SOFIA UNIVERSITY “ST. KLIMENT OHRIDSKI”

FACULTY OF MATHEMATICS AND INFORMATICS
DEPARTMENT OF PROBABILITY, OPERATIONS RESEARCH AND STATISTICS

DISSERTATION FOR PH. D. DEGREE IN MATHEMATICS

Professional direction: 4.5. Mathematics
Doctoral programme: Probability Theory and Mathematical Statistics

Branching processes - optimization and applications

Kaloyan Nikolaev Vitanov, M. Sc.

Ph. D. supervisor: prof. Maroussia Slavtchova-Bojkova

SOFIA, BULGARIA
2022

Contents

CHAPTER 1

Introduction

1.1 Short history of branching processes

Branching processes are a class of stochastic processes that capture some of the fundamental aspects of division and propagation observed in nature. Branching processes model the evolution of a population of objects (these objects can correspond to real-world elementary particles, photons, electrons, atoms, molecules, cells, viruses, bacteria, animals, people, information, finances, and other entities) through time and study various characteristics of this evolution. It comes as no surprise that the areas of application of branching processes are diverse and numerous - physics, chemistry, biology, demography, ecology, economy, etc. This diversity of contexts stimulates the development of various kinds of branching processes, adapted towards answering particular questions of interest found within these contexts. Indeed, there are branching processes in discrete-time and continuous-time, with one or multiple types of objects, branching processes with random immigration, branching diffusion processes, spatial branching processes, controlled branching processes, and others. For examples of these processes, the interested reader is invited to scroll through the references within the dissertation.

The study of branching processes begins around the middle of the 19th century with the question of explaining the disappearance of aristocratic family lines in Europe. In 1845, the French mathematician and statistician Bienaymé first studied the process of extinction of the French noble families, [83], and created the first branching process model (for many more historical facts about Bienaymé and his contributions to mathematics, the interested reader is referred to [84]). Unfortunately, as Bienaymé left no students, his name, as well as his results, gradually faded into obscurity.

A few decades later, in 1873, concerns regarding the extinction of noble family names reappeared among the scientific community, as Francis Galton formulated Problem 4001 in *Educational Times*:

"...PROBLEM 4001: A large nation, of whom we will only concern ourselves with adult males, N in number, and who each bear separate surnames colonise a district. Their law of population is such that, in each generation, a_0 per cent of the adult males have no male children who reach adult life; a_1 have one such male child; a_2 have two; and so on up to a_5 who have five. Find (1) what proportion of their surnames will have become extinct after r generations; and (2) how many instances there will be of the surname being held by m persons..."

Upon receiving only one solution to the problem that "...was from a correspondent who wholly failed to perceive its intricacy, and his results were totally erroneous..." (see [85]), Galton invited H. W. Watson to work on Problem 4001. One year later, in 1874, Galton and Watson published their famous work on [85] that is considered by many to mark the beginning of

Mr. Galton then read the following paper by the Rev. H. W. Watson and himself:

***On the* PROBABILITY *of the* EXTINCTION *of* FAMILIES. By the Rev. H. W. WATSON. With PREFATORY REMARKS, by FRANCIS GALTON, F.R.S.**

Figure 1.1: The title of the article of Galton and Watson [85].

branching process theory.

Within [85], Watson determined the extinction probability as a fixed point of the reproduction generation function f. He observed that 1 is always such a fixed point, $f(1) = 1$, and from this he and Galton concluded that "...All the surnames, therefore, tend to extinction in an indefinite time, and this result might have been anticipated generally, for a surname lost can never be recovered, and there is an additional chance of loss in every successive generation. This result must not be confounded with that of the extinction of the male population; for in every binomial case where q is greater than 2 we have $t_1 + 2t_2 + \cdots + qt_q > 1$, and, therefore an indefinite increase of male population...". This conclusion, however, is false as it is now known that it is always the smallest root within $[0, 1]$ of the fixed point equation that gives the correct answer. It took more that 50 years for the correct solution to be published by the Danish mathematician J. F. Steffensen within [88] in 1930. It took 40 more years for Heyde and Seneta to first note in 1972 that Bienaymé already had the correct statement of the Criticality Theorem back in 1845 (see [84]). For more interesting details about the early years of branching process theory, see [86], [87].

After World War II, branching processes and their applications in physics became intensely researched, leading to the rapid development of the field. In the Soviet Union, the work of A. N. Kolmogorov and two of his students, B. A. Sevastyanov and N. Dmitriev, paved the way for branching processes with the seminal papers of [89] and [90]. Indeed, the term *branching process* itself is considered to have been coined by A. N. Kolmogorov and N. A. Dmitriev in their work [89] from 1947. Historically, the legacy of B. A. Sevastyanov is one of the most prominent ones within the field (at least with respect to the Soviet sphere of influence), see, e.g., monograph [8] and the references therein. According to [91], as early as his dissertation, Sevastyanov was already laying the foundation of modern branching process theory. At the time, the last two chapters of his dissertation were declared as classified and for five years (see [91]) he was not allowed to keep them, publish them, or discuss them with anyone but his supervisor, Kolmogorov. Developing the results of his dissertation, Sevastyanov eventually obtained fundamental results in almost all principal directions in the theory of branching processes: branching processes with immigration, general branching processes with arbitrary distribution of particle lifetimes, transition phenomena in branching processes, diffusion branching processes, controlled branching processes, and regularity conditions. Sevastyanov's monograph Branching Processes [8], published in 1971, accumulates the majority of the results of the theory of branching processes at the time. It was translated later on into Japanese and German.

We will also shed a bit more light on the historically more obscure person and work, at least for the time being, of Nikolai Aleksandrovich Dmitriev (see [92] for more interesting details). His grandfather Kostadin Popdimitrov Hadjikostov was a member of the regiment of Hristo Botev during the time of the April uprising in Bulgaria in 1876. Dmitriev showed his mathematical talent very early at the age 14 by winning one of the first mathematical olympiads in the USSR. He was later a Ph. D. student of

Figure 1.2: B. A. Sevastyanov

A. N. Kolmogorov in Steklov Mathematical Institute. Dmitriev played an important role in the creation of the atomic and hydrogen bombs, developing the theory of incomplete atomic explosion. He applied the theory of branching processes for modeling chain reactions with nuclear degradation, such as conditions under which energy release will decrease critically and degenerate ([93]). Due to the nature of his research, N. A. Dmitriev worked most of his life in secret scientific organizations within the Soviet Union. After his joint publication with Kolmogorov of the seminal paper [89], the name of Dmitriev disappears from the public scene, although it is now known that he has over 80 classified papers. Reportedly, N. A. Dmitriev had very high authority among his colleagues. In an instance when A. N. Kolmogorov was asked to assist in the the process of installing an electronic computer, he joked: "You do not need this computer. You have Kolya". Dmitriev also significantly contributed for the creation and development of impulse nuclear reactors, which are still being used.

Meanwhile in the west, mainly in the USA, the theory of branching processes advanced perhaps most notably due to the work of T. Harris (see monograph [58], published 1963) on electron-photon cascades and Galton-Watson processes with continuous type space (energy). In 1948, Bellman and Harris [94] considered branching processes with time structure and called their result an age-dependent branching process. Such processes are generally not Markovian and can model populations where individuals can have variable lifespans, but split into a random number of children at

Figure 1.3: N. A. Dmitriev

death, independently of age. Sevastyanov introduced truly age-dependent branching processes, where, the reproduction probabilities are possibly affected by the mother's age at splitting (see, e.g. [8]). Crump, Mode, and Jagers in [95], [96], [34] (1968 - 1969) introduced general branching process where members of the population can give birth repeatedly, in streams of events modeled by a point process. Another monograph of note is [57] from Athreya and Ney (1972).

In time, the areas of research and application of branching processes gradually shift more towards biology, demography, genetics, and related contexts - see [11], [59], [19], [105], [14], [13], [10], [61], [62], [64], [65], [63], [66], [20], [100], [104], [36] among many others. The immediate predecessors of this dissertation, [1] - [7], are also styled in the setting of biological populations escaping extinction, with the focus of [1] - [7] being on cancer modeling (see also [104]). In recent years controlled branching processes also enjoy notable development (see [115], [124], [181], [182], [120], [121]).

In Bulgaria, questions related to branching process theory were first considered in the textbooks of Obreshkov [97] and Obretenov [98] in 1963 and 1974 respectively. From then onward, Bulgarian authors have contributed much to the field, developing various areas of the theory and applications, see [19], [22], [23], [37], [18], [26], [24], [39], [183], [25], [32], [18], [26], [24], [39], [32], [12], [107], [139], [140], [141], [142], [143], [144],

[145], [48], [53], [54], [55], [56], [193], [194], [195], [149], [150], [151], [152], [153], [50], [51], [129], [1], [2], [100], among many others. It should come as no surprise that first The World Congress on Branching Processes was organized in Varna in 1993 by Bulgarian scientists with chairman of the organizing committee N. Yanev, the results of which were published in [101]. We also note the publishing of the textbook on branching processes by M. Slavtchova-Bojkova & N. Yanev [9] in 2018.

1.2 Review of branching process literature

There is an abundance of literature within the field of branching processes and their application in various areas.

We begin our review with the book from Haccou, Jagers, and Vatutin, [10], devoted to biological applications of branching processes. The book considers discrete-time branching processes and branching in continuous time. The discussed mathematical results are applied to models of large populations, development of populations and their extinction. Two monographs on the application of branching processes to biological problems are the books of Mode [60] (applications in demography), [33] (applications in epidemiology). The classic book of Jagers [11] provides much information about the stochastic models of populations. The last chapters of the book are devoted to application of branching process theory to problems in demography (growth of population, age of childbearing, length of generations), cell kinetics (estimation of cell death, cycle time distribution, etc.). Another monograph related to biology is [21]. There, the theory of branching processes is applied to the problem of receptor clustering in transmembrane signaling. In addition to that, branching process theory is applied to the phenomenon of antigen-antibody interactions and to aggregate size distributions on cell surface. For more examples of application of branching stochastic processes as models of population dynamics of particles having different nature (from elementary particles to cells, microorganisms, plants, animals, information, individuals, etc.) see [19], [35], [111], [112], [22], [23], [37]. The book of Kimmel and Axelrod [14] presents applications of branching processes theory in the areas of molecular biology, cellular biology, human evolution and medicine. Another book devoted to biological application of branching processes is [104]. There, Durrett uses multi-type branching processes to model cancer. The results on the branching processes are applied to model metastasis, ovarian cancer, and tumor heterogeneity. The monograph of [105] presents the application of single-type branching processes to a particular area of mathematical genetics: neutral evolution. There are numerous interesting articles articles discussing application of branching processes in biology and other fields of science and practice. In [22] there is an application of a Bellman-Harris process to biological systems with cells proliferation. This research was continued in [18], [26], [24], [39], [183], [25], [32], [12]. Interesting applications in epidemiology can be seen in [27], [28], [29], [30], [31], [198]. Many other discussions can easily be found: cancer modeling [40], populations escaping extinction [61], [62], [64], [65], [63], [5], [7], sand avalanches [184], biodiversity [185], analysis of heights of trees [186], evolutionary rescue [188], population dynamics [187], information cascades [189], quantum electronics [190], polymer degradation [191], etc. Several topics connected to multi-type branching processes are: epidemiological risk estimation [172], disease outbreak [173], genetics [15], mast cells modelling [16], polynucleotide evolution [174], forest fires [175], population dynamics [38], models of transposable elements in haploid populations [17], diffusion [176], spreading dynamics of populations [177], queuing [170], optimization [77], identification of multi-type branching processes [171]. Results on the behaviour of the multi-type branching processes in the supercritical case can be found in [57], [165], [166], [167], [168], [169]. For application of branching processes in economics, we mention the Epps model and its generalizations [192], [193], [194], [195], as well as interest rate modeling [196].

A contemporary book devoted to classical and modern branching processes is [106]. Some of the topics covered within are: (i) Tree structures and branching processes; (ii) Branching random walks; (iii) Measure valued branching processes; (iv) Branching with dependence; (v) Large deviations in processes; (vi) Classical branching processes. This book also enjoys Bulgarian contribution [107]. Other books which are more orientated to the mathematical aspects of the theory of branching processes are [108], [35], [57].

An application of the theory of branching processes in physics in presented in the book [110]. The mathematical tools used in describing branching processes are used to derive a large number of properties of neutron distribution in multiplying systems with or without an external source. Then, the theory is applied to the description of the neutron fluctuations in nuclear reactor cores as well as in small samples of fissile material.

In the monograph [109], the authors discuss branching processes in varying and random environments. Within branching processes in random environment model, an additional environmental stochasticity is incorporated and consequently the conditions for reproduction may vary in a random fashion from one generation to the next. Single and multi-type BPREs are discussed. Various authors have studied the probability for extinction for branching processes in random environments. One result is that if the random environmental sequence of the process is stationary and ergodic then the probability for extinction has values 0 or 1 [178], [179], [180].

Another interesting class of branching processes are the regenerative branching processes [46], [47]. These processes are Bienaymé-Galton-Watson (BGW) processes with state-dependent immigration. Processes of this type allow immigration of new particles only in state zero, which means that the population regenerates when it becomes extinct. The state zero is no longer an absorbing state and becomes a reflecting barrier instead. The process evolution consists of a sequence of cycles which are independent and stochastically equivalent. The model is generalized for Bellman-Harris processes [48], [49], [50], [51]. The case of non-homogeneous migration was studied in, [53], [54], [55], [56].

Limiting distributions for BGW processes with an increasing random number of ancestors are studied in [139], [140], [141], [142], [143], [144], [145].

The problem of extreme values arising in branching processes has numerous practical applications. Selected results on the study of extremes in branching processes are reviewed in [158]. The maximum family size of BGW processes is studied in [159], [160]. Maxima related to the offspring size for different classes of branching processes are discussed in [161], [162], [163], [164].

Branching processes can also be considered in the context of spatial evolution. For example, particles born in a given generation can be distributed in R^d space according to a given locally-bounded random measure, independently of the position of the mother-particle [154]. Results (in German) for such a setting can be seen in [155], [156], [157].

Another class of branching processes of note is the class of branching processes with diffusion [149], [150], [151], [152], [153].

There are numerous studies on statistical problems connected to branching processes [41], [42], [43], [44], [138], [45]. For example [43] discusses problems of statistical inference that occur when the standard assumption of independent observations is relaxed. There is much literature on the statistical problem of estimating the mean m and other parameters of a supercritical branching process with and without immigration. Selected references are [43], [113], [114]. Other papers on various topics relates to branching process estimation are [134], [135], [136], [137], [146], [147], [148].

Age-dependent branching processes are studied in [8], [50], [51]. Such processes have also been studied in the context of regeneration [126], [127]. Processes with two types of immigration are investigated in [129], [130], [131]. Sevastyanov branching processes with general immigration are considered in [132], [133]. There, asymptotic results are obtained for the moments of the process and the limiting distributions in all three cases: subcritical, critical, and supercritical.

Controlled branching processes (CBPs) are integer-valued discrete-time Markov processes where the population size in every generation can be randomly regulated before reproduction by emigration of part of the population, or after reproduction by immigration of individuals. Migration can be non-homogeneous [123], [124], [125]. Controlled branching processes with random migration can be seen in [119]. Controlled branching processes with multi-type random control functions are investigated in [181], [182]. Results on the asymptotic behavior of the probability of extinction are present in [117], [118]. Random control functions can be seen in [116]. A contemporary monograph on controlled branching processes is [120]. Recently, continuous-time controlled branching processes have been introduced in [121].

We conclude with quick notes about the evolution of the research preceding this dissertation ([1] - [7]):

1. In [1] (2016), M. Slavtchova-Bojkova considers a multi-type Bellman-Harris decomposable branching process consisting of two particle types, Type 0 is supercritical and Type 1 is subcritical. The basic functional equation is obtained. The paper then proceeds to study the probability generating function for the random variable describing the number of mutations towards the supercritical Type 0, the probability of extinction of the process, the time to escaping extinction, and the immediate risk of escaping extinction.

2. In [2] (2017), M. Slavtchova-Bojkova, P. Trayanov, and S. Dimitrov expand the context of [1], obtain new results, and introduce a numer-

ical scheme for calculating the integral equations derived. Further, a simulation study is done.

3. In [3] (2017), K. Vitanov and M. Slavtchova-Bojkova introduce a more general model. That is the multi-type Bellman-Harris decomposable branching process with an arbitrary number of subcritical types that can produce mutants towards the single supercritical Type, Type 0. No backward mutation from Type 0 is allowed. The basic functional equations are obtained and the p.g.f.s for the random variables describing the number of mutations towards the supercritical Type 0 are studied.

4. In [5] (2019) M. Slavtchova-Bojkova and K. Vitanov expand the results from [3] by adding results for the probabilities of extinction of the process, the time until occurrence of the first mutant towards Type 0 that initiates a non-extinction process, and the immediate risk of escaping extinction.

5. In [6] (2019) K. Vitanov and M. Slavtchova-Bojkova expand the numerical scheme from [2] for the integral equations obtained for the p.g.f.s of the random variables describing the number of mutations towards the supercritical Type 0 found in [3] and the equations for the probabilities of extinction found and [5].

6. In [7] (2022) the model is further expanded with the introduction of two classes of particle types $\mathbb{W}_e$ and $\mathbb{W}_0$ where particles with types from $\mathbb{W}_e$ can produce mutants with types from $\mathbb{W}_0$, however particles with types from $\mathbb{W}_0$ cannot have offspring with types from $\mathbb{W}_e$. The model from [5] is a particular case in that regard, with $\mathbb{W}_0$ consisting only of Type 0. A further extension is the introduction of dependence of particle reproduction from particle age, i.e., the branching processes considered are no longer multi-type Bellman-Harris, but multi-type Sevastyanov. Within [7], systems of integral equations for the p.g.f.s of the process are obtained, followed by results for the probabilities of extinction of the process, the p.g.f.s for particle production from $\mathbb{W}_e$ towards $\mathbb{W}_0$, the time until occurrence of the first mutant from $\mathbb{W}_e$ towards $\mathbb{W}_0$, and the immediate risk of escaping extinction. A numerical scheme, extending the one from [6], for computing all obtained system of equations is provided as well.

1.3 Notes on multi-type Sevastyanov branching processes

We present some of the work within the classical monograph of B. A. Sevastyanov, [8], that is of immediate interest for our investigation of the novel Multi–type Sevastyanov Branching Processes through probabilities of Mutation between types (MSBPM) that we define in Chapter 2. Within this Section, we follow Chapter VIII of [8] and use (most of) the notation from there. As [8] is written in Russian, our translation below has some minor adaptations. In addition to that, we have taken the liberty of modestly reordering parts of the exposition, without losing content. Note that we also use **bold** when denoting vectors.

1.3.1 Classical definition of the multi-type Sevastyanov branching process

We consider n types of particles : $T_1, T_2, \ldots, T_n$. This system of particles is under the following assumptions:

1. We assume the process starts at $t = 0$ with one particle of type T_i of age 0.
2. Each particle of the type T_i has random lifespan τ^i with probability distribution $\mathbb{P}\{\tau^i \leq t\} = G^i(t)$ where $G^i(-0) = 0$ and $G^i(+0) < 1$.
3. At the end of its life, each particle converts to a set (this set can contain 0 or more particles) consisting of α_1 particles of type $T_1, \ldots,$ α_n particles of type T_n, and these particles have age 0 at the moment of conversion.
4. $p^i_{\boldsymbol{\alpha}}(u)$ is the conditional probability that a type T_i particle converts into $\boldsymbol{\alpha}$ particles, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)^\top \in \mathbb{N}^n$ (the conditions being that the particle has age u at the time of conversion and that the conversion actually happens).
5. Each of the new particles evolves as described above.
6. The evolution of a particle of the type T_i is determined by the joint distribution of its lifespan τ^i and of the vector quantity $\boldsymbol{v}^i = (v^i_1, \ldots, v^i_n)^\top$ which describes the offspring of the particle. The joint

distribution is given by

$$\mathbb{P}\{\tau^i \in B, \boldsymbol{v}^i = \boldsymbol{\alpha}\} = \int\limits_B \boldsymbol{p}^i_{\boldsymbol{\alpha}}(u) dG^i(u). \tag{1.1}$$

In (1.1) $\boldsymbol{\alpha} \in \mathbb{N}^n$ and B is a Borel set on the straight line.

7. If some generation is empty (no particles there) then all following generations are empty (the process has become extinct/degenerate).
8. The basic assumption for the process is that the particles have independent evolution: the evolution of a particle within a generation is independent form the evolution of other particles within the same generation and the conditional distribution of the evolution of all particles of a generation depends only on the composition of this generation, not on given evolution of previous generations. If the composition of the generation is $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_n)^\top$ then the distribution of the evolution of all particles of this generation is the set of the independent distributions $P\{\tau^i \in B, \boldsymbol{v}^i = \boldsymbol{\alpha}\}$ given by (1.1).

We introduce probability generating functions (p.g.f.s)

$$h^i(u; \boldsymbol{s}) = \sum_{\boldsymbol{\alpha}} p^i_{\boldsymbol{\alpha}}(u) \boldsymbol{s}^{\boldsymbol{\alpha}}. \tag{1.2}$$

We can organize $h^i(u; \boldsymbol{s})$ per each type into the vector:

$$\boldsymbol{h}(u; \boldsymbol{s}) = \Big(h^1(u; \boldsymbol{s}), \ldots, h^n(u; \boldsymbol{s})\Big)^\top \tag{1.3}$$

We further introduce p.g.fs

$$F^i(t; \boldsymbol{s}) = \sum_{\boldsymbol{\alpha}} \mathbb{P}\{\boldsymbol{\mu}^i(t) = \boldsymbol{\alpha}\} \boldsymbol{s}^{\boldsymbol{\alpha}}, \quad i = 1, \ldots, n, \tag{1.4}$$

where $\boldsymbol{\mu}^i(t) = \big(\mu^i_1(t), \ldots, \mu^i_n(t)\big)^\top$ means that there are $\mu^i_j(t)$ particles of the type T_j at moment t. The superscript i indicates that we are considering a process that started at 0 with one particle of type T_i with age 0. We can organize $F^i(t; \boldsymbol{s})$ per each type into the vector:

$$\boldsymbol{F}(t; \boldsymbol{s}) = \Big(F^1(t; \boldsymbol{s}), \ldots, F^n(t; \boldsymbol{s})\Big)^\top. \tag{1.5}$$

We assume that the probabilities $p^i_{\boldsymbol{\alpha}}(u)$ are measurable with respect to u and satisfy the condition $\sum_{\boldsymbol{\alpha}\in\mathbb{N}^n} p^i_{\boldsymbol{\alpha}}(u) = 1$. Then, $h^i(u;\boldsymbol{s})$ are measurable with respect to u and are generation functions with respect to $\boldsymbol{s} = (s^1,\dots,s^n)^\top$.

1.3.2 System of integral equations for the probability generating functions of the process

We present without proof:

Theorem 1.1. *The generating functions $F^i(t;\boldsymbol{s})$ satisfy the following system of non-linear integral equations*

$$F^i(t;\boldsymbol{s}) = \int_0^t h^i\big(u;\boldsymbol{F}(t-u;\boldsymbol{s})\big)dG^i(u) + s^i\big(1-G^i(t)\big),\ i=1,\dots,n, \tag{1.6}$$

where $t \geq 0$ and $|\boldsymbol{s}| \leq 1$.

Let S_T, $0 < T < \infty$ be the complete metric space of vector functions $g(t;\boldsymbol{s}) = \Big(g^1(t;\boldsymbol{s}),\dots,g^n(t;\boldsymbol{s})\Big)^\top$, $\boldsymbol{s} = (s^1,\dots,s^n)^\top$, with components $g^i(t;\boldsymbol{s})$ defined in $0 \leq t \leq T$ with $|\boldsymbol{s}| \leq 1$ and measurable with respect to t. Let $g(t;\boldsymbol{s})$ represent multidimensional probability generating functions with respect to $\boldsymbol{s}$. Further, denote

$$a^i_j(u) = h^i_j(u;\mathbf{1}) = \frac{\partial h^i(u;\boldsymbol{s})}{\partial s^j}\,|_{\boldsymbol{s}=\mathbf{1}}, \quad A^i_j = \int_0^\infty a^i_j(u)dG^i(u) \tag{1.7}$$

Then the following theorem is in effect:

Theorem 1.2. *If $G^i(0) = 0$, $i = 1,\dots,n$, and A^i_j, $i,j = 1,\dots,n$, are finite, then the system of integral equations (1.6) has, at $|\boldsymbol{s}| \leq 1$, a unique solution $F(t;\boldsymbol{s})$, which for $0 \leq t \leq T$, belongs to S_T for all $T > 0$.*

1.3.3 Probabilities of extinction

We can set $\boldsymbol{s} = \mathbf{0}$ in the generating functions $F^i(t;\boldsymbol{s})$ from (1.4) and by doing so obtain the probabilities $P^i(t) = F^i(t;\mathbf{0})$ for the event that at the moment of time t there are no particles within the process. We can define probabilities P^i for degeneration (extinction) of the process as

$$P^i = \lim_{t\to\infty} P^i(t),\ \boldsymbol{P} = (P^1, \dots, P^n)^\top. \tag{1.8}$$

We say that the process eventually *degenerates* (becomes *extinct*) if all $P^i = 1$.

where $\boldsymbol{h}(s) = \Big(h^1(\boldsymbol{s}), \dots, h^n(\boldsymbol{s})\Big)^\top$ and

$$h^i(\boldsymbol{s}) = \int_0^\infty h^i(u; \boldsymbol{s}) dG^i(u). \tag{1.9}$$

Theorem 1.3. *The probabilities for extinction P^i satisfy the system of equations*

$$\boldsymbol{s} = \boldsymbol{h}(\boldsymbol{s}), \tag{1.10}$$

and are equal to the coordinates of the root of (1.10) which is closest to the origin and is in within the cube $\mathbf{0} \le \boldsymbol{s} \le \mathbf{1}$.

Before proceeding further, we note that we have given the Perron-Frobenius theorem as Theorem 4 as well as the definition of a Perron root via Definition 2 within the Appendix. These elements can also be found in Section 5 of Chapter IV within [8] in Russian.

Theorem 1.4. *Assume that each particle type has non-zero probability for producing* 0 *in its offspring. The process degenerates if and only if the Perron root R of the matrix $\|\ A^i_j\ \|$ is not larger than 1.*

Remark 1.1. *Within the original Theorem 2 from page 238 in [8], there is no assumptions that each particle can have an empty offspring with non-zero probability. If the assumption is dropped, then we can have classes of particle types that have probabilities of extinction* 0 *(those are the "final classes" mentioned within the original Theorem 2, final classes are defined on page 137 of [8]) which is a case that we are not interested in within the dissertation.*

The Perron root R determines the criticality of the process:

Definition 1.1. *We say that the process is subcritical if $R < 1$, is critical if $R = 1$, and is supercritical if $R > 1$.*

1.4 A remark on the use of "particles" and "mutation" within the dissertation

As already noted, branching process models can be applied in many, vastly different, contexts. Hence, when devising general models (such as the models from Chapter 2), it makes sense to choose the key terms used carefully.

Throughout the dissertation, we will refer to the objects being modeled through branching processes as *particles*. We prefer to use "particles" instead of "objects", because "objects" is a too general term that may tempt us to include in it entities that are not suitable for modeling via branching processes. Although we could use "cells" instead of "particles", the term "cells" fixates our attention too much on biological contexts. "Cells" is beneficial, for example, when modeling cancer evolution and escape from extinction, however the repeated use of "cells" may conceal the fact that branching process used for modeling could be applicable outside of biology. We find the term "particle" as having an appropriate balance. Indeed, we can interpret "cells" as biological "particles". "Particles" ensures that we will think carefully before trying to apply our branching process models to entities such as animals and humans, which may not necessarily conform in their behavior to the propagation assumptions of the model. With respect to "objects", "particles" allows us to have a more clear perception of what we are attempting to model.

We will also be using the term *mutation* in order to indicate that a particle within the offspring of a type i particle is of type j, $i \neq j$. Unfortunately "mutation" draws us again closer to biological contexts, however alternatives such as "alteration", "modification", "transformation", introduce more confusion than clarity. Further, unlike "mutation", the alternatives are not in line with the already biology-compatible terminology commonly used within branching processes (e.g., the usage of terms such as "offspring", "progeny", "mother"/"daughter" particle, are common). We will be using "mutation" rather loosely. As an example, consider an animal population (our "particles" will be individual animals - the use of "particles" encourages us to be careful about what we are trying to model) spread across multiple geographical locations, with conditions in these locations influencing the reproductive capabilities of animals, thus inducing types. Within this dummy example, we will consider as "mutation" the relocation of the offspring of an animal, upon its creation, from one location to another.

1.5 Notes on stochastic sequential decision problems

Within Chapter 3 of the dissertation, we will be investigating stochastic sequential decision problems in the context of systems with underlying branching process dynamics. Here, we give an informal description of what a "sequential decision problem" is.

Description 1.1. Assume that there is a dynamic system with which we can interact. We observe the system at specified moments in time, called *decision epochs*. We will denote the set of decision epochs with $\hat{t}_T$, $\hat{t}_T = \{0 = t_0, t_1, t_2, \ldots, t_T = \hat{T}\}$, the distance between two neighboring decision epochs can vary but cannot be 0 or ∞. At each decision epoch $t_i \in \hat{t}_T$, except at $t_T = \hat{T}$, we make a decision (interaction with the system) that affects how the system evolves from t onward. Upon making a decision, we collect rewards (or incur costs), with the exception of $t_T = \hat{T}$ where we only collect predefined rewards (or incur predefined costs). A *sequential decision problem* is a problem of choosing such decisions so that the cumulative expected reward, after collecting the rewards at $t_T = \hat{T}$, is maximized (or the cumulative expected cost is minimized). The problem can be deterministic or stochastic.

The informal Description 1.1 serves the needs of this dissertation well, however, slight variations of the concept can exists in different mathematical communities. We formalize the idea in Definition 3.2 within Chapter 3.

Evidently, the benefit of solving a sequential problem is obtaining the "best" course of action to be applied at the current moment. This course of action, however, is "best" only in the context of the, also contained in the solution, "best" courses of action at future decision epochs. Thus, the overarching benefit of solving sequential decision problems is directly connected to the adequacy of our understanding of the dynamics of the system we are trying to model. In the case of us having a poor understanding, solutions obtained from sequential decision problems can prove to be worse compared to the strategy of taking the immediately best course of action at each decision epoch without regard of possible consequences in the future.

The real world is abundant with decision related situations that can be cast as sequential decision problems. Sequential decision problems can be easily found in industry, management, finance, and government planning, among various possible contexts. We list some concrete examples:

1. Dynamic assignment - consider a limited resource such as engineers or programmers with different qualifications as well as a series of possible tasks that need to be completed with varying priority. How should our limited resources be distributed in time given the currently pending tasks and our expectations for the occurrence of future tasks?
2. Storage problems - how much stock should we buy at today's prices so that we can meet random demand in the coming days, should we delay a part of our purchase for another day, when the conditions are possibly better, or should we stockpile additional reserves today?
3. Finance - how do we rebalance our portfolio today, considering possible further rebalancing and anticipating information at predefined future moments, so that we maximize the chance of it achieving a desired goal at the end of a desire time period?
4. Government policy - How the government should plan subsidizing agricultural production each year, given observed weather patterns and anticipating global market prices?

Mathematically approaching the topic of optimization in a stochastic setting is a complicated matter and over the years a myriad of mathematical communities have evolved, providing various perspectives and solution methods. [82] (page 10), as well as [80] and [81], identify at least 15 distinct communities that deal with some variant of deterministic or stochastic sequential decision problems. We list these communities below.

1. Derivative-based stochastic search
2. Derivative-free stochastic search
3. Decision trees
4. Markov decision processes
5. Optimal control
6. Approximate dynamic programming
7. Reinforcement learning
8. Optimal stopping
9. Stochastic programming

10. Multi-armed bandit problem
11. Simulation optimization
12. Active learning
13. Chance constrained programming
14. Model predictive control
15. Robust optimization

Unfortunately, these communities are not unified in their approach to solving sequential decision problems. According to [82] (page 8), the 15 communities listed use "roughly eight notational systems" and also "...The fragmentation of the communities (and their differing notational systems) disguises common approaches developed in different areas of practice, and challenges cross-fertilization of ideas. A problem that starts off simple (like the inventory problem) lends itself to a particular solution strategy, such as dynamic programming. As the problem grows in realism (and complexity), the original technique will no longer work, and we need to look to other communities to find a suitable method...". This state of affairs is not so surprising given the complexity (and possibly scale) of the problems being investigated and also the richness of possible approaches which may serve one setting excellently but fail in others.

A brief overview of the problem formulations tackled by each of the communities listed above, as well as some of the basic notation they use, can be found in Chapter 2 of [82] and in [80]. In order to keep our presentation manageable, we will briefly remark here that the discussion in Chapter 3 of the dissertation is most closely related to Markov decision processes ([67], [68], [69], [70], [71]), Optimal control ([204], [205], [206], [207], [208]), and Approximate dynamic programming ([72], [73], [74], [75], [76], [78]).

1.6 Conceptual organization of the dissertation

This dissertation is conceptually divided into two topics explored in Chapter 2 and Chapter 3 respectively. Chapter 2 defines the novel Multi–type Sevastyanov Branching Processes through probabilities of Mutation between types (MSBPM) and obtains results of interest in the context of populations escaping extinction. Chapter 3 is dedicated towards the incorporation of branching processes, including the MSBPM, into optimization problems known as Sequential Decision Problems (SDPs).

The novel MSBPM from Chapter 2 is connected to the classic multi-type Sevastyanov branching process, however, within the MSBPM, probabilities for mutation are used for writing down expressions of interest. This makes the novel MSBPM well adapted towards modeling biological populations under stress that escape extinction. Within Chapter 2, we obtain various systems of equations for the MSBPM as well as for quantities relevant in the context of populations escaping extinction. To the best of our knowledge, such an in-depth investigation of the topic has not been done previously (excluding our earlier work in [7] as well as preceding papers [1] - [6]) for multi-type, continuous-time branching processes. We explore the case of the MSBPM starting with one particle of age 0 and the case of the MSBPM starting with one particle of age a, $a \neq 0$. The latter case, to the best of our knowledge, has not been explored previously in a systematic manner within the context of branching processes. Numerical schemes for calculating all obtained systems of equations are also developed within Chapter 2.

In Chapter 3, we begin with an introduction of the "Universal Modeling Framework" developed by Warren B. Powell in [82] (2022). The choice of modeling framework within which we specify our Sequential Decision Problems (SDPs) is of paramount importance for our perspective on the systems we attempt to model as well as for the ease of possible future extensions of our results. Our choice of framework allows us to utilize Bellman's optimality equation for finding solutions of SDPs, provided that our models conform to the assumptions of the framework. We proceed with our novel considerations and results as follows. We recast the multi-type Bienaymé-Galton-Watson (BGW) branching process optimization problem, considered in [77], as a SDP within the "Universal Modeling Framework". In Theorem 3.2 from Subsection 3.4.3, we provide a novel proof for Theorem 3.1 from [77] that is based on Bellman's optimality equation. Theorem 3.2 enables us to efficiently find the solution of SDPs with underlying BGW dynamics. Next, we incorporate the multi-type Bellman-Harris branching process with exponential lifespan distributions as well as the Multi-type Bellman-Harris Branching Process through probabilities of Mutation between types (MBHBPM; a particular case of the MSBPM) with exponential lifespan distributions into a SDP and prove that a result, similar to

Theorem 3.2, holds. We then shown that, with respect to a novel state space, the MSBPM and the multi-type Sevastyanov branching process can also be incorporated into SDPs. Unfortunately, an analogue of Theorem 3.2 is not available for these processes. Regardless, Bellman's optimality equation allows us to consider the Approximate Dynamic Programming (ADP) approach for finding the solution of obtained SDPs within future research. We conclude our investigations by outlining a general ADP algorithm based on post-decision state variables that may serve as a starting point for the future development of a specialized ADP algorithm for SDPs with branching process based dynamics.

Within the Appendix, we have provided some standard results regarding the Perron-Frobenius theorem. We reference these results in some of our MSBPM related discussions.

More detailed description of the structure of Chapter 2 and Chapter 3, as well as relevant remarks and discussions, can be seen in the corresponding "Chapter overview and organization" sections within these chapters.

CHAPTER 2

Multi-type continuous-time branching processes through probabilities of mutation between types

2.1 Chapter overview and organization

In this Chapter, we define the novel continuous time branching process model that will play the central role within this dissertation - the Multi–type Sevastyanov Branching Processes through probabilities of Mutation between types (MSBPM). The MSBPM can be considered as a relative of the classical multi-type Sevastyanov branching process as defined in Chapter VIII in [8]. The novel characteristic of the MSBPM, with respect to the classical formulation in [8], is the use of probabilities of mutation (a particle is a "mutant" if it is of type that is different from the type of its mother particle) between types. More specifically, through the use of probabilities of mutation, effectively, we decompose the classical probabilities $p^i_{\boldsymbol{\alpha}}(u)$ for a particle of type i of age u to transform into $\boldsymbol{\alpha}$ particles at the end of its lifespan (see page 229 in [8] or Subsection 1.3.1) into two components: 1) Probabilities $p_{ik}(u)$ for the total k number of offspring, regardless of offspring type, of a type i particle of age u; 2) Probabilities for mutation of an offspring particle of a type i particle towards type j, u_{ij}.

The use of probabilities for mutation opens the way for applications of the MSBPM into many biological contexts. Most notably, the MSBPM is well suited for modeling biological populations under stress that face certain extinction unless a "beneficial" mutation occurs (or a combination of mutations occur), leading to supercritical behavior. Such situations are of interest in the areas of cancer modeling and treatment, spread of viruses, vaccination campaigns, control over agricultural pests and others (see, e.g., [61], [62], [63], [64], [65], [66], [104], [1] - [7]). In biological contexts it is easier to estimate the probabilities for the total number of offspring, $p_{ik}(u)$, and the probabilities of mutation, u_{ij}, found within the MSBPM, than the more abstract $p^i_{\boldsymbol{\alpha}}(u)$ used in the multi-type Sevastyanov branching process. The use of $p_{ik}(u)$ and u_{ij} often provides us with a model with more clear and straightforward interpretations.

Within this Chapter, we concentrate our efforts towards obtaining results for the MSBPM regarding quantities that are of interest in the context of populations escaping extinction. Other authors, see [64], [65], [66], discuss similar topics to the ones explored within the dissertation, however, their discussions is based on multi-type Bienaymé-Galton-Watson (BGW) processes. The BGW is a discrete time branching process while the MSBPM is in continuous time - a more difficult theoretical setting. [61], [62], [64], [65], explore populations escaping extinction under the assumption that probabilities of mutation are small quantities. Such an assumption is not made for obtaining the results for the MSBPM, making the model more general in terms of possible applications. [61] and [62] assume Poisson and geometric offspring distributions for obtaining their results. The results obtained for the MSBPM do not rely on particular assumptions of offspring distributions. [64], [65] consider only one supercritical type, in contrast the MSBPM can accommodate an arbitrary number of supercritical types, in addition to that almost all of our results do not depend on type criticality. Further, almost all results obtained for the MSBPM do not rely on an assumption about the process being non-decomposable, this assumption being central for many results valid for the classical multi-type Sevastyanov branching process. We also note that our results for the MSBPM do not rely on particular assumptions about the lifespan distributions among types. All mentioned features of the MSBPM and the results obtained within the current Chapter, highlight the flexibility and wide area of applicability of the MSBPM in modeling populations escaping extinction. However, the MSBPM is not to be understood as exclusively tied to biology - the model can be applied in other areas as well, provided a proper interpretation of u_{ij}.

This Chapter is a continuation and generalization of our previous work in [1] - [7] where the focus is on cancer modeling as well as modeling escape from extinction. More specifically, with respect to our latest work in Vitanov & Slavtchova-Bojkova [7] (2022), the MSBPM provides an extension in the following directions: 1) The MSBPM can be non-decomposable; 2) The "emitting" class and the target class can intersect. The process discussed in [7] is a particular case of the MSBPM, that is, the decomposable MSBPM (DMSBPM) explored in Subsection 2.3.1 of this dissertation. The DMSBPM is of particular interest for modeling mutation as it describes an irreversible path in the evolution of a population of particles. We note that the development of cancer resistance towards medical treatment in many situations can be attributed to biological mutations. Thus, the MSBPM and its decomposable variants rise as well suited candidates for modeling the risk of cancer reemerging even when an apparently successful treatment is applied. We note that the work in the following Chapter 3 can be considered as a further continuation of [1] - [7]. A real-world example of a sequential decision problem (we discuss these problems in Chapter 3) is the planning of cancer treatment administration throughout time with respect to cost and benefit considerations. While the results from Chapter 3 are not yet ready for handling the nuances of this particular example, Chapter 3 is to be understood as a step towards solving such problems.

Within this Chapter, we obtain systems of integral equations for the probability generating functions (p.g.f.s) of the MSBPM as well as for the probabilities of extinction within the MSBPM. We obtain p.g.f.s for the production of particles from one class of particle types within the process to another. For the general case the particles produced need not necessarily be mutants, however, for particular cases of interest of the MSBPM, such as the decomposable MSBPM (DMSBPM), we investigate only the production of mutants. The DMSBPM can be used to model, for example, a "beneficial" mutation that is reachable only after certain preceding mutations have occurred (see Figure 2.12 is Subsection 2.3.1.1 and Figure 2 within [61]) or other relevant mutation schemes (Figure 2.11 on page 83). We obtain the distribution of the random variable "time until first 'successful' particle"/"time until first 'successful' mutant" within the MSBPM, i.e., the time until the occurrence of a particle/mutant that initiates a non-extincting process. We also obtain expressions for the hazard function that is with respect to the occurrence of the first "successful" particle/mutant. We stress that the proofs of the results within this Chapter do not depend on assumptions whether a process is decomposable or not (although some statements are valid only for a decomposable process), decomposability is treated as a particular case where a particular set of probabilities of mutation contains only zeros.

In addition to obtaining results for the case where the MSBPM starts with one particle of age 0, we also obtain novel variants of our results for the case where the process starts with one particle of age a, $a \neq 0$. We have so far not detected other authors that consider initial particles with non-zero age. As can be seen from the various figures within the Chapter, significant difference in the behavior of the investigated quantities can be observed when we look at those t that are close to the beginning of the process. This observation has the potential to be very useful with respect future research stemming from Chapter 3, where optimization problems related to decision making are investigated in the context of branching processes.

All results obtained within the Chapter can be computed with the help of the novel Numerical Scheme 1 and Numerical Scheme 2 constructed in Subsection 2.2.7.

This Chapter is organized as follows. In Section 2.2, we define the Multi–type Sevastyanov Branching Processes through probabilities of Mutation between types (MSBPM) and obtain various systems of equations for quantities of interest in the context of populations escaping extinction. More specifically, in Subsection 2.2.1 we define the MSBPM. We then obtain the system of integral equations for the probability generating functions (p.g.f.s) of the process in Subsection 2.2.2 as well as results the probabilities for extinction within Subsection 2.2.3. Next, in Subsection 2.2.4 we investigate the p.g.f.s for the number of particles produced within the process from a class of particle types towards all types within the process. We then continue with results concerning the occurrence of the first "successful" particle produced from any type within a class of types within the MSBPM in Subsection 2.2.5. In Subsection 2.2.6 we obtain expressions for the hazard function defined with respect to the occurrence of the first "successful" particle. In Subsection 2.2.7 we provide two numerical schemes that can be used for computing obtained systems of equations throughout Chapter 2. We finish this Section with specifications of the example MSBPMs that we use, in conjunction with the constructed numerical schemes, for demonstrating results obtained within the Chapter. In Section 2.3, we investigate two particular cases of the MSBPM. In Subsection 2.3.1, we consider the decomposable MSBPM (DMSBPM). Within the Subsection, we obtain variants of the novel results for the MSBPM that are valid for the DMSBPM and also explore some additional results that stem from the enforced decomposability. In Subsection 2.3.2, we consider the particular case of the DMSBPM where there is no dependence of particle reproduction from particle age. We write down the results for this important case as corollaries of the results obtained previously.

2.2 Multi–type Sevastyanov Branching Processes through probabilities of Mutation between types (MSBPM)

The theoretical journey preceding the formulation of the Multi–type Sevastyanov Branching Processes through probabilities of Mutation between types (MSBPM), defined in Subsection 2.2.1, can be traced with the gradual expansion of the continuous-time model with two types discussed in [1] (2016) via the extensions from [2] (2017), [3] (2017), [5] (2019), [6] (2019), and [7] (2022). Notes about the contents of these papers can be seen at the end of Section 1.2 in the Introduction. We highlighted the novelty and features of the MSBPM as well as the relation of the process to previous work from other authors in Section 2.1. The current Section contains novel, yet unpublished, results that extend our recent publication Vitanov & Slavtchova-Bojkova [7] (March 2022).

In what follows, we will be providing figures that illustrate the results we obtain. The formal specification of the example processes, to which these figures correspond, is given in Subsection 2.2.8. For the purpose of a more fluid exposition, some of the figures are given before Subsection 2.2.8. All computations within the dissertation are done via code written in Python 3.8.13 [209]. The code uses the NumPy 1.20.3 [210] and SciPy 1.6.2 [211] libraries. Figures, that do not contain nodes, are created with Matplotlib 3.5.1 [212]. Figures that contain nodes are created with yEd 3.20.1 [213].

2.2.1 Notation and definition of the MSBPM

We begin with the introduction of some of the notation and prerequisites that we extensively use throughout the dissertation:

1. Let $\mathbb{W} = \{1, 2, \ldots, n\}$. $\mathbb{W}$ denotes the set of possible particle types.

2. Denote $\boldsymbol{\delta}^i = \left(\delta_1^i, \ldots, \delta_n^i\right)^\top$, where $\delta_j^i = 0$ if $i \neq j$ and $\delta_j^i = 1$ if $i = j$. We will use $\boldsymbol{\delta}^i$ to specify a single initial particle of type i that is of age 0. For a single initial particle of type i that is of age a, $a \neq 0$, we will use $\boldsymbol{\delta}_a^i$. Again, we set $\boldsymbol{\delta}_a^i = \left(\delta_1^i, \ldots, \delta_n^i\right)^\top$ with $\delta_j^i = 0$ if $i \neq j$ and $\delta_j^i = 1$ if $i = j$, however, the subscript "a" in $\boldsymbol{\delta}_a^i$ now specifies the age of the initial particle.

3. We will be denoting the lifespan cumulative distribution function (c.d.f.) at t for type i particles, of age 0, with $G_i(t)$. If a type i particle is of age a, we will denote the corresponding c.d.f., conditioned on the age of the particle, with $G_{i,a}(t)$.

4. If X is some random variable (r.v.), we denote with $\widetilde{X}$ an identical and independent copy of X. Also, if $\boldsymbol{X} = \left(X_1, \ldots, X_n\right)^\top$ is a random vector, then $\widetilde{\boldsymbol{X}}$ is an identical and independent copy of $\boldsymbol{X}$.

5. The probability generating function (p.g.f.) of a discrete r.v. X is given by $\mathbb{E}\left[s^X\right] = \sum_{x=0}^{\infty} p_x s^x$, where $|s| \leq 1$. The p.g.f. of a random vector $\boldsymbol{X} = \left(X_1, \ldots, X_n\right)^\top$, comprised of discrete r.v.s, is given by

$$\mathbb{E}\left[\prod_{i=1}^{n} s_i^{X_i}\right] = \sum_{x_1,\ldots,x_n=0}^{\infty} \left[p(x_1, \ldots, x_n) \prod_{i=1}^{n} s_i^{x_i}\right],$$

where $max\{|s_1|, \ldots, |s_n|\} \leq 1$. The last requirement can be written as $|\boldsymbol{s}| \leq 1$, where $\boldsymbol{s} = (s_1, \ldots, s_n)^\top$.

We define the novel branching process of main interest within the dissertation:

Definition 2.1. *Define the Multi-type Sevastyanov Branching Process through probabilities of Mutation between types (MSBPM) as the multi-type branching process satisfying:*

1. *Each particle type is uniquely associated with an integer from $\mathbb{W}$ and conforms to:*

   (a) *The lifespan of particles of type i, $i \in \mathbb{W}$, is modeled by a (continuous) r.v. τ_i. The corresponding cumulative distribution function (c.d.f.) is denoted by $G_i(t) = \mathbb{P}(\tau_i \le t)$, also $G_i(0^+) = 0$.*

   (b) *The number of particles in the offspring of a type i, $i \in \mathbb{W}$, particle of age a is modeled by a (discrete) r.v. $\nu_i(a)$. We denote with $p_{ik}(a)$ the probability that a type i particle of age a has k, $k \in \mathbb{N}_0$, offspring particles (regardless of their type). Thus, $\nu_i(a)$ is specified by given $\big\{p_{ik}(a)\big\}_{k=0}^{\infty}$, $\sum_{k=0}^{\infty} p_{ik}(a) = 1$. We denote the corresponding p.g.f. of $\nu_i(a)$ with $f_i(a; s) = \mathbb{E}\big[s^{\nu_i(a)}\big] = \sum_{k=0}^{\infty} p_{ik}(a)s^k$, $|s| \le 1$.*

2. *Each daughter particle of a type i particle can be of any type $j \in \mathbb{W}$. The type of a daughter particle is determined at birth. If $i \ne j$ we say that a "mutation" occurs. The probability that a daughter particle of a type i particle is a type j particle is denoted by u_{ij}, $u_{ij} \ge 0$, $\sum_{j=1}^{n} u_{ij} = 1$. Further:*

   (a) *If type i cannot have daughters of type j we consider the corresponding u_{ij} as $u_{ij} = 0$.*

   (b) *Particles are not allowed to change their type within their lifespan.*

3. *All particles from all particle types evolve independently from one another, irrespective of generation.*

4. *Formally $\Big\{\boldsymbol{Z}(t) = \big(Z_1(t), Z_2(t), \ldots, Z_n(t)\big)^{\top}\Big\}_{t\ge 0}$, where $\boldsymbol{Z}(t)$ stands for the MSBPM at t and $Z_i(t)$ is the number of particles of type i that exist at t.*

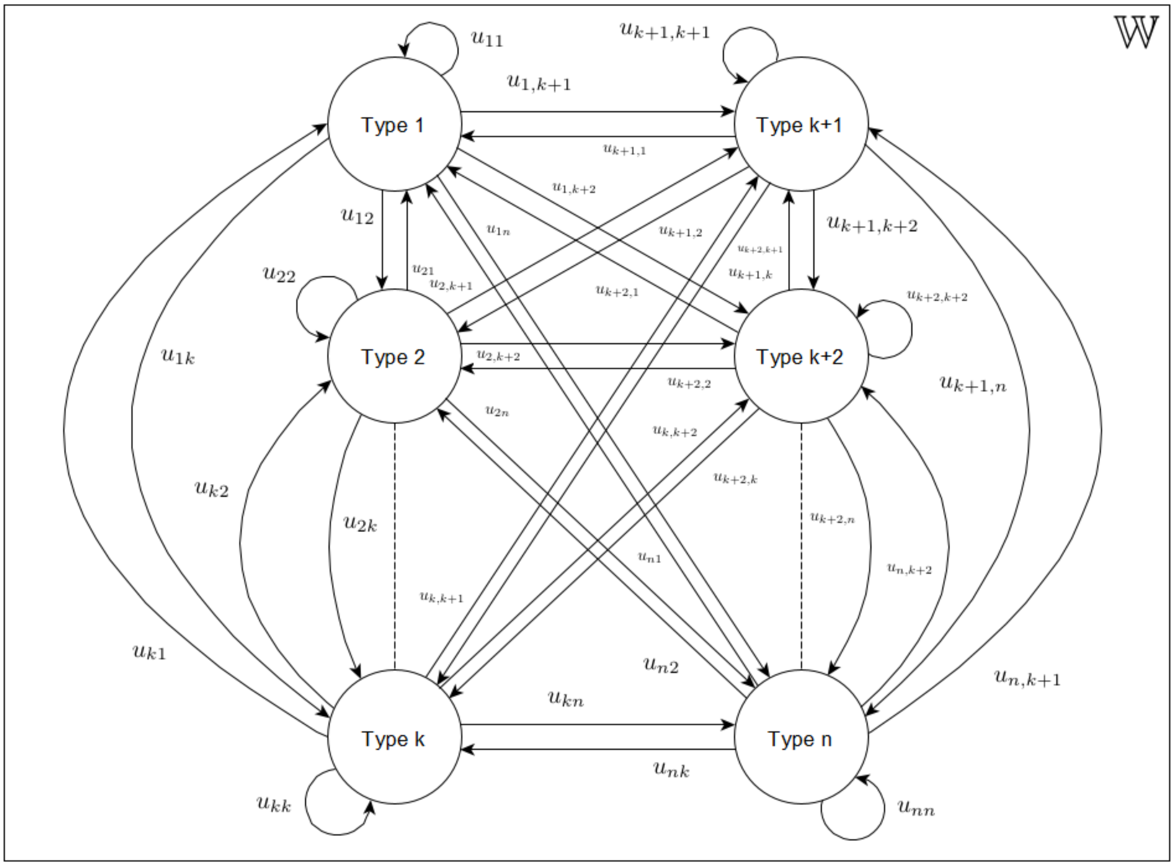

Figure 2.1: A diagram of the MSBPM depicting all possible paths of mutation within the process. Note that some of the u_{ij} may be equal to 0 depending on context. In such cases the corresponding arrows are removed from the diagram. See Figure 2.11 and Figure 2.12 in Subsection 2.3.1.1 for two possible realizations of interest of the MSBPM.

From Definition 2.1, we can see the connection between the MSBPM and the multi-type Sevastyanov branching process defined in Chapter VIII of [8]. Through $p_{ik}(a)$ and u_{ij}, as specified in Definition 2.1, we can construct an analogue of $p^i_{\boldsymbol{\alpha}}(a)$ (see Chapter VIII of [8] page 229 or Subsection 1.3.1) that has the same interpretation. This is done by setting $\sum_{j=1}^n \alpha_j = k$ and $p^i_{\boldsymbol{\alpha}}(a) := p_{ik}(a)\frac{k!}{\alpha_1!\ldots\alpha_n!}u_{i1}^{\alpha_1}\ldots u_{in}^{\alpha_n}$.

Definition 2.1 bears no explicit or implicit assumptions about the decomposability of the process. Decomposability can be modeled by setting an appropriate combination of u_{ij} to 0 and can be avoided either by setting all $u_{ij} \neq 0$ or by careful selection which u_{ij} are being set to 0. We note that some of the results within [8] (e.g., Theorem 1.4 within the Introduction) rely on the Perron–Frobenius theorem (see Theorem 4 within the Appendix) and the calculation of Perron roots, which implies non-decomposability. However, within the base definition of the classical multi-type Sevastyanov branching process, given in Subsection 1.3.1 (or Chapter VIII from [8]), nothing explicitly forbids for the branching process to be decomposable. If we take the definition from [8] as having no implicit assumption for decomposability, then the MSBPM (starting with a particle of age 0) can be viewed as the multi-type Sevastyanov branching process from [8] rewritten via probabilities of mutation. If we take the definition from [8] as having an implicit assumption of non-decomposability, then the MSBPM is not exactly the multi-type Sevastyanov branching process from [8], however, it is still closely connected to it. Either way, we argue that naming the process from Definition 2.1 as "Multi-type Sevastyanov Branching Process through probabilities of Mutation between types" is appropriate as it is evident that the MSBPM (when starting with one particle of age 0) enjoys all classical multi-type Sevastyanov branching process theorems that do not rely on an assumption of non-decomposability. The MSBPM also enjoys those results that do have this assumption, provided that the non-zero u_{ij} within the MSBPM make it non-decomposable.

Throughout Chapter 2, we obtain various novel results for quantities of interest in context of populations escaping extinction. The novelty of these results is partially due to the fact that they are valid within the framework set by the novel MSBPM. To the best of our knowledge, an in-depth investigation of the topic has not been done previously for a continuous-time branching process model of such high sophistication as the MSBPM (here we exclude our earlier work in [7] as well as preceding papers [1] - [6]).

We also explore the case where the MSBPM starts with a particle that has non-zero age. To the best of our knowledge, this has also not been done previously in a systematic manner for a continuous-time branching process. Processes starting from particles with non-zero age are important in the context of sequential decision problems (see Chapter 3), where at a given decision epoch we have a collection of particles, some of them with non-zero age, that will continue to evolve after we apply a decision. As we will see, the analogues of our results for the case when the MSBPM starts with a particle with non-zero age rely on the results for the case when the process starts with a particle with age 0.

Below, whenever we consider a MSBPM with no dependence of the reproductive capabilities of particles from their age, we will be referring to the process as *Multi-type Bellman-Harris Branching Process through* *probabilities of Mutation between types (MBHBPM).*

2.2.2 Probability generating functions for the MSBPM

We turn our attention to the p.g.f.s of the MSBPM. Although there exists a theorem in [8] (Theorem 1, Chapter VIII, page 231) about the p.g.f.s of the multi-type Sevastyanov branching process that uses $p^i_{\boldsymbol{\alpha}}(a)$, to the best of our knowledge, there is no previously proven analogue of the theorem in [8] that uses u_{ij} and $p_{ik}(a)$.

