

# IMPERIAL

IMPERIAL COLLEGE LONDON

FACULTY OF NATURAL SCIENCES  
DEPARTMENT OF MATHEMATICS

---

Studies in Branching Processes  
*Mathematical Modelling of Structured Populations  
and Epidemics*

Tresnia Berah

Supervisor: Dr. Mikko Pakkanen,  
Co-Supervisor: Dr. Riccardo Passeggeri

---

Thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy and  
the Diploma of the Imperial College of Science, Technology and Medicine

ACADEMIC YEAR 2024-2025

## Statement of originality

I, Tresnia Berah, declare that the work contained in this thesis is my own, or else is appropriately referenced. Any material for which I am not responsible is explicitly indicated by the appropriate use of references and citations.

## Copyright declaration

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non Commercial 4.0 International Licence (CC BY-NC). Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose. When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes. Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

# Abstract

This thesis treats of various types of branching processes, with some applications to epidemiology. Chapters 1 and 2 introduce a novel time-varying Crump-Mode-Jagers (CMJ) branching process whose governing integral equations are then relied upon in order to provide a rigorous mathematical justification to the renewal incidence equation in infectious disease modelling. Chapter 3 focuses on the asymptotic behaviour of a structured branching population where each individual in the population is characterised by a trait or a type whose dynamics follow a Markov process. The branching process is then expected to be driven by a positive triplet of first eigenvalue problem of the first moment semigroup. A strong law of large numbers for super-critical branching Markov processes is proven assuming convergence of the renormalized semigroup in weighted total variation norm. Convergence is obtained under an  $L \log L$  condition which provides a new Kesten-Stigum result in infinite dimension and relaxes the uniform convergence assumption of the renormalized first moment semigroup required in the work of Asmussen and Hering in 1976. Finally Chapter 4 treats of a novel variational auto-encoder, called  $\pi$ -VAE, which has the property of being a proper stochastic process. As the combination of a generative model and a stochastic process that can be used as a prior,  $\pi$ -VAE is then used to do full Bayesian inference. We provide particular examples with geospatial or epidemiological data.

## Preface

This thesis brings together the work I have covered during my four years as a PhD student at the Centre for Doctoral Training in Modern Statistics and Machine Learning at Imperial College, London. The presentation of the different chapters of this thesis tends to prioritise linearity of content over chronological order. Thus, while the main thread along Chapters 1, 2 and 3 relates to studies in general branching processes, Chapter 4 presents an earlier article on variational autoencoder to which I contributed in the first six months of my doctoral training.

Chapters 1 and 2, forming in effect the first part of this thesis, primarily treat of a time-varying extension of a Crump-Mode-Jagers branching process (TV-CMJ) introduced originally in a publication by Pakkanen et al. [Pak+23] in the Journal of Mathematical Biology (JOMB) in August 2023. These chapters cover the first two years of my PhD under the direct supervision of Dr. Mikko Pakkanen at Imperial College. The original reason for introducing a time-varying Crump-Mode-Jagers process was to study the renewal-type integral equations derived from the dynamics of such a generalised branching process and to eventually relate them to the well-known incidence renewal equation used in epidemiology. As a way of introduction to the extended time-varying CMJ, let us recall that a classical assumption behind general branching processes, traditionally called Crump-Mode-Jagers (CMJ) branching processes as originally introduced by Crump and Mode [CM68; CM69] and Jagers [Jag69a], is based on the fact that the life-length distribution and the reproduction process do not depend directly on calendar time. Branching processes in random environments (see e.g. [AK71; Ban09]) and for instance generation dependent branching processes (see e.g. [Fil72], [Edl78]) have been introduced in the literature, but, to the extent of our current knowledge, a strict continuous dependence on calendar time for the behaviour of branching individuals has not yet fully been considered. Thus, here we introduce a CMJ branching process whose underlying probability measure in effect depends on calendar time.

The work on the time-varying extension of a CMJ branching process stemmed from a series of questions originally raised in January 2020 by Samir Bhatt, Professor of Statistics and Public Health at Imperial College and today also Professor of Machine Learning and Public Health at the University of Copenhagen. In fact, this time-varying extension of a CMJ branching process was originally proposed during the COVID pandemic in 2020-2021, at a time when the modelling of  $R_0$ , *the basic reproduction number*, had become crucially important not only for estimating and quantifying the spread of infection, but also in practice for implementing optimal preventive nonpharmaceutical solutions [Fla+20]. In the context of infectious disease modelling, the reproduction number  $R_0$  is classically assumed to remain piecewise constant over certain periods of time (as opposed to being continuously changing with time), but the latter represents a significant simplifying and limiting hypothesis, since in reality the reproduction number does vary continuously and its variations have significant effects on the evolution of an infectious disease. Therefore, having a mathematical framework whereby the reproduction number is modelled as being calendar-time dependent would lead to more realistic estimations and simulations.

Traditionally, the  $R_0$  reproduction number is introduced through the incidence renewal equation commonly used in epidemiology. This equation is most often written as follows (see

e.g. [Fra07; CDE18])

$$I(t) = R_0 \int_0^\infty I(t - \tau) \gamma(\tau) d\tau, \quad t > 0, \quad (1)$$

where  $I(t)$  represents the number of new infections at time  $t$  and  $\gamma(\tau)$  is the *intrinsic generation-interval distribution*, in other words the time distribution between the infection of an individual and the infection of others by that individual in the context of disease transmission. In practice, this generation-interval is intrinsic to the biology and to the transmission characteristics of the considered pathogen and is often modelled probabilistically as a distribution. Note that the generation-interval often needs to be differentiated from the *serial interval*, which corresponds to the time between the onset of symptoms. Equation (1) assumes that  $\gamma(\tau)$  describes the time-interval between successive infections in a transmission chain.

In mathematical modelling of infectious diseases, the intrinsic generation-interval is used as a probability distribution  $\gamma(\tau)$ , where  $\tau$  represents the time since an individual was infected. In that sense,  $\gamma(\tau)$  quantifies the likelihood that a secondary infection occurs at time  $\tau$ . Typically, an initial condition such as  $I(0) = 1$  is assumed for the first infected individual to ensure that the incidence process is well defined. In the literature on infectious disease modelling, and to our knowledge, this renewal equation appears either as *a priori* given, or *de facto* accepted. It is at best presented using various heuristic justifications. But in all cases, it does not seem to have been given a proper mathematical foundation.

Nonetheless, the incidence renewal equation is ubiquitous in epidemiology, as it contributes to estimating the effective reproduction number, forecasting epidemics, or at times evaluating interventions as well. Thus, and beyond other questions about the mathematical relationship between the renewal incidence equation in epidemiology and integral equations derived from age-dependent branching processes, Prof. Bhatt wanted to study more specifically how the form of the incidence renewal equation such as (1) would evolve if the reproduction number  $R_0$  became  $R_t$ , i.e. if it were assumed to evolve with calendar time. This questioning, combined with the observation that the common incidence equation had only been heuristically derived in the literature, led us to go back in 2020-21 to the classical derivations of first-order integral renewal-type equations for age-dependent branching processes made by Harris [Har63], Bellman and Harris [BH52] and subsequently by Crump Mode and Jagers [CM68; CM69; Jag69a; Jag75] for generalised branching processes.

The derivations of integral equations related to branching processes are most often based, on one hand, on the branching property, commonly called *first-generation principle*, originally defined by Harris [Har63], who states that a branching process is composed of the (independent) superposition of processes initiated by the first-generation progeny, and, on the other hand, on the representation of a branching process as a sum of random characteristics, as introduced by Crump, Mode, and Jagers (see e.g. [CM68; CM69; Jag75]). We mainly rely on these classical tools and techniques to construct *canonically* a TV-CMJ branching process in Chapter 1. We provide sufficient conditions for the TV-CMJ to not explode in finite time almost surely (in other words, to be *regular*). Next in Chapter 2, we study some of its first-order integral equations. A particular lens on infectious disease modelling is then adopted towards the end of Chapter 2, where we compare the renewal incidence equation commonly used in epidemiology with the integral equation derived for the expected birth rate in a TV-CMJ branching process. At this point, we link branching populations back to infectious disease modelling and provide mathematical grounding to the renewal incidence equation. In particular, we derive a more general form of the renewal incidence equation (1)

based on a reproduction number that is calendar-time dependent.

The second part of this thesis, composed of Chapter 3, while staying in the domain of branching processes, steers away from the TV-CMJ process, its dynamics and the renewal-type equations, to study the asymptotic behaviour of a structured super-critical branching Markov process. This object will be formally defined in Chapter 3. We provide sufficient conditions on the first-moment semigroup of the renormalised process to ensure convergence almost sure and in  $L^1$  of the empirical measure of the process. Chapter 3 is primarily based on an article co-authored with Prof. Vincent Bansaye and Dr. Bertrand Cloez\*. This third chapter in practice covers most of the work achieved in the last 20 months of my doctoral training between April 2023 and December 2024. During this period of time, I was mainly under the supervision of Prof. Bansaye at Ecole Polytechnique in Paris and Dr. Cloez at INRAe (MISTEA) in Montpellier. As an explanatory background for this change of lead supervision at this stage of my thesis, I should mention that since I wanted to continue working specifically on branching processes, I took the initiative to reach out to various experts in this field outside of Imperial College. This led me to working in the spring of 2022, in particular with Prof. Kimmel, Professor of Statistics at Rice University, on a couple of preliminary projects. Prof. Kimmel then adopted me informally into his reading group, which gave me the opportunity to present my work a few times to a targeted audience in a more formal manner. Finally, since June 2023, I have been working full time with Prof. Bansaye and Dr. Cloez on a project that Chapter 3 aims at summarising.

Going back to the content of Chapter 3, it treats of a branching population where each individual is characterised (structured) by a trait belonging to an abstract space. This trait follows Markovian dynamics. We focus on the supercritical recurrent case where the population may survive and grow and the trait distribution converges to a certain measure that happens to be the left eigenvector of the first-moment semigroup associated with the first eigenvalue problem. When the trait space is of finite dimension, the main result that describes the asymptotic behaviour of a multitype Markov branching process is classically due to Kesten and Stigum [KS66] who showed an  $L \log L$  condition on the reproduction process to be necessary and sufficient for a supercritical branching process to converge almost surely. A decade later, Asmussen and Hering [AH76] extended the latter finite-dimensional result to an infinite-dimensional (abstract) space and provided a set of conditions of the first-moment semigroup of the branching process to converge in the supercritical case. However, the latter set of conditions requires uniform convergence of the first-moment semigroup and is intrinsically linked to the knowledge of the leading eigentriple of the first-moment semigroup. Those two conditions are rarely verified, and even when they are, tend to be challenging to prove in practice.

With this context in mind, Chapter 3 presents a strong law of large numbers for the empirical measure associated with a structured branching process in an abstract space. In particular, we provide a set of convergence conditions for the first-moment semigroup that does not require uniformity and relies on knowledge of a Lyapunov-type function instead of the full leading eigentriple. Adding to this a slight adaptation of the classical  $L \log L$  condition for structured populations, it can be said that we provide a new Kesten-Stigum result [KS66] in infinite dimension and relax the uniform convergence assumption for the normalised first-moment semigroup required in the classical works of Asmussen and Hering in 1976 [AH76].

---

\*submitted on HAL and ArXiv in March 2025.

Finally, the third part of this thesis covers Chapter 4 and refers to an article by Mishra et al. [Mis+22], published in October 2022 in Statistics and Computing. The article, with a particular example related to geo-spatial data, presents a novel variational autoencoder (called  $\pi$ -VAE) whose objective is to learn low-dimensional embeddings of function classes. Between October 2020 and June 2021, I worked on the  $\pi$ -VAE model before focussing on branching processes, and the former deep learning model became the subject of my first *mini-project* as a first-year CDT student. My main contributions to the article consisted of building a prototype  $\pi$ -VAE in Python, clarifying the training procedure of the model, and in particular clearly distinguishing and rewriting the three different parts of the loss functions of the  $\pi$ -VAE. This work is presented in Chapter 4 with a brief introduction to variational inference and auto-encoders. I also mention towards the end of Chapter 4 two sister models to  $\pi$ -VAE, namely AggVAE [Sem+23a] and PriorCVAE [Sem+23b] which were then showcased in their respective articles with applications to infectious disease mapping.

Finally, as a CDT student in Modern Statistics and Machine Learning between Imperial College and Oxford University, I also had the great privilege to follow a few short modules, partly taught, partly research-based, which gave me the opportunity to learn and deepen my understanding (to cite but a few) about Bayesian modelling and computations, statistical machine learning, advanced Monte Carlo methods, or generative modelling. For each of these modules, we were required to present a research question based on specific pieces of literature related to the topic. This led me to regularly work in teams with other CDT students and to learn about various topics in modern statistics and machine learning.

---

## Acknowledgements

The time has finally come for me to write this page. I have sometimes thought about what I would write here, especially when I was losing confidence that I could ever actually write this page one day... The past four years have been very rewarding and have also naturally brought their challenges. I started on this journey in October 2020 with the clear intention of doing a doctorate in mathematics, but little did I know how much I would learn about many other things than mathematics along the way. I grew so much during this period and learnt invaluable lessons that I will carry with me wherever I go next and whatever I do next.

I have many people to thank for being in the position today to write these words. I may depart from common usage here by first thanking my five incredible children, Naomi, Tobias, Elyana, Talitha, and Shani. Today, all teens and for two of them very young adults, they give me immense strength and continuous joy. They have been unconditionally supportive, and it is unashamedly with incredible pride that I can say that I am the mother of five very unique individuals, for whom my deepest wish is that they love learning throughout their lives as much as I have always loved learning and growing. Their humour, their wit, their fresh and honest perspectives on life are an infinite source of inspiration to me. They are my models as much as I hope to have been theirs in some ways and more particularly during this adventure of the last four years.

I would like to thank Seth Flaxman and Mikko Pakkanen who vouched for me to embark on this journey, which, to be fair, from the outside, may not look like the obvious path a mother of five young children in her late forties would take. I also thank Prof. Gandy who played a decisive role in offering me a place in the StatML CDT. Mikko has since become my supervisor and I very warmly thank him for his support, his patience, and his very wise guidance. Among many other things, Mikko taught me to communicate more effectively and with greater perspective.

This whole journey into branching processes started thanks to Prof. Samir Bhatt without whom I may never have considered those mathematical objects. I am forever grateful to Sam for entrusting me with his preliminary questions, for involving me in his team, and for sharing his incredible intuition, which eventually led me to genuinely feel passionate about branching processes and their applications.

My life as a PhD student has not been linear though, and as I wished to continue working on branching processes, I took the initiative in 2022 to contact experts in the field outside Imperial College. This led me to work with Prof. Kimmel on a few projects and discuss various ideas. I am immensely grateful to Marek Kimmel for being so generous with his time. I remember emailing him for the first time very late at night one weekend, convinced that I would never get any reply. My inbox was proving differently the next morning with a very inspiring and encouraging reply! Since then, I took part in Prof. Kimmel's seminar on branching processes and have had many discussions about their related renewal-like integral equations.

However, it is not an understatement to say that my life as a PhD student, who was stubbornly wanting to focus on branching processes, turned out for the better the day Prof. Vincent Bansaye replied to my first email. I would almost like to say *the rest is history*, but this probably sounds a bit cliché. However, there is no doubt in my mind that I am able today to submit my thesis thanks to Vincent, his unlimited passion for anything related to branching processes and mathematics in general, and also his genuine kindness. Vincent has

---

been instrumental in guiding me with the literature at first, and then progressively helping me consider various project ideas related to structured branching populations. Since June 2023, I have had the immense privilege of working directly and closely with him and with Bertrand Cloez on what has now become the third chapter of this thesis. At the beginning of our collaboration, Bertrand kindly invited me a few times to work with him in person at INRAe (MISTEA) in Montpellier. We spent many hours in front of his blackboard. I am immensely grateful for his guidance over the past 18 months. On this note, I would like to thank the whole MISTEA team in Montpellier who welcomed me very warmly every time I was there. I also had the chance to go and work with Vincent at Ecole Polytechnique in Palaiseau. Going down memory lane, it felt surreal to me to see how much the campus had changed since my years as a *Polytechnicienne* there between 1992 and 1995. This PhD is in some ways long overdue as I was meant to start it in 2003 under Prof. Mark Davis, and I feel extremely grateful to Vincent for reintroducing me to my old school and showing me indirectly how much I owe my original education in France.

Back to the UK and Imperial College...I would like to thank the two current StatML CDT directors, Sarah Filippi and Ed Cohen for their incredible support in the past two years. Their responsiveness and constant guidance have been critical in my progress toward my goal. In particular, I will forever remember Sarah and be grateful to her for agreeing to meet me on Christmas Eve last year at the last minute despite her being ill.

There are a few people I have had the privilege of getting to know while roaming the corridors of the Huxley building. Heather Battey is one of them. Heather has also been very supportive and has given great advice at times. She helped me gain confidence in myself and become more independent in my research. Now, as the year 2021 was just starting with yet another lockdown due to Covid 19, and thus the prospect of homeschooling my five children for a few more months was dwelling on me, I had the incredible luck to meet Prof. Andrew Parry whom I now consider a close friend. Andrew gave me continued support. His unique personality and experience are without a doubt legendary, and I have been lucky enough to benefit from his advice and wisdom an almost uncountable number of times in the past three years.

Finally, I would also like to mention my co-office PhD students: Augusto whose daily presence was a source of comfort and laughter, Amber always so kind, Elliott who would bring his mountaineering kit on a regular basis, and Carlos who tried not to bang the door so much in the end..... Sadly, I never had lunch or coffee with any of them, as they could testify that caring for five children and doing a PhD in Maths at Imperial did not leave much free time to socialise. However, I have met incredible people during these formative four years. The other StatML students from Oxford and Imperial are among them. I will end this acknowledgement with a very special mention to Alex Larionov who not only has helped me grasp computational problems with more perspective, but who has also been kind enough to offer his time to become a role model to my five children. It has been written and said many times before, but I confirm: learning is living ...

## Contents

|          |                                                                                      |           |
|----------|--------------------------------------------------------------------------------------|-----------|
| <b>0</b> | <b>Introduction</b>                                                                  | <b>14</b> |
| 0.1      | Structure of the thesis . . . . .                                                    | 14        |
| 0.2      | Brief literature review on branching processes . . . . .                             | 17        |
| 0.3      | Synthetic overview of mathematical modelling in epidemiology . . . . .               | 20        |
| <b>1</b> | <b>Time-Varying Crump-Mode-Jagers branching process</b>                              | <b>23</b> |
| 1.1      | Introduction . . . . .                                                               | 23        |
| 1.2      | Probability space for a CMJ process . . . . .                                        | 24        |
| 1.2.1    | Ulam-Harris-Neveu notations and set of individuals . . . . .                         | 25        |
| 1.2.2    | Family histories . . . . .                                                           | 26        |
| 1.2.3    | Events space . . . . .                                                               | 26        |
| 1.2.4    | Product probability measure . . . . .                                                | 28        |
| 1.3      | Probability space for a time-varying CMJ . . . . .                                   | 28        |
| 1.3.1    | Family histories and events space . . . . .                                          | 29        |
| 1.3.2    | Recursive conditional probability measures . . . . .                                 | 30        |
| 1.4      | Counting by random characteristics . . . . .                                         | 32        |
| 1.4.1    | Formal definition of a time-varying CMJ branching process . . . . .                  | 32        |
| 1.4.2    | Examples of applications in epidemiology . . . . .                                   | 34        |
| 1.4.3    | Time-varying Bellman Harris branching process . . . . .                              | 35        |
| 1.4.4    | Time-varying inhomogeneous Poisson branching process . . . . .                       | 37        |
| 1.5      | Regularity conditions . . . . .                                                      | 38        |
| 1.5.1    | Classical regular CMJ branching process . . . . .                                    | 38        |
| 1.5.2    | Time-varying CMJ branching process . . . . .                                         | 40        |
| 1.5.3    | Time-varying Bellman-Harris branching process . . . . .                              | 43        |
| 1.5.4    | Time-varying Inhomogeneous Poisson branching process . . . . .                       | 44        |
| <b>2</b> | <b>Integral equations for a time-varying CMJ branching process</b>                   | <b>47</b> |
| 2.1      | First-moment integral equation of a TV-CMJ process . . . . .                         | 48        |
| 2.2      | A generic renewal-like integral equation . . . . .                                   | 51        |
| 2.3      | Integral equation for the expected birth rate . . . . .                              | 53        |
| 2.3.1    | From cumulative incidence to expected birth rate . . . . .                           | 53        |
| 2.3.2    | Absolutely continuous part of cumulative incidence . . . . .                         | 55        |
| 2.3.3    | Cumulative incidence described as a generalised function . . . . .                   | 57        |
| 2.3.4    | Expected birth rate as derivative of cumulative incidence . . . . .                  | 58        |
| 2.4      | Discretisation algorithm of integral equations . . . . .                             | 59        |
| 2.4.1    | Discretisation scheme . . . . .                                                      | 59        |
| 2.4.2    | Convergence of discretisation algorithm . . . . .                                    | 61        |
| 2.5      | The renewal incidence equation . . . . .                                             | 65        |
| 2.5.1    | Incidence in infectious disease modelling . . . . .                                  | 65        |
| 2.5.2    | Existence and uniqueness of a solution to a generalised incidence equation . . . . . | 67        |

|          |                                                                                                |            |
|----------|------------------------------------------------------------------------------------------------|------------|
| 2.5.3    | Extension and reformulation of the incidence renewal equation . . . . .                        | 69         |
| 2.5.4    | Equivalence between expected birth rate and incidence . . . . .                                | 71         |
| 2.6      | Backward-forward duality of branching processes . . . . .                                      | 74         |
| 2.6.1    | Brief literature context . . . . .                                                             | 74         |
| 2.6.2    | Forward equation for cumulative incidence . . . . .                                            | 75         |
| 2.6.3    | Forward and backward incidence equations . . . . .                                             | 78         |
| 2.7      | Consistency with backcalculation of incidence . . . . .                                        | 79         |
| 2.7.1    | Epidemiology context . . . . .                                                                 | 79         |
| 2.7.2    | Consistency . . . . .                                                                          | 80         |
| <b>3</b> | <b>On the strong law of large numbers for supercritical general branching processes</b>        | <b>83</b>  |
| 3.1      | Overview of structured branching Markov processes . . . . .                                    | 83         |
| 3.2      | Strong law of large of numbers in discrete time . . . . .                                      | 89         |
| 3.2.1    | Construction and definitions . . . . .                                                         | 89         |
| 3.2.2    | Assumptions and main theorem . . . . .                                                         | 92         |
| 3.2.3    | Contraction operator and decomposition of the process . . . . .                                | 94         |
| 3.2.4    | $L \log L$ moment propagation . . . . .                                                        | 97         |
| 3.2.5    | Martingale decomposition . . . . .                                                             | 102        |
| 3.2.6    | Proof of convergence in discrete time . . . . .                                                | 112        |
| 3.3      | Strong law of large numbers in continuous time . . . . .                                       | 115        |
| 3.3.1    | Assumptions and main theorem in continuous time . . . . .                                      | 115        |
| 3.3.2    | Proof of the main theorem in continuous time . . . . .                                         | 116        |
| 3.3.3    | Application to branching elliptic diffusion . . . . .                                          | 119        |
| 3.4      | Conclusion and future work . . . . .                                                           | 121        |
| <b>4</b> | <b>A stochastic process prior for Bayesian deep learning with applications in epidemiology</b> | <b>123</b> |
| 4.1      | Introduction . . . . .                                                                         | 123        |
| 4.2      | Rapid overview on Bayesian inference . . . . .                                                 | 124        |
| 4.2.1    | The Bayesian paradigm . . . . .                                                                | 124        |
| 4.2.2    | Challenges of Bayesian inference . . . . .                                                     | 125        |
| 4.2.3    | Variational inference approaches and latent variable models . . . . .                          | 126        |
| 4.3      | Variational autoencoders . . . . .                                                             | 127        |
| 4.3.1    | Autoencoder . . . . .                                                                          | 127        |
| 4.3.2    | Variational autoencoder . . . . .                                                              | 128        |
| 4.3.3    | VAEs in a Bayesian inference setting . . . . .                                                 | 129        |
| 4.4      | Encoding stochastic processes with $\pi$ VAE . . . . .                                         | 131        |
| 4.4.1    | A two-step approach . . . . .                                                                  | 131        |
| 4.4.2    | Architecture of $\pi$ VAE . . . . .                                                            | 132        |
| 4.4.3    | Simulation and inference with $\pi$ VAE . . . . .                                              | 135        |
| 4.4.4    | $\pi$ VAE is a stochastic process . . . . .                                                    | 136        |
| 4.4.5    | Applications to spatial Bayesian inference . . . . .                                           | 136        |
| 4.5      | Discussion and Conclusion . . . . .                                                            | 138        |
| <b>5</b> | <b>Conclusion</b>                                                                              | <b>139</b> |
|          | <b>Appendices</b>                                                                              | <b>141</b> |
| <b>A</b> | <b>Appendices to Chapter 1</b>                                                                 | <b>142</b> |
| A.1      | Miscellaneous measure theoretic results . . . . .                                              | 142        |

---

|          |                                                              |            |
|----------|--------------------------------------------------------------|------------|
| A.1.1    | Product sigma-field . . . . .                                | 142        |
| A.1.2    | Transition kernels . . . . .                                 | 142        |
| A.1.3    | Ionescu-Tulcea theorem . . . . .                             | 143        |
| A.2      | First-order stochastic dominance . . . . .                   | 143        |
| A.3      | Fubini and Tonelli Theorems . . . . .                        | 144        |
| A.4      | Unsymmetric Fubini Theorem . . . . .                         | 145        |
| A.5      | Auxiliary lemma to prove Propositions 1.1 and 2.3 . . . . .  | 145        |
| <b>B</b> | <b>Appendices to Chapter 2</b>                               | <b>147</b> |
| B.1      | Grönwall's inequalities . . . . .                            | 147        |
| B.1.1    | Grönwall's inequality for measurable functions . . . . .     | 147        |
| B.1.2    | Two particular forms . . . . .                               | 148        |
| B.2      | Proof of Proposition 2.1 . . . . .                           | 148        |
| B.2.1    | Existence of a solution . . . . .                            | 148        |
| B.2.2    | Uniqueness of the solution . . . . .                         | 153        |
| B.3      | Useful results on Schwartz distributions . . . . .           | 154        |
| B.4      | Auxiliary result to prove Theorem 2.1 . . . . .              | 156        |
| B.5      | Leibniz rule for differentiation under an integral . . . . . | 157        |
| <b>C</b> | <b>Appendices to Chapter 3</b>                               | <b>158</b> |
| C.1      | Perron-Frobenius theorem . . . . .                           | 158        |
| C.2      | Kesten-Stigum theorem . . . . .                              | 158        |
| C.3      | Asmussen-Hering Law of Large Numbers . . . . .               | 159        |
| C.4      | Extended convexity result . . . . .                          | 161        |
| C.5      | Useful bounds related to the $\log^* x$ function . . . . .   | 162        |
| C.6      | Auxiliary result for the martingale decomposition . . . . .  | 163        |

# CHAPTER 0

## Introduction

This introductory chapter presents the structure of the thesis and provides a high-level description of the main objectives, challenges and results related to each of the subjects covered. This chapter then continues with a brief historical and literature review of the major developments in the theory of branching processes, with a particular focus on the theoretical extensions most relevant to the content of Chapters 1 to 3. This brief introduction ends with a broad overview of the most common mathematical modelling choices in epidemiology with a reminder that branching processes happen to be a realistic model of a spread of infection when non-linear effects are negligible.

### 0.1 Structure of the thesis

Chapter 1 presents a time-varying extension of a generalised Crump-Mode-Jagers (CMJ) branching process [CM68; CM69; Jag75]. Whilst more details on the *classical* CMJ generalised branching process will be provided in Section 0.2, we note here that the time-varying extension introduced in Chapters 1 and 2 was proposed with the objective of being able to model more accurately (and in particular on a continuous basis) the variations with calendar time of the so-called  $R_0$  *reproduction number* commonly used in epidemiology. This reproduction number is generally assumed, at best, to be piece-wise constant and is interpreted as *the average number of secondary cases that would arise from a primary case infected at time  $t$  if conditions remained the same after time  $t$*  [Fra07].

Typically, the reproduction number is used to calculate various quantities of interest in the study of infectious diseases, such as the total number of infected cases present at a certain time (also called *prevalence* in epidemiology), the total number of infected cases up to a certain time (*cumulative incidence*), or the number of infections appearing at a point in time (*incidence*). Accurate knowledge of  $R_0$  at any time drastically improves inference and prevention capabilities. Recalling that epidemics can be described by branching processes when non-linear effects (such as e.g. susceptible depletion) are negligible, we extended the construction of the generalised CMJ branching process to introduce a continuous and flexible calendar-time reproduction process  $(R_t)_{t \geq 0}$  that would then provide more modelling flexibility. Compared to branching processes such as Galton-Watson (GW) or Bellman-Harris (BH) (whose main features will be described in Section 0.2), CMJ process was chosen as it relies on more realistic assumptions: in particular, individuals in the branching population (e.g. infection cases) can give birth at any time during their life, and the reproduction law may also depend on

the life-length distribution. This choice of modelling led us to introduce a novel time-varying CMJ (TV-CMJ) model whose probabilistic construction is presented in Chapter 1.

The canonical probabilistic construction of Chapter 1 emerged as we were trying to adapt the classical probabilistic constructions of a CMJ process presented in the literature. These later constructions are all based on event spaces  $\Omega$  whose single elements  $\omega$  (building blocks) are called *family histories* (originally introduced by Harris [Har63]). In short, to fully describe a branching population, family histories must include at the minimum basic sequences of events such as ages at childbearing (e.g.  $(\tau(k))_{k \in \mathbb{N}}$  for the parent's age when its  $k$ -th child is born) and lifelengths real-valued nonnegative random variables  $L$ . Family histories are in essence sequences of random variables and stochastic processes that allow reconstructing the entire tree of a branching population. In Chapter 1, we first review the canonical probabilistic construction, starting from a classical CMJ process, but decided to depart from the traditional construction. Indeed, in the ordinary CMJ, the reproduction processes are independent and identically distributed (i.i.d.). This latter property ensures that one can specify an instance of the reproduction process and assert that there exists a countably infinite collection of copies of it on a product space. In contrast, in the TV-CMJ, because of the assumed dependence on the time of birth, the reproduction processes are neither mutually independent, nor identically distributed, thus instead of relying on a Cartesian product of individual measures to construct the overall measure, we rely on a recursive construction of transition probability kernels indexed by times of birth of the parent. The construction of the probability measure of a TV-CMJ will be fully detailed in Section 1.3.2.

Continuing in Chapter 1, we then introduce the concept of regularity of a branching process: a branching process is said to be regular if it does not explode in finite time almost surely. This leads us to adapting the regularity conditions introduced originally by Crump, Mode, and Jagers ([CM68; CM69; Jag75]). Again, because of the dependence on the time of birth of the parent of all random quantities in the TV-CMJ process, the original regularity assumptions of the latter authors could not be directly applied in the case for our time-varying extension. Thus, our regularity proof, whilst still relying on the self-similarity property of the TV-CMJ, combines and adapts, on one hand, the classical proof of Crump, Mode and Jagers with the ones of Edler [Edl78] and Fildes [Fil72] who introduced generation-dependent CMJ processes. We end the chapter by providing two specific examples of TV-CMJ processes, namely, a time-varying Bellman-Harris (TV-BH) process and a time-varying inhomogeneous Poisson (TV-IP) process.

Chapter 2 continues to treat a TV-CMJ branching process and presents integral *renewal-type* equations that can be derived from its dynamics. The main objectives of introducing a TV-CMJ process were to be able to derive integral equations using a time-varying reproduction number and to ultimately provide mathematical grounding to the derivation of the well-known *incidence renewal equation* ubiquitously used in epidemiology. Thus in this chapter, we first carefully derive integral equations respectively for the number of individuals alive at a given time point (corresponding to *prevalence* in epidemiology), the total number of individuals that have been born up to a given time point (i.e. equivalent to *cumulative incidence* in an epidemiological context), and then eventually for the *expected birth-rate* of a TV-CMJ process. The latter quantity, in an epidemiological context, is directly equivalent to the number of new infections at a given time point, that is, *incidence*.

Next, to provide a rigorous mathematical proof of the incidence renewal equation, we go back to the classical incidence equation (see, e.g. [Fra07]) based on  $R_0$  and first extend it naturally to introduce a time-varying reproduction number  $R_t$ ,  $t \geq 0$ . One of the objectives

of Chapter 2 is to prove equivalence between the derived equation for the expected birth rate of a TV-CMJ process and the proposed extended incidence equation. However, considering the fact that a quantity such as cumulative incidence needs to be interpreted as a generalised function (as it is not fully differentiable), we first make a detour through the world of Schwartz distributions [Sch51] to prove the equivalence. A priori, these two equations have different forms and our equivalence proof relies on decomposing the considered quantities (i.e. on one hand the expected birth rate of a TV-CMJ and on the other hand, incidence of an infectious disease) between respectively their absolutely continuous parts and their singular parts. The singular parts being identical, we are directly left with comparing the absolutely continuous parts, whose equations we then discretise by relying on the algorithm originally introduced in [Pak+23, Section 3.1]. We prove that the discretised versions of the absolutely continuous parts converge toward each other, and in particular, our convergence proof provides an error rate of  $O(n^{-1})$ . Finally, we end this chapter with a mathematical proof of the commonly used convolution formula between prevalence and survival to infer incidence in epidemiology.

In Chapter 3, steering away from TV-CMJ branching processes, we move toward general branching populations structured by a Markovian trait. In particular, we present works that will form the basis of an article co-written with Prof. Vincent Bansaye and Dr. Bertrand Cloez (and which will be submitted on ArXiv at the latest in March 2025). An observation here on a change of perspective between Chapters 1 and 2 and Chapter 3 may be appropriate. Indeed, in Chapters 1 and 2, we formulate branching processes as sums of random characteristics to study the dynamics of the branching process and to focus on various quantities of interest that describe the branching population in any random phase of life. In contrast, and for reasons clarified further down, we adopt in Chapter 3 a different representation based on describing a branching process as a random-valued measure consisting of a sum of Dirac deltas (this will be defined formally in Section 3.2.1). Thus, in Chapter 3, we represent general branching processes as random counting measures, whereby individuals in the branching population are characterised by a Markovian trait belonging to an abstract (infinite-dimensional) space.

Both perspectives or representations have their advantages. Although different in nature, they are however equivalent and the choice of one over the other is linked to the objective of study. Note that Cohn and Jagers [CJ94] provide a clear description of the advantages of counting by random characteristics which can adopt various forms and thus lead to account for different quantities of interest linked to the whole population by only measuring independent contributions of individuals that have already been born and ignoring the unborn ones and their progeny's. And in fact, our main objective in Chapter 2 was to derive integral equations that describe the dynamics of the branching process. On the other hand, and as just mentioned in the above paragraph, the structured Markov branching process studied in Chapter 3 is written as a Dirac measure. In this later chapter, as we focus more particularly on the asymptotic behaviour of the branching model, the point of view of random-valued measures is more relevant: each individual is characterised by a trait that follows Markovian dynamics and lives in an abstract space.

The literature contains several well-known references, in particular Kesten and Stigum [KS66], when the trait space is finite dimensional, and Assmussen and Herring [AH76]) when the trait space is infinite dimensional. Classically, the convergence criteria for the empirical measure describing the branching process are based on an eigenvalue decomposition of the first-moment semigroup related to the renormalised branching process. Those criteria tend however to be challenging to verify in practice as they are related to the knowledge of eigenelements, themselves rarely accessible in real life. In contrast, Chapter 3 introduces a novel set of assumptions based on the knowledge of a certain Lyapunov function of the first-moment

semi-group. In this context, we extend the classical result of Asmussen and Hering [AH76] and we prove a Law of Large Numbers for this type of processes in the supercritical recurrent case, i.e. when the population may survive and grow, and the trait distribution converges. The branching process is then expected to be driven by the first eigenvalue problem of the first-moment semigroup. We prove a strong convergence result and nondegeneracy of the limit under a condition on the first-moment semigroup and an  $L \log L$  condition reminiscent of the Kesten-Stigum condition [KS66]. This relaxes the uniform convergence of the first moment required by Asmussen and Hering [AH76].

The techniques of proofs are interesting in themselves, using martingale families, contraction properties of the semigroup, and the truncation argument of [AH76]. In particular, we also prove  $L^1$  convergence and unify different results in the literature. It can then be said that results of Chapter 3 complement and extend current literature and provide in particular more accessible and verifiable convergence assumptions. Almost sure convergence is proven in discrete and continuous time, whilst  $L^1$  convergence is proven in discrete time (note that the proof of  $L^1$  convergence in continuous time will be finalised in the article to be submitted in March 2025).

Finally, the last chapter of this thesis, Chapter 4, based on an article by Mishra et al. [Mis+22], relates to an extension of a variational autoencoder (VAE) called the *prior encoding variational autoencoder* ( $\pi$ VAE). Traditional VAEs, although useful in practice as they extend auto-encoder into generative models, are however limited by their inference capabilities, in the sense that they represent finite dimensional random variables. They are not stochastic processes per se. In contrast, we show that  $\pi$ VAE is a new continuous stochastic process and therefore, it can be used to learn low-dimensional embeddings of function classes by combining a trainable feature mapping with a generative model using a VAE. It is shown that the framework can accurately learn expressive function classes such as Gaussian processes, but also properties of functions such as their integrals. For popular tasks, such as spatial interpolation, [Mis+22] shows that  $\pi$ VAE achieved state-of-the-art performance in terms of both accuracy and computational efficiency. This work introduces an elegant and scalable means of fully performing Bayesian inference for stochastic processes within probabilistic programming languages such as Stan [GLG15]. Chapter 4 concludes with a brief mention of two related VAE-based models, subsequent to  $\pi$ VAE, namely AggVAE [Sem+23a] and PriorCVAE [Sem+23b], which were trained for direct applications to the mapping of infectious diseases.

I worked on  $\pi$ VAE in the first 6 months of my Ph.D. and contributed originally by coding a  $\pi$ VAE in Pytorch, by clarifying its training procedure, whilst working on synthetic data to estimate its performance and educate our understanding of the various parts of the loss function.

## 0.2 Brief literature review on branching processes

Branching processes are mathematical models that prove to be most useful in describing the random dynamics of evolving populations whose *individuals*, assumed to live independently, can reproduce over time and die according to various probability laws. Depending on the field of application, the individuals composing the population can be seen, for instance, as particles in physics, cells in biology, infectious cases in epidemiology, or even as family names.

Indeed, the theory of branching processes has a long history in mathematics, whilst

traditionally and originally being rooted in the mid-nineteenth century in the problem of estimating the probability of extinction of family names. The solution of this latter problem is known to have been treated simultaneously by Francis Galton and Henry Watson in England and in parallel by the French probabilist Irénée-Jules Bienaymé. It led to the famous Bienaymé-Galton-Watson branching process, most commonly known as the Galton-Watson (GW) process, where each individual lives for one unit of time (or a fixed but given time period, equal for all generations) and then reproduces independently and randomly according to a given reproduction law. Athreya and Ney [AN04, Ch.1] provide a detailed treatment of Galton-Watson processes and, in particular, of their limit theorems.

Next, in the middle of the twentieth century, and as a general extension of Galton-Watson processes, Bellman and Harris [BH48], then Harris [Har63, Ch.VI] and Sevast'yanov [Sev67] meaningfully strengthened and extended the theory of branching processes towards *age-dependent* branching processes: individuals still reproduce at time of death like in a GW process, but the concept of age is introduced in the structure of the branching process. The most notorious age-dependent branching process has become known as the Bellman-Harris (BH) process. In such a process, each individual lives for a random age according to a life-length random variable  $L \geq 0$ , and then produces a random number of children at death time, according to a reproduction law  $\xi$ . The random variables  $L$  and  $\xi$  are assumed independent. The Sevast'yanov branching process is defined similarly, except that in this case,  $\xi$  and  $L$  can be dependent, which means that the number of progeny of an individual may depend on its lifetime. Many important results related either to the asymptotic behaviour or to the general dynamics of these age-dependent branching processes have since been proven. Given the focus of the next two chapters of this thesis, we cite here in particular Bellman [BH48] who shows that the expectation of a BH process follows a renewal-type equation.

A decade or two later, in particular, in 1968-1969, Crump and Mode [CM68] and [CM69] and independently Jagers [Jag69a], introduced *generalised branching processes*. Indeed, and as noted by Crump and Mode [CM68], age-dependent branching processes are restrictive in their modelling capacities, as they must abide by two constraints that greatly limit their field of application. In fact, not only age-dependent branching processes such as BH must assume that parents die at the time they reproduce. In addition, these age-dependent models assume that the probability that an individual of a given age  $\tau \geq 0$  gives birth within the age interval  $(\tau, \tau + d\tau)$  does not depend on the history of the individual considered. However, in practice, births happen during the lifetime, and the probability of giving birth in a certain age interval depends on the history of the individual. These two latter restrictions led Crump, Mode, and Jagers to introduce a generalised branching process where individuals could not only give birth at various time points during their lives but also where the life-length random variable and the reproduction point process could depend on each other. This dependency opens the door to cases where the probability of giving birth within a certain age interval may depend on the history of the considered individual. Therefore, historically one of the main reasons for introducing generalised branching models was the need to generalise age-dependent branching processes that had been introduced and studied in particular from the point of view of integral renewal-type equations by Harris [Har63] and Sevast'yanov [Sev67; Sev68]

It is with these limitations in mind that Crump and Mode [CM68] but also Jagers in various articles (e.g. [JN84]) and in his reference book [Jag75], introduced a more general age-dependent branching process, traditionally named Crump-Mode-Jagers (CMJ). In a CMJ branching process, each individual lives for a random time distributed according to a random variable  $L$  and reproduces at random ages according to a point process  $\xi$ . To be more specific and as described by [Bal+14], an individual denoted  $u$ , born at time  $b_u \geq 0$ , lives

and reproduces according to its individual laws  $L_u$  and  $\xi_u$ . Then, denoting the successive ordered points of  $\xi_u$  by  $0 \leq \tau_{u1} \leq \tau_{u2} \leq \dots \leq L_u$ , the branching model assumes that the individual labelled  $u$  gives birth to one child at each of the time points (which do not have to be distinct) denoted  $b_u + \tau_{u1}, b_u + \tau_{u2}, \dots, b_u + L_u$ . (Note that we will detail the proper labelling of individuals in the population in Chapter 1). Thus, CMJ branching processes generalise age-dependent branching processes since individuals can give birth at random times during their lifetime. BH processes can be viewed as forming a particular case of CMJ processes where reproduction can only occur at death time.

CMJ processes have been studied from various angles and by many authors, and in particular by Jagers [Jag75, Ch. 6]. Similarly to Crump and Mode, Jagers relies on a description of the process based on *populations counted with random characteristic*. A perspective based on random characteristic allows us to describe a branching population in any random phase of life that might or might not depend on the life length or the reproduction process of the individuals. Extensions such as multi-type CMJ branching processes have also been treated from different perspectives and with various objectives. We very basically here only cite the reference book of Mode [Mod71], Cohn's article [Coh89] and a more recent work by Kolesko [KS23].

As previously mentioned, Chapters 1 and 2 of this thesis introduce a time-varying extension of a CMJ branching process, denoted TV-CMJ. The latter specifically brings calendar-time dependence to the probability measure that underlies a classical CMJ branching process. It was originally presented in the article by Pakkanen et al. [Pak+23]. At first, it could be argued that the introduction of calendar-time dependence could possibly be viewed as a particular case of branching processes in varying environments. Those were first introduced in discrete time by [SW69] who assumed that the sequence of environments was identically and independently distributed (i.i.d.). Then, more general cases of randomly distributed environments, i.e. not just i.i.d., were further introduced, for instance, by [AK71; KV17; CJ94; Ban09; BB13]. They became called *branching processes in random environment*, in short BPRE. However, to the best of our knowledge, when BPRE are studied, the environments from one generation to another are constrained to be sampled from a *given and identical* distribution, whereas, in a TV-CMJ branching process, the probability measure underlying the process is time dependent and can therefore continuously vary with calendar time without having to be sampled from a given distribution.

Moving away from general random environments but focussing on the literature on branching processes and time dependence, we note that Fildes [Fil72] introduced an age-dependent branching process with a variable life-time distribution that can vary at each successive generation. Observe also that Edler [Edl78] studied a generation-dependent CMJ branching process where each generation is determined by its own common probability distribution with respect to the life-length and reproduction process. Again, our time-varying CMJ model differs from these examples in the sense that we assume that the reproduction law can change continuously with time and not discretely per generation. We assume in effect that the underlying probability measure that governs the dynamics of the TV-CMJ process can vary at any time  $t \geq 0$ .

A common choice in the literature of branching processes, especially when it comes to the study of their asymptotic behaviour, consists of writing them as random counting measures [Dyn94; Daw+94], or more specifically as the sum of Dirac measures. In fact, this point of view will be adopted in Chapter 3 when we study the  $L \log L$  condition and convergence of the first-moment semigroup to deduce a law of large numbers for structured

branching processes. Random measures, in other words point processes, have proven to be very efficient tools to describe and analyse the long-term dynamics of generalised branching processes, where particles behave independently and are characterised by a trait valued in an abstract measurable space and which follows a Markovian dynamics. Each individual is fully characterised and represented by the Dirac measure at its trait value.

### 0.3 Synthetic overview of mathematical modelling in epidemiology

The mention of the incidence renewal equation at the beginning of this introductory chapter brings us to clarify the perspective adopted in this thesis and also leads us to provide a very high-level overview of mathematical modelling in epidemiology. Indeed, Chapters 1 to 3 study branching processes, with a focus either on their governing integral equations (Chapters 1 and 2) or their asymptotic behaviour (i.e. long-term behaviour in the supercritical case in Chapter 3). We will also observe that the deep learning model presented in Chapter 4,  $\pi$ VAE, or more precisely its sister models AggVAE [Sem+23a] and PriorCVAE [Sem+23b], although based on Bayesian variational inference and, therefore, not on branching processes per se, provide direct practical and useful applications in the field of infectious disease mapping.

Consequently, even if slightly remote when it comes to Chapter 4, a common denominator or thread between all models introduced throughout the four chapters of this thesis is *the potential for applications to epidemiology*. The choice had to be made between adopting a perspective based on a branching population or on epidemiology. We have chosen the former and therefore follow the conventions of branching processes as a rule throughout the thesis. However, at some points, especially when we discuss the incidence renewal equation in Section 2.5, we may temporarily switch to a more specialized lexicon for epidemiology.

In order to give some idea of where branching processes stand within the palette of various mathematical models applied to model epidemics, we now provide a very rapid and non-exhaustive overview of the most commonly used models in this very large field. In fact, there exist different streams or approaches when it comes to the mathematical modelling of infectious diseases. The main dichotomy could possibly be seen as deterministic versus stochastic models. Starting with deterministic models, Ross in 1915 [Ros15], and then, Kermack and McKendrick in 1927 [KM27], introduced the idea of splitting the population into *compartments* and using transition rates to govern the dynamics from one compartment to another. They also introduced a threshold quantity, the so-called *reproduction number*, to summarise the state of an outbreak. The main basic compartments model is known as the SIR model. In brief summary, Kermack and McKendrick consider a closed population of size  $N$  and introduce for any time  $t \geq 0$ , the number  $S(t)$  of individuals susceptible to the disease, the number  $I(t)$  of infected individuals (capable of spreading the disease) and the number  $R(t)$  of individuals recovered from the disease. They structure the infected population by *age-since-infection*, denoted  $s$ , which corresponds to the time elapsed since infection occurred. Transmission rates, generally denoted  $\beta(s)$ , for  $s \geq 0$ , and defined as effective contact rates between an infected of infection age  $s$  and a susceptible, allow one to account for changes in infectiousness during the infection period. Furthermore, removal rates usually denoted  $\gamma(s)$  allow us to describe dynamics between infected and recovered groups. Kermack and McKendrick [KM27] derived a system of integro-differential equations that has subsequently been declined in several variants that in particular introduce more refined sets of compartments (e.g. SEIR models extend SIR models to add an *exposure* stage; SIS models are used in cases

when immunity is not permanent; SIVR models take into account a group of vaccinated individuals ...).

Beyond compartmental models, but still within the deterministic framework, we can mention *network models* that use graph theory to represent individuals as nodes and interactions as edges. This type of model tends to be useful for studying diseases that spread through close contact and social interactions. Another type of model, called *agent-based models (ABMs)*, simulates the actions and interactions of individual agents to assess the effects on the system as a whole, allowing for heterogeneous behaviour.

As our focus in Chapter 2 is to provide mathematical justification to the incidence renewal equation, we mention the meaningful contributions of Odo Diekmann who wrote extensively on the connections between epidemic models and renewal or delay equations; see e.g. [DGG08], [Die+10], [DG12] and his monograph on delay equations [Die+12]. Note also that in [Bre+12], the authors, including Diekmann, show that a large class of epidemic models can be mathematically formulated as a scalar renewal equation for the force of infection.

Finally, still in the context of deterministic models and, importantly, without any pretension to exhaustiveness, we mention *time-since-infection models* which were also originally derived by Kermack and McKendrick [KM27] and can be viewed as an enhancement to SIR models. More recently, Brauer [Bra05; Bra08] and Yang and Brauer [YB08] contributed to presenting and unifying the many variants of time-since-infection models. Given the resonance of the latter models with renewal-type equations, we now follow Yang and Brauer's formulation and present more specifically the time-since-infection model with infection-age-varying transmission rate. We see that this model leads to a renewal equation. *An assumption* is made that if  $i(t, s)$  is the number of infected individuals of infection-age  $s$  at time  $t$ , then it can be written as

$$i(t, s) = c(t - s)\mathcal{F}(s),$$

where  $c(t - s)$  is the number of new infected individuals at time  $t - s$  and  $\mathcal{F}(s)$  is the probability for an infected individual of infection-age  $s$  to stay in the infected population (survival probability). Assuming that  $\mathcal{F}(s)$  has the required properties of a cumulative distribution function (CDF), it follows that the number of infected individuals  $\mathcal{P}(t)$  at time  $t$ , in other words prevalence, is given by

$$\mathcal{P}(t) = \int_0^\infty c(t - s)\mathcal{F}(s)ds,$$

whereas the number  $c(t)$  of new infected individuals at time  $t$  (incidence) is governed (notwithstanding initial and boundaries conditions which we omit in this brief review) by

$$c(t) = \frac{\mathcal{S}(t)}{N} \int_0^\infty \beta(s)c(t - s)\mathcal{F}(s)ds,$$

where  $\mathcal{S}(t)$  is the number of individuals susceptible to infection at time  $t$ .

This time-since-infection model therefore leads to a Volterra or renewal integral equation which in essence states that the number of newly infected individuals (i.e. incidence) at time  $t$  is the result of a convolution between the transmission rate  $\beta(s)$  and the number of cases alive during the period  $[0, t]$  (i.e. prevalence). We will provide additional mathematical grounding to the incidence renewal equation and to this convolution backcalculation in Section 2.5.

This non-exhaustive list of deterministic models leads us now naturally to mentioning stochastic types of epidemic models. In the 1920s, Reed and Frost introduced a stochastic epidemic model. There is no formal historical literature reference for the Reed-Frost model as it actually was never published and only taught in biostatistics classes at the time. In essence, the Reed-Frost model is an example of a *chain binomial model*: it assumes that each infected individual in a given generation independently infects each susceptible individual in the population with some fixed probability  $p$ . Despite being somewhat simplistic, this binomial model has historically proven effective in the context of moderately large populations. Among other assumptions, the Reed-Frost chain-binomial model assumes that only one new case would be produced if a particular susceptible came into contact with more than one infected case. Following the unpublished works of Reed and Frost of the 1920s, Bailey in 1953 [Bai53] introduced a stochastic analogue to the Kermack and McKendrick compartment model [KM27] mentioned above. Bailey's *general stochastic epidemic* assumes homogeneous mixing of the susceptibles and, in particular, a partial differential equation for the probability generating function of the reproduction process is derived. Additionally, we mention Frank Ball who analysed non-Markovian stochastic epidemic processes in [Bal86]. We should also observe that most of the deterministic models cited above have also been given a stochastic variant which means that models such as stochastic SIR or SEIR models or random network models complement the growing panoply of mathematical models for infectious disease.

Having briefly mentioned these stochastic epidemic models, we now turn to the usage of branching processes to represent and study mathematically the spread of infectious diseases. In such a case, we assume that an infection or more generally an epidemic, at least at its onset or in its early phases, can be modelled as a branching process whereby individuals (e.g. infectious cases) behave identically and independently of each other. This modelling approximation was first mentioned in the work of Bartlett [Bar78] and Kendall [Ken56] and given general mathematical grounding, at least asymptotically, through the works of Ball [Bal83], Ball and Donnelly [BD95] and Metz [Met78] who showed that an epidemic process converges to a limiting branching process when the number of susceptible individuals in the population tends to infinity. More widely, this connection has often been made in practice and in the literature (see, e.g. [KM27; Bar67; Jac10; FKG03; LM23; Hau+15; VM11; Ish05]). In particular, we cite Ball et al. [Bal+14] who explain that the process representing the number of infectious individuals in an epidemic model can be satisfactorily described mathematically by a branching process if we can assume that the population is homogeneously mixing and that the number of infected people is small relative to the total size of susceptible individuals in the population.

In summary, it can be said that branching processes represent an effective mathematical tool when it comes to modelling epidemics stochastically in large populations when non-linear effects like susceptible depletion can be neglected.

# CHAPTER 1

## Time-Varying Crump-Mode-Jagers branching process

### 1.1 Introduction

In this chapter, we consider and study a *time-varying* extension of a single-type *classical* CMJ branching process counted by random characteristic. This model will be denoted TV-CMJ. The calendar-time dependence introduced into the CMJ branching process will eventually lead to the derivation of integral renewal-type equations in Chapter 2, which will themselves provide full mathematical justification to the incidence renewal equation commonly used in epidemiology.

Before we present the specificities of our time-varying extension of a CMJ process, let us note that CMJ branching processes in random or changing environments (see, e.g. [CJ94]) have been extensively treated in the literature, while more specifically, some time variations of CMJ processes have also already been considered, such as generation-dependent CMJ [Edl78] and [BG87], or CMJ with variable life distributions [Fil72]. The TV-CMJ process considered here can however be distinguished from the latter models in the sense that it assumes that the entire behaviour (in particular, the lifelength distribution and the reproduction law) of each individual continuously varies with calendar time. We draw attention of the reader to the fact that the time-dependence comes directly from the probability measure underlying the branching process. Indeed, we will see that the lifelength random variable  $L$  and the reproduction point process  $(\xi(t))_{t \geq 0}$  defined in the subsequent sections depend on time only through their distributions or laws. The calendar-time dependence has to be followed through the probability measure underlying the construction of the branching process and therefore has consequences on the form of the renewal-type integral equations that govern the dynamics of a TV-CMJ branching process. This is the result we wished to obtain to generalise the incidence renewal equation in epidemiology to a time-varying reproduction process. Indeed the inclusion of calendar time into the various governing integral equations that describe the dynamics of the branching process proves very useful in real-life applications such as infectious disease modelling where dependence on calendar time is a crucial ingredient to ensure models are realistic. This specific angle of study, to the best of our knowledge, has not yet been treated in detail in the literature.

Thus in summary, in a TV-CMJ branching process, the whole behaviour of each

individual is time-varying. This calendar-time dependence will simply arise from the particular assumption that the original ancestor denoted 0 is born at a given nonnegative time denoted  $\sigma_0 := \sigma \geq 0$  on which the law of each random variable and stochastic process underlying the TV-CMJ branching process will depend.

We place ourselves in the *simpler* case of a single-type branching process without immigration, since our focus in the rest of the chapter will be to determine how the classical integral renewal-type equations governing a CMJ branching process evolve when a calendar-time dependence is introduced. Assuming immigration and/or a multitype setting would only cloud the focus of our study on the effect of time dependence.

This chapter is structured as follows. In Section 1.2, we provide a canonical probabilistic construction of a classical CMJ branching process which will serve as a basis for the construction of a TV-CMJ branching process in Section 1.3. A TV-CMJ branching process is then formally defined in Section 1.4. In passing, we propose two examples of TV-CMJ branching processes as extensions of well-known classical types, namely the time-varying Bellman-Harris (TV-BH) branching process and the time-varying Inhomogeneous Poisson branching process (TV-IP). Finally, in Section 1.5, we recall regularity conditions for a traditional CMJ branching process, and then adapt the latter to provide general regularity conditions ensuring that a TV-CMJ branching process is well defined and does not explode in finite time almost surely.

## 1.2 Probability space for a CMJ process

Various canonical probabilistic constructions of CMJ branching processes can be found in the literature; see e.g. [CM68; Jag75]. However, they have rarely been presented in the detailed manner of this current section, mainly because of the underlying assumption that individual reproduction processes of a CMJ are i.i.d., which in turn implies that the measure on the population space can simply be defined as a product of individual Cartesian measures.

As we will see, in a TV-CMJ process, we depart from this i.i.d. assumption for the reproduction process and thus have to pay more attention to how the overall probability measure needs to be constructed. This is mainly why we present below a relatively detailed measure-theoretic construction of a CMJ branching process, which will serve as a basis to be adapted in Section 1.3 in order to define the probability space underlying a TV-CMJ.

Although already briefly mentioned in Chapter 0, we recall in slightly more detail the concept of *family histories* originally introduced by Harris [Har63]. Family histories are meant to represent the basic sequences of events (i.e. childbearing ages and lengths of life) that can be realised by individuals who make up the population of a branching process. Generally speaking, a family history fully describes all of the life-cycles, also called life careers, of each of the *conceivable* individuals in the entire population. More precisely, a life career of an individual must encompass, at the minimum, all the necessary and sufficient information about its random lifelength, its random number of children, and its random childbearing ages. In particular, Harris observes that family histories often include some redundant information, especially with respect to offspring that is never born (i.e. realised). However, despite their redundancy, family histories prove to be very practical building blocks for uniquely determining a specific family tree of a branching process. Moreover, it is on this space of family histories that a probability measure underlying the branching process can be fully determined.

In the same way that Harris [Har63] constructed a probability space to underlie the

rigorous measure theoretic definition of an age-dependent branching process, Crump and Mode [CM68] also provided a construction of a probability space on which any random function related to a CMJ process can be defined. Somewhat similarly, but without directly invoking family histories, Jagers relies on the underlying dynamics of the process to define the probability space of a CMJ process in Section 2 of [Jag75].

We have observed though that these classical constructions cannot be fully mirrored and extracted from the literature as such to describe the specific probability space underlying a TV-CMJ process and more precisely to describe its associated probability measure. On the one hand, in the classical CMJ process, reproduction laws are i.i.d. and a simple Cartesian product space can be relied upon, whereas on the other hand, in a TV-CMJ process, the time dependence implies that the reproduction processes are not i.i.d. anymore and the overall probability measure needs to be constructed using transition kernels indexed by the time of birth of parents.

### 1.2.1 Ulam-Harris-Neveu notations and set of individuals

Let us first describe the population of branching individuals. We denote by  $\mathcal{U}$  the entire population of *all conceivable* individuals (i.e. realised or not), initiated by a unique individual, the *ancestor*, denoted 0, whose progeny is labelled using the classical Ulam-Harris-Neveu notations, originally introduced by Harris [Har63].

We briefly recall this commonly used labelling system by defining  $\mathbb{N} := \{1, 2, \dots\}$  and writing  $\mathbb{N}^0 := \{0\}$ , so that the ancestor 0 is the unique individual belonging to  $\mathbb{N}^0$ . Children of the ancestor are then naturally represented by single positive integers (elements of  $\mathbb{N}$ ), while grandchildren are represented by elements of  $\mathbb{N}^2 \dots$

More generally, any individual  $u \in \mathcal{U}^* := \mathcal{U} \setminus \{0\}$  can be described by a unique  $n$ -tuple ( $n \in \mathbb{N}$ ) of positive integers  $u = \langle u_1 u_2 \dots u_n \rangle \in \mathbb{N}^n$ , where  $u$  is the  $u_n$ -th child of the  $u_{n-1}$ -th child  $\dots$  of the  $u_1$ -th child of the ancestor. We say that an individual  $u \in \mathcal{U}$  is *realised* if it was ever born. Thus, we will distinguish at times between  $\mathbb{N}^n$  which represents the set of all *conceivable* individuals of length  $n$ , and  $\mathbb{G}_n \subset \mathbb{N}^n$  which will represent the generation  $n$  of individuals of length  $n$  that have actually been born. In other words, a realised individual  $u = \langle u_1 u_2 \dots u_n \rangle$  is then said to belong to generation  $\mathbb{G}_n \subset \mathbb{N}^n$ , and we write  $|u| := n$ , whereas the set of all conceivable individuals of potential length (ancestry) equal to  $n$  is simply  $\mathbb{N}^n$ . Note that the ancestor is the only individual in generation 0, which we denote  $\mathbb{G}_0 := \{0\}$ .

As specified by Jagers and Nerman [JN84], if  $u = \langle u_1 \dots u_n \rangle \in \mathbb{N}^n$  and  $v = \langle v_1 \dots v_m \rangle \in \mathbb{N}^m$ , then, regardless of whether  $u, v$  or  $uv$  are realised, we write  $uv = \langle u_1 \dots u_n v_1 \dots v_m \rangle$  for the (potentially hypothetical) concatenation of both individuals. By convention, we assume  $\langle 0u \rangle = \langle u0 \rangle = u$ . Additionally, for any positive integer  $u_{n+1} \in \mathbb{N}$ , and any individual  $u \in \mathbb{N}^n$ , the  $u_{n+1}$ -th child of individual  $u$  will simply be denoted  $\langle uu_{n+1} \rangle$ , and we will obviously have that  $\langle uu_{n+1} \rangle \in \mathbb{N}^{n+1}$ . In such a case, individual  $u$  is called the *parent* of  $\langle uu_{n+1} \rangle$  and  $\langle uu_{n+1} \rangle$  is the  $u_{n+1}$ -th *descendant or child* of  $u$ . This leads us naturally to describe the entire population of all *conceivable* individuals as

$$\mathcal{U} := \bigcup_{n=0}^{\infty} \mathbb{N}^n.$$

A graphic representation of the Ulam-Harris-Neveu notations is shown in figure 1.1.

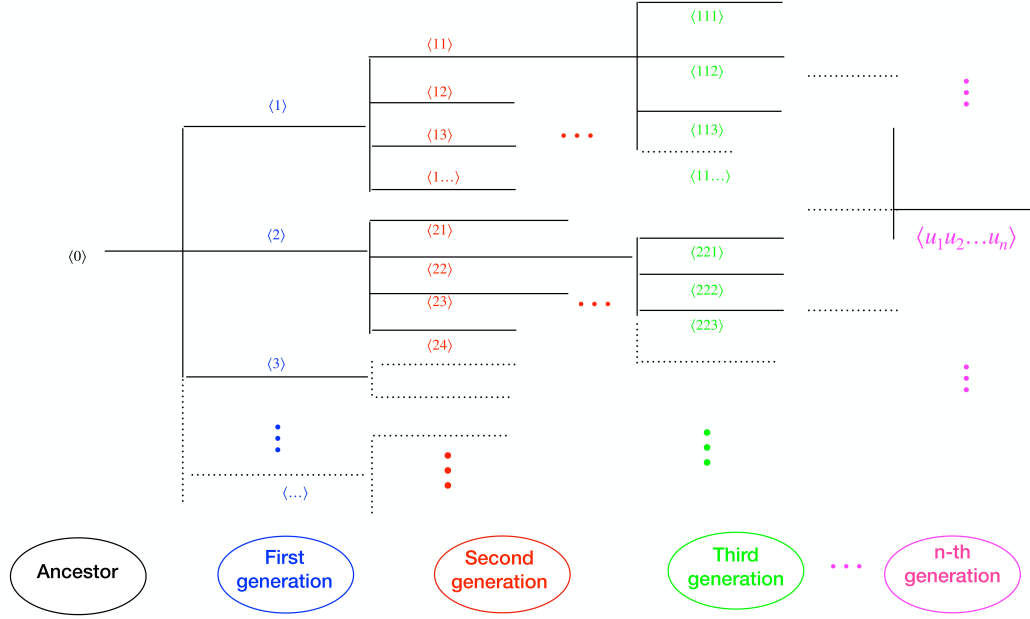


Figure 1.1: Graphic illustration of the Ulam-Harris-Neveu indexation of a branching population. At this stage, we are only depicting how the individuals in their respective generations are labelled, without accounting for potential differences in lifelengths or childbearing ages. The labelling scheme shows how the naming of a new offspring simply consists in appending a new *index* to the parent's tuple. For instance, any tuple of the form  $\langle u_1 u_2 u_3 \rangle$  represents an individual belonging to the third generation.

### 1.2.2 Family histories

Thus far, we have introduced the population set. But to construct a branching population, each individual  $u \in \mathcal{U}$  needs to be characterised, at the minimum, by its random lifelength, represented say by a non-negative real-valued random variable  $L_u$  and by its point reproduction process  $(\xi_u(t))_{t \geq 0}$ , whose points  $(\tau_u(k))_{k \in \mathbb{N}}$  are assumed to represent the successive childbearing ages when its  $k$ -th child is born. We will define these quantities more formally in the subsequent sections of this chapter, but for the moment, let us note that infinite sequences made up of these previous individual random quantities lead to defining individual family histories or individual life-careers, which we choose to denote  $(\omega_u)_{u \in \mathcal{U}}$ . When practical, if  $u \in \mathbb{N}^n$ , we may also denote  $\omega_u := \omega_n$  to emphasise that we are considering the life career of an individual in  $\mathbb{N}^n$ . Each  $\omega_u$  is generally defined as an infinite sequence of random lifelengths and childbearing ages.

Individual life-careers make up family histories which are traditionally considered to be the building blocks composing any event space of a branching population. Finally, each family history can be thought of as a realisation of a branching process that leads to a specific branching tree.

### 1.2.3 Events space

Here, we try to remain as general as possible and simply start by giving ourselves an abstract probability space  $\Omega_0$ .

We assume that the potential life careers  $\omega_0$  of the ancestor 0 live on  $\Omega_0$ . We associate the natural sigma-algebra  $\mathcal{S}_0$  to  $\Omega_0$ . Any given element  $\omega_0 \in \Omega_0$  must contain sufficient information to describe the life-cycle of the ancestor.

Moving forward from generation 0, we observe that each of the first-generation individuals, simply labelled  $i$ , where  $i \in \mathbb{N}$ , can be similarly associated with a random sequence  $\omega_i$  in  $\Omega_0$  that represents a realisation of its own life career. Thus, to describe the life-cycle of each of the conceivable first-generation individuals in  $\mathbb{N}$ , we require an identical copy of  $\Omega_0$ . It follows that the space  $\Omega^{(1)}$  describing the life-cycle of all conceivable first generation individuals is given by

$$\Omega^{(1)} := \Omega_0^{\mathbb{N}}.$$

More generally, letting  $k \in \mathbb{N}$ , for each conceivable individual in  $\mathbb{N}^k$ , another independent copy of  $\Omega_0$  is required to represent this individual's possible life careers. It follows that the events space related to  $\mathbb{N}^k$  is

$$\Omega^{(k)} := (\Omega_0)^{\mathbb{N}^k}.$$

By convention, note that for  $k = 0$ , we define  $\Omega^{(0)} := \Omega_0$ . In parallel, for any  $k \geq 1$ , we endow the events space  $\Omega^{(k)}$  with its natural sigma-algebra  $\mathcal{S}^{(k)}$ , which by Appendix A.1.1 is given by the product

$$\mathcal{S}^{(k)} := (\mathcal{S}_0)^{\otimes \mathbb{N}^k}.$$

Next, in order to determine the events space related to any individual in the entire population  $\mathcal{U}$ , we introduce for any  $n \in \mathbb{N}$  the *cumulative* product space of events, denoted  $\Omega_n$ , such that

$$\Omega_n := \prod_{k=0}^n \Omega^{(k)}.$$

In other words,  $\Omega_n$  is the events space of family histories associated to all individuals belonging to  $\bigcup_{k=0}^n \mathbb{N}^k$ . In contrast, the events space  $\Omega^{(k)}$  is associated to the set of individuals belonging strictly to  $\mathbb{N}^k$  where  $k \leq n$ . (Note that we have already assumed that  $\Omega_0 = \Omega^{(0)}$  to ensure that this definition is consistent with  $n = 0$ .)

Again by Appendix A.1.1, the natural product sigma-algebra associated to  $\Omega_n$  is given by

$$\mathcal{S}_n := \bigotimes_{k=0}^n \mathcal{S}^{(k)}.$$

Finally, the entire events space of family histories pertaining to any individual  $x \in \mathcal{U}$  and its associated natural product sigma-algebra are given by

$$\Omega := \prod_{k=0}^{\infty} \Omega^{(k)}, \quad \mathcal{S} := \bigotimes_{k=0}^{\infty} \mathcal{S}^{(k)}.$$

An element  $\omega \in \Omega$  can be viewed as a countably infinite sequence of nonnegative real numbers  $\omega = (\omega_u)_{u \in \mathcal{U}}$  where, in effect,  $\omega_u$  describes the life cycle of an individual  $u$ .

We have now set up a measurable space  $(\Omega, \mathcal{S})$  and we are left with defining a measure on this space associated to the generalised branching process we wish to consider.

