process is called *positive regular* (see [AthNey72] Section 5.2). Moreover, the process satisfies the **(XlogX)** assumption: for all $i, j = 1 \dots d$,

$$\mathbb{E}\left[X_{1,j} \ln X_{1,j} | \mathbf{X}_0 = \mathbf{e_i}\right] < \infty. \tag{4.1.7}$$

Indeed, for all $i = 1 \dots d$,

$$p_i(\mathbf{k}) = \begin{cases} \frac{(\Psi_i)^k}{k!} e^{-\Psi_i} & \text{if } \mathbf{k} = (k, \delta_{2,i+1}, \dots, \delta_{d,i+1}) \\ 0 & \text{otherwise.} \end{cases}$$

Hence the following holds. If $j = 1$, then

$$\sum_{\substack{\mathbf{k}\in\mathbb{N}^d \\ \mathbf{k}\neq\mathbf{0}}} k_j \ln k_j \, p_i(\mathbf{k}) = \sum_{k\geqslant 1} k \ln k \, \frac{(\Psi_i)^k}{k!} e^{-\Psi_i}.$$

Denoting $u_k = k \ln k \, (\Psi_i)^k / k!$, the series $\sum_{k\geqslant 1} u_k$ converges because $(u_k)_{k\geqslant 1}$ satisfies d'Alembert's criterion $\lim_{k\to\infty} u_{k+1} u_k^{-1} = \lim_{k\to\infty} \ln(k+1)\Psi_i (k \ln k)^{-1} = 0 < 1$. If $j \neq 1$, we have

$$\sum_{\substack{\mathbf{k}\in\mathbb{N}^d \\ \mathbf{k}\neq\mathbf{0}}} k_j \ln k_j \, p_i(\mathbf{k}) = 0.$$

The theory of multitype positive regular and non simple BGW processes implies that the extinction of the process $(\mathbf{X}_n)_{n\geqslant 0}$ occurs almost surely if and only if the Perron's root ρ of the mean matrix $\mathbf{M}$ is smaller than or equal to 1 ([AthNey72] Theorem 5.3.2). By definition, any eigenvalue λ of the matrix $\mathbf{M}$ is solution of $\det(\mathbf{M} - \lambda\mathbf{I}) = 0$. Let us define $D_d(\lambda) := \det(\mathbf{M} - \lambda\mathbf{I})$. By induction on $d \geqslant 1$, we prove that

$$D_d(\lambda) = (-1)^{d+1} \left[\sum_{k=1}^{d} \Psi_k \lambda^{d-k} - \lambda^d\right].$$

We have indeed $D_1(\lambda) = \Psi_1 - \lambda$, and for all $d \geqslant 1$, by expanding the determinant $D_{d+1}(\lambda)$ along the last row, we obtain that

$$D_{d+1}(\lambda) = (-1)^{d+2} \Psi_d - \lambda D_d(\lambda) = (-1)^{d+2} \left[\sum_{k=1}^{d+1} \Psi_k \lambda^{d+1-k} - \lambda^{d+1}\right].$$

Hence

$$\det(\mathbf{M}-\lambda\mathbf{I})=0 \iff \sum_{k=1}^{d}\Psi_k\lambda^{d-k}-\lambda^d=0 \iff \sum_{k=1}^{d}\Psi_k\lambda^{-k}=1.$$

The Perron's root ρ of the matrix $\mathbf{M}$ is thus the largest real solution of the equation

$$\sum_{k=1}^{d}\Psi_k\rho^{-k}=1. \tag{4.1.8}$$

But

$$\rho\leqslant 1 \iff \rho^{-1},\dots,\rho^{-d}\geqslant 1 \iff 1=\sum_{k=1}^{d}\Psi_k\rho^{-k}\geqslant\sum_{k=1}^{d}\Psi_k,$$

hence we have the following result ([Jac10A, JacPe10]):

Proposition 4.1.1. *Let us define*

$$R:=\sum_{k=1}^{d}\Psi_k. \tag{4.1.9}$$

Then the process $(\boldsymbol{X}_n)_{n\geqslant 0}$ *defined by* (4.1.1) *and* (4.1.4) *is*

- *subcritical* ($\rho<1$) *if* $R<1$,
- *critical* ($\rho=1$) *if* $R=1$,
- *supercritical* ($\rho>1$) *if* $R>1$.

4.2 Estimation of the unknown parameter

Throughout this section we consider the BGW branching process $(\mathbf{X}_n)_{n\geqslant 0}$ introduced in Section 4.1, with generating function (4.1.5) and mean matrix (4.1.6).

We assume that the Ψ_k's affinely depend on some unknown parameter θ_0:

$$\Psi_k(\theta_0)=a_k\theta_0+b_k,\quad k=1\dots d, \tag{4.2.1}$$

where a_k and b_k are known. The purpose of this section is to provide estimators for θ_0. We will see in **Chapter 5** that in the epidemiological context of rare $SEIR$ diseases in large populations, θ_0 could correspond either to the horizontal or to the vertical infection parameter. In the following, we denote by $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{\Psi}(\theta_0)$ the d-dimensional vectors with coordinates a_k, b_k and $\Psi_k(\theta_0)$ respectively.

In what follows, we shall adopt a notation more suitable for parametric statistical studies, and denote with a subscript on $\mathbb{P}$ and $\mathbb{E}$ the parameter of the distribution. The initial distribution which was indicated by a subscript in the previous chapters, will from now on be written explicitly. For example,

$$\mathbb{P}_{\theta_0}\left[\mathbf{X}_n=\mathbf{j}|\mathbf{X}_0=\mathbf{i}\right].$$

Our aim is to provide estimators of θ_0 based on the observations $(\mathbf{X}_0,\dots,\mathbf{X}_n)$, with asymptotic properties corresponding to interesting characteristics, notably in the epidemiological context. We are thus looking for estimators suitable in the subcritical and/or supercritical cases, with asymptotic properties, as the initial population size grows to infinity, or as the number of observations n tends to infinity. We would thus entirely cover the problem of estimating the infection parameter in the growth and decay phases of the population, offering moreover several alternatives depending on which asymptotic is suitable regarding the available data.

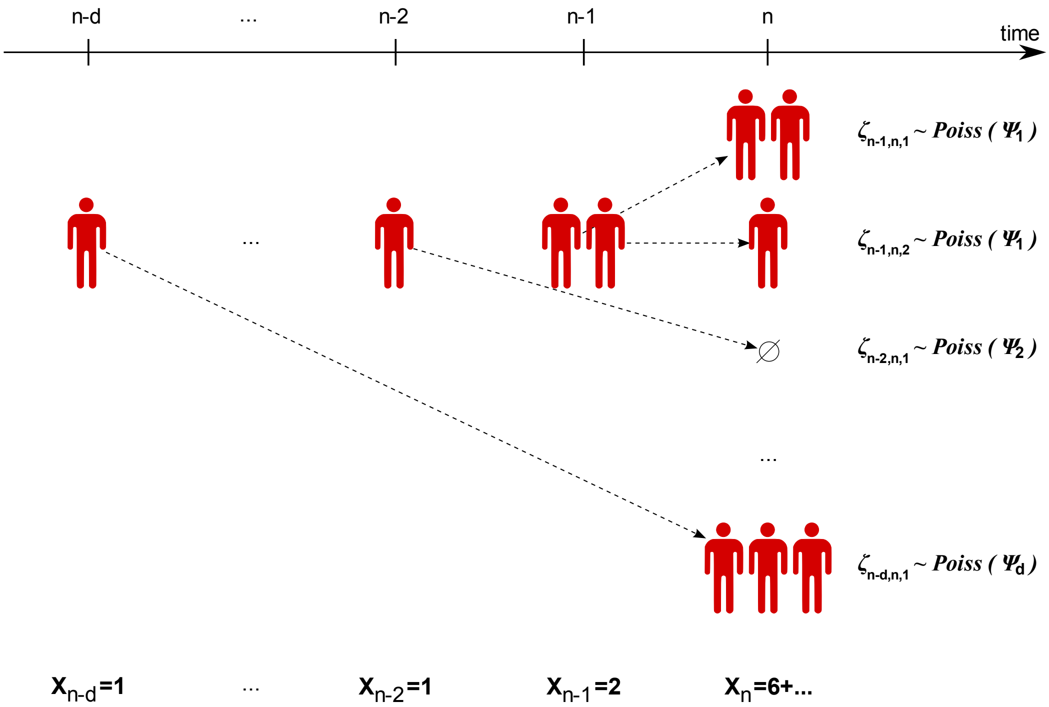

Figure 4.1: An illustration of the model (4.1.1).

We first point out that, as an immediate consequence of Proposition 4.1.1 and (4.2.1) , the criticality of the process depends on the sign of $\theta_0 - \theta_{crit.}$, where

$$\theta_{crit.} := \frac{1 - \sum_{k=1}^{d} b_k}{\sum_{k=1}^{d} a_k}. \tag{4.2.2}$$

More precisely,

$$\text{if } \begin{cases} \theta_0 < \theta_{crit.} \\ \theta_0 = \theta_{crit.} \\ \theta_0 > \theta_{crit.} \end{cases} \quad \text{the process is } \begin{cases} \text{subcritical} \\ \text{critical.} \\ \text{supercritical} \end{cases} \tag{4.2.3}$$

As mentioned in the introduction, we first investigate the numerous results in the literature dedicated to the estimation theory for general branching processes, in order to find an appropriate estimator for our model. In its early paper [Har48] in 1948, T. E. Harris provided an estimator for the mean value m_0 of a single-type BGW process $X_0, \ldots, X_n$. It is a maximum likelihood estimator, now referred to as the *Harris estimator*, based on observed values of the individual offspring size for each individual in each generation. The estimator is

$$\widehat{m}_n^{MLE} := \frac{X_1 + \ldots + X_n}{X_0 + \ldots + X_{n-1}}, \tag{4.2.4}$$

and Harris proved the consistency of $\widehat{m}_n^{MLE}$ as $n \to \infty$ in the supercritical case, on the set of non-extinction. Note that the estimator $\widehat{m}_n^{MLE}$ only involves $X_0, \ldots, X_n$. It is actually proved in [Dion72] that $\widehat{m}_n^{MLE}$ is also the maximum likelihood estimator of m_0 based on the observed values of $X_0, \ldots, X_n$ only. It is straightforward to show that $\widehat{m}_n^{MLE}$ is also the weighted conditional least squares estimator (CLSE) based on the process $X_k/\sqrt{X_{k-1}}$, defined as follows

$$\widehat{m}_n^{CLSE} := \arg\min_{m \geqslant 0} \sum_{k=1}^{n} \frac{(X_k - mX_{k-1})^2}{X_{k-1}}. \tag{4.2.5}$$

Similar estimation problems have been considered in the multitype case. In [AsmKei78], S. Asmussen and N. Keiding proposed a maximum likelihood estimator of the Perron's root ρ_0 based on the observations of the whole genealogy of the population (*i.e.* each offspring vector produced by every individual). It is proved in [KeiLau78] that this estimator is also the maximum likelihood estimator solely based on the observations at each generation of the total number of individuals of type j whose parents were of type i, for every $i, j = 1 \ldots d$. However in epidemiology this kind of variables are generally not observable. For our model this would imply indeed that, considering the number of individuals at a given time, we could say how many of them are stemming from an individual alive k time-units earlier. We are thus more interested in estimations based on the generations, or on the total size of the generations, such as the other estimator presented in [AsmKei78],

$$\widetilde{\rho}_n = \frac{|\mathbf{X}_1| + \ldots + |\mathbf{X}_n|}{|\mathbf{X}_0| + \ldots + |\mathbf{X}_{n-1}|}. \tag{4.2.6}$$

For $d = 1$, $\widetilde{\rho}_n$ clearly reduces to the Harris estimator defined in (4.2.4). Note that the relation (4.1.8) implies under assumption (4.2.1) that

$$\theta_0 = \frac{1 - \sum_{k=1}^{d} b_k \rho_0^{-k}}{\sum_{k=1}^{d} a_k \rho_0^{-k}}. \tag{4.2.7}$$

Hence an estimation of ρ_0 would provide an estimation of θ_0. However, the opposite is not true since ρ_0 cannot in general be expressed as an explicit function of θ_0. In the supercritical case, the estimator $\widetilde{\rho}_n$ was shown to be consistent, as $n \to \infty$, on the set of non-extinction, with an explicit asymptotic distribution ([AsmKei78]).

Due to the Poissonian character of the transitions of the process $(\mathbf{X}_n)_{n \geqslant 0}$, it is possible, in our setting, to express the joint probability function of the observations $X_0, \ldots, X_n$, without involving the whole or partial genealogy of the process. The likelihood function is indeed given by two factors, one of which is independent of θ_0, the logarithm of the other being

$$L(\theta_0) := -\theta_0 \sum_{k=1}^{n} \boldsymbol{a} \cdot \mathbf{X}_{k-1} + \sum_{k=1}^{n} X_k \ln\left(\boldsymbol{\Psi}(\theta_0) \cdot \mathbf{X}_{k-1}\right),$$

The MLE of θ_0 based on the observations $X_0, \ldots, X_n$ is thus a solution of $L'(\theta) = 0$, where

$$L'(\theta) = -\sum_{k=1}^{n} \boldsymbol{a} \cdot \mathbf{X}_{k-1} + \sum_{k=1}^{n} X_k \frac{\boldsymbol{a} \cdot \mathbf{X}_{k-1}}{\boldsymbol{\Psi}(\theta) \cdot \mathbf{X}_{k-1}}.$$

This equation has in general no explicit solution, except for simple cases such as the one-dimensional case $d = 1$, or the linear case $\boldsymbol{b} = \mathbf{0}$. The MLE is then, respectively,

$$\begin{aligned} \widehat{\theta}_n^{MLE} &\overset{d=1}{=} \frac{\sum_{k=1}^{n} (X_k - bX_{k-1})}{\sum_{k=1}^{n} aX_{k-1}}, \\ \widehat{\theta}_n^{MLE} &\overset{\boldsymbol{b}=\mathbf{0}}{=} \frac{\sum_{k=1}^{n} X_k}{\sum_{k=1}^{n} \boldsymbol{a} \cdot \mathbf{X}_{k-1}}. \end{aligned} \tag{4.2.8}$$

As shown later (see (4.2.13)), it corresponds in these cases to the CLSE of θ_0. It is however in general not the case, and we choose to focus on the CLSE.

In **Subsection 4.2.1** we first study the weighted CLSE

$$\widehat{\theta}_{|\mathbf{X}_0|}^{X} := \arg\min_{\theta \in \Theta} \sum_{k=1}^{n} \frac{\left[X_k - \mathbb{E}_\theta\left(X_k | \mathbf{X}_{k-1}\right)\right]^2}{\boldsymbol{a} \cdot \mathbf{X}_{k-1}}, \tag{4.2.9}$$

and its asymptotic properties, as $|\mathbf{X}_0| \to \infty$, for any class of criticality. Since we are only interested in the asymptotic in $|\mathbf{X}_0|$, we omit for the sake of clarity the subscript n in the estimator.

In a second instance, since we aim at finding an estimator with asymptotic properties, as $n \to \infty$, holding in the subcritical case, we consider in **Subsection 4.2.2** the homogeneous Markov chain $(\mathbf{Z}_k)_{k \geqslant 0}$ defined by

$$\mathbb{P}\left(Z_k = j | \mathbf{Z}_{k-1} = \mathbf{i}\right) = \mathbb{P}\left(X_k = j | \mathbf{X}_{k-1} = \mathbf{i}, \mathbf{X}_k \neq \mathbf{0}\right),$$

and study the associated weighted CLSE

$$\widehat{\theta}_n^{Z} := \arg\min_{\theta \in \Theta} \sum_{k=1}^{n} \frac{\left[Z_k - \mathbb{E}_\theta\left(Z_k | \mathbf{Z}_{k-1}\right)\right]^2}{\boldsymbol{a} \cdot \mathbf{Z}_{k-1}}. \tag{4.2.10}$$

Finally, thanks to relation (4.2.7), we derive from the estimator (4.2.6) a third estimator of θ_0,

$$\widetilde{\theta}_n^{X} := \frac{1 - \sum_{k=1}^{d} b_k \widetilde{\rho}_n^{\,-k}}{\sum_{k=1}^{d} a_k \widetilde{\rho}_n^{\,-k}}, \tag{4.2.11}$$

and deduce in **Subsection 4.2.3** from [AsmKei78] asymptotic properties of $\widetilde{\theta}_n^{X}$, as $n \to \infty$, in the supercritical case on the set of non-extinction.

In the following, we denote $\Theta :=]\theta_{min}, \theta_{max}[$ with $\theta_{max} > \theta_{min} > 0$.

4.2.1 A CLSE with asymptotic properties, as $|\mathbf{X}_0| \to \infty$

In this section, we provide an estimator with asymptotic properties, as the initial population size $|\mathbf{X}_0|$ tends to infinity, holding for any class of criticality. We consider the weighted CLSE based on the process $Y_k := X_k/\sqrt{\boldsymbol{a}\cdot\mathbf{X}_{k-1}}$,

$$\begin{aligned}\widehat{\theta}^X_{|\mathbf{X}_0|} &:= \arg\min_{\theta\in\Theta}\sum_{k=1}^{n}\left(Y_k-\mathbb{E}_\theta\left(Y_k|\mathbf{X}_{k-1}\right)\right)^2\\ &= \arg\min_{\theta\in\Theta}\sum_{k=1}^{n}\frac{\left(X_k-\boldsymbol{\Psi}(\theta)\cdot\mathbf{X}_{k-1}\right)^2}{\boldsymbol{a}\cdot\mathbf{X}_{k-1}}.\end{aligned}\tag{4.2.12}$$

We easily derive the following explicit form

$$\widehat{\theta}^X_{|\mathbf{X}_0|} = \frac{\sum_{k=1}^{n}\left(X_k-\boldsymbol{b}\cdot\mathbf{X}_{k-1}\right)}{\sum_{k=1}^{n}\boldsymbol{a}\cdot\mathbf{X}_{k-1}}.\tag{4.2.13}$$

The normalization of the process X_k by $\sqrt{\boldsymbol{a}\cdot\mathbf{X}_{k-1}}$ appears to be the most natural and suitable for the following reasons. First, this normalization generalizes the normalization $X_k/\sqrt{aX_{k-1}}$ in the monotype case, which is the one leading to the Harris estimator of $m_0 = a\theta_0 + b$. We have indeed, for $d=1$,

$$a\widehat{\theta}^X_{X_0} + b = \widehat{m}_n^{MLE}.$$

As mentioned in (4.2.8), it also corresponds, in the linear case, to the MLE of θ_0. In addition, defining for any vector $\mathbf{u}$, $\underline{\mathbf{u}} := \min_i u_i$ and $\overline{\mathbf{u}} := \max_i u_i$, we have

$$\theta_0 + \frac{\underline{\boldsymbol{b}}}{\overline{\boldsymbol{a}}} \leqslant \mathbb{E}_{\theta_0}\left(\left(Y_k-\mathbb{E}_{\theta_0}\left(Y_k|\mathbf{X}_{k-1}\right)\right)^2|\mathbf{X}_{k-1}\right) = \theta_0 + \frac{\boldsymbol{b}\cdot\mathbf{X}_{k-1}}{\boldsymbol{a}\cdot\mathbf{X}_{k-1}} \leqslant \theta_0 + \frac{\overline{\boldsymbol{b}}}{\underline{\boldsymbol{a}}},\tag{4.2.14}$$

hence the conditional variance of the error term $Y_k - \mathbb{E}_{\theta_0}\left(Y_k|\mathbf{X}_{k-1}\right)$ in the stochastic regression equation

$$Y_k = \mathbb{E}_{\theta_0}\left(Y_k|\mathbf{X}_{k-1}\right) + Y_k - \mathbb{E}_{\theta_0}\left(Y_k|\mathbf{X}_{k-1}\right)$$

is invariant under multiplication of the whole process, and bounded respectively to $(\mathbf{X}_k)_{k\geqslant 0}$.

We provide asymptotical results for the estimator $\widehat{\theta}^X_{|\mathbf{X}_0|}$ defined by (4.2.13), as the initial population size tends to infinity. We introduce the following notation. For every $i,j=1\dots d$ and $k\geqslant 1$, $m_{ij}^{(k)}(\theta)$ denotes the (i,j)-th entry in the k-th power of the matrix $\mathbf{M}(\theta)$ given by (4.1.6). We define

$$\sigma^2\left(\theta\right) := \theta + \frac{\sum_{k=1}^{n}\sum_{j=1}^{d}\sum_{i=1}^{d}\alpha_j b_i m_{ji}^{(k-1)}(\theta)}{\sum_{k=1}^{n}\sum_{j=1}^{d}\sum_{i=1}^{d}\alpha_j a_i m_{ji}^{(k-1)}(\theta)}.\tag{4.2.15}$$

We can now express the main result of this section.

Theorem 4.2.1. *Let us assume that there exist some $\alpha_i \in [0,1]$, $i=1\dots d$, such that, for all $i=1\dots d$,*

$$\lim_{|\boldsymbol{X}_0|\to\infty}\frac{X_{0,i}}{|\mathbf{X}_0|} \overset{a.s.}{=} \alpha_i.\tag{4.2.16}$$

Then $\widehat{\theta}^X_{|\boldsymbol{X}_0|}$ is strongly consistent:

$$\lim_{|\boldsymbol{X}_0|\to\infty}\widehat{\theta}^X_{|\boldsymbol{X}_0|} \overset{a.s.}{=} \theta_0,\tag{4.2.17}$$

and is asymptotically normally distributed:

$$\lim_{|\boldsymbol{X}_0|\to\infty}\sqrt{\frac{\sum_{k=1}^{n}\boldsymbol{a}\cdot\boldsymbol{X}_{k-1}}{\sigma^2(\widehat{\theta}^X_{|\boldsymbol{X}_0|})}}\left(\widehat{\theta}^X_{|\boldsymbol{X}_0|}-\theta_0\right) \overset{\mathcal{D}}{=} \mathcal{N}\left(0,1\right).\tag{4.2.18}$$

Note that assumption (4.2.16) allows several coordinates (at most $d-1$) of the initial vector $\mathbf{X}_0$ to be null. If for all $i=1\dots d$, $\alpha_i = \frac{1}{d}$, then all the coordinates of $\mathbf{X}_0$ grow proportionally to infinity.

In order to prove Theorem 4.2.1, we first show the following lemma, which takes advantage of the branching property of the process $(\mathbf{X}_k)_{k\geqslant 0}$, and uses the strong law of large numbers.

Lemma 4.2.2. *Assuming* (4.2.16)*, the following holds for all $k = 1 \dots n$ and all $i = 1 \dots d$,*

$$\lim_{|\boldsymbol{X}_0| \to \infty} \frac{X_{k,i}}{|\boldsymbol{X}_0|} \overset{a.s.}{=} \sum_{j=1}^{d} \alpha_j m_{ji}^{(k)}(\theta_0). \tag{4.2.19}$$

Proof of Lemma 4.2.2. Using the branching property of the process $(\mathbf{X}_k)_{k \geqslant 0}$, we write

$$X_{k,i} = \sum_{j=1}^{X_{0,1}} X_{k,i,j}^{(1)} + \ldots + \sum_{j=1}^{X_{0,d}} X_{k,i,j}^{(d)},$$

where for all $l = 1 \dots d$ and $j = 1 \dots X_{0,l}$, $X_{k,i,j}^{(l)}$ is the i-th coordinate of a d-type branching process at time k initiated by a single individual of type l. For k, i and l fixed, the random variables $\{X_{k,i,j}^{(l)}\}_j$ are i.i.d. with mean value $m_{li}^{(k)}(\theta_0)$. According to the strong law of large numbers and under (4.2.16), we have, for every $l = 1 \dots d$ such that $X_{0,l} \neq 0$,

$$\lim_{|\mathbf{X}_0| \to \infty} \frac{\sum_{j=1}^{X_{0,l}} X_{k,i,j}^{(l)}}{X_{0,l}} \overset{a.s.}{=} m_{li}^{(k)}(\theta_0),$$

which together with (4.2.16) leads to (4.2.19). □

Proof of Theorem 4.2.1. To prove the consistency of $\widehat{\theta}_{|\mathbf{X}_0|}^X$ we apply Lemma 4.2.2 to (4.2.13), using the fact that $X_k = X_{k,1}$ and $X_{k-i} = X_{k-1,i}$, and obtain

$$\lim_{|\mathbf{X}_0| \to \infty} \widehat{\theta}_{|\mathbf{X}_0|}^X \overset{a.s.}{=} \frac{\sum_{k=1}^{n} \sum_{j=1}^{d} \alpha_j \left(m_{j1}^{(k)}(\theta_0) - \sum_{i=1}^{d} b_i m_{ji}^{(k-1)}(\theta_0) \right)}{\sum_{k=1}^{n} \sum_{i=1}^{d} \sum_{j=1}^{d} a_i \alpha_j m_{ji}^{(k-1)}(\theta_0)}. \tag{4.2.20}$$

By definition,

$$m_{j1}^{(k)}(\theta_0) = \sum_{i=1}^{d} m_{ji}^{(k-1)}(\theta_0) m_{i1}(\theta_0) = \sum_{i=1}^{d} m_{ji}^{(k-1)}(\theta_0) \left(a_i \theta_0 + b_i \right),$$

hence (4.2.20) immediately leads to (4.2.17).

We are now interested in the asymptotic distribution of $\widehat{\theta}_{|\mathbf{X}_0|}^X - \theta_0$. We derive from (4.2.13) that

$$\sqrt{\sum_{k=1}^{n} \boldsymbol{a} \cdot \mathbf{X}_{k-1}} \left(\widehat{\theta}_{|\mathbf{X}_0|}^X - \theta_0 \right) = \frac{\sum_{k=1}^{n} \left(X_k - \boldsymbol{\Psi}(\theta_0) \cdot \mathbf{X}_{k-1} \right)}{\sqrt{\sum_{k=1}^{n} \boldsymbol{a} \cdot \mathbf{X}_{k-1}}}. \tag{4.2.21}$$

By (4.1.1),

$$X_k - \boldsymbol{\Psi}(\theta_0) \cdot \mathbf{X}_{k-1} = \sum_{i=1}^{d} \sum_{j=1}^{X_{k-i}} \left(\zeta_{k-i,k,j} - \Psi_i\left(\theta_0\right) \right) =: \sum_{i=1}^{d} \sum_{j=1}^{X_{k-i}} \mathring{\zeta}_{k-i,k,j}, \tag{4.2.22}$$

where the $\left\{\zeta_{k-i,k,j}\right\}_j$ are i.i.d. given $\mathcal{F}_{k-1}$, following a Poisson distribution with parameter $\Psi_i(\theta_0)$, and the $\left\{\zeta_{k-i,k,j}\right\}_{i,j}$ are independent given $\mathcal{F}_{k-1}$. Renumbering the $\mathring{\zeta}_{k-i,k,j}$ we then obtain

$$\sum_{k=1}^{n} \left(X_k - \boldsymbol{\Psi}(\theta_0) \cdot \mathbf{X}_{k-1} \right) = \sum_{i=1}^{d} \sum_{j=1}^{\sum_{k=1}^{n} X_{k-i}} \mathring{\zeta}_{k-i,k,j}. \tag{4.2.23}$$

Applying a central limit theorem for the sum of a random number of independent random variables (see *e.g.* [Blum63]), we obtain that for all $i = 1 \dots d$,

$$\lim_{|\mathbf{X}_0| \to \infty} \frac{\sum_{j=1}^{\sum_{k=1}^{n} X_{k-i}} \mathring{\zeta}_{k-i,k,j}}{\sqrt{\sum_{k=1}^{n} X_{k-i}}} \overset{\mathcal{D}}{=} \mathcal{N}\left(0, a_i \theta_0 + b_i\right). \tag{4.2.24}$$

We have used the fact that $|\mathbf{X}_0|$ is a real positive sequence growing to infinity, and $\sum_{k=1}^{n} X_{k-i}$ a sequence of integered-valued random variables such that $\sum_{k=1}^{n} X_{k-i}/\,|\mathbf{X}_0|$ converges in probability to a finite random variable. In our case this last limit is actually deterministic, since we have shown in Lemma 4.2.2 that

$$\lim_{|\mathbf{X}_0|\to\infty} \frac{\sum_{k=1}^{n} X_{k-i}}{|\mathbf{X}_0|} \overset{a.s.}{=} \sum_{k=1}^{n}\sum_{j=1}^{d} \alpha_j m_{ji}^{(k-1)}(\theta_0).$$

Using (4.2.23) in (4.2.21), we write

$$\sqrt{\sum_{k=1}^{n} \boldsymbol{a}\cdot\mathbf{X}_{k-1}}\left(\widehat{\theta}^X_{|\mathbf{X}_0|}-\theta_0\right) = \sum_{i=1}^{d} \frac{\sum_{j=1}^{\sum_{k=1}^{n} X_{k-i}} \mathring{\zeta}_{k-i,k,j}}{\sqrt{\sum_{k=1}^{n} X_{k-i}}} \frac{\sqrt{\sum_{k=1}^{n} X_{k-i}}}{\sqrt{\sum_{k=1}^{n} \boldsymbol{a}\cdot\mathbf{X}_{k-1}}}. \tag{4.2.25}$$

Using again Lemma 4.2.2,

$$\lim_{|\mathbf{X}_0|\to\infty} \frac{\sqrt{\sum_{k=1}^{n} X_{k-i}}}{\sqrt{\sum_{k=1}^{n} \boldsymbol{a}\cdot\mathbf{X}_{k-1}}} \overset{a.s.}{=} \sqrt{\frac{\sum_{k=1}^{n}\sum_{j=1}^{d} \alpha_j m_{ji}^{(k-1)}(\theta_0)}{\sum_{k=1}^{n}\sum_{j=1}^{d}\sum_{l=1}^{d} \alpha_j a_l m_{jl}^{(k-1)}(\theta_0)}},$$

which, combined to (4.2.24) and (4.2.25), implies by Slutsky's Lemma that

$$\lim_{|\mathbf{X}_0|\to\infty} \sqrt{\sum_{k=1}^{n} \boldsymbol{a}\cdot\mathbf{X}_{k-1}}\left(\widehat{\theta}^X_{|\mathbf{X}_0|}-\theta_0\right) \overset{\mathcal{D}}{=} \mathcal{N}\left(0,\sigma^2\left(\theta_0\right)\right). \tag{4.2.26}$$

By (4.2.15) and (4.2.17),

$$\lim_{|\mathbf{X}_0|} \frac{\sqrt{\sigma^2(\theta_0)}}{\sqrt{\sigma^2(\widehat{\theta}^X_{|\mathbf{X}_0|})}} \overset{a.s.}{=} 1,$$

from which we deduce (4.2.18). □

Remark 4.2.3. We point out that we do not use the Poissonian character of the transitions of the process (4.1.1) to derive the properties of $\widehat{\theta}^X_{|\mathbf{X}_0|}$, but we simply need its first and second order moments. This estimator can thus be applied to any process of the form (4.1.1), where the $\{\zeta_{n-k,n,i}\}_i$ do not necessarily follow a Poisson distribution, but satisfy $\mathbb{E}_{\theta_0}\left(\zeta_{n-k,n,i}|\mathcal{F}_{n-1}\right) = \Psi_k(\theta_0)$. The variance should be either known, or previously estimated, and the process should be normalized accordingly such that the error term in the stochastic regression equation remains bounded.

4.2.2 A CLSE with asymptotic properties, as $n\to\infty$

In this section we are interested in building an estimator of θ_0 suitable for populations modeled by (4.1.1), in the subcritical case, hence for *vanishing populations*, or populations in their *decay phase*. The estimator $\widehat{\theta}^X_{|\mathbf{X}_0|}$ introduced in Subsection 4.2.1 might be suitable, since its asymptotic properties are valid for any class of criticality. However, we might not be in the case where $|\mathbf{X}_0|$ is large enough to assume that these asymptotic properties are applicable. We thus whish to build an estimator with asymptotic properties as the number of observations n tends to infinity.

For this purpose, and because of the almost sure extinction of the process $(\mathbf{X}_k)_{k\geqslant 0}$ in the subcritical case, we consider instead of $(\mathbf{X}_k)_{k\geqslant 0}$, the associated process *conditioned on non-extinction at each time-step*. By this we mean the homogeneous Markov chain $(\mathbf{Z}_k)_{k\geqslant 0}$ defined by the following transition probabilities: for all $\mathbf{i},\mathbf{j}\in\mathbb{N}^d\backslash\{\mathbf{0}\}$,

$$\begin{aligned} Q(\mathbf{i},\mathbf{j}) :=& \mathbb{P}_{\theta_0}\left(\mathbf{Z}_k=\mathbf{j}|\mathbf{Z}_{k-1}=\mathbf{i}\right) \\ =& \mathbb{P}_{\theta_0}\left(\mathbf{X}_k=\mathbf{j}|\mathbf{X}_{k-1}=\mathbf{i},\mathbf{X}_k\neq\mathbf{0}\right). \end{aligned} \tag{4.2.27}$$

We thus take into account the information that at each time-step k, the population does not become extinct $\mathbf{0}$, which is the case for every $k \leqslant n$ if the population is still extant at time n (by this we mean the population with its memory, *i.e.* $\mathbf{X}_n \neq \mathbf{0}$). Denoting by $P(\mathbf{i}, \mathbf{j})$ the transition probabilities of the process $(\mathbf{X}_k)_{k\geqslant 0}$, we thus have,

$$Q(\mathbf{i}, \mathbf{j}) = \frac{P(\mathbf{i}, \mathbf{j})}{1 - P(\mathbf{i}, \mathbf{0})}, \quad \mathbf{i}, \mathbf{j} \in \mathbb{N}^d \backslash \{\mathbf{0}\}. \tag{4.2.28}$$

By definition of $(\mathbf{X}_k)_{k\geqslant 0}$, $P(\mathbf{i}, \mathbf{j}) = 0$ as soon as $(j_2, \ldots, j_d) \neq (i_1, \ldots, i_{d-1})$. Hence the same holds for $Q(\mathbf{i}, \mathbf{j})$, and the process $(\mathbf{Z}_k)_{k\geqslant 0}$ consequently satisfies for all $i = 2 \ldots d$ and $k \geqslant 0$,

$$Z_{k,i} \overset{a.s.}{=} Z_{k-1,i-1}.$$

We define the one-dimensional d-Markovian process $(Z_k)_{k\geqslant 0}$ corresponding to the first coordinate, for all $k \geqslant 0$,

$$Z_k := Z_{k,1}. \tag{4.2.29}$$

Let us compute explicitly its transition probabilities. For all d-dimensional vector $\mathbf{u}$, we define the truncated sum

$$\lceil \mathbf{u} \rceil := u_1 + \ldots + u_{d-1}. \tag{4.2.30}$$

By definition,

$$\begin{aligned} \mathbb{P}_{\theta_0}\left(Z_k = j | \mathbf{Z}_{k-1}\right) &= Q\left(\mathbf{Z}_{k-1}, (j, Z_{k-1}, Z_{k-2}, \ldots, Z_{k-d+1})\right) \\ &= \frac{P\left(\mathbf{Z}_{k-1}, (j, Z_{k-1}, Z_{k-2}, \ldots, Z_{k-d+1})\right)}{1 - P\left(\mathbf{Z}_{k-1}, \mathbf{0}\right)}, \end{aligned}$$

with

$$P\left(\mathbf{Z}_{k-1}, \mathbf{0}\right) = \begin{cases} e^{-\Psi_d(\theta_0) Z_{k-d}} \text{ if } Z_{k-1} = \ldots = Z_{k-d+1} = 0, \\ 0 \text{ otherwise.} \end{cases}$$

Hence

$$\mathbb{P}_{\theta_0}\left(Z_k = j | \mathbf{Z}_{k-1}\right) = \frac{\left(\mathbf{\Psi}(\theta_0) \cdot \mathbf{Z}_{k-1}\right)^j e^{-\mathbf{\Psi}(\theta_0) \cdot \mathbf{Z}_{k-1}}}{j! \left(1 - \mathbf{1}_{\{\lceil \mathbf{Z}_{k-1} \rceil = 0\}} e^{-\Psi_d(\theta_0) Z_{k-d}}\right)}. \tag{4.2.31}$$

In this section we study a CLSE associated with the process $(Z_k)_{k\geqslant 0}$, and obtain asymptotic properties as the number of observations n tends to infinity, in the subcritical case. Hence this estimator is particularly adapted for the study of the decay phase of the population, even if the number of individuals at the beginning of this phase is not very large.