Definition 2.2. *We denote the p.g.f. of a MSBPM, starting with one particle of type i, $i \in \mathbb{W}$, that is of age 0, with:*

$$F_i(t;\boldsymbol{s}) = \mathbb{E}\Big(\prod_{j\in\mathbb{W}} s_j^{Z_j(t)} |\ \boldsymbol{Z}(0) = \boldsymbol{\delta}^i\Big),$$

where $|\boldsymbol{s}| \leq 1$. We denote the p.g.f. of a MSBPM, starting with one particle of type i, $i \in \mathbb{W}$, that is of age a, $a \neq 0$, with

$$F_{i,a}(t;\boldsymbol{s}) = \mathbb{E}\Big(\prod_{j\in\mathbb{W}} s_j^{Z_j(t)} |\ \boldsymbol{Z}(0) = \boldsymbol{\delta}^i_a\Big),$$

where $|\boldsymbol{s}| \leq 1$.

Theorem 2.1. *The following system of integral equations holds for the MSBPM, $i \in \mathbb{W}$:*

$$F_i(t;\, \boldsymbol{s}) = s_i\big(1 - G_i(t)\big) + \int_0^t f_i\Big(y; \sum_{r\in\mathbb{W}} u_{ir} F_r(t-y;\, \boldsymbol{s})\Big) dG_i(y). \tag{2.1}$$

Proof. Let the MSBPM start with one particle of type i. We can expand the expectation in Definition 2.2 as follows:

$$\begin{aligned} F_i(t;\boldsymbol{s}) &= \mathbb{E}\Big(\prod_{j\in\mathbb{W}} s_j^{Z_j(t)}\ |\ \boldsymbol{Z}(0) = \boldsymbol{\delta}^i\Big) \\ &= \mathbb{E}\Bigg[\mathbb{E}\Big(\prod_{j\in\mathbb{W}} s_j^{Z_j(t)}\ |\ \boldsymbol{Z}(0) = \boldsymbol{\delta}^i, \big(\tau_i, \nu_i(\tau_i)\big)\Big)\Bigg]. \end{aligned}$$

Note that the assumption of independent evolution of all particles from all types allows us to consider each daughter particle, at its moment of birth, as starting a new independent copy of the MSBPM.

The assumption of independent evolution proves useful when considering the possible outcomes with respect to τ_i:

1. The initial particle of type i, $i \in \mathbb{W}$, dies/reproduces at some moment y, $y > t$. The probability for this event is $\big(1 - G_i(t)\big)$. In this case, we have $\mathbb{E}\Big(\prod_{j \in \mathbb{W}} s_j^{Z_j(t)} | \, \boldsymbol{Z}(0) = \boldsymbol{\delta}^i \Big) = s_i^1 = s_i$.

2. The initial particle of type i, $i \in \mathbb{W}$, dies and reproduces at moment y, $y \leq t$. In this case, for each offspring particle of type $m \in \mathbb{W}$, we obtain a new independent process starting at y with a corresponding
$$\mathbb{E}\Big(\prod_{j \in \mathbb{W}} s_j^{\widetilde{Z}_j(t-y)} \mid \widetilde{\boldsymbol{Z}}(0) = \boldsymbol{\delta}^m \Big).$$

As for $\nu_i(\tau_i)$, we keep in mind that, if reproduction occurs before t, the initial particle will have k, $k \in \mathbb{N}_0$, offspring particles. We further note that the distribution of these k offspring particles, among types, follows the multinomial distribution.

We are ready to proceed with the core of our proof:

$$\begin{aligned}
&F_i(t; \boldsymbol{s}) = \\
&= \mathbb{E}\Bigg[\mathbb{E}\Big(\prod_{j \in \mathbb{W}} s_j^{Z_j(t)} \mid \boldsymbol{Z}(0) = \boldsymbol{\delta}^i, \big(\tau_i, \nu_i(\tau_i)\big) \Big) \Bigg] \\
&= s_i \big(1 - G_i(t)\big) + \\
&+ \int_0^t \sum_{k=0}^{\infty} p_{ik}(y) \sum_{\sum_{l \in \mathbb{W}} k_l = k} \Bigg[\frac{k!}{\prod_{v \in \mathbb{W}} k_v!} \prod_{r \in \mathbb{W}} u_{ir}^{k_r} \cdot \\
&\qquad \cdot \prod_{m \in \mathbb{W}} \Big[\mathbb{E}\Big(\prod_{j \in \mathbb{W}} s_j^{\widetilde{Z}_j(t-y)} \mid \widetilde{\boldsymbol{Z}}(0) = \boldsymbol{\delta}^m \Big) \Big]^{k_m} \Bigg] dG_i(y) \\
&= s_i \big(1 - G_i(t)\big) + \\
&+ \int_0^t \sum_{k=0}^{\infty} p_{ik}(y) \sum_{\sum_{l \in \mathbb{W}} k_l = k} \Bigg[\frac{k!}{\prod_{v \in \mathbb{W}} k_v!} \cdot
\end{aligned}$$

$$\cdot \prod_{r\in\mathbb{W}} \Big[u_{ir}\mathbb{E}\Big(\prod_{j\in\mathbb{W}} s_j^{\widetilde{Z}_j(t-y)} \mid \widetilde{\boldsymbol{Z}}(0)=\boldsymbol{\delta}^r\Big)\Big]^{k_r}\Bigg]dG_i(y)$$

$$= s_i\big(1-G_i(t)\big)+$$

$$+\int_0^t \sum_{k=0}^{\infty} p_{ik}(y) \sum_{\sum_{l\in\mathbb{W}} k_l=k} \left[\frac{k!}{\prod_{v\in\mathbb{W}} k_v!}\cdot \prod_{r\in\mathbb{W}} \big[u_{ir}F_r(t-y;\boldsymbol{s})\big]^{k_r}\right]dG_i(y)$$

$$= s_i\big(1-G_i(t)\big)+\int_0^t \sum_{k=0}^{\infty} p_{ik}(y)\Big[\sum_{r\in\mathbb{W}} u_{ir}F_r(t-y;\boldsymbol{s})\Big]^k dG_i(y)$$

$$= s_i\big(1-G_i(t)\big)+\int_0^t f_i\Big(y;\sum_{r\in\mathbb{W}} u_{ir}F_r(t-y;\boldsymbol{s})\Big)dG_i(y).$$

In the expressions above whenever $u_{ir}=0$, since there can be no mutations from type i towards type r, we have that the corresponding k_r is also $k_r=0$. Thus, in this case, $u_{ir}^{k_r}=1$. □

Remark 2.1. *It is evident that the steps of the proof of Theorem 2.1 do not depend on the configuration of non-zero u_{ij} nor on $\boldsymbol{s}$.*

Systems of integral equations for $F_i(t;\boldsymbol{s})$, corresponding to particular cases of the MSBPM where some of the u_{ij} are 0 and/or where we are interested only into particles from a particular class of types, can be immediately obtained from Theorem 2.1. More specifically, this is achieved by setting the necessary u_{ij} to 0 and also setting those coordinates of $\boldsymbol{s}$, that do not correspond to types from the class of interest, to 1. We see how this is done in Subsection 2.3.1 and Subsection 2.3.2.

Corollary 2.1. *Let a MSBPM start with one particle of type i, $i\in\mathbb{W}$, that is of age a, $a\neq 0$. Then*

$$F_{i,a}(t;\,\boldsymbol{s}) = s_i\big(1-G_{i,a}(t)\big)+\int_0^t f_i\Big(a+y;\sum_{r\in\mathbb{W}} u_{ir}F_r(t-y;\,\boldsymbol{s})\Big)dG_{i,a}(y). \tag{2.2}$$

Proof. The proof is completely analogous to the proof of Theorem 2.1, with the exceptions being that when we consider the cases, with respect to the moment of death of the initial particle of type i, we have $G_{i,a}(t)$ instead of $G_i(t)$, and also when writing $p_{ik}(\cdot)$, we must take into account the age, a, of the initial particle. □

Remark 2.2. *We can see from Corollary 2.1 that* $F_{i,a}(t;\boldsymbol{s})$, *for a fixed* i, $i \in \mathbb{W}$, *does not depend on* $F_{k,a}(t;\boldsymbol{s})$, $k \in \mathbb{W}$. $F_{i,a}(t;\boldsymbol{s})$, *however, depends on equations* (2.1).

We can obtain the expected number of particles, per type, that exist at t within a MSBPM starting with one particle of type i, that is of age 0, via differentiating $F_i(t;\boldsymbol{s})$. Note that this is different form (1.7), on page 19, where we have the number of particles, per type, within the *next generation* of a type i particle. For obtaining the expected number of particles of type j at t, we calculate the left-hand, with respect to $\mathbf{1}$, partial derivative:

$$\mathbb{E}\Big(Z_j(t) \mid \boldsymbol{Z}(0) = \boldsymbol{\delta}^i\Big) = \lim_{\Delta s_j \to 0^+} \frac{F_i(t;\,\mathbf{1}) - F_i(t;\,\mathbf{1} - \Delta\boldsymbol{s}_j)}{\Delta s_j}, \tag{2.3}$$

where $\Delta\boldsymbol{s}_j = (0, \ldots, \Delta s_j, \ldots, 0)^\top$. The same line of reasoning is valid when the initial particle is of age a, $a \neq 0$ - for obtaining the expected number of particles of type j at t, we calculate the left-hand, with respect to $\mathbf{1}$, partial derivative:

$$\mathbb{E}\Big(Z_j(t) \mid \boldsymbol{Z}(0) = \boldsymbol{\delta}^i_a\Big) = \lim_{\Delta s_j \to 0^+} \frac{F_{i,a}(t;\,\mathbf{1}) - F_{i,a}(t;\,\mathbf{1} - \Delta\boldsymbol{s}_j)}{\Delta s_j}. \tag{2.4}$$

Numerical Scheme 1 and Numerical Scheme 2, presented in Section 2.2.7, can be applied to the numerators in the right sides of equations (2.3) and (2.4) respectively, thus we will not develop these equations any further.

2.2.3 Probabilities of extinction for the MSBPM

Definition 2.3. *We define the probability of extinction until time t of a MSBPM beginning with one particle of type i, $i \in \mathbb{W}$, that is of age 0, as:*

$$q_i(t) = \mathbb{P}\left(\sum_{j \in \mathbb{W}} Z_j(t) = 0 \mid \boldsymbol{Z}(0) = \boldsymbol{\delta}^i\right).$$

Similarly, if the initial particle is of age a, $a \neq 0$:

$$q_{i,a}(t) = \mathbb{P}\left(\sum_{j \in \mathbb{W}} Z_j(t) = 0 \mid \boldsymbol{Z}(0) = \boldsymbol{\delta}^i_a\right).$$

Remark 2.3. *If the event* $\left\{\sum_{j\in\mathbb{W}} Z_j(t^*) = 0 \mid \boldsymbol{Z}(0) = \boldsymbol{\delta}^i\right\}$ *has occurred for some* t^**, since there are no particles left in the process, the event* $\left\{\sum_{j\in\mathbb{W}} Z_j(t) = 0 \mid \boldsymbol{Z}(0) = \boldsymbol{\delta}^i\right\}$ *necessarily occurs for all* t*,* $t^* < t$*. Thus,* $\left\{\sum_{j\in\mathbb{W}} Z_j(t_1) = 0 \mid \boldsymbol{Z}(0) = \boldsymbol{\delta}^i\right\} \subseteq \left\{\sum_{j\in\mathbb{W}} Z_j(t_2) = 0 \mid \boldsymbol{Z}(0) = \boldsymbol{\delta}^i\right\}$ *and* $q_i(t_1) \leq q_i(t_2)$ *for all* $t_1 < t_2$*,* $i \in \mathbb{W}$*. Similarly, it is also true that* $q_{i,a}(t_1) \leq q_{i,a}(t_2)$ *for all* $t_1 < t_2$*,* $i \in \mathbb{W}$*.*

Theorem 2.2. *The following system of integral equations holds for the MSBPM,* $i \in \mathbb{W}$*:*

$$q_i(t) = \int_0^t f_i\Big(y; \sum_{r\in\mathbb{W}} u_{ir} q_r(t-y)\Big) dG_i(y). \tag{2.5}$$

Proof. We recognize that $F_i(t;\boldsymbol{s})$ are p.g.f.s and apply $\boldsymbol{s} = \boldsymbol{0}$ into the result of Theorem 2.1. □

Corollary 2.2. *The following system of integral equations holds for the MSBPM,* $i \in \mathbb{W}$*:*

$$q_{i,a}(t) = \int_0^t f_i\Big(a + y; \sum_{r\in\mathbb{W}} u_{ir} q_r(t-y)\Big) dG_{i,a}(y). \tag{2.6}$$

Proof. We recognize that $F_{i,a}(t;\boldsymbol{s})$ are p.g.f.s and apply $\boldsymbol{s} = \boldsymbol{0}$ into the result of Corollary 2.1. □

Remark 2.4. *We can see from Corollary 2.2 that* $q_{i,a}(t)$*, for a fixed* i*,* $i \in \mathbb{W}$*, does not depend on* $q_{k,a}(t)$*,* $k \in \mathbb{W}$*.* $q_{i,a}(t)$*, however, depends on equations* (2.5)*.*

Definition 2.4. *We denote the probability of extinction of a MSBPM beginning with one particle of type* i*,* $i \in \mathbb{W}$*, that is of age* 0*, as:*

$$q_i = \mathbb{P}\left(\sum_{j\in\mathbb{W}} Z_j(t) = 0 \text{ for some } t > 0 \mid \boldsymbol{Z}(0) = \boldsymbol{\delta}^i\right).$$

Similarly, if the initial particle is of age a*,* $a \neq 0$*:*

$$q_{i,a} = \mathbb{P}\left(\sum_{j\in\mathbb{W}} Z_j(t) = 0 \text{ for some } t > 0 \mid \boldsymbol{Z}(0) = \boldsymbol{\delta}^i_a\right).$$

Remark 2.5. *As a consequence of Remark 2.3, we have the inclusion* $\Big\{ \sum_{j\in\mathbb{W}} Z_j(t^*) = 0 \mid \boldsymbol{Z}(0) = \boldsymbol{\delta}^i \Big\} \subseteq \Big\{ \sum_{j\in\mathbb{W}} Z_j(\infty) = 0 \mid \boldsymbol{Z}(0) = \boldsymbol{\delta}^i \Big\}$ *for any fixed* t^**. Thus, we can write*

$$q_i = \mathbb{P}\Bigg(\sum_{j\in\mathbb{W}} Z_j(\infty) = 0 \mid \boldsymbol{Z}(0) = \boldsymbol{\delta}^i \Bigg),$$

therefore $q_i = \lim_{t\to\infty} q_i(t)$*. The same reasoning provides us with*

$$q_{i,a} = \mathbb{P}\Bigg(\sum_{j\in\mathbb{W}} Z_j(\infty) = 0 \mid \boldsymbol{Z}(0) = \boldsymbol{\delta}^i_a \Bigg),$$

and $q_{i,a} = \lim_{t\to\infty} q_{i,a}(t)$

Before we give the next theorem, we note that we follow the interpretation $\lim_{y\to\infty} \lim_{t\to\infty} q_i(t-y) = q_i$. This interpretation basically means that whatever the moment of birth, y, of a type i particle, the independent process that starts with this particle has infinite time to develop as $t \to \infty$. This interpretation is natural and more realistic than the alternative - if we assume $\lim_{y\to\infty} \lim_{t\to\infty} q_i(t-y) = q_i(b)$, where b is a constant, we will be, arbitrarily, setting the length of the time interval, b, that the process starting from the particle at $y \to \infty$ has at its disposal to develop until $t \to \infty$. We discuss an analogous situation, in more detail, in Remark 2.10. For now, we give

Definition 2.5. *We define* $\lim_{y\to\infty} \lim_{t\to\infty} q_i(t-y) = q_i$ *and* $\lim_{y\to\infty} \lim_{t\to\infty} q_{i,a}(t-y) = q_{i,a}$.

Theorem 2.3. *The following system of integral equations holds for the MSBPM,* $i \in \mathbb{W}$*:*

$$q_i = \int_0^\infty f_i\Big(y; \sum_{r\in\mathbb{W}} u_{ir} q_r\Big) dG_i(y). \tag{2.7}$$

Proof. Due to considerations analogous to those in Remark 2.10, we adopt Definition 2.5. Taking also into account the result of Theorem 2.2, as well as Remark 2.5, the proof of the current theorem is analogous to the proof of *Case* 1 within Theorem 2.5. □

Corollary 2.3. *The following system of integral equations holds for the MSBPM, $i \in \mathbb{W}$:*

$$q_{i,a} = \int_0^\infty f_i\Big(a+y; \sum_{r\in\mathbb{W}} u_{ir} q_r\Big) dG_{i,a}(y). \tag{2.8}$$

Proof. Due to considerations analogous to those in Remark 2.10, we adopt Definition 2.5. Taking also into account the result of Corollary 2.2, as well as Remark 2.5, the proof of the current theorem is analogous to the proof of *Case* 1 within Theorem 2.5. □

Remark 2.6. *We can see from Corollary 2.3 that $q_{i,a}$, for a fixed i, $i \in \mathbb{W}$, does not depend on $q_{k,a}$, $k \in \mathbb{W}$. $q_{i,a}$, however, depends on equations* (2.7).

Corollary 2.4. *In the particular case where there is no dependence of particle reproduction from particle age, i.e., $f_i(y; s) = f_i(s)$, $i \in \mathbb{W}$, the systems of integral equations* (2.7) *and* (2.8) *become the following system of equations, $i \in \mathbb{W}$:*

$$q_i = q_{i,a} = f_i\Big(\sum_{r\in\mathbb{W}} u_{ir} q_r\Big). \tag{2.9}$$

Proof. To prove for q_i, we take the result from Theorem 2.3 and drop the dependence from y within f_i:

$$\begin{aligned} q_i &= \int_0^\infty f_i\Big(\sum_{r\in\mathbb{W}} u_{ir} q_r\Big) dG_i(y) \\ &= f_i\Big(\sum_{r\in\mathbb{W}} u_{ir} q_r\Big) \int_0^\infty dG_i(y) \\ &= f_i\Big(\sum_{r\in\mathbb{W}} u_{ir} q_r\Big). \end{aligned}$$

To prove for $q_{i,a}$, we take the result from Corollary 2.3 and drop the dependence from y (and a) within f_i.

$$\begin{aligned} q_{i,a} &= \int_0^\infty f_i\Big(\sum_{r\in\mathbb{W}} u_{ir} q_r\Big) dG_{i,a}(y). \\ &= f_i\Big(\sum_{r\in\mathbb{W}} u_{ir} q_r\Big) \int_0^\infty dG_{i,a}(y) \end{aligned}$$

$$= f_i\Big(\sum_{r\in\mathbb{W}} u_{ir}q_r\Big).$$

$\square$

We complete this subsection with two figures that illustrate the results obtained. We note that the calculations done for Figure 2.2 and Figure 2.3 at $t = 5000$ suggest that there could be a more general result than Corollary 2.4.

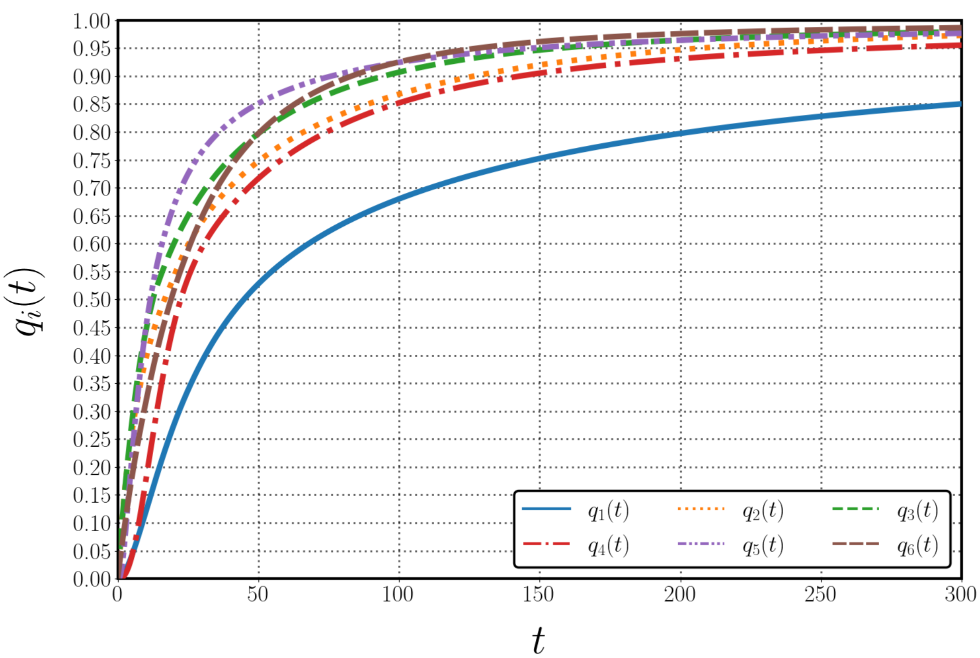

Figure 2.2: An application of Theorem 2.3 - probabilities of extinction for the example MSBPM (Table 2.11, Table 2.6) with mutation scheme "$\mathbb{W}$ towards $\mathbb{W}$" (Table 2.7) starting with one particle of age 0. Displayed values are cut at $t = 300$ ($h = 10^{-2}$) so that the dynamics of the different $q_i(t)$ within $[0, 300]$ is visible. The slightly above 1 criticality of the process (see the preliminary analysis in Subsection 2.2.8.4) forces us to calculate $q_i(t)$ for large t in order to obtain q_i. At $t = 5000$ ($h = 10^{-1}$), we have $q_1 \approx 0.96009038$, $q_2 \approx 0.99744155$, $q_3 \approx 0.99602606$, $q_4 \approx 0.99015646$, $q_5 \approx 0.99464378$, $q_6 \approx 0.99762545$, those values practically not changing when calculating with $t = 3500, 4000, 4500, 5000$. The large values of q_i are in agreement with the low criticality of the process as per the preliminary analysis of Subsection 2.2.8.4.

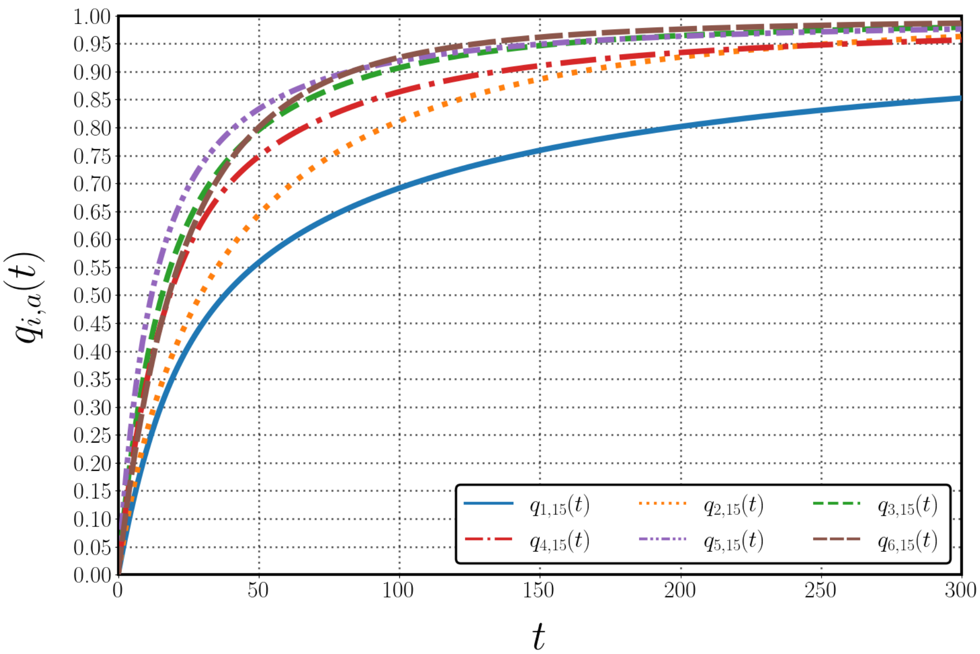

Figure 2.3: An application of Corollary 2.3 - probabilities of extinction for the example MSBPM (Table 2.11, Table 2.6) with mutation scheme "$\mathbb{W}$ towards $\mathbb{W}$" (Table 2.7) starting with one particle of age $a = 15$. Displayed values are cut at $t = 300$ ($h = 10^{-2}$) so that the dynamics of the different $q_i(t)$ within $[0, 300]$ is visible. We note that, with respect to Figure 2.2, $q_{1,15}(t)$ is slightly more curved when compared to $q_1(t)$. At $t = 5000$ ($h = 10^{-1}$), we have $q_{1,15} \approx 0.96008794$, $q_{2,15} \approx 0.99743786$, $q_{3,15} \approx 0.99602587$, $q_{4,15} \approx 0.99015601$, $q_{5,15} \approx 0.99464342$, $q_{6,15} \approx 0.99762525$. The values of $q_{i,a}(t)$ at, $t = 5000$, practically coincide with the values of $q_i(t)$ from Figure 2.2, although for smaller t this is not the case.

2.2.4 Number of particles produced from $\mathbb{W}_e$ towards $\mathbb{W}$ within the MSBPM

Let $\mathbb{W}_e \subseteq \mathbb{W}$ be a subset of types within the MSBPM (the subscript "e" stands for "emit"). Note that we allow $\mathbb{W}_e = \mathbb{W}$. The number of occurred mutations from $\mathbb{W}_e$ towards types in $\mathbb{W} \setminus \mathbb{W}_e$ is a crucial quantity in the context of populations escaping extinction as the types that have supercritical reproduction are usually modeled to be outside of $\mathbb{W}_e$. We investigate the production of mutants from $\mathbb{W}_e \subset \mathbb{W}$ towards $\mathbb{W}_0 = \mathbb{W} \setminus \mathbb{W}_e$

in Subsection 2.3.1 and Subsection 2.3.2. In the current Subsection, we derive more general results. These general results concern general particle production, that is, the particles produced from $\mathbb{W}_e$ can be of any type within $\mathbb{W}$ and are not necessarily mutants. Particular cases of these results that correspond to particle production from $\mathbb{W}_e$ towards any subclass of $\mathbb{W}$ within the MSBPM, can be straightforwardly obtained by appropriately setting $u_{ij} = 0$, appropriately setting coordinates of $\boldsymbol{s}$ ot 1, and realizing in some occasions that $s^X = 1$ due to a relevant random variable X being always 0.

Definition 2.6. *Denote with $I_j^{\mathbb{W}_e}(t)$ the number of particles (mutant or not) of type j, $j \in \mathbb{W}$, produced from particles with types from $\mathbb{W}_e$ until t within a MSBPM. We do not count the initial particle within any of the $I_j^{\mathbb{W}_e}(t)$. For a MSBPM starting with one particle of type i, $i \in \mathbb{W}$, that is of age 0, denote with $h_i^{\mathbb{W}_e}(t;\boldsymbol{s})$ the following p.g.f.*

$$h_i^{\mathbb{W}_e}(t;\boldsymbol{s}) = \mathbb{E}\Big(\prod_{j\in\mathbb{W}} s_j^{I_j^{\mathbb{W}_e}(t)} \mid \boldsymbol{Z}(0) = \boldsymbol{\delta}^i\Big),$$

where $|\boldsymbol{s}| \leq 1$. We denote the corresponding p.g.f., when the MSBPM starts with one particle of type i, $i \in \mathbb{W}$, that is of age a, $a \neq 0$, with

$$h_{i,a}^{\mathbb{W}_e}(t;\boldsymbol{s}) = \mathbb{E}\Big(\prod_{j\in\mathbb{W}} s_j^{I_j^{\mathbb{W}_e}(t)} \mid \boldsymbol{Z}(0) = \boldsymbol{\delta}_a^i\Big),$$

where $|\boldsymbol{s}| \leq 1$.

We note that unlike $F_i(t;\boldsymbol{s})$, which compactly contain information about the number of particles, per type, that exist *at* t, $h_i^{\mathbb{W}_e}(t;\boldsymbol{s})$ contain information about the number of particles that have been produced *until* t (with respect to t some of the produced particles may no longer exist).

Theorem 2.4. *The following system of integral equations holds within the MSBPM:*

1. *For $i \in \mathbb{W}_e$*

   (2.10)
$$h_i^{\mathbb{W}_e}(t;\boldsymbol{s}) = \big(1 - G_i(t)\big) + \int_0^t f_i\Big(y; \sum_{r\in\mathbb{W}} u_{ir} s_r h_r^{\mathbb{W}_e}(t-y;\boldsymbol{s})\Big) dG_i(y).$$

2. For $i \notin \mathbb{W}_e$

$$h_i^{\mathbb{W}_e}(t;\boldsymbol{s}) = \big(1-G_i(t)\big) + \int_0^t f_i\Big(y; \sum_{r\in\mathbb{W}} u_{ir} h_r^{\mathbb{W}_e}(t-y;\boldsymbol{s})\Big) dG_i(y). \tag{2.11}$$

Proof. The proof of the theorem is similar in idea with the proof of Theorem 2.1. We begin with expanding the expectation in Definition 2.6 as follows:
For each $i \in \mathbb{W}$, we have

$$\begin{aligned} h_i^{\mathbb{W}_e}(t;\boldsymbol{s}) &= \mathbb{E}\Big(\prod_{j\in\mathbb{W}} s_j^{I_j^{\mathbb{W}_e}(t)} \mid \boldsymbol{Z}(0)=\boldsymbol{\delta}^i\Big) \\ &= \mathbb{E}\left[\mathbb{E}\Big(\prod_{j\in\mathbb{W}} s_j^{I_j^{\mathbb{W}_e}(t)} \mid \boldsymbol{Z}(0)=\boldsymbol{\delta}^i, \big(\tau_i, \nu_i(\tau_i)\big)\Big)\right]. \end{aligned}$$

Note that the assumption of independent evolution of all particles from all types allows us to consider each daughter particle, at its moment of birth, as starting a new independent copy of the MSBPM.

The possible outcomes, with respect to τ_i, are:

1. The initial particle of type i, $i \in \mathbb{W}$, dies/reproduces at some moment y, $y > t$. The probability for this event is $\big(1-G_i(t)\big)$. In this case, we have $\mathbb{E}\Big(\prod_{j\in\mathbb{W}} s_j^{I_j^{\mathbb{W}_e}(t)} \mid \boldsymbol{Z}(0)=\boldsymbol{\delta}^i\Big) = \prod_{j\in\mathbb{W}} s_j^0 = 1$.

2. The initial particle of type i, $i \in \mathbb{W}$, dies and reproduces at moment y, $y \le t$.

   (a) If $i \notin \mathbb{W}_e$, then for each offspring particle of type $m \in \mathbb{W}$, we have a new independent process starting at y with a corresponding
   $$\mathbb{E}\Big(\prod_{j\in\mathbb{W}} s_j^{\widetilde{I}_j^{\mathbb{W}_e}(t-y)} | \widetilde{\boldsymbol{Z}}(0)=\boldsymbol{\delta}^m\Big).$$

   (b) If $i \in \mathbb{W}_e$, then for each offspring particle of type $m \in \mathbb{W}$, we obtain a new independent process starting at y with a corresponding
   $$\mathbb{E}\Big(s_m^{\left(1+\widetilde{I}_m^{\mathbb{W}_e}(t-y)\right)} \cdot \prod_{j\in\mathbb{W}, j\neq m} s_j^{\widetilde{I}_j^{\mathbb{W}_e}(t-y)} \mid \widetilde{\boldsymbol{Z}}(0)=\boldsymbol{\delta}^m\Big) =$$

$$= s_m \mathbb{E}\Big(\prod_{j \in \mathbb{W}} s_j^{\widetilde{I}_j^{\mathbb{W}_e}(t-y)} \mid \widetilde{\boldsymbol{Z}}(0) = \boldsymbol{\delta}^m \Big).$$

We now proceed with the core of the proof:

1. Let $i \in \mathbb{W}_e$. We have

$$\begin{aligned}
&h_i^{\mathbb{W}_e}(t;\boldsymbol{s}) = \\
&= \mathbb{E}\bigg[\mathbb{E}\Big(\prod_{j \in \mathbb{W}} s_j^{I_j^{\mathbb{W}_e}(t)} \mid \boldsymbol{Z}(0) = \boldsymbol{\delta}^i, \big(\tau_i, \nu_i(\tau_i)\big) \Big) \bigg] \\
&= \big(1 - G_i(t)\big) + \\
&+ \int_0^t \sum_{k=0}^{\infty} p_{ik}(y) \sum_{\sum_{l \in \mathbb{W}} k_l = k} \bigg[\frac{k!}{\prod_{v \in \mathbb{W}} k_v!} \prod_{r \in \mathbb{W}} u_{ir}^{k_r} \cdot \\
&\qquad \cdot \prod_{m \in \mathbb{W}} \Big[s_m \mathbb{E}\Big(\prod_{j \in \mathbb{W}} s_j^{\widetilde{I}_j^{\mathbb{W}_e}(t-y)} \mid \widetilde{\boldsymbol{Z}}(0) = \boldsymbol{\delta}^m \Big) \Big]^{k_m} \bigg] dG_i(y) \\
&= \big(1 - G_i(t)\big) + \\
&+ \int_0^t \sum_{k=0}^{\infty} p_{ik}(y) \sum_{\sum_{l \in \mathbb{W}} k_l = k} \bigg[\frac{k!}{\prod_{v \in \mathbb{W}} k_v!} \cdot \\
&\qquad \cdot \prod_{r \in \mathbb{W}} \Big[u_{ir} s_r \mathbb{E}\Big(\prod_{j \in \mathbb{W}} s_j^{\widetilde{I}_j^{\mathbb{W}_e}(t-y)} \mid \widetilde{\boldsymbol{Z}}(0) = \boldsymbol{\delta}^r \Big) \Big]^{k_r} \bigg] dG_i(y) \\
&= \big(1 - G_i(t)\big) + \\
&+ \int_0^t \sum_{k=0}^{\infty} p_{ik}(y) \sum_{\sum_{l \in \mathbb{W}} k_l = k} \bigg[\frac{k!}{\prod_{v \in \mathbb{W}} k_v!} \cdot \prod_{r \in \mathbb{W}} \big[u_{ir} s_r h_r^{\mathbb{W}_e}(t-y;\boldsymbol{s}) \big]^{k_r} \bigg] dG_i(y) \\
&= \big(1 - G_i(t)\big) + \int_0^t \sum_{k=0}^{\infty} p_{ik}(y) \Big[\sum_{r \in \mathbb{W}} u_{ir} s_r h_r^{\mathbb{W}_e}(t-y;\boldsymbol{s}) \Big]^k dG_i(y) \\
&= \big(1 - G_i(t)\big) + \int_0^t f_i\Big(y; \sum_{r \in \mathbb{W}} u_{ir} s_r h_r^{\mathbb{W}_e}(t-y;\boldsymbol{s}) \Big) dG_i(y).
\end{aligned}$$

2. Now, let $i \notin \mathbb{W}_e$. We have

$$h_i^{\mathbb{W}_e}(t;\boldsymbol{s}) =$$

$$
\begin{aligned}
&= \mathbb{E}\Bigg[\mathbb{E}\Big(\prod_{j\in\mathbb{W}} s_j^{I_j^{\mathbb{W}_e}(t)} \mid \boldsymbol{Z}(0)=\boldsymbol{\delta}^i, \big(\tau_i, \nu_i(\tau_i)\big)\Big)\Bigg] \\
&= \big(1-G_i(t)\big)+ \\
&+ \int_0^t \sum_{k=0}^{\infty} p_{ik}(y) \sum_{\sum_{l\in\mathbb{W}} k_l = k} \Bigg[\frac{k!}{\prod_{v\in\mathbb{W}} k_v!} \prod_{r\in\mathbb{W}} u_{ir}^{k_r} \cdot \\
&\qquad \cdot \prod_{m\in\mathbb{W}} \Big[\mathbb{E}\Big(\prod_{j\in\mathbb{W}} s_j^{\widetilde{I}_j^{\mathbb{W}_e}(t-y)} | \widetilde{\boldsymbol{Z}}(0)=\boldsymbol{\delta}^m\Big)\Big]^{k_m}\Bigg] dG_i(y) \\
&= \big(1-G_i(t)\big)+ \\
&+ \int_0^t \sum_{k=0}^{\infty} p_{ik}(y) \sum_{\sum_{l\in\mathbb{W}} k_l = k} \Bigg[\frac{k!}{\prod_{v\in\mathbb{W}} k_v!} \cdot \\
&\qquad \cdot \prod_{r\in\mathbb{W}} \Big[u_{ir}\mathbb{E}\Big(\prod_{j\in\mathbb{W}} s_j^{\widetilde{I}_j^{\mathbb{W}_e}(t-y)} | \widetilde{\boldsymbol{Z}}(0)=\boldsymbol{\delta}^r\Big)\Big]^{k_r}\Bigg] dG_i(y) \\
&= \big(1-G_i(t)\big)+ \\
&+ \int_0^t \sum_{k=0}^{\infty} p_{ik}(y) \sum_{\sum_{l\in\mathbb{W}} k_l = k} \Bigg[\frac{k!}{\prod_{v\in\mathbb{W}} k_v!} \cdot \\
&\qquad \cdot \prod_{r\in\mathbb{W}} \big[u_{ir} h_r^{\mathbb{W}_e}(t-y;\boldsymbol{s})\big]^{k_r}\Bigg] dG_i(y) \\
&= \big(1-G_i(t)\big) + \int_0^t \sum_{k=0}^{\infty} p_{ik}(y) \Big[\sum_{r\in\mathbb{W}} u_{ir} h_r^{\mathbb{W}_e}(t-y;\boldsymbol{s})\Big]^k dG_i(y) \\
&= \big(1-G_i(t)\big) + \int_0^t f_i\Big(y; \sum_{r\in\mathbb{W}} u_{ir} h_r^{\mathbb{W}_e}(t-y;\boldsymbol{s})\Big) dG_i(y).
\end{aligned}
$$

In the expressions above whenever $u_{ir}=0$, since there can be no mutations from type i towards type r, we have that the corresponding k_r is also $k_r=0$. Thus, in this case, $u_{ir}^{k_r}=1$. □

Remark 2.7. *It is evident that the steps of the proof of Theorem 2.4 do not depend on the configuration of non-zero u_{ij} nor on $\boldsymbol{s}$. This observation makes it straightforward to derive p.g.f.s for particle production within particular cases of the MSBPM.*

We obtain the p.g.f.s for the entirety of the particle production within a MSBPM via:

Corollary 2.5. *Let $\mathbb{W}_e = \mathbb{W}$. The following system of integral equations hold within the MSBPM, $i \in \mathbb{W}$:*

$$h_i^{\mathbb{W}}(t;\boldsymbol{s}) = \big(1 - G_i(t)\big) + \int_0^t f_i\Big(y; \sum_{r\in\mathbb{W}} u_{ir} s_r h_r^{\mathbb{W}}(t-y;\boldsymbol{s})\Big) dG_i(y). \tag{2.12}$$

Proof. The proof follow immediately from the proof of Theorem 2.4 by setting $\mathbb{W}_e = \mathbb{W}$. □

Systems of integral equations for $h_i^{\mathbb{W}_e}(t;\boldsymbol{s})$, corresponding to particular cases of the MSBPM, where some of the u_{ij} are 0 and/or where we are interested only in the production of particles from $\mathbb{W}_e$ towards a subclass of $\mathbb{W}$, can be immediately obtained from Theorem 2.4. More specifically, this is achieved by setting the necessary u_{ij} to 0 and also setting those coordinates of $\boldsymbol{s}$, that do not correspond to types from the target subclass of $\mathbb{W}$, to 1. We will see how this is done in Subsection 2.3.1 and Subsection 2.3.2.

Corollary 2.6. *Let a MSBPM start with one particle of type i, $i \in \mathbb{W}$, that is of age a, $a \neq 0$. Then the following system of integral equations holds:*

1. *For $i \in \mathbb{W}_e$*
$$h_{i,a}^{\mathbb{W}_e}(t;\boldsymbol{s}) = \big(1 - G_{i,a}(t)\big) + \int_0^t f_i\Big(a+y; \sum_{r\in\mathbb{W}} u_{ir} s_r h_r^{\mathbb{W}_e}(t-y;\boldsymbol{s})\Big) dG_{i,a}(y). \tag{2.13}$$

2. *For $i \notin \mathbb{W}_e$*
$$h_{i,a}^{\mathbb{W}_e}(t;\boldsymbol{s}) = \big(1 - G_{i,a}(t)\big) + \int_0^t f_i\Big(a+y; \sum_{r\in\mathbb{W}} u_{ir} h_r^{\mathbb{W}_e}(t-y;\boldsymbol{s})\Big) dG_{i,a}(y). \tag{2.14}$$

Proof. The proof is completely analogous to the proof of Theorem 2.4, with the exceptions being that when we consider the cases, with respect to the moment of death of the initial particle of type i, we have $G_{i,a}(t)$ instead of $G_i(t)$, and also when writing $p_{ik}(\cdot)$, we must take into account the age, a, of the initial particle. □

Remark 2.8. *We can see from Corollary 2.6 that* $h_{i,a}^{\mathbb{W}_e}(t;\boldsymbol{s})$*, for a fixed* i*,* $i \in \mathbb{W}$*, does not depend on* $h_{k,a}^{\mathbb{W}_e}(t;\boldsymbol{s})$*,* $k \in \mathbb{W}$*.* $h_{i,a}^{\mathbb{W}_e}(t;\boldsymbol{s})$*, however, depends on equations* (2.10) *and* (2.11)*.*

Corollary 2.7. *Let* $\mathbb{W}_e = \mathbb{W}$*. The following system of integral equations hold within the MSBPM,* $i \in \mathbb{W}$*:*

$$h_{i,a}^{\mathbb{W}}(t;\boldsymbol{s}) = \big(1-G_{i,a}(t)\big)+\int_0^t f_i\Big(a+y;\sum_{r\in\mathbb{W}} u_{ir}s_r h_r^{\mathbb{W}}(t-y;\boldsymbol{s})\Big)dG_{i,a}(y). \tag{2.15}$$

Proof. The proof follows immediately from the proof of Corollary 2.6 by setting $\mathbb{W}_e = \mathbb{W}$. □

We can obtain the expected number of particles produced until t, per type, within a MSBPM starting with one particle of type i, that is of age 0, via differentiating $h_i^{\mathbb{W}_e}(t;\boldsymbol{s})$. For obtaining the expected number of produced particles of type j until t, we calculate the left-hand, with respect to $\mathbf{1}$, partial derivative:

$$\mathbb{E}\Big(I_j^{\mathbb{W}_e}(t) \mid \boldsymbol{Z}(0) = \boldsymbol{\delta}^i\Big) = \lim_{\Delta s_j \to 0^+} \frac{h_i^{\mathbb{W}_e}(t;\mathbf{1}) - h_i^{\mathbb{W}_e}(t;\mathbf{1}-\Delta\boldsymbol{s}_j)}{\Delta s_j}, \tag{2.16}$$

where $\Delta\boldsymbol{s}_j = (0,\dots,\Delta s_j,\dots,0)^\top$. The same line of reasoning is valid when the initial particle is of age a, $a \neq 0$ - for obtaining the expected number of produced particles of type j until t, we calculate the left-hand, with respect to $\mathbf{1}$, partial derivative:

$$\mathbb{E}\Big(I_j^{\mathbb{W}_e}(t) \mid \boldsymbol{Z}(0) = \boldsymbol{\delta}_a^i\Big) = \lim_{\Delta s_j \to 0^+} \frac{h_{i,a}^{\mathbb{W}_e}(t;\mathbf{1}) - h_{i,a}^{\mathbb{W}_e}(t;\mathbf{1}-\Delta\boldsymbol{s}_j)}{\Delta s_j}. \tag{2.17}$$

Numerical Scheme 1 and Numerical Scheme 2, presented in Section 2.2.7, can be applied to the numerators in the right sides of equations (2.16) and (2.17) respectively, thus we will not develop these equations any further.

Next, we investigate $I_j^{\mathbb{W}_e}(t)$ and $h_i^{\mathbb{W}_e}(t;\boldsymbol{s})$ as $t \to \infty$.

Definition 2.7. *Denote with* $I_j^{\mathbb{W}_e}$ *the number of particles (mutants or not) of type* j*,* $j \in \mathbb{W}$*, produced from particles with types from* $\mathbb{W}_e$ *during the whole MSBPM. We do not count the initial particle within any of the*

$I_j^{\mathbb{W}_e}$. *For a MSBPM starting with one particle of type i, $i \in \mathbb{W}$, that is of age* 0*, denote with* $h_i^{\mathbb{W}_e}(\boldsymbol{s})$ *the following p.g.f.*

$$h_i^{\mathbb{W}_e}(\boldsymbol{s}) = \mathbb{E}\Big(\prod_{j\in\mathbb{W}} s_j^{I_j^{\mathbb{W}_e}} \mid \boldsymbol{Z}(0) = \boldsymbol{\delta}^i \Big),$$

where $|\boldsymbol{s}| \le 1$*. We denote the corresponding p.g.f., when the MSBPM starts with one particle of type i, $i \in \mathbb{W}$, that is of age a, $a \neq 0$, with*

$$h_{i,a}^{\mathbb{W}_e}(\boldsymbol{s}) = \mathbb{E}\Big(\prod_{j\in\mathbb{W}} s_j^{I_j^{\mathbb{W}_e}} \mid \boldsymbol{Z}(0) = \boldsymbol{\delta}_a^i \Big),$$

where $|\boldsymbol{s}| \le 1$*.*

Remark 2.9. *From Definition 2.7 it is evident that* $I_j^{\mathbb{W}_e} := \lim_{t\to\infty} I_j^{\mathbb{W}_e}(t)$ *almost surely. Considering this and the fact that there is a one-to-one correspondence between r.v.s and p.g.f.s, it follows that* $h_i^{\mathbb{W}_e}(\boldsymbol{s}) = \lim_{t\to\infty} h_i^{\mathbb{W}_e}(t;\boldsymbol{s})$ *and* $h_{i,a}^{\mathbb{W}_e}(\boldsymbol{s}) = \lim_{t\to\infty} h_{i,a}^{\mathbb{W}_e}(t;\boldsymbol{s})$*.*

Definition 2.8. *We define* $\lim_{y\to\infty}\lim_{t\to\infty} I_j^{\mathbb{W}_e}(t-y) = I_j^{\mathbb{W}_e}$*. Consequently* $\lim_{y\to\infty}\lim_{t\to\infty} h_i^{\mathbb{W}_e}(t-y;\boldsymbol{s}) = h_i^{\mathbb{W}_e}(\boldsymbol{s})$ *and* $\lim_{y\to\infty}\lim_{t\to\infty} h_{i,a}^{\mathbb{W}_e}(t-y;\boldsymbol{s}) = h_{i,a}^{\mathbb{W}_e}(\boldsymbol{s})$*.*

Remark 2.10. *Definition 2.8 is relevant for the proof we provide for Theorem 2.5 below. We now give more details about our considerations. Let us begin by investigating a MSBPM at some finite moment t, let a particle be created within the process at some finite y, $y \le t$. The particle created at y starts an independent copy process that has (finite) time interval of length $t-y$ in order to produce particles from $\mathbb{W}_e$ towards $j \in \mathbb{W}$, thus we work with* $\widetilde{I}_j^{\mathbb{W}_e}(t-y)$ *- no problems to account for in this case. Next, as per Remark 2.9, if we let $t \to \infty$ while y remains fixed, we now consider* $\widetilde{I}_j^{\mathbb{W}_e} = \lim_{t\to\infty} \widetilde{I}_j^{\mathbb{W}_e}(t-y)$ *regardless of the value of our fixed y - in this case as well no additional problems arise. However, if we let $y \to \infty$ in addition to $t \to \infty$, we now face the question of how to interpret, with respect to $t \to \infty$, a process that starts from a particle created at $y \to \infty$. If we set* $\lim_{y\to\infty}\lim_{t\to\infty} \widetilde{I}_j^{\mathbb{W}_e}(t-y) = \widetilde{I}_j^{\mathbb{W}_e}(b)$*, where $b \in [0,\infty)$, we would, quite arbitrarily, be giving the length of the time interval during which the copy process can produce particles from $\mathbb{W}_e$ towards $j \in \mathbb{W}$. On the other hand,* *if we set* $\lim_{y\to\infty}\lim_{t\to\infty}\widetilde{I}_j^{\mathbb{W}_e}(t-y)=\widetilde{I}_j^{\mathbb{W}_e}$, *we would be stating that regardless of the starting moment of the copy process, the length of the time interval, with respect to* $t\to\infty$, *during which the copy process produces particles from* $\mathbb{W}_e$ *towards* $j\in\mathbb{W}$, *is infinite. We deem that the latter interpretation is more appropriate in the context of the MSBPM.*

Theorem 2.5. *The following system of equations holds within the MSBPM,* $i\in\mathbb{W}$*:*

1. *Let* $i\in\mathbb{W}_e$. *Then*

$$h_i^{\mathbb{W}_e}(\boldsymbol{s})=\int_0^\infty f_i\Big(y;\sum_{r\in\mathbb{W}}u_{ir}s_rh_r^{\mathbb{W}_e}(\boldsymbol{s})\Big)dG_i(y). \tag{2.18}$$

2. *Let* $i\notin\mathbb{W}_e$. *Then*

$$h_i^{\mathbb{W}_e}(\boldsymbol{s})=\int_0^\infty f_i\Big(y;\sum_{r\in\mathbb{W}}u_{ir}h_r^{\mathbb{W}_e}(\boldsymbol{s})\Big)dG_i(y). \tag{2.19}$$

Proof.

1. Let $i\in\mathbb{W}_e$. We use the result of Theorem 2.4:

$$\begin{aligned}h_i^{\mathbb{W}_e}(\boldsymbol{s})&=\lim_{t\to\infty}h_i^{\mathbb{W}_e}(t;\boldsymbol{s})\\&=\lim_{t\to\infty}\big(1-G_i(t)\big)+\lim_{t\to\infty}\int_0^t f_i\Big(y;\sum_{r\in\mathbb{W}}u_{ir}s_rh_r^{\mathbb{W}_e}(t-y;\boldsymbol{s})\Big)dG_i(y)\\&=\lim_{t\to\infty}\int_{-\infty}^\infty I_{[0,t]}(y)\cdot f_i\Big(y;\sum_{r\in\mathbb{W}}u_{ir}s_rh_r^{\mathbb{W}_e}(t-y;\boldsymbol{s})\Big)dG_i(y),\end{aligned}$$

where $I_{[0,t]}(y)$ is the indicator function of $[0,t]$. Since (for any $i\in\mathbb{W}_e$) it is true that the absolute value of the integrand is ≤ 1, by virtue of the dominated convergence theorem, we can pass the limit inside the integral. Thus, taking into account Definition 2.8 (and Remark 2.10), we obtain

$$h_i^{\mathbb{W}_e}(\boldsymbol{s})=\int_0^\infty f_i\Big(y;\sum_{r\in\mathbb{W}}u_{ir}s_rh_r^{\mathbb{W}_e}(\boldsymbol{s})\Big)dG_i(y).$$

2. Let $i \notin \mathbb{W}_e$. The proof is completely analogous to the proof for the case of $i \in \mathbb{W}$.

$\square$

We obtain the p.g.f.s for the entirety of the particle production within a MSBPM via:

Corollary 2.8. *Let $\mathbb{W}_e = \mathbb{W}$. The following system of integral equations holds within the MSBPM, $i \in \mathbb{W}$:*

$$h_i^{\mathbb{W}}(\boldsymbol{s}) = \int_0^\infty f_i\Big(y; \sum_{r\in\mathbb{W}} u_{ir} s_r h_r^{\mathbb{W}}(\boldsymbol{s})\Big) dG_i(y). \tag{2.20}$$

Proof. The proof follow immediately from the proof of Theorem 2.5 by setting $\mathbb{W}_e = \mathbb{W}$. $\square$

Corollary 2.9. *Let a MSBPM start with one particle of type i, $i \in \mathbb{W}$, that is of age a, $a \neq 0$. Then the following system of integral equations holds:*

1. *Let $i \in \mathbb{W}_e$. Then*

$$h_{i,a}^{\mathbb{W}_e}(\boldsymbol{s}) = \int_0^\infty f_i\Big(a + y; \sum_{r\in\mathbb{W}} u_{ir} s_r h_r^{\mathbb{W}_e}(\boldsymbol{s})\Big) dG_{i,a}(y). \tag{2.21}$$

2. *Let $i \notin \mathbb{W}_e$. Then*

$$h_{i,a}^{\mathbb{W}_e}(\boldsymbol{s}) = \int_0^\infty f_i\Big(a + y; \sum_{r\in\mathbb{W}} u_{ir} h_r^{\mathbb{W}_e}(\boldsymbol{s})\Big) dG_{i,a}(y). \tag{2.22}$$

Proof. We take the results from Corollary 2.6 and follow the steps of the proof of Theorem 2.5. $\square$

Remark 2.11. *We can see from Corollary 2.9 that $h_{i,a}^{\mathbb{W}_e}(\boldsymbol{s})$, for a fixed i, $i \in \mathbb{W}$, does not depend on $h_{k,a}^{\mathbb{W}_e}(\boldsymbol{s})$, $k \in \mathbb{W}$. $h_{i,a}^{\mathbb{W}_e}(\boldsymbol{s})$, however, depends on equations* (2.18) *and* (2.19)*.*

Corollary 2.10. *Let $\mathbb{W}_e = \mathbb{W}$. The following system of integral equations holds within the MSBPM, $i \in \mathbb{W}$:*

$$h_{i,a}^{\mathbb{W}}(\boldsymbol{s}) = \int_0^\infty f_i\Big(a + y; \sum_{r\in\mathbb{W}} u_{ir} s_r h_r^{\mathbb{W}}(\boldsymbol{s})\Big) dG_{i,a}(y). \tag{2.23}$$

Proof. The proof follow immediately from the proof of Corollary 2.9 by setting $\mathbb{W}_e = \mathbb{W}$. □

Corollary 2.11. *In the particular case where there is no dependence of particle reproduction from particle age, i.e., $f_i(y; s) = f_i(s)$, $i \in \mathbb{W}$, the system of integral equations* (2.18), (2.19) *from Theorem 2.5, and the system of integral equations* (2.21), (2.22) *from Corollary 2.9, become the following system of equations:*

1. *Let $i \in \mathbb{W}_e$. Then*

$$h_i^{\mathbb{W}_e}(\boldsymbol{s}) = h_{i,a}^{\mathbb{W}_e}(\boldsymbol{s}) = f_i\Big(\sum_{r\in\mathbb{W}} u_{ir} s_r h_r^{\mathbb{W}_e}(\boldsymbol{s})\Big). \tag{2.24}$$

2. *Let $i \notin \mathbb{W}_e$. Then*

$$h_i^{\mathbb{W}_e}(\boldsymbol{s}) = h_{i,a}^{\mathbb{W}_e}(\boldsymbol{s}) = f_i\Big(\sum_{r\in\mathbb{W}} u_{ir} h_r^{\mathbb{W}_e}(\boldsymbol{s})\Big). \tag{2.25}$$

Proof. In order to complete the proof we must simply take the results from Theorem 2.5 and Corollary 2.9, and we drop the dependence from y within f_i:

1. Let $i \in \mathbb{W}_e$. We have

$$\begin{aligned} h_i^{\mathbb{W}_e}(\boldsymbol{s}) &= \int_0^\infty f_i\Big(\sum_{r\in\mathbb{W}} u_{ir} s_r h_r^{\mathbb{W}_e}(\boldsymbol{s})\Big) dG_i(y) \\ &= f_i\Big(\sum_{r\in\mathbb{W}} u_{ir} s_r h_r^{\mathbb{W}_e}(\boldsymbol{s})\Big) \int_0^\infty dG_i(y) \\ &= f_i\Big(\sum_{r\in\mathbb{W}} u_{ir} s_r h_r^{\mathbb{W}_e}(\boldsymbol{s})\Big) \end{aligned}$$

and

$$\begin{aligned} h_{i,a}^{\mathbb{W}_e}(\boldsymbol{s}) &= \int_0^\infty f_i\Big(\sum_{r\in\mathbb{W}} u_{ir} s_r h_r^{\mathbb{W}_e}(\boldsymbol{s})\Big) dG_{i,a}(y) \\ &= f_i\Big(\sum_{r\in\mathbb{W}} u_{ir} s_r h_r^{\mathbb{W}_e}(\boldsymbol{s})\Big) \int_0^\infty dG_{i,a}(y) \\ &= f_i\Big(\sum_{r\in\mathbb{W}} u_{ir} s_r h_r^{\mathbb{W}_e}(\boldsymbol{s})\Big). \end{aligned}$$

2. Let $i \notin \mathbb{W}_e$. The steps of the proof are completely analogous to the steps of the proof for the case of $i \in \mathbb{W}_e$.