### 1.2.4 Product probability measure

To define a measure  $\mathbb{P}$  in  $(\Omega, \mathcal{S})$ , we first revert to the space  $(\Omega_0, \mathcal{S}_0)$  associated to the ancestor 0, for which we assume given a probability measure  $\pi_0$  that describes the distribution of any event (i.e. family history)  $\omega_0 \in \Omega_0$ . Introduced as such, the probability space  $(\Omega_0, \mathcal{S}_0, \pi_0)$  must then support at the minimum a random variable (r.v.) that represents the lifelength of an individual and also a point process whose points represent a nondecreasing sequence of childbearing ages. In other words, the space  $(\Omega_0, \mathcal{S}_0, \pi_0)$ , at least, supports the existence of the following random quantities

1. A real-valued nonnegative r.v.  $L \geq 0$ , such that  $L(\omega_0)$  represents the lifelength of the ancestor in the realisation  $\omega_0$ . Omitting reference to the random life career  $\omega_0$ , we will denote by  $G(t)$  the cumulative distribution function (CDF) of  $L$  such that

$$G(t) := \pi_0(L \leq t).$$

2. A point process  $(\xi(t))_{t \geq 0}$  whose ordered points, denoted  $(\tau(k; \omega_0))_{k \in \mathbb{N}}$ , represent a nondecreasing sequence of random childbearing ages, in other words, the successive ages of the ancestor 0 when its  $k$ -th child is born in the realisation  $\omega_0$ . We define an associated reproduction process

$$\begin{aligned} N : \mathbb{R}_+ \times \Omega_0 &\rightarrow \mathbb{N} \\ (t; \omega_0) &\mapsto N(t; \omega_0) := \#\{k \in \mathbb{N}; \tau(k; \omega_0) \leq t\}, \end{aligned}$$

which counts the number of children born by time  $t \geq 0$  from ancestor 0 in the realisation  $\omega_0$ . Note that by convention, if the  $k$ -th child is never born when considering a given life career  $\omega_0$ , then we assume that  $\tau(k; \omega_0) := \infty$ . Omitting again reference to  $\omega_0$ , the family of CDFs  $(F_k(t))_{k \in \mathbb{N}}$  for the ages of childbearing will be denoted

$$F_k(t) := \pi_0(\tau(k) \leq t).$$

*Remark.* Although each  $\omega_0 \in \Omega_0$  represents a complete development of the family composed of the ancestor 0 and its descendants, Crump and Mode [CM68] observe that each life career supports more information than the minimum required. In fact, it is possible that  $\tau(k) > l$ , where  $l$  is a realisation of  $L$  in  $\omega_0$ , in which case the  $k$ -th offspring is not realised.

In a *classical (i.e. non-time varying)* CMJ branching process, individuals reproduce and live independently. This independence implies that the probability measure associated with  $(\Omega, \mathcal{S})$  is simply the Cartesian product of  $\pi_0$  on the population  $\mathcal{U}$ , that is,

$$\mathbb{P} := (\pi_0)^{\mathcal{U}},$$

and  $(\Omega, \mathcal{S}, \mathbb{P})$  represents the probability space underlying a traditional CMJ process. We will see in the next section that one has to depart from this construction for the probability measure of a TV-CMJ process.

## 1.3 Probability space for a time-varying CMJ

We adapt the previous description of the probability space underlying a traditional CMJ branching process to a time-varying context. Recall that the distinguishing feature of

a *time-varying CMJ (TV-CMJ)* branching process, compared to a classical CMJ branching process (see e.g. [Jag69a]) relies on the fact that the behaviour of individuals is assumed calendar-time dependent. This additional assumption brings forth general calendar-time dependencies to the probability measure that determines such a process, whereby independence from one generation to the next does not hold any more, compared to the time-invariant case.

Jagers and Nerman [JN84] emphasised that branching populations are traditionally defined by the almost *self-defining* property that individual lives must be i.i.d. However, in a TV-CMJ process, we deviate from this i.i.d. assumption. Let us denote by  $\sigma_0 \geq 0$  the time of birth of the ancestor 0, and more generally  $\sigma_u$  for the time of birth of any other individual  $u$ . The departing assumption of a TV-CMJ compared to a CMJ process is that the time of birth of a parent is sufficient to fix the probability distributions related to its progeny. This calendar-time dependence of a TV-CMJ process from one individual to its progeny leads to the following Markov property: as future individuals depend on the past only through the time of birth, which we now generally denote  $\sigma$ , of their parent, the conditional distributions of future individuals given the past individuals and their related information are equal to the conditional distributions of these future individuals given  $\sigma$ . This Markovian property for the TV-CMJ branching process will prove repeatedly useful in the following derivations.

In particular, the canonical probabilistic construction in this section provides the necessary foundations for the measure-theoretic arguments used in the subsequent proofs of this chapter.

### 1.3.1 Family histories and events space

The construction of the space and its associated sigma-algebra  $(\Omega, \mathcal{S})$  according to Section 1.2 still holds for a TV-CMJ branching process, as it is not affected by the times of birth  $\sigma \geq 0$  of the original ancestor and its progeny. However, now the probability measure must take into account the time-varying element introduced by  $\sigma$ . To do so, we replace  $\pi_0$  of the previous section by  $\pi_0^\sigma$  to describe the distribution of any event  $\omega_0 \in \Omega_0$ .

Starting with the ancestor 0, similarly to before, the space  $(\Omega_0, \mathcal{S}_0, \pi_0^\sigma)$  must support a lifespan random variable  $L$  and a point process  $\xi$  whose ordered points, denoted  $(\tau(k))_{k \in \mathbb{N}}$ , represent a nondecreasing sequence of random childbearing ages. However, in contrast to the classical CMJ, the CDF of  $L$  must now depend on  $\sigma$ . We denote

$$G^\sigma(t) := \pi_0^\sigma(L \leq t). \quad (1.1)$$

Similarly, for all  $k \in \mathbb{N}$ , the CDFs of the childbearing ages must also depend on  $\sigma$  and are denoted

$$F_k^\sigma(t) := \pi_0^\sigma(\tau(k) \leq t). \quad (1.2)$$

*Remark.* Note here importantly that it is not the random variables themselves, i.e.  $L$  or  $\tau(k)$  that depend on  $\sigma$  but their distributions.

### 1.3.2 Recursive conditional probability measures

Recall that in the case of a CMJ branching process whose underlying probability measure is traditionally invariant in time, we could simply define the probability measure as

$$\mathbb{P} = (\pi_0)^{\mathcal{U}}.$$

Now, in a TV-CMJ process, due to the assumed continuous dependency of the probability measure on the time of birth  $\sigma$  of a parent, we cannot simply rely on Cartesian products to build up the probability measure over the whole space. Instead, we recursively construct on the cumulative space  $\Omega_n$ , for  $n \in \mathbb{N}$ , a sequence of related probability measures, which we will denote  $(\mathbb{P}_n^\sigma)_{n \in \mathbb{N}}$ , and which will be built as compositions of transition probability kernels  $(\kappa_k^\sigma)_{k \leq n}$  from  $\Omega_k$  to  $\Omega^{(k+1)}$ , for  $k \leq n$ . These kernels allow the bridging of successive generations.

*Remark.* In this section, we are mostly interested in how to transition from a parent in  $\mathbb{N}^n$  to one of its direct progeny in  $\mathbb{N}^{n+1}$ . We will therefore adopt shortened notations for individual life-careers and family histories. Instead of writing  $\omega_u$  for  $u \in \mathbb{N}^n$ , we may at times and when relevant write  $\omega_n$ , which can possibly then relate to any individual  $u \in \mathbb{N}^n$ .

Recalling the definition of a probability transition kernel (see Appendix A.1.2), we proceed by recursion, whereas for  $n = 0$ , we first let  $\Omega_0 = \Omega^{(0)}$  and

$$\mathbb{P}_0^\sigma := \pi_0^\sigma.$$

Next for any  $n \geq 1$ , we introduce transition probability kernels

$$\kappa_n : (\Omega_{n-1}, \mathcal{S}_{n-1}) \rightarrow (\Omega^{(n)}, \mathcal{S}^{(n)}),$$

defined, as we will make more precise below, as countably infinite products of the original measure  $\pi_0^\sigma$ . For any  $n \in \mathbb{N}$ , we (recursively) index the kernel  $\kappa_n$  with respect to the times of birth of individuals belonging to the  $n$ -th generation. Indexing is determined by induction on  $n \in \mathbb{N}$ , for any individual  $u \in \mathbb{N}^n$ , through nonnegative, real-valued, random functions  $\sigma_u(\omega_{n-1})$ , defined in the event space  $\Omega_{n-1}$ . These functions, in effect, compute the time of birth of  $u$ .

More precisely starting with  $n = 1$ , any first-generation individual labelled  $i$ , where  $i \in \mathbb{N}$ , can be associated with a random function  $\sigma_i(\cdot)$  defined on  $\Omega_0 = \Omega^{(0)}$  that computes the time of birth of the  $i$ -th child of ancestor 0 such that for any  $\omega_0 \in \Omega_0$ ,

$$\sigma_i(\omega_0) := \sigma + \tau(i; \omega_0).$$

$\sigma_i(\omega_0)$  is the time of birth of child  $i$  in the realisation  $\omega_0$ . Thanks to this formalism, a first transition probability kernel  $\kappa_1^\sigma$  from  $\Omega_0$  to  $\Omega^{(1)}$  can be defined as a product over  $\mathbb{N}$  of independent measures  $\pi_0^\sigma$  over-scripted by the times of birth of  $i$ , namely for any  $A \in \mathcal{S}^{(1)}$ ,

$$\kappa_1^\sigma(\omega_0, A) := \prod_{i \in \mathbb{N}} \pi_0^{\sigma_i(\omega_0)}(A).$$

Note that  $\kappa_1^\sigma(\omega_0, A)$  is a measure on  $\mathcal{S}^{(1)}$ . Next, appealing to the definition of kernels composition (recalled in Appendix A.1.3), we introduce a probability measure  $\mathbb{P}_1^\sigma$  on  $\Omega_1 = \Omega^{(0)} \times \Omega^{(1)}$

given by

$$\mathbb{P}_1^\sigma := \pi_0^\sigma \otimes \kappa_1^\sigma,$$

where the measure  $\pi_0^\sigma$  can be viewed as a probability kernel, and the composition is defined for any  $A \in \mathcal{S}^{(1)}$  by

$$\mathbb{P}_1^\sigma(A) := (\pi_0^\sigma \otimes \kappa_1^\sigma)(A) = \int_{\Omega_0} \pi_0^\sigma(d\omega_0) \int_{\Omega^{(1)}} \kappa_1^\sigma(\omega_0, d\omega^{(1)}) \mathbf{1}_A(\omega_0, \omega^{(1)}).$$

More generally, for any  $n \geq 1$  and any  $u = \langle u_1 \dots u_n \rangle \in \mathbb{N}^n$ , we assume defined on  $\Omega_{n-1}$  similar random functions  $\sigma_u(\omega_{n-1})$ , that compute the time of birth of  $u$  when  $\omega_{n-1}$  is realised. By induction, the  $u_{n+1}$ -th child of  $u$ , denoted  $\langle u_1 \dots u_n u_{n+1} \rangle := \langle u u_{n+1} \rangle \in \mathbb{N}^{n+1}$ , is then naturally associated with a random function  $\sigma_{uu_{n+1}}(\cdot)$  on  $\Omega_n$  that computes its time of birth recursively as follows

$$\sigma_{uu_{n+1}}(\omega_n) = \sigma_u(\omega_n) + \tau(u_{n+1}; \omega_n),$$

where we recall that  $\omega_n := \omega_u$  and  $\tau(u_{n+1}; \omega_n)$  is the age of  $u \in \mathbb{N}^n$  when its  $u_{n+1}$ -th child is born (realised or not).

Note that we slightly abused notations here as the function  $\sigma_u$  is originally only defined on  $\Omega_{n-1}$  and therefore when we write,  $\sigma_u(\omega_n)$  the realisation  $\omega_n$  must correspond to the restriction of  $\omega_n$  in  $\Omega_{n-1}$ . Finally

However, note that if the  $u_{n+1}$ -th child is not realised, we have by construction  $\tau(u_{n+1}; \omega_n) := \infty$ , which implies that any unrealised child will also be associated with an infinite time of birth.

*Remark.* Although the functions  $\sigma_u(\cdot)$  have so far been defined in  $\Omega_n$  for  $n \geq 1$  and  $u \in \mathbb{N}^n$ , we observe that we could extend their definition to  $n = 0$  and therefore to the ancestor 0, by introducing the (non-random) constant function  $\sigma_0(\omega_0) := \sigma$ .

We now return to the construction of the probability transition kernels and proceed identically to the case  $n = 1$ . In particular, we use the previously defined random functions  $\sigma_u(\cdot)$  as upper-script indices for the product of measures  $\pi_0^\sigma$ . We recursively define a kernel  $\kappa_n : \Omega_{n-1} \rightarrow \Omega^{(n)}$  as a product of independent measures, such as for any  $A^{(n)} \in \mathcal{S}^{(n)}$ ,

$$\kappa_n^\sigma(\omega_{n-1}, A^{(n)}) = \prod_{u \in \mathbb{N}^n} \pi_0^{\sigma_u(\omega_{n-1})}(A^{(n)}).$$

Using once again the result on kernels composition [Kal17, Ch.I p.33], we define recursively a probability measure  $\mathbb{P}_n^\sigma$  on  $\Omega_n$

$$\mathbb{P}_n^\sigma := \mathbb{P}_{n-1}^\sigma \otimes \kappa_n^\sigma,$$

which is given for any  $A_n \in \mathcal{S}_n$  by

$$\mathbb{P}_n^\sigma(A_n) = \int_{\Omega_{n-1}} \mathbb{P}_{n-1}^\sigma(d\omega_{n-1}) \int_{\Omega^{(n)}} \kappa_n^\sigma(\omega_{n-1}, d\omega^{(n)}) \mathbf{1}_{A_n}(\omega_{n-1}, \omega^{(n)}).$$

Observe that we could equally write

$$\mathbb{P}_n^\sigma = \pi_0^\sigma \otimes \bigotimes_{k=1}^n \kappa_k^\sigma.$$

Recall that the probability distribution  $\pi_0^\sigma$  in  $\Omega^0$  has been assumed to support, for any  $\omega_0 \in \Omega^0$ , the random quantities  $L(\omega_0)$ , and  $\tau(k; \omega_0)$ , from which the reproduction process  $N(t; \omega_0)$  could then be inferred. More generally now, for any  $n \in \mathbb{N}$ , the sequence of probability measures  $\mathbb{P}_n^\sigma$  defined in  $\Omega_n$  by the above construction will also support, at the minimum, for any individual  $u \in \mathcal{U}$ , with a life-career  $\omega_u$ , the lifelength r.v.  $L(\omega_u)$  and the nondecreasing sequence of childbearing ages  $(\tau(k; \omega_u))_{k \in \mathbb{N}}$ . The reproduction point process  $(N(t; \omega_u))_{t \geq 0}$  will then be similarly inferred.

In a final step, we need to extend the sequence of probability measures  $(\mathbb{P}_n^\sigma)_{n \in \mathbb{N}}$  to the whole space  $\Omega = \prod_{k \geq 0} \Omega^{(k)}$ . To do this, we rely on the Ionescu-Tulcea theorem (see Appendix A.1.3) which was originally proven by Ionescu [Ion49], and more recently detailed by Neveu [NFF65, Prop V.I.I p.162] and Pollard [Pol02, th. 49 p.101]. Typically, in a context where probability events are dependent, the Ionescu-Tulcea theorem can be invoked to ensure that not only the limiting probability measure  $\mathbb{P}^\sigma$  exists on  $\Omega$ , but is also *consistent* with the sequence of conditional probability measures  $\mathbb{P}_n^\sigma$  on each of the  $\Omega^{(n)}$  making up  $\Omega$ . In other words, Ionescu-Tulcea guarantees the existence of a probability measure  $\mathbb{P}^\sigma$  on  $\Omega$  such that for any  $A_n \in \mathcal{S}_n$ ,

$$\mathbb{P}^\sigma(A_n \times \prod_{k=n+1}^{\infty} \Omega^{(k)}) := \mathbb{P}_n^\sigma(A_n).$$

## 1.4 Counting by random characteristics

### 1.4.1 Formal definition of a time-varying CMJ branching process

In the previous section, we have defined the probability space  $(\Omega, \mathcal{S}, \mathbb{P}^\sigma)$  and, in particular, it was observed that the time dependence is accounted for through the construction of the probability measure  $\mathbb{P}^\sigma$ . We now closely follow and adapt the formalism introduced by Jagers [Jag75] for a CMJ to define a TV-CMJ branching process underlined by the probability space  $(\Omega, \mathcal{S}, \mathbb{P}^\sigma)$ .

Jagers [Jag75] notes that generalised branching processes rely on the idea of counting individuals along stochastic characteristics whose values depend on individual properties. In this context, a *random characteristic* is defined as follows:

**Definition 1.4.1** (Random characteristic). A random characteristic is any real-valued, non-negative random process  $(\chi(t, \omega))_{t \in \mathbb{R}_+}$ , measurable with respect to the sigma-algebra  $\mathcal{B}^+ \times \mathcal{S}$ , where  $\mathcal{B}^+$  is the sigma-algebra associated with  $\mathbb{R}_+$ . Furthermore, the paths of  $(\chi(\cdot, \omega))_{t \in \mathbb{R}_+}$  are assumed to be right-continuous with left-limits (càdlàg).

*Remark* (Normalisation). Given that the realisations of the paths  $\chi(t, \omega)$  have to be bounded and are used for counting, we will naturally always assume that  $\chi(t, \omega)$  is normalised, i.e.  $\chi(t, \omega) \in [0, 1]$ ,  $\forall t \geq 0$ , and  $\omega \in \Omega$ .

Next, we introduce for any individual  $u \in \mathcal{U}$  a translation operator  $S_u : \Omega \rightarrow \Omega$ , called

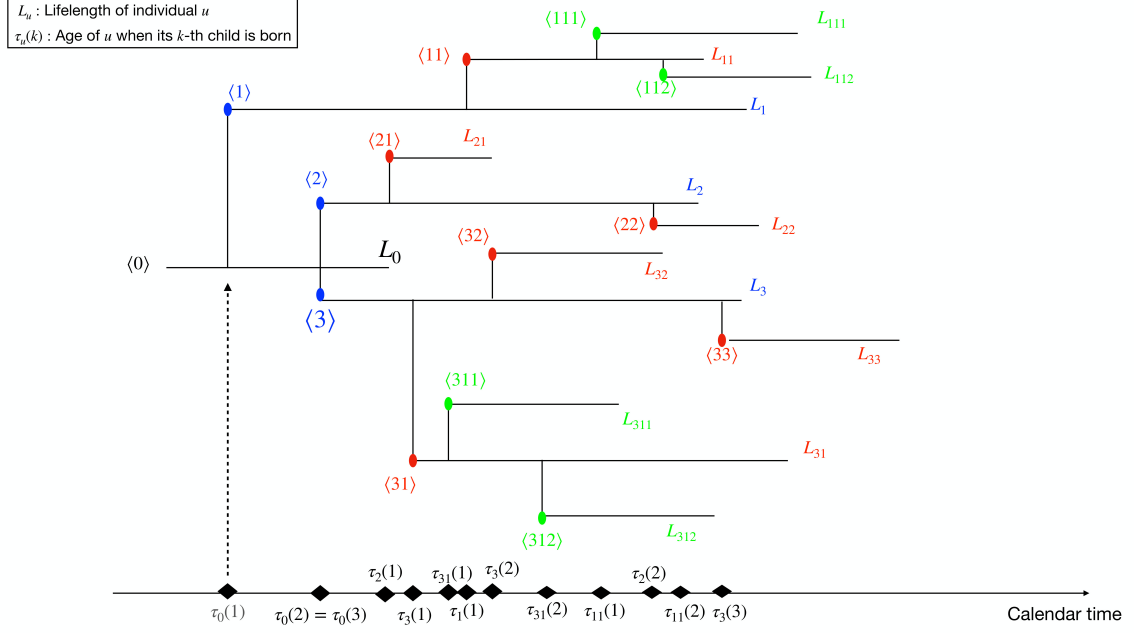


Figure 1.2: Realisation of a TV-CMJ process

by Jagers [Jag75] the *daughter process* of  $u$ , and which in effect renders  $u$  the possible ancestor of any element in  $\Omega$ . It could also be said that the process  $S_u$  gives roots to any element of  $\Omega$  with individual  $u$ . More precisely,  $S_u$  is a map that transforms any sequence of conceivable events  $(\omega_v; v \in \mathcal{U})$  into the shifted sequence  $(\omega_{uv}; v \in \mathcal{U})$  of potential events. Then, composing the daughter process  $S_x$  with a random characteristic  $\chi(t, \cdot)$  naturally leads to an *individual* characteristic pertaining to  $x$

$$\chi_u(t, \omega_u) := \chi \circ S_u(t, \omega). \quad (1.3)$$

The *individual* random characteristic  $\chi_u(t, \omega_u)$  will now be denoted  $\chi(t)$ , thus omitting references to  $u \in \mathcal{U}$  and to  $\omega_u \in \Omega$ .

**Definition 1.4.2** (TV-CMJ branching process). Let  $(\Omega, \mathcal{S}, \mathbb{P}^\sigma)$  be the probability space constructed in Section 1.3 and let  $(\chi(t; \cdot))_{t \geq 0}$  be a random characteristic according to definition 1.4.1, whose individual version  $(\chi_u(t, \cdot))_{t \geq 0}$  for any  $u \in \mathcal{U}$  is given by the composition in equation (1.3). Omitting references to  $\omega \in \Omega$ , a *time-varying generalised branching process*,  $Z^\chi(t; \sigma)$  counted by the random characteristic  $\chi(t)$ , also called a time-varying Crump-Mode-Jagers branching process (TV-CMJ), is defined as

$$Z^\chi(t; \sigma) := \sum_{u \in \mathcal{U}} \chi_u(t - \sigma_u), \quad t \in \mathbb{R}_+.$$

Observe that this infinite sum is always well-defined (i.e. it corresponds to the sum of a finite number of contributions related to individuals  $u \in \mathcal{U}$  whose times of birth  $\sigma_u$  happen to be finite).

As an example, Figure 1.2 provides a possible realisation of a TV-CMJ branching process.

*Remark.* From now on, unless necessary and for clarity purposes, we will omit any reference to  $\omega$  when we manipulate any of the random functions, such as times of birth, random

characteristics, childbearing ages, or reproduction process. In particular, we will denote  $\sigma_u := \sigma_u(\omega)$ , and  $\tau(k) := \tau(k; \omega)$ .

Note also that for any  $u = (u_1 \dots u_n) \in \mathbb{N}^n$ , the time of birth  $\sigma_u$  of individual  $u$  depends inductively on the time of birth  $\sigma$  of the ancestor 0 and on the successive childbearing ages as follows

$$\sigma_{u_1 \dots u_n} = \sigma + \tau(u_1) + \tau(u_1 u_2) + \dots + \tau(u_1 \dots u_n).$$

In addition, we will introduce the time-varying reproduction number as the expectation of the reproduction point process,

$$\begin{aligned} R : \mathbb{R}_+ \times \mathbb{R}_+ &\rightarrow \mathbb{R}_+ \\ (t; \sigma) &\mapsto R(t; \sigma) := \mathbb{E}^\sigma(\xi(t)), \end{aligned}$$

and the counting reproduction process as

$$\begin{aligned} N : \mathbb{R}_+ &\rightarrow \mathbb{R}_+ \\ t &\mapsto N(t) := \#\{k \in \mathbb{N}; \tau(k) \leq t\}. \end{aligned}$$

Finally, we define the expected reproduction process as

$$\begin{aligned} \Lambda : \mathbb{R}_+ \times \mathbb{R}_+ &\rightarrow \mathbb{R}_+ \\ (t; \sigma) &\mapsto \Lambda(t; \sigma) := \mathbb{E}^\sigma(N(t)), \end{aligned} \tag{1.4}$$

which will play an important role in the derivation of integral equations describing the dynamics of the branching process.

### 1.4.2 Examples of applications in epidemiology

Depending on the choice of the random characteristic  $\chi(t)$ , the resulting branching process  $Z^\chi(t; \sigma)$  may account for different quantities of interest. We give here two specific examples which have a very direct interpretation in the context of infectious disease modelling. In this section, we thus use a lexicon belonging to the field of epidemiology and we call individuals of a branching population *infectious cases*.

#### Prevalence

Prevalence in epidemiology is defined as the expected number of infectious cases present in an outbreak at any given time point. We assume that a TV-CMJ branching process is used to model the dynamics of an infectious disease. Then we observe that an infectious case  $u \in \mathcal{U}$  is present at time  $t \geq 0$  if it has already appeared but has not yet been eradicated, i.e. if

$$\sigma_u \leq t < \sigma_u + L.$$

We take as individual characteristic the indicator function (where  $t - \sigma_u$  is in effect the age of case  $u$  at  $t$ ) given by

$$\chi_u^{pr}(t - \sigma_u) := \begin{cases} 1, & \text{if } 0 \leq t - \sigma_u < L, \\ 0, & \text{otherwise.} \end{cases}$$

This random characteristic determines the following generalised TV-CMJ process

$$Z^{pr}(t; \sigma) = \sum_{u \in \mathcal{U}} \chi_u^{pr}(t - \sigma_u) = \sum_{x \in \mathcal{U}} \mathbf{1}_{[0, L)}(t - \sigma_u).$$

The process  $Z^{pr}(t; \sigma)$  accounts for the number of cases currently infectious at time  $t \geq 0$ . In the particular context of infectious disease modelling or more generally epidemiology, *the expectation of such a process is called prevalence*, which we denote here

$$Pr(t; \sigma) := \mathbb{E}^\sigma(Z^{pr}(t; \sigma)).$$

It corresponds to the expected number of individuals alive at time  $t \geq 0$  in the TV-CMJ branching process.

### Cumulative incidence

We now define another commonly used individual characteristic which leads to accounting for the cumulative number of births up to a certain time  $t \geq 0$  when considering a TV-CMJ branching process. It is described by the indicator function

$$\chi_u^B(t - \sigma_u) := \begin{cases} 1, & \text{if } \sigma_u \leq t, \\ 0, & \text{otherwise.} \end{cases} \quad (1.5)$$

The associated process counted by random characteristic is given by

$$Z^B(t; \sigma) = \sum_{u \in \mathcal{U}} \chi_u^B(t - \sigma_u) = \sum_{u \in \mathcal{U}} \mathbf{1}_{[0, \infty)}(t - \sigma_u).$$

From the perspective of an infectious disease, this process records the number of infectious cases that have ever appeared up to time  $t \geq 0$ . In epidemiology, taking the expectation of such a process would correspond to considering *cumulative incidence* which we denote

$$B(t; \sigma) := \mathbb{E}^\sigma(Z^B(t; \sigma)).$$

As the characteristic function  $\chi(t)$  is normalised in  $[0, 1]$ , the process  $B(t; \sigma)$  represents in effect the largest observable quantity when counting individuals by random characteristic and in particular will serve to define a regular TV-CMJ process (non-explosion in finite time).

If, on the one hand, the choice of random characteristic  $\chi(t)$  specifies the observable quantities in relation to a certain type of TV-CMJ process, then, on the other hand, the respective choices of probability measures for the lifespan r.v. and for the childbearing ages allow us in turn to determine the *type* of branching process we wish to consider. In particular, in the next section, specifying the properties of the lifespan distributions and/or of the reproduction process allows us to present two examples of TV-CMJ branching processes, extended from their well-known time-invariant cases which have proved popular and practical in the literature.

#### 1.4.3 Time-varying Bellman Harris branching process

In the classical Bellman-Harris (BH) branching process, originally fully described in [Har63], individuals can only give birth at the time of death. We describe here the time-varying extension of this commonly used model of branching process. Note that despite (or maybe

because of) the apparently simplifying assumption that births can only occur at times of death, we have to depart slightly from the general exposition proposed above of a TV-CMJ process.

In fact, recall that in the construction of the probability space in Section 1.3, a prespecified probability measure  $\pi_0^\sigma$  is assumed to support at the minimum the lifelength r.v.  $L$  and a point process  $(\xi(t))_{t \geq 0}$ , whose points represent the sequence of childbearing ages  $(\tau(k))_{k \in \mathbb{N}}$ . However, in a BH process, the childbearing ages are trivially equal by construction to  $\sigma + L$ , where  $\sigma$  is the time of birth of the mother. Therefore, it seems redundant to rely on the distribution of childbearing ages to specify a TV-BH branching process, as this latter distribution can be inferred from the lifelength distribution.

Thus, in order to describe the reproduction process in a TV-BH process, we depart from the description given by (1.1) and (1.2), and instead rely on an integer valued, exogenous (e.g. possibly environmental) point process  $\xi_{BH}(t)$ , defined on  $\mathbb{R}_+$ . This reproduction point process  $\xi_{BH}(t)$  does not depend on  $L$ , and its distribution, assumed independent on  $\sigma$ , is described by a probability mass function (pmf) given for any  $k \in \mathbb{N}$  by

$$\mathbb{P}^\sigma(\xi_{BH}(t) = k) := q_t(k).$$

The process  $(\xi_{BH}(t))_{t \geq 0}$  is not observable at any time  $t \geq 0$ , but it is, however, observable at any time of death of the form  $\sigma + L$ . The definition of the counting reproduction process  $N_{BH}(t)$  associated with the TV-BH follows for any  $t \geq 0$ ,

$$N_{BH}(t) = \xi_{BH}(\sigma + L) \mathbf{1}_{t \geq \sigma + L}.$$

Similarly to the TV-CMJ, we define the expected reproduction process  $\Lambda_{BH}(t; \sigma) = E^\sigma(N_{BH}(t))$  and the time-varying reproduction number  $R(t) := \mathbb{E}^\sigma(\xi_{BH}(t))$ . Recalling the independence of the process  $\xi_{BH}(t)$  with  $\sigma$ , observe here that the reproduction number  $R(t)$ , despite being measured with the probability  $\mathbb{P}^\sigma$ , is in effect independent of  $\sigma$ . In fact, we can also rewrite the reproduction number as a function of the pmf of the point process  $\xi_{BH}(t)$  such that for any  $t \geq 0$ ,

$$R(t) = \sum_{k=0}^{\infty} k q_t(k).$$

*Remark.* Whereas the dynamics of a TV-CMJ process is specified by the lifelength random variable  $L$  and the sequence of childbearing ages  $(\tau(k))_{k \in \mathbb{N}}$ , the dynamics of a TV-BH process is instead specified by  $L$  and a point process  $\xi_{BH}(t)$  observable at times  $\sigma + L$ . However, childbearing ages (whose binary values, depending on whether an individual is ever born or not, are given by  $L$  or  $\infty$ ) can be recovered from the point process as follows

$$\tau(k) := \begin{cases} L, & \text{if } k \leq \xi_{BH}(\sigma + L), \\ \infty, & \text{if } k > \xi_{BH}(\sigma + L). \end{cases}$$

To end this section, we derive a relationship between the expected reproduction process  $\Lambda_{BH}(t; \sigma)$  and the time-varying reproduction number  $R(t)$  that is particular to a TV-BH process. This formula will prove useful in the following sections.

**Lemma 1.4.1.** *In a TV-BH branching process, the expected reproduction process  $\Lambda(t; \sigma)$  is*

the convolution of the time-varying reproduction number  $R(\cdot)$  with the lifelength CDF  $G^\sigma(\cdot)$ .

$$\Lambda_{BH}(t; \sigma) = \int_{\sigma}^t R(s) dG^\sigma(s - \sigma).$$

*Proof.* We write, successively using the Total Law of Probability and the independence of  $\xi_{BH}(t)$  and  $L$ ,

$$\begin{aligned} \Lambda_{BH}(t; \sigma) &= \mathbb{E}^\sigma(N_{BH}(\sigma; t)), \\ &= \mathbb{E}^\sigma(\xi_{BH}(\sigma + L) \mathbf{1}_{t \geq \sigma + L}), \\ &= \mathbb{E}^\sigma \left( \mathbb{E}^\sigma \left( \xi_{BH}(\sigma + L) \mid L \right) \mathbf{1}_{t \geq \sigma + L} \right), \\ &= \mathbb{E}^\sigma \left( \mathbb{E}^\sigma \left( \xi_{BH}(\sigma + l)_{l=L} \right) \mathbf{1}_{t \geq \sigma + L} \right), \\ &= \int_0^{t-\sigma} R(\sigma + s) dG^\sigma(s). \end{aligned}$$

□

*Example* (TV-BH with a Poisson reproduction point process). Finally, as a practical and useful example of the TV-BH branching process, observe that we can assume that the discrete reproduction pmf is described by a Poisson distribution, with parameter  $R(t)$ . In this case, we have  $q_t(k) \sim \text{Poisson}_{R(t)}(k)$ .

#### 1.4.4 Time-varying inhomogeneous Poisson branching process

The TV-BH branching process presented in the previous section lacks in realism as, most often, we need to model populations whose individuals give birth at various times *during* their lives. With this lack of realism in mind, in this section, following [Pak+23], we aim to define a time-varying model of branching process which, in particular, is more adapted to realistically describe an infectious disease propagation, where individuals (i.e. infectious cases) reproduce randomly throughout their lifetime.

As in equation (1.1), we still assume given a nonnegative real-valued lifelength r.v.  $L$  and its CDF  $G^\sigma(t)$ . Here, similarly to how we proceeded to specify the TV-BH process, we will also determine the reproduction process rather than the distribution of childbearing ages as in (1.2). To that effect, we introduce two nonnegative real-valued functions which will bring flexibility to model separately on the one hand, the population-level, and on the other hand, the individual-level characteristics of the process. More precisely, we introduce

- a nonnegative measurable real-valued function  $t \mapsto \rho(t)$ , defined on  $\mathbb{R}_+$ , whose purpose is to model *the population-level* variation in transmissibility, and
- another nonnegative measurable real-valued function  $t \mapsto k(t)$  defined on  $\mathbb{R}_+$ , whose aim is to model how *individual-level* infectiousness varies over time.

In turn, for  $t \geq \sigma$ , the quantity defined as

$$r^\sigma(t) := \rho(t)k(t - \sigma)$$

can be interpreted as the infinitesimal rate at time  $t$  of new individuals (e.g. infectious cases) generated by a single individual originally born (e.g. infected) at time  $\sigma$ .

Finally, we introduce the unit rate homogeneous Poisson process denoted  $(\Phi(t))_{t \geq 0}$  which is assumed here to be independent of the lifetime random variable  $L$ . We now define the reproduction process  $N_{IP}(t; \sigma)$  as

$$N_{IP}(t; \sigma) := \begin{cases} \Phi\left(\int_0^t \rho(s + \sigma)k(s)ds\right), & t < L, \\ \Phi\left(\int_0^L \rho(s + \sigma)k(s)ds\right), & t \geq L. \end{cases} \quad (1.6)$$

Then it can be said that the so defined counting process  $N_{IP}(t; \sigma)$  and the lifelength random variable  $L$ , together fully determine a *time-varying Inhomogeneous Poisson (TV-IP) process*. This process can usefully be applied in the context of infectious disease modelling, whereby individuals can give birth at random points during their lifetime according to a Poisson process, whose time-varying rate depends both on a population-level factor of transmissibility and on an individual level of infectiousness.

## 1.5 Regularity conditions

Let us first recall the general definition of a regular branching process.

**Definition 1.5.1** (Regular branching process). A branching process  $(Z(t; \cdot))_{t \geq 0}$  defined on a probability space  $(\Omega, \Sigma, \mathbb{P})$  is said to be *regular* if it does not explode in finite time almost surely (a.s.), in other words if for any realisation  $\omega \in \Omega$ ,

$$\mathbb{P}(Z(t; \omega) < \infty) = 1, \quad \text{for all } 0 \leq t < \infty.$$

In this section, sufficient conditions are determined to ensure more specifically that a TV-CMJ branching process  $(Z^\chi(t; \cdot))_{t \geq 0}$  defined as per section 1.4 on the space  $(\Omega, \Sigma, \mathbb{P}^\sigma)$  is regular. Based on the above definition, we want to guarantee the well-definiteness of the TV-CMJ process through the following condition

$$\mathbb{P}^\sigma(Z^\chi(t; \omega) < \infty) = 1, \quad \text{for all } 0 \leq t < \infty, \quad \omega \in \Omega.$$

### 1.5.1 Classical regular CMJ branching process

Crump and Mode provide a definition of a *regular* generalised CMJ branching process  $Z^\chi(t)$  defined on a probability space  $(\Omega, \mathbb{P})$ , in the traditional case where  $\mathbb{P}$  does not depend on the time of birth  $\sigma$  of the ancestor. Their definition of regularity involves the largest possible quantity counted by individual characteristic, namely the one associated with the random characteristic  $\chi^B(t) := \mathbf{1}_{[\sigma, \infty)}(t)$  introduced in equation (1.5), which records the total number of individuals  $Z^B(t)$  that have cumulatively been born by time  $t \geq 0$ .

**Definition 1.5.2** (Regular CMJ process [CM68]). A generalised branching process  $Z^\chi(t)$ , associated with any individual characteristic  $\chi(t)$ , and defined in a probability space  $(\Omega, \mathbb{P})$ , is called *regular* if

$$\mathbb{P}(Z^B(t) < \infty) = 1, \quad \forall t \geq 0,$$

where  $Z^B(t) = \sum_{u \in \mathcal{U}} \mathbf{1}_{[0, \infty)}(t - \sigma_u)$  represents the cumulative number of individuals born by time  $t \geq 0$ .  $L$  is the lifelength random variable and  $\sigma_u$  is the time of birth of individual  $x$  in the population  $\mathcal{U}$ .

*Remark.* Note that to show regularity (non-explosion in finite time) of a TV-CMJ process on  $(\Omega, \mathcal{S}, \mathbb{P}^\sigma)$ , it suffices in our case to show that  $\mathbb{E}(Z^B(t)) < \infty$ . Indeed, recall that the process is nonnegative. If we were to assume (by contradiction) that  $\mathbb{P}(Z^B(t) < \infty) < 1$ , then there would exist a subset  $A \subset \Omega$  of strictly positive measure on which  $Z^B(t) = \infty$ . By the non-negativity of the process, we would then deduce that  $\mathbb{E}(Z^B(t)) = \infty$ . Thus if  $\mathbb{E}(Z^B(t)) < \infty$  then  $\mathbb{P}(Z^B(t) < \infty) = 1$ . In effect, it can be said that although  $L^1$  convergence does not necessarily imply a.s. convergence, the conclusion is different when it comes to finiteness. In fact, the finiteness of the expectation is sufficient to deduce a.s. finiteness. Consequently, and also as the proof seems easier to handle, rather than trying to prove directly that  $\mathbb{P}(Z^B(t) < \infty) = 1, \forall t \geq 0$ , we will show instead that  $\mathbb{E}(Z^B(t)) < \infty$  to ensure non-explosion in finite time.

Now recall that Crump and Mode [CM68] define the expectation of the reproduction process under a time-invariant probability measure  $\mathbb{P}$  for a CMJ branching process as  $\Lambda(t) := \mathbb{E}(N(t)), t \geq 0$ , and the authors prove that it is sufficient to assume that

$$\forall t \geq 0, \Lambda(t) < \infty \text{ and } \Lambda(0) < 1, \quad (1.7)$$

to ensure regularity of a CMJ process.

In parallel, but in the more specific context of a BH process, Sevast'yanov [Sev67] introduces another regularity condition based on the generating function of the process. Recall that the latter is defined for  $t \in [0, \infty)$  and  $s \in (-1, 1)$  by

$$F(t; s) := \sum_{n=0}^{\infty} \mathbb{P}(Z^{\text{Pr}}(t) = n) s^n.$$

Denoting  $F(t; 1) := \lim_{s \uparrow 1} F(t; s)$ , Sevast'yanov [Sev67] calls a BH branching process regular if the generating function verifies the following condition

$$F(t; 1) = 1. \quad (1.8)$$

Although these respective regularity conditions (1.7) and (1.8) may seem slightly different at first glance, it can be easily checked that in their respective contexts (i.e. classical CMJ or BH), they are equivalent to each other. Indeed, Sevast'yanov's condition (1.8) is equivalent to assuming that

$$Z^{\text{Pr}}(t) < \infty \quad \text{a.s.} \quad t \geq 0.$$

But in the specific case of a BH process, where individuals give birth at time of death, the above condition is itself equivalent to

$$Z^B(t) < \infty \quad \text{a.s.} \quad t \geq 0,$$

which is exactly the condition provided by Crump and Mode.

*Remark* (Probability of explosion). Noting that it is always true that  $F(t; 1) \leq 1$ , then when  $F(t; 1) < 1$ , the positive quantity  $q(t) := 1 - F(t; 1) > 0$  can be interpreted as *the probability*

of explosion, in which event infinitely many particles are a.s. generated in a finite time interval.

### 1.5.2 Time-varying CMJ branching process

We rely on the previous definition of a regular CMJ process to obtain sufficient regularity conditions for a TV-CMJ branching process. However, the proof of [CM68, Th.4.1] establishing sufficient regularity conditions for a CMJ branching process is not directly adaptable to a TV-CMJ process. In fact, a direct adaptation of the proof of [CM68, Th.4.1] would involve invoking the independence of random childbearing ages, whereas this independence does not hold any more in the time-varying case, since we need to account for the fact that the behaviour of any offspring depends on the time of birth of its parent.

Despite the fact that the regularity conditions found in [CM68] cannot be directly adapted to the time-varying case, we have found that we can nevertheless adapt the regularity proof of Edler [Edl78, Th.2.1], who happens to treat the case of generation-dependent CMJ processes and who also eventually relies on [CM68, Lemma 4.1 and 4.2].

To do so, we introduce a notation similar to [CM68] and [Edl78], albeit adapted to the time-varying case. Consider an individual  $u = (u_1 \dots u_n) \in \mathbb{N}^n$ . Recall that  $\sigma_{u_1 \dots u_n}$  represents its time of birth, and for any  $t \geq 0$  and  $k \in \mathbb{N}$  the CDFs of the childbearing ages  $F_k^\sigma(t) := \mathbb{P}^\sigma(\tau(k) \leq t)$  were defined in equation (1.2). We introduce a set of assumptions that will prove sufficient to ensure regularity.

**Assumptions** (Sufficient regularity conditions for a TV-CMJ branching process).

(R-1) For any  $k \in \mathbb{N}$ , there exists a CDF defined on  $\mathbb{R}_+$ , independent of  $\sigma$ , given by  $t \mapsto \tilde{F}_k(t)$ , and which upper-bounds  $F_k^\sigma(t)$  for any  $\sigma, t \geq 0$ . In other words  $\tilde{F}_k(t)$  is a nondecreasing, right-continuous function such that  $\tilde{F}_k(0) = 1$ ,  $\lim_{t \rightarrow \infty} \tilde{F}_k(t) = 1$ , and

$$F_k^\sigma(t) \leq \tilde{F}_k(t), \quad \forall \sigma, t \geq 0.$$

(R-2) Given (R-1), there exists a nonnegative real-valued, nondecreasing, right-continuous function  $t \mapsto \tilde{\Lambda}(t)$ , defined on  $\mathbb{R}_+$ , such that  $\tilde{\Lambda}(0) < 1$ ,  $\tilde{\Lambda}(t) < \infty$ , for  $t \geq 0$  and

$$\sum_{k=1}^{\infty} \tilde{F}_k(t) \leq \tilde{\Lambda}(t), \quad t \geq 0.$$

Before showing that conditions (R-1) and (R-2) are sufficient to ensure that a TV-CMJ branching process is regular, we prove the following intermediary lemma.

**Lemma 1.5.1.** Assume condition (R-1) holds. Consider an individual  $u = \langle u_1, \dots, u_k \rangle \in \mathbb{N}^k$ ,  $k \in \mathbb{N}$  and let  $\sigma_{u_1 \dots u_k}$  denote its random time of birth. Then there exist  $k$  nondecreasing, right-continuous functions denoted  $(\tilde{F}_{u_j})_{j=1, \dots, k}$ , independent of  $\sigma$ , such that for any  $\sigma \geq 0$  and  $t \geq 0$

$$\mathbb{P}^\sigma(\sigma_{u_1 \dots u_k} \leq t) \leq \tilde{F}_{u_1} * \dots * \tilde{F}_{u_k}(t), \quad (1.9)$$

where the operator  $*$  is the convolution operator defined such that for any real-valued non-

negative Lebesgue measurable functions  $f, g$  we have

$$f * g(t) := \int_0^t f(s)g(t-s)ds.$$

*Proof.* Take any  $\sigma \geq 0$ . We prove by induction on  $k \in \mathbb{N}$  that, under assumption (R-1), inequality (1.9) holds. Let us first show that it holds for  $k = 1$ . Indeed, observe that

$$\begin{aligned} \mathbb{P}^\sigma(\sigma_{u_1} \leq t) &= \mathbb{P}^\sigma(\sigma + \tau(u_1) \leq t), \\ &= \mathbb{P}^\sigma(\tau(u_1) \leq t - \sigma), \\ &= F_{u_1}^\sigma(t - \sigma), \\ &\leq F_{u_1}^\sigma(t), \end{aligned}$$

where  $t \mapsto F_{u_1}^\sigma(t) := \mathbb{P}^\sigma(\tau(u_1) \leq t)$  is non-decreasing. However, thanks to assumption (R-1), we know that there exists a CDF  $\tilde{F}_{u_1}(t)$  independent of  $\sigma$  such that

$$F_{u_1}^\sigma(t) \leq \tilde{F}_{u_1}(t).$$

We therefore have shown that

$$\mathbb{P}^\sigma(\sigma_{u_1} \leq t) \leq \tilde{F}_{u_1}(t),$$

which is the inequality we wished to prove at rank  $k = 1$ . Classically we now assume that inequality (1.9) holds at rank  $k - 1$ . Recalling the simple relationship between times of birth and childbearing ages, i.e.  $\sigma_{u_1 \dots u_k} = \sigma_{u_1 \dots u_{k-1}} + \tau(u_k)$ , and applying the Law of Total Probability, we get

$$\begin{aligned} \mathbb{P}^\sigma(\sigma_{u_1 \dots u_k} \leq t) &= \mathbb{P}^\sigma(\sigma_{u_1 \dots u_{k-1}} + \tau(u_k) \leq t), \\ &= \int_{\mathbb{R}^+} \mathbb{P}^\sigma(\tau(u_k) \leq t - y \mid y = \sigma_{u_1 \dots u_{k-1}}) d\mathbb{P}^\sigma(y). \end{aligned} \quad (1.10)$$

Considering a TV-CMJ process initiated by an individual born at time  $\sigma_{u_1 \dots u_{k-1}}$ , the self-similarity property of the branching process allows us to write

$$\mathbb{P}^\sigma(\tau(u_k) \leq t - y \mid y = \sigma_{u_1 \dots u_{k-1}}) = \mathbb{P}^{\sigma_{u_1 \dots u_{k-1}}}(\tau(u_k) \leq t - y) =: F_{u_k}^{\sigma_{u_1 \dots u_{k-1}}}(t - y), \quad (1.11)$$

where, on the right-hand side above, we recognised the CDF  $F_{u_k}^{\sigma_{u_1 \dots u_{k-1}}}(t)$  of the childbearing age of the ancestor born at time  $\sigma_{u_1 \dots u_{k-1}}$ . Next, given that Assumption (R-1) is valid for any  $\sigma \geq 0$ , we deduce that there exists another CDF  $\tilde{F}_{u_k}(t)$  independent of  $\sigma_{u_1 \dots u_{k-1}}$  such that

$$F_{u_k}^{\sigma_{u_1 \dots u_{k-1}}}(t) \leq \tilde{F}_{u_k}(t), \quad \forall t \geq 0. \quad (1.12)$$

Combining (1.10), (1.11) and (1.12), we obtain

$$\mathbb{P}^\sigma(\sigma_{u_1 \dots u_k} \leq t) \leq \int_{\mathbb{R}^+} \tilde{F}_{u_k}(t - y) d\mathbb{P}^\sigma(y).$$

We now apply the induction hypothesis at rank  $k - 1$  that tells us that there exist  $k - 1$  CDFs, namely  $\tilde{F}_{u_1}, \dots, \tilde{F}_{u_{k-1}}(t)$  such that

$$\mathbb{P}^\sigma(\sigma_{u_1 \dots u_{k-1}} \leq t) \leq \tilde{F}_{u_1} \star \dots \star \tilde{F}_{u_{k-1}}(t).$$

This inequality can be interpreted as a first-order stochastic dominance between CDFs. We can then apply Lemma (A.2.1) given that the function  $y \mapsto \tilde{F}_{u_k}(t - y)$  is non-increasing, i.e.

$$\int_{\mathbb{R}^+} \tilde{F}_{u_k}(t - y) d\mathbb{P}^\sigma(y) \leq \int_{\mathbb{R}^+} \tilde{F}_{u_k}(t - y) d\left(\tilde{F}_{u_1} \star \dots \star \tilde{F}_{u_{k-1}}\right)(y).$$

This latter inequality corresponds exactly to the desired result

$$\mathbb{P}^\sigma(\sigma_{u_1 \dots u_k} \leq t) \leq \tilde{F}_{u_1} \star \dots \star \tilde{F}_{u_k}(t).$$

□

We now apply the intermediary result provided by Lemma 1.5.1 to prove that assumptions (R-1) and (R-2) are sufficient to ensure that a TV-CMJ is well defined, in other words, that the process a.s. does not explode in finite time. More formally,

**Proposition 1.1.** *Let  $Z^\chi(t; \sigma)$  be a TV-CMJ process. Assume that conditions (R-1) and (R-2) are satisfied. Then  $Z^\chi(t; \sigma)$  is regular.*

*Proof.* Here we adapt simultaneously the regularity proofs of Crump and Mode [CM68] (their Lemma 4.1 and its proof are transcribed in Appendix A.5.1) and have also found inspiration from Edler [Edl78, Th.2.1]. Returning to the definition of regularity, we need to show that, when we take  $\chi(t) = B(t)$ , which is the random characteristic defined in Section 1.4.2 that leads to cumulative incidence, then  $\mathbb{P}^\sigma(Z^B(t; \sigma) < \infty) = 1$ .

We consider an individual  $u = (u_1 \dots u_k)$ , whose time of birth is denoted  $\sigma_u := \sigma_{u_1 \dots u_k}$ . The characteristic function  $\chi_u^B(t; \sigma_u) := \mathbf{1}_{\sigma_u \leq t}$  determines whether individual  $x$  has been born by time  $t \geq 0$ . We define the process counting the number of individuals born up to time  $t \geq 0$  as  $Z^B(t; \sigma) = \sum_{u \in \mathcal{U}} \chi_u^B(t; \sigma_u)$ , and compute its expectation

$$\begin{aligned} \mathbb{E}^\sigma(Z^B(t; \sigma)) &= \sum_{u \in \mathcal{U}} \mathbb{E}^\sigma(\mathbf{1}_{\sigma_u \leq t}), \\ &= \sum_{u \in \mathcal{U}} \mathbb{P}^\sigma(\sigma_u \leq t), \\ &= 1 + \sum_{k=1}^{\infty} \sum_{u_1=1}^{\infty} \dots \sum_{u_k=1}^{\infty} \mathbb{P}^\sigma(\sigma_{u_1 \dots u_k} \leq t). \end{aligned}$$

Assumption (R-1) holds, therefore we can apply Lemma 1.5.1 and obtain a first upper-bound, independent of  $\sigma$

$$\mathbb{E}^\sigma(Z^B(t; \sigma)) \leq 1 + \sum_{k=1}^{\infty} \sum_{u_1=1}^{\infty} \dots \sum_{u_k=1}^{\infty} \tilde{F}_{u_1} * \tilde{F}_{u_2} * \dots * \tilde{F}_{u_k}(t).$$

Finally, condition (R-2) gives us a second upper-bound

$$\mathbb{E}^\sigma(Z^B(t; \sigma)) \leq 1 + \sum_{k=1}^{\infty} \tilde{\Lambda}^{k*}(t),$$

where  $\tilde{\Lambda}^{k*}(t)$  represents the  $k$ -th convolution of  $\tilde{\Lambda}(t)$  with itself. This inequality has led us back to the (non-time-varying) classical CMJ process treated by Crump and Mode [CM68].

The last step of the proof is similar to their proof of regularity [CM68, Th 4.1.] and consists of applying directly [CM68, Lemma 4.1 ] to  $\tilde{\Lambda}(t)$ . We thus obtain

$$\sum_{k=1}^{\infty} \tilde{\Lambda}^{k*}(t) < \infty,$$

which implies that  $\mathbb{E}^{\sigma}(Z^B(t; \sigma)) < \infty$ , i.e. a sufficient condition of non-explosion.  $\square$

*Remark.* Observe that in effect the above proof relies on a simple technique by which the TV-CMJ process is upper-bounded by a non-time-varying process, which itself satisfies sufficient regularity conditions.

From now on, even if not systematically mentioned, conditions (R-1) and (R-2) will be assumed to hold to ensure that the TV-CMJ branching process under consideration is regular.

### 1.5.3 Time-varying Bellman-Harris branching process

We first briefly recall for clarity of exposition the basic properties of a TV-BH process introduced in Section 1.4.3. The number of children born from each individual in a TV-BH process is quantified by an exogenous point process  $(\xi_{BH}(t))_{t \geq 0}$ , observable at times of death determined by  $\sigma + L$ . This point reproduction process is itself determined by a pmf  $(q_t(k))_{k \in \mathbb{N}}$ , independent of  $\sigma$ , defined by  $\mathbb{P}^{\sigma}(\xi_{BH}(t) = k) = q_t(k)$ . Finally, we recall that  $L$  and  $\xi_{BH}(t)$  are independent of each other.

As already observed, childbearing ages are equal to  $\sigma + L$ . Consequently, imposing a specific regularity assumption in relation to their CDFs as in (R-1) would in effect be equivalent to making another assumption for the CDF  $G^{\sigma}(t)$  of the lifelength r.v. which is the *raison-d'être* of (R-2). To adapt to this specificity, our first assumption here relates more directly to the reproduction process  $\xi_{BH}(t)$ .

**Assumption** (Sufficient regularity conditions for a TV-BH process).

(BH-1) For any  $k \in \mathbb{N}$  and  $t \geq 0$ , the function  $t \mapsto \tilde{F}_k(t)$  defined on  $\mathbb{R}_+$  as

$$\tilde{F}_k(t) = \sup_{\sigma \geq 0} \mathbb{E}^{\sigma} \left( \mathbf{1}_{L \leq t} \mathbb{P}^{\sigma} \left( \xi_{BH}(\sigma + l) \geq k \right)_{l=L} \right)$$

has the properties of a CDF.

(BH-2) There exists a nonnegative real-valued, nondecreasing, and right-continuous function  $\tilde{G}(t)$ , independent of  $\sigma$ , such that  $\tilde{G}(0) < 1$ ,  $\tilde{G}(t) < \infty$  for any  $t \geq 0$  and

$$G^{\sigma}(t) \leq \tilde{G}(t), \quad t \geq 0.$$

*Remark* (Non-explosion at  $t = 0$ ). Note that (BH-2) implies that  $G^{\sigma}(0) \leq \tilde{G}(0) < 1$  and, therefore, given that  $L$  is nonnegative,  $\mathbb{P}^{\sigma}(L \leq 0) = \mathbb{P}^{\sigma}(L = 0) < 1$ , that is  $\mathbb{P}^{\sigma}(L = 0) \neq 1$ .

**Lemma 1.5.2.** A TV-BH branching process verifying conditions (BH-1) and (BH-2) is regular.

*Proof.* Recall that  $\tau(k) = L$  if  $\xi_{BH}(\sigma + L) \geq k$ , and otherwise  $\tau(k) = \infty$  by convention. Thus,

it is very direct to verify that

$$F_k(t) = \mathbb{P}^\sigma(\tau(k) \leq t) = \mathbb{E}^\sigma(\mathbf{1}_{L \leq t} \mathbb{P}^\sigma(\xi_{BH}(\sigma + L) \geq k)) \quad (1.13)$$

$$\leq \sup_{\sigma \geq 0} \mathbb{E}^\sigma \left( \mathbf{1}_{L \leq t} \mathbb{P}^\sigma \left( \xi_{BH}(\sigma + l) \geq k \right)_{l=L} \right) =: \tilde{F}_k(t), \quad (1.14)$$

which guarantees that condition (R-1) is satisfied. Then, recalling the notation of the pmf  $q_t(k) := \mathbb{P}^\sigma(\xi_{BH}(t) = k)$ , in order to verify (R-2), we write successively

$$\begin{aligned} \sum_{k=0}^{\infty} \tilde{F}_k(t) &= \sum_{k=0}^{\infty} \sup_{\sigma \geq 0} \mathbb{E}^\sigma \left( \mathbf{1}_{L \leq t} \mathbb{P}^\sigma \left( \xi_{BH}(\sigma + l) \geq k \right)_{l=L} \right), \\ &\leq \sup_{\sigma \geq 0} \mathbb{E}^\sigma \left( \mathbf{1}_{L \leq t} \sum_{k=0}^{\infty} \mathbb{P}^\sigma \left( \xi_{BH}(\sigma + l) \geq k \right)_{l=L} \right), \\ &\leq \sup_{\sigma \geq 0} \mathbb{E}^\sigma \left( \mathbf{1}_{L \leq t} \sum_{k=0}^{\infty} k \mathbb{P}^\sigma \left( \xi_{BH}(\sigma + l) = k \right)_{l=L} \right), \\ &\leq G^\sigma(t) \sum_{k=0}^{\infty} k q_t(k), \\ &\leq \tilde{G}(t), \end{aligned}$$

where we applied Tonelli's theorem in the second line and (BH-2) in the last inequality.

It now suffices to define  $\tilde{\Lambda}(t) := \tilde{G}(t)$  which is a real-valued, nonnegative, nondecreasing, right-continuous function defined on  $\mathbb{R}_+$ , such that  $\tilde{\Lambda}(0) < 1$ , and for any  $t \geq 0$ ,  $\tilde{\Lambda}(t) < \infty$ , and

$$\sum_{k=1}^{\infty} \tilde{F}_k(t) \leq \tilde{\Lambda}(t).$$

Thus, (R-2) is satisfied and we can conclude that a TV-BH branching process under assumptions (BH-1) and (BH-2) also verifies (R-1) and (R-2) and therefore, by Theorem 1.1, such a TV-BH process is regular.

□

#### 1.5.4 Time-varying Inhomogeneous Poisson branching process

For any  $t \geq 0$ , let  $\Phi(t)$  represent the unit rate homogeneous Poisson process. Recall the definition of the reproduction process  $N_{IP}(t)$  of a time-varying inhomogeneous Poisson branching process as per equation (1.6)

$$N_{IP}(t; \sigma) := \begin{cases} \Phi \left( \int_0^t \rho(s + \sigma) k(s) ds \right), & t < L, \\ \Phi \left( \int_0^L \rho(s + \sigma) k(s) ds \right), & t \geq L. \end{cases}$$

We introduce the following two conditions that will prove to be sufficient for such a process to not explode in finite time almost surely.

**Assumptions** (Sufficient regularity conditions for a TV-IP branching process).

(IP-1) The transmissibility function  $t \mapsto \rho(t)$  is bounded on  $\mathbb{R}_+$ , i.e. there exists a nonnegative finite real-valued constant  $\bar{\rho} \in [0, \infty)$  such that

$$\bar{\rho} := \sup_{t \geq 0} \rho(t) < \infty,$$

(IP-2) The function  $t \mapsto k(t)$  representing individual infectiousness is bounded on  $\mathbb{R}_+$ , i.e. there exists a nonnegative finite real-valued constant  $\bar{k} \in [0, \infty)$  such that

$$\bar{k} := \sup_{t \geq 0} k(t) < \infty.$$

We then define

$$R := \bar{\rho}\bar{k} < \infty,$$

and introduce a linear-time-changed Poisson process based on the unit rate homogeneous Poisson process such that

$$\bar{\Phi}(t) := \Phi(Rt).$$

We now observe that we can simply upper-bound the expected reproduction process  $\Lambda_{IP}(t; \sigma)$  of a TV-IP branching process by appealing to the independence of the lifelength random variable  $L$  with the unit rate  $\Phi(t)$ . Indeed, we get

$$\begin{aligned} \Lambda_{IP}(t; \sigma) &= \mathbb{E}^\sigma(N_{IP}(t)), \\ &= \mathbb{E}^\sigma\left(\Phi\left(\int_0^t \rho(u + \sigma)k(u)du\right)\mathbf{1}_{\{t < L\}}\right) + \mathbb{E}^\sigma\left(\Phi\left(\int_0^L \rho(u + \sigma)k(u)du\right)\mathbf{1}_{\{t \geq L\}}\right), \\ &\leq \mathbb{E}^\sigma\left(\Phi\left(\int_0^t \rho(u + \sigma)k(u)du\right)\mathbf{1}_{\{t < L\}}\right) + \mathbb{E}^\sigma\left(\Phi\left(\int_0^t \rho(u + \sigma)k(u)du\right)\mathbf{1}_{\{t \geq L\}}\right), \\ &\leq \mathbb{E}^\sigma\left(\Phi\left(\int_0^t \rho(u + \sigma)k(u)du\right)\right), \\ &\leq \mathbb{E}^\sigma(\Phi(Rt)) = \mathbb{E}^\sigma(\bar{\Phi}(t)). \end{aligned}$$

Let us now consider a simple homogeneous Poisson process  $\bar{\Phi}(t) = \Phi(Rt)$  with time-invariant rate  $R$  and a branching process whose reproduction process is given by  $\bar{\Phi}(t)$ . Such a branching process with homogeneous rate is known to be regular. However, in addition, observing that that  $\Phi(t)$  is a nondecreasing function, we deduce that the intensity given by  $\bar{\Phi}(t)$  upper bounds the intensity of a TV-IP branching process whose reproduction process is  $N_{IP}(t; \sigma)$ . Therefore, regularity of this dominating time-invariant process ensures regularity of the TV-IP branching process.

Now that the regularity conditions for a TV-IP branching process have been clarified, we can give a more precise expression for the associated expected reproduction process  $\Lambda_{IP}(t; \sigma)$ , although to do so, we need to add another condition with respect to the existence of a density of the lifelength CDF.

**Assumption** (Additional assumption for the CDF in a TV-IP case).

(G-density) The CDF  $G^\sigma(t)$  of the lifelength random variable admits a measurable density

$g^\sigma(t)$  such that

$$G^\sigma(t) = \int_0^t g^\sigma(u) du, \quad t \geq 0.$$

**Lemma 1.5.3.** *Assume condition (G-density) holds as well as regularity conditions (IP-1) and (IP-2). Then the expected reproduction process of a regular TV-IP branching process is given by*

$$\Lambda_{IP}(t; \sigma) = \int_0^t \rho(v + \sigma) k(v) \overline{G}^\sigma(v) dv,$$

where  $\overline{G}^\sigma(t) := 1 - G^\sigma(t)$  represents the survival function.

*Proof.* Recall

$$\Lambda_{IP}(t; \sigma) = \mathbb{E}^\sigma \left( \Phi \left( \int_0^t \rho(v + \sigma) k(v) dv \right) \mathbf{1}_{\{t < L\}} \right) + \mathbb{E}^\sigma \left( \Phi \left( \int_0^L \rho(v + \sigma) k(v) dv \right) \mathbf{1}_{\{t \geq L\}} \right). \quad (1.15)$$

We rewrite the first term on the right-hand side above by relying on the independence between the unit rate homogeneous Poisson process  $\Phi(t)_{t \geq 0}$  and the lifelength r.v.  $L$ . Indeed,

$$\begin{aligned} \mathbb{E}^\sigma \left( \Phi \left( \int_0^t \rho(v + \sigma) k(v) dv \right) \mathbf{1}_{\{t < L\}} \right) &= \mathbb{E}^\sigma \left( \Phi \left( \int_0^t \rho(v + \sigma) k(v) dv \right) \right) \mathbb{P}(t < L), \\ &= \overline{G}^\sigma(t) \int_0^t \rho(v + \sigma) k(v) dv. \end{aligned}$$

Now, integrating by parts the right-hand side term in the above equation, we get

$$\begin{aligned} \overline{G}^\sigma(t) \int_0^t \rho(v + \sigma) k(v) dv &= \int_0^t \int_0^s \rho(v + \sigma) k(v) dv d\overline{G}^\sigma(s) + \int_0^t \overline{G}^\sigma(s) \rho(s + \sigma) k(s) ds \\ &= - \int_0^t \int_0^s \rho(v + \sigma) k(v) dv g^\sigma(s) ds + \int_0^t \rho(s + \sigma) k(s) \overline{G}^\sigma(s) ds, \end{aligned}$$

since it was assumed that  $G^\sigma(\cdot)$  has a density and  $\frac{d\overline{G}^\sigma(u)}{du} = -\frac{dG^\sigma(u)}{du} = -g^\sigma(u)$ . Next applying the Law of Total Expectation on the second term in (1.15), we get

$$\begin{aligned} \mathbb{E}^\sigma \left( \Phi \left( \int_0^L \rho(v + \sigma) k(v) dv \right) \mathbf{1}_{\{t \geq L\}} \right) &= \mathbb{E}^\sigma \left( \mathbb{E}^\sigma \left( \Phi \left( \int_0^L \rho(v + \sigma) k(v) dv \right) \mathbf{1}_{\{t \geq L\}} \middle| L \right) \right), \\ &= \mathbb{E}^\sigma \left( \mathbb{E}^\sigma \left( \Phi \left( \int_0^\ell \rho(v + \sigma) k(v) dv \right) \right)_{\ell=L} \mathbf{1}_{\{t \geq L\}} \right), \\ &= \int_0^t \int_0^s \rho(v + \sigma) k(v) dv g^\sigma(s) ds. \end{aligned}$$

This indeed simplifies to

$$\Lambda_{IP}(t; \sigma) = \int_0^t \rho(v + \sigma) k(v) \overline{G}^\sigma(v) dv,$$

which is the expected form for the reproduction process of a regular TV-IP process with density.  $\square$

## CHAPTER 2

# Integral equations for a time-varying CMJ branching process

One of the primary objectives of this chapter is to provide a rigorous mathematical foundation to the incidence renewal equation commonly used in epidemiology. We do this by connecting two worlds: the world of infectious disease modelling and the world of generalised branching processes. Indeed, we assume that an infection or more generally an epidemic, at least at its onset or in its early phases, can be modelled as a branching process whereby individuals (e.g. infectious cases) behave identically and independently of each other. This modelling approximation was first mentioned in the work of Bartlett [Bar78] and Kendall [Ken56] and given general mathematical grounding, at least asymptotically, through the works of Ball [Bal83], Ball and Donnelly [BD95] and Metz [Met78] who showed that an epidemic process converges to a limiting branching process when the number of susceptible individuals in the population tends to infinity. More widely, this connection has often been made in practice and in the literature (see, e.g. [KM27; Bar67; Jac10; FKG03; LM23; Hau+15; VM11; Ish05]). In particular, we cite Ball et al. [Bal+14] who explain that the process representing the number of infectious individuals in an epidemic model can be satisfactorily described mathematically by a branching process if we can assume that the population is homogeneously mixing and that the number of infected people is small relative to the total size of susceptible individuals in the population.

To mathematically derive the incidence renewal equation used in infectious disease modelling, our starting point will be the TV-CMJ process as introduced in the previous chapter. We use this former modelling framework to derive an integral renewal-type equation for the first moment of a TV-CMJ process. Next, based on an existence and uniqueness result of a generic integral equation derived from a TV-CMJ (see Appendix B.2), and relying in parallel on the manipulation of generalised functions (see Appendix B.3), we successively derive an integral equation for the cumulative number of births of a TV-CMJ and then for its expected birth rate. We are then in a position to compare, on one hand, the later equation related to the dynamics of the expected birth rate of a TV-CMJ process with, on the other hand, a time-varying extension of the incidence equation commonly applied heuristically in epidemiology.

We finally prove that, once discretised according to Algorithm I in [Pak+23], these two apparently different integral renewal-type equations converge towards each other, with an error rate of order  $\mathcal{O}(n^{-1})$ . In effect, it can thus be said that this convergence result provides

mathematical foundation to the celebrated incidence renewal equation so commonly used in practice in epidemiology and in infectious disease modelling.

Finally, we also provide an alternative perspective based on the duality forward-backward of the previously derived integral equations, whereby the incidence renewal equation of epidemiology can be seen as the forward formulation of the integral equation of the expected birth rate of a TV-CMJ process.

## 2.1 First-moment integral equation of a TV-CMJ process

A classical approach introduced by Harris [Har63] to study the dynamics and long-term behaviour of an age-dependent branching process consists of deriving renewal-like integral equations for various quantities of interest related to the branching process and its moments. In particular, Bellman and Harris in the case of binary splitting [BH52] and [BH48], and then Harris in a more general case [Har63, Th.7.1], derived the well-known Bellman-Harris integral equation that governs the evolution of the generating function of an age-dependent branching process. The renewal-like integral equation derived by Bellman and Harris for the generating function is then shown to extend to other functionals related to the process and in particular to its first and second moment. Harris [Har63, Ch.6] provides a detailed derivation of these equations. Although this is not the focus of this chapter, it should be noted that the study of integral equations derived from the dynamics of branching processes has contributed to many results with respect to their asymptotic behaviour. We cite here, among many more, Sevast'yanov [Sev68], but also Athreya and Ney [AN04] for a classical reference on this topic, and Jagers [Jag69b] who in particular studies the almost sure convergence of an age-dependent branching process by using results of renewal theory (see also, e.g. [Dur71; Pak72; Chi72]). These connections between generalised branching processes and renewal-like integral equations have also proven useful from an application perspective. We can cite Tratnyanov [Tra16] who studies population age structures using a numerical approach to solving renewal equations.

The renewal aspect of all these integral equations is invariably based on the same principle originally called *the first generation principle* by Harris [Har63, Th.6.1, Ch.V]. This principle states that the branching process is the superposition of the processes initiated by the first generation progeny originating from the ancestor. This first generation principle, as it applies more generally to any succession of generations, is also commonly summarised as *branching property* and reflects in effect a self-similarity behaviour between individuals in a branching population.

In the more specific context of an epidemic infection, Kimmel [Kim83] relies on the first generation principle when he considers a branching process to model the development of an infectious disease where each new infectious case can be viewed as the ancestor of its own subprocess. The branching property can be summarised as follows. The number of infections alive within the whole process at time  $t$  is equal to the sum of the numbers of infections coming from each of the subprocesses.

In this section, mirroring the derivation of Pakkanen et al. [Pak+23, Section 2.2], we apply the *first generation principle* to the TV-CMJ process in order to single out the first generation of progeny in the computation of the sum of characteristic functions making up  $Z(t; \sigma)$ . This technique classically leads to a generic integral equation governing the first moment of a TV-CMJ.