We consider the CLSE corresponding to the normalized process $Z_k / \sqrt{\boldsymbol{a} \cdot \mathbf{Z}_{k-1}}$,

$$\begin{aligned} \widehat{\theta}_n^Z &:= \arg\min_{\theta \in \Theta} S_n(\theta), \\ S_n(\theta) &:= \sum_{k=1}^n \left(\frac{Z_k}{\sqrt{\boldsymbol{a} \cdot \mathbf{Z}_{k-1}}} - f(\theta, \mathbf{Z}_{k-1}) \right)^2, \end{aligned} \tag{4.2.32}$$

where

$$\begin{aligned} f(\theta_0, \mathbf{Z}_{k-1}) &:= \mathbb{E}_{\theta_0}\left(\frac{Z_k}{\sqrt{\boldsymbol{a} \cdot \mathbf{Z}_{k-1}}} \Big| \mathbf{Z}_{k-1} \right) \\ &= \frac{\mathbf{\Psi}(\theta_0) \cdot \mathbf{Z}_{k-1}}{\sqrt{\boldsymbol{a} \cdot \mathbf{Z}_{k-1}} \left(1 - \mathbf{1}_{\{\lceil \mathbf{Z}_{k-1} \rceil = 0\}} e^{-\Psi_d(\theta_0) Z_{k-d}}\right)}. \end{aligned} \tag{4.2.33}$$

Denoting by f' the derivative of f with respect to θ, we thus have, for all $\theta \in \Theta$ and all $\mathbf{j} \in \mathbb{N}^d$, $\mathbf{j} \neq \mathbf{0}$,

$$f'(\theta, \mathbf{j}) = \begin{cases} \sqrt{a_d j_d} \frac{1 - (1 + \Psi_d(\theta) j_d) e^{-\Psi_d(\theta) j_d}}{\left(1 - e^{-\Psi_d(\theta) j_d}\right)^2} & \text{if } \lceil \mathbf{j} \rceil = 0, \\ \sqrt{\boldsymbol{a} \cdot \mathbf{j}} & \text{otherwise.} \end{cases} \tag{4.2.34}$$

We finally define

$$\varepsilon_k := \frac{Z_k}{\sqrt{\boldsymbol{a} \cdot \mathbf{Z}_{k-1}}} - f(\theta_0, \mathbf{Z}_{k-1}), \tag{4.2.35}$$

which implies that

$$\mathbb{E}_{\theta_0}\left(\varepsilon_k^2 | \mathbf{Z}_{k-1}\right) = \frac{\boldsymbol{\Psi}(\theta_0) \cdot \mathbf{Z}_{k-1}}{\boldsymbol{a} \cdot \mathbf{Z}_{k-1}\left(1 - \mathbf{1}_{\{\lceil \mathbf{Z}_{k-1} \rceil = 0\}} e^{-\Psi_d(\theta_0) Z_{k-d}}\right)}, \tag{4.2.36}$$

and the conditional variance of the error term ε_k in the stochastic regression equation is consequently bounded:

$$\theta_0 + \frac{\underline{\boldsymbol{b}}}{\overline{\boldsymbol{a}}} \leqslant \mathbb{E}_{\theta_0}\left(\varepsilon_k^2 | \mathbf{Z}_{k-1}\right) \leqslant \left(1 - e^{-\Psi_d(\theta_0)}\right)^{-1} \left(\theta_0 + \frac{\overline{\boldsymbol{b}}}{\underline{\boldsymbol{a}}}\right). \tag{4.2.37}$$

Theorem 4.2.4. *The estimator $\widehat{\theta}_n^Z$ is strongly consistent:*

$$\lim_{n \to \infty} \widehat{\theta}_n^Z \overset{a.s.}{=} \theta_0. \tag{4.2.38}$$

Proof. According to Proposition 3.1 in [Jac10B], sufficient conditions for the strong consistency of $\widehat{\theta}_n^Z$ are that $f(., \mathbf{Z}_{k-1})$ is Lipschitz, in the sense that there exists a nonnegative $\sigma(\mathbf{Z}_0, \ldots, \mathbf{Z}_{k-1})$-measurable function C_k satisfying for all θ_1, $\theta_2 \in \Theta$, $|f(\theta_1, \mathbf{Z}_{k-1}) - f(\theta_2, \mathbf{Z}_{k-1})| \overset{a.s.}{\leqslant} C_k\, |\theta_1 - \theta_2|$, that $\overline{\lim}_{k \to \infty} \mathbb{E}_{\theta_0}\left(\varepsilon_k^2 | \mathbf{Z}_{k-1}\right) \overset{a.s.}{<} \infty$, and that

$$\lim_{n \to \infty} \inf_{|\theta - \theta_0| \geqslant \delta} \sum_{k=1}^{n} \left(f(\theta_0, \mathbf{Z}_{k-1}) - f(\theta, \mathbf{Z}_{k-1})\right)^2 \overset{a.s.}{=} \infty. \tag{4.2.39}$$

The Lipschitz condition is satisfied thanks to (4.2.34), which shows that $f'(., \mathbf{Z}_{k-1})$ is bounded on Θ.

The second condition follows from (4.2.37).

Let $\delta > 0$ and $\theta \in \Theta$ such that $|\theta - \theta_0| \geqslant \delta$. We assume for convenience that $\theta_0 > \theta$. In order to prove (4.2.39), we apply the mean value theorem to the function $f(., \mathbf{Z}_{k-1})$, and obtain that there exists some $\tilde{\theta}_k \in]\theta, \theta_0[$ such that

$$f'(\tilde{\theta}_k, \mathbf{Z}_{k-1}) = \frac{f(\theta_0, \mathbf{Z}_{k-1}) - f(\theta, \mathbf{Z}_{k-1})}{\theta_0 - \theta}.$$

Consequently,

$$\begin{aligned}
\sum_{k=1}^{n} \left(f(\theta_0, \mathbf{Z}_{k-1}) - f(\theta, \mathbf{Z}_{k-1})\right)^2 &= (\theta_0 - \theta)^2 \sum_{k=1}^{n} \left(f'(\tilde{\theta}_k, \mathbf{Z}_{k-1})\right)^2 \\
&= (\theta_0 - \theta)^2 \sum_{k=1}^{n} \boldsymbol{a} \cdot \mathbf{Z}_{k-1} \frac{\left(1 - \mathbf{1}_{\{\lceil \mathbf{Z}_{k-1} \rceil = 0\}} \left(1 + \Psi_d(\tilde{\theta}_k) Z_{k-d}\right) e^{-\Psi_d(\tilde{\theta}_k) Z_{k-d}}\right)^2}{\left(1 - \mathbf{1}_{\{\lceil \mathbf{Z}_{k-1} \rceil = 0\}} e^{-\Psi_d(\tilde{\theta}_k) Z_{k-d}}\right)^4} \\
&\geqslant (\theta_0 - \theta)^2 \left(1 - \left(1 + \Psi_d(\theta_1)\right) e^{-\Psi_d(\theta_1)}\right)^2 \sum_{k=1}^{n} \boldsymbol{a} \cdot \mathbf{Z}_{k-1} \\
&\geqslant \delta^2 \left(1 - \left(1 + \Psi_d(\theta_1)\right) e^{-\Psi_d(\theta_1)}\right)^2 \underline{\boldsymbol{a}}\, n,
\end{aligned}$$

which implies (4.2.39). □

In order to study the asymptotic distribution of $\widehat{\theta}_n^Z$, we shall prove the positive recurrence of the Markov chain $(\mathbf{Z}_k)_{k \geqslant 0}$, and the finiteness of the first and second-order moments of its stationary distribution.

Remark 4.2.5. We point out that the Yaglom distribution ν_{θ_0} associated with the subcritical process $(\mathbf{X}_k)_{k\geqslant 0}$, defined as follows (see *e.g.* [AthNey72], Theorem 5.4.2),

$$\lim_{n\to\infty} \mathbb{P}_{\theta_0}\left(\mathbf{X}_n = \mathbf{i}|\mathbf{X}_0 = \mathbf{j},\, \mathbf{X}_n \neq \mathbf{0}\right) = \nu_{\theta_0}(\mathbf{i}), \quad \mathbf{i},\mathbf{j} \in \mathbb{N}^d\backslash\{\mathbf{0}\}, \tag{4.2.40}$$

and which is a *quasi-stationary distribution* for the Markov chain $(\mathbf{X}_k)_{k\geqslant 0}$, is not a *stationary distribution* for $(\mathbf{Z}_k)_{k\geqslant 0}$. In other words (denoting just for the purpose of this remark the initial distribution as a subscript), we have

$$\mathbb{P}_{\theta_0,\nu_{\theta_0}}\left(\mathbf{X}_1 = \mathbf{j}\,|\,\mathbf{X}_1 \neq \mathbf{0}\right) = \nu_{\theta_0}\left(\mathbf{j}\right), \tag{4.2.41}$$

which does not imply in general that

$$\mathbb{P}_{\theta_0,\nu_{\theta_0}}\left(\mathbf{Z}_1 = \mathbf{j}\right) = \nu_{\theta_0}\left(\mathbf{j}\right). \tag{4.2.42}$$

Indeed, (4.2.41) means that

$$\frac{\sum_{\mathbf{i}\in\mathbb{N}^d\backslash\{\mathbf{0}\}} \nu_{\theta_0}(\mathbf{i}) P\left(\mathbf{i},\mathbf{j}\right)}{1 - \sum_{\mathbf{i}\in\mathbb{N}^d\backslash\{\mathbf{0}\}} \nu_{\theta_0}\left(\mathbf{i}\right) P\left(\mathbf{i},\mathbf{0}\right)} = \nu_{\theta_0}\left(\mathbf{j}\right), \tag{4.2.43}$$

while (4.2.42) would mean

$$\sum_{\mathbf{i}\in\mathbb{N}^d\backslash\{\mathbf{0}\}} \nu_{\theta_0}(\mathbf{i}) Q\left(\mathbf{i},\mathbf{j}\right) = \nu_{\theta_0}\left(\mathbf{j}\right),$$

i.e.

$$\sum_{\mathbf{i}\in\mathbb{N}^d\backslash\{\mathbf{0}\}} \nu_{\theta_0}(\mathbf{i}) \frac{P\left(\mathbf{i},\mathbf{j}\right)}{1 - P\left(\mathbf{i},\mathbf{0}\right)} = \nu_{\theta_0}\left(\mathbf{j}\right). \tag{4.2.44}$$

In general, (4.2.43) does not imply (4.2.44), except if the Yaglom distribution is reduced to a Dirac measure.

However, we can prove that the Markov chain $(\mathbf{Z}_k)_{k\geqslant 0}$ has a unique stationary probability measure, *a priori* distinct from the Yaglom distribution.

Proposition 4.2.6. *Let us assume that the process $\left(\boldsymbol{X}_k\right)_{k\geqslant 0}$ is subcritical. Then the homogeneous Markov chain $(\boldsymbol{Z}_k)_{k\geqslant 0}$ is irreducible positive recurrent, and its stationary distribution λ_{θ_0} satisfies for all $i, j = 1 \ldots d$,*

$$\sum_{\boldsymbol{k}\in\mathbb{N}^d} k_i \lambda_{\theta_0}(\boldsymbol{k}) < \infty, \tag{4.2.45}$$

$$\sum_{\boldsymbol{k}\in\mathbb{N}^d} k_i k_j \lambda_{\theta_0}(\boldsymbol{k}) < \infty. \tag{4.2.46}$$

Proof. Clearly, the chain is irreducible: due to the Poisson random variables coming in play, any nonzero state is attainable from any other nonzero state in a finite time.

The positive recurrence then follows from a criterion given *e.g.* in [Twe75], Theorem 3.1: positive recurrence is implied if there exists a finite set $A \subset \mathbb{N}^d\backslash\{\mathbf{0}\}$ and a non-negative function g on $\mathbb{N}^d\backslash\{\mathbf{0}\}$ such that

$$\sum_{\mathbf{j}\in\mathbb{N}^d\backslash\{\mathbf{0}\}} Q(\mathbf{i},\mathbf{j}) g(\mathbf{j}) \leqslant g(\mathbf{i}) - 1, \quad \mathbf{i} \notin A. \tag{4.2.47}$$

We define $g(\mathbf{j}) := \sum_{k=1}^d a_k j_k$, where $a_1 = 1$, and for all $k = 2 \ldots d$,

$$0 < a_k < \min\left(a_{k-1} - \Psi_{k-1}\left(\theta_0\right), \Psi_k\left(\theta_0\right)\right). \tag{4.2.48}$$

The existence of such a_k is ensured by the fact that in the subcritical case, $\sum_{k=1}^d \Psi_k(\theta_0) < 1$ (see Proposition 4.1.1), and in particular $1 - \Psi_1(\theta_0) > 0$. We then have (see (4.2.33)) for all $\mathbf{i} \neq \mathbf{0}$

$$\begin{aligned} g(\mathbf{i}) - 1 - \sum_{\mathbf{j} \in \mathbb{N}^d \setminus \{\mathbf{0}\}} Q(\mathbf{i}, \mathbf{j}) g(\mathbf{j}) &= g(\mathbf{i}) - 1 - \mathbb{E}_{\theta_0}(Z_1 | \mathbf{Z}_0 = \mathbf{i}) - \sum_{k=2}^d a_k i_{k-1} \\ &= g(\mathbf{i}) - 1 - \frac{\boldsymbol{\Psi}(\theta_0) \cdot \mathbf{i}}{1 - \mathbf{1}_{\{\lceil \mathbf{i} \rceil = 0\}} e^{-\Psi_d(\theta_0) i_d}} - \sum_{k=2}^d a_k i_{k-1} \\ &\sim_{|\mathbf{i}| \to \infty} g(\mathbf{i}) - 1 - \boldsymbol{\Psi}(\theta_0) \cdot \mathbf{i} - \sum_{k=2}^d a_k i_{k-1} \\ &= \sum_{k=1}^{d-1} (a_k - \Psi_k(\theta_0) - a_{k+1}) i_k + (a_d - \Psi_d(\theta_0)) i_d - 1. \end{aligned}$$

Hence

$$\lim_{|\mathbf{i}| \to \infty} \big(g(\mathbf{i}) - 1 - \sum_{\mathbf{j} \in \mathbb{N}^d \setminus \{\mathbf{0}\}} Q(\mathbf{i}, \mathbf{j}) g(\mathbf{j}) \big) = +\infty.$$

Consequently there exists some finite set A satisfying (4.2.47), and the chain $(\mathbf{Z}_k)_{k \geqslant 0}$ is positive recurrent.

Let us now prove that its stationary measure λ_{θ_0} admits finite first-order moments. First, we point out that by definition of $\mathbf{Z}_n$, we have for all $i = 1 \dots d$,

$$\sum_{\mathbf{k} \in \mathbb{N}^d} k_i \lambda_{\theta_0}(\mathbf{k}) = \mathbb{E}_{\theta_0}\left(\lim_{n \to \infty} Z_{n,i} \right) = \mathbb{E}_{\theta_0}\left(\lim_{n \to \infty} Z_{n-i+1,1} \right) = \sum_{\mathbf{k} \in \mathbb{N}^d} k_1 \lambda_{\theta_0}(\mathbf{k}) =: m^{\lambda_{\theta_0}}. \tag{4.2.49}$$

It is thus enough to demonstrate (4.2.45) for $i = 1$. According to [Twe83], Theorem 1, in order that $\sum_{\mathbf{j} \in \mathbb{N}^d \setminus \{\mathbf{0}\}} \varphi(\mathbf{j}) \lambda_{\theta_0}(\mathbf{j}) < \infty$ for some given non-negative function φ, it suffices that for some non-empty finite set B and some function h with $h(\mathbf{j}) \geqslant \varphi(\mathbf{j})$, $\mathbf{j} \notin B$,

$$\sum_{\mathbf{j} \in \mathbb{N}^d \setminus \{\mathbf{0}\}} Q(\mathbf{i}, \mathbf{j}) h(\mathbf{j}) \leqslant h(\mathbf{i}) - \varphi(\mathbf{i}), \quad \mathbf{i} \notin B. \tag{4.2.50}$$

Taking $\varphi(\mathbf{j}) := j_1$ and $h(\mathbf{j}) := \sum_{k=1} b_k j_k$ with

$$b_1 = \frac{2}{1 - (\Psi_1(\theta_0) + \ldots + \Psi_d(\theta_0))}, \quad b_k = b_1 (\Psi_k(\theta_0) + \ldots + \Psi_d(\theta_0)), \ k = 2 \dots d, \tag{4.2.51}$$

we have for all $\mathbf{i} \neq \mathbf{0}$, $h(\mathbf{j}) \geqslant \varphi(\mathbf{j})$, and

$$\begin{aligned} h(\mathbf{i}) - \varphi(\mathbf{i}) - \sum_{\mathbf{j} \in \mathbb{N}^d \setminus \{\mathbf{0}\}} Q(\mathbf{i}, \mathbf{j}) h(\mathbf{j}) &\sim_{|\mathbf{i}| \to \infty} \sum_{k=1}^d b_k i_k - i_1 - b_1 \boldsymbol{\Psi}(\theta_0) \cdot \mathbf{i} - \sum_{k=2}^d b_k i_{k-1} \\ &= \sum_{k=1}^{d-1} (b_k - b_1 \Psi_k(\theta_0) - b_{k+1}) i_k - i_1 - (b_d - b_1 \Psi_d(\theta_0)) i_d \\ &= (b_1 - b_1 \Psi_1(\theta_0) - b_2 - 1) i_1 \\ &= \frac{1}{1 - (\Psi_1(\theta_0) + \ldots + \Psi_d(\theta_0))} i_1. \end{aligned}$$

Hence

$$\lim_{|\mathbf{i}| \to \infty} \big(h(\mathbf{i}) - f(\mathbf{i}) - \sum_{\mathbf{j} \in \mathbb{N}^d \setminus \{\mathbf{0}\}} Q(\mathbf{i}, \mathbf{j}) h(\mathbf{j}) \big) = +\infty,$$

and there exists some finite set B satisfying (4.2.50). Consequently, $\sum_{\mathbf{j} \in \mathbb{N}^d \setminus \{\mathbf{0}\}} j_1 \lambda_{\theta_0}(\mathbf{j}) < \infty$, and the quantity $m^{\lambda_{\theta_0}}$ defined in (4.2.49) is finite.

Due to the specific properties of the process $(\mathbf{Z}_k)_{k\geqslant 0}$, it is possible to deduce from this that the second-order moments of the stationary measure λ_{θ_0} are finite as well. Indeed, for all $i = 1\dots d-1$, using the inequality $x\,(1-e^{-x})^{-1} \leqslant 1+x$, $x \geqslant 0$,

$$\mathbb{E}_{\theta_0}\left(Z_n Z_{n-i}\right) = \mathbb{E}_{\theta_0}\left[Z_{n-i}\frac{\mathbf{\Psi}(\theta_0)\cdot\mathbf{Z}_{n-1}}{1-\mathbf{1}_{\{\lceil \mathbf{Z}_{n-1}\rceil=0\}}e^{-\Psi_d(\theta_0)Z_{n-d}}}\right] \leqslant \mathbb{E}_{\theta_0}\left[Z_{n-i}\Big(1+\mathbf{\Psi}(\theta_0)\cdot\mathbf{Z}_{n-1}\Big)\right],$$

hence

$$\varlimsup_{n\to\infty} \mathbb{E}_{\theta_0}\left(Z_n Z_{n-i}\right) \leqslant m^{\lambda_{\theta_0}} + \max_{k=0\dots d-1}\varlimsup_{n\to\infty} \mathbb{E}_{\theta_0}\left(Z_n Z_{n-k}\right)\sum_{j=1}^{d}\Psi_j(\theta_0). \tag{4.2.52}$$

Similarly,

$$\begin{aligned}
\mathbb{E}_{\theta_0}\left(Z_n^2\right) &= \mathbb{E}_{\theta_0}\left[\frac{\mathbf{\Psi}(\theta_0)\cdot\mathbf{Z}_{n-1}}{1-\mathbf{1}_{\{\lceil \mathbf{Z}_{n-1}\rceil=0\}}e^{-\Psi_d(\theta_0)Z_{n-d}}}\left(1+\frac{\mathbf{\Psi}(\theta_0)\cdot\mathbf{Z}_{n-1}}{1-\mathbf{1}_{\{\lceil \mathbf{Z}_{n-1}\rceil=0\}}e^{-\Psi_d(\theta_0)Z_{n-d}}}\right)\right]\\
&\leqslant \mathbb{E}_{\theta_0}\left[\Big(2+\mathbf{\Psi}(\theta_0)\cdot\mathbf{Z}_{n-1}\Big)^2\right]\\
&= 4+4\sum_{j=1}^{d}\Psi_j(\theta_0)\mathbb{E}_{\theta_0}\left(Z_{n-j}\right)+\sum_{j=1}^{d}\sum_{l=1}^{d}\Psi_j(\theta_0)\Psi_l(\theta_0)\mathbb{E}_{\theta_0}\left(Z_{n-j}Z_{n-l}\right),
\end{aligned}$$

which by Fatou's lemma and (4.2.49) leads to (using the fact that $\sum_{j=1}^{d}\Psi_j(\theta_0) < 1$)

$$\varlimsup_{n\to\infty} \mathbb{E}_{\theta_0}\left(Z_n^2\right) \leqslant 4+4m^{\lambda_{\theta_0}} + \max_{k=0\dots d-1}\varlimsup_{n\to\infty} \mathbb{E}_{\theta_0}\left(Z_n Z_{n-k}\right)\sum_{j=1}^{d}\Psi_j(\theta_0).$$

Together with (4.2.52) this implies that

$$\max_{k=0\dots d-1}\varlimsup_{n\to\infty} \mathbb{E}_{\theta_0}\left(Z_n Z_{n-k}\right) \leqslant 4+4m^{\lambda_{\theta_0}} + \max_{k=0\dots d-1}\varlimsup_{n\to\infty} \mathbb{E}_{\theta_0}\left(Z_n Z_{n-k}\right)\sum_{j=1}^{d}\Psi_j(\theta_0),$$

and thus

$$\max_{k=0\dots d-1}\varlimsup_{n\to\infty} \mathbb{E}_{\theta_0}\left(Z_n Z_{n-k}\right) \leqslant \frac{4+4m^{\lambda_{\theta_0}}}{1-\sum_{j=1}^{d}\Psi_j(\theta_0)} < \infty.$$

We then obtain by means of Fatou's lemma that for every $i, j = 1\dots d$,

$$\begin{aligned}
\sum_{\mathbf{k}\in\mathbb{N}^d} k_i k_j \lambda_{\theta_0}(\mathbf{k}) &= \mathbb{E}_{\theta_0}\left(\lim_{n\to\infty} Z_{n,i}Z_{n,j}\right) = \mathbb{E}_{\theta_0}\left(\lim_{n\to\infty} Z_n Z_{n-|i-j|}\right)\\
&\leqslant \varlimsup_{n\to\infty} \mathbb{E}_{\theta_0}\left(Z_n Z_{n-|i-j|}\right) \leqslant \max_{k=0\dots d-1}\varlimsup_{n\to\infty} \mathbb{E}_{\theta_0}\left(Z_n Z_{n-k}\right) < \infty.
\end{aligned}$$

□

Let us finally quote from [Bil61], Theorem 1.1 and 1.3, the following strong law of large numbers for homogeneous irreducible positive recurrent Markov chains, which can be applied here. For every λ_{θ_0}-integrable function $g : \mathbb{N}^d\backslash\{\mathbf{0}\} \to \mathbb{R}$,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1} g(\mathbf{Z}_k) \overset{a.s.}{=} \sum_{\mathbf{j}\in\mathbb{N}^d} g(\mathbf{j})\lambda_{\theta_0}(\mathbf{j}). \tag{4.2.53}$$

We can now prove the following theorem, which states the asymptotic normality of the estimator $\widehat{\theta}_n^Z$ as $n\to\infty$.

Theorem 4.2.7. *Let us assume that the process $(\boldsymbol{X}_k)_{k\geqslant 0}$ is subcritical. Then the estimator $\widehat{\theta}_n^Z$ is asymptotically normally distributed:*

$$\lim_{n\to\infty}\sqrt{n\frac{\left(\sum_{\boldsymbol{j}\in\mathbb{N}^d}\left(f'(\theta_0,\boldsymbol{j})\right)^2\lambda_{\theta_0}(\boldsymbol{j})\right)^2}{\sum_{\boldsymbol{j}\in\mathbb{N}^d}\left(f'(\theta_0,\boldsymbol{j})\right)^2 f(\theta_0,\boldsymbol{j})\left(\boldsymbol{a}\cdot\boldsymbol{j}\right)^{-1/2}\lambda_{\theta_0}(\boldsymbol{j})}}\left(\widehat{\theta}_n^Z-\theta_0\right)\overset{\mathcal{D}}{=}\mathcal{N}\left(0,1\right), \tag{4.2.54}$$

where f is given by (4.2.33).

Proof. We follow the steps of the proof of Proposition 6.1 in [Jac10B], which provides sufficient conditions to obtain the asymptotic behavior of a general conditional least squares estimator.

Writing the Taylor expansion of $S_n'(\theta)$ in the neighborhood of θ_0, we obtain that

$$\widehat{\theta}_n^Z-\theta_0=-\frac{S_n'(\theta_0)}{S_n''(\tilde{\theta}_n)}, \tag{4.2.55}$$

for some $\tilde{\theta}_n=\theta_0+t_n\left(\widehat{\theta}_n^Z-\theta_0\right)$, with $t_n\in\,]0,1[$. Since $S_n'(\theta_0)=-2\sum_{k=1}^n\varepsilon_k f'(\theta_0,\mathbf{Z}_{k-1})$, we can write

$$\sqrt{n}\left(\widehat{\theta}_n^Z-\theta_0\right)=\frac{\sum_{k=1}^n\varepsilon_k f'(\theta_0,\mathbf{Z}_{k-1})}{\sqrt{n}}\left(\frac{F_n}{n}\right)^{-1}\left(\frac{1}{2}\frac{S_n''(\tilde{\theta}_n)}{F_n}\right)^{-1}, \tag{4.2.56}$$

where

$$F_n:=\sum_{k=1}^n\left(f'(\theta_0,\mathbf{Z}_{k-1})\right)^2.$$

By (4.2.34), for all $\mathbf{j}\in\mathbb{N}^d$, $\mathbf{j}\neq\mathbf{0}$,

$$0\leqslant f'(\theta_0,\mathbf{j})\leqslant\frac{\sqrt{\boldsymbol{a}\cdot\mathbf{j}}}{(1-e^{-\Psi_d(\theta_0)})^2},$$

hence we deduce by means of (4.2.53) and (4.2.45) that

$$\lim_{n\to\infty}\frac{F_n}{n}\overset{a.s.}{=}\sum_{\mathbf{j}\in\mathbb{N}^d}\left(f'(\theta_0,\mathbf{j})\right)^2\lambda_{\theta_0}(\mathbf{j}). \tag{4.2.57}$$

In view of (4.2.56), we now prove that

$$\lim_{n\to\infty}\frac{S_n''(\tilde{\theta}_n)}{F_n}\overset{a.s.}{=}2. \tag{4.2.58}$$

Computing S_n'' thanks to the formula $S_n(\theta)=\sum_{k=1}^n\left(\varepsilon_k+f(\theta_0,\mathbf{Z}_{k-1})-f(\theta,\mathbf{Z}_{k-1})\right)^2$, it appears that (4.2.58) is true, as soon as the following holds:

$$\lim_{n\to\infty}\sup_{\theta\in\Theta}\frac{\left|\sum_{k=1}^n\varepsilon_k\,f''(\theta,\mathbf{Z}_{k-1})\right|}{F_n}\overset{a.s.}{=}0, \tag{4.2.59}$$

$$\lim_{n\to\infty}\frac{\sum_{k=1}^n\left(f'(\tilde{\theta}_n,\mathbf{Z}_{k-1})\right)^2}{F_n}\overset{a.s.}{=}1, \tag{4.2.60}$$

and

$$\lim_{n\to\infty}\frac{\sum_{k=1}^n\left(f(\theta_0,\mathbf{Z}_{k-1})-f(\tilde{\theta}_n,\mathbf{Z}_{k-1})\right)f''(\tilde{\theta}_n,\mathbf{Z}_{k-1})}{F_n}\overset{a.s.}{=}0. \tag{4.2.61}$$

Let us prove (4.2.59)-(4.2.61). Note that, for every $\mathbf{j} \neq \mathbf{0}$, $f''(\theta, \mathbf{j}) = 0$ if $\lceil \mathbf{j} \rceil \neq 0$, and

$$f''(\theta, \mathbf{j}) = \frac{(a_d j_d)^{3/2} \, e^{-\Psi_d(\theta) j_d} \left[e^{-\Psi_d(\theta) j_d} \left(\Psi_d(\theta) \, j_d + 2 \right) + \Psi_d(\theta) \, j_d - 2 \right]}{\left(1 - e^{-\Psi_d(\theta) j_d}\right)^3}$$

otherwise.

First, (4.2.59) is given by a strong law of large numbers proved in [Jac10B], Proposition 5.1. The latter can be indeed applied since $f''(., \mathbf{Z}_{k-1})$ fulfills the required Lipschitz condition, and $\lim_n F_n \overset{a.s.}{=} \infty$ (as an immediate consequence of the stronger result (4.2.57)).

In view of (4.2.60) we consider the function $\left(f'(\theta, \mathbf{j})\right)^2$ and its derivative $2f'(\theta, \mathbf{j}) f''(\theta, \mathbf{j})$. For all $\theta \in \overline{\Theta}$ and all $\mathbf{j} \neq \mathbf{0}$ with $\lceil \mathbf{j} \rceil = 0$,

$$\begin{aligned} |2f'(\theta, \mathbf{j}) f''(\theta, \mathbf{j})| &\leqslant 4 \frac{(a_d j_d)^2 \, e^{-\Psi_d(\theta) j_d} \left(\Psi_d(\theta) \, j_d + 2 \right)}{\left(1 - e^{-\Psi_d(\theta) j_d}\right)^5} \\ &\leqslant \frac{4 \max_{x \geqslant 0} (x+2)^3 \, e^{-x}}{\left(1 - e^{-\Psi_d(\theta_{min})}\right)^5} =: c_1. \end{aligned} \tag{4.2.62}$$

Consequently,

$$\frac{\left| \sum_{k=1}^n f'(\tilde{\theta}_n, \mathbf{Z}_{k-1})^2 - f'(\theta_0, \mathbf{Z}_{k-1})^2 \right|}{F_n} \leqslant c_1 \left| \widehat{\theta}_n^Z - \theta_0 \right| \left(\frac{F_n}{n} \right)^{-1}, \tag{4.2.63}$$

which by (4.2.57) and the strong consistency of $\widehat{\theta}_n^Z$ almost surely tends to 0. Writing

$$\frac{\sum_{k=1}^n f'(\tilde{\theta}_n, \mathbf{Z}_{k-1})^2}{F_n} = 1 + \frac{\sum_{k=1}^n \left(f'(\tilde{\theta}_n, \mathbf{Z}_{k-1})^2 - f'(\theta_0, \mathbf{Z}_{k-1})^2 \right)}{F_n},$$

this implies (4.2.60).

It now remains to prove (4.2.61). With similar computations as above, one shows that there exists a deterministic constant $c_2 > 0$ such that

$$\frac{\left| \sum_{k=1}^n \left(f(\theta_0, \mathbf{Z}_{k-1}) - f(\tilde{\theta}_n, \mathbf{Z}_{k-1}) \right) f''(\tilde{\theta}_n, \mathbf{Z}_{k-1}) \right|}{F_n} \leqslant c_2 |\widehat{\theta}_n^Z - \theta_0| \left(\frac{F_n}{n} \right)^{-1},$$

which thanks to (4.2.57) and the strong consistency of $\widehat{\theta}_n^Z$ implies (4.2.61).

Finally, in order to prove (4.2.56), we want to show that $\sum_{k=1}^n \varepsilon_k f'(\theta_0, \mathbf{Z}_{k-1}) / \sqrt{n}$ converges in distribution, and for this purpose we make use of the following central limit theorem for sequences of martingales (see *e.g.* [Reb80] or [PraRao99]).

Proposition 4.2.8. *Let $\{M_k^{(n)}, \mathcal{F}_k^{(n)}, 1 \leqslant k \leqslant n\}$, $n \geqslant 1$ be a sequence of square integrable martingales. For each $n \geqslant 1$, we denote by $\langle M \rangle^{(n)} = (\langle M \rangle_k^{(n)})_{1 \leqslant k \leqslant n}$ the associated Meyer process. We assume that there exists a constant c such that $\lim_{n \to \infty} \langle M_n \rangle^{(n)} \overset{\mathcal{P}}{=} c^2$, and assume moreover that for all $\varepsilon > 0$,*

$$\lim_{n \to \infty} \sum_{k=1}^n \mathbb{E} \left[\left| M_k^{(n)} - M_{k-1}^{(n)} \right|^2 \mathbf{1}_{\left\{ \left| M_k^{(n)} - M_{k-1}^{(n)} \right| \geqslant \varepsilon \right\}} \middle| \mathcal{F}_{k-1}^{(n)} \right] \overset{\mathcal{P}}{=} 0.$$

Then $\lim_{n \to \infty} M_n^{(n)} \overset{\mathcal{D}}{=} \mathcal{N}(0, c^2)$.