$\square$

The results, displayed in Figure 2.4 and Figure 2.5 below, suggest that a general theorem for $h_i^{\mathbb{W}_e}(\boldsymbol{s}) = h_{i,a}^{\mathbb{W}_e}(\boldsymbol{s})$ may exist. For Figure 2.4 and Figure 2.5 we have used $\mathbb{W}_e = \{4, 5, 6\}$.

We further give Figure 2.6 and Figure 2.7 in order to demonstrate that the result of Proposition 2.1 (Section 2.3.1.4) is generally not true for an arbitrary $\mathbb{W}_0$. For Figure 2.6 and Figure 2.7 below, we have also used $\mathbb{W}_e = \{4, 5, 6\}$ and with respect to Proposition 2.1 we consider $\mathbb{W}_0 = \mathbb{W}$ and $\boldsymbol{q}_{\mathbb{W}_0} = \boldsymbol{q}$. We note that although largely similar, Figure 2.6 and Figure 2.7 depict slightly different behavior for $t \in [0, 100]$.

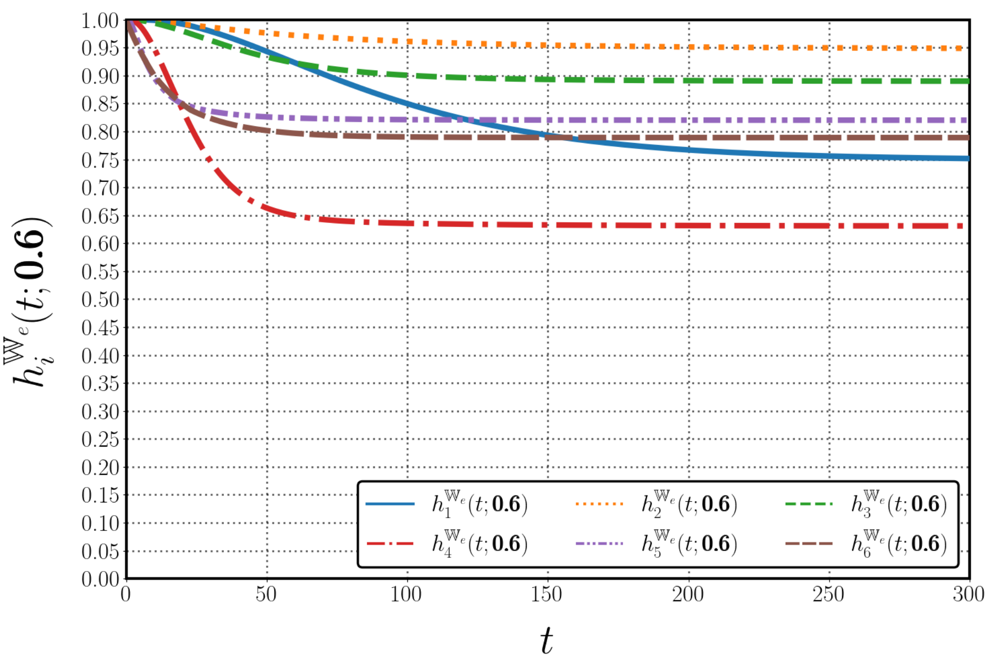

Figure 2.4: An application of Theorem 2.4 for the example MSBPM (Table 2.11, Table 2.6) with mutation scheme "$\mathbb{W}$ towards $\mathbb{W}$" (Table 2.7) with $\mathbb{W}_e = \{4, 5, 6\}$ and starting with one particle of age 0. At $t = 5000$ ($h = 10^{-1}$), we have $h_1^{\mathbb{W}_e}(\mathbf{0.6}) \approx 0.74851297$, $h_2^{\mathbb{W}_e}(\mathbf{0.6}) \approx 0.94742302$, $h_3^{\mathbb{W}_e}(\mathbf{0.6}) \approx 0.88973182$, $h_4^{\mathbb{W}_e}(\mathbf{0.6}) \approx 0.63076144$, $h_5^{\mathbb{W}_e}(\mathbf{0.6}) \approx 0.82006882$, $h_6^{\mathbb{W}_e}(\mathbf{0.6}) \approx 0.78891955$.

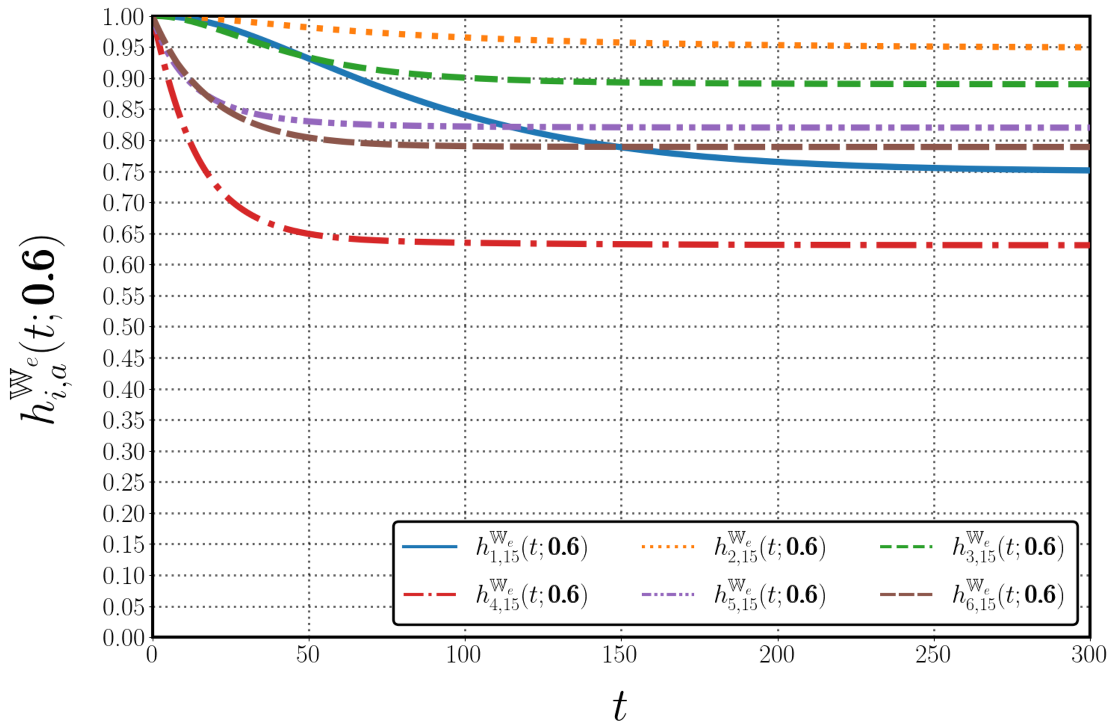

Figure 2.5: An application of Corollary 2.6 - probabilities for extinction for the example MSBPM (Table 2.11, Table 2.6) with mutation scheme "$\mathbb{W}$ towards $\mathbb{W}$" (Table 2.7) with $\mathbb{W}_e = \{4, 5, 6\}$ and starting with one particle of age $a = 15$. At $t = 5000$ ($h = 10^{-1}$), we have $h^{\mathbb{W}_e}_{1,15}(\mathbf{0.6}) \approx 0.74851297$, $h^{\mathbb{W}_e}_{2,15}(\mathbf{0.6}) \approx 0.94742315$, $h^{\mathbb{W}_e}_{3,15}(\mathbf{0.6}) \approx 0.88973182$, $h^{\mathbb{W}_e}_{4,15}(\mathbf{0.6}) \approx 0.63076144$, $h^{\mathbb{W}_e}_{5,15}(\mathbf{0.6}) \approx 0.82006882$, $h^{\mathbb{W}_e}_{6,15}(\mathbf{0.6}) \approx 0.78891955$.

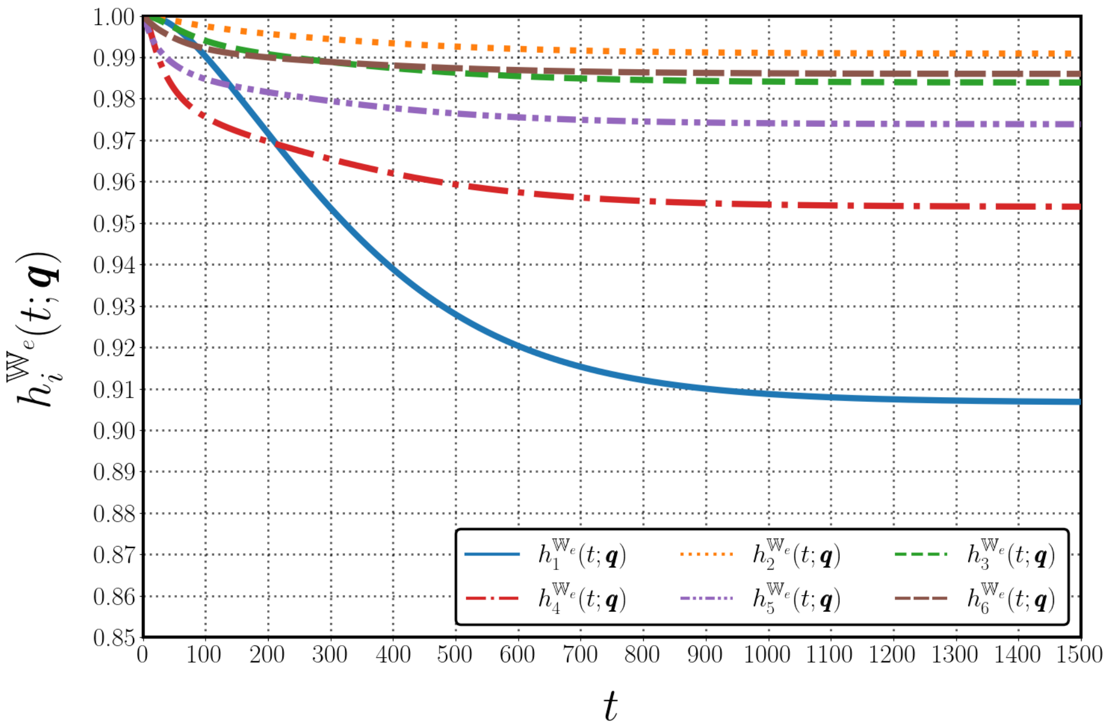

Figure 2.6: Calculations ($h = 10^{-2}$) for $\boldsymbol{s} = \boldsymbol{q}$ for the example MSBPM (Table 2.11, Table 2.6) with mutation scheme "$\mathbb{W}$ towards $\mathbb{W}$" (Table 2.7) with $\mathbb{W}_e = \{4,5,6\}$, $\mathbb{W}_0 = \mathbb{W}$, and starting with one particle of age 0. $\boldsymbol{q}$ is the vector with probabilities of extinction calculated for Figure 2.2, i.e., $\boldsymbol{q} = (0.96009038, 0.99744155,\ 0.99602606,\ 0.99015646, 0.99464378, 0.99762545)^\top$. At $t = 1500$, we have $h_1^{\mathbb{W}_e}(\boldsymbol{q}) = 0.9068164$, $h_2^{\mathbb{W}_e}(\boldsymbol{q}) = 0.99088022$, $h_3^{\mathbb{W}_e}(\boldsymbol{q}) = 0.98390969$, $h_4^{\mathbb{W}_e}(\boldsymbol{q}) = 0.95394249$, $h_5^{\mathbb{W}_e}(\boldsymbol{q}) = 0.97383695$, $h_6^{\mathbb{W}_e}(\boldsymbol{q}) = 0.98599877$. Evidently, the statement $\boldsymbol{q}_i = h_i^{\mathbb{W}_e}(\boldsymbol{q})$ for $i \in \mathbb{W}_e$ is not true.

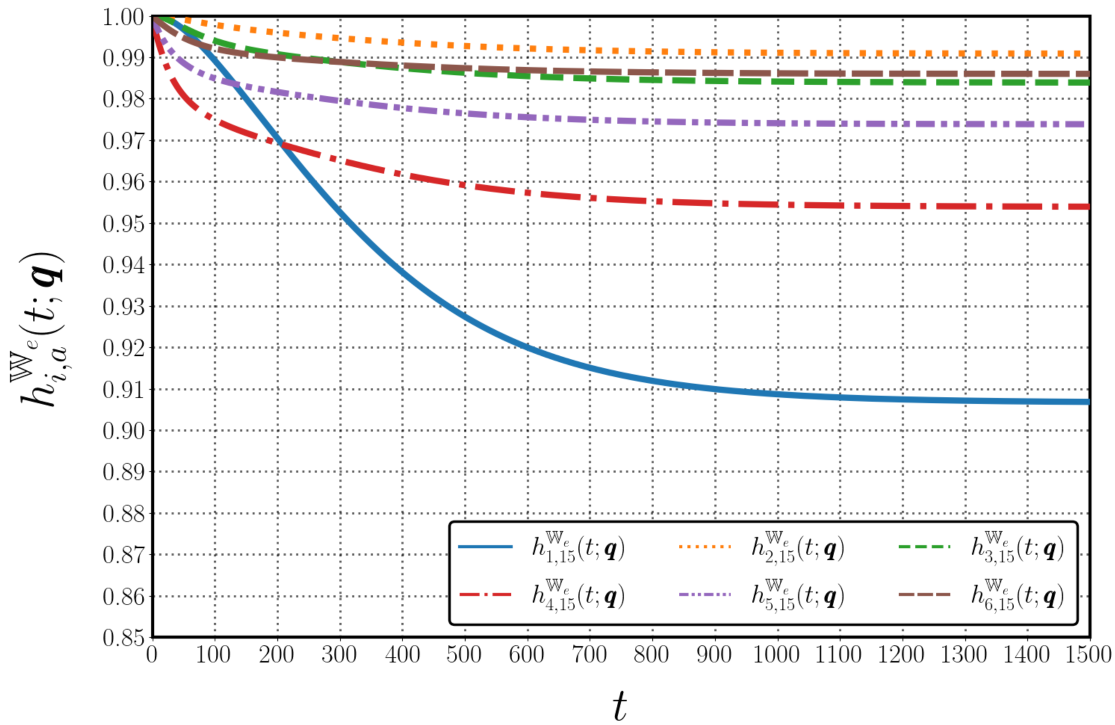

Figure 2.7: Calculations ($h = 10^{-2}$) for $\boldsymbol{s} = \boldsymbol{q}$ for the example MSBPM (Table 2.11, Table 2.6) with mutation scheme "$\mathbb{W}$ towards $\mathbb{W}$" (Table 2.7) with $\mathbb{W}_e = \{4, 5, 6\}$, $\mathbb{W}_0 = \mathbb{W}$, and starting with one particle of age $a = 15$. $\boldsymbol{q}$ is the vector with probabilities of extinction calculated for Figure 2.2, i.e., $\boldsymbol{q} = (0.96009038, 0.99744155,\ 0.99602606,\ 0.99015646, 0.99464378, 0.99762545)^\top$. At $t = 1500$, we have $h_{1,15}^{\mathbb{W}_e}(\boldsymbol{q}) = 0.90681099$, $h_{2,15}^{\mathbb{W}_e}(\boldsymbol{q}) = 0.99088474$, $h_{3,15}^{\mathbb{W}_e}(\boldsymbol{q}) = 0.98390969$, $h_{4,15}^{\mathbb{W}_e}(\boldsymbol{q}) = 0.95394063$, $h_{5,15}^{\mathbb{W}_e}(\boldsymbol{q}) = 0.97383717$, $h_{6,15}^{\mathbb{W}_e}(\boldsymbol{q}) = 0.98599877$. Evidently, the statement $\boldsymbol{q}_i = h_{i,a}^{\mathbb{W}_e}(\boldsymbol{q})$ for $i \in \mathbb{W}_e$ is not true.

2.2.5 Time until occurrence of the first "successful" particle produced from $\mathbb{W}_e$ towards $\mathbb{W}$ within the MSBPM

We call a particle produced from $\mathbb{W}_e$ towards $\mathbb{W}$ "successful" if it initiates a non-extincting MSBPM.

Definition 2.9. *Denote with* $T_{\mathbb{W}}^{\mathbb{W}_e}$ *the r.v. that is the time until occurrence of the first "successful" particle produced from a type within* $\mathbb{W}_e$ *towards a type within* $\mathbb{W}$ *in a MSBPM starting with some combination of particles with types within* $\mathbb{W}_e$*. Without loss of generality, we set the starting number of particles per type* $r \in \mathbb{W}_e$ *to be* k_r *and the starting number of particles per type* $r \in \mathbb{W} \setminus \mathbb{W}_e$ *to be* 0*. We denote the so specified initial state of the process as* $\boldsymbol{Z}(0) = \boldsymbol{\alpha}^*$*. We define* $T_{\mathbb{W}}^{\mathbb{W}_e} = \infty$ *as the event that no "successful" particles have been produced from* $\mathbb{W}_e$ *towards* $\mathbb{W}$ *in a MSBPM beginning with an initial state* $\boldsymbol{\alpha}^*$*. Thus, we may write* $T_{\mathbb{W}}^{\mathbb{W}_e} \in (0, \infty]$*. If the MSBPM starts with a single particle of type* i*,* $i \in \mathbb{W}_e$*, of age* 0*, we use* $T_{\mathbb{W},i}^{\mathbb{W}_e}$ *as a shortcut notation. If the initial particle is of age* a*,* $a \neq 0$*, we use* $T_{\mathbb{W},i,a}^{\mathbb{W}_e}$*.*

Theorem 2.6. *Let the MSBPM start with* k_r *particles per type* r*,* $r \in \mathbb{W}_e$*. Let all particles form* $\boldsymbol{\alpha}^*$ *have age* 0*. The distribution of* $T_{\mathbb{W}}^{\mathbb{W}_e}$ *has the following properties:*

(i) $\mathbb{P}\Big(T_{\mathbb{W}}^{\mathbb{W}_e} > t \mid \boldsymbol{Z}(0) = \boldsymbol{\alpha}^*\Big) = \prod_{r \in \mathbb{W}_e} \Big[h_r^{\mathbb{W}_e}(t; \boldsymbol{q})\Big]^{k_r}$.

(ii) $\mathbb{P}\Big(T_{\mathbb{W}}^{\mathbb{W}_e} = \infty \mid \boldsymbol{Z}(0) = \boldsymbol{\alpha}^*\Big) = \prod_{r \in \mathbb{W}_e} \Big[h_r^{\mathbb{W}_e}(\boldsymbol{q})\Big]^{k_r}$.

(iii) If at least one particle type within $\mathbb{W}$ *is supercritical, we have*

$$\mathbb{E}\Big[T_{\mathbb{W}}^{\mathbb{W}_e} \mid T_{\mathbb{W}}^{\mathbb{W}_e} < \infty,\ \boldsymbol{Z}(0) = \boldsymbol{\alpha}^*\Big] =$$

$$= \frac{1}{1 - \prod_{r \in \mathbb{W}_e} \Big[h_r^{\mathbb{W}_e}(\boldsymbol{q})\Big]^{k_r}} \int_0^\infty \left[\prod_{r \in \mathbb{W}_e} \Big[h_r^{\mathbb{W}_e}(t; \boldsymbol{q})\Big]^{k_r} - \right.$$

$$\left. - \prod_{r \in \mathbb{W}_e} \Big[h_r^{\mathbb{W}_e}(\boldsymbol{q})\Big]^{k_r} \right] dt,$$

if not, then the expectation does not exist.

Proof.
Property (i): Let the process start with a single particle of type i, $i \in \mathbb{W}_e$. The event $\Big\{T_{\mathbb{W},i}^{\mathbb{W}_e} > t\Big\}$ means that if we consider the separate MSBPM stemming from particles produced from $\mathbb{W}_e$ towards $\mathbb{W}$, that have come

into existence prior to t or at t, all those processes become extinct. Thus, by the law of total probability:

$$\mathbb{P}\Big(T^{\mathbb{W}_e}_{\mathbb{W},i} > t\Big) =$$
$$= \sum_{k=0}^{\infty} \Bigg[\sum_{\sum_{r\in\mathbb{W}} k_r = k} \Big[\mathbb{P}\Big(I^{\mathbb{W}_e}_j(t) = k_j,\ j \in \mathbb{W} \mid \boldsymbol{Z}(0) = \boldsymbol{\delta}^i\Big) \cdot \prod_{m\in\mathbb{W}} q_m^{k_m}\Big]\Bigg]$$
$$= h_i^{\mathbb{W}_e}(t;\boldsymbol{q}).$$

The result for $\boldsymbol{\alpha}^*$ follows from the assumption of independent evolution.
Property (ii): Let the process start with a single particle of type i, $i \in \mathbb{W}_e$. We have

$$\mathbb{P}\Big(T^{\mathbb{W}_e}_{\mathbb{W},i} = \infty\Big) = \lim_{t\to\infty} \mathbb{P}\Big(T^{\mathbb{W}_e}_{\mathbb{W},i} > t\Big) = \lim_{t\to\infty} h_i^{\mathbb{W}_e}(t;\boldsymbol{q}) = h_i^{\mathbb{W}_e}(\boldsymbol{q}).$$

The result for $\boldsymbol{\alpha}^*$ follows from the assumption of independent evolution.
Property (iii):

$$\mathbb{E}\Big[T^{\mathbb{W}_e}_{\mathbb{W}} \mid T^{\mathbb{W}_e}_{\mathbb{W}} < \infty,\ \boldsymbol{Z}(0) = \boldsymbol{\alpha}^*\Big] =$$
$$= \int_0^\infty \Big[1 - \mathbb{P}\big(T^{\mathbb{W}_e}_{\mathbb{W}} \le t | T^{\mathbb{W}_e}_{\mathbb{W}} < \infty, \boldsymbol{Z}(0) = \boldsymbol{\alpha}^*\big)\Big] dt$$
$$= \int_0^\infty \Bigg[1 - \frac{\mathbb{P}\big(T^{\mathbb{W}_e}_{\mathbb{W}} \le t, T^{\mathbb{W}_e}_{\mathbb{W}} < \infty | \boldsymbol{Z}(0) = \boldsymbol{\alpha}^*\big)}{1 - \prod_{r\in\mathbb{W}_e}\Big[h_r^{\mathbb{W}_e}(\boldsymbol{q})\Big]^{k_r}}\Bigg]$$
$$= \frac{1}{1 - \prod_{r\in\mathbb{W}_e}\Big[h_r^{\mathbb{W}_e}(\boldsymbol{q})\Big]^{k_r}} \int_0^\infty \Bigg[1 - \prod_{r\in\mathbb{W}_e}\Big[h_r^{\mathbb{W}_e}(\boldsymbol{q})\Big]^{k_r} - \mathbb{P}\big(T^{\mathbb{W}_e}_{\mathbb{W}} \le t | \boldsymbol{Z}(0) = \boldsymbol{\alpha}^*\big)\Bigg] dt$$
$$= \frac{1}{1 - \prod_{r\in\mathbb{W}_e}\Big[h_r^{\mathbb{W}_e}(\boldsymbol{q})\Big]^{k_r}} \int_0^\infty \Bigg[\mathbb{P}\big(T^{\mathbb{W}_e}_{\mathbb{W}} > t | \boldsymbol{Z}(0) = \boldsymbol{\alpha}^*\big) - \prod_{r\in\mathbb{W}_e}\Big[h_r^{\mathbb{W}_e}(\boldsymbol{q})\Big]^{k_r}\Bigg] dt$$
$$= \frac{1}{1 - \prod_{r\in\mathbb{W}_e}\Big[h_r^{\mathbb{W}_e}(\boldsymbol{q})\Big]^{k_r}} \int_0^\infty \Bigg[\prod_{r\in\mathbb{W}_e}\Big[h_r^{\mathbb{W}_e}(t;\boldsymbol{q})\Big]^{k_r} - \prod_{r\in\mathbb{W}_e}\Big[h_r^{\mathbb{W}_e}(\boldsymbol{q})\Big]^{k_r}\Bigg] dt.$$

Note that if each type from $\mathbb{W}$ is either subcritical or critical, then $\boldsymbol{q} = \boldsymbol{1}$ and $1 - \prod_{r\in\mathbb{W}_e}\Big[h_r^{\mathbb{W}_e}(\boldsymbol{q})\Big]^{k_r} = 0$. $\square$

Now that we have proven Theorem 2.6 for the notationally more simple case of all particles in $\boldsymbol{\alpha}^*$ having age 0, we can more easily prove the more general:

Theorem 2.7. *Let the MSBPM start with k_r particles per type r, $r \in \mathbb{W}_e$, let the starting particles in $\boldsymbol{\alpha}^*$ have ages $a_{r,c}$, $c \in \{1, 2, \dots, k_r\}$, where $a_{r,c}$ is the age of the c-th particle of type r. We allow $a_{r,c}$ to be 0. The distribution of $T_{\mathbb{W}}^{\mathbb{W}_e}$ has the following properties:*

(i) $\mathbb{P}\Big(T_{\mathbb{W}}^{\mathbb{W}_e} > t \mid \boldsymbol{Z}(0) = \boldsymbol{\alpha}^*\Big) = \prod_{r\in\mathbb{W}_e} \Big[\prod_{c=1}^{k_r} h_{r,a_{r,c}}^{\mathbb{W}_e}(t;\boldsymbol{q})\Big]$.

(ii) $\mathbb{P}\Big(T_{\mathbb{W}}^{\mathbb{W}_e} = \infty \mid \boldsymbol{Z}(0) = \boldsymbol{\alpha}^*\Big) = \prod_{r\in\mathbb{W}_e} \Big[\prod_{c=1}^{k_r} h_{r,a_{r,c}}^{\mathbb{W}_e}(\boldsymbol{q})\Big]$.

(iii) If at least one particle type within $\mathbb{W}$ is supercritical, we have

$$\mathbb{E}\Big[T_{\mathbb{W}}^{\mathbb{W}_e} \mid T_{\mathbb{W}}^{\mathbb{W}_e} < \infty,\ \boldsymbol{Z}(0) = \boldsymbol{\alpha}^*\Big] =$$
$$= \frac{1}{1 - \prod_{r\in\mathbb{W}_e}\Big[\prod_{c=1}^{k_r} h_{r,a_{r,c}}^{\mathbb{W}_e}(\boldsymbol{q})\Big]} \int_0^\infty \Bigg[\prod_{r\in\mathbb{W}_e}\Big[\prod_{c=1}^{k_r} h_{r,a_{r,c}}^{\mathbb{W}_e}(t;\boldsymbol{q})\Big] -$$
$$- \prod_{r\in\mathbb{W}_e}\Big[\prod_{c=1}^{k_r} h_{r,a_{r,c}}^{\mathbb{W}_e}(\boldsymbol{q})\Big]\Bigg] dt,$$

if not, then the expectation does not exist.

Proof. Property (i): Let the process start with a single particle of type i, $i \in \mathbb{W}_e$, that is of age $a_{i,c}$, $c \in \{1, 2, \dots, k_i\}$. The event $\Big\{T_{\mathbb{W},i,a_{i,c}}^{\mathbb{W}_e} > t\Big\}$ means that if we consider the separate MSBPM stemming from particles produced from $\mathbb{W}_e$ towards $\mathbb{W}$, that have come into existence prior to t or at t, all those processes become extinct. Thus, by the law of total probability:

$$\mathbb{P}\Big(T_{\mathbb{W},i,a_{i,c}}^{\mathbb{W}_e} > t\Big) =$$
$$= \sum_{k=0}^{\infty}\Bigg[\sum_{\sum_{r\in\mathbb{W}} k_r = k}\Big[\mathbb{P}\Big(I_j^{\mathbb{W}_e}(t) = k_j,\ j\in\mathbb{W} \mid \boldsymbol{Z}(0) = \boldsymbol{\delta}_{a_{i,c}}^i\Big)\cdot \prod_{m\in\mathbb{W}} q_m^{k_m}\Big]\Bigg]$$
$$= h_{i,a_{i,c}}^{\mathbb{W}_e}(t;\boldsymbol{q}).$$

The result for $\boldsymbol{\alpha}^*$ follows from the assumption of independent evolution. The proofs for Property (ii) and Property (iii) are analogous to the proofs of the corresponding properties in Theorem 2.6. $\square$

We illustrate Theorem 2.6 and Theorem 2.7 with Figure 2.8 and Figure 2.9 below. In our experimental setup the age of the initial particle does not play a significant role in the long run. However, there is some difference in the behavior for t close to 0.

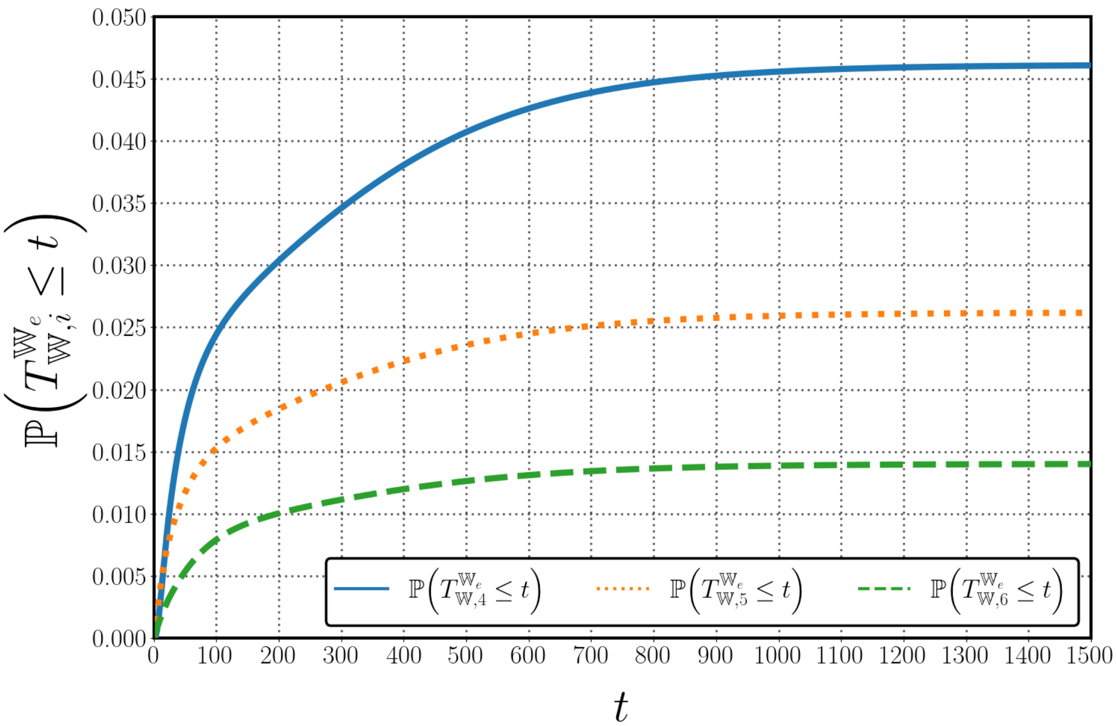

Figure 2.8: An application of Theorem 2.6 - calculations ($h = 10^{-2}$) for $\boldsymbol{s} = \boldsymbol{q}$ for the example MSBPM (Table 2.11, Table 2.6) with mutation scheme "$\mathbb{W}$ towards $\mathbb{W}$" (Table 2.7) with $\mathbb{W}_e = \{4, 5, 6\}$ starting with one particle of age 0. $\boldsymbol{q}$ is the vector with probabilities of extinction calculated for Figure 2.2, i.e., $\boldsymbol{q} = (0.96009038, 0.99744155, 0.99602606, 0.99015646, 0.99464378, 0.99762545)^\top$. $P\left(T_{\mathbb{W},i}^{\mathbb{W}_e} \leq t\right) = 1 - h_i^{\mathbb{W}_e}(t; \boldsymbol{q})$, we can reuse the calculations done for Figure 2.6. At $t = 1500$, we have $P\left(T_{\mathbb{W},4}^{\mathbb{W}_e} \leq t\right) = 0.04605751$, $P\left(T_{\mathbb{W},5}^{\mathbb{W}_e} \leq t\right) = 0.02616305$, $P\left(T_{\mathbb{W},6}^{\mathbb{W}_e} \leq t\right) = 0.01400123$. As $t = 1500$ is sufficiently large as to conclude that $h_i^{\mathbb{W}_e}(1500; \boldsymbol{q}) = h_i^{\mathbb{W}_e}(\boldsymbol{q})$, we also have $P\left(T_{\mathbb{W},4}^{\mathbb{W}_e} = \infty\right) = h_4^{\mathbb{W}_e}(\boldsymbol{q}) = 0.95394249$, $P\left(T_{\mathbb{W},5}^{\mathbb{W}_e} = \infty\right) = h_5^{\mathbb{W}_e}(\boldsymbol{q}) = 0.97383695$, $P\left(T_{\mathbb{W},6}^{\mathbb{W}_e} = \infty\right) = h_6^{\mathbb{W}_e}(\boldsymbol{q}) = 0.98599877$.

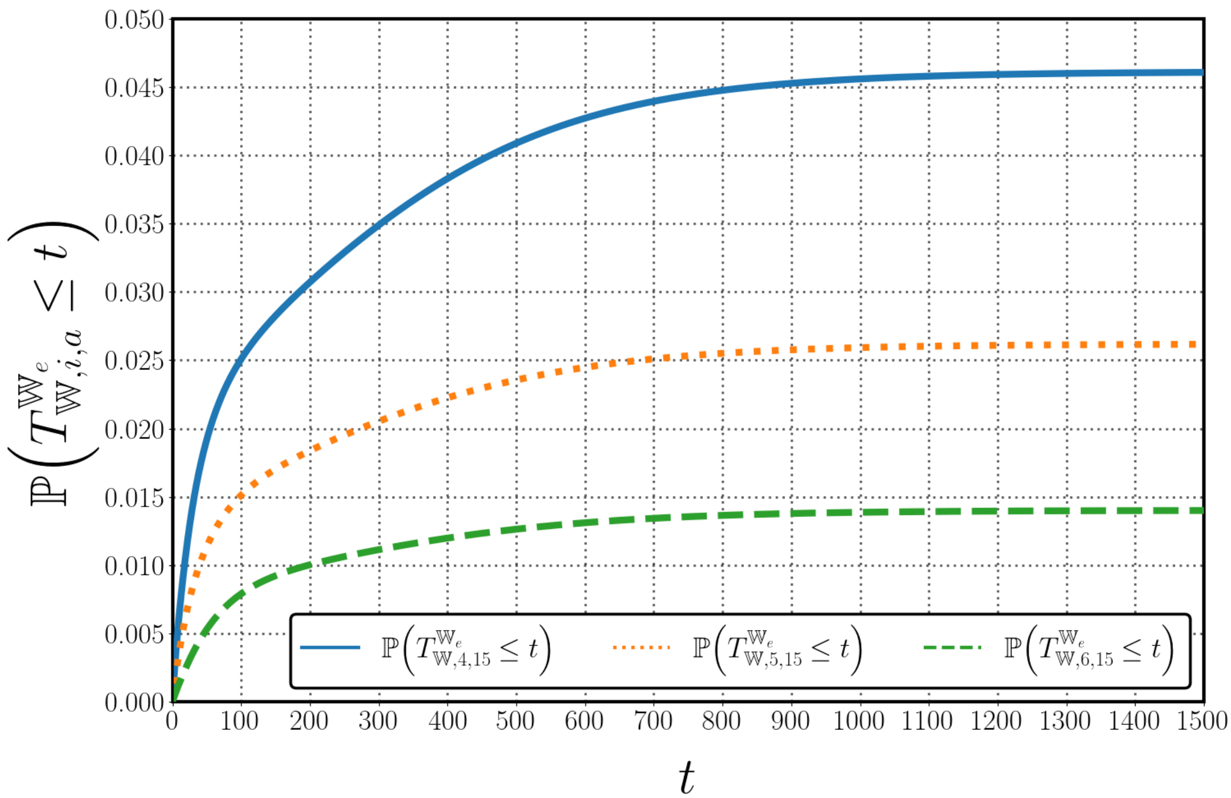

Figure 2.9: An application of Theorem 2.7 - calculations ($h = 10^{-2}$) for $\boldsymbol{s} = \boldsymbol{q}$ for the example MSBPM (Table 2.11, Table 2.6) with mutation scheme "$\mathbb{W}$ towards $\mathbb{W}$" (Table 2.7) with $\mathbb{W}_e = \{4, 5, 6\}$ starting with one particle of age $a = 15$. $\boldsymbol{q}$ is the vector with probabilities of extinction calculated for Figure 2.2, i.e., $\boldsymbol{q} = (0.96009038, 0.99744155,\ 0.99602606,\ 0.99015646,\ 0.99464378, 0.99762545)^\top$. $P\left(T^{\mathbb{W}_e}_{\mathbb{W},i,15} \leq t\right) = 1 - h^{\mathbb{W}_e}_{i,15}(t; \boldsymbol{q})$, we can reuse the calculations done for Figure 2.7. At $t = 1500$, we have $P\left(T^{\mathbb{W}_e}_{\mathbb{W},4,15} \leq t\right) = 0.04605937$, $P\left(T^{\mathbb{W}_e}_{\mathbb{W},5,15} \leq t\right) = 0.02616283$, $P\left(T^{\mathbb{W}_e}_{\mathbb{W},6,15} \leq t\right) = 0.01400123$. As $t = 1500$ is sufficiently large as to conclude that $h^{\mathbb{W}_e}_{i,15}(1500; \boldsymbol{q}) = h^{\mathbb{W}_e}_{i,15}(\boldsymbol{q})$, we also have $P\left(T^{\mathbb{W}_e}_{\mathbb{W},4,15} = \infty\right) = h^{\mathbb{W}_e}_{4,15}(\boldsymbol{q}) = 0.95394063$, $P\left(T^{\mathbb{W}_e}_{\mathbb{W},5,15} = \infty\right) = h^{\mathbb{W}_e}_{5,15}(\boldsymbol{q}) = 0.97383717$, $P\left(T^{\mathbb{W}_e}_{\mathbb{W},6,15} = \infty\right) = h^{\mathbb{W}_e}_{6,15}(\boldsymbol{q}) = 0.98599877$.

2.2.6 Immediate risk of producing a "successful" particle from $\mathbb{W}_e$ towards $\mathbb{W}$ within the MSBPM

We will study the immediate risk of escape facilitated by the particles with types within $\mathbb{W}_e$ via the following hazard function:

Definition 2.10. *Define for an initial particle of type i, $i \in \mathbb{W}_e$, the following hazard function:*

1. *If the initial particle is of age* 0

$$g_{\mathbb{W},i}^{\mathbb{W}_e}(t)dt = \mathbb{P}\Big(T_{\mathbb{W},i}^{\mathbb{W}_e} \in (t, t+dt] \mid T_{\mathbb{W},i}^{\mathbb{W}_e} > t\Big). \tag{2.26}$$

2. *If the initial particle if of age a, $a \neq 0$*

$$g_{\mathbb{W},i,a}^{\mathbb{W}_e}(t)dt = \mathbb{P}\Big(T_{\mathbb{W},i,a}^{\mathbb{W}_e} \in (t, t+dt] \mid T_{\mathbb{W},i,a}^{\mathbb{W}_e} > t\Big). \tag{2.27}$$

It is clear that for $i \in \mathbb{W}_e$

$$g_{\mathbb{W},i}^{\mathbb{W}_e}(t)dt = \frac{\mathbb{P}\Big(T_{\mathbb{W},i}^{\mathbb{W}_e} \in (t, t+dt],\, T_{\mathbb{W},i}^{\mathbb{W}_e} > t\Big)}{\mathbb{P}\Big(T_{\mathbb{W},i}^{\mathbb{W}_e} > t\Big)}.$$

Thus,

$$g_{\mathbb{W},i}^{\mathbb{W}_e}(t) = \frac{F^{(1)}_{T_{\mathbb{W},i}^{\mathbb{W}_e}}(t)}{\mathbb{P}\Big(T_{\mathbb{W},i}^{\mathbb{W}_e} > t\Big)}, \tag{2.28}$$

where $F^{(1)}_{T_{\mathbb{W},i}^{\mathbb{W}_e}}(t)$ is the probability density function of $T_{\mathbb{W},i}^{\mathbb{W}_e}$. We can find the c.d.f. of $T_{\mathbb{W},i}^{\mathbb{W}_e}$ via Theorem 2.6 and then approximate $F^{(1)}_{T_{\mathbb{W},i}^{\mathbb{W}_e}}(t)$, for example, with a forward difference.

It is evident that the same line of thought outlined above, applied for a starting particle of age a, $a \neq 0$, leads us to

$$g_{\mathbb{W},i,a}^{\mathbb{W}_e}(t) = \frac{F^{(1)}_{T_{\mathbb{W},i.a}^{\mathbb{W}_e}}(t)}{\mathbb{P}\Big(T_{\mathbb{W},i,a}^{\mathbb{W}_e} > t\Big)}. \tag{2.29}$$

A particular case of interest arises when the MSBPM is decomposable into two classes - $\mathbb{W}_e$ and $\mathbb{W}_0 = \mathbb{W} \setminus \mathbb{W}_e$, where we consider the particles produced from $\mathbb{W}_e$ towards $\mathbb{W}_0$ but $\mathbb{W}_0$ cannot produce particles towards $\mathbb{W}_e$. Within this setting it is beneficial to consider a modification of the hazard function from Definition 2.10 that includes the condition that at t, when we consider the immediate risk of escape, it is true that $\sum_{c\in\mathbb{W}_e} Z_c(t) > 0$. Indeed if there are no particles from $\mathbb{W}_e$ at t then there can be no "successful" particles produced from $\mathbb{W}_e$ towards $\mathbb{W}_0$. For an exploration of this case, see Subsection 2.3.1.6.

We illustrate the behavior of $g^{\mathbb{W}_e}_{\mathbb{W},i}(t)$ and $g^{\mathbb{W}_e}_{\mathbb{W},i}(t)$ in Figure 2.10. We take note of the very small values on the vertical axis.

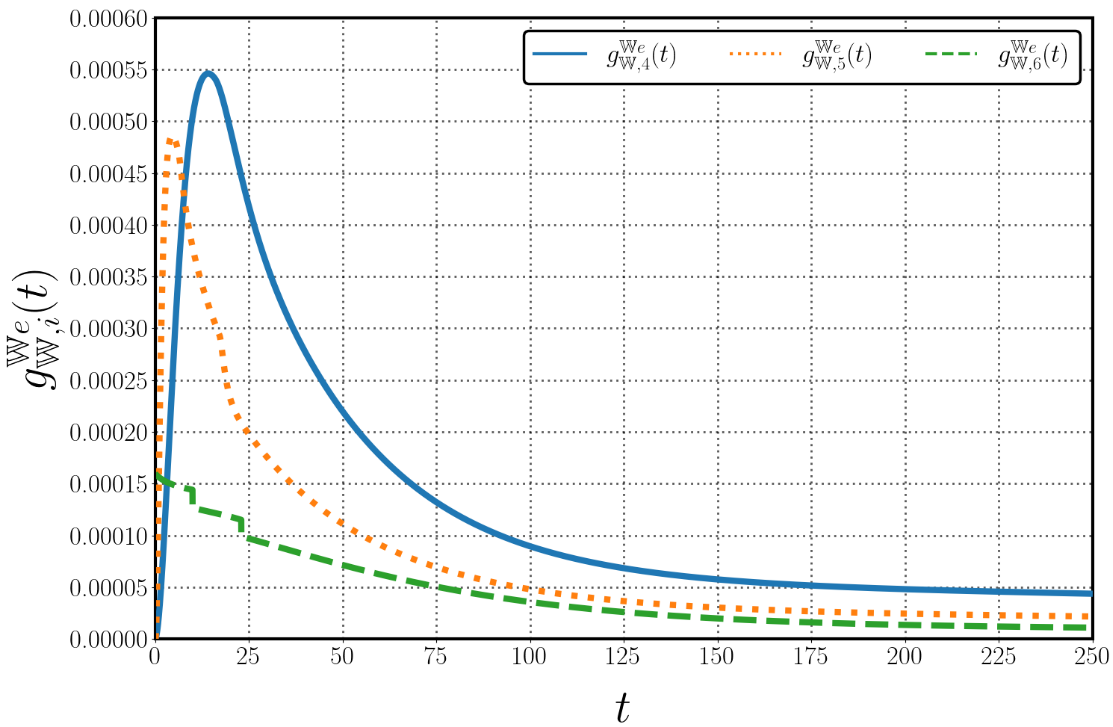

Figure 2.10: Calculations for equation (2.28), with $h = 10^{-2}$, for the example MSBPM (Table 2.11, Table 2.6) with mutation scheme "$\mathbb{W}$ towards $\mathbb{W}$" (Table 2.7) with $\mathbb{W}_e = \{4, 5, 6\}$ starting with one particle of age 0. At $t = 1000$, we have $g^{\mathbb{W}_e}_{\mathbb{W},4}(1000) = 2.71785074e - 06$, $g^{\mathbb{W}_e}_{\mathbb{W},5}(1000) = 1.28172130e - 06$, $g^{\mathbb{W}_e}_{\mathbb{W},6}(1000) = 6.87975348e - 07$.

2.2.7 Numerical schemes for computing obtained systems of integral equations for the MSBPM

We organize the integral equations obtained so far into two tables. Table 2.1 contains all integral equations that are for a MSBPM starting with a particle of age 0. We also put into Table 2.1 the result of Theorem 2.8, i.e., equation (2.53), as it conforms to the same pattern. Let us denote $B_i(t;\boldsymbol{s}) = \int_0^t f_i\Big(y;\, C_i(t-y;\boldsymbol{s})\Big)dG_i(y)$, where $C_i(t-y;\boldsymbol{s})$ is the corresponding second argument of f_i with respect to the entry of interest in Table 2.1. Through Numerical Scheme 1, derived below, we provide a general numerical method immediately applicable to the integral equations listed. We note that Numerical Scheme 1 can trace its origin to [2], where a model consisting of two particle types is discussed.

Eq.	$L_i(t;\boldsymbol{s})$		$A_i(t;\boldsymbol{s})$		$B_i(t;\boldsymbol{s})$	
(2.1)	$F_i(t;\,\boldsymbol{s})$	$=$	$s_i\big(1-G_i(t)\big)$	$+$	$\int_0^t f_i\Big(y;\sum_{r\in\mathbb{W}} u_{ir}F_r(t-y;\,\boldsymbol{s})\Big)dG_i(y)$	$i\in\mathbb{W}$
(2.5)	$q_i(t)$	$=$	0	$+$	$\int_0^t f_i\Big(y;\sum_{r\in\mathbb{W}} u_{ir}q_r(t-y)\Big)dG_i(y)$	$i\in\mathbb{W}$
(2.7)	q_i	$=$	0	$+$	$\int_0^\infty f_i\Big(y;\sum_{r\in\mathbb{W}} u_{ir}q_r\Big)dG_i(y)$	$i\in\mathbb{W}$
(2.10)	$h_i^{\mathbb{W}_e}(t;\boldsymbol{s})$	$=$	$\big(1-G_i(t)\big)$	$+$	$\int_0^t f_i\Big(y;\sum_{r\in\mathbb{W}} u_{ir}s_r h_r^{\mathbb{W}_e}(t-y;\boldsymbol{s})\Big)dG_i(y)$	$i\in\mathbb{W}_e$
(2.11)	$h_i^{\mathbb{W}_e}(t;\boldsymbol{s})$	$=$	$\big(1-G_i(t)\big)$	$+$	$\int_0^t f_i\Big(y;\sum_{r\in\mathbb{W}} u_{ir}h_r^{\mathbb{W}_e}(t-y;\boldsymbol{s})\Big)dG_i(y)$	$i\notin\mathbb{W}_e$
(2.12)	$h_i^{\mathbb{W}}(t;\boldsymbol{s})$	$=$	$\big(1-G_i(t)\big)$	$+$	$\int_0^t f_i\Big(y;\sum_{r\in\mathbb{W}} u_{ir}s_r h_r^{\mathbb{W}}(t-y;\boldsymbol{s})\Big)dG_i(y)$	$i\in\mathbb{W}$
(2.18)	$h_i^{\mathbb{W}_e}(\boldsymbol{s})$	$=$	0	$+$	$\int_0^\infty f_i\Big(y;\sum_{r\in\mathbb{W}} u_{ir}s_r h_r^{\mathbb{W}_e}(\boldsymbol{s})\Big)dG_i(y)$	$i\in\mathbb{W}_e$
(2.19)	$h_i^{\mathbb{W}_e}(\boldsymbol{s})$	$=$	0	$+$	$\int_0^\infty f_i\Big(y;\sum_{r\in\mathbb{W}} u_{ir}h_r^{\mathbb{W}_e}(\boldsymbol{s})\Big)dG_i(y)$	$i\notin\mathbb{W}_e$
(2.20)	$h_i^{\mathbb{W}}(\boldsymbol{s})$	$=$	0	$+$	$\int_0^\infty f_i\Big(y;\sum_{r\in\mathbb{W}} u_{ir}s_r h_r^{\mathbb{W}}(\boldsymbol{s})\Big)dG_i(y)$	$i\in\mathbb{W}$
(2.53)	$V_i(t)$	$=$	0	$+$	$\int_0^t f_i\Bigg(y;\Big[\sum_{m\in\mathbb{W}_e} u_{im}V_m(t-y)\Big]+\Big[\sum_{r\in\mathbb{W}_0} u_{ir}q_r\Big]\Bigg)dG_i(y)$	$i\in\mathbb{W}_e$

Table 2.1: Systems of integral equations for the case of a MSBPM starting with a particle of age 0.

Numerical Scheme 1. *Let $L_i(t;\boldsymbol{s})$ be from Table 2.1. The corresponding system of integral equations can be numerically computed via the following steps:*

1. *Let $t=0$. For every i that participates in the corresponding system of integral equations, compute the initial point $L_i(0;\boldsymbol{s}) = A_i(0;\boldsymbol{s})$.*

2. *Let $t = kh$, $k = 1, 2, \ldots$, where h is the chosen step size. For every i that participates in the corresponding system of integral equations compute*

$$L_i(kh; \boldsymbol{s}) \approx A_i(kh; \boldsymbol{s}) + \sum_{j=1}^{k} f_i\Big(jh;\, C_i\big((k-j)h; \boldsymbol{s}\big)\Big) \cdot \Big(G_i\big(jh\big) - G_i\big((j-1)h\big)\Big).$$

Proof. We note that $B_i(0; \boldsymbol{s}) = 0$ regardless of $L_i(0; \boldsymbol{s})$. Thus, the initial point at $k = 0$, i.e., at $t = 0$, is given by $L_i(0; \boldsymbol{s}) = A_i(0; \boldsymbol{s})$. Next, let $t = kh$, $k = 1, 2, \ldots$. For every i that participates in the corresponding system of integral equations, we have

$$\begin{aligned} L_i(kh; \boldsymbol{s}) &= A_i(kh; \boldsymbol{s}) + B_i(kh; \boldsymbol{s}) \\ &= A_i(kh; \boldsymbol{s}) + \int_0^{kh} f_i\Big(y;\, C_i(kh - y; \boldsymbol{s})\Big) dG_i(y) \\ &= A_i(kh; \boldsymbol{s}) + \sum_{j=1}^{k} \int_{(j-1)h}^{jh} f_i\Big(y;\, C_i(kh - y; \boldsymbol{s})\Big) dG_i(y). \end{aligned}$$

Approximating the integrals in the sum through the right rectangle rule, we obtain:

$$L_i(kh; \boldsymbol{s}) \approx A_i(kh; \boldsymbol{s}) + \sum_{j=1}^{k} f_i\Big(jh;\, C_i\big((k-j)h; \boldsymbol{s}\big)\Big) \cdot \Big(G_i\big(jh\big) - G_i\big((j-1)h\big)\Big),$$

where we note that when computing $C_i\big((k-j)h; \boldsymbol{s}\big)$, we use the already obtained approximated values for the $L_r\big((k-j)h; \boldsymbol{s}\big)$ that are within $C_i\big((k-j)h; \boldsymbol{s}\big)$. $\square$

Table 2.2, contains all integral equations that are obtained for a MSBPM starting with a particle of age a, $a \neq 0$. We also put into Table 2.2 the result of Corollary 2.22, i.e., equation (2.54), as it conforms to the same pattern. We denote $B_{i,a}(t; \boldsymbol{s}) = \int_0^t f_i\Big(a + y;\, C_i(t - y; \boldsymbol{s})\Big) dG_{i,a}(y)$, however, we stress that all $C_i(t - y; \boldsymbol{s})$ remain as in Table 2.1. Numerical Scheme 2, derived below, is applicable to all integral equations listed within Table 2.2.

Eq.	$L_{i,a}(t;\boldsymbol{s})$		$A_{i,a}(t;\boldsymbol{s})$		$B_{i,a}(t;\boldsymbol{s})$	
(2.2)	$F_{i,a}(t;\boldsymbol{s})$	$=$	$s_i\big(1-G_{i,a}(t)\big)$	$+$	$\int_0^t f_i\Big(a+y;\sum_{r\in\mathbb{W}}u_{ir}F_r(t-y;\boldsymbol{s})\Big)dG_{i,a}(y)$	$i\in\mathbb{W}$
(2.6)	$q_{i,a}(t)$	$=$	0	$+$	$\int_0^t f_i\Big(a+y;\sum_{r\in\mathbb{W}}u_{ir}q_r(t-y)\Big)dG_{i,a}(y)$	$i\in\mathbb{W}$
(2.8)	$q_{i,a}$	$=$	0	$+$	$\int_0^\infty f_i\Big(a+y;\sum_{r\in\mathbb{W}}u_{ir}q_r\Big)dG_{i,a}(y)$	$i\in\mathbb{W}$
(2.13)	$h_{i,a}^{\mathbb{W}_e}(t;\boldsymbol{s})$	$=$	$\big(1-G_{i,a}(t)\big)$	$+$	$\int_0^t f_i\Big(a+y;\sum_{r\in\mathbb{W}}u_{ir}s_rh_r^{\mathbb{W}_e}(t-y;\boldsymbol{s})\Big)dG_{i,a}(y)$	$i\in\mathbb{W}_e$
(2.14)	$h_{i,a}^{\mathbb{W}_e}(t;\boldsymbol{s})$	$=$	$\big(1-G_{i,a}(t)\big)$	$+$	$\int_0^t f_i\Big(a+y;\sum_{r\in\mathbb{W}}u_{ir}h_r^{\mathbb{W}_e}(t-y;\boldsymbol{s})\Big)dG_{i,a}(y)$	$i\notin\mathbb{W}_e$
(2.15)	$h_{i,a}^{\mathbb{W}}(t;\boldsymbol{s})$	$=$	$\big(1-G_{i,a}(t)\big)$	$+$	$\int_0^t f_i\Big(a+y;\sum_{r\in\mathbb{W}}u_{ir}s_rh_r^{\mathbb{W}}(t-y;\boldsymbol{s})\Big)dG_{i,a}(y)$	$i\in\mathbb{W}$
(2.21)	$h_{i,a}^{\mathbb{W}_e}(\boldsymbol{s})$	$=$	0	$+$	$\int_0^\infty f_i\Big(a+y;\sum_{r\in\mathbb{W}}u_{ir}s_rh_r^{\mathbb{W}_e}(\boldsymbol{s})\Big)dG_{i,a}(y)$	$i\in\mathbb{W}_e$
(2.22)	$h_{i,a}^{\mathbb{W}_e}(\boldsymbol{s})$	$=$	0	$+$	$\int_0^\infty f_i\Big(a+y;\sum_{r\in\mathbb{W}}u_{ir}h_r^{\mathbb{W}_e}(\boldsymbol{s})\Big)dG_{i,a}(y)$	$i\notin\mathbb{W}_e$
(2.23)	$h_{i,a}^{\mathbb{W}}(\boldsymbol{s})$	$=$	0	$+$	$\int_0^\infty f_i\Big(a+y;\sum_{r\in\mathbb{W}}u_{ir}s_rh_r^{\mathbb{W}}(\boldsymbol{s})\Big)dG_{i,a}(y)$	$i\in\mathbb{W}$
(2.54)	$V_{i,a}(t)$	$=$	0	$+$	$\int_0^t f_i\Bigg(a+y;\Big[\sum_{m\in\mathbb{W}_e}u_{im}V_m(t-y)\Big]+\Big[\sum_{r\in\mathbb{W}_0}u_{ir}q_r\Big]\Bigg)dG_{i,a}(y)$	$i\in\mathbb{W}_e$

Table 2.2: Systems of integral equations for the case of a MSBPM starting with a particle of age a, $a\neq 0$.