**Lemma 2.1.1.** *Let  $Z(t; \sigma) := \sum_{x \in \mathcal{U}} \chi(t - \sigma_x)$  be a TV-CMJ branching process as per Definition*

1.4.2. Denote  $f(t; \sigma) := \mathbb{E}^\sigma(Z(t; \sigma))$  its first moment and  $\Lambda(t; \sigma) := \mathbb{E}^\sigma(N(t))$ , its expected reproduction function defined in equation (1.4). Assume that the sufficient regularity conditions (R-1) and (R-2) hold. Then  $f(t; \sigma)$  satisfies the following renewal-like integral equation

$$f(t; \sigma) = \mathbb{E}^\sigma(\chi(t - \sigma)) + \int_{(0, t-\sigma]} f(t; \sigma + s) d\Lambda(s; \sigma), \quad t \geq \sigma \geq 0. \quad (2.1)$$

*Proof.* Recall that by construction of the time-varying probability space  $(\Omega, \mathcal{S}, \mathbb{P}^\sigma)$  underlying a TV-CMJ process, we have assumed that the initial probability measure  $\pi^\sigma$  supports the reproduction point process  $\xi(t)$  whose points represent the sequence of childbearing ages  $\tau(k)$ . We can assume without loss of generality that these points are ordered in a monotonic (non-decreasing) manner, which in turn implies that any first generation individual indexed by  $k \in \mathbb{N}$  can be chronologically labelled thanks to the associated childbearing age  $\tau(k)$  of the ancestor. Note that if and when offspring are born at the same time, an identical childbearing age will be associated as many times as there are offspring born at that time. Additionally, and by convention, we set  $\tau(k) := \infty$  if the ancestor's  $k$ -th child is not realised. We recall that the times of birth of the first generation progeny follow the obvious relationship

$$\sigma_k = \sigma + \tau(k), \quad k \in \mathbb{N}.$$

We denote  $\mathcal{U}_k \subset \mathcal{U}$  the population subset initiated by the first generation individual  $k \in \mathbb{N}$ . (Trivially it is assumed that  $\mathcal{U}_k = \emptyset$  if  $k$  is not realised.) When  $k$  is realised though, the set  $\mathcal{U}_k$  includes individual  $k$  and all its progeny over the successive generations. The branching property applied to the first generation therefore leads directly to the decomposition

$$\begin{aligned} Z(t; \sigma) &= \chi(t - \sigma) + \sum_{\substack{k \in \mathbb{N} \\ \sigma_k \leq t}} \sum_{x \in \mathcal{U}_k} \chi(t - \sigma_x), \\ &= \chi(t - \sigma) + \sum_{\substack{k \in \mathbb{N} \\ \sigma_k \leq t}} Z_k(t; \sigma), \end{aligned}$$

where we introduced for any  $k \in \mathbb{N}$  the branching processes  $Z_k(t; \sigma) := \sum_{x \in \mathcal{U}_k} \chi(t - \sigma_x)$ , initiated by the first generation individuals labelled  $k \in \mathbb{N}$ .

As already mentioned in chapter 1, if we condition the processes  $(Z_k(t; \sigma))_{k \in \mathbb{N}}$  by the times of births  $(\sigma_k)_{k \in \mathbb{N}}$ , the branching property implies that  $(Z_k(t; \sigma))_{k \in \mathbb{N}}$  are conditionally mutually independent and distributed identically to processes denoted  $\tilde{Z}(t; \sigma_k)$ , whereby  $(\tilde{Z}(\cdot, \sigma))$  is itself an identical copy of  $Z(\cdot, \sigma)$ . In other words,

$$(Z_k(t; \sigma) | \sigma_1, \sigma_2, \dots) \sim \tilde{Z}(t; \sigma_k) \sim Z(t; \sigma_k).$$

We now take expectation in the equation giving the above decomposition of  $Z(t; \sigma)$ , and simply apply the Tower property\*, relying on the regularity of the process, which means that sums

---

\*Recall that the Tower Property states that  $\mathbb{E}(X) = \mathbb{E}(\mathbb{E}(X|Y))$  whenever the expectations are properly defined.

and expectations can be interchanged or on the Tonelli-Fubini theorem (see Appendix A.3)

$$\begin{aligned}
f(t; \sigma) &= \mathbb{E}^\sigma(Z(t; \sigma)) = \mathbb{E}^\sigma(\chi(t - \sigma)) + \mathbb{E}^\sigma \left( \sum_{k: \sigma_k \leq t} Z_k(t; \sigma) \right), \\
&= \mathbb{E}^\sigma(\chi(t - \sigma)) + \mathbb{E}^\sigma \left( \mathbb{E}^\sigma \left( \sum_{\substack{k \geq 1 \\ \sigma_k \leq t}} Z_k(t; \sigma) \middle| \sigma_1, \sigma_2, \dots \right) \right), \\
&= \mathbb{E}^\sigma(\chi(t - \sigma)) + \mathbb{E}^\sigma \left( \sum_{\substack{k \geq 1 \\ \sigma_k \leq t}} \mathbb{E}^\sigma \left( Z_k(t; \sigma) \middle| \sigma_1, \sigma_2, \dots \right) \right), \\
&= \mathbb{E}^\sigma(\chi(t - \sigma)) + \mathbb{E}^\sigma \left( \sum_{\substack{k \geq 1 \\ \sigma_k \leq t}} f(t; \sigma_k) \right). \tag{2.2}
\end{aligned}$$

Recall that, by construction, the reproduction counting process  $N(t)$  possesses non-negative jumps of integer size, which we denote

$$\Delta N(t) := N(t) - \lim_{v \downarrow t^-} N(v) \geq 0.$$

We can thus transform the discrete countable sum on the right hand-side of (2.2) into a continuous one,

$$\begin{aligned}
\mathbb{E}^\sigma \left( \sum_{\substack{k \geq 1 \\ \sigma_k \leq t}} f(t; \sigma_k) \right) &= \mathbb{E}^\sigma \left( \sum_{v \in (\sigma, t]} f(t; \sigma) \Delta N(v - \sigma) \right), \\
&= \mathbb{E}^\sigma \left( \sum_{s \in (0, t - \sigma]} f(t; s + \sigma) \Delta N(s) \right), \\
&= \sum_{s \in (0, t - \sigma]} f(t; s + \sigma) \mathbb{E}^\sigma(\Delta N(s)), \\
&= \int_{(0, t - \sigma]} f(t; s + \sigma) \mathbb{E}^\sigma(dN(s)), \\
&= \int_{(0, t - \sigma]} f(t; s + \sigma) d\Lambda(s; \sigma),
\end{aligned}$$

where, in the third line above we have been able to interchange the expectation and the sum operators because the number of jumps is actually finite thanks to the regularity assumption of the process. In summary, we have shown that the first moment represented by the function  $(t; \sigma) \mapsto f(t; \sigma) := \mathbb{E}(Z(t; \sigma))$  satisfies for any  $t \geq \sigma \geq 0$ ,

$$f(t; \sigma) = \mathbb{E}^\sigma(\chi(t - \sigma)) + \int_{(0, t - \sigma]} f(t; s + \sigma) d\Lambda(s; \sigma),$$

which is the required integral equation given in (2.1).  $\square$

Next, we briefly provide a couple of specific examples of the above renewal-like integral equation using the characteristic functions  $\chi^B(t; \sigma)$  and  $\chi^{Pr}(t; \sigma)$  introduced in Section 1.4.2. We also apply each of these equations to the TV-BH and the TV-IP branching processes described in Sections 1.5.3 and 1.5.4.

### Integral equation for the cumulative number of births

We can apply the previous result to the random characteristic  $\chi^B(t - \sigma) := \mathbf{1}_{[0, \infty)}(t - \sigma)$  and obtain

$$B(t; \sigma) = \begin{cases} 0, & t < \sigma, \\ \mathbf{1}_{[0, \infty)}(t - \sigma) + \int_{(0, t - \sigma]} B(t; s + \sigma) d\Lambda(s; \sigma), & t \geq \sigma. \end{cases} \quad (2.3)$$

In the particular case of a regular TV-BH process, assuming conditions (BH-1) and (BH-2) hold, the equation becomes

$$B_{BH}(t; \sigma) = \mathbf{1}_{[0, \infty)}(t - \sigma) + \int_{(0, t - \sigma]} B_{BH}(t; s + \sigma) R(\sigma + s) dG^\sigma(s),$$

where trivially for  $t < \sigma$ ,  $B_{BH}(t; \sigma) = 0$ . Similarly, assuming conditions (IP-2) and (IP-1) hold (that is, the TV-IP branching process is regular), then by Lemma 1.5.3,  $d\Lambda(u; \sigma) = \rho(u + \sigma)k(u)\overline{G}^\sigma(u)du$ . Again trivially for  $t < \sigma$ ,  $B_{IP}(t; \sigma) = 0$ , and for  $t \geq \sigma$ ,

$$B_{IP}(t; \sigma) = \mathbf{1}_{[0, \infty)}(t - \sigma) + \int_{(0, t - \sigma]} B_{IP}(t; s + \sigma) \rho(s + \sigma)k(s)\overline{G}^\sigma(s)ds.$$

### Integral equation for the expected number of individuals alive at a given time

In the case where the individual characteristic is given by  $\chi^{Pr}(t - \sigma) := \mathbf{1}_{[0, L)}(t - \sigma)$ , we obtain

$$\Pr(t; \sigma) = 1 - G^\sigma(t - \sigma) + \int_{(0, t - \sigma]} \Pr(t; s + \sigma) d\Lambda(s; \sigma), \quad t \geq \sigma.$$

More specifically the equation for the expected number of individuals alive in the populations respectively of a TV-BH and of a TV-IP process are given for  $t \geq \sigma$ ,

$$\begin{aligned} \Pr_{BH}(t; \sigma) &= 1 - G^\sigma(t - \sigma) + \int_{(0, t - \sigma]} \Pr_{BH}(t; s + \sigma) R(\sigma + s) dG^\sigma(s), \\ \Pr_{IP}(t; \sigma) &= 1 - G^\sigma(t - \sigma) + \int_{(0, t - \sigma]} \Pr_{IP}(t; s + \sigma) \rho(s + \sigma)k(s)\overline{G}^\sigma(s)ds. \end{aligned}$$

## 2.2 A generic renewal-like integral equation

As already mentioned, one of the primary objectives of this chapter is to provide a rigorous mathematical foundation to the incidence renewal equation commonly used in epidemiology. We will encounter quite a few integral renewal-like equations in the next sections of this chapter, and the incidence equation will be one of them. However, most integral equations derived from a TV-CMJ branching process can be classified under the same generic

umbrella, which is, in essence, mirrored with the first-moment equation (2.1) derived in the previous section. We can write such a generic integral equation in its most general form, albeit under specific technical assumptions with regards to the functions  $(t; \sigma) \mapsto v(t; \sigma)$  and  $(t; \sigma) \mapsto \Lambda(t; \sigma)$  defined on  $\mathbb{R}_+ \times \mathbb{R}_+$ , as

$$u(t; \sigma) = \begin{cases} 0, & 0 \leq t < \sigma, \\ v(t; \sigma) + \int_{(0, t-\sigma]} u(t; \sigma + s) d\Lambda(s; \sigma), & t \geq \sigma. \end{cases}$$

When it comes to traditional CMJ processes, a similar generic renewal-like integral equation plays a crucial role in the literature where the dynamics and asymptotic behaviour of such processes are studied. We cite, among others, Jagers [Jag67] who studies integral equations of the form  $Y(t) = \int_0^t X(s) ds$ , related to an age-dependent branching process denoted  $X(t)$ . Jagers derives an integral equation for the moment-generating function of the coupled process  $(X(t), Y(t))$  on which he then relies to analyse the asymptotic behaviour of the first two moments of  $(X(t), Y(t))$  and the limit distribution of  $Y(t)$ . Crump and Mode also investigate integral equations for the first and second moments of generalised branching processes in their foundational papers [CM68] and [CM69]. Ney [Ney64] studies integral equations related to the law of total probability and the moments of a generalised branching process. We will also see in Section 2.6 that Kimmel [Kim82; Kim83], and later Kimmel and Arino [KA94] have done extensive work on the derivation of integral equations related to generalised branching processes.

The above system of equations will also be de facto ubiquitous in the study of our time-varying extension of a CMJ branching process. In this section, we show that under some sufficient assumptions related to the expected reproduction function  $(t; \sigma) \mapsto \Lambda(t; \sigma)$  and to the generic term  $(t; \sigma) \mapsto v(t; \sigma)$ , there exists a unique solution to this system. The system can be viewed in effect as representing a generic umbrella for the type of integral equations governing the dynamics of generalised branching processes. Some aspects of our existence and uniqueness proof rely partly on the work of Levinson [Lev60] who originally showed the existence and uniqueness of a solution to the integral equation governing the moment-generating function of a Bellman-Harris branching process and then showed how those equations can also be very usefully manipulated to deduce limiting theorems for age-dependent branching processes. However, as mentioned already, we will not be studying limiting theorems in this chapter as we will mainly be focussing on the derivation of an integral equation for a quantity that in practice is representative of an expected birth rate of a branching process and which corresponds to what is commonly known as incidence in epidemiology.

Let us now state a set of assumptions which will prove sufficient to ensure the existence and uniqueness of a solution to the above system.

**Assumptions** (Used to show existence and uniqueness).

(EU-1) For any fixed  $\sigma \geq 0$ , the function  $t \mapsto \Lambda(t; \sigma)$  defined on  $\mathbb{R}_+$  is continuously differentiable, i.e. there exists a continuous density function  $t \mapsto \lambda(t; \sigma)$  such that

$$d\Lambda(t; \sigma) = \lambda(t; \sigma) dt.$$

Additionally, for any  $t \geq 0$ , the function  $\sigma \mapsto \lambda(t; \sigma)$ , defined on  $\mathbb{R}_+$ , is bounded on any compact interval of  $\mathbb{R}$ .

(EU-2) The function  $(t; \sigma) \mapsto v(t; \sigma)$  defined on  $\mathbb{R}_+ \times \mathbb{R}_+$  is right-continuous with respect to

$(t; \sigma)$ , and has at most first-kind discontinuities\*. Moreover  $v(t; \sigma) = 0$  when  $0 \leq t < \sigma$ .

We state the main result of this section.

**Proposition 2.1.** *Suppose that Assumptions (EU-1) and (EU-2) hold. Then the generic integral equation given by*

$$u(t; \sigma) = \begin{cases} 0, & 0 \leq t < \sigma, \\ v(t; \sigma) + \int_{(0, t-\sigma]} u(t; \sigma + s) \lambda(s; \sigma) ds, & t \geq \sigma, \end{cases} \quad (\text{Generic Integral Eq.})$$

*admits a unique non-negative solution  $u(t; \sigma)$  defined on  $\mathbb{R}_+ \times \mathbb{R}_+$ . This solution is right-continuous with respect to  $(t; \sigma)$ , non-decreasing with respect to  $t$  for any  $\sigma \geq 0$  fixed and bounded on any compact interval with respect to  $\sigma$  for any  $t \geq 0$  fixed.*

*Proof.* (Appendix B.2) Classically, the proof of Proposition 2.1 is divided into two parts. We first show that there exists a solution that satisfies the properties stated in Proposition 2.1, and then we prove its uniqueness. The proof of existence is based on a procedure called *Picard iteration* which consists in generating a sequence of functions that approximate the solution we wish to show the existence of, while the proof of uniqueness relies on a Grönwall argument.  $\square$

## 2.3 Integral equation for the expected birth rate

This section presents a rigorous derivation of the integral equation satisfied by a quantity defined in Proposition 2.2 below, which in effect coincides with the expected birth rate in a TV-CMJ process. Even though we are still in the general context of branching populations (as opposed to more specifically the framework of infectious disease modelling), we may progressively and at times throughout this section be using a lexicon drawn from epidemiology, such as incidence and cumulative incidence. This is so because the objective of Section 2.5 is to provide a mathematical basis for the renewal incidence equation commonly applied in epidemiology.

### 2.3.1 From cumulative incidence to expected birth rate

Chapter 1 presented a measure-theoretic construction of a time-varying extension of a classical CMJ branching process whose underlying probability measure is calendar-time dependent. In the literature on integral equations derived from CMJ branching processes, up to our knowledge, the main quantities of interest tend to be, most of the time, the moment generating functions (see e.g. [Har63], [BH52], [Jag67], [CM69]), the cumulative number of births or deaths or the number of individuals alive at a given time point (see e.g. [Kim82; Kim83]).

In contrast, in this section, we seek to obtain an integral equation for an expected birth rate, which, when well defined, would correspond to the derivative of the cumulative number of births (the latter being otherwise known as cumulative incidence in epidemiology). Our strategy is to rely on the integral equation derived for the cumulative number of deaths in a TV-CMJ branching process, but reformulate it as a distributional integral equation, in other

---

\*In other words,  $v(t; \sigma)$  does not have infinite discontinuities with respect to  $(t; \sigma)$ , and both the right-hand and left-hand limits exist and are finite for any  $(t; \sigma) \in \mathbb{R}_+ \times \mathbb{R}_+$ .

words, an equation involving generalised Schwartz distributions [Sch51]. This extension allows us to mathematically define the derivative of the cumulative number of births in a TV-CMJ process and derive its integral equation, which we will then relate directly in Section 2.5 to the epidemiology incidence equation.

To provide practical evidence of the relevance of the following derivation of an integral equation for the expected birth rate in a TV-CMJ branching process, we may find it helpful to project ourselves away from the more theoretical framework of generalised branching processes and navigate towards a more physical setting. As mentioned at the beginning of Section 2.2, the early dynamics of an infection disease propagation can be reliably modelled using a branching process. Epidemiology is not the only context that can be modelled by branching processes. Individuals in a branching population can at times represent cases of new infections but also equally, for instance, neutrons in particle physics, cancer cells in biology . . . . Any setting where dynamics of populations need studying (and where individuals can be assumed for a certain period of time to behave identically and independently, either unconditionally or in a Markovian context) can usefully be modelled by branching processes.

Various quantities of interest are of relevance, but especially in the epidemiological context that followed the onset of COVID-19 in January 2020, it appeared even more valuable to rigorously derive an integral equation for the *expected birth rate*, rather than being somewhat limited to integral equations for cumulative incidence or prevalence. The expected birth rate corresponds mathematically to the *derivative* of the cumulative number of births which is, in effect, the average number of births at a given point in time  $t \geq 0$ . Being able to model and simulate such a quantity in practice allows us to obtain a much finer understanding of the dynamics of the epidemics.

With this background in mind, in this section we study in more detail the integral renewal-like equation (2.3) that was derived in Section 2.1 for the cumulative incidence in a TV-CMJ branching process. We first reformulate the cumulative incidence of a TV-CMJ branching process by redefining it as a generalised function, which in turn leads to the derivation of an equation for its distributional derivative in Proposition 2.2. This distributional derivative can then be seen to represent in essence the expected number of births that occur at any time point  $t \geq 0$  in a TV-CMJ branching process. We call it *expected birth rate*.

More precisely, let us go back to the general equation (2.3) for cumulative incidence, in other words, the expected cumulative number of births up to a certain time  $t \geq 0$  in a TV-CMJ process. We first need to ensure the existence and uniqueness of a solution to this equation. To do so, it suffices to assume that condition (EU-1) holds. Indeed, letting  $v(t; \sigma) \equiv \mathbf{1}_{[0, \infty)}(t - \sigma)$  guarantees that condition (EU-2) is de facto true. Consequently, the latter condition is not needed in this case. We then rely on Proposition 2.1 which states that there exists a unique solution to the system of equations that determines the cumulative number of births in a TV-CMJ process. We rewrite this system here for convenience using the assumption that  $\Lambda(t)$  has a density

$$B(t; \sigma) = \begin{cases} 0, & t < \sigma, \\ \mathbf{1}_{[0, \infty)}(t - \sigma) + \int_{(0, t - \sigma]} B(t; \sigma + s) \lambda(s; \sigma) ds, & t \geq \sigma. \end{cases} \quad (2.4)$$

Our objective in this section is to determine an integral equation for a quantity that represents, if well-defined, the expected birth rate of new individuals in a TV-CMJ branching process. To do so, we proceed as follows: in Section 2.3.2 we first introduce the Lebesgue decomposition of the cumulative incidence function uniquely defined as the solution to (2.4). Next in Section

2.3.3 we will redefine cumulative incidence as a generalised function (otherwise called *Schwartz distribution*, see [Sch51]). This technical apparatus will allow us to properly define in Section 2.3.4 a distributional derivative for the cumulative incidence, which we will then adopt as the mathematical representation of the expected birth rate in a TV-CMJ branching process. The main result of this section is summarised in the following proposition.

**Notation.** We denote by  $\mathcal{D}'(\mathbb{R})$  the space of Schwartz distributions also known as generalised functions. (See Appendix B.3 for a very brief and targeted review of the results on this topic that will be needed in the following derivations).

**Proposition 2.2** (Integral equation for the expected birth rate). *Assume that conditions (R-1), (R-2) and (EU-1) hold. Let  $B_\sigma \in \mathcal{D}'(\mathbb{R})$  represent the generalised function that extends, according to Definition B.3.2, the cumulative incidence function  $(t; \sigma) \mapsto B(t; \sigma)$  of a TV-CMJ. Then the distributional derivative  $f_\sigma := B'_\sigma \in \mathcal{D}'(\mathbb{R})$  is solution to the following system*

$$\langle f_\sigma, \varphi \rangle = \langle \delta_\sigma, \varphi \rangle + \int_{\mathbb{R}_+} \varphi(t) \tilde{f}(t; \sigma) dt, \quad \varphi \in \mathcal{C}_c^\infty(\mathbb{R}), \quad (2.5)$$

where  $\tilde{f} : \mathbb{R}_+ \times \mathbb{R}_+ \rightarrow \mathbb{R}_+$  is the unique solution to the integral equation

$$\tilde{f}(t; \sigma) = \begin{cases} 0, & t < \sigma, \\ \lambda(t - \sigma; \sigma) + \int_{(0, t - \sigma]} \tilde{f}(t; \sigma + s) \lambda(s; \sigma) ds, & t \geq \sigma. \end{cases} \quad (2.6)$$

### 2.3.2 Absolutely continuous part of cumulative incidence

We place ourselves in the context of a TV-CMJ branching process which describes the dynamics of a population according to Definition 1.4.2, and assume conditions (R-1) and (R-2) hold to ensure that the process is regular and does not explode in finite time almost surely.

We denote by  $B(t; \sigma)$  the unique solution to the system of equations (2.4). The function  $t \mapsto B(t; \sigma)$  is locally integrable in  $\mathbb{R}_+$ . Thus, by the Lebesgue Decomposition Theorem (e.g. [Hall13, ch. VI, p.134]), such a locally integrable function is uniquely decomposable with respect to the Lebesgue measure as a sum of a singular and an absolutely continuous measures. To simplify the following exposition, we will assume that  $t \mapsto B(t; \sigma)$  admits a unique point of discontinuity at  $t = \sigma$ . More precisely,

**Assumption** (Absolute continuity of  $t \mapsto B(t; \sigma)$  for  $t > \sigma$ ).

(A1) The function  $t \mapsto B_{ac}(t; \sigma) := B(t; \sigma) - \mathbf{1}_{[\sigma, \infty)}(t)$  defined as

$$B_{ac}(t; \sigma) := \begin{cases} 0, & 0 \leq t \leq \sigma, \\ B(t; \sigma) - 1, & t > \sigma, \end{cases} \quad (2.7)$$

is absolutely continuous for any  $t \in \mathbb{R}_+$ .

*Remark.* Given the system of equations (2.4), Assumption (A1) in effect assumes that the point  $t = \sigma$  is the unique discontinuity point of the cumulative incidence function  $t \mapsto B(t; \sigma)$  in  $\mathbb{R}_+$ . In practice, the cumulative incidence of a TV-CMJ branching process could potentially have subsequent discontinuity points. However, if this is the case, the analysis in the following would still hold but would need to be restricted to each successive interval of continuity.

We now derive the system of integral equations in which the absolutely continuous part  $B_{ac}(t; \sigma)$  is the unique solution.

**Lemma 2.3.1.** *The absolutely continuous function  $t \mapsto B_{ac}(t; \sigma) = B(t; \sigma) - \mathbf{1}_{[\sigma, \infty)}(t)$  defined on  $\mathbb{R}_+$  is the unique solution to the integral equation*

$$B_{ac}(t; \sigma) = \begin{cases} 0, & 0 \leq t \leq \sigma, \\ \int_{(\sigma, t]} \lambda(s - \sigma; \sigma) ds + \int_{(\sigma, t]} B_{ac}(t; s) \lambda(s - \sigma; \sigma) ds, & t > \sigma. \end{cases} \quad (2.8)$$

*Proof.* Indeed, for  $t > \sigma$ , the definition of  $B_{ac}(t; \sigma)$  yields

$$\begin{aligned} B_{ac}(t; \sigma) &= B(t; \sigma) - \mathbf{1}_{[\sigma, \infty)}(t), \\ &= \int_{(\sigma, t]} B(t; s) \lambda(s - \sigma; \sigma) ds, \\ &= \int_{(\sigma, t]} \left( \mathbf{1}_{[s, \infty)}(t) + B_{ac}(t; s) \right) \lambda(s - \sigma; \sigma) ds, \end{aligned}$$

where we simply relied on (2.4) in the second line and on (2.7) in the last line. The above simplifies

$$B_{ac}(t; \sigma) = \int_{(\sigma, t]} \lambda(s - \sigma; \sigma) ds + \int_{(\sigma, t]} B_{ac}(t; s) \lambda(s - \sigma; \sigma) ds.$$

It now suffices to observe that the above equation is of the form of the (Generic Integral Eq.) whereby

$$v(t; \sigma) := \int_{(\sigma, t]} \lambda(s - \sigma; \sigma) ds.$$

Such a function clearly satisfies the properties required in Assumption (EU-2), in other words  $v(t; \sigma)$  is right-continuous with respect to  $(t; \sigma)$  and has at most first-kind discontinuities. Since Assumption (EU-1) is also satisfied, Proposition 2.1 ensures that the absolutely continuous part of  $B(t; \sigma)$ , denoted  $B_{ac}(t; \sigma)$ , is the unique solution to the integral equation (2.7).  $\square$

The next step is to verify that the partial derivative of  $B_{ac}(t; \sigma)$  with respect to  $t \geq 0$  also satisfies the (Generic Integral Eq.). Let us define

$$\tilde{f}(t; \sigma) := \frac{\partial}{\partial t} B_{ac}(t; \sigma).$$

**Lemma 2.3.2.** *The function  $(t; \sigma) \mapsto \tilde{f}(t; \sigma)$  verifies the integral equation*

$$\tilde{f}(t; \sigma) = \begin{cases} 0, & 0 \leq t \leq \sigma, \\ \lambda(t - \sigma; \sigma) + \int_{(0, t - \sigma]} \tilde{f}(t; \sigma + s) \lambda(s; \sigma) ds, & t > \sigma. \end{cases} \quad (2.9)$$

*Remark.* Observe that  $t \mapsto \tilde{f}(t; \sigma)$  is not necessarily continuous at  $t = \sigma$  when  $\lambda(0; \sigma) \neq 0$ .

*Proof.* We start from (2.9) which is of the generic form given by (Generic Integral Eq.) with  $v(t; \sigma) := \lambda(t - \sigma; \sigma)$ . By Proposition 2.1, this integral equation admits a unique right-continuous solution which we denote  $\tilde{f}(t; \sigma)$ . As it is right-continuous for  $t \geq \sigma$ ,  $t \mapsto \tilde{f}(t; \sigma)$  is measurable and in particular, admits primitives. We consider here the primitive  $\tilde{F}(t; \sigma)$  defined in such a

way that it is equal to 0 when  $t = \sigma$ , that is,

$$\tilde{F}(t; \sigma) := \int_{\sigma}^t \tilde{f}(u; \sigma) du.$$

Substituting  $\tilde{f}(u; \sigma)$  by its expression in equation (2.9), we rewrite the above equality which leads to an integral equation for  $\tilde{F}(t; \sigma)$ . For  $t > \sigma$ , we get

$$\begin{aligned} \tilde{F}(t; \sigma) &= \int_{\sigma}^t \left( \lambda(u - \sigma; \sigma) + \int_{(0, u - \sigma]} \tilde{f}(u; s + \sigma) \lambda(s; \sigma) ds \right) du, \\ &= \int_{\sigma}^t \lambda(u - \sigma; \sigma) du + \int_{\sigma}^t du \int_{(\sigma, u]} \tilde{f}(u; \nu) \lambda(\nu - \sigma; \sigma) d\nu, \\ &= \int_{\sigma}^t \lambda(u - \sigma; \sigma) du + \int_{(\sigma, t]} \lambda(\nu - \sigma; \sigma) \left( \int_{\nu}^t \tilde{f}(u; \nu) du \right) d\nu, \end{aligned}$$

where the Unsymmetric Fubini theorem (see Appendix A.4) was applied in the last line. In summary,  $\tilde{F}(t; \sigma)$ , verifies

$$\tilde{F}(t; \sigma) = \begin{cases} 0, & 0 \leq t \leq \sigma, \\ \int_{\sigma}^t \lambda(u - \sigma; \sigma) du + \int_{(\sigma, t]} \tilde{F}(t; \nu) \lambda(\nu - \sigma; \sigma) d\nu, & t > \sigma, \end{cases}$$

which we recognise as being equation (2.8). Therefore, by uniqueness of solution,

$$\tilde{F}(t; \sigma) = B_{ac}(t; \sigma),$$

which yields, taking the partial derivative on each side,

$$\frac{\partial}{\partial t} B_{ac}(t; \sigma) = \tilde{f}(t; \sigma).$$

We conclude that  $\frac{\partial}{\partial t} B_{ac}(t; \sigma)$  verifies (2.9). □

### 2.3.3 Cumulative incidence described as a generalised function

We now return to the cumulative incidence function  $(t; \sigma) \mapsto B(t; \sigma)$  with the objective of reformulating it as a generalised function [Sch51], which will therefore de facto admit a distributional derivative. We generally refer to Appendix B.3 for a succinct overview of the results needed to support the proofs of this section. Based on Definition B.3.1 in Appendix, we recall that a *distribution on  $\mathbb{R}$*  denoted  $T : \mathcal{D}(\mathbb{R}) \rightarrow \mathbb{R}$  is any continuous linear functional on  $\mathcal{D}(\mathbb{R})$ , where the space  $\mathcal{D}(\mathbb{R})$  is the set  $\mathcal{C}_c^{\infty}(\mathbb{R})$  of infinitely differentiable functions with compact support on  $\mathbb{R}$ . The linear space of all distributions on  $\mathbb{R}$  is denoted  $\mathcal{D}'(\mathbb{R})$ . It is the topological dual of  $\mathcal{D}(\mathbb{R})$  and is often called the space of test functions on  $\mathbb{R}$ . More specifically, when there exists a locally integrable function  $f_T$  on  $\mathbb{R}$  such that

$$\langle T, \varphi \rangle = \langle f_T, \varphi \rangle = \int_{\mathbb{R}} f_T(x) \varphi(x) dx, \quad \varphi \in \mathcal{D}(\mathbb{R}).$$

then the distribution  $T$  is said to be *regular*.

Now observing that the function  $t \mapsto B(t; \sigma)$  is locally integrable in  $\mathbb{R}_+$  for any fixed  $\sigma$ , the above result implies that the function  $t \mapsto B(t; \sigma)$  can naturally be extended into a

(regular) generalised function. For any  $\sigma \geq 0$ , we will denote this Schwartz distribution by  $B_\sigma$  : it is a continuous linear functional defined on the space of test functions  $\varphi \in \mathcal{C}_c^\infty(\mathbb{R}_+)$ . We define its action on test functions more formally in the following lemma.

**Lemma 2.3.3** (Definition of  $B_\sigma$ ). *Let  $B_\sigma : \mathcal{C}_c^\infty(\mathbb{R}_+) \rightarrow \mathbb{R}_+$  be the generalised function defined for any  $\varphi \in \mathcal{C}_c^\infty(\mathbb{R}_+)$  by*

$$\langle B_\sigma, \varphi \rangle := \int_{\mathbb{R}_+} B(t; \sigma) \varphi(t) dt = \int_{\mathbb{R}_+} \varphi(t) dt + \int_{\mathbb{R}_+} B_{ac}(t; \sigma) \varphi(t) dt, \quad (2.10)$$

where  $B_{ac}(t; \sigma)$  is the unique solution of (2.8). This defines a unique distribution on  $\mathcal{D}'(\mathbb{R})$ , which we call cumulative incidence.

*Proof.* We start with the definition, namely  $B_\sigma := \int_{\mathbb{R}_+} B(t; \sigma) \varphi(t) dt$  where we substitute  $B(t; \sigma)$  by its Lebesgue decomposition,

$$\begin{aligned} \langle B_\sigma, \varphi \rangle &= \int_{\mathbb{R}_+} \left( \mathbf{1}_{[\sigma, \infty)}(t) + B_{ac}(t; \sigma) \right) \varphi(t) dt, \\ &= \int_{[\sigma, \infty)} \varphi(t) dt + \int_{\mathbb{R}_+} B_{ac}(t; \sigma) \varphi(t) dt. \end{aligned}$$

The linearity and continuity of  $B_\sigma$  with respect to  $\varphi$  are immediate. Uniqueness also follows immediately. Indeed, on the one hand, the first term  $\varphi \mapsto \int_{[\sigma, \infty)} \varphi(t) dt$  defines a unique continuous linear functional on  $\mathcal{C}_c^\infty(\mathbb{R}_+)$ . On the other hand, as  $B_{ac}(t; \sigma)$  is already uniquely defined as the solution to the integral equation (2.7), it follows directly that the second term above defines a linear functional on  $\mathcal{C}_c^\infty(\mathbb{R}_+)$ .  $\square$

### 2.3.4 Expected birth rate as derivative of cumulative incidence

This short excursion in the space of Schwartz distributions has led us to redefine cumulative incidence as a generalised function  $B_\sigma$ . The objective of this reformulation is to ensure that such a function has a well-defined distributional derivative, which we denote  $B'_\sigma$ .

We now return to the distribution  $f_\sigma : \mathcal{C}_c^\infty(\mathbb{R}) \rightarrow \mathbb{R}$  defined by the system of equations (2.5) and (2.6). We show that the distributional derivative  $B'_\sigma$  is identical to  $f_\sigma$ . To do so, it suffices to prove that the actions of each of these distributions on the set of test functions are similar.

**Lemma 2.3.4.** *For any  $\sigma \geq 0$ , the distributional derivative  $B'_\sigma$  of the generalised function  $B_\sigma$  defined in (2.10) is the unique solution to the system of equations (2.5) and (2.6), in other words,*

$$B'_\sigma = f_\sigma.$$

*Proof.* Let us first recall the action of a distributional derivative  $T \in \mathcal{D}'(\mathbb{R})$  on any test function  $\varphi \in \mathcal{D}(\mathbb{R})$ . For any exponent  $\alpha \in (\mathbb{N}^*)^n$ , we have

$$\langle d^{(\alpha)} T, \varphi \rangle := (-1)^{|\alpha|} \langle T, d^{(\alpha)} \varphi \rangle.$$

This result (also recalled in Appendix B.3.3) can directly be applied in our context for  $\alpha = 1$ . We then get

$$\begin{aligned}
\langle B'_\sigma, \varphi \rangle &= -\langle B_\sigma, \varphi' \rangle, \\
&= -\int_{[\sigma, \infty)} \varphi'(t) dt - \int_{\mathbb{R}_+} B_{ac}(t; \sigma) \varphi'(t) dt, \\
&= -[\varphi(t)]_\sigma^\infty - \left[ \frac{\partial}{\partial t} B_{ac}(t; \sigma) \varphi(t) \right]_{\mathbb{R}_+} + \int_{\mathbb{R}_+} \frac{\partial}{\partial t} B_{ac}(t; \sigma) \varphi(t) dt, \\
&= \varphi(\sigma) - [\tilde{f}(t; \sigma) \varphi(t)]_{\mathbb{R}_+} + \int_{\mathbb{R}_+} \tilde{f}(t; \sigma) \varphi(t) dt, \\
&= \langle \delta_\sigma, \varphi \rangle + \int_{\mathbb{R}_+} \tilde{f}(t; \sigma) \varphi(t) dt.
\end{aligned}$$

The first line is simply the definition of a distributional derivative. The second line relies on integral equation (2.10) satisfied by  $B_\sigma$ . A simple integration by parts leads to the third line, whilst the last two lines rely on one hand on the fact that the partial derivative of the absolutely continuous part of  $B(t; \sigma)$  is  $\tilde{f}(t; \sigma)$ , and on the other hand on the compactness of the support of any test functions. Comparing with (2.5), uniqueness of the solution yields

$$B'_\sigma = f_\sigma,$$

which means that the solution to (2.5) and (2.6) is the distributional derivative of cumulative incidence.  $\square$

*Remark.* From now on,  $B'_\sigma = f_\sigma$  will be called *the expected birth rate* of a TV-CMJ branching process. This denomination will help us to lexically distinguish it from what is commonly called *incidence* in epidemiology.

So far, we have derived the integral renewal-type equation satisfied by  $f_\sigma$ . In section 2.5.4, we will want to prove equivalence between the integral equation for  $f_\sigma$  and the extended incidence renewal equation (defined formally in subsequent section). However, before we embark on this equivalence proof, we make a slight detour to recall a discretisation algorithm originally presented in [Pak+23] and which was used to estimate rates of transmission from serological prevalence of SARS-CoV-2 in the UK and historical incidence data on Influenza, Measles, SARS and Smallpox. In particular, we prove that this discretisation algorithm converges to order  $O(n^{-1})$  as  $n \rightarrow \infty$ .

## 2.4 Discretisation algorithm of integral equations

In this section, our starting point is the (Generic Integral Eq.). We introduce a discretisation scheme directly linked to [Pak+23, Algorithm I] and prove that this algorithm converges to the order  $O(n^{-1})$  under specific assumptions.

### 2.4.1 Discretisation scheme

We assume in this section that conditions (EU-1) and (EU-2) hold to ensure the existence and uniqueness of the solution to the (Generic Integral Eq.) of Proposition 2.1. As mentioned in Section 2.2, the integral equations encountered in this chapter all share a similar form. For convenience of the reader, we recall below the (Generic Integral Eq.), whose unique solution

is the function denoted here  $(t; \sigma) \mapsto f(t; \sigma)$  in  $\mathbb{R}_+ \times \mathbb{R}_+$ ,

$$f(t; \sigma) = \begin{cases} 0, & 0 \leq t < \sigma, \\ v(t; \sigma) + \int_0^{t-\sigma} f(t; \sigma + s) \lambda(s; \sigma) ds, & t \geq \sigma. \end{cases} \quad (\text{Generic})$$

Particular choices for  $v(t; \sigma)$  lead to different quantities of interest with respect to the considered branching process. So for instance, if we take the perspective of infectious disease modelling, we have

$$v(t, \sigma) := \begin{cases} \mathbf{1}_{[0; \infty)}(t - \sigma, ) & f(t; \sigma) = B(t; \sigma), & (\text{Cumulative incidence}) \\ 1 - G^\sigma(t - \sigma), & f(t; \sigma) = \text{Pr}(t; \sigma), & (\text{Prevalence}) \\ \lambda(t - \sigma; \sigma), & f(t; \sigma) = \tilde{f}(t; \sigma). & (\text{Incidence for } t > \sigma) \end{cases}$$

When trying to simulate and solve numerically ([Generic Integral Eq.](#)), a key hurdle is encountered. Indeed, on the right-hand side, we get  $f(t; u)$  for  $\sigma \leq u \leq t$  and not  $f(u; \sigma)$  for  $\sigma \leq u \leq t$ . This means in practice that, to solve, for example,  $f(t; 0)$  for  $t \geq 0$ , we need to actually solve  $f(t; \sigma)$  for any pair  $(t; \sigma)$  such that  $0 \leq \sigma \leq t$ .

To overcome this implementation challenge, a discretisation algorithm was introduced in [\[Pak+23, Section 3\]](#). In this section, we only present the simpler version of the algorithm (i.e. Algorithm 1 of [\[Pak+23\]](#)) as the details of the simulations for the vectorised version are provided in the appendix of [\[Pak+23\]](#). Our focus here is not on implementation, but on proving convergence of the algorithm toward the solution of ([Generic Integral Eq.](#)). We will provide full details in Theorem 2.1 and in its subsequent proof of what is actually meant by convergence here.

However, before we present Algorithm 1, let us observe that numerically solving for ([Generic Integral Eq.](#)) is greatly facilitated if we introduce an auxiliary quantity defined for any  $t \geq s \geq 0$  by

$$f_t(s) := f(t; t - s).$$

More precisely, starting from the ([Generic Integral Eq.](#)), we proceed with a simple change of variable and observe that, for any fixed  $t \geq 0$ , the *single-argument* function  $s \mapsto f_t(s)$  defined on the interval  $[0, t]$  is governed by the renewal-like integral equation

$$f_t(s) = v(t; t - s) + \int_0^s f_t(s - u) \lambda(u; t - s) du, \quad 0 \leq s \leq t. \quad (2.11)$$

We then recover  $f(t, 0)$  for  $t \geq 0$  via  $f(t, 0) = f_t(t)$ .

Let us now briefly detail how we proceed to discretise the space  $\mathbb{R}_+ \times \mathbb{R}_+$ . We simply introduce a grid step of size  $\frac{1}{n}$  for  $n \in \mathbb{N}^*$ , and define a subset  $\mathbb{T}_n \subset \mathbb{R}_+ \times \mathbb{R}_+$  such that

$$\mathbb{T}_n := \{(t; \sigma) \in \left(\frac{1}{n}\mathbb{Z}\right)^2 : 0 \leq \sigma \leq t\}.$$

Then, for any given pair of integers  $0 \leq i \leq j$ , we introduce a discretised equation mirroring

our generic integral equation and whose solution, denoted here  $\hat{f} : \mathbb{T}_n \rightarrow \mathbb{R}_+$  satisfies

$$\hat{f}\left(\frac{j}{n}, \frac{i}{n}\right) = \begin{cases} v\left(\frac{j}{n}, \frac{j}{n}\right), & j = i, \\ v\left(\frac{j}{n}, \frac{i}{n}\right) + \frac{1}{n} \sum_{k=1}^{j-i-1} \hat{f}\left(\frac{j}{n}, \frac{k+i}{n}\right) \lambda\left(\frac{k}{n}, \frac{i}{n}\right), & 0 \leq i < j. \end{cases}$$

We now operate a similar change of variable as the one that led to equation (2.11),

$$\hat{f}\left(\frac{j}{n}, \frac{i}{n}\right) = \begin{cases} v\left(\frac{j}{n}, \frac{j}{n}\right), & j = i, \\ v\left(\frac{j}{n}, \frac{j-i}{n}\right) + \frac{1}{n} \sum_{k=1}^{i-1} \hat{f}\left(\frac{j}{n}, \frac{j-(i-k)}{n}\right) \lambda\left(\frac{k}{n}, \frac{j-i}{n}\right) & 0 \leq i < j. \end{cases}$$

Next, if we fix  $j \in \mathbb{N}$ , we can define for any  $0 \leq i \leq j$  the *single-argument* function

$$\hat{f}_{\frac{j}{n}}\left(\frac{i}{n}\right) := \hat{f}\left(\frac{j}{n}, \frac{j-i}{n}\right),$$

and rewrite the discretised system

$$\hat{f}_{\frac{j}{n}}\left(\frac{i}{n}\right) = \begin{cases} v\left(\frac{j}{n}, \frac{j}{n}\right), & i = j, \\ v\left(\frac{j}{n}, \frac{j-i}{n}\right) + \frac{1}{n} \sum_{k=1}^{i-1} \hat{f}_{\frac{j}{n}}\left(\frac{i-k}{n}\right) \lambda\left(\frac{k}{n}, \frac{j-i}{n}\right), & i < j. \end{cases}$$

If we let  $N \in \mathbb{N}$  and  $n \in \mathbb{N}^*$ , and consider the discrete time points  $t_j = \frac{j}{n}$  for  $j = 0, \dots, N$ , then, in practice, we are interested in solving  $f(t_j, 0) = f_{t_j}(t_j)$ . This leads us to defining the pseudo-code given in Algorithm 1 (originally implemented in [Pak+23]).

---

**Algorithm 1** Discretisation of integral equations

---

**Require:** functions  $v(\cdot; \cdot)$  and  $\lambda(\cdot; \cdot)$

**Require:** step size  $\frac{1}{n} > 0$  for  $n \in \mathbb{N}^*$

**Require:** number of time steps  $N \in \mathbb{N}$

```

1: for  $j = 0, \dots, N$  do
2:   for  $i = 1, \dots, j - 1$  do
3:     if  $i = 0$  then
4:        $\hat{f}_{\frac{j}{n}}\left(\frac{i}{n}\right) \leftarrow v\left(\frac{j}{n}, \frac{j}{n}\right)$ 
5:     else
6:        $\hat{f}_{\frac{j}{n}}\left(\frac{i}{n}\right) \leftarrow v\left(\frac{j}{n}, \frac{j-i}{n}\right) + \frac{1}{n} \sum_{k=1}^{i-1} \hat{f}_{\frac{j}{n}}\left(\frac{i-k}{n}\right) \lambda\left(\frac{k}{n}, \frac{j-i}{n}\right)$ 
7:     end if
8:   end for
9: end for
10: return  $\hat{f}_{\frac{j}{n}}\left(\frac{j}{n}\right)$  ( $\approx f\left(\frac{j}{n}, 0\right)$ ) for any  $j = 0, \dots, N$ 

```

---

### 2.4.2 Convergence of discretisation algorithm

In this section, we prove that the error term of the discretisation Algorithm 1 is of order  $O(n^{-1})$ . First, however, we need to introduce a set of sufficient assumptions to guarantee the validity (in other words, the convergence) of the discretisation algorithm. Note that in the following, we mention  $K$ -Lipschitz continuous functions without specifying the value of  $K > 0$ .

This value may change from one function to another, but this does not affect the arguments of the proof.

**Assumption** (Used to prove convergence of discretisation algorithm to order  $O(n^{-1})$ ).

(A2) *There exists a positive constant  $K > 0$  such that the function  $(t; s) \mapsto v(t; s)$  is  $K$ -Lipschitz continuous for  $t \geq 0$  and  $s \in [0, t]$ ;*

(A3) *There exists a positive constant  $K > 0$  such that the function  $(t; s) \mapsto \lambda(t; s)$  is  $K$ -Lipschitz continuous for  $t \geq 0$  and  $s \in [0, t]$ ;*

**Theorem 2.1** (Error term of Algorithm 1). *Assume conditions (EU-1) and (EU-2) hold. Moreover, assume that (A2) and (A3) hold. Then for any  $0 \leq s \leq t$  and for any  $n \in \mathbb{N}^*$ , let*

$$s_n := \frac{\lfloor ns \rfloor}{n} \quad \text{and} \quad t_n := \frac{\lfloor nt \rfloor}{n}.$$

Define on the one hand the sequence  $(\hat{f}_{t_n}(s_n))_{n \in \mathbb{N}^*}$  as solution to the system

$$\hat{f}_{t_n}(s_n) = \begin{cases} v(t_n; t_n), & s = 0, \\ v(t_n; t_n - s_n) + \frac{1}{n} \sum_{k=1}^{ns_n-1} \hat{f}_{t_n}\left(s_n - \frac{k}{n}\right) \lambda\left(\frac{k}{n}; t_n - s_n\right), & 0 < s \leq t. \end{cases} \quad (2.12)$$

And on the other hand, let the function  $f_t(s)$  be the unique solution of

$$f_t(s) = \begin{cases} v(t; t), & s = 0, \\ v(t; t - s) + \int_0^s f_t(s - u) \lambda(u; t - s) du, & 0 < s \leq t. \end{cases} \quad (2.13)$$

Then,

$$|\hat{f}_{t_n}(s_n) - f_t(s)| \sim O(n^{-1}), \quad \text{as } n \rightarrow \infty.$$

*Proof.* Let  $0 \leq s \leq t$  and denote for any  $n \in \mathbb{N}^*$ ,  $j := \lfloor nt \rfloor$ ,  $i := \lfloor ns \rfloor$  such that

$$t_n := \frac{\lfloor nt \rfloor}{n} = \frac{j}{n}, \quad s_n := \frac{\lfloor ns \rfloor}{n} = \frac{i}{n}.$$

We can rewrite (2.12) as follows

$$\hat{f}_{t_n}(s_n) = v(t_n; t_n - s_n) + \sum_{k=1}^{ns_n-1} \int_{k/n}^{(k+1)/n} \hat{f}_{t_n}\left(s_n - \frac{k}{n}\right) \lambda\left(\frac{k}{n}; t_n - s_n\right) du,$$

where we simply used that  $\int_{k/n}^{(k+1)/n} du = 1/n$ .

Next, we substitute  $k$  with  $\lfloor nu \rfloor$  for  $u \in [\frac{k}{n}, \frac{k+1}{n}]$  and obtain, denoting  $u_n := \frac{\lfloor nu \rfloor}{n}$ ,

$$\begin{aligned} \hat{f}_{t_n}(s_n) &= v(t_n; t_n - s_n) + \int_{1/n}^{s_n} \hat{f}_{t_n}\left(s_n - \frac{\lfloor nu \rfloor}{n}\right) \lambda\left(\frac{\lfloor nu \rfloor}{n}; t_n - s_n\right) du, \\ &= v(t_n; t_n - s_n) + \int_{1/n}^{s_n} \hat{f}_{t_n}(s_n - u_n) \lambda(u_n; t_n - s_n) du. \end{aligned}$$

Taking the difference between the preceding equality and (2.13), we obtain

$$\begin{aligned} |\hat{f}_{t_n}(s_n) - f_t(s)| &\leq |v(t; t-s) - v(t_n; t_n - s_n)| \\ &\quad + \left| \int_0^s f_t(s-u)\lambda(u; t-s)du - \int_{1/n}^{s_n} \hat{f}_{t_n}(s_n - u_n)\lambda(u_n; t_n - s_n)du \right|, \end{aligned}$$

which we rewrite by splitting and re-arranging the integrals,

$$\begin{aligned} |\hat{f}_{t_n}(s_n) - f_t(s)| &\leq |v(t; t-s) - v(t_n; t_n - s_n)| + \int_0^{1/n} f_t(s-u)\lambda(u; t-s)du \\ &\quad + \int_{s_n}^s f_t(s-u)\lambda(u; t-s)du + \int_0^s \left| \hat{f}_{t_n}(s_n - u_n)\lambda(u_n; t_n - s_n) - f_t(s-u)\lambda(u; t-s) \right| du. \end{aligned} \tag{2.14}$$

In what follows, the value of the positive constant  $C > 0$  may vary from one line to another but it will always be finite and independent of  $n$ . We will also use (implicitly at times) the somewhat vague expression *for  $n$  large enough* to avoid clouding the argument with unnecessary notations.

Returning to the first three terms on the right-hand side of (2.14), we observe that they can be upper-bounded together, appealing respectively to the right-continuity of the function  $(t; s) \mapsto v(t; t-s)$  for  $0 \leq s \leq t$  and to the boundedness of  $u \mapsto \lambda(u; t-s)$  and of  $s \mapsto f_t(s)$  on any compact. We have, for  $n$  large enough,

$$|v(t; t-s) - v(t_n; t_n - s_n)| + \int_0^{1/n} f_t(s-u)\lambda(u; t-s)du + \int_{s_n}^s f_t(s-u)\lambda(u; t-s)du \lesssim \frac{1}{n},$$

where the symbol  $\lesssim$  means that for  $n$  large enough, there exists a constant  $C > 0$  such that the quantity on the left hand side is upper-bounded by  $\frac{C}{n}$ . Next we consider the fourth term in (2.14) and simply apply the triangular inequality,

$$\begin{aligned} &\int_0^s \left| \hat{f}_{t_n}(s_n - u_n)\lambda(u_n; t_n - s_n) - f_t(s-u)\lambda(u; t-s) \right| du \\ &\leq \int_0^s \left| \hat{f}_{t_n}(s_n - u_n) - f_t(s-u) \right| \lambda(u_n; t_n - s_n) du + \int_0^s f_t(s-u) \left| \lambda(u_n; t_n - s_n) - \lambda(u; t-s) \right| du. \end{aligned}$$

Here the Lipschitz property of  $\lambda(t; \sigma)$  in Assumption (A3) and boundedness of  $f_t$  yield

$$\begin{aligned} &\int_0^s \left| \hat{f}_{t_n}(s_n - u_n)\lambda(u_n; t_n - s_n) - f_t(s-u)\lambda(u; t-s) \right| du \\ &\lesssim \frac{s}{n} + C \int_0^s \left| \hat{f}_{t_n}(s_n - u_n) - f_t(s-u) \right| du. \end{aligned}$$

We now define  $(s-u)_n := \frac{\lfloor n(s-u) \rfloor}{n}$  and again simply apply the triangular inequality twice to

get

$$\begin{aligned} \int_0^s |\hat{f}_{t_n}(s_n - u_n) - f_t(s - u)| du &\leq \int_0^s |\hat{f}_{t_n}(s_n - u_n) - \hat{f}_{t_n}((s - u)_n)| du \\ &\quad + \int_0^s |\hat{f}_{t_n}((s - u)_n) - f_t((s - u)_n)| du \\ &\quad + \int_0^s |f_t((s - u)_n) - f_t(s - u)| du. \end{aligned}$$

It was proven in Appendix B.4.1 that  $f_t(s)$  and  $\hat{f}_{t_n}(s_n)$  are  $K$ -Lipschitz continuous. Thus

$$\int_0^s |\hat{f}_{t_n}(s_n - u_n) - f_t(s - u)| du \lesssim \frac{s}{n} + \int_0^s |\hat{f}_{t_n}((s - u)_n) - f_t((s - u)_n)| du.$$

Next, splitting again the integrand on the right-hand side, we get

$$\begin{aligned} \int_0^s |\hat{f}_{t_n}(s_n - u_n) - f_t(s - u)| du &\lesssim \frac{s}{n} + \int_0^s |\hat{f}_{t_n}((s - u)_n) - f_t(s - u)| du \\ &\quad + \int_0^s |f_t(s - u) - f_t((s - u)_n)| du. \end{aligned}$$

Finally appealing once again to the  $K$ -Lipschitzness of  $f_t(s)$ ,

$$\int_0^s |\hat{f}_{t_n}(s_n - u_n) - f_t(s - u)| du \lesssim \frac{s}{n} + \int_0^s |\hat{f}_{t_n}((s - u)_n) - f_t(s - u)| du.$$

Combining all these upper-bounds together, we have obtained

$$|\hat{f}_{t_n}(s_n) - f_t(s)| \lesssim \frac{s}{n} + C \int_0^s |\hat{f}_{t_n}((s - u)_n) - f_t(s - u)| du. \quad (2.15)$$

For  $n \in \mathbb{N}^*$  and  $t \in \mathbb{R}_+$  fixed, we now introduce the following functions defined for  $s \in [0, t]$  by

$$\begin{aligned} s &\mapsto h_n(s) := \frac{s}{n}, \\ s &\mapsto g_{n,t}(s) := |\hat{f}_{t_n}(s_n) - f_t(s)|. \end{aligned}$$

Note in particular that

$$g_{n,t}(s - u) = |\hat{f}_{t_n}((s - u)_n) - f_t(s - u)|,$$

which means that we can rewrite (2.15) by substituting these functions, and get

$$g_{n,t}(s) \lesssim h_n(s) + C \int_0^s g_{n,t}(s - u) du.$$

The functions  $g_{n,t}(s)$  and  $h_n(s)$  are measurable and the conditions of Lemma B.1.1 are satisfied. We can apply Grönwall's Proposition B.1 for measurable functions and deduce that there exists a constant  $C > 0$  such that

$$g_{n,t}(s) \leq h_n(s) e^{Cs}.$$

Now it suffices to note that for a given  $s > 0$ ,  $\lim_{n \rightarrow \infty} h_n(s) = 0$ . Hence, for any  $t \geq s \geq 0$ , it

follows that  $\lim_{n \rightarrow \infty} g_{n,t}(s) = 0$  which translates into

$$\lim_{n \rightarrow \infty} \hat{f}_{t_n}(s_n) = f_t(s).$$

Since  $t_n \rightarrow t$  and  $s_n \rightarrow s$  as  $n \rightarrow \infty$ , we have shown that the discretised version  $\hat{f}_{t_n}(s_n)$  tends to  $f_t(s)$  as  $n \rightarrow \infty$ . Algorithm 1 is therefore a valid discretisation scheme when adopting this type of convergence. In other words, Theorem 2.1 mathematically validates the practical usage of Algorithm 1 whose purpose is to approximate  $f_t(s)$  by its discrete counterpart  $\hat{f}_{t_n}(s_n)$ .  $\square$

## 2.5 The renewal incidence equation

### 2.5.1 Incidence in infectious disease modelling

In infectious disease modelling and, more generally, in the context of epidemiology, the study of the dynamics of populations is a crucial ingredient to better understand the causes and spreads of diseases, whether on a geographical and/or on a temporal basis. Typically, integral equations of renewal type prove a very useful mathematical tool to study not only the asymptotic behaviour but also *stationary regime*, when it exists, of population dynamics.

As described in the introduction of [Pak+23], renewal equations are commonly used in modelling the number of new infections, i.e. incidence, arising at a given time. We mention here Fraser [Fra07] who presents an estimator for the *instantaneous reproduction number*  $R(t)$  for  $t \geq 0$ , defined as a function of calendar time and that represents the expectation of the number of secondary infections, assuming that nothing else changes in the environment. The function  $t \mapsto R(t)$  is mathematically defined on  $\mathbb{R}_+$  by Fraser as the integral of a *transmissibility rate*  $\beta(t; \tau)$  such that

$$R(t) = \int_0^\infty \beta(t; \tau) d\tau,$$

where  $t$  represents calendar time and  $\tau$  is the age of infection (i.e. time since infection). In [Fra07], the link between reproduction number and incidence is thus made through the transmissibility rate  $\beta(t; \tau)$ , wherein incidence is commonly reported in practice as a discrete time series of the number of incident cases in a short time interval  $[t; t + dt]$ , and its average, denoted  $I(t)$ , is assumed to follow the renewal equation

$$I(t) = \int_0^\infty \beta(t; \tau) I(t - \tau) d\tau, \quad t > 0, \quad (2.16)$$

with appropriate initial conditions at  $t = 0$  to avoid the obvious null solution. The function  $t \mapsto I(t)$  is defined as *the average number of new infections appearing at a given time point*  $t \geq 0$ .

From a mathematical point of view, incidence can equivalently be seen as the derivative of cumulative incidence (assuming that the function representing cumulative incidence is differentiable). However, it is interesting to note that, as far as our knowledge of the literature extends, an incidence renewal equation such as (2.16) seems to have only been heuristically justified, while being accepted as a common modelisation tool for infectious disease. Despite the seemingly lack of mathematical evidence for the derivation of the incidence equation in epidemiology, this equation proves extremely useful in practice to forecast or simulate the

incidence of a disease. In fact, there are many other studies in the literature that reinforce this argument. We can cite [Cor+13] who present a model to estimate the  $R$  number from incidence time-series, but also [Nou+18] who simulate future incidence using a branching process model based on the renewal equation, given originally in [Fra07].

In summary, on the one hand, the ability to simulate incidence (as opposed, for example, to being limited to model cumulative incidence or prevalence) opens many doors to a better understanding and modelisation of infectious disease dynamics, but on the other hand, the incidence renewal equation, which happens to be ubiquitously used in epidemiology, is only supported by heuristic arguments, such as the one based on transmissibility in [Fra07]. With this in mind, the work presented on the TV-CMJ branching processes in Section 1.3 can now be linked to a more formal derivation of a renewal-like incidence equation. The link we make in this section and the proof of equivalence between what we have called the expected birth rate  $f_\sigma$  of a TV-CMJ branching process and incidence, in effect, provide a more formal mathematical basis for the derivation of the governing incidence equation in epidemiology.

Before we go further, though, and to enable us to navigate between the framework of TV-CMJ branching processes and the one of epidemiology, we introduce here a specific set of assumptions that will suffice to underlie the proofs in the rest of this entire section. These new technical conditions in effect combine the existence and uniqueness conditions (EU-1) and (EU-2) into one condition, which we denote (A4). But we also add the condition that the density  $\lambda(t; \sigma)$  of the expected reproduction process can be written as the product of two continuous separable functions, namely  $r(t + \sigma)$  and  $\gamma(t)$ . The proof of Proposition 2.4 highlights that this last assumption on  $\lambda(t; \sigma)$  is necessary to directly link the framework of the TV-CMJ branching process with the formulation of incidence in epidemiology.

**Assumptions** (Sufficient conditions for equivalence between incidence and expected birth rate).

(A4) Assume condition (EU-1) holds and there exist two continuous and non-negative functions  $t \mapsto r(t)$  and  $t \mapsto \gamma(t)$  defined on  $\mathbb{R}_+$ , independent of  $\sigma$ , such that for any  $(t; \sigma) \in \mathbb{R}_+ \times \mathbb{R}_+$ ,

$$\lambda(t; \sigma) = r(t)\gamma(t - \sigma), \quad \text{with} \quad \Gamma(t) := \int_0^t \gamma(s)ds,$$

For any  $(t; \sigma) \in \mathbb{R}_+ \times \mathbb{R}_+$  such that  $0 \leq \sigma \leq t$ , let  $\bar{\gamma}(t; \sigma) := \sup_{u \in [0, t-\sigma]} \gamma(u) < \infty$ .

(A5) For any  $(t; \sigma) \in \mathbb{R}_+ \times \mathbb{R}_+$  such that  $0 \leq \sigma \leq t$ , define  $\rho(t; \sigma) := \sup_{u \in [\sigma, t]} r(u) < \infty$ , and assume that

$$\rho(0; \sigma)\Gamma(0^+) < 1.$$

We proceed as follows. We first prove in Section 2.5.2 the existence and uniqueness of a solution to a renewal-like integral equation, which will then lead in Section 2.5.3 to the redefinition of incidence as a generalised function, which we will denote  $I_\sigma$ . The incidence equation of which  $I_\sigma$  is the solution can be viewed as an extension of the commonly used renewal incidence equation in epidemiology. Next in Section 2.5.4, we compare the renewal-like equation derived in Proposition 2.2 for the expected birth rate  $f_\sigma$  of a TV-CMJ branching process with the newly defined and extended incidence equation. We prove that even though, at first sight, the two equations to which  $f_\sigma$  and  $I_\sigma$  are solutions, are of different forms, they are nevertheless equivalent.

*Remark.* Section 2.6 will show that this equivalence is based on a Kolmogorov-like duality forward-backward.

### 2.5.2 Existence and uniqueness of a solution to a generalised incidence equation

In this section, since our aim is to extend the definition of incidence as a generalised function, we first show in Proposition 2.3 existence and uniqueness of a solution to an equation which will determine the absolute continuous part of incidence. Our proof here relies again, similarly to the proof of Proposition 2.1, on an argument based on Picard iteration.

**Proposition 2.3.** *Assume that conditions (A4) and (A5) hold. Let  $(\tilde{I}_n(t; \sigma))_{n \in \mathbb{N}}$  be defined for  $n = 0$  by  $\tilde{I}_0(t; \sigma) = 0$  when  $0 \leq t < \sigma$  and  $\tilde{I}_0(t; \sigma) = r(t)\gamma(t - \sigma)$  when  $t \geq \sigma \geq 0$ , and for  $n \geq 1$  by the system*

$$\tilde{I}_{n+1}(t; \sigma) = \begin{cases} 0, & 0 \leq t < \sigma, \\ r(t)\gamma(t - \sigma) + r(t) \int_0^{t-\sigma} \tilde{I}_n(t-s; \sigma)\gamma(s)ds, & t \geq \sigma. \end{cases} \quad (2.17)$$

*Then for any  $t, \sigma \in \mathbb{R}_+ \times \mathbb{R}_+$ , the sequence of functions  $(\tilde{I}_n(t; \sigma))_{n \in \mathbb{N}}$  converges towards the function  $\tilde{I} : \mathbb{R}_+ \times \mathbb{R}_+ \rightarrow \mathbb{R}_+$ , which is the unique solution of*

$$\tilde{I}(t; \sigma) = \begin{cases} 0, & 0 \leq t < \sigma, \\ r(t)\gamma(t - \sigma) + r(t) \int_0^{t-\sigma} \tilde{I}(t-s; \sigma)\gamma(s)ds, & t \geq \sigma. \end{cases}$$

*Proof.* Well-definiteness of the sequence  $(\tilde{I}_n(t; \sigma))_{n \in \mathbb{N}}$  is automatic. It is very straightforward to verify (simple induction) that  $0 \leq \tilde{I}_n(t; \sigma) \leq \tilde{I}_{n+1}(t; \sigma)$  for any  $t, \sigma \geq 0$ . Thus  $(\tilde{I}_n(t; \sigma))_{n \in \mathbb{N}}$  is a monotonous sequence. We now show that  $\tilde{I}_n(t; \sigma)$  converges pointwise. Then a direct application of the Monotone Convergence Theorem will yield existence of a solution to (2.17). In a last step, we will show that this solution is unique using a classical Grönwall argument.

Let us start with the existence part. To do so, we define for any given pair  $(t; \sigma) \in \mathbb{R}_+ \times \mathbb{R}_+$  another related sequence of functions  $(N_j(t; \sigma))_{j \in \mathbb{N}}$ , representing the increments of  $\tilde{I}_j(t; \sigma)$  such that  $N_0(t; \sigma) := r(t)\gamma(t - \sigma)$  and for  $j \geq 1$ , and  $t, \sigma \in \mathbb{R}_+ \times \mathbb{R}_+$ ,

$$N_j(t; \sigma) := \tilde{I}_j(t; \sigma) - \tilde{I}_{j-1}(t; \sigma).$$

In other words, for any  $n \in \mathbb{N}$  and  $\sigma, t \geq 0$  we can also write

$$\tilde{I}_n(t; \sigma) = \sum_{j=0}^n N_j(t; \sigma).$$

We show by induction on  $n \in \mathbb{N}$  that for a given pair  $t \geq \sigma \geq 0$ , the sequence  $(N_n(t; \sigma))_{n \in \mathbb{N}}$  is upper-bounded as per (Ind-H3). Our induction hypothesis is as follows

$$N_n(t; \sigma) \leq \bar{\gamma}(t; \sigma) \rho^{*(n+1)}(t; \sigma) \Gamma^{*n}(t - \sigma). \quad (\text{Ind-H3})$$

Let  $n = 0$ . Then,

$$N_0(t; \sigma) = r(t)\gamma(t - \sigma) \leq \bar{\gamma}(t; \sigma) \rho(t; \sigma),$$

ans thus (Ind-H3) holds at rank 0. Next take  $n \geq 1$ . We have

$$\begin{aligned} N_{n+1}(t; \sigma) &= r(t) \int_0^{t-\sigma} N_n(t-s, \sigma) \gamma(s) ds, \\ &\leq r(t) \int_0^{t-\sigma} \bar{\gamma}(t-s; \sigma) \rho^{*(n+1)}(t-s; \sigma) \Gamma^{*n}(t-\sigma-s) \gamma(s) ds, \end{aligned}$$

where we appealed to (Ind-H3) at rank  $n$ . Now it suffices to observe that

$$\bar{\gamma}(t-s; \sigma) = \sup_{u \in [0, t-s-\sigma]} \gamma(u) \leq \sup_{u \in [0; t-\sigma]} \gamma(u) = \bar{\gamma}(t; \sigma),$$

and equally,

$$\rho(t-s; \sigma) = \sup_{u \in [\sigma; t-s]} r(u) \leq \sup_{u \in [\sigma; t]} r(u) = \rho(t; \sigma).$$

Thus, returning to the upper-bound of  $N_{n+1}(t; \sigma)$ , we can write

$$\begin{aligned} N_{n+1}(t; \sigma) &\leq \bar{\gamma}(t; \sigma) \rho^{*(n+2)}(t; \sigma) \int_0^{t-\sigma} \Gamma^{*n}(t-\sigma-s) \gamma(s) ds, \\ &= \bar{\gamma}(t; \sigma) \rho^{*(n+2)}(t; \sigma) \Gamma^{*(n+1)}(t-\sigma), \end{aligned}$$

which ends the proof by induction of (Ind-H3).

Finally, let us define for any fixed  $\sigma \geq 0$ , the function  $t \mapsto \Delta(t) := \rho(t; \sigma) \Gamma(t)$ , and refer to Appendix A.5. The function  $\Delta(t)$  so defined satisfies the required assumptions of Lemma A.5.1. Indeed,  $t \mapsto \rho(t; \sigma) \Gamma(t)$  is non-decreasing and right-continuous. And furthermore, by Assumption (A5), we have  $\rho(0; \sigma) \Gamma(0) < 1$  and  $\rho(t; \sigma) \Gamma(t) < \infty$ . Thus we can apply Lemma A.5.1 and we get

$$\sum_{n=1}^{\infty} \Delta(t)^{*n} < \infty, \quad \forall t \geq 0,$$

in other words,

$$\sum_{n=0}^{\infty} N_n(t; \sigma) < \infty,$$

which can be rewritten, for  $(t; \sigma)$  fixed, in terms of limit for the sequence  $\tilde{I}_n(t; \sigma)$  as

$$\lim_{n \rightarrow \infty} \tilde{I}_n(t; \sigma) < \infty.$$

In summary, the sequence of functions  $(\tilde{I}_n(t; \sigma))_{n \in \mathbb{N}}$  is non-decreasing, uniformly bounded and point-wise convergent. An application of the Monotone Convergence Theorem allows us to pass to the limit in equation (2.17) and obtain

$$\lim_{n \rightarrow \infty} \tilde{I}_n(t; \sigma) = \tilde{I}(t; \sigma),$$

where  $\tilde{I}(t; \sigma)$  is solution of

$$\tilde{I}(t; \sigma) = \begin{cases} 0, & 0 \leq t < \sigma, \\ r(t)\gamma(t - \sigma) + r(t) \int_0^{t-\sigma} \tilde{I}(t - s; \sigma)\gamma(s)ds, & t \geq \sigma. \end{cases}$$

Existence of a solution now being proven, we show uniqueness of the solution by appealing to a standard Grönwall argument. Let us assume that there exist two solutions  $\tilde{I}^{(1)}(t; \sigma)$  and  $\tilde{I}^{(2)}(t; \sigma)$  to equation (2.17) and denote  $\tilde{J}(t; \sigma) := \tilde{I}^{(2)}(t; \sigma) - \tilde{I}^{(1)}(t; \sigma)$ . This function then satisfies  $\tilde{J}(t; \sigma) = 0$  for  $0 \leq t < \sigma$  and for  $t \geq \sigma$ ,

$$\tilde{J}(t; \sigma) = r(t) \int_0^{t-\sigma} \tilde{J}(t - s; \sigma)\gamma(s)ds.$$

A direct application of Proposition B.1 - more precisely equations (B.1.2) and (B.1.3) in appendix - leads to  $\tilde{J}(t; \sigma) = 0$  for  $t \geq \sigma$ . This provides uniqueness of solution to (2.17).  $\square$

### 2.5.3 Extension and reformulation of the incidence renewal equation

Strong of the previous existence and uniqueness result above, we now return to the renewal incidence equation (2.16) commonly used in epidemiology. In this section, we wish to generalise this renewal equation, with the view in the next section to compare it with the system of equations verified by the distributional derivative of cumulative incidence in the context of a TV-CMJ branching process.