We will apply this Proposition to $\{M_k^{(n)}, \mathcal{F}_k^{(n)}, 1 \leqslant k \leqslant n\}$, defined as follows: for every $k \leqslant n$,

$$M_k^{(n)} := \frac{1}{\sqrt{n}} \sum_{l=1}^{k} \varepsilon_l f'(\theta_0, \mathbf{Z}_{l-1}).$$

First, for every $k \leqslant n$, $\mathbb{E}_{\theta_0}\left(\varepsilon_k f'(\theta_0, \mathbf{Z}_{k-1})/\sqrt{n} | \mathbf{Z}_{k-1}\right) = 0$. Second,

$$\mathbb{E}_{\theta_0}\left(\left(\frac{\varepsilon_k f'(\theta_0, \mathbf{Z}_{k-1})}{\sqrt{n}}\right)^2 \Big| \mathbf{Z}_{k-1}\right) = \frac{\left(f'(\theta_0, \mathbf{Z}_{k-1})\right)^2 f\left(\theta_0, \mathbf{Z}_{k-1}\right)}{n\sqrt{\boldsymbol{a} \cdot \mathbf{Z}_{k-1}}},$$

hence $M_k^{(n)}$ is a sequence of square integrable martingales. Moreover, by (4.2.45),

$$\sum_{\mathbf{j} \in \mathbb{N}^d} \frac{\left(f'\left(\theta_0, \mathbf{j}\right)\right)^2 f\left(\theta_0, \mathbf{j}\right)}{\sqrt{\boldsymbol{a} \cdot \mathbf{j}}} \lambda_{\theta_0}(\mathbf{j}) \leqslant \frac{1}{\left(1 - e^{-\Psi_d(\theta_0)}\right)^5} \sum_{\mathbf{j} \in \mathbb{N}^d} \boldsymbol{\Psi}(\theta_0) \cdot \mathbf{j} \lambda_{\theta_0}(\mathbf{j}) < \infty, \tag{4.2.64}$$

so by means of (4.2.53),

$$\begin{aligned}
\lim_{n \to \infty} \langle M_n \rangle^{(n)} &= \lim_{n \to \infty} \sum_{k=1}^{n} \mathbb{E}_{\theta_0}\left(\left(\frac{\varepsilon_k f'(\theta_0, \mathbf{Z}_{k-1})}{\sqrt{n}}\right)^2 \Big| \mathbf{Z}_{k-1}\right) \\
&\overset{a.s.}{=} \sum_{\mathbf{j} \in \mathbb{N}^d} \frac{\left(f'\left(\theta_0, \mathbf{j}\right)\right)^2 f\left(\theta_0, \mathbf{j}\right)}{\sqrt{\boldsymbol{a} \cdot \mathbf{j}}} \lambda_{\theta_0}(\mathbf{j}).
\end{aligned}$$

Third, using Cauchy-Schwarz and Bienaymé-Chebyshev inequalities,

$$\begin{aligned}
&\sum_{k=1}^{n} \mathbb{E}_{\theta_0}\left[\left|\frac{\varepsilon_k f'(\theta_0, \mathbf{Z}_{k-1})}{\sqrt{n}}\right|^2 \mathbf{1}_{\left\{\left|\frac{\varepsilon_k f'(\theta_0, \mathbf{Z}_{k-1})}{\sqrt{n}}\right| \geqslant \varepsilon\right\}} \Big| \mathbf{Z}_{k-1}\right] \\
&\quad \leqslant \sum_{k=1}^{n} \left(\mathbb{E}_{\theta_0}\left[\left|\frac{\varepsilon_k f'(\theta_0, \mathbf{Z}_{k-1})}{\sqrt{n}}\right|^4 \Big| \mathbf{Z}_{k-1}\right]\right)^{\frac{1}{2}} \left(\mathbb{P}_{\theta_0}\left[\left|\frac{\varepsilon_k f'(\theta_0, \mathbf{Z}_{k-1})}{\sqrt{n}}\right| \geqslant \varepsilon \Big| \mathbf{Z}_{k-1}\right]\right)^{\frac{1}{2}} \\
&\quad \leqslant \frac{1}{n^{\frac{3}{2}} \varepsilon} \sum_{k=1}^{n} |f'(\theta_0, \mathbf{Z}_{k-1})|^3 \left(\mathbb{E}_{\theta_0}\left[\varepsilon_k^4 | \mathbf{Z}_{k-1}\right]\right)^{\frac{1}{2}} \left(\mathbb{E}_{\theta_0}\left[\varepsilon_k^2 | \mathbf{Z}_{k-1}\right]\right)^{\frac{1}{2}}.
\end{aligned} \tag{4.2.65}$$

We have

$$|f'(\theta_0, \mathbf{Z}_{k-1})| \leqslant \frac{\sqrt{\boldsymbol{a} \cdot \mathbf{Z}_{k-1}}}{\left(1 - e^{-\Psi_d(\theta_0)}\right)^2}, \quad \mathbb{E}_{\theta_0}\left(\varepsilon_k^2 | \mathbf{Z}_{k-1}\right) \leqslant \frac{\boldsymbol{\Psi}(\theta_0) \cdot \mathbf{Z}_{k-1}}{\boldsymbol{a} \cdot \mathbf{Z}_{k-1}\left(1 - e^{-\Psi_d(\theta_0)}\right)},$$

and

$$\mathbb{E}_{\theta_0}\left(\varepsilon_k^4 | \mathbf{Z}_{k-1}\right) = \frac{\boldsymbol{\Psi}(\theta_0) \cdot \mathbf{Z}_{k-1}\left(1 + 3\boldsymbol{\Psi}(\theta_0) \cdot \mathbf{Z}_{k-1}\right)}{\left(\boldsymbol{a} \cdot \mathbf{Z}_{k-1}\right)^2 \left(1 - e^{-\Psi_d(\theta_0)}\right)},$$

hence

$$\begin{aligned}
&|f'(\theta_0, \mathbf{Z}_{k-1})|^3 \left(\mathbb{E}_{\theta_0}\left[\varepsilon_k^4 | \mathbf{Z}_{k-1}\right]\right)^{\frac{1}{2}} \left(\mathbb{E}_{\theta_0}\left[\varepsilon_k^2 | \mathbf{Z}_{k-1}\right]\right)^{\frac{1}{2}} \\
&\qquad \leqslant \frac{\boldsymbol{\Psi}(\theta_0) \cdot \mathbf{Z}_{k-1}\left(1 + 3\boldsymbol{\Psi}(\theta_0) \cdot \mathbf{Z}_{k-1}\right)^{\frac{1}{2}}}{\left(1 - e^{-\Psi_d(\theta_0)}\right)^7} \\
&\qquad \leqslant \frac{\boldsymbol{\Psi}(\theta_0) \cdot \mathbf{Z}_{k-1} + \sqrt{3}\left(\boldsymbol{\Psi}(\theta_0) \cdot \mathbf{Z}_{k-1}\right)^{\frac{3}{2}}}{\left(1 - e^{-\Psi_d(\theta_0)}\right)^7}.
\end{aligned} \tag{4.2.66}$$

Since the stationary distribution λ_{θ_0} has finite second-order moments (see Proposition 4.2.6), we deduce from (4.2.65) and (4.2.66) by virtue of (4.2.53) that

$$\lim_{n\to\infty}\sum_{k=1}^{n}\mathbb{E}_{\theta_0}\left[\left|\frac{\varepsilon_k f'(\theta_0,\mathbf{Z}_{k-1})}{\sqrt{n}}\right|^2 \mathbf{1}_{\left\{\left|\frac{\varepsilon_k f'(\theta_0,\mathbf{Z}_{k-1})}{\sqrt{n}}\right|\geqslant\varepsilon\right\}}\middle|\mathbf{Z}_{k-1}\right]\overset{a.s.}{=}0.$$

It then ensues from Proposition 4.2.8 that

$$\lim_{n\to\infty}\frac{\sum_{k=1}^{n}\varepsilon_k f'(\theta_0,\mathbf{Z}_{k-1})}{\sqrt{n}}\overset{\mathcal{D}}{=}\mathcal{N}\left(0,\sum_{\mathbf{j}\in\mathbb{N}^d}\frac{\left(f'\left(\theta_0,\mathbf{j}\right)\right)^2 f\left(\theta_0,\mathbf{j}\right)}{\sqrt{\boldsymbol{a}\cdot\mathbf{j}}}\lambda_{\theta_0}(\mathbf{j})\right). \tag{4.2.67}$$

Finally, (4.2.56) together with (4.2.57), (4.2.58), (4.2.67) and Slutsky's Lemma imply that

$$\lim_{n\to\infty}\sqrt{n}\left(\widehat{\theta}_n^Z-\theta_0\right)\overset{\mathcal{D}}{=}\mathcal{N}\left(0,\frac{\sum_{\mathbf{j}\in\mathbb{N}^d}\left(f'\left(\theta_0,\mathbf{j}\right)\right)^2 f\left(\theta_0,\mathbf{j}\right)\left(\boldsymbol{a}\cdot\mathbf{j}\right)^{-1/2}\lambda_{\theta_0}(\mathbf{j})}{\left(\sum_{\mathbf{j}\in\mathbb{N}^d}\left(f'\left(\theta_0,\mathbf{j}\right)\right)^2\lambda_{\theta_0}(\mathbf{j})\right)^2}\right). \tag{4.2.68}$$

□

Since the stationary distribution λ_{θ_0} is not explicitly known, Theorem 4.2.7 is not directly applicable. We can however deduce the following more practical result:

Corollary 4.2.9. *Let us assume that the process $(\boldsymbol{X}_k)_{k\geqslant 0}$ is subcritical. Then the estimator $\widehat{\theta}_n^Z$ has the following asymptotic distribution*

$$\lim_{n\to\infty}\frac{\sum_{k=0}^{n}\left(f'\left(\widehat{\theta}_n^Z,\boldsymbol{Z}_k\right)\right)^2}{\sqrt{\sum_{k=0}^{n}\left(f'\left(\widehat{\theta}_n^Z,\boldsymbol{Z}_k\right)\right)^2 f\left(\widehat{\theta}_n^Z,\boldsymbol{Z}_k\right)\left(\boldsymbol{a}\cdot\boldsymbol{Z}_k\right)^{-1/2}}}\left(\widehat{\theta}_n^Z-\theta_0\right)\overset{\mathcal{D}}{=}\mathcal{N}\left(0,1\right). \tag{4.2.69}$$

Proof. The result is immediate as soon as we prove that

$$\lim_{n\to\infty}\frac{1}{n+1}\sum_{k=0}^{n}\frac{\left(f'\left(\widehat{\theta}_n^Z,\mathbf{Z}_k\right)\right)^2 f\left(\widehat{\theta}_n^Z,\mathbf{Z}_k\right)}{\sqrt{\boldsymbol{a}\cdot\mathbf{Z}_k}}\overset{a.s.}{=}\sum_{\mathbf{j}\in\mathbb{N}^d}\frac{\left(f'\left(\theta_0,\mathbf{j}\right)\right)^2 f\left(\theta_0,\mathbf{j}\right)}{\sqrt{\boldsymbol{a}\cdot\mathbf{j}}}\lambda_{\theta_0}\left(\mathbf{j}\right), \tag{4.2.70}$$

as well as the equivalent result for the numerator. For this purpose, we write

$$\begin{aligned}\sum_{k=0}^{n}\frac{\left(f'\left(\widehat{\theta}_n^Z,\mathbf{Z}_k\right)\right)^2 f\left(\widehat{\theta}_n^Z,\mathbf{Z}_k\right)}{\sqrt{\boldsymbol{a}\cdot\mathbf{Z}_k}}&=\sum_{k=0}^{n}\frac{\left(f'\left(\theta_0,\mathbf{Z}_k\right)\right)^2 f\left(\theta_0,\mathbf{Z}_k\right)}{\sqrt{\boldsymbol{a}\cdot\mathbf{Z}_k}}\\&+\sum_{k=0}^{n}\left[\frac{\left(f'\left(\widehat{\theta}_n^Z,\mathbf{Z}_k\right)\right)^2 f\left(\widehat{\theta}_n^Z,\mathbf{Z}_k\right)}{\sqrt{\boldsymbol{a}\cdot\mathbf{Z}_k}}-\frac{\left(f'\left(\theta_0,\mathbf{Z}_k\right)\right)^2 f\left(\theta_0,\mathbf{Z}_k\right)}{\sqrt{\boldsymbol{a}\cdot\mathbf{Z}_k}}\right],\end{aligned} \tag{4.2.71}$$

and show that $\left(f'\left(.,\mathbf{j}\right)\right)^2 f\left(.,\mathbf{j}\right)\left(\boldsymbol{a}\cdot\mathbf{j}\right)^{-1/2}$ has a bounded derivative and is thus Lipschitz:

$$\left|\frac{2f''(\theta,\mathbf{j})f'(\theta,\mathbf{j})f(\theta,\mathbf{j})+(f'(\theta,\mathbf{j}))^3}{\sqrt{\boldsymbol{a}\cdot\mathbf{j}}}\right|\leqslant 2c_1\frac{\boldsymbol{\Psi}(\theta_{max})\cdot\mathbf{j}}{\left(1-e^{-\Psi_d(\theta_{min})}\right)^3}.$$

This enables us to write

$$\begin{aligned}\frac{1}{n+1}\sum_{k=0}^{n}&\left|\frac{\left(f'\left(\widehat{\theta}_n^Z,\mathbf{Z}_k\right)\right)^2 f\left(\widehat{\theta}_n^Z,\mathbf{Z}_k\right)}{\sqrt{\boldsymbol{a}\cdot\mathbf{Z}_k}}-\frac{\left(f'\left(\theta_0,\mathbf{Z}_k\right)\right)^2 f\left(\theta_0,\mathbf{Z}_k\right)}{\sqrt{\boldsymbol{a}\cdot\mathbf{Z}_k}}\right|\\&\leqslant\left|\widehat{\theta}_n^Z-\theta_0\right|\frac{2c_1}{\left(1-e^{-\Psi_d(\theta_{min})}\right)^3}\frac{1}{n+1}\sum_{k=0}^{n}\boldsymbol{\Psi}(\theta_{max})\cdot\mathbf{Z}_k.\end{aligned} \tag{4.2.72}$$

By the strong consistency of $\widehat{\theta}_n^Z$ together with (4.2.53) and (4.2.45), (4.2.72) almost surely tends to zero. Combined with (4.2.53) and (4.2.64) in (4.2.71), this implies (4.2.70). $\square$

Remark 4.2.10. Let us define the subset of $\mathbb{N}^d$

$$\mathcal{S} := \{(0,\ldots,0,x)\,,\ x \in \mathbb{N}^*\}\,.$$

Then, by construction of the estimators $\widehat{\theta}_{|\mathbf{X}_0|}^X$ and $\widehat{\theta}_n^Z$, we have for any $\mathbb{N}^d$-valued sequence $(\boldsymbol{x}_0,\ldots,\boldsymbol{x}_n)$ such that for all $k \leqslant n$, $x_{k,i} = x_{k-1,i-1}$ and such that $\boldsymbol{x}_k \notin \mathcal{S}$ (*i.e.* such that there is no sequence of $d-1$ successive zeros),

$$\widehat{\theta}_{|\mathbf{X}_0|}^X(\boldsymbol{x}_0,\ldots,\boldsymbol{x}_n) = \widehat{\theta}_n^Z(\boldsymbol{x}_0,\ldots,\boldsymbol{x}_n)\,. \tag{4.2.73}$$

Using this together with the fact that, defining

$$T^X := \inf\{k \geqslant 1, \mathbf{X}_k \in \mathcal{S}\}\,, \quad \text{and} \quad T^Z := \inf\{k \geqslant 1, \mathbf{Z}_k \in \mathcal{S}\}\,,$$

we have

$$\lim_{\substack{|\boldsymbol{x}_0|\to\infty \\ \boldsymbol{x}_0\notin\mathcal{S}}} \mathbb{P}\left(T^Z \leqslant n|\mathbf{Z}_0 = \boldsymbol{x}_0\right) = \lim_{\substack{|\boldsymbol{x}_0|\to\infty \\ \boldsymbol{x}_0\notin\mathcal{S}}} \mathbb{P}\left(T^X \leqslant n|\mathbf{X}_0 = \boldsymbol{x}_0\right) = 0,$$

we can deduce from Proposition 4.2.1 (but omit here the detailed proof) that, for $n \in \mathbb{N}$ fixed,

$$\lim_{|\mathbf{Z}_0|\to\infty} \widehat{\theta}_n^Z \overset{a.s.}{=} \theta_0, \tag{4.2.74}$$

and

$$\lim_{|\mathbf{Z}_0|\to\infty} \sqrt{\frac{\sum_{k=1}^n \boldsymbol{a}\cdot\mathbf{Z}_{k-1}}{\sigma^2(\widehat{\theta}_n^Z)}}\left(\widehat{\theta}_n^Z - \theta_0\right) \overset{\mathcal{D}}{=} \mathcal{N}(0,1)\,. \tag{4.2.75}$$

4.2.3 An explicit estimator with asymptotic properties, as $n \to \infty$

The aim of this section is to provide an estimator with asymptotic properties, as time n tends to infinity, in the supercritical case, which thus corresponds to the growth phase of the population.

For this purpose, we use the following estimator of the Perron's root introduced in [AsmKei78],

$$\widetilde{\rho}_n := \frac{|\mathbf{X}_1| + \ldots + |\mathbf{X}_n|}{|\mathbf{X}_0| + \ldots + |\mathbf{X}_{n-1}|}, \tag{4.2.76}$$

which has asymptotic properties as n tends to infinity in the supercritical case, on the set of nonextinction. In order to estimate θ_0, we use relation (4.2.7) and deduce the following explicit estimator of θ_0:

$$\widetilde{\theta}_n^X := \frac{1 - \sum_{k=1}^d b_k \widetilde{\rho}_n^{-k}}{\sum_{k=1}^d a_k \widetilde{\rho}_n^{-k}}. \tag{4.2.77}$$

All what follows can be applied to any process of the form (4.1.1), where the $\{\zeta_{n-k,n,i}\}_i$ do not necessarily follow a Poisson distribution, but satisfy $\mathbb{E}_{\theta_0}(\zeta_{n-k,n,i}|\mathcal{F}_{n-1}) = \Psi_k(\theta_0)$.

For each $i = 1\ldots d$ and $n \in \mathbb{N}$, we define the covariance matrix $\mathbf{V}_n^i$ (to avoid heavy notation we do not explicitly write the dependence in θ_0 of $\mathbf{V}_n^i$ and of the next introduced objects) with entries

$$\left[\mathbf{V}_n^i\right]_{jk} := \mathbb{E}_{\theta_0}(X_{n,j}X_{n,k}|\mathbf{X}_0 = \mathbf{e}_i) - \mathbb{E}_{\theta_0}(X_{n,j}|\mathbf{X}_0 = \mathbf{e}_i)\,\mathbb{E}_{\theta_0}(X_{n,k}|\mathbf{X}_0 = \mathbf{e}_i)\,.$$

In particular, $\left[\mathbf{V}_1^i\right]_{jk} = \Psi_i(\theta_0)$ if $j = k = i$, and is null otherwise.

Let $\boldsymbol{\xi}(\theta_0)$ and $\boldsymbol{\eta}(\theta_0)$ be the right and left eigenvector of $\mathbf{M}(\theta_0)$ for its Perron's root ρ_0, with normalization $\boldsymbol{\xi}(\theta_0)\cdot\mathbf{1} = \boldsymbol{\xi}(\theta_0)\cdot\boldsymbol{\eta}(\theta_0) = 1$. The equality $\mathbf{M}(\theta_0)\boldsymbol{\xi}(\theta_0)^T = \rho_0\boldsymbol{\xi}(\theta_0)^T$ implies that

$$\begin{aligned}
&\rho_0\xi_j(\theta_0) = \Psi_1(\theta_0)\xi_1(\theta_0) + \xi_{j+1}(\theta_0), \quad j = 1\ldots d-1,\\
&\rho_0\xi_d(\theta_0) = \Psi_d(\theta_0)\xi_1(\theta_0),
\end{aligned}$$

hence, for all $j = 1 \dots d$,

$$\xi_j(\theta_0) = \rho_0^{j-1} \sum_{i=j}^{d} \Psi_i(\theta_0)\rho_0^{-i}\xi_1.$$

Finally, $\boldsymbol{\xi}(\theta_0) \cdot \mathbf{1} = 1$ leads to

$$\xi_1(\theta_0) = \frac{1}{\sum_{k=1}^{d} \rho_0^{k-1} \sum_{i=k}^{d} \Psi_i(\theta_0)\rho_0^{-i}}.$$

Consequently, for each $j = 1 \dots d$,

$$\xi_j(\theta_0) = \frac{\rho_0^j \sum_{l=j}^{d} \rho_0^{-l}\Psi_l(\theta_0)}{\sum_{k=1}^{d} \rho_0^k \sum_{i=k}^{d} \rho_0^{-i}\Psi_i(\theta_0)}. \tag{4.2.78}$$

Similarly, we prove that $\boldsymbol{\eta}(\theta_0)$ is given by

$$\eta_k(\theta_0) = \rho_0^{-(k-1)} \frac{\sum_{k=1}^{d} \rho_0^{k-1} \sum_{i=k}^{d} \Psi_i(\theta_0)\rho_0^{-i}}{\sum_{k=1}^{d} \sum_{i=k}^{d} \Psi_i(\theta_0)\rho_0^{-i}}. \tag{4.2.79}$$

The basic limit theorem in the supercritical case states that there exists a random variable W_0 such that (see *e.g.* Theorem 5.6.1 in [AthNey72])

$$\lim_{n\to\infty} \rho_0^{-n}\mathbf{X}_n \overset{a.s.}{=} \boldsymbol{\eta}(\theta_0)W_0. \tag{4.2.80}$$

Let us recall the results obtained by S. Asmussen and N. Keiding in [AsmKei78], Theorem 6.1. First, as pointed out by N. Becker in [Beck77], the estimator $\widetilde{\rho}_n$ is strongly consistent on the set of non-extinction $\{W_0 > 0\}$.

Second, once adequately normalized, $\widetilde{\rho}_n - \rho_0$ is asymptotically normal. However, the asymptotic behavior of $\widetilde{\rho}_n - \rho_0$ depends qualitatively on the relative sizes of $\rho(\theta_0)$ and λ^2, where λ is the absolute value of a certain eigenvalue of $\mathbf{M}(\theta_0)$. More precisely, let $\{\lambda_i\}_{i=1\dots s}$ be the spectrum of $\mathbf{M}(\theta_0)$, and for each $i = 1 \dots s$, let r_i be the algebraic multiplicity of λ_i. We denote by $\mathcal{B} = \{\mathbf{u}_{i,j},\ i = 1 \dots s,\ j = 1 \dots r_i\}$ the base of the Jordan canonical decomposition of $\mathbf{M}(\theta_0)$, *i.e.* such that for all $i = 1 \dots s$,

$$\mathbf{M}(\theta_0)\mathbf{u}_{i,1} = \lambda_i\mathbf{u}_{i,1}, \quad \mathbf{M}(\theta_0)\mathbf{u}_{i,j} = \mathbf{u}_{i,j-1} + \lambda_i\mathbf{u}_{i,j},\ j = 2 \dots r_i.$$

Let us define the vector $\boldsymbol{\varsigma} := \mathbf{1} - \boldsymbol{\xi}(\theta_0)$ and denote $(\varsigma_{i,j})_{\substack{i=1\dots s,\\ j=1\dots r_i}}$ its coordinates in $\mathcal{B}$:

$$\boldsymbol{\varsigma} = \sum_{i=1}^{s}\sum_{j=1}^{r_i} \varsigma_{i,j}\mathbf{u}_{i,j}.$$

Then λ is defined as follows,

$$\lambda = \lambda(\boldsymbol{\varsigma}) := \max_{i=1\dots s} \{|\lambda_i| : \exists j = 1 \dots r_i \text{ such that } \varsigma_{i,j} \neq 0\}, \tag{4.2.81}$$

and

$$\gamma = \gamma(\boldsymbol{\varsigma}) := \max_{\substack{i=1\dots s:\\ |\lambda_i|=\lambda}} \{j = 1 \dots r_i : \varsigma_{i,j} \neq 0\}. \tag{4.2.82}$$

We similarly define $\lambda(\boldsymbol{x})$ and $\gamma(\boldsymbol{x})$ for any complex vector $\boldsymbol{x} \in \mathbb{C}^d$. As detailed in [AsmKei78],

$$\widetilde{\rho}_n - \rho_0 = \frac{S_n + T_n}{|\mathbf{X}_0| + \dots + |\mathbf{X}_{n-1}|},$$

where (to avoid heavy notation, when no confusion is possible, we do not write differently column and row vectors when multiplied by a matrix)

$$\begin{aligned} S_n &:= \sum_{k=0}^{n-1} \left(\mathbf{X}_{k+1} - \mathbf{X}_k \mathbf{M}(\theta_0)\right) \cdot \mathbf{1}, \\ T_n &:= \sum_{k=0}^{n-1} \mathbf{X}_k \cdot \boldsymbol{\kappa}, \\ \boldsymbol{\kappa} &:= \left(\mathbf{M}(\theta_0) - \rho_0 \mathbf{I}\right) \mathbf{1}. \end{aligned}$$

It appears that S_n and T_n are of the same order of magnitude when $\lambda^2 < \rho_0$, while T_n dominates S_n if $\lambda^2 \geqslant \rho_0$. In order to deal with the case $\lambda^2 < \rho_0$, we define for all $n \in \mathbb{N}$,

$$\boldsymbol{\nu}_n := \mathbf{1} + \sum_{k=0}^{n-1} \left[\mathbf{M}(\theta_0)\right]^k \boldsymbol{\kappa},$$

and

$$\begin{aligned} C_1 &:= (\rho_0 - 1) \sum_{n=1}^{\infty} \rho_0^{-n} \sum_{i=1}^{d} \eta_i(\theta_0) \boldsymbol{\nu}_n \mathbf{V}_1^i \boldsymbol{\nu}_n \\ &= (\rho_0 - 1) \sum_{n=1}^{\infty} \rho_0^{-n} \sum_{i=1}^{d} \eta_i(\theta_0) \Psi_i(\theta_0) \nu_{n,1}^2. \end{aligned} \tag{4.2.83}$$

If $\lambda^2 \geqslant \rho_0$, then there exist vectors $\boldsymbol{\varsigma}^1$ and $\boldsymbol{\varsigma}^2$ such that $\left(\mathbf{M}(\theta_0) - \rho_0 \mathbf{I}\right) \boldsymbol{\varsigma} = \left(\mathbf{M}(\theta_0) - \mathbf{I}\right) \boldsymbol{\varsigma}^1 + \boldsymbol{\varsigma}^2$, with $\lambda(\boldsymbol{\varsigma}^1) = \lambda$, $\gamma(\boldsymbol{\varsigma}^1) = \gamma$ and $\lambda(\boldsymbol{\varsigma}^2) \leqslant 1$. If $\lambda^2 = \rho_0$, we set moreover

$$C_2 := \left(1 - \frac{1}{\rho_0}\right) \lim_{n\to\infty} \frac{\sum_{i=1}^{d} \eta_i(\theta_0) \boldsymbol{\varsigma}^1 \mathbf{V}_n^i \boldsymbol{\varsigma}^1}{\rho_0^n n^{2\gamma-1}}. \tag{4.2.84}$$

We can now quote Theorem 6.1 of [AsmKei78]. Note that the **(XlogX)** assumption implies that (see *e.g.* [AthNey72], Theorem 5.6.1)

$$\mathbb{P}(W_0 > 0) > 0.$$

For notational convenience, it is assumed in this section just as in [AsmKei78] that $\mathbb{P}(W_0 = 0) = 0$. The results stated here are thus valid on the set of non-extinction.

Theorem 4.2.11 ([AsmKei78] Thm 6.1-6.3)**.** *Let us assume that the process $(\boldsymbol{X}_k)_{k\geqslant 0}$ is supercritical. Then, on the set of non-extinction, the estimator $\widetilde{\rho}_n$ is consistent:*

$$\lim_{n\to\infty} \widetilde{\rho}_n \overset{a.s.}{=} \rho_0, \tag{4.2.85}$$

and has the following asymptotic distribution.

If $\lambda^2 < \rho_0$,

$$\lim_{n\to\infty} \sqrt{W_0 \left(1 + \ldots + \rho_0^{n-1}\right)} \left(\widetilde{\rho}_n - \rho_0\right) \overset{\mathcal{D}}{=} \mathcal{N}\left(0, C_1\right). \tag{4.2.86}$$

If $\lambda^2 = \rho_0$ and $C_2 > 0$,

$$\lim_{n\to\infty} \sqrt{\frac{W_0 \left(1 + \ldots + \rho_0^{n-1}\right)}{n^{2\gamma-1}}} \left(\widetilde{\rho}_n - \rho_0\right) \overset{\mathcal{D}}{=} \mathcal{N}\left(0, C_2\right). \tag{4.2.87}$$

If $\lambda^2 > \rho_0$, there exist random variables H_n with $\overline{\lim} |H_n| < \infty$, such that

$$\lim_{n\to\infty} \left[\frac{W_0 \left(1 + \ldots + \rho_0^{n-1}\right)}{\lambda^{n-1}(n-1)^{\gamma-1}} \left(\widetilde{\rho}_n - \rho_0\right) - H_{n-1}\right] \overset{a.s.}{=} 0. \tag{4.2.88}$$

Remark 4.2.12. We point out that W_0, γ, λ, C_1, C_2 and H_n depend on θ_0.

Let us now deduce from this theorem the asymptotic properties of the estimator of θ_0 defined by (4.2.77). For this purpose, we define the constant

$$C_0 := \left(\frac{\sum_{k=1}^d k a_k \rho_0^{-k} + \sum_{k=1}^d a_k \rho_0^{-k} \sum_{k=1}^d k b_k \rho_0^{-k} - \sum_{k=1}^d b_k \rho_0^{-k} \sum_{k=1}^d k a_k \rho_0^{-k}}{\rho_0 \left(\sum_{k=1}^d a_k \rho_0^{-k} \right)^2} \right)^2. \tag{4.2.89}$$

We then obtain the following result.

Theorem 4.2.13. *Let us assume that the process $(\boldsymbol{X}_k)_{k\geqslant 0}$ is supercritical. Then, on the set of non-extinction, the estimator $\widetilde{\theta}_n^X$ is consistent:*

$$\lim_{n\to\infty} \widetilde{\theta}_n^X \overset{a.s.}{=} \theta_0, \tag{4.2.90}$$

and has the following asymptotic distribution.

If $\lambda^2 < \rho_0$,

$$\lim_{n\to\infty} \sqrt{\frac{W_0 \left(1 + \ldots + \rho_0^{n-1}\right)}{C_1 C_0}} \left(\widetilde{\theta}_n^X - \theta_0 \right) \overset{\mathcal{D}}{=} \mathcal{N}(0,1). \tag{4.2.91}$$

If $\lambda^2 = \rho_0$ and $C_2 > 0$,

$$\lim_{n\to\infty} \sqrt{\frac{W_0 \left(1 + \ldots + \rho_0^{n-1}\right)}{n^{2\gamma-1} C_2 C_0}} \left(\widetilde{\theta}_n^X - \theta_0 \right) \overset{\mathcal{D}}{=} \mathcal{N}(0,1). \tag{4.2.92}$$

If $\lambda^2 > \rho_0$, there exist random variables H_n with $\overline{\lim}\, |H_n| < \infty$, such that

$$\lim_{n\to\infty} \left[\frac{W_0 \left(1 + \ldots + \rho_0^{n-1}\right)}{\lambda^{n-1}(n-1)^{\gamma-1}} \left(\widetilde{\theta}_n^X - \theta_0 \right) - H_{n-1} \right] \overset{a.s.}{=} 0. \tag{4.2.93}$$

Proof. The strong consistency is immediate from (4.2.85).

We then express $\widetilde{\theta}_n^X - \theta_0$ as a function of $\widetilde{\rho}_n - \rho_0$, in order to deduce its asymptotic distribution from (4.2.86)-(4.2.88). We write

$$\begin{aligned} &\widetilde{\theta}_n^X - \theta_0 \\ &= \frac{\sum_k a_k \left(\rho_0^{-k} - \widetilde{\rho}_n^{-k}\right) + \sum_k a_k \rho_0^{-k} \sum_k b_k \left(\rho_0^{-k} - \widetilde{\rho}_n^{-k}\right) - \sum_k b_k \rho_0^{-k} \sum_k a_k \left(\rho_0^{-k} - \widetilde{\rho}_n^{-k}\right)}{\sum_k a_k \widetilde{\rho}_n^{-k} \sum_k a_k \rho_0^{-k}}, \end{aligned} \tag{4.2.94}$$

and use the fact that, for all $k = 1 \ldots d$,

$$\rho_0^{-k} - \widetilde{\rho}_n^{-k} = \left(\widetilde{\rho}_n - \rho_0\right) \frac{\sum_{l=1}^k \rho_0^{l-k} \widetilde{\rho}_n^{1-l}}{\widetilde{\rho}_n \rho_0},$$

in order to obtain

$$\begin{aligned} \widetilde{\theta}_n^X - \theta_0 = \left(\widetilde{\rho}_n - \rho_0\right) \Bigg[& \frac{\sum_{k=1}^d a_k \sum_{l=1}^k \rho_0^{l-k} \widetilde{\rho}_n^{1-l}}{\widetilde{\rho}_n \rho_0 \sum_{k=1}^d a_k \widetilde{\rho}_n^{-k} \sum_{k=1}^d a_k \rho_0^{-k}} \\ & + \frac{\sum_{k=1}^d a_k \rho_0^{-k} \sum_{k=1}^d b_k \sum_{l=1}^k \rho_0^{l-k} \widetilde{\rho}_n^{1-l} - \sum_{k=1}^d b_k \rho_0^{-k} \sum_{k=1}^d a_k \sum_{l=1}^k \rho_0^{l-k} \widetilde{\rho}_n^{1-l}}{\widetilde{\rho}_n \rho_0 \sum_{k=1}^d a_k \widetilde{\rho}_n^{-k} \sum_{k=1}^d a_k \rho_0^{-k}} \Bigg]. \end{aligned} \tag{4.2.95}$$

By (4.2.85), the square bracket in (4.2.95) almost surely converges to $\sqrt{C_0}$, and (4.2.91)-(4.2.93) are immediately deduced from (4.2.86)-(4.2.88). □

Unfortunately this theorem is seldom of direct practical applicability, in particular because of the differentiation between the three cases $\lambda^2 < \rho_0$, $\lambda^2 = \rho_0$ and $\lambda^2 > \rho_0$. We can not provide here an asymptotic confidence interval solely based on the observations, as we did for the estimators $\widehat{\theta}^X_{|\mathbf{X}_0|}$ and $\widehat{\theta}^Z_n$ (see Theorem 4.2.1 and Corollary 4.2.9).