Numerical Scheme 2. *Let $L_{i,a}(t;\boldsymbol{s})$ be from Table 2.2. The corresponding system of integral equations can be numerically computed via the following steps:*

1. *Let $t=0$. For every i that participates in the corresponding system of integral equations, compute the initial point $L_{i,a}(0;\boldsymbol{s})=A_{i,a}(0;\boldsymbol{s})$.*
2. *Let $t=kh$, $k=1,2,\dots$, where h is the chosen step size. For every i that participates in the corresponding system of integral equations compute*

$$L_{i,a}(kh;\boldsymbol{s})\approx$$
$$\approx A_{i,a}(kh;\boldsymbol{s})+\sum_{j=1}^{k}f_i\Big(a+jh;\,C_i\big((k-j)h;\boldsymbol{s}\big)\Big)\cdot\Big(G_{i,a}\big(jh\big)-G_{i,a}\big((j-1)h\big)\Big).$$

Proof. We note that $B_{i,a}(0;\boldsymbol{s})=0$ regardless of $L_{i,a}(0;\boldsymbol{s})$. Thus, the initial point at $k=0$, i.e., at $t=0$, is given by $L_{i,a}(0;\boldsymbol{s})=A_{i,a}(0;\boldsymbol{s})$. Next, let $t=kh$, $k=1,2,\dots$. For every i that participates in the corresponding system of integral equations, we have

$$\begin{aligned}L_{i,a}(kh;\boldsymbol{s})&=A_{i,a}(kh;\boldsymbol{s})+B_{i,a}(kh;\boldsymbol{s})\\&=A_{i,a}(kh;\boldsymbol{s})+\int_0^{kh}f_i\Big(a+y;\,C_i(kh-y;\boldsymbol{s})\Big)dG_{i,a}(y)\end{aligned}$$

$$= A_{i,a}(kh;\boldsymbol{s}) + \sum_{j=1}^{k} \int_{(j-1)h}^{jh} f_i\Big(a+y;\, C_i(kh-y;\boldsymbol{s})\Big) dG_{i,a}(y).$$

Approximating the integrals in the sum through the right rectangle rule, we obtain:

$$L_{i,a}(kh;\boldsymbol{s}) \approx$$

$$\approx A_{i,a}(kh;\boldsymbol{s}) + \sum_{j=1}^{k} f_i\Big(a+jh;\, C_i\big((k-j)h;\boldsymbol{s}\big)\Big)\cdot\Big(G_{i,a}(jh) - G_{i,a}\big((j-1)h\big)\Big),$$

where we note that when computing $C_i\big((k-j)h;\boldsymbol{s}\big)$ we use the approximated values for the $L_r\big((k-j)h;\boldsymbol{s}\big)$ that are within $C_i\big((k-j)h;\boldsymbol{s}\big)$. These approximated values are obtained via the application of Numerical Scheme 1 to the corresponding L_r from Table 2.1. $\square$

We note that our implementation of Numerical Scheme 1 and Numerical Scheme 2, through which we have created the various figures found throughout Chapter 2, is done in Python 3.8.13 [209] by using the NumPy 1.20.3 [210] and SciPy 1.6.2 [211] libraries.

The theoretical examination of the properties of Numerical Scheme 1 and Numerical Scheme 2 is hindered by two difficulties:

1. The schemes are applied onto Riemann–Stieltjes integrals.
2. When computing the approximations of $L_i(kh;\boldsymbol{s})$ and $L_{i,a}(kh;\boldsymbol{s})$, within $C_i\big((k-j)h;\boldsymbol{s}\big)$ we use previously obtained values that are already approximations.

At this point, we are not able to provide a formal expression for the error of the schemes, however we can note the following considerations. First, it is evident from the proofs that at $k=0$ the solutions are exact. At $k=1$, since we use the results obtained at $k=0$, it is clear that providing smaller values for the step size h reduces the error for the computation at $k=1$, the same reasoning is then transferred for $k=2,3\dots$. Thus, smaller values of h do reduce the error within the schemes. Second, $C_i\big((k-j)h;\boldsymbol{s}\big)$ is within f_i, which is a well behaved, with respect to its second argument, p.g.f. with values in $[0,1]$. Thus, using previously obtained approximated values within $C_i\big((k-j)h;\boldsymbol{s}\big)$ is not expected to lead to big deviations from the value of f_i obtained with no approximations involved, especially for small h. Third, below we provide computed values for $h = 10^{-1}$, $h = 10^{-2}$, and $h = 10^{-3}$ for the probabilities for extinction of the example MSBPM, starting with one particle of age 0, that is the result of using Table 2.11 and Table 2.6, with mutation scheme "$\mathbb{W}$ towards $\mathbb{W}$" (i.e., Table 2.7):

t	$q_1(t)$	$q_2(t)$	$q_3(t)$	$q_4(t)$	$q_5(t)$	$q_6(t)$
0.5	5.44098264e-4	2.40789763e-2	3.41406470e-2	6.79641172e-5	6.07436015e-4	2.02888829e-2
1	0.00210551	0.06113254	0.06662528	0.00051097	0.00652419	0.03999338
5	0.04063088	0.26054039	0.27669537	0.03925087	0.21429338	0.17896027
20	0.27549534	0.54060739	0.59974974	0.45539839	0.66750029	0.51619682
50	0.52765209	0.74709026	0.79684177	0.71632983	0.85075051	0.79677134
100	0.6799033	0.86746777	0.90674343	0.85175863	0.92425437	0.92499966
300	0.85007145	0.97266778	0.97910123	0.95519509	0.97635273	0.98651223

Table 2.3: $h = 10^{-1}$.

t	$q_1(t)$	$q_2(t)$	$q_3(t)$	$q_4(t)$	$q_5(t)$	$q_6(t)$
0.5	5.44098912e-4	2.40794063e-2	3.41410757e-2	6.79641172e-5	6.07436015e-4	2.02890491e-2
1	0.00210552	0.061138	0.06662699	0.00051097	0.00652419	0.03999408
5	0.04063234	0.26068007	0.27672715	0.03925087	0.21429357	0.17897572
20	0.27555781	0.54091172	0.60002973	0.45546386	0.66756675	0.51630158
50	0.52788389	0.74736755	0.7972164	0.71664095	0.85087272	0.79694082
100	0.68019356	0.86766493	0.90700384	0.85201402	0.92436393	0.92513654
300	0.85027741	0.97272033	0.9791542	0.95527849	0.97639584	0.98654333

Table 2.4: $h = 10^{-2}$.

t	$q_1(t)$	$q_2(t)$	$q_3(t)$	$q_4(t)$	$q_5(t)$	$q_6(t)$
0.5	5.44098994e-4	2.40794623e-2	3.41411215e-2	6.79641172e-5	6.07436015e-4	2.02890669e-2
1	0.00210552	0.0611386	0.06662717	0.00051097	0.00652419	0.03999415
5	0.04063249	0.26069416	0.27673034	0.03925087	0.2142936	0.17897727
20	0.27556408	0.54094221	0.60005773	0.45547044	0.66757342	0.51631206
50	0.5279071	0.74739528	0.79725384	0.71667209	0.85088495	0.79695776
100	0.68022261	0.86768464	0.90702985	0.85203956	0.92437488	0.92515022
300	0.850298	0.97272558	0.97915948	0.95528682	0.97640015	0.98654643

Table 2.5: $h = 10^{-3}$.

It is evident that differences between step sizes $h = 10^{-1}$, $h = 10^{-2}$, and $h = 10^{-3}$, are practically negligible. The same conclusion is true for a process starting with one particle of age a, $a \neq 0$, and quantities other than the probabilities of extinction. Fourth, theoretical conclusions such as Proposition 2.1 are satisfied when applying Numerical Scheme 1 and Numerical Scheme 2.

In terms of computational strain, we note that as k increases so do the number of computations that we have to do in order to obtain the result for

k. Thus, the numerical schemes slow down as they advance at large values of k (respectively t). Using $h = 10^{-3}$, or smaller h, is not advisable since the algorithm slows down significantly. All of our observations indicate that for computations with large t even a step size of $h = 10^{-1}$ works sufficiently well.

We take special note on the topic of calculating expectations via Numerical Scheme 1 and Numerical Scheme 2. In the case of a MSBPM with exploding population it is clear that in order to obtain the precise value of the expectation for type i particles at t, we need to set Δs_j in equation (2.3) (or (2.4), (2.16), (2.17)) sufficiently small. As the exploding number of type i particles continues to increase with t, the necessary Δs_j eventually becomes so small so that our numerical schemes (or any other non-specialized for such situations numerical method) cannot register it adequately. Hence, care must be exercised upon inspecting calculated expectations.

Remark 2.12. *Note that when applying Numerical Scheme 1 to any of the L_i from Table 2.1, we must simultaneously calculate approximate values for all L_r expressions that are in the corresponding system of integral equations within C_i. As Numerical Scheme 2 uses the values computed by Numerical Scheme 1, provided that we have these values, we can compute $L_{i,a}$ from Table 2.2 individually. We further note that when calculating for t, we also obtain the values at each previous point in the grid given by $t = kh$, $k = 0, 1, \dots$.*

Remark 2.13. *Numerical Scheme 1 and Numerical Scheme 2 are applicable also to all results within Section 2.3 as the expressions obtained within the Section are particular cases of the expressions within Table 2.1 and Table 2.2.*

Remark 2.14. *All limit quantities within this Chapter, obtained as $t \to \infty$, can be numerically computed via Numerical Scheme 1 or Numerical Scheme 2 by computing the appropriate equations from Table 2.1 or Table 2.2 for a sufficiently large t.*

2.2.8 Setups of example processes

We now give the setups of the example processes, that we will use for demonstrating our results within this Chapter. All of the processes given here consist of 6 types of particles. We employ in total 3 different mutation schemes - one scheme ensures a non-decomposable process, while the other two enforce decomposability. With respect to the dependence of the reproduction capabilities of particles from particle age, we consider a Bellman-Harris type dependence (i.e. the reproductive capabilities of particles do not depend on age) and a Sevastyanov type dependence. This makes for 6 different processes, irrespective of initial particle type and initial particle age.

For simplicity of exposition, instead of an everywhere continuous dependence from particle age, all of the reproduction p.g.f.s for the Sevastyanov type dependence can be broken into 3 distinct p.g.f.s, each having a fixed form within a predefined age interval specific for each particle type. This choice can also be viewed as an emulation of a real-world scenario where we may have to work with discrete estimations valid within some interval.

2.2.8.1 Lifespan distributions within the example processes

We will work with the following lifespan distributions:

1. Under $Exp(\lambda)$, we understand the Exponential distribution with probability density function (p.d.f.) of $f(x;\lambda) = \frac{1}{\lambda}e^{-x/\lambda}$, $x \in [0,\infty)$, $\lambda > 0$.

2. Under $Lognorm(\mu,\sigma)$, we understand the Lognormal distribution with p.d.f. of $f(x;\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}}exp\Big(-\frac{(ln(x)-\mu)^2}{2\sigma^2}\Big)$, $x \in (0,\infty)$, $\mu \in (-\infty,+\infty)$, $\sigma > 0$.

3. Under $Gamma(\alpha,\beta)$ we understand the Gamma distribution with p.d.f. of $f(x;\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}$, $x \in (0,\infty)$, $\alpha,\beta > 0$.

Throughout all of our experiments each particle type will have the lifespan distribution, specified in Table 2.6.

Type i	Lifespan distribution	Expected lifespan	Variance
Type 1	$Gamma(2, 0.1)$	20	200
Type 2	$Lognorm(1.712, 1.415)$	15.08	1455.89
Type 3	$Exp(10)$	10	100
Type 4	$Gamma(3, 1/6)$	18	108
Type 5	$Lognorm(2.158, 0.9)$	12.97	209.89
Type 6	$Exp(17)$	17	289

Table 2.6: Lifespan distributions per type.

2.2.8.2 Mutation schemes within the example processes

We will use the following mutation schemes:

1. Mutation scheme "$\mathbb{W}$ towards $\mathbb{W}$". This mutation scheme is illustrated in Figure 2.1. The corresponding matrix is irreducible (see Definition 1) and the process is non-decomposable.

Type i	u_{i1}	u_{i2}	u_{i3}	u_{i4}	u_{i5}	u_{i6}
Type 1	0.70	0.10	0.10	0.05	0.03	0.02
Type 2	0.01	0.90	0.03	0.02	0.03	0.01
Type 3	0.01	0.02	0.80	0.05	0.04	0.08
Type 4	0.10	0.05	0.05	0.60	0.10	0.10
Type 5	0.05	0.03	0.02	0.10	0.70	0.10
Type 6	0.01	0.05	0.03	0.06	0.05	0.80

Table 2.7: Mutation scheme "$\mathbb{W}$ towards $\mathbb{W}$".

2. Mutation scheme "$\mathbb{W}_e$ towards $\mathbb{W}_0$", where $\mathbb{W}_e = \{4, 5, 6\}$, $\mathbb{W}_0 = \{1, 2, 3\}$. Types from $\mathbb{W}_e$ can produce types from $\mathbb{W}$, however, types from $\mathbb{W}_0$ can only produce types from $\mathbb{W}_0$. This mutation scheme is illustrated in Figure 2.11. The corresponding matrix is reducible (see Definition 1) and the process is decomposable.

Type i	u_{i1}	u_{i2}	u_{i3}	u_{i4}	u_{i5}	u_{i6}
Type 1	0.80	0.10	0.10	0.00	0.00	0.00
Type 2	0.10	0.90	0.00	0.00	0.00	0.00
Type 3	0.10	0.20	0.70	0.00	0.00	0.00
Type 4	0.10	0.05	0.05	0.60	0.10	0.10
Type 5	0.05	0.03	0.02	0.10	0.70	0.10
Type 6	0.01	0.05	0.03	0.06	0.05	0.80

Table 2.8: Mutation scheme "$\mathbb{W}_e$ towards $\mathbb{W}_0$".

3. Mutation scheme "$\mathbb{W}_e$ towards $\mathbb{W}_0, \mathbb{W}_0$ forms a chain", where $\mathbb{W}_e = \{4,5,6\}$, $\mathbb{W}_0 = \{1,2,3\}$ and types from $\mathbb{W}_e$ can produce mutants only towards Type 3 from $\mathbb{W}_0$, Type 3 can produce mutants only towards Type 2, Type 2 can produce mutants only towards Type 1, and Type 1 can only produce Type 1 particles. This mutation scheme is illustrated in Figure 2.12. The corresponding matrix is reducible (see Definition 1) and the process is decomposable.

Type i	u_{i1}	u_{i2}	u_{i3}	u_{i4}	u_{i5}	u_{i6}
Type 1	1.00	0.00	0.00	0.00	0.00	0.00
Type 2	0.10	0.90	0.00	0.00	0.00	0.00
Type 3	0.00	0.15	0.85	0.00	0.00	0.00
Type 4	0.00	0.00	0.10	0.60	0.15	0.15
Type 5	0.00	0.00	0.05	0.10	0.70	0.15
Type 6	0.00	0.00	0.07	0.06	0.07	0.80

Table 2.9: Mutation scheme "$\mathbb{W}_e$ towards $\mathbb{W}_0$, $\mathbb{W}_0$ forms a chain".

2.2.8.3 Probability generating functions for the total number of offspring within the example processes

We give two tables for the two types of dependence of reproduction capabilities from particle age. In the Sevastyanov setting, we have deliberately constructed the individual p.g.f.s, for each type, to become "weaker" in terms of reproduction intensity, as particles become older. Arguably, such a setting can be appropriate in biological contexts.

In both cases below, we let only Type 1 to be supercritical, while all other types are subcritical or critical. This setting is of particular interest for us, as it can be used to model the escape from extinction of a population under stress, such as cancer cells under treatment (see [61], [62], [64], [65], [7], [6]).

1. For the Bellman-Harris type dependence from particle age, we consider the following p.g.f.s for the total number of offspring. Note that we are actually considering a particular case of the MSBPM, i.e., a Multi-type Bellman-Harris Branching Process through probabilities of Mutation between types (MBHBPM). If the mutation scheme used is either "$\mathbb{W}_e$ towards $\mathbb{W}_0$" or "$\mathbb{W}_e$ towards $\mathbb{W}_0$, $\mathbb{W}_0$ forms a chain", we have a decomposable MBHBPM, or DMBHBPM for short. The DMBHBPM is investigated in Subsection 2.3.2.

Type i	p.g.f.	Expected offspring	Variance
Type 1	$0.45s^0 + 0.30s^2 + 0.25s^4$	1.60	2.64
Type 2	$0.54s^0 + 0.46s^2$	0.92	0.99
Type 3	$0.70s^0 + 0.12s^2 + 0.18s^4$	0.96	2.44
Type 4	$0.75s^0 + 0.25s^4$	1.00	3.00
Type 5	$0.79s^0 + 0.21s^4$	0.84	2.65
Type 6	$0.70s^0 + 0.20s^2 + 0.10s^4$	0.80	1.76

Table 2.10: Probaility generating functions for the total number of particles in the offspring of a Type i particle within an example Multi-type Bellman-Harris Branching Process through probabilities of Mutation between types (MBHBPM).

2. For the Sevastyanov type dependence from particle age, we consider the following p.g.f.s. Note that in this particular configuration, the expectation of the total number of particles in an offspring, as well as

the variance, decrease as particles (regardless of type) become older. Further, note that the within interval $(0, X]$, where X varies from particle type to particle type, the p.g.f.s are actually the ones within Table 2.10. If the mutation scheme used is either "$\mathbb{W}_e$ towards $\mathbb{W}_0$" or "$\mathbb{W}_e$ towards $\mathbb{W}_0$, $\mathbb{W}_0$ forms a chain", we have a decomposable MSBPM, or DMSBPM for short. The DMSBPM is investigated within Subsection 2.3.1.

Type i, τ_i	Corresponding p.g.f.	Expected offspring	Variance
Type 1, $\tau_1 \in (0, 11]$	$0.45s^0 + 0.30s^2 + 0.25s^4$	1.60	2.64
Type 1, $\tau_1 \in (11, 24]$	$0.45s^0 + 0.35s^2 + 0.20s^4$	1.50	2.35
Type 1, $\tau_1 \in (24, \infty]$	$0.45s^0 + 0.40s^2 + 0.15s^4$	1.40	2.04
Type 2, $\tau_2 \in (0, 7]$	$0.54s^0 + 0.46s^2$	0.92	0.99
Type 2, $\tau_2 \in (7, 19]$	$0.57s^0 + 0.43s^2$	0.86	0.98
Type 2, $\tau_2 \in (19, \infty]$	$0.60s^0 + 0.40s^2$	0.80	0.96
Type 3, $\tau_3 \in (0, 8]$	$0.70s^0 + 0.12s^2 + 0.18s^4$	0.96	2.44
Type 3, $\tau_3 \in (8, 13]$	$0.64s^0 + 0.26s^2 + 0.1s^4$	0.92	1.79
Type 3, $\tau_3 \in (13, \infty]$	$0.56s^0 + 0.44s^2$	0.88	0.99
Type 4, $\tau_4 \in (0, 8]$	$0.75s^0 + 0.25s^4$	1.00	3.00
Type 4, $\tau_4 \in (8, 16]$	$0.64s^0 + 0.24s^2 + 0.12s^4$	0.96	1.96
Type 4, $\tau_4 \in (16, \infty]$	$0.55s^0 + 0.45s^2$	0.90	0.99
Type 5, $\tau_5 \in (0, 6]$	$0.79s^0 + 0.21s^4$	0.84	2.65
Type 5, $\tau_5 \in (6, 17]$	$0.80s^0 + 0.20s^4$	0.80	2.56
Type 5, $\tau_5 \in (17, \infty]$	$0.81s^0 + 0.19s^4$	0.76	2.46
Type 6, $\tau_6 \in (0, 10]$	$0.70s^0 + 0.20s^2 + 0.10s^4$	0.80	1.76
Type 6, $\tau_6 \in (10, 23]$	$0.73s^0 + 0.17s^2 + 0.10s^4$	0.74	1.73
Type 6, $\tau_6 \in (23, \infty]$	$0.75s^0 + 0.15s^2 + 0.10s^4$	0.70	1.71

Table 2.11: Probaility generating functions for the total number of particles in the offspring of a Type i particle within an example Multi-type Sevastyanov Branching Process through probabilities of Mutation between types (MSBPM).

2.2.8.4 Preliminary analysis

Recall Theorem 1.4 (or Theorem 2 from [8], page 238). Evidently, the probability generating functions discussed above do not allow particles to have 0 offspring with probability 0. Thus, we may apply Theorem 1.4 and gain some preliminary insight about the behavior of the example processes. Note that we obtain probabilities $p^i_{\boldsymbol{\alpha}}(u)$ necessary for computing $||A^i_j||$ by calculating $p^i_{\boldsymbol{\alpha}}(a) = p_{ik}(a)\frac{k!}{\alpha_1!\ldots\alpha_n!}u_{i1}^{\alpha_1}\ldots u_{in}^{\alpha_n}$, as discussed below Definition 2.1.

1. For mutation scheme "$\mathbb{W}$ towards $\mathbb{W}$", assuming that a MSBPM starts with particles of age 0, we can immediately apply Theorem 1.4.

   (a) For Bellman-Harris type dependence on age, using Table 2.7 and Table 2.10, we calculate $||A^i_j||$. We take into account that, for Bellman-Harris, $a^i_j(u) = a^i_j$ (see equation (1.7) on page 19).

$$||A^i_j|| = \begin{pmatrix} 1.12 & 0.16 & 0.16 & 0.08 & 0.048 & 0.032 \\ 0.0092 & 0.828 & 0.0276 & 0.0184 & 0.0276 & 0.0092 \\ 0.0096 & 0.0192 & 0.768 & 0.048 & 0.0384 & 0.0768 \\ 0.1 & 0.05 & 0.05 & 0.6 & 0.1 & 0.1 \\ 0.042 & 0.0252 & 0.0168 & 0.084 & 0.588 & 0.084 \\ 0.008 & 0.04 & 0.024 & 0.048 & 0.04 & 0.64 \end{pmatrix}$$

   $||A^i_j||$ has a Perron root of 1.1655915. Hence, by Definition 1.1, the process is supercritical.

   (b) For Sevastyanov type dependence on age, we can exploit the specifics of our example p.g.f.s for the total number of offspring in order to save some computations. Indeed, as all particles decrease their reproductive capabilities with age, we can take the worst available p.g.f.s for the total number of offspring and check the criticality of the corresponding MBHBPM - the criticality of our MSBPM will be higher. Using the last entries, with respect to type, from Table 2.11, we calculate:

$$||A_j^i|| = \begin{pmatrix} 0.98 & 0.14 & 0.14 & 0.07 & 0.042 & 0.028 \\ 0.008 & 0.72 & 0.024 & 0.016 & 0.024 & 0.008 \\ 0.0088 & 0.0176 & 0.704 & 0.044 & 0.0352 & 0.0704 \\ 0.09 & 0.045 & 0.045 & 0.54 & 0.09 & 0.09 \\ 0.038 & 0.0228 & 0.0152 & 0.076 & 0.532 & 0.076 \\ 0.007 & 0.035 & 0.021 & 0.042 & 0.035 & 0.56, \end{pmatrix}$$

with a Perron root of 1.02362078. Thus, by Definition 1.1, the initially considered MSBPM supercritical. Note also that at its best (i.e., considering the first entries, with respect to type, from Table 2.11), the MSBPM corresponds to a criticality of 1.1655915.

We note that criticality of 1.1655915 is not very high, thus we expect to see large values for the probabilities for extinction.

2. For mutation scheme "$\mathbb{W}_e$ towards $\mathbb{W}_0$", assuming that a MSBPM starts with particles of age 0, we proceed in the same way.

   (a) For Bellman-Harris type dependence on age, using Table 2.8 and Table 2.10, we calculate the $||A_j^i||$.

$$||A_j^i|| = \begin{pmatrix} 1.28 & 0.16 & 0.16 & 0.0 & 0.0 & 0.0 \\ 0.092 & 0.828 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.096 & 0.192 & 0.672 & 0.0 & 0.0 & 0.0 \\ 0.1 & 0.05 & 0.05 & 0.6 & 0.1 & 0.1 \\ 0.042 & 0.0252 & 0.0168 & 0.084 & 0.588 & 0.084 \\ 0.008 & 0.04 & 0.024 & 0.048 & 0.04 & 0.64 \end{pmatrix},$$

   $||A_j^i||$ is reducible and we cannot take advantage of Theorem 1.4. However, we can concentrate our attention of the upper-left corner of $||A_j^i||$. We denote

$$||A_j^i||_{3\times 3} = \begin{pmatrix} 1.28 & 0.16 & 0.16 \\ 0.092 & 0.828 & 0.0 \\ 0.096 & 0.192 & 0.672 \end{pmatrix}.$$

   $||A_j^i||_{3\times 3}$ is non-reducible and the branching process corresponding to it benefits from Theorem 1.4. $||A_j^i||_{3\times 3}$ has a Perron root of 1.34000669, thus if the initial particle of the decomposable MSBPM is from $\mathbb{W}_0 = \{1, 2, 3\}$ the process observed is supercritical.

(b) For Sevastyanov type dependence on age, again, we take the worst available p.g.f.s for the total number of offspring and check the criticality of the corresponding MBHBPM.

$$||A^i_j|| = \begin{pmatrix} 1.12 & 0.14 & 0.14 & 0.0 & 0.0 & 0.0 \\ 0.08 & 0.72 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.088 & 0.176 & 0.616 & 0.0 & 0.0 & 0.0 \\ 0.09 & 0.045 & 0.045 & 0.54 & 0.09 & 0.09 \\ 0.038 & 0.0228 & 0.0152 & 0.076 & 0.532 & 0.076 \\ 0.007 & 0.035 & 0.021 & 0.042 & 0.035 & 0.56 \end{pmatrix}$$

Again, we inspect

$$||A^i_j||_{3\times 3} = \begin{pmatrix} 1.12 & 0.14 & 0.14 \\ 0.08 & 0.72 & 0.0 \\ 0.088 & 0.176 & 0.616 \end{pmatrix}.$$

$||A^i_j||_{3\times 3}$ has a Perron root of 1.17447077, thus if the initial particle of the decomposable MSBPM is from $\mathbb{W}_0 = \{1, 2, 3\}$ the process observed is supercritical.

3. For mutation scheme "$\mathbb{W}_e$ towards $\mathbb{W}_0$, $\mathbb{W}_0$ forms a chain" with a MSBPM starting with a particles of age 0, we can evaluate the critically only with respect to type 1. Regardless, we will give our computations for $||A^i_j||$.

(a) For Bellman-Harris type dependence on age, using Table 2.9 and Table 2.10, we calculate the $||A^i_j||$.

$$||A^i_j|| = \begin{pmatrix} 1.6 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.092 & 0.828 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.144 & 0.816 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.1 & 0.6 & 0.15 & 0.15 \\ 0.0 & 0.0 & 0.042 & 0.084 & 0.588 & 0.126 \\ 0.0 & 0.0 & 0.056 & 0.048 & 0.056 & 0.64 \end{pmatrix}$$

Evidently, if the process begins with one particle of type 1, we will have a supercritical single-type Sevastyanov branching process with criticality 1.6.

(b) For Sevastyanov type dependence on age, we take the worst available p.g.f.s for the total number of offspring and calculate for the corresponding MBHBPM.

$$||A_j^i|| = \begin{pmatrix} 1.4 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.08 & 0.72 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.132 & 0.748 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.09 & 0.54 & 0.135 & 0.135 \\ 0.0 & 0.0 & 0.038 & 0.076 & 0.532 & 0.114 \\ 0.0 & 0.0 & 0.049 & 0.042 & 0.049 & 0.56 \end{pmatrix}$$

If the process begins with one particle of type 1, we will have a supercritical single-type Sevastyanov branching process with criticality 1.4.

2.3 Particular cases of the MSBPM

2.3.1 Decomposable Multi-type Sevastyanov Branching Process through probabilities of Mutation between types (DMSBPM)

Within this Subsection, we focus on the Decomposable Multi-type Sevastyanov Branching Process through probabilities of Mutation between types (DMSBPM). This process is the Decomposable Multi-type Sevastyanov Branching Process (DMSBP) discussed in Vitanov & Slavtchova-Bojkova [7] (2022), however, within the dissertation, we deem appropriate to add an extra “M” at the end in order to stress that the use of mutation probabilities within the definition of the process is a distinguishing feature. We will show that the systems of equations for the DMSBPM (a.k.a. DMSBP), obtained in [7], can be derived from the systems of equations for the MSBPM. We pay special attention to the topic of immediate escape from extinction as there is a particularity stemming from the fact that the DMSBPM is by definition decomposable. We also provide the systems of equations for the DMSBPM in the case of a process starting with a particle that is of age a, $a \neq 0$. We note that the DMSBPM can be used to model an irreversible path in the evolution of a population of particles.

2.3.1.1 Definition of the DMSBPM

Recall that $\mathbb{W} = \{1, 2, \ldots, n\}$. In what follows, we let $\mathbb{W}_0 \subset \mathbb{W}$ and $\mathbb{W}_e = \mathbb{W} \backslash \mathbb{W}_0$. In order to avoid ambiguity, without loss of generality, we impose the following ordering - if $|\mathbb{W}_0| = b$, then $\mathbb{W}_0 = \{1, 2, \ldots, b\}$ and $\mathbb{W}_e = \{b+1, b+2, \ldots, n\}$. We denote $\boldsymbol{s}_{\mathbb{W}_0} = (s_1, \ldots, s_b, 1, \ldots, 1)^\top$ and $\boldsymbol{q}_{\mathbb{W}_0} = (q_1, \ldots, q_b, 1, \ldots, 1)^\top$. In our considerations below, we assume that the i-th coordinate of $\boldsymbol{s}$ is always equal to the i-th coordinate of $\boldsymbol{s}_{\mathbb{W}_0}$, $i \in \mathbb{W}_0$. Analogously, the i-th coordinate of $\boldsymbol{q}$ is always equal to the i-th coordinate of $\boldsymbol{q}_{\mathbb{W}_0}$, $i \in \mathbb{W}_0$.

Definition 2.11. *Define the Decomposable Multi-type Sevastyanov Branching Process through probabilities of Mutation between types (DMSBPM) as the multi-type branching process satisfying:*

1. *Each particle type is uniquely associated with an integer from $\mathbb{W}$ and conforms to:*
   (a) *The lifespan of particles of type i, $i \in \mathbb{W}$, is modeled by a (continuous) r.v. τ_i. The corresponding cumulative distribution function (c.d.f.) is denoted by $G_i(t) = \mathbb{P}(\tau_i \leq t)$, also $G_i(0^+) = 0$.*
   (b) *The number of particles in the offspring of a type i, $i \in \mathbb{W}$, particle of age a is modeled by a (discrete) r.v. $\nu_i(a)$. We denote with $p_{ik}(a)$ the probability that a type i particle of age a has k, $k \in \mathbb{N}_0$, offspring particles (regardless of their type). Thus, $\nu_i(a)$ is specified by given $\left\{p_{ik}(a)\right\}_{k=0}^{\infty}$, $\sum_{k=0}^{\infty} p_{ik}(a) = 1$. We denote the corresponding p.g.f. of $\nu_i(a)$ with $f_i(a; s) = \mathbb{E}\left[s^{\nu_i(a)}\right] = \sum_{k=0}^{\infty} p_{ik}(a) s^k$, $|s| \leq 1$.*
2. *Each daughter particle of a type i particle, $i \in \mathbb{W}_e$, can be of any type $j \in \mathbb{W}$, however, daughter particles of type i particles, $i \in \mathbb{W}_0$, can only be of type $j \in \mathbb{W}_0$. The type of a daughter particle is determined at birth. If $i \neq j$ we say that a "mutation" occurs. The probability that a daughter particle of a type i particle is a type j particle is denoted by u_{ij}, $u_{ij} \geq 0$, $\sum_{j=1}^{n} u_{ij} = 1$. Further:*
   (a) *If type i cannot have daughters of type j we consider the corresponding u_{ij} as $u_{ij} = 0$.*
   (b) *Particles are not allowed to change their type within their lifespan.*

3. *All particles from all particle types evolve independently from one another, irrespective of generation.*

4. *Formally* $\left\{ \boldsymbol{Z}(t) = \left(Z_1(t), Z_2(t), \ldots, Z_n(t)\right)^\top \right\}_{t \geq 0}$*, where* $\boldsymbol{Z}(t)$ *stands for the DMSBPM at* t *and* $Z_i(t)$ *is the number of particles of type* i *that exist at* t*.*

For illustrative purposes, Figure 2.11 depicts the most general DMSBPM, while Figure 2.12 shows a particular case of interest in biology.

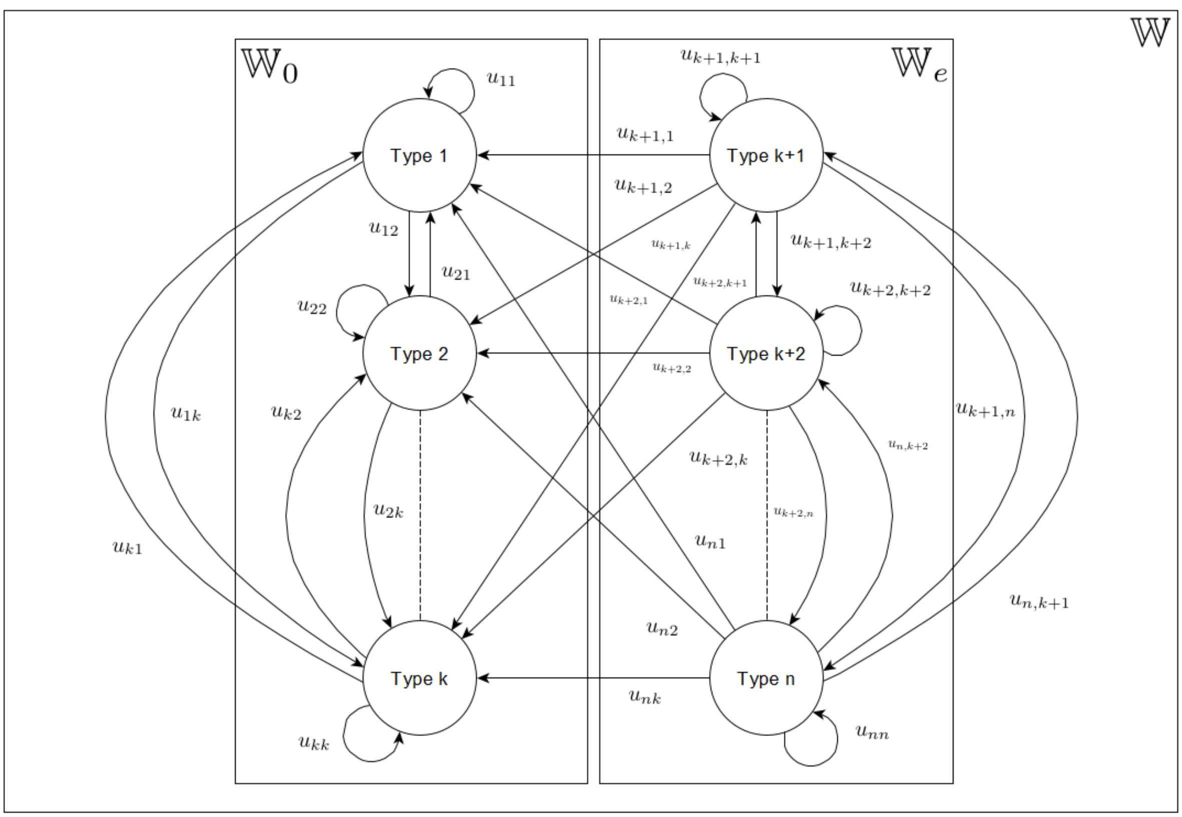

Figure 2.11: A diagram of the DMSBPM depicting all possible paths of mutation within the process. Note that some of the u_{ij} may be equal to 0 depending on context. In such cases the corresponding arrows are removed from the diagram.

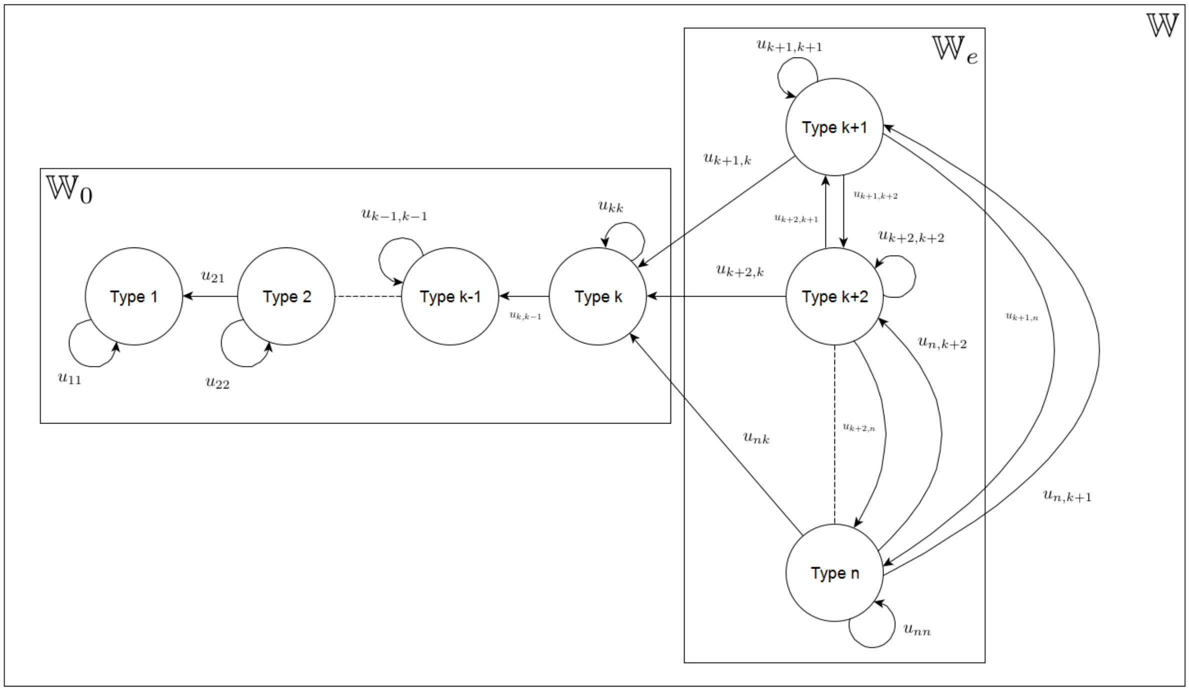

Figure 2.12: A particular case of the DMSBPM where type 1 is reachable only after mutations leading to type k, type $k-1$, $\ldots$, type 2 occur (assuming the process begins with particles with types from $\mathbb{W}_e$). A subcase of special interest arises when type 1 is the only supercritical type - types within $\mathbb{W}_e$ can be used to model an existing badly adapted population approaching extinction, while the mutation path towards type 1, provided by the types within $\mathbb{W}_0$, leads to possible escape from extinction.

2.3.1.2 Probability generating functions for the DMSBPM

Recall Definition 2.2, valid for the MSBPM, where we had, $i \in \mathbb{W}$,

$$F_i(t;\boldsymbol{s}) = \mathbb{E}\Big(\prod_{j\in\mathbb{W}} s_j^{Z_j(t)} |\ \boldsymbol{Z}(0) = \boldsymbol{\delta}^i \Big),$$

and

$$F_{i,a}(t;\boldsymbol{s}) = \mathbb{E}\Big(\prod_{j\in\mathbb{W}} s_j^{Z_j(t)} |\ \boldsymbol{Z}(0) = \boldsymbol{\delta}_a^i \Big).$$

For a DMSBPM starting with one particle of type i, $i \in \mathbb{W}_0$, since there can be no mutations from $\mathbb{W}_0$ towards $\mathbb{W}_e$, the DMSBPM will have only particles with types from $\mathbb{W}_0$. Thus, if $i \in \mathbb{W}_0$, we effectively have

$$F_i(t;\boldsymbol{s}) = \mathbb{E}\Big(\prod_{j\in\mathbb{W}_0} s_j^{Z_j(t)} |\ \boldsymbol{Z}(0) = \boldsymbol{\delta}^i \Big),$$

and

$$F_{i,a}(t;\boldsymbol{s}) = \mathbb{E}\Big(\prod_{j\in\mathbb{W}_0} s_j^{Z_j(t)} |\ \boldsymbol{Z}(0) = \boldsymbol{\delta}_a^i\Big),$$

which is the same as substituting $\boldsymbol{s} = \boldsymbol{s}_{\mathbb{W}_0}$ into $F_i(t;\cdot)$ and $F_{i,a}(t;\cdot)$ from Definition 2.2:

$$F_i(t;\boldsymbol{s}_{\mathbb{W}_0}) = \mathbb{E}\Big(\prod_{j\in\mathbb{W}_0} s_j^{Z_j(t)} |\ \boldsymbol{Z}(0) = \boldsymbol{\delta}^i\Big),$$

$$F_{i,a}(t;\boldsymbol{s}_{\mathbb{W}_0}) = \mathbb{E}\Big(\prod_{j\in\mathbb{W}_0} s_j^{Z_j(t)} |\ \boldsymbol{Z}(0) = \boldsymbol{\delta}_a^i\Big).$$

In the context of the DMSBPM, for clarity and convenience, we give the following definition.

Definition 2.12. *Given Definition 2.2, denote the p.g.f. of a DMSBPM starting with one particle of type $i \in \mathbb{W}$ as*

1. *For $i \in \mathbb{W}_e$*

$$F_i(t;\boldsymbol{s}) = \mathbb{E}\Big(\prod_{j\in\mathbb{W}} s_j^{Z_j(t)} |\ \boldsymbol{Z}(0) = \boldsymbol{\delta}^i\Big),$$

$$F_{i,a}(t;\boldsymbol{s}) = \mathbb{E}\Big(\prod_{j\in\mathbb{W}} s_j^{Z_j(t)} |\ \boldsymbol{Z}(0) = \boldsymbol{\delta}_a^i\Big),$$

where $|\boldsymbol{s}| \le 1$.

2. *For $i \in \mathbb{W}_0$, due to the fact that there can be no mutations from $\mathbb{W}_0$ towards $\mathbb{W}_e$*

$$F_i(t;\boldsymbol{s}_{\mathbb{W}_0}) = \mathbb{E}\Big(\prod_{j\in\mathbb{W}_0} s_j^{Z_j(t)} |\ \boldsymbol{Z}(0) = \boldsymbol{\delta}^i\Big),$$

$$F_{i,a}(t;\boldsymbol{s}_{\mathbb{W}_0}) = \mathbb{E}\Big(\prod_{j\in\mathbb{W}_0} s_j^{Z_j(t)} |\ \boldsymbol{Z}(0) = \boldsymbol{\delta}_a^i\Big),$$

where $|\boldsymbol{s}_{\mathbb{W}_0}| \le 1$.

Corollary 2.12. *For the DMSBPM, the following system of integral equations holds:*

1. *For* $i \in \mathbb{W}_e$

(2.30)

$$F_i(t;\boldsymbol{s}) = s_i\big(1 - G_i(t)\big) + \int_0^t f_i\Bigg(y; \Big[\sum_{m\in\mathbb{W}_e} u_{im}F_m(t-y;\boldsymbol{s})\Big] + \\ + \Big[\sum_{r\in\mathbb{W}_0} u_{ir}F_r(t-y;\boldsymbol{s}_{\mathbb{W}_0})\Big]\Bigg)dG_i(y).$$

2. *For* $i \in \mathbb{W}_0$

(2.31)

$$F_i(t;\boldsymbol{s}_{\mathbb{W}_0}) = s_i\big(1 - G_i(t)\big) + \int_0^t f_i\Bigg(y; \sum_{r\in\mathbb{W}_0} u_{ir}F_r(t-y;\boldsymbol{s}_{\mathbb{W}_0})\Bigg)dG_i(y).$$

Proof. Considering Remark 2.1, the proof follows directly from the result of Theorem 2.1. For $i \in \mathbb{W}_e$, by virtue of Definition 2.12, we can decompose the second argument of f_i from equation (2.1)

$$F_i(t;\boldsymbol{s}) = s_i\big(1 - G_i(t)\big) + \int_0^t f_i\Big(y; \sum_{r\in\mathbb{W}} u_{ir}F_r(t-y;\boldsymbol{s})\Big)dG_i(y)$$

into

$$F_i(t;\boldsymbol{s}) = s_i\big(1 - G_i(t)\big) + \int_0^t f_i\Bigg(y; \Big[\sum_{m\in\mathbb{W}_e} u_{im}F_m(t-y;\boldsymbol{s})\Big] + \\ + \Big[\sum_{r\in\mathbb{W}_0} u_{ir}F_r(t-y;\boldsymbol{s}_{\mathbb{W}_0})\Big]\Bigg)dG_i(y).$$

For $i \in \mathbb{W}_0$, in addition to the decomposition above, we can further remove $\sum_{m\in\mathbb{W}_e} u_{im}F_m(t-y;\boldsymbol{s})$, since for all $m \in \mathbb{W}_e$, we have $u_{im} = 0$ because there can be no mutations from $\mathbb{W}_0$ towards $\mathbb{W}_e$. Thus, for $i \in \mathbb{W}_0$, we have

$$F_i(t;\boldsymbol{s}_{\mathbb{W}_0}) = s_i\big(1 - G_i(t)\big) + \int_0^t f_i\Bigg(y; \sum_{r\in\mathbb{W}_0} u_{ir}F_r(t-y;\boldsymbol{s}_{\mathbb{W}_0})\Bigg)dG_i(y).$$

□

Corollary 2.13. *For the DMSBPM, the following system of integral equations holds:*

1. *For* $i \in \mathbb{W}_e$

$$F_{i,a}(t;\boldsymbol{s}) = s_i\big(1-G_{i,a}(t)\big)+\int_0^t f_i\Bigg(a+y;\Big[\sum_{m\in\mathbb{W}_e} u_{im}F_m(t-y;\boldsymbol{s})\Big]+ + \Big[\sum_{r\in\mathbb{W}_0} u_{ir}F_r(t-y;\boldsymbol{s}_{\mathbb{W}_0})\Big]\Bigg)dG_{i,a}(y).$$

2. *For* $i \in \mathbb{W}_0$

$$F_{i,a}(t;\boldsymbol{s}_{\mathbb{W}_0}) = s_i\big(1-G_{i,a}(t)\big)+\int_0^t f_i\Bigg(a+y;\sum_{r\in\mathbb{W}_0} u_{ir}F_r(t-y;\boldsymbol{s}_{\mathbb{W}_0})\Bigg)dG_{i,a}(y).$$

Proof. The proof is analogous to the proof of Corollary 2.12, however we use as a starting point for the proof the result of Corollary 2.1. □

Expectations for $Z_i(t)$, $i \in \mathbb{W}$, can be handled as in Subsection 2.2.2.

2.3.1.3 Probabilities of extinction for the DMSBPM

Within the DMSBPM the definitions for $q_i(t)$, $q_{i,a}(t)$ (Definition 2.3) and q_i, $q_{i,a}$ (Definition 2.4) do not need additional elaborations. The same is true for Remark 2.3, Remark 2.5, and Definition 2.5, which remain in effect for the DMSBPM.

Corollary 2.14. *The following system of integral equations holds for the DMSBPM:*

1. *For* $i \in \mathbb{W}_e$
   (2.32)

$$q_i(t) = \int_0^t f_i\Bigg(y;\Big[\sum_{m\in\mathbb{W}_e} u_{im}q_m(t-y)\Big]+\Big[\sum_{r\in\mathbb{W}_0} u_{ir}q_r(t-y)\Big]\Bigg)dG_i(y),$$

Which can be rewritten as

$$q_i(t) = \int_0^t f_i\Big(y;\sum_{r\in\mathbb{W}} u_{ir}q_r(t-y)\Big)dG_i(y).$$

2. For $i \in \mathbb{W}_0$

$$q_i(t) = \int_0^t f_i\Big(y; \sum_{r\in\mathbb{W}_0} u_{ir} q_r(t-y)\Big) dG_i(y). \tag{2.33}$$

Proof. We recognize that $F_i(t;\boldsymbol{s})$ and $F_i(t;\boldsymbol{s}_{\mathbb{W}_0})$ are p.g.f.s and apply $\boldsymbol{s} = \mathbf{0}$ into the result of Corollary 2.12. □

Corollary 2.15. *The following system of integral equations holds for the DMSBPM:*

1. For $i \in \mathbb{W}_e$

$$q_{i,a}(t) = \int_0^t f_i\left(a+y; \Big[\sum_{m\in\mathbb{W}_e} u_{im} q_m(t-y)\Big] + \Big[\sum_{r\in\mathbb{W}_0} u_{ir} q_r(t-y)\Big]\right) dG_{i,a}(y), \tag{2.34}$$

Which can be rewritten as

$$q_{i,a}(t) = \int_0^t f_i\Big(a+y; \sum_{r\in\mathbb{W}} u_{ir} q_r(t-y)\Big) dG_{i,a}(y).$$

2. For $i \in \mathbb{W}_0$

$$q_{i,a}(t) = \int_0^t f_i\Big(a+y; \sum_{r\in\mathbb{W}_0} u_{ir} q_r(t-y)\Big) dG_{i,a}(y). \tag{2.35}$$

Proof. We recognize that $F_{i,a}(t;\boldsymbol{s})$ and $F_i(t;\boldsymbol{s}_{\mathbb{W}_0})$ are p.g.f.s and apply $\boldsymbol{s} = \mathbf{0}$ into the result of Corollary 2.13. □

It is also easy to see that:

Corollary 2.16. *The following system of equations holds for the DMS-BPM:*

1. For $i \in \mathbb{W}_e$

$$q_i = \int_0^\infty f_i\left(y; \Big[\sum_{m\in\mathbb{W}_e} u_{im} q_m\Big] + \Big[\sum_{r\in\mathbb{W}_0} u_{ir} q_r\Big]\right) dG_i(y). \tag{2.36}$$

Which can be rewritten as

$$q_i = \int_0^\infty f_i\Big(y; \sum_{r\in\mathbb{W}} u_{ir} q_r\Big) dG_i(y).$$

2. For $i \in \mathbb{W}_0$

$$q_i = \int_0^\infty f_i\Big(y; \sum_{r \in \mathbb{W}_0} u_{ir} q_r\Big) dG_i(y). \tag{2.37}$$

Proof. The proof follows from the result of Theorem 2.3. For $i \in \mathbb{W}_0$, we have set all $u_{ij} = 0, j \in \mathbb{W}_e$. □

Corollary 2.17. *The following system of equations holds for the DMS-BPM:*

1. For $i \in \mathbb{W}_e$

$$q_{i,a} = \int_0^\infty f_i\left(a + y; \Big[\sum_{m \in \mathbb{W}_e} u_{im} q_m\Big] + \Big[\sum_{r \in \mathbb{W}_0} u_{ir} q_r\Big]\right) dG_{i,a}(y). \tag{2.38}$$

Which can be rewritten as

$$q_{i,a} = \int_0^\infty f_i\Big(a + y; \sum_{r \in \mathbb{W}} u_{ir} q_r\Big) dG_{i,a}(y).$$

2. For $i \in \mathbb{W}_0$

$$q_{i,a} = \int_0^\infty f_i\Big(a + y; \sum_{r \in \mathbb{W}_0} u_{ir} q_r\Big) dG_{i,a}(y). \tag{2.39}$$

Proof. The proof follows from the result of Corollary 2.3. For $i \in \mathbb{W}_0$, we have set all $u_{ij} = 0, j \in \mathbb{W}_e$. □

For our final result concerting probabilities of extinction, we direct the reader to Proposition 2.1 within Subsection 2.3.1.4. This results is not necessarily true for a general MSBPM as we saw from Figure 2.6 and Figure 2.7 in Subsection 2.2.4.

We illustrate the behavior of the probabilities of extinction of the DMS-BPM for mutation scheme “$\mathbb{W}_e$ towards $\mathbb{W}_0$” (Table 2.8) in Figure 2.13 and 2.14, as well as for mutation scheme “$\mathbb{W}_e$ towards $\mathbb{W}_0, \mathbb{W}_0$ forms a chain” (Table 2.9) in Figure 2.15 and Figure 2.16.

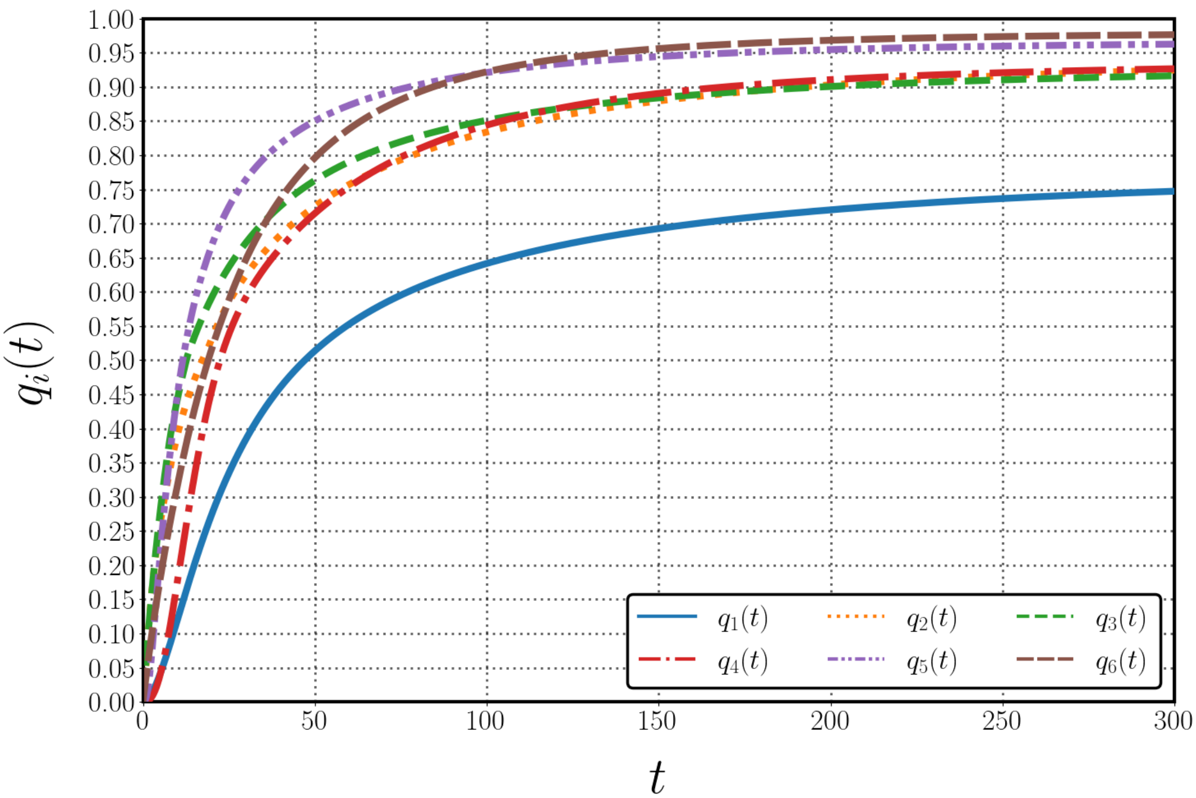

Figure 2.13: An application of Corollary 2.14 - probabilities for extinction of the example DMSBPM (Table 2.11, Table 2.6) with mutation scheme "$\mathbb{W}_e$ towards $\mathbb{W}_0$" (Table 2.8) starting within one particle of age 0. Displayed values are cut at $t = 300$ ($h = 10^{-2}$) so that the dynamics of the different $q_i(t)$ within $[0, 300]$ is visible. In accordance with the preliminary analysis from Subsection 2.2.8.4, if the DMSBPM starts with a particle with type from $\mathbb{W}_0$, the criticality of the process is higher in comparison with the case of Figure 2.2, thus q_1, q_2, and q_3, are smaller in comparison with Figure 2.2. More specifically, at $t = 1500$ ($h = 10^{-1}$), we have $q_1 \approx 0.77206078$, $q_2 \approx 0.94263439$, $q_3 \approx 0.92948039$, $q_4 \approx 0.93797045$, $q_5 \approx 0.96806635$, $q_6 \approx 0.98125857$.