Our starting point is based on works such as those of Champredon [CDE18], for instance, who rely on an incidence renewal equation which happens to be somewhat only heuristically justified. We aim at introducing a more generic extension to the *basic* renewal incidence equation given by

$$I(t) = R_0 \int_0^\infty I(t - s)\gamma(s)ds, \quad t > 0, \quad (2.18)$$

where  $R_0$  is the constant or *basic reproduction number* and  $\gamma(s)$  is the *intrinsic generation-interval distribution* representing the probability that an infection not only survives but also gets transmitted a number  $s$  days after occurrence [CDE18]. Note that, obviously, to obtain a solution that is not trivially null, an initial condition must be given in practice. Here, for simplicity, we assume that  $I(0) = 1$ , which means that the infection starts with a single *index case* at  $t = 0$ .

A first seemingly natural extension of equation (2.18), which actually proves to be very useful in practice, consists of rendering the reproduction number calendar-time dependent; in other words, we simply and directly substitute the  $R_0$  number by a function  $r(t)$  on  $\mathbb{R}_+$  within equation (2.18). Furthermore, and to establish the basis for a consistent comparison between the incidence renewal equation used in the modelling of infectious diseases and the renewal-like equation for the expected birth rate  $f_\sigma$  derived from a TV-CMJ branching process, we also now assume that the index case of the infectious disease can arise at any given time  $\sigma \geq 0$ .

Thus, to represent the incidence of an infectious disease, we are therefore led to introducing a slightly more general renewal equation than (2.18) for a newly defined function

$I(t; \sigma) : \mathbb{R}_+ \times \mathbb{R}_+ \rightarrow \mathbb{R}_+$  such that

$$I(t; \sigma) = \begin{cases} 1, & t = \sigma, \text{ (initial condition)} \\ r(t) \int_{(0, t-\sigma]} I(t-s; \sigma) \gamma(s) ds, & t > \sigma. \end{cases}$$

The next and final level of generalisation of the incidence equation (2.18) consists of reformulating it within the space of generalised functions, in order to stand on the same foot as the formalism of the previous section with respect to cumulative incidence  $B_\sigma$  and its distributional derivative  $f_\sigma$ .

**Definition 2.5.1** (Reformulation of incidence as a generalised function). Let  $I_\sigma$  represent the generalised function in  $\mathcal{D}'(\mathbb{R})$  uniquely defined for any test function  $\varphi \in \mathcal{C}_c^\infty(\mathbb{R})$  as

$$\langle I_\sigma, \varphi \rangle = \langle \delta_\sigma, \varphi \rangle + \int_0^\infty \varphi(t) \tilde{I}(t; \sigma) dt, \quad (2.19)$$

where  $\tilde{I} : \mathbb{R}_+ \times \mathbb{R}_+ \rightarrow \mathbb{R}_+$  is defined as the unique solution to

$$\tilde{I}(t; \sigma) = \begin{cases} 0, & 0 \leq t < \sigma \\ r(t) \gamma(t - \sigma) + r(t) \int_0^{t-\sigma} \tilde{I}(t-s; \sigma) \gamma(s) ds, & t \geq \sigma. \end{cases}$$

$I_\sigma$  is said to represent the incidence of an infectious disease whose calendar-time-dependent reproduction number is given by  $r(t)$  and whose intrinsic generation-interval distribution is given by  $\gamma(t)$ .

*Proof.* We verify that Definition 2.5.1 is well posed, in other words, it defines a unique generalised function  $I_\sigma$ . Indeed, equation (2.19) is the unique decomposition of a generalised function as the sum of a singular and a regular distributions. The singular part is uniquely (well defined) by the Dirac delta distribution. Besides, it has already been shown that there exists a unique solution  $\tilde{I}(t; \sigma)$  verifying the integral equation for the absolute part above. Thus, existence and uniqueness of  $\tilde{I}(t; \sigma)$  imply those of  $I_\sigma$ .  $\square$

*Remarks.* Before we move on to the proof of equivalence between  $f_\sigma$  and  $I_\sigma$  in Section 2.5.4, and now that incidence has been formally redefined as a generalised function, we briefly go back to assumption (A4) and make a couple of remarks. Let us note that the first term in equation (2.21), namely  $r(t) \gamma(t - \sigma)$ , as a consequence of condition (A4), satisfies condition (EU-2). Therefore, the latter condition is no longer needed. Next, in order to make the connection between the intrinsic generation interval distribution  $\gamma(\cdot)$  in epidemiology and the lifelength density  $g(\cdot)$  of a TV-CMJ branching process, we had to further assume that  $\gamma(\cdot)$  and thus  $\Gamma(\cdot)$  do not depend on  $\sigma \geq 0$ .

*Examples* (TV-BH and TV-IP branching processes). As possible examples of TV-CMJ branching processes we briefly verify here that condition (A4) holds under the two examples introduced in Chapter 1.

- **TV-BH process:** Here the density  $g(t)$  of the life-length random variable corresponds to the generation interval distribution  $\gamma(t)$  and the expected reproduction number  $R(t)$  corresponds to  $r(t)$ . Thus, to guarantee that the TV-BH process satisfies Assumption (A4), we need

$$\gamma(t) := g(t), \quad \Gamma(t) := G(t) \quad \text{and} \quad r(t) := R(t). \quad (\text{TV-BH})$$

- TV-IP process: Equally condition (A4) is satisfied with the following correspondence

$$\gamma(t) := \overline{G}(t), \quad \Gamma(t) = \int_0^t \overline{G}(s) ds \quad \text{and} \quad r(t) := \rho(t). \quad (\text{TV-IP})$$

■

#### 2.5.4 Equivalence between expected birth rate and incidence

The objective of this section is to prove equivalence between the two generalised functions we have introduced so far, namely the expected birth rate  $f_\sigma$  of a TV-CMJ process and incidence  $I_\sigma$ , solution to the extended epidemiology renewal equation. Our strategy of proof consists of comparing the system of equations verified on the one hand by  $I_\sigma \in \mathcal{D}'(\mathbb{R})$  determined in Definition 2.5.1 and on the other hand by  $f_\sigma \in \mathcal{D}'(\mathbb{R})$  of Proposition 2.2. We then prove under certain technical conditions that  $I_\sigma$  and  $f_\sigma$  are identical.

**Proposition 2.4** (Equivalence between expected birth rate and incidence equations). *Assume that conditions (A4) and (A5) hold. Let  $I_\sigma \in \mathcal{D}'(\mathbb{R})$  represent the generalised function uniquely determined by*

$$\langle I_\sigma, \varphi \rangle = \langle \delta_\sigma, \varphi \rangle + \int_0^\infty \varphi(t) \tilde{I}(t; \sigma) dt, \quad \varphi \in \mathcal{C}_c^\infty(\mathbb{R}_+)$$

where

$$\tilde{I}(t; \sigma) = \begin{cases} 0, & 0 \leq t < \sigma, \\ r(t)\gamma(t - \sigma) + r(t) \int_0^{t-\sigma} \tilde{I}(t - s; \sigma) \gamma(s) ds, & t \geq \sigma. \end{cases}$$

and let  $f_\sigma \in \mathcal{D}'(\mathbb{R})$  be the expected birth rate of a TV-CMJ uniquely determined by

$$\langle f_\sigma, \varphi \rangle = \langle \delta_\sigma, \varphi \rangle + \int_0^\infty \varphi(t) \tilde{f}(t; \sigma) dt, \quad \varphi \in \mathcal{C}_c^\infty(\mathbb{R}_+),$$

where

$$\tilde{f}(t; \sigma) = \begin{cases} 0, & 0 \leq t < \sigma, \\ r(t)\gamma(t - \sigma) + \int_{[0, t-\sigma]} \tilde{f}(t; s + \sigma) r(s + \sigma; \sigma) \gamma(s) ds, & t \geq \sigma. \end{cases}$$

Then for any  $\sigma \geq 0$ , and any  $\varphi \in \mathcal{C}_c^\infty(\mathbb{R})$ ,

$$\langle f_\sigma, \varphi \rangle = \langle I_\sigma, \varphi \rangle. \quad (2.20)$$

Although it was necessary to introduce Schwartz distributions to formally define the expected birth rate and incidence and, in particular, ensure that they correspond to differentiable mathematical quantities, we can, however, avoid unnecessary formal dealings in the space  $\mathcal{D}'(\mathbb{R})$  to prove equivalence. Indeed, the respective singular parts of  $f_\sigma$  and  $I_\sigma$  are identical and equal to  $\langle \delta_\sigma, \cdot \rangle$ , the Dirac delta distribution at  $\sigma$ . Thus, by the uniqueness of these decompositions, we can rephrase the *distributional equality* in (2.20) and simply prove that the respective solutions of (2.21) and (2.22) are identical.

**Theorem 2.2.** Let  $\tilde{I} : \mathbb{R}_+ \times \mathbb{R}_+ \rightarrow \mathbb{R}_+$  be the unique solution of

$$\tilde{I}(t; \sigma) = \begin{cases} 0, & 0 \leq t < \sigma, \\ r(t)\gamma(t - \sigma) + r(t) \int_0^{t-\sigma} \tilde{I}(t - s; \sigma)\gamma(s)ds, & t \geq \sigma. \end{cases} \quad (2.21)$$

and  $\tilde{f} : \mathbb{R}_+ \times \mathbb{R}_+ \rightarrow \mathbb{R}_+$  be the unique solution of

$$\tilde{f}(t; \sigma) = \begin{cases} 0, & 0 \leq t < \sigma, \\ r(t)\gamma(t - \sigma) + \int_{(0, t-\sigma]} \tilde{f}(t; s + \sigma)r(s + \sigma)\gamma(s)ds, & t \geq \sigma. \end{cases} \quad (2.22)$$

Then for any  $t \geq \sigma$ ,

$$\tilde{f}(t; \sigma) = \tilde{I}(t; \sigma).$$

*Proof.* To prove Theorem 2.2, we first introduce a sequence of functions  $(\tilde{f}_n(t; \sigma))_{n \in \mathbb{N}}$  whose successive terms verify an equation which mirrors equation (2.22). For  $n = 0$ , we let  $\tilde{f}_0(t; \sigma) = 0$  when  $0 \leq t < \sigma$  and  $\tilde{f}_0(t; \sigma) = r(t)\gamma(t - \sigma)$  when  $t \geq \sigma \geq 0$ , and for  $n \geq 1$  we define

$$\tilde{f}_{n+1}(t; \sigma) = \begin{cases} 0, & 0 \leq t < \sigma, \\ r(t)\gamma(t - \sigma) + \int_0^{t-\sigma} \tilde{f}_n(t; s + \sigma)r(s + \sigma)\gamma(s)ds, & t \geq \sigma. \end{cases} \quad (2.23)$$

In fact  $(\tilde{f}_n(t; \sigma))_{n \in \mathbb{N}}$  corresponds exactly to the sequence of functions  $(u_n(t; \sigma))_{n \in \mathbb{N}}$  introduced in the proof of Proposition 2.1, in the particular case where  $v(t; \sigma) := r(t)\gamma(t - \sigma)$  and  $\lambda(t; \sigma) := r(t)\gamma(t + \sigma)$ . Thus we already know by the proof of Proposition 2.1 that  $(\tilde{f}_n(t; \sigma))_{n \in \mathbb{N}}$  is a Picard iterate that converges uniformly to  $\tilde{f}(t; \sigma)$ , solution to (2.22).

We have two sequences of functions  $(\tilde{I}_n(t; \sigma))_{n \in \mathbb{N}}$  and  $(\tilde{f}_n(t; \sigma))_{n \in \mathbb{N}}$  that respectively converge towards the unique solutions  $\tilde{I}(t; \sigma)$  and  $\tilde{f}(t; \sigma)$  of equations (2.21) and (2.22). Thus to show that  $\tilde{I}(t; \sigma) = \tilde{f}(t; \sigma)$ , it suffices to show that both sequences  $(\tilde{f}_n(t; \sigma))_{n \in \mathbb{N}}$  and  $(\tilde{I}_n(t; \sigma))_{n \in \mathbb{N}}$  are identical term by term.

Let us show by induction on  $n \in \mathbb{N}$  that

$$\forall n \in \mathbb{N}, \quad \forall i \leq n \quad \tilde{f}_i(t; \sigma) = \tilde{I}_i(t; \sigma), \quad t, \sigma \geq 0. \quad (\text{Ind-H4})$$

We first verify that induction hypothesis (Ind-H4) holds for the base cases  $n = 0$  and  $n = 1$ . The case  $n = 0$  is true by construction. Then at rank  $n = 1$ , on the one hand, we rewrite (2.23) recalling the definition of  $\tilde{f}_0(t; \sigma) = r(t)\gamma(t - \sigma)$  and obtain

$$\tilde{f}_1(t; \sigma) = \begin{cases} 0, & 0 \leq t < \sigma, \\ r(t)\gamma(t - \sigma) + r(t) \int_0^{t-\sigma} r(s + \sigma)\gamma(s + \sigma)\gamma(t - s - \sigma)ds, & t \geq \sigma. \end{cases}$$

And on the other hand, we rewrite (2.17) substituting  $\tilde{I}_0(t; \sigma) = r(t)\gamma(t - \sigma)$  and obtain

$$\tilde{I}_1(t; \sigma) = \begin{cases} 0, & 0 \leq t < \sigma, \\ r(t)\gamma(t - \sigma) + r(t) \int_0^{t-\sigma} r(t - s)\gamma(t - s - \sigma)\gamma(s)ds, & t \geq \sigma. \end{cases}$$

An obvious change of variable  $\nu = t - s - \sigma$  in the above equation shows that  $\tilde{f}_1(t; \sigma) = \tilde{I}_1(t; \sigma)$  for any  $t, \sigma \geq 0$ . The induction hypothesis (Ind-H4) is verified for the two base cases  $n = 0$  and  $n = 1$ .

Now let us take  $n \geq 1$  and assume that for any  $i \leq n$ ,  $\tilde{f}_i(t; \sigma) = \tilde{I}_i(t; \sigma)$ . When  $0 \leq t < \sigma$ ,

we trivially have  $\tilde{f}_{n+1}(t; \sigma) = \tilde{I}_{n+1}(t; \sigma) = 0$ . Thus take  $t \geq \sigma$  and let us again rewrite (2.23) by using (Ind-H4) at rank  $n$  which lets us substitute  $\tilde{f}_n(t; \sigma)$  by  $\tilde{I}_n(t; \sigma)$ , such that

$$\tilde{f}_{n+1}(t; \sigma) = r(t)\gamma(t - \sigma) + \int_0^{t-\sigma} \tilde{I}_n(t; s + \sigma)r(s + \sigma)\gamma(s)ds. \quad (2.24)$$

Next we rely on (2.17) satisfied at rank  $n$ , to rewrite an expression for  $\tilde{I}_n(t; s + \sigma)$ , i.e.,

$$\tilde{I}_n(t; s + \sigma) = r(t)\gamma(t - s - \sigma) + r(t) \int_0^{t-s-\sigma} \tilde{I}_{n-1}(t - \nu; s + \sigma)\gamma(\nu)d\nu, \quad t \geq \sigma,$$

which we substitute in (2.24) to get

$$\begin{aligned} \tilde{f}_{n+1}(t; \sigma) &= r(t)\gamma(t - \sigma) + r(t) \int_0^{t-\sigma} \left( \gamma(t - s - \sigma) \right. \\ &\quad \left. + \int_0^{t-s-\sigma} \tilde{I}_{n-1}(t - \nu; s + \sigma)\gamma(\nu)d\nu \right) r(s + \sigma)\gamma(s)ds, \\ &= r(t)\gamma(t - \sigma) + r(t) \int_0^{t-\sigma} \gamma(t - s - \sigma)r(s + \sigma)\gamma(s)ds \\ &\quad + r(t) \underbrace{\int_0^{t-\sigma} \left( \int_0^{t-s-\sigma} \tilde{I}_{n-1}(t - \nu; s + \sigma)\gamma(\nu)d\nu \right) r(s + \sigma)\gamma(s)ds}_A. \end{aligned} \quad (2.25)$$

We now rewrite the integral term  $A$  above,

$$\begin{aligned} A &= \int_0^{t-\sigma} \left( \int_0^{t-s-\sigma} \tilde{f}_{n-1}(t - \nu; s + \sigma)\gamma(\nu)d\nu \right) r(s + \sigma)\gamma(s)ds, \\ &= \int_0^{t-\sigma} \left( \int_0^{t-\nu-\sigma} r(s + \sigma)\tilde{f}_{n-1}(t - \nu; s + \sigma)\gamma(s)ds \right) \gamma(\nu)d\nu, \end{aligned}$$

where we first relied on the induction hypothesis (i.e.  $\tilde{I}_{n-1}(t; \sigma) = \tilde{f}_{n-1}(t; \sigma)$ ) and on the ordinary Fubini theorem in the second line. At this stage, we recognise part of the recursive relationship for  $\tilde{f}_n(t - \nu; \sigma)$ . Indeed (2.23) gets rewritten

$$\int_0^{t-\nu-\sigma} r(s + \sigma)\tilde{f}_{n-1}(t - \nu; s + \sigma)\gamma(s)ds = \tilde{f}_n(t - \nu; \sigma) - r(t - \nu)\gamma(t - \nu - \sigma).$$

Thus  $A$  becomes

$$A = \int_0^{t-\sigma} \left( \tilde{f}_n(t - \nu; \sigma) - r(t - \nu)\gamma(t - \nu - \sigma) \right) \gamma(\nu)d\nu.$$

Returning now to the expression we were left with for  $\tilde{f}_{n+1}(t; \sigma)$  in (2.25), we get

$$\begin{aligned} \tilde{f}_{n+1}(t; \sigma) &= r(t)\gamma(t - \sigma) + r(t) \int_0^{t-\sigma} \gamma(t - s - \sigma)r(s + \sigma)\gamma(s)ds, \\ &\quad + r(t) \int_0^{t-\sigma} \left( \tilde{f}_n(t - \nu; \sigma) - r(t - \nu)\gamma(t - \nu - \sigma) \right) \gamma(\nu)d\nu. \end{aligned}$$

It suffices to notice that

$$r(t) \int_0^{t-\sigma} r(t-\nu) \gamma(t-\nu-\sigma) \gamma(\nu) d\nu = -r(t) \int_{t-\sigma}^0 r(s+\sigma) \gamma(s) \gamma(t-s-\sigma) ds,$$

to deduce that

$$\tilde{f}_{n+1}(t; \sigma) = r(t) \gamma(t-\sigma) + r(t) \int_0^{t-\sigma} \tilde{f}_n(t-\nu; \sigma) \gamma(\nu) d\nu,$$

which happens to be exactly the integral equation verified by  $\tilde{I}_n(t; \sigma)$ .

Finally, the uniqueness property of the respective solutions to equations (2.23) and (2.17) lets us conclude that for any  $n \in \mathbb{N}$  and  $t \geq \sigma$ ,

$$\tilde{f}_n(t; \sigma) = \tilde{I}_n(t; \sigma).$$

We can now end the proof of Theorem 2.2. Passage to the limit in (2.23) relies on the Dominated Convergence Theorem since it was proven that  $\tilde{f}_n(t; \sigma)$  converges uniformly (see proof of Lemma B.2.3), whereas passage to the limit in (2.17) appeals to the Monotone Convergence theorem, since it was proven that  $\tilde{I}_n(t; \sigma)$  is monotonous and converges pointwise. This ends the proof that the (absolute continuous part of) the expected birth rate of a TV-CMJ branching process as solution to equation (2.22) and the (absolute continuous part of) the incidence process as solution to equation (2.21) coincide. Finally as the singular parts of each of these distributions are equal, we can conclude that  $I_\sigma = f_\sigma$ .  $\square$

## 2.6 Backward-forward duality of branching processes

### 2.6.1 Brief literature context

One of our main references to support the study of integral equations derived from a TV-CMJ branching process was the early works of Prof. Kimmel [Kim82; Kim83]. Indeed, Kimmel in the early 1980's worked on renewal-like integral equations with applications to branching processes, where he treated the general case of multitype age-dependent branching processes in varying environment. He derived renewal-like integral equations for three main quantities in expectation, namely, the expected cumulative number of deaths, the expected cumulative number of births (cumulative incidence), and the expected number of particles alive at a given time (prevalence). Kimmel applied his results to the descriptions of the dynamics of malignant cell populations under the action of cytostatic agents [Kim83].

Despite the obvious analogies with [Kim82] and [Kim83], our set-up in this chapter and the previous one can be distinguished from Kimmel's along three dimensions: we consider an extension of a CMJ branching process, whereas Kimmel treats a less general age-dependent model; our TV-CMJ process is single type, when Kimmel's age-dependent process is multitype; the variations of our environment originate from a calendar-time dependence when Kimmel assumes a more general (abstract) varying environment. Despite these differences, we have found that our equations (see Section 2.1) for cumulative incidence and prevalence in the case of a TV-BH (and therefore age-dependent) branching process are fully consistent with those derived by Kimmel in the works mentioned above [Kim82, Eqs. 2.2, 2.7 and 3.4].

Kimmel's approach [Kim82] relies on viewing a branching process, in expectation, as a

dynamical system with control terms. He proves an equivalence result between two renewal-like integral equations verified by the expected cumulative number of deaths. As the author and one of his referees mentioned at the time, these two types of equations are reminiscent of the backward and forward Kolmogorov equations. However, interestingly enough, although Kimmel [Kim82] derives a backward equation for cumulative incidence, he does not derive at the time a forward equation for this quantity. The question can then be asked if we could complement Kimmel's set of renewal-like equations with a forward equation for cumulative incidence.

In parallel, Kimmel [Kim83] adopts a point process approach in which, as just mentioned, he derives three backward integral equations, respectively, for the expected cumulative number of deaths, for cumulative incidence, and for prevalence in a multitype age-dependent branching process in a varying environment. To derive these backward equations, he relies on a principle that mirrors the first-generation principle, which he calls the *principle of last generation*. Note as well that in the late 1980's and early 1990's, Kimmel and Arino (see, e.g. [AK93] and [KAA95]) studied the backward and forward duality of renewal-like integral equations derived from a branching process in the context of cell populations dynamics.

We have just noted that Kimmel [Kim82] did not derive a forward equation for cumulative incidence. Nor was it his objective at the time to fully derive dual equations for incidence (although it could be argued that Eq.(16) in his later work [KAA95] can be seen as a backward incidence equation). Thus, one of the objectives of this section is to link back some of Kimmel's earlier works to our present work on a TV-CMJ branching process. We show in Section 2.6.2 that, under some technical assumptions, it is actually possible to directly derive from Kimmel's work a forward equation for cumulative incidence. In effect, this latter equation nicely complements the sets of backward-forward equations originally described in [Kim82] and [Kim83]. Next in Section 2.6.3, (and again under some specific assumptions, in particular requiring absolute continuity), on the one hand, we rederive the equation for the expected birth rate of a TV-CMJ process as was done in Section 2.3.4 but this time, by simply and directly differentiating Kimmel's backward equation of cumulative incidence. And, on the other hand, starting instead from the newly derived forward equation for cumulative incidence, we in effect close the loop as we obtain, by direct differentiation, the renewal-incidence equation used in epidemiology.

This complementary perspective on duality backward-forward allows us to validate from another angle the form for the renewal incidence equation used in epidemiology. In other words, the incidence renewal equation can be viewed as a forward equation, dual to the backward equation, which was derived for the expected birth rate of a TV-CMJ process in the context of branching populations.

## 2.6.2 Forward equation for cumulative incidence

In this section, we complement the set of equations derived by Kimmel [Kim82; Kim83] by deriving a forward equation for the cumulative incidence of a TV-CMJ branching process. First, we recall some of the notations, assumptions, and results in [Kim82] and [Kim83]. Kimmel's main assumptions were stated in [Kim82, Theorem 1]. Transcribing them into the notations of this chapter, we summarise them below.

**Assumptions** (Based on Theorem 1 of [Kim82]).

(Ki-1)  $\Gamma(t; \sigma)$  defined on  $\mathbb{R}_+ \times \mathbb{R}_+$  is right-continuous with respect to  $(t; \sigma)$  and is non-decreasing

in  $t \geq 0$ . Additionally, for any  $\sigma \geq 0$ ,  $\Gamma(0; \sigma) = 0$ , and  $\lim_{t \rightarrow \infty} \Gamma(t; \sigma) = 1$  (that is,  $t \mapsto \Gamma(t; \sigma)$  is CDF).

(Ki-2) There exists a right-continuous, non-decreasing function  $t \mapsto \Gamma_0(t)$ , independent of  $\sigma$ , such that

$$\Gamma(t; \sigma) \leq \Gamma_0(t), \quad t, \sigma \geq 0.$$

(Ki-3) The function  $t \mapsto r(t)$  defined on  $\mathbb{R}_+$  is non-negative, measurable and bounded.

In the context of this section, cumulative incidence will, as before, be denoted  $B(t; \sigma)$  but we additionally introduce the function  $(t; \sigma) \mapsto D(t; \sigma)$ , defined on  $\mathbb{R}_+ \times \mathbb{R}_+$ , which represents the expected cumulative number of deaths in a TV-CMJ branching process, initiated by a single individual born at time  $\sigma \geq 0$ . One of the main results in [Kim82] is given in Theorem 1. It states and proves equivalence between the forward and backward equations derived for the cumulative number of deaths  $D(t; \sigma)$ . We recall these two equations here. First we let  $D(t; \sigma) = 0$  for  $0 \leq t < \sigma$  and for  $t \geq \sigma \geq 0$ ,

$$D(t; \sigma) := \Gamma(t - \sigma) + \int_{\sigma}^t r(u) \Gamma(u - \sigma; \sigma) d_u D(t; \sigma), \quad (\text{Forward-D})$$

$$D(t; \sigma) := \Gamma(t - \sigma) + \int_{\sigma}^t r(u) D(t; u) d_u \Gamma(u - \sigma; \sigma). \quad (\text{Backward-D})$$

With regards to cumulative incidence, Kimmel [Kim82, Theorem 2] only proved a backward equation which we recall below. Again we let  $B(t; \sigma) = 0$  for  $0 \leq t < \sigma$  and for  $t \geq \sigma \geq 0$ ,

$$B(t; \sigma) = 1 + \int_{\sigma}^t r(u) B(t, u) d_u \Gamma(u - \sigma; \sigma). \quad (\text{Backward-B})$$

If Kimmel does not derive a forward equation for  $B(t, \sigma)$ , he, however, additionally provides two equations that directly link the expected cumulative number of deaths  $D(t; \sigma)$  with cumulative incidence  $B(t; \sigma)$ . We give here these two equations that will be needed for our derivation of the forward equation for  $B(t; \sigma)$ . For  $t \geq \sigma \geq 0$ ,

$$D(t; \sigma) = \int_{\sigma}^t \Gamma(t - s; \sigma) d_s B(s; \sigma), \quad (\text{D-B})$$

$$B(t; \sigma) = 1 + \int_{\sigma}^t r(u) d_u D(u; \sigma). \quad (\text{B-D})$$

We now introduce absolute continuity assumptions with respect to  $t$  for  $t > \sigma$  in order to differentiate directly and derive on the one hand, the forward equation for cumulative incidence and on the other hand, in the next section, the backward and forward equations for incidence. These assumptions are as follows.

**Assumptions** (Directly adapted from [Kim82]).

(Ki-4) For any  $\sigma \geq 0$ , the function  $t \mapsto D(t; \sigma)$  defined on  $\mathbb{R}_+$  such that  $D(t; \sigma) = 0$  for  $t \in [0, \sigma)$  and solution to (Forward-D) and (Backward-D) for  $t \geq \sigma$  admits a continuous density  $d(t; \sigma)$  for  $t > \sigma$ , i.e.,

$$d_t D(t; \sigma) = d(t; \sigma) dt, \quad \text{for } t > \sigma.$$

(Ki-5) For any  $\sigma \geq 0$ , the function  $t \mapsto B(t; \sigma)$  defined on  $\mathbb{R}_+$  such that  $B(t; \sigma) = 0$  for  $t \in [0, \sigma)$  and solution to (Backward-B) for  $t \geq \sigma$  admits a continuous density  $b(t; \sigma)$  for  $t > \sigma$ , i.e.,

$$d_t B(t; \sigma) = b(t; \sigma) dt, \quad \text{for } t > \sigma.$$

(Ki-6) The CDF  $t \mapsto \Gamma(t; \sigma)$  defined on  $\mathbb{R}_+$  is independent of  $\sigma$ , i.e.,

$$\Gamma(t; \sigma) \equiv \Gamma(t), \quad \forall t \in \mathbb{R}_+,$$

and admits a continuous density  $\gamma(t)$  on  $\mathbb{R}_+$

$$d_t \Gamma(t) := \gamma(t) dt.$$

We now have all the elements to derive the forward integral equation for cumulative incidence.

**Lemma 2.6.1** (Forward equation for  $B(t; \sigma)$ ). Assume conditions (Ki-1) to (Ki-6) hold. Then for any  $t > \sigma$ ,

$$B(t; \sigma) = 1 + \int_{\sigma}^t r(u) \gamma(u - \sigma) du + \int_{\sigma}^t r(u) \int_{\sigma}^u \gamma(u - s) b(s; \sigma) ds du. \quad (\text{Forward-B})$$

*Proof.* Let us rewrite equation (Forward-D) and equation (B-D) using condition (Ki-4),

$$D(t; \sigma) = \Gamma(t - \sigma) + \int_{\sigma}^t \Gamma(t - u) r(u) d(u; \sigma) du, \quad t > \sigma, \quad (2.26)$$

$$B(t; \sigma) = 1 + \int_{\sigma}^t r(u) d(u; \sigma) du, \quad t > \sigma. \quad (2.27)$$

A simple application of the Leibniz rule (see Appendix B.5) to equation (2.27) leads to

$$b(t; \sigma) = r(t) d(t; \sigma), \quad t > \sigma,$$

which, in turn, allows us to rewrite equation (2.26) as

$$D(t; \sigma) = \Gamma(t - \sigma) + \int_{\sigma}^t \Gamma(t - u) b(u; \sigma) du.$$

Next we differentiate the above equation and again apply the Leibniz rule. We obtain for  $t > \sigma$

$$\begin{aligned} d(t; \sigma) &= \gamma(t - \sigma) + \Gamma(0) b(t; \sigma) + \int_{\sigma}^t \gamma(t - u) b(u; \sigma) du, \\ &= \gamma(t - \sigma) + \int_{\sigma}^t \gamma(t - u) b(u; \sigma) du, \end{aligned}$$

where we used that  $\Gamma(0) = 0$ . Finally, we return once again to equation (2.27) where we substitute  $d(u; \sigma)$  by its expression found in the previous equation. It follows for  $t > \sigma$

$$\begin{aligned} B(t; \sigma) &= 1 + \int_{\sigma}^t r(u) \left[ \gamma(u - \sigma) + \int_{\sigma}^u \gamma(u - s) b(s; \sigma) ds \right] du, \\ &= 1 + \int_{\sigma}^t r(u) \gamma(u - \sigma) du + \int_{\sigma}^t r(u) \int_{\sigma}^u \gamma(u - s) b(s; \sigma) ds du, \end{aligned}$$

which gives us the desired forward equation for cumulative incidence.  $\square$

### 2.6.3 Forward and backward incidence equations

To recap, in the previous section, we started from the works of Kimmel in the 1980s [Kim82; Kim83], but placed ourselves in the setup of a TV-CMJ branching process. We reviewed the main equations for cumulative incidence and the expected cumulative number of deaths and completed the set of equations originally derived by Kimmel with a new forward equation for  $B(t; \sigma)$ .

As already mentioned in the introduction of Section 2.6, Kimmel and Arino, in the mid-1990s, then treated the forward-backward duality of branching processes (see [KAA95], [KA94]). In particular, they mentioned two different types of modelling approaches [KA94], by which they compare systems of equations for *flux of just born [and] flux of just dividing cells*. These fluxes are reminiscent of the derivatives of the expected cumulative number of births and deaths, and the two systems can be viewed as backward and forward duals of each other. In [KAA95], the duality of these systems is even more clearly highlighted.

In this section, our starting point is the dual system of equations for cumulative incidence, given, respectively, by (Backward-B) and (Forward-B). We see that by simply differentiating each equation in parallel, we get back on the one hand the equation derived for the expected birth rate in a TV-CMJ branching process and, on the other hand, the renewal incidence equation used in epidemiology. This confirms the validity of the incidence renewal equation as the forward dual of the equation derived when a branching process viewpoint is taken.

Let us naturally denote for the remainder of this section

$$I(t; \sigma) := b(t; \sigma).$$

Basically, incidence  $I(t; \sigma)$  as introduced in Section 2.5 corresponds to the density  $b(t; \sigma)$  of this section.

**Lemma 2.6.2** (Backward-Forward incidence equations). *Assume conditions (Ki-1) to (Ki-6) hold. Then  $I(t; \sigma)$  verifies the following two dual backward and forward equations,*

$$I(t; \sigma) = r(t)\gamma(t - \sigma) + r(t) \int_{\sigma}^t I(s; \sigma)\gamma(t - s)ds, \quad t > \sigma, \quad (\text{Forward-I})$$

$$I(t; \sigma) = r(t)\gamma(t - \sigma) + \int_{\sigma}^t r(u)I(t; u)\gamma(u - \sigma)du, \quad t > \sigma. \quad (\text{Backward-I})$$

*Proof.* The derivation of the forward and backward equations for incidence arises now very naturally from directly applying the Leibniz rule, respectively, to the forward and backward equations for cumulative incidence. Let us start with equation (Forward-B) in Lemma 2.6.1. By applying the Leibniz rule, we get

$$I(t; \sigma) = r(t)\gamma(t - \sigma) + r(t) \int_{\sigma}^t \gamma(t - s)I(s; \sigma)ds, \quad t > \sigma.$$

This gives us exactly the incidence renewal equation of Section 2.5.4. Similarly, let us also

apply the Leibniz rule to equation (Backward-B). We obtain in this case

$$I(t; \sigma) = B(t, t)r(t)\gamma(t - \sigma) + \int_{\sigma}^t I(t, u)r(u)\gamma(u - \sigma)du, \quad t > \sigma.$$

Recalling that in our context we have assumed a single initial birth at time  $t = \sigma$ , it follows that  $B(t, t) = 1$ , which leads to the desired backward incidence equation.  $\square$

In summary, this detour into Kimmel's work ([Kim82; Kim83; KAA95]) has allowed us to observe that the equivalence between the expected birth rate and the incidence proven in Section 2.5.4 can simply be viewed as a duality forward-backward for the number of expected births at time  $t > \sigma$  in a TV-CMJ branching process. This duality was derived in this section from Kimmel's original work related to integral equations issued from multi-type age-dependent branching processes in random environment, which we generalised to TV-CMJ processes. The equivalence/duality shown in Section 2.5.4, on the other hand, was based mainly on analytical tools (Picard Iterates, Schwartz distribution ...).

Finally, in Section 2.7, and as a useful and practical application of the previous derivations, we demonstrate how the result of equivalence provided in Section 2.5.4 leads to a direct mathematical justification of the *backcalculations* commonly applied in the modelling of infectious diseases. In the following, we will first recall the basics of these backcalculations.

## 2.7 Consistency with backcalculation of incidence

### 2.7.1 Epidemiology context

The method of backcalculation in epidemiology and infectious disease modelling consists of recovering prevalence through a convolution of incidence data with the survival function of the generation interval. In broad terms, we can write

$$\text{Prevalence} = \text{Survival} \star \text{Incidence},$$

where  $\star$  denotes the convolution operator. As described in Birrell et al. [Bir+12], backcalculation is a technique used to reconstruct the history of infections based on the data provided by the time series of observed disease time points. This method was originally introduced and applied to estimate HIV and AIDS incidence data by Brookmeyer and Gail [BG88] and Cui and Becker [CB00] and later, to cite but a few, by Birrell et al. [Bir+12] and Brizzi et al. [Bri+19]. Note also Mezzetti and Robertson [MR99] who backcalculated cancer incidence rates, Deuffic et al. [Deu+99] who modelled the hepatitis C epidemic in France, or Crump [CM15] who backcalculated the incidence of leprosy infection. We are only citing very few references here, but this shortlist does not give credit to how widely backcalculation of incidence is used in practice and to how helpful this inference technique has proven to be, in particular, when combined with the Bayesian modelling framework.

In this section, we rely once more on the modelling analogy between the dynamics of an infectious disease (at least in its early phases) and those of a TV-CMJ branching process. We rely on the integral equations previously derived for the number of alive individuals at time  $t$  (prevalence) and the expected birth rate (incidence) of a TV-CMJ branching process to prove that the backcalculation so often applied in practice is consistent with those renewal-like integral equations.

### 2.7.2 Consistency

We assume that conditions (A4) and (A5) are true to ensure that Proposition 2.4 holds, which means that we have equivalence between the expected birth rate  $f_\sigma$  of a TV-CMJ process and incidence  $I_\sigma$  of an infectious disease. Prevalence at time  $t \geq 0$  in epidemiology corresponds to the expected number of individuals present (alive) in the population at that time. Recall the integral equation derived in Section 2.1 for prevalence of a TV-CMJ branching process initiated at time  $\sigma \geq 0$  by a single individual,

$$\Pr(t; \sigma) = 1 - G^\sigma(t - \sigma) + \int_0^{t-\sigma} \Pr(t; \sigma + u) d\Lambda(u; \sigma), \quad t > \sigma.$$

Note that here, by Assumption (A4),  $G^\sigma(t) \equiv \Gamma(t; \sigma) \equiv \Gamma(t) = \int_0^t \gamma(s) ds$  does not depend on  $\sigma$  and, in addition, we also have  $d\Lambda(t; \sigma) = r(t)\gamma(t - \sigma)$ . Thus we rewrite

$$\Pr(t; \sigma) = \bar{\Gamma}(t - \sigma) + \int_0^{t-\sigma} \Pr(t; \sigma + u) r(u + \sigma) \gamma(u) du, \quad t > \sigma,$$

where  $\bar{\Gamma}(t) := 1 - \Gamma(t)$  denotes the survival function, and again is independent of  $\sigma$ . On the other hand, recall that incidence, equivalently, the expected birth rate  $f_\sigma$ , has been defined as a generalised function in Proposition 2.2 and follows the system of equations

$$\langle f_\sigma, \varphi \rangle = \langle \delta_\sigma, \varphi \rangle + \int_{\mathbb{R}_+} \varphi(t) \tilde{f}(t; \sigma) dt, \quad \varphi \in \mathcal{C}_c^\infty(\mathbb{R}_+),$$

where  $\tilde{f} : \mathbb{R}_+ \times \mathbb{R}_+$  is the unique solution to the integral equation

$$\tilde{f}(t; \sigma) = \begin{cases} 0, & t < \sigma, \\ r(t)\gamma(t - \sigma) + \int_0^{t-\sigma} \tilde{f}(t; u + \sigma) r(u + \sigma) \gamma(u) du, & t \geq \sigma. \end{cases}$$

We wish to prove that the convolution of the survival function  $\bar{\Gamma}(t)$  with the Schwartz distribution  $f_\sigma$  gives back prevalence, that is,

$$f_\sigma * \bar{\Gamma}(t) = \Pr(t; \sigma), \quad t > \sigma.$$

Our strategy of proof is by now well tested: we rely on one hand on the integral equations verified by each of these latter quantities and on the other hand, on the uniqueness of solution of each of these equations. In particular, we show here that the integral equation satisfied by  $f_\sigma * \bar{\Gamma}(t)$  is identical to the integral equation verified by  $\Pr(t; \sigma)$  for any  $t > \sigma$ .

The proof of consistency of the backcalculation formula therefore relies on the convolution of a Schwartz distribution (a.k.a. generalised function) with a *classical* function given by the survival function. Here we refer the reader to Appendix B.1 which gives the formula for a convolution of a generalised function with an infinitely differentiable function with compact support. Thus, in our case, it is sufficient to assume that  $\bar{\Gamma}(t)$  is infinitely differentiable in  $\mathbb{R}_+$  to ensure that the convolution between  $f_\sigma$  and  $\bar{\Gamma}(t)$  is well defined. To enable us to apply this formula, we thus add the following assumption with respect to the differentiability of  $\Gamma(t)$ .

**Assumption** (Consistency of backcalculation).

(A6) The CDF  $t \mapsto \Gamma(t)$  is infinitely differentiable with compact support on  $\mathbb{R}_+$ , i.e.  $\Gamma(t) \in \mathcal{C}_c^\infty(\mathbb{R}_+)$ .

**Proposition 2.5** (Consistency with backcalculation). *Under Assumptions (A4), (A5) and (A6), we have*

$$Pr(t; \sigma) = (\bar{\Gamma} * f_\sigma)(t), \quad (2.28)$$

for any  $t > \sigma \geq 0$ .

*Proof.* First, by Assumption (A6),  $\Gamma(t) \in \mathcal{C}_c^\infty(\mathbb{R}_+)$ , and thus the function  $t \mapsto \bar{\Gamma} * f_\sigma(t)$  is well defined (see Appendix B.1). Next, recalling that all the integral equations under consideration admit a unique solution, we observe that to prove equality in (2.28), it suffices to prove that the function  $t \mapsto (\bar{\Gamma} * f_\sigma)(t)$  satisfies the same integral equation as  $Pr(t; \sigma)$ . In other words, we wish to prove that

$$(\bar{\Gamma} * f_\sigma)(t) = \bar{\Gamma}(t - \sigma) + \int_0^{t-\sigma} (\bar{\Gamma} * f_{\sigma+u})(t) r(u + \sigma) \gamma(u) du, \quad t > \sigma. \quad (2.29)$$

Let us first focus on the right hand-side of equation (2.29) and start by rewriting  $(\bar{\Gamma} * f_{\sigma+u})(t)$ . Indeed we can directly apply the result of Appendix B.1 on the convolution of a generalised function  $T$  with a function  $\psi \in \mathcal{C}_c^\infty(\mathbb{R}_+)$ , which we recall here for convenience,

$$(T * \psi)(t) = \langle T, \tau_t \check{\psi} \rangle = \langle T, \psi(t - \cdot) \rangle,$$

where  $\check{\psi}(t) = \psi(-t)$  and  $\tau_u \psi(t) = \psi(t - u)$ . Thus,

$$\begin{aligned} (\bar{\Gamma} * f_{\sigma+u})(t) &= f_{\sigma+u}(\bar{\Gamma}(t - \cdot)), \\ &= \langle \delta_{\sigma+u}, \bar{\Gamma}(t - \cdot) \rangle + \int_{\mathbb{R}_+} \bar{\Gamma}(t - s) \tilde{f}(s, \sigma + u) ds, \\ &= \bar{\Gamma}(t - \sigma - u) + \int_0^t \bar{\Gamma}(t - s) \tilde{f}(s, \sigma + u) ds. \end{aligned}$$

Next we rewrite the integral term on the right hand side of equation (2.29) substituting  $(\bar{\Gamma} * f_{\sigma+u})(t)$  with the above equality. We get

$$\begin{aligned} \int_0^{t-\sigma} (\bar{\Gamma} * f_{\sigma+u})(t) r(u + \sigma) \gamma(u) du &= \int_0^{t-\sigma} r(u + \sigma) \gamma(u) \left( \bar{\Gamma}(t - \sigma - u) \right. \\ &\quad \left. + \int_0^t \bar{\Gamma}(t - s) \tilde{f}(s, \sigma + u) ds \right) du, \\ &= \int_0^{t-\sigma} r(u + \sigma) \gamma(u) \bar{\Gamma}(t - \sigma - u) du \\ &\quad + \int_0^{t-\sigma} r(u + \sigma) \gamma(u) \left( \int_0^t \bar{\Gamma}(t - s) \tilde{f}(s, \sigma + u) ds \right) du. \end{aligned}$$

Focusing now on the left-hand side of (2.29), we rely on equation (2.5) satisfied by the generalised function  $f_\sigma$  in Proposition 2.2,

$$\begin{aligned} (\bar{\Gamma} * f_\sigma)(t) &= \langle f_\sigma, \bar{\Gamma}(t - \cdot) \rangle, \\ &= \bar{\Gamma}(t - \sigma) + \int_0^t \bar{\Gamma}(t - \nu) \tilde{f}(\nu, \sigma) d\nu. \end{aligned}$$

We can trivially cancel out the first term  $\bar{\Gamma}(t - \sigma)$  on each side. We are therefore left to prove

$$\begin{aligned} \int_0^t \bar{\Gamma}(t - \nu) \tilde{f}(\nu, \sigma) d\nu &= \int_0^{t-\sigma} r(u + \sigma) \gamma(u) \bar{\Gamma}(t - \sigma - u) du \\ &\quad + \int_0^{t-\sigma} r(u + \sigma) \gamma(u) \left( \int_0^t \bar{\Gamma}(t - s) \tilde{f}(s, \sigma + u) ds \right) du. \end{aligned} \quad (2.30)$$

To rewrite the left-hand-side of equation (2.30), we first proceed to a simple change of variable in the integral equation (2.22) satisfied by  $\tilde{f}(t; \sigma)$ ,

$$\tilde{f}(\nu; \sigma) = r(\nu) \gamma(\nu - \sigma) + \int_\sigma^\nu \tilde{f}(\nu; p) r(p) \gamma(p - \sigma) dp.$$

Then we substitute  $\tilde{f}(\nu, \sigma)$  and get

$$\int_0^t \bar{\Gamma}(t - \nu) \tilde{f}(\nu, \sigma) d\nu = \int_0^t \bar{\Gamma}(t - \nu) \left( r(\nu) \gamma(\nu - \sigma) + \int_\sigma^\nu \tilde{f}(\nu; p) r(p) \gamma(p - \sigma) dp \right) d\nu.$$

Next we trivially expand and operate another change of variable in each of the integral terms,

$$\begin{aligned} \int_0^t \bar{\Gamma}(t - \nu) \tilde{f}(\nu, \sigma) d\nu &= \int_0^t \bar{\Gamma}(t - \nu) r(\nu) \gamma(\nu - \sigma) d\nu + \int_0^t \bar{\Gamma}(t - \nu) \int_\sigma^\nu \tilde{f}(\nu; p) r(p) \gamma(p - \sigma) dp d\nu, \\ &= \int_0^{t-\sigma} \bar{\Gamma}(t - s - \sigma) r(s + \sigma) \gamma(s) ds \\ &\quad + \int_0^t \bar{\Gamma}(t - \nu) \int_0^{\nu-\sigma} \tilde{f}(\nu; s + \sigma) r(s + \sigma) \gamma(s) ds d\nu. \end{aligned}$$

We then keep the first integral above unchanged as we recognise the first term in the sum of (2.30) and focus on the second double integral where by we eventually apply the Unsymmetric Fubini theorem (see Appendix A.4) in the last line below,

$$\begin{aligned} &\int_0^t \bar{\Gamma}(t - \nu) \int_0^{\nu-\sigma} \tilde{f}(\nu; s + \sigma) r(s + \sigma) \gamma(s) ds d\nu \\ &= \int_0^t \bar{\Gamma}(t - \nu) \mathbf{1}_{0 \leq \nu \leq t} \int_0^{\nu-\sigma} \mathbf{1}_{0 \leq s \leq \nu-\sigma} \tilde{f}(\nu; s + \sigma) r(s + \sigma) \gamma(s) ds d\nu, \\ &= \int_0^t \int_0^{\nu-\sigma} \mathbf{1}_{0 \leq \nu \leq t} \mathbf{1}_{0 \leq s \leq \nu-\sigma} \bar{\Gamma}(t - \nu) \tilde{f}(\nu; s + \sigma) r(s + \sigma) \gamma(s) ds d\nu, \\ &= \int_0^t r(s + \sigma) \gamma(s) \left( \int_0^{t-\sigma} \mathbf{1}_{0 \leq \nu \leq t} \mathbf{1}_{0 \leq s \leq \nu-\sigma} \bar{\Gamma}(t - \nu) \tilde{f}(\nu; s + \sigma) d\nu \right) ds, \\ &= \int_0^{t-\sigma} r(s + \sigma) \gamma(s) \left( \int_0^t \bar{\Gamma}(t - \nu) \tilde{f}(\nu; s + \sigma) d\nu \right) ds. \end{aligned}$$

Combining the different parts, we have obtained that

$$\begin{aligned} \int_0^t \bar{\Gamma}(t - \nu) \tilde{f}(\nu, \sigma) d\nu &= \int_0^{t-\sigma} \bar{\Gamma}(t - s - \sigma) r(s + \sigma) \gamma(s) ds \\ &\quad + \int_0^{t-\sigma} r(s + \sigma) \gamma(s) \left( \int_0^t \bar{\Gamma}(t - \nu) \tilde{f}(\nu; s + \sigma) d\nu \right) ds, \end{aligned}$$

which we recognise as equality (2.30) and which is what was left to end the proof of Proposition 2.5.  $\square$

# CHAPTER 3

## On the strong law of large numbers for supercritical general branching processes<sup>†</sup>

### 3.1 Overview of structured branching Markov processes

In this chapter, we consider branching processes for structured populations. More specifically, we will focus on the asymptotic behaviour of a structured branching population where each individual in the population is characterised by a trait or a type whose dynamics follow a Markov process. We will place ourselves in the most general case where the reproduction rate depends on the individual trait (non-neutral branching) and the transmission law of the trait also depends on the trait of the mother. Moreover, the trait is assumed to belong to a general and abstract measurable space  $\mathcal{X}$ .

Many models in biology and ecology rely on structured branching processes and in particular benefit from relaxing the assumption of non-neutrality of the reproduction law. For example, in the case of spatial models, the reproduction law may depend on the location through birth. Mortality can also be seen as sensitive to the available quantity of resources or, more generally, to the environment (e.g. exposure to light, temperature ...) or to the presence of predators. Similarly in biology, when considering cell division, traits such as size, age, weight, or pathogen load of the cell can affect its survival and division. Typically, in biology, the trait could be phenotypical, but since branching processes for structured populations are of relevant practical importance in other scientific domains (e.g. neutron transport theory [HKV20], queuing theory [BC18], epidemiology, ecology, demography, finance - see workshop [Mol+10] ...), we will not necessarily restrict ourselves to specific types of traits in this chapter and consider a case with a general abstract type space.

We wish to study more particularly the asymptotic behaviour of branching Markov processes in an abstract space in the case where branching is supercritical, i.e. the mean number of progeny is strictly higher than 1, and the underlying Markov process representing the trait dynamics is recurrent, which means that the trait distribution converges towards a stationary probability measure. Our aim is to provide a convergence result that in part unifies but also extends the current literature. With the latter in mind, and in order to provide context to the results presented further down in this chapter, we will first (although in a very

---

<sup>†</sup>This chapter is based on an article that will be submitted on ArXiv in the early part of 2025. It is co-authored with Prof. Bansaye and Dr. Cloez.

brief and summarised manner) recount some of the main results proven in the past 50 years regarding asymptotic behaviour and the laws of large numbers for Markov branching processes. As it is impossible, in the bare framework of this overview, to give full credit to the depth and width of the literature in this field, we will limit ourselves to the main models and results on which this chapter relies.

Starting from *simplest models* and to the best of our knowledge, the very first works on the asymptotic behaviour of Galton-Watson processes in the supercritical case can be attributed to Harris [Har63], Kesten and Stigum [KS66], and Athreya [AK71]. They concern single-type or  $p$ -multitype Galton-Watson processes, where the trait space is of finite dimension  $p \in \mathbb{N}$ . To allow us to discuss classical convergence results and extinction criteria in technical terms, we first introduce a few notations in the simpler case of finite-dimensional space.

In the multitype finite-dimensional case, the trait space  $\mathcal{X}$  can be seen as a finite subset of integers  $\{1, \dots, p\}$ . Denoting  $(e_i, 1 \leq i \leq p)$  the canonical basis of  $\mathbb{R}^p$ , a  $p$ -multitype Galton-Watson process is a Markov chain  $(Z_n = (Z_{n,1}, \dots, Z_{n,p}))_{n \in \mathbb{N}}$  valued in  $\mathbb{N}^p$  whose probability kernel is given for  $\mathbf{j} \in \mathbb{N}^p$  and  $1 \leq i \leq p$  by the transition probabilities

$$\mathbb{P}(Z_1 = \mathbf{j} \mid Z_0 = e_i) = p_i(\mathbf{j}).$$

The chain verifies the following branching property

$$Z_{n+1} = \sum_{i=1}^p \sum_{k=0}^{Z_{n,i}} \xi_{k,i}, \quad \text{for any } n \geq 0,$$

where  $(\xi_{k,i}, 1 \leq i \leq p, k \geq 0)$  are independent random variables in  $\mathbb{N}^p$  whose distribution is given by

$$\mathbb{P}(\xi_{k,i} = \mathbf{j}) = p_i(\mathbf{j}), \quad \forall k \geq 0.$$

Here, the  $i$ -th coordinate  $Z_{n,i}$  of the  $p$ -dimensional vector  $Z_n$  corresponds to the number of individuals of type  $i$  in the  $n$ -th generation.

The purpose of introducing these technical notations in our introduction is to allow us to present the classical nonextinction criteria for a multitype Galton-Watson process and, in passing, emphasise the spectral approach, which, besides the spinal approach (see e.g. [BM15], [Clo17], [BW20], [EW06], [Mar19]) is the technique most commonly adopted in the literature to obtain convergence results of branching processes. Athreya and Ney [AK71, Ch. 3] gave a thorough description of the limit behaviour of one-dimensional continuous-time Markov branching processes. In the single-type case, it is well known that the branching population becomes extinct if the mean number of progeny per individual  $m$  is smaller than or equal to 1. In the multitype case, extinction criteria classically depend on the maximum eigenvalue of the first-moment matrix  $S$  whose coefficients are defined as

$$S_{i,j} = \mathbb{E}(Z_{1,j}^{(i)}), \quad 1 \leq i, j \leq p.$$

where  $Z_{n,j}^{(i)}$  represents the number of progeny of type  $j$  in the  $n$ -th generation of an individual of type  $i$ . In particular, the coefficients  $\mathbb{E}(Z_{1,j}^{(i)})$  correspond respectively to the mean number of progeny of type  $j$  in the first generation coming from an individual of type  $i$ . When the mean matrix  $S$  of the reproduction law is primitive (i.e. irreducible aperiodic), the Perron-Frobenius theorem (Appendix C.1) ensures that there exists a unique eigentriple  $(\lambda, h, \gamma)$  where  $h$  is

a positive eigenfunction on  $\mathcal{X}$ ,  $\gamma$  is an eigenmeasure on  $\mathcal{X}$  and  $\lambda \in (0, \infty)$  is the maximal eigenvalue such that

$$Sh = \lambda h, \quad \gamma S = \lambda \gamma,$$

and

$$\gamma(h) = \sum_{i \in \mathcal{X}} h(x) \gamma_i = 1. \quad (\text{Normalisation})$$

Furthermore, there is exponential convergence

$$S_{i,j}^n \stackrel{n \rightarrow \infty}{\sim} \lambda^n h(i) \gamma_j, \quad 1 \leq i, j \leq p.$$

We note here that the finiteness of  $\mathcal{X}$  ensures uniformity of the convergence. Extinction criteria in the multidimensional Galton-Watson case were proved in [Har63, Ch. II]. If  $\lambda \leq 1$ , the process goes almost surely extinct, whereas if  $\lambda > 1$ , the probability of extinction is strictly smaller than 1. Typically, it can be proven that the probability of extension (or of explosion) is determined by the fixed point of the reproduction generating function.

When the type or trait space is of finite dimension, the main result describing the asymptotic behaviour of a multitype Markov branching process is due to Kesten and Stigum [KS66], which we state in more detail in the Appendix C.2. Kesten and Stigum in effect provide a necessary and sufficient condition for a multidimensional Galton-Watson process  $(Z_n)_{n \in \mathbb{N}}$  with a mean number of offspring  $m$  to not be degenerate. In particular, they prove that when  $\lambda > 1$ , the martingale  $W_n = Z_n/m^n$  is non-degenerate and converges to a deterministic limit given by  $W\gamma$ . The necessary and sufficient condition of non-extinction (whereby non-extinction is defined as  $\mathbb{P}(W > 0) > 0$ ) is given by

$$\mathbb{E} \left( Z_{1,j}^{(i)} \log Z_{1,j}^{(i)} \right) < \infty, \quad 1 \leq i, j \leq p.$$

This condition has since been celebrated as the  *$L \log L$  or Kesten-Stigum condition* (where  $L$  represents the reproduction law of the process). An alternative proof of the Kesten-Stigum Theorem in a single-type setting was given by Lyons, Pemantle, and Peres [LPP95], which was then extended to a multitype setting by Kurtz, Lyons, Pemantle, and Peres [Kur+97].

Similarly to the spectral approach based on the mean reproduction matrix in the finite-dimensional case, in this chapter we adopt a semigroup and linear operator framework, which becomes especially relevant when  $\mathcal{X}$  is infinite. We will construct more rigorously the measure-valued Markov branching process in Section 3.2.1. For the moment, let us simply mention that measure-valued processes have been extensively used in the literature to study structured populations. We note in particular Fournier and Méléard [FM04] who study a spatially structured static population, or Bansaye and Méléard's exposition [BM15] on stochastic models for structured populations. We also cite Tran [Chi06] and Cloez [Clo17] who studied large population limits of measure-valued structured processes. Although not the focus of study in this chapter, we observe briefly here that measure-valued processes are in effect the microscopic version of macroscopic models, and one can find an abundant literature related to the connections between these two populations scales (see e.g. [FM04; Ban+11; Clo17]).

In the framework of this introduction, we will remain somewhat informal with respect to the definition of measure-valued processes, while more details will be presented in Section 3.2.1. Here, at this point, we consider that the branching process  $Z$  can be represented by its

empirical measure and at an individual level. Denoting  $\mathcal{U}$  the entire population, and  $\mathbb{G}_n \subset \mathcal{U}$  the subset of individuals belonging to generation  $n$ , the branching process can be represented as the following random measure

$$Z_n = \sum_{u \in \mathbb{G}_n} \delta_{Z(u)},$$

where  $Z(u)$  is the type of individual  $u$  and  $\delta_{Z(u)}$  is the Dirac mass (distribution) at trait value  $Z(u) \in \mathcal{X}$ . Thus, for any type  $x \in \mathcal{X}$ , denoting  $Z_n^{(x)} = Z_n(\{x\})$ , and for any non-negative function  $f$ , we have

$$Z_n(f) = \sum_{u \in \mathbb{G}_n} f(Z(u)) = \sum_{x \in \mathcal{X}} f(x) Z_n^{(x)}.$$

In the above equation, the first representation is individual-based whilst the second is based on the trait structure of the population.

Strong of this representation of a Markov branching process as a measure-valued process, we now consider the first-moment semigroup of a multitype branching process defined for  $x \in \mathcal{X}$  and  $f$  non-negative function on  $\mathcal{X}$  by

$$S_n f(x) = \mathbb{E}_{\delta_x}(Z_n(f)) = \mathbb{E} \left( \sum_{x \in \mathcal{X}} f(x) Z_n^{(x)} \mid Z_0 = \delta_{x_0} \right).$$

We observe that for  $x_0, y \in \mathcal{X}$ ,  $S_{x_0, y} = S_1 \mathbf{1}_y(x_0) = \mathbb{E}(Z_1^{(y)} \mid Z_0 = \delta_{x_0})$ . In finite setting, uniform convergence in the Perron-Frobenius result can be written as

$$\sup_{\substack{x \in \mathcal{X}, \\ |f| \leq h}} \left| \frac{S_n f(x)}{\lambda^n h(x) \gamma(f)} - 1 \right| \xrightarrow{n \rightarrow \infty} 0. \quad (3.1)$$

Note that above, we must have  $|f| \leq h$ , where  $h$  is the leading right eigenfunction of the first-moment semigroup and is often difficult (impossible) to know.

In infinite type spaces, several versions of the Kesten-Stigum theorem have been obtained; however, the extension to the infinite-dimensional case is delicate and has attracted a lot of attention until now. The law of large numbers for branching structured processes takes a more explicit meaning when we consider measure-valued branching processes, given that measure-valued processes in effect represent the empirical distribution. We recall that the law of large numbers for such processes has been proven in different contexts and in different versions. In particular in discrete time, Athreya and Kang [AK98a; AK98b] prove the convergence of the proportions of individuals with a certain type, where the set of types (traits) can be either discrete or continuous. However, and to our knowledge, the single general statement that extends the finite-dimensional case was derived more than 50 years ago (more precisely a decade after Kesten and Stigum [KS66]) by Asmussen and Hering [AH76]. We refer to Appendix C.3 for an overview of their main results. They prove, in both cases, i.e. in discrete and continuous time, non-degeneracy and an almost sure strong law of large numbers for a branching Markov process in an abstract space, under specific conditions for the first-moment semigroup and an  $L \log L$  condition. Their proofs rely on a spectral decomposition of the first-moment semigroup of the branching Markov process, and they apply their results in particular cases of branching diffusions.

The results of Asmussen and Hering [AH76] with respect to their law of large numbers

can be summed up as follows: starting from one single individual with type  $x \in \mathcal{X}$ , Asmussen and Hering prove in the supercritical case ( $\lambda > 1$ ), in discrete time at first, that

$$\mathbb{E}_{\delta_x}(W) = h(x), \quad \lim_{n \rightarrow \infty} \frac{Z_n(f)}{\lambda^n} = \gamma(f) W \quad \text{a.s.},$$

where  $W$  is a nonnegative random variable that only depends on the initial condition  $Z_0 = \delta_x$ , and the result holds for any  $f$  such that  $\sup_{\mathcal{X}} |f|/h < \infty$  under their set of conditions denoted  $(M)$  as detailed in Appendices C.3.1-C.3.4, which correspond to a uniform convergence assumption for the first-moment semigroup similar to the one given in (3.1).

Their result also relies on the following Kesten-Stigum-like moment condition

$$\mathbb{E}_{\gamma}(Z_1(h) \log^* Z_1(h)) := \int_{\mathcal{X}} Z_1(h) \log^* Z_1(h) d\gamma < \infty, \quad (3.2)$$

where the continuous function  $x \mapsto \log^* x$  on  $[0, \infty)$  is defined as

$$\log^* x = \begin{cases} x/e, & 0 \leq x \leq e, \\ \log x, & x > e. \end{cases} \quad (3.3)$$

On the one hand, the previous  $L \log L$  condition (3.2) of Asmussen and Hering is optimal. On the other hand, the uniform convergence (3.1) may be difficult to check, even when it is satisfied. Indeed, it requires direct knowledge of the eigentriple  $(\lambda, h, \gamma)$ , which is rare in practice. In this direction, let us mention recent works [HKV20; GHK22] and references therein, in particular regarding applications to neutron transport, which seek conditions that would be more easily verifiable in practice.

Observe that the set of assumptions of Asmussen and Hering denoted  $(M)$  (see Appendix C.3, equations C.3.1 to C.3.4) is similar to (3.1) which is itself reminiscent of the finite-dimensional case. In particular, Asmussen and Hering's first-moment semigroup condition  $(M)$  forces the second eigenfunction to be dominated by the first. Moreover, the uniform convergence assumed by Asmussen and Hering is actually not satisfied in general for branching processes with non-bounded type space or with bounded space but degenerate behaviour on the boundary. For example, for neutral (non-local) models where the mean number of offspring does not depend on the type  $x \in \mathcal{X}$ , the eigenfunction  $h$  is the identity, i.e.  $h = \mathbf{1}_{\mathbb{R}_+}$ , and condition (3.1) implies that the Markov chain following the typical type along the spine *comes down from infinity* [Ban+11; Clo17]. This means that the Markov chain returns to compact sets very fast when it starts from large values, which excludes behaviours like random walks with negative drift or subcritical branching processes for large values.

Among the research topics that have motivated the relaxation of uniform convergence of the renormalised first-moment, we cite (non-exhaustively) supercritical branching Brownian motion with absorption [LS20], branching diffusion [EHK10], super-diffusion [Eng09], Markov chain along non-homogeneous genealogies [BH15], or growth-fragmentation process [BW20; TBV22; BGY23; HW20]. Studies in these papers obtain strong laws of large numbers for various types of branching processes, and generally provide a counterpart to the Kesten-Stigum theorem for a given subclass of branching processes. However, they generally rely on some  $L^p$  moment condition, with  $p > 1$ , and also require quantitative estimates of the first-moment semigroup (including a good knowledge of its eigentriple).

Our aim in this chapter, based on a paper co-written with Prof. Bansaye and Dr. Cloez <sup>\*</sup>, is to present a general Kesten-Stigum theorem that unifies and extends the literature. We want to relax the uniformity in condition (3.1) whilst keeping minimal moment conditions and prove strong convergence of the empirical measure. We also complement Asmussen and Hering [AH76] by proving  $L^1$  convergence in discrete time. Almost sure convergence is proven in both discrete and continuous time. On this note, we believe that  $L^1$  convergence in continuous time can also be derived from our work <sup>†</sup>. We present a martingale decomposition into, respectively, an  $L^1$  and an  $L^2$  part that relies on the subtle truncation argument of [AH76]. But our proof also involves new ingredients. Indeed, it relies simultaneously on a family of martingales and on contraction results associated to the semigroup.

Let us explain and motivate here more specifically this contraction result, which leads to our assumption on the ergodic profile of the first-moment semigroup. In the neutral case, the existence of a stationary regime for the trait distribution is directly linked to the ergodicity of the Markov chain of a typical individual and to the existence of a Lyapunov function  $V^*$  for the associated Markov kernel [Ban+11; Clo17]. Spine techniques have provided various general meanings to a typical individual reflecting the distribution of the traits in the population, see, e.g. [Mar17; HKK10]. In the same vein, many Markovian approaches have recently been developed to analyse the long-term behaviour of nonconservative (non-Markovian) semigroups [CV16; Vel23; Ban+22; KM03; Ban+11]. They provide sufficient conditions for contraction in total weight variation. These contraction conditions involve in particular the existence of a Lyapunov function  $V^*$  that quantifies the role of the initial type in the speed of convergence of the renormalised semigroup. More generally, different approaches have allowed to quantify the convergence of non-conservative semigroups and to obtain quantitative convergence in (possibly weighted) total variation norm which relaxes the uniformity [KM03; KM12; Del04; DM02; DHJ23; BCG20; Ban+22; Vel23; CV23; CV16].

These latter works all provide sufficient (and sometimes necessary) conditions on the semigroup so that for any  $x \in \mathcal{X}$  and any  $n \in \mathbb{N}$ ,

$$\sup_{|f| \leq V^*} \left| \frac{S_n f(x)}{\lambda^n} - h(x) \gamma(f) \right| \leq C V^*(x) \eta^n,$$

where  $\eta \in (0, 1)$  and  $V^* \geq h$  is a Lyapunov function which quantifies the impact of the initial type in the speed of convergence and  $C$  is a constant. Note here that condition (3.1) of [AH76] amounts to requiring  $V^* = h$ . Among the new situations covered by our results, the case where  $V^* = 1$  and where  $h$  vanishes at a boundary is interesting; see Section 3.3.3 for a high-level description of potential applications of our results. The case of a nonbounded domain  $\mathcal{X}$ , where  $V^*$  grows to infinity faster than  $h$  provides another relevant class of examples; see, for instance, [Ban+22; TBV22; BGY23]. We also want to allow subgeometric convergence of the renormalised semigroup, as in [AH76]. We are motivated in particular by polynomial speed and refer to [CM23; CG24] and Section 3.3.3 for a presentation of possible applications of our results.

This chapter is organised as follows. In Section 3.2, we focus on the discrete-time framework. In particular in Section 3.2.1, we first give a general construction of our Markov branching process, whereas in Section 3.2.2, we introduce our main set of assumptions and the general result in discrete time. Then in Section 3.2.3 we define the contraction operator and prove the key decomposition of the renormalised measure. In Section 3.2.4, we show that the

<sup>\*</sup>The article will be submitted on ArXiv in the early part of 2025.

<sup>†</sup>Current work to be added to the article includes presenting a proof of  $L^1$  convergence in continuous time.

propagation of the  $L \log L$  condition holds under some additional assumptions with respect to the Lyapunov function  $V^*$  introduced in Theorem 3.1. In Section 3.2.5 we study the families of martingales involved in the decomposition of the normalised branching process. We then prove the main convergence result (a.s. and in  $L^1$ ) in discrete time, and in Section 3.2.6 we provide some additional estimates. Section 3.3 is devoted to the continuous-time framework. In particular, Section 3.3.1 presents the assumptions and the main theorem in continuous time. Next, from the discrete result just obtained, we prove in Section 3.3.2 the counterpart theorem in continuous time, namely the non-degenerescence of the limiting martingale and the a.s. convergence of the renormalised empirical measure. Our proofs and techniques are partly inspired here by the approach originally adopted by Asmussen and Hering [AH76]. Finally, we briefly present an example related to branching diffusions in Section 3.3.3 that illustrates the main results. (Note that more specific examples and applications will be detailed in the article to be submitted on ArXiv in the early part of 2025.)