Remark 4.2.14. $\widetilde{\rho}_n$ seems to have no interesting asymptotical properties for n fixed, as $|\mathbf{X}_0| \to \infty$, when $d > 1$ (if $d = 1$ then it reduces to the Harris estimator which is also the CLSE, hence Subsection 4.2.1 can be applied), unless we assume that, for all $i = 1 \dots d$,

$$\lim_{|\mathbf{X}_0| \to \infty} \frac{X_{0,i}}{|\mathbf{X}_0|} \overset{a.s.}{=} \frac{\eta_i(\theta_0)}{\boldsymbol{\eta}(\theta_0) \cdot \mathbf{1}}. \tag{4.2.96}$$

If this holds, then for any multitype branching process $(\mathbf{X}_k)_{k \geqslant 0}$ of any class of criticality,

$$\lim_{|\mathbf{X}_0| \to \infty} \widetilde{\rho}_n \overset{a.s.}{=} \rho_0$$

It is however obvious that assumption (4.2.96) is much too strong and nearly never applicable.

4.2.4 Comparison of the estimators and illustration of the asymptotic

All the following simulations and computations have been done with the numerical computing and programming environment Matlab.

Comparison of the estimators

In this section we compare the three estimators introduced in Subsections 4.2.1-4.2.3 on a set of simulated trajectories, for several values of $|\mathbf{X}_0|$ and n. As a context of simulation, we choose the BSE epidemic in Great-Britain which, as detailed in **Section 5.1**, can be modeled by a multitype branching process $(\mathbf{X}_k)_{k \geqslant 0}$ of the form (4.1.1), with $d = 9$. For each $k = 1 \dots 9$, Ψ_k is of the form (4.2.1), *i.e.* depends affinely on the unknown parameter θ_0, which corresponds in this context to the horizontal infection parameter (see (5.1.2) in Section **Section 5.1**). The numerical values a_k and b_k are given in Table5.3. Then, by (4.2.3), the process is subcritical (resp. supercritical) for $\theta_0 < \theta_{crit.}$ (resp. $\theta_0 > \theta_{crit.}$), with $\theta_{crit.} \simeq 23$.

We focus on the three following set of trajectories. Fixing the parameter $\theta_0 = 15$, we first simulate trajectories of the unconditioned subcritical process $(\mathbf{X}_k)_{k \geqslant 0}$, and then of the conditioned process $(\mathbf{Z}_k)_{k \geqslant 0}$. We finally simulate, with the parameter $\theta_0 = 35$, trajectories of the unconditioned supercritical process $(\mathbf{X}_k)_{k \geqslant 0}$. We consider different values of $|\mathbf{X}_0|$ and n, namely $|\mathbf{X}_0| = 10, 100, 1000$ and $n = 10, 50, 100$. For every couple $(|\mathbf{X}_0|, n)$, we simulate, in each of the previously mentioned cases, 100 trajectories of length n (*i.e* not extinct at time n), initiated by $\mathbf{X}_0 = (0, \dots, 0, |\mathbf{X}_0|)$, and compute the corresponding empirical means and standard deviations of the estimators. These are reported in Tables 4.1-4.3, which allow to compare the three different estimators in each of these situations.

The empty entries in Table 4.1 are due to the fact that for some given couples $(|\mathbf{X}_0|, n)$, trajectories of the subcritical process initiated by $\mathbf{X}_0$ with an extinction time greater than n occur only with a very small probability. We recall that the estimator $\widehat{\theta}^Z_n$ has no explicit form. Its precision thus depends on the optimization method which is chosen, while the precision for $\widehat{\theta}^X_{|\mathbf{X}_0|}$ and $\widetilde{\theta}^X_n$ solely depends on the computing program. As a consequence, the estimations obtained with $\widehat{\theta}^X_{|\mathbf{X}_0|}$ and $\widehat{\theta}^Z_n$ might slightly differ from each other, even when they are in theory equal, *i.e.* on trajectories with no sequence of $d - 1 = 8$ zeros. We can see however in Table 4.1 and Table 4.3 that, in our example, this approximation error remains very small.

Table 4.1 enables to compare $\widehat{\theta}^X_{|\mathbf{X}_0|}$ and $\widetilde{\theta}^X_n$. As just mentioned, $\widehat{\theta}^Z_n$ is, in this case, equal to $\widehat{\theta}^X_{|\mathbf{X}_0|}$ since no trajectory contains 8 consecutive zeros. Obviously, $\widehat{\theta}^X_{|\mathbf{X}_0|}$ provides an estimation of the parameter much closer to θ_0 than $\widetilde{\theta}^X_n$, which is of no surprise, since $\widetilde{\theta}^X_n$ is not proved to be consistent in the subcritical case. This table provides moreover an illustration of the consistency

n \ $\lvert\mathbf{X}_0\rvert$		10 mean	10 std. dev.	100 mean	100 std. dev.	1000 mean	1000 std. dev.
10	$\widehat{\theta}^X_{\lvert\mathbf{X}_0\rvert}$	14.7179	4.8811	14.7806	1.5794	14.9930	0.4526
	$\widehat{\theta}^Z_n$	14.7181	4.8805	14.7806	1.5794	14.9930	0.4526
	$\widetilde{\theta}^X_n$	22.2960	3.5440	22.1341	1.1615	22.3036	0.3438
50	$\widehat{\theta}^X_{\lvert\mathbf{X}_0\rvert}$	/	/	15.1834	0.9552	14.9675	0.3370
	$\widehat{\theta}^Z_n$	/	/	15.1834	0.9551	14.9675	0.3371
	$\widetilde{\theta}^X_n$	/	/	19.0956	0.4860	18.9621	0.1803
100	$\widehat{\theta}^X_{\lvert\mathbf{X}_0\rvert}$	/	/	/	/	/	/
	$\widehat{\theta}^Z_n$	/	/	/	/	/	/
	$\widetilde{\theta}^X_n$	/	/	/	/	/	/

Table 4.1: Empirical means and standard deviations of $\widehat{\theta}^X_{|\mathbf{X}_0|}$, $\widehat{\theta}^Z_n$ and $\widetilde{\theta}^X_n$ corresponding to 100 trajectories of length n of the unconditioned subcritical process $(\mathbf{X}_k)_{k\geqslant 0}$ initiated by $\mathbf{X}_0 = (0,\dots,0,|\mathbf{X}_0|)$ and simulated with the infection parameter $\theta_0 = 15$, for different couples $(|\mathbf{X}_0|, n)$.

n \ $\lvert\mathbf{X}_0\rvert$		10 mean	10 std. dev.	100 mean	100 std. dev.	1000 mean	1000 std. dev.
10	$\widehat{\theta}^X_{\lvert\mathbf{X}_0\rvert}$	14.4306	5.1569	14.8198	1.5400	14.9138	0.5442
	$\widehat{\theta}^Z_n$	14.4306	5.1568	14.8198	1.5400	14.9138	0.5442
	$\widetilde{\theta}^X_n$	22.0041	3.6403	22.1723	1.1272	22.2378	0.4094
50	$\widehat{\theta}^X_{\lvert\mathbf{X}_0\rvert}$	16.0774	2.2719	15.0800	1.0376	15.0420	0.3276
	$\widehat{\theta}^Z_n$	14.6195	3.3079	15.0595	1.0550	15.0428	0.3248
	$\widetilde{\theta}^X_n$	19.7192	1.1291	19.0371	0.5284	18.9985	0.1714
100	$\widehat{\theta}^X_{\lvert\mathbf{X}_0\rvert}$	17.7708	1.3873	15.1534	0.9573	15.0346	0.4027
	$\widehat{\theta}^Z_n$	14.7098	2.6979	14.8563	1.0287	15.0208	0.4047
	$\widetilde{\theta}^X_n$	20.4621	0.7545	19.0074	0.4620	18.9211	0.1943

Table 4.2: Empirical means and standard deviations of $\widehat{\theta}^X_{|\mathbf{X}_0|}$, $\widehat{\theta}^Z_n$ and $\widetilde{\theta}^X_n$ corresponding to 100 trajectories of length n of the conditioned process $(\mathbf{Z}_k)_{k\geqslant 0}$ in the subcritical case, initiated by $\mathbf{X}_0 = (0,\dots,0,|\mathbf{X}_0|)$ and simulated with the infection parameter $\theta_0 = 15$, for different couples $(|\mathbf{X}_0|, n)$.

n \ $\lvert\mathbf{X}_0\rvert$		10 mean	10 std. dev.	100 mean	100 std. dev.	1000 mean	1000 std. dev.
10	$\widehat{\theta}^X_{\lvert\mathbf{X}_0\rvert}$	35.3485	6.2014	35.2777	1.6247	34.9629	0.6258
	$\widehat{\theta}^Z_n$	35.3611	6.1672	35.2777	1.6295	34.9630	0.6271
	$\widetilde{\theta}^X_n$	38.4696	5.3626	38.5015	1.4765	38.2363	0.5670
50	$\widehat{\theta}^X_{\lvert\mathbf{X}_0\rvert}$	34.7898	1.2210	34.9792	0.2760	35.0008	0.0860
	$\widehat{\theta}^Z_n$	34.7898	1.2205	34.9792	0.2764	35.0008	0.0953
	$\widetilde{\theta}^X_n$	34.8578	1.2613	35.0580	0.2816	35.0816	0.0877
100	$\widehat{\theta}^X_{\lvert\mathbf{X}_0\rvert}$	34.9942	0.1014	35.0056	0.0302	35.0021	0.0107
	$\widehat{\theta}^Z_n$	34.9943	0.1042	35.0056	0.0300	35.0000	0.0000
	$\widetilde{\theta}^X_n$	34.9930	0.1025	35.0053	0.0313	35.0032	0.0116

Table 4.3: Empirical means and standard deviations of $\widehat{\theta}^X_{|\mathbf{X}_0|}$, $\widehat{\theta}^Z_n$ and $\widetilde{\theta}^X_n$ corresponding to 100 trajectories of length n of the unconditioned supercritical process $(\mathbf{X}_k)_{k\geqslant 0}$ initiated by $\mathbf{X}_0 = (0,\dots,0,|\mathbf{X}_0|)$ and simulated with the infection parameter $\theta_0 = 35$, for different couples $(|\mathbf{X}_0|, n)$.

of $\widehat{\theta}^X_{|\mathbf{X}_0|}$ and a probable non consistency of $\widetilde{\theta}^X_n$, as $|\mathbf{X}_0|$ tends to infinity, which appears clearly for $n = 10$.

Table 4.2 illustrates again the fact that $\widetilde{\theta}^X_n$ is not accurate when the process is not supercritical. This table is however very interesting to compare $\widehat{\theta}^X_{|\mathbf{X}_0|}$ and $\widehat{\theta}^Z_n$ on trajectories of the process conditioned on non-extinction at each step, which might for n large enough present one or several sequences of 8 zeros (see Figure 4.2). It appears that for long trajectories (*e.g.* $n = 50$ or $n = 100$), we obtain a better empirical mean with the estimator $\widehat{\theta}^Z_n$, but a larger standard deviation. This is particularly obvious when the initial size of the clinical population is small ($|\mathbf{X}_0| = 10$). In order to better illustrate this phenomenon, we represent in Figure 4.3 the estimations obtained with the estimators $\widehat{\theta}^X_{|\mathbf{X}_0|}$ and $\widehat{\theta}^Z_n$ for the 100 simulated trajectories, used in Table 4.2, of the conditioned process with the infection parameter $\theta_0 = 15$, with $|\mathbf{X}_0| = 10$, respectively for $n = 50$ and for $n = 100$. It appears that, as n increases, $\widehat{\theta}^X_{|\mathbf{X}_0|}$ tends to overestimate θ_0, while $\widehat{\theta}^Z_n$ remains close to θ_0 but with a larger standard deviation. Moreover, Table 4.2 illustrates the consistency of $\widehat{\theta}^Z_n$, as n tends to infinity, as well as its consistency, as $|\mathbf{X}_0|$ tends to infinity (see Remark 4.2.10).

Finally, Table 4.3 allows to compare $\widehat{\theta}^X_{|\mathbf{X}_0|}$ with $\widetilde{\theta}^X_n$ in the supercritical case (again, $\widehat{\theta}^Z_n$ is here in theory equal to $\widehat{\theta}^X_{|\mathbf{X}_0|}$). On those examples, $\widehat{\theta}^X_{|\mathbf{X}_0|}$ provides a better estimation when the number of observations is small ($n = 10$). However, when n increases, the consistency of $\widetilde{\theta}^X_n$ comes in play, and it appears that $\widetilde{\theta}^X_n$ seems as good as the CLSE $\widehat{\theta}^X_{|\mathbf{X}_0|}$, although $\widetilde{\theta}^X_n$ is originally not built as an estimator with usual characteristics (CLSE, MLE, moments estimator...), but rather as an explicit estimator based upon realistic data.

Asymptotic normal distribution

The aim of this section is to illustrate the asymptotic normal distribution of each of the three estimators, namely (4.2.18), (4.2.69) and (4.2.91) (occurring in the case $\lambda^2 < \rho_0$).

To this end, we first simulate 1000 trajectories of length $n = 10$ of the supercritical unconditioned process $(\mathbf{X}_k)_{k\geqslant 0}$, initiated by $\mathbf{X}_0 = (0,\dots,0,1000)$, with the infection parameter $\theta_0 = 35$. We represent in Figure 4.4.1 the empirical distribution of

$$\sqrt{\frac{\sum_{k=1}^n \boldsymbol{a}\cdot\mathbf{X}_{k-1}}{\sigma^2(\widehat{\theta}^X_{|\mathbf{X}_0|})}}\left(\widehat{\theta}^X_{|\mathbf{X}_0|} - \theta_0\right) \tag{4.2.97}$$

corresponding to the 1000 trajectories. Drawing the empirical Gaussian distribution corresponding

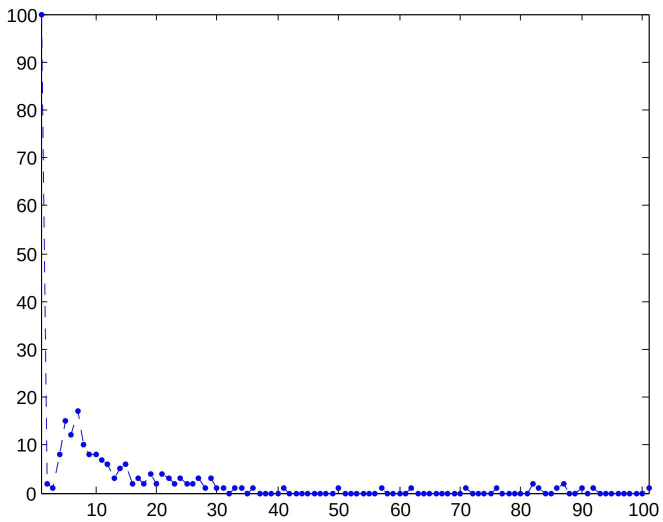

Figure 4.2: Simulation of a trajectory of length $n = 100$ of the conditioned process $(\mathbf{Z}_k)_{k\geqslant 0}$ in the subcritical case, initiated by $(0,\ldots,0,100)$, with the infection parameter $\theta_0 = 15$.

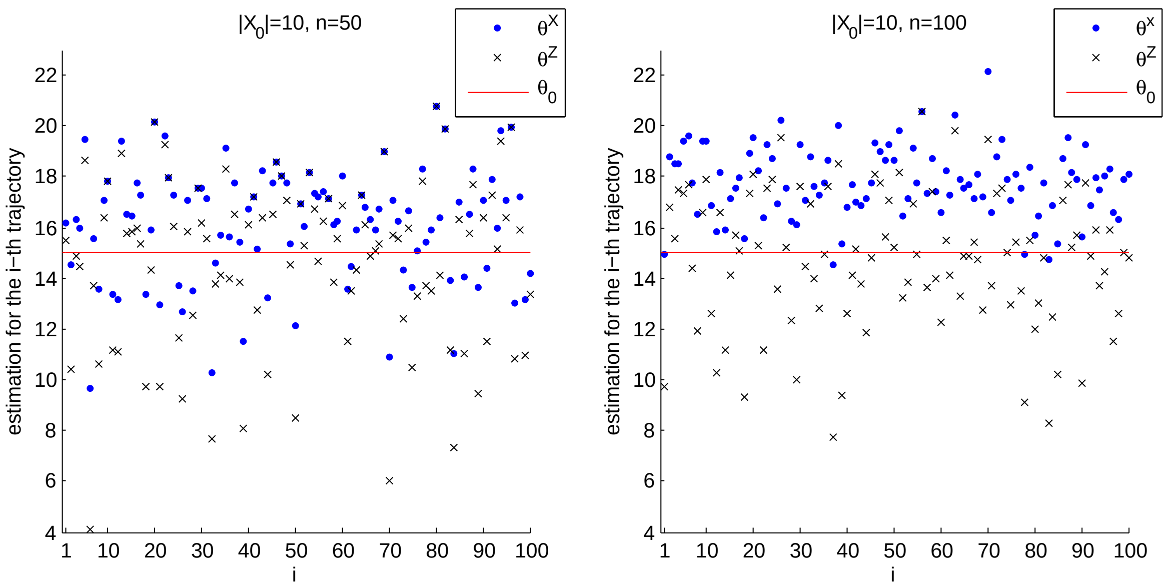

Figure 4.3: Estimations with $\widehat{\theta}^X_{|\mathbf{X}_0|}$ and $\widehat{\theta}^Z_n$ for 100 simulated trajectories (used in Table 4.2) of length $n = 50$ (resp. $n = 100$) of the conditioned process $(\mathbf{Z}_k)_{k\geqslant 0}$ in the subcritical case with the infection parameter $\theta_0 = 15$ and initiated by $\mathbf{X}_0 = (0,\ldots,0,10)$.

to 1000 realizations, we see that the two histograms are indeed very similar.

Second, we simulate 1000 trajectories of length $n = 100$ of the process $(\mathbf{Z}_k)_{k\geqslant 0}$ in the subcritical case, initiated by $\mathbf{X}_0 = (0, \ldots, 0, 10)$, with the infection parameter $\theta_0 = 15$. We then represent in Figure 4.4.2 the empirical distribution of

$$\frac{\sum_{k=0}^{n}\left(f'\left(\widehat{\theta}_n^Z, \mathbf{Z}_k\right)\right)^2}{\sqrt{\sum_{k=0}^{n}\left(f'\left(\widehat{\theta}_n^Z, \mathbf{Z}_k\right)\right)^2 f\left(\widehat{\theta}_n^Z, \mathbf{Z}_k\right)(\boldsymbol{a}\cdot\mathbf{Z}_k)^{-1/2}}}\left(\widehat{\theta}_n^Z - \theta_0\right), \tag{4.2.98}$$

corresponding to the 1000 trajectories. It appears to be very similar to the empirical Gaussian distribution, however with a larger negative support, which could mean that $\widehat{\theta}_n^Z$ has a tendency to underestimate the real parameter. This is confirmed by Table 4.2, where for $(|\mathbf{X}_0|, n) = (10, 100)$ the empirical mean equals 14.7098 and the empirical standard deviation is rather large, equal to 2.6979.

We finally choose to illustrate the asymptotic normal distribution (4.2.91) occurring in the case $\lambda^2 < \rho_0$. For the sake of simplicity and to be ensured that no numerical approximation errors interfere in the Jordan decomposition of the mean matrix, we work in a simpler context as the one considered until now: we choose $d = 2$, and $\boldsymbol{a} = (2, 1)$, $\boldsymbol{b} = (1/2, 0)$. The process is thus supercritical if and only if $\theta_0 > 1/6$. Taking $\theta_0 = 1/4$, we have

$$\mathbf{M} = \begin{pmatrix} 1 & 1 \\ \frac{1}{4} & 0 \end{pmatrix},$$

with eigenvalues $\lambda_1 = \rho_0 = \frac{1+\sqrt{2}}{2}$ and $\lambda_2 = \frac{1-\sqrt{2}}{2}$. Then necessarily $\lambda = |\lambda_2|$, and we check that $\lambda^2 < \rho_0$. We compute

$$C_0 = \frac{1}{4}\left(\frac{4\rho_0^2 + 4\rho_0 - 1}{4\rho_0^2 + 4\rho_0 + 1}\right)^2 = \left(\sqrt{2} - 1\right)^2.$$

We simulate 1000 trajectories of length $n = 100$ of the supercritical process $(\mathbf{X}_k)_{k\geqslant 0}$ initiated by $\mathbf{X}_0 = (0, 1000)$, with the infection parameter $\theta_0 = 1/4$. As an approximation of the random variable W_0, we take $\rho_0^{-n}\mathbf{X}_n \cdot \boldsymbol{\xi}(\theta_0)$ (see (4.2.80)), and using the fact that (see (4.2.83))

$$\lim_{n\to\infty} \frac{\rho_0^n \beta_n}{1 + \ldots + \rho_0^{n-1}} = C_1,$$

where

$$\beta_n := \sum_{k=1}^{n} \rho_0^{-k} \sum_{i=1}^{2} \eta_i(\theta_0)\Psi_i(\theta_0)\nu_{k,1}^2,$$

we approximate C_1 by $\rho_0^n\beta_n\left(1 + \ldots + \rho_0^{n-1}\right)^{-1}$. We then represent in Figure 4.4.3 the empirical distribution of

$$\sqrt{\frac{\mathbf{X}_n \cdot \boldsymbol{\xi}(\theta_0)\left(1 + \ldots + \rho_0^{n-1}\right)^2}{\rho_0^{2n}\beta_n C_0}}\left(\widetilde{\theta}_n^X - \theta_0\right) \tag{4.2.99}$$

corresponding to the 1000 trajectories. It appears to be not so close to the Gaussian distribution, which can be due to the several approximations just mentioned.

Conclusion

According to the simulations presented in Subsection 4.2.4, the conditional least squares estimators $\widehat{\theta}_{|\mathbf{X}_0|}^X$ and $\widehat{\theta}_n^Z$ appear to be accurate and equivalent estimators of θ_0 at finite distance $(|\mathbf{X}_0|, n)$ in the model introduced in Section 4.1, with moreover good asymptotic properties for any class of criticality. In addition, the estimator $\widehat{\theta}_n^Z$, which takes into account more information, provides for long trajectories with sequences of zeros, estimations which are, according to the simulations, better in mean but which have a larger standard deviation. The estimator $\widetilde{\theta}_n^X$ derived from

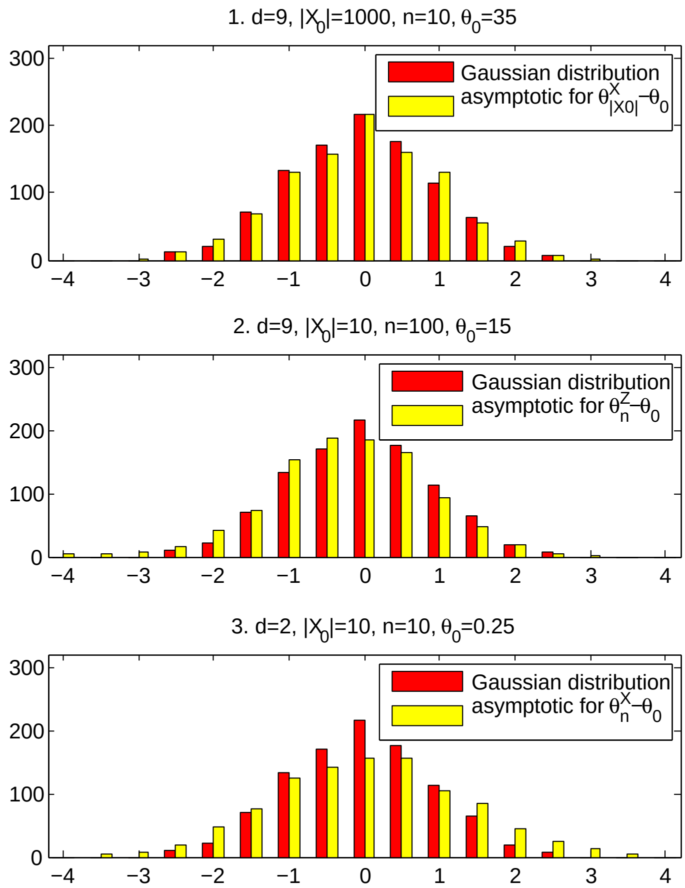

Figure 4.4: 1) Empirical distribution of (4.2.97) for 1000 trajectories of length $n = 10$ of the supercritical process $(\mathbf{X}_k)_{k\geqslant 0}$ initiated by $\mathbf{X}_0 = (0, \ldots, 0, 1000)$, with $\theta_0 = 35$. 2) Empirical distribution of (4.2.98) for 1000 trajectories of length $n = 100$ of the conditioned process $(\mathbf{Z}_k)_{k\geqslant 0}$ in the subcritical case initiated by $\mathbf{X}_0 = (0, \ldots, 0, 10)$, with $\theta_0 = 15$. 3) Empirical distribution of (4.2.99) for 1000 trajectories of length $n = 100$ of the supercritical process $(\mathbf{X}_k)_{k\geqslant 0}$ initiated by $\mathbf{X}_0 = (0, 10)$, with $\theta_0 = 0.25$. Comparison with the empirical Gaussian distribution.

the explicit estimator $\widetilde{\rho}_n$ of the Perron's root introduced in [AsmKei78], only provides satisfying estimations in the supercritical case, which are, in this case, not as good or are equivalent to the ones obtained with $\widehat{\theta}^X_{|\mathbf{X}_0|}$. Due to the differentiation of three cases and to the presence of unexplicitly known random variables in its asymptotic distribution, it is not possible to build an asymptotic confidence interval of θ_0 based on $\widetilde{\theta}^X_n$. The use of $\widetilde{\theta}^X_n$ is thus in our specific case less appropriate than the use of the CLSE. However, we point out that $\widehat{\theta}^X_{|\mathbf{X}_0|}$ and $\widehat{\theta}^Z_n$ are of an extremely more limited use that $\widetilde{\rho}_n$, since they do not provide an estimation of the Perron's root and only concern very specific processes, while $\widetilde{\rho}_n$ is suitable for any multitype BGW process.

4.3 Study of the very late extinction case

The aim of this section is to provide tools for the prediction of the evolution of the population modeled by (4.1.1), in case of a very late extinction of this population. In particular, in the epidemiological context, if this population represents the number of cases, this would correspond to the prediction of the evolution of the epidemic in the *worst-case scenario*.

To this end, we introduce the Q-process associated with the BGW process (4.1.4) , *i.e.* the process conditioned on "not being extinct in the distant future". We then build an estimator $\widehat{\theta}_n^{X^*}$ of θ_0 with asymptotic properties, as n tends to infinity, in the setting of the conditioned process. This enables us to predict the evolution of the population size in the very late extinction case, thanks to simulations of trajectories of the Q-process, where θ_0 is estimated by $\widehat{\theta}_n^{X^*}$. This will be illustrated in **Subsection 5.3.2**, where we handle with the problem of the propagation of the BSE epidemic in Great-Britain, in case of a very late extinction of the disease.

First, we compute, in **Subsection 4.3.1**, the transition law of the Q-process. Then we provide in **Subsection 4.3.2**, a CLSE of θ_0 in the setting of the conditioned process, and show its strong consistency and asymptotic normality, as time tends to infinity.

4.3.1 Q-process associated with the model

In order to study the evolution of the population size in its decay phase, assuming that the extinction of the population will occur very late, we consider the conditioned distribution

$$\mathbb{P}_{\theta_0}\left(\mathbf{X}_n = \,.\,|\ \mathbf{X}_{n+k} \neq \mathbf{0}\right), \text{ for any } k \text{ very large.} \tag{4.3.1}$$

We approximate this law with by the following limit

$$\lim_{k\to\infty} \mathbb{P}_{\theta_0}\left(\mathbf{X}_n = \,.\,|\ \mathbf{X}_{n+k} \neq \mathbf{0}\right). \tag{4.3.2}$$

As already mentioned (see (2.1.18)), it has been proved in [DalJof08] that this limit defines a Markov process (the Q-process associated with $(\mathbf{X}_n)_{n\geqslant 0}$), denoted $(\mathbf{X}_n^*)_{n\geqslant 0}$ in what follows. This Markov process has the following transition probability: for every $n \geqslant 1$, $\mathbf{i},\mathbf{j} \in \mathbb{N}^d$, $\mathbf{i} \neq \mathbf{0}$,

$$\mathbb{P}_{\theta_0}\left(\mathbf{X}_n^* = \mathbf{j} | \mathbf{X}_{n-1}^* = \mathbf{i}\right) = \frac{1}{\rho_0} \frac{\mathbf{j}\cdot\boldsymbol{\xi}(\theta_0)}{\mathbf{i}\cdot\boldsymbol{\xi}(\theta_0)} \mathbb{P}_{\theta_0}\left(\mathbf{X}_n = \mathbf{j} | \mathbf{X}_{n-1} = \mathbf{i}\right), \tag{4.3.3}$$

where $\boldsymbol{\xi}(\theta_0) := (\xi_1(\theta_0), \ldots, \xi_d(\theta_0))$ is the normalized right eigenvector of the mean matrix $\mathbf{M}(\theta_0)$ defined by (4.1.6), associated to its Perron's root ρ_0. Its explicit expression is given by (4.2.78).

Note that, by (4.3.3) and by definition of $(\mathbf{X}_n)_{n\geqslant 0}$, $\mathbb{P}_{\theta_0}\left(\mathbf{X}_n^* = \mathbf{j} | \mathbf{X}_{n-1}^* = \mathbf{i}\right) = 0$ as soon as $(j_2, \ldots, j_d) \neq (i_1, \ldots, i_{d-1})$. Hence the process $(\mathbf{X}_n^*)_{n\geqslant 0}$ satisfies for all $i = 2 \ldots d$ and $n \geqslant 0$, denoting $\mathbf{X}_n^* = (X_{n,1}^*, \ldots, X_{n,d}^*)$,

$$X_{n,i}^* \overset{a.s.}{=} X_{n-1,i-1}^*.$$

We define the one-dimensional d-Markovian process $(X_n^*)_{n\geqslant 0}$ corresponding to the first coordinate of $\mathbf{X}_n^*$, for all $n \geqslant 0$,

$$X_n^* := X_{n,1}^*. \tag{4.3.4}$$

According to [DalJof08] and under the **(XlogX)** assumption (see (4.1.7)), the conditioned process $\mathbf{X}_n^*$ is positive recurrent with stationary probability measure π_{θ_0} given by,

$$\pi_{\theta_0}(\mathbf{i}) := \frac{\mathbf{i}\cdot\boldsymbol{\xi}\nu_{\theta_0}(\mathbf{i})}{\sum_{\mathbf{k}\in\mathbb{N}^d} \mathbf{k}\cdot\boldsymbol{\xi}\nu_{\theta_0}(\mathbf{k})}, \ \mathbf{i} \in \mathbb{N}^d, \tag{4.3.5}$$

where ν_{θ_0} is the Yaglom distribution of the process $(\mathbf{X}_n)_{n\geqslant 0}$, defined in (4.2.40).

Let us compute explicitly the law of the Q-process associated with the process (4.1.1).

Proposition 4.3.1. *Conditionally on $\boldsymbol{X}^*_{n-1}$, X^*_n is distributed as the sum of two independent Poisson and Bernoulli random variables:*

$$X^*_n|\boldsymbol{X}^*_{n-1} \overset{\mathcal{D}}{\sim} \mathcal{P}oisson\left(\boldsymbol{X}^*_{n-1}\cdot\boldsymbol{\Psi}(\theta_0)\right) * \mathcal{B}\left(p\left(\theta_0, \boldsymbol{X}^*_{n-1}\right)\right), \tag{4.3.6}$$

where

$$p\left(\theta_0, \boldsymbol{X}^*_{n-1}\right) := \frac{\xi_1(\theta_0)\boldsymbol{X}^*_{n-1}\cdot\boldsymbol{\Psi}(\theta_0)}{\xi_1(\theta_0)\boldsymbol{X}^*_{n-1}\cdot\boldsymbol{\Psi}(\theta_0) + \sum_{k=2}^d X^*_{n-k+1}\xi_k(\theta_0)}. \tag{4.3.7}$$

Proof. For the sake of simplicity, let us write $\boldsymbol{\xi}$, and $\boldsymbol{\Psi}$, instead of $\boldsymbol{\xi}(\theta_0)$, and $\boldsymbol{\Psi}(\theta_0)$. Applying (4.1.2) and (4.3.3), we obtain that

$$\begin{aligned}
&\mathbb{P}_{\theta_0}\left(X^*_n = j|\mathbf{X}^*_{n-1}\right)\\
&= \mathbb{P}_{\theta_0}\left(\mathbf{X}^*_n = (j, X^*_{n-1}, \ldots, X^*_{n-(d-1)})|\mathbf{X}^*_{n-1}\right)\\
&= \frac{1}{\rho_0}\frac{j\xi_1 + \sum_{k=2}^d X^*_{n-k+1}\xi_k}{\mathbf{X}^*_{n-1}\cdot\boldsymbol{\xi}}\mathbb{P}_{\theta_0}\left(\mathbf{X}_n = (j, X^*_{n-1}, \ldots, X^*_{n-(d-1)})|\mathbf{X}_{n-1} = \mathbf{X}^*_{n-1}\right)\\
&= \frac{1}{\rho_0}\frac{j\xi_1 + \sum_{k=2}^d X^*_{n-k+1}\xi_k}{\mathbf{X}^*_{n-1}\cdot\boldsymbol{\xi}}\frac{\left(\mathbf{X}^*_{n-1}\cdot\boldsymbol{\Psi}\right)^j}{j!}e^{-\mathbf{X}^*_{n-1}\cdot\boldsymbol{\Psi}}\\
&= \frac{\xi_1\mathbf{X}^*_{n-1}\cdot\boldsymbol{\Psi}}{\rho_0\mathbf{X}^*_{n-1}\cdot\boldsymbol{\xi}}\frac{\left(\mathbf{X}^*_{n-1}\cdot\boldsymbol{\Psi}\right)^{j-1}}{(j-1)!}e^{-\mathbf{X}^*_{n-1}\cdot\boldsymbol{\Psi}} + \frac{\sum_{k=2}^d X^*_{n-k+1}\xi_k}{\rho_0\mathbf{X}^*_{n-1}\cdot\boldsymbol{\xi}}\frac{\left(\mathbf{X}^*_{n-1}\cdot\boldsymbol{\Psi}\right)^j}{j!}e^{-\mathbf{X}^*_{n-1}\cdot\boldsymbol{\Psi}}.
\end{aligned}$$

The equality $\mathbf{M}\boldsymbol{\xi}^T = \rho_0\boldsymbol{\xi}^T$ implies that, for all $k = 1\ldots d-1$, $\rho_0\xi_k = \Psi_k\xi_1 + \xi_{k+1}$, and that $\rho_0\xi_d = \Psi_d\xi_1$. Consequently,

$$\rho_0\mathbf{X}^*_{n-1}\cdot\boldsymbol{\xi} = \xi_1\mathbf{X}^*_{n-1}\cdot\boldsymbol{\Psi} + \sum_{k=2}^d X^*_{n-k+1}\xi_k,$$

and

$$\begin{aligned}
\mathbb{P}_{\theta_0}\left(X^*_n = j|\mathbf{X}^*_{n-1}\right) &= \frac{\xi_1\mathbf{X}^*_{n-1}\cdot\boldsymbol{\Psi}}{\xi_1\mathbf{X}^*_{n-1}\cdot\boldsymbol{\Psi} + \sum_{k=2}^d X^*_{n-k+1}\xi_k}\frac{\left(\mathbf{X}^*_{n-1}\cdot\boldsymbol{\Psi}\right)^{j-1}}{(j-1)!}e^{-\mathbf{X}^*_{n-1}\cdot\boldsymbol{\Psi}}\\
&\quad + \left(1 - \frac{\xi_1\mathbf{X}^*_{n-1}\cdot\boldsymbol{\Psi}}{\xi_1\mathbf{X}^*_{n-1}\cdot\boldsymbol{\Psi} + \sum_{k=2}^d X^*_{n-k+1}\xi_k}\right)\frac{\left(\mathbf{X}^*_{n-1}\cdot\boldsymbol{\Psi}\right)^j}{j!}e^{-\mathbf{X}^*_{n-1}\cdot\boldsymbol{\Psi}}\\
&= \left[\mathcal{P}oisson\left(\mathbf{X}^*_{n-1}\cdot\boldsymbol{\Psi}\right) * \mathcal{B}\left(\frac{\xi_1\mathbf{X}^*_{n-1}\cdot\boldsymbol{\Psi}}{\xi_1\mathbf{X}^*_{n-1}\cdot\boldsymbol{\Psi} + \sum_{k=2}^d X^*_{n-k+1}\xi_k}\right)\right](j).
\end{aligned}$$

□

Remark 4.3.2. The conditioned process $(\mathbf{X}^*_n)_{n\geqslant 0}$ thus behaves at each time-step like the unconditioned process $(\mathbf{X}_n)_{n\geqslant 0}$, according to a Poisson distribution, except that it has, at each time, the possibility to add one unit or not, according to a Bernoulli random variable. Moreover, for every $n \geqslant 1$, if $X^*_{n-1} = \ldots = X^*_{n-(d-1)} = 0$, then according to (4.3.7), $p(\theta_0, \mathbf{X}^*_{n-1}) = 1$, which implies that at time n, the probability to add one unit is equal to one. Hence for every $n \geq 1$, $\mathbb{P}_{\theta_0}(\mathbf{X}^*_n = \mathbf{0}) = 0$, and we obtain again the result that the Q-process $(\mathbf{X}^*_n)_{n\geqslant 0}$ can never become extinct.