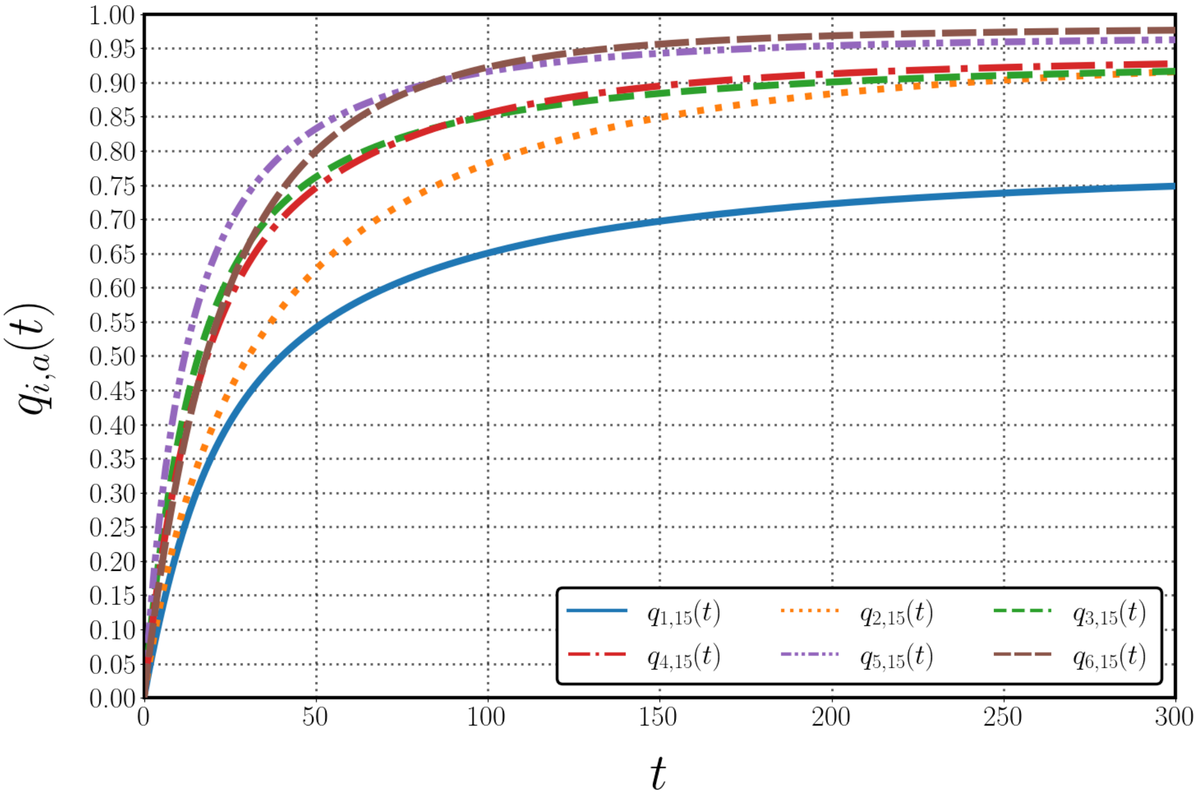

Figure 2.14: An application of Corollary 2.15 - probabilities of extinction for the example DMSBPM (Table 2.11, Table 2.6) with mutation scheme "$\mathbb{W}_e$ towards $\mathbb{W}_0$" (Table 2.8) starting within one particle of age $a = 15$. Displayed values are cut at $t = 300$ ($h = 10^{-2}$) so that the dynamics of the different $q_i(t)$ within $[0, 300]$ is visible. We note that, with respect to Figure 2.13, $q_{1,15}(t)$ is slightly more curved when compared to $q_1(t)$. At $t = 1500$ ($h = 10^{-1}$), we have $q_{1,15} \approx 0.77206255$, $q_{2,15} \approx 0.94251251$, $q_{3,15} \approx 0.92948039$, $q_{4,15} \approx 0.93797146$, $q_{5,15} \approx 0.96806622$, $q_{6,15} \approx 0.98125857$.

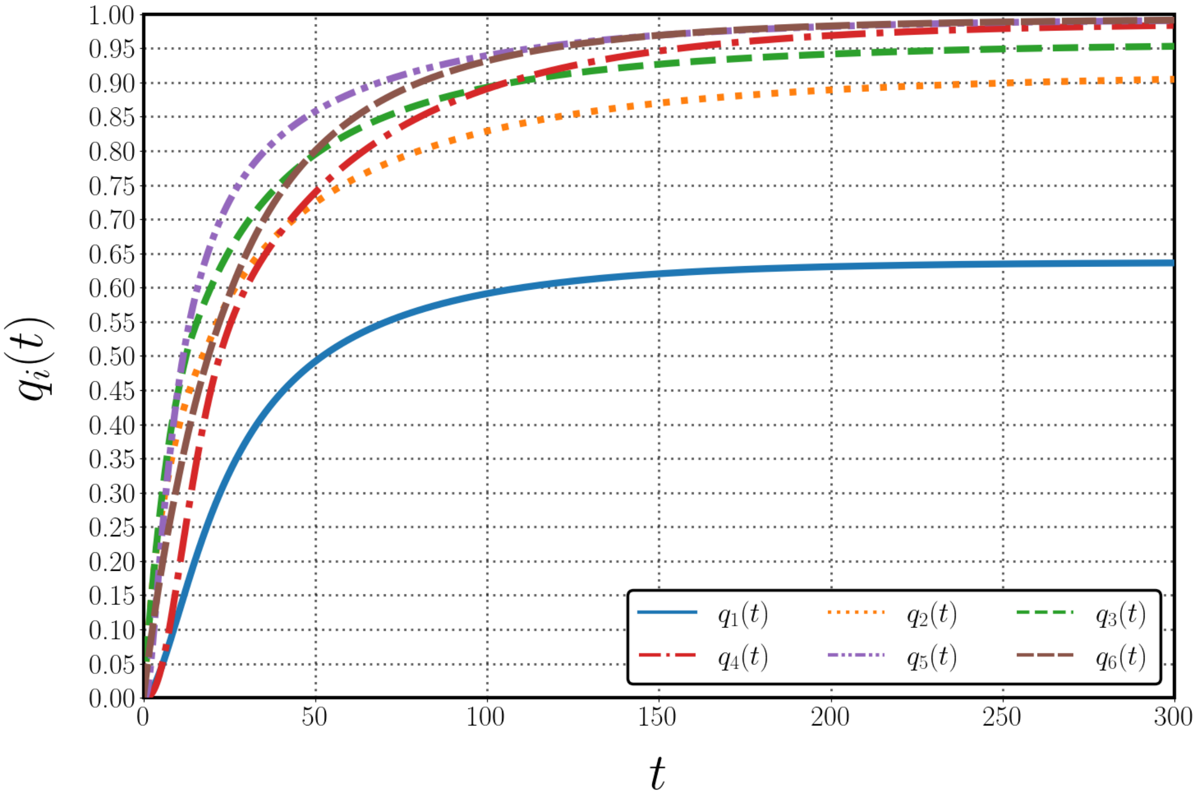

Figure 2.15: An application of Corollary 2.14 - probabilities for extinction of the example DMSBPM (Table 2.11, Table 2.6) with mutation scheme "$\mathbb{W}_e$ towards $\mathbb{W}_0, \mathbb{W}_0$ forms a chain" (Table 2.9) starting within one particle of age 0. Displayed values are cut at $t = 300$ ($h = 10^{-2}$) so that the dynamics of the different $q_i(t)$ within $[0, 300]$ is visible. Unlike the case in Figure 2.13, here type 1 does not have to "share" offspring with other types, leading to q_1 being smaller. Type 2, being one mutation away from type 1, has the second smallest q_i. Type 3, being two mutations away from type 1 has the third smallest q_i. More specifically, at $t = 1500$ ($h = 10^{-1}$), we have $q_1 \approx 0.63701453$, $q_2 \approx 0.91447893$, $q_3 \approx 0.95914728$, $q_4 \approx 0.98841861$, $q_5 \approx 0.99346401$, $q_6 \approx 0.99375408$.

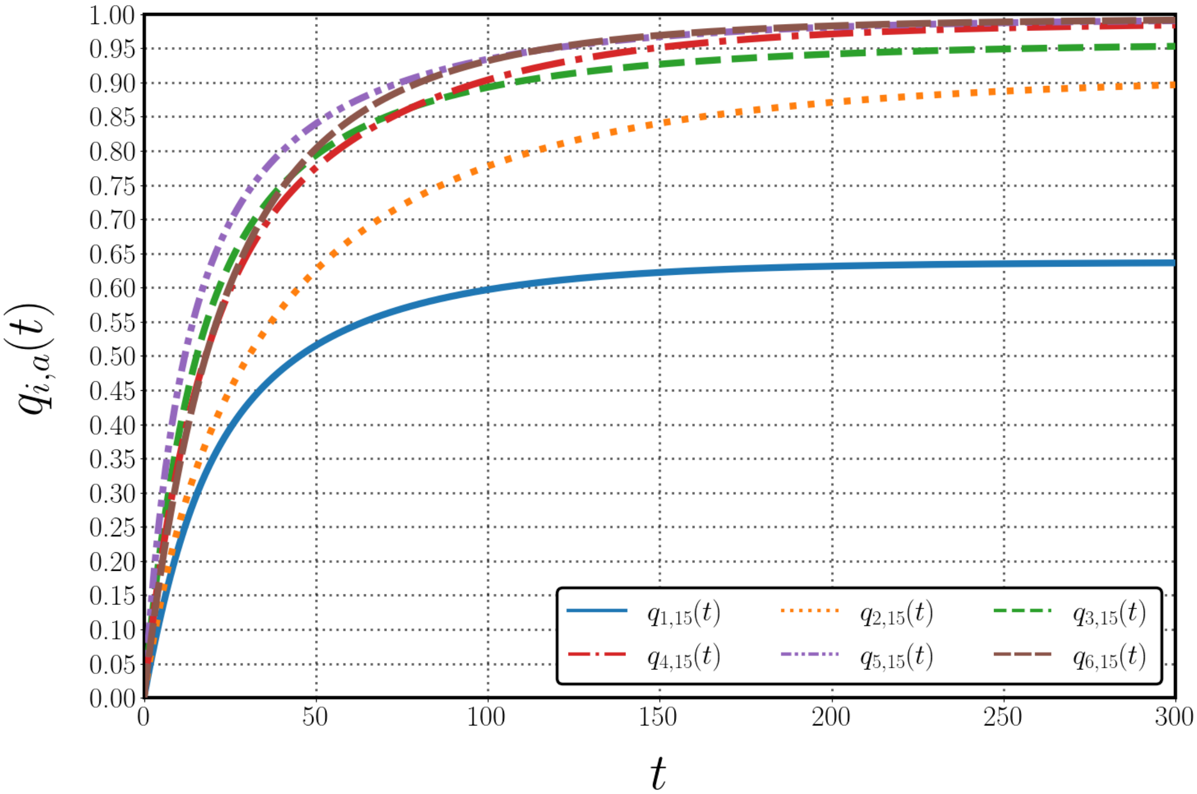

Figure 2.16: An application of Corollary 2.15 - probabilities of extinction for the example DMSBPM (Table 2.11, Table 2.6) with mutation scheme "$\mathbb{W}_e$ towards $\mathbb{W}_0, \mathbb{W}_0$ forms a chain" (Table 2.9) starting within one particle of age $a = 15$. Displayed values are cut at $t = 300$ ($h = 10^{-2}$) so that the dynamics of the different $q_i(t)$ within $[0, 300]$ is visible. With respect to Figure 2.15, the evolution of $q_{2,15}(t)$ in $[0, 150]$ is significantly altered by the age of the initial particle. Regardless, we see that for large t all $q_{i,15}$ approach the values of the corresponding q_i from Figure 2.15. At $t = 1500$ ($h = 10^{-1}$), we have $q_{1,15} \approx 0.63701453$, $q_{2,15} \approx 0.91436551$, $q_{3,15} \approx 0.95914728$, $q_{4,15} \approx 0.98841889$, $q_{5,15} \approx 0.99346397$, $q_{6,15} \approx 0.99375408$.

2.3.1.4 Number of mutants produced from $\mathbb{W}_e$ towards $\mathbb{W}_0$ within the DMSBPM

If a population consists only of particles with types from $\mathbb{W}_e$ and all particle types from $\mathbb{W}_e$ are subcritical or critical, the only hope for the population to escape extinction is to produce a mutant towards $\mathbb{W}_0$ (assuming that at least one type from $\mathbb{W}_0$ is supercritical). Thus, counting the number of mutations from $\mathbb{W}_e$ towards $\mathbb{W}_0$ is of primary importance in this setting. We already have the general Definition 2.6, where we count all particles produced (mutant or not) from $\mathbb{W}_e$ towards $\mathbb{W}$, now, within the DMSBPM, we will be interested into counting only the mutants produced from $\mathbb{W}_e$ towards $\mathbb{W}_0$. From Definition 2.6, we have

$$h_i^{\mathbb{W}_e}(t;\boldsymbol{s}) = \mathbb{E}\Big(\prod_{j\in\mathbb{W}} s_j^{I_j^{\mathbb{W}_e}(t)} \mid \boldsymbol{Z}(0) = \boldsymbol{\delta}^i\Big)$$

and

$$h_{i,a}^{\mathbb{W}_e}(t;\boldsymbol{s}) = \mathbb{E}\Big(\prod_{j\in\mathbb{W}} s_j^{I_j^{\mathbb{W}_e}(t)} \mid \boldsymbol{Z}(0) = \boldsymbol{\delta}_a^i\Big),$$

where $|\boldsymbol{s}| \le 1$. We can obtain the information regarding mutants produced from $\mathbb{W}_e$ towards $\mathbb{W}_0$ by setting $\boldsymbol{s} = \boldsymbol{s}_{\mathbb{W}_0}$. Further, it is clear that if the initial particle is of type i, $i \in \mathbb{W}_0$, then no mutants from $\mathbb{W}_e$ towards $\mathbb{W}_0$ can occur, hence, for this case, $h_i^{\mathbb{W}_e}(t;\boldsymbol{s}) = h_{i,a}^{\mathbb{W}_e}(t;\boldsymbol{s}) = 1$. Following these considerations, in the context of the DMSBPM, for clarity and convenience, we give the following definition.

Definition 2.13. *Given Definition 2.6, for a DMSBPM starting with one particle of type i, $i \in \mathbb{W}$, denote the p.g.f.s for the numbers of mutants produced from $\mathbb{W}_e$ towards $\mathbb{W}_0$, until t, as*

1. *For $i \in \mathbb{W}_e$*

$$h_i^{\mathbb{W}_e}(t;\boldsymbol{s}_{\mathbb{W}_0}) = \mathbb{E}\Big(\prod_{j\in\mathbb{W}_0} s_j^{I_j^{\mathbb{W}_e}(t)} \mid \boldsymbol{Z}(0) = \boldsymbol{\delta}^i\Big),$$

$$h_{i,a}^{\mathbb{W}_e}(t;\boldsymbol{s}_{\mathbb{W}_0}) = \mathbb{E}\Big(\prod_{j\in\mathbb{W}_0} s_j^{I_j^{\mathbb{W}_e}(t)} \mid \boldsymbol{Z}(0) = \boldsymbol{\delta}_a^i\Big),$$

where $|\boldsymbol{s}| \le 1$.

2. *For $i \in \mathbb{W}_0$, due to the fact that there can be no mutations from $\mathbb{W}_0$ towards $\mathbb{W}_e$*

$$h_i^{\mathbb{W}_e}(t;\boldsymbol{s}_{\mathbb{W}_0}) = h_{i,a}^{\mathbb{W}_e}(t;\boldsymbol{s}_{\mathbb{W}_0}) = 1.$$

Corollary 2.18. *For the DMSBPM the following system of integral equations holds:*

1. *For $i \in \mathbb{W}_e$*

(2.40)
$$h_i^{\mathbb{W}_e}(t;\boldsymbol{s}_{\mathbb{W}_0}) = \big(1-G_i(t)\big) + \int_0^t f_i\Bigg(y;\Big[\sum_{m\in\mathbb{W}_e} u_{im}h_m^{\mathbb{W}_e}(t-y;\boldsymbol{s}_{\mathbb{W}_0})\Big]+ + \Big[\sum_{r\in\mathbb{W}_0} u_{ir}s_r\Big]\Bigg)dG_i(y).$$

2. *For $i \in \mathbb{W}_0$*
$$h_i^{\mathbb{W}_e}(t;\boldsymbol{s}_{\mathbb{W}_0}) = 1.$$

Proof. From Definition 2.13, we know that $h_i^{\mathbb{W}_e}(t;\boldsymbol{s}_{\mathbb{W}_0}) = 1$, for $i \in \mathbb{W}_0$. Recall Theorem 2.4 and Remark 2.1 (the statement of Remark 2.1 is valid with respect to Theorem 2.4 as well). From Theorem 2.4 we have equations (2.10)
$$h_i^{\mathbb{W}_e}(t;\boldsymbol{s}) = \big(1-G_i(t)\big) + \int_0^t f_i\Big(y;\sum_{r\in\mathbb{W}} u_{ir}s_r h_r^{\mathbb{W}_e}(t-y;\boldsymbol{s})\Big)dG_i(y),\ i\in\mathbb{W}_e.$$
Substituting $\boldsymbol{s}=\boldsymbol{s}_{\mathbb{W}_0}$ yields
$$h_i^{\mathbb{W}_e}(t;\boldsymbol{s}_{\mathbb{W}_0}) = \big(1-G_i(t)\big)+\int_0^t f_i\Bigg(y;\Big[\sum_{m\in\mathbb{W}_e} u_{im}h_m^{\mathbb{W}_e}(t-y;\boldsymbol{s}_{\mathbb{W}_0})\Big]+ + \Big[\sum_{r\in\mathbb{W}_0} u_{ir}s_r h_r^{\mathbb{W}_e}(t-y;\boldsymbol{s}_{\mathbb{W}_0})\Big]\Bigg)dG_i(y),\ i\in\mathbb{W}_e.$$
Taking into account that $h_i^{\mathbb{W}_e}(t;\boldsymbol{s}_{\mathbb{W}_0}) = 1, i\in\mathbb{W}_0$, we obtain the statement of the theorem. □

Corollary 2.19. *For the DMSBPM the following system of integral equations holds:*

1. *For $i \in \mathbb{W}_e$*

(2.41)
$$h_{i,a}^{\mathbb{W}_e}(t;\boldsymbol{s}_{\mathbb{W}_0}) = \big(1-G_{i,a}(t)\big) + \int_0^t f_i\Bigg(a+y;\Big[\sum_{m\in\mathbb{W}_e} u_{im}h_m^{\mathbb{W}_e}(t-y;\boldsymbol{s}_{\mathbb{W}_0})\Big]+ + \Big[\sum_{r\in\mathbb{W}_0} u_{ir}s_r\Big]\Bigg)dG_{i,a}(y).$$

2. *For $i \in \mathbb{W}_0$*
$$h_{i,a}^{\mathbb{W}_e}(t; \boldsymbol{s}_{\mathbb{W}_0}) = 1.$$

Proof. The proof is analogous to the proof of Corollary 2.18. During the proof, we use the result of Corollary 2.6. □

Definition 2.14. *Given Definition 2.7, for a DMSBPM starting with one particle of type i, $i \in \mathbb{W}$, denote the p.g.f.s for the numbers of mutants produced from $\mathbb{W}_e$ towards $\mathbb{W}_0$, during the entire process, as*

1. *For $i \in \mathbb{W}_e$*

$$h_i^{\mathbb{W}_e}(\boldsymbol{s}_{\mathbb{W}_0}) = \mathbb{E}\Big(\prod_{j \in \mathbb{W}_0} s_j^{I_j^{\mathbb{W}_e}} \mid \boldsymbol{Z}(0) = \boldsymbol{\delta}^i \Big),$$

$$h_{i,a}^{\mathbb{W}_e}(\boldsymbol{s}_{\mathbb{W}_0}) = \mathbb{E}\Big(\prod_{j \in \mathbb{W}_0} s_j^{I_j^{\mathbb{W}_e}} \mid \boldsymbol{Z}(0) = \boldsymbol{\delta}_a^i \Big),$$

where $|\boldsymbol{s}| \leq 1$.

2. *For $i \in \mathbb{W}_0$, due to the fact that there can be no mutations from $\mathbb{W}_0$ towards $\mathbb{W}_e$*
$$h_i^{\mathbb{W}_e}(\boldsymbol{s}_{\mathbb{W}_0}) = h_{i,a}^{\mathbb{W}_e}(\boldsymbol{s}_{\mathbb{W}_0}) = 1.$$

Remark 2.15. *As a consequence of Remark 2.8, we also have that $\lim\limits_{y\to\infty} \lim\limits_{t\to\infty} h_i^{\mathbb{W}_e}(t-y; \boldsymbol{s}_{\mathbb{W}_0}) = h_i^{\mathbb{W}_e}(\boldsymbol{s}_{\mathbb{W}_0})$ and $\lim\limits_{y\to\infty} \lim\limits_{t\to\infty} h_{i,a}^{\mathbb{W}_e}(t-y; \boldsymbol{s}_{\mathbb{W}_0}) = h_{i,a}^{\mathbb{W}_e}(\boldsymbol{s}_{\mathbb{W}_0})$.*

Corollary 2.20. *The following system of equations holds within the DMSBPM, $i \in \mathbb{W}$:*

1. *For $i \in \mathbb{W}_e$*

(2.42)
$$h_i^{\mathbb{W}_e}(\boldsymbol{s}_{\mathbb{W}_0}) = \int_0^\infty f_i\Bigg(y; \Big[\sum_{m \in \mathbb{W}_e} u_{im} h_m^{\mathbb{W}_e}(\boldsymbol{s}_{\mathbb{W}_0}) \Big] + \Big[\sum_{r \in \mathbb{W}_0} u_{ir} s_r \Big] \Bigg) dG_i(y).$$

2. *For $i \in \mathbb{W}_0$*

$$h_i^{\mathbb{W}_e}(\boldsymbol{s}_{\mathbb{W}_0}) = 1. \tag{2.43}$$

Proof. From Definition 2.14, we have that $h_i^{\mathbb{W}_e}(\boldsymbol{s}_{\mathbb{W}_0}) = 1$, for $i \in \mathbb{W}_0$. Recall Theorem 2.5 and Remark 2.1 (the statement of Remark 2.1 is valid with respect to Theorem 2.5 as well). From Theorem 2.5 we have equations (2.18)

$$h_i^{\mathbb{W}_e}(\boldsymbol{s}) = \int_0^\infty f_i\Big(y; \sum_{r\in\mathbb{W}} u_{ir} s_r h_r^{\mathbb{W}_e}(\boldsymbol{s})\Big) dG_i(y), \ i \in \mathbb{W}_e.$$

Substituting $\boldsymbol{s} = \boldsymbol{s}_{\mathbb{W}_0}$ and applying $h_i^{\mathbb{W}_e}(\boldsymbol{s}_{\mathbb{W}_0}) = 1$, when $i \in \mathbb{W}_0$, yields the statement of the theorem. □

Corollary 2.21. *For the DMSBPM the following system of integral equations holds:*

1. *For* $i \in \mathbb{W}_e$
(2.44)
$$h_{i,a}^{\mathbb{W}_e}(\boldsymbol{s}_{\mathbb{W}_0}) = \int_0^\infty f_i\left(a+y; \Big[\sum_{m\in\mathbb{W}_e} u_{im} h_m^{\mathbb{W}_e}(\boldsymbol{s}_{\mathbb{W}_0})\Big] + \Big[\sum_{r\in\mathbb{W}_0} u_{ir} s_r\Big]\right) dG_{i,a}(y).$$

2. *For* $i \in \mathbb{W}_0$
$$h_{i,a}^{\mathbb{W}_e}(\boldsymbol{s}_{\mathbb{W}_0}) = 1.$$

Proof. The proof is analogous to the proof of Corollary 2.20. During the proof, we use the result of Corollary 2.9. □

Expectations for $I_j^{\mathbb{W}_e}(t)$, $i \in \mathbb{W}$, can be handled as in Subsection 2.2.4.

Proposition 2.1. *Let each particle type from* $\mathbb{W}_e$*, within the DMSBPM, be either subcritical or critical. Then for* $i \in \mathbb{W}_e$
(2.45)

$$q_i = h_i^{\mathbb{W}_e}(\boldsymbol{q}_{\mathbb{W}_0}) = \int_0^\infty f_i\left(y; \Big[\sum_{m\in\mathbb{W}_e} u_{im} h_m^{\mathbb{W}_e}(\boldsymbol{q}_{\mathbb{W}_0})\Big] + \Big[\sum_{r\in\mathbb{W}_0} u_{ir} q_r\Big]\right) dG_i(y)$$

and
(2.46)

$$q_{i,a} = h_{i,a}^{\mathbb{W}_e}(\boldsymbol{q}_{\mathbb{W}_0}) = \int_0^\infty f_i\left(a+y; \Big[\sum_{m\in\mathbb{W}_e} u_{im} h_m^{\mathbb{W}_e}(\boldsymbol{q}_{\mathbb{W}_0})\Big] + \Big[\sum_{r\in\mathbb{W}_0} u_{ir} q_r\Big]\right) dG_{i,a}(y).$$

Proof. Under these assumptions, in order for a DMSBPM starting with one particle of type $i \in \mathbb{W}_e$, of age 0, to become extinct, each occurring mutant from $\mathbb{W}_e$ towards $\mathbb{W}_0$ must initiate a process that becomes extinct. Using the total probability formula we obtain

$$\begin{aligned} q_i &= \sum_{k=0}^{\infty}\left[\sum_{\sum_{m\in\mathbb{W}_0} k_m = k}\left[\mathbb{P}\Big(I_j^{\mathbb{W}_e} = k_j, j \in \mathbb{W}_0 \mid \boldsymbol{Z}(0) = \boldsymbol{\delta}^i\Big)\cdot\prod_{r\in\mathbb{W}_0} q_r^{k_r}\right]\right] \\ &= h_i^{\mathbb{W}_e}(\boldsymbol{q}_{\mathbb{W}_0}). \end{aligned}$$

Analogously, if a DMSBPM starts with a particle that is of age a, $a \neq 0$

$$\begin{aligned} q_{i,a} &= \sum_{k=0}^{\infty}\left[\sum_{\sum_{m\in\mathbb{W}_0} k_m = k}\left[\mathbb{P}\Big(I_j^{\mathbb{W}_e} = k_j, j \in \mathbb{W}_0 \mid \boldsymbol{Z}(0) = \boldsymbol{\delta}_a^i\Big)\cdot\prod_{r\in\mathbb{W}_0} q_r^{k_r}\right]\right] \\ &= h_{i,a}^{\mathbb{W}_e}(\boldsymbol{q}_{\mathbb{W}_0}). \end{aligned}$$

□

In the following figures, we illustrate the results obtained and highlight the result of Proposition 2.1 in Figure 2.21. Again, the figures indicate that perhaps more general results, than the ones of Corollary 2.4 and Corollary 2.11, exist.

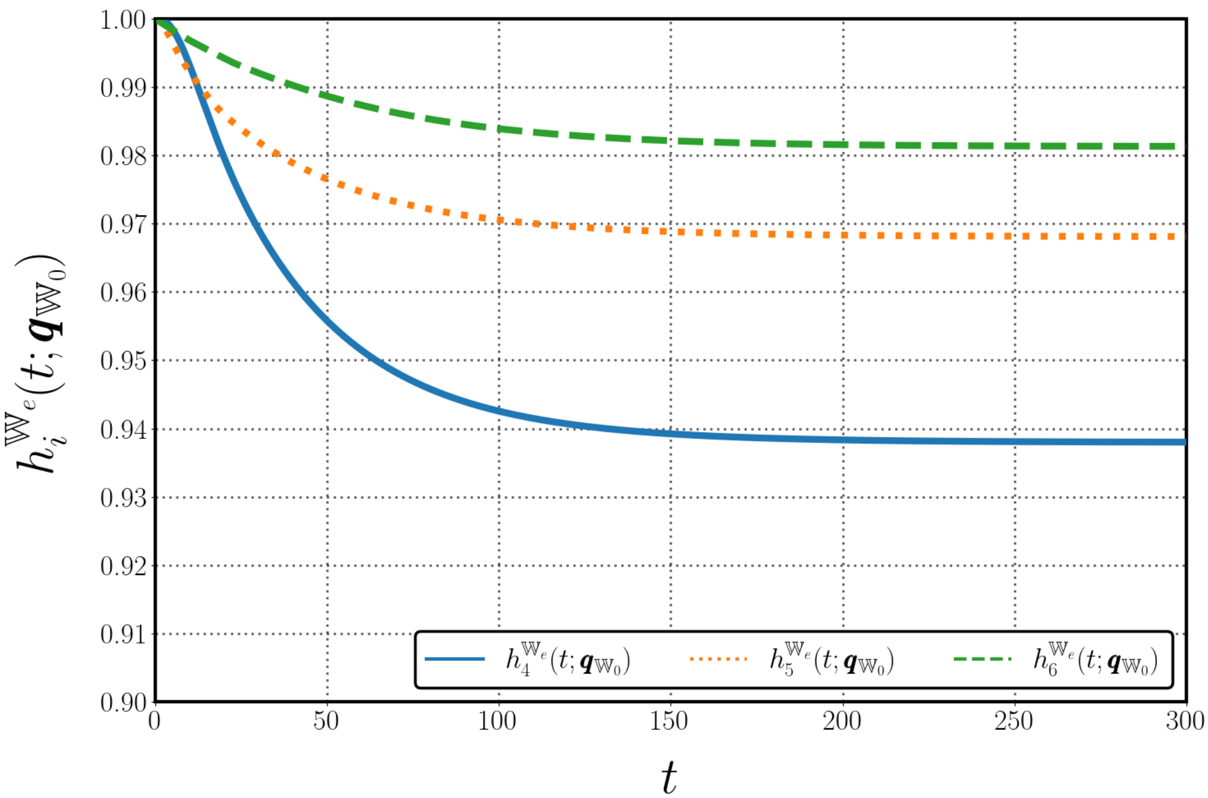

Figure 2.17: Calculations ($h = 10^{-2}$) for $\boldsymbol{s} = \boldsymbol{q}_{\mathbb{W}_0}$ for the example DMS-BPM (Table 2.11, Table 2.6) with mutation scheme "$\mathbb{W}_e$ towards $\mathbb{W}_0$" (Table 2.8) starting with one particle of age 0. Recall from Figure 2.13 that $\boldsymbol{q} = (0.77206078, 0.94263439, 0.92948039, 0.93797045, 0.96806635, 0.98125857)^\top$. Thus, $\boldsymbol{q}_{\mathbb{W}_0} = (0.77206078, 0.94263439, 0.92948039, 1.0, 1.0, 1.0)^\top$. At $t = 1500$, we have $h_4^{\mathbb{W}_e}(\boldsymbol{q}_{\mathbb{W}_0}) = 0.93797589$, $h_5^{\mathbb{W}_e}(\boldsymbol{q}_{\mathbb{W}_0}) = 0.96806891$, $h_6^{\mathbb{W}_e}(\boldsymbol{q}_{\mathbb{W}_0}) = 0.98126128$. Evidently, $\boldsymbol{q}_i = h_i^{\mathbb{W}_e}(\boldsymbol{q}_{\mathbb{W}_0})$ is true for $i \in \mathbb{W}_e$, as per Proposition 2.1.

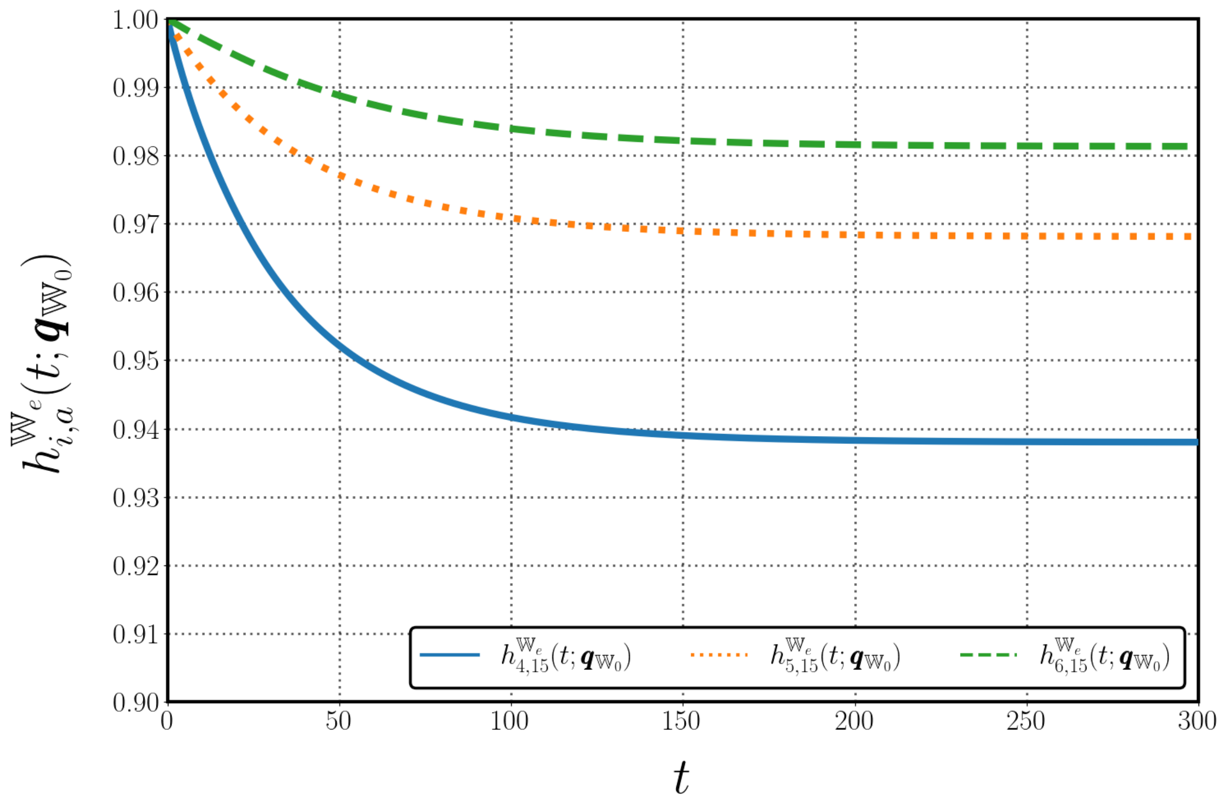

Figure 2.18: Calculations ($h = 10^{-2}$) for $\boldsymbol{s} = \boldsymbol{q}_{\mathbb{W}_0}$ for the example DMSBPM (Table 2.11, Table 2.6) with mutation scheme "$\mathbb{W}_e$ towards $\mathbb{W}_0$" (Table 2.8) starting with one particle of age $a = 15$. Recall from Figure 2.13 that $\boldsymbol{q} = (0.77206078, 0.94263439, 0.92948039, 0.93797045, 0.96806635, 0.98125857)^\top$, consequently $\boldsymbol{q}_{\mathbb{W}_0} = (0.77206078, 0.94263439, 0.92948039, 1.0, 1.0, 1.0)^\top$. At $t = 1500$, we have $h_{4,15}^{\mathbb{W}_e}(\boldsymbol{q}_{\mathbb{W}_0}) = 0.93797589$, $h_{5,15}^{\mathbb{W}_e}(\boldsymbol{q}_{\mathbb{W}_0}) = 0.96806891$, $h_{6,15}^{\mathbb{W}_e}(\boldsymbol{q}_{\mathbb{W}_0}) = 0.98126128$. Recall from Figure 2.14 that $\boldsymbol{q}_{15} = (0.77206255, 0.94251251, 0.92948039, 0.93797146, 0.96806622, 0.98125857)^\top$. Evidently, $q_{i,a} = h_{i,a}^{\mathbb{W}_e}(\boldsymbol{q}_{\mathbb{W}_0})$ is true for $i \in \mathbb{W}_e$, as per Proposition 2.1.

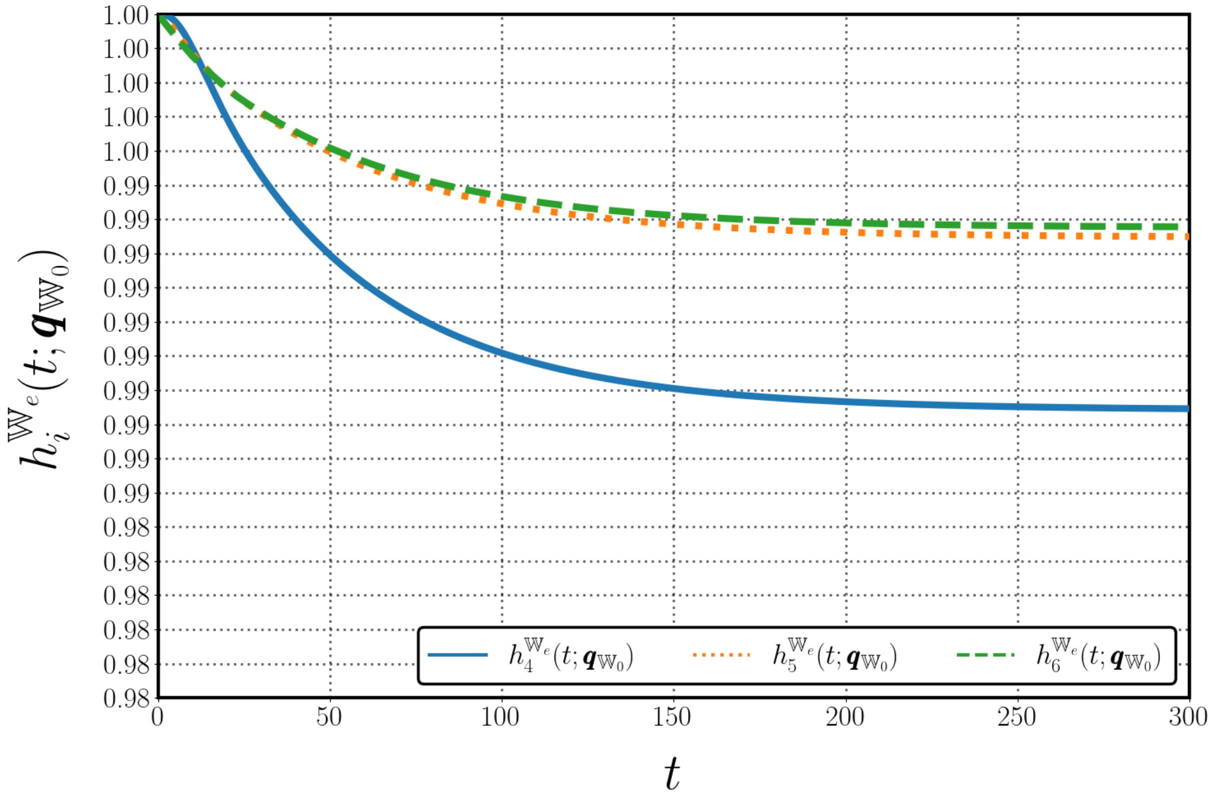

Figure 2.19: Calculations ($h = 10^{-2}$) for $\boldsymbol{s} = \boldsymbol{q}_{\mathbb{W}_0}$ for the example DMSBPM (Table 2.11, Table 2.6) with mutation scheme "$\mathbb{W}_e$ towards $\mathbb{W}_0, \mathbb{W}_0$ forms a chain" (Table 2.9) starting with one particle of age 0. Recall from Figure 2.15 that $\boldsymbol{q} = (0.63701453, 0.91447893, 0.95914728, 0.98841861, 0.99346401, 0.99375408)^\top$. Thus, $\boldsymbol{q}_{\mathbb{W}_0} = (0.63701453, 0.91447893, 0.95914728, 1.0, 1.0, 1.0)^\top$. At $t = 1500$, we have $h_4^{\mathbb{W}_e}(\boldsymbol{q}_{\mathbb{W}_0}) = 0.98842052$, $h_5^{\mathbb{W}_e}(\boldsymbol{q}_{\mathbb{W}_0}) = 0.99346507$, $h_6^{\mathbb{W}_e}(\boldsymbol{q}_{\mathbb{W}_0}) = 0.99375507$. Evidently, $q_i = h_i^{\mathbb{W}_e}(\boldsymbol{q}_{\mathbb{W}_0})$ is true for $i \in \mathbb{W}_e$, as per Proposition 2.1.

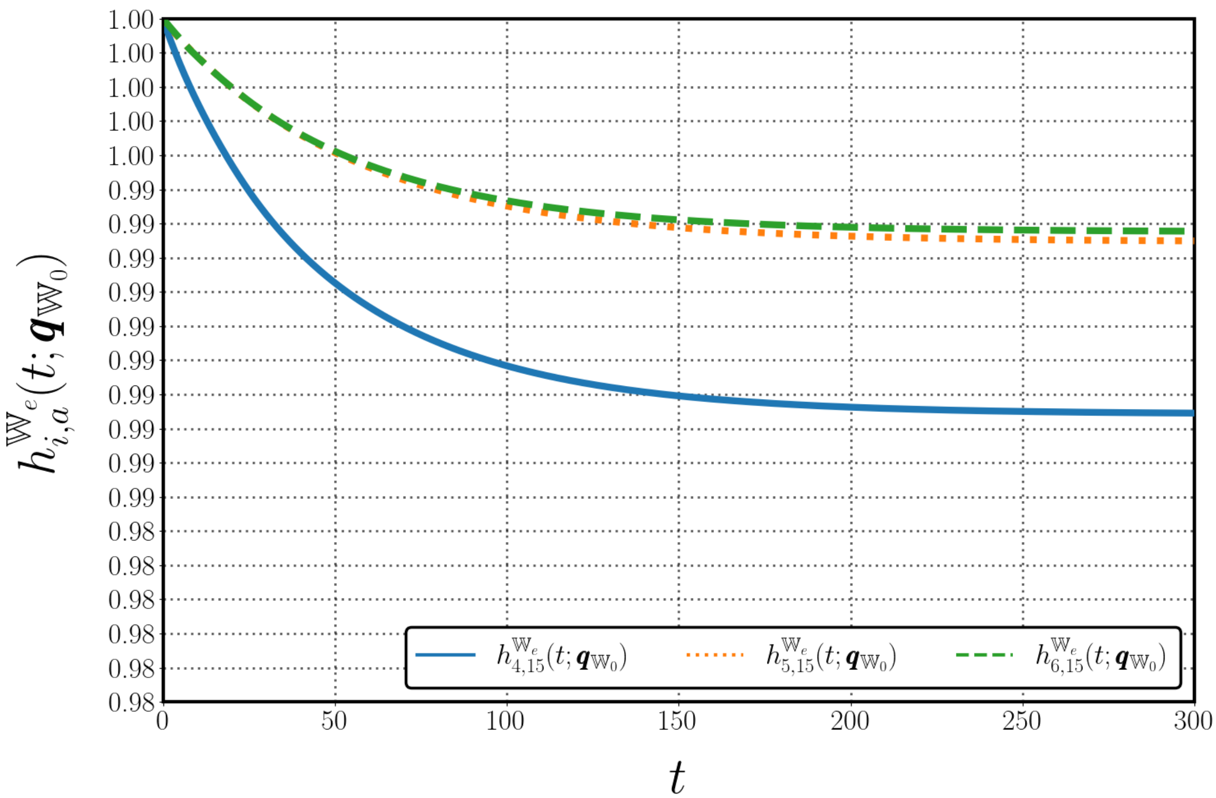

Figure 2.20: Calculations ($h = 10^{-2}$) for $\boldsymbol{s} = \boldsymbol{q}_{\mathbb{W}_0}$ for the example DMSBPM (Table 2.11, Table 2.6) with mutation scheme "$\mathbb{W}_e$ towards $\mathbb{W}_0, \mathbb{W}_0$ forms a chain" (Table 2.9) starting with one particle of age $a = 15$. Recall from Figure 2.15 that $\boldsymbol{q} = (0.63701453, 0.91447893, 0.95914728, 0.98841861, 0.99346401, 0.99375408)^\top$, consequently $\boldsymbol{q}_{\mathbb{W}_0} = (0.63701453, 0.91447893, 0.95914728, 1.0, 1.0, 1.0)^\top$. At $t = 1500$, we have $h^{\mathbb{W}_e}_{4,15}(\boldsymbol{q}_{\mathbb{W}_0}) = 0.98842052$, $h^{\mathbb{W}_e}_{5,15}(\boldsymbol{q}_{\mathbb{W}_0}) = 0.99346507$, $h^{\mathbb{W}_e}_{6,15}(\boldsymbol{q}_{\mathbb{W}_0}) = 0.99375507$. Recall from Figure 2.16 that $\boldsymbol{q}_{15} = (0.63701453, 0.91436551, 0.95914728, 0.98841889, 0.99346397, 0.99375408)^\top$. Evidently, $\boldsymbol{q}_{i,a} = h^{\mathbb{W}_e}_{i,a}(\boldsymbol{q}_{\mathbb{W}_0})$ is true for $i \in \mathbb{W}_e$, as per Proposition 2.1.

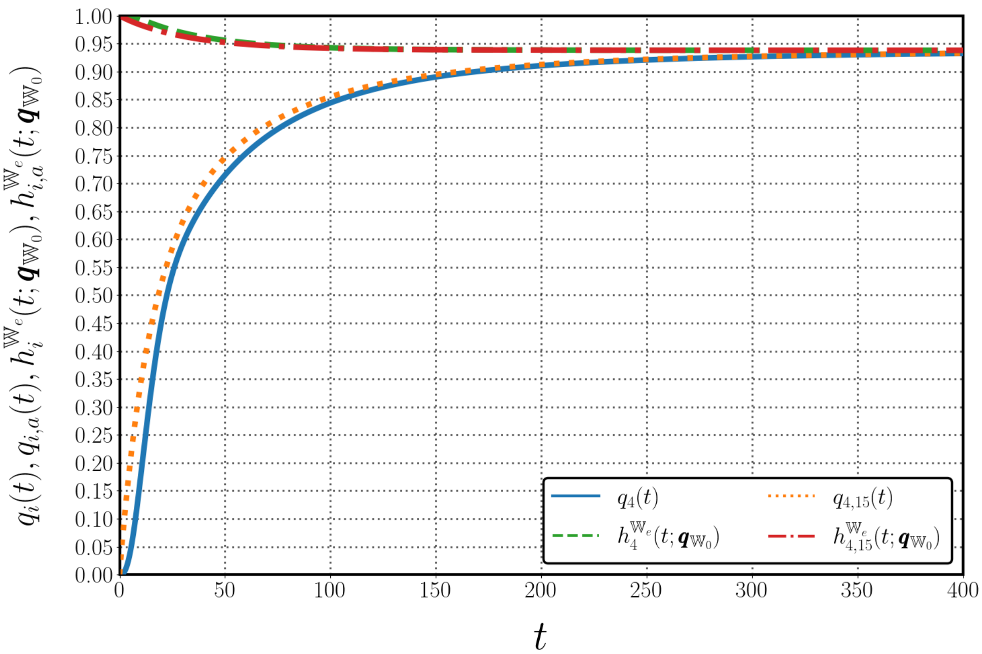

Figure 2.21: Examination of the result of Proposition 2.1 for the example DMSBPM (Table 2.11, Table 2.6) with mutation scheme "$\mathbb{W}_e$ towards $\mathbb{W}_0$" (Table 2.8). Displayed lines are for type 4. The figures corresponding to type 5 and type 6 are analogous. The limit value, computed with $t = 1500$ ($h = 10^{-2}$), is 0.93797589.

2.3.1.5 Time until occurrence of the first "successful" mutant produced from $\mathbb{W}_e$ towards $\mathbb{W}_0$ within the DMSBPM

We call a mutant from $\mathbb{W}_e$ towards $\mathbb{W}_0$ "successful" if it initiates a non-extincting DMSBPM.

Definition 2.15. *Denote with $T_{\mathbb{W}_0}^{\mathbb{W}_e}$ the r.v. that is the time until occurrence of the first "successful" mutant produced from a type within $\mathbb{W}_e$ towards a type within $\mathbb{W}_0$ in a DMSBPM starting with some combination of particles with types within $\mathbb{W}_e$. Without loss of generality, we set the starting number of particles per type $r \in \mathbb{W}_e$ to be k_r and denote the initial state of the process as $\boldsymbol{Z}(0) = \boldsymbol{\alpha}^* = \big(0, \ldots, 0, Z_{b+1}(0) = k_{b+1}, \ldots, Z_n(0) = k_n\big)^\top$. In $\boldsymbol{\alpha}^*$, we have taken into account the arrangements made at the beginning of Subsection 2.3.1.1 and, without loss of generality, we have set $|\mathbb{W}_0| = b$. We define $T_{\mathbb{W}_0}^{\mathbb{W}_e} = \infty$ as the event that no "successful" mutants occur dur-*

ing a DMSBPM beginning with an initial state $\boldsymbol{\alpha}^*$. *Thus, we may write* $T_{\mathbb{W}_0}^{\mathbb{W}_e} \in (0, \infty]$. *If the DMSBPM starts with a single particle of type* i, $i \in \mathbb{W}_e$, *of age* 0, *we use* $T_{\mathbb{W}_0,i}^{\mathbb{W}_e}$ *as a shortcut notation. If the initial particle is of age* a, $a \neq 0$, *we use* $T_{\mathbb{W}_0,i,a}^{\mathbb{W}_e}$.

Proposition 2.2. *Let the DMSBPM start with* k_r *particles per type* r, $r \in \mathbb{W}_e$. *Let all particles form* $\boldsymbol{\alpha}^*$ *have age* 0. *The distribution of* $T_{\mathbb{W}_0}^{\mathbb{W}_e}$ *has the following properties:*

(i) $\mathbb{P}\Big(T_{\mathbb{W}_0}^{\mathbb{W}_e} > t \mid \boldsymbol{Z}(0) = \boldsymbol{\alpha}^*\Big) = \prod_{r\in\mathbb{W}_e} \Big[h_r^{\mathbb{W}_e}(t; \boldsymbol{q}_{\mathbb{W}_0})\Big]^{k_r}$.

(ii) $\mathbb{P}\Big(T_{\mathbb{W}_0}^{\mathbb{W}_e} = \infty \mid \boldsymbol{Z}(0) = \boldsymbol{\alpha}^*\Big) = \prod_{r\in\mathbb{W}_e} \Big[h_r^{\mathbb{W}_e}(\boldsymbol{q}_{\mathbb{W}_0})\Big]^{k_r}$.

(iii) If at least one particle type $r \in \mathbb{W}_0$ *is supercritical, we have*

$$\mathbb{E}\Big[T_{\mathbb{W}_0}^{\mathbb{W}_e} \mid T_{\mathbb{W}_0}^{\mathbb{W}_e} < \infty,\ \boldsymbol{Z}(0) = \boldsymbol{\alpha}^*\Big] =$$
$$= \frac{1}{1 - \prod_{r\in\mathbb{W}_e} \Big[h_r^{\mathbb{W}_e}(\boldsymbol{q}_{\mathbb{W}_0})\Big]^{k_r}} \int_0^\infty \Bigg[\prod_{r\in\mathbb{W}_e} \Big[h_r^{\mathbb{W}_e}(t; \boldsymbol{q}_{\mathbb{W}_0})\Big]^{k_r} -$$
$$- \prod_{r\in\mathbb{W}_e} \Big[h_r^{\mathbb{W}_e}(\boldsymbol{q}_{\mathbb{W}_0})\Big]^{k_r}\Bigg] dt,$$

if not, then the expectation does not exist.

Proof.
Property (i): Let the process start with a single particle of type i, $i \in \mathbb{W}_e$. The event $\Big\{T_{\mathbb{W}_0,i}^{\mathbb{W}_e} > t\Big\}$ means that if we consider the separate DMSBPM stemming from particles produced from $\mathbb{W}_e$ towards $\mathbb{W}_0$, that have come into existence prior to t or at t, all those processes become extinct. Thus, by the law of total probability:

$$\mathbb{P}\Big(T_{\mathbb{W}_0,i}^{\mathbb{W}_e} > t\Big) =$$
$$= \sum_{k=0}^{\infty} \Bigg[\sum_{\sum_{r\in\mathbb{W}_0} k_r = k} \Big[\mathbb{P}\Big(I_j^{\mathbb{W}_e}(t) = k_j,\ j \in \mathbb{W}_0 \mid \boldsymbol{Z}(0) = \boldsymbol{\delta}^i\Big) \cdot \prod_{m\in\mathbb{W}_0} q_m^{k_m}\Big]\Bigg]$$
$$= h_i^{\mathbb{W}_e}(t; \boldsymbol{q}_{\mathbb{W}_0}).$$

The result for $\boldsymbol{\alpha}^*$ follows from the assumption of independent evolution. The proofs for Property (ii) and Property (iii) are analogous to the proofs found in Theorem 2.6. □

Proposition 2.3. *Let the DMSBPM start with k_r particles per type r, $r \in \mathbb{W}_e$, let the starting particles in $\boldsymbol{\alpha}^*$ have ages $a_{r,c}$, $c \in \{1, 2, \dots, k_r\}$ where $a_{r,c}$ is the age of the c-th particle of type r. We allow $a_{r,c}$ to be 0. The distribution of $T_{\mathbb{W}_0}^{\mathbb{W}_e}$ has the following properties:*

(i) $\mathbb{P}\Big(T_{\mathbb{W}_0}^{\mathbb{W}_e} > t \mid \boldsymbol{Z}(0) = \boldsymbol{\alpha}^*\Big) = \prod_{r\in\mathbb{W}_e}\Big[\prod_{c=1}^{k_r} h_{r,a_{r,c}}^{\mathbb{W}_e}(t; \boldsymbol{q}_{\mathbb{W}_0})\Big].$

(ii) $\mathbb{P}\Big(T_{\mathbb{W}_0}^{\mathbb{W}_e} = \infty \mid \boldsymbol{Z}(0) = \boldsymbol{\alpha}^*\Big) = \prod_{r\in\mathbb{W}_e}\Big[\prod_{c=1}^{k_r} h_{r,a_{r,c}}^{\mathbb{W}_e}(\boldsymbol{q}_{\mathbb{W}_0})\Big].$

(iii) If at least one particle type within $\mathbb{W}$ is supercritical, we have

$$\mathbb{E}\Big[T_{\mathbb{W}_0}^{\mathbb{W}_e} \mid T_{\mathbb{W}_0}^{\mathbb{W}_e} < \infty,\ \boldsymbol{Z}(0) = \boldsymbol{\alpha}^*\Big] =$$

$$= \frac{1}{1 - \prod_{r\in\mathbb{W}_e}\Big[\prod_{c=1}^{k_r} h_{r,a_{r,c}}^{\mathbb{W}_e}(\boldsymbol{q}_{\mathbb{W}_0})\Big]} \int_0^\infty \Bigg[\prod_{r\in\mathbb{W}_e}\Big[\prod_{c=1}^{k_r} h_{r,a_{r,c}}^{\mathbb{W}_e}(t; \boldsymbol{q}_{\mathbb{W}_0})\Big] - \prod_{r\in\mathbb{W}_e}\Big[\prod_{c=1}^{k_r} h_{r,a_{r,c}}^{\mathbb{W}_e}(\boldsymbol{q}_{\mathbb{W}_0})\Big]\Bigg] dt,$$

if not, then the expectation does not exist.

Proof. Property (i): Let the process start with a single particle of type i, $i \in \mathbb{W}_e$, that is of age $a_{i,c}$, $c \in \{1, 2, \dots, k_i\}$. The event $\Big\{T_{\mathbb{W}_0,i,a_{i,c}}^{\mathbb{W}_e} > t\Big\}$ means that if we consider the separate DMSBPM stemming from particles produced from $\mathbb{W}_e$ towards $\mathbb{W}_0$, that have come into existence prior to t or at t, all those processes become extinct. Thus, by the law of total probability:

$$\begin{aligned}
&\mathbb{P}\Big(T_{\mathbb{W}_0,i,a_{i,c}}^{\mathbb{W}_e} > t\Big) = \\
&= \sum_{k=0}^{\infty}\Bigg[\sum_{\sum_{r\in\mathbb{W}_0} k_r = k}\Big[\mathbb{P}\Big(I_j^{\mathbb{W}_e}(t) = k_j,\ j \in \mathbb{W}_0 \mid \boldsymbol{Z}(0) = \boldsymbol{\delta}_{a_{i,c}}^i\Big) \cdot \prod_{m\in\mathbb{W}_0} q_m^{k_m}\Big]\Bigg] \\
&= h_{i,a_{i,c}}^{\mathbb{W}_e}(t; \boldsymbol{q}_{\mathbb{W}_0}).
\end{aligned}$$

The result for $\boldsymbol{\alpha}^*$ follows from the assumption of independent evolution. The proofs for Property (ii) and Property (iii) are analogous to the proofs of the corresponding properties in Theorem 2.6. $\square$

We illustrate the behavior of $\mathbb{P}\left(T^{\mathbb{W}_e}_{\mathbb{W}_0,i} > t\right)$ and $\mathbb{P}\left(T^{\mathbb{W}_e}_{\mathbb{W}_0,i,a} > t\right)$ in Figure 2.22 and Figure 2.23.