## 3.2 Strong law of large of numbers in discrete time

### 3.2.1 Construction and definitions

Henceforth, we work on a probability space denoted  $(\Omega, \mathcal{F}, \mathbb{P})$  that satisfies the usual conditions. Let  $\mathcal{X}$  be a measurable abstract space endowed with its natural sigma-algebra  $\mathcal{B}_{\mathcal{X}}$ . We proceed now with the general construction of a discrete-time Markov branching process with measurable state space  $\mathcal{X}$  and, in particular, refer to the lecture notes by Bansaye [Ban24] for additional material concerning the construction of branching processes for population dynamics.

Similarly to Chapters 1 and 2, we rely on Ulam-Harris-Neveu notation to describe the branching population. Thus, we let

$$\mathcal{U} = \bigcup_{n=0}^{\infty} \mathbb{N}^n$$

represent the set of all possible individual labels. The initial individual is denoted  $\emptyset$ . An individual  $u \in \mathcal{U}$  may have  $K \in \mathbb{N}$  descendants, labelled, respectively  $u1, u2, \dots, uK$ . These notations allow identifying individuals while simultaneously retaining their genealogical background.

Here, each individual  $u \in \mathcal{U}$  of trait  $x \in \mathcal{X}$  is assumed to live independently and give birth to a random number  $L^x$  of progeny. The reproduction law of  $L^x$  is trait dependent. In addition, the distribution of the offspring traits depends on the mother's trait. More precisely, let  $X_k^x$  be the random variable representing the trait of the  $k$ -th child of an individual with trait  $x$ . Then trait transmission to children is modelled by a family of random vectors  $(\Theta_x)_{x \in \mathcal{X}}$ , whereby for any  $x \in \mathcal{X}$ ,  $\Theta_x = (X_k^x : k = 1, \dots, L^x)$ , is a random variable valued in  $\mathbb{X} = \bigcup_{k=0}^{\infty} \mathcal{X}^k$ , endowed with the sigma-algebra  $\bigcup_{k=0}^{\infty} (\mathcal{B}_{\mathcal{X}})^k$ .

If for each label  $u \in \mathcal{U}$ , we independently choose a family of random vectors  $(\Theta_x(u))_{x \in \mathcal{X}}$  sampled from the common distribution  $(\Theta_x)_{x \in \mathcal{X}}$ , then we obtain a random variable  $\Theta_x(u) = (X_k^x(u) : k = 1, \dots, L^x(u))$  where more specifically  $X_k^x(u)$  is the trait of the  $k$ -th offspring of the individual  $u$  whose trait is  $x$ . Finally, we also assume that  $\Theta : \mathcal{X} \times \Omega \rightarrow \mathbb{X}$  is measurable, where  $\mathcal{X} \times \Omega$  is endowed with the natural product sigma-algebra. We have now introduced all the elements sufficient to define recursively a Markov branching process, whereby we construct in

parallel the set of living individuals  $\bigcup_{n=0}^{\infty} \mathbb{G}_n$  and their traits. We start with a single individual.

**Definition 3.2.1** (Branching Markov process). The family of random variables  $(Z(u))_{u \in \mathbb{G}_n, n \in \mathbb{N}}$  initiated with one single individual denoted  $\emptyset$  of trait  $x_0 \in \mathcal{X}$  and with transmission random variables  $(\Theta_x(u))_{x \in \mathcal{X}, u \in \mathcal{U}}$ , is defined recursively by

$$\mathbb{G}_0 = \{\emptyset\}, \quad Z(\emptyset) = x_0,$$

and for all  $n \geq 0$ , for all  $u \in \mathbb{G}_n$  and  $k \in \llbracket 1, L^{Z(u)}(u) \rrbracket$ ,

$$\mathbb{G}_{n+1} = \{uk : u \in \mathbb{G}_n, k \in \llbracket 1, L^{Z(u)}(u) \rrbracket\}, \quad Z(uk) = X_k^{Z(u)}(u).$$

$\mathbb{G}_n$  is the set of individuals in generation  $n$  and  $Z(u)$  is the trait of individual  $u$ .

The above construction of the branching process  $(Z(u))_{u \in \mathbb{G}_n, n \in \mathbb{N}}$  is inductive and simultaneously defines the generations  $\mathbb{G}_n$  of living individuals at time  $n$  and the Markovian trait process  $Z(u)$  of each individual in  $\mathbb{G}_n$ . This definition implies that the law of  $(Z(u))_{u \in \mathbb{G}_n, n \in \mathbb{N}}$  is determined by the law of  $(\Theta_x)_{x \in \mathcal{X}}$  and by the value of the initial trait  $x_0$ . Observe that such a definition easily extends to any initial condition, including several individuals and random traits.

As briefly discussed in the overview (Section 3.1), we adopt the perspective of representing a branching Markov process as a measure-valued process. Such a representation allows one to simultaneously follow the dynamics of the entire population whilst keeping track of each individual dynamics. In order to introduce measure-valued processes, we first define the set  $\mathcal{M}_F(\mathcal{X})$  of non-negative finite measures on  $\mathcal{X}$  and then define  $\mathcal{M}_P(\mathcal{X}) \subset \mathcal{M}_F(\mathcal{X})$  as the subset of point processes on  $\mathcal{X}$  such as

$$\mathcal{M}_P(\mathcal{X}) := \left\{ \sum_{i=1}^n \delta_{x_i}; \ n \geq 0, \ x_1, \dots, x_n \in \mathcal{X} \right\},$$

where  $\delta_x$  represents the Dirac distribution at point  $x$ . We now define for any  $n \in \mathbb{N}$  the random point measure given by

$$Z_n = \sum_{u \in \mathbb{G}_n} \delta_{Z(u)}.$$

Such a process  $(Z_n)_{n \in \mathbb{N}}$  verifies both a Markov property and a branching property, which we state in a combined manner below as we introduce the process rooted in an individual  $u$ . Note that the branching property is directly inherited from the independence of the random variable  $(\Theta_x(u))_{x \in \mathcal{X}}$  for  $u \in \mathcal{U}$ .

**Corollary 3.1** (Process  $Z^{(u)}$  rooted in  $u$ ). *For any  $n \geq 0$ , conditionally on  $\mathcal{F}_n$ , the natural filtration associated to  $Z_n$ , the processes  $(Z^{(u)})_{u \in \mathbb{G}_n}$ , which we will call the processes rooted in  $u$ , defined as*

$$Z^{(u)} = (Z_p^{(u)})_{p \in \mathbb{N}}, \quad \text{where} \quad Z_p^{(u)} = \sum_{v: uv \in \mathbb{G}_{n+p}} \delta_{Z(uv)}, \quad (3.4)$$

*are independent of each other. Moreover,  $Z^{(u)}$  is distributed as the original process  $(Z_n)_{n \in \mathbb{N}}$  but started from  $\delta_{Z(u)}$ , where  $Z(u)$  is the trait of individual  $u$ .*

We now define for any measurable and non-negative function  $f$  on  $\mathcal{X}$  the quantity

$$S_1 f(x) = \mathbb{E}_{\delta_x} \left( \sum_{u \in \mathbb{G}_1} f(Z(u)) \right) = \mathbb{E}_{\delta_x}(Z_1(f)),$$

which is non-negative but can be infinite. To ensure that we can define the first-moment semigroup as a linear operator on a more general set of functions, we introduce a (*Lyapunov condition*). This condition will become redundant once the more specific moment Assumption (A2) will be introduced since the latter is in effect stronger than this preliminary (*Lyapunov condition*). Nevertheless, this Lyapunov-like condition below allows us at this stage of our presentation to properly define the first-moment semigroup on a space of functions.

Thus, for the time being, and in order to properly define the first-moment semigroup, we assume that there exists a positive function  $V^*$  on  $\mathcal{X}$  such that there exists a positive constant  $C > 0$  which satisfies for any  $x \in \mathcal{X}$ ,

$$S_1 V^*(x) \leq C V^*(x). \quad (\text{Lyapunov condition})$$

Now if we define the functional space  $\mathcal{B}^+(V^*)$  such as

$$\mathcal{B}^+(V^*) := \left\{ f \text{ measurable on } \mathcal{X}, \text{ non negative: } \sup_{x \in \mathcal{X}} \frac{f(x)}{V^*(x)} < \infty \right\},$$

the (*Lyapunov condition*) above guarantees that  $\mathcal{B}^+(V^*)$  is stable by  $S_1$ . Next, more generally for any  $n \geq 0$ , we define the non-negative and possibly infinite quantity

$$S_n f(x) = \mathbb{E}_{\delta_x} \left( \sum_{u \in \mathbb{G}_n} f(Z(u)) \right) = \mathbb{E}_{\delta_x}(Z_n(f)), \quad f \in \mathcal{B}^+(V^*).$$

In particular, we have  $S_n = S_1 \circ \dots \circ S_1 = S_1^n$ . Again  $\mathcal{B}^+(V^*)$  is stable by  $S_n$ . Indeed if  $f \in \mathcal{B}^+(V^*)$ , then  $S_1 f \in \mathcal{B}^+(V^*)$  and by iteration  $S_n f \in \mathcal{B}^+(V^*)$  for any  $n \in \mathbb{N}$ . We can then consider the space  $\mathcal{B}(V^*)$  of measurable functions  $f : \mathcal{X} \rightarrow \mathbb{R}$  which are dominated by  $V^*$ , in other words,

$$\mathcal{B}(V^*) := \left\{ f \text{ measurable on } \mathcal{X} : \sup_{x \in \mathcal{X}} \frac{|f(x)|}{V^*(x)} < \infty \right\},$$

which we endow with the norm  $\|\cdot\|_{\mathcal{B}(V^*)}$  defined for  $f \in \mathcal{B}(V^*)$  as

$$\|f\|_{\mathcal{B}(V^*)} := \sup_{x \in \mathcal{X}} \frac{|f(x)|}{V^*(x)}.$$

The normed space  $(\mathcal{B}(V^*), \|\cdot\|_{\mathcal{B}(V^*)})$  is then a Banach space. We now use a classical measure-theoretic argument and extend the definition of the linear operator  $S_1$  from the cone  $\mathcal{B}^+(V^*)$  to  $\mathcal{B}(V^*)$  by relying on the positive and negative parts of a function. In fact, we write

$$S_1 f(x) = S_1 f_+(x) - S_1 f_-(x),$$

where  $f_+(x) := f(x)\mathbf{1}_{f(x) \geq 0}$  and  $f_-(x) := -f(x)\mathbf{1}_{f(x) \leq 0}$ . The space  $\mathcal{B}(V^*)$  is stable by  $S_1$ . We now show that

**Lemma 3.2.1.**  $(S_n)_{n \in \mathbb{N}}$  is a positive semigroup, i.e. for any  $x \in \mathcal{X}$  and  $n \in \mathbb{N}$ ,

$$S_{n+1}f(x) = S_n(S_1f)(x) = S_1(S_nf)(x), \quad f \in \mathcal{B}(V^*).$$

*Proof.* Let us denote  $v \succ u$  when  $v$  is a descendant of  $u$ . We can then write any  $f \in \mathcal{B}(V^*)$ ,

$$\mathbb{E} \left( \sum_{u \in \mathbb{G}_{n+1}} f(Z(u)) \mid \mathcal{F}_n \right) = \mathbb{E} \left( \sum_{\substack{u \in \mathbb{G}_n, v \in \mathbb{G}_{n+1}, \\ v \succ u}} f(Z(v)) \mid \mathcal{F}_n \right) = \sum_{u \in \mathbb{G}_n} \mathbb{E} \left( \sum_{\substack{v \in \mathbb{G}_{n+1}, \\ v \succ u}} f(Z(v)) \mid \mathcal{F}_n \right).$$

But the branching property (3.4) ensures that

$$\mathbb{E} \left( \sum_{\substack{v \in \mathbb{G}_{n+1}, \\ v \succ u}} f(Z(v)) \mid \mathcal{F}_n \right) = S_1f(Z(u)),$$

and therefore, by the Law of Total Probability,

$$\begin{aligned} S_{n+1}f(x) &= \mathbb{E}_{\delta_x} \left( \mathbb{E} \left( \sum_{u \in \mathbb{G}_{n+1}} f(Z(u)) \mid \mathcal{F}_n \right) \right), \\ &= \mathbb{E}_{\delta_x} \left( \sum_{u \in \mathbb{G}_n} S_1f(Z(u)) \right) = S_n(S_1f)(x). \end{aligned}$$

The other identity is proved in a similar way by conditioning with respect to  $\mathcal{F}_1$ , or by using the associativity of the composition. It also follows directly that the space  $\mathcal{B}(V^*)$  is stable by  $S_n$ .  $\square$

### 3.2.2 Assumptions and main theorem

Now that we have presented the major elements needed to construct a measure-valued Markov branching process, let us recall that the aim of this chapter is to prove a strong law of large numbers for the empirical measure of a supercritical Markov branching process defined on an abstract space. We will indeed give a convergence result towards a deterministic limit under a set of newly derived assumptions. As explained above, our main set of assumptions is reminiscent of the conditions denoted  $(M)$  by Asmussen and Hering [AH76] (see Appendices (C.3.1)-(C.3.4)) since it also concerns the first-moment semigroup  $(S_n)_{n \in \mathbb{N}}$  associated with the branching process which we introduced in the previous section. However, we eliminated the need for uniform convergence and instead appealed to Lyapunov-like functions.

**Assumptions** (First-moment semigroup convergence).

(A1) There exists a positive measurable function  $V^* : \mathcal{X} \rightarrow (0, \infty)$  and an eigentriple  $(\lambda, \gamma, h)$  such that  $\lambda > 1$ ,  $\gamma$  is a probability measure on  $\mathcal{X}$ ,  $h : \mathcal{X} \rightarrow (0, \infty)$  is a positive measurable function such that for any  $n \in \mathbb{N}^*$ ,

$$\gamma S_n = \lambda^n \gamma, \quad S_n h = \lambda^n h, \quad \gamma(h) = 1,$$

and

$$\sup_{x \in \mathcal{X}} \frac{h(x)}{V^*(x)} < \infty, \quad \gamma(V^*) < \infty.$$

(A2) There exists a sequence of non-negative real numbers  $(a_n)_{n \in \mathbb{N}}$  such that for any  $n \geq 0$  and any  $x \in \mathcal{X}$ ,

$$\sup_{|f| \leq V^*} \left| \frac{S_n f(x)}{\lambda^n} - h(x) \gamma(f) \right| \leq V^*(x) a_n, \quad (3.5)$$

$$\sum_{k=1}^{\infty} \frac{a_k}{k} < \infty. \quad (3.6)$$

*Remarks.* Let us add a couple of remarks about this set of assumptions.

- Since each side of the inequality (3.5) is used repeatedly in the following proofs, we expand it below to potentially obtain a more direct visualisation of the respective lower and upper bounds that will constantly be relied upon in our proofs. For any  $n \in \mathbb{N}$  and  $x \in \mathcal{X}$  and for  $|f| \leq V^*$ ,

$$h(x) \gamma(f) - V^*(x) a_n \leq \lambda^{-n} S_n f(x) \leq h(x) \gamma(f) + V^*(x) a_n.$$

And more generally when  $f \in \mathcal{B}(V^*)$ , we will use

$$h(x) \gamma(f) - V^*(x) a_n \|f\|_{\mathcal{B}(V^*)} \leq \lambda^{-n} S_n f(x) \leq h(x) \gamma(f) + V^*(x) a_n \|f\|_{\mathcal{B}(V^*)}. \quad (3.7)$$

- We observe here that Asmussen and Hering [AH76] (see Appendix (C.3.3)) only need to assume that  $a_n \rightarrow 0$ , while we need slightly more. In effect, condition (3.6) implies that we need  $a_n \rightarrow 0$  faster than  $1/\log n$ . On the other hand, compared to [AH76], we relax the uniform convergence over  $\mathcal{X}$  by allowing the function  $V^*$  to be large compared to the eigenfunction  $h$ .
- Writing Assumption (3.5) for  $f = h$ , using that  $h$  is dominated by  $V^*$ , that  $\gamma(V^*) < +\infty$  and taking  $n = 1$ , we get

$$S_n h(x) \leq \lambda V^*(x) (1 + a_1),$$

which shows that Assumption (3.5) is stronger than the (*Lyapunov condition*) introduced above to ensure that the first-moment semigroup was well defined. Thus, in effect, assuming (3.5) implies that the first-moment semigroup  $(S_n)_{n \in \mathbb{N}}$  is stable and well defined in  $\mathcal{B}(V^*)$ .

We now state our main theorem in discrete-time which provides a law of large numbers for the empirical measure of the Markov branching process in the supercritical case.

**Theorem 3.1.** *Let Assumptions (A1) and (A2) hold. We further suppose that there exists a measurable function  $V : \mathcal{X} \rightarrow (0, \infty)$  such that  $h \in \mathcal{B}(V)$ ,  $V \leq V^*$  and for any  $k \in \mathbb{N}^*$ ,*

$$\sup_{x \in \mathcal{X}} \frac{\mathbb{E}_{\delta_x}(Z_k(V) \log^* Z_k(V))}{V^*(x)} < \infty. \quad (3.8)$$

Then for any  $x \in \mathcal{X}$  and  $Z_0 = x$ ,

$$\lim_{n \rightarrow \infty} \frac{Z_n(h)}{\lambda^n} = W^{(x)}, \quad \text{a.s. and in } L^1,$$

and the non-negative random variable  $W^{(x)} \geq 0$  is such that

$$\mathbb{E}_{\delta_x}(W^{(x)}) = h(x).$$

Besides, for any  $f \in \mathcal{B}(V)$ ,

$$\lim_{n \rightarrow \infty} \frac{Z_n(f)}{\lambda^n} = \gamma(f) W^{(x)}, \quad \text{a.s. and in } L^1.$$

We have denoted the limiting random variable  $W^{(x)}$  to emphasise that it depends only on the initial condition and not on  $f$ .

In practice, it is often sufficient to verify the  $L \log L$  condition (3.8) for  $k = 1$ . In that direction, we refer to Proposition 3.1 for a rigorous statement that ensures propagation over time of this moment condition. Our  $L \log L$  moment condition (3.8) is equivalent to the classical  $L \log L$  criterion of Kesten-Stigum for a Galton-Watson process with a single type or a finite number of types. More precisely, it is equivalent to Asmussen and Hering's condition (3.2), i.e.  $\mathbb{E}_\gamma(Z_1(h) \log^* Z_1(h)) < \infty$ , as soon as  $V^*$  is bounded. In general, as  $\gamma(V^*) < \infty$ , our  $L \log L$  condition (3.8) implies

$$\mathbb{E}_\gamma(Z_1(V) \log^* Z_1(V)) < \infty, \quad (3.9)$$

which itself implies the Asmussen-Hering condition (3.2) since  $h$  is dominated by  $V$ . In effect, we believe that the  $L \log L$  conditions (3.8) and (3.9) are close. More precisely, and to our knowledge, when it comes to applications, checking (3.9) often equates to proving (3.8).

The finite-dimensional case and the uniform case (3.1) of Asmussen and Hering [AH76] equate to choosing  $V^* = V = h$  in which all functions' spaces coincide. However, in our work here, new cases are considered that will prove useful in applications. In particular we consider the case  $h \ll V = V^* = 1$  where uniformity (3.1) fails, but where the equality  $V = V^*$  ensures that not only some property of coming down from infinity still holds, but also propagation in time of our moment condition holds. The case  $h \leq V \ll V^* \sim V \log(V)$  will also be relevant, where the decrease from infinity is only linear or geometric, excluding coming down from infinity. This situation matters in particular for branching diffusions and growth fragmentations. We refer to Section 3.3.3 for an application to branching diffusions with absorption.

### 3.2.3 Contraction operator and decomposition of the process

In this section we introduce a linear operator over the functions space  $\mathcal{B}(V^*)$  which will be very useful for the subsequent proofs and in particular when writing the decomposition of the renormalised process.

Let  $V^*$  be the function introduced by Assumptions (A1) and (A2). We define for  $r \in \mathbb{N}$

the linear operator

$$\begin{aligned} T_r : \mathcal{B}(V^*) &\rightarrow \mathcal{B}(V^*) \\ f &\mapsto T_r f := \lambda^{-r} S_r f - \gamma(f)h, \end{aligned}$$

as well as for  $n \in \mathbb{N}$ , its  $n$ -th iteration

$$T_r^n = T_r \circ \cdots \circ T_r,$$

with the convention that  $T_r^0 = \text{ID}$ , the identity operator. Observe that Assumption (A2) and in particular inequality (3.7) on the semigroup writes

$$\|T_r f\|_{\mathcal{B}(V^*)} \leq a_r \|f\|_{\mathcal{B}(V^*)},$$

and implies

$$\|T_r^n f\|_{\mathcal{B}(V^*)} \leq a_r^n \|f\|_{\mathcal{B}(V^*)}.$$

Given that  $a_r \rightarrow 0$  as  $r \rightarrow \infty$ , we obtain a contraction when  $a_r < 1$ . In practice, we will use the following exponential decrease

$$|T_r^n f(x)| \leq a_r^n V^*(x),$$

that holds for any  $f$  such that  $|f| \leq V^*$ ,  $x \in \mathcal{X}$  and  $n \geq 0$ . Finally recalling that  $\gamma$  is a left eigenmeasure for the semigroup, we observe that for  $r \geq 1$ ,

$$\gamma(T_r f) = \lambda^{-r} \gamma(S_r f) - \gamma(f)\gamma(h) = 0,$$

where we used the properties of the normalised eigentriple. We next introduce the normalised process defined for any  $f \in \mathcal{B}(V^*)$  by

$$X_n^{(r)}(f) := \frac{Z_{nr}(f)}{\lambda^{nr}}, \quad n \geq 0, \quad r \geq 1.$$

Observe that  $(X_n^{(1)}(h))_{n \in \mathbb{N}}$  is the classical non-negative martingale associated with the eigenfunction  $h$ . The actual target of our study is more specifically the process  $X_n^{(1)}(f)$  for  $f \in \mathcal{B}(V)$ , whose asymptotic behaviour we wish to describe.

We denote  $(\mathcal{F}_n)_{n \in \mathbb{N}}$  the natural filtration of the branching process  $(Z_n)_{n \in \mathbb{N}}$  and  $\mathcal{F}_n^{(r)} = \mathcal{F}_{nr}$  for  $n \geq 0$ ,  $r \geq 1$ . Our proofs will involve the martingale increments  $(\Delta_n^{(r)}(f))_{n \in \mathbb{N}}$  defined for any  $f \in \mathcal{B}(V^*)$  and  $n, r \geq 1$  by

$$\Delta_n^{(r)}(f) := X_n^{(r)}(f) - \mathbb{E}(X_n^{(r)}(f) | \mathcal{F}_{n-1}^{(r)}).$$

We can equivalently write

$$\Delta_n^{(r)}(f) = M_n^{(r)}(f) - M_{n-1}^{(r)}(f),$$

where  $(M_n^{(r)}(f))_{n \in \mathbb{N}}$  is the associated martingale starting from 0 given for  $n, r \geq 1$  by

$$M_n^{(r)}(f) = \sum_{i=1}^n \Delta_i^{(r)}(f). \tag{3.10}$$

*Remark.* Taking  $f = h$ , observe that the classical martingale gets written as

$$X_n^{(r)}(h) = X_0^{(r)}(h) + M_n^{(r)}(h).$$

Our next lemma will show that the normalised empirical measure  $X_n^{(r)}$  can be simply decomposed using the contraction operator  $T_r$ . The remainder term in the below decomposition, on the other hand, depends on the contraction operator  $T_r$  and on the martingales  $M_n^{(r)}$  introduced above. Our overall proof technique for the law of large numbers relies partly on this decomposition.

**Lemma 3.2.2.** *Under Assumptions (A1) and (A2), for any  $f \in \mathcal{B}(V^*)$ ,  $r \geq 1$  and  $n \geq 0$ , we have*

$$X_n^{(r)}(f) = X_0^{(r)}(T_r^n f) + \gamma(f) X_{n-1}^{(r)}(h) + R_n^{(r)}(f),$$

where the remainder  $R_n^{(r)}(f)$  is given by

$$R_n^{(r)}(f) = \sum_{i=1}^n \Delta_i^{(r)}(T_r^{n-i} f) = M_n^{(r)}(f) + \sum_{i=1}^{n-1} M_{n-i}^{(r)}(T_r^{i-1}(T_r - Id)f).$$

*Proof.* We start by decomposing the population belonging to generation  $\mathbb{G}_{nr}$  in terms of its ancestors in the generation  $\mathbb{G}_{(n-1)r}$ . We write  $v \succ u$  when  $v$  is a descendant of  $u$ . By relying on the branching property in the first equality, we obtain

$$\begin{aligned} \mathbb{E}(X_n^{(r)}(f) | \mathcal{F}_{n-1}^{(r)}) &= \lambda^{-(n-1)r} \sum_{u \in \mathbb{G}_{(n-1)r}} \mathbb{E} \left( \lambda^{-r} \sum_{v \in \mathbb{G}_{nr}: v \succ u} f(Z(v)) | \mathcal{F}_{n-1}^{(r)} \right), \\ &= \lambda^{-(n-1)r} \sum_{u \in \mathbb{G}_{(n-1)r}} \lambda^{-r} S_r f(Z(u)), \\ &= X_{n-1}^{(r)}(\lambda^{-r} S_r f), \\ &= X_{n-1}^{(r)}(T_r f) + \gamma(f) X_{n-1}^{(r)}(h), \end{aligned}$$

where we successively used the definition of the semigroup  $S_r$  in the second line, the definition of the normalised process  $X_{n-1}^{(r)}$  in the penultimate line and the definition of the contraction operator in the last line. Thus on one hand, we have just obtained

$$\mathbb{E}(X_n^{(r)}(f) | \mathcal{F}_{n-1}^{(r)}) = X_{n-1}^{(r)}(T_r f) + \gamma(f) X_{n-1}^{(r)}(h). \quad (3.11)$$

And, on the other hand, the definition of martingale increment  $\Delta_n^{(r)}(f)$  can be rearranged as follows

$$X_n^{(r)}(f) = \mathbb{E}(X_n^{(r)}(f) | \mathcal{F}_{n-1}^{(r)}) + \Delta_n^{(r)}(f).$$

Substituting the conditional expectation in the above equality by its value from (3.11) yields

$$X_n^{(r)}(f) = X_{n-1}^{(r)}(T_r f) + \gamma(f) X_{n-1}^{(r)}(h) + \Delta_n^{(r)}(f). \quad (3.12)$$

Next we rewrite equation (3.12) but at rank  $n - 1$  and substituting  $f$  with  $T_r f$ ,

$$X_{n-1}^{(r)}(T_r f) = X_{n-2}^{(r)}(T_r^2 f) + \gamma(T_r f) X_{n-2}^{(r)}(h) + \Delta_{n-1}^{(r)}(T_r f).$$

Going back to equation (3.12) and substituting yet again  $X_{n-1}^{(r)}(T_r f)$  with its expression provided on the right-hand side above, we obtain

$$X_n^{(r)}(f) = X_{n-2}^{(r)}(T_r^2 f) + \gamma(f) X_{n-1}^{(r)}(h) + \gamma(T_r f) X_{n-2}^{(r)}(h) + \Delta_{n-1}^{(r)}(T_r f) + \Delta_n^{(r)}(f),$$

after some re-ordering of the various terms. This succession of substitutions can be iterated until we get

$$X_n^{(r)}(f) = X_0^{(r)}(T_r^n f) + \sum_{i=1}^n \gamma(T_r^{i-1} f) X_{n-i}^{(r)}(h) + R_n^{(r)}(f),$$

where the *remainder* term,  $R_n^{(r)}(f)$ , is given by

$$R_n^{(r)}(f) = \sum_{i=1}^n \Delta_i^{(r)}(T_r^{n-i} f) = \sum_{i=1}^n M_i^{(r)}(T_r^{n-i} f) - M_{i-1}^{(r)}(T_r^{n-i} f).$$

Note that we trivially used the relationship between  $\Delta_i^{(r)}$  and  $M_i^{(r)}$  in the second equality. We then expand the previous sum to help us re-arrange and simplify the term  $R_n^{(r)}(f)$ ,

$$\begin{aligned} R_n^{(r)}(f) &= M_1^{(r)}(T_r^{n-1} f) - M_0^{(r)}(T_r^{n-1} f) + M_2^{(r)}(T_r^{n-2} f) - M_1^{(r)}(T_r^{n-2} f) \\ &\quad + \vdots \\ &\quad + M_{n-1}^{(r)}(T_r f) - M_{n-2}^{(r)}(T_r f) + M_n^{(r)}(T_r^0 f) - M_{n-1}^{(r)}(T_r^0 f), \end{aligned}$$

and eventually rely on the linearity of the operator  $f \mapsto M_i^{(r)}(f)$  and on the identity  $T_r^0 = \text{ID}$  to get

$$R_n^{(r)}(f) = M_n^{(r)}(f) + \sum_{i=1}^{n-1} M_{n-i}^{(r)}(T_r^{i-1}(T_r - \text{ID})f).$$

The result follows by recalling that  $\gamma T_r = 0$ , which implies  $\gamma(T_r^{i-1} f) = 0$  for  $i \geq 2$ .  $\square$

*Remark.* Observe that the decomposition in Lemma 3.2.2, does not depend on Assumption (A2) as the only element upon which we relied was  $\gamma(T_r f) = 0$ , for any  $f \in \mathcal{B}(V^*)$ , which is a consequence of Assumption (A1).

### 3.2.4 $L \log L$ moment propagation

In this section, we introduce another function  $V \leq V^*$  and the associated moment

$$I_n = \sup_{x \in \mathcal{X}} \frac{\mathbb{E}_{\delta_x}(Z_n(V) \log^* Z_n(V))}{V^*(x)}. \quad (3.13)$$

We present a useful technical result to ensure that, under certain conditions, this moment propagates over time and grows as the first eigenvalue.

**Proposition 3.1.** *Let Assumptions (A1) and (A2) hold. Let  $V : \mathcal{X} \rightarrow (0, \infty)$  be a measurable*

function such that  $V \leq V^*$  and consider for  $n \geq 1$ ,

$$I_n = \sup_{x \in \mathcal{X}} \frac{\mathbb{E}_{\delta_x}(Z_n(V) \log^* Z_n(V))}{V^*(x)}. \quad (3.14)$$

(i) If  $S_1 V \in \mathcal{B}(V)$  and  $I_1 < \infty$ , then  $I_n < \infty$  for any  $n \in \mathbb{N}^*$  (propagation).

(ii) If we further assume that there exists  $C \geq 0$  such that for any  $n \in \mathbb{N}$ ,  $S_n V \leq C \lambda^n V$ , then additionally,

$$\lim_{n \rightarrow \infty} \frac{\log I_n}{n} = \log \lambda.$$

This preliminary result not only proves that (the first assertion of) [AH76, Theorem 2] (see Theorem C.4) holds in our setting, but also gives a sharper estimate of  $I_n$ , which seems to be new in this context.

*Proof.* Let us start with statement (i). Consider  $n \in \mathbb{N}^*$  such that  $n = mq + r$ , with  $m, q, r \in \mathbb{N}^*$  and  $0 \leq r \leq m$ . We first show an intermediary result: if  $S_1 V \in \mathcal{B}(V)$ , then there exists a constant  $0 < K_V < \infty$  such that

$$I_n \leq K_V^{n-r} \left( I_r + \lambda^r (\gamma(V^*) I_q + 1) \right).$$

For any  $f \in \mathcal{B}(V)$ , let us first define the function  $x \mapsto i_n(f)(x)$  on  $\mathcal{X}$  by

$$i_n(f)(x) := \mathbb{E}_{\delta_x}(Z_n(f) \log^* Z_n(f)),$$

which we will simply denote  $i_n(x)$  when  $f = V$ . Observe that, in particular, we have by (3.14) that

$$I_n = \sup_{x \in \mathcal{X}} \left( \frac{i_n(x)}{V^*(x)} \right).$$

We wish to upper bound  $i_n(x)$ . A direct application of the branching property leads to

$$Z_n(V) = \sum_{u \in \mathbb{G}_{n-q}} Z_q^{(u)}(V).$$

We now write (directly using  $i_n(V)(x) := i_n(x)$ )

$$\begin{aligned} i_n(x) &= \mathbb{E}_{\delta_x}(Z_n(V) \log^* Z_n(V)), \\ &= \mathbb{E}_{\delta_x} \left( \mathbb{E} \left( \sum_{u \in \mathbb{G}_{n-q}} Z_q^{(u)}(V) \log^* \sum_{u \in \mathbb{G}_{n-q}} Z_q^{(u)}(V) \middle| \mathcal{F}_{n-q} \right) \right), \\ &\leq \mathbb{E}_{\delta_x} (\mathbb{E}(Z_n(V) | \mathcal{F}_{n-q}) \log^* \mathbb{E}(Z_n(V) | \mathcal{F}_{n-q})), \\ &+ \mathbb{E}_{\delta_x} \left( \sum_{u \in \mathbb{G}_{n-q}} \mathbb{E}(Z_q^{(u)}(V) \log^* Z_q^{(u)}(V) | \mathcal{F}_{n-q}) \right), \end{aligned} \quad (3.15)$$

where we simply relied on the Tower property in the second equality and then we applied the conditional version of the extended Jensen convexity result (see Lemma C.4.1) to  $i_n(x)$  since

$Z_n(V)$  is a sum of independent variables. Next, observing that

$$\mathbb{E}(Z_n(V)|\mathcal{F}_{n-q}) = Z_{n-q}(S_q V),$$

we can upper bound the first term on the right-hand-side of inequality (3.15). Indeed we recognise it to be of the form  $i_{n-q}(f)$  for  $f = S_q(V)$ ,

$$\mathbb{E}_{\delta_x}(Z_{n-q}(S_q V) \log^* Z_{n-q}(S_q V)) \leq i_{n-q}(S_q V)(x). \quad (3.16)$$

Now, by Assumption (i), recall that  $S_1 V \in \mathcal{B}(V)$ . Thus there exists a finite constant, say  $K_V > 0$ , such that  $S_q V \leq K_V^q V$  for any  $q \in \mathbb{N}^*$ , which leads to

$$\mathbb{E}_{\delta_x}(Z_{n-q}(S_q V) \log^* Z_{n-q}(S_q V)) \leq K_V^q i_{n-q}(x).$$

Next we re-write the second term in (3.15), using the branching process  $Z^{(u)}$  rooted in  $u$ ,

$$\begin{aligned} \mathbb{E}_{\delta_x} \left( \sum_{u \in \mathbb{G}_{n-q}} \mathbb{E}(Z_q^{(u)}(V) \log^* Z_q^{(u)}(V) | \mathcal{F}_{n-q}) \right) &= \mathbb{E}_{\delta_x} \left( \sum_{u \in \mathbb{G}_{n-q}} \mathbb{E}_{Z^{(u)}}(Z_q(V) \log^* Z_q(V)) \right), \\ &\leq S_{n-q} i_q(x), \end{aligned}$$

where we appealed to the definition of  $S_{n-q}$  and of  $i_q(V) = i_q$ . Observe that assuming  $I_q < \infty$  at this point implies  $i_q \in \mathcal{B}(V^*)$ , which in turns means that we can apply Condition (3.5) to the function  $x \mapsto i_q(x)$  and obtain

$$\begin{aligned} \mathbb{E}_{\delta_x} \left( \sum_{u \in \mathbb{G}_{n-q}} \mathbb{E}(Z_q^{(u)}(V) \log^* Z_q^{(u)}(V) | \mathcal{F}_{n-q}) \right) &\leq \lambda^{n-q} \left( \gamma(i_q) h(x) + V^*(x) a_{n-q} \right), \\ &\leq \lambda^{n-q} V^*(x) \left( \gamma(V^*) I_q + a_{n-q} \right), \end{aligned} \quad (3.17)$$

where we used that  $h \in \mathcal{B}(V^*)$  and  $\gamma(i_q) \leq I_q \gamma(V^*)$ . Combining these upper bounds, inequality (3.15) becomes

$$i_n(x) \leq K_V^q i_{n-q}(x) + \lambda^{n-q} V^*(x) (I_q \gamma(V^*) + a_{n-q}).$$

Iterating the above, we obtain

$$\begin{aligned} i_n(x) &\leq K_V^{mq} i_{n-mq}(x) + V^*(x) \gamma(V^*) I_q \sum_{k=1}^m \lambda^{n-kq} K_V^{q(k-1)} + V^*(x) \sum_{k=1}^m K_V^{q(k-1)} \lambda^{n-kq} a_{n-kq}, \\ &\leq K_V^{mq} i_r(x) + V^*(x) \gamma(V^*) I_q \frac{\lambda^n}{K_V^q} \sum_{k=1}^m \left( \frac{K_V}{\lambda} \right)^{qk} + V^*(x) \frac{\lambda^n}{K_V^q} \sum_{k=1}^m \left( \frac{K_V}{\lambda} \right)^{kq} a_{n-kq}, \\ &\leq K_V^{mq} i_r(x) + C V^*(x) \frac{\lambda^n}{K_V^q} \left( \frac{K_V}{\lambda} \right)^{q(m+1)} (\gamma(V^*) I_q + 1), \\ &\leq K_V^{mq} i_r(x) + C V^*(x) \lambda^{r-q} K_V^{qm} (\gamma(V^*) I_q + 1), \end{aligned}$$

where we used that  $a_n$  is bounded and the value of the constant  $C$  may vary from one line to another but is bounded and independent of  $n$ . Dividing by  $V^*$ , taking the sup over  $x \in \mathcal{X}$  and recalling that  $\lambda > 1$  yields

$$I_n \leq C K_V^{mq} (I_r + \lambda^r (\gamma(V^*) I_q + 1)).$$

It suffices to take  $q, r = 1$  (i.e.  $n = m + 1$ ) to get

$$I_n \leq CK_V^{n-1} (I_1 + \lambda(\gamma(V^*)I_1 + 1)).$$

The fact that  $I_n$  is finite as soon as  $I_1$  is finite is a direct consequence of the above inequality. Let us also note that there is no circularity in the above reasoning. Indeed it was mentioned that inequality (3.17) holds if  $I_q < \infty$  (as we need  $i_q \in \mathcal{B}(V^*)$ ). But if  $I_q = \infty$ , (3.17) would also trivially hold.

Let us now prove statement (ii). Again consider  $n \in \mathbb{N}^*$  such that  $n = mq + r$ , where  $m, q, r \in \mathbb{N}^*$  and  $0 \leq r \leq m$ . Under (ii), there exists  $C \geq 0$  such that for any  $n \in \mathbb{N}$ ,  $S_n V \leq C\lambda^n V$ . We show here first that there exists a finite constant  $K > 0$  such that

$$I_n \leq K\lambda^n C^m (I_r + I_q + 1). \quad (3.18)$$

Observe that the exact same arguments of the proof of statement (i) can be applied and hold up until equation (3.16), which we rewrite here for clarity purposes.

$$\mathbb{E}_{\delta_x}(Z_{n-q}(S_q V) \log^* Z_{n-q}(S_q V)) \leq i_{n-q}(S_q V)(x).$$

Additionally (3.15) also still holds. Now, by (ii), there exists  $C \geq 0$  such that

$$S_q V(x) \leq C\lambda^q V(x).$$

This assumption leads to

$$\mathbb{E}_{\delta_x}(Z_{n-q}(S_q V) \log^* Z_{n-q}(S_q V)) \leq C\lambda^q i_{n-q}(x).$$

We pursue with the second term in (3.15), where the same upper bound (3.17) still holds under the conditions of (ii), i.e.,

$$\mathbb{E}_{\delta_x} \left( \sum_{u \in \mathbb{G}_{n-q}} \mathbb{E}(Z_q^{(u)}(V) \log^* Z_q^{(u)}(V) | \mathcal{F}_{n-q}) \right) \leq \lambda^{n-q} V^*(x) (\gamma(V^*)I_q + a_{n-q}).$$

Combining these two upper bounds, we obtain

$$i_n(x) \leq C\lambda^q i_{n-q}(x) + \lambda^{n-q} V^*(x) (\gamma(V^*)I_q + a_{n-q}),$$

which becomes by iteration, relying on the boundedness of  $(a_n)_n$  and on  $\lambda > 1$

$$\begin{aligned} i_n(x) &\leq \lambda^{mq} C^m i_{n-mq}(x) + V^*(x) \lambda^{n-q} \frac{C^m - 1}{C - 1} (\gamma(V^*)I_q + a_{n-q}), \\ &\leq C^m \lambda^{n-r} i_r(x) + C^m \lambda^{n-q} V^*(x) (I_q \gamma(V^*) + 1), \\ &\leq C^m \lambda^n (i_r(x) + V^*(x) (I_q \gamma(V^*) + 1)). \end{aligned}$$

Dividing by  $V^*$  and taking the sup on each side, we obtain

$$I_n \leq \lambda^n C^m (I_r + \gamma(V^*)I_q + 1).$$

We can simplify this last inequality further by introducing a constant  $K > 0$  such that

$$I_n \leq K\lambda^n C^m (I_r + I_q + 1).$$

In the case where  $q = r = 1$ , we obtain (for another  $K > 0$  bounded)

$$I_n \leq K \lambda^n C^{n-1} (I_1 + 1). \quad (3.19)$$

The fact that  $I_n$  is finite as soon as  $I_1$  is finite is again a direct consequence of the above inequality. The first part of statement (ii) is proved and we turn to the asymptotic analysis of  $\log I_n/n$ .

A lower bound for  $\log I_n/n$  can be derived as a consequence of the semigroup Condition (3.5) applied to the function  $V$ . Indeed, the latter condition yields

$$\frac{\mathbb{E}_{\delta_x}(Z_n(V))}{V^*(x)} = \frac{S_n V(x)}{V^*(x)} \geq \lambda^n \left( \gamma(V) \frac{h(x)}{V^*(x)} - a_n \right).$$

Then, considering the event where  $Z_n(V) > e$  and its complementary set, we observe on one hand that

$$\begin{aligned} \mathbb{E}_{\delta_x}(Z_n(V) \log^* Z_n(V)) &\geq \mathbb{E}_{\delta_x}(Z_n(V) \log^* Z_n(V) \mathbf{1}_{Z_n(V) > e}), \\ &\geq \mathbb{E}_{\delta_x}(Z_n(V) \mathbf{1}_{Z_n(V) > e}), \end{aligned}$$

and on the other hand, that

$$\begin{aligned} \mathbb{E}_{\delta_x}(Z_n(V)) &= \mathbb{E}_{\delta_x}(Z_n(V) \mathbf{1}_{Z_n(V) > e}) + \mathbb{E}_{\delta_x}(Z_n(V) \mathbf{1}_{Z_n(V) \leq e}), \\ &\leq \mathbb{E}_{\delta_x}(Z_n(V) \mathbf{1}_{Z_n(V) > e}) + e. \end{aligned}$$

Thus,

$$\frac{\mathbb{E}_{\delta_x}(Z_n(V) \log^* Z_n(V))}{V^*(x)} \geq \frac{\mathbb{E}_{\delta_x}(Z_n(V))}{V^*(x)} - \frac{e}{V^*(x)},$$

and

$$I_n \geq \sup_{x \in \mathcal{X}} \left\{ \lambda^n \left( \gamma(V) \frac{h(x)}{V^*(x)} - a_n \right) - \frac{e}{V^*(x)} \right\}.$$

Fixing now  $x$  on the right hand side and recalling that  $a_n \rightarrow 0$  yields the lower bound

$$\liminf_{n \rightarrow \infty} \frac{\log I_n}{n} \geq \lambda.$$

To upper bound  $\frac{\log I_n}{n}$ , we proceed in two steps. Let us take log on each side of inequality (3.19) and get an a priori upper bound first,

$$\limsup_{n \rightarrow \infty} \frac{\log I_n}{n} \leq \log \lambda + \log C := C'. \quad (3.20)$$

In other words, we can write

$$\log I_n \lesssim C' n.$$

In a second step, we use again inequality (3.18) and choose specifically  $m := \lfloor \sqrt{n} \rfloor$  such that now both integers  $r, q$  are such that  $r, q \leq \lfloor \sqrt{n} \rfloor$  and  $n = \lfloor \sqrt{n} \rfloor q + r$ . Inequality (3.18) becomes

$$I_n \leq K \lambda^n C^{\lfloor \sqrt{n} \rfloor} (I_r + I_q + 1).$$

We now combine the last bound with the a priori upper bound (3.20), appealing to the facts that  $C, C' < \infty$ , and  $\gamma(V^*) < \infty$ . We get

$$\limsup_{n \rightarrow \infty} \frac{\log I_n}{n} \leq \log \lambda + \limsup_{n \rightarrow \infty} \frac{\sqrt{n} \log C}{n} + 2C' \limsup_{n \rightarrow \infty} \frac{\sqrt{n}}{n} = \log \lambda,$$

which ends the proof of statement (ii).  $\square$

Note here that Theorem 3.1 assumes that  $I_k < \infty$  for any  $k \in \mathbb{N}$ . What Proposition 3.1 tells us is that if we were to assume that  $S_1 V \in \mathcal{B}(V)$  as in Statement (i), then it would suffice to assume  $I_1 < \infty$  in Theorem 3.1. Equally, if we were to assume that  $S_n V \leq C \lambda^n V$  as in Statement (ii), then not only would it suffice to assume  $I_1 < \infty$  in Theorem 3.1 but we would also have a more precise estimate of the asymptotic moment  $\log I_n/n$ .

### 3.2.5 Martingale decomposition

We want to determine and fine-tune the long-term behaviour of the martingales  $M^{(r)}(f)$  defined in (3.10). This is achieved by exploiting a martingale decomposition, respectively, into  $L^2$  and  $L^1$  contributions. Classically, the  $L^1$  part gathers large jumps. The decomposition here is inspired by the subtle truncation argument put forward by [AH76]. More precisely, the martingale increment writes

$$\begin{aligned} \Delta_{n+1}^{(r)}(f) &= X_{n+1}^{(r)}(f) - \mathbb{E}(X_{n+1}^{(r)}(f) | \mathcal{F}_n^{(r)}), \\ &= \lambda^{-nr} \sum_{u \in \mathbb{G}_{nr}} \left( \lambda^{-r} Z_r^{(u)}(f) - \mathbb{E}_{Z(u)}(X^{(r)}(f)) \right) = A_{n+1}^{(r)}(f) + B_{n+1}^{(r)}(f), \end{aligned} \quad (3.21)$$

where  $Z^{(u)}$ , defined in (3.4), is the branching process rooted in  $u$  and the two contributions  $A$  and  $B$  are defined for any  $n \in \mathbb{N}$  by

$$A_{n+1}^{(r)}(f) = \lambda^{-nr} \sum_{u \in \mathbb{G}_{nr}} \left( \lambda^{-r} Z_r^{(u)}(f) \mathbf{1}_{Z_r^{(u)}(V) \leq \lambda^{nr}} - \mathbb{E}_{Z(u)}(X^{(r)}(f) \mathbf{1}_{Z_r(V) \leq \lambda^{nr}}) \right), \quad (3.22)$$

$$B_{n+1}^{(r)}(f) = \lambda^{-nr} \sum_{u \in \mathbb{G}_{nr}} \left( \lambda^{-r} Z_r^{(u)}(f) \mathbf{1}_{Z_r^{(u)}(V) > \lambda^{nr}} - \mathbb{E}_{Z(u)}(X^{(r)}(f) \mathbf{1}_{Z_r(V) > \lambda^{nr}}) \right). \quad (3.23)$$

Observe that in this section, and in particular to prove Proposition 3.2, we need to consider functions  $f$  that belong to the functional space  $\mathcal{B}(V)$ , not simply  $\mathcal{B}(V^*)$ , and where  $V \leq V^*$  is the Lyapunov function introduced in Theorem 3.1.

**Proposition 3.2.** *Let Assumptions of Theorem 3.1 hold. Then for any  $f \in \mathcal{B}(V)$ ,  $r \geq 1$  and  $n \in \mathbb{N}$ , we have the decomposition*

$$M_n^{(r)}(f) = \sum_{i=1}^n A_i^{(r)}(f) + \sum_{i=1}^n B_i^{(r)}(f), \quad \text{where}$$

- $\left( \sum_{i=1}^n A_i^{(r)}(f) \right)_{n \in \mathbb{N}}$  is a martingale, bounded in  $L^2$ ,
- $\left( \sum_{i=1}^n B_i^{(r)}(f) \right)_{n \in \mathbb{N}}$  is a uniformly integrable martingale.

As a consequence,

$$(M_n^{(r)}(f))_{n \in \mathbb{N}} \xrightarrow{n \rightarrow \infty} M_\infty^{(r)}(f), \quad \text{a.s. and in } L^1,$$

and in particular,

$$(X_n(h))_{n \in \mathbb{N}} \xrightarrow{n \rightarrow \infty} W^{(x)}, \quad \text{a.s. and in } L^1,$$

where the random variable  $W^{(x)} \geq 0$  depends on the initial condition  $Z_0 = x$  and is defined so that  $\mathbb{E}_{\delta_x}(W^{(x)}) = h(x)$ .

*Proof.* We first focus on the  $L^1$  part of the martingale decomposition which involves  $\sum_{i=1}^n B_i^{(r)}(f)$ . To that effect, let us introduce for any  $i, r \in \mathbb{N}^*$ ,

$$B_{i+1}^{(r)+} := \lambda^{-ir} \sum_{u \in \mathbb{G}_{ir}} \left( \lambda^{-r} Z_r^{(u)}(V) \mathbf{1}_{Z_r^{(u)}(V) > \lambda^{ir}} + \mathbb{E}_{Z(u)}(X_r(V) \mathbf{1}_{Z_r(V) > \lambda^{ir}}) \right),$$

and observe that for any  $f \in \mathcal{B}(V)$ , we have

$$|B_i^{(r)}(f)| \leq \|f\|_{\mathcal{B}(V)} B_i^{(r)+}. \quad (3.24)$$

Moreover, if we define for any  $i, r \in \mathbb{N}^*$  the function  $y \mapsto \phi_i^{(r)}(y)$  on  $\mathcal{X}$  by

$$\phi_i^{(r)}(y) := \lambda^{-r} \mathbb{E}_{\delta_y}(Z_r(V) \mathbf{1}_{Z_r(V) > \lambda^{ir}}),$$

then we can write

$$\phi_i^{(r)}(Z(u)) = \lambda^{-r} \mathbb{E}_{Z(u)}(Z_r(V) \mathbf{1}_{Z_r(V) > \lambda^{ir}}) = \lambda^{-r} \mathbb{E}_{\delta_x}(Z_r^{(u)}(V) \mathbf{1}_{Z_r^{(u)}(V) > \lambda^{ir}}),$$

where we used that  $Z^{(u)}$  is the branching process rooted in  $u$ . Thus, taking the expectation of  $B_{i+1}^{(r)+}$  leads to

$$\mathbb{E}_{\delta_x}(B_{i+1}^{(r)+}) = 2\lambda^{-ir} \mathbb{E}_{\delta_x} \left( \sum_{u \in \mathbb{G}_{ir}} \phi_i^{(r)}(Z(u)) \right) = 2\lambda^{-ir} S_{ir} \phi_i^{(r)}(x),$$

where we simply appealed to the definition of the first-moment semigroup in the second equality. We now rely on Assumption (A2) (and more precisely on (3.7)) to get the upper bound

$$\mathbb{E}_{\delta_x}(B_{i+1}^{(r)+}) \leq 2 \left( h(x) \gamma(\phi_i^{(r)}) + V^*(x) a_{ir} \|\phi_i^{(r)}\|_{\mathcal{B}(V^*)} \right).$$

Our objective is to upper bound the sum  $\sum_{i=0}^{\infty} \mathbb{E}_{\delta_x}(B_{i+1}^{(r)+})$ , so we compute, recalling in the first inequality below that  $h$  is dominated by  $V$  and thus by  $V^*$ ,

$$\begin{aligned} & \sum_{i=0}^{\infty} \left( h(x) \gamma(\phi_i^{(r)}) + V^*(x) a_{ir} \|\phi_i^{(r)}\|_{\mathcal{B}(V^*)} \right) \\ & \leq \lambda^{-r} \|h\|_{\mathcal{B}(V^*)} V^*(x) \mathbb{E}_\gamma \left( \sum_{i=0}^{\infty} Z_r(V) \mathbf{1}_{Z_r(V) > \lambda^{ir}} \right) + \lambda^{-r} V^*(x) \sum_{i=0}^{\infty} a_{ir} \sup_{y \in \mathcal{X}} \frac{\mathbb{E}_{\delta_y}(Z_r(V) \mathbf{1}_{Z_r(V) > \lambda^{ir}})}{V^*(y)}, \end{aligned}$$

where we relied on Tonelli's theorem to exchange expectations and infinite sums. We now split the sum on the right-hand side of the above inequality into two terms which we will study separately. More precisely, based on the above inequality, we can introduce a constant  $C > 0$  such that

$$\begin{aligned} & \sum_{i=0}^{\infty} \left( h(x) \gamma(\phi_i^{(r)}) + V^*(x) a_{ir} \|\phi_i^{(r)}\|_{\mathcal{B}(V^*)} \right) \\ & \leq C \lambda^{-r} V^*(x) \left( \underbrace{\mathbb{E}_{\gamma} \left( \sum_{i=0}^{\infty} Z_r(V) \mathbf{1}_{Z_r(V) > \lambda^{ir}} \right)}_{\text{Term 1}} + \underbrace{\sum_{i=0}^{\infty} a_{ir} \sup_{y \in \mathcal{X}} \frac{\mathbb{E}_{\delta_y}(Z_r(V) \mathbf{1}_{Z_r(V) > \lambda^{ir}})}{V^*(y)}}_{\text{Term 2}} \right). \end{aligned} \quad (3.25)$$

Note that in the rest of this proof and subsequent proofs, we will denote  $C$  a constant independent of  $n$  or  $i$ , but whose value can change from line to line.

We want to control both expectation terms on the right hand side of the inequality (3.25). First, a Markov-type inequality allows us to upper bound Term 2. Indeed, observe that

$$\sup_{y \in \mathcal{X}} \frac{\mathbb{E}_{\delta_y}(Z_r(V) \mathbf{1}_{Z_r(V) > \lambda^{ir}})}{V^*(y)} \leq \sup_{y \in \mathcal{X}} \frac{1}{V^*(y)} \mathbb{E}_{\delta_y} \left( Z_r(V) \frac{\log^* Z_r(V)}{\log^* \lambda^{ir}} \right).$$

But since  $\log^* \lambda^u \geq Cu \log \lambda$ , (see Appendix C.5.1), we obtain

$$\sup_{y \in \mathcal{X}} \frac{\mathbb{E}_{\delta_y}(Z_r(V) \mathbf{1}_{Z_r(V) > \lambda^{ir}})}{V^*(y)} \leq \frac{C}{ir \log \lambda} I_r.$$

Thus, overall, Term 2 in (3.25) is bounded as follows

$$\text{Term 2} := \sum_{i=0}^{\infty} a_{ir} \sup_{y \in \mathcal{X}} \frac{\mathbb{E}_{\delta_y}(Z_r(V) \mathbf{1}_{Z_r(V) > \lambda^{ir}})}{V^*(y)} \leq \frac{C}{\log \lambda} I_r \sum_{i=0}^{\infty} \frac{a_{ir}}{ir}. \quad (3.26)$$

Next, to upper bound Term 1 in (3.25), we define

$$N_r = \sup \{ i \in \mathbb{N} : Z_r(V) > \lambda^{ir} \} = \sup \left\{ i \in \mathbb{N} : \frac{\log^*(Z_r(V))}{\log^* \lambda^{ir}} > 1 \right\},$$

which then allows us to write

$$\begin{aligned} \sum_{i=0}^{\infty} Z_r(V) \mathbf{1}_{Z_r(V) > \lambda^{ir}} & \leq \sum_{i=0}^{\infty} Z_r(V) \frac{\log^*(Z_r(V))}{\log^*(\lambda^{N_r r})} \mathbf{1}_{Z_r(V) > \lambda^{ir}}, \\ & \leq N_r \frac{Z_r(V) \log^*(Z_r(V))}{\log^*(\lambda^{N_r r})}, \\ & \leq \frac{C}{r \log \lambda} Z_r(V) \log^*(Z_r(V)), \end{aligned}$$

where we used again in the last line that  $\log^*(\lambda^{N_r r}) \geq CN_r r \log \lambda$ . (See Appendix C.5.1).

Observing now that

$$\mathbb{E}_{\gamma}(Z_r(V) \log^* Z_r(V)) \leq \gamma(V^*) I_r,$$

it then suffices to take the expectation on each side with respect to the initial distribution  $\gamma$  to obtain the following upper bound for Term 1,

$$\begin{aligned} \text{Term 1} &:= \mathbb{E}_\gamma \left( \sum_{i=0}^{\infty} Z_r(V) \mathbf{1}_{Z_r(V) > \lambda^{ir}} \right) \leq \frac{C}{r \log \lambda} \mathbb{E}_\gamma (Z_r(V) \log^*(Z_r(V))), \\ &\leq \frac{C}{r \log \lambda} \gamma(V^*) I_r. \end{aligned} \quad (3.27)$$

Finally, gathering estimates (3.26) and (3.27), we obtain

$$\sum_{i=0}^{\infty} \mathbb{E}_{\delta_x} (B_{i+1}^{(r)+}) \leq C V^*(x) \frac{\lambda^{-r}}{r} I_r \left( \gamma(V^*) + \sum_{i=0}^{\infty} \frac{a_{ir}}{i} \right).$$

Now by the assumptions of Theorem 3.1, we have that  $\gamma(V^*) < \infty$ ,  $\sum_{i=0}^{\infty} \frac{a_{ir}}{ir} < \infty$  and  $I_r < \infty$ . Thus,

$$\sum_{i=1}^{\infty} \mathbb{E}_{\delta_x} (B_i^{(r)+}) < \infty.$$

But, the inequality

$$\sum_{i=1}^n |B_i^{(r)}(f)| \leq \|f\|_{\mathcal{B}(V)} \sum_{i=1}^n B_i^{(r)+}$$

ensures that  $\left( \sum_{n=1}^n B_n^{(r)}(f) \right)_{n \in \mathbb{N}}$  is uniformly integrable.

Let us now turn towards the  $L^2$  part of the decomposition of  $M_n^{(r)}(f)$ . We use below the semi-martingale decomposition of the square of  $\left( \sum_{i=1}^n A_i^{(r)}(f) \right)_{n \in \mathbb{N}}$ . More precisely, we rely on the quadratic variation and recall that

$$\left( \sum_{i=1}^n A_i^{(r)}(f) \right)^2 - \sum_{i=1}^n \mathbb{E} \left( A_i^{(r)}(f)^2 \mid \mathcal{F}_{i-1}^{(r)} \right)$$

is a martingale starting from 0. We wish to compute

$$\begin{aligned} &\mathbb{E} \left( A_{i+1}^{(r)}(f)^2 \mid \mathcal{F}_i^{(r)} \right) \\ &= \mathbb{E} \left( \lambda^{-2ir} \left( \sum_{u \in \mathbb{G}_{ir}} \left( \lambda^{-r} Z_r^{(u)}(f) \mathbf{1}_{Z_r^{(u)}(V) \leq \lambda^{ir}} - \mathbb{E}_{Z(u)}(X^{(r)}(f) \mathbf{1}_{Z_r(V) \leq \lambda^{ir}}) \right) \right)^2 \mid \mathcal{F}_i^{(r)} \right). \end{aligned}$$

We introduce the function defined on  $\mathcal{X}$  for integers  $r \geq 1$ ,  $i \geq 0$  and for  $f \in \mathcal{B}(V)$  by

$$\psi_i^{(r,f)}(y) := \mathbb{E}_{\delta_y} \left( (X_r(f) \mathbf{1}_{Z_r(V) \leq \lambda^{ir}} - \mathbb{E}_{\delta_y}(X_r(f) \mathbf{1}_{Z_r(V) \leq \lambda^{ir}}))^2 \right).$$

Observing that the random variables composing  $A_{i+1}^{(r)}(f)$  are centred and independent conditionally on  $\mathcal{F}_i^{(r)}$ , we can first expand the sum over  $\mathbb{G}_{ir}$  and then rely on the definition of  $\psi_i^{(r,f)}$ .

We get

$$\begin{aligned}
& \mathbb{E} \left( A_{i+1}^{(r)}(f)^2 | \mathcal{F}_i^{(r)} \right) \\
&= \mathbb{E} \left( \lambda^{-2ir} \sum_{u \in \mathbb{G}_{ir}} \left( \lambda^{-r} Z_r^{(u)}(f) \mathbf{1}_{Z_r^{(u)}(V) \leq \lambda^{ir}} - \mathbb{E}_{Z(u)}(X^{(r)}(f) \mathbf{1}_{Z_r(V) \leq \lambda^{ir}}) \right)^2 \middle| \mathcal{F}_i^{(r)} \right), \\
&= \lambda^{-ir} X_i^{(r)}(\psi_i^{(r,f)}). \tag{3.28}
\end{aligned}$$

Moreover, introducing now the function  $y \mapsto V_i^{(r)}(y)$  given for  $i \geq 0$ ,  $r \geq 1$ , by

$$V_i^{(r)}(y) = 2\mathbb{E}_{\delta_y} (X_r(V)^2 \mathbf{1}_{Z_r(V) \leq \lambda^{ir}}),$$

we can write for any  $f \in \mathcal{B}(V)$ ,

$$\psi_i^{(r,f)}(y) \leq 2\mathbb{E}_{\delta_y} (X_r(f)^2 \mathbf{1}_{Z_r(V) \leq \lambda^{ir}}) \leq \|f\|_{\mathcal{B}(V)}^2 V_i^{(r)}(y).$$

Returning to (3.28), we get

$$\sum_{i=1}^n \mathbb{E} \left( A_i^{(r)}(f)^2 | \mathcal{F}_{i-1}^{(r)} \right) \leq \|f\|_{\mathcal{B}(V)}^2 \sum_{i=0}^{\infty} \lambda^{-ir} X_i^{(r)}(V_i^{(r)}). \tag{3.29}$$

To conclude, let us prove that the martingale  $\left( \sum_{i=1}^n A_i^{(r)}(f) \right)_{n \in \mathbb{N}}$  is bounded in  $L^2$  by computing the expectation on the right-hand side of (3.29). Indeed,

$$\begin{aligned}
\mathbb{E}_{\delta_x} \left( \sum_{i=0}^{\infty} \lambda^{-ir} X_i^{(r)}(V_i^{(r)}) \right) &= \sum_{i=0}^{\infty} \lambda^{-ir} \lambda^{-ir} S_{ir} V_i^{(r)}(x), \\
&\leq \underbrace{h(x) \sum_{i=0}^{\infty} \lambda^{-ir} \gamma(V_i^{(r)})}_{\text{Term 3}} + \underbrace{V^*(x) \sum_{i=0}^{\infty} a_{ir} \lambda^{-ir} \|V_i^{(r)}\|_{\mathcal{B}(V^*)}}_{\text{Term 4}}, \tag{3.30}
\end{aligned}$$

where, once again, we used Assumption (A2) (more specifically (3.5)) to upper bound the first-moment semigroup. We follow similar arguments as for the  $L^1$  component to upper bound Terms 3 and 4. We will see below that whilst trying to get an upper bound for Term 4, we will obtain an upper bound for Term 3. Let us first consider  $\mathbb{E}_{\delta_x}(Z_r(V)^2 \mathbf{1}_{Z_r(V) \leq \lambda^{ir}})$  by splitting the interval of the indicator function as follows

$$\{Z_r(V) \leq \lambda^{ir}\} = \{Z_r(V) < e \leq \lambda^{ir}\} \cup \{e \leq Z_r(V) \leq \lambda^{ir}\}. \tag{3.31}$$

We define for any  $r \geq 1$

$$N'_r := \inf\{i \in \mathbb{N} : Z_r(V) \leq \lambda^{ir}\}.$$

Observe now that we can write

$$\begin{aligned}
\sum_{i=0}^{\infty} \lambda^{-ir} \mathbf{1}_{Z_r(V) \leq \lambda^{ir}} &= \sum_{i=N'_r}^{\infty} \lambda^{-ir} \mathbf{1}_{Z_r(V) \leq \lambda^{ir}}, \\
&\leq \lambda^{-rN'_r} \frac{1}{1 - \lambda^{-r}} = \frac{C}{2Z_r(V)} = \frac{C}{2\lambda^r X_r(V)}, \tag{3.32}
\end{aligned}$$

where we recall that the finite constant  $C$  does not depend on  $r$ . Then for any  $y \in \mathcal{X}$ , appealing to Tonelli's theorem and the definition of  $V_i^{(r)}$ ,

$$\begin{aligned} \sum_{i=0}^{\infty} \lambda^{-ir} V_i^{(r)}(y) &= 2\mathbb{E}_{\delta_y} \left( X_r(V)^2 \sum_{i=0}^{\infty} \lambda^{-ir} \mathbf{1}_{Z_r(V) \leq \lambda^{ir}} \right), \\ &\leq C\lambda^{-r} \mathbb{E}_{\delta_y}(X_r(V)) = C\lambda^{-2r} S_r V(y), \end{aligned}$$

where we also relied on (3.32) to obtain the first inequality and on the definition of the first-moment semigroup in the last equality. This yields, in turn, taking expectation with respect to the measure  $\gamma$  on each side,

$$\sum_{i=0}^{\infty} \lambda^{-ir} \gamma(V_i^{(r)}) \leq C\lambda^{-2r} \gamma(S_r V).$$

Recall that, since  $V \leq V^*$ , Assumption (A2) (i.e. (3.5)) implies

$$\begin{aligned} S_r V(x) &\leq \lambda^r (h(x) \gamma(V) + V^*(x) a_r), \\ &\leq \lambda^r V^*(x) (\gamma(V^*) + a_r), \\ &\leq C\lambda^r V^*(x), \end{aligned}$$

where we used that  $h \leq V^*$ ,  $\gamma(V^*) < \infty$  and  $a_r$  bounded. This finally gives us an upper bound for Term 3,

$$\text{Term 3} = h(x) \sum_{i=0}^{\infty} \lambda^{-ir} \gamma(V_i^{(r)}) \leq C\lambda^{-r} \gamma(V^*) V^*(x).$$

Observe next, since the function  $u \mapsto u/\log(u)$  is increasing for  $u > e$ ,

$$a^2 \mathbf{1}_{a \leq b} = a \log(a) \frac{a}{\log(a)} \mathbf{1}_{a/\log(a) \leq b/\log(b)} \leq a \log(a) \frac{b}{\log(b)}, \quad \text{for } a, b > e.$$

Thus,

$$\mathbb{E}_{\delta_y}(Z_r(V)^2 \mathbf{1}_{e \leq Z_r(V) \leq \lambda^{ir}}) \leq \frac{\lambda^{ir}}{ir \log \lambda} \mathbb{E}_{\delta_y}(Z_r(V) \log(Z_r(V))).$$

But using the fact that  $\log^* u = \log u$  when  $u > e$ , we can equivalently write

$$\mathbb{E}_{\delta_y}(Z_r(V)^2 \mathbf{1}_{e \leq Z_r(V) \leq \lambda^{ir}}) \leq \frac{\lambda^{ir}}{ir \log \lambda} \mathbb{E}_{\delta_y}(Z_r(V) \log^*(Z_r(V))). \quad (3.33)$$

Next note that for  $u < e$ , we have by definition of  $u \mapsto \log^* u$  that  $u^2 = eu \log^*(u)$ . So we can write in this instance

$$E_{\delta_y} \left( Z_r(V)^2 \mathbf{1}_{Z_r(V) < e \leq \lambda^{ir}} \right) \leq E_{\delta_y} \left( e Z_r(V) \log^* Z_r(V) \mathbf{1}_{Z_r(V) < e \leq \lambda^{ir}} \right). \quad (3.34)$$

Additionally, by Lemma C.5.2 in appendix (that in effect simply serves to recall that the logarithm grows more slowly than a polynomial), we know that the inequality

$$\frac{C\lambda^{ir}}{ir \log \lambda} \geq 1,$$

holds for any  $C \geq \frac{1}{e}$ . We can then write

$$E_{\delta_y}(eZ_r(V) \log^* Z_r(V) \mathbf{1}_{Z_r(V) < e \leq \lambda^{ir}}) \leq eC \frac{\lambda^{ir}}{ir \log \lambda} E_{\delta_y}(Z_r(V) \log^* Z_r(V)). \quad (3.35)$$

Combining (3.33) and (3.34)-(3.35), we obtain

$$\begin{aligned} E_{\delta_y}(Z_r(V)^2 \mathbf{1}_{Z_r(V) \leq \lambda^{ir}}) &= E_{\delta_y}(Z_r(V)^2 \mathbf{1}_{Z_r(V) < e \leq \lambda^{ir}}) + E_{\delta_y}(Z_r(V)^2 \mathbf{1}_{e < Z_r(V) \leq \lambda^{ir}}), \\ &\leq C \frac{\lambda^{ir}}{ir \log \lambda} \mathbb{E}_{\delta_y}(Z_r(V) \log^*(Z_r(V))). \end{aligned}$$

Now observe that

$$\|V_i^{(r)}\|_{\mathcal{B}(V^*)} = \frac{2}{\lambda^{2r}} \sup_{y \in \mathcal{X}} \frac{\mathbb{E}_{\delta_y}(Z_r(V)^2 \mathbf{1}_{Z_r(V) \leq \lambda^{ir}})}{V^*(y)},$$

which yields

$$\sum_{i=0}^{\infty} \lambda^{-ir} a_{ir} \|V_i^{(r)}\|_{\mathcal{B}(V^*)} = \frac{2}{\lambda^{2r}} \sum_{i=0}^{\infty} \lambda^{-ir} a_{ir} \sup_{y \in \mathcal{X}} \frac{\mathbb{E}_{\delta_y}(Z_r(V)^2 \mathbf{1}_{Z_r(V) \leq \lambda^{ir}})}{V^*(y)}.$$

Again we decompose the interval  $\{Z_r(V) \leq \lambda^{ir}\}$  as per (3.31). Relying first on inequality (3.33), we get

$$\begin{aligned} \frac{2}{\lambda^{2r}} \sum_{i=0}^{\infty} \lambda^{-ir} a_{ir} \sup_{y \in \mathcal{X}} \frac{\mathbb{E}_{\delta_y}(Z_r(V)^2 \mathbf{1}_{e \leq Z_r(V) \leq \lambda^{ir}})}{V^*(y)} &\leq \frac{2\lambda^{-2r}}{\log \lambda} \sum_{i=0}^{\infty} \frac{a_{ir}}{ir} \sup_{y \in \mathcal{X}} \frac{\mathbb{E}_{\delta_y}(Z_r(V) \log^* Z_r(V))}{V^*(y)}, \\ &\leq \frac{2\lambda^{-2r}}{\log \lambda} I_r \sum_{i=0}^{\infty} \frac{a_{ir}}{ir}. \end{aligned}$$

Next relying on inequalities (3.34)-(3.35), we get

$$\begin{aligned} \sum_{i=0}^{\infty} \lambda^{-ir} a_{ir} \sup_{y \in \mathcal{X}} \frac{\mathbb{E}_{\delta_y}(Z_r(V)^2 \mathbf{1}_{Z_r(V) < e \leq \lambda^{ir}})}{V^*(y)} &\leq \frac{eC}{\log \lambda} \sum_{i=0}^{\infty} \lambda^{-ir} \left(\frac{\lambda^{ir}}{ir}\right) a_{ir} \sup_{y \in \mathcal{X}} \frac{E_{\delta_y}(Z_r(V) \log^* Z_r(V))}{V^*(y)}, \\ &\leq \frac{eC}{\log \lambda} I_r \sum_{i=0}^{\infty} \frac{a_{ir}}{ir}. \end{aligned}$$

Combining these together, we obtain

$$\sum_{i=0}^{\infty} \lambda^{-ir} a_{ir} \|V_i^{(r)}\|_{\mathcal{B}(V^*)} \leq C \frac{\lambda^{-2r}}{\log \lambda} I_r \sum_{i=0}^{\infty} \frac{a_{ir}}{ir}.$$

Thus,

$$\text{Term 4} = V^*(x) \sum_{i=0}^{\infty} \lambda^{-ir} a_{ir} \|V_i^{(r)}\|_{\mathcal{B}(V^*)} < CV^*(x) \lambda^{-2r} I_r \sum_{i=0}^{\infty} \frac{a_{ir}}{ir}.$$

Finally, applying once again the Tower Property in the first line below and recalling inequality

(3.29) in the second line, we get

$$\begin{aligned} \mathbb{E}_{\delta_x} \left( \sum_{i=1}^n A_i^{(r)}(f) \right)^2 &= \mathbb{E}_{\delta_x} \left( \sum_{i=1}^n \mathbb{E} \left( A_i^{(r)}(f)^2 \mid \mathcal{F}_{i-1}^{(r)} \right) \right), \\ &\leq \|f\|_{\mathcal{B}(V)}^2 \mathbb{E}_{\delta_x} \left( \sum_{i=0}^{\infty} \lambda^{-ir} X_i^{(r)}(V_i^{(r)}) \right), \\ &\leq C \|f\|_{\mathcal{B}(V)}^2 \lambda^{-r} V^*(x) \left( \gamma(V^*) + \lambda^{-r} I_r \sum_{i=0}^{\infty} \frac{a_{ri}}{ri} \right), \end{aligned}$$

where we used in the last line the upper bounds obtained for Term 3 and Term 4.