4.3.2 CLSE for the Q-process

In order to make predictions of the evolution of the population size in case of a very late extinction, *i.e.* in order to make predictions of the behavior of the Q-process $(\mathbf{X}^*_n)_{n\geqslant 0}$ introduced in Subsection 4.3.1, we need to estimate the parameter θ_0 for this conditioned process. We point out that θ_0

does not play the same role in the conditioned process as in the unconditioned process, since, as shown in Proposition 4.3.1, this parameter interferes not only in the Poisson random variable but also in the Bernoulli one. It would thus be irrelevant to estimate θ_0 with an estimator built for the unconditioned process, for instance one of the estimators introduced in Section 4.2.

Let us notice that, according to (4.3.6), the process $(\mathbf{X}_n^*)_{n\geqslant 0}$ could be written as a multitype branching process with state-dependent immigration. Because of this state-dependence, and since the parameter θ_0 acts in a nonlinear way in the immigration, the methods developed in estimation theory for branching processes with immigration can not be directly applied here (see for example [QuiDur77]). Consequently, we build a CLSE of θ_0 in the setting of the conditioned process. Similarly as in Subsection 4.2.1 relative to the unconditioned process, we consider the weighted CLSE based on the process $Y_k^* := X_k^*/\sqrt{\boldsymbol{a}\cdot\mathbf{X}_{k-1}^*}$,

$$\widehat{\theta}_n^{X^*} := \arg\min_{\theta\in\Theta} S_n^*(\theta), \quad S_n^*(\theta) := \sum_{k=1}^{n}\left(Y_k^* - f^*(\theta,\mathbf{X}_{k-1}^*)\right)^2, \tag{4.3.8}$$

where Θ is as in Subsection 4.2.1, and

$$f^*(\theta_0,\mathbf{X}_{k-1}^*) := \mathbb{E}_{\theta_0}\left(Y_k^*\middle|\mathbf{X}_{k-1}^*\right) = \frac{\mathbf{X}_{k-1}^*\cdot\mathbf{\Psi}(\theta_0) + p\left(\theta_0,\mathbf{X}_{k-1}^*\right)}{\sqrt{\boldsymbol{a}\cdot\mathbf{X}_{k-1}^*}}. \tag{4.3.9}$$

The normalization of the process X_k^* by $\sqrt{\boldsymbol{a}\cdot\mathbf{X}_{k-1}^*}$ appears to be the most natural and suitable. On the one hand, it corresponds to the same normalization as for the unconditioned process (see (4.2.12)), which, as detailed in Subsection 4.2.1, generalizes the normalization leading in the monotype case to the Harris estimator ([Har48]). On the other hand, defining

$$\varepsilon_k^* := Y_k^* - f^*(\theta_0,\mathbf{X}_{k-1}^*),$$

the error term between the normalized process and its conditional expectation, we obtain that

$$g\left(\theta_0,\mathbf{X}_{k-1}^*\right) := \mathbb{E}_{\theta_0}\left((\varepsilon_k^*)^2\,|\mathbf{X}_{k-1}^*\right) = \frac{\mathbf{X}_{k-1}^*\cdot\mathbf{\Psi}(\theta_0) + p\left(\theta_0,\mathbf{X}_{k-1}^*\right)\left(1-p\left(\theta_0,\mathbf{X}_{k-1}^*\right)\right)}{\boldsymbol{a}\cdot\mathbf{X}_{k-1}^*}, \tag{4.3.10}$$

which implies

$$\theta_0 + \frac{\underline{\boldsymbol{b}}}{\overline{\boldsymbol{a}}} \leqslant \mathbb{E}_{\theta_0}\left((\varepsilon_k^*)^2\,|\mathbf{X}_{k-1}^*\right) \leqslant \theta_0 + \frac{\overline{\boldsymbol{b}}+1}{\underline{\boldsymbol{a}}}. \tag{4.3.11}$$

Hence the conditional variance of the error term ε_k^* in the stochastic regression equation

$$Y_k^* = f^*(\theta_0,\mathbf{X}_{k-1}^*) + \varepsilon_k^*$$

is invariant under multiplication of the whole process by any constant, and bounded respectively to $(\mathbf{X}_k^*)_{k\geqslant 0}$.

The following theorem states the strong consistency and asymptotic normality of the estimator $\widehat{\theta}_n^{X^*}$, as n tends to infinity. We denote by $f^{*'}$ the derivative of f^* with respect to θ.

Theorem 4.3.3. *Let us assume that the process $(\boldsymbol{X}_n)_{n\geqslant 0}$ defined by* (4.1.4) *is subcritical. Then the estimator $\widehat{\theta}_n^{X^*}$ is strongly consistent:*

$$\lim_{n\to\infty}\widehat{\theta}_n^{X^*} \overset{a.s.}{=} \theta_0, \tag{4.3.12}$$

and asymptotically normally distributed:

$$\lim_{n\to\infty}\sqrt{n\frac{\left(\sum_{\boldsymbol{j}\in\mathbb{N}^d}\left(f^{*'}(\theta_0,\boldsymbol{j})\right)^2\pi_{\theta_0}(\boldsymbol{j})\right)^2}{\sum_{\boldsymbol{j}\in\mathbb{N}^d}\left(f^{*'}(\theta_0,\boldsymbol{j})\right)^2 g(\theta_0,\boldsymbol{j})\pi_{\theta_0}(\boldsymbol{j})}}\left(\widehat{\theta}_n^{X^*}-\theta_0\right) \overset{\mathcal{D}}{=} \mathcal{N}\left(0,1\right), \tag{4.3.13}$$

where f^* and g are given by (4.3.9) *and* (4.3.10).

In order to prove this theorem, we need the following technical results.

First, the proofs of Theorem 4.3.3 and of its Corollary 4.3.6 strongly rely on the following strong law of large numbers for homogeneous irreducible positive recurrent Markov chains ([Bil61], Theorem 1.1 and 1.3), applied to the conditioned process $(\mathbf{X}_n^*)_{n\geqslant 0}$ and its stationary distribution π_{θ_0} (see (4.3.5)): for every π_{θ_0}-integrable function $h : \mathbb{N}^d\backslash\{\mathbf{0}\} \to \mathbb{R}$,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} h(\mathbf{X}_k^*) \overset{a.s.}{=} \sum_{\mathbf{j}\in\mathbb{N}^d} h(\mathbf{j})\pi_{\theta_0}(\mathbf{j}). \tag{4.3.14}$$

Next, we prove that the stationary measure π_{θ_0} admits finite first and second-order moments.

Proposition 4.3.4. *The stationary measure π_{θ_0} satisfies for all $i, j = 1\ldots d$,*

$$\sum_{\boldsymbol{k}\in\mathbb{N}^d} k_i\pi_{\theta_0}(\boldsymbol{k}) < \infty, \tag{4.3.15}$$

$$\sum_{\boldsymbol{k}\in\mathbb{N}^d} k_ik_j\pi_{\theta_0}(\boldsymbol{k}) < \infty. \tag{4.3.16}$$

Proof. By using Proposition 4.3.1,

$$\begin{aligned}\mathbb{E}_{\theta_0}\left(X_n^*\right) &= \mathbb{E}_{\theta_0}\left[\mathbb{E}_{\theta_0}\left(X_n^*|\mathbf{X}_{n-1}^*\right)\right]\\ &= \mathbb{E}_{\theta_0}\left[\mathbf{X}_{n-1}^*\cdot\mathbf{\Psi}(\theta_0) + p(\theta_0,\mathbf{X}_{n-1}^*)\right]\\ &\leqslant \sum_{k=1}^d \mathbb{E}_{\theta_0}\left(X_{n-k}^*\right)\Psi_k(\theta_0) + 1,\end{aligned}$$

which implies that

$$\varlimsup_{n\to\infty}\ \mathbb{E}_{\theta_0}\left(X_n^*\right) \leqslant \frac{1}{1-\sum_{k=1}^d \Psi_k(\theta_0)} < \infty.$$

We consequently obtain by means of Fatou's lemma that, for every $i = 1\ldots d$,

$$\sum_{\mathbf{k}\in\mathbb{N}^d} k_i\pi_{\theta_0}(\mathbf{k}) = \mathbb{E}_{\theta_0}\left(\lim_{n\to\infty} X_{n,i}^*\right) = \mathbb{E}_{\theta_0}\left(\lim_{n\to\infty} X_{n-i+1}^*\right) = \mathbb{E}_{\theta_0}\left(\lim_{n\to\infty} X_n^*\right) \leqslant \varlimsup_{n\to\infty}\ \mathbb{E}_{\theta_0}\left(X_n^*\right) < \infty.$$

We similarly prove that π_{θ_0} has finite second-order moments by writing

$$\begin{aligned}\mathbb{V}ar_{\theta_0}\left(X_n^*|\mathbf{X}_{n-1}^*\right) &= \mathbf{X}_{n-1}^*\cdot\mathbf{\Psi}(\theta_0) + \frac{\xi_1(\theta_0)\mathbf{X}_{n-1}^*\cdot\mathbf{\Psi}(\theta_0)\sum_{k=2}^d X_{n-k+1}^*\xi_k(\theta_0)}{\left(\xi_1(\theta_0)\mathbf{X}_{n-1}^*\cdot\mathbf{\Psi}(\theta_0)+\sum_{k=2}^d X_{n-k+1}^*\xi_k(\theta_0)\right)^2}\\ &\leqslant \mathbf{X}_{n-1}^*\cdot\mathbf{\Psi}(\theta_0) + \frac{1}{4}.\end{aligned}$$

□

We finally need the following analytical result. For every $\theta\in\Theta$, we denote by $\rho(\theta)$ the Perron's root of the mean matrix $\mathbf{M}(\theta)$. We thus have in particular $\rho_0 = \rho(\theta_0)$. For the sake of simplicity, for any $k\in\mathbb{R}$, we will write $\rho(\theta)^k$ instead of $[\rho(\theta)]^k$.

Proposition 4.3.5. *Let $\gamma > 0$ such that $\theta_1 - \gamma \geq 0$, and let us write $(\theta_{1,\gamma}, \theta_{2,\gamma}) := (\theta_1-\gamma, \theta_2+\gamma)$. Then $\rho(\theta)$ and is a $\mathcal{C}^3$-function of θ on $\Theta_\gamma :=]\theta_{1,\gamma}, \theta_{2,\gamma}[\ \supset \overline{\Theta}$.*

Moreover, for each $j = 1\ldots d$, the jth coordinate $\xi_j(\theta)$ of $\boldsymbol{\xi}(\theta)$ is a $\mathcal{C}^3$-function of θ on $\Theta_\gamma \supset \overline{\Theta}$.

Proof. It is well-known that the (unordered) spectrum of a matrix depends continuously on the coefficients of the matrix. Since the entries of $\mathbf{M}(\theta)$ are smooth functions of θ, its *unordered* spectrum can be written $\operatorname{Spec}\mathbf{M}(\theta) := \{\lambda_1(\theta),\ldots,\lambda_d(\theta)\}$, where each $\lambda_i(\theta)$ is a complex-valued and continuous function of θ, and the numbering $1,\ldots,d$ does not correspond to any specific order.

Let us first show that $\rho(\theta)$ also depends continuously on θ. Let $\theta \in \Theta_\gamma$. There exists a subscript $1 \leqslant i_0 \leqslant d$ (depending on θ) such that $\rho(\theta) = \lambda_{i_0}(\theta) = \Re(\lambda_{i_0}(\theta))$. We define

$$\tau(\theta) := \min_{i \neq i_0} \left\{ \rho(\theta) - \Re(\lambda_i(\theta)) \right\}.$$

Let $\varepsilon > 0$ such that $\tau(\theta) > \varepsilon$. By continuity of $\lambda_i(\theta)$ and of $\Re(\lambda_i(\theta))$, there exists $\delta > 0$ such that, for all $|\theta' - \theta| \leqslant \delta$ and all $i \neq i_0$, $|\Re(\lambda_i(\theta)) - \Re(\lambda_i(\theta'))| < \tau(\theta) - \varepsilon$, and such that

$$|\rho(\theta) - \Re(\lambda_{i_0}(\theta'))| < \varepsilon. \tag{4.3.17}$$

Then, for all $|\theta' - \theta| \leqslant \delta$, we have, for all $i \neq i_0$,

$$\begin{aligned} \Re(\lambda_{i_0}(\theta')) - \Re(\lambda_i(\theta')) &= \Re(\lambda_{i_0}(\theta')) - \rho(\theta) + \rho(\theta) - \Re(\lambda_i(\theta)) + \Re(\lambda_i(\theta)) - \Re(\lambda_i(\theta')) \\ &> -\varepsilon + \tau(\theta) - (\tau(\theta) - \varepsilon) = 0, \end{aligned}$$

which by definition of $\rho(\theta')$ means that $\rho(\theta') = \Re(\lambda_{i_0}(\theta'))$. Relation (4.3.17) then becomes $|\rho(\theta) - \rho(\theta')| < \varepsilon$, which proves the continuity of $\rho(\theta)$, on Θ_γ.

Let us now prove that $\rho(\theta)$ is differentiable at any $\theta \in \Theta_\gamma$. For this purpose, we use the relation $\sum_{k=1}^d \Psi_k(\theta)\rho(\theta)^{-k} = 1$ (see (4.1.8)), from which it ensues that for every $h \neq 0$,

$$\sum_{k=1}^{d} (a_k\theta + b_k) \frac{\rho(\theta)^{-k} - \rho(\theta+h)^{-k}}{h} - \sum_{k=1}^{d} a_k \rho(\theta+h)^{-k} = 0. \tag{4.3.18}$$

Using the relation

$$\begin{aligned} &\rho(\theta)^{-k} - \rho(\theta+h)^{-k} \\ &\quad = \left(\rho(\theta)^{-1} - \rho(\theta+h)^{-1}\right)\left(\rho(\theta)^{-k+1} + \rho(\theta)^{-k+2}\rho(\theta+h)^{-1} + \ldots + \rho(\theta+h)^{-k+1}\right) \\ &\quad = (\rho(\theta+h) - \rho(\theta)) \frac{\rho(\theta)^{-k+1} + \rho(\theta)^{-k+2}\rho(\theta+h)^{-1} + \ldots + \rho(\theta+h)^{-k+1}}{\rho(\theta+h)\rho(\theta)} \\ &\quad =: (\rho(\theta+h) - \rho(\theta)) Q_k(h), \end{aligned}$$

equation (4.3.18) becomes

$$\frac{\rho(\theta+h) - \rho(\theta)}{h} \sum_{k=1}^{d} (a_k\theta + b_k) Q_k(h) - \sum_{k=1}^{d} a_k \rho(\theta+h)^{-k} = 0,$$

and thus, by continuity of $\rho(\theta)$,

$$\lim_{h \to 0} \frac{\rho(\theta+h) - \rho(\theta)}{h} = \lim_{h \to 0} \frac{\sum_{k=1}^d a_k \rho(\theta+h)^{-k}}{\sum_{k=1}^d (a_k\theta + b_k) Q_k(h)} = \frac{\sum_{k=1}^d a_k \rho(\theta)^{-k}}{\sum_{k=1}^d k(a_k\theta + b_k)\rho(\theta)^{-k-1}} = \rho'(\theta).$$

The last relation implies that $\rho'(\theta)$ is continuous and differentiable on Θ_γ, and we prove by iterating the same method that ρ is $\mathcal{C}^3$ on Θ_γ (we would prove actually the same way that ρ is $\mathcal{C}^\infty$).

The last assertion of Proposition 4.3.5 is then an immediate consequence of (4.2.78). □

Let us now prove Theorem 4.3.3.

Proof of Theorem 4.3.3. The proof of this theorem is quite similar to the one of Theorem 4.2.7. According to Proposition 3.1 in [Jac10B], sufficient conditions for the strong consistency of $\widehat{\theta}_n^{X^*}$ are that 1) $f^*(., \mathbf{X}_{k-1}^*)$ is Lipschitz, in the sense that there exists a nonnegative $\mathcal{F}_{k-1}^*$-measurable function C_k (where $\mathcal{F}_{k-1}^* := \sigma(\mathbf{X}_0^*, \ldots, \mathbf{X}_{k-1}^*)$), satisfying for all $\theta_1, \theta_2 \in \Theta$,

$$\left| f^*(\theta_1, \mathbf{X}_{k-1}^*) - f^*(\theta_2, \mathbf{X}_{k-1}^*) \right| \overset{a.s.}{\leqslant} C_k \left| \theta_1 - \theta_2 \right|,$$

2) that $\overline{\lim}_{k\to\infty}\mathbb{E}_{\theta_0}\left((\varepsilon_k^*)^2\,|\mathbf{X}_{k-1}^*\right)\overset{a.s.}{<}\infty$, and 3) that

$$\lim_{n\to\infty}\inf_{\substack{\theta\in\Theta\\|\theta-\theta_0|\geqslant\delta}}\sum_{k=1}^{n}\left(f^*(\theta_0,\mathbf{X}_{k-1}^*)-f^*(\theta,\mathbf{X}_{k-1}^*)\right)^2\overset{a.s.}{=}\infty. \tag{4.3.19}$$

First, for all $\theta\in\Theta$ and $\mathbf{j}\in\mathbb{N}^d$, $\mathbf{j}\neq\mathbf{0}$, $f^{*'}(\theta,\mathbf{j})=(\boldsymbol{a}\cdot\mathbf{j}+p'(\theta,\mathbf{j}))\,(\boldsymbol{a}\cdot\mathbf{j})^{-1/2}$, where p' denotes the derivative of p with respect to θ, which thanks to (4.3.7) and Proposition 4.3.5 is bounded on Θ. The Lipschitz condition for $f^*(.,\mathbf{X}_{k-1}^*)$ is thus satisfied.

The second condition follows from (4.3.11).

The last condition (4.3.19) comes from the fact that, for every $\delta>0$ and every $\theta\in\Theta$ such that $|\theta-\theta_0|\geqslant\delta$, applying the mean value theorem to the $\mathcal{C}^1$-function $p(.,\mathbf{X}_{k-1}^*)$,

$$\begin{aligned}\sum_{k=1}^{n}\left(f^*(\theta_0,\mathbf{X}_{k-1}^*)-f^*(\theta,\mathbf{X}_{k-1}^*)\right)^2&=(\theta_0-\theta)^2\sum_{k=1}^{n}\boldsymbol{a}\cdot\mathbf{X}_{k-1}^*\left(1+\frac{p\left(\theta_0,\mathbf{X}_{k-1}^*\right)-p\left(\theta,\mathbf{X}_{k-1}^*\right)}{(\theta_0-\theta)\,\boldsymbol{a}\cdot\mathbf{X}_{k-1}^*}\right)^2\\&\geqslant\delta^2\sum_{k=1}^{n}\boldsymbol{a}\cdot\mathbf{X}_{k-1}^*\inf_{\theta\in\overline{\Theta}}\left(1+\frac{p'(\theta,\mathbf{X}_{k-1}^*)}{\boldsymbol{a}\cdot\mathbf{X}_{k-1}^*}\right)^2.\end{aligned}$$

Let us show that the function

$$\mathbf{j}\mapsto\boldsymbol{a}\cdot\mathbf{j}\inf_{\theta\in\overline{\Theta}}\left(1+\frac{p'(\theta,\mathbf{j})}{\boldsymbol{a}\cdot\mathbf{j}}\right)^2$$

is π_{θ_0}-integrable. For every $\theta\in\overline{\Theta}$, $\mathbf{j}\in\mathbb{N}^d$ and $\mathbf{j}\neq\mathbf{0}$, denoting

$$\Xi:=\sup_{i,\theta\in\overline{\Theta}}\{\xi_i(\theta),|\xi_i'(\theta)|\}<\infty$$

(by continuity of $\xi_i'(\theta)$ on $\Theta_\gamma\supset\overline{\Theta}$), and $\underline{\boldsymbol{\xi}}:=\min_{i,\theta\in\overline{\Theta}}\xi_i(\theta)>0$,

$$\begin{aligned}|p'(\theta,\mathbf{j})|&=\left|\frac{\left(\xi_1'(\theta)\mathbf{j}\cdot\boldsymbol{\Psi}(\theta)+\xi_1(\theta)\mathbf{j}\cdot\boldsymbol{a}\right)\sum_{i=2}^d j_i\xi_i(\theta)-\xi_1(\theta)\mathbf{j}\cdot\boldsymbol{\Psi}(\theta)\sum_{i=2}^d j_i\xi_i'(\theta)}{\left(\xi_1(\theta)\mathbf{j}\cdot\boldsymbol{\Psi}(\theta)+\sum_{i=2}^d j_i\xi_i(\theta)\right)^2}\right|\\&\leqslant\frac{\Xi^2}{\underline{\boldsymbol{\xi}}^2}\frac{3\mathbf{j}\cdot\boldsymbol{\Psi}(\theta)\sum_{i=2}^d j_i}{\left(\mathbf{j}\cdot\boldsymbol{\Psi}(\theta)+\sum_{i=2}^d j_i\right)^2}\leqslant\frac{3\Xi^2}{4\underline{\boldsymbol{\xi}}^2}=:B_1.\end{aligned}\tag{4.3.20}$$

Hence, for all $\mathbf{j}\neq\mathbf{0}$,

$$\left|\boldsymbol{a}\cdot\mathbf{j}\inf_{\theta\in\overline{\Theta}}\left(1+\frac{p'(\theta,\mathbf{j})}{\boldsymbol{a}\cdot\mathbf{j}}\right)^2\right|\leqslant\left(1+\frac{B_1}{\underline{\boldsymbol{a}}}\right)^2\boldsymbol{a}\cdot\mathbf{j},$$

and applying (4.3.14) together with (4.3.15) we obtain that

$$\lim_{n\to\infty}\frac{1}{n+1}\sum_{k=1}^{n}\boldsymbol{a}\cdot\mathbf{X}_{k-1}^*\inf_{\theta\in\overline{\Theta}}\left(1+\frac{p'(\theta,\mathbf{X}_{k-1}^*)}{\boldsymbol{a}\cdot\mathbf{X}_{k-1}^*}\right)^2\overset{a.s.}{=}\sum_{\mathbf{j}\in\mathbb{N}^d}\boldsymbol{a}\cdot\mathbf{j}\inf_{\theta\in\overline{\Theta}}\left(1+\frac{p'(\theta,\mathbf{j})}{\boldsymbol{a}\cdot\mathbf{j}}\right)^2\pi_{\theta_0}(\mathbf{j}).\tag{4.3.21}$$

Let $\mathbf{j}\neq\mathbf{0}$ fixed. Since for all $\theta\in\overline{\Theta}$, $p'(\theta,\mathbf{j})\neq-\boldsymbol{a}\cdot\mathbf{j}$, the extreme value theorem implies that

$$\inf_{\theta\in\overline{\Theta}}\left(1+\frac{p'(\theta,\mathbf{j})}{\boldsymbol{a}\cdot\mathbf{j}}\right)^2>0.$$

Hence the right term in (4.3.21) is strictly positive, which together with (4.3.21) leads to (4.3.19).

Let us now consider the asymptotic distribution of $\widehat{\theta}_n^{X^*}-\theta_0$. For this purpose, we follow the steps of the proof of Proposition 6.1 in [Jac10B]. Writing the Taylor expansion of $S_n^{*'}(\theta)$ in the neighborhood of θ_0 we obtain that $\widehat{\theta}_n^{X^*} - \theta_0 = -S_n^{*'}(\theta_0)/S_n^{*''}(\tilde{\theta}_n)$, for some $\tilde{\theta}_n = \theta_0 + t_n\left(\widehat{\theta}_n^{X^*} - \theta_0\right)$, with $t_n \in]0,1[$. Since $S_n^{*'}(\theta_0) = -2\sum_{k=1}^n \varepsilon_k^* f^{*'}(\theta_0, \mathbf{X}_{k-1}^*)$, we can write

$$\sqrt{n}\left(\widehat{\theta}_n^{X^*} - \theta_0\right) = \frac{\sum_{k=1}^n \varepsilon_k^* f^{*'}(\theta_0, \mathbf{X}_{k-1}^*)}{\sqrt{n}} \left(\frac{F_n^*}{n}\right)^{-1} \left(\frac{1}{2}\frac{S_n^{*''}(\tilde{\theta}_n)}{F_n^*}\right)^{-1}, \tag{4.3.22}$$

where

$$F_n^* := \sum_{k=1}^n \left(f^{*'}(\theta_0, \mathbf{X}_{k-1}^*)\right)^2.$$

Let us first show that

$$\lim_{n\to\infty} \frac{F_n^*}{n} \overset{a.s.}{=} \sum_{\mathbf{j}\in\mathbb{N}^d} \left(f^{*'}(\theta_0, \mathbf{j})\right)^2 \pi_{\theta_0}(\mathbf{j}). \tag{4.3.23}$$

This is an application of (4.3.14) and (4.3.15), since for all $\mathbf{j} \in \mathbb{N}^d$, $\mathbf{j} \neq \mathbf{0}$,

$$\left(f^{*'}(\theta_0, \mathbf{j})\right)^2 \leqslant \boldsymbol{a}\cdot\mathbf{j} + 2B_1 + \frac{B_1^2}{\underline{\boldsymbol{a}}}. \tag{4.3.24}$$

In view of (4.3.22), we now prove that

$$\lim_{n\to\infty} \frac{S_n^{*''}(\tilde{\theta}_n)}{F_n^*} \overset{a.s.}{=} 2. \tag{4.3.25}$$

Computing $S_n^{*''}$ thanks to the formula $S_n^*(\theta) = \sum_{k=1}^n \left(\varepsilon_k^* + f^*(\theta_0, \mathbf{X}_{k-1}^*) - f^*(\theta, \mathbf{X}_{k-1}^*)\right)^2$, it appears that (4.3.25) is true, as soon as the following holds:

$$\lim_{n\to\infty}\sup_{\theta\in\Theta} \frac{\left|\sum_{k=1}^n \varepsilon_k^* f^{*''}(\theta, \mathbf{X}_{k-1}^*)\right|}{F_n^*} \overset{a.s.}{=} 0, \tag{4.3.26}$$

$$\lim_{n\to\infty} \frac{\sum_{k=1}^n \left(f^{*'}(\tilde{\theta}_n, \mathbf{X}_{k-1}^*)\right)^2}{F_n^*} \overset{a.s.}{=} 1, \tag{4.3.27}$$

and

$$\lim_{n\to\infty} \frac{\sum_{k=1}^n \left(f^*(\theta_0, \mathbf{X}_{k-1}^*) - f^*(\tilde{\theta}_n, \mathbf{X}_{k-1}^*)\right) f^{*''}(\tilde{\theta}_n, \mathbf{X}_{k-1}^*)}{F_n^*} \overset{a.s.}{=} 0. \tag{4.3.28}$$

Let us prove (4.3.26)-(4.3.28). First, (4.3.26) is given by a strong law of large numbers proved in [Jac10B], Proposition 5.1. The latter can be indeed applied since $\lim_n F_n^* \overset{a.s.}{=} \infty$ (as an immediate consequence of the stronger result (4.3.23)), and since $f^{*''}(., \mathbf{X}_{k-1}^*)$ fulfills the required Lipschitz condition. Indeed, by Proposition 4.3.5, $\xi_i'''(\theta)$ is continuous on the compact set $\overline{\Theta}$ and is thus bounded on $\overline{\Theta}$, which implies that $f^{*'''}(., \mathbf{X}_{k-1}^*) = p'''\left(., \mathbf{X}_{k-1}^*\right)(\boldsymbol{a}\cdot\mathbf{X}_{k-1}^*)^{-1/2}$ is bounded by a $\mathcal{F}_{k-1}^*$-measurable function.

In view of (4.3.27), we consider the function $(f^*(\theta,\mathbf{j}))^2$ and its derivative $2f^{*'}(\theta,\mathbf{j})f^{*''}(\theta,\mathbf{j})$. Similarly as for (4.3.20), one can show that there exists a constant $B_2 > 0$ such that for all $\theta \in \overline{\Theta}$, and all $\mathbf{j} \neq \mathbf{0}$, $|p''(\theta,\mathbf{j})| \leqslant B_2$. This implies

$$\left|2f^{*'}(\theta,\mathbf{j})f^{*''}(\theta,\mathbf{j})\right| \leqslant 2\left|\frac{(\boldsymbol{a}\cdot\mathbf{j} + p'(\theta,\mathbf{j}))\,p''(\theta,\mathbf{j})}{\boldsymbol{a}\cdot\mathbf{j}}\right| \leqslant 2B_2\left(1 + \frac{B_1}{\underline{\boldsymbol{a}}}\right). \tag{4.3.29}$$

Consequently,

$$\frac{\left|\sum_{k=1}^n f^{*'}(\tilde{\theta}_n, \mathbf{X}_{k-1}^*)^2 - f^{*'}(\theta_0, \mathbf{X}_{k-1}^*)^2\right|}{F_n^*} \leqslant 2B_2\left(1 + \frac{B_1}{\underline{\boldsymbol{a}}}\right)\left|\widehat{\theta}_n^{X^*} - \theta_0\right|\left(\frac{F_n^*}{n}\right)^{-1}, \tag{4.3.30}$$

which by (4.3.23) and the strong consistency of $\widehat{\theta}_n^{X^*}$ almost surely tends to 0. Writing

$$\frac{\sum_{k=1}^n f^{*'}(\tilde{\theta}_n, \mathbf{X}_{k-1}^*)^2}{F_n^*} = 1 + \frac{\sum_{k=1}^n \left(f^{*'}(\tilde{\theta}_n, \mathbf{X}_{k-1}^*)^2 - f^{*'}(\theta_0, \mathbf{X}_{k-1}^*)^2 \right)}{F_n^*},$$

this implies (4.3.27).

It now remains to prove (4.3.28). We write

$$\begin{aligned}
&\frac{\left| \sum_{k=1}^n \left(f^*(\theta_0, \mathbf{X}_{k-1}^*) - f^*(\tilde{\theta}_n, \mathbf{X}_{k-1}^*) \right) f^{*''}(\tilde{\theta}_n, \mathbf{X}_{k-1}^*) \right|}{F_n^*} \\
&\leqslant \frac{1}{F_n^*} \sum_{k=1}^n \frac{\boldsymbol{a} \cdot \mathbf{X}_{k-1}^* |\theta_0 - \tilde{\theta}_n| + |p(\theta_0, \mathbf{X}_{k-1}^*) - p(\tilde{\theta}_n, \mathbf{X}_{k-1}^*)|}{\boldsymbol{a} \cdot \mathbf{X}_{k-1}^*} |p''(\tilde{\theta}_n, \mathbf{X}_{k-1}^*)| \\
&\leqslant \frac{1}{F_n^*} \sum_{k=1}^n \left(|\theta_0 - \tilde{\theta}_n| + \frac{B_1 |\theta_0 - \tilde{\theta}_n|}{\underline{a}} \right) B_2 \leqslant |\theta_0 - \widehat{\theta}_n^{X^*}| B_2 \left(1 + \frac{B_1}{\underline{a}} \right) \left(\frac{F_n^*}{n} \right)^{-1},
\end{aligned}$$

which thanks to (4.3.23) and the strong consistency of $\widehat{\theta}_n^{X^*}$ implies (4.3.28).