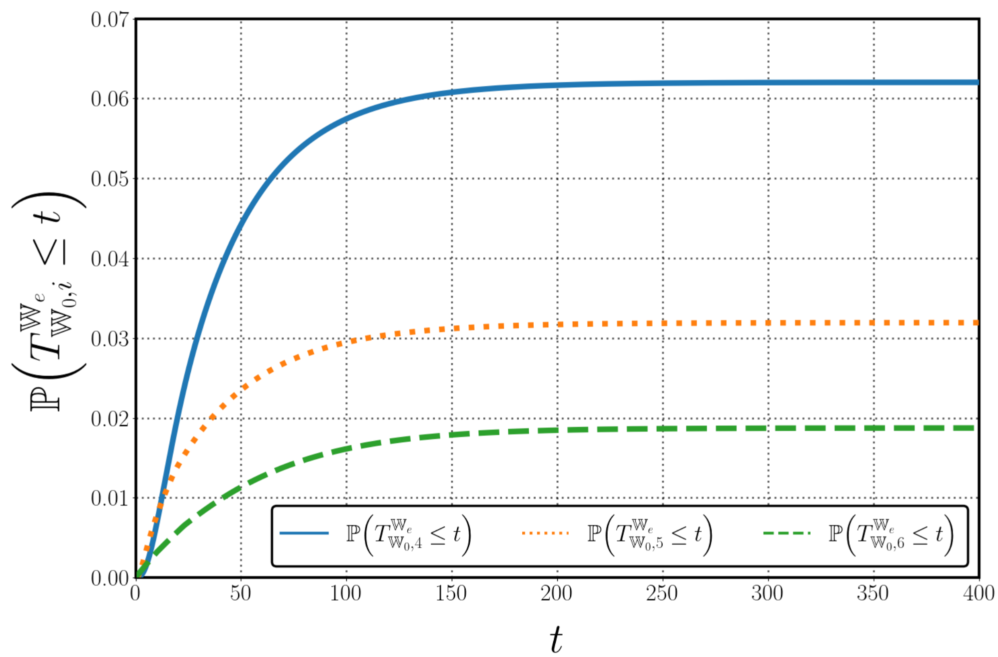

Figure 2.22: An application of Proposition 2.2 - calculations ($h = 10^{-2}$) for $\boldsymbol{s} = \boldsymbol{q}_{\mathbb{W}_0}$ for the example MSBPM (Table 2.11, Table 2.6) with mutation scheme "$\mathbb{W}_e$ towards $\mathbb{W}_0$" (Table 2.8) starting with one particle of age 0. Recall from Figure 2.13 that $\boldsymbol{q} = (0.77206078, 0.94263439, 0.92948039, 0.93797045, 0.96806635, 0.98125857)^\top$, consequently $\boldsymbol{q}_{\mathbb{W}_0} = (0.77206078, 0.94263439, 0.92948039, 1.0, 1.0, 1.0)^\top$. $P\left(T^{\mathbb{W}_e}_{\mathbb{W},i} \leq t\right) = 1 - h^{\mathbb{W}_e}_i(t; \boldsymbol{q}_{\mathbb{W}_0})$, we can reuse the calculations done for Figure 2.17. At $t = 1500$, we have $P\left(T^{\mathbb{W}_e}_{\mathbb{W}_0,4} \leq t\right) = 0.062024108$, $P\left(T^{\mathbb{W}_e}_{\mathbb{W}_0,5} \leq t\right) = 0.0319310927$, $P\left(T^{\mathbb{W}_e}_{\mathbb{W}_0,6} \leq t\right) = 0.0187387245$. As $t = 1500$ is sufficiently large as to conclude that $h^{\mathbb{W}_e}_i(1500; \boldsymbol{q}_{\mathbb{W}_0}) = h^{\mathbb{W}_e}_i(\boldsymbol{q}_{\mathbb{W}_0})$, we also have $P\left(T^{\mathbb{W}_e}_{\mathbb{W}_0,4} = \infty\right) = h^{\mathbb{W}_e}_4(\boldsymbol{q}_{\mathbb{W}_0}) = 0.93797589$, $P\left(T^{\mathbb{W}_e}_{\mathbb{W}_0,5} = \infty\right) = h^{\mathbb{W}_e}_5(\boldsymbol{q}_{\mathbb{W}_0}) = 0.96806891$, $P\left(T^{\mathbb{W}_e}_{\mathbb{W}_0,6} = \infty\right) = h^{\mathbb{W}_e}_6(\boldsymbol{q}_{\mathbb{W}_0}) = 0.98126128$.

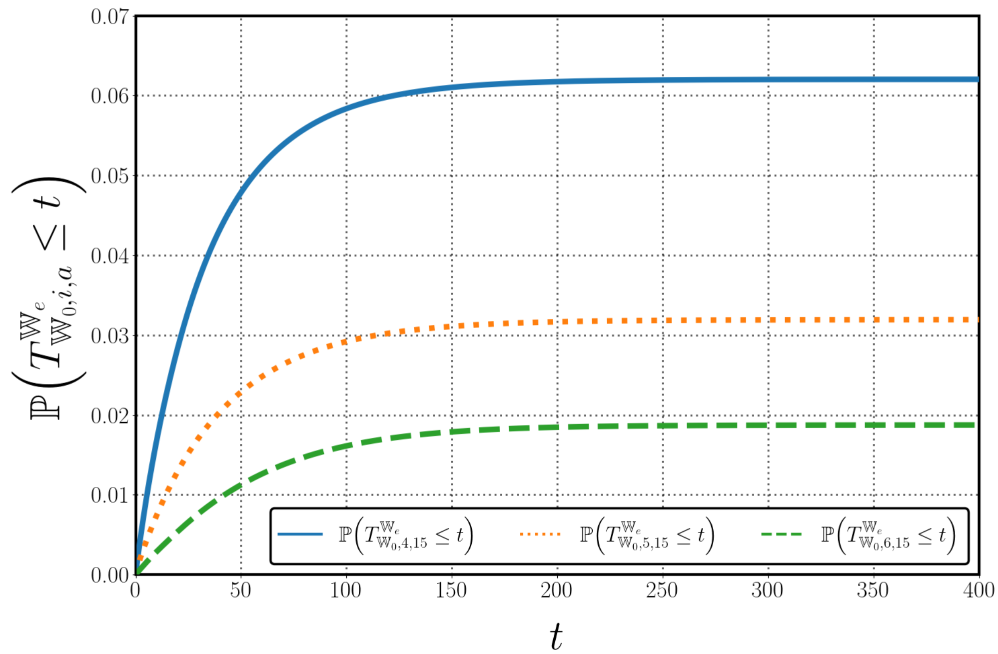

Figure 2.23: An application of Proposition 2.3 - calculations ($h = 10^{-2}$) for $\boldsymbol{s} = \boldsymbol{q}_{\mathbb{W}_0}$ for the example MSBPM (Table 2.11, Table 2.6) with mutation scheme "$\mathbb{W}_e$ towards $\mathbb{W}_0$" (Table 2.8) starting with one particle of age $a = 15$. Recall from Figure 2.13 that $\boldsymbol{q} = (0.77206078, 0.94263439, 0.92948039, 0.93797045, 0.96806635, 0.98125857)^\top$, consequently $\boldsymbol{q}_{\mathbb{W}_0} = (0.77206078, 0.94263439, 0.92948039, 1.0, 1.0, 1.0)^\top$. $P\left(T^{\mathbb{W}_e}_{\mathbb{W}_0,i,15} \leq t\right) = 1 - h^{\mathbb{W}_e}_{i,15}(t; \boldsymbol{q}_{\mathbb{W}_0})$, we can reuse the calculations done for Figure 2.18. At $t = 1500$, we have $P\left(T^{\mathbb{W}_e}_{\mathbb{W}_0,4,15} \leq t\right) = 0.062024108$, $P\left(T^{\mathbb{W}_e}_{\mathbb{W}_0,5,15} \leq t\right) = 0.0319310922$, $P\left(T^{\mathbb{W}_e}_{\mathbb{W}_0,6} \leq t\right) = 0.0187387245$. As $t = 1500$ is sufficiently large as to conclude that $h^{\mathbb{W}_e}_{i,15}(1500; \boldsymbol{q}_{\mathbb{W}_0}) = h^{\mathbb{W}_e}_{i,15}(\boldsymbol{q}_{\mathbb{W}_0})$, we also have $P\left(T^{\mathbb{W}_e}_{\mathbb{W}_0,4,15} = \infty\right) = h^{\mathbb{W}_e}_{4,15}(\boldsymbol{q}_{\mathbb{W}_0}) = 0.93797589$, $P\left(T^{\mathbb{W}_e}_{\mathbb{W}_0,5,15} = \infty\right) = h^{\mathbb{W}_e}_{5,15}(\boldsymbol{q}_{\mathbb{W}_0}) = 0.96806891$, $P\left(T^{\mathbb{W}_e}_{\mathbb{W}_0,6,15} = \infty\right) = h^{\mathbb{W}_e}_{6,15}(\boldsymbol{q}_{\mathbb{W}_0}) = 0.98126128$.

2.3.1.6 Immediate risk of producing a "successful" mutant from $\mathbb{W}_e$ towards $\mathbb{W}_0$ within the DMSBPM

Recall Definition 2.10 from Subsection 2.2.6, which is also valid for the case of the DMSBPM. Due to our interest in $\mathbb{W}_0$, we specify

Definition 2.16. *Define for an initial particle of type i, $i \in \mathbb{W}_e$, the following hazard function:*

1. *If the initial particle is of age* 0

$$g^{\mathbb{W}_e}_{\mathbb{W}_0,i}(t)dt = \mathbb{P}\Big(T^{\mathbb{W}_e}_{\mathbb{W}_0,i} \in (t, t+dt] \mid T^{\mathbb{W}_e}_{\mathbb{W}_0,i} > t\Big). \tag{2.47}$$

2. *If the initial particle if of age a, $a \neq 0$*

$$g^{\mathbb{W}_e}_{\mathbb{W}_0,i,a}(t)dt = \mathbb{P}\Big(T^{\mathbb{W}_e}_{\mathbb{W}_0,i,a} \in (t, t+dt] \mid T^{\mathbb{W}_e}_{\mathbb{W}_0,i,a} > t\Big). \tag{2.48}$$

It is clear that equations (2.28) and (2.29), from Subsection 2.2.6, hold with $\mathbb{W}$ being substituted with $\mathbb{W}_0$.

If there are no particles from types from $\mathbb{W}_e$ left in the population, the probability of occurrence of a "successful" mutant from $\mathbb{W}_e$ towards $\mathbb{W}_0$ is 0. Therefore, in addition to Definition 2.16 above, we can also investigate the following modification of the standard formulation of the hazard function:

Definition 2.17. *Define for an initial particle of type i, $i \in \mathbb{W}_e$, the following modified hazard function:*

1. *If the initial particle is of age* 0

$$\hat{g}^{\mathbb{W}_e}_{\mathbb{W}_0,i}(t)dt = \mathbb{P}\Big(T^{\mathbb{W}_e}_{\mathbb{W}_0,i} \in (t, t+dt] \mid T^{\mathbb{W}_e}_{\mathbb{W}_0,i} > t,\ \sum_{c \in \mathbb{W}_e} Z_c(t) > 0\Big). \tag{2.49}$$

2. *If the initial particle if of age a, $a \neq 0$*

$$\hat{g}^{\mathbb{W}_e}_{\mathbb{W}_0,i,a}(t)dt = \mathbb{P}\Big(T^{\mathbb{W}_e}_{\mathbb{W}_0,i,a} \in (t, t+dt] \mid T^{\mathbb{W}_e}_{\mathbb{W}_0,i,a} > t,\ \sum_{c \in \mathbb{W}_e} Z_c(t) > 0\Big). \tag{2.50}$$

In other words, we will consider the probability of the first "successful" mutant occurring immediately after time t, under the additional condition that at time t the population has at least one particle from a type from $\mathbb{W}_e$.

Let us inspect $\hat{g}^{\mathbb{W}_e}_{\mathbb{W}_0,i}(t)dt$, $i \in \mathbb{W}_e$:

$$\begin{aligned}\hat{g}^{\mathbb{W}_e}_{\mathbb{W}_0,i}(t)dt &= \frac{\mathbb{P}\Big(T^{\mathbb{W}_e}_{\mathbb{W}_0,i} \in (t, t+dt],\, T^{\mathbb{W}_e}_{\mathbb{W}_0,i} > t,\, \sum_{c\in\mathbb{W}_e} Z_c(t) > 0\Big)}{\mathbb{P}\Big(T^{\mathbb{W}_e}_{\mathbb{W}_0,i} > t,\, \sum_{c\in\mathbb{W}_e} Z_c(t) > 0\Big)} \\ &= \frac{\mathbb{P}\Big(T^{\mathbb{W}_e}_{\mathbb{W}_0,i} \in (t, t+dt]\Big)}{\mathbb{P}\Big(T^{\mathbb{W}_e}_{\mathbb{W}_0,i} > t\Big) - \mathbb{P}\Big(T^{\mathbb{W}_e}_{\mathbb{W}_0,i} > t,\, \sum_{c\in\mathbb{W}_e} Z_c(t) = 0\Big)},\end{aligned}$$

which can be rewritten as

$$\hat{g}^{\mathbb{W}_e}_{\mathbb{W}_0,i}(t) = \frac{F^{(1)}_{T^{\mathbb{W}_e}_{\mathbb{W}_0,i}}(t)}{\mathbb{P}\Big(T^{\mathbb{W}_e}_{\mathbb{W}_0,i} > t\Big) - \mathbb{P}\Big(T^{\mathbb{W}_e}_{\mathbb{W}_0,i} > t,\, \sum_{c\in\mathbb{W}_e} Z_c(t) = 0\Big)}, \tag{2.51}$$

where $F^{(1)}_{T^{\mathbb{W}_e}_{\mathbb{W}_0,i}}(t)$ is the probability density function of $T^{\mathbb{W}_e}_{\mathbb{W}_0,i}$. We can find the c.d.f. of $T^{\mathbb{W}_e}_{\mathbb{W}_0,i}$ via Proposition 2.2 and then approximate $F^{(1)}_{T^{\mathbb{W}_e}_{\mathbb{W}_0,i}}(t)$, for example, with a forward difference.

It is evident that the same line of thought outlined above, applied for a starting particle of age a, $a \neq 0$, leads us to

$$\hat{g}^{\mathbb{W}_e}_{\mathbb{W}_0,i,a}(t) = \frac{F^{(1)}_{T^{\mathbb{W}_e}_{\mathbb{W}_0,i,a}}(t)}{\mathbb{P}\Big(T^{\mathbb{W}_e}_{\mathbb{W}_0,i,a} > t\Big) - \mathbb{P}\Big(T^{\mathbb{W}_e}_{\mathbb{W}_0,i,a} > t,\, \sum_{c\in\mathbb{W}_e} Z_c(t) = 0\Big)}. \tag{2.52}$$

To simplify notations we introduce:

Definition 2.18. *For $i \in \mathbb{W}_e$, denote*

$$V_i(t) = \mathbb{P}\Big(T^{\mathbb{W}_e}_{\mathbb{W}_0,i} > t,\, \sum_{c\in\mathbb{W}_e} Z_c(t) = 0\Big),$$

$$V_{i,a}(t) = \mathbb{P}\Big(T^{\mathbb{W}_e}_{\mathbb{W}_0,i,a} > t,\, \sum_{c\in\mathbb{W}_e} Z_c(t) = 0\Big).$$

Theorem 2.8. *The probability $V_i(t)$ of the event that jointly the first "successful" mutant does not occur before or at t and there are no particles from $\mathbb{W}_e$ left at t, for a DMSBPM starting with a particle of age 0, satisfies the following system of integral equations:*

(2.53)

$$V_i(t) = \int_0^t f_i\Bigg(y; \Big[\sum_{m\in\mathbb{W}_e} u_{im}V_m(t-y)\Big] + \Big[\sum_{r\in\mathbb{W}_0} u_{ir}q_r\Big]\Bigg) dG_i(y), \quad i \in \mathbb{W}_e.$$

Proof. Again, we condition with respect to the moment of death of the initial particle:

$$\begin{aligned}
V_i(t) &= \mathbb{P}\Big(T^{\mathbb{W}_e}_{\mathbb{W}_0,i} > t,\ \sum_{c\in\mathbb{W}_e} Z_c(t) = 0\Big)\\
&= \int_0^t \sum_{k=0}^{\infty} p_{ik}(y) \sum_{\sum_{l\in\mathbb{W}} k_l = k} \Bigg[\frac{k!}{\prod_{v\in\mathbb{W}} k_v!} \prod_{r\in\mathbb{W}} u_{ir}^{k_r} \cdot \\
&\qquad \cdot \prod_{r\in\mathbb{W}} \Big[\mathbb{P}\Big(\widetilde{T}^{\mathbb{W}_e}_{\mathbb{W}_0,r} > t-y,\ \sum_{c\in\mathbb{W}_e} \widetilde{Z}_c(t-y) = 0\Big)\Big]^{k_r}\Bigg] dG_i(y)\\
&= \int_0^t \sum_{k=0}^{\infty} p_{ik}(y) \sum_{\sum_{l\in\mathbb{W}} k_l = k} \Bigg[\frac{k!}{\prod_{v\in\mathbb{W}} k_v!} \cdot \\
&\qquad \cdot \prod_{m\in\mathbb{W}_e} \Big[u_{im}\mathbb{P}\Big(\widetilde{T}^{\mathbb{W}_e}_{\mathbb{W}_0,m} > t-y,\ \sum_{c\in\mathbb{W}_e} \widetilde{Z}_c(t-y) = 0\Big)\Big]^{k_m} \cdot \\
&\qquad \cdot \prod_{r\in\mathbb{W}_0} \Big[u_{ir}q_r\Big]^{k_r}\Bigg] dG_i(y)\\
&= \int_0^t \sum_{k=0}^{\infty} p_{ik}(y) \sum_{\sum_{l\in\mathbb{W}} k_l = k} \Bigg[\frac{k!}{\prod_{v\in\mathbb{W}} k_v!} \cdot \\
&\qquad \cdot \prod_{m\in\mathbb{W}_e} \Big[u_{im}V_m(t-y)\Big]^{k_m} \cdot \prod_{r\in\mathbb{W}_0} \Big[u_{ir}q_r\Big]^{k_r}\Bigg] dG_i(y)\\
&= \int_0^t \sum_{k=0}^{\infty} p_{ik}(y) \Bigg[\Big[\sum_{m\in\mathbb{W}_e} u_{im}V_m(t-y)\Big] + \Big[\sum_{r\in\mathbb{W}_0} u_{ir}q_r\Big]\Bigg]^k dG_i(y)
\end{aligned}$$

$$= \int_0^t f_i\Bigg(y; \Big[\sum_{m\in\mathbb{W}_e} u_{im}V_m(t-y)\Big] + \Big[\sum_{r\in\mathbb{W}_0} u_{ir}q_r\Big]\Bigg)dG_i(y).$$

□

Corollary 2.22. *The probability $V_{i,a}(t)$ of the event that jointly the first "successful" mutant does not occur before or at t and there are no particles from $\mathbb{W}_e$ left at t, for a DMSBPM starting with a particle of age a, $a \neq 0$, satisfies the following system of integral equations:*

(2.54)

$$V_{i,a}(t) = \int_0^t f_i\Bigg(a+y; \Big[\sum_{m\in\mathbb{W}_e} u_{im}V_m(t-y)\Big]+\Big[\sum_{r\in\mathbb{W}_0} u_{ir}q_r\Big]\Bigg)dG_{i,a}(y), \quad i \in \mathbb{W}_e.$$

Proof. The proof is completely analogous to the proof of Theorem 2.8. □

We report that, unfortunately, we experienced precision-related numerical difficulties, when trying to calculate $\hat{g}^{\mathbb{W}_e}_{\mathbb{W}_0,i}(t)$ and $\hat{g}^{\mathbb{W}_e}_{\mathbb{W}_0,i,a}(t)$ for our experimental setups from Subsection 2.2.8, hence we cannot provide illustrative figures at this point. This is not very surprising considering the small values already observed in Figure 2.10 from Subsection 2.2.6. However, we note that in "simpler", or more fine-tuned, setups, such as the setup of two types within [2], $\hat{g}^{\mathbb{W}_e}_{\mathbb{W}_0,i}(t)$ can be computed less problematically. The elimination of the numerical difficulties encountered is a to be considered in future research.

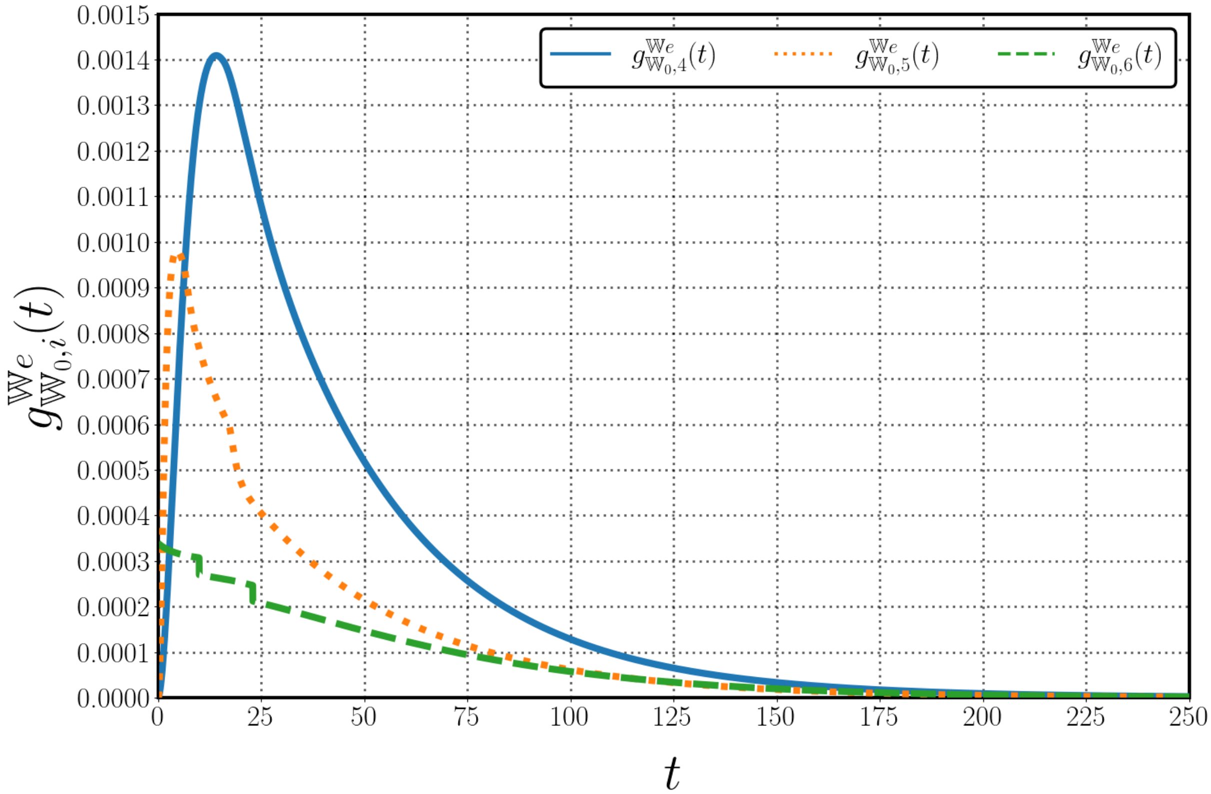

Figure 2.24: Calculations for $g_{\mathbb{W}_0,i}^{\mathbb{W}_e}(t)$, with $h = 10^{-2}$, for the example DMSBPM (Table 2.11, Table 2.6) with mutation scheme "$\mathbb{W}_e$ towards $\mathbb{W}_0$" (Table 2.8) starting with one particle of age 0. At $t = 1000$, we have $g_{\mathbb{W}_0,4}^{\mathbb{W}_e}(1000) = 6.32378172e - 12$, $g_{\mathbb{W}_0,5}^{\mathbb{W}_e}(1000) = 1.18904764e - 11$, $g_{\mathbb{W}_0,6}^{\mathbb{W}_e}(1000) = 3.11698760e - 12$.

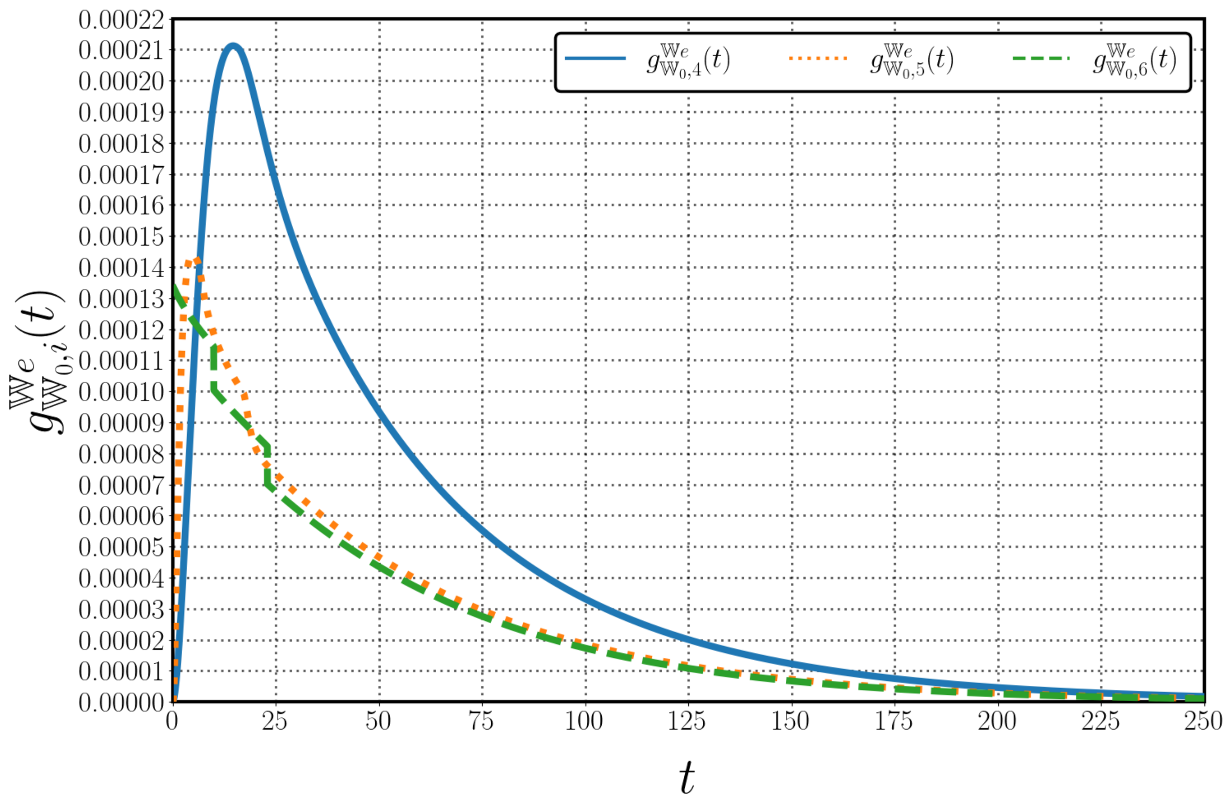

Figure 2.25: Calculations for $g^{\mathbb{W}_e}_{\mathbb{W}_0,i}(t)$, with $h = 10^{-2}$, for the example DMSBPM (Table 2.11, Table 2.6) with mutation scheme "$\mathbb{W}_e$ towards $\mathbb{W}_0, \mathbb{W}_0$ forms a chain" (Table 2.9) starting with one particle of age 0. At $t = 1000$, we have $g^{\mathbb{W}_e}_{\mathbb{W}_0,4}(1000) = 9.49276919e - 12$, $g^{\mathbb{W}_e}_{\mathbb{W}_0,5}(1000) = 3.60108692e - 11$, $g^{\mathbb{W}_e}_{\mathbb{W}_0,6}(1000) = 4.94432496e - 12$.

2.3.2 Decomposable Multi-type Bellman-Harris Branching Process through probabilities of Mutation between types (DMBHBPM)

Within this Subsection, we introduce the Decomposable Multi-type Bellman-Harris Branching Process through probabilities of Mutation between types (DMBHBPM). As the DMBHBPM is a particular case of the DMSBPM, discussed in Subsection 2.3.1, we limit ourselves to stating the process definition and writing without detailed proof all statements as corollaries of previous results. We note that the DMBHBPM presented here is a generalized version of the multi-type Bellman-Harris branching process (multi-type BHBP) discussed in our previous work Slavtchova-Bojkova & Vitanov [5] (2019). Within [5] the class $\mathbb{W}_0$ consists of only one type. A further extension, with respect to [5], are the results for the case of a process starting with a particle of age a, $a \neq 0$.

2.3.2.1 Definition of the DMBHBPM

Definition 2.19. *Define the Decomposable Multi-type Bellman-Harris Branching Process through probabilities of Mutation between types (DMB-HBPM) as the multi-type branching process satisfying:*

1. *Each particle type is uniquely associated with an integer from $\mathbb{W}$ and conforms to:*
   (a) *The lifespan of particles of type i, $i \in \mathbb{W}$, is modeled by a (continuous) r.v. τ_i. The corresponding cumulative distribution function (c.d.f.) is denoted by $G_i(t) = \mathbb{P}(\tau_i \leq t)$, also $G_i(0^+) = 0$.*
   (b) *The number of particles in the offspring of a type i, $i \in \mathbb{W}$, particle is modeled by a (discrete) r.v. ν_i. We denote with p_{ik} the probability that a type i particle has k, $k \in \mathbb{N}_0$, offspring particles (regardless of their type). Thus, ν_i is specified by given $\{p_{ik}\}_{k=0}^{\infty}$, $\sum_{k=0}^{\infty} p_{ik} = 1$. We denote the corresponding p.g.f. of ν_i with $f_i(s) = \mathbb{E}\left[s^{\nu_i}\right] = \sum_{k=0}^{\infty} p_{ik} s^k$, $|s| \leq 1$.*
2. *Each daughter particle of a type i particle, $i \in \mathbb{W}_e$, can be of any type $j \in \mathbb{W}$, however, daughter particles of type i particles, $i \in \mathbb{W}_0$, can only be of type $j \in \mathbb{W}_0$. The type of a daughter particle is determined at birth. If $i \neq j$ we say that a "mutation" occurs. The probability that a daughter particle of a type i particle is a type j particle is denoted by u_{ij}, $u_{ij} \geq 0$, $\sum_{j=1}^{n} u_{ij} = 1$. Further:*
   (a) *If type i cannot have daughters of type j we consider the corresponding u_{ij} as $u_{ij} = 0$.*
   (b) *Particles are not allowed to change their type within their lifespan.*
3. *All particles from all particle types evolve independently from one another, irrespective of generation.*

4. *Formally* $\left\{\boldsymbol{Z}(t) = \big(Z_1(t), Z_2(t), \ldots, Z_n(t)\big)^\top\right\}_{t\geq 0}$, *where* $\boldsymbol{Z}(t)$ *stands for the DMBHBPM at* t *and* $Z_i(t)$ *is the number of particles of type* i *that exist at* t.

It is clear from Definition 2.19 that the DMBHBPM is obtained by removing the dependence on particle age from $v_i(a)$, $p_{ik}(a)$, and $f_i(a;s)$, within the DMSBPM. As all results from Section 2.2 and Subsection 2.3.1 are proven for the more general case where dependence on particle age is included and also the steps of all proofs remain the same if this dependence is removed, it is straightforward to obtain corresponding results for the DMBHBPM. We note that within the DMBHBPM, as in the DMSBPM, particle types are divided into two classes $\mathbb{W}_0$ and $\mathbb{W}_e$ for which $\mathbb{W}_0 \cap \mathbb{W}_e = \emptyset$ and $\mathbb{W}_0 \cup \mathbb{W}_e = \mathbb{W}$. Particles with types from $\mathbb{W}_e$ can produce particles with types from $\mathbb{W}$ and particles with types from $\mathbb{W}_0$ can only produce particles with types form $\mathbb{W}_0$. Figure 2.11 and Figure 2.12 can also be interpreted as depicting particular cases of the DMBHBPM.

2.3.2.2 Results for the DMBHBPM

All definitions from Subsection 2.3.1 are valid for the DMBHBPM.

Corollary 2.23. *For the DMBHBPM, the following system of integral equations holds:*

1. *For* $i \in \mathbb{W}_e$

$$F_i(t;\boldsymbol{s}) = s_i\big(1-G_i(t)\big) + \int_0^t f_i\Bigg(\Big[\sum_{m\in\mathbb{W}_e} u_{im}F_m(t-y;\boldsymbol{s})\Big] + \\ + \Big[\sum_{r\in\mathbb{W}_0} u_{ir}F_r(t-y;\boldsymbol{s}_{\mathbb{W}_0})\Big]\Bigg)dG_i(y).$$

2. *For* $i \in \mathbb{W}_0$

$$F_i(t;\boldsymbol{s}_{\mathbb{W}_0}) = s_i\big(1-G_i(t)\big) + \int_0^t f_i\Bigg(\sum_{r\in\mathbb{W}_0} u_{ir}F_r(t-y;\boldsymbol{s}_{\mathbb{W}_0})\Bigg)dG_i(y).$$

Proof. Having the result of Corollary 2.12, we remove age dependence within f_i. $\square$

Note that in the next corollary, we can drop age dependence within f_i, however, the age of the initial particle still affects the remaining lifespan distribution of the initial particle.

Corollary 2.24. *For the DMBHBPM, the following system of integral equations holds:*

1. *For* $i \in \mathbb{W}_e$

$$F_{i,a}(t;\boldsymbol{s}) = s_i\big(1-G_{i,a}(t)\big) + \int_0^t f_i\Bigg(\Big[\sum_{m\in\mathbb{W}_e} u_{im}F_m(t-y;\boldsymbol{s})\Big]+ \\ +\Big[\sum_{r\in\mathbb{W}_0} u_{ir}F_r(t-y;\boldsymbol{s}_{\mathbb{W}_0})\Big]\Bigg)dG_{i,a}(y).$$

2. *For* $i \in \mathbb{W}_0$

$$F_{i,a}(t;\boldsymbol{s}_{\mathbb{W}_0}) = s_i\big(1-G_{i,a}(t)\big)+\int_0^t f_i\Bigg(\sum_{r\in\mathbb{W}_0} u_{ir}F_r(t-y;\boldsymbol{s}_{\mathbb{W}_0})\Bigg)dG_{i,a}(y).$$

Proof. Having the result of Corollary 2.13, we remove age dependence within f_i. □

Corollary 2.25. *The following system of integral equations holds for the DMBHBPM:*

1. *For* $i \in \mathbb{W}_e$

$$q_i(t) = \int_0^t f_i\Bigg(\Big[\sum_{m\in\mathbb{W}_e} u_{im}q_m(t-y)\Big] + \Big[\sum_{r\in\mathbb{W}_0} u_{ir}q_r(t-y)\Big]\Bigg)dG_i(y),$$

Which can be rewritten as

$$q_i(t) = \int_0^t f_i\Big(\sum_{r\in\mathbb{W}} u_{ir}q_r(t-y)\Big)dG_i(y).$$

2. *For* $i \in \mathbb{W}_0$

$$q_i(t) = \int_0^t f_i\Big(\sum_{r\in\mathbb{W}_0} u_{ir}q_r(t-y)\Big)dG_i(y).$$

Proof. Having the result of Corollary 2.14, we remove age dependence within f_i. □

Again, we can drop age dependence within f_i, however, the age of the initial particle still affects the remaining lifespan distribution of the initial particle.

Corollary 2.26. *The following system of integral equations holds for the DMBHBPM:*

1. *For* $i \in \mathbb{W}_e$

$$q_{i,a}(t) = \int_0^t f_i\Bigg(\Big[\sum_{m\in\mathbb{W}_e} u_{im}q_m(t-y)\Big] + \Big[\sum_{r\in\mathbb{W}_0} u_{ir}q_r(t-y)\Big]\Bigg)dG_{i,a}(y),$$

Which can be rewritten as

$$q_{i,a}(t) = \int_0^t f_i\Big(\sum_{r\in\mathbb{W}} u_{ir}q_r(t-y)\Big)dG_{i,a}(y).$$

2. *For* $i \in \mathbb{W}_0$

$$q_{i,a}(t) = \int_0^t f_i\Big(\sum_{r\in\mathbb{W}_0} u_{ir}q_r(t-y)\Big)dG_{i,a}(y). \tag{2.55}$$

Proof. Having the result of Corollary 2.15, we remove age dependence within f_i. □

Following Corollary 2.4, valid for the MSBPM, we can merge Corollary 2.16 and Corollary 2.17, valid for the DMSBPM, into:

Corollary 2.27. *The following system of equations holds for the DMB-HBPM:*

1. *For* $i \in \mathbb{W}_e$

$$q_i = q_{i,a} = f_i\Big(\sum_{r\in\mathbb{W}} u_{ir}q_r\Big).$$

2. *For* $i \in \mathbb{W}_0$

$$q_i = q_{i,a} = f_i\Big(\sum_{r\in\mathbb{W}_0} u_{ir}q_r\Big).$$

Proof. We apply Corollary 2.4 onto Corollary 2.16 and Corollary 2.17, by dropping age dependence within f_i. □

Corollary 2.28. *For the DMBHBPM the following system of integral equations holds:*

1. *For* $i \in \mathbb{W}_e$

(2.56)
$$h_i^{\mathbb{W}_e}(t;\boldsymbol{s}_{\mathbb{W}_0}) = \big(1 - G_i(t)\big) + \int_0^t f_i\Bigg(\Big[\sum_{m\in\mathbb{W}_e} u_{im} h_m^{\mathbb{W}_e}(t-y;\boldsymbol{s}_{\mathbb{W}_0})\Big] + \\ + \Big[\sum_{r\in\mathbb{W}_0} u_{ir} s_r\Big]\Bigg) dG_i(y).$$

2. *For* $i \in \mathbb{W}_0$
$$h_i^{\mathbb{W}_e}(t;\boldsymbol{s}_{\mathbb{W}_0}) = 1.$$

Proof. Having the result of Corollary 2.18, we remove age dependence within f_i. □

Once again, we can drop age dependence within f_i, however, the age of the initial particle still affects the remaining lifespan distribution of the initial particle.

Corollary 2.29. *For the DMBHBPM the following system of integral equations holds:*

1. *For* $i \in \mathbb{W}_e$
$$h_{i,a}^{\mathbb{W}_e}(t;\boldsymbol{s}_{\mathbb{W}_0}) = \big(1 - G_{i,a}(t)\big) + \int_0^t f_i\Bigg(\Big[\sum_{m\in\mathbb{W}_e} u_{im} h_m^{\mathbb{W}_e}(t-y;\boldsymbol{s}_{\mathbb{W}_0})\Big] + \\ + \Big[\sum_{r\in\mathbb{W}_0} u_{ir} s_r\Big]\Bigg) dG_{i,a}(y).$$

2. *For* $i \in \mathbb{W}_0$
$$h_{i,a}^{\mathbb{W}_e}(t;\boldsymbol{s}_{\mathbb{W}_0}) = 1.$$

Proof. Having the result of Corollary 2.19, we remove age dependence within f_i. □

Corollary 2.30. *The following system of equations holds within the DMBHBPM:*

1. *For* $i \in \mathbb{W}_e$

$$h_i^{\mathbb{W}_e}(\boldsymbol{s}_{\mathbb{W}_0}) = f_i\left(\Big[\sum_{m\in\mathbb{W}_e} u_{im} h_m^{\mathbb{W}_e}(\boldsymbol{s}_{\mathbb{W}_0})\Big] + \Big[\sum_{r\in\mathbb{W}_0} u_{ir} s_r\Big]\right).$$

2. *For* $i \notin \mathbb{W}_e$

$$h_i^{\mathbb{W}_e}(\boldsymbol{s}_{\mathbb{W}_0}) = 1.$$

Proof. Having Corollary 2.20, we remove age dependence within f_i. □

Corollary 2.31. *For the DMBHBPM the following system of integral equations holds:*

1. *For* $i \in \mathbb{W}_e$

$$h_{i,a}^{\mathbb{W}_e}(\boldsymbol{s}_{\mathbb{W}_0}) = f_i\left(\Big[\sum_{m\in\mathbb{W}_e} u_{im} h_m^{\mathbb{W}_e}(\boldsymbol{s}_{\mathbb{W}_0})\Big] + \Big[\sum_{r\in\mathbb{W}_0} u_{ir} s_r\Big]\right).$$

2. *For* $i \in \mathbb{W}_0$

$$h_{i,a}^{\mathbb{W}_e}(\boldsymbol{s}_{\mathbb{W}_0}) = 1.$$

Proof. Having the result of Corollary 2.21, we remove age dependence within f_i. □

Corollary 2.32. *Let each particle type from* $\mathbb{W}_e$*, within the DMBHBPM, be either subcritical or critical. Then for* $i \in \mathbb{W}_e$

$$q_i = h_i^{\mathbb{W}_e}(\boldsymbol{q}_{\mathbb{W}_0}) = f_i\left(\Big[\sum_{m\in\mathbb{W}_e} u_{im} h_m^{\mathbb{W}_e}(\boldsymbol{q}_{\mathbb{W}_0})\Big] + \Big[\sum_{r\in\mathbb{W}_0} u_{ir} q_r\Big]\right)$$

and

$$q_{i,a} = h_{i,a}^{\mathbb{W}_e}(\boldsymbol{q}_{\mathbb{W}_0}) = f_i\left(\Big[\sum_{m\in\mathbb{W}_e} u_{im} h_m^{\mathbb{W}_e}(\boldsymbol{q}_{\mathbb{W}_0})\Big] + \Big[\sum_{r\in\mathbb{W}_0} u_{ir} q_r\Big]\right).$$

Proof. Having the result of Proposition 2.1, we remove age dependence within f_i. □

Proposition 2.2 and Proposition 2.3 for the time until occurrence of the first "successful" mutant from $\mathbb{W}_e$ towards $\mathbb{W}_0$, $T_{\mathbb{W}_0}^{\mathbb{W}_e}$, can be applied directly for the case of the DMBHBPM.

Equation (2.51) and equation (2.52) for the modified hazard functions $\hat{g}_{\mathbb{W}_0,i}^{\mathbb{W}_e}(t)$ and $\hat{g}_{\mathbb{W}_0,i,a}^{\mathbb{W}_e}(t)$ respectively, are also valid for the DMBHBPM. We also have

Corollary 2.33. *The probability of the event that jointly the first "successful" mutant does not occur before or at* t *and there are no particles from* $\mathbb{W}_e$ *left at* t*, for a DMBHBPM starting with a particle of age* 0*, satisfies the following system of integral equations:*

$$V_i(t) = \int_0^t f_i\Bigg(\Big[\sum_{m\in\mathbb{W}_e} u_{im}V_m(t-y)\Big] + \Big[\sum_{r\in\mathbb{W}_0} u_{ir}q_r\Big]\Bigg)dG_i(y), \quad i \in \mathbb{W}_e.$$

Proof. Having the result of Theorem 2.8, we remove age dependence within f_i. □

And lastly, as the age of the initial particle affects its remaining lifespan,

Corollary 2.34. *The probability of the event that jointly the first "successful" mutant does not occur before or at* t *and there are no particles from* $\mathbb{W}_e$ *left at* t*, for a DMBHBPM starting with a particle of age* a*,* $a \neq 0$*, satisfies the following system of integral equations:*

$$V_{i,a}(t) = \int_0^t f_i\Bigg(\Big[\sum_{m\in\mathbb{W}_e} u_{im}V_m(t-y)\Big] + \Big[\sum_{r\in\mathbb{W}_0} u_{ir}q_r\Big]\Bigg)dG_{i,a}(y), \quad i \in \mathbb{W}_e.$$

Proof. Having the result of Corollary 2.22, we remove age dependence within f_i. □

CHAPTER 3

Sequential decision problems with branching process based dynamics

3.1 Chapter overview and organization

In this Chapter, we incorporate branching processes into optimization problems known as Sequential Decision Problems (SDPs). For our modeling of SDPs, we use Warren B. Powell's "Universal Modeling Framework" developed in [82] (2022). Our motivation for this choice of framework can be summarized as follows: 1) The "Universal Modeling Framework" is an attempt to unify the 15 communities discussed in the Introduction (recall Section 1.5, pages 22 - 24). This may prove beneficial in future research, where we may consider complicating the branching process based SDPs that we investigate; 2) The "Universal Modeling Framework" is straightforwardly connected to Approximate Dynamic Programming (ADP; see [74], [76], [78]) and Reinforcement Learning (RL; see [203], [82]). ADP and RL rely on simulations in order to produce solutions (of varying quality) for complex SDPs. We envision future research stemming from this dissertation, regarding SDPs with branching process based dynamics, as simulations based; 3) For our purposes, the "Universal Modeling Framework" is conceptually and notationally close to the discussions within the Markov decision processes community (see Puterman [70] (2005)). This is a good starting point for considering SDPs with Bienaymé-Galton-Watson (BGW) branching process dynamics as the BGW process is Markovian within standard definitions.

Our modeling of SDPs within this dissertation is heavily based on the ideas developed in [82] and [78] and as such shares the strengths and weaknesses of the “Universal Modeling Framework”. We note that, with respect to our purposes, we have made some minor contributions to the presentation in [82], those contributions being Proposition 3.1 and Proposition 3.2 from Subsection 3.2.7, as well as the inclusion of the discount factor γ in some equations and statements.

In Section 3.2 and Section 3.3, we present the “Universal Modeling Framework” by adapting parts of the presentation in [82]. In Section 3.4, Section 3.5, and Section 3.6, we obtain our novel results that incorporate certain branching processes into SDPs within the “Universal Modeling Framework”. The results from Section 3.4, Section 3.5, and Section 3.6, have not been published yet. In Section 3.7, we outline, but do not apply or investigate the properties of, a general ADP algorithm that can be used as a starting point for developing a specialized ADP algorithm for finding the solution of the SDP discussed within Section 3.6. We stress that in the dissertation, we do not consider stochastic differential equations within the optimization problems investigated.

The standard notion of “state” within the branching processes community postulates that the state of a branching process at t is to be understood as the number of particles, per type, that exist at t. With respect to a so defined “state” the BGW branching process as well as the Bellman-Harris branching process with exponential lifespans are Markovian. The scarce literature dedicated to combining specifically branching processes and SDPs, see [77], [199], [200], [201], [202], concentrates its efforts on branching processes that are Markovian under the standard notion of state. The papers listed, effectively, discuss multi-type Bienaymé-Galton-Watson branching processes. Our novel idea within Section 3.6 is to consider a novel definition of the state of the MSBPM (the MSBPM is generally non-Markovian under the standard notion of “state” since particle reproduction can depend on particle age). Under the newly defined “state”, we prove that the MSBPM is Markovian. This and the following considerations within Section 3.6 formally allow us to apply ADP and RL for the purpose of finding solutions of the corresponding SDP.

We note that [202] develops a model-free RL algorithm for a SDP with BGW branching process based dynamics. Contrary to [202], our agenda is to exploit a specified branching process model (such as the MSBPM) as much as possible. We note that while RL algorithms are usually model-free, ADP algorithms are usually model-based. This is why we outline a general ADP algorithm in Section 3.7 as an illustration of possible future research

- although we successfully incorporate the (generally non-Markovian) MSBPM into SDPs within the "Universal Modeling Framework", devising practical computational algorithms requires substantial further research. We hope that the computational tractability of the MSBPM, via Numerical Scheme 1 and Numerical Scheme 2, together with the theoretical foundation laid within the dissertation, will facilitate future success.

Within Section 3.4, we consider the paper of S. R. Pliska from 1976, [77]. [77] discusses a SDP with BGW branching process based dynamics and provides a theorem that allows us to efficiently obtain the solution of the (finite-horizon) SDP, described within Section 2 and Section 3 of the paper, via a Dynamic Programming algorithm (see, [67], [68], [69], [70]). Although [77] acknowledges that the algorithm obtained is a Dynamic Programming algorithm, the proof of Theorem 3.1. from [77] uses conditional expectations and does not use Bellman's optimality equation (see Section 3.3). Within Section 3.4, we recast the discussion in [77] into the more contemporary "Universal Modeling Framework" and provide a novel proof of Theorem 3.1 from [77] that is based on Bellman's optimality equation.

In Section 3.5, we consider the Multi-type Bellman-Harris Branching Process through probabilities of Mutation between types (MBHBPM; a special case of the MSBPM) with exponential lifespan distributions. The MBHBPM with exponential lifespan distributions is Markovian under the classical definition of state. Our novel contribution for this case is that we formally incorporate the process in SDPs within the "Universal Modeling Framework" and show that a result analogous to the Theorem 3.1. from [77] holds.

We stress that the algorithms obtained in Section 3.4 and Section 3.5, that allow for efficiently finding the solutions of the corresponding SDPs, discussed within these sections, have a limited scope of application. More precisely, these algorithms easily become non-applicable upon introducing further (appropriately modeled with respect to the "Universal Modeling Framework") dependencies within the discussed SDPs. However, for such complex SDPs, we can still consider the ADP and/or RL approach.

We note that the field of Controlled Branching Processes (CBPs; see [115], [124], [181], [182], [120], [121]) contains ideas that are close to the ideas found within the discussion of sequential decision problems. The relationship between CBPs and SDPs is that a CBP is a branching process and as such can be used as a model of uncertainty within a SDP.

This Chapter can also be viewed as a continuation of the efforts within

[1]-[7], as well as Chapter 2, to model cancer evolution and populations escaping extinction. Indeed, this context can benefit much from an introduction of SDPs as SDPs provide a way for planning appropriate actions in advance. This can be very beneficial, for example, within the case of administering cancer therapies as the different costs and expected results associated with different therapies have to be considered by the recipient and medical personnel. SDPs model the outcomes of the choices available to us, thus, depending on our objectives, they have the potential to become a useful tool for finding the best way available for forcing a population into extinction or for maximizing its chance of survival.

The Chapter is organized as follows. In Section 3.2, we introduce relevant concepts from the "Universal Modeling Framework" proposed by Warren B. Powell in [82]. In Section 3.3, we discuss Bellman's optimality equation. In Section 3.4, we recast the model from [77] into the "Universal Modeling Framework" and utilize Bellman's optimality equation to provide a novel proof for Theorem 3.1 from [77]. In Section 3.5, we consider the MBHBPM with exponential lifespan distributions, construct a corresponding SDP and prove a result similar to Theorem 3.1 from [77] for this case. In Section 3.6, we consider the MSBPM and construct a SDP with MSBPM based dynamics. In Section 3.7, we outline an ADP approach for solving the constructed SDP with MSBPM based dynamics. We finish with Section 3.8, where we give illustrative examples of SDPs with branching process based dynamics.

3.2 Modeling of Sequential Decision Problems (SDPs)

Within this Section, we introduce the "Universal Modeling Framework" developed by Warren B. Powell in [82] (2022) and [80]. We follow primarily Chapter 9 (page 467) from [82], however, we have streamlined the presentation in accordance with our purposes.

Recall the 15 mathematical communities, discussed in Section 1.5, that consider Sequential Decision Problems (SDPs). Each of these communities has its distinctive toolset and perspective when working with deterministic and/or stochastic optimization. The choice of community/communities within which to develop and associate our work is thus of paramount importance. The "Universal Modeling Framework" aims at covering the particularities encountered within all of the 15 communities that work with SDPs in a unified way. This may facilitate a more easy incorporation of ideas from the aforementioned communities into subsequent future research that continues the work within this dissertation. The framework (as well as the presentation in [82] and [78]) is also oriented towards the use of simulations and computer resources - we believe this approach towards solving complicated SDPs to hold great promise, considering that closed-form solutions are rare for SDPs.

The notation used within the "Universal Modeling Framework" is close to the notation used within the starting point of our considerations within Section 3.4, that is [77], and also is a refinement of the notation found in [78] from where we draw the idea of using ADP in future research. The "Universal Modeling Framework" is also notationally close to Optimal control ([204], [205], [206], [207], [208]) and Markov Decision Processes (MDP; [67], [68], [69], [70], [71]). We stress that working within the "Universal Modeling Framework" allows us to utilize Bellman's equality equation, discussed in Section 3.3, which formally opens the gate for techniques such as Approximate Dynamic Programming (ADP) and Reinforcement Learning (RL) to be applied onto SDPs with underlying branching process based dynamics.

Within the dissertation we do not engage into the full extent of the "Universal Modeling Framework", presented in [82]. As we concentrate the majority of our attention on problems close to MDP, we only provide concepts that we use and refer the reader to Chapter 9 in [82] for the entirety of the framework.

Following [82] (page 470), there are 5 components when modeling any sequential decision problem:

1. State variables
2. Decision/action/control variables
3. Exogenous information variables
4. Transition function
5. Objective function

Before we begin defining these components, recall our informal Description 1.1 of a SDP from Chapter 1 (page 22 of the dissertation). From now on, in order to avoid unnecessary notational clutter, we adopt:

Notational Choice 1. We index the decision epochs with t, $t = 0, 1, 2, \ldots$ Any variable indexed with t is understood as a variable corresponding to decision epoch with index t. When we talk about intervals, e.g. interval (t, t+1), we understand the interval between epochs with index t and index $t+1$. We assume that the distance between two decision epochs can vary between any two neighboring epochs but cannot be 0 or ∞. If there is a final epoch, its index is T.

Also:

Remark 3.1. *Below, we assume that any variable indexed with t is known at t. With respect to Notational Choice 1, the variable is known at the decision epoch indexed by t.*

We note that, due to the high level of interconnection between the aforementioned components, it is rather difficult to define them entirely separately from one other. In order to improve our presentation, we discuss the components in a different order than the one listed above. The concept of a "state variable" is central for all other concepts, however, it is the last one defined. We take this approach as the definition of a "state variable", provided within [82], actually involves the other four concepts. Until we arrive at Subsection 3.2.5, where we define them properly, let S_t denote the state at t and let $\mathcal{S}$ (or $\mathcal{S}_t$) be the state space , i.e., the set that contains all possible states. After the definitions below, in Subsection 3.2.6, we make remarks and discuss implications.

3.2.1 Exogenous information

All new information that enters the system within the interval between two decision epochs is considered *exogenous information* (see Section 9.6 from [82] on page 506). Even for a purely stochastic system, such as a branching process for example, the realization of the system at a particular decision epoch is considered exogenous information. The word "exogenous" expresses the notion that this kind of information is not a result of (although it can be affected by) other information already available in the system. Within the "Universal Modeling Framework" the exogenous information, together with the initial state S_0 (if it is stochastic), model all sources of uncertainty. We use W_t as generic notation for exogenous information. W_{t+1} denotes the exogenous information that becomes available during $(t, t+1)$.

W_t may represent multiple random variables or complicated constructs of random variables.

3.2.2 Decision variables and Policies/Decision functions

Decision variables (also known as actions or control variables; see Section 9.5 from [82] on page 500) represent how we control/interact with the evolving system. Decisions applied at epoch t may affect the evolution of the system from t onward. We denote the set of all possible decisions, the decision space, with $\mathcal{X}$ (or $\mathcal{X}_t$). We use x to index the elements of $\mathcal{X}$ (or $\mathcal{X}_t$). The decision made at decision epoch t is denoted by x_t where the subscript is meant to stress the fact that a particular element of $\mathcal{X}$ (or $\mathcal{X}_t$) is to be associated with t. Decisions can often be written as multi-dimensional vectors although they can also be more complicated mathematical objects, such as matrices.

A *policy* (see Subsection 9.5.5 from [82] on page 505) is a rule that outputs a decision x_t given a state S_t. We denote the set of all available policies with Π and index the elements of the set with π. Whenever we want to formalize the application of policy π, we may write $X^\pi(S_t)$ (or $X_t^\pi(S_t)$). $X^\pi(S_t)$ is called a *decision function* and having a decision function is equivalent to having a policy π. π can be thought of as the information that describes the decision function $X^\pi(S_t)$ (or $X_t^\pi(S_t)$). In literature the terms "policy" and "decision function" are often interchangeable. Constraints related to our decision making can enter our modeling via $\mathcal{X}$ or the information contained within Π.

3.2.3 Transition function

The *transition function* (also known as "system model" or "state transition model"; see Section 9.7 from [82] on page 515) determines how the system evolves from state S_t to state S_{t+1} given the decision that was made at time t and the exogenous information that arrives between t and $t+1$. We denote the transition function with $S^M(S_t, x_t, W_{t+1})$ (the $S^M(\cdot)$ stands for "state model"), thus $S_{t+1} = S^M(S_t, x_t, W_{t+1})$. The transition function encompasses all of the dynamics of the system, including the updating of estimates and beliefs.

Contrary to this general and simplistic description, usually the proper formal definition of the transition function for a particular problem is a very difficult task. Additional care must be taken so that the definitions of state variables, decision variables, and transition function, agree with each other.

The transition function $S^M(S_t, x_t, W_{t+1})$ can be viewed as a close relative of the transition function f used in Optimal control. We note that in the dissertation, we do not consider stochastic differential equations within the optimization problems investigated.

3.2.4 Contribution function and Objective function

The *objective function* (see Section 9.8 from [82] on page 518) specifies a relevant performance metric. We write the objective function via the *contribution function* $C(S_t, x_t)$ (or $C_t(S_t, x_t)$) which outputs the result, with respect to our metric, of applying decision x_t onto state S_t. Let γ be a discount factor (usually $\gamma \leq 1$). If the initial state S_0 is probabilistic, the objective function can be written as

$$F^\pi(S_0) = \mathbb{E}_{S_0}\mathbb{E}_{W_t,\ldots,W_T|S_0}\left\{\sum_{t=0}^{T}\gamma^t C_t\big(S_t, X_t^\pi(S_t)\big)|S_0\right\},$$

if not then we may write

$$F^\pi(S_0) = \mathbb{E}_{W_t,\ldots,W_T|S_0}\left\{\sum_{t=0}^{T}\gamma^t C_t\big(S_t, X_t^\pi(S_t)\big)|S_0\right\},$$

or more compactly

$$F^\pi(S_0) = \mathbb{E}\left\{\sum_{t=0}^{T}\gamma^t C_t\big(S_t, X_t^\pi(S_t)\big)|S_0\right\}. \tag{3.1}$$

Within the dissertation we are interested in finding

$$\max_{\pi\in\Pi} F^\pi(S_0).$$

3.2.5 State variables

Unfortunately even the definition of "state", a concept central to sequential decision problems, is not universally agreed upon within the 15 communities discussed in the Introduction (see Section 1.5 on page 22 of the dissertation, [79], [80], [81], [82]).