Now recall that by the assumptions of Theorem 3.1, we have that  $I_n < \infty$  for any  $n \in \mathbb{N}^*$ ,  $\sum_{i=1}^{\infty} \frac{a_{ir}}{ir} < \infty$  and  $\gamma(V^*) < \infty$ . So

$$\mathbb{E}_{\delta_x} \left( \sum_{i=1}^n A_i^{(r)}(f) \right)^2 < \infty.$$

We have therefore obtained the  $L^2$  boundedness of  $\sum_{i=1}^n A_i^{(r)}(f)$ . Note that the above also provides a more quantitative estimate of the bound, which is useful for instance when we let  $r$  go to infinity to prove that this  $L^2$  component disappears. The first part of the proposition has thus been proven.

We now obtain directly the convergence of the martingale  $(M_n^{(r)}(f))_{n \in \mathbb{N}}$ , as it is the sum of a uniformly integrable martingale and a martingale bounded in  $L^2$ . Indeed, referring to [Wil91, Ch. 14 p.134], recall that a uniformly integrable martingale converges a.s. and in  $L^1$ . Besides, a martingale bounded in  $L^2$  converges a.s. and in  $L^2$  (thus in  $L^1$  as well).

Finally recalling that  $X_n(h) = h(x) + M_n^{(1)}(h)$  proves the last part of Proposition 3.2.  $\square$

To prove Proposition 3.3 in the next section which itself leads to our main result in discrete time, i.e. Theorem 3.1, we need to control the series of martingales introduced in Lemma 3.2.2. This control is achieved by exploiting the contraction  $T_r$ .

**Lemma 3.2.3.** *Under Assumptions of Theorem 3.1, for any  $f \in \mathcal{B}(V)$  and any  $r$  such that  $a_r < 1$ , we have*

$$\sup_{n > N} \left| \sum_{i=N}^{n-1} M_{n-i}^{(r)}(T_r^{i-1}(T_r - Id)f) \right| \xrightarrow{N \rightarrow \infty} 0, \quad \text{a.s. and in } L^1.$$

*Proof.* In order to lighten notations in this proof, we introduce a second contraction operator. For any  $n \in \mathbb{N}$ ,  $i \leq n$  and any  $N \in \mathbb{N}$ , we define

$$Q_r^{n,i,N} := T_r^{n-i} - T_r^{N-1},$$

such that for any  $f \in \mathcal{B}(V)$ , we can write

$$\begin{aligned} \|Q_r^{n,i,N} f\|_{\mathcal{B}(V)} &\leq 2\|T_r^{n-i} f\|_{\mathcal{B}(V)}, \\ &\leq 2a_r^{n-i} \|f\|_{\mathcal{B}(V)}. \end{aligned} \quad (3.36)$$

We now compute for any  $n > N$

$$\begin{aligned} \sum_{i=N}^{n-1} M_{n-i}^{(r)}(T_r^{i-1}(T_r - \text{Id})f) &= \sum_{i=1}^{n-N} M_i^{(r)}(T_r^{n-i} f - T_r^{n-(i+1)} f), \\ &= M_1^{(r)}(T_r^{n-1} f) + \sum_{i=2}^{n-N} (M_i^{(r)} - M_{i-1}^{(r)})(T_r^{n-i} f) - M_{n-N}^{(r)}(T_r^{N-1} f), \\ &= \sum_{i=1}^{n-N} \Delta_i^{(r)}(Q_r^{n,i,N} f), \end{aligned}$$

where we appealed to equality C.6.1 (see Appendix C.6) in the first line. Let us exploit the martingale decomposition provided in equation (3.21) by writing

$$\sum_{i=1}^{n-N} \Delta_i^{(r)}(Q_r^{n,i,N} f) = \sum_{i=1}^{n-N} A_i^{(r)}(Q_r^{n,i,N} f) + \sum_{i=1}^{n-N} B_i^{(r)}(Q_r^{n,i,N} f). \quad (3.37)$$

First, focusing on the second term above which corresponds to the  $L^1$  part, we get the following bound

$$\begin{aligned} \sup_{n > N} \left| \sum_{i=1}^{n-N} B_i^{(r)}(Q_r^{n,i,N} f) \right| &\leq \sum_{n=N+1}^{\infty} \sum_{i=1}^{n-N} |B_i^{(r)}(Q_r^{n,i,N} f)|, \\ &\leq 2 \sum_{i=1}^{\infty} B_i^{(r)+} \sum_{n=i+N}^{\infty} \|T_r^{n-i} f\|_{\mathcal{B}(V)}, \\ &\leq 2\|f\|_{\mathcal{B}(V)} \frac{a_r^N}{1-a_r} \sum_{i=1}^{\infty} B_i^{(r)+}, \end{aligned}$$

where we used the upper bound (3.24) in the second line. Recalling the  $a_r$  contraction of  $T_r$ , this proves a.s. and  $L^1$  convergence of the  $L^1$  component corresponding to  $B$ .

Let us now turn to the  $L^2$  part. Observing again that the random variables composing  $A_i^{(r)}(f)$  are centred, we re-write the sum of squares as follows

$$\mathbb{E}_{\delta_x} \left( \sum_{i=1}^{n-N} A_i^{(r)}(Q_r^{n,i,N} f) \right)^2 \leq \sum_{i=1}^{n-N} \mathbb{E}_{\delta_x} \left( A_i^{(r)}(Q_r^{n,i,N} f) \right)^2.$$

Next, given that by (3.29), we have for any  $f \in \mathcal{B}(V^*)$ ,

$$\mathbb{E} \left( A_i^{(r)}(f)^2 \mid \mathcal{F}_{i-1}^{(r)} \right) \leq \|f\|_{\mathcal{B}(V)}^2 \lambda^{-ir} X_i^{(r)}(V_i^{(r)}),$$

we can make similar estimates to evaluate  $\|A_i^{(r)}(Q_r^{n,i,N} f)\|_{\mathcal{B}(V)}^2$ . Indeed, applying the Tower

property in the first line, we get

$$\begin{aligned}\mathbb{E}_{\delta_x} \left( A_i^{(r)}(Q_r^{n,i,N} f) \right)^2 &= \mathbb{E}_{\delta_x} \left( \mathbb{E} \left( A_i^{(r)}(Q_r^{n,i,N} f) \right)^2 \middle| \mathcal{F}_{i-1}^{(r)} \right), \\ &\leq \|Q_r^{n,i,N} f\|_{\mathcal{B}(V)}^2 \lambda^{-ir} \mathbb{E}_{\delta_x} \left( X_i^{(r)}(V_i^{(r)}) \right).\end{aligned}$$

Next, by Assumption (A2),

$$\mathbb{E}_{\delta_x} \left( A_i^{(r)}(Q_r^{n,i,N} f) \right)^2 \leq \|Q_r^{n,i,N} f\|_{\mathcal{B}(V)}^2 \lambda^{-ir} \left( h(x) \gamma(V_i^{(r)}) + a_{ir} V^*(x) \|V_i^{(r)}\|_{\mathcal{B}(V^*)} \right).$$

If we now introduce for any  $i \in \mathbb{N}$  the function  $x \mapsto u_i(x)$  given by

$$u_i(x) := \lambda^{-ir} \left( h(x) \gamma(V_i^{(r)}) + a_{ir} V^*(x) \|V_i^{(r)}\|_{\mathcal{B}(V^*)} \right),$$

we obtain, by using the contraction given in (3.36) for  $\|Q_r^{n,i,N} f\|_{\mathcal{B}(V)}$ ,

$$\mathbb{E}_{\delta_x} \left( A_i^{(r)}(Q_r^{n,i,N} f) \right)^2 \leq 4a_r^{2(n-i)} \|f\|_{\mathcal{B}(V)}^2 u_i(x).$$

Thus, summing over  $i$ , we have

$$\mathbb{E}_{\delta_x} \left( \sum_{i=1}^{n-N} A_i^{(r)}(Q_r^{n,i,N} f) \right)^2 \leq 4\|f\|_{\mathcal{B}(V)}^2 \sum_{i=1}^{n-N} a_r^{2(n-i)} u_i(x).$$

It now suffices to note that equation (3.30) can be re-written

$$\mathbb{E}_{\delta_x} \left( \sum_{i=1}^{\infty} \lambda^{-ir} X_i^{(r)}(V_i^{(r)}) \right) \leq \sum_{i=1}^{\infty} u_i(x),$$

and recall that the proof of Proposition (3.2) has led us to showing that  $\sum_{i=0}^{\infty} u_i(x) < \infty$ . Finally since  $a_r < 1$ , we have

$$\sum_{N=1}^{\infty} \sum_{n=N+1}^{\infty} \sum_{i=1}^{n-N} a_r^{2(n-i)} u_i(x) = \left( \sum_{i=1}^{\infty} u_i(x) \right) \sum_{N=1}^{\infty} \sum_{k=N+1}^{\infty} a_r^{2k} < \infty.$$

We therefore have now shown that

$$\sum_{N=1}^{\infty} \sum_{n=N+1}^{\infty} \mathbb{E}_{\delta_x} \left( \sum_{i=1}^{n-N} A_i^{(r)}(Q_r^{n,i,N} f) \right)^2 < \infty. \quad (3.38)$$

Observing that

$$\sup_{n>N} \left( \sum_{i=1}^{n-N} A_i^{(r)}(Q_r^{n,i,N} f) \right)^2 \leq \sum_{n=N+1}^{\infty} \left( \sum_{i=1}^{n-N} A_i^{(r)}(Q_r^{n,i,N} f) \right)^2,$$

we get by taking expectation on each side in the above inequality and summing over  $N$

$$\sum_{N=1}^{\infty} \mathbb{E}_{\delta_x} \left( \sup_{n>N} \left( \sum_{i=1}^{n-N} A_i^{(r)}(Q_r^{n,i,N} f) \right)^2 \right) \leq \sum_{N=1}^{\infty} \mathbb{E}_{\delta_x} \left( \sum_{n=N+1}^{\infty} \left( \sum_{i=1}^{n-N} A_i^{(r)}(Q_r^{n,i,N} f) \right)^2 \right).$$

At this point, we can rely on the bound of inequality (3.38) and on Tonelli's theorem for the right-hand side of the above inequality to conclude that

$$\sum_{N=1}^{\infty} \mathbb{E}_{\delta_x} \left( \sup_{n>N} \left( \sum_{i=1}^{n-N} A_i^{(r)}(Q_r^{n,i,N} f) \right)^2 \right) < \infty \quad \text{a.s.} \quad (3.39)$$

This ensures the following  $L^2$  convergence result

$$\mathbb{E}_{\delta_x} \left( \sup_{n>N} \left( \sum_{i=1}^{n-N} A_i^{(r)}(Q_r^{n,i,N} f) \right)^2 \right) \xrightarrow{N \rightarrow \infty} 0,$$

as this is trivially the term of a convergent series in (3.39), and

$$\sum_{N=1}^{\infty} \sup_{n>N} \left( \sum_{i=1}^{n-N} A_i^{(r)}(Q_r^{n,i,N} f) \right)^2 < \infty \quad \text{a.s.},$$

by the non-negativity of the terms in the series. Finally the above yields

$$\sup_{n>N} \left| \sum_{i=1}^{n-N} A_i^{(r)}(Q_r^{n,i,N} f) \right| \xrightarrow{N \rightarrow \infty} 0 \quad \text{a.s.},$$

which ensures the a.s. and  $L^1$  convergence to 0 of the  $L^2$  part involving  $\sum_{i=1}^{n-N} A_i^{(r)}(Q_r^{n,i,N} f)$  in (3.37). Recalling that these two types of convergence also hold for the  $L^1$  component represented by  $\sum_{i=1}^{n-N} B_i^{(r)}(Q_r^{n,i,N} f)$  ends this proof.  $\square$

### 3.2.6 Proof of convergence in discrete time

We are now in a position to prove Theorem 3.1. We start with the following proposition.

**Proposition 3.3.** *Let the assumptions of Theorem 3.1 hold. Let  $Z_0 = x$ . Then for any  $f \in \mathcal{B}(V)$  and any  $r \in \mathbb{N}$  such that  $a_r < 1$ , the following convergence holds*

$$\lim_{n \rightarrow \infty} X_n^{(r)}(f) = \gamma(f) W^{(x)}, \quad \text{a.s. and in } L^1,$$

where  $W^{(x)} \geq 0$  is a non-negative random variable that does not depend on  $f$  nor on  $r \in \mathbb{N}^*$ .

*Proof.* We use Lemma 3.2.2 to write

$$X_n^{(r)}(f) = X_0^{(r)}(T_r^n f) + \gamma(f) X_{n-1}^{(r)}(h) + R_n^{(r)}(f),$$

and split the rest  $R_n^{(r)}$  as follows for any  $N \geq 1$ ,

$$R_n^{(r)}(f) = M_n^{(r)}(f) + \sum_{i=1}^{N-1} M_{n-i}^{(r)}(T_r^{i-1}(T_r - \text{Id})f) + \sum_{i=N}^{n-1} M_{n-i}^{(r)}(T_r^{i-1}(T_r - \text{Id})f).$$

Here we proceed in successive steps as we need to treat both almost sure convergence and  $L^1$  convergence. We first rely on Lemma 3.2.3 to choose a random integer  $N$  large enough so that the last term is uniformly small for any  $n > N$  a.s. Next, we choose a (non-random) integer  $N$  large enough so that the  $L^1$  norm of the last term is small enough. Finally, we use Proposition 3.2 to obtain the convergence a.s. and in  $L^1$  of  $M_n^{(r)}(f)$  to  $M_\infty^{(r)}(f)$  and the convergence of  $M_{n-i}^{(r)}(T_r^{i-1}(T_r - \text{Id})f)$  to  $M_\infty^{(r)}(T_r^{i-1}(T_r - \text{Id})f)$  as  $n$  tends to infinity for  $i \leq N$ . Gathering these results leads to

$$\begin{aligned} \lim_{n \rightarrow \infty} R_n^{(r)}(f) &= M_\infty^{(r)}(f) + \sum_{i=1}^{\infty} M_\infty^{(r)}(T_r^{i-1}(T_r - \text{Id})f), \\ &= M_\infty^{(r)}(f) + \sum_{i=1}^{\infty} M_\infty^{(r)}(T_r^i f) - M_\infty^{(r)}(T_r^{i-1} f) = 0, \end{aligned}$$

as the sum in the last line is telescoping. Finally, the convergence holds a.s. and in  $L^1$ .

To recap, we have now shown that when  $n \rightarrow \infty$ , the remainder term  $R_n^{(r)}(f) \rightarrow 0$  a.s. and in  $L^1$ . It then suffices to recall as well that when  $n \rightarrow \infty$ ,  $T_r^n(f) \rightarrow 0$  (contraction argument) to see that  $X_0^{(r)} T_r^n(f) \rightarrow 0$ . Finally, by Proposition 3.2, we recall  $X_{n-1}^{(r)}(h) \rightarrow W^{(x)}$  a.s. and in  $L^1$  when  $n \rightarrow \infty$ . Thus, the limiting random variable does not depend on  $f$  or on  $r$  and is determined by  $W^{(x)}$ . This ends the proof.  $\square$

*End of proof of Theorem 3.1.* We can now finalise the proof of the last part of Theorem 3.1 which, given Proposition 3.3, becomes very direct. Indeed, we now wish to prove that for any  $f \in \mathcal{B}(V)$ ,

$$\frac{Z_n(f)}{\lambda^n} \xrightarrow{n \rightarrow \infty} \gamma(f)W^{(x)}, \quad \text{a.s. and in } L^1.$$

But we have just proven in Proposition 3.3 that for  $r \in \mathbb{N}$  large enough such that  $a_r < 1$ , the following convergence holds

$$\frac{Z_{nr}(f)}{\lambda^{nr}} \xrightarrow{n \rightarrow \infty} \gamma(f)W^{(x)}, \quad \text{a.s. and in } L^1.$$

Let us next introduce for any integer  $k = 0, 1, \dots, r-1$ , the process

$$X_n^{(r,k)}(f) = X_{nr+k}(f).$$

Obviously, for  $k = 0$ , we are back to  $X_n^{(r,0)}(f) = X_n^{(r)}(f)$ . Thus assuming an initial condition  $Z_0 = \delta_x$ , Proposition 3.3 tells us that

$$X_n^{(r,0)}(f) \xrightarrow{n \rightarrow \infty} \gamma(f)W^{(x)}, \quad \text{a.s. and in } L^1.$$

Now, when  $k = 1$ , we observe that the process  $X_n^{(r,1)}(f) = X_{nr+1}(f)$  can still be viewed as a process of the form  $X_n^{(r)}(f)$  but whose initial condition is translated from  $Z_0 = x$  to a new

initial condition determined by the number of individuals in the first generation with traits vector  $X_1$ . Since this number of individuals is finite, we can apply the results of Proposition 3.3 a finite number of times to each of the individuals composing the first generation, which leads to

$$X_{nr+1}(f) \xrightarrow{n \rightarrow \infty} \gamma(f)W^{(r,1,x)}, \quad \text{a.s. and in } L^1.$$

Here we have purposefully re-introduced a dependence on  $r$  and  $k = 1$  to the limiting random variable which consequently gets written for the moment  $W^{(r,1,x)}$  but note that  $W^{(r,1,x)}$  still does not depend on  $f$ . Proposition 3.3 has shown us that the limiting martingale does not depend on  $r$  when  $k = 0$ . However when  $k = 1, \dots, r-1$ , the possible dependence on  $k$  may implicitly lead to a dependence on  $r$ , hence why we keep the notations as above.

However, the same process, just done with  $k = 1$ , can equally be applied to  $k = 2, \dots$  up until  $k = r-1$ . In each case, we simply apply Proposition 3.3 a finite number of times to the translated initial conditions (which always correspond to a finite number of initial individuals). Thus we conclude that, for any  $r \in \mathbb{N}$  and any  $k = 0, \dots, r-1$ ,

$$X_{nr+k}(f) \xrightarrow{n \rightarrow \infty} \gamma(f)W^{(r,k,x)}, \quad \text{a.s. and in } L^1,$$

where the random variable  $W^{(r,k,x)}$  does not depend on  $f$  but depends on the initial condition and may depend on  $r$  and  $k$ .

At this point, it suffices to observe that when taking  $r = 1$ ,  $k = 0$  and  $f = h$ , we obtain the martingale  $X_n(h) = \lambda^{-n}Z_n(h)$  which we know converges to  $W^{(x)}$  (since  $\gamma(h) = 1$ ). Thus,  $W^{(r,k,x)} = W^{(x)}$  depends neither on  $r$  nor on  $k$  and only on the initial condition.

We can finally conclude that the following result holds

$$X_{nk+r}(f) \xrightarrow{n \rightarrow \infty} \gamma(f)W^{(x)}, \quad \text{a.s. and in } L^1.$$

Noting that the set of integers of the form  $\{kr \mid k = 0, \dots, r-1; r \in \mathbb{N}\}$  covers the entire set  $\mathbb{N}$  and appealing to the fact that all the subsequences  $X_{nr+k}(f)$  converge to the same limit, namely  $\gamma(f)W^{(x)}$ , ends the proof.  $\square$

*Remark.* Let us mention that the results of the previous section should allow us to go a bit farther and provide estimates not only of the probability that  $W$  is positive (recalling that  $\mathbb{P}(W > 0)$  represents the probability of non-extinction), but also of the way this probability of non-extinction depends on the initial type. We believe that these potential estimates could be achieved using the Cauchy-Schwarz inequality. However, to our knowledge, in practice, it is sufficient to simply describe the survival event, and it seems that additional conditions are needed in specific cases to prove that when the population globally survives, then  $W$  is positive.

Before moving on to the continuous-time setting in Section 3.3, we complement our discrete-time results with the following classical limit result. It is a consequence of our above estimates and will be useful, in particular, for the proof of Theorem 3.2 in continuous time.

**Lemma 3.2.4.** *Under Assumptions of Theorem 3.1, for any  $f \in \mathcal{B}(V)$  and  $r \geq 1$ ,*

$$\lim_{n \rightarrow \infty} \Delta_n^{(r)}(f) = \lim_{n \rightarrow \infty} X_n^{(r)}(f) - \mathbb{E}\left(X_n^{(r)}(f) \mid \mathcal{F}_n\right) = 0 \quad \text{a.s.}$$

*Proof.* Recall the decomposition

$$\Delta_n^{(r)}(f) = A_n^{(r)}(f) + B_n^{(r)}(f),$$

where  $A_n^{(r)}(f)$  and  $B_n^{(r)}(f)$  are defined in (3.22) and (3.23). From the proof of Proposition 3.2, we know that

$$\mathbb{E} \left( \sum_{n=1}^{\infty} |B_n^{(r)}(f)| \right) < \infty, \quad \mathbb{E} \left( \sum_{n=1}^{\infty} A_n^{(r)}(f)^2 \right) < \infty.$$

Thus

$$\sum_{n=1}^{\infty} |B_n^{(r)}(f)| < \infty, \quad \sum_{n=1}^{\infty} A_n^{(r)}(f)^2 < \infty, \quad \text{a.s.},$$

which ensure that  $B_n^{(r)}(f)$  and  $A_n^{(r)}(f)^2$  and thus  $A_n^{(r)}(f)$  go to 0 a.s. as  $n$  tends to infinity.  $\square$

### 3.3 Strong law of large numbers in continuous time

#### 3.3.1 Assumptions and main theorem in continuous time

This section briefly handles the continuous-time setting. Here we further assume that  $\mathcal{X}$  is a separable metric space. This additional condition related to the abstract space  $\mathcal{X}$  also appears in [AH76] and is used at the end of the proof of Theorem 3.1. We will not go into as much detail as in the discrete-time setting and simply consider (assume the existence of) a continuous-time measure-valued càdlàg Markov branching process  $(Z_t)_{t \geq 0}$  in  $\mathcal{X}$ . Then, for any  $x \in \mathcal{X}$ , we define its first-moment semigroup

$$S_t f(x) = \mathbb{E}_{\delta_x}(Z_t(f)),$$

which is nonnegative but could be infinite.

Let us first state the set of assumptions needed to prove Theorem 3.2.

**Assumptions** (Continuous time).

(A4) *There exists a measurable function  $V : \mathcal{X} \rightarrow (0, \infty)$  and a positive triplet  $(\gamma, h, \lambda)$  of eigenelements such that  $\lambda > 0$ ,  $\gamma$  is a probability measure on  $\mathcal{X}$ ,  $h : \mathcal{X} \rightarrow (0, \infty)$  is lower semi-continuous almost everywhere with respect to  $\gamma$  and*

$$\gamma S_t = e^{\lambda t} \gamma, \quad S_t h = e^{\lambda t} h, \quad \gamma(h) = 1, \quad \sup_{\mathcal{X}} h/V < \infty, \quad \gamma(V) < \infty.$$

(A5) *For any  $t \geq 0$  and  $x \in \mathcal{X}$ ,*

$$S_t V(x) < \infty.$$

(A6) *There exists a decreasing function  $t \mapsto a(t)$  on  $\mathbb{R}^+$  such that for any  $t \geq 0$ ,*

$$\sup_{|f| \leq V} \left| e^{-\lambda t} S_t f(x) - h(x) \gamma(f) \right| \leq a(t) V(x), \quad \int_1^\infty \frac{a(s)}{s} ds < \infty.$$

(A7) There exists  $t_0 > 0$  such that for any  $t \in (0, t_0]$ ,

$$\sup_{x \in \mathcal{X}} \frac{\mathbb{E}_{\delta_x}(Z_t(V) \log^* Z_t(V))}{V(x)} < \infty.$$

*Remark.* Note that Assumption (A5) implies that  $S_t f(x) < \infty$  for any  $x \in \mathcal{X}$  and any function  $f$  such that  $|f| \leq V$ .

We now show the continuous-time analogue to Theorem 3.1 (and observe that the second assertion of Theorem 3.2 is valid only for  $f \in \mathcal{B}(h)$ ). Note also that the  $L^1$  convergence in the continuous-time setting is currently work in progress.

**Theorem 3.2.** *Let  $x \in \mathcal{X}$ . Assume  $Z_0 = \delta_x$  and let Assumptions (A4) to (A7) hold. Then the limiting martingale  $W^{(x)}$  satisfies*

$$\mathbb{E}_{\delta_x}(W^{(x)}) = h(x) \quad \text{and} \quad \lim_{t \rightarrow \infty} e^{-\lambda t} Z_t(h) = W^{(x)}, \quad a.s.$$

Moreover the following convergence holds almost surely for any  $f \in \mathcal{B}(h)$ ,

$$\lim_{t \rightarrow \infty} e^{-\lambda t} Z_t(f) = \gamma(f) W^{(x)}, \quad a.s.$$

### 3.3.2 Proof of the main theorem in continuous time

As mentioned above, for the moment, we prove almost sure convergence in continuous time while our  $L^1$  proof is being finalised and will be detailed in the article to be submitted on ArXiv in the early part of 2025. Let us first prove here the following intermediary lemma on which is based the proof of Theorem 3.2.

**Lemma 3.3.1.** *Let  $f \in \mathcal{B}_+(V)$  and  $A \subset \mathcal{X}$  be a measurable subset such that  $\gamma(\partial A) = 0$ . Then,*

$$\liminf_{t \rightarrow \infty} e^{-\lambda t} Z_t(f \mathbf{1}_A) \geq \gamma(f \mathbf{1}_A) W^{(x)}.$$

*Proof.* For any  $\epsilon > 0$ ,  $x \in \mathcal{X}$ ,  $A \subset \mathcal{X}$  and  $f \in \mathcal{B}_+(V)$ , we introduce the subset of  $A$  defined as

$$A_f^\epsilon(x) = \left\{ y \in A \mid f(y) > \frac{1}{1+\epsilon} f(x) \right\}.$$

In words, points in  $A_f^\epsilon(x)$  are such that their image by  $f$  is strictly greater than  $f(x)$ . Next, for any  $t \geq 0$ , we write

$$t = n\delta + s, \quad n \in \mathbb{N}, \quad \delta > 0, \quad s \in [0, \delta).$$

We denote respectively by  $\mathcal{U}_{n\delta}$  the set of individuals alive at time  $n\delta$ , and  $\mathcal{U}_{n\delta+s}(u)$  the set of individuals alive at time  $n\delta + s$  but issued from individual  $u$  alive at  $n\delta$ .

Then, for  $n \in \mathbb{N}$ ,  $\delta > 0$  and  $\epsilon > 0$ , we introduce the function  $F_u^{n,\delta,\epsilon} : x \mapsto F_u^{n,\delta,\epsilon}(x)$  on  $\mathcal{X}$  defined by

$$F_u^{n,\delta,\epsilon}(x) := (1 + \epsilon)^{-1} f(x) \mathbf{1}_{\{Z_{n\delta+s}(v) \in A_f^\epsilon(x), \forall v \in \mathcal{U}_{n\delta+s}(u), \forall s \in [0, \delta)\}}.$$

The rationale behind this function is to select individuals  $u \in \mathcal{U}_{n\delta}$  whose descendants  $v$  at time  $n\delta + s$  are such that  $f(Z_{n\delta+s}(v)) > \frac{1}{1+\epsilon} f(Z_{n\delta}(u))$ . We will however simplify notations

and write  $F_u(x) := F_u^{n,\delta,\epsilon}(x)$ , as  $n, \delta$  and  $\epsilon$  are fixed in our context. By the branching property, we have

$$\begin{aligned} Z_{n\delta+s}(f\mathbf{1}_A) &= \sum_{u \in \mathcal{U}_{n\delta}} \sum_{v \in \mathcal{U}_{n\delta+s}(u)} f\mathbf{1}_A(Z(v)), \\ &\geq \sum_{u \in \mathcal{U}_{n\delta}} (1+\epsilon)^{-1} f(Z_{n\delta}(u)) \mathbf{1}_{\{Z_{n\delta+s}(v) \in A_f^\epsilon(Z(u)), \forall v \in \mathcal{U}_{n\delta+s}(u), \forall s \in [0, \delta]\}}, \\ &\geq \sum_{u \in \mathcal{U}_{n\delta}} F_u(Z_{n\delta}(u)), \end{aligned} \quad (3.40)$$

where the first inequality relies on the definition of the subset  $A_f^\epsilon(Z(u))$  and the second one relies on the definition of  $F_u$ . Now, we introduce for any  $\delta, \epsilon > 0$  the function  $y \mapsto \xi^{\delta,\epsilon}(y)$  on  $\mathcal{X}$  defined as

$$\xi^{\delta,\epsilon}(y) = \mathbb{P}_{\delta_y} (Z_s(v) \in A_f^\epsilon(x), \forall v \in \mathcal{U}_s, \forall s \in [0, \delta]),$$

and observe

$$\begin{aligned} \xi^{\delta,\epsilon}(Z_{n\delta}(u)) &= \mathbb{P}_{Z_{n\delta}(u)} (Z_s(v) \in A_f^\epsilon(Z_{n\delta}(u)), \forall v \in \mathcal{U}_s, \forall s \in [0, \delta]), \\ &= \mathbb{P}_{\delta_x} (Z_{n\delta+s}(v) \in A_f^\epsilon(Z_{n\delta}(u)), \forall v \in \mathcal{U}_{n\delta+s}(u), \forall s \in [0, \delta] \mid \mathcal{F}_{n\delta}), \end{aligned} \quad (3.41)$$

where we once again appealed to the branching property in the second equality. We now consider, relying on the definition of  $F_u$  in the first equality,

$$\begin{aligned} &\mathbb{E}(\mathbf{1}_{u \in \mathcal{U}_{n\delta}} F_u(Z_{n\delta}(u)) \mid \mathcal{F}_{n\delta}), \\ &= \mathbf{1}_{u \in \mathcal{U}_{n\delta}} (1+\epsilon)^{-1} \mathbb{E} \left( f(Z_{n\delta}(u)) \mathbf{1}_{\{Z_{n\delta+s}(v) \in A_f^\epsilon(Z_{n\delta}(u)), \forall v \in \mathcal{U}_{n\delta+s}(u), \forall s \in [0, \delta]\}} \mid \mathcal{F}_{n\delta} \right), \\ &\geq \mathbf{1}_{u \in \mathcal{U}_{n\delta}} (1+\epsilon)^{-1} f(Z_{n\delta}(u)) \xi^{\delta,\epsilon}(Z_{n\delta}(u)), \end{aligned}$$

where we used expression (3.41) for  $\xi^{\delta,\epsilon}(Z_{n\delta}(u))$  in the last line. Now, taking conditional expectation on each side of (3.40), we can write

$$\begin{aligned} \mathbb{E}(Z_{n\delta+s}(f\mathbf{1}_A) \mid \mathcal{F}_{n\delta}) &\geq \mathbb{E} \left( \sum_{u \in \mathcal{U}_{n\delta}} \mathbf{1}_{u \in \mathcal{U}_{n\delta}} F_u(Z_{n\delta}(u)) \mid \mathcal{F}_{n\delta} \right), \\ &\geq (1+\epsilon)^{-1} \sum_{u \in \mathcal{U}_{n\delta}} (f \cdot \xi^{\delta,\epsilon})(Z_{n\delta}(u)). \end{aligned}$$

If we define

$$X_t(f) := e^{-\lambda t} Z_t(f),$$

we get

$$\mathbb{E}(X_{n\delta+s}(f\mathbf{1}_A) \mid \mathcal{F}_{n\delta}) \geq \lambda^{-s} (1+\epsilon)^{-1} \sum_{u \in \mathcal{U}_{n\delta}} (f \cdot \xi^{\delta,\epsilon})(X_{n\delta}(u)).$$

But recalling Lemma 3.2.4, we have

$$\liminf_{n \rightarrow \infty} X_{n\delta+s}(f\mathbf{1}_A) = \liminf_{n \rightarrow \infty} \mathbb{E}(X_{n\delta+s}(f\mathbf{1}_A) \mid \mathcal{F}_{n\delta}),$$

and thus obtain for any  $\delta, \epsilon > 0$ ,

$$\begin{aligned} \liminf_{n \rightarrow \infty} X_{n\delta+s}(f\mathbf{1}_A) &\geq \liminf_{n \rightarrow \infty} \left( \lambda^{-s}(1+\epsilon)^{-1} \sum_{u \in \mathcal{U}_{n\delta}} (f \cdot \xi^{\delta, \epsilon})(X_{n\delta}(u)) \right), \\ &\geq \lambda^{-s}(1+\epsilon)^{-1} \gamma(f \cdot \xi^{\delta, \epsilon})W^{(x)}, \end{aligned}$$

where we were able to apply Theorem 3.1 to the function  $f \cdot \xi^{\delta, \epsilon}$  in the last inequality. Here observe that by the simple definition of the subset  $A_f^\epsilon(x)$ ,  $Z_s(v) \in A_f^\epsilon(x)$  is equivalent to

$$Z_s(v) \in A, \quad \text{and} \quad f(Z_s(v)) > \frac{1}{1+\epsilon} f(x).$$

Now it suffices to recall that the process  $Z_t$  is right-continuous and appeal to the assumption that the boundary of  $A$  has a null  $\gamma$ -measure (i.e.  $\gamma(\delta A) = 0$ ) to obtain

$$\begin{aligned} \lim_{\delta \rightarrow 0} \xi^{\delta, \epsilon}(x) &= \lim_{\delta \rightarrow 0} \mathbb{P}_{\delta x} (Z_s(v) \in A_f^\epsilon(x), \forall v \in \mathcal{U}_s, \forall s \in [0, \delta]), \\ &\geq \mathbf{1}_A(x), \end{aligned}$$

for  $\gamma$ -almost all  $x \in \mathcal{X}$ . So we have, since  $\lambda^{-s} \rightarrow 1$  as  $\delta$  and thus  $s \rightarrow 0$ ,

$$\liminf_{n \rightarrow \infty} X_{n\delta+s}(f\mathbf{1}_A) \geq (1+\epsilon)^{-1} \gamma(f\mathbf{1}_A)W^{(x)}.$$

Finally, letting  $\epsilon \rightarrow 0$  leads to

$$\liminf_{t \rightarrow \infty} \lambda^{-t} Z_t(f\mathbf{1}_A) \geq \gamma(f\mathbf{1}_A)W^{(x)},$$

which is the limiting result we were set to prove.  $\square$

*Proof of Theorem 3.2.* We are now in a position to finalise the proof of Theorem 3.2. To do so, let us first apply Lemma 3.3.1 to both  $A$  and its complementary  $A^c$ . We thus have on one hand,

$$\liminf_{t \rightarrow \infty} e^{-\lambda t} Z_t(h\mathbf{1}_A) \geq \gamma(h\mathbf{1}_A)W^{(x)},$$

and on the other hand,

$$\begin{aligned} \limsup_{t \rightarrow \infty} e^{-\lambda t} Z_t(h\mathbf{1}_A) &= \limsup_{t \rightarrow \infty} e^{-\lambda t} (Z_t(h) - Z_t(h\mathbf{1}_{A^c})), \\ &\leq \limsup_{t \rightarrow \infty} e^{-\lambda t} Z_t(h) - \liminf_{t \rightarrow \infty} e^{-\lambda t} Z_t(h\mathbf{1}_{A^c}). \end{aligned}$$

But since  $\limsup_{t \rightarrow \infty} e^{-\lambda t} Z_t(h) = \lim_{t \rightarrow \infty} e^{-\lambda t} Z_t(h) = W^{(x)}$ , we also get (recalling  $\gamma(h) = 1$ )

$$\begin{aligned} \limsup_{t \rightarrow \infty} e^{-\lambda t} Z_t(h\mathbf{1}_A) &\leq \gamma(h)W^{(x)} - \gamma(h\mathbf{1}_{A^c})W^{(x)}, \\ &= \gamma(h\mathbf{1}_A)W^{(x)}. \end{aligned}$$

The latter ensures that

$$\lim_{t \rightarrow \infty} e^{-\lambda t} Z_t(h\mathbf{1}_A) = \gamma(h\mathbf{1}_A)W^{(x)}. \quad (3.42)$$

Let us now take  $f \in \mathcal{B}(h)$  and introduce the sequence of random measures  $(\mu_t(f))_{t \geq 0}$  given by

$$\mu_t(f) := \frac{e^{-\lambda t} Z_t(fh)}{e^{-\lambda t} Z_t(h)}.$$

We also define for  $f \in \mathcal{B}(h)$  and any  $A \subset \mathcal{X}$  the measure

$$\gamma^f(A) := \int_A f d\gamma = \gamma(f \mathbf{1}_A).$$

Then for each continuity set  $A$  of  $\gamma^h$ , (i.e. for each set such that  $\gamma(\delta A) = 0$ ), the limiting result in (3.42) becomes

$$\lim_{t \rightarrow \infty} \frac{e^{-\lambda t} Z_t(h \mathbf{1}_A)}{e^{-\lambda t} Z_t(h)} = \lim_{t \rightarrow \infty} \mu_t(\mathbf{1}_A) = \gamma^h(A), \quad \text{a.s.}$$

We now appeal to the càdlàg property of  $Z_t$ , and to the result of weak convergence of measures given in [Bil13, Theorem 2.3]. The latter requires that the space  $\mathcal{X}$  be a separable metric space. [Bil13, Theorem 2.3] allows us to say that for any lower semi-continuous function  $f \in \mathcal{B}(h)$ , the sequence of measures  $(\mu_t(f))_{t \geq 0}$  converges to  $\gamma^f$ . In other words,

$$\lim_{t \rightarrow \infty} e^{-\lambda t} Z_t(f) = \gamma(f) W^{(x)}, \quad \text{a.s.},$$

which is the convergence result we were set to prove in the continuous case.  $\square$

### 3.3.3 Application to branching elliptic diffusion

In this section, we briefly evoke a possible useful application of our results. However, we will not go into all the intricate details involved in the proofs, as the latter would require introducing a lot of technicalities about the underlying example. Thus, we will limit ourselves to presenting the main points to build an example based on branching elliptic diffusion [CV23]. This should hopefully give a sense of how our results can be more generally applied to other contexts involving various branching Markov processes.

As mentioned above, our example is based mainly on the works of Champagnat and Villemonais [CV23]. We consider a branching Markov process in which each particle  $u \in \mathcal{U}$  is characterised at time  $t \geq 0$  by a trait  $X_t^u$  that belongs to  $\mathcal{X} = D$ , an open connected subset of  $\mathbb{R}^d$ ,  $d \geq 1$ . Between branching events, the dynamics of the trait are described by a process  $(X_t)_{t \geq 0}$  solution to the following stochastic differential equation (SDE)

$$dX_t = b(X_t)dt + \sigma(X_t)dB_t, \quad (3.43)$$

where  $(B_t)_{t \geq 0}$  is a standard  $r$ -dimensional Brownian motion,  $b : D \rightarrow \mathbb{R}^d$  and  $\sigma : D \rightarrow \mathbb{R}^{d \times r}$  are both locally Hölder functions and  $\sigma$  is locally uniformly elliptic in  $D$ . The diffusion process  $(X_t)_{t \geq 0}$  is assumed to be immediately absorbed at a cemetery point  $\delta \notin D$  at the first exit point  $\tau_{\text{exit}}$  of  $D$  defined as

$$\tau_{\text{exit}} = \inf\{t \geq 0 : X_t \in D^c\}.$$

As mentioned in [CV23], the proof of existence, as well as the construction of such a diffusion process both necessitate some work since the coefficients  $b(\cdot)$  and  $\sigma(\cdot)$  are only defined on the open set  $D$  but not at the boundary point  $\delta$ . However, we refer the reader to [CV23, Section

12.1] for the construction of a process  $(X_t)_{t \geq 0}$  as a weak solution to (3.43) up to the first exit time

$$\tau_{K_k^c} := \inf\{t \in \mathbb{R}_+ : X_t \in K_k^c\}$$

of each compact subset  $K_k \subset D$  defined for any  $k \in \mathbb{N}^*$  as

$$K_k := \{x \in D : |x| \leq k \text{ and } d(x, D^c) \geq 1/k\}.$$

In this case, the first exit time is given by

$$\tau_{\text{exit}} = \sup_{k \geq 1} \tau_{K_k^c}.$$

Let us now describe the branching events. Any particle  $u \in \mathcal{U}$  with trait  $X_t^u = x$  branches at a rate  $r(x)$  and produces  $k \in \mathbb{N}$  offspring with probability  $p_k(x)$ . These offspring have the same trait  $x$ . We denote the mean number of offspring  $m(x) := \sum_{k=1}^{\infty} k p_k(x)$  for any  $x \in D$ . We now introduce a sufficient condition to ensure the existence of such a process.

**Assumptions** (Sufficient to ensure existence of the process).

(A8) For any  $k \in \mathbb{N}$  the functions  $x \mapsto p_k(x)$  and  $x \mapsto r(x)$  are continuous over  $D$ .

(A9) The following bound holds

$$\sup_{x \in \mathcal{X}} r(x) \sum_{k=1}^{\infty} k \log(k) p_k(x) < +\infty.$$

We will not go into the details of the proof here, but it can be proven that Assumptions (A8) and (A9) ensure the existence and uniqueness of such a branching process and additionally guarantee that its first-moment semigroup is well defined for any bounded function.

Now that we have introduced the branching Markov process  $Z_t$  whose trait  $X_t$  is a weak solution to (3.43), we wish to study its asymptotic behaviour in the supercritical case. To do so, we place ourselves in the framework of absorbed diffusion processes with killing as per [CV23, Section 4.4] and obtain sufficient conditions for the branching process to converge a.s. towards its limiting martingale. However, to properly cite our theorem, we still need to introduce a couple of mathematical objects. Let  $x \mapsto \kappa(x)$  on  $D$  be the measurable locally bounded function given by

$$\kappa(x) := \bar{r} - r(x)(m(x) - 1) \geq 0,$$

where

$$\bar{r} := \sup_{x \in D} (r(x)(m(x) - 1)) < \infty.$$

We will consider  $(X^\kappa)_{t \geq 0}$  corresponding to the diffusion process  $(X_t)_{t \geq 0}$ , weak solution to (3.43) but killed at rate  $\kappa$ . More precisely, for an independent exponential random variable  $\zeta$  with parameter 1, we set

$$\tau_\delta = \tau_{\text{exit}} \wedge \inf \left\{ t \geq 0, \int_0^t \kappa(X_s) ds > \zeta \right\},$$

and let  $X_t^\kappa = X_t$  for  $t \leq \tau_\delta$  and  $X_t^\kappa = \delta$  for  $t \geq \tau_\delta$ . We also define as in [CV23] for some  $x \in D$  and open ball  $B$  such that  $\bar{B} \subset D$ , the constant

$$\lambda_0 := \inf\{\ell > 0, \text{ s.t. } \liminf_{t \rightarrow \infty} e^{\ell t} \mathbb{P}_x(X_t \in B) > 0\}.$$

Again, referring to [CV23, Section 12.2], it is proven that in the above context  $\lambda_0 < \infty$  and  $\lambda_0$  does not depend on  $x$  or on  $B$ .

**Assumptions** (Related to the diffusion process).

(A10) *There exists a subset  $D_0 \subsetneq D$  such that*

$$\inf_{x \in D \setminus D_0} \kappa(x) > \lambda_0.$$

(A11) *There exists a time  $s_1 > 0$  such that*

$$\sup_{x \in D_0} \mathbb{P}_x(s_1 < \tau_\delta \wedge \tau_{K_k}) \xrightarrow{k \rightarrow \infty} 0,$$

where  $D_0$  is determined in (A10).

As specified in [CV23, Remark 11], an alternative simple and sufficient condition is also given by  $\lim_{k \rightarrow \infty} \inf_{x \in D \setminus K_k} \kappa(x) = \infty$ . Lastly, we consider the growth rate of the mean population size, defined by

$$\lambda = \limsup_{t \rightarrow \infty} \frac{\log(\mathbb{E}(Z_t(\mathbf{1}_{\mathbb{R}_+})))}{t}.$$

We focus on the supercritical regime when  $\lambda$  is positive. Our results allow us to state the following law of large numbers for branching diffusion.

**Theorem 3.3.** *Under Assumptions (A8) to (A11), let  $Z_0 = x$  and  $\lambda > 0$ . The limiting martingale  $W^{(x)}$  satisfies*

$$\mathbb{E}_{\delta_x}(W^{(x)}) = h(x) \quad \text{and} \quad \lim_{t \rightarrow \infty} e^{-\lambda t} Z_t(h) = W^{(x)} \quad a.s.$$

Moreover the following convergence holds for any  $f \in \mathcal{B}(h)$

$$\lim_{t \rightarrow \infty} e^{-\lambda t} Z_t(f) = \gamma(f) W^{(x)}, \quad a.s.$$

*Proof.* The proof is not the subject of this chapter as its details would appeal to many intricate intermediate results proven in [CV23], as well as to Assumption (A) and Theorem 2.1 in [Ban+22].  $\square$

### 3.4 Conclusion and future work

The results presented in this chapter make up an article that will be submitted in the early part of 2025. This work tends to unify and extend the current asymptotic results with respect to supercritical branching Markov processes. As already mentioned, our starting points were mainly the works of Asmussen and Hering [AH76] and Kesten and Stigum [KS66]. We obtained here a strong law of large numbers for structured branching populations that, in

particular, relies on a set of assumptions that we believe is easier to verify in practice than the assumptions of the first-moment semigroup denoted  $(M)$  introduced by Asmussen and Hering (see Appendices C.3.1-C.3.4). We also relax the uniformity condition required in [AH76].

However, we would like to add the following elements which are still in preparation at the moment: we need to finalise the  $L^1$  convergence proof in continuous time to complement our almost-sure result; we also would like to present a general approach to a main class of examples. This approach will guide in particular the various steps to be taken to ensure that the conclusion of Theorem 3.2 can be applied in various contexts. Each example will always present itself with its own specificities, but the idea will be to present general guidelines to be followed to verify the assumptions of Theorem 3.2.

# CHAPTER 4

## A stochastic process prior for Bayesian deep learning with applications in epidemiology<sup>†</sup>

### 4.1 Introduction

This chapter presents  $\pi$ VAE, a type of variational auto-encoder (VAE) that is proven to be a stochastic process rather than a mere random variable, as a standard VAE usually is. Therefore,  $\pi$ VAE, as a generative model also allows to do proper Bayesian inference and use MCMC techniques and software.

Recalling that a central task in machine learning is to specify a function or a set of functions that best generalise to new data, stochastic processes [Ros96; Pav14] provide a mathematically elegant way to define a class of functions, where each element of a stochastic process is a (usually infinite) collection of random variables. Popular examples of stochastic processes in computational statistics and machine learning are Gaussian processes [RW06], Dirichlet processes [Ant74], log-Gaussian Cox processes [MSW98], Hawkes processes [Haw71], Mondrian processes [RT09], and Gauss-Markov processes [LRL11]. Many of these processes are intimately connected with popular techniques in machine learning and deep learning; for example, both the infinite-width limit of a single-layer neural network and the evolution of a deep neural network by gradient descent are Gaussian processes [Nea96; JGH18].

However, although stochastic processes have many favourable properties and provide a mathematically elegant way to model complex data, providing, in theory at least, flexible priors over function classes that can encode a wide range of interesting assumptions, in practice, efficient inference from a stochastic process by optimisation or marginalisation is difficult, a problem further exacerbated with big data and high-dimensional input spaces. For example, inference and prediction using a Gaussian process requires matrix inversions that scale in a cubic manner with data size, log-Gaussian Cox processes require the evaluation of an intractable integral, and Markov processes are often highly correlated. Bayesian inference can be even more challenging due to complex high-dimensional posterior topologies. One of the standard methods to evaluate posterior expectations relies on Markov Chain Monte Carlo (MCMC) sampling, but high autocorrelation, narrow typical sets and poor scalability have prevented use in big data and complex model settings. A plethora of approximation

---

<sup>†</sup>This chapter, primarily based on an article [Mis+22] published in Statistics and Computing in October 2022, treats a novel variational autoencoder called  $\pi$ VAE, the prior-encoding variational autoencoder.

algorithms exists [Min01; RBB18; LPB17; WT11; Blu+15], but few actually yield accurate posterior estimates [Yao+18; Hug+19; Hof+13; Yao+19].

Rather than relying on approximate Bayesian inference to solve complex models,  $\pi$ VAE [Mis+22] extends the palette of variational autoencoders (VAE) [KW14; RMW14] to develop a more portable model that, in particular, can work with state-of-the-art Bayesian MCMC software such as Stan [Car+17].  $\pi$ VAE is used to learn low-dimensional embeddings of function classes. More precisely,  $\pi$ VAE’s framework is shown to not only accurately learn expressive function classes such as Gaussian processes, but also properties of functions to enable statistical inference (such as the integral of a log-Gaussian process). Inference on the resulting models is tractable and yields accurate posterior expectations and uncertainty. For popular tasks, such as spatial interpolation, Mishra et al. [Mis+22] show that  $\pi$ VAE achieves state-of-the-art performance both in terms of accuracy and computational efficiency.

We note here that the three real-world examples of applications of spatial interpolation provided in [Mis+22] are not directly related to infectious disease. Indeed, the first example relates to the prediction of the deviation in land surface temperature in East Africa [Ton+18], the second example is linked to the prediction of the distance of a robotic arm from the target [ST03], and finally the third example presents a reconstruction of MNIST digits. However, it should also be noted that  $\pi$ VAE’s *sister models* such as aggVAE [Sem+23a] and priorCVAE [Sem+23b] present spatio-temporal applications of inference, respectively, of malaria prevalence in Kenya and HIV prevalence in Zimbabwe. Thus, this chapter will introduce the general architecture of  $\pi$ VAE and the underlying concepts behind it, which are very similar to those behind aggVAE [Sem+23a] and priorCVAE [Sem+23b] and we will also briefly conclude with examples of applications of the latter models, as they were directly related to epidemiology.

This chapter is organised as follows: first, and as a way of introduction to the topic, in Section 4.2, we present a very brief review of some of the major concepts of Bayesian inference that underlie the theory behind the construction of the  $\pi$ VAE. Then in Section 4.3, we gradually detail the architectures and principles of auto-encoders and of VAEs. We then explain how VAEs can be used for Bayesian inference and, in particular, introduce PriorVAE [Sem+22], which is a parent of  $\pi$ VAE. Section 4.4 focusses on  $\pi$ VAE: in Section 4.4.1, we present the two-stage approach on which  $\pi$ VAE is based; its architecture is given in Section 4.4.2. Explanations of its simulation and inference processes are found in Section 4.4.3. We show in Section 4.4.4 that  $\pi$ VAE is indeed a stochastic process. Finally in Section 4.4.5, we briefly review examples of applications of  $\pi$ VAE but focus on applications of its sister models, namely AggVAE [Sem+23a] and PriorCVAE [Sem+23b], which happen to relate more directly to spatiotemporal data sets for infectious diseases.

## 4.2 Rapid overview on Bayesian inference

From now on we assume given a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  with the usual conditions.

### 4.2.1 The Bayesian paradigm

The *Bayesian paradigm* relies on learning from data by updating our prior knowledge about some variables or parameters which we denote here  $\theta$ . Our initial beliefs or current knowledge about  $\theta$  is encoded in a prior distribution  $\mathbb{P}(\theta)$ . Then, given a set of observed data  $x$ , we characterise, thanks to a likelihood function  $\mathbb{P}(x \mid \theta)$ , how likely these observed data should have arisen, given specific values of the parameters (or variables)  $\theta$ . The objective is

to determine a posterior distribution  $\mathbb{P}(\theta | x)$  for the parameters of interest. This posterior represents our updated beliefs about  $\theta$  and is classically derived using Bayes' formula

$$\mathbb{P}(\theta | x) = \frac{\mathbb{P}(x | \theta)\mathbb{P}(\theta)}{\int_{\Theta} \mathbb{P}(x | \theta)\mathbb{P}(\theta)d\theta} = \frac{\mathbb{P}(x | \theta)\mathbb{P}(\theta)}{\mathbb{P}(x)}.$$

The above rule can be summarised in an even simpler form: the posterior is proportional to the product of the prior and the likelihood, in other words,

$$\mathbb{P}(\theta | x) \propto \mathbb{P}(x | \theta)\mathbb{P}(\theta).$$

This equation embodies the essence of the Bayesian approach: on the one hand, the likelihood  $\mathbb{P}(x | \theta)$  can be seen as a function of the parameters  $\theta$  for fixed data  $x$ . The likelihood is the channel by which the information contained in the data gets transmitted to the posterior. On the other hand, the prior expresses current beliefs about the parameters and contains the uncertainty about the parameters inherent before seeing additional data.

A common task in Bayesian machine learning is to use the posterior distribution  $\mathbb{P}(\theta | x)$  to make predictions on unseen data  $x^*$ . An informative probability distribution  $\mathbb{P}(x^* | x)$  of unobserved data can be obtained by incorporating information on the parameter  $\theta$  learnt from observed data. The probability distribution of unobserved data  $\mathbb{P}(x^* | x)$  is known as the posterior predictive distribution and is calculated as follows

$$\mathbb{P}(x^* | x) = \mathbb{E}_{\mathbb{P}(\theta|x)}\mathbb{P}(x^* | \theta) = \int_{\Theta} \mathbb{P}(x^* | \theta)\mathbb{P}(\theta | x)d\theta.$$

The posterior predictive distribution can be seen as an average of the conditional predictions  $\mathbb{P}(x^* | \theta)$  over the posterior distribution of  $\theta$ .

### 4.2.2 Challenges of Bayesian inference

Bayesian inference comes mostly with two major challenges: computing the normalising constant  $\mathbb{P}(x)$  which often involves the computation of a very intractable integral, and directly drawing samples from the posterior. However, although knowing  $\mathbb{P}(x)$  is often not required in practice, knowing the posterior, on the other hand, is necessary, for instance, to compute expectations. In addition, even if and when we know exactly the density of the posterior, it still proves challenging to directly draw samples from it. Indeed, in practice, inference requires sampling the stochastic process to *mix well*, and this makes the problem of posterior inference computationally challenging.

There exist various classical Bayesian inference methods to approximate (posterior) distributions. Among the most common ones, we shall mention Markov Chain Monte Carlo (MCMC), which relies on the construction of a Markov chain whose stationary distribution happens to be the posterior from which we wish to sample parameter values iteratively.

Variational Bayesian inference, the method underlying  $\pi$ VAE, is also an approximation technique that greatly simplifies inference by turning a marginalisation problem into an optimisation problem. It aims at approximating the posterior by optimising a so-called lower bound on the evidence with respect to a certain variational distribution. We provide more details on this technique in our next section.

### 4.2.3 Variational inference approaches and latent variable models

We now turn the focus of this brief introductory review to the large topic of Bayesian variational inference. We cover below the bare details to allow a reader not necessarily acquainted with the latter to smoothly read the rest of the chapter.

In a single sentence, variational inference transforms the problem of posterior inference from an intractable model into a more tractable optimisation problem. It is one of the reasons why variational inference is often proven very efficient in high-dimensional problems. Additionally, variational techniques provide a unified framework not only for model inference but also to learn new tasks. In fact, they rely on optimising certain lower bounds on the model evidence (i.e. on the data) simultaneously with respect to the parameters of the variational approximation and with respect to the deterministic parameters of the model.

Let us first recall a few basics about latent variable models. Indeed latent variables, i.e. variables that are not directly observed but inferred through a mathematical model from other variables that are themselves observed, are inherent to variational methods. For example, in an epidemiological context, the actual number of daily cases, often unobserved due to limited testing capacity, can be considered as a latent variable and inferred from the more limited but reported daily cases, or even from other disease surveillance data. The goal of latent variable methods is most often to model the data generating process using lower-dimensional representations of the data. Let  $\mathcal{X}$  be the input space where the data  $x$  live and let  $\mathcal{Z}$  be a latent space of random variables  $z$  of a lower dimension than  $x$ . We can view  $z$  as a set of points  $z = (z_i)_{i=1,\dots,n}$  associated with the inputs  $x = (x_i)_{i=1,\dots,n}$  whereby the dimension of each latent variable is smaller than the original input. Then a latent variable model (LVM) is a probabilistic generative model where each  $z_i \in \mathcal{Z}$  is assumed to be a useful lower-dimensional representation of its associated data point  $x_i \in \mathcal{X}$  by relying on some global parameters  $\theta$  that are usually assumed to be deterministic. Data and latent variables are most of the time i.i.d. given  $\theta$ . The LVM model is described by

$$\mathbb{P}_\theta(x) = \int_{\mathcal{Z}} \mathbb{P}_\theta(x | z) \mathbb{P}(z) dz.$$

Two major tasks can be tackled in the context of latent variable models: inference, which consists in computing or, most of the time, approximating the posterior distribution over latent variables, and then learning, which consists in estimating the model parameters  $\theta$ . Typically both inference and learning can be achieved in theory through maximising the marginal log-likelihood, i.e.  $\log \mathbb{P}_\theta(x)$ , but as this optimisation process often proves very challenging (e.g. presence of many local optima), we tend to rely in practice on a general approach based on the formulation of an objective called the *Evidence Lower BOund (ELBO)*, which in effect represents the lower bound on the marginal log-likelihood. Maximising the ELBO allows us to learn and infer models simultaneously.

Let us give a more specific formulation of the ELBO. Following the perspective described, for example, in [Bis06, Ch.10] or [Rai22, Section 9.3.1], we adopt a Bayesian view on both the latent variables  $z$  and the global parameters  $\theta$ . Thus, the joint model is given by  $\mathbb{P}(x, z, \theta)$  and, most importantly, we can then treat latent variables  $z$  and parameters  $\theta$  in a similar way. We consider a joint distribution for  $z$  and  $\theta$ , which we denote  $\mathbb{Q}(z, \theta)$ . It is called a *variational distribution*. We then introduce the quantity

$$\text{ELBO}(\mathbb{Q}) = \mathbb{E}_{\mathbb{Q}}(\log \mathbb{P}(x, z, \theta)) - \mathbb{E}_{\mathbb{Q}}(\log \mathbb{Q}(z, \theta)),$$

which, as we will see just below, is a lower-bound on the log-evidence  $\log \mathbb{P}(x)$ . Indeed, recalling the definition of the Kullback-Leibler (KL) divergence  $D_{KL}(\mathbb{P} \parallel \mathbb{Q})$  between two continuous probabilities  $\mathbb{P}$  and  $\mathbb{Q}$  with respective densities  $p$  and  $q$  [KL51],

$$D_{KL}(\mathbb{P} \parallel \mathbb{Q}) = \int_{\mathcal{X}} p(x) \log \frac{p(x)}{q(x)} dx,$$

we re-write the ELBO as follows

$$\begin{aligned} \text{ELBO}(\mathbb{Q}) &= \log \mathbb{P}(x) + \mathbb{E}_{\mathbb{Q}}(\log \mathbb{P}(z, \theta) \mid x) - \mathbb{E}_{\mathbb{Q}}(\log \mathbb{Q}(z, \theta)), \\ &= \log \mathbb{P}(x) - D_{KL}(\mathbb{Q}(z, \theta) \parallel \mathbb{P}(z, \theta \mid x)). \end{aligned}$$

Then, Gibbs inequality [Gib02] states that for any probability distributions  $\mathbb{P}$  and  $\mathbb{Q}$  defined on the same probability space, the KL divergence is always non-negative,

$$D_{KL}(\mathbb{Q} \parallel \mathbb{P}) \geq 0.$$

Therefore  $\text{ELBO}(\mathbb{Q})$  is maximised (and then equal to the log-evidence) when the KL divergence between the variational distribution  $\mathbb{Q}(Z, \theta)$  and the joint conditional distribution  $\mathbb{P}(Z, \theta \mid X)$  is null.

Let us now summarise the essence of the variational inference approach: the joint posterior  $\mathbb{P}(Z, \theta \mid X)$  is the distribution that we wish to compute. But, since it is most of the time (if not always) intractable, we approximate this posterior with a certain family  $\mathcal{Q}$  of tractable variational distributions, denoted  $\mathbb{Q}(z, \theta)$ . Then the learning objective is achieved by minimising the divergence  $D_{KL}(\mathbb{Q}(z, \theta) \parallel \mathbb{P}(z, \theta \mid x))$  over the tractable family of distributions  $\mathcal{Q}$ . This latter objective can also be translated into maximising  $\text{ELBO}(\mathbb{Q})$ , i.e. finding the tightest lower bound on log evidence.

## 4.3 Variational autoencoders

Before presenting the structure of the  $\pi$ VAE model in Section 4.4, we provide a more specific background review of its building blocks in this section. We therefore gradually introduce the autoencoder in Section 4.3.1, the variational autoencoder (VAE) in Section 4.3.2, and finally another model called PriorVAE [Sem+22] in Section 4.3.3, which is also a novel application of VAEs in a Bayesian inference setting and that is based on a two-stage approach closely related to the one adopted to construct  $\pi$ VAE.

### 4.3.1 Autoencoder

An autoencoder [HS06] is a model composed of two component networks:

- The encoder function  $e : \mathcal{X} \rightarrow \mathcal{Z}$  encodes data or inputs  $x$  from an input space  $\mathcal{X}$  into a latent space  $\mathcal{Z}$  of a lower dimension than  $\mathcal{X}$ . The encoder function can be mathematically expressed as  $z = e(x; \gamma)$ , where  $\gamma$  are learnable parameters, and typically the function  $e$  consists of multiple layers of neurones. The encoder captures the essential features of the input and is generally designed to compress the input data into a lower-dimensional latent representation.
- The decoder function  $d : \mathcal{Z} \rightarrow \mathcal{X}$  reconstructs the original input  $x$  into  $\hat{x} \in \mathcal{X}$  from the latent representation  $z \in \mathcal{Z}$ . The decoder decompresses the latent variables with

the objective of reconstructing an output as similar to the input as possible. It can be expressed as a function  $\hat{x} = d(z; \psi)$  parameterised by learnable parameters  $\psi$ , and again  $d$  generally comprises multiple layers.

The objective of training an autoencoder is to minimise reconstruction loss, which measures how well the output  $\hat{x}$  approximates the original input  $x$ . The parameters of  $e$  and  $d$  are learnt through the minimisation process. A common choice for the loss function is the Mean Squared Error (MSE) given by

$$L(x, \hat{x}) = \frac{1}{N} \sum_{i=1}^N (x_i - \hat{x}_i)^2,$$

where  $N$  is the number of input features. The autoencoder training process can be summarised into three distinct phases: feeding the input data  $x$  through the encoder to obtain a latent representation  $z$ ; passing  $z$  through the decoder to produce the reconstructed output  $\hat{x}$ , and calculating the reconstruction loss to update the parameters  $\gamma$  and  $\psi$  using gradient descent or other optimisation algorithms.

Although autoencoders leverage deep learning architectures to capture complex patterns and compress information effectively, they are limited in practice because they are not generative models.

### 4.3.2 Variational autoencoder

A variational autoencoder (VAE), on the other hand, extends an autoencoder into a generative model [KW14] by incorporating probabilistic elements into the encoding process. Thus, unlike standard autoencoders that encode input data in *fixed points* in the latent space, VAEs learn to encode input as *distributions* (typically Gaussian). This means that, for each input or each data point, instead of producing fixed latent points, the VAE produces a set of parameters (e.g. the mean and variance of a Gaussian distribution) that define a probability distribution in the latent space  $\mathcal{Z}$ . Once these variational parameters are produced, a variational approximation to the posterior can then be estimated.

Before we provide slightly more technical details on the architecture of VAEs, we note that, in a variety of applications, VAEs have already done a superb job of reconstructing training datasets and enabling the generation of new data: samples from the latent space are decoded to generate synthetic data [KW19]. In the simplest case, we are given input  $x \in \mathcal{X}$  such as a flattened image or a discrete time series. A standard VAE has three components, two of them being the main components of a regular autoencoder, but a VAE additionally incorporates probabilistic modelling. Note that the encoder  $e(x, \gamma)$  and the decoder  $d(z, \psi)$  below are fully connected neural networks (although they could include convolution or recurrent layers).

- **Encoder (Inference Model):** an encoder network  $e(x; \gamma)$  encodes inputs  $x \in \mathcal{X}$  using learnable parameters  $\gamma$ . Contrary to the encoder of an autoencoder that produces a fixed latent representation, the encoder of a VAE compresses the input data  $x$  into parameters of a distribution in the latent space, specifically the mean and standard deviation  $z_\mu$  and  $z_{sd}$  of a multivariate Gaussian distribution. The output of the encoder network are thus vectors of mean and standard deviation parameters  $z_\mu$  and  $z_{sd}$ , and these vectors

are used to define the random variable  $z$  in the latent space as follows

$$[z_\mu, z_{sd}]^\top = e(x; \gamma).$$

Note that the encoder outputs in practice  $\log(z_{sd})$  to ensure positive semi-definiteness.

- The latent subspace is then defined by the random variables  $z$  sampled from  $\mathcal{N}(z_\mu, z_{sd}^2 \mathbb{I})$ , i.e.,

$$z \sim \mathcal{N}(z_\mu, z_{sd}^2 \mathbb{I}).$$

- Decoder (Generative Model): The decoder reconstructs the original input  $x$  into  $\hat{x}$  from a latent variable  $z$  sampled from the encoded distribution by producing  $\hat{x} = d(z, \psi)$ .

To train a VAE, a variational approximation is used to estimate the posterior distribution

$$\mathbb{P}(z \mid x, \gamma, \psi) \propto \mathbb{P}(x \mid z, \gamma, \psi) \mathbb{P}(z).$$

Following Kingma [KB14], the optimal parameters for the encoder and decoder are found by maximising the evidence lower bound (ELBO)

$$\arg \max_{\gamma, \psi} \mathbb{E}_z (\log \mathbb{P}(x \mid z, \gamma, \psi) - D_{KL}(z \parallel \mathcal{N}(0, \mathbb{I}))).$$

The first term in the above equation is the likelihood quantifying how well  $\hat{x}$  matches  $x$ . In practice, we can simply directly adopt the mean squared error (MSE) loss, referred to as *the reconstruction loss*, without taking a probabilistic perspective. The second term is a Kullback-Leibler divergence to ensure that the latent variable  $z$  is as similar as possible to the prior distribution, a standard normal. Again, this second term can be specified directly without the evidence of the lower-bound derivation. Generally, we view the Kullback-Leibler divergence as a measure of how much the learnt distribution  $\mathbb{P}(z \mid x, \gamma, \psi)$  diverges from the prior distribution  $\mathbb{P}(z)$  (assumed to be  $\mathcal{N}(0, I)$ ). Here we take a slightly different view, in which the KL divergence is seen as a regularisation penalty to ensure that the latent parameters are approximately uncorrelated by penalising how far they deviate from  $\mathcal{N}(0, \mathbb{I})$ .

Once training is complete, we fix the values of the parameters  $\psi$ , and use the decoder as a generative model. To simplify subsequent notations, we may refer to a fully trained decoder as  $d(z)$ . Generating a new sample from a trained VAE is simple: first, we draw a random variable  $z \sim \mathcal{N}(0, \mathbb{I})$  and then apply the decoder, which is a deterministic transformation to obtain  $d(z)$ . We immediately see that  $d(z)$  is itself a random variable. In the next section, we will use this generative model as a prior in a Bayesian framework by linking it to a likelihood to obtain a posterior.