In view of (4.3.22), we finally want to prove that $\sum_{k=1}^n \varepsilon_k^* f^{*'}(\theta_0, \mathbf{X}_{k-1}^*)/\sqrt{n}$ converges in distribution, and to this end we make again use of Proposition 4.2.8. Let us define, for every $k \leqslant n$,

$$M_k^{(n)} := \frac{1}{\sqrt{n}} \sum_{l=1}^k \varepsilon_l^* f^{*'}(\theta_0, \mathbf{X}_{l-1}^*).$$

First, for every $k \leqslant n$, $\mathbb{E}_{\theta_0} \left(\varepsilon_k^* f^{*'}(\theta_0, \mathbf{X}_{k-1}^*)/\sqrt{n} | \mathbf{X}_{k-1}^* \right) = 0$. Second,

$$\mathbb{E}_{\theta_0} \left(\left(\frac{\varepsilon_k^* f^{*'}(\theta_0, \mathbf{X}_{k-1}^*)}{\sqrt{n}} \right)^2 \Big| \mathbf{X}_{k-1}^* \right) = \frac{\left(f^{*'}(\theta_0, \mathbf{X}_{k-1}^*) \right)^2 g\left(\theta_0, \mathbf{X}_{k-1}^*\right)}{n},$$

where $g\left(\theta_0, \mathbf{X}_{k-1}^*\right)$ is given by (4.3.10), and thus $M_k^{(n)}$ is a sequence of square integrable martingales. Moreover, using inequalities (4.3.11) and (4.3.24), we obtain by (4.3.15),

$$\sum_{\mathbf{j} \in \mathbb{N}^d} \left(f^{*'}(\theta_0, \mathbf{j}) \right)^2 g(\theta_0, \mathbf{j}) \pi_{\theta_0}(\mathbf{j}) \leqslant \left(\theta_0 + \frac{\overline{\boldsymbol{b}} + 1}{\underline{\boldsymbol{a}}} \right) \left(\sum_{\mathbf{j} \in \mathbb{N}^d} \boldsymbol{a} \cdot \mathbf{j}\, \pi_{\theta_0}(\mathbf{j}) + 2B_1 + \frac{B_1^2}{\underline{\boldsymbol{a}}} \right) < \infty. \quad (4.3.31)$$

So, by means of (4.3.14),

$$\begin{aligned}
\lim_{n \to \infty} \langle M_n \rangle^{(n)} &= \lim_{n \to \infty} \sum_{k=1}^n \mathbb{E}_{\theta_0} \left(\left(\frac{\varepsilon_k f^{*'}(\theta_0, \mathbf{X}_{k-1}^*)}{\sqrt{n}} \right)^2 \Big| \mathbf{X}_{k-1}^* \right) \\
&\overset{a.s.}{=} \sum_{\mathbf{j} \in \mathbb{N}^d} \left(f^{*'}(\theta_0, \mathbf{j}) \right)^2 g(\theta_0, \mathbf{j}) \pi_{\theta_0}(\mathbf{j}).
\end{aligned}$$

Third, using Cauchy-Schwarz and Bienaymé-Chebyshev inequalities,

$$\begin{aligned}
&\sum_{k=1}^n \mathbb{E}_{\theta_0} \left[\left| \frac{\varepsilon_k^* f^{*'}(\theta_0, \mathbf{X}_{k-1}^*)}{\sqrt{n}} \right|^2 \mathbf{1}_{\left\{ \left| \frac{\varepsilon_k^* f^{*'}(\theta_0, \mathbf{X}_{k-1}^*)}{\sqrt{n}} \right| \geqslant \varepsilon \right\}} \Big| \mathbf{X}_{k-1}^* \right] \\
&\leqslant \sum_{k=1}^n \left(\mathbb{E}_{\theta_0} \left[\left| \frac{\varepsilon_k^* f^{*'}(\theta_0, \mathbf{X}_{k-1}^*)}{\sqrt{n}} \right|^4 \Big| \mathbf{X}_{k-1}^* \right] \right)^{\frac{1}{2}} \left(\mathbb{P}_{\theta_0} \left[\left| \frac{\varepsilon_k^* f^{*'}(\theta_0, \mathbf{X}_{k-1}^*)}{\sqrt{n}} \right| \geqslant \varepsilon \Big| \mathbf{X}_{k-1}^* \right] \right)^{\frac{1}{2}} \\
&\leqslant \frac{1}{n^{\frac{3}{2}} \varepsilon} \sum_{k=1}^n |f^{*'}(\theta_0, \mathbf{X}_{k-1}^*)|^3 \left(\mathbb{E}_{\theta_0} \left[(\varepsilon_k^*)^4 | \mathbf{X}_{k-1}^* \right] \right)^{\frac{1}{2}} \left(\mathbb{E}_{\theta_0} \left[(\varepsilon_k^*)^2 | \mathbf{X}_{k-1}^* \right] \right)^{\frac{1}{2}}. \quad (4.3.32)
\end{aligned}$$

Let us compute $\mathbb{E}_{\theta_0}((\varepsilon_k^*)^4 \,|\mathbf{X}_{k-1}^*)$. We can show that the 4th central moment of the independent sum of a Poisson and a Bernoulli random variables, equals $\mu_4 + 6\mu_2\gamma_2 + \gamma_4$, where μ_i and γ_i denote the ith central moment of the Poisson and of the Bernoulli variable. If these variables have parameter λ and p, respectively, then $\mu_4 = \lambda(1+3\lambda)$, $\mu_2 = \lambda$, $\gamma_4 = p(1-p)(3p^2 - 3p + 1) \in [0,1]$, and $\gamma_2 = p(1-p) \in [0,1]$. We thus obtain

$$\left|\mathbb{E}_{\theta_0}\left((\varepsilon_k^*)^4 \,|\mathbf{X}_{k-1}^*\right)\right| \leqslant \frac{\boldsymbol{\Psi}(\theta_0)\cdot\mathbf{X}_{k-1}^*\left(7 + 3\boldsymbol{\Psi}(\theta_0)\cdot\mathbf{X}_{k-1}^*\right) + 1}{\left(\boldsymbol{a}\cdot\mathbf{X}_{k-1}^*\right)^2}.$$

Hence

$$\begin{aligned}
&|f^{*'}(\theta_0, \mathbf{X}_{k-1}^*)|^3 \left(\mathbb{E}_{\theta_0}\left[\varepsilon_k^4|\mathbf{X}_{k-1}^*\right]\right)^{\frac{1}{2}} \left(\mathbb{E}_{\theta_0}\left[\varepsilon_k^2|\mathbf{X}_{k-1}^*\right]\right)^{\frac{1}{2}} \\
&\quad\leqslant \left(\sqrt{\boldsymbol{a}\cdot\mathbf{X}_{k-1}^*} + \frac{B_1}{\sqrt{\underline{\boldsymbol{a}}}}\right)^3 \frac{\sqrt{\boldsymbol{\Psi}(\theta_0)\cdot\mathbf{X}_{k-1}^*\left(7 + 3\boldsymbol{\Psi}(\theta_0)\cdot\mathbf{X}_{k-1}^*\right) + 1}}{\boldsymbol{a}\cdot\mathbf{X}_{k-1}^*} \left(\theta_0 + \frac{\overline{\boldsymbol{b}} + 1}{\underline{\boldsymbol{a}}}\right)^{\frac{1}{2}}. \qquad (4.3.33)
\end{aligned}$$

Since the highest power of X_n^* involved in (4.3.33) is 3/2, and since the stationary distribution π_{θ_0} has finite second-moments (see Proposition 4.3.4), we can apply (4.3.14) to (4.3.32) and obtain that

$$\lim_{n\to\infty} \sum_{k=1}^{n} \mathbb{E}_{\theta_0}\left[\left|\frac{\varepsilon_k^* f^{*'}(\theta_0, \mathbf{X}_{k-1}^*)}{\sqrt{n}}\right|^2 \mathbf{1}_{\left\{\left|\frac{\varepsilon_k^* f^{*'}(\theta_0, \mathbf{X}_{k-1}^*)}{\sqrt{n}}\right| \geqslant \varepsilon\right\}} \middle| \mathbf{X}_{k-1}^*\right] \overset{a.s.}{=} 0.$$

It then ensues from Proposition 4.2.8 that

$$\lim_{n\to\infty} \frac{\sum_{k=1}^n \varepsilon_k^* f^{*'}(\theta_0, \mathbf{X}_{k-1}^*)}{\sqrt{n}} \overset{\mathcal{D}}{=} \mathcal{N}\left(0, \sum_{\mathbf{j}\in\mathbb{N}^d} \left(f^{*'}(\theta_0, \mathbf{j})\right)^2 g(\theta_0, \mathbf{j})\, \pi_{\theta_0}(\mathbf{j})\right). \qquad (4.3.34)$$

Finally, (4.3.22) together with (4.3.23), (4.3.25), (4.3.34) and Slutsky's Lemma imply that

$$\lim_{n\to\infty} \sqrt{n}\left(\widehat{\theta}_n^{X^*} - \theta_0\right) \overset{\mathcal{D}}{=} \mathcal{N}\left(0, \frac{\sum_{\mathbf{j}\in\mathbb{N}^d} \left(f^{*'}(\theta_0, \mathbf{j})\right)^2 g(\theta_0, \mathbf{j})\, \pi_{\theta_0}(\mathbf{j})}{\left(\sum_{\mathbf{j}\in\mathbb{N}^d} \left(f^{*'}(\theta_0, \mathbf{j})\right)^2 \pi_{\theta_0}(\mathbf{j})\right)^2}\right). \qquad (4.3.35)$$

□

Theorem 4.3.3 is clearly not directly applicable: first, the stationary distribution π_{θ_0} is in general not explicitly known, and second, (4.3.13) involves the function $f^{*'}$, and thus requires the knowledge of the derivative of the function $\xi_j(\theta_0)$, which is not an explicit function of θ_0 since ρ_0 itself is not an explicit function of θ_0. In order to solve this problem, we deduce from (4.3.13) the following more practical result.

Corollary 4.3.6. *The estimator $\widehat{\theta}_n^{X^*}$ has the following asymptotic distribution*

$$\lim_{n\to\infty} \frac{\sum_{k=0}^n \left(f^{*'}\left(\widehat{\theta}_n^{X^*}, \boldsymbol{X}_k^*\right)\right)^2}{\sqrt{\sum_{k=0}^n \left(f^{*'}\left(\widehat{\theta}_n^{X^*}, \boldsymbol{X}_k^*\right)\right)^2 g\left(\widehat{\theta}_n^*, \boldsymbol{X}_k^*\right)}} \left(\widehat{\theta}_n^{X^*} - \theta_0\right) \overset{\mathcal{D}}{=} \mathcal{N}(0, 1). \qquad (4.3.36)$$

Proof. The result is immediate as soon as we prove that

$$\lim_{n\to\infty} \frac{1}{n+1} \sum_{k=0}^{n} \left(f^{*'}\left(\widehat{\theta}_n^{X^*}, \mathbf{X}_k^*\right)\right)^2 g\left(\widehat{\theta}_n^{X^*}, \mathbf{X}_k^*\right) \overset{a.s.}{=} \sum_{\mathbf{j}\in\mathbb{N}^d} \left(f^{*'}(\theta_0, \mathbf{j})\right)^2 g(\theta_0, \mathbf{j})\, \pi_{\theta_0}(\mathbf{j}), \qquad (4.3.37)$$

as well as the equivalent result for the numerator. For this purpose, we write

$$\sum_{k=0}^{n}\left(f^{*'}\left(\widehat{\theta}_n^{X^*},\mathbf{X}_k^*\right)\right)^2 g\left(\widehat{\theta}_n^{X^*},\mathbf{X}_k^*\right)=\sum_{k=0}^{n}\left(f^{*'}\left(\theta_0,\mathbf{X}_k^*\right)\right)^2 g\left(\theta_0,\mathbf{X}_k^*\right)$$
$$+\sum_{k=0}^{n}\left[\left(f^{*'}\left(\widehat{\theta}_n^{X^*},\mathbf{X}_k^*\right)\right)^2 g\left(\widehat{\theta}_n^{X^*},\mathbf{X}_k^*\right)-\left(f^{*'}\left(\theta_0,\mathbf{X}_k^*\right)\right)^2 g\left(\theta_0,\mathbf{X}_k^*\right)\right], \quad (4.3.38)$$

and show that $(f^{*'}(.,\mathbf{j}))^2 g(.,\mathbf{j})$ has a bounded derivative and is thus Lipschitz. We have indeed $|g'(\theta,\mathbf{j})| = |\boldsymbol{a}\cdot\mathbf{j}+p'(\theta,\mathbf{j})|\,(\boldsymbol{a}\cdot\mathbf{j})^{-1} \leqslant 1+B_1\underline{\boldsymbol{a}}^{-1}$, hence

$$\left|2f^{*''}(\theta,\mathbf{j})f^{*'}(\theta,\mathbf{j})g(\theta,\mathbf{j})+(f^{*'}(\theta,\mathbf{j}))^2 g'(\theta,\mathbf{j})\right|$$
$$\leqslant 2B_2\left(1+B_1\underline{\boldsymbol{a}}^{-1}\right)\underline{\boldsymbol{a}}^{-1}\left(\boldsymbol{\Psi}(\theta)\cdot\mathbf{j}+1\right)+\left(1+B_1\underline{\boldsymbol{a}}^{-1}\right)^2\left(\boldsymbol{a}\cdot\mathbf{j}+B_1\right)\leqslant B_3\boldsymbol{\Psi}(\theta_2)\cdot\mathbf{j},$$

for some constant $B_3>0$. This enables to write

$$\frac{1}{n+1}\sum_{k=0}^{n}\left|\left(f^{*'}\left(\widehat{\theta}_n^{X^*},\mathbf{X}_k^*\right)\right)^2 g\left(\widehat{\theta}_n^{X^*},\mathbf{X}_k^*\right)-\left(f^{*'}\left(\theta_0,\mathbf{X}_k^*\right)\right)^2 g\left(\theta_0,\mathbf{X}_k^*\right)\right|$$
$$\leqslant\left|\widehat{\theta}_n^{X^*}-\theta_0\right| B_3\frac{1}{n+1}\sum_{k=0}^{n}\boldsymbol{\Psi}(\widehat{\theta}_{max})\cdot\mathbf{X}_k^*. \quad (4.3.39)$$

By the strong consistency of $\widehat{\theta}_n^{X^*}$ together with (4.3.14) and (4.3.15), (4.3.39) almost surely tends to zero. Combined with (4.3.14) and (4.3.31) in (4.3.38), this implies (4.3.37). □

Remark 4.3.7. We point out that $\xi_j'(\theta)$ is a function of $\rho(\theta)$ and $\rho'(\theta)$, because $\xi_j(\theta)$ is an explicit function of $\rho(\theta)$ (see (4.2.78)). Moreover $\rho'(\theta)$ satisfies

$$\rho'(\theta)=\frac{\sum_{k=1}^{d}a_k\rho(\theta)^{-k}}{\sum_{k=1}^{d}k(a_k\theta+b_k)\rho(\theta)^{-k-1}}$$

(see Proposition 4.3.5). Consequently (4.3.36) can be used as soon as $\rho(\widehat{\theta}_n^{X^*})$ is known. In general the largest solution $\rho(\theta)$ of $\sum_{k=1}^{d}\Psi_k(\theta)\rho(\theta)^{-k}=1$ is not an explicit function of θ, but can be easily numerically approximated.

Chapter 5

BSE epidemic in Great-Britain

The bovine spongiform encephalopathy (BSE, or "mad cow disease") is a fatal neurodegenerative transmissible disease in cattle due to self-replicating proteins, the prions. The epidemic in Great Britain was first officially identified in 1986 ([Wel87]), and reached its peak in 1992 (36682 cases). The epidemic is nowadays in its decay phase, and since only a very few cases were recently reported (33 in 2008 and 9 from January to September 2009, see [OIE]), the spread of the disease should *a priori* "soon" come to an end.

Until the main feed ban regulation, introduced by the British government in July 1988, on the feeding of protein derived from ruminants to any ruminant, the main routes of transmission of BSE were horizontal *via* protein supplements (Meat and Bone Meal, milk replacers), and maternal from a cow to its calf. A previous statistical study ([Jac10A]) concluded to the full efficiency of the 1988 ban. Since most of cattle are slaughtered before the age of 10 years, the fact that cases of BSE are still observed more than 20 years later could suggest the existence of a remaining source of infection, either *via* a maternal transmission route, or *via* a horizontal one, for example *via* the ingestion of excreted prions from alive infected animals. The prediction of the future disease spread thus strongly relies on the intensity of this remaining infection.

Our goal is therefore to estimate this remaining infection and to predict in a fine way the future epidemic evolution, based on a stochastic model taking into account the variability of transmission, of incubation time and of survival. More precisely, besides the estimation of the remaining infection, we plan to predict the incidences of cases and of infected cattle, to estimate the distributions of the epidemic extinction time and that of the total number of cases until extinction (epidemic size), and finally to analyze the behavior of the future disease spread in case of a very late extinction. To this end, we use a stochastic model on the yearly incidences of cases of the form (4.1.1). This model was elaborated in [Jac10A] as the limit process, as the initial population size increases to infinity, of a stochastic age and population-dependent branching process taking into account all the health stages of the disease and the variability of the infection process, under the assumption that the disease is rare at the initial time of the epidemic, and that the probability for an animal to be infected follows a Reed-Frost model. We assume that the process starting from 1989 is time-homogeneous, that is, we neglect the different regulations and breeding changes occurring from 1989 which could have some influence on the probability for an animal to catch the disease.

The limit model of the form (4.1.1) is a recursive process with Poissonian transitions and memory of order $d := a_M - 1$, where a_M is the largest reported survival age with a nonnegligible probability (here $a_M = 10$). The recursivity of the process allows long-term predictions contrary to usual back-calculation methods, and the branching stochastic and integer-valued character of the process allows a realistic random and finite extinction time, contrary to more classical deterministic models that predict the extinction only as the time tends to infinity.

The purpose of this chapter is to quantify the infection of an epidemic in its different phases (growth phase, decay phase, and decay phase assuming a very late extinction) by using appropriate estimators of the infection parameter for each of these phases, previously introduced in Section

Year	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998
Cases	7137	14181	25032	36682	34370	23945	14302	8016	4312	3179
Year	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
Cases	2274	1355	1113	1044	549	309	203	104	53	33

Table 5.1: Yearly number of cases of BSE reported in Great-Britain from 1989 to 2008 ([OIE]).

4.2. The epidemic model which is used here is suitable for any rare transmissible $SEIR$ disease in a large branching population following a Reed-Frost model for the infection. The process corresponds to the incidence of the clinical cases, which are assumed to be the only available observations.

Our study is based on the yearly number of cases of BSE reported in Great Britain from 1989 to 2008 (Figure 5.1 and Table 5.1) provided by the World Organisation for Animal Health ([OIE]). No data are yet available after September 2009.

The outline of the chapter is as follows. We first present in **Section 5.1** the epidemic model. **Section 5.2** is devoted to the prediction of the disease spread. In **Subsection 5.2.1**, we estimate the remaining current infection parameter θ_0, the other parameters of the model being estimated from previous works ([Jac10A] for the incubation distribution, [Don97] for the survival probabilities, and [Brad96, Don97] for the maternal infection). We use the CLSE introduced in Subsection 4.2.1. It has the advantage to rely only on the distribution of the process, contrary to the Bayesian estimator provided in [Jac10A] that also requires an arbitrary prior distribution of θ_0. Moreover, it is strongly consistent and asymptotically normal, when the initial population size tends to infinity, which enables the elaboration of a confidence interval. Furthermore, it is shown in Section 4.2.4, on the base of simulations, that this estimator provides accurate estimations in the setting of the epidemic phase starting from 1989. Since in [Jac10A] some correlations were shown between the infection parameter and the incubation distribution parameters, and between the two kinds of infection parameters (maternal and horizontal), we analyze the sensitivity of the estimator to the incubation and maternal infection parameters. Then we predict in **Subsection 5.2.2**, by means of simulations, the future yearly incidences of cases, as well as the future yearly incidences of infected cattle. This last quantity is an important piece of information, because non diagnosed infected cattle could enter the human food process, and this proportion of animals could be much higher than the incidence of cases, even in the decay phase, because of the large proportion of cattle slaughtered at low ages and the possibility for those that are infected to have not achieve their incubation. We next predict in **Subsection 5.2.3** and in **Subsection 5.2.4**, the extinction time distribution and the total epidemic size distribution, not simply by means of simulations, but also using their exact theoretical distribution and the confidence interval obtained for θ_0. The final **Section 5.3** is devoted to the very late extinction case, which means that we study and predict the future incidence of cases in case of a very late extinction of the disease. We know, thanks to the prediction of the extinction time in **Subsection 5.2.3**, that such a scenario is very unlikely. It must however be carefully studied in order to provide a complete risk analysis of the future spread of the BSE epidemic, since it corresponds to the most dangerous scenario. We estimate the infection parameter for the Q-process associated with the epidemic model, using the estimator introduced in Section 4.3.2, in order to predict the evolution of the incidence of cases and of the incidence of infected cattle, in case of a very late extinction.

5.1 The epidemic model

5.1.1 Description

The epidemic model used here has been elaborated by C. Jacob and introduced in [Jac10A]. It is obtained as the limit, as the initial size of the population tends to infinity, of an age and population-dependent branching process based on the random health state evolution of each animal, taking into account the different types of transmission routes. We will briefly recall this

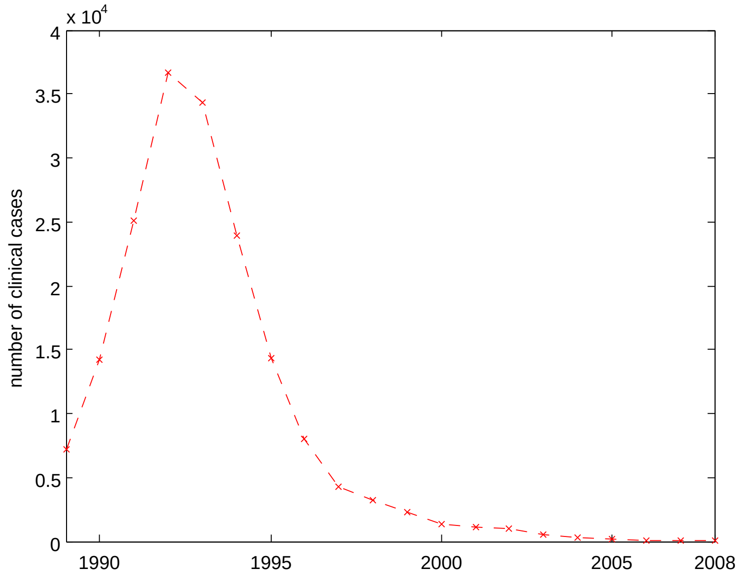

Figure 5.1: Yearly number of cases of BSE reported in Great Britain from 1989 to 2008 ([OIE]).

result in **Subsection 5.1.4**.

The BSE can be considered as a $SEIR$ disease, which means that the population can be classified into those who are *susceptible* (S) to the disease, those who have been infected but are not yet infectious themselves (E for *exposed*), those who are *infectious* (I), and finally those who are *removed* (R) of the susceptible population, either because they recover and gets immune or because of their death. The latter is what happens for the BSE, which is a fatal disease.

However, in order to fit the observations available in Great-Britain for the BSE epidemic, which are the yearly incidences of cases (*i.e.* animals with clinical symptoms at the very end of their incubation period, which can last several years), the classification of the health status slightly differs in [Jac10A] from the standard convention labels. The classification is the following (see Figure 5.2) and consists of four states:

- S susceptible animals,
- E infected animals without clinical symptoms but possibly infectious,
- I clinical cases,
- R slaughtered animals.

One of the difficulty of the model is that S and E animals seem identical, *i.e.* are apparently healthy. The only available observations are the incidence of cases I (Figure 5.3). As already mentioned in the introduction, the different types of transmission routes are maternal from a cow to its calf, horizontal from supplemental feeding made of animal proteins, and possibly from excreting infectious animals.

The epidemic model elaborated in [Jac10A] is the following. It is a random function of the only yearly incidences of cases $(X_n)_n$. More precisely, it is a Markovian process of order $d := a_M - 1 = 9$, where $a_M = 10$ is the largest reported survival age with a nonnegligible probability:

$$X_n = \sum_{k=1}^{d} \sum_{i=1}^{X_{n-k}} \zeta_{n-k,n,i},\ n \geqslant 1. \tag{5.1.1}$$

The $\{\zeta_{n-k,n,i}\}_i$, that represent the number of secondary cases produced at time n with a delay k (incubation time) from each case i of time $n-k$, are i.i.d. given $\mathcal{F}_{n-1} := \sigma\big(\{X_{n-k}\}_{k\geqslant 1}\big)$, with a

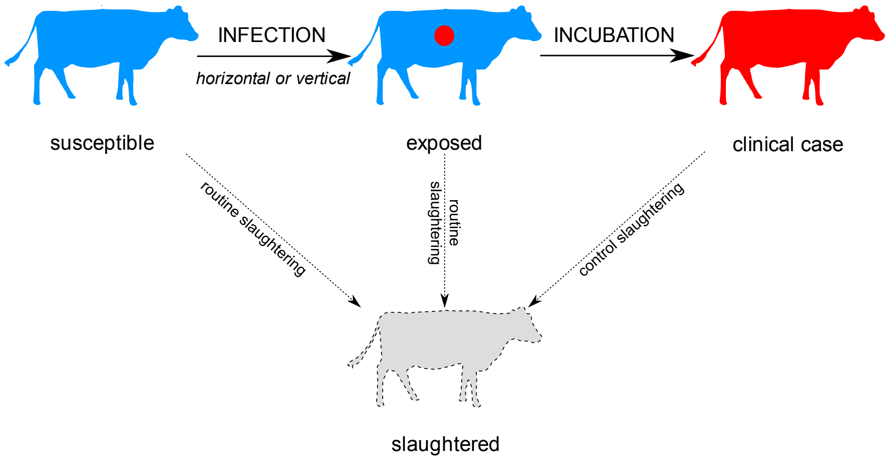

Figure 5.2: Evolution of the health status of an animal in the BSE epidemic.

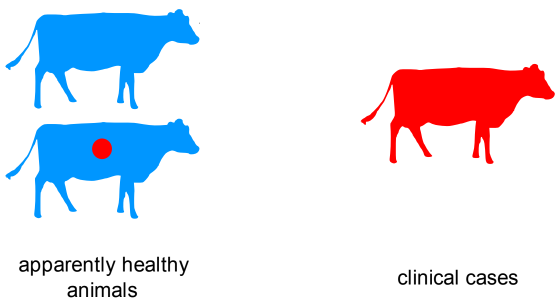

Figure 5.3: Available observations.

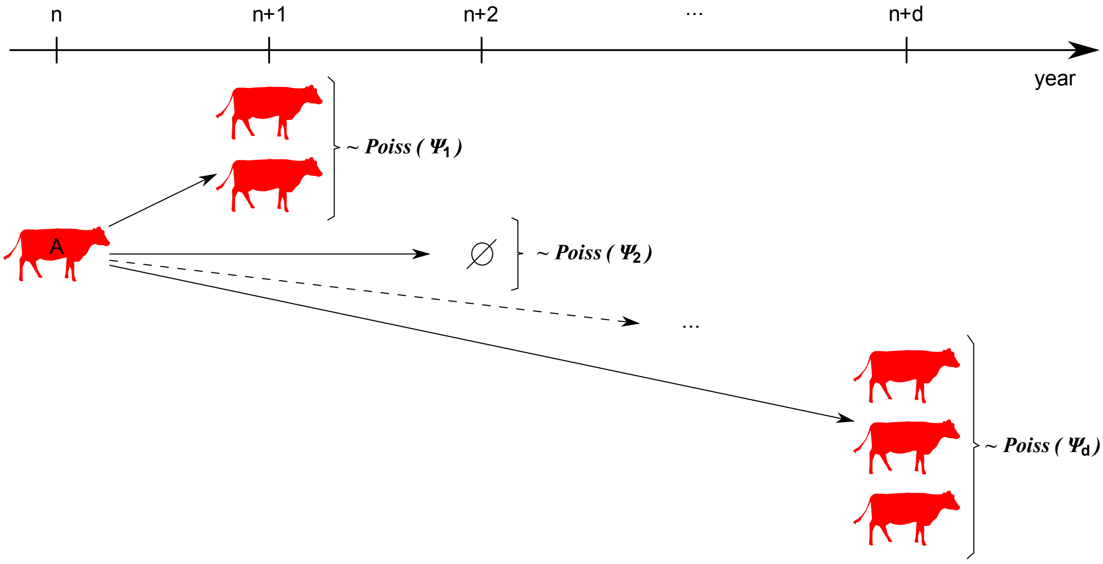

Figure 5.4: Stochastic model for the BSE epidemic.

common Poisson distribution with parameter $\Psi_k(\theta_0)$ independent of n, and the $\{\zeta_{n-k,n,i}\}_{i,k}$ are independent given $\mathcal{F}_{n-1}$. The epidemic model thus belongs to the class of processes of the form (4.1.1) introduced in Section 4.1. Figure 5.4 illustrates how a clinical case alive at time n (animal A in the figure) generates secondary cases several years later. This will be justified in **Subsection 5.1.4**, notably with Figure 5.6.

The first time $n = 1$ of the model is chosen here in order that the model covers the period starting from 1989, that we consider here as a time-homogeneous period, that is $X_{1-d} = X^{obs}_{1989}$, and therefore $n = 1$ corresponds to $1989 + d = 1998$.

In addition, for $k = 1 \ldots d$,

$$\Psi_k(\theta_0) := \theta_0 \frac{\sum_{i=k+1}^{d+1} S_i}{\sum_{j=1}^{d+1} S_j} P_{inc.}(k) + p_{mat.} \frac{S_{k+1}}{\sum_{j=1}^{d+1} S_j} P_{inc.}(k), \tag{5.1.2}$$

denoted from now on

$$\Psi_k(\theta_0) = a_k \theta_0 + b_k, \tag{5.1.3}$$

where

- θ_0 is the unknown remaining infection parameter from 1989, *via* a horizontal route of transmission, *i.e.* the mean number per infective and per year of newly infected cattle *via*, for example, the ingestion of excreted prions from other alive infected cattle;

- $p_{mat.}$ is the maternal infection parameter, *i.e.* the probability for a calf with an infectious mother to be infected by its mother during the first year of its life;

- S_k is the probability of survival for apparently healthy animals; the quantity

$$\frac{S_{k+1}}{\sum_{j=1}^{d+1} S_j}$$

  represents the percentage of animals aged $k + 1$ years, at any year;

- $P_{inc.}(k)$ denotes the probability that the intrinsic incubation period equals k years. Following [Jac10A], we assume here a discretized Weibull distribution with parameters α, β,

$$P_{inc.}(k) := e^{-\frac{\alpha-1}{\alpha\beta^\alpha}(k-1)^\alpha} - e^{-\frac{\alpha-1}{\alpha\beta^\alpha}k^\alpha},$$

where α is a shape parameter, and β is the mode of the probability density of the corresponding Weibull distribution.

We assume here that $\Psi_i(\theta_0) > 0$, for $i = 1 \dots d$, which means that at least either $p_{mat.} > 0$ or $\theta_0 > 0$. The epidemic process defined by (5.1.1) can thus be written as a positive regular and non simple d-type BGW process (see Section 4.1).

5.1.2 Theoretical results

In addition to the properties satisfied by the process (5.1.1) which have already been mentioned in Section 4.1, we quote here some theoretical results that will be used later, namely the distribution of the extinction time and the distribution of the total size of the process until extinction.

As usual in stochastic processes, these quantities are calculated conditionally on the value of $\mathbf{X}_0$, but for the sake of simplicity, we do not make it appear in the notations.

Extinction time

Let

$$T := \inf\{n \geqslant 1, \mathbf{X}_n = \mathbf{0}\} \tag{5.1.4}$$

denote the extinction time of the process $(\mathbf{X}_n)_{n\geqslant 0}$, and

$$\boldsymbol{f}_n^{\theta_0}(\mathbf{r}) := \boldsymbol{f}^{\theta_0}[\boldsymbol{f}_{n-1}^{\theta_0}(\mathbf{r})]$$

the nth iterate of the generating function $\boldsymbol{f}^{\theta_0} := (f_1^{\theta_0}, \dots, f_d^{\theta_0})$ given by (4.1.5). Then

$$\begin{aligned}\mathbb{P}_{\theta_0}(T \leqslant n) = \mathbb{P}_{\theta_0}(\mathbf{X}_n = \mathbf{0}) &= \left(\boldsymbol{f}_n^{\theta_0}(\mathbf{0})\right)^{\mathbf{X}_0} \\ &= \left(f_{n,1}^{\theta_0}(\mathbf{0})\right)^{X_{0,1}} \dots \left(f_{n,d}^{\theta_0}(\mathbf{0})\right)^{X_{0,d}}. \end{aligned} \tag{5.1.5}$$

Tree size

Let

$$N := \sum_{n=1}^{T} X_n \tag{5.1.6}$$

be the size of the tree generated by the process $(\mathbf{X}_n)_{n\geqslant 1}$, thus excluding $\mathbf{X}_0$. Then, in the subcritical case (see [JacPe10]),

$$N \overset{\mathcal{D}}{=} \oplus_{k=1}^{d} \oplus_{i=1}^{X_{-k+1}} \left(\oplus_{j=1}^{Y_{k,i}} N_{k,i,j}\right), \tag{5.1.7}$$

where $\oplus$ denotes the mutual independence (with the convention $\oplus_{k=1}^{0} := 0$), where the $\{Y_{k,i}\}_{k,i}$ are independent, the $\{Y_{k,i}\}_i$ i.i.d with

$$Y_{k,i} \sim \mathcal{P}oisson(\sum_{l=k}^{d} \Psi_l(\theta_0)),$$

and the $\{N_{k,i,j}\}_{k,i,j}$ i.i.d. with

$$N_{k,i,j} \sim \mathcal{B}orel - \mathcal{T}anner(\sum_{l=1}^{d} \Psi_l(\theta_0), 1),$$

that is,

$$\mathbb{P}(N_{k,i,j} = n) = e^{-n\sum_{l=1}^{d}\Psi_l(\theta_0)} \frac{\left(n\sum_{l=1}^{d}\Psi_l(\theta_0)\right)^{n-1}}{n!}, \; n \geqslant 1. \tag{5.1.8}$$

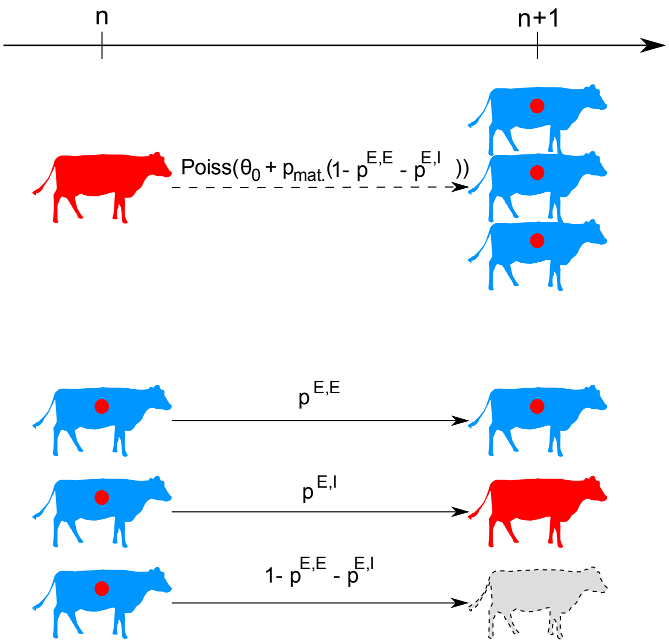

Figure 5.5: 2-type epidemic model, in the case of a geometric incubation time and life span.

This implies

$$\mathbb{E}_{\theta_0}(N) = \sum_{k=1}^{d} X_{-(k-1)} \left(\frac{\sum_{l=k}^{d} \Psi_l(\theta_0)}{1 - \sum_{l=1}^{d} \Psi_l(\theta_0)} \right), \tag{5.1.9}$$

$$\mathbb{V}ar_{\theta_0}(N) = \sum_{k=1}^{d} X_{-(k-1)} \frac{\sum_{l=k}^{d} \Psi_l(\theta_0)}{\left(1 - \sum_{l=1}^{d} \Psi_l(\theta_0)\right)^3}. \tag{5.1.10}$$

5.1.3 The geometrical case

If we assume the particular (unrealistic) case of a geometric incubation time and life span with $d = \infty$, then it is also possible to rewrite the epidemic model according to a 2-type BGW, by considering the hidden Markov process keeping track of the animals in the incubation period, which constitute the memory of the process.