Within the dissertation, we adopt the definition of "state" variables given in [82] (more precisely the "Policy-dependent" version in Section 9.4 of [82] on page 481). This choice has certain implication discussed in Subsection 3.2.6.

Definition 3.1. *A state variable is a function of history that, combined with the exogenous information (and a policy), is necessary and sufficient to compute the cost/contribution function, the decision function (the policy), and any information required by the transition function to model the information needed for the cost/contribution and decision functions.*

Definition 3.1 appears a bit convoluted at first, however this is due to the generality that it tries to achieve. For a more thorough discussion of the presented definition, the interested reader is directed to Chapter 9 from [82], pages 481-484.

For the purpose of clarity of the discussion within the dissertation, we (independently of [82]) provide a more concrete yet more descriptive statement of Definition 3.1 as follows: The state of the system at t is understood as the information, that we know and keep record of, about the system at t. Further, in order for our modeling to work, the information encoded as the "state" at t together with decisions related information encoded as "policy", must be necessary and sufficient for finding a decision (i.e., calculate the decision function) to be applied to the system at t and also to calculate the contribution function at t (which may depend on the decision obtained). In addition to that, the state at t, together with the decision made at t, and the exogenous information that arrives at $t+1$, must be necessary and sufficient for the transition function to produce the state at $t+1$.

We denote the set of all possible states, the state space, with $\mathcal{S}$ (or $\mathcal{S}_t$). We use s to index the elements of $\mathcal{S}$ (or $\mathcal{S}_t$) at t and s' to index the elements of $\mathcal{S}$ (or $\mathcal{S}_{t+1}$) at $t+1$. States can often be represented as vectors, however this is not always the case.

The initial state S_0 is allowed to be stochastic (note that the indexing of W_t starts from $t = 1$).

3.2.6 Discussion

The following discussion is not directly borrowed from [82] and contains some of our thoughts regarding the newly obtained concepts. This discussion serves for clarifying these concepts, as well as presenting implications and considerations relevant for the incorporation of branching processes into SDPs within the "Universal Modeling Framework".

The definitions given in the preceding Subsections are general and provide descriptive information about the required properties of the 5 components necessary for modeling a SDP within the "Universal Modeling Framework". The level of generality employed is appropriate, considering the vast variety of possible problems the framework tries to encompass. We direct the reader to Section 3.8, within the dissertation, for concrete example specifications of SDPs with branching processes based dynamics.

As stated in [82], page 482, (vi), Definition 3.1 that defines the state variables has the following crucial implication:

Remark 3.2. *As a consequence of Definition 3.1, all properly modeled, with respect to the "Universal Modeling Framework", dynamic systems are Markovian by construction.*

This implication complicates actual modeling within the "Universal Modeling Framework" as it is often the case that appropriately discerning relevant dependencies within a dynamic system is not a trivial task. On the other hand, forcing our modeling to seek a Markovian description of a designated system may help us spend the necessary time to properly understand the dynamics of the system that we try to model and thus save us from venturing into inefficient or erroneous modeling. In addition to that, a SDP with Markovian dynamics can benefit from Bellman's optimality equation (see Section 3.3) and consequently can benefit from approximate dynamic programming and reinforcement learning as well.

Remark 3.2 relates to the exogenous information in the following way:

Remark 3.3. *The exogenous information evolving within $(t, t+1)$ can depend on states and decisions at t, but not on prior moments, or can be completely independent. Any knowledge about the exogenous information process within $(t, t+1)$, such as parameters for probability distributions, is explicitly or implicitly encoded within the state of the system and/or the decision made at t.*

We also need to be careful when modeling decisions:

Remark 3.4. *As the decision function depends only on the state at t, it cannot functionally depend on states and decisions prior to t. When doing our modeling, we must be careful - we must further ensure that the outcome from applying a decision upon the state at t does not depend on information for states and decisions prior to t.*

In a general setting, the transition function $S^M(S_t, x_t, W_{t+1})$ can have stochastic as well as deterministic components. Our focus within the dissertation is incorporating branching processes within SDPs, thus we will not be considering transition functions with deterministic components. For our purposes the transition function between t and $t+1$ is equated to the branching process between t and $t+1$ that is set by the decision at t and that has as an initial state all necessary information about the particles that exist at t. We note that in the dissertation, we do not consider stochastic differential equations within our SDPs. We stress that, since a branching process is in fact an interplay of the realizations of various underlying probability distributions (see, e.g., the definition of the MSBPM - Definition 2.1), we can prove relevant statements for a branching process through proving them for its underlying distributions.

We finish this discussion with a comment concerning the incorporation of branching process based dynamics into SDPs. Definition 3.1 and Remark 3.2 raise the question about the applicability of the "Universal Modeling Framework" outside of the well known cases where branching process can be viewed as Markov chains. Indeed, the Bienaymé-Galton-Watson branching process and Bellman-Harris branching processes with exponential lifespan are Markovian with respect to states defined as the number of particles, per type, that exist at a specified moment in time t. In order to apply the "Universal Modeling Framework" outside of these cases, it will be necessary to redefine the usual state space associated with a branching process in such a way so that the Markov property holds. In Section 3.6, we will do precisely that for the, generally non-Markovian under the standard notion of "state", MSBPM defined in Chapter 2.

3.2.7 Formal definition of a SDP within the "Universal Modeling Framework"

We are now ready to formalize Description 1.1 from Section 1.5 from the Introduction (recall page 22 of the dissertation). Definition 3.2, that we give below, constitutes an aggregation of the discussion within Chapter 9 from [82]. When writing Definition 3.2, we keep in mind Notational Choice 1.

Definition 3.2. *A finite-horizon Sequential Decision Problem (finite-horizon SDP) within the "Universal Modeling Framework", with final decision epoch T, is characterized by the sequence*

$$(S_0, x_0, W_1, S_1, \ldots, S_t, x_t, W_{t+1}, S_{t+1}, \ldots, S_T).$$

The objective of a finite-horizon SDP is to find a policy that satisfies

$$\max_{\pi \in \Pi} \mathbb{E}_{S_0} \mathbb{E}_{W_1,\ldots,W_T|S_0} \left\{ \sum_{t=0}^{T} \gamma^t C_t\big(S_t, X_t^\pi(S_t)\big) | S_0 \right\}.$$

An infinite-horizon Sequential Decision Problem (infinite-horizon SDP) within the "Universal Modeling Framework" is characterized by the sequence

$$(S_0, x_0, W_1, S_1, \ldots, S_t, x_t, W_{t+1}, S_{t+1}, \ldots).$$

The objective of an infinite-horizon SDP is to find a policy that satisfies

$$\max_{\pi \in \Pi} \mathbb{E}_{S_0} \mathbb{E}_{W_1,W_2,\ldots|S_0} \left\{ \sum_{t=0}^{\infty} \gamma^t C_t\big(S_t, X_t^\pi(S_t)\big) | S_0 \right\}.$$

If the set of all policies, Π, is infinite, we write "sup" instead of "max".

Within the dissertation, we will be interested only in finite-horizon SDPs. We leave the topic of incorporating branching processes into infinite-horizon SDPs for future considerations.

Next, we provide Proposition 3.1 and Proposition 3.2 independently of the discussion within [82]. We will actively use these propositions, as well as the considerations following them, in subsequent Sections.

Proposition 3.1. *For any fixed policy π, a SDP within the "Universal Modeling Framework" constitutes a discrete-time (possibly non-stationary) Markov chain with respect to $t = 0, 1, \ldots, T$.*

Proof. Recall that the transition function S^M connects the state at t and the state at $t+1$ via $S_{t+1} = S^M(S_t, x_t, W_{t+1})$. For a fixed policy π, since $X_t^\pi(\cdot)$ is functionally dependent only on S_t, we effectively have $S_{t+1} =$

$S^M(S_t, W_{t+1})$. For a fixed π, under the same reasoning, if the stochastic W_{t+1} is not independent, then it is functionally dependent (recall Remark 3.3) only on S_t. As S^M is a transition function for a SDP within the "Universal Modeling Framework", S^M is conditionally independent from states and decisions prior to t. Thus, the probability for obtaining S_{t+1} at $t+1$ depends only on S_t and the we have a Markov chain between t and $t+1$. The Markov chain is non-stationary if the transition probabilities, under π, are different between different neighboring decision epochs. □

The following is also true

Proposition 3.2. *A discrete-time, with respect to $t = 0, 1, \ldots, T$, possibly non-stationary, Markov chain can be viewed as a SDP within the "Universal Modeling Framework".*

Proof. A discrete-time, with respect to $t = 0, 1, \ldots, T$, Markov chain corresponds to a SDP within the "Universal Modeling Framework" that has only one possible policy. Indeed, the states of the Markov chain, the exogenous information that can be derived from the Markov chain, the singular policy, and the transition function of the Markov chain, satisfy the necessary definitions within the "Universal Modeling Framework". Without loss of generality, we can set to objective and contribution functions to 0. □

Perhaps the most difficult moment when modeling a dynamic system as a SDP within the "Universal Modeling Framework" is verifying that the defined state variables, decision variables, exogenous information variables, contribution and objective functions, and transition function, all satisfy the assumptions of a SDP within the framework. More specifically, proving the conditional independence of the transition function between t and $t+1$ from states and decisions prior to t can be especially problematic in the general case. Proposition 3.1 and Proposition 3.2 provide us with a way of checking that our model is indeed a SDP within the "Universal Modeling Framework" - what we need to do is verify that for every fixed policy π the resulting process is a discrete time Markov chain with respect to $t = 0, 1, \ldots, T$.

3.3 Bellman's optimality equation for SDPs within the "Universal Modeling Framework"

Within this Section, we adapt the discussions in Chapter 14 from [82] and Chapter 3 from [78]. Similarly to the presentation for Markov Decision Processes (MDP) within Section 4.3 in [70], we will first define Bellman's optimality equation within the "Universal Modeling Framework" and then, we will show its relevance with respect to solving SPDs.

3.3.1 Definition of Bellman's optimality equation for SDPs within the "Universal Modeling Framework"

Bellman's optimality equation for SDPs within the "Universal Modeling Framework" is discussed within Section 14.2 from [82]. In the current Subsection, we focus only on aspects of the discussion that we will need when incorporating branching processes into SDPs.

Definition 3.3. *In the context of a SDP within the "Universal Modeling Framework", as given by Definition 3.2, define (the expectation form of) Bellman's optimality equation at t as*

$$V_t(S_t) = \max_{x_t \in \mathcal{X}_t} \Big(C_t(S_t, x_t) + \gamma \mathbb{E}\big\{ V_{t+1}(S_{t+1}) | S_t, x_t \big\} \Big). \tag{3.2}$$

$V_t(S_t)$ is also known as the *value function* as it gives us the value of being in state S_t at t. If we write equation (3.2) for consecutive t it is evident that the Markov property is implied, which is to be expected knowing that all dynamic systems considered within the "Universal Modeling Framework" are Markovian. When using (3.2), we keep in mind that the exogenous information, if any, may or may not depend on S_t and x_t (and does not depend on states and decisions prior to t). If there is exogenous information affecting our system, its influence is captured by the expectation in (3.2).

Whenever the state space is discrete, we can write Bellman's optimality equation in the following form that is considered standard within the MDP community:

$$V_t(S_t) = \max_{x_t \in \mathcal{X}_t} \big(C_t(S_t, x_t) + \gamma \sum_{s' \in \mathcal{S}_{t+1}} \mathbb{P}(S_{t+1} = s' | S_t, x_t) V_{t+1}(s')\big). \tag{3.3}$$

If there is an exogenous information affecting our system, its influence is captured by $\mathbb{P}(S_{t+1} = s' | S_t, x_t)$.

If a policy π is fixed, we can substitute x_t in (3.2) and (3.3) with $X_t^\pi(S_t)$ and write

$$V_t^\pi(S_t) = C_t\big(S_t, X_t^\pi(S_t)\big) + \gamma \mathbb{E}\big\{V_{t+1}^\pi(S_{t+1}) | S_t, X_t^\pi(S_t)\big\}$$

and

$$V_t^\pi(S_t) = C_t\big(S_t, X_t^\pi(S_t)\big) + \gamma \sum_{s' \in \mathcal{S}_{t+1}} \mathbb{P}\big(S_{t+1} = s' | S_t, X_t^\pi(S_t)\big) V_{t+1}^\pi(s').$$

$V_t^\pi(S_t)$ is called the *value function under policy* π as it provides us with the value of being in state S_t following policy π. When π is fixed, since $X_t^\pi(\cdot)$ is a function of S_t, we can remove $X_t^\pi(S_t)$ from the conditions, thus arriving at

$$V_t^\pi(S_t) = C_t\big(S_t, X_t^\pi(S_t)\big) + \gamma \mathbb{E}\big\{V_{t+1}^\pi(S_{t+1}) | S_t\big\},$$

$$V_t^\pi(S_t) = C_t\big(S_t, X_t^\pi(S_t)\big) + \gamma \sum_{s' \in \mathcal{S}_{t+1}} \mathbb{P}\big(S_{t+1} = s' | S_t\big) V_{t+1}^\pi(s').$$

Bellman's optimality equation is linked to the solution of the SDP given by Definition 3.2. Since our primary interest are finite-horizon SDPs, we limit ourselves to formally verifying this connection only for the case of finite-horizon SDPs within Theorem 3.1 in the next Subsection.

3.3.2 Bellman's optimality equation and the solution of a finite-horizon SDP within the "Universal Modeling Framework"

Within this Subsection, we will be considering finite-horizon Sequential Decision Problems (SDPs). The proofs presented here are adapted versions of the proofs found in Subsection 14.12.1 of [82] (page 770). Here "adapted" indicates that we have added the discount factor γ in the statements and proofs within Subsection 14.12.1 from [82]. We note that due to Notational Choice 1, the statements below are, effectively, proven for decision epochs with (possibly) varying distance from one another.

Recall that the objective function of a finite-horizon SDP is given by equation (3.1), i.e.,

$$F^\pi(S_0) = \mathbb{E}\left\{ \sum_{t=0}^{T} \gamma^t C_t(S_t, X_t^\pi(S_t)) | S_0 \right\}.$$

We begin by investigating

$$F_t^\pi(S_t) = \mathbb{E}\left\{ \sum_{t'=t}^{T-1} \gamma^{t'-t} C_{t'}\big(S_{t'}, X_{t'}^\pi(S_{t'})\big) + \gamma^{T-t} C_T(S_T) | S_t \right\}. \tag{3.4}$$

It is clear that $F^\pi(S_0) = F_0^\pi(S_0)$. Instead of tackling (3.4) directly, we consider for a fixed π

$$V_t^\pi(S_t) = C_t(S_t, X_t^\pi(S_t)) + \gamma \mathbb{E}\big\{ V_{t+1}^\pi(S_{t+1}) | S_t \big\}. \tag{3.5}$$

The following proposition (Proposition 14.12.1 from [82]) is an application of the law of total expectation.

Proposition 3.3.

$$F_t^\pi(S_t) = V_t^\pi(S_t)$$

Proof. We provide a proof by induction. At T it is clear that $F_T^\pi(S_T) = V_T^\pi(S_T) = C_T(S_T)$. Next, assume the statement holds for $t+1, t+2, \dots, T$. We show that it is true for t. From the induction hypothesis and from equations (3.4) and (3.5), we have

$$\begin{aligned}
V_t^\pi(S_t) &= \\
&= C_t(S_t, X_t^\pi(S_t)) + \gamma \mathbb{E}\Bigg\{ \mathbb{E}\Big\{ \sum_{t'=t+1}^{T-1} \gamma^{t'-(t+1)} C_{t'}(S_{t'}, X_{t'}^\pi(S_{t'})) + \\
&\qquad\qquad + \gamma^{T-(t+1)} C_T(S_T) \Big| S_{t+1} \Big\} \Big| S_t \Bigg\} \\
&= C_t(S_t, X_t^\pi(S_t)) + \gamma \mathbb{E}\Big\{ \sum_{t'=t+1}^{T-1} \gamma^{t'-(t+1)} C_{t'}(S_{t'}, X_{t'}^\pi(S_{t'})) + \gamma^{T-(t+1)} C_T(S_T) \Big| S_t \Big\}
\end{aligned}$$

$$= \mathbb{E}\Big\{ \sum_{t'=t}^{T-1} \gamma^{t'-t} C_{t'}(S_{t'}, X^{\pi}_{t'}(S_{t'})) + \gamma^{T-t} C_T(S_T) \Big| S_t \Big\}$$
$$= F^{\pi}_t(S_t),$$

which completes the proof. □

Next, we turn our attention to

$$F^*_t(S_t) = \max_{\pi \in \Pi} F^{\pi}_t(S_t).$$

If Π is infinite, we replace "max" with "sup". The following theorem corresponds to Theorem 14.12.1 from [82].

Theorem 3.1. *Let $V_t(S_t)$ be a solution to Bellman's optimality equation* (3.2)

$$V_t(S_t) = \max_{x_t \in \mathcal{X}_t} \Big(C_t(S_t, x_t) + \gamma \mathbb{E}\big\{ V_{t+1}(S_{t+1}) | S_t, x_t \big\} \Big).$$

Then, for a finite-horizon SDP

$$F^*_t(S_t) = V_t(S_t).$$

Proof. The proof consists of two parts. First, we show by induction that $V_t(S_t) \geq F^*_t(S_t)$ for all $S_t \in \mathcal{S}$ and $t = 0, 1, \ldots, T-1$. Then, we show that the reverse inequality is true.

Part 1:
Since $V_T(S_T) = C_T(S_T) = F^{\pi}_T(S_T)$ for all S_T and all $\pi \in \Pi$, we get that $V_T(S_T) = F^*_T(S_T)$.
Let our induction hypothesis be that $V_{t'}(S_{t'}) \geq F^*_{t'}(S_{t'})$ for $t' = t+1, t+2, \ldots, T$, and let π be an arbitrary policy. For $t' = t$, Bellman's optimality equation (3.2) is

$$V_t(S_t) = \max_{x_t \in \mathcal{X}_t} \Big(C_t(S_t, x_t) + \gamma \mathbb{E}\big\{ V_{t+1}(S_{t+1}) | S_t, x_t \big\} \Big).$$

By the induction hypothesis, $V_{t+1}(S_{t+1}) \geq F^*_{t+1}(S_{t+1})$, so we get

$$V_t(S_t) \geq \max_{x_t \in \mathcal{X}_t} \Big(C_t(S_t, x_t) + \gamma \mathbb{E}\big\{ F^*_{t+1}(S_{t+1}) | S_t, x_t \big\} \Big).$$

By definition, we have that $F^*_{t+1}(S_{t+1}) \geq F^\pi_{t+1}(S_{t+1})$ for an arbitrary π. Let $X^\pi(S_t)$ be the decision that is chosen by policy π when in state S_t. Then with the help of Proposition 3.3

$$\begin{aligned} V_t(S_t) &\geq \max_{x_t \in \mathcal{X}_t} \Big(C_t(S_t, x_t) + \gamma \mathbb{E}\big\{ F^\pi_{t+1}(S_{t+1}) | S_t, x_t \big\} \Big) \\ &\geq C_t(S_t, X^\pi(S_t)) + \gamma \mathbb{E}\big\{ F^\pi_{t+1}(S_{t+1}) | S_t, X^\pi(S_t) \big\} \\ &= F^\pi_t(S_t). \end{aligned}$$

Thus

$$V_t(S_t) \geq F^\pi_t(S_t) \text{ for all } \pi \in \Pi,$$

which proves Part 1.

Part 2:
Now we are going to prove the inequality from the other side. More specifically, we want to show that for any $\varepsilon > 0$ there exists a policy π that satisfies

$$F^\pi_t(S_t) + (T - t)\varepsilon \geq V_t(S_t). \tag{3.6}$$

To do this, we again recall Bellman's optimality equation

$$V_t(S_t) = \max_{x_t \in \mathcal{X}_t} \Big(C_t(S_t, x_t) + \gamma \mathbb{E}\big\{ V_{t+1}(S_{t+1}) | S_t, x_t \big\} \Big). \tag{3.7}$$

In general, the set $\mathcal{X}_t$ may be infinite, whereupon we have to replace "max" with a "sup" and handle the case where an optimal decision may not exist. For this case, we know that we can design a decision rule that returns a decision x_t that satisfies

$$V_t(S_t) \leq C_t(S_t, x_t) + \gamma \mathbb{E}\big\{ V_{t+1}(S_{t+1}) | S_t, x_t \big\} + \varepsilon. \tag{3.8}$$

We can prove (3.6) by induction. We first note that (3.6) is true for $t = T$ since $F^\pi_T(S_T) = V_T(S_T)$. Now assume that it is true for $t' = t+1, t+2, \ldots, T$. We know from equation (3.5) and Proposition 3.3 that

$$F^\pi_t(S_t) = C_t(S_t, X^\pi(S_t)) + \gamma \mathbb{E}\big\{ F^\pi_{t+1}(S_{t+1}) | S_t, X^\pi(S_t) \big\}.$$

We can use our induction hypothesis, which is $F^\pi_{t+1}(S_{t+1}) \geq V_{t+1}(S_{t+1}) - (T - (t+1))\varepsilon$, to obtain

$$F^\pi_t(S_t) \geq C_t(S_t, X^\pi(S_t)) + \gamma \mathbb{E}\big\{ V_{t+1}(S_{t+1}) - (T - (t+1))\varepsilon | S_t, X^\pi(S_t) \big\}$$

$$= C_t(S_t, X^\pi(S_t)) + \gamma\mathbb{E}\big\{V_{t+1}(S_{t+1})|S_t, X^\pi(S_t)\big\}-$$
$$-\gamma\mathbb{E}\big\{(T-t-1)\varepsilon|S_t, X^\pi(S_t)\big\}$$
$$= \Big\{C_t(S_t, X^\pi(S_t)) + \gamma\mathbb{E}\big\{V_{t+1}(S_{t+1})|S_t, X^\pi(S_t)\big\} + \varepsilon\Big\} - (T-t)\varepsilon.$$

Now, using (3.8), we replace the term in brackets with $V_t(S_t)$:

$$F_t^\pi(S_t) \geq V_t(S_t) - (T-t)\varepsilon,$$

which completes Part 2.

Combining the results of Part 1 and Part 2, we have shown that

$$F_t^*(S_t) + (T-t)\varepsilon \geq F_t^\pi(S_t) + (T-t)\varepsilon \geq V_t(S_t) \geq F_t^*(S_t).$$

Since ε can be arbitrary small, the proof of the theorem is complete. □

3.3.3 Bellman's optimality equation and the three curses of dimensionality

In what follows, we adopt the term "dynamic programming" as viewed in [70], page 3 - we will use "dynamic programming" to describe an approach for solving SDPs based on inductive computation. Dynamic programming algorithms are algorithms that utilize Bellman's optimality (3.2) or equation (3.3).

If the state and decision spaces are discrete and finite (these are some of the classical assumptions within the Markov decision processes community, see Subsection 2.1.2 from [70] on page 18), we can try using the following algorithm, often referred to as "Backward Dynamic Programming" or "Backward Induction Algorithm" (see page 65 from [78] and page 92 from [70]), in order to solve a SDP.

Algorithm 3.1.

Step 0. *Initialization: Initialize the terminal contribution $V_T(S_T)$. Set $t = T-1$.*

Step 1. *Calculate* (3.3) *(or* (3.2)*):*

$$V_t(S_t) = \max_{x_t}\Big\{C_t(S_t, x_t) + \gamma\sum_{s'\in\mathcal{S}}\mathbb{P}(s'|S_t, x_t)V_{t+1}(s')\Big\},$$

for all $S_t \in \mathcal{S}$.

Step 2. *If $t > 0$, decrement t and return to Step 1. Else STOP.*

Unfortunately, Algorithm 3.1, as well as other dynamic programming algorithms such as "Value iteration" (see Section 3.4 from [78]) and "Policy iteration" (see Section 3.5 from [78]) for infinite-horizon SDPs, usually works well only for (relatively) small problems with discrete state and action spaces. This is because at some point in the algorithm (for Algorithm 3.1 this is Step 1), we have to do computations for all $S_t \in \mathcal{S}$. Algorithm 3.1 illustrates well the so-called curse of dimensionality - generic dynamic programming algorithms can easily (but not always) become non-applicable with respect to multi-dimensional and/or countably infinite $\mathcal{S}$. A second curse of dimensionality exists in relation to x_t since in some settings the decision space can also be multi-dimensional and/or countably infinite. There is potentially an additional third curse of dimensionality if we try to decompose $\mathbb{P}(s'|S_t, x_t)$ with respect to the exogenous information which can be specified by multiple random variables.

Dropping the assumptions for discrete and finite state and decision spaces, we can encounter multi-dimensional, continuous, state and decision spaces, further the exogenous information can also be specified by multiple continuous random variables. In such cases it is often not possible to apply the dynamic programming paradigm directly. However, the core ideas of the paradigm can be used within approaches such as Approximate Dynamic Programming (ADP; see [78], [76], [82]) and Reinforcement Learning (RL; see [203], [82]) that are adapted towards handling the curses of dimensionality. We will touch upon ADP in Section 3.7 of the dissertation.

We note that branching processes are stochastic processes where the state is usually defined as the number of particles, per type, that exist at a specified t. Such a state space is countably infinite and possibly multi-dimensional. Thus, applying algorithms based on Bellman's optimality equation in the context of branching processes is not trivial.

3.4 SDPs with underlying BGW branching process dynamics

To the best of our knowledge, stochastic sequential decision problems, where the dynamics is generated specifically by a branching process (the case of the Bienaymé-Galton-Watson branching process is investigated), are considered for the first time in [77] (1976). Theorem 3.1 from [77] provides us with a dynamic programming algorithm that, unlike Algorithm 3.1 from Subsection 3.3.3, does not require from us to iterate all states within $\mathcal{S}$.

Since we will be considering Bienaymé-Galton-Watson (BGW) branching processes, Notational Choice 1 is redundant - the moments in time when the decision epochs occur coincide with the indexing given ty t.

Our contributions within this Section are: 1) We recast the Markov decision process from [77] into the "Universal Modeling Framework"; 2) We provide a novel proof for Theorem 3.1 from [77] that is based on Bellman's optimality equation. These contributions have not been published yet.

3.4.1 Informal description of the SDP with underlying BGW branching process dynamics

Before jumping into notational particularities, in order to have a clear picture of the dynamic system we model, we give an informal description of the finite-horizon SDP discussed in Section 2 and Section 3 of [77]. The sketch of this description is applicable to all SDPs discussed within the dissertation.

Consider a finite set of decision epochs indexed by t, $t = 0, 1, \ldots, T$. Next, consider a multi-type BGW branching process that is consecutively observed at the decision epochs indexed by t. At each observation, starting with $t = 0$:

1. For each particle that exists at t, from each particle type, a decision is made from among a discrete set of possible decisions. Decisions affect the reproductive capabilities of particles that exist at t, thus the characteristics of the BGW branching process may change whenever a decision is made. Because of this, the evolution of the system of particles is actually modeled by a sequence of (possibly) different BGW branching processes, each existing for 1 time unit and each passing its offspring as the initial state of the next BGW branching process. If $t = T$ no decisions are made.

2. Upon making a decision for a particle, a reward specific to the combination of particle type and decision made is collected. If $t = T$, as stated, no decisions are made, instead a predefined "terminal reward" is collected per each particle with respect to type.

3. After making decisions and collecting rewards at t, the (possibly modified) BGW branching process is left to evolve for 1 time unit until $t+1$, where we make new decisions and collect new rewards. If $t = T$ we terminate after collecting the terminal rewards.

Our objective is to maximize the cumulative expected reward collected by choosing appropriate decisions at each t.

3.4.2 Definition of SDP Model 1 - SDP with underlying BGW branching process dynamics

Before proceeding, we must take note of the following circumstance within [77]:

Remark 3.5. *In [77] it is originally allowed for different decisions to be made per each particle per each type. However, as proven within [77] itself, the optimal policy is one that produces the same decision for each particle within a type. Actually, the assumption of independent evolution of particles ensures that once we have found an optimal decision at t (corresponding to the optimal policy) for one particle of type i, then this decision is optimal at t for all particles of type i. This in effect allows us to redefine the decision space used in [77] as a space consisting of discrete tuples where each element gives the decision for all particles of the corresponding type. We do take this approach as doing otherwise introduces unnecessary notational clutter.*

For convenience, we also adopt the following:

Remark 3.6. *For notational convenience, we will not explicitly write the probability distributions and p.g.f.s, per type, that govern a branching process between t and $t+1$, within the state variables. We do so because, we want to take advantage of writing the state of the system as a vector where each vector coordinate corresponds to a type. We will deem probability distributions and p.g.f.s associated with type i to be implicitly known when we are considering particles of type i.*

We are now ready to specify SDP Model 1 which is the finite-horizon model discussed in Section 3 of [77].

Definition of SDP Model 1. Define SDP Model 1 as the finite-horizon Sequential Decision Problem that satisfies:

1. We observe a BGW branching process at successive times indexed with $t = 0, 1, 2, \dots, T$.

2. Let k be the number of particles types within the BGW branching process. The state space $\mathcal{S}_t$ consists of all k-dimensional vectors whose coordinates are non-negative integers. The t index in $\mathcal{S}_t$ indicates that there are probability distributions and p.g.f.s associated with each type that may change with t (after a decision has been made). The state of the process at t is also called the “generation” or the “population” at t. The state of the process at t is given by $\boldsymbol{S}_t = (S_{1,t}, S_{2,t}, \dots, S_{k,t})^\top$, where all $S_{i,t}$ are with values in $\mathbb{N}_0$ and $S_{i,t}$ is the (non-negative) number of particles of type i that exist at t. The initial state $\boldsymbol{S}_0$ is deterministic.

3. Each particle type i has a specific finite set of possible decisions (actions) $\widetilde{\mathcal{X}}_i$ associated with it. Hence, the decision space is given by $\mathcal{X} = \widetilde{\mathcal{X}}_1 \times \widetilde{\mathcal{X}}_2 \times \cdots \times \widetilde{\mathcal{X}}_k$. We denote the decisions made at t with $\boldsymbol{x}_t = (x_{1,t}, x_{2,t}, \dots, x_{k,t})^\top$.

4. Let $c_i(x_{i,t})$ be the individual contribution (reward) received for a type i particle after making decision $x_{i,t}$ for all particles of type i at t. We assume that $-\infty < c_i(x_{i,t}) < \infty$ for all i and that $c_i(\cdot)$ do not depend on t. If we let $\boldsymbol{c}(\boldsymbol{x}_t) = \big(c_1(x_{1,t}), \dots, c_k(x_{k,t})\big)^\top$, then the generation contribution at t is $C_t\big(\boldsymbol{S}_t, \boldsymbol{x}_t\big) = \sum_{i=1}^k S_{i,t} \cdot c_i(x_{i,t}) = \boldsymbol{S}_t^\top \boldsymbol{c}(\boldsymbol{x}_t)$. At $t = T$ no decisions are made, instead a terminal $\boldsymbol{c}_T = \big(c_1, \dots, c_k\big)^\top$ is collected, hence the generation contribution at $t = T$ is $C_T(\boldsymbol{S}_T) = \boldsymbol{S}_T^\top \boldsymbol{c}_T$.

5. The decision selected for a particle affects the number of offspring, per type, that the particle has in the next generation.

   (a) For each k-dimensional vector $\boldsymbol{q} = (q_1, \dots, q_k)^\top$ of non-negative integers, let $p_i(\boldsymbol{q}, x_{i,t})$ be the probability that a type i particle, whose corresponding decision is $x_{i,t}$, will produce exactly q_1 type 1 offspring, …, q_k type k offspring, $\sum_{\boldsymbol{q}} p_t(\boldsymbol{q}, x_{i,t}) = 1$.

   (b) Corresponding to each $p_i(\cdot, x_{i,t})$ is the row vector $\boldsymbol{m}_i(x_{i,t}) = \big(m_{i1}(x_{i,t}), \dots, m_{ik}(x_{i,t})\big)$, where $m_{ij}(x_{i,t})$ equals the expected number of type j offspring produced by a single particle of type i under decision $x_{i,t}$. We assume that $m_{ij}(x_{i,t}) < \infty$ for all $x_{i,t} \in \widetilde{\mathcal{X}}_i$ and $i, j = 1, \dots, k$. Given $\boldsymbol{x}_t \in \mathcal{X}$, we organize the expectations into matrix $M(\boldsymbol{x}_t) = (\boldsymbol{m}_1(x_{1,t}), \ldots, \boldsymbol{m}_k(x_{k,t}))^\top$.

6. Denote a policy by π. The set of possible policies, with respect to $\mathcal{X}$, is Π. Denote the decision function at t, corresponding to policy π, with $X_t^\pi(\cdot)$.
7. A discount factor γ is given.
8. We want to obtain the maximum expected T-period discounted reward given by

$$\max_{\pi \in \Pi} \mathbb{E}\left\{ \sum_{t=0}^{T-1} \gamma^t C_t(\boldsymbol{S}_t, X_t^\pi(\boldsymbol{S}_t)) + \gamma^T C_T(\boldsymbol{S}_T) | \boldsymbol{S}_0 \right\}.$$

In the context of Proposition 3.1 and Proposition 3.2 from Subsection 3.2.7, let us verify that for any fixed π SDP Model 1 is a (possibly non-stationary) Markov chain. It has already been noticed within [77] that the model defined there gives rise to a non-stationary Markov chain for each policy, although no details are provided. Indeed, given $\boldsymbol{S}_t$, the decision dictated by π, $\boldsymbol{x}_t$, modifies all (offspring) distributions that are relevant for the dynamics of the system. As within a BGW branching process all particles at t are of age 0, $\boldsymbol{x}_t$ sets the dynamics between t and $t+1$ (we explicitly note that particles have age 0 as the question regarding particle age is central to Section 3.6 and the SDP defined there). Thus we have a Markov chain since the dynamics is dependent only on $\boldsymbol{S}_t$ (the policy π being applied is fixed) and there is no dependence from prior states. When considering $\boldsymbol{S}_{t+1}$ and $(t+1, t+2)$, everything is as in the case of $(t, t+1)$, however, $\boldsymbol{x}_{t+1}$ (dictated by π) sets possibly different underlying distributions. When the distributions being set are different, the transition function between $t+1$ and $t+2$ is also different, thus the resulting Markov chain can be non-stationary. This completes our check and verifies that SDP Model 1 is indeed a SDP within the "Universal Modeling Framework".

We can formally characterize the transition function of SDP Model 1 in the following way. Let $\boldsymbol{Z}_t = \big(Z_{1,t}, \ldots, Z_{k,t}\big)^\top$ denote the multi-type BGW process at t, where $Z_{i,t}$ is the number of particles of type i that exist at t. Then, the transition function corresponds to $\mathbb{P}(\boldsymbol{Z}_{t+1} = \boldsymbol{S}_{t+1} | \boldsymbol{Z}_t = \boldsymbol{S}_t, X_t^\pi(\boldsymbol{S}_t) = \boldsymbol{x}_t)$, where we keep Remark 3.6 in mind. Having checked that SDP Model 1 is truly a SDP, we have Bellman's optimality equation at our disposal.

Remark 3.7. *Note that for $\boldsymbol{S}_{t+1}$ the associated probability distributions and p.g.f.s for each type are the ones that come into effect after decision $\boldsymbol{x}_t$ was made at t for $\boldsymbol{S}_t$.*

3.4.3 Solution of SDP Model 1

The algorithmic procedure to be used for efficiently obtaining the solution of SDP Model 1 has been established in Theorem 3.1 within [77] (1976). The proof of Theorem 3.1 from [77] is based on conditional expectations and although the author recognizes that the optimal policy can be computed from Theorem 3.1 "...with the usual iterative methods of dynamic programming...", Bellman's optimality equation is not mentioned within the paper.

In this Subsection, we provide a novel proof of Theorem 3.1 from [77] that is centered around Bellman's optimality equation (3.2). More specifically, since SDP Model 1 is a SDP within the "Universal Modeling Framework", we can utilize Theorem 3.1 from Subsection 3.3.2 which allows us to apply Bellman's optimality equation. We use Bellman's optimality equation as a starting point within our proof. Our motivation for developing a new proof can be summarized as follows: 1) The new proof is in the more contemporary context of the "Universal Modeling Framework"; 2) A proof that utilizes Bellman's optimality equation can prove to be very useful as a guideline in future research where more complicated SDPs may be considered; 3) This new proof also happens to be applicable in the context of SDP Model 2, defined in Subsection 3.5.1, and, to the best of our knowledge, SDP Model 2 has not been considered before in the literature.

The definition of the *maximum return operator*, identified in [77], will allow us to write our expressions more compactly:

Definition 3.4. *Let $\boldsymbol{X}$ be a k-dimensional vector. The maximum return operator $\mathcal{R}$ is given by*

$$\mathcal{R}\boldsymbol{X} = \max_{\boldsymbol{x}\in\mathcal{X}} \left\{ \boldsymbol{c}(\boldsymbol{x}) + \gamma M(\boldsymbol{x})\boldsymbol{X} \right\}.$$

We denote the n-fold composition of $\mathcal{R}$ as $\mathcal{R}^n$ and $\mathcal{R}^0$ is understood as the identity operator.

We are ready for the main results of Section 3.4.

Theorem 3.2. *For SDP Model 1, the value function $V_t(\boldsymbol{S}_t)$ satisfies*

$$V_t(\boldsymbol{S}_t) = \boldsymbol{S}_t^\top \mathcal{R}^{T-t}\boldsymbol{c}_T, \qquad t = 0, 1, \ldots, T-1. \tag{3.9}$$

Policy π, with corresponding $X_t^\pi(\cdot)$, that computes decisions $\boldsymbol{x}_t$ satisfying

$$\boldsymbol{c}(\boldsymbol{x}_t) + \gamma M(\boldsymbol{x}_t)\mathcal{R}^{T-t-1}\boldsymbol{c}_T = \mathcal{R}^{T-t}\boldsymbol{c}_T, \qquad t = 0, 1, \ldots, T-1, \tag{3.10}$$

is optimal.

Proof. Let us look at the expectation form of Bellman's optimality equation, i.e., equation (3.2) for $t \neq T$. We write it down, using vector notation, and take into account that, within SPD Model 1, $\mathcal{X}$ and the individual contributions received per particle do not depend on t

$$V_t(\boldsymbol{S}_t) = \max_{\boldsymbol{x}_t \in \mathcal{X}} \Big(C_t(\boldsymbol{S}_t, \boldsymbol{x}_t) + \gamma \mathbb{E}\big\{ V_{t+1}(\boldsymbol{S}_{t+1}) | \boldsymbol{S}_t, \boldsymbol{x}_t \big\} \Big).$$

Let $\boldsymbol{S}_t = (S_{1,t}, S_{2,t}, \ldots, S_{k,t})^\top$. The assumption of independent evolution of particles within SDP Model 1 allow us to view the BGW branching process at t with state $\boldsymbol{S}_t$ as consisting of $S_{1,t}$ number of independent BGW branching processes starting at t with one particle of type 1, plus $S_{2,t}$ number of independent BGW branching processes starting at t with one particle of type 2, ..., plus $S_{k,t}$ number of independent BGW branching processes starting at t with one particle of type k. Thus, for $\boldsymbol{S}_t = (S_{1,t}, S_{2,t}, \ldots, S_{k,t})^\top$ we may write

$$\begin{aligned}
V_t(\boldsymbol{S}_t) = \max_{\boldsymbol{x}_t \in \mathcal{X}} \Bigg(&\boldsymbol{S}_t^\top \boldsymbol{c}(\boldsymbol{x}_t) + \gamma \Big(S_{1,t} \cdot \mathbb{E}\big\{ V_{t+1}(\boldsymbol{S}_{t+1}) | (1, 0, \ldots, 0)^\top, \boldsymbol{x}_t \big\} + \\
&+ S_{2,t} \cdot \mathbb{E}\big\{ V_{t+1}(\boldsymbol{S}_{t+1}) | (0, 1, \ldots, 0)^\top, \boldsymbol{x}_t \big\} + \ldots \\
&+ S_{k,t} \cdot \mathbb{E}\big\{ V_{t+1}(\boldsymbol{S}_{t+1}) | (0, 0, \ldots, 1)^\top, \boldsymbol{x}_t \big\} \Big) \Bigg) \\
= \max_{\boldsymbol{x}_t \in \mathcal{X}} \Bigg(&\boldsymbol{S}_t^\top \boldsymbol{c}(\boldsymbol{x}_t) + \gamma \boldsymbol{S}_t^\top \cdot \Big(\mathbb{E}\big\{ V_{t+1}(\boldsymbol{S}_{t+1}) | \boldsymbol{S}_t = \boldsymbol{\delta}^1, \boldsymbol{x}_t \big\}, \ldots, \\
&\mathbb{E}\big\{ V_{t+1}(\boldsymbol{S}_{t+1}) | \boldsymbol{S}_t = \boldsymbol{\delta}^k, \boldsymbol{x}_t \big\} \Big)^\top \Bigg).
\end{aligned}$$

If we denote
(3.11)
$$\boldsymbol{E}_t(\boldsymbol{x}_t) = \Big(\mathbb{E}\big\{V_{t+1}(\boldsymbol{S}_{t+1})|\boldsymbol{S}_t = \boldsymbol{\delta}^1, \boldsymbol{x}_t\big\}, \dots, \mathbb{E}\big\{V_{t+1}(\boldsymbol{S}_{t+1})|\boldsymbol{S}_t = \boldsymbol{\delta}^k, \boldsymbol{x}_t\big\}\Big)^\top,$$

we arrive at the compact expression

$$\begin{aligned} V_t(\boldsymbol{S}_t) &= \max_{\boldsymbol{x}_t \in \mathcal{X}} \Big(\boldsymbol{S}_t^\top \cdot \big(\boldsymbol{c}(\boldsymbol{x}_t) + \gamma \boldsymbol{E}_t(\boldsymbol{x}_t)\big)\Big) \\ &= \boldsymbol{S}_t^\top \cdot \max_{\boldsymbol{x}_t \in \mathcal{X}} \big(\boldsymbol{c}(\boldsymbol{x}_t) + \gamma \boldsymbol{E}_t(\boldsymbol{x}_t)\big). \end{aligned} \tag{3.12}$$

We can now see that in (3.12) the choice of best $\boldsymbol{x}_t \in \mathcal{X}$ actually does not depend on $\boldsymbol{S}_t$.

Let us now look at $t = T$ where by the definition of SDP Model 1 we have

$$V_T(\boldsymbol{S}_T) = C_T(\boldsymbol{S}_T) = \boldsymbol{S}_T^\top \boldsymbol{c}_T. \tag{3.13}$$

Applying (3.11) for $t = T-1$, we obtain

$$\begin{aligned} \boldsymbol{E}_{T-1}(\boldsymbol{x}_{T-1}) &= \Big(\mathbb{E}\big\{\boldsymbol{S}_T^\top \boldsymbol{c}_T|\boldsymbol{S}_{T-1} = \boldsymbol{\delta}^1, \boldsymbol{x}_{T-1}\big\}, \dots, \mathbb{E}\big\{\boldsymbol{S}_T^\top \boldsymbol{c}_T|\boldsymbol{S}_{T-1} = \boldsymbol{\delta}^k, \boldsymbol{x}_{T-1}\big\}\Big)^\top \\ &= M(\boldsymbol{x}_{T-1})\boldsymbol{c}_T. \end{aligned}$$

Substituting into (3.12) leads to

$$V_{T-1}(\boldsymbol{S}_{T-1}) = \boldsymbol{S}_{T-1}^\top \cdot \max_{\boldsymbol{x}_{T-1} \in \mathcal{X}} \big(\boldsymbol{c}(\boldsymbol{x}_{T-1}) + \gamma M(\boldsymbol{x}_{T-1})\boldsymbol{c}_T\big) \tag{3.14}$$
$$= \boldsymbol{S}_{T-1}^\top \mathcal{R} \boldsymbol{c}_T, \tag{3.15}$$

which proves (3.9) for $t = T-1$. We continue by induction. Let (3.9) be true for $t' = t+1, t+2, \dots, T-1$, and let us investigate $t' = t$. By virtue of (3.12), we have

$$V_t(\boldsymbol{S}_t) = \boldsymbol{S}_t^\top \cdot \max_{\boldsymbol{x}_t \in \mathcal{X}} \big(\boldsymbol{c}(\boldsymbol{x}_t) + \gamma \boldsymbol{E}_t(\boldsymbol{x}_t)\big).$$

Writing down (3.11) for t, we obtain

$$\begin{aligned} \boldsymbol{E}_t(\boldsymbol{x}_t) &= \Big(\mathbb{E}\big\{V_{t+1}(\boldsymbol{S}_{t+1})|\boldsymbol{S}_t = \boldsymbol{\delta}^1, \boldsymbol{x}_t\big\}, \dots, \mathbb{E}\big\{V_{t+1}(\boldsymbol{S}_{t+1})|\boldsymbol{S}_t = \boldsymbol{\delta}^k, \boldsymbol{x}_t\big\}\Big)^\top \\ &= \Big(\mathbb{E}\big\{\boldsymbol{S}_{t+1}^\top \mathcal{R}^{T-t-1} \boldsymbol{c}_T|\boldsymbol{S}_t = \boldsymbol{\delta}^1, \boldsymbol{x}_t\big\}, \dots, \mathbb{E}\big\{\boldsymbol{S}_{t+1}^\top \mathcal{R}^{T-t-1} \boldsymbol{c}_T|\boldsymbol{S}_t = \boldsymbol{\delta}^k, \boldsymbol{x}_t\big\}\Big)^\top \end{aligned}$$

$$= M(\boldsymbol{x}_t)\mathcal{R}^{T-t-1}\boldsymbol{c}_T,$$

thus

$$\begin{aligned} V_t(\boldsymbol{S}_t) &= \boldsymbol{S}_t^\top \cdot \max_{\boldsymbol{x}_t \in \mathcal{X}} \left(\boldsymbol{c}(\boldsymbol{x}_t) + \gamma M(\boldsymbol{x}_t)\mathcal{R}^{T-t-1}\boldsymbol{c}_T\right) \\ &= \boldsymbol{S}_t^\top \mathcal{R}^{T-t}\boldsymbol{c}_T, \end{aligned}$$

which completes the proof of (3.9). The proof of (3.10) follows from (3.9) by noticing that at any $t = 0, 1, \ldots, T-1$, the best $\boldsymbol{x}_t$ is derived by maximizing $\left(\boldsymbol{c}(\boldsymbol{x}_t) + \gamma M(\boldsymbol{x}_t)\mathcal{R}^{T-t-1}\boldsymbol{c}_T\right)$, which, by the definition of $\mathcal{R}$, is equivalent to it $\mathcal{R}^{T-t}\boldsymbol{c}_T$. □

We note the following with respect to the result obtained within Theorem 3.2:

Remark 3.8. *Note that Theorem 3.2 in effect says that finding the best policy π actually does not depend on the state at any t. Finding the best decision at t amounts to applying the maximum return operator, which means that we still have to go through all possible x_t at t and make corresponding computations using matrices $M(\boldsymbol{x}_t)$. We note that enumerating all possible decisions and computing their corresponding matrices $M(\boldsymbol{x}_t)$ can prove to be a challenge for problems that have many types and many possible decisions per type, thus the curse of dimensionality related to the decision space persists.*

3.5 SDPs with underlying exponential lifespan MBHBPM dynamics

Recall that the multi-type Bellman-Harris branching process is a particular case of a multi-type Sevastyanov branching process where particle reproduction does not depend on particle age. Analogously, the Multi-type Bellman-Harris Branching Process through probabilities of Mutation between types (MBHBPM) is a particular case of the MSBPM from Definition 2.1 where particle reproduction does not depend on particle age.

Throughout this Section, let $\boldsymbol{Z}_t = \left(Z_{1,t}, \ldots, Z_{k,t}\right)^\top$ denote either the multi-type Bellman-Harris branching process or the MBHBPM at t, where $Z_{i,t}$ is the number of particles of type i that exist at t. We will be following Remark 3.6 and keeping Remark 3.7 and Notational Choice 1 in mind.

Our contributions within this Section are: 1) We provide a proof that the multi-type Bellman-Harris branching process with exponential lifespans, as well as the MBHBPM with exponential lifespans, are discrete-time Markov chains with respect to $t = 0, 1, \ldots, T$; 2) For these processes, we construct SDPs within the "Universal Modeling Framework"; 3) We show that a theorem similar to Theorem 3.2 holds for the newly constructed SDPs. These contributions have not been published yet.

3.5.1 Definition of SDP Model 2 - SDP with underlying exponential lifespan MBHBPM dynamics

It is well known, [57], that a multi-type Bellman-Harris branching process with exponential lifespan distributions for all particle types can be viewed as a continuous-time Markov chain due to the "memorylessness" property of the exponential distribution. Since we will be considering SDPs, however, we are interested whether the multi-type Bellman-Harris branching process with exponential lifespan distributions for all particle types can be viewed as a discrete-time Markov chain with respect to a discrete set of moments in time indexed by $t = 0, 1, \ldots, T$.

We could not find an instance in the literature where the multi-type Bellman-Harris branching process with exponential lifespans is considered as a discrete-time Markov chain, hence we provide our own proof.

Proposition 3.4. *A multi-type Bellman-Harris branching process with exponential lifespan distributions for all particle types, with states defined as the number of particles, per type, that exist at moment t, is a discrete-time Markov chain with respect to the moments in time indexed by $t = 0, 1, \ldots, T$.*

Proof. With respect to $t = 0, 1, \ldots, T$, we need to prove that the transition function between the state at t and the state at $t + 1$, for the multi-type Bellman-Harris branching process with exponential lifespan distributions for all particle types, is conditionally independent from states prior to t and depends at most on the state at t, i.e., that $\mathbb{P}(\boldsymbol{Z}_{t+1} = \boldsymbol{s}_{t+1} | \boldsymbol{Z}_t = \boldsymbol{s}_t, \ldots, \boldsymbol{Z}_0 = \boldsymbol{s}_0) = \mathbb{P}(\boldsymbol{Z}_{t+1} = \boldsymbol{s}_{t+1} | \boldsymbol{Z}_t = \boldsymbol{s}_t)$.

Due to the memorylessness property of the exponential distribution all particles that exist at t can be considered as having age 0. Having the assumption of independent evolution of particles in mind, the dynamics of a multi-type Bellman-Harris branching process with exponential lifespans, between t and $t+1$, is set by:

1. The lifespan distribution associated with type i, $i = 1, \ldots, k$.

2. The offspring distribution among types associated with type i, $i = 1, \ldots, k$.

We must prove conditional independence from states prior to t of the aforementioned distributions. Conveniently, by definition, these distributions do not depend on any state. Consequently, the transition function between t and $t+1$ depends only on the state at t. □

Since the MBHBPM with exponential lifespan distributions corresponds to the multi-type Bellman-Harris branching process with exponential lifespan distributions (see the discussion below Definition 2.1), by virtue of Proposition 3.4, the MBHBPM with exponential lifespan distributions, as well as the DMBHBPM with exponential lifespan distributions, can also be viewed as a discrete-time Markov chain with respect to $t = 0, 1, \ldots, T$.

We now turn to defining a SDP with underlying exponential lifespan MBHBPM dynamics. When defining a new SDP, however, we must make sure that the transition function for the new SDP from t to $t+1$ is conditionally independent not only from states at moments prior to t but also from decisions made prior to t. Having in mind Proposition 3.1 and Proposition 3.2, we must check if for every fixed π, π being with respect from the class of decisions we want to consider, the new model, with respect to its transition function and $t = 0, 1, \ldots, T$, is a (possibly non-stationary) discrete-time Markov chain. We first give the definition of the new SDP and then check that it is indeed a SDP within the "Universal Modeling Framework".

Definition of SDP Model 2. Define SDP Model 2 as the finite-horizon SDP that corresponds to the definition of SDP Model 1 upon which the following modifications are applied:

1. We observe a MBHBPM at epochs indexed with t, $t = 0, 1, 2, \ldots, T$. Regardless of t, lifespan distributions for particles of each type must be exponential.

2. Let k be the number of types of particles within the MBHBPM. The state space $\mathcal{S}_t$ consists of all k-dimensional vectors whose coordinates

are non-negative integers. The t index in $\mathcal{S}_t$ indicates that there are probability distributions and p.g.f.s associated with each type that may change with t (after a decision has been made). However, although lifespan distribution may change their parameters they must continue to be exponential. The state of the process at t is also called the "generation" or the "population" at t. The state of the process at t is given by $\boldsymbol{S}_t = (S_{1,t}, S_{2,t}, \dots, S_{k,t})^\top$, where all $S_{i,t}$ are with values in $\mathbb{N}_0$ and $S_{i,t}$ is the (non-negative) number of particles of type i that exist at t. The initial state $\boldsymbol{S}_0$ is deterministic.

5. The chosen decision $\boldsymbol{x}_t$ affects the lifespan distributions (however the distributions remain exponential), the distributions for the number of particles in the offspring, and the probabilities for mutation within the offspring, of all particles that exist at t. Thus, the types of all particles that exist at t are modified as a consequence of $\boldsymbol{x}_t$. Particles that exist at t can create only particles that are of the modified types, hence only the modified types are being propagated until $t+1$.

   (a) Corresponding to the i-th coordinate of $\boldsymbol{x}_t$ is the row vector $\boldsymbol{m}_i(x_{i,t}) = \big(m_{i1}(x_{i,t}), \dots, m_{ik}(x_{i,t})\big)$, where $m_{ij}(x_{i,t})$ denotes the expected number of type j particles at $t+1$ within a MBHBPM with exponential lifespan distributions starting at t with a single particle of type i under decision $x_{i,t}$. We assume that $m_{ij}(x_{i,t}) < \infty$ for all $x_{i,t} \in \widetilde{\mathcal{X}}_i$ and $i, j = 1, \dots, k$. Given $\boldsymbol{x}_t \in \mathcal{X}$, we organize the expectations into matrix $M(\boldsymbol{x}_t) = (\boldsymbol{m}_1(x_{1,t}), \dots, \boldsymbol{m}_k(x_{k,t}))^\top$.

Proposition 3.5. *SDP Model 2 is a SDP within the "Universal Modeling Framework".*

Proof. Similarly to SDP Model 1, SDP Model 2 considers a class of decisions at each t that modify, without dependence from states and decisions prior to t, all of the underlying distributions associated with each particle type. Lifespan distributions continue to be exponential, however, their parameters can change. Under these circumstances, for a fixed π, we can duplicate the proof of Proposition 3.4 for the case of SDP Model 2, with the following exceptions: 1) The offspring distribution among types associated with type i, $i = 1, \dots, k$, is now written via probabilities of mutation, i.e., $p^i_{\boldsymbol{\alpha}} = p_{i\hat{k}} \frac{\hat{k}!}{\alpha_1! \dots \alpha_k!} u_{i1}^{\alpha_1} \dots u_{ik}^{\alpha_k}$ (see the discussion after Definition 2.1); 2) After implementing a decision at decision epoch t, the distributions associated with the process change; 3) If a particle existing at t survives until $t+1$, the argument is restarted with respect to $t+1$. Consequently, for every fixed π, SDP Model 2 can be viewed as a (possibly non-stationary) discrete-time Markov chain with respect to its transition function and $t = 0, 1, \ldots, T$. $\square$

Since SDP Model 2 is a truly a SDP within the "Universal Modeling Framework", similarly to Section 3.4.2, we can formally characterize the transition function via $\mathbb{P}(\boldsymbol{Z}_{t+1} = \boldsymbol{S}_{t+1} | \boldsymbol{Z}_t = \boldsymbol{S}_t, X_t^\pi(\boldsymbol{S}_t) = \boldsymbol{x}_t)$. Further, we can now utilize Bellman's optimality equation (3.2).

Evidently, an analogous SDP can be defined for the case of the standard multi-type Bellman-Harris branching process with exponential lifespans as well (we use the standard $p^i_{\boldsymbol{\alpha}}$ instead of $p_{i\hat{k}} \frac{\hat{k}!}{\alpha_1! \ldots \alpha_k!} u_{i1}^{\alpha_1} \ldots u_{ik}^{\alpha_k}$).

3.5.2 Solution of SDP Model 2

As can be seen from the definition, SDP Model 2 is very similar to SDP Model 1. Indeed, Theorem 3.2 can be replicated in the context of SDP Model 2.

Theorem 3.3. *For SDP Model 2, the value function $V_t(\boldsymbol{S}_t)$ satisfies*

$$V_t(\boldsymbol{S}_t) = \boldsymbol{S}_t^\top \mathcal{R}^{T-t} \boldsymbol{c}_T, \qquad t = 0, 1, \ldots, T-1. \tag{3.16}$$

Policy π, with corresponding $X_t^\pi(\cdot)$, that computes decisions $\boldsymbol{x}_t$ satisfying

$$\boldsymbol{c}(\boldsymbol{x}_t) + \gamma M(\boldsymbol{x}_t) \mathcal{R}^{T-t-1} \boldsymbol{c}_T = \mathcal{R}^{T-t} \boldsymbol{c}_T, \qquad t = 0, 1, \ldots, T-1, \tag{3.17}$$

is optimal.

Proof. Due to the memorylessness property of the exponential distribution, we can consider all particles that exist at t as having age 0. This and the assumption of independent evolution of particles allow us to use matrices $M(\boldsymbol{x}_t)$ regardless of the state of the system. The rest of the proof of Theorem 3.3 is completely analogous to the proof of Theorem 3.2, with the exception that, instead of a BGW branching process, we consider a MBHBPM with exponential lifespan distributions. $\square$

Remark 3.9. *When regard to the steps necessary for finding a solution for SDP Model 2, the situation is analogous to Remark 3.8 valid for SDP Model 1. There is, however, an additional complication, namely computing $M(\boldsymbol{x}_t) = (\boldsymbol{m}_1(x_{1,t}), \ldots, \boldsymbol{m}_k(x_{k,t}))^\top$ for the MBHBPM with exponential lifespan distributions. Fortunately, we can do all necessary computations* *via Numerical Scheme 1 applied to the appropriate p.g.f.* $F_i(t;\boldsymbol{s})$*. Recall Notational Choice 1, more specifically that decision epochs can have varying distance from one another. Recall also that within Numerical Scheme 1 in order to compute the values at* t *we have to compute the values (that are on the grid) before* t*. Hence, we can save computational time by computing necessary quantities only once for the largest distance between any two neighboring decision epochs.*

Remark 3.10. *Due to the correspondence between the multi-type Bellman-Harris branching process and the MBHBPM it is clear that for a SDP with dynamics based on the multi-type Bellman-Harris branching process with exponential lifespan distributions a result analogous to Theorem 3.3 will hold.*

3.6 SDPs with underlying MSBPM dynamics

Throughout this Section, we will consider that Remark 3.6 is in effect. Also, we keep Remark 3.7 and Notational Choice 1 in mind.

As we saw within the proof of Proposition 3.4, due to the memorylessness property of the exponential lifespan distribution, we were able to consider all particles at t as having age 0, thus their remaining lifespan was not affected by their true age. If we drop the assumption that lifespan distributions are exponential, the remaining lifespan of particles that exist at t depends on their lifespan distribution as well as their respective age. In the context of the usual definition for states of a branching process, the states being defined as the number of particles, per type, that exist at t, the information within the state (and its associated distributions) at t is not sufficient for a transition function to produce the state at $t+1$. Indeed, so defined states, that do not account for the age of the particles that exist at $t = 0, 1, \ldots, T$, cannot fully model by themselves the dynamics between any two t and $t+1$ as this dynamics (more specifically the remaining lifespan of particles that exist at t) depends on the ages of the particles that exist at t. Considering a MSBPM instead of a MBHBPM, not only the remaining lifespan of particles depends on their age at t but also their reproductive capabilities depend on the moment of death of a particle, where the moment of death itself is influenced by the remaining lifespan distribution. A proper transition function between t and $t+1$ has to take into account the ages of particles that exist at t and since any fixed, in terms of number of particles per type, state at t can have many realizations with respect to the ages of the individual particles, it follows that, with respect to the "Universal Modeling Framework", we have to redefine the state space in order to be able to construct a SDP where the Markov property holds.

Our contributions within this Section are: 1) We construct a novel state space and show that, with respect to it, the multi-type Sevastyanov branching process, as well as the MSBPM, constitute discrete-time Markov chains with respect to $t = 0, 1, \dots, T$; 2) For these processes, we construct SDPs within the "Universal Modeling Framework". The contributions of this Section have not been published yet.

Unfortunately, unlike Section 3.4 and Section 3.5, there is no theorem that facilitates an efficient way for finding the solution of a MSBPM (or multi-type Sevastyanov branching process) based SDP. However, the fact that we can utilize, by virtue of Theorem 3.1, Bellman's optimality equation (3.2), allows us to consider approaches such as Approximate Dynamic Programming (ADP) and Reinforcement Learning (RL) for finding the solution of interest. We leave the development of a specialized ADP (or RL) algorithm for future research.

We highlight that stochastic SDPs (and generally stochastic problems) are one of the most difficult optimization problems within the field of optimization. Nice and compact results are seldom available and different problems may require specialized algorithms solely designed for them. The mere existence of the 15 fragmented communities, discussed in Section 1.5, that deal with sequential decision problems, testifies to the lack of an overarching approach that can handle a sufficiently large class of problems. In this context, the successful incorporation of the MSBPM (and other branching processes) into SDPs within the "Universal Modeling Framework" is a significant development, as it provides us with Bellman's optimality equation as a possible tool to be used when searching for solutions.

3.6.1 Definition of SDP Model 3 - SDP with underlying MSBPM dynamics

We now consider a general MSBPM where the lifespan distributions need not be exponential and reproductive capabilities of particles can depend on their age at their moment of death. In this case the ages of the particles that exist at t influence the dynamics of the system - in addition to the lifespan distribution and offspring distribution, among types, associated with each particle type, we also need the remaining lifespan of the particles that exist at t, which can be obtained by further knowing the ages of the particles at t. We construct a novel state space to be used for defining a SDP with either multi-type Sevastyanov branching processes or MSBPM dynamics within the "Universal Modeling Framework". To the best of our knowledge, multi-type Sevastyanov branching processes, with a state space that holds information about particle age, have not been discussed previously.

Definition 3.5. *Let there be k types. For each type denote with $\mathcal{D}_i$ the set of all 2-tuples of the following form:*

1. *The first element of the tuple is an integer. We denote this integer with r, $r \in \mathbb{N}_0$.*
2. *The second element of the 2-tuple is a r-tuple. We denote this r-tuple with $\boldsymbol{l}$. Each element l_i of $\boldsymbol{l}$ is a non-negative real number, i.e., $l_i \in \mathbb{R}_+$. The numbers within $\boldsymbol{l}$ are ordered from smallest to largest. Duplication is allowed in which case duplicating numbers are written next to each other.*

At t, associate with each $\mathcal{D}_i$ probability distributions. Following Remark 3.6, we will not write these distributions explicitly, but will consider them implicitly known. Denote $\mathcal{D}_i$ with associated distributions at t as $\mathcal{D}_{i,t}$. For the collection of k types at t, denote $\mathcal{D}_t^k = \mathcal{D}_{1,t} \times \mathcal{D}_{2,t} \times \cdots \times \mathcal{D}_{k,t}$.

Evidently, each element $d \in \mathcal{D}_{i,t}$ can be interpreted as carrying information about the number of particles of type i that exist at t (the first component of the 2-tuple) and also information about the age of each particle of type i that exists at t (the second component of the 2-tuple). The ordering from smallest to largest number within the second component of the 2-tuple is in place to ensure that an actual state cannot be written in more than one way. Note that the distributions associated with $\mathcal{D}_{i,t}$ can be the same for each t.

With respect to $\mathcal{D}_t^k$ and $t = 0, 1, \ldots, T$, we can formally write the multi-type Sevastyanov branching process, as well as the MSBPM, as $\boldsymbol{Z}_t^{\mathcal{D}_t^k} = \left(Z_{1,t}^{\mathcal{D}_{1,t}}, \ldots, Z_{k,t}^{\mathcal{D}_{k,t}}\right)^\top$, where $Z_{i,t}^{\mathcal{D}_{i,t}}$ denotes the number of particles of type i, as well as their age, that exist at t.

Proposition 3.6. *A multi-type Sevastyanov branching process, with states defined as the elements of $\mathcal{D}_t^k$, is a discrete-time Markov chain with respect to $t = 0, 1, \ldots, T$.*

Proof. With respect to $t = 0, 1, \ldots, T$, we need to prove that the transition function between the state at t and the state at $t + 1$, for the multi-type Sevastyanov branching process, is conditionally independent from states prior to t and depends at most on the state at t, i.e., that $\mathbb{P}\Big(\boldsymbol{Z}_{t+1}^{\mathcal{D}_{t+1}^k} = \boldsymbol{s}_{t+1} | \boldsymbol{Z}_t^{\mathcal{D}_t^k} = \boldsymbol{s}_t, \ldots, \boldsymbol{Z}_0^{\mathcal{D}_0^k} = \boldsymbol{s}_0\Big) = \mathbb{P}\Big(\boldsymbol{Z}_{t+1}^{\mathcal{D}_{t+1}^k} = \boldsymbol{s}_{t+1} | \boldsymbol{Z}_t^{\mathcal{D}_t^k} = \boldsymbol{s}_t\Big)$.

Let us consider a multi-type Sevastyanov branching process starting at t with a single particle from an arbitrary type that has arbitrary age a. By virtue of the assumption of independent evolution of particles, conclusions drawn for such a process are applicable to the independent processes staring at t from each particle within the general multi-type Sevastyanov branching process with many particles, possibly with varying types and ages.

Following a specified lifespan distribution, the remainder of the particle's lifespan is given by the lifespan distribution of the particle conditioned on its age at t. The dynamics of a system that consists of this singular particle is completely determined, until the moment of death $\bar{t}'$ ($\bar{t}'$ does not index a decision epoch, it is a genuine point in time) of the particle, by its remaining lifespan distribution and its distribution of number of particles, per type, in the offspring (this distribution depends on the age of the particle at its moment of death).

From $\bar{t}'$ onward the system is described by a collection of independent multi-type Sevastyanov branching processes corresponding to each of the particles in the offspring, each process form the collection treats $\bar{t}'$ as its initial moment. Thus, the lifespan distributions, the distributions of number of particles, per type, in the offspring, together with the age of the particle that exists at t (all of those are encoded in or associated with the state at t that is an element of $\mathcal{D}_t^k$), describe all of the dynamics from t onward within a system consisting of this one particle of age a that exists at t.

It follows that the dynamics of a multi-type Sevastyanov branching process (with general continuous lifespan distributions) starting at t, possibly with multiple particles of various types and ages, is set by:

1. The lifespan distribution associated with type i, $i = 1, \ldots, k$.
2. The offspring distribution (this distribution depends on the age of a

particle at its moment of death) among types associated with type i, $i = 1, \ldots, k$.

3. For each particle that exists at t - the age a of the particle.

We must now prove conditional independence from states prior to t of the aforementioned components. By definition, the lifespan distributions for all particle types do not depend on any state. The specification itself of the distribution, for each particle type, of the number of offspring among types, also does not depend on any state. Since the age a of a particle at t is a component of the state at t and is thus considered known at t, it cannot be viewed, with respect to the transition function, as functionally dependent on states prior to t. Hence, the transition function between t and $t+1$ depends only on the state at t. $\square$

As the MSBPM corresponds to the multi-type Sevastyanov branching process (see the discussion below Definition 2.1), by virtue of Proposition 3.6, the MSBPM, as well as the DMSBPM, can also be viewed as a discrete-time Markov chain with respect to $\mathcal{D}_t^k$ and $t = 0, 1, \ldots, T$.

Next, we will define a SDP with underlying MSBPM dynamics. Similarly to Subsection 3.5.1, we first give the definition of the new SDP and then check that it is indeed a SDP within the "Universal Modeling Framework".

Definition of SDP Model 3. Define SDP Model 3 as the finite-horizon SDP that corresponds to the definition of SDP Model 1 upon which the following modifications are applied:

1. We observe a MSBPM, as defined in Definition 2.1, at epochs indexed with t, $t = 0, 1, 2, \ldots, T$.

2. Let k be the number of types of particles within the MSBPM. The state space is $\mathcal{D}_t^k$. The t index in $\mathcal{D}_t^k$ indicates that there are probability distributions and p.g.f.s associated with each type that may change with t (after a decision has been made). The state of the process at t is also called the "generation" or the "population" at t. The state of the process at t is given by $\boldsymbol{S}_t = (S_{1,t}, S_{2,t}, \ldots, S_{k,t})^\top$, where $S_{i,t} \in \mathcal{D}_{i,t}$ with the interpretation of the first component of $S_{i,t}$ being the number of particles of type i that exist at t and the interpretation of the second component of $S_{i,t}$ being the ages of each particle of type i that exists at t. The initial state $\boldsymbol{S}_0$ is deterministic.

5. The chosen decision $\boldsymbol{x}_t$ affects the lifespan distributions, the distributions for the number of particles in the offspring, and the probabilities for mutation within the offspring, of all particles that exist at t. Thus, the types of all particles that exist at t are modified as a consequence of $\boldsymbol{x}_t$. Particles that exist at t can create only particles that are of the modified types, hence only the modified types are being propagated until $t+1$. $\boldsymbol{x}_t$ does not affect the age of particles that exist at t.

In the context of Proposition 3.1 and Proposition 3.2, let us verify that for any fixed π SDP Model 3 is a (non-stationary) Markov chain with respect to its transition function.

Proposition 3.7. *SDP Model 3 is a SDP within the "Universal Modeling Framework".*

Proof. Similarly to SDP Model 1, SDP Model 3 considers a class of decisions at each t that modify, without dependence from states and decisions prior to t, all of the underlying distributions associated with each particle type. Decisions do not affect the age of particles that exist at t. Under these circumstances, for a fixed π, we can duplicate the proof of Proposition 3.6 for the case of SDP Model 3, with the following exceptions: 1) The offspring distribution among types associated with type i, $i = 1, \ldots, k$, is now written via probabilities of mutation, i.e., $p^i_{\boldsymbol{\alpha}}(a) = p_{i\hat{k}}(a)\frac{\hat{k}!}{\alpha_1!\ldots\alpha_k!}u_{i1}^{\alpha_1}\ldots u_{ik}^{\alpha_k}$ (see the discussion after Definition 2.1); 2) After implementing a decision at decision epoch t, the distributions associated with the process change; 3) If a particle existing at t survives until $t+1$, the argument is restarted with respect to $t+1$. Consequently, SDP Model 3, can be viewed as a (possibly non-stationary) discrete-time Markov chain with respect to its transition function, $\mathcal{D}^k_t$, and $t = 0, 1, \ldots, T$. $\square$

Since SDP Model 3 is a truly a SDP within the "Universal Modeling Framework", we can formally characterize the transition function via $\mathbb{P}\Big(\boldsymbol{Z}_{t+1}^{\mathcal{D}^k_{t+1}} = \boldsymbol{S}_{t+1} | \boldsymbol{Z}_t^{\mathcal{D}^k_t} = \boldsymbol{S}_t, X^\pi_t(\boldsymbol{S}_t) = \boldsymbol{x}_t\Big)$. Further, we have Bellman's optimality equation (3.2) at our disposal.

Evidently, an analogous SDP can be defined for the case of the standard multi-type Sevastyanov branching process as well (we use the standard $p^i_{\boldsymbol{\alpha}}(a)$ instead of $p_{i\hat{k}}(a)\frac{\hat{k}!}{\alpha_1!\ldots\alpha_k!}u_{i1}^{\alpha_1}\ldots u_{ik}^{\alpha_k}$).

Unfortunately, we cannot define for SDP Model 3 a matrix of expectations $M(\boldsymbol{x}_t)$ similar to the matrices of expectations within SDP Model 1 and SDP Model 2. This is because for SDP Model 1 and SDP Model 2 we were able to consider particles at t as having age 0 and from there we were also able to compute the expectation for each type given a $\boldsymbol{x}_t$. However, within SDP Model 3, we cannot consider the particles that exist at t as having age 0, they have an age that is encoded in the state $\boldsymbol{S}_t$. Consequently, the matrix of expectations depends not only on $\boldsymbol{x}_t$ but also on $\boldsymbol{S}_t$, i.e., we have $M(\boldsymbol{S}_t, \boldsymbol{x}_t)$. This circumstance breaks the proof of Theorem 3.2, thus we cannot replicate the theorem for the case of SDP Model 3.

3.7 Notes on Approximate Dynamic Programming (ADP)

Within this Section, we follow Subsection 9.4.5 (page 490) from [82], as well as Subsection 4.6.4 (page 138), Subsection 4.6.5 (page 139), and Subsection 4.8.3 (page 147), from [78].

3.7.1 Post-decision state variables and their connection with Bellman's optimality equation

Let us recall Definition 3.2, more specifically that a finite-horizon SDP is characterized by the sequence

$$(S_0, x_0, W_1, \ldots, S_t, x_t, W_t, \ldots, S_T). \tag{3.18}$$

Since state S_t is what we know just before we make a decision, x_t, we will refer to it as the *pre-decision state.*

In some settings, it is useful to model the state immediately after a decision is made. We denote the state at t under the effect of decision x_t as S_t^x. We call S_t^x the *post-decision state.* Including the post-decision state into our considerations we can rewrite the sequence given by (3.18) as

$$(S_0, x_0, S_0^x, W_1, S_1, x_1, S_1^x, W_2, S_2, x_2, S_2^x, \ldots, x_{T-1}, S_{T-1}^x, W_T, S_T).$$

Since there is no new exogenous information between making the decision x_t and the observation of the post-decision state S_t^x, the post-decision state is a deterministic function of the pre-decision state S_t and x_t.

Remark 3.11. *Within Section 3.4, Section 3.5, Section 3.6, we can see that the post-decision state $\boldsymbol{S}_t^{\boldsymbol{x}}$ for SDP Model 1, SDP Model 2, and*

SDP Model 3, consists of the original state at t, $\boldsymbol{S}_t$*, however, with new distributions associated with each particle type, that correspond to* $\boldsymbol{x}_t$*.*

Having defined the notion of post-decision state, we can consider decomposing the transition function $S^M(S_t, x_t, W_{t+1})$ into two steps, these steps being first

$$S_t^x = S^{M,x}(S_t, x_t) \tag{3.19}$$

and then

$$S_{t+1} = S^{M,W}(S_t^x, W_{t+1}). \tag{3.20}$$

The structure of $S^{M,x}(\cdot)$ and $S^{M,W}(\cdot)$, if such a decomposition is possible, generally is highly problem-dependent.

Remark 3.12. *For SDP Model 1, SDP Model 2, and SDP Model 3, this decomposition is possible. This follows from Remark 3.11 where it is clear that computing* $\boldsymbol{S}_t^{\boldsymbol{x}}$ *is trivial.*

With respect to the post-decision state variable S_t^x it is useful to introduce:

Definition 3.6. *Define* $V_t^x(S_t^x)$ *as the value of being in the post-decision state* S_t^x*, i.e.,*

$$V_t^x(S_t^x) = \mathbb{E}\big\{V_{t+1}(S_{t+1})|S_t^x\big\} \tag{3.21}$$

Let us explore the he relationship between $V_t(S_t)$ and $V_t^x(S_t^x)$ that arises from Definition 3.6. The following equations are true

$$V_{t-1}^x(S_{t-1}^x) = \mathbb{E}\big\{V_t(S_t)|S_{t-1}^x\big\}, \tag{3.22}$$

$$V_t(S_t) = \max_{x_t \in \mathcal{X}_t} \big(C_t(S_t, x_t) + \gamma V_t^x(S_t^x)\big), \tag{3.23}$$

$$V_t^x(S_t^x) = \mathbb{E}\big\{V_{t+1}(S_{t+1})|S_t^x\big\}. \tag{3.24}$$

We note that since S_t^x is a deterministic function of S_t and x_t, in (3.23), $V_t(S_t)$ is a solution of a deterministic optimization problem. Substituting (3.24) into (3.23) leads us back to the expectation form of Bellman's optimality equation (3.2)

$$V_t(S_t) = \max_{x_t \in \mathcal{X}_t} \Big(C_t(S_t, x_t) + \gamma \mathbb{E}\big\{V_{t+1}(S_{t+1})|S_t, x_t\big\}\Big).$$

On the other hand, substituting (3.23) into (3.22) provides us with

$$V_{t-1}^x(S_{t-1}^x) = \mathbb{E}\Big\{ \max_{x_t \in \mathcal{X}_t} \big(C_t(S_t, x_t) + \gamma V_t^x(S_t^x) \big) | S_{t-1}^x \Big\}, \tag{3.25}$$

It is important to note that the expectation in (3.25) is outside the "max" operator, hence (3.25) can be approximated via probabilistic techniques.

3.7.2 An ADP algorithm using post-decision state variables

In what follows, the core idea to update estimates of the expectation in (3.25) iteratively over a number of iterations, each iteration being a Monte Carlo simulation of the SDP we are interested in. With respect to (3.25), we denote the estimates of $V_{t-1}^x(S_{t-1}^x)$ as $\overline{V}_{t-1}^n(S_{t-1}^x)$, where n marks the number of the iteration we are currently on. We also use S_t^n, x_t^n, and $S_t^{x,n}$ to denote the state, decision, and post-decision state, during iteration n. We denote the value of the approximation of (3.23), that we will use at iteration n, with $\hat{v}_t^n$.

Approximate dynamic programming (ADP) algorithms conceptually revolve around the following steps (per each t) for each n:

1. Given $S_{t-1}^{x,n}$, simulate S_t^n. This step is motivated by the condition in (3.25).
2. Given S_t^n, use an approximation of deterministic problem (3.23) to obtain x_t^n and $\hat{v}_t^n$. This step is motivated by the fact that (3.23) is within the expectation in (3.25).
3. Given S_t^n, x_t^n, $\hat{v}_t^n$, and the old $\overline{V}_{t-1}^{n-1}(S_{t-1}^{x,n})$, use some rule to improve the existing approximation of the expectation in (3.25), i.e., update $\overline{V}_{t-1}^{n-1}(S_{t-1}^{x,n})$ to $\overline{V}_{t-1}^n(S_{t-1}^{x,n})$.

Within Step 2, given S_t^n, we have to solve the following deterministic approximation of (3.23)

$$\hat{v}_t^n = \max_{x_t \in \mathcal{X}_t} \big(C_t(S_t^n, x_t) + \gamma \overline{V}_t^{n-1}(S_t^{x,n}) \big), \tag{3.26}$$

which can be written more explicitly as

$$\hat{v}_t^n = \max_{x_t \in \mathcal{X}_t} \big(C_t(S_t^n, x_t) + \gamma \overline{V}_t^{n-1}(S^{M,x}(S_t^n, x_t)) \big).$$

Solving (3.26) gives us x_t^n and $\hat{v}_t^n$, where x_t^n corresponds to $\hat{v}_t^n$.

We can formally write the updating rule for Step 3 as

$$\overline{V}_{t-1}^n(S_{t-1}^{x,n}) \leftarrow U^V\left(\overline{V}_{t-1}^{n-1}(S_{t-1}^{x,n}), S_{t-1}^{x,n}, \hat{v}_t^n\right).$$

We will use the simple updating rule

$$\overline{V}_{t-1}^n(S_{t-1}^{x,n}) = (1-\alpha_{n-1})\overline{V}_{t-1}^{n-1}(S_{t-1}^{x,n}) + \alpha_{n-1}\hat{v}_t^n. \tag{3.27}$$

We outline the ADP algorithm on page 147 from [78]. We have chosen to present this ADP algorithm due to our observation in Remark 3.11. We keep Notational Choice 1 in mind.

Algorithm 3.2.

Step 0. *Set* $n = 1$*. For an initial State* S_0^1 *provide the value function approximation* $\overline{V}_t^0(S_t)$ *for all* S_t *and* t*.*

Step 1. *Choose a sample path* ω^n*.*

Step 2. *For* $t = 0, 1, 2, \ldots, T$*, do*

> ***Step 2a.*** *Optimization: Compute a decision* $x_t^n = X_t^\pi(S_t)$ *and find the post-decision state* $S_t^{x,n} = S^{M,x}(S_t^n, x_t^n)$*.*
>
> ***Step 2b.*** *Simulation: Find the next pre-decision state using* $S_t^n = S^M(S_t^n, x_t^n, W_{t+1}(\omega^n))$*.*

Step 3. *Update the value function approximation (e.g., via* (3.27)*) in order to obtain* $\overline{V}_t^n(S_t)$ *for all* t*.*

Step 4. *If we have not met our stopping rule, let* $n = n + 1$ *and go to step 1. Else terminate*

We note that Algorithm 3.2 is very basic and needs to be refined in order to be applicable in practice. For example, in Step 0 we have to compute $\overline{V}_t^0(S_t)$ for all S_t which is susceptible to curses of dimensionality. For the case of the MSBPM, we have to devise an appropriate strategy that allows us to compute/update multiple states simultaneously. Evidently, such a strategy is an approximation, however, this approximation can prove sufficient for practical purposes. For an in-depth exploration of ADP algorithms, the reader is referred to the entirety of [78].

We further note that the initial value function approximation plays an important role when applying Alogrithm 3.2 (and its subsequent refinements). More specifically, it can significantly influence the decisions we make and consequently the states we visit during the execution of the algorithm (a bad initial approximation can permanently lead us astray from the truly optimal decisions). Recall our results from Chapter 2 for processes starting with one particle of age a, $a \neq 0$. These results may prove useful for the purpose of constructing an appropriate $\overline{V}_t^0(S_t)$. Further, for the case of SDP Model 3, we can use the result from applying Theorem 3.3 for the corresponding SDP Model 2, that results from dropping the dependence of particle reproduction on particle age, as an ingredient for the initial value function approximation. A thorough methodology for constructing good value function approximations and obtaining quality solutions for SDP Model 3 remains to be developed in future work.

3.8 Example SDPs with branching process based dynamics

We illustrate concepts and results within this Chapter with a few example sequential decision problems. These examples are purely illustrative and are not connected to real-world data.

3.8.1 Informal description of the example model

We consider a population consisting of three types of particles - Bachelor's degree students (type 1), Master's degree students (type 2), and Ph. D. degree students (type 3). From now own we will use "individuals" instead of "particles". This population is a subset of the larger population of a country that we will refer to as the "environment". The attraction and sustenance of students does not exhaust the environment.

The three types of individuals are related to each other in the following way: 1) Bachelor's degree students can "produce" more Bachelor's degree students (type 1) as well as Master's degree students (type 2); 2) Master's degree students can produce either Master's degree students (type 2) or Ph.

D. degree students (type 3); 3) Ph. D. degree students can only produce Ph. D. degree students (type 3).

The "end of life" of an individual, i.e., the moment when "reproduction" occurs is to be understood as the end of an "agitation" cycle. Indeed, it takes time for the idea to start a Bachelor's degree to grow and bear fruit within an individual from the environment. It is also true that for the three types of students the shadow of doubt, whether they should quit or continue, reemerges periodically. The event of having zero offspring is to be understood as an individual either dropping out of the education system or obtaining the desired degree and not continuing with a pursuit of higher degrees. The event of producing one individual is to be understood as the same individual continuing education (in which case his/hers commitment is reaffirmed and the "lifespan" associated with the agitation cycle is restarted). The event of an individual producing multiple individuals within the system is to be understood as the individual continuing education and also attracting new individuals. "Mutation" probabilities u_{ij}, in this context, correspond to individuals changing the degree they try to obtain. When $i = j$ the individual continues with his/hers current degree, when $i \neq j$ the individual obtains the degree corresponding to type i and is now pursuing the degree that corresponds to type j. The, rather unlikely, event that an individual produces multiple mutants corresponds to that individual being able to draw individuals that can pursue the corresponding degree from within the environment back into the education system.

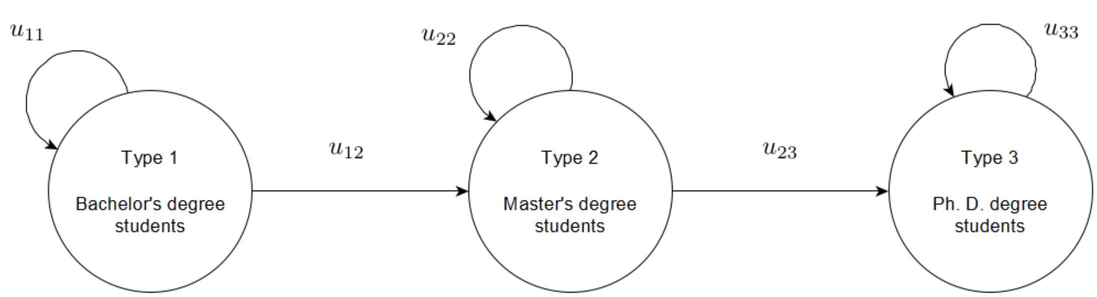

Figure 3.1: Mutation scheme valid for the example SDP.

The government can regulate the population of students via different advertisement strategies. Different media and events have different impact on different types of students as well as different cost. For example social media can be most impactful with respect to the spread of the ideas of Bachelor's degree students towards young people within the environment.

On the other hand conferences and conventions can be most motivating for Ph. D. students to continue pursuing their degree. The government needs to balance its limited budget throughout time so that it minimizes its costs and still achieve a target objective. For our illustrative purposes, we will consider that the government only incurs costs for implementing chosen advertisement policies, however, at the terminal, with respect to our considerations, moment in time the European Union provides financing for each active Ph. D. student (the European Union presumably considers the number of active Ph. D. students as an indicator for the well being of the institutions within the country).

The above example model, although sounding somewhat realistic, has many fallacies with respect to properly reflecting all intricacies of its real-world counterpart. For example, the real-world limitation that each degree is to be obtained within a specified amount of time is not well defined within our model. Another example is the real-world circumstance that even if there are no individuals left in the education system it is still possible for Bachelor's degree students to enroll in it. Regardless, this model can be used as a starting point for an exploration of the topic. The model is open for upgrades and refinements, as an example we could consider controlled branching process for providing a mechanism for removing individuals after a certain period of time, or a branching process with immigration in order to handle the case of zero Bachelor's degree students in the educational system. Such extensions, however, need to be incorporated into a proper definition of a SDP within the "Universal Modeling Framework" in order to be able to utilize Bellman's optimality equation.

We note that our implementation of the result of Theorem 3.2 and Theorem 3.3, as well as the application of Numerical Scheme 1 and Numerical Scheme 2 to all relevant, within this example, quantities from Chapter 2, is done in Python 3.8.13 [209] by using the NumPy 1.20.3 [210] and SciPy 1.6.2 [211] libraries.

3.8.2 SDP with underlying Bienaymé-Galton-Watson branching processes dynamics

Let us assume that the underlying dynamics is set by a Bienaymé-Galton-Watson (BGW) branching processes. This assumption is reasonable if the government collects statistics and reconsiders its advertisement strategy once per year (or some other time interval). BGW branching processes are also in line the fact that students are (usually) accepted into the system only once per year. By considering BGW, we equate the completion of an "agitation" cycle with the arrival of a new decision epoch.

The tables below give the correspondence between the decision of the government to implement a particular advertisement strategy ("Decision i" corresponds to implementing "strategy i") towards a type of individuals and the effects of that decision on the individuals.

We have followed certain rules when building these tables (again, this example is purely illustrative and is not based on real-world data). For each type we have decomposed the probability of producing 0 individuals into two components. The first component represent the, presumably, standard rate of individuals either obtaining a degree and dropping from the education system or simply dropping. The second component is the effect from applying a certain strategy on the rate of dropping out of the system regardless of whether a degree has been obtained. Evidently less funding from the government leads to greater dropping rates. We have assumed that the rate of "mutation" of degrees pursued remains fixed regardless of the actions of the government, i.e., an individual determined to continue education completes his/hers degree and advances to the next degree with an unchanged pace.

The following tables utilize some of the notation used within SDP Model 1

$x_{1,t}$	$c_1(x_{1,t})$	p.g.f. for total offspring in a generation	Expected offspring in a generation	u_{11}	u_{12}	u_{13}
Decision 1	-1.6	$(0.45 + 0.00)s^0 + 0.15s^1 + 0.20s^2 + 0.20s^3$	1.15	0.70	0.30	0.00
Decision 2	-1.1	$(0.45 + 0.15)s^0 + 0.20s^1 + 0.15s^2 + 0.05s^3$	0.65	0.70	0.30	0.00
Decision 3	-0.7	$(0.45 + 0.30)s^0 + 0.15s^1 + 0.10s^2 + 0.00s^3$	0.35	0.70	0.30	0.00

Table 3.1: Effects from implementing possible decisions for type 1 individuals.

$x_{2,t}$	$c_2(x_{2,t})$	p.g.f. for total offspring in a generation	Expected offspring in a generation	u_{21}	u_{22}	u_{23}
Decision 1	-1.1	$(0.30+0.10)s^0+0.60s^1$	0.6	0.00	0.85	0.15
Decision 2	-2.56	$(0.30+0.00)s^0+0.70s^1$	0.7	0.00	0.85	0.15
Decision 3	-1.8	$(0.30+0.10)s^0+0.60s^1$	0.6	0.00	0.85	0.15

Table 3.2: Effects from implementing possible decisions for type 2 individuals.

$x_{3,t}$	$c_3(x_{3,t})$	p.g.f. for total offspring in a generation	Expected offspring in a generation	u_{31}	u_{32}	u_{33}
Decision 1	-0.5	$(0.15+0.30)s^0+0.55s^1$	0.55	0.00	0.00	1.00
Decision 2	-2.1	$(0.15+0.20)s^0+0.65s^1$	0.65	0.00	0.00	1.00
Decision 3	-3.3	$(0.15+0.00)s^0+0.85s^1$	0.85	0.00	0.00	1.00

Table 3.3: Effects from implementing possible decisions for type 2 individuals.

We consider four decision epochs (the counting starts from "Epoch 0") before the final epoch at which we collect the terminal reward, i.e. $T=4$. The terminal reward is given by $(0,0,325)$.

The discount factor is $\gamma=0.65$. The discount factor indicates that the government has more pressing spending at Epoch 0 and its near future as opposed to waiting for extra funds coming in the far future.

Following SDP Model 1 and considering that the initial state is deterministic, we have to find

$$\max_{\pi\in\Pi}\mathbb{E}\left\{\sum_{t=0}^{T-1}\gamma^t C_t(\boldsymbol{S}_t, X_t^\pi(\boldsymbol{S}_t))+\gamma^T C_T(\boldsymbol{S}_T)|\boldsymbol{S}_0\right\}.$$

At each decision epoch we have 27 different decision tuples, each decision tuple has its corresponding matrix of expectations $M(\boldsymbol{x}_t)$. We omit writing here these matrices, however we note that for the BGW they can be obtained via the p.g.f. of the process. The 27 possible decision tuples amount to $27^4 = 531\,441$ possible policies π. By applying Theorem 3.2, we only have to check the 27 possible decision tuples with respect to the optimal return operator, once per each decision epoch. This means that the application of Theorem 3.2 not only allows us to not iterate all possible states, but also reduces the necessary computations quite dramatically. In addition to that, the optimal policy does not depend on the initial state $\boldsymbol{S}_0$.

The optimal policy obtained via Theorem 3.2 is (the i-th coordinate corresponds to type i, the number at this coordinate is the number of the decision) Epoch 0: (1, 1, 3), Epoch 1: (1, 2, 3), Epoch 2: (1, 2, 3), Epoch 3: (3, 2, 3). Not surprisingly, as decision 1 for type 1 is the only decision in the problem that leads to supercritical reproduction, we chose it throughout Epoch 0 - 2. At Epoch 3, since there can be no direct production from type 1 towards type 3, it is best to cut the funding for Bachelor's degree students as much as possible. We also see that, in this setting, it is always best to spend the maximum amount on Ph. D. students.

As the optimal policy does not depend on the initial state $\boldsymbol{S}_0$, we can construct the optimal discounted reward associated with the optimal policy for any initial state $\boldsymbol{S}_0$ by appropriately summing the optimal discounted rewards corresponding to processes that start with one individual from type $i = 1, 2, 3$. In our case the optimal discounted rewards per individual are: $(2.5761181, 5.21493921, 23.59679382)$.

3.8.3 SDP with underlying exponential lifespan MB-HBPM dynamics

For the model described in Subsection 3.8.1, we can consider a MBHBPM with exponential lifespan distributions. In this case we can interpret decision epochs as the moments at which the individuals enroll in the education system officially and the government registers them. However, in the meantime, individuals may experience multiple "agitation" cycles leading to multiple chances for dropping from the system as well as multiple chances for drawing individuals from the environment into starting a degree. We consider that individuals, even if they are not yet officially enrolled into the education system but have been successfully agitated, start behaving as type i individuals (i.e., production of individuals from the three types that we consider can occur between decision epochs). We note that in this continuous setting it is perfectly possible (and rather unlikely) for a Bachelor's degree student to ultimately produce, e.g., two Ph. D. students between two neighboring decision epochs. We assume all such possibilities are reflected appropriately within the probabilities for the total number of offspring as well as the probabilities for mutation.

In the context of SDP Model 2 (keeping Notational Choice 1 in mind), we consider again Table 3.1, Table 3.2, and Table 3.8.2 from the previous Subsection 3.8.2.

We will again consider four decision epochs (the counting starts from "Epoch 0") before the final epoch at which we collect the terminal reward, i.e. $T = 4$. The terminal reward is again $(0, 0, 325)$. The discount factor is once more $\gamma = 0.65$.

Under $Exp(\lambda)$, we understand the Exponential distribution with probability density function (p.d.f.) of $f(x; \lambda) = \frac{1}{\lambda}e^{-x/\lambda}$, $x \in [0, \infty)$, $\lambda > 0$. We use the following lifespan distributions with respect to decision made:

Decisions for type 1	Lifespan distribution	Expected lifespan	Variance
Decision 1	$Exp(4)$	4	16
Decision 2	$Exp(7)$	7	49
Decision 3	$Exp(8)$	8	64

Table 3.4: Correspondence between lifespan distribution and decision for type 1.

Decisions for type 2	Lifespan distribution	Expected lifespan	Variance
Decision 1	$Exp(7)$	7	49
Decision 2	$Exp(4.5)$	4.5	20.25
Decision 3	$Exp(8)$	8	64

Table 3.5: Correspondence between lifespan distribution and decision for type 2.

Decisions for type 3	Lifespan distribution	Expected lifespan	Variance
Decision 1	$Exp(9)$	9	81
Decision 2	$Exp(8)$	8	64
Decision 3	$Exp(7)$	7	49

Table 3.6: Correspondence between lifespan distribution and decision for type 3.

Following SDP Model 2 and considering that the initial state is deterministic, we have to find

$$\max_{\pi\in\Pi} \mathbb{E}\left\{ \sum_{t=0}^{T-1} \gamma^t C_t(\boldsymbol{S}_t, X_t^\pi(\boldsymbol{S}_t)) + \gamma^T C_T(\boldsymbol{S}_T)|\boldsymbol{S}_0 \right\}.$$

Despite reusing Subsection 3.8.2, the SDP from Subsection 3.8.2 and the SDP from the current section are not directly comparable because the exponential lifespan distributions need to be appropriately scaled. We will not attempt to do a scaling here and for our illustrative purposes, we assume that any two neighboring decisions epochs are spaced 10 time units apart from one another.

At each decision epoch we again have 27 different decision tuples, each decision tuple has its corresponding matrix of expectations $M(\boldsymbol{x}_t)$. However, recalling Remark 3.9, we have to utilize Numerical Scheme 1 and equation (2.3) from Subsection 2.2.2 in order to compute the 27 different matrices $M(\boldsymbol{x}_t)$. We omit writing down all of the 27 expectation matrices here, but we note that we have used $h = 10^2$ and $\Delta s_j = 10^{-8}$ when computing them via equation (2.3) from Subsection 2.2.2.

The optimal policy obtained via Theorem 3.3 is (the i-th coordinate corresponds to type i, the number at this coordinate is the number of the decision) Epoch 0: (1, 1, 3), Epoch 1: (1, 2, 3), Epoch 2: (1, 2, 3), Epoch 3: (1, 2, 3). The optimal discounted rewards per individual are: $(7.7624472, 4.28270503, 18.10259452)$. Although we should be very careful about comparing the total values of the individual optimal discounted rewards yielded by the solutions of the SDP from Subsection 3.8.2 and the current SDP, we can still make the following observation. For the current SDP, the individual optimal discounted reward for type 1 individuals has improved considerably when compared those for type 2 and type 3 individuals. This is not surprising. As the expected lifespan for all types, regardless of government decision, is smaller that the distance of 10 time units between any two neighboring decision epochs, within the current SDP there are more opportunities for Master's degree and Ph. D. degree students to leave the system in comparison to the SDP from Subsection 3.8.2. On the other hand there are also more opportunities for the Bachelor's degree students to draw people from the environment.

3.8.4 Notes on SDPs with underlying MSBPM dynamics

With respect to the setting of Subsection 3.8.3 we can change the lifespan distributions to be different from the exponential distribution. This can make the model more adequate as it is likely that the behavior of individuals cannot be described well with only one kind of probability distribution. From the figures within Chapter 2, we have seen that for times close to the start of a MSBPM, there can be significant difference in the behavior of the process when comparing the case of the initial particle having age 0 and the case of the initial particle having age a, $a \neq 0$. As we have set the distance between any two neighboring decision epochs to be 10 time units, we can expect significant difference in the evolution of the system.

Unfortunately, as there is no analogue of Theorem 3.2 or Theorem 3.3 for the case of MSBPM, we have no easy way to obtain a solution for the corresponding SDP. Fortunately, we have proven that the MSBPM can be incorporated into SDPs within the "Universal Modeling Framework", which allows us to use Bellman's optimality equation in our considerations. This formally opens the gate for Approximate Dynamic Programming (ADP; see [78]) algorithms, which are often model-based, as well as Reinforcement Learning (RL; see [82], [203]) algorithms, which are often model-free. With respect to ADP, we can attempt developing Algorithm 3.2 outlined in Subsection 3.7.2. For this purpose, we can try obtaining an initial value function approximation via the results from Chapter 2, or we can take as an initial value function approximation the solution of the corresponding SDP Model 2. We can also attempt developing a RL algorithm first and then try to refine it with APD. In all of these cases, we will need to be able to simulate the evolution of the system. The development of ADP and RL algorithms is left for future research.

Conclusion

Approbation

Results from the dissertation have been presented at: FMI Spring Scientific Session (March 2019, 2021, Sofia, Bulgaria), National Seminar on Probability and Statistics (June 2019, Sofia, Bulgaria), 21st European Young Statisticians Meeting (29 July - 02 August 2019, Belgrade, Serbia), Sofia University Young Researchers Conference (February 2020, Sofia, Bulgaria), The 19th Conference of the Applied Stochastic Models and Data Analysis International Society ASMDA2021 and DEMOGRAPHICS2021 WORKSHOP (June 2021, Athens, Greece), The 5th International Workshop on Branching Processes and their Applications - IWBPA 2021 (April 2021, Badajoz, Spain).

The following publications were written during the writing of the dissertation:

1. M. Slavtchova-Bojkova, K. Vitanov. Modelling cancer evolution by multi-type age-dependent branching processes. Comptes rendus de l'Acade'mie bulgare des Sciences, **71**, 10, 1297-1305, (2018).
2. M. Slavtchova-Bojkova, K. Vitanov. Multi-type age-dependent branching processes as models of metastasis evolution. Stochastic Models, **35**, 284-299, (2019), https://doi.org/10.1080/15326349.2019.1600410.
3. K. Vitanov, M. Slavtchova-Bojkova. On decomposable multi-type Bellman-Harris branching process for modeling cancer cell populations with mutations. 21st European Young Statisticians meeting - Proceedings, 113-118, (2019).
4. K. Vitanov, M. Slavtchova-Bojkova. Modeling escape from extinction with decomposable multi-type Sevastyanov branching processes.

Stochastic Models, (2022), https://doi.org/10.1080/15326349.2022.2041037.

Scientific contributions

Within this dissertation the novel Multi–type Sevastyanov Branching Processes through probabilities of Mutation between types (MSBPM) is developed and explored in the context of populations escaping extinction. Unlike previous works in the field, the MSBPM and the results obtained do not depend on assumptions about mutations being small quantities or particular lifespan distributions nor on assumptions of non-decomposability or particular reproduction rates. As such, the novel MSBPM and the associated novel results constitute a continuous-time extension and/or generalization of previously obtained results by other authors concerning populations escaping extinction (see, e.g., [61], [62], [64], [65]) as well as a continuation of our previous results in the same field within Vitanov & Slavtchova-Bojkova [7] (2022) as well as preceding papers [1] - [6]. Various systems of equations have been obtained - systems of equations for the probability generating functions (p.g.f.s) of the process, for the probabilities of extinction, for the p.g.f.s of particle production from one class of particle types to another. Results concerning the time until occurrence of the first "successful" particle as well as the immediate risk of a "successful" particle emerging have also been obtained. To the best of our knowledge, such an in-depth investigation of the topic has not been done previously for multi-type, continuous-time branching processes (excluding our earlier work in [7] as well as preceding papers [1] - [6]). Aforementioned results have been obtained for the case of the MSBPM starting with one particle of age 0 and for the case of the MSBPM starting with one particle of age a, $a \neq 0$. The latter case, to the best of our knowledge, has not been explored previously in a systematic manner within the context of branching processes. Particular cases of decomposable MSBPMs have also been considered in the manner described above. Numerical schemes for calculating all obtained systems of equations have been developed.

Multi-type Bienaymé-Galton-Watson (BGW) branching processes, multi-type Bellman-Harris branching process with exponential lifespan distributions, multi-type Sevastyanov branching process, as well the MSBPM and its variants, have been successfully incorporated into Sequential Decision Problems (SDPs) within the "Universal Modeling Framework" developed in [82]. To the best of our knowledge, with the exception of the

BGW branching process, branching processes have not been considered in the context of SDPs (within the “Universal Modeling Framework” or in other modeling frameworks). This incorporation formally opens the gate for techniques such as Approximate Dynamic Programming (ADP) and Reinforcement Learning (RL) to be applied onto SDPs with underlying branching process based dynamics. A novel proof of Theorem 3.1 from [77], concerning an efficient algorithm for finding the solution of SDPs with multi-type BGW branching processes dynamics, that uses Bellman’s optimality equation, has been obtained. An analogous novel result for the case of the multi-type Bellman-Harris branching process with exponential lifespan distributions, as well as for the case of the Multi-type Bellman-Harris Branching Process through probabilities of Mutation between types (MBHBPM) with exponential lifespan distributions, has been identified. A novel state space has been constructed for the purpose of successfully incorporating the MSBPM and the multi-type Sevastyanov branching process into SDPs within the “Universal Modeling Framework”.

Note regarding used software

All computations within the dissertation are done via code written in Python 3.8.13 [209]. The code uses the NumPy 1.20.3 [210] and SciPy 1.6.2 [211] libraries. Figures, that do not contain nodes, are created with Matplotlib 3.5.1 [212]. Figures that contain nodes are created with yEd 3.20.1 [213].

Declaration for originality of obtained results

I declare that the current dissertation “Branching processes - optimization and applications” contains original results obtained as a product of my research (supported by my supervisor). Results that have been obtained, published, or described by other authors are appropriately cited within the Bibliography.

The dissertation has not been applied for the purpose of obtaining a scientific degree from another school, university, or scientific institute.

Gratitude

I would like to thank my supervisor, prof. Maroussia Slavtchova-Bojkova, for our discussions, fruitful exchange of ideas, and support throughout the years! I am also grateful to my colleagues at the Faculty of Mathematics and Informatics of Sofia University "St. Kliment Ohridski" for always being kind and helpful!

Appendix

A.1. Perron–Frobenius theorem

In this section we present without proof basic results about non-negative and positive square matrices. For what follows, we have used Chapter 17 from [197].

We say a real matrix A is *non-negative* and write $A \geq 0$, if $a_{ij} \geq 0$ for all entries of A. We say A is *positive* and write $A > 0$ when $a_{ij} > 0$ for all entries of A.

Theorem 1. *(Perron) Let A be a $n \times n$ positive matrix. The the following hold.*

1. $\rho = \rho(A) > 0$ *and* $\rho(A) \in spec(A)$.
2. *There is a positive eigenvector* $\boldsymbol{p}$ *such that* $A\boldsymbol{p} = \rho\boldsymbol{p}$. *Furthermore, any other positive eigenvector is a multiple of* $\boldsymbol{p}$.
3. *The eigenvalue ρ is a simple eigenvalue - i.e., it has algebraic multiplicity 1. Consequently, if* $A\boldsymbol{v} = \rho\boldsymbol{v}$, *then* $\boldsymbol{v}$ *is a scalar multiple of* $\boldsymbol{p}$.
4. *If λ is an eigenvalue of A and $\lambda \neq \rho$, then $|\lambda| < \rho$.*
5. *Let* $B = \frac{1}{\rho}A$, *so that* $\rho(B) = 1$. *Then* $\lim_{k\to\infty} B^k = L$ *exists and has rank* 1. *Each column of L is a positive multiple of* $\boldsymbol{p}$.
6. *Let* $\boldsymbol{x} \geq \mathbf{0}$, $\boldsymbol{x} \neq \mathbf{0}$. *Then* $\lim_{k\to\infty} B^k\boldsymbol{x} = L\boldsymbol{x}$ *is a positive multiple of* $\boldsymbol{p}$.

General non-negative matrices provide us with more scarce results:

Theorem 2. *(Perron–Frobenius) If A is an $n \times n$, non-negative matrix with spectral radius ρ. Then the following hold.*

1. *The number ρ is an eigenvalue of A.*
2. *There exists a non-zero vector $\boldsymbol{p} \geq \mathbf{0}$ such that $A\boldsymbol{p} = \rho\boldsymbol{p}$.*
3. *There exists a non-zero vector $\boldsymbol{y} \geq \mathbf{0}$ such that $\boldsymbol{y}^\top A = \rho\boldsymbol{y}^\top$.*

Now, we concentrate our attention on the class of "irreducible" non-negative square matrices.

Definition 1. *An $n \times n$ matrix A is said to be reducible if there exists a permutation matrix P such that $P^\top AP = \begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22}, \end{pmatrix}$ where A_{11} is $k \times k$, A_{22} is $(n-k) \times (n-k)$, and $1 \leq k \leq n-1$. If A is not reducible, we say A is irreducible.*

The following theorem is useful for checking irreducibility:

Theorem 3. *Let A be an $n \times n$ non-negative matrix. Then A is irreducible if and only of $(I + A)^{n-1} > 0$.*

We arrive at the Perron–Frobenius theorem:

Theorem 4. *(Perron–Frobenius) If A is an $n \times n$, non-negative, irreducible matrix, then the following hold.*

1. *The spectral radius $\rho = \rho(A)$ of A is positive and is an eigenvalue of A.*
2. *The eigenvalue ρ is a simple root of the characteristic polynomial $p_A(x)$ of A.*
3. *There is a positive vector $\boldsymbol{p}$ such that $A\boldsymbol{p} = \rho\boldsymbol{p}$. Furthermore, $A\boldsymbol{x} = \rho\boldsymbol{x}$ if and only if $\boldsymbol{x}$ is a scalar multiple of $\boldsymbol{p}$.*
4. *There is a positive vector $\boldsymbol{q}$ such that $\boldsymbol{q}^\top A = \rho\boldsymbol{q}^\top$, and $\boldsymbol{y}^\top A = \rho\boldsymbol{y}^\top$ if and only if $\boldsymbol{y}$ is a scalar multiple of $\boldsymbol{q}$.*

Definition 2. *The positive eigenvalue ρ and positive eigenvector $\boldsymbol{p}$ from Theorem 1 and Theorem 4 are called Perron root and Perron eigenvector respectively.*

Bibliography

[1] M. Slavtchova-Bojkova. On two-type decomposable branching processes in continuous time and time to escape extinction. In Branching Processes and their Applications, Lecture Notes in Statistics, **219**, Puerto, I.M., et al. Eds., Springer International Publishing: Switzerland, 319-329, (2016).

[2] M. Slavtchova-Bojkova, P. Trayanov, S. Dimitrov. Branching processes in continuous time as models of mutations: Computational approaches and algorithms. Computational Statistics and Data Analysis, **113**, 111–124, (2017).

[3] K. Vitanov, M. Slavtchova-Bojkova. Multitype branching processes in continuous time as models of cancer. Annuaire de l'Universite de Sofia "St. Kl. Ohridski", Fac. Math and Inf., **104**, 193-200, (2017).

[4] M. Slavtchova-Bojkova, K. Vitanov. Modelling cancer evolution by multi-type age-dependent branching processes. Comptes rendus de l'Acade'mie bulgare des Sciences, **71**, 10, 1297-1305, (2018).

[5] M. Slavtchova-Bojkova, K. Vitanov. Multi-type age-dependent branching processes as models of metastasis evolution. Stochastic Models, **35**, 284-299, (2019), https://doi.org/10.1080/15326349.2019.1600410.

[6] K. Vitanov, M. Slavtchova-Bojkova. On decomposable multi-type Bellman-Harris branching process for modeling cancer cell populations with mutations. 21st European Young Statisticians meeting - Proceedings, 113-118, (2019).

[7] K. Vitanov, M. Slavtchova-Bojkova. Modeling escape from extinction with decomposable multi-type Sev-

astyanov branching processes. Stochastic Models, (2022), https://doi.org/10.1080/15326349.2022.2041037.

[8] B. A. Sevastyanov. Branching processes. Mir, Moscow, (1971) (in Russian).

[9] M. Slavtchova-Bojkova, N. M. Yanev. Branching stochastic processes, 2nd ed. University Publ. "St. Kl.Ohridski", (2018) (in Bulgarian).

[10] P. Haccou, P. Jagers, V. Vatutin. Branching processes: Variation, Growth and Extinction of Populations. Cambridge University Press, (2007).

[11] P. Jagers. Branching Processes with Biological Applications (1st ed.). John Wiley & Sons Ltd.: New York, (1975).

[12] N. M. Yanev. Stochastic models of cell proliferation kinetics based on branching processes. pp. 3 - 20 in A. Almudevar, D. Oakes, J. Hall. (Editors). Statistical Modeling for biological systems. Springer Nature, Switzerland (2020).

[13] Pakes, A. (2003). Biological applications of branching processes. Handbook of Statistics Vol. 21 Stochastic Processes: Modelling and Simulation (Shanbhag, D.N. and Rao, C.R., eds.), Chapter 18, 693–773, Elsevier Science B.V.

[14] M. Kimmel, D. Axelrod. Branching Processes in Biology. Springer: New York, (2002).

[15] E. Pollak. Survival probabilities for some multitype branching processes in genetics. Journal of mathematical biology, **30**, No.6, 583 –596 (1992).

[16] J. Nedelman, H. Downs, P. Pharr. Inference for an age - dependent, multitype branching - process model of mast cells. Journal of Mathematical Biology, bf 25, No.2, 203 –226 (1987).

[17] M. E. Moody. A branching process model for the evolution of transposable elements. Journal of Mathematical Biology, **26**, No.3, 347 – 357 (1988).

[18] N. M. Yanev, A. Yu. Yakovlev. Dynamics of induced cell proliferation systems within a framework of a branching process model: 2. Some characteristics of the cell cycle temporal organization. Cytology **25**, 818 – 826 (1983).

[19] A. Yakovlev, N. Yanev. Transient processes in cell proliferation kinetics. Lecture Notes in Biomathematics, v.82, Springer, Berlin (1989).

[20] N. Yanev. Branching processes in cell proliferation kinetics. Ch.12 in: Workshop on Branching Processes and Their Application, Lecture Notes in Statistics - Proceedings, M. Gonzales, M. Molina, I. del Puerto, M. Mota, R. Martinez, A. Ramos (Editors), Springer, Berlin 159 – 178, (2010).

[21] C. A. Macken, A. S. Perelson. Branching processes applied to cell surface aggregation phenomena. Springer, Berlin (1985).

[22] N. M. Yanev, A. Yu. Yakovlev. Dynamics of induced cell proliferation systems within a framework of a branching process model: 1. Numbers of cells in successive generations. Cytology 22, 945 – 953 (1980)(in Russian).

[23] N. M. Yanev, A. Yu. Yakovlev. Dynamics of induced cell proliferation systems within a framework of a branching process model: 2. Some characteristics of the cell cycle temporal organization. Cytology **25**, 818 – 826 (1983) (in Russian).

[24] N. M. Yanev, A. Yu. Yakovlev, M. Tanoushev. A method for estimation of the probability of the cell reproductive death. Studia Biophysica **123**, No. 2, 17 – 124 (1988) (in Russian).

[25] A. Yu. Yakovlev, N. M. Yanev. Branching stochastic processes with immigration in analysis of renewing cell-populations. Math. Biosci. **203**, No. 1 , 37 – 63 (2006).

[26] N. M. Yanev, A. Yu. Yakovlev. On the distribution of marks over a proliferating sell population obeying the Bellman-Harris branching process. Mathematical Biosciences **75**, 159 – 173 (1985).

[27] C. P. Farrington , M. N. Kanaan, N. J. Gay. Branching process models for surveillance of infectious diseases controlled by mass vaccination. Biostatistics, **4**, 279 - 295 (2003).

[28] C. Jacob. Branching processes: Their role in epidemiology. Int. J. Environ. Res. Public Health.,**7**, 1186 – 1204 (2010).

[29] R. Bartoszynsk. On a certain model of an epidemic. Applicationes Mathematicae, **13**, 139 – 151 (1972).

[30] N. Becker. Estimation for discrete time branching processes with application to epidemics. Biometrics, **33**, 515 - 522 (1977).

[31] N. G. Becker, J. L. Hopper. Assessing the heterogeneity of disease spread through a community. American. J. Epidemiol. **117** 362 – 74 (1983).

[32] A. Yu. Yakovlev, N. M. Yanev. Branching populations of cells bearing continuous labels. Pliska Stud. Math. Bulg. **18**, 387 – 400 (2007).

[33] C. J. Mode. Stochastic processes in epidemiology - HIV/AIDS and other infectious diseases. World Scientific, Singapore (2000).

[34] P. Jagers. A general stochastic model for population development. Skand. Aktuarietidskr. **52**, 84