### 4.3.3 VAEs in a Bayesian inference setting

VAEs have been commonly used in the literature to create or learn a generative model of observed data [KW14], such as images. In particular, Semenova et al. [Sem+22] introduced a novel application of VAEs in a Bayesian inference setting, using a two-stage approach that is closely related to the approach adopted in the construction of  $\pi$ VAE. In summary, Semenova et al. [Sem+22] propose approximating a class of Gaussian process priors in a spatial setting through prior sampling and subsequent fitting of a VAE. Thus, in a first stage, a VAE is trained to encode and decode a large dataset of vectors consisting of samples drawn from a

specified Gaussian process prior  $\mathbb{P}(\theta)$  over random vectors. In a second stage, the original prior is replaced with the approximate prior such that  $\theta := d(z)$  where  $z \sim \mathcal{N}(0, \mathbb{I})$ . The resultant decoder allows for very efficient spatial inference, due to the low-dimensional and independent distributed latent Gaussian space representation of the VAE.

To see how all this works in a Bayesian inference setting, consider a likelihood  $\mathbb{P}(x \mid \theta)$  linking the parameter  $\theta$  to the data  $x$ . Bayes' rule gives the unnormalised posterior

$$\mathbb{P}(\theta \mid x) \propto \mathbb{P}(x \mid \theta)\mathbb{P}(\theta).$$

The trained decoder, now denoted  $d(z)$  since the learnt parameters  $\psi$  can be fixed, serves as a drop-in replacement for the original prior class in a Bayesian setting

$$\mathbb{P}(z \mid x, d) \propto \mathbb{P}(x \mid d(z))\mathbb{P}(z). \quad (4.1)$$

The implementation within a probabilistic programming language is very straightforward: a standard normal prior and a deterministic function (the decoder) are all that is needed.

It is useful to contrast the inference task described by equation (4.1) which follows from Bayesian inference using a VAE with the inference task underlying, for instance, a Bayesian neural network (BNN) [Nea96] or a Gaussian process in primal form [RR08]. In a BNN with parameters  $\omega$  and hyperparameters  $\lambda$ , the unnormalised posterior would be given by

$$\mathbb{P}(\omega, \lambda \mid x) \propto \mathbb{P}(x \mid \omega, \lambda)\mathbb{P}(\omega \mid \lambda)\mathbb{P}(\lambda). \quad (4.2)$$

The key difference between equation (4.2) and equation (4.1) is the term  $\mathbb{P}(\omega \mid \lambda)$ . First, the dimension of  $\omega$  is typically huge, sometimes in the millions, whereas in the equation (4.1) the latent dimension of  $z$  is typically small ( $< 50$ ). Furthermore, observe that  $\mathbb{P}(\omega \mid \lambda)$  is conditional on  $\lambda$ , while  $\mathbb{P}(z)$  is uncorrelated and unconditioned.

Full batch MCMC training is typically prohibitive for BNNs due to large datasets and the high dimensionality of  $\omega$ . Approximate Bayesian inference algorithms tend to poorly capture the complex posterior [Yao+19; Yao+18]. Additionally,  $\omega$  tends to be highly correlated, making efficient MCMC nearly impossible. Finally, as dimensions and depth increase, the posterior distribution suffers from complex multimodality and concentration in a narrow typical set [Bet+17]. In contrast, off-the-shelf MCMC methods are very effective in manipulating equation (4.1) because the prior space that needs to be explored is as simple as it could be: a standard normal distribution, while the complexity of the model lives within the deterministic (and differentiable) decoder.

Another similar approach is the neural process [Gar+18]. Like Gaussian processes, neural processes encode a distribution over functions. A neural process is a stochastic process that fits to the data while simultaneously learning the function class (estimating the prior and posterior simultaneously). In functional form, it is similar to a VAE except for two elements: the KL divergence penalty term of a NP is equal to the divergence from training and testing latent variables (as opposed to a standard normal in a VAE), and an aggregation step is included in neural processes to allow for exchangeability and consistency, which are required for it to be a valid stochastic process. However, a key theoretical limitation of neural processes is that their accuracy is hindered by variational inference, which provides no guarantee that the estimated posterior distribution is correct. In practice, neural processes often learn functions that substantially underfit and, therefore, tend to predict both point estimates and uncertainty inaccurately [Kim+19].  $\pi$ VAE does not suffer from this pitfall as it benefits from the MCMC

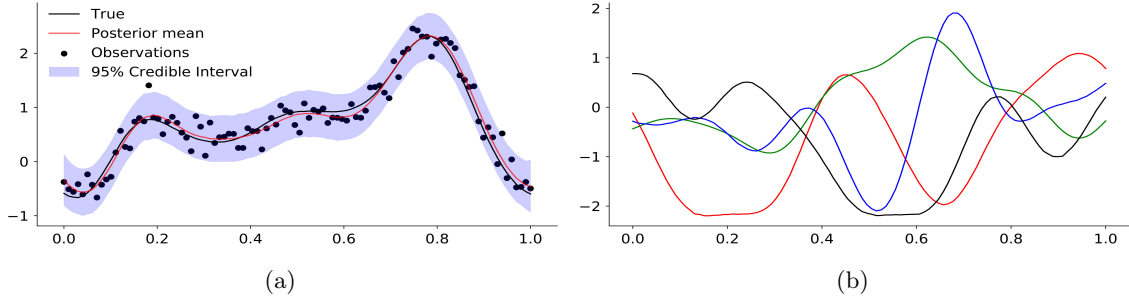


Figure 4.1: Learning functions with VAE: (a) Prior samples from a VAE trained on Gaussian process samples (b) we fit our VAE model to data drawn from a Gaussian process (blue) plus noise (black points). The posterior mean of our model is in red with the 95% epistemic credible intervals shown in purple.

approach.

*Examples.* In a challenging spatial statistics setting, Semenova et al. [Sem+22] used a VAE-based approach with Bayesian inference and achieved MCMC effective sample sizes *exceeding* actual sample sizes, due to the incredible efficiency of the MCMC sampler. Another example taken from [Mis+22] using VAEs to perform inference is shown in [Mis+22, Fig.1] where a VAE with latent dimensionality 10 is trained on samples drawn from a zero-mean Gaussian process with an RBF kernel ( $K(\delta) = e^{-\delta^2/8^2}$ ) observed on the grid  $0, 0.01, 0.02, \dots, 1.0$ . In [Mis+22, Fig.1], it can be observed that the true function is closely recovered and the data noise parameter are correctly estimated. The MCMC samples showed virtually no autocorrelation, and all diagnostic checks were excellent (see appendix in [Mis+22]). Solving the equivalent problem using a Gaussian process prior would not only be considerably more expensive ( $\mathcal{O}(n^{2.7})$  - see [RW06]) but correlations in the parameter space would complicate MCMC sampling and necessitate very long chains to achieve even modest effective sample sizes.

These examples demonstrate the promise that VAEs hold in improving Bayesian inference by encoding function classes in a two-stage process. However, as described above, a VAE provides a novel prior over random vectors, but inference and prediction are not possible at new input locations because a VAE is not a stochastic process. In the following, we thus take the next step by introducing  $\pi$ VAE, a new stochastic process capable of approximating useful and widely used priors over function classes, such as Gaussian processes.

## 4.4 Encoding stochastic processes with $\pi$ VAE

### 4.4.1 A two-step approach

Similarly to [Sem+22] we propose a two-step approach to proceed to VAE-based Bayesian inference: we first learn low-dimensional representations of samples from a given function class (e.g. sample paths from a Gaussian process prior). Then we use the resulting low-dimensional representation and the decoder to perform Bayesian inference.

One key benefit of the two-step approach of VAE-based Bayesian inference is that we decouple prior from inference to encode arbitrarily complex prior function classes, without needing to calculate any data likelihoods. A second key benefit is that when inference is performed, our sampler operates in a low-dimensional uncorrelated latent space that greatly helps efficiency and computation, as demonstrated in the spatial statistics setting

in PriorVAE [Sem+22]. One limitation of this approach (and of PriorVAE) is that we are restricted to encoding finite-dimensional priors, because VAEs are not stochastic processes. To overcome the limitation of PriorVAE, we take as inspiration the Karhunen-Loève decomposition of a stochastic process as a random linear combination of basis functions and introduce a new VAE called *the prior-encoding-VAE* ( $\pi$ VAE).  $\pi$ VAE is a valid stochastic process by construction, it is capable of learning a set of basis functions, and it incorporates a VAE, enabling simulation and highly effective fully Bayesian inference.

We can summarise  $\pi$ VAE’s two-step approach as follows: first, we encode the prior using our novel architecture (described in Section 4.4.2 below); second, we use the learnt basis and decoder network, a new stochastic process in its own right, as a prior, combining it with a likelihood in a fully Bayesian modelling framework, and use MCMC to fit our model and infer the posterior.

We believe our framework’s novel decoupling into two stages is critically important for many complex scenarios because we do not need to compromise in terms of either the expressiveness of deep learning or accurately characterising the posterior fully using Bayesian inference. We thus avoid some of the drawbacks of other Bayesian deep learning approaches which rely solely on variational inference, and the drawbacks of standard MCMC methods for stochastic processes which are inefficient and suffer from poor convergence.

#### 4.4.2 Architecture of $\pi$ VAE

To create a model with the ability to perform inference on a wide range of problems, we have to ensure that it is a valid stochastic process. Previous attempts in deep learning in this direction have been inspired by the Kolmogorov Extension Theorem and have focused on extending from a finite-dimensional distribution to a stochastic process. Specifically, [Gar+18] introduced an aggregation step (typically an average) to create a global distribution of fixed dimension. However, as noted in [Kim+19], this can lead to underfitting.

We take a different approach with  $\pi$ VAE, inspired by the Karhunen-Loève expansion [Kar47; Loe48]. Recall that a centred stochastic process  $f(s)$  can be written as an infinite sum

$$f(s) = \sum_{j=1}^{\infty} \beta_j \phi_j(s), \quad (4.3)$$

for pairwise uncorrelated random variables  $\beta_j$  and continuous real-valued functions forming an orthonormal basis  $\phi_j(s)$ . The random variables  $\beta_j$  provide a linear combination of a fixed set of basis functions  $\phi_j$ . This perspective has a long history in neural networks (e.g. radial basis function networks).

We now consider a deep learning trainable parameterisation of the Karhunen-Loève expansion (4.3) as inspiration. The objective is to learn deterministic basis functions  $\phi_j(\cdot)$  while allowing the coefficients  $\beta_j$  to be random. To do so, we let  $s \mapsto \Phi(s)$  be a feature mapping with weights  $w$ , i.e. a feed-forward neural network architecture over the input space, representing the basis functions. Let  $\beta$  be a vector of weights on the basis functions such that  $f(s) = \beta^\top \Phi(s)$ . We then use a VAE architecture to encode and decode the random vectors  $\beta$ , which means that on the one hand, we maintain the random variable perspective, but at the same time, we learn a flexible low-dimensional nonlinear generative model.

This model needs to be specified and trained. As with the VAE in the previous section,

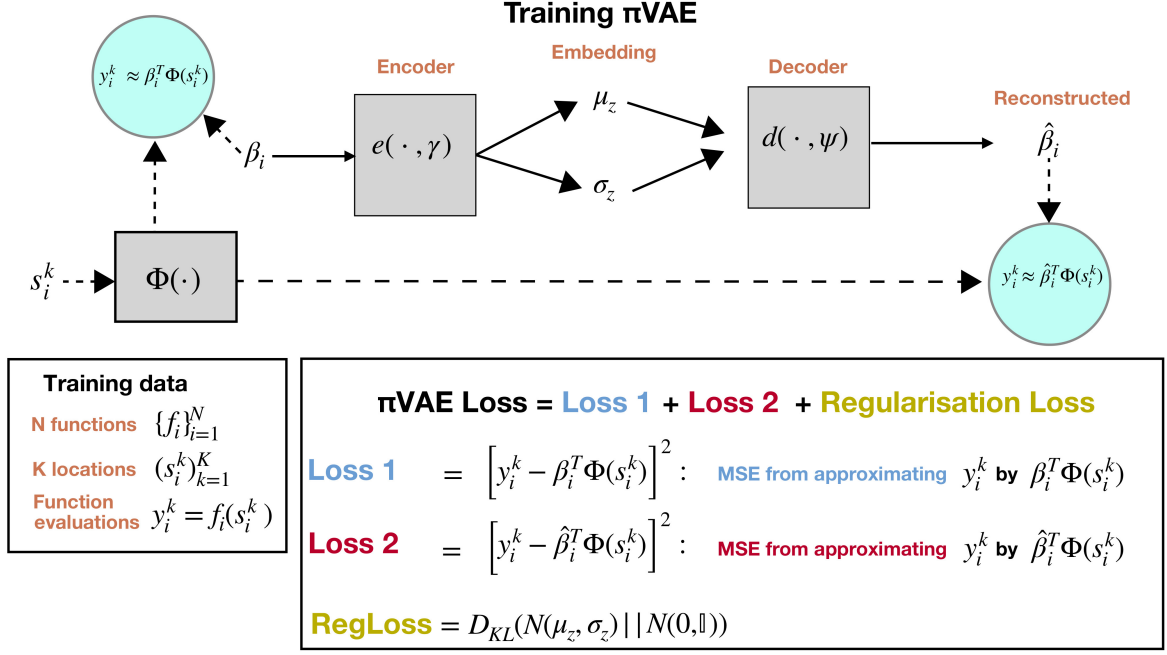


Figure 4.2: Schematic description of the end-to-end training procedure for  $\pi$ VAE, including the reconstruction loss. Dashed arrows contribute to the loss, blue circles are reconstructions, and grey boxes are functions.

$\pi$ VAE is trained on draws from a prior. Our goal is to encode a stochastic process prior  $\Pi$ . To achieve this, we consider  $N$  function realisations which we denote  $(f_i(s))_{i=1}^N$ . Each function realisation  $f_i(s)$  can be viewed as an infinite-dimensional object, as it is a function defined for any  $s \in \mathbb{R}$ . For each realisation  $i = 1, \dots, N$ , we further assume that we are given a finite set of  $K_i$  observation locations. However, here, for simplicity of implementation, we take  $K_i = K$ , in other words, we assume that the number  $K$  of evaluations for each function  $f_i$  is constant across each draw  $i$ . We denote the observed values as  $y_i^k := f_i(s_i^k)$ , for  $k = 1, \dots, K$ . The training data set therefore consists of  $N$  sets of  $K$  observation locations and function values

$$\mathcal{D}_{\text{train}} := \{(s_i^1, y_i^1) \dots, (s_i^K, y_i^K)\}_{i=1}^N.$$

Note that the set  $(s_i^k)_{k=1}^K$  of  $K$  observation locations varies in each of the  $N$  realisations.

We now turn more precisely to the architecture of  $\pi$ VAE (Fig. 4.2). The feature mapping  $s_i^k \mapsto \Phi(s_i^k)$  (for  $k = 1, \dots, K$ ) is shared and learnt across all the  $i = 1, \dots, N$  function draws. It consists of a feedforward neural network and is parameterised by a set of global parameters  $w$  that must be learnt. Then a particular random function realization  $f_i(\cdot)$  is represented by a random vector  $\beta_i$  applied to the feature mapping  $\Phi(\cdot)$  such that at each location  $k = 1, \dots, K$ ,

$$f_i(s_i^k) = \beta_i^T \Phi(s_i^k).$$

The random vector  $\beta_i$  is learnt with a *standard* VAE architecture. However, we note the following non-standard setup: on one hand,  $\beta_i$  is a learnable parameter of our model (it is learnt by a standard VAE), but it is also the input to the encoder of the VAE. The decoder

attempts to reconstruct  $\beta_i$  with an output  $\hat{\beta}_i$ . We denote the encoder and decoder as

$$\begin{aligned} [z_\mu, z_{sd}]^\top &= e(\beta; \gamma), \\ z &\sim \mathcal{N}(z_\mu, z_{sd}^2 \mathbb{I}), \\ \hat{\beta} &= d(z; \psi). \end{aligned}$$

We can now express the loss, which combines the two parts (namely encoder and decoder) of the network, summing across all observations (i.e  $k = 1, \dots, K$  and  $i = 1, \dots, N$ ). Rather than deriving an evidence lower bound, we proceed directly to specify a loss function in three parts. In the first, we use MSE to check the fit of the random vectors  $\beta_i$  and the feature mapping  $\Phi(\cdot)$  to the data  $(y_i^k; i = 1, \dots, N; k = 1, \dots, K)$ ,

$$\text{Loss 1} := \frac{1}{NK} \sum_{i=1}^N \sum_{k=1}^K (y_i^k - \beta_i^\top \Phi(s_i^k))^2.$$

This first loss term corresponds to the approximation of the function by their Karhunen-Loève expansion. In the second part of the loss, we use MSE to check the fit of the reconstructed vectors  $\hat{\beta}_i$  and  $\Phi(\cdot)$  to the input data

$$\text{Loss 2} := \frac{1}{NK} \sum_{i=1}^N \sum_{k=1}^K (y_i^k - \hat{\beta}_i^\top \Phi(s_i^k))^2.$$

This is partly the classical loss term of a VAE to which we have to add the standard variational loss

$$\text{RegLoss} := D_{KL}(z \parallel \mathcal{N}(0, \mathbb{I})).$$

*Remark.* We did not consider reconstruction loss  $\|\beta_i - \hat{\beta}_i\|^2$  because in practice we observed that it did not improve training.

Let us now provide some additional intuition to the overall architecture. The feature map  $\Phi(s)$  transforms each observed location into a fixed feature space that is shared for all locations  $k = 1, \dots, K$  across all functions  $i = 1, \dots, N$ . In practice,  $\Phi(s)$  could be an explicit feature representation for a Reproducing Kernel Hilbert Space (RKHS) (e.g. a radial basis function (RBF) network or a random Fourier feature basis [RR08]), a neural network of arbitrary construction or, as used in the examples in [Mis+22], a combination of both. Following the transformation of the observed locations  $s_i^k$  by the feature map (i.e.  $s_i^k$  is replaced by  $\Phi(s_i^k)$ ), a linear basis  $\beta$  (which we obtain from a trained nonlinear decoder network) is used to predict function evaluations at an arbitrary location. The intuition behind these two transformations is to learn the association between locations and observations while allowing for randomness:  $\Phi$  provides the spatial correlation structure and  $\beta$  the randomness.

*Remark.* Note that explicit choices can lead to existing stochastic processes. For instance, we can obtain a Gaussian process with kernel  $k(\cdot, \cdot)$  using a single-layer linear VAE for  $\beta$  (meaning the random vectors  $\beta$  are simply standard normally distributed) and also setting  $\Phi(s) = L^\top s$  where  $L$  is the Cholesky decomposition of the Gram matrix  $K$  such that  $K_{ij} = k(s_i, s_j)$ .

In contrast to a standard VAE encoder that takes as input the data to be encoded,  $\pi$ VAE first transforms the input data (locations) into a higher-dimensional feature space through  $\Phi$ , and then connects this feature space to the outputs,  $y$ , through a linear mapping,  $\beta$ . The  $\pi$ VAE decoder takes the output from the encoder and attempts to recreate  $\beta$  from

a lower-dimensional probabilistic embedding. This recreation,  $\hat{\beta}$ , is then used as a linear mapping with the *same* feature  $\Phi$  to obtain a reconstruction of the outputs  $y$ . It is crucial to note that a single global  $\beta$  vector is *not* learnt. Instead, a specific  $\beta_i$  is learnt for each function realisation  $i = 1, \dots, N$ .

In terms of the number of parameters, we need to learn the weights  $w$  of the feedforward network that learn the feature mapping;  $\gamma$  and  $\psi$ , the learnable parameters of the encoder and decoder, and also the coefficients  $\beta_1, \dots, \beta_N$  that are learnt by the VAE. Although this may seem like a huge computational task,  $K$  is typically quite small ( $< 200$ ) and learning can be relatively quick (dominated by matrix multiplication of hidden layers). The appendix in [Mis+22] provides an algorithm that presents the step-by-step training process of  $\pi$ VAE.

#### 4.4.3 Simulation and inference with $\pi$ VAE

Given a trained embedding  $\Phi(\cdot)$  and a trained decoder  $d(z)$ , we can use  $\pi$ VAE as a generative model to simulate sample paths  $f$  as follows. First a single function  $f$  is obtained by drawing

$$z \sim \mathcal{N}(0, \mathbb{I}),$$

and then defining

$$f(s) := d(z)^\top \Phi(s).$$

For a fixed  $z$ ,  $f(s)$  is a deterministic function - a sample path from  $\pi$ VAE defined for all  $s$ . Varying  $z$  produces different sample paths. Computationally,  $f$  can be efficiently evaluated at any arbitrary location  $s$  using matrix algebra:  $f(s) = d(z)^\top \Phi(s)$ . We remark that the stochastic process perspective is readily apparent: for a random variable  $z$ ,  $d(z)^\top \Phi(s)$  is a random variable defined on the same probability space for all  $s$ .

Besides the simulation part,  $\pi$ VAE can also be used for inference on new pairs (location, observation) of data  $(s_j, y_j)$ , where the unnormalised posterior distribution is

$$\mathbb{P}(z \mid d, y_j, s_j, \Phi) \propto \mathbb{P}(y_j \mid d, s_j, z, \Phi) \mathbb{P}(z). \quad (4.4)$$

The likelihood is given by  $\mathbb{P}(y_j \mid d, s_j, z, \Phi)$  and the prior by  $\mathbb{P}(z)$ . MCMC can be used to efficiently obtain samples from the posterior distribution over  $z$  using equation (4.4). An implementation in probabilistic programming languages such as Stan [Car+17] is also very straightforward; see [Mis+22].

Additionally, the posterior predictive distribution of  $y_j$  at a location  $s_j$  is given by

$$\mathbb{P}(y_j \mid d, s_j, \Phi) = \int \mathbb{P}(y_j \mid d, s_j, \Phi, z) \mathbb{P}(z \mid d, y_j, s_j, \Phi) dz. \quad (4.5)$$

While equations (4.4)-(4.5) are written for a single location  $s_j$ , we can extend them to any arbitrary collection of locations without loss of generality, a necessary condition for  $\pi$ VAE to be a valid stochastic process. Further, the distinguishing point between equation (4.1) related to BNN and equations (4.4)-(4.5) is the conditioning on input locations and on the map  $\Phi$ . *It is  $\Phi$  that ensures  $\pi$ VAE is a valid stochastic process.* We formally prove this in the next short section.

*Remark.* The step-by-step processes for simulation and inference with  $\pi$ VAE are presented in

the appendix of [Mis+22].

#### 4.4.4 $\pi$ VAE is a stochastic process

**Lemma 4.4.1.**  *$\pi$ VAE is a stochastic process.*

First, recall that mathematically a stochastic process is defined as a collection  $\{f(s) : s \in S\}$ , where  $f(s)$  for each location  $s \in S$  is a random variable in a common probability space  $(\Omega, \mathcal{F}, P)$ ; see, for example, [Pav14, Definition 1.1]. This technical requirement is necessary to ensure that, for any location  $s_1, \dots, s_n \in S$ , the random variables  $f(s_1), \dots, f(s_n)$  have a well-defined joint distribution. Subsequently, it also ensures consistency. Namely, writing  $f_i := f(s_i)$  and integrating  $f_n$  out, we get

$$\mathbb{P}(f_1, \dots, f_{n-1}) = \int_{f_n} \mathbb{P}(f_1, \dots, f_n) df_n.$$

*Proof.* For  $\pi$ VAE, we have  $f(\cdot) := d(z)^\top \Phi(\cdot)$ , where  $z$  is a multivariate Gaussian random variable, hence defined in some probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ . Since  $d$  and  $\Phi$  are deterministic (measurable) functions, it follows that  $f(s_i) := d(z)^\top \Phi(s_i)$  for any  $i = 1, \dots, n$ , is a random variable on  $(\Omega, \mathcal{F}, P)$ , whereby  $\{f(s) : s \in S\}$  is finitely exchangeable and Kolmogorov consistent, and therefore a stochastic process.  $\square$

We remark here that  $\pi$ VAE is a new stochastic process. If  $\pi$ VAE is trained on samples from a zero mean Gaussian process with a squared exponential covariance function, and if we choose  $\Phi$  to have the same covariance function, and the decoder  $d$  to be linear, then  $\pi$ VAE will be a Gaussian process. But for a non-positive definite  $\Phi$  and/or a nonlinear decoder function  $d$ , even if  $\pi$ VAE is trained on samples from a Gaussian process, it will not truly be a Gaussian process, but some other stochastic process which approximates a Gaussian process. We do not know the theoretical conditions under which  $\pi$ VAE will perform better or worse than existing classes of stochastic processes; its general construction means that theoretical results will be challenging to prove in full generality. However, [Mis+22] showed that, in practice,  $\pi$ VAE performs very well.

#### 4.4.5 Applications to spatial Bayesian inference

Mishra et.al. [Mis+22] apply  $\pi$ VAE to three real-world data sets, namely, prediction of deviation in land surface temperature in East Africa [Ton+18], prediction of the distance of a robotic arm from the target [ST03], and reconstruction of the MNIST digits.

The first example mentioned above shows that  $\pi$ VAE is able to capture small-scale features and in particular produces a far better reconstruction than both the full rank and the low rank Gaussian processes. A possible reason for this outperformance could be that the extra hidden layers in the feature map  $\Phi$  create a much richer structure that could capture elements of nonstationarity [Ton+18]. It is also possible that the ability to use state-of-the-art MCMC and estimate a reliable posterior expectation might create resilience to overfitting [Mis+22]. The second example compares  $\pi$ VAE on the Kin40K [ST03] data set with state-of-the-art full and approximate Gaussian process with results extracted from [Wan+19]. Even though the full rank Gaussian process results reported in [Wan+19] are better than those from  $\pi$ VAE, the latter model is still competitive and in particular far better than the approximate Gaussian

process methods. One advantage of  $\pi$ VAE is that it performs full Bayesian inference, whereas exact Gaussian process is estimated by maximising the log marginal likelihood in closed form. Finally, the third example consists of applying  $\pi$ VAE to the task of reconstructing MNIST digits using a subset of pixels from each image. Image completion is seen here as a 2D regression task. This task is to predict the intensity of the pixels given the location of the pixels.  $\pi$ VAE is compared to neural processes (see [Gar+18]) trained on full MNIST digits. The performance of  $\pi$ VAE increases with increase in pixel locations available during prediction: with only 10% pixels,  $\pi$ VAE is still able to learn a decent approximation of the image. Note that results from neural processes are more competitive in this case, as they were specifically trained on full MNIST, whereas  $\pi$ VAE, trained on a more general prior class of 2D Gaussian processes, still provides decent approximations to the images.

We now steer away from  $\pi$ VAE itself to briefly mention two of its sister models, namely the AggVAE [Sem+23a] and the PriorCVAE [Sem+23b]. Those models are slight variants of  $\pi$ VAE and were applied directly to real-world data sets related to epidemiology and infectious diseases. Indeed AggVAE, introduced shortly after  $\pi$ VAE, is applied in [Sem+23a] to infer spatial prevalence of Malaria in Kenya, whereas PriorCVAE [Sem+23b] provides an example of inference of HIV prevalence in Zimbabwe.

Similarly to  $\pi$ VAE, AggVAE was built upon PriorVAE [Sem+22] which, as previously mentioned, can encode spatial priors over small areas of VAEs. AggVAE jointly encodes Gaussian process aggregates using PriorVAE and then proves very useful in a context of disease mapping over fine resolution scales. More particularly, AggVAE [Sem+23a] was used to encode aggregated latent Gaussian process evaluations over administrative units (district, province) in the context of mapping Malaria prevalence in Kenya to overcome the limitation of more commonly used adjacency-based approaches. Malaria is one of the main causes of mortality in sub-Saharan Africa and district-level disease mapping is one of the major surveillance techniques. But administrative boundaries changed in Kenya in 2010, making the task of disease mapping more challenging. Similarly to PriorVAE, a regular grid was chosen to cover the domain of interest, and Gaussian process prior realisations were drawn over an artificial grid. Then, as a next step, Gaussian process evaluations to the district level were aggregated by viewing each district as a polygon, using both the old and new district boundaries. The change of support problem was then handled jointly to encode the old and new Gaussian process aggregates. Finally, results from aggregated Gaussian process priors and AggVAE were compared: the model using aggregated Gaussian process ran 10,000 times longer [Sem+23a] and was barely converging, while AggVAE was much faster and presented better  $\hat{R}$  and average ESS (effective sample size) statistics (see Table 1 [Sem+23a]).

PriorCVAE [Sem+23b] is also another VAE-based model similar to PriorVAE and  $\pi$ VAE. One of the objectives of introducing PriorCVAE was to overcome a known limitation of traditional VAEs. Indeed, although VAEs are learnt generators and can be used as drop-in replacements for priors in MCMC inference, they tend, on the other hand, to lose information about the model hyperparameters. PriorCVAE introduces a conditional VAE architecture which allows us to explicitly encode hyperparameters. Additionally, conditioning the VAE upon the hyperparameters of the encoded stochastic processes allows us to estimate those hyperparameters at the inference stage. Relying on the conditional VAE framework shifts learning emphasis towards priors. The model was applied in [Sem+23b] to real-life spatial data of HIV prevalence in Zimbabwe and was also used in a temporal context to encode a solution of the SIR model. In the former application, PriorCVAE performance was compared with the INLA method [LR15] (R-INLA is a popular R package for Bayesian spatial modelling). A direct comparison was not easy as R-INLA's interface is not flexible enough. However,

it was shown that R-INLA’s posteriors were not sensitive to the choice of priors, whereas PriorCVAE’s posteriors were.

## 4.5 Discussion and Conclusion

This chapter, based primarily on [Mis+22], proposes a novel VAE formulation of a stochastic process, with the ability to learn function classes and properties of functions.  $\pi$ VAE is comparable to PriorVAE [Sem+22] in the sense that it is a VAE-based model that can encode continuous stochastic processes and their finite realisations. They both are able to use a trained decoder to approximate high-dimensional Gaussian processes for Bayesian inference using MCMC techniques. The main difference between  $\pi$ VAE and PriorVAE lies in the type of priors they can each encode (low-dimensional embedding of function classes for  $\pi$ VAE versus fixed-dimensional Gaussian process realisations for PriorVAE). If PriorVAE needs to be trained on a predefined spatial grid,  $\pi$ VAE on the other hand, can handle off-grid locations. However,  $\pi$ VAE is heavier to train than PriorVAE as it requires to learn a feature map. We have also mentioned two relative models to  $\pi$ VAE, AggVAE and PriorCVAE, which were recently shown to perform well in the task of mapping infectious diseases.

Typically  $\pi$ VAEs have a small (5-50) uncorrelated latent dimension of parameters, so Bayesian inference with MCMC is straightforward and highly effective in successfully exploring the posterior distribution. On this point, it should be observed that an accurate estimate of uncertainty is essential in many areas such as medical decision making.

Let us now also add a few words about the complexity of the  $\pi$ VAE. Once a  $\pi$ VAE is trained and defined, the complexity of the decoder scales linearly to the size of the largest hidden layer. Additionally, because the latent variables are penalised via the KL term from deviating from a standard normal distribution, the latent space is approximately uncorrelated, leading to high effective sample sizes in MCMC sampling.

There are costs to using  $\pi$ VAE, especially the large upfront cost in training. For complex priors, training could take days or weeks and will invariably require the heuristics and parameter searches inherent in applied deep learning to achieve a good performance. However, once trained, a  $\pi$ VAE network is applicable on a wide range of problems, with the Bayesian MCMC inference step taking seconds or minutes.

In essence,  $\pi$ VAE combines the power of deep learning to create high-capacity function classes, while ensuring tractable inference using fully Bayesian MCMC approaches. A potential exciting use of  $\pi$ VAE is to incorporate domain knowledge about the problem, such as monotonicity or complicated dynamics, which can be encoded directly into the prior [CDS18] on which  $\pi$ VAE is trained. The log-Gaussian Cox process example in [Mis+22] shows that not just functions can be modelled, but also properties of functions such as integrals. The most surprising result in [Mis+22] is perhaps the performance of  $\pi$ VAE on spatial interpolation. Despite being trained on samples from a Gaussian process,  $\pi$ VAE substantially outperforms a full-rank Gaussian process. It is conjectured that this outperformance may be due to the more complex structure of the feature representation  $\Phi$  and due to a resilience to over-fitting. Overall, taken together, this work is relevant for advance in the field of Bayesian deep learning. It provides a practical framework combining, on one hand, the expressive capability of deep neural networks to encode stochastic processes, with, on the other hand, the effectiveness of fully Bayesian and highly efficient gradient-based MCMC inference to fit to data while fully characterising uncertainty.

# CHAPTER 5

## Conclusion

We briefly conclude here by summarising the main highlights of this thesis.

In Chapter 1 we introduced a time-varying generalised branching process, a.k.a. TV-CMJ branching process. This branching process is a variant of the classical CMJ branching process. Its distinguishing feature, compared to the classical generalised branching process, relies on a calendar-time dependence of its underlying probability measure. Chapter 1 provides a measure-theoretic construction of the TV-CMJ branching process and gives a set of sufficient regularity conditions that guarantees that a TV-CMJ branching process does not explode in finite time. We also present two particular examples of TV-CMJ, adapted from age-dependent branching processes, namely the time-varying Bellman-Harris (TV-BH) and the inhomogeneous Poisson branching process (TV-IP).

Chapter 2 still relates to the TV-CMJ branching process and focusses on the integral renewal-like equations that govern the dynamics of the process. We start by deriving an integral equation for the first moment of the process and then introduce a generic integral equation that plays an umbrella role for all of the subsequent integral equations derived from a TV-CMJ process. We prove the existence and uniqueness of a solution to this generic integral equation. We then more specifically treat the case of the integral equation that leads to the process representing the cumulative number of new individuals up to a certain time, a.k.a. cumulative incidence in epidemiology. Our objective is to derive a similar equation for incidence, i.e. the expected number of births at any time-point. In effect, this quantity mathematically corresponds, when it exists, to the derivative of cumulative incidence. To derive such an equation, we extend the definition of the latter quantity as a generalised function (i.e. a Schwartz distribution). This extension allows us to separate the singular and regular parts of the generalised function and derive an equation for the distributional derivative of cumulative incidence. This leads us to an integral renewal-type equation for the expected birth rate of a TV-CMJ process. We then turn to the incidence renewal equation commonly used in epidemiology, which we similarly reformulate in the framework of generalised functions. This puts us on similar mathematical grounds and allows us to show the equivalence between the expected birth rate of a TV-CMJ branching process and the incidence of an infectious disease as commonly defined in the epidemiology literature. It could be said that our equivalence proof provides mathematical foundations to the incidence equation in epidemiology.

Chapter 3 steers away from the TV-CMJ branching process and treats Markov branching processes for structured populations, where each individual is assumed to be characterised

(structured) by a trait belonging to an abstract space and whose dynamics follow a Markovian process. We endeavour to study the most general case where the reproduction law is trait dependent and trait transmission may depend on the mother’s trait. We place ourselves in the super-critical case, i.e. the mean number of progeny is strictly higher than 1, and the underlying Markov process that describes the trait dynamics is recurrent, which means that the trait distribution converges towards a stationary probability measure. We provide a convergence result that both unifies and extends the current literature. More precisely, our starting point was the convergence result of Asmussen and Hering [AH76] that has limitations in practice, as the conditions of convergence rely on the knowledge of the eigentriple of the first-moment semigroup of the branching process. However, the eigentriple is rarely or never known in applications. We therefore introduced new technical conditions that instead require knowledge of a Lyapunov-type function. We also managed to dispense with the uniform convergence condition of [AH76]. In summary, and under technical conditions clarified in Chapter 3, we managed to prove, in discrete time, a.s. and  $L^1$  convergence of the empirical measure of a branching Markov process in the supercritical case. In continuous time, we proved almost sure convergence, whilst still working on finalising a proof for the  $L^1$  convergence. Note that the literature provides examples of  $L^p$  convergence in the case of  $1 < p < \infty$ .

Finally, Chapter 4 does not specifically focus on branching process as it is based on a novel type of VAE that is capable of encoding a prior for Bayesian deep learning. However, if the model itself is not a branching process, it has potential applications in epidemiology. We briefly mention in particular its sister models, namely AggVAE [Sem+23a] and PriorCVAE [Sem+23b], which were applied to the mapping of infectious diseases. We provide a brief overview of the concepts underlying Bayesian variational inference and present the architecture of  $\pi$ VAE which is proven to be a stochastic process from which tractable inference using fully Bayesian MCMC approaches can be applied.

# Appendices

# APPENDIX A

## Appendices to Chapter 1

### A.1 Miscellaneous measure theoretic results

In this section we recall various results or definitions in measure and probability theory that are used at various points in the thesis, in particular in Chapter 1. Most results, unless otherwise stated originate from [Kal97].

#### A.1.1 Product sigma-field

We cite here [Kal97, Lemma 1.2] which defines the product sigma-field of a Cartesian product of metric spaces.

Recall that we can endow a metric or topological space  $S$  with its Borel  $\sigma$ -field  $\mathcal{B}_S$  generated by the topology (i.e. class of open sets) in  $S$ .

Let  $S_t, t \in T$  be a family of spaces for which we define the Cartesian product  $\prod_{t \in T} S_t$ . Typically when  $T = \{1, \dots, n\}$  or  $T = \mathbb{N}$ , we write the product space  $S_1 \times \dots \times S_n$  or  $S_1 \times S_2 \times \dots$  and if  $S_t = S$  for all  $t$ , we write  $S^T$  or  $S^n$  or  $S^\infty$ . When  $S_t$  are topological spaces, the product  $\prod_{t \in T} S_t$  is endowed with the product topology.

**Lemma A.1.1.** *For any separable metric spaces  $S_1, S_2, \dots$ , we have*

$$\mathcal{B}(S_1 \times S_2 \times \dots) = \mathcal{B}_{S_1} \otimes \mathcal{B}_{S_2} \otimes \dots$$

#### A.1.2 Transition kernels

The construction of the probability measure underlying the TV-CMJ branching process in Section 1.3.2 of Chapter 1 relies in part on the composition of probability kernels. Here, based on [Kal17, Ch.I], we first start by recalling the definition of a kernel, then we specify a probability transition kernel and finally cite the composition result [Kal17, Ch.I Lemma 1.17].

**Definition A.1.1** (Transition kernel). ([Kal97, Ch.I pg.55]) Given some measurable spaces  $(S, \mathcal{S})$  and  $(T, \mathcal{T})$ , we define a transition kernel from  $S$  to  $T$  as a measurable function  $\kappa$  from  $S$  to the space of measures on  $T$ , often written as  $\kappa : S \rightarrow T$ . Thus,  $\kappa$  is a function of two variables  $s \in S$  and  $B \in \mathcal{T}$ , such that  $\kappa(s, B)$  is  $S$ -measurable in  $s \in S$  for fixed  $B \in \mathcal{T}$  and is a measure with respect to  $B \in \mathcal{T}$  for fixed  $s \in S$ .

**Definition A.1.2** (Probability transition kernel). A transition kernel is a probability transition kernel if all measures  $\kappa_s$  for  $s \in S$  are probability measures.

**Definition A.1.3** (Compositions of kernels). ([Kal97, Ch.I pg.58]) For any  $s$ -finite kernel  $\mu : S \rightarrow T$  and  $\nu : S \times T \rightarrow U$ , we define their *composition* as the kernel  $k \otimes \nu : S \rightarrow T \times U$  given by

$$(\mu \otimes \nu)_s f = \int \mu_s(dt) \int \nu_{s,t}(du) f(t, u).$$

### A.1.3 Ionescu-Tulcea theorem

This theorem is applied in Section 1.3.2 to construct the recursive conditional probability measures underlying a TV-CMJ branching process.

**Theorem A.1** ([Ion49]). Suppose that  $(\Omega^0, \mathcal{A}^0, \mathbb{P}^0)$  is a probability space and  $(\Omega^i, \mathcal{A}^i)_{i \in \mathbb{N}}$  is a sequence of measurable spaces. For each  $i \in \mathbb{N}$ , let

$$\Omega_i := \prod_{k=0}^i \Omega^k, \quad \text{and} \quad \mathcal{A}_i := \bigotimes_{k=0}^i \mathcal{A}^k,$$

and introduce for any  $i \geq 1$

$$\kappa_i : (\Omega_{i-1}, \mathcal{A}_{i-1}) \rightarrow (\Omega^i, \mathcal{A}^i),$$

the Markov kernel from  $(\Omega_{i-1}, \mathcal{A}_{i-1})$  to  $(\Omega^i, \mathcal{A}^i)$ . Then there exists a sequence  $(\mathbb{P}_i)_{i \in \mathbb{N}}$  of probability measures

$$\mathbb{P}_i := \mathbb{P}_0 \otimes \bigotimes_{k=1}^i \kappa_k,$$

defined on the product space for the sequence  $(\Omega_i, \mathcal{A}_i)_{i \in \mathbb{N}}$  and there exists a uniquely defined probability measure  $\mathbb{P}$  on  $(\prod_{k=0}^{\infty} \Omega^k, \bigotimes_{k=0}^{\infty} \mathcal{A}^k)$  such that

$$\mathbb{P}_i(A) = \mathbb{P}\left(A \times \prod_{k=i+1}^{\infty} \Omega^k\right)$$

is satisfied for each  $A \in \mathcal{A}_i$  and  $i \in \mathbb{N}$ . The measure  $\mathbb{P}$  has conditional probabilities equal to the stochastic kernels.

## A.2 First-order stochastic dominance

The notion of stochastic dominance is in particular described by Levy [Lev90] within a mathematical finance context, but it applies more generally. Here we simply recall the generic definition of first-order stochastic dominance between two CDFs, as this result is used in the proof of equation (1.5.1). We also state and prove a very basic lemma that allows comparing Stieltjes integrals with respect to two CDFs that dominate each other.

**Definition A.2.1** (First-order stochastic dominance (FSD)). A random variable  $X_1$  defined on  $\mathbb{R}$  has first-order stochastic dominance over another random variable  $X_2$  defined on  $\mathbb{R}$  if for

any outcome  $x \in \mathbb{R}$

$$\mathbb{P}(X_1 \geq x) \geq \mathbb{P}(X_2 \geq x).$$

In terms of cumulative distribution functions, this translates into

$$F_{X_1}(x) \leq F_{X_2}(x),$$

with strict inequality for some values of  $x \in \mathbb{R}$

**Lemma A.2.1.** *Let  $F_1 : \mathbb{R} \rightarrow \mathbb{R}$  and  $F_2 : \mathbb{R} \rightarrow \mathbb{R}$  be two non-increasing right continuous functions such that for any  $x \in \mathbb{R}$ ,  $F_1(x) \leq F_2(x)$ , and  $F_1(0) = F_2(0)$ . Then for any non-negative, non-increasing measurable function  $h : \mathbb{R} \rightarrow \mathbb{R}$*

$$\int_0^t h(x) dF_1(x) \leq \int_0^t h(x) dF_2(x), \quad \forall t \geq 0.$$

*Proof.* It suffices to apply integration by parts

$$\int_0^t h(x) dF_1(x) - \int_0^t h(x) dF_2(x) = \underbrace{(F_1(t) - F_2(t))h(t)}_{\leq 0} - \underbrace{\int_0^t (F_1(x) - F_2(x)) dh(x)}_{\leq 0} \leq 0.$$

□

### A.3 Fubini and Tonelli Theorems

The content of this section is directly taken from [Bil13]. We suppose that  $(X, \mathcal{X}, \mu)$  and  $(Y, \mathcal{Y}, \nu)$  are  $\sigma$ -finite measure spaces and  $\pi$ , the product measure is defined such that

$$\pi(E) = \int_X \nu(y : (x, y) \in E) \mu(dx) = \int_Y \mu(x : (x, y) \in E) \nu(dy), \quad E \in \mathcal{X} \times \mathcal{Y}.$$

We cite the classical and ubiquitous Fubini and Tonelli theorem.

**Theorem A.2** (Fubini and Tonelli). [Bil13] *For any nonnegative function  $f$  measurable on  $\mathcal{X} \times \mathcal{Y}$ , the functions  $\int_Y f(x, y) \nu(dy)$  and  $\int_X f(x, y) \mu(dx)$  are measurable respectively on  $\mathcal{X}$  and  $\mathcal{Y}$  and*

$$\int_{X \times Y} f(x, y) \pi(dx, dy) = \int_X \left( \int_Y f(x, y) \nu(dy) \right) \mu(dx), \quad (\text{A.3.1})$$

$$\int_{X \times Y} f(x, y) \pi(dx, dy) = \int_Y \left( \int_X f(x, y) \mu(dx) \right) \nu(dy), \quad (\text{A.3.2})$$

where the left side is a double integral whereas the right sides correspond to iterated integrals. Additionally if  $f$  (not necessarily nonnegative) is integrable with respect to  $\pi$ , then the two functions  $\int_Y f(x, y) \nu(dy)$  and  $\int_X f(x, y) \mu(dx)$  are finite and measurable on  $A_0$  and  $B_0$ , respectively, where  $\mu(X - A_0) = \nu(Y - B_0) = 0$  and (A.3.1) and (A.3.2) hold.

## A.4 Unsymmetric Fubini Theorem

Although the classical Fubini theorem for Lebesgue integrals is well known, we apply in Sections 2.3.2, 2.5.4 and 2.7.2 a version of Fubini that relies on Lebesgue-Stieltjes integrals. It was proven by Cameron and Martin [CM41] and it is now commonly called the Unsymmetric Fubini theorem. We state it here.

**Theorem A.3.** *Let  $k(x)$  be a function of bounded variation in every finite interval of  $\mathbb{R}$ . Let  $p(x, u)$  be Borel measurable in  $(x, u)$ . For almost all  $x \in \mathbb{R}$  with respect to  $k(x)$ , let it be of bounded variation in  $u$  over every finite  $u$ -interval. Denote by  $V(x, u)$  the variation  $V(x, u) = \int_{0+0}^{u+0} |d_v p(x, v)|$ . Assume that  $\int_{-\infty}^{\infty} V(x, u) |dk(x)|$  exists (i.e. is finite) for all  $u \in \mathbb{R}$ . Let  $s(u)$  be Borel measurable on  $\mathbb{R}$ . Then the existence (finiteness) of either*

$$\int_{-\infty}^{\infty} |s(u)| d_u \int_{-\infty}^{\infty} V(x, u) |dk(x)|,$$

or

$$\int_{-\infty}^{\infty} |dk(x)| \int_{-\infty}^{\infty} |s(u)| |d_u p(x, u)|,$$

implies the existence of the other, their equality and the existence and equality of the following integrals

$$\int_{-\infty}^{\infty} s(u) d_u \int_{-\infty}^{\infty} p(x, u) dk(x) = \int_{-\infty}^{\infty} dk(x) \int_{-\infty}^{\infty} s(u) d_u p(x, u).$$

## A.5 Auxiliary lemma to prove Propositions 1.1 and 2.3

Crump and Mode [CM68] introduce sufficient regularity conditions for a CMJ branching process. We present here the statement and the original proof of [CM68, Lemma 4.1] as this result is used in the last steps of the proofs of our Propositions 1.1 and 2.3.

**Lemma A.5.1** (Lemma 4.1 [CM68]). *If  $\Delta(t)$  is a non-decreasing function, right-continuous such that  $\Delta(0) < 1$  and  $\Delta(t) < \infty$  for all  $t \geq 0$  then*

$$\sum_{k=1}^{\infty} \Delta^{k*}(t) < \infty, \quad \forall t \geq 0,$$

where  $\Delta^{k*}(t)$  stands for the  $k$ -th convolution of  $\Delta(t)$  with respect to itself.

*Proof.* Let  $t \geq 0$  be fixed. We first treat the straightforward case where  $\Delta(t) < 1$ . Then,

$$\sum_{k=1}^{\infty} \Delta^{k*}(t) \leq \sum_{k=1}^{\infty} \Delta^k(t) = \frac{1}{1 - \Delta(t)} - 1 < \infty.$$

Let us now assume that  $\Delta(t) \geq 1$ . Then there exists  $\alpha > 0$  such that

$$\bar{\Delta}(u) := \int_0^u e^{-\alpha y} d\Delta(y) < 1, \quad 0 \leq u \leq t.$$

We can then apply the result of the preceding case to get

$$\sum_{k=1}^{\infty} \bar{\Delta}^{k*}(t) < \infty.$$

Using Laplace transforms, it can be shown that

$$\bar{\Delta}^{k*}(u) = \int_0^u e^{-\alpha y} d\Delta^{k*}(y).$$

It follows that

$$\sum_{k=1}^{\infty} \Delta^{k*}(t) \leq e^{\alpha t} \sum_{k=1}^{\infty} \bar{\Delta}^{k*}(t) < \infty.$$

Thus, the series  $\sum_{k=1}^{\infty} \Delta^{k*}(t)$  converges in both cases, when  $\Delta(t) < 1$  and when  $\Delta \geq 1$ .  $\square$

# APPENDIX B

## Appendices to Chapter 2

### B.1 Grönwall's inequalities

#### B.1.1 Grönwall's inequality for measurable functions

The original work of Grönwall [Gro19] led to many variants of the so-called Grönwall inequality. A useful source in particular for finding specific forms of the Grönwall inequality is [Dra03]. We first state a general integral form that can be described as a Grönwall inequality for measurable functions that do not have to be continuous. This result is adapted from [Hor96, Th.3.1.d]

**Lemma B.1.1** (Reformulation of Theorem 3.1 [Hor96]). *Let  $I$  denote an interval in  $\mathbb{R}$  of the form  $[a, \infty)$ , or  $[a, b]$  or  $[a, b)$  with  $a < b$ . Let  $\alpha$  and  $u$  be measurable functions from  $I$  to  $\mathbb{R}$ . Let  $\mu$  be a continuous non-negative measure on the Borel  $\sigma$ -algebra of  $I$  satisfying  $\mu([a, t]) < \infty$ , for  $t \in I$ . Assume that  $u$  is integrable with respect to  $\mu$ , i.e.  $\int_{[a, t]} |u(s)| \mu(ds) < \infty$ , for  $t \in I$ , and assume that the following inequality holds.*

$$u(t) \leq \alpha(t) + \int_a^t \beta(s) u(s) ds, \quad t \in I.$$

*If in addition either*

- $\alpha(t) \geq 0$  on  $I$ , or
  - $t \mapsto \mu([a, t])$  is continuous on  $I$  and  $\alpha$  is integrable with respect to  $\mu$ ,
- then,  $u(t)$  satisfies the following Grönwall inequality*

$$u(t) \leq \alpha(t) + \int_a^t \alpha(s) \exp(\mu(I_{s,t})) \mu(ds), \quad t \in I.$$

*In the special case where  $\mu$  has a density  $\beta$  with respect to the Lebesgue measure, the above inequality becomes*

$$u(t) \leq \alpha(t) + \int_a^t \alpha(s) \beta(s) \exp\left(\int_s^t \beta(r) dr\right) ds, \quad t \in I.$$

### B.1.2 Two particular forms

In the following, we describe the two main cases in which the latter result is invoked at various points in Chapter 2 or in [Pak+23].

**Proposition B.1** (Extended Grönwall's inequalities). *Let the function  $(t, \sigma) \mapsto f(t; \sigma)$  be non-negative and locally bounded on  $\mathbb{R}_+ \times \mathbb{R}_+$ . Assume that there exists a bounded non-negative function  $(t, \sigma, u) \mapsto b(t; \sigma; u)$  on  $\mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}_+$  such that either*

$$f(t; \sigma) \leq \int_0^{t-\sigma} f(t; u + \sigma) b(t; \sigma, u) du, \quad t \geq \sigma \geq 0, \quad (\text{B.1.1})$$

or

$$f(t; \sigma) \leq \int_0^{t-\sigma} f(t - u; \sigma) b(t; \sigma, u) du, \quad t \geq \sigma \geq 0. \quad (\text{B.1.2})$$

Then

$$f(t; \sigma) = 0 \quad \text{for any } t \geq \sigma \geq 0. \quad (\text{B.1.3})$$

*Proof.* Writing  $\bar{b} := \sup_{t, \sigma, u} b(t; \sigma; u)$ , inequality (B.1.1) implies

$$f(t; \sigma) \leq \bar{b} \int_0^{t-\sigma} f(t; u + \sigma) du, \quad t \geq \sigma \geq 0.$$

If we now consider  $f_c(t) := f(c; c - t)$  for any  $c \geq t \geq 0$ , then

$$f_c(t) \leq \bar{b} \int_0^t f(c; u + c - t) du = \bar{b} \int_0^t f_c(t - u) du = \bar{b} \int_0^t f_c(u) du.$$

We can now apply Lemma B.1.1 which does not require the function  $f_c$  to be continuous and deduce that  $f_c(t) = 0$  for any  $c \geq t \geq 0$ , from which (B.1.3) easily follows. Next, similarly, starting from inequality (B.1.2), we have

$$f(t; \sigma) \leq \bar{b} \int_0^{t-\sigma} f(t - u; \sigma) du \leq \bar{b} \int_0^t f(t - u; \sigma) du \leq \bar{b} \int_0^t f(u; \sigma) du,$$

and (B.1.3) follows by applying again Lemma B.1.1.  $\square$

## B.2 Proof of Proposition 2.1

### B.2.1 Existence of a solution

Our proof of existence is constructive. More precisely, for any given pair  $(t; \sigma) \in \mathbb{R}_+ \times \mathbb{R}_+$ , we introduce a sequence of non-negative functions  $(u_n(t; \sigma))_{n \in \mathbb{N}}$  whose successive elements satisfy the following recursive relationship, mirroring the (Generic Integral Eq.)

$$\begin{cases} u_0(t; \sigma) = v(t; \sigma), & n = 0, \\ u_n(t; \sigma) = v(t; \sigma) + \int_{(0, t-\sigma]} u_{n-1}(t; \sigma + s) \lambda(s; \sigma) ds, & n \geq 1. \end{cases} \quad (\text{B.2.1})$$

We observe that, given conditions (EU-1) and (EU-2), the sequence is trivially well defined for any pair  $(t; \sigma) \in \mathbb{R}_+ \times \mathbb{R}_+$ . We first show that the sequence of functions  $(u_n(t; \sigma))_{n \in \mathbb{N}}$

converges uniformly as  $n \rightarrow \infty$ . Then, it will suffice to appeal to the Dominated Convergence Theorem to pass to the limit in the recursive relationship and deduce that a solution to the (Generic Integral Eq.) exists as the limit function of the sequence  $(u_n(t; \sigma))_{n \in \mathbb{N}}$ .

We start by proving the following lemma that describes the main analytical properties of the functions  $(t; \sigma) \mapsto u_n(t; \sigma)$  defined on  $\mathbb{R}_+ \times \mathbb{R}_+$  when  $n \in \mathbb{N}$  is fixed.

**Lemma B.2.1.** *For any  $n \in \mathbb{N}$ , let*

$$\begin{aligned} u_n : \mathbb{R}_+ \times \mathbb{R}_+ &\rightarrow \mathbb{R}_+ \\ (t; \sigma) &\mapsto u_n(t; \sigma), \end{aligned}$$

*be a function defined inductively by the system of equations (B.2.1). Then the function  $(t; \sigma) \mapsto u_n(t; \sigma)$  is right-continuous with respect to  $(t; \sigma)$  on  $\mathbb{R}_+ \times \mathbb{R}_+$ , and for any given  $\sigma \geq 0$  fixed, the function  $t \mapsto u_n(t; \sigma)$  is nondecreasing on  $\mathbb{R}_+$ .*

*Proof.* The properties of monotonicity with respect to  $t$  and boundedness with respect to  $\sigma$  can be proven by direct induction proofs without any difficulty. We will therefore only focus our attention here on proving the right-continuity property of  $t \mapsto u_n(t; \sigma)$ .

To do so we fix  $\sigma \geq 0$ . First, considering  $t < \sigma$ , by Assumption (EU-2), we have  $u_0(t; \sigma) = v(t; \sigma) = 0$  and then, by recursion, for any  $n \geq 1$ ,  $u_n(t; \sigma) = v(t; \sigma) = 0$  when  $t < \sigma$ . Thus, for  $0 \leq t < \sigma$ ,  $u_n(t; \sigma) \equiv 0$  is constant and de facto continuous with respect to  $t$ .

Let us now consider  $t \geq \sigma$ . To prove by induction on  $n \in \mathbb{N}$ , for a given  $\sigma \geq 0$ , that  $t \mapsto u_n(t; \sigma)$  is right-continuous, we introduce  $T > 0$  such that  $0 \leq \sigma \leq t < T$  to enable us to work on a compact subset of  $\mathbb{R}_+$ . We prove the following induction hypothesis

$$\forall s \in [\sigma, t], \quad t \mapsto u_n(t; s) \text{ is right-continuous with respect to } t \in [\sigma, T]. \quad (\text{Ind-H1})$$

Firstly, by Assumption (EU-2), we observe that (Ind-H1) holds true for  $n = 0$  whereby  $u_0(t; \sigma) = v(t; \sigma)$ . Then assume that  $n \geq 1$  and (Ind-H1) holds true for  $n \geq 1$ . By a change of variable, the recursive relationship (B.2.1) becomes (since it is assumed to also hold for any  $s \in [\sigma, t]$ )

$$u_n(t; s) = v(t; s) + \int_{(s, t]} u_{n-1}(t; p) \lambda(p - s; s) ds.$$

Now let us take any  $\epsilon > 0$  and compute for  $s \in [\sigma, t]$ ,

$$\begin{aligned} u_{n+1}(t + \epsilon; s) - u_{n+1}(t; s) &= v(t + \epsilon; s) + \int_{(s, t + \epsilon]} u_n(t + \epsilon; p) \lambda(p - s; s) dp, \\ &\quad - v(t; s) - \int_{(s, t]} u_n(t; p) \lambda(p - s; s) dp, \\ &= v(t + \epsilon; s) - v(t; s) + \int_{(s, t]} \left( u_n(t + \epsilon; p) - u_n(t; p) \right) \lambda(p - s; s) dp, \\ &\quad + \int_{(t, t + \epsilon]} u_n(t + \epsilon, p) \lambda(p - s; s) dp. \end{aligned} \quad (\text{B.2.2})$$

We show that each of the terms in equation (B.2.2) tends to 0 as  $\epsilon \rightarrow 0$ . Indeed, by Assumption

(EU-2), that is by right-continuity of  $v(t; s)$  with respect to  $t$ , we have

$$\lim_{\epsilon \rightarrow 0^+} \left( v(t + \epsilon; s) - v(t; s) \right) = 0.$$

Next, the application of (Ind-H1) to  $u_n(t; p)$  for  $p \in (s, t] \subset [\sigma, t]$  guarantees that  $t \mapsto u_n(t; p)$  is right continuous. Hence,

$$\lim_{\epsilon \rightarrow 0^+} \left( u_n(t + \epsilon; p) - u_n(t; p) \right) = 0.$$

Furthermore, since  $t \mapsto u_n(t; p)$  is monotonous, it is of finite variation with respect to  $t$ , which means that the difference  $\left( u_n(t + \epsilon; p) - u_n(t; p) \right)$  is bounded for any  $p \in (s, t]$ . Finally by Assumption (EU-1),  $p \mapsto \lambda(p - s; s)$  is continuous for  $s$  fixed. It is therefore bounded for  $p \in (s, t]$ . In summary, the entire integrand  $\left( u_n(t + \epsilon; p) - u_n(t; p) \right) \lambda(p - s; s)$  is not only bounded but also tends to 0. By the Dominated Convergence Theorem, we have

$$\lim_{\epsilon \rightarrow 0^+} \int_{(s, t]} \left( u_n(t + \epsilon; p) - u_n(t; p) \right) \lambda(p - s; s) dp = 0.$$

Let us deal with the last term in equation (B.2.2) where we chose  $\epsilon > 0$  small enough to guarantee that  $\epsilon \leq T - t$  which means that  $t + \epsilon \in [\sigma, T]$ . We can then apply the induction hypothesis (Ind-H1), that is  $t \mapsto u_n(t + \epsilon; p)$  is right-continuous. Combined with the boundedness of  $s \mapsto \lambda(p - s; s)$ , we get

$$\lim_{\epsilon \rightarrow 0^+} \int_{(t, t + \epsilon]} u_n(t + \epsilon, p) \lambda(p - s; s) dp = 0.$$

This ends the proof of (Ind-H1). Thus  $t \mapsto u_n(t; \sigma)$  is right-continuous on  $[0, T]$  for any  $T > 0$ . Right-continuity on  $\mathbb{R}_+$  follows.  $\square$

**Lemma B.2.2.** *For any given  $(t; \sigma) \in \mathbb{R}_+ \times \mathbb{R}_+$ , the sequence of functions  $(u_n(t; \sigma))_{n \in \mathbb{N}}$  defined recursively in (B.2.1) is non-decreasing with respect to  $n \in \mathbb{N}$ , that is,  $u_n(t; \sigma) \leq u_{n+1}(t; \sigma)$ .*

*Proof.* The proof is very straightforward by induction. We will omit stating the induction hypothesis and just observe that the case  $n = 0$  relies trivially on the non-negativity of  $v(t; \sigma)$  and of  $\Lambda(t; \sigma)$ . Next for  $n \geq 1$  and any  $s \in [\sigma, t]$ ,

$$u_{n+1}(t; s) - u_n(t; s) = \int_{(s, t]} \left( u_n(t; p) - u_{n-1}(t; p) \right) \lambda(p - s; s) dp,$$

and we observe that the integrand term is nonnegative by the induction hypothesis given for the previous rank.  $\square$

Let us now consider  $\mathcal{C}_b([0, T])$  the space of non-negative bounded right continuous functions on  $[0, T]$  which we endow with the uniform distance defined as

$$d(f, g) := \sup_{t \in [0, T]} |f(t) - g(t)|, \quad f, g \in \mathcal{C}_b([0, T]).$$

The space  $(\mathcal{C}_b([0, T], d)$  so defined is a complete metric space. We show in Lemma B.2.3 that the sequence of functions  $(u_n(t; \sigma))_{n \in \mathbb{N}}$  converges uniformly in the function space  $\mathcal{C}_b([0, T])$ .

Indeed we will first show that the sequence of functions  $t \mapsto u_n(t; \sigma)$  satisfies for any  $\sigma \geq 0$  and  $t \in [0, T]$ ,

$$d(u_n(t; \sigma), u_{n-1}(t; \sigma)) \longrightarrow 0 \text{ as } n \rightarrow \infty,$$

which means that the sequence of functions  $(u_n(t; \sigma))_{n \in \mathbb{N}}$  is uniformly Cauchy for  $\sigma \geq 0$  and  $t \in [0, T]$ . Then, uniform convergence follows simply by recalling that any uniformly Cauchy sequence in a complete metric space does converge uniformly.

**Lemma B.2.3.** *For any given  $(t; \sigma) \in \mathbb{R}_+ \times \mathbb{R}_+$ , the sequence of functions  $(u_n(t; \sigma))_{n \in \mathbb{N}}$  defined recursively in (B.2.1) converges uniformly as  $n \rightarrow \infty$  toward a limit function which we denote  $u(t; \sigma)$ , i.e.,*

$$u(t; \sigma) := \lim_{n \rightarrow \infty} u_n(t; \sigma). \quad (\text{B.2.3})$$

*Proof.* Glossing over the trivial case where  $t < \sigma$ , let us assume  $t \geq \sigma$  and fix  $T > t$ . By Assumption (EU-2),  $v(t; \sigma)$  is bounded with respect to both variables  $t$  and  $\sigma$  on any compact of  $\mathbb{R}_+ \times \mathbb{R}_+$ . In particular, for any  $T > 0$ , there exists a non-negative constant  $K_T$  such that

$$K_T := \sup_{\substack{t, \sigma \in [0, T] \\ \sigma \leq t}} v(t; \sigma) < \infty.$$

Additionally by the second part of Assumption (EU-1),  $s \mapsto \lambda(p; s)$  is bounded on  $[0, T]$  for any  $p \in [0, T]$ . Thus there exists a finite non-negative constant  $L_T$  such that

$$L_T := \sup_{\substack{s, p \in [0, T] \\ s \leq p}} \lambda(p; s) < \infty.$$

We now prove the following induction hypothesis, which will allow us to deduce that for any given  $(t; \sigma) \in \mathbb{R}_+ \times \mathbb{R}_+$ , the sequence  $(u_n(t; \sigma))_{n \in \mathbb{N}}$  is Cauchy. We have fixed  $T > 0$  and  $\sigma, t \in [0, T]$ . Let us show that

$$\forall s \in [\sigma, t], \quad n \geq 1, \quad |u_n(t; s) - u_{n-1}(t; s)| \leq K_T \frac{L_T^n (t - s)^n}{n!}. \quad (\text{Ind-H2})$$

By Assumption (EU-1), for any  $s \in [0, T]$ , the function  $p \mapsto \lambda(p; s)$  is continuous on  $[0, T]$  and therefore bounded on the compact  $[0, T]$ . We verify (Ind-H2) holds for  $n = 1$ . Indeed,

$$\begin{aligned} u_1(t; s) - u_0(t; s) &= \int_{(s, t]} u_0(t; p) \lambda(p - s; s) dp, \\ &= \int_{(s, t]} v(t; p) \lambda(p - s; s) dp, \\ &\leq K_T L_T (t - s). \end{aligned}$$

We next assume that (Ind-H2) is valid for  $n \geq 1$ . We have

$$\begin{aligned}
 |u_{n+1}(t; s) - u_n(t; s)| &= \left| \int_{(s, t]} \left( u_n(t; p) - u_{n-1}(t; p) \right) \lambda(p - s; s) dp \right|, \\
 &\leq L_T \int_{(s, t]} |u_n(t; p) - u_{n-1}(t; p)| dp, \\
 &\leq L_T \int_{(s, t]} K_T \frac{L_T^n (t - p)^n}{n!} dp, \\
 &\leq K_T L_T^{n+1} \frac{(t - s)^{n+1}}{(n + 1)!},
 \end{aligned}$$

where induction hypothesis (Ind-H2) at rank  $n$  was appealed to in the third line above. The last inequality confirms that (Ind-H2) is valid for any  $n \geq 1$ . It now suffices to apply it to  $s = \sigma$ . We have shown that for any  $T > 0$  and  $t, \sigma \in [0, T]$ ,

$$|u_n(t; \sigma) - u_{n-1}(t; \sigma)| \leq K_T \frac{L_T^n (t - \sigma)^n}{n!}, \quad \forall n \in \mathbb{N},$$

which implies, given the factorial term on the right-hand side in the denominator that

$$\lim_{n \rightarrow \infty} |u_n(t; \sigma) - u_{n-1}(t; \sigma)| = 0, \quad \forall t, \sigma \in [0, T].$$

The latter ensures that  $(u_n(t; \sigma))_{n \in \mathbb{N}}$  is a Cauchy sequence in  $(\mathcal{C}_b([0, T], d))$ . Now since  $(\mathcal{C}_b([0, T], d))$  is a complete metric space, it follows that  $(u_n(t; \sigma))_{n \in \mathbb{N}}$  converges uniformly. We denote

$$u(t; \sigma) := \lim_{n \rightarrow \infty} u_n(t; \sigma).$$

Finally, the above convergence was proved for any  $T > 0$  and for any  $t \in [0, T]$ , thus it holds for any  $t \in \mathbb{R}_+$ . This ends of the proof of Lemma B.2.3.  $\square$

It now remains to show that the limit function  $u(t; \sigma)$  defined in (B.2.3) is a solution to the (Generic Integral Eq.).

**Lemma B.2.4** (Existence). *Under assumptions (EU-1) and (EU-2), the function  $u(t; \sigma)$  determined in equation (B.2.3) on  $\mathbb{R}_+ \times \mathbb{R}_+$  is a non-negative solution to the (Generic Integral Eq.) which satisfies the properties stated in Lemma B.2.1, namely  $u(t; \sigma)$  is non-decreasing and right-continuous with respect to  $t$  for  $\sigma$  fixed and bounded in  $\sigma$  for  $t$  fixed.*

*Proof.* We fix  $t, \sigma \in \mathbb{R}_+ \times \mathbb{R}_+$ . Uniform convergence of the sequence  $(u_n(t; \sigma))_{n \in \mathbb{N}}$  allows us to appeal to the Dominated Convergence Theorem, in which case the recursive relationship (B.2.1) yields

$$u(t; \sigma) = v(t; \sigma) + \int_{(0, t - \sigma]} u(t; \sigma + s) \lambda(s; \sigma) ds.$$

This equation is valid for any  $(t; \sigma) \in \mathbb{R}_+ \times \mathbb{R}_+$  and guarantees the existence of a solution to the (Generic Integral Eq.). Finally, preservation of the properties of  $u_n(t; \sigma)$  by uniform convergence and passage to the limit gives us the desired properties for  $u(t; \sigma)$ .  $\square$

### B.2.2 Uniqueness of the solution

Uniqueness is proven by appealing to a classical argument based on Grönwall inequalities (see Appendix B.1 where we present a specific version of Grönwall's inequality in measure spaces for functions that do not have to be a priori continuous [Hor96, Theorem 3.1.(d)]).