Let us denote by $p^{E,E}$ the probability for a E animal to stay one time unit more in the E state, and $p^{E,I}$ its probability to become I at the following time (see Figure 5.5).

Let $\mathbf{Y}_n := \left(Y_n^I, Y_n^E\right)$, where the two coordinates correspond respectively to the incidence of clinical cases I and to the number of E animals, at time n. The process $\mathbf{Y}_n$ may be written as follows:

$$\begin{aligned} Y_n^I &= \sum_{i=1}^{Y_{n-1}^E} \delta_{n,i}^{E,I}, \\ Y_n^E &= \sum_{i=1}^{Y_{n-1}^I} \zeta_{n,i}^E + \sum_{i=1}^{Y_{n-1}^E} \delta_{n,i}^{E,E}, \end{aligned} \tag{5.1.11}$$

where

$$\zeta_{n,i}^E | \mathcal{F}_{n-1}^Y \sim \mathcal{P}oisson\left(\theta_0 + p_{mat}\left(1 - p^{E,E} - p^{E,I}\right)\right),$$
$$\delta_{n,i}^{E,E} | \mathcal{F}_{n-1}^Y \sim \mathcal{B}\left(p^{E,E}\right),$$

$$\delta^{E,I}_{n,i} | \mathcal{F}^Y_{n-1} \sim \mathcal{B}\left(p^{E,I}\right).$$

We can show that the first coordinate Y^I_n then well corresponds to the epidemic process X^n. Indeed, let us define

$$\zeta_{n-k,n,i} := \sum_{j=1}^{\zeta^E_{n-k+1,i}} \delta^{E,E}_{n-k+1,n-k+2,i,j} \dots \delta^{E,I}_{n-1,n,i,j}.$$

Using (5.1.11), we obtain that

$$\begin{aligned} Y^I_n &= \sum_{k=2}^{\infty} \sum_{i=1}^{Y^I_{n-k}} \zeta_{n-k,n,i}, \\ \zeta_{n-k,n,i} | \mathcal{F}^Y_{n-1} &\sim \mathcal{P}oisson\left(\Psi^I_k(\theta_0)\right), \end{aligned} \tag{5.1.12}$$

with, for $k \geqslant 2$,

$$\Psi^I_k\left(\theta_0\right) = \left(\theta_0 + p_{mat}\left(1 - p^{E,E} - p^{E,I}\right)\right)\left(p^{E,E}\right)^{k-2} p^{E,I}.$$

Moreover, since under the geometric distributions, the survival law $(S_k)_{k \geqslant 1}$ and the intrinsic incubation period distribution $(P_{inc.}(k))_{k \geqslant 1}$ are given by

$$\begin{aligned} S_k &= \left(p^{E,E} + p^{E,I}\right)^k, \\ P_{inc.}(k) &= \left(\frac{p^{E,E}}{p^{E,E} + p^{E,I}}\right)^{k-1} \frac{p^{E,I}}{p^{E,E} + p^{E,I}}, \end{aligned}$$

then, for every $k \geqslant 2$,

$$\Psi^I_k\left(\theta_0\right) = \theta_0 \frac{\sum_{i=k}^{\infty} S_i}{\sum_{j=1}^{\infty} S_j} P_{inc.}(k-1) + p_{mat.} \frac{S_k}{\sum_{j=1}^{\infty} S_j} P_{inc.}(k-1). \tag{5.1.13}$$

Comparing (5.1.12)-(5.1.13) with (5.1.1)-(5.1.2), we see that we obtain similar models. The only difference lies in the fact that for every $k \geqslant 2$, $\Psi^I_k = \Psi_{k-1}$. Indeed, in the 2-type model (5.1.11), a clinical case I cannot generate secondary cases one year later (see Figure 5.5) and we thus have $\Psi^I_1 = 0$.

In this setting of a geometric incubation time and life span, the study of the incidence of cases thus reduces to the study of (5.1.11), which is a 2-type BGW process with mean matrix

$$\begin{pmatrix} 0 & \theta + p_{mat}(1 - p^{E,E} - p^{E,I}) \\ p^{E,I} & p^{E,E} \end{pmatrix}.$$

Its Perron's root is

$$\rho = \frac{p^{E,E} + \sqrt{(p^{E,E})2 + 4(\theta + p_{mat}(1 - p^{E,E} - p^{E,I}))p^{E,I}}}{2},$$

and we obtain that

$$\rho \leqslant 1 \iff \frac{\left(\theta + p_{mat}\left(1 - p^{E,E} - p^{E,I}\right)\right) p^{E,I}}{1 - p^{E,E}} \leqslant 1,$$

which is equivalent to $\sum_{k=2}^{\infty} \Psi^I_k \leqslant 1$.

5.1.4 Origin of the model

It is shown in [JacPe10] that the age-dependent process (4.1.1) can be obtained as the limit of a more complex age and population-dependent process $\mathbf{N}_n := \left(N_n^k\right)_{k\in\mathcal{T}}$, describing the population-size at time n for each type $k \in \mathcal{T}$ (corresponding *e.g.* to health stages, locations, ages *etc.*).

We first very briefly recall this result (which is due to C. Jacob and thus will not be exposed in this thesis), and then describe in which way this has been applied in order to obtain the epidemic model (5.1.1).

The number of k individuals at time n is given by

$$N_n^k = \sum_{l=1}^{a_M} \sum_{h\in\mathcal{T}} \sum_{i=1}^{N_{n-l}^h} Y_{n-l,n,i}^{(h),k}, \tag{5.1.14}$$

where a_M is the largest survival age, and $Y_{n-l,n,i}^{(h),k}$ is the number of k individuals generated at time n by individual i belonging to the type h at time $n-l$. This number $Y_{n-l,n,i}^{(h),k}$ of "mathematical" offsprings depends on the individual transition of i from h to k, as well as on the number of its "true" offsprings and their respective transition from their initial type to k.

Assuming the existence of a subset $\mathcal{K} \subset \mathcal{T}$ of *rare types*, and eventually of a subset $\mathcal{K}' \subset \mathcal{K}$ corresponding to rare types "of interest", it is shown in [JacPe10] (Proposition 5.1) that, under technical assumptions (which are rather weak in the epidemiological context), the process N_n^k converges in distribution, for all $k \in \mathcal{K}$, as the initial population size $N_0 := \sum_{k\in\mathcal{T}} N_0^k$ tends to infinity. Moreover, the process summed on the rare types of interest, $X_n := \sum_{k\in\mathcal{K}'} X_n^k$, where $X_n^k \overset{\mathcal{D}}{:=} \lim_{N_0} N_n^k$, is a process of the form (4.1.1).

In the epidemiological context of a $SEIR$ disease with horizontal and vertical infection routes, where the types are the health states, if $\mathcal{K}$ concerns all the infected states and $\mathcal{K}' \subset \mathcal{K}$, the clinical state, then we obtain that the process

$$X_n := \lim_{N_0\to\infty} \sum_{k\in\mathcal{K}'} N_n^k,$$

corresponding to the incidence of the clinical cases at time n, assuming that the initial population size N_0 is very large, is of the form (5.1.1)-(5.1.2). We illustrate in Figure 5.6 in which way a clinical case alive at time n (animal A in the figure) can actually generate secondary cases (animals B and C in the figure) several years later, through the processes of infection, incubation and slaughtering. This also very roughly justifies the presence in $\Psi_k(\theta_0)$ of the infection terms θ_0 and $p_{mat.}$, as well as of the survival and incubation terms S_k and $P_{inc.}(k)$.

5.2 Prediction of the disease spread

The recursivity of the epidemic model (5.1.1) enables long-term predictions of the process. For this purpose we need to estimate the parameters of the model, namely the incubation period distribution $P_{inc.}(.)$, the maternal infection parameter $p_{mat.}$, the survival distribution in the apparently healthy state $(S_k)_k$, and the infection parameter θ_0 corresponding to the remaining horizontal infection route.

The survival distribution is derived from [Don97] (see Table 5.2). The incubation period distribution, that can be estimated only on the whole epidemic series (growth and decay) because of a very bad identifiability of these parameters with the infection parameters on a sole monotonous phase, is estimated by the Bayesian maximum a posteriori (MAP) estimations realized on the whole epidemic, under the modeling of this distribution by a Weibull distribution with parameters α, β (see Subsection 5.1.1). We set

$$\alpha = \widehat{\alpha}^{MAP} = 3.84$$

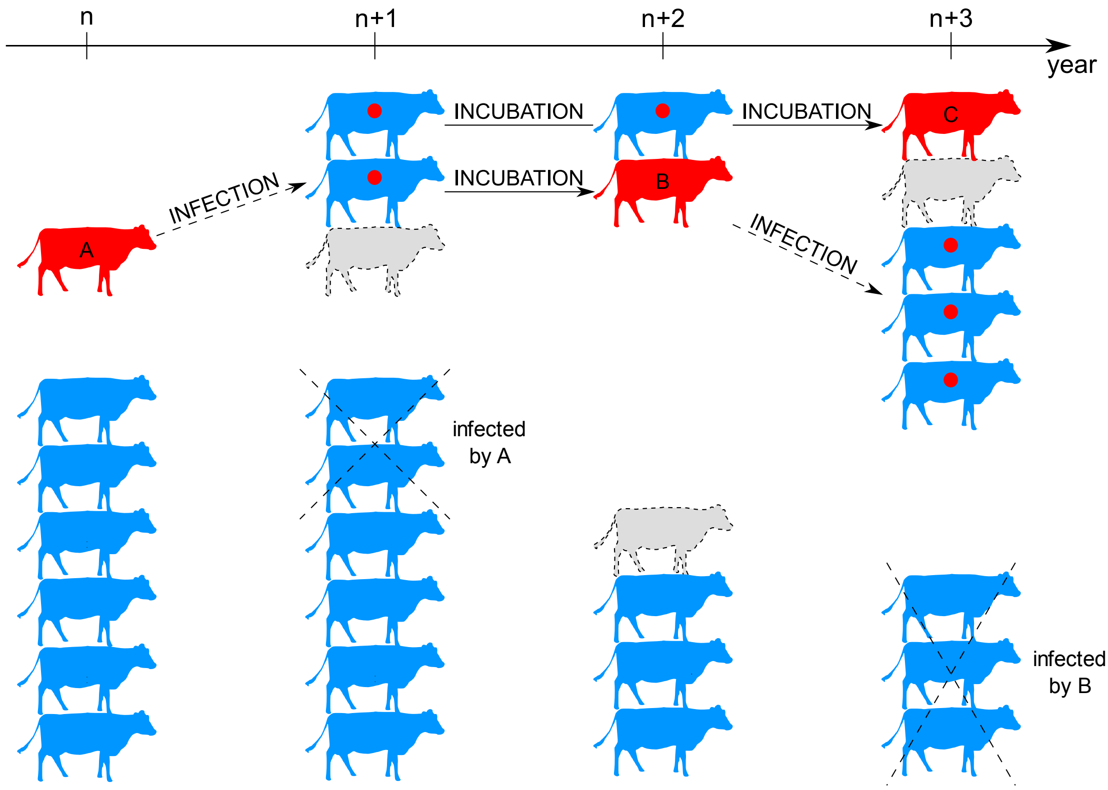

Figure 5.6: Model taking account all the health status of the disease.

and

$$\beta = \widehat{\beta}^{MAP} = 7.46,$$

where $\widehat{\alpha}^{MAP}$ and $\widehat{\beta}^{MAP}$ are the MAP Bayesian estimations obtained in [Jac10A]. Finally, we set $p_{mat.} = 0.1$, which is the largest commonly admitted value for this parameter ([Brad96, Don97]). Concerning θ_0, we could also use the MAP estimator

$$\widehat{\theta}^{MAP} = 2.43 \tag{5.2.1}$$

got on the whole epidemic until 2007 ([Jac10A]), but since it is the parameter of interest, we reestimate this parameter using here a conditional least squares estimation approach, which does not require an arbitrary prior distribution relative of this parameter as in the Bayesian setting.

The CLSE and a confidence interval of θ_0 is calculated in **Subsection 5.2.1** in the setting of the model (5.1.1), where

$$\mathbf{X}_0 = \mathbf{X}^{obs}_{1997} = (X^{obs}_{1997}, \ldots, X^{obs}_{1989}),$$

given in Table 5.1. We study moreover the sensitivity of this estimation to the values of $(p_{mat.}, \alpha, \beta)$.

In **Subsection 5.2.2**, we predict the future disease evolution from 2009. For the predictions from 2009, we set $\mathbf{X}_0 = \mathbf{X}^{obs}_{2008}$.

All the following simulations and computations have been done with the numerical computing and programming environment MATLAB.

5.2.1 Estimation of the infection parameter

Estimation

In order to estimate the infection parameter θ_0, we use in model (5.1.1)-(5.1.2) the weighted CLSE $\widehat{\theta}^{X}_{|\mathbf{X}_0|}$ introduced in Subsection 4.2.1, where $\{p_{mat.}, \alpha, \beta, \{S_j\}\}$ correspond to the values given in

S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}
0.97	0.65	0.36	0.30	0.25	0.18	0.10	0.06	0.02	0.01

Table 5.2: Observed survival probabilities of cattle in Great Britain (deduced from [Don97]). S_k is the probability for an (apparently) healthy animal to survive at least until k years.

k	1	2	3	4	5	6	7	8	9
$a_k.10^3$	0.2192	1.9315	5.5275	9.2323	10.4353	8.3260	5.1357	1.8642	0.5569
$b_k.10^4$	0.0738	0.5432	1.8025	3.7227	5.0766	4.3821	3.4238	1.2428	0.5569

Table 5.3: Values of a_k and b_k defined by (5.1.2)-(5.1.3).

the introduction of Section 5.2. The corresponding numerical values of a_k and b_k defined by (5.1.2)-(5.1.3) are given in Table 5.3.

According to Table 5.1,

$$|\mathbf{X}_0| = |\mathbf{X}_{1997}^{obs}| = 167977.$$

We are thus close to the asymptotic $|\mathbf{X}_0| \to \infty$. The estimator (4.2.12) provides the following estimation,

$$\widehat{\theta}_{|\mathbf{X}_0|}^{X} = 2.4486. \tag{5.2.2}$$

We point out that this estimation is very close to the maximum *a posteriori* Bayesian estimation (5.2.1), got on the whole epidemic until 2007.

Using the asymptotic normality of $\widehat{\theta}_{|\mathbf{X}_0|}^{X}$ given by (4.2.18), we obtain the following confidence interval with asymptotic probability 0.95:

$$[\widehat{\theta}_{min}, \widehat{\theta}_{max}],$$

where

$$\widehat{\theta}_{min} := \widehat{\theta}_{|\mathbf{X}_0|}^{X} - \frac{1.96}{\widehat{c}_1}, \quad \text{and} \quad \widehat{\theta}_{max} := \widehat{\theta}_{|\mathbf{X}_0|}^{X} + \frac{1.96}{\widehat{c}_1}.$$

The value

$$\widehat{c}_1 := \frac{\sum_{k=1}^{n} \boldsymbol{a} \cdot \mathbf{X}_{k-1}}{\sigma^2\left(\widehat{\theta}_{|\mathbf{X}_0|}^{X}\right)}$$

is computed assuming (see (4.2.16))

$$\alpha_i = \frac{X_{1997,i}^{obs}}{|\mathbf{X}_{1997}^{obs}|}.$$

Therefore $\widehat{c}_1 = 40.3938$, and

$$\begin{aligned} &\mathbb{P}\Big(\theta_0 \in \Big[\widehat{\theta}_{min}, \widehat{\theta}_{max}\Big]\Big) \simeq 0.95, \\ &\widehat{\theta}_{min} = 2.4000, \quad \widehat{\theta}_{max} = 2.4971. \end{aligned} \tag{5.2.3}$$

Let us mention that a credibility interval of θ_0 of probability 0.95 calculated from the quantiles at 0.025 and 0.975 of the posterior distribution of θ in the Bayesian setting was $[2.231, 2.728]$ [Jac10A], which is much larger than the confidence interval given here. Although this confidence interval is an asymptotic one, as $|\mathbf{X}_0| \to \infty$, it is a very good approximation of the true confidence interval since $|\mathbf{X}_0|$ is here very large.

Sensitivity analysis

Let us evaluate the sensitivity of the estimator to the other parameters of the model, namely the maternal infection parameter $p_{mat.}$ and the incubation parameters α and β. To this end, we

$p_{mat.}$	α	β	$\widehat{\theta}^X_{\lvert\mathbf{X}_0\rvert}$	$[\widehat{\theta}_{min}, \widehat{\theta}_{max}]$
0.1	3.84	7.46	2.4486	$[2.4000, 2.4971]$
0	3.84	7.46	2.5022	$[2.4537, 2.5507]$
1	3.84	7.46	1.9658	$[1.9172, 2.0143]$
0.1	2	7.46	2.7976	$[2.7425, 2.8528]$
0.1	20	7.46	4.0664	$[3.9862, 4.1465]$
0.1	3.84	1	1.0143	$[0.9940, 1.0345]$
0.1	3.84	10	6.2579	$[6.1355, 6.3803]$
0.1	3	6	1.5487	$[1.5177, 1.5797]$
0.1	4	5	1.0277	$[1.0069, 1.0486]$

Table 5.4: Estimation of the environmental infection parameter θ_0, and its confidence interval $[\widehat{\theta}_{min}, \widehat{\theta}_{max}]$ with asymptotic probability 0.95, for different values of the maternal infection parameter $p_{mat.}$ and of the incubation parameters (α, β). The values $(3.84, 7.46)$ correspond to the Bayesian MAP estimations $(\widehat{\alpha}^{MAP}, \widehat{\beta}^{MAP})$, and 0.1 to the largest commonly admitted value for $p_{mat.}$.

compute the estimation of θ_0 and the associated confidence interval with asymptotic probability 0.95, for different values of $(p_{mat.}, \alpha, \beta)$, listed in Table 5.4.

The first line of the table corresponds to the parameters chosen for the model. In each of the four following lines, we fix two coordinates and choose an extremal (unrealistic) value for the third one. It appears that the estimation of θ_0 does not strongly depend on the value of the maternal infection parameter: we obtain estimations of the same order, even when assuming that no maternal infection occur, or that every newborn with an infectious mother is infected. However, the estimation differs a bit more when we change the parameters of the incubation period distribution. We obtain for example an estimation of the order of 4 if the shape parameter α of the Weibull distribution equals 20, and around 6 if the mode β equals 10. Since these cases correspond to unrealistic incubation period distributions (see Figure 5.7), we also compute the estimation of θ_0 for two other more realistic sets of (α, β), namely $(\alpha, \beta) = (3, 6)$ and $(\alpha, \beta) = (4, 5)$. The results are given in the last two lines of Table 5.4, and we obtain estimations of the order of 1 to 2.

Conclusion

Even for very unrealistic values (α, β), all the estimations of θ_0 remain in the same order of magnitude of several units, which is really small compared to the estimations obtained in [Jac10A] for the infection *via* Meat and Bone Meal or lactoreplacers (before 1989), which are of the order of 1000. However, although these estimations are all very small, θ_0 seems nonnull. This could suggest the existence of a minor but nonnull infection source which is not of maternal type.

5.2.2 Prediction of the incidences of cases and infected cattle

In this subsection, we predict the spread of the disease from 2009 by means of simulations of the epidemic process, where θ_0 is replaced by its previous estimation $\widehat{\theta}^X_{\lvert\mathbf{X}_0\rvert} = 2.4486$, and where the initial time of the model is 2008, that is

$$\mathbf{X}_0 = \mathbf{X}^{obs}_{2008} = (X^{obs}_{2008}, \dots, X^{obs}_{2000}).$$

The simulations are done recursively using the transition law (4.1.2).

The expected value of cases in 2009 given the past is, by definition,

$$\mathbb{E}_{\widehat{\theta}^X_{\lvert\mathbf{X}_0\rvert}} \left(X_{2009} \lvert \mathbf{X}^{obs}_{2008} \right) = \mathbf{\Psi} \left(\widehat{\theta}^X_{\lvert\mathbf{X}_0\rvert} \right) \cdot \mathbf{X}^{obs}_{2008} = 46.46,$$

which is surprisingly higher than the number of reported cases in 2008 ($X^{obs}_{2008} = 33$), and larger than the 9 reported cases in 2009 from January to September (see [OIE]). The simulations initiated

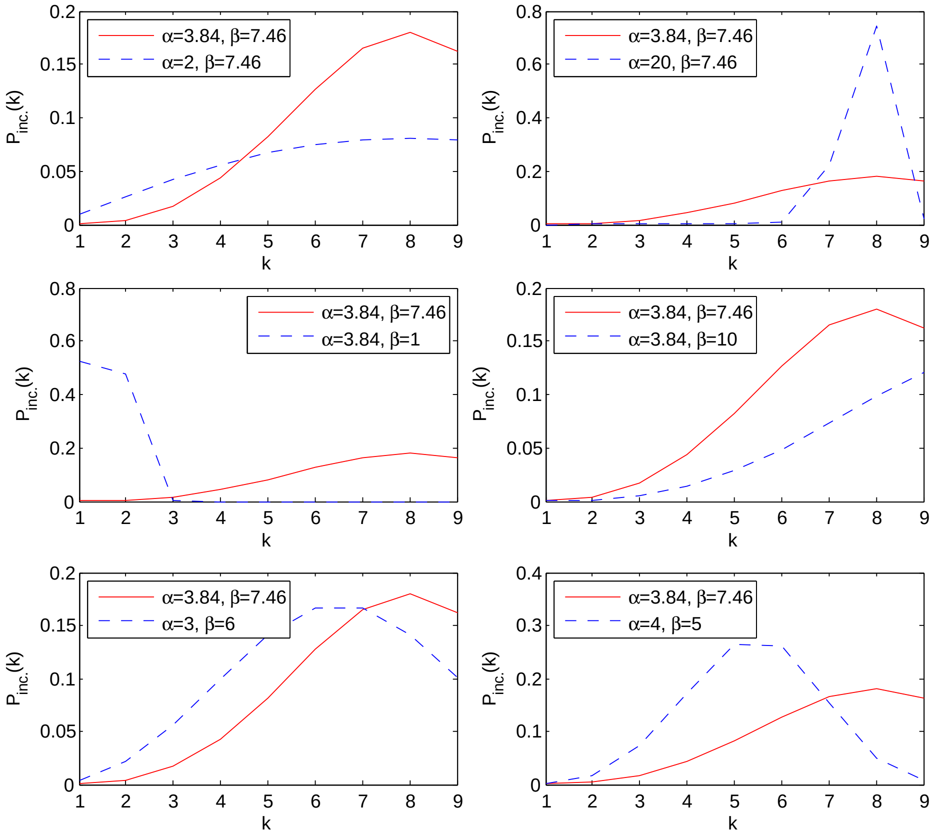

Figure 5.7: Incubation period distribution for different values of the shape parameter α and of the mode β of the Weibull distribution. In red, the distribution computed with $(\alpha, \beta) = (\widehat{\alpha}^{MAP}, \widehat{\alpha}^{MAP}) = (3.84, 7.46)$ (values chosen for our model). For each figure we use an adapted scale in order to compare the two distributions.

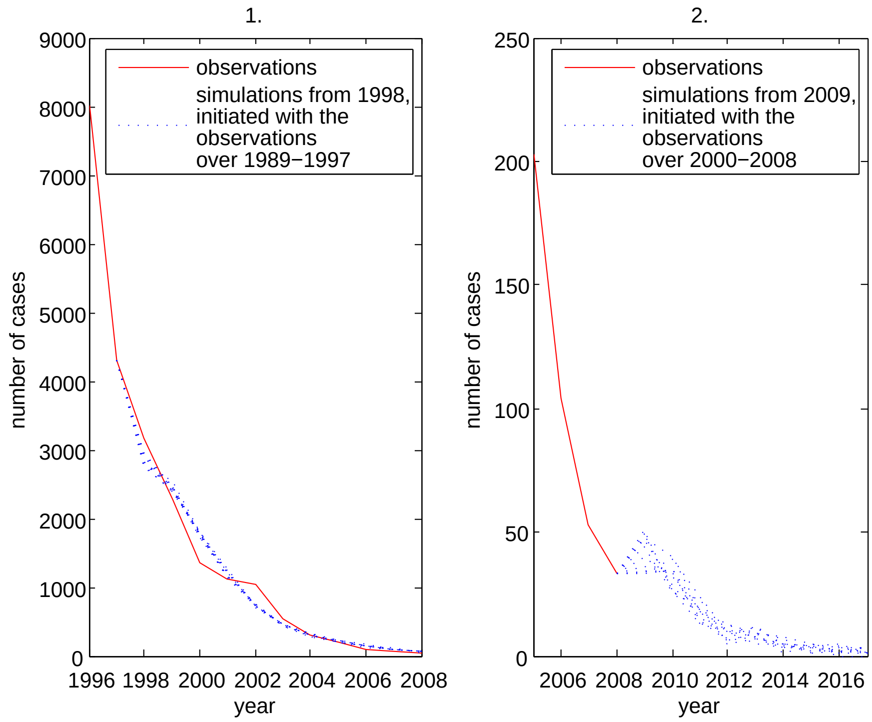

Figure 5.8: On Figure 5.8.1, 10 simulations of the epidemic process with the infection parameter $\widehat{\theta}^X_{|\mathbf{X}_0|} = 2.4486$ on the period 1998-2008, and comparison with the observations; on Figure 5.8.2, 10 simulations of the same epidemic process, on the period 2009-2017.

by $\mathbf{X}_0 = \mathbf{X}^{obs}_{2008}$ present indeed a "step" between 2008 and 2009, as illustrated in Figure 5.8.2. The model thus does not completely fit the reality at this point, which can be explained by the choice of a two-parameter incubation period distribution, the use of several assumptions to obtain a simplified model, and an over or under-declaration of the number of cases during the past 20 years. Nevertheless, the height of this step is extremely small compared to the values of the process, which are in majority of the order of 10000 to 1000 since 1997, and we point out that the model initiated by $\mathbf{X}_0 = \mathbf{X}^{obs}_{1997}$ provides quite realistic simulations on the period 1998-2008 compared to the real observations on the same period, as illustrated in Figure 5.8.1. The epidemic model thus seems to provide a satisfying prediction of the *overall* evolution of the real epidemic. We must however be cautious when interpreting predictions at a given year.

Incidences of cases

We simulate 1000 trajectories of the epidemic model $(\mathbf{X}_n)_{n\geqslant 0}$ initiated by the observed values $\mathbf{X}^{obs}_{2008}$. We illustrate in Figure 5.9.1, for each year from 2009, the maximum, minimum, median, 0.025 and 0.975 quantiles associated with the 1000 realizations.

Incidences of infected cattle

As mentioned in the introduction, it is also crucial to study and predict the evolution of the incidence of infected cattle in the population. These could not only infect other animals later on, when they are in the infectious phase of their incubation period, but also, since they have no clinical symptoms, they could enter the human food process if they do not undergo a diagnostic test. The incidence E_n of infected cattle at time n, conditionally on the number X_n of cases at

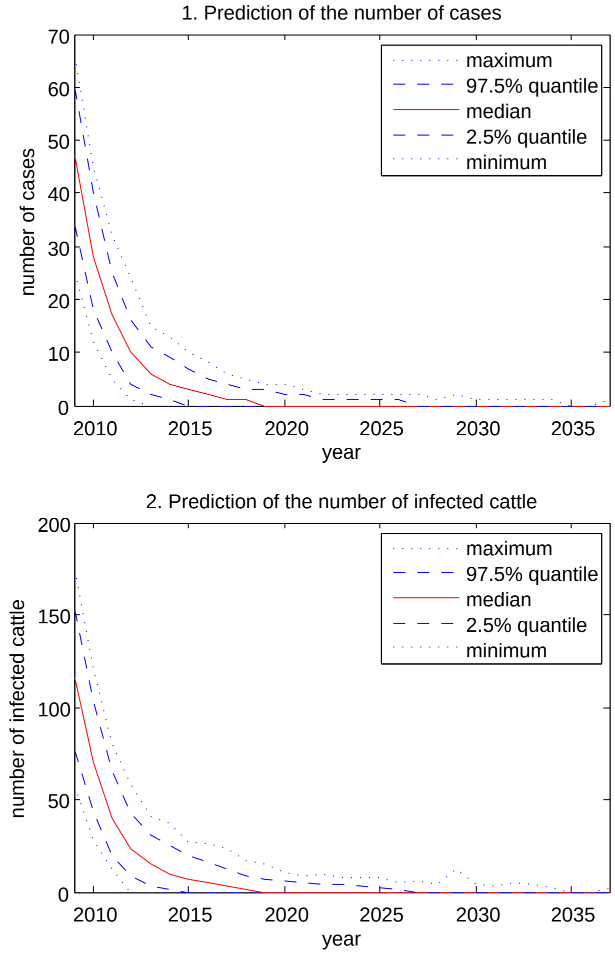

Figure 5.9: Prediction, based on 1000 simulations of the epidemic process with the infection parameter $\widehat{\theta}^{X}_{|\mathbf{X}_0|} = 2.4486$, of the yearly incidences of cases (1) and of infected cattle (2) from 2009. 95% of the trajectories remain in the band delimited by the blue dashed lines.

that time, is given by the following Poisson distribution (see [JacPe10]):

$$E_n|X_n \sim \mathcal{P}oisson\left(\left(\theta_0 + p_{mat.}\frac{S_1}{\sum_{j=1}^{d+1} S_j}\right) X_n\right), \, n \geqslant 0. \tag{5.2.4}$$

For every $n \geqslant 0$ and for each of the 1000 simulated values X_n, we make one realization of E_n, according to (5.2.4). We then illustrate in Figure 5.9.2, the yearly maximum, minimum, median, 0.025 and 0.975 quantiles associated with the 1000 realizations. As expected, the incidence of infected cattle is much larger than the incidence of cases.

5.2.3 Prediction of the extinction time

We know thanks to (4.2.3) that the process is subcritical if and only if $\theta_0 < \theta_{crit.} \simeq 23$. The epidemic process observed here is thus obviously subcritical, and will die out almost surely in a finite time.

Let

$$T := 2008 + \inf\{n \geqslant 1, \mathbf{X}_n = \mathbf{0}\}$$

n	$\mathbb{P}_{\widehat{\theta}_{max}}(T \leqslant n)$	$\mathbb{P}_{\widehat{\theta}_{min}}(T \leqslant n)$	n	$\mathbb{P}_{\widehat{\theta}_{max}}(T \leqslant n)$	$\mathbb{P}_{\widehat{\theta}_{min}}(T \leqslant n)$
2009	0.0000	0.0000	**2036**	**0.9560**	**0.9622**
...	...	...	2037	0.9703	0.9747
2023	0.0002	0.0004	2038	0.9800	0.9831
2024	0.0032	0.0049	2039	0.9866	0.9888
2025	0.0226	0.0307	2040	0.9910	0.9925
2026	0.0837	0.1037	2041	0.9940	0.9950
2027	0.1941	0.2260	2042	0.9960	0.9967
2028	0.3319	0.3702	2043	0.9973	0.9978
2029	0.4737	0.5123	2044	0.9982	0.9985
2030	**0.6040**	**0.6386**	2045	0.9988	0.9990
2031	0.7137	0.7421	2046	0.9992	0.9993
2032	0.7990	0.8211	2047	0.9995	0.9996
2033	0.8613	0.8778	2048	0.9996	0.9997
2034	0.9050	0.9171	2049	0.9998	0.9998
2035	0.9352	0.9439	2050	0.9998	0.9999

Table 5.5: Cumulative distribution function of the year of extinction for the infection parameters $\widehat{\theta}_{min}$ and $\widehat{\theta}_{max}$ defined by (5.2.3). The year 2030 (resp. 2036) corresponds to the smallest n such that $\mathbb{P}_{\widehat{\theta}_{max}}(T \leqslant n) \geqslant 0.5$ (resp. $\mathbb{P}_{\widehat{\theta}_{max}}(T \leqslant n) \geqslant 0.95$).

denote the extinction year of the epidemic model. By (5.1.5), for every $n \geqslant 1$,

$$\mathbb{P}_{\theta_0}(T \leqslant 2008 + n) = \left(f_{n,1}^{\theta_0}(\mathbf{0})\right)^{X_{2008}^{obs}} \cdots \left(f_{n,d}^{\theta_0}(\mathbf{0})\right)^{X_{2000}^{obs}},$$

which by iterating the generating function $\boldsymbol{f}^{\theta_0}$ can be computed explicitly.

For every $i = 1 \dots d$, $n \in \mathbb{N}$ and $\mathbf{r} \in [0,1]^d$, $\theta \mapsto f_{n,i}^{\theta}(\mathbf{r})$ is a decreasing function, hence $\theta \mapsto \mathbb{P}_\theta(T \leqslant n)$ is decreasing as well, and we obtain, thanks to (5.2.3), that for $n \geqslant 2009$,

$$\mathbb{P}\left(\mathbb{P}_{\theta_0}(T \leqslant n) \in \left[\mathbb{P}_{\widehat{\theta}_{max}}(T \leqslant n), \mathbb{P}_{\widehat{\theta}_{min}}(T \leqslant n)\right]\right) \simeq 0.95. \tag{5.2.5}$$

We collect in Table 5.5 the values of the confidence interval $\left[\mathbb{P}_{\widehat{\theta}_{max}}(T \leqslant n), \mathbb{P}_{\widehat{\theta}_{min}}(T \leqslant n)\right]$ for each $n \geqslant 2009$, and obtain in particular that

$$\mathbb{P}\Big(\mathbb{P}_{\theta_0}(T \leqslant 2030) \geqslant 0.5\Big) \simeq 0.95,$$
$$\mathbb{P}\Big(\mathbb{P}_{\theta_0}(T \leqslant 2036) \geqslant 0.95\Big) \simeq 0.95.$$

We illustrate in Figure 5.10 the cumulative distribution function of the extinction time for the parameters $\widehat{\theta}_{min}$ and $\widehat{\theta}_{max}$, corresponding to the confidence band, as well as for two extremal values of the infection parameter ($\theta = 1$ and $\theta = 6$). We see in Figure 5.10 that the confidence band $[\mathbb{P}_{\widehat{\theta}_{max}}(T \leqslant n), \mathbb{P}_{\widehat{\theta}_{min}}(T \leqslant n)]$ is very narrow, leading to an accurate estimation of $\mathbb{P}_{\theta_0}(T \leqslant n)$.

5.2.4 Prediction of the total size of the epidemic

Let

$$N := \sum_{n=1}^{T-2008} X_n$$

be the total size of the future epidemic from 2009 (total number of cases from 2009 until the extinction of the epidemic).

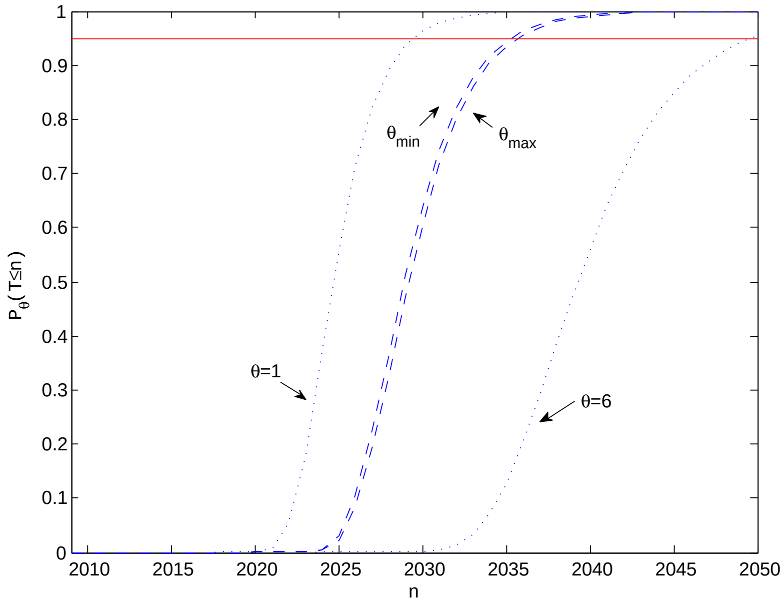

Figure 5.10: In dashed lines, cumulative distribution function of the year of extinction for the infection parameters $\widehat{\theta}_{min}$ and $\widehat{\theta}_{max}$ defined by (5.2.3). In dotted lines, cumulative distribution function of the year of extinction for two extremal values of the infection parameter ($\theta = 1$ and $\theta = 6$).

We compute the distribution of N using (5.1.7), (5.1.8), conditionally on the event $\{\mathbf{X}_0 = \mathbf{X}_{2008}^{obs}\}$.

We first compute the first two moments of N, using (5.1.9) and (5.1.10),

$$\begin{aligned} \mathbb{E}_{\widehat{\theta}_{min}}(N) &= 119.3170, \quad \mathbb{V}ar_{\widehat{\theta}_{min}}(N) = 149.2324, \\ \mathbb{E}_{\widehat{\theta}_{max}}(N) &= 124.6133, \quad \mathbb{V}ar_{\widehat{\theta}_{max}}(N) = 157.3303. \end{aligned} \tag{5.2.6}$$

Let us point out that, according to (5.1.9) and (5.1.10), $\mathbb{E}_{\theta_0}(N)$ and $\mathbb{V}ar_{\theta_0}(N)$ are increasing functions of θ_0.

Then defining $\sum_{j=1}^{0} \cdot = 0$ and $\prod_{j=1}^{0} \cdot = 1$, we can explicitly compute, for every $n \in \mathbb{N}$,

$$\begin{aligned} \mathbb{P}_{\theta_0}(N = n) = \sum_{\substack{\{0 \leqslant y_{k,i} \leqslant n, \{1 \leqslant n_{k,i,j} \leqslant n\}_j\}_{i,k}: \\ \sum_{k=1}^{d} \sum_{i=1}^{X_{-k+1}} \sum_{j=1}^{y_{k,i}} n_{k,i,j} = n}} \prod_{k=1}^{d} \prod_{i=1}^{X_{-k+1}} e^{-\sum_{l=k}^{d} \Psi_l(\theta_0)} \frac{\left(\sum_{l=k}^{d} \Psi_l(\theta_0)\right)^{y_{k,i}}}{y_{k,i}!} \\ \times \prod_{j=1}^{y_{k,i}} e^{-n_{k,i,j} \sum_{l=1}^{d} \Psi_l(\theta_0)} \frac{\left(n_{k,i,j} \sum_{l=1}^{d} \Psi_l(\theta_0)\right)^{n_{k,i,j}-1}}{n_{k,i,j}!}, \end{aligned} \tag{5.2.7}$$

and obtain thanks to (5.2.3), for every $n \in \mathbb{N}$,

$$\mathbb{P}\left(\mathbb{P}_{\theta_0}(N \leqslant n) \in \left[\mathbb{P}_{\widehat{\theta}_{max}}(N \leqslant n), \mathbb{P}_{\widehat{\theta}_{min}}(N \leqslant n)\right]\right) \simeq 0.95. \tag{5.2.8}$$

Using (5.2.7), the confidence interval (5.2.8) should be theoretically computable for any value of n, even large ones, and according to (5.2.6), n of the order of hundred is quite likely. However, because of the numerous partitions and combinatoric terms involved in (5.2.7), and because the values $X_{2000}^{obs}, \ldots, X_{2008}^{obs}$ are very large, formula (5.2.7) is not easily computable for large values

n	$\mathbb{P}_{\widehat{\theta}_{max}}(N \leqslant n)$	$\mathbb{P}_{\widehat{\theta}_{min}}(N \leqslant n)$	n	$\mathbb{P}_{\widehat{\theta}_{max}}(N \leqslant n)$	$\mathbb{P}_{\widehat{\theta}_{min}}(N \leqslant n)$
80	0.0000	0.0000	130	0.6830	0.8130
90	0.0030	0.0070	140	0.9090	0.9440
100	0.0240	0.0600	141	0.9220	0.9520
110	0.1170	0.2510	142	0.9310	0.9600
120	0.3800	0.5180	143	0.9420	0.9660
121	0.4100	0.5550	144	0.9490	0.9710
122	0.4430	0.5840	**145**	**0.9570**	**0.9740**
123	0.4710	0.6140	146	0.9660	0.9810
124	**0.5000**	**0.6510**	150	0.9820	0.9930
125	0.5320	0.6850	160	0.9980	0.9990
126	0.5610	0.7140	170	1.0000	1.0000

Table 5.6: Empirical cumulative distribution function (based on 1000 simulated trajectories) of the total size of the epidemic for the infection parameters $\widehat{\theta}_{min}$ and $\widehat{\theta}_{max}$ defined by (5.2.3). The value 124 (resp. 145) corresponds to the smallest n such that $\mathbb{P}_{\widehat{\theta}_{max}}(N \leqslant n) \geqslant 0.5$ (resp. $\mathbb{P}_{\widehat{\theta}_{max}}(N \leqslant n) \geqslant 0.95$).

of n. Consequently, we rather choose to compute the empirical cumulative distribution functions $\mathbb{P}_{\widehat{\theta}_{min}}(N \leqslant n)$ and $\mathbb{P}_{\widehat{\theta}_{max}}(N \leqslant n)$, based on 1000 simulated trajectories of the epidemic model from 2009 (with memory over the years 2000-2008).

The numerical values are reported in Table 5.6. We obtain in particular that

$$\mathbb{P}\Big(\mathbb{P}_{\theta_0}(N \leqslant 124) \geqslant 0.5\Big) \simeq 0.95,$$
$$\mathbb{P}\Big(\mathbb{P}_{\theta_0}(N \leqslant 145) \geqslant 0.95\Big) \simeq 0.95.$$

We illustrate in Figure 5.11 the empirical cumulative distribution function of the extinction time for the parameters $\widehat{\theta}_{min}$ and $\widehat{\theta}_{max}$ corresponding to the confidence band, as well as for two extremal values of the infection parameter ($\theta = 1$ and $\theta = 6$).

5.3 Prediction of the disease spread in case of a very late extinction

In this section we focus on the *worst-case scenario*, which means a very late extinction of the epidemic. In order to predict the evolution of the epidemic in such a case, we use the Q-process associated with the branching process, introduced in Subsection 4.3.1. We first estimate in **Subsection 5.3.1** the infection parameter θ_0 with the estimator $\widehat{\theta}_n^{X^*}$ built for the Q-process and defined by (4.3.8), and then predict in **Subsection 5.3.2** the spread of the disease by means of simulations.

5.3.1 Estimation of the infection parameter

Let us compute the value of $\widehat{\theta}_n^{X^*}$ based on the data in Great Britain presented in Table 5.1. Unfortunately, since the observations stop in 2008, the number of available observations is only $n = 11$. We are thus far from the asymptotic setting $n \to \infty$ of the previous section. However, the large value of $|\mathbf{X}_0|$ can make us hope a good accuracy. We point out that, by making use of the estimator $\widehat{\theta}_n^{X^*}$ on the real data, we make an unverifiable assumption on the future of the epidemic: we consider the data as if they were the beginning of a trajectory with very late extinction. This should have the following consequence: the estimation provided by $\widehat{\theta}_n^{X^*}$ should *a priori* be a bit

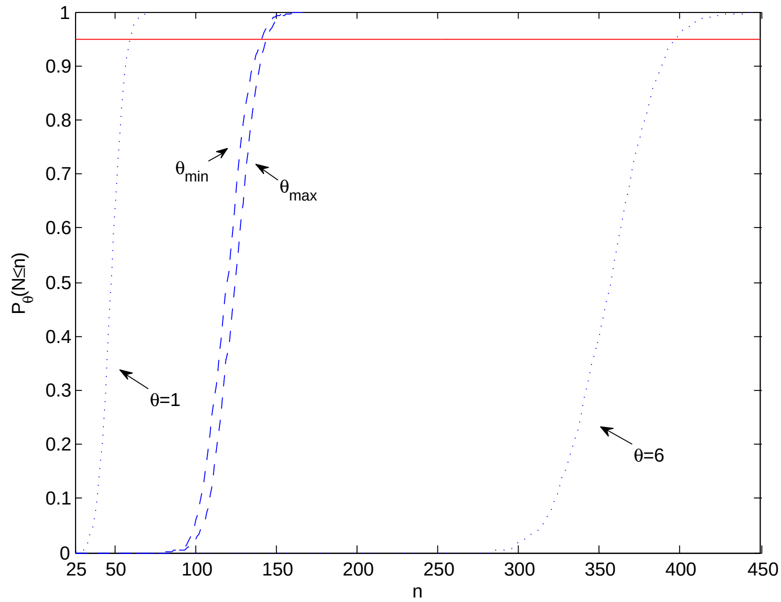

Figure 5.11: In dashed lines, the empirical cumulative distribution functions (based on 1000 simulated trajectories) of the total size of the epidemic for the infection parameters $\widehat{\theta}_{min}$ and $\widehat{\theta}_{max}$ defined by (5.2.3). In dotted lines, the empirical cumulative distribution functions of the total size of the epidemic for two extremal values of the infection parameter ($\theta = 1$ and $\theta = 6$).

smaller than the value 2.4486 provided by $\widehat{\theta}^{X}_{|\mathbf{X}_0|}$. Indeed we obtain:

$$\widehat{\theta}^{X^*}_n = 2.4472. \tag{5.3.1}$$

Using this value, we obtain the following rate of convergence in (4.3.36) to the asymptotic normal distribution,

$$\widehat{c}_2 := \frac{\sum_{k=0}^{n}\left(f^{*'}\left(\widehat{\theta}^{X^*}_n, \mathbf{X}^*_k\right)\right)^2}{\sqrt{\sum_{k=0}^{n}\left(f^{*'}\left(\widehat{\theta}^{X^*}_n, \mathbf{X}^*_k\right)\right)^2 g\left(\widehat{\theta}^{*}_n, \mathbf{X}^*_k\right)}} = 40.4967. \tag{5.3.2}$$

We deduce the confidence interval of θ_0 with asymptotic probability 0.95, $[\widehat{\theta}^*_{min}, \widehat{\theta}^*_{max}]$, where

$$\widehat{\theta}^*_{min} := \widehat{\theta}^{X^*}_n - \frac{1.96}{\widehat{c}_2}, \quad \text{and} \quad \widehat{\theta}^*_{max} := \widehat{\theta}^{X^*}_n + \frac{1.96}{\widehat{c}_2}.$$

We thus obtain

$$\begin{gathered}\mathbb{P}\Big(\theta_0 \in \left[\widehat{\theta}^*_{min}, \widehat{\theta}^*_{max}\right]\Big) \simeq 0.95, \\ \widehat{\theta}^*_{min} = 2.3988, \quad \widehat{\theta}^*_{max} = 2.4956,\end{gathered} \tag{5.3.3}$$

which is of the same magnitude order as the confidence interval $[2.4000, 2.4971]$ obtained with the unconditioned process.

5.3.2 Prediction of the disease spread

In order to predict the behavior of the "most dangerous" evolution of the epidemic, we can now use the transition law of the conditioned process given by Proposition 4.3.1.

First, we see thanks to Figure 5.12.1 that the simulations provided by the conditioned process initiated by $\mathbf{X}_0 = \mathbf{X}^{obs}_{1997}$, and where θ_0 is estimated by $\widehat{\theta}^{X^*}_n = 2.4472$, are quite realistic, compared

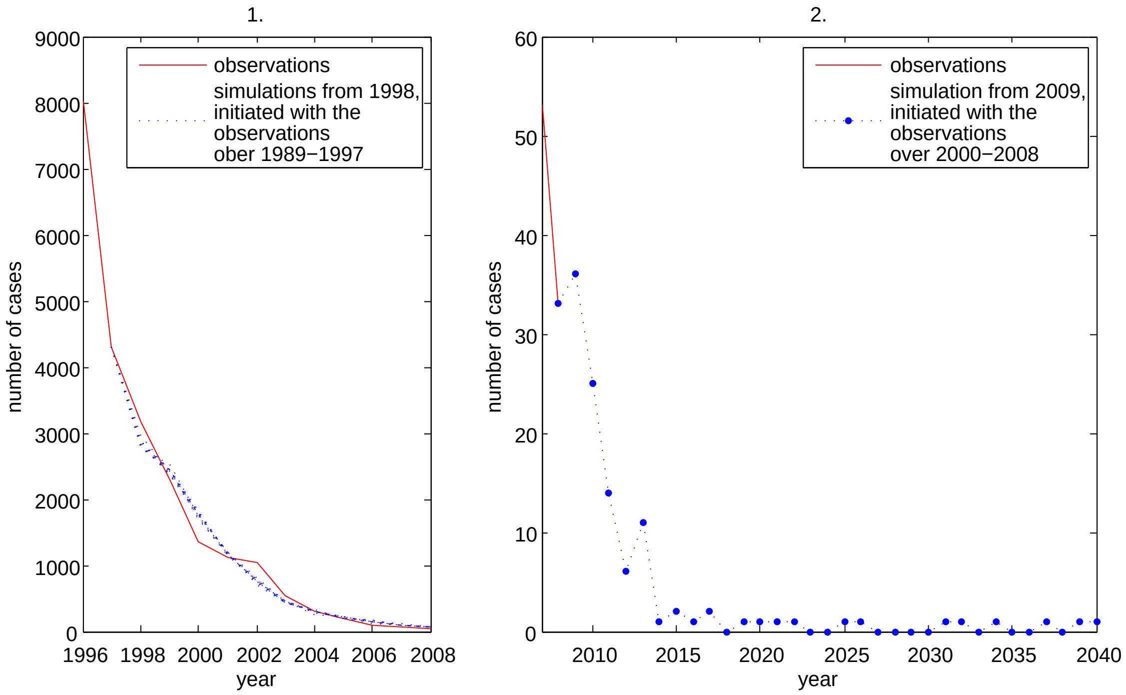

Figure 5.12: On Figure 5.12.1, 10 simulations of the conditioned process with the infection parameter $\widehat{\theta}_n^{X^*} = 2.4472$ from 1998 to 2008, and comparison with the observations; on Figure 5.12.2, one simulation of the conditioned process with the infection parameter $\widehat{\theta}_n^{X^*} = 2.4472$ from 2009 to 2040.

to the real observations on the period 1998-2008. Figure 5.12.2 is an example of one simulation on the period 2009-2040 of the conditioned process, for $\mathbf{X}_0^* = \mathbf{X}_{2008}^{obs}$. It appears that the values of this simulated trajectory are rapidly very small, and of course never equals $\mathbf{0}$.

For a finer prediction, we simulate 1000 realizations of the Q-process from 2009 until 2050, with $\mathbf{X}_0^* = \mathbf{X}_{2008}^{obs}$ and $\theta_0 = \widehat{\theta}_n^{X^*}$. Moreover, for every $k \geqslant 0$ and for each of the 1000 simulated values X_k^*, we make one realization of the incidence E_k of infected cattle at time k, according to the law given by (5.2.4). Figure 5.13 represents the yearly maximum, minimum, median, 0.25 and 0.975 quantiles associated with the 1000 realizations of respectively, the incidence of cases and the incidence of infected cattle, in case of a very late extinction.

Conclusion

It appears thanks to this last study that the supposedly "most dangerous" trajectories, corresponding to a very late extinction of the epidemic, nevertheless do not reach high values and do not present a new peak of epidemic.

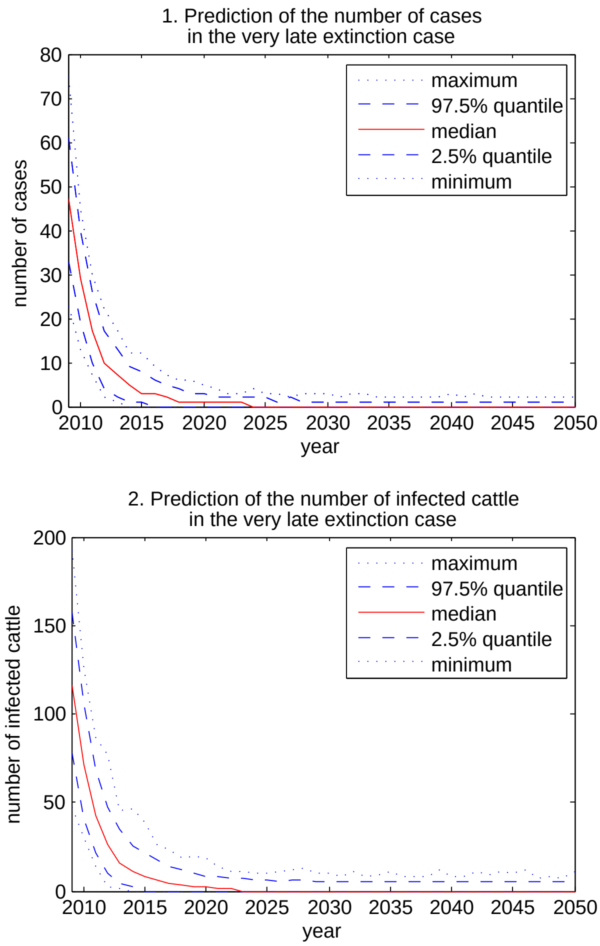

Figure 5.13: Prediction, based on 1000 simulations of the conditioned process $(\mathbf{X}^*_n)_{n\geqslant 0}$, with the parameter $\widehat{\theta}_n^{X^*} = 2.4472$, of the yearly incidences of cases (1) and of infected cattle (2) from 2009. 95% of the trajectories remain in the band delimited by the blue dashed lines.

Publications list

1. Pénisson, S. (2010). Continuous-time multitype branching processes conditioned on very late extinction. *Accepted for publication in ESAIM Probability and Statistics.*

2. Jacob, C. and Pénisson S. (2010). A general branching process with age, memory and population dependence. *Submitted for publication to Advances in Applied Probability*

3. Pénisson, S. (2010). Estimation of the infection parameter in the different phases of an epidemic modeled by a branching process. *Submitted for publication to Advances in Applied Probability*

4. Pénisson S. and Jacob, C. (2010). BSE epidemic in Great-Britain: prediction of the disease spread and study of the very late extinction case scenario, based on a stochastic branching model. *Submitted for publication to Risk Analysis.*

Bibliography

[AsmKei78] ASMUSSEN, S. and KEIDING, N. (1978). Martingale central limit theorems and asymptotic estimations theory for multitype branching processes. *Adv. Appl. Prob.* **10**, 109–129.

[Ath68] ATHREYA, K. B. (1968). Some results on multitype continuous time Markov branching processes. *Ann. Math. Statist.* **39**, 347–357.

[AthKar71] ATHREYA, K. B. and KARLIN, S. (1971). On branching processes with random environments, I and II. *Ann. Math. Statist.* **42**, 1499–1520 and 1843-1858.

[AthNey72] ATHREYA, K. B. and NEY, P. E. (1972). *Branching processes.* Springer-Verlag, New-York.

[AthNey73] ATHREYA, K. B. and NEY, P. E. (1973). *Limit theorems for the means of branching random walks.* Transactions of the Sixth Prague Conference on Information Theory, Statistical Decision Functions, Random Processes. 63–72. Academia, Prague.

[Bar57] BARTLETT, M. S. (1957). On theoretical models for competitive and predatory biological systems. *Biometrika.* **44**, 27–42.

[Bar60] BARTLETT, M. S. (1960). *Stochastic Population Models.* Methuen, London.

[Beck77] BECKER, N. (1977). Estimation for discrete time branching processes with application to epidemics. *Biometrics.* **33**, 515–522.

[Bil61] BILLINGSLEY, P. (1961). Statistical inference for Markov processes. *Statistical Research Monographs, Vol. II. The University of Chicago Press, Chicago, Ill.*

[Bin76] BINGHAM, N. H. (1976). Continuous branching processes and spectral positivity. *Stochastic Process. Appl.* **4**, 17–242.

[Blum63] BLUM, J. R., HANSON, D. L. and ROSENBLATT, J. I. (1963). On the central limit theorem for the sum of a random number of independent random variables. *Prob. Theory Rel. Fields.* **1**, 389–393.

[Brad96] BRADBURY, J. (1996). Maternal transmission of BSE demonstrated in cattle. *Lancet* **9024**, 393.

[Cat09] CATTIAUX, P., COLLET, P., LAMBERT, A., MARTÍNEZ, S., MÉLÉARD, S. and SAN MARTÍN, J. (2009). Quasi-stationary distributions and diffusion models in population dynamics. *Ann. Probab.* **37**, 1926–1969.

[ChaRoe08] CHAMPAGNAT, N. and RŒLLY, S. (2008). Limit theorems for conditioned multitype Dawson-Watanabe processes and Feller diffusions. *Electron. J. Probab.* **13**, 777–810.

[Con67] CONNER, H. (1967). A note on limit theorems for Markov branching processes. *Proc. Amer. Math. Soc.* **18** , 76–86.

[CruMod68] Crump, K. S. and Mode, C. J. (1968-1969). A general age-dependent branching process, I and II. *J. Math. Anal. Appl.* **24**, 494–508 and **25**, 8–17.

[DalJof08] Dallaporta, S. and Joffe, A. (2008). The Q-process in a multitype branching process. *Int. J. Pure Appl. Math.* **42**, 235–240.

[Dion72] Dion, J. P. (1972). Estimation des probabilités initiales et de la moyenne d'un processus de Galton-Watson. Thèse de doctorat, Université de Montréal.

[DioYa97] Dion, J. P. and Yanev, N. M. (1997). Limit theorems and estimation theory for branching processes with an increasing random number of ancestors. *J. Appl. Probab.* **34**, 309–327.

[Don97] Donnelly, C. A., Ferguson, N. M., Ghani, A. C., Woolhouse, M. E. J., Watt, C. J. and Anderson, R. M. (1997). The epidemiology of BSE in cattle herds in Great Britain. I. Epidemiological processes, demography of cattle and approaches to control by culling. *Phil. Trans. R. Soc. Lond. B.* **352**, 781–804.

[DuLeGa02] Duquesne, T. and Le Gall, J. F. (2002). Random trees, Lévy processes and spatial branching processes. *Astérisque.* **281**.

[Ev93] Evans, S. N. (1993). Two representations of a conditioned superprocess. In *Proceedings of the Royal Society of Edinburgh, Section A.* **123**, 959–971.

[EvPer90] Evans, S. N. and Perkins, E. (1990). Measure-valued Markov branching processes conditioned on non-extinction. *Israel J. Math.* **71**, 329–337.

[EthKur86] Ethier, S. N. and Kurtz, T. G (1986). *Markov processes: characterization and convergence.* Wiley.

[Fel51] Feller, W. (1951). Diffusion processes in genetics. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, 1950.* University of California Press, Berkeley and Los Angeles, 227–246.

[Fer95] Ferrari, P. A., Kesten, H., Martinez, S. and Picco, P. (1995). Existence of quasi-stationary distributions. A renewal dynamical approach. *Ann. Probab.* **23**, 501–521.

[FlePre74] Fleischmann, K. and Prehn, U. (1974). Ein Grenzwertsatz für subkritische Verzweigungsprozesse mit endlich vielen Typen von Teilchen. *Math. Nachr.* **64**, 357–362.

[FleiSieg77] Fleischmann, K. and Siegmund-Schultze, R. (1977). The structure of reduced critical Galton-Watson processes. *Math. Nachr.* **79** 233–241.

[Frob1908] Frobenius, G. (1908–1909). Über Matrizen aus positiven Elementen. *Sitzungsberichte der königlich preussischen Akademie der Wissenschaften.*

[Gal1873] Galton, F. (1873). Problem 4001. *Educational Times* **1**, 17.

[GeoBa03] Georgii, H. O. and Baake, E. (2003). Supercritical multitype branching processes: the ancestral types of typical individuals. *Adv. in Appl. Probab.* **35**, 1090–1110.

[Gin82] Ginzburg, L. R., Slobodkin, L. B., Johnson, K. and Bindman, A. G. (1982). Quasiextinction probabilities as a measure of impact on population growth. *Risk Anal.* **2**, 171–181.

[Gos01] Gosselin, F. (2001). Asymptotic behavior of absorbing Markov chains conditional on nonabsorption for applications in conservation biology. *Ann. Appl. Probab.* **11**, 261–284.

[Gut95] Guttorp, P. (1995). Three papers on the history of branching processes. *Int. Stat. Rev.* **63**, 234–345.

[Hac05] Haccou, P., Jagers, P. and Vatutin, V. A. (2005). *Branching processes: variation, growth and extinction of populations.* Cambridge Studies in Adaptive Dynamics, Cambridge.

[Har48] Harris, T. E. (1948). Branching processes. *Ann. Math. Statistics.* **19**, 474–494.

[Har51] Harris, T. E. (1951). Some mathematical models for branching processes. *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, 1950*, 305–328. University of California Press, Berkeley and Los Angeles.

[Har63] Harris, T. E. (1963). *The theory of branching processes.* Springer-Verlag.

[Hea65] Heathcote, C. R. (1965). A branching process allowing immigration. *J. Roy. Statist. Soc. Ser. B.* **27**, 138–143.

[Jac10A] Jacob, C., Maillard-Teyssier, L., Denis, J. B. and Bidot, C. (2010). A branching process approach for the propagation of the Bovine Spongiform Encephalopathy in Great-Britain. In *Branching processes and their Applications*, Lecture Notes in Statistics, Springer-Verlag, 227–242.

[Jac10B] Jacob, C. (2010). Conditional least squares estimation in nonstationary nonlinear stochastic regression models. *Ann. Statist.* **38**, 566–597.

[JacPe10] Jacob, C. and Pénisson, S. (2010). A general multitype branching process with age, memory and population dependence. Submitted to *Adv. Appl. Prob.*

[Jag75] Jagers, P. (1975). *Branching processes with biological applications.* Wiley Series in Probability and Mathematical Statistics - Applied Probability and Statistics. Wiley-Interscience, London-New York-Sydney.

[Jag09] Jagers, P. (2009). Some notes on the history of branching processes, from my perspective. *Lecture at the Oberwolfach Symposium on Random Trees*, 18–24 January 2009. `http://www.math.chalmers.se/~jagers/Branching%20History.pdf`

[Jag07] Jagers, P., Klebaner, F. C. and Sagitov, S. (2007). On the path to extinction. *PNAS* **104**, 6107–6111.

[JagLag08] Jagers, P. and Lagerås, A. N. (2008). General branching processes conditioned on extinction are still branching processes. *Electron. Commun. Probab.* **13**, 540–547.

[Jir58] Jiřina, M. (1958). Stochastic branching processes with continuous state space. *Czech. Math. J.* **8**, 292–312.

[Jir64] Jiřina, M. (1964). Branching processes with measure-valued states. *Trans. Third Prague Conf. Information Theory, Statist. Decision Functions, Random Processes (Liblice, 1962).* 333–357. Publ. House Czech. Acad. Sci., Prague.

[Jir69] Jiřina, M. (1969). On Feller's branching diffusion processes. *Časopis Pěst. Mat.* **94**, 84–90.

[Jof67] Joffe, A. (1967). On the Galton-Watson branching process with mean less than one.*Ann. Math. Statist.* **38**, 264–266.

[JofMet86] JOFFE, A. and MÉTIVIER, M. (1986). Weak convergence of sequences of semimartingales with applications to multitype branching processes. *Adv. in Appl. Probab.* **18**, 20–65.

[JofSpit67] JOFFE, A. and SPITZER, F. (1967). On multitype branching processes with $\rho \leqslant 1$. *J. Math. Anal. Appl.* **19**, 409–430.

[KaWat71] KAWAZU, K. and WATANABE, S. (1971). Branching processes with immigration and related limit theorems. *Theory Probab. Appl.* **16**, 34–51.

[KeiLau78] KEIDING, N. and LAURITZEN, S. L. (1978). Marginal maximum likelihood estimates and estimation of the offsprings mean in a branching process. *Scand. J. Statist.* **5**, 106–110.

[Ken66] KENDALL, D. G. (1966). Branching processes since 1873. *J. London Math. Soc* **41**, 385–486.

[Kolm38] KOLMOGOROV, A. N. (1938). Zur Lösung einer biologischen Aufgabe. *Comm. Math. Mech. Chebyshev Univ. Tomsk 2.* **1**, 1–12.

[KolDmi47] KOLMOGOROV, A. N. and DMITRIEV, N. A. (1947). Branching stochastic processes. *C. R. (Doklady) Acad. Sci. URSS (N.S.)* **56**, 5–8.

[Lamb07] LAMBERT, A. (2007). Quasi-stationary distributions and the continuous-state branching process conditioned to be never extinct. *Electron. J. Probab.* **12**, 420–446.

[Lamp67] LAMPERTI, J. (1967). The limit of a sequence of branching processes. *Z. Wahrsch. Verw. Gebiete* **7**, 271–288.

[LaNey68] LAMPERTI, J. and NEY, P. (1968). Conditioned branching process and their limiting diffusions. *Theory Probab. Appl.* **13**, 126–137.

[LeGaJan98] LE GALL, J. F. and LE JAN, Y. (1998). Branching processes in Lévy processes: the exploration process. *Ann. Probab.* **26**, 213–252.

[Li00] LI, Z. H. (2000). Asymptotic behaviour of continuous time and state branching processes. *J. Aust. Math. Soc. Serie A.* **68**, 68–84.

[Lin72] LINDVALL, T. (1972). Convergence of critical Galton-Watson branching processes. *J. Appl. Probab.* **9**, 445–450.

[Man61] MANDL, P. (1961). Spectral theory of semi-groups connected with diffusion processes and its application. *Czechoslovak Math. J.* **11**, 558–569.

[Mod71] MODE, C. (1971). *Multitype Branching Processes. Theory and Applications.* Modern Analytic and Computational Methods in Science and Mathematics, No. 34 American Elsevier Publishing Co., Inc., New York.

[Ogu75] OGURA, Y. (1975). Asymptotic behavior of multitype Galton-Watson processes. *J. Math. Kyoto Univ.* **15**, 251–302.

[OIE] WORLD ORGANISATION FOR ANIMAL HEALTH (OIE). `http://www.oie.int/eng/info/en_esbru.htm`.

[Pak03] PAKES, A. G. (2003). *Biological applications of branching processes.* Stochastic processes: modelling and simulation, 693–773, Handbook of Statist., 21, North-Holland, Amsterdam.

[Pak09] PAKES, A. G. (2009). Conditional limit theorems for continuous time and state branching processes. In *Records and Branching Processes*, Ahsanullah, M., Yanev, G.P., Nova Science Publishers, 63–103.

[Per1907] PERRON, O. (1907). Zur Theorie der Matrices. *Math. Ann.* **64**, 248–263.

[PraRao99] PRAKASA RAO, B. L. S. (1999). *Semimartingales and their statistical inference.* Monographs on Statistics and Applied Probability 83, Chapmann & Hall/CRC.

[QuiDur77] QUINE, M. P. and DURHAMN, P. (1977). Estimation for multitype branching processes. *J. Appl. Probab.* **14**, 829–835.

[Reb80] REBOLLEDO, R. (1980). Central limit theorems for local martingales. *Probab. Theory Relat. Fields* **51**, 269–286.

[RoeRou89] RŒLLY, S. and ROUAULT, A. (1989). Processus de Dawson-Watanabe conditionné par le futur lointain. *C. R. Acad. Sci. Paris Sér. I Math.* **309**, 867–872.

[RySko70] RYŽOV, J. M. and SKOROHOD, A. V. (1970) Homogeneous branching processes with a finite number of types and with continuously changing mass. *Teor. Verojatnost. i Primenen.* **15**, 722–726.

[Sen67] SENETA, E. (1967). The Galton-Watson process with mean one. *J. Appl. Probab.* **4**, 489–495.

[Sen73] SENETA, E. (1973). *Non-negative matrices - An introduction to theory and applications.* Halsted Press.

[SenVer66] SENETA, E. and VERE-JONES, D. (1966). On quasi-stationary distributions in discrete-time Markov chains with a denumerable infinity of states. *J. Appl. Probab.* **3**, 403–434.

[SenVer68] SENETA, E. and VERE-JONES, D. (1968). On the asymptotic behaviour of subcritical branching processes with continuous state space. *Z. Wahrsch. Verw. Gebiete.* **10**, 212–225.

[Sew51] SEWASTJANOW, B. A. (1951). The theory of branching random processes. *Uspehi Matem. Nauk (N.S.)* **46**, 47–99.

[Sew75] SEWASTJANOW, B. A. (1975). *Verzweigungsprozesse.* R. Oldenbourg Verlag.

[SmiWi69] SMITH, W. L. and WILKINSON, W. E. (1969). On branching processes in random environments. *Ann. Math. Statist.* **40**, 814–827.

[Twe75] TWEEDIE, R. L. (1975). Sufficient conditions for ergodicity and recurrence of Markov chains on a general state space. *Stochastic Process. Appl.* **3**, 385–403.

[Twe83] TWEEDIE, R. L. (1983). The existence of moments for stationary Markov chains. *J. Appl. Probab.* **20**, 191–196.

[Wel87] WELLS, G. A., SCOTT, A. C., JOHNSON, C. T., GUNNING, R. F., HANCOCK, R. D., JEFFREY, M., DAWSON, M. and BRADLEY, R. (1987). A novel progressive spongiform encephalopathy in cattle. *Vet Rec.* **121**, 419–420.

[Yag47] YAGLOM A. M. (1947). Certain limit theorems of the theory of branching random processes. *Doklady Akad. Nauk SSSR (N.S.).* **56**, 795–798.

[YaWat71] YAMADA, T. and WATANABE, S. (1971). On the uniqueness of solutions of stochastic differential equations. *J. Math. Kyoto Univ.* **11**, 155–167.