**Lemma B.2.5** (Uniqueness). *Under Assumptions (EU-1) and (EU-2), the (Generic Integral Eq.) admits a unique solution.*

*Proof.* Let us assume the existence of another possible solution, which we denote  $\tilde{u}(t; \sigma)$  to the (Generic Integral Eq.). The function  $(t; \sigma) \mapsto \tilde{u}(t; \sigma)$  is thus assumed to satisfy the exact same properties as  $u(t; \sigma)$ . We define the right-continuous function  $w : \mathbb{R}_+ \times \mathbb{R}_+ \rightarrow \mathbb{R}$

$$w(t; \sigma) := \begin{cases} 0, & t < \sigma, \\ u(t; \sigma) - \tilde{u}(t; \sigma), & t \geq \sigma, \end{cases}$$

which verifies the system of equations

$$w(t; \sigma) = \begin{cases} 0, & t < \sigma, \\ \int_{(\sigma, t]} w(t; p) \lambda(p - \sigma; \sigma) dp, & t \geq \sigma. \end{cases} \quad (\text{B.2.4})$$

Note that the function  $t \mapsto w(t; \sigma)$  defined on  $\mathbb{R}_+$  is actually continuous at  $t = \sigma$  with  $w(\sigma; \sigma) = 0$ . We fix  $(t; \sigma) \in \mathbb{R}_+ \times \mathbb{R}_+$  such that  $t > \sigma$  and introduce the univariate function  $\nu \mapsto w_t(\nu)$  defined on  $(0, t]$  by

$$w_t(\nu) := w(t; t - \nu).$$

Obviously the latter can be rewritten  $w_t(t - s) = w(t; s)$  for any  $s \in [0, t]$ . We now rewrite (B.2.4) substituting  $w(t; \cdot)$  with the univariate function  $w_t(t - \cdot)$ . We get

$$\begin{aligned} w(t; s) = w_t(t - s) &= \int_{(s, t]} w_t(t - p) \lambda(p - s; s) dp, \\ &= \int_{(0, t-s]} w_t(t - s - \nu) \lambda(\nu; s) d\nu. \end{aligned}$$

Next, let  $T > 0$  be such that  $T > t > \sigma \geq 0$ . Writing  $\zeta = t - s$  for  $s \in [0, t]$ , we get

$$\begin{aligned} w_t(\zeta) &= \int_{(0, \zeta]} w_t(\zeta - \nu) \lambda(\nu; s) d\nu, \\ &\leq L_T \int_{(0, \zeta]} w_t(\zeta - \nu) d\nu, \\ &\leq L_T \int_{(0, \zeta]} w_t(p) dp. \end{aligned}$$

Right-continuity of  $s \mapsto w(t; s)$  induces left-continuity of  $\nu \mapsto w_t(\nu)$  on  $(0, t]$ . Furthermore, condition (EU-2) ensures that  $w_t(\nu)$  has at most first-kind discontinuities (i.e. if there exist discontinuities, none of them is of infinite type). The function  $\nu \mapsto w_t(\nu)$  is therefore measurable with respect to  $\nu \in [0, t]$  and the conditions for applying Lemma B.1.1 (see Appendix B.1) are satisfied. We obtain

$$w_t(\zeta) \leq 0.$$

Given that a symmetric mirroring argument can be applied to the function  $-w(t; s)$ , we deduce that  $w_t(\zeta) \equiv 0$ , in other words,

$$w_t(t - s) = w(t; s) = u(t; s) - \tilde{u}(t; s) = 0,$$

which proves uniqueness of the solution to the (Generic Integral Eq.).  $\square$

### B.3 Useful results on Schwartz distributions

In many fields of applied mathematics appears the need to differentiate functions which according to the classical theory of functional analysis are not continuous and therefore not differentiable. The theory of distributions formalised by Laurent Schwartz ([HS52]) provides a clear mathematical framework within which functions can be generalised as continuous linear functionals and therefore extended as *differentiable functions*. Since Schwartz' original introduction, the literature surrounding generalised functions has flourished. Here we present the main results we rely upon in Chapter 2. Our main source is based on [Al-92]. In particular, we only consider real-valued generalised functions.

**Definition B.3.1** (Distributions and test functions). A *distribution on  $\mathbb{R}$*  denoted  $T : \mathcal{D}(\mathbb{R}) \rightarrow \mathbb{R}$  is any continuous linear functional on  $\mathcal{D}(\mathbb{R})$ , where the space  $\mathcal{D}(\mathbb{R})$  is the set  $\mathcal{C}_c^\infty(\mathbb{R})$  of infinitely differentiable functions with compact support on  $\mathbb{R}$ . The linear space of all distributions on  $\mathbb{R}$  is denoted  $\mathcal{D}'(\mathbb{R})$ . It is the topological dual of  $\mathcal{D}(\mathbb{R})$  and is often called the space of test functions on  $\mathbb{R}$ .

**Definition B.3.2** (Regular distribution). A distribution  $T$  is said to be *regular* if there exists a locally integrable function  $f_T$  on  $\mathbb{R}$  such that

$$T(\varphi) = \langle f_T, \varphi \rangle = \int_{\mathbb{R}} f_T(x) \varphi(x) dx, \quad \varphi \in \mathcal{D}(\mathbb{R}).$$

Otherwise the distribution is said to be *singular*.

**Definition B.3.3** (Differentiation of distributions). For any  $T \in \mathcal{D}'(\mathbb{R})$ , we define

$$d^{(\alpha)}T(\varphi) := (-1)^{|\alpha|}T(d^{(\alpha)}\varphi), \quad \varphi \in \mathcal{D}(\mathbb{R}), \quad \alpha \in (\mathbb{N}^*)^n.$$

We recall the definition of the convolution of a function in  $C_c^\infty(\mathbb{R})$  with a locally integrable function.

**Definition B.3.4.** For any  $f \in L_{\text{loc}}^1(\mathbb{R})$  and  $\varphi \in \mathcal{C}_K^\infty(\mathbb{R})$ , where  $K$  is a compact subset of  $\mathbb{R}$ , we define the *convolution* of  $f$  and  $\varphi$  as the function

$$(f * \varphi)(x) := \int_{\mathbb{R}} f(x - y) \varphi(y) dy = \int_{\mathbb{R}} f(y) \varphi(x - y) dy.$$

Definition B.3.4 can be extended to the convolution of two distributions  $T_1$  and  $T_2$  on  $\mathbb{R}$  by the formula

$$\begin{aligned} (T_1 * T_2)(\varphi) &= (T_1 \otimes T_2)(\varphi(x + y)), \\ &= T_1(T_2(\varphi(x + y))), \quad \varphi \in \mathcal{D}(\mathbb{R}). \end{aligned}$$

**Definition B.3.5** (Translation of a distribution). We define the translation of the distribution  $T \in \mathcal{D}'(\mathbb{R})$  for  $\varphi \in \mathcal{D}(\mathbb{R})$  by

$$\begin{aligned} (\tau_h T)(\varphi) &= T(\tau_{-h} \varphi), \\ &= T(\varphi(x+h)), \end{aligned}$$

which is again a distribution in  $\mathbb{R}$ .

**Lemma B.3.1.** For any  $T \in \mathcal{D}'(\mathbb{R})$ ,

$$\tau_h T = \delta_h * T. \quad (\text{B.3.1})$$

**Definition B.3.6** (Reflection). For any function  $f$  on  $\mathbb{R}$ , we define its *reflection* in 0 as the function  $\check{f}$  defined on  $\mathbb{R}$  by  $\check{f}(x) = f(-x)$ . We extend this definition to  $\mathcal{D}'(\mathbb{R})$  by duality

$$\check{T}(\varphi) = T(\check{\varphi}), \quad \varphi \in \mathcal{D}(\mathbb{R}).$$

The following theorem states that the convolution of a distribution with a smooth function is a smooth function.

**Theorem B.1.** For all  $T \in \mathcal{D}'(\mathbb{R})$  and all  $\psi \in C_c^\infty(\mathbb{R})$ , the convolution

$$(T * \psi)(x) = T(\tau_x \check{\psi}).$$

*Proof of theorem (B.1).* For any  $\varphi \in \mathcal{D}(\mathbb{R})$ , we have

$$(T * \psi)(\varphi) = T_x(\langle \psi(y), \varphi(x+y) \rangle).$$

But we can write

$$\begin{aligned} \langle \psi(y), \varphi(x+y) \rangle &= \int \psi(y) \varphi(x+y) dy, \\ &= \int \psi(\xi - x) \varphi(\xi) d\xi, \\ &= \langle \psi(\xi - x), \varphi(\xi) \rangle, \\ &= \langle \check{\psi}(x - \xi), \varphi(\xi) \rangle, \\ &= \langle \tau_\xi \check{\psi}(x), \varphi(\xi) \rangle. \end{aligned}$$

Hence we obtain

$$\begin{aligned} (T * \psi)(\varphi) &= T_x(\langle \tau_\xi \check{\psi}(x), \varphi(\xi) \rangle), \\ &= \langle T(\tau_\xi \check{\psi}), \varphi(\xi) \rangle. \end{aligned}$$

□

We recall the following result, which ensures that the convolution of a distribution with a function in  $C_c^\infty(\mathbb{R})$  stays in  $C_c^\infty(\mathbb{R})$ .

**Corollary B.1.** If  $\varphi(x, y) \in C_c^\infty(\mathbb{R}^2)$  has compact support as a function of  $x$  and  $y$  separately, then  $T_1(\varphi)(y) \in C_c^\infty(\mathbb{R})$  for every  $T_1 \in \mathcal{D}'(\mathbb{R})$  and  $T_2(\varphi)(x) \in C_c^\infty(\mathbb{R})$  for every  $T_2 \in \mathcal{D}'(\mathbb{R})$ .

By Corollary B.1, the function  $(T * \psi)(x) = T(\tau_x \check{\psi}) = T_y(\psi(x - y))$  is thus still in  $C_c^\infty(\mathbb{R})$ .

## B.4 Auxiliary result to prove Theorem 2.1

Recalling the functions  $s_n \mapsto \hat{f}_{t_n}(s_n)$  on  $[0, t_n]$  and  $s \mapsto f_t(s)$  on  $[0, t]$ , which were defined in Theorem 2.1, we prove here that they are both  $K$ -Lipschitz-continuous (where the constant  $K > 0$  may differ).

**Lemma B.4.1.** *Let  $n \in \mathbb{N}$ ,  $t \geq s \geq 0$ , and let  $s_n \mapsto \hat{f}_{t_n}(s_n)$  on  $[0, t_n]$  and  $s \mapsto f_t(s)$  on  $[0, t]$  be defined according to Theorem 2.1. Then there exists a positive constant  $K > 0$  such that  $s_n \mapsto \hat{f}_{t_n}(s_n)$  is  $K$ -Lipschitz on  $[0, t_n]$  and  $s \mapsto f_t(s)$  is  $K$ -Lipschitz on  $[0, t]$ .*

*Proof.* It suffices to prove here that for  $s_n \in [0, t_n]$ , we have

$$\left| \hat{f}_{t_n} \left( s_n + \frac{1}{n} - \hat{f}_{t_n}(s_n) \right) \right| \leq \frac{K}{n},$$

and for  $s \in [0, t]$ ,

$$\left| f_t \left( s + \frac{1}{n} \right) - f_t(s) \right| \leq \frac{K}{n}.$$

Recall first that

$$\hat{f}_{t_n}(s_n) = v(t_n; t_n - s_n) + \frac{1}{n} \sum_{k=1}^{ns_n-1} \hat{f}_{t_n} \left( s_n - \frac{k}{n} \right) \lambda \left( \frac{k}{n}, t_n - s_n \right).$$

Thus,

$$\begin{aligned} \hat{f}_{t_n} \left( s_n + \frac{1}{n} \right) &= v \left( t_n; t_n - \left( s_n + \frac{1}{n} \right) \right) + \frac{1}{n} \sum_{k=1}^{ns_n} \hat{f}_{t_n} \left( s_n - \frac{k-1}{n} \right) \lambda \left( \frac{k}{n}, t_n - \left( s_n + \frac{1}{n} \right) \right), \\ &= v \left( t_n; t_n - \left( s_n + \frac{1}{n} \right) \right) + \frac{1}{n} \sum_{k=0}^{ns_n-1} \hat{f}_{t_n} \left( s_n - \frac{k}{n} \right) \lambda \left( \frac{k+1}{n}, t_n - \left( s_n + \frac{1}{n} \right) \right). \end{aligned}$$

This leads to

$$\begin{aligned} \left| \hat{f}_{t_n} \left( s_n + \frac{1}{n} \right) - \hat{f}_{t_n}(s_n) \right| &\leq \left| v \left( t_n; t_n - \left( s_n + \frac{1}{n} \right) \right) - v(t_n; t_n - s_n) \right| \\ &\quad + \frac{1}{n} \sum_{k=1}^{ns_n-1} \left| \hat{f}_{t_n} \left( s_n - \frac{k}{n} \right) \lambda \left( \frac{k+1}{n}, t_n - \left( s_n + \frac{1}{n} \right) \right) - \lambda \left( \frac{k}{n}, t_n - s_n \right) \right| \\ &\quad + \frac{1}{n} \left| \hat{f}_{t_n}(s_n) \lambda \left( \frac{1}{n}; t_n - \left( s_n + \frac{1}{n} \right) \right) \right|, \\ &\leq \frac{K}{n}, \end{aligned}$$

where we relied on the Lipschitz property of the functions  $(t; s) \mapsto v(t, s)$  and  $(t; s) \mapsto \lambda(t; s)$  in the last line.

Similarly for  $f_t(s)$ , let us write first

$$\begin{aligned} f_t\left(s + \frac{1}{n}\right) &= v\left(t; t - \left(s + \frac{1}{n}\right)\right) + \int_0^{s+1/n} f_t\left(s + \frac{1}{n} - u\right) \lambda\left(u; t - \left(s + \frac{1}{n}\right)\right) du, \\ &= v\left(t; t - \left(s + \frac{1}{n}\right)\right) + \int_{-1/n}^s f_t(s - u) \lambda\left(u + \frac{1}{n}; t - \left(s + \frac{1}{n}\right)\right) du. \end{aligned}$$

Thus,

$$\begin{aligned} \left|f_t\left(s + \frac{1}{n}\right) - f_t(s)\right| &\leq \left|v\left(t; t - \left(s + \frac{1}{n}\right)\right) - v(t; t - s)\right| \\ &\quad + \int_0^s f_t(s - u) \left|\lambda\left(u + \frac{1}{n}; t - \left(s + \frac{1}{n}\right)\right) - \lambda(u; t - s)\right| du \\ &\quad + \int_{-1/n}^0 f_t(s - u) \lambda\left(u + \frac{1}{n}; t - \left(s + \frac{1}{n}\right)\right) du, \\ &\leq \frac{K}{n}, \end{aligned}$$

where we again relied on the Lipschitz property of the functions  $(t; s) \mapsto v(t, s)$  and  $(t; s) \mapsto \lambda(t; s)$  in the last line.

□

## B.5 Leibniz rule for differentiation under an integral

This classical differentiation rule is applied repeatedly in Section 2.6.

**Lemma B.5.1.** [AR80] *Let  $-\infty < x_0 < x_1 < \infty$  and let  $\alpha(x)$  and  $\beta(x)$  be continuously differentiable functions of  $x$  for  $x_0 < x < x_1$ . Let  $f(x, t)$  and  $\frac{\partial f}{\partial x}(x, t)$  be continuous functions of  $x \in (x_0, x_1)$  and  $t \in (\alpha(x), \beta(x))$ . Then for any  $x \in (x_0, x_1)$*

$$\frac{d}{dx} \int_{\alpha(x)}^{\beta(x)} f(x, t) dt = \beta'(x) f(x, \beta(x)) - \alpha'(x) f(x, \alpha(x)) + \int_{\alpha(x)}^{\beta(x)} \frac{\partial f}{\partial x}(x, t) dt.$$

# APPENDIX C

## Appendices to Chapter 3

### C.1 Perron-Frobenius theorem

We recall the Perron-Frobenius theorem for positive matrices as stated in [AN04, Ch.5 Th.1]

**Theorem C.1** (Perron-Frobenius). *A strictly positive matrix  $S = (S_{ij})_{i,j=1}^p$  has a maximal eigenvalue  $\lambda$  which is positive, simple and has associated positive right and left eigenvectors  $\mathbf{u}$  and  $\mathbf{v}$ . If these vectors are normalised such that  $\langle \mathbf{u}, \mathbf{v} \rangle = 1$  and  $\langle \mathbf{u}, \mathbf{1} \rangle = 1$  then one can write*

$$S^n = \lambda^n P + R^n,$$

where  $P$  is the matrix whose  $(i, j)$ -th entry is  $P_{ij} = u_i v_j$  and where

(i)  $PR = RP = 0$ , and

(ii) The  $(i, j)$ -th entry of  $R^n$ , namely  $R_{ij}^{(n)}$  is such that  $R_{ij}^{(n)} \leq c\lambda_0^n$ ,  $i, j = 1, \dots, p$  for some  $c < \infty$  and  $0 < \lambda_0 < \lambda$ .

### C.2 Kesten-Stigum theorem

We cite here the celebrated Kesten-Stigum theorem [KS66] originally proven in 1966. It states that the asymptotic proportions of individuals of each type in a multitype  $p$  branching population are deterministic and given by the maximal left eigenvalue of the first-moment matrix  $S = (\mathbb{E}(Z_{1,j}^{(i)}))_{1 \leq i, j \leq p}$ . The matrix  $S$  is assumed to be strictly positive. We associate the left and right eigenvectors  $\mathbf{u}$  and  $\mathbf{v}$  with its maximum eigenvalue  $\lambda > 0$ . In particular  $\mathbf{u}$  is positive.

**Assumption** (Kesten-Stigum condition). *Let  $p \in \mathbb{N}$ .*

$$\mathbb{E} \left( Z_{1,j}^{(i)} \log Z_{1,j}^{(i)} \right) < \infty, \quad 1 \leq i, j \leq p. \quad (\text{C.2.1})$$

**Theorem C.2** (Kesten-Stigum [KS66]). *Consider a multidimensional Galton-Watson process  $Z_n = (Z_{n,1}, \dots, Z_{n,p})_{n \in \mathbb{N}}$ , and denote its first-moment matrix  $S = (\mathbb{E}(Z_{1,j}^{(i)}))_{1 \leq i, j \leq p}$ . Assume that the process is not degenerate, i.e. there exist  $1 \leq i, j \leq p$  such that  $S_{i,j} \neq 1$  and assume*

that  $S$  is positive. Then if  $\lambda > 1$ , we have

$$\lim_{n \rightarrow \infty} \frac{Z_n}{\lambda^n} = \mathbf{v}W, \quad \text{a.s.}$$

where  $W$  is random vector. Additionally there is no extinction if and only if condition (C.2.1) holds, i.e.  $\mathbb{P}(W > 0) > 0 \iff$  (C.2.1) holds. Also, if (C.2.1) is satisfied then for all  $1 \leq i \leq p$ ,

- $\mathbb{E}(W \mid Z_0 = \mathbf{e}_i) = u_i$ ,
- $\mathbb{P}(W = 0 \mid Z_0 = \mathbf{e}_i) = q_i$ , where  $q_i$  is the  $i$ -th coordinate of the extinction probability of the multitype Galton-Watson process started with an individual of type  $i$ .

### C.3 Asmussen-Hering Law of Large Numbers

We state here the main results (in discrete and continuous time) related to strong limit theorems for general supercritical branching processes which were proven by Asmussen and Hering [AH76]. Our work in Chapter 3 aims in particular at extending their results and other, more recent, existing results where the state space is infinite dimensional. Among other things, in Chapter 3, under some technical conditions related to the convergence of the first-moment semigroup and the existence of appropriate Lyapunov functions for the underlying Markov process, we prove almost sure convergence of the branching process. Our sufficient set of technical conditions on the convergence of the first-moment semigroup should be easier to verify in practice than, for instance, the set of assumptions denoted  $(M)$  by Asmussen and Hering [AH76]. In particular, we will observe that our conditions require knowledge of a Lyapunov-type function instead of full knowledge of the eigentriplet, as is the case in [AH76]. Most importantly, we will also see that our set of assumptions does not require uniform convergence.

Let us first recall the set of assumptions denoted  $(M)$  in [AH76] and on which all their results are based. Let  $(\mathcal{X}, \mathcal{A})$  be a measurable space and  $\mathcal{B}$  be the Banach algebra of all bounded and  $\mathcal{A}$  measurable functions  $f$  in  $\mathcal{X}$  with the norm  $\|f\| := \sup_{x \in \mathcal{X}} |f(x)|$ . Let  $\mathcal{B}_+$  be the nonnegative cone on  $\mathcal{B}$ . Below  $\mathbb{T}$  represents either  $\mathbb{N}$  or  $\mathbb{R}_+$ .

**Assumptions** (Denoted  $(M)$  in [AH76]). *The first moment semigroup  $(M_t)_{t \in \mathbb{T}}$  of a branching Markov process  $(Z_t)_{t \in \mathbb{T}}$  exists and can be represented in the following form.*

- (i) *There exists a nonnegative number  $\lambda \in \mathbb{R}_+$ , a nonnegative function  $h \in \mathcal{B}_+$  and a nonnegative bounded linear functional  $\gamma$  on  $\mathcal{B}$  such that for all  $x \in \mathcal{X}$ ,  $t \in \mathbb{T}$ , and  $f \in \mathcal{B}$ , we can write, assuming  $\gamma$  and  $h$  are normalised (i.e.  $\gamma(h) = 1$ )*

$$M_t f(x) = \lambda^t \gamma(f) h(x) + Q_t f(x). \quad (\text{C.3.1})$$

- (ii)  *$\gamma$  and  $h$  satisfy the property of eigenvectors for  $M_t$ , which is equivalent to*

$$\gamma Q_t = 0, \quad Q_t h = 0, \quad \gamma M_t = \lambda^t \gamma, \quad M_t h = \lambda^t h. \quad (\text{C.3.2})$$

- (iii) *The remainder  $Q_t$  is uniformly bounded with respect to  $x \in \mathcal{X}$ , in the sense that there exists a function  $\alpha : \mathbb{T} \rightarrow [0, \infty)$  satisfying*

$$\lambda^{-t} \alpha_t \rightarrow 0, \quad t \rightarrow \infty, \quad (\text{C.3.3})$$

and such that for any  $x \in \mathcal{X}$ ,  $f \in \mathcal{B}_+$ ,  $t > 0$ ,

$$|Q_t f(x)| \leq \alpha_t \gamma(f) h(x). \quad (\text{C.3.4})$$

This set of assumptions (M) implies that  $\gamma$  is a measure and that  $\gamma$  and  $h$  are, respectively, the left and right eigenvectors of the first moment semigroup  $M_t$  corresponding to the eigenvalue  $\lambda$ . In particular,  $\gamma$  is the invariant distribution of the types.

Note that Asmussen and Hering provided an alternative representation of the preceding assumptions on which they rely for most of their proofs. For any  $n \in \mathbb{N}$ ,  $\exists c_n^-, c_n^+ \in \mathbb{R}$ , such that  $c_n^-, c_n^+ \rightarrow 1$ ,  $n \rightarrow \infty$ , and for all  $x \in \mathcal{X}$ ,  $f \in \mathcal{B}$ ,

$$\lambda^n c_n^- \gamma(f) h(x) \leq M_n f(x) \leq \lambda^n c_n^+ \gamma(f) h(x).$$

They then prove the following lemma.

**Lemma C.3.1** (Martingale  $W_t$ ). *The process defined as*

$$W_t := \lambda^{-t} Z_t(h), \quad t \in \mathbb{T},$$

*is a martingale with respect to the filtration generated by the Markov process  $\mathcal{F}_t := \sigma(X_s; 0 \leq s \leq t)$ .*

They then introduce the quantity

$$I_t := \mathbb{E}_\gamma (Z_t(h) \log Z_t(h)) = \int_{\mathcal{X}} \mathbb{E}_{\delta_x} (Z_t(h) \log Z_t(h)) \gamma(dx),$$

which represents the  $L \log L$  moment of the process  $Z_t(h)$ , provided that the type of the original particle is distributed according to the measure  $\gamma$  (i.e. given that the initial distribution of the population is  $\gamma$ ) and obtain a first result in the supercritical case ( $\lambda > 1$ ) in discrete time.

**Lemma C.3.2.** *Assume  $\lambda > 1$ . The martingale  $(W_n)_{n \in \mathbb{N}}$  converges a.s. towards a limit denoted  $W$ , that is,*

$$\lim_{n \rightarrow \infty} \lambda^{-n} Z_n(h) \xrightarrow{n \rightarrow \infty} W, \quad a.s.$$

Next, Asmussen and Hering prove  $L^1$  convergence of the process in the supercritical case under the set of conditions described by (M).

**Theorem C.3** (Theorem (1) [AH76]). *Assume conditions (M) hold and  $\lambda > 1$ . Then*

$$\lambda^{-n} Z_n(f) \xrightarrow{n \rightarrow \infty} \gamma(f) W, \quad a.s. \quad f \in L_\gamma^1, \quad (\text{C.3.5})$$

where  $L_\gamma^1$  represents the space of functions measurable with respect to the eigenmeasure  $\gamma$ .

Recalling Asmussen-Hering's nondegeneracy condition of Kesten-Stigum type

$$I_n = \mathbb{E}_\gamma (Z_n(h) \log Z_n(h)) < \infty, \quad (\text{C.3.6})$$

the authors next prove propagation of the  $L \log L$  condition.

**Theorem C.4** (Theorem (2) [AH76]).

$$\forall n > 0, \quad \text{either} \quad I_n < \infty \quad \text{or} \quad I_n = \infty. \quad (\text{First assertion})$$

Furthermore,

$$\begin{cases} \text{If } I_1 < \infty, & \text{then} \quad \mathbb{E}_{\delta_x} W = h(x), \\ \text{If } I_1 = \infty, & \text{then} \quad W \equiv 0 \quad a.s. \end{cases} \quad (\text{Second assertion})$$

Finally, we end this brief summary of [AH76] by stating their main result in continuous time.

**Theorem C.5** (Theorem (1)' [AH76]). *Let  $\mathcal{X}$  be a separable metric space,  $\mathcal{U}$  the topological Borel algebra and  $Z_t$ , a right-continuous branching Markov process satisfying condition (M) with  $\lambda > 1$ . If  $g \in \mathcal{B}_+$  is lower semi-continuous almost everywhere with respect to  $\gamma$ , and if, in addition,*

$$\lambda^{-t} Z_t(g) \xrightarrow{a.s.} \gamma(g)W, \quad t \rightarrow \infty,$$

then,

$$\lambda^{-t} Z_t(gf) \xrightarrow{a.s.} \gamma(gf)W, \quad t \rightarrow \infty, \quad (\text{C.3.7})$$

for every  $f \in \mathcal{B}$  which is continuous almost everywhere with respect to  $\gamma$ .

## C.4 Extended convexity result

We present a convexity result originally stated in [AH76, Lemma 1]. This result and its conditional version are used at multiple points throughout the proofs of chapter 3. We call it *extended Jensen or extended convexity result* because it is reminiscent of the classical Jensen inequality.

First let us recall the definition (3.3) given in Section 3.1 of the non-negative, continuous function  $x \mapsto \log^* x$  defined on  $\mathbb{R}_+$  such that

$$\log^* x = \begin{cases} x/e, & 0 \leq x \leq e, \\ \log x, & x > e. \end{cases}$$

**Lemma C.4.1** (Lemma 1 of [AH76]). *The function  $x \mapsto x \log^* x$  defined for  $x \in \mathbb{R}_+$  is non-negative, non-decreasing and convex. Moreover, for  $n \in \mathbb{N}^*$ , if we let  $S_n = X_1 + \dots + X_n$  be the sum of  $n$  independent non-negative random variables, then*

$$\mathbb{E} S_n \log^* S_n \leq \mathbb{E} S_n \log^* \mathbb{E} S_n + \sum_{i=1}^n \mathbb{E} X_i \log^* X_i.$$

*Proof.* Observe that the function  $x \mapsto \log^* x$  is concave. Thus, for any  $a, b \geq 0$ ,

$$\log^*(a + b) \leq \log^* a + \log^* b. \quad (\text{C.4.1})$$

The first part of the lemma is a direct consequence of (C.4.1) and of the properties and definition of  $\log^* x$ , and can be thus deduced immediately. Next we write for any  $S_n = \sum_{i=1}^n X_i$  where

$X_i$  are non-negative independent random variables,

$$\begin{aligned}
\mathbb{E} S_n \log^* S_n &= \mathbb{E} \left( \sum_{i=1}^n X_i \log^* \sum_{i=1}^n X_i \right), \\
&\leq \sum_{i=1}^n \left( \mathbb{E} X_i \log^* \sum_{\substack{j=1 \\ j \neq i}}^n X_j + \mathbb{E} X_i \log^* X_i \right), \\
&\leq \sum_{i=1}^n \left( \mathbb{E} X_i \log^* \mathbb{E} \sum_{\substack{j=1 \\ j \neq i}}^n X_j + \mathbb{E} X_i \log^* X_i \right), \\
&\leq \mathbb{E} S_n \log^* \mathbb{E} S_n + \sum_{i=1}^n \mathbb{E} X_i \log^* X_i,
\end{aligned}$$

where we used property (C.4.1) in the the second line, the classical Jensen inequality in the third line and finally the non-decreasing property of  $x \mapsto x \log^* x$  in the last line.  $\square$

*Remark.* Note that in practice we will be using a conditional version of the above inequality. Let  $(\mathcal{F}_n)_n$  be the natural filtration associated to the random variables  $(X_n)_{n \in \mathbb{N}}$ . Then for any  $n \in \mathbb{N}^*$  and any non-negative integer  $0 \leq q \leq n$ ,

$$\mathbb{E}(S_n \log^* S_n | \mathcal{F}_{n-q}) \leq \mathbb{E}(S_n | \mathcal{F}_{n-q}) \log^* \mathbb{E}(S_n | \mathcal{F}_{n-q}) + \sum_{i=1}^n \mathbb{E}(X_i \log^* X_i | \mathcal{F}_{n-q}).$$

## C.5 Useful bounds related to the $\log^* x$ function

**Lemma C.5.1.** *Let  $C \in (0, 1)$ . Then we have for any  $x \geq 0$  and  $n \in \mathbb{N}$ ,*

$$\log^*(x^n) \geq Cn \log x.$$

*Proof.* We introduce the function defined on  $\mathbb{R}_+$  by  $g(x) := \log^*(x^n) - nC \log x$  where it is assumed that  $0 < C < 1$ . By definition of  $\log^* x$  in (3.3),

$$g(x) := \begin{cases} \frac{x^n}{e} - nC \log x, & 0 \leq x < e^{1/n}, \\ n(1 - C) \log x, & x \geq e^{1/n}. \end{cases}$$

For  $x \in [0, e^{1/n}]$ ,

$$g'(x) = \frac{n}{ex} (x^n - eC),$$

which is negative for  $0 \leq x < (Ce)^{1/n}$ , and non-negative for  $(Ce)^{1/n} < x < e^{1/n}$ . Next, for  $x \geq e^{1/n}$ ,

$$g'(x) = \frac{n(1 - C)}{x} > 0.$$

Thus  $g(x)$  admits a minimum at  $x = (Ce)^{1/n}$ . Since  $g((Ce)^{1/n}) = -C \log C \geq 0$ , it follows

that  $g(x) \geq 0$  for any  $x$  such that  $x^n \in [0, e]$ .  $\square$

**Lemma C.5.2.** *Let  $C > \frac{1}{e}$ . Then for any  $x \geq 0$  and  $n \in \mathbb{N}$ , we have  $n \log x \leq Cx^n$ .*

*Proof.* We prove this inequality by introducing the function defined on  $\mathbb{R}_+$  given by  $g(x) := Cx^n - n \log x$ . Then trivially  $g'(x) = \frac{n}{x}(Cx^n - 1)$ , and  $g(x)$  attains its minimum at  $x = (\frac{1}{C})^{1/n}$  where  $g\left((\frac{1}{C})^{1/n}\right) = 1 + \log C > 0$  since  $C > \frac{1}{e}$ .  $\square$

## C.6 Auxiliary result for the martingale decomposition

Recall here that we introduced a remainder term  $R_n^{(r)}(f)$  in Lemma 3.2.2 for the decomposition of  $X_n^{(r)}(f)$  whereby

$$R_n^{(r)}(f) = M_n^{(r)}(f) + \sum_{i=1}^{n-1} M_{n-i}^{(r)}(T_r^{i-1}(T_r - \text{Id})f).$$

Here we show that we can re-write the second term in the sum above in terms of  $\Delta_i^{(r)}$ 's. We will then eventually rely upon this result to show that the remainder term  $R_n^{(r)}(f)$  converges to 0 as  $n \rightarrow \infty$ .

**Lemma C.6.1.** *For any  $n \geq 1$ ,  $N \leq n$  and  $f \in \mathcal{B}(V)$ , we have*

$$\sum_{i=N}^{n-1} M_{n-i}^{(r)}(T_r^i - T_r^{i-1})(f) = \sum_{i=1}^{n-N} \Delta_i^{(r)}(T_r^{n-i}(f) - T_r^{N-1}(f)). \quad (\text{C.6.1})$$

First observe trivially that by definition of the martingale and its increment, we have

$$M_{n-i}^{(r)}(f) = \sum_{j=1}^{n-i} \Delta_j^{(r)}(f).$$

Next, we use the above expression for  $M$  and then expand the second sum in  $\Delta$ 's

$$\begin{aligned} \sum_{i=N}^{n-1} M_{n-i}^{(r)}(T_r^i - T_r^{i-1})(f) &= \sum_{i=N}^{n-1} \sum_{j=1}^{n-i} \Delta_j^{(r)}(T_r^i - T_r^{i-1})(f), \\ &= \sum_{j=1}^{n-N} \Delta_j^{(r)}(T_r^N - T_r^{N-1})(f) \\ &\quad + \sum_{j=1}^{n-N-1} \Delta_j^{(r)}(T_r^{N+1} - T_r^N)(f) \\ &\quad + \sum_{j=1}^{n-N-2} \Delta_j^{(r)}(T_r^{N+2} - T_r^{N+1})(f) \\ &\quad + \vdots \\ &\quad + \Delta_1^{(r)}(T_r^{n-1} - T_r^{n-2}). \end{aligned}$$

We then develop each term line-by-line to obtain

$$\begin{aligned}
\sum_{i=N}^{n-1} M_{n-i}^{(r)}(T_r^i - T_r^{i-1})(f) &= \Delta_1^{(r)} T_r^{n-1}(f) - \sum_{j=1}^{n-N} \Delta_j^{(r)} (T_r^{N-1})(f) \\
&\quad + \sum_{j=1}^{n-N} \Delta_j^{(r)} (T_r^N(f)) - \sum_{j=1}^{n-N-1} \Delta_j^{(r)} (T_r^N(f)) \\
&\quad + \sum_{j=1}^{n-N-1} \Delta_j^{(r)} (T_r^{N+1}(f)) - \sum_{j=1}^{n-N-2} \Delta_j^{(r)} (T_r^{N+1}(f)) \\
&\quad \vdots \\
&\quad + \sum_{j=1}^2 \Delta_j^{(r)} (T_r^{n-2}(f)) - \Delta_1^{(r)} (T_r^{n-2}(f)).
\end{aligned}$$

But we observe that each line above (from the second one onwards) simplifies and one term remains. More precisely we get

$$\begin{aligned}
\sum_{i=N}^{n-1} M_{n-i}^{(r)}(T_r^i - T_r^{i-1})(f) &= \Delta_1^{(r)} T_r^{n-1}(f) - \sum_{j=1}^{n-N} \Delta_j^{(r)} (T_r^{N-1})(f) \\
&\quad + \Delta_{n-N}^{(r)} (T_r^N(f)) + \Delta_{n-N-1}^{(r)} (T_r^{N+1}(f)) + \dots + \Delta_2^{(r)} T_r^{n-2}(f).
\end{aligned}$$

We finally obtain

$$\begin{aligned}
\sum_{i=N}^{n-1} M_{n-i}^{(r)}(T_r^i - T_r^{i-1})(f) &= \sum_{i=1}^{n-N} \Delta_i^{(r)} (T_r^{n-i}(f)) - \sum_{j=1}^{n-N} \Delta_j^{(r)} (T_r^{N-1})(f), \\
&= \sum_{i=1}^{n-N} \Delta_i^{(r)} (T_r^{n-i}(f) - T_r^{N-1}(f)).
\end{aligned}$$

## Bibliography

- [AR80] John Amazigo and Lester Rubinfeld. “Advanced calculus and its applications to the engineering and physical sciences”. In: *John Wiley & Sons* (1980).
- [Ant74] Charles Antoniak. “Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems”. In: *The annals of statistics* (1974), pp. 1152–1174.
- [AK93] Ovide Arino and Marek Kimmel. “Comparison of approaches to modeling of cell population dynamics”. In: *SIAM Journal on Applied Mathematics* 53.5 (1993), pp. 1480–1504.
- [AH76] Søren Asmussen and Heinrich Hering. “Strong limit theorems for general supercritical branching processes with applications to branching diffusions”. In: *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 36.3 (1976), pp. 195–212.
- [AK98a] Krishna Athreya and Hye-Jeong Kang. “Some limit theorems for positive recurrent branching Markov chains: I”. In: *Advances in Applied Probability* 30.3 (1998), pp. 693–710.
- [AK98b] Krishna Athreya and Hye-Jeong Kang. “Some limit theorems for positive recurrent branching Markov chains: II”. In: *Advances in Applied Probability* 30.3 (1998), pp. 711–722.
- [AK71] Krishna Athreya and Samuel Karlin. “On branching processes with random environments: I: Extinction probabilities”. In: *The Annals of Mathematical Statistics* 42.5 (1971), pp. 1499–1520.
- [AN04] Krishna Athreya and Peter Ney. *Branching processes*. Courier Corporation, 2004.
- [Bai53] Norman TJ Bailey. “The total size of a general stochastic epidemic”. In: *Biometrika* (1953), pp. 177–185.
- [Bal86] Frank Ball. “A unified approach to the distribution of total size and total area under the trajectory of infectives in epidemic models”. In: *Advances in Applied Probability* 18.2 (1986), pp. 289–310.
- [Bal83] Frank Ball. “The threshold behaviour of epidemic models”. In: *Journal of Applied Probability* 20.2 (1983), pp. 227–241.
- [BD95] Frank Ball and Peter Donnelly. “Strong approximations for epidemic models”. In: *Stochastic processes and their applications* 55.1 (1995), pp. 1–21.
- [Bal+14] Frank Ball, Miguel González, Rodrigo Martínez, and Maroussia Slavtchova-Bojkova. “Stochastic monotonicity and continuity properties of functions defined on Crump–Mode–Jagers branching processes, with application to vaccination in epidemic modelling”. In: (2014).

- [Ban09] Vincent Bansaye. “Cell contamination and branching processes in a random environment with immigration”. In: *Advances in Applied Probability* 41.4 (2009), pp. 1059–1081.
- [Ban24] Vincent Bansaye. *Processus de branchement et populations structurées*. Accessed: 15-11-2024. Jan. 2024.
- [BB13] Vincent Bansaye and Christian Böinghoff. “Lower large deviations for supercritical branching processes in random environment”. In: *Proceedings of the Steklov Institute of Mathematics* 282 (2013), pp. 15–34.
- [BC18] Vincent Bansaye and Alain Camanes. “Queueing for an infinite bus line and aging branching process”. In: *Queueing Systems* 88.1 (2018), pp. 99–138.
- [BCG20] Vincent Bansaye, Bertrand Cloez, and Pierre Gabriel. “Ergodic behavior of non-conservative semigroups via generalized Doeblin’s conditions”. In: *Acta Applicandae Mathematicae* 166 (2020), pp. 29–72.
- [Ban+22] Vincent Bansaye, Bertrand Cloez, Pierre Gabriel, and Aline Marguet. “A non-conservative Harris ergodic theorem”. In: *Journal of the London Mathematical Society* 106.3 (2022), pp. 2459–2510.
- [Ban+11] Vincent Bansaye, Jean-François Delmas, Laurence Marsalle, and Viet Chi Tran. “Limit theorems for Markov processes indexed by continuous time Galton–Watson trees”. In: *Annals of Applied Probability* (2011).
- [BGY23] Vincent Bansaye, Chenlin Gu, and Linglong Yuan. “A growth-fragmentation-isolation process on random recursive trees and contact tracing”. In: *The Annals of Applied Probability* 33.6B (2023), pp. 5233–5278.
- [BH15] Vincent Bansaye and Chunmao Huang. “Weak law of large numbers for some Markov chains along non homogeneous genealogies”. In: *ESAIM: Probability and Statistics* 19 (2015), pp. 307–326.
- [BM15] Vincent Bansaye and Sylvie Méléard. *Stochastic models for structured populations*. Vol. 16. Springer, 2015.
- [Bar78] Maurice Bartlett. *An introduction to stochastic processes: with special reference to methods and applications*. CUP Archive, 1978.
- [Bar67] Robert Bartoszyński. “Branching processes and the theory of epidemics”. In: *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability*. Vol. 4. University of California Press. 1967, p. 259.
- [BH52] Richard Bellman and Theodore Harris. “On age-dependent binary branching processes”. In: *The Annals of Mathematics* 55 (1952), pp. 280–295.
- [BH48] Richard Bellman and Theodore Harris. “On the theory of age-dependent stochastic branching processes”. In: *Proceedings of the National Academy of Sciences of the United States of America* 34 (1948), pp. 601–604.
- [BW20] Jean Bertoin and Alexander Watson. “The strong Malthusian behavior of growth-fragmentation processes”. In: *Annales Henri Lebesgue* 3 (2020), pp. 795–823.
- [Bet+17] Michael Betancourt, Simon Byrne, Sam Livingstone, and Mark Girolami. “The geometric foundations of Hamiltonian Monte Carlo”. In: *Bernoulli* (2017).
- [BG87] John Biggins and Thomas Götz. “Expected population size in the generation-dependent branching process”. In: *Journal of applied probability* 24.2 (1987), pp. 304–314.

- [Bil13] Patrick Billingsley. *Convergence of probability measures*. John Wiley & Sons, 2013.
- [Bir+12] Paul Birrell, Tim Chadborn, Noel Gill, Valerie Delpech, and Daniela De Angelis. “Estimating trends in incidence, time-to-diagnosis and undiagnosed prevalence using a CD4-based Bayesian back-calculation”. In: *Statistical Communications in Infectious Diseases* 4.1 (2012).
- [Bis06] Christopher Bishop. “Pattern recognition and machine learning”. In: *Springer google schola* 2 (2006), pp. 1122–1128.
- [Blu+15] Charles Blundell, Julien Cornebise, Koray Kavukcuoglu, and Daan Wierstra. “Weight uncertainty in neural networks”. In: *32nd International Conference on Machine Learning, ICML 2015*. 2015. ISBN: 9781510810587. arXiv: [1505.05424](#).
- [Bra05] Fred Brauer. “Age of infection in epidemiology models.” In: *Electronic Journal of Differential Equations (EJDE)[electronic only]* 2005 (2005), pp. 29–37.
- [Bra08] Fred Brauer. “Age-of-infection and the final size relation”. In: *Mathematical Biosciences & Engineering* 5.4 (2008), pp. 681–690.
- [Bre+12] Dimitri Breda, Odo Diekmann, WF De Graaf, Andrea Pugliese, and Rossana Vermiglio. “On the formulation of epidemic models (an appraisal of Kermack and McKendrick)”. In: *Journal of biological dynamics* 6.sup2 (2012), pp. 103–117.
- [Bri+19] Francesco Brizzi, Paul Birrell, Martyn T Plummer, Peter Kirwan, Alison E Brown, Valerie C Delpech, O Noel Gill, and Daniela De Angelis. “Extending Bayesian back-calculation to estimate age and time specific HIV incidence”. In: *Lifetime data analysis* 25 (2019), pp. 757–780.
- [BG88] Ron Brookmeyer and Mitchell H Gail. “A method for obtaining short-term projections and lower bounds on the size of the AIDS epidemic”. In: *Journal of the American Statistical Association* 83.402 (1988), pp. 301–308.
- [CM41] Robert Cameron and William Martin. “An unsymmetric Fubini theorem”. In: (1941).
- [CM23] José Cañizo and Stéphane Mischler. “Harris-type results on geometric and subgeometric convergence to equilibrium for stochastic semigroups”. In: *Journal of Functional Analysis* 284.7 (2023), p. 109830.
- [Car+17] Bob Carpenter, Andrew Gelman, Matthew Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. “Stan : A Probabilistic Programming Language”. In: *Journal of Statistical Software* 76.1 (2017), pp. 1–32.
- [CDS18] Anthony Caterini, Arnaud Doucet, and Dino Sejdinovic. “Hamiltonian variational auto-encoder”. In: *Advances in Neural Information Processing Systems*. 2018. arXiv: [1805.11328](#).
- [CV16] Nicolas Champagnat and Denis Villemonais. “Exponential convergence to quasi-stationary distribution and Q-process”. In: *Probability Theory and Related Fields* 164.1 (2016), pp. 243–283.
- [CV23] Nicolas Champagnat and Denis Villemonais. “General criteria for the study of quasi-stationarity”. In: *Electronic Journal of Probability* 28 (2023), pp. 1–84.
- [CDE18] David Champredon, Jonathan Dushoff, and David JD Earn. “Equivalence of the Erlang-distributed SEIR epidemic model and the renewal equation”. In: *SIAM Journal on Applied Mathematics* 78.6 (2018), pp. 3258–3278.

- [Chi06] Tran Viet Chi. “Large Population Limit and Time Behaviour of a Stochastic Particle Model Describing an Age-structured Population”. In: (2006).
- [Chi72] Vladimir Pavlovich Chistyakov. “Limit Theorems for Age-Dependent Branching Processes”. In: *Theory of Probability & Its Applications* 17.1 (1972), pp. 54–71.
- [Clo17] Bertrand Cloez. “Limit theorems for some branching measure-valued processes”. In: *Advances in Applied Probability* 49.2 (2017), pp. 549–580.
- [CG24] Bertrand Cloez and Pierre Gabriel. “Fast, slow convergence, and concentration in the house of cards replicator-mutator model”. In: *Differential and Integral Equations* 37.7/8 (2024), pp. 547–584.
- [Coh89] Harry Cohn. “On the growth of the multitype supercritical branching process in a random environment”. In: *The Annals of Probability* 17.3 (1989), pp. 1118–1123.
- [CJ94] Harry Cohn and Peter Jagers. “General branching processes in varying environment”. In: *The Annals of Applied Probability* (1994), pp. 184–193.
- [Cor+13] Anne Cori, Neil M Ferguson, Christophe Fraser, and Simon Cauchemez. “A new framework and software to estimate time-varying reproduction numbers during epidemics”. In: *American journal of epidemiology* 178.9 (2013), pp. 1505–1512.
- [CM68] Kenny Crump and Charles Mode. “A general age-dependent branching process. I”. In: *Journal of mathematical analysis and applications* 24.3 (1968), pp. 494–508.
- [CM69] Kenny Crump and Charles Mode. “A general age-dependent branching process. II”. In: *Journal of mathematical analysis and applications* 25.1 (1969), pp. 8–17.
- [CM15] Ronald Crump and Graham Medley. “Back-calculating the incidence of infection of leprosy in a Bayesian framework”. In: *Parasites & vectors* 8 (2015), pp. 1–9.
- [CB00] Jisheng Cui and Niels Becker. “Estimating HIV incidence using dates of both HIV and AIDS diagnoses”. In: *Statistics in medicine* 19.9 (2000), pp. 1165–1177.
- [Daw+94] Donald Dawson et al. *Measure-valued processes, stochastic partial differential equations, and interacting systems*. Vol. 5. American Mathematical Soc., 1994.
- [Del04] Pierre Del Moral. *Feynman-kac formulae*. Springer, 2004.
- [DHJ23] Pierre Del Moral, Emma Horton, and Ajay Jasra. “On the stability of positive semigroups”. In: *The Annals of Applied Probability* 33.6A (2023), pp. 4424–4490.
- [DM02] Pierre Del Moral and Laurent Miclo. “On the stability of nonlinear Feynman-Kac semigroups”. In: *Annales de la Faculté des sciences de Toulouse: Mathématiques*. Vol. 11. 2. 2002, pp. 135–175.
- [Deu+99] Sylvie Deuffic, Laurent Buffat, Thierry Poynard, and Alain-Jacques Valleron. “Modeling the hepatitis C virus epidemic in France”. In: *Hepatology* 29.5 (1999), pp. 1596–1601.
- [DGG08] Odo Diekmann, Philipp Getto, and Mats Gyllenberg. “Stability and bifurcation analysis of Volterra functional equations in the light of suns and stars”. In: *SIAM journal on Mathematical Analysis* 39.4 (2008), pp. 1023–1069.
- [DG12] Odo Diekmann and Mats Gyllenberg. “Equations with infinite delay: blending the abstract and the concrete”. In: *Journal of differential equations* 252.2 (2012), pp. 819–851.
- [Die+10] Odo Diekmann, Mats Gyllenberg, Johan AJ Metz, Shinji Nakaoka, and Andre M de Roos. “Daphnia revisited: local stability and bifurcation theory for physiologically structured population models explained by way of an example”. In: *Journal of mathematical biology* 61.2 (2010), pp. 277–318.

- [Die+12] Odo Diekmann, Stephan A Van Gils, Sjoerd MV Lunel, and Hans-Otto Walther. *Delay equations: functional-, complex-, and nonlinear analysis*. Vol. 110. Springer Science & Business Media, 2012.
- [Dra03] Sever Silvestru Dragomir. “Some Gronwall type inequalities and applications”. In: *Science Direct Working Paper S1574-0358* (2003), p. 04.
- [Dur71] Stephen Durham. “Limit theorems for a general critical branching process”. In: *Journal of Applied Probability* 8.1 (1971), pp. 1–16.
- [Dyn94] Evgenii Borisovich Dynkin. *An introduction to branching measure-valued processes*. 6. American Mathematical Soc., 1994.
- [Edl78] L Edler. “Strict supercritical generation-dependent Crump–Mode–Jagers branching processes”. In: *Advances in Applied Probability* 10.4 (1978), pp. 744–763.
- [Eng09] János Engländer. “Law of large numbers for superdiffusions: the non-ergodic case”. In: *Annales de l’IHP Probabilités et statistiques*. Vol. 45. 1. 2009, pp. 1–6.
- [EHK10] János Engländer, Simon Harris, and Andreas Kyprianou. “Strong law of large numbers for branching diffusions”. In: *Annales de l’IHP Probabilités et statistiques*. Vol. 46. 1. 2010, pp. 279–298.
- [EW06] János Engländer and Anita Winter. “Law of large numbers for a class of superdiffusions”. In: *Annales de l’Institut Henri Poincaré (B) Probability and Statistics*. Vol. 42. 2. Elsevier. 2006, pp. 171–185.
- [FKG03] Paddy Farrington, Mona Kanaan, and Nigel Gay. “Branching process models for surveillance of infectious diseases controlled by mass vaccination”. In: *Biostatistics* 4.2 (2003), pp. 279–295.
- [Fil72] Robert Fildes. “An age dependent branching process with variable lifetime distribution”. In: *Advances in Applied Probability* 4.3 (1972), pp. 453–474.
- [Fla+20] Seth Flaxman, Swapnil Mishra, Axel Gandy, H Juliette T Unwin, Thomas A Mellan, Helen Coupland, Charles Whittaker, Harrison Zhu, Tresnia Berah, Jeffrey W Eaton, et al. “Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe”. In: *Nature* 584.7820 (2020), pp. 257–261.
- [FM04] Nicolas Fournier and Sylvie Méléard. “A microscopic probabilistic description of a locally regulated population and macroscopic approximations”. In: *The Annals of Applied Probability* (2004).
- [Fra07] Christophe Fraser. “Estimating individual and household reproduction numbers in an emerging epidemic”. In: *PloS one* 2.8 (2007), e758.
- [Gar+18] Marta Garnelo, Jonathan Schwarz, Dan Rosenbaum, Fabio Viola, Danilo Rezende, SM Eslami, and Yee Whye Teh. “Neural processes”. In: *preprint arXiv:1807.01622* (2018).
- [GLG15] Andrew Gelman, Daniel Lee, and Jiqiang Guo. “Stan: A probabilistic programming language for Bayesian inference and optimization”. In: *Journal of Educational and Behavioral Statistics* 40.5 (2015), pp. 530–543.
- [Gib02] Josiah Willard Gibbs. *Elementary principles in statistical mechanics: developed with especial reference to the rational foundations of thermodynamics*. C. Scribner’s sons, 1902.
- [GHK22] Isaac Gonzalez, Emma Horton, and Andreas Kyprianou. “Asymptotic moments of spatial branching processes”. In: *Probability Theory and Related Fields* 184.3 (2022), pp. 805–858.

- [Gro19] Thomas Hakon Gronwall. “Note on the derivatives with respect to a parameter of the solutions of a system of differential equations”. In: *Annals of Mathematics* 20.4 (1919), pp. 292–296.
- [Al-92] Mohammed Al-Gwaiz. *Theory of distributions*. CRC Press, 1992.
- [Hal13] Paul Halmos. *Measure theory*. Vol. 18. Springer, 2013.
- [HS52] Israel Halperin and Laurent Schwartz. *Introduction to the Theory of Distributions*. University of Toronto Press, 1952.
- [HKK10] Simon Harris, Robert Knobloch, and Andreas Kyprianou. “Strong law of large numbers for fragmentation processes”. In: *Annales de l’IHP Probabilités et statistiques*. Vol. 46. 1. 2010, pp. 119–134.
- [Har63] Theodore Harris. *The theory of branching processes*. Vol. 6. Springer Berlin, 1963.
- [Hau+15] Sophie Hautphenne, Gautier Krings, Jean-Charles Delvenne, and Vincent D Blondel. “Sensitivity analysis of a branching process evolving on a network with application in epidemiology”. In: *Journal of Complex Networks* 3.4 (2015), pp. 606–641.
- [Haw71] Alan Hawkes. “Spectra of some self-exciting and mutually exciting point processes”. In: *Biometrika* 58.1 (1971), pp. 83–90.
- [HS06] Geoffrey Hinton and Ruslan Salakhutdinov. “Reducing the dimensionality of data with neural networks”. In: *science* 313.5786 (2006), pp. 504–507.
- [Hof+13] Matthew Hoffman, David Blei, Chong Wang, and John Paisley. “Stochastic variational inference”. In: *The Journal of Machine Learning Research* 14.1 (2013), pp. 1303–1347.
- [HKV20] Emma Horton, Andreas Kyprianou, and Denis Villemonais. “Stochastic methods for the neutron transport equation I”. In: *The Annals of Applied Probability* 30.6 (2020), pp. 2573–2612.
- [HW20] Emma Horton and Alexander Watson. “Strong laws of large numbers for a growth-fragmentation process with bounded cell sizes”. In: *preprint arXiv:2012.03273* (2020).
- [Hor96] Laszlo Horvath. “Gronwall–Bellman type integral inequalities in measure spaces”. In: *Journal of mathematical analysis and applications* 202.1 (1996), pp. 183–193.
- [Hug+19] Jonathan Huggins, Mikołaj Kasprzak, Trevor Campbell, and Tamara Broderick. “Practical Posterior Error Bounds from Variational Objectives”. In: *arXiv preprint arXiv:1910.04102* (2019).
- [Ion49] Cassius Ionescu Tulcea. “Mesures dans les espaces produits”. In: *Atti Accad. Naz. Lincei Rend* 7 (1949), pp. 208–211.
- [Ish05] Valerie Isham. “Stochastic models for epidemics”. In: *Oxford statistical science series* 33 (2005), p. 27.
- [Jac10] Christine Jacob. “Branching processes: their role in epidemiology”. In: *International journal of environmental research and public health* 7.3 (2010), pp. 1186–1204.
- [JGH18] Arthur Jacot, Franck Gabriel, and Clément Hongler. “Neural tangent kernel: Convergence and generalization in neural networks”. In: *Advances in Neural Information Processing Systems*. 2018.
- [Jag69a] Peter Jagers. “A general stochastic model for population development”. In: *Scandinavian Actuarial Journal* 1969.1-2 (1969), pp. 84–103.

- [Jag75] Peter Jagers. “Branching processes with biological applications”. In: *John Wiley & Sons* (1975).
- [Jag67] Peter Jagers. “Integrals of branching processes”. In: *Biometrika* 54.1-2 (1967), pp. 263–271.
- [Jag69b] Peter Jagers. “Renewal theory and the almost sure convergence of branching processes”. In: *Arkiv för Matematik* 7.6 (1969), pp. 495–504.
- [JN84] Peter Jagers and Olle Nerman. “The growth and composition of branching populations”. In: *Advances in applied probability* 16.2 (1984), pp. 221–259.
- [Kal97] Olav Kallenberg. *Foundations of modern probability*. Vol. 2. Springer, 1997.
- [Kal17] Olav Kallenberg. *Random measures, theory and applications*. Vol. 1. Springer, 2017.
- [Kar47] Kari Karhunen. “On linear methods in probability theory”. In: *Annales Academiae Scientiarum Fennicae, Ser. Al*. Vol. 37. 1947, pp. 3–79.
- [Ken56] David Kendall. “Deterministic and stochastic epidemics in closed populations”. In: *Proceedings of the third Berkeley symposium on mathematical statistics and probability*. Vol. 4. University of California Press Berkeley, CA, USA. 1956, pp. 149–165.
- [KM27] William Kermack and Anderson McKendrick. “A contribution to the mathematical theory of epidemics”. In: *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character* 115.772 (1927), pp. 700–721.
- [KV17] Götz Kersting and Vladimir Vatutin. *Discrete time branching processes in random environment*. John Wiley & Sons, 2017.
- [KS66] Harry Kesten and Bernt Stigum. “A limit theorem for multidimensional Galton-Watson processes”. In: *The Annals of Mathematical Statistics* 37.5 (1966), pp. 1211–1223.
- [Kim+19] Hyunjik Kim, Andriy Mnih, Jonathan Schwarz, Marta Garnelo, S M Ali Eslami, Dan Rosenbaum, Oriol Vinyals, and Yee Whye Teh. “Attentive Neural Processes”. In: *CoRR* abs/1901.0 (2019).
- [Kim82] Marek Kimmel. “An equivalence result for integral equations with application to branching processes”. In: *Bulletin of Mathematical Biology* 44.1 (1982), pp. 1–15.
- [Kim83] Marek Kimmel. “The point-process approach to age-and time-dependent branching processes”. In: *Advances in Applied Probability* 15.1 (1983), pp. 1–20.
- [KA94] Marek Kimmel and Ovide Arino. “Two simple models of almost the same population with very different dynamics”. In: *Mathematical Biosciences* 122.2 (1994), pp. 183–200.
- [KAA95] Marek Kimmel, Ovide Arino, and David Axelrod. “Backward/forward duality of branching processes and cell population dynamics”. In: *Differential Equations And Applications To Biology And To Industry-Proceedings Of The Claremont International Conference Dedicated To The Memory Of Starvros Busenberg (1941-1993)*. World Scientific. 1995, p. 233.
- [KB14] Diederik Kingma and Jimmy Ba. “Adam: A Method for Stochastic Optimization”. In: *arXiv preprint arXiv:1412.6980* (Dec. 2014).
- [KW19] Diederik Kingma and Max Welling. “An introduction to variational autoencoders”. In: *Foundations and Trends® in Machine Learning* 12.4 (2019), pp. 307–392.

- [KW14] Diederik Kingma and Max Welling. “Auto-Encoding Variational Bayes (VAE, reparameterization trick)”. In: *ICLR 2014* (2014).
- [KS23] Konrad Kolesko and Ecaterina Sava-Huss. “Limit theorems for discrete multitype branching processes counted with a characteristic”. In: *Stochastic Processes and their Applications* 162 (2023), pp. 49–75.
- [KM12] Ioannis Kontoyiannis and Sean Meyn. “Geometric ergodicity and the spectral gap of non-reversible Markov chains”. In: *Probability Theory and Related Fields* 154.1 (2012), pp. 327–339.
- [KM03] Ioannis Kontoyiannis and Sean Meyn. “Spectral theory and limit theorems for geometrically ergodic Markov processes”. In: *The Annals of Applied Probability* 13.1 (2003), pp. 304–362.
- [KL51] Solomon Kullback and Richard Leibler. “On information and sufficiency”. In: *The annals of mathematical statistics* 22.1 (1951), pp. 79–86.
- [Kur+97] Thomas Kurtz, Russell Lyons, Robin Pemantle, and Yuval Peres. “A conceptual proof of the Kesten-Stigum theorem for multi-type branching processes”. In: *Classical and modern branching processes* (1997), pp. 181–185.
- [LM23] Arnab Kumar Laha and Sourav Majumdar. “A multi-type branching process model for epidemics with application to COVID-19”. In: *Stochastic Environmental Research and Risk Assessment* 37.1 (2023), pp. 305–325.
- [LPB17] Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. “Simple and scalable predictive uncertainty estimation using deep ensembles”. In: *Advances in Neural Information Processing Systems*. 2017. arXiv: [1612.01474](https://arxiv.org/abs/1612.01474).
- [Lev60] Norman Levinson. “Limiting theorems for age-dependent branching processes”. In: *Illinois Journal of Mathematics* 4.1 (1960), pp. 100–118.
- [Lev90] Haim Levy. “Stochastic dominance”. In: *Utility and probability* (1990), pp. 251–254.
- [LR15] Finn Lindgren and Håvard Rue. “Bayesian spatial modelling with R-INLA”. In: *Journal of statistical software* 63.19 (2015).
- [LRL11] Finn Lindgren, Håvard Rue, and Johan Lindström. “An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach”. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 73.4 (2011), pp. 423–498.
- [Loe48] Michel Loeve. “Fonctions aléatoires du second ordre”. In: *Processus stochastique et mouvement Brownien* (1948), pp. 366–420.
- [LS20] Oren Louidor and Santiago Saglietti. “A strong law of large numbers for supercritical branching Brownian motion with absorption”. In: *Journal of Statistical Physics* 181.4 (2020), pp. 1112–1137.
- [LPP95] Russell Lyons, Robin Pemantle, and Yuval Peres. “Conceptual proofs of  $L \log L$  criteria for mean behavior of branching processes”. In: *The Annals of Probability* (1995), pp. 1125–1138.
- [Mar19] Aline Marguet. “A law of large numbers for branching Markov processes by the ergodicity of ancestral lineages”. In: *ESAIM: Probability and Statistics* 23 (2019), pp. 638–661.
- [Mar17] Aline Marguet. “Branching processes for structured populations and estimators for cell division”. PhD thesis. Université Paris Saclay (COMUE), 2017.

- [Met78] Johannes Metz. “The epidemic in a closed population with all susceptibles equally vulnerable; some results for large susceptible populations and small initial infections”. In: *Acta biotheoretica* 27.1-2 (1978), pp. 75–123.
- [MR99] Maura Mezzetti and Chris Robertson. “A hierarchical Bayesian approach to age-specific back-calculation of cancer incidence rates”. In: *Statistics in medicine* 18.8 (1999), pp. 919–933.
- [Min01] Thomas Minka. “Expectation Propagation for Approximate Bayesian Inference”. In: *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*. UAI’01. Seattle, Washington: Morgan Kaufmann Publishers Inc., 2001, pp. 362–369.
- [Mis+22] Swapnil Mishra, Seth Flaxman, Tresnia Berah, Harrison Zhu, Mikko Pakkanen, and Samir Bhatt. “ $\pi$  VAE: a stochastic process prior for Bayesian deep learning with MCMC”. In: *Statistics and Computing* 32.6 (2022), p. 96.
- [Mod71] Charles Mode. “Multitype branching processes: theory and applications”. In: *American Elsevier Publishing Co.* (1971).
- [Mol+10] Manuel Molina, Manuel Mota, Alfonso Ramos, Inés Maria Puerto, and Rodrigo Martínez. *Workshop on Branching Processes and Their Applications*. Springer-Verlag Berlin Heidelberg, 2010.
- [MSW98] Jesper Møller, Anne Randi Syversveen, and Rasmus Plenge Waagepetersen. “Log gaussian cox processes”. In: *Scandinavian journal of statistics* 25.3 (1998), pp. 451–482.
- [Nea96] Robert Neal. “Bayesian Learning for Neural Networks”. In: *Lecture Notes in Statistics -New York- Srpinger Verlag* (1996). ISSN: 0930-0325.
- [NFF65] Jacques Neveu, Amiel Feinstein, and Robert Fortet. “Mathematical foundations of the calculus of probability”. In: *Holden-Day, Inc* (1965).
- [Ney64] Peter Ney. “Generalized branching processes I: Existence and uniqueness theorems”. In: *Illinois Journal of Mathematics* 8.2 (1964), pp. 316–331.
- [Nou+18] Pierre Nouvellet, Anne Cori, Tini Garske, Isobel M Blake, Ilaria Dorigatti, Wes Hinsley, Thibaut Jombart, Harriet L Mills, Gemma Nedjati-Gilani, Maria D Van Kerkhove, et al. “A simple approach to measure transmissibility and forecast incidence”. In: *Epidemics* 22 (2018), pp. 29–35.
- [Pak72] Anthony Pakes. “Limit theorems for an age-dependent branching process with immigration”. In: *Mathematical Biosciences* 14.3-4 (1972), pp. 221–234.
- [Pak+23] Mikko Pakkanen, Xenia Miscouridou, Matthew Penn, Charles Whittaker, Tresnia Berah, Swapnil Mishra, Thomas Mellan, and Samir Bhatt. “Unifying incidence and prevalence under a time-varying general branching process”. In: *Journal of Mathematical Biology* 87.2 (2023), p. 35.
- [Pav14] Grigorios Pavliotis. *Stochastic processes and applications: diffusion processes, the Fokker-Planck and Langevin equations*. Vol. 60. Springer, 2014.
- [Pol02] David Pollard. *A user’s guide to measure theoretic probability*. 8. Cambridge University Press, 2002.
- [RR08] Ali Rahimi and Benjamin Recht. “Random features for large-scale kernel machines”. In: *Advances in neural information processing systems*. 2008, pp. 1177–1184.

- [Rai22] Tom Rainforth. *Advanced topics in statistical machine learning*. [www.robots.ox.ac.uk/twgr/assets/teaching/notes.pdf](http://www.robots.ox.ac.uk/twgr/assets/teaching/notes.pdf). Accessed: 2024-11-12. 2022.
- [RW06] Carl Rasmussen and Christopher Williams. *Gaussian processes for machine learning*. Adaptive computation and machine learning. MIT Press, 2006.
- [RMW14] Danilo Jimenez Rezende, Shakir Mohamed, and Daan Wierstra. “Stochastic Backpropagation and Approximate Inference in Deep Generative Models”. In: *International Conference on Machine Learning*. 2014, pp. 1278–1286.
- [RBB18] Hippolyt Ritter, Aleksandar Botev, and David Barber. “A scalable laplace approximation for neural networks”. In: *6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings*. 2018.
- [Ros15] Ronald Ross. “Some a priori pathometric equations”. In: *British medical journal* 1.2830 (1915), p. 546.
- [Ros96] Sheldon Ross. *Stochastic processes*. Vol. 2. John Wiley & Sons, 1996.
- [RT09] Daniel Roy and Yee Whye Teh. “The Mondrian process”. In: *Advances in Neural Information Processing Systems 21 - Proceedings of the 2008 Conference*. 2009. ISBN: 9781605609492.
- [ST03] Anton Schwaighofer and Volker Tresp. “Transductive and inductive methods for approximate Gaussian process regression”. In: *Advances in Neural Information Processing Systems*. 2003, pp. 977–984.
- [Sch51] Laurent Schwartz. *Théorie des distributions. 2 (1951)*. Hermann, 1951.
- [Sem+23a] Elizaveta Semenova, Swapnil Mishra, Samir Bhatt, Seth Flaxman, and Juliette Unwin. “Deep learning and MCMC with aggVAE for shifting administrative boundaries: mapping malaria prevalence in Kenya”. In: *International Workshop on Epistemic Uncertainty in Artificial Intelligence*. Springer. 2023, pp. 13–27.
- [Sem+23b] Elizaveta Semenova, Prakhar Verma, Max Cairney-Leeming, Arno Solin, Samir Bhatt, and Seth Flaxman. “PriorCVAE: scalable MCMC parameter inference with Bayesian deep generative modelling”. In: *arXiv preprint arXiv:2304.04307* (2023).
- [Sem+22] Elizaveta Semenova, Yidan Xu, Adam Howes, Theo Rashid, Samir Bhatt, Swapnil Mishra, and Seth Flaxman. “PriorVAE: encoding spatial priors with variational autoencoders for small-area estimation”. In: *Journal of the Royal Society Interface* 19.191 (2022), p. 20220094.
- [Sev67] Boris Sevast’yanov. “On the regularity of branching processes”. In: *Mathematical notes of the Academy of Sciences of the USSR* 1.1 (1967), pp. 34–40.
- [Sev68] Boris Sevast’yanov. “Renewal equations and moments of branching processes”. In: *Mathematical notes of the Academy of Sciences of the USSR* 3 (1968), pp. 3–10.
- [SW69] Walter Smith and William Wilkinson. “On branching processes in random environments”. In: *The Annals of Mathematical Statistics* (1969), pp. 814–827.
- [TBV22] Milica Tomašević, Vincent Bansaye, and Amandine Véber. “Ergodic behaviour of a multi-type growth-fragmentation process modelling the mycelial network of a filamentous fungus”. In: *ESAIM: Probability and Statistics* 26 (2022), pp. 397–435.
- [Ton+18] Jean-Francois Ton, Seth Flaxman, Dino Sejdinovic, and Samir Bhatt. “Spatial mapping with Gaussian processes and nonstationary Fourier features”. In: *Spatial Statistics* (2018). ISSN: 2211-6753.

- [Tra16] Plamen Trayanov. “Crump-Mode-Jagers branching process: a numerical approach”. In: *Branching Processes and Their Applications* (2016), pp. 167–182.
- [VM11] Behrouz Fathi Vajargah and Mojtaba Moradi. “Period dependent branching process and its applications in epidemiology”. In: *Infection, Genetics and Evolution* 11.6 (2011), pp. 1225–1228.
- [Vel23] Aurélien Velleret. “Exponential quasi-ergodicity for processes with discontinuous trajectories”. In: *ESAIM: Probability and Statistics* 27 (2023), pp. 867–912.
- [Wan+19] Ke Wang, Geoff Pleiss, Jacob Gardner, Stephen Tyree, Kilian Q Weinberger, and Andrew Gordon Wilson. “Exact Gaussian processes on a million data points”. In: *Advances in Neural Information Processing Systems*. 2019, pp. 14622–14632.
- [WT11] Max Welling and Yee Whye Teh. “Bayesian learning via stochastic gradient langevin dynamics”. In: *Proceedings of the 28th International Conference on Machine Learning, ICML 2011*. 2011. ISBN: 9781450306195.
- [Wil91] David Williams. *Probability with martingales*. Cambridge university press, 1991.
- [YB08] Christine Yang and Fred Brauer. “Calculation of  $R_0$  for age-of-infection models”. In: *Mathematical Biosciences & Engineering* 5.3 (2008), pp. 585–599.
- [Yao+19] Jiayu Yao, Weiwei Pan, Soumya Ghosh, and Finale Doshi-Velez. “Quality of uncertainty quantification for Bayesian neural network inference”. In: *arXiv preprint arXiv:1906.09686* (2019).
- [Yao+18] Yuling Yao, Aki Vehtari, Daniel Simpson, and Andrew Gelman. “Yes, but did it work?: Evaluating variational inference”. In: *35th International Conference on Machine Learning, ICML 2018*. 2018. ISBN: 9781510867963. arXiv: [1802.02538](https://arxiv.org/abs/1802.02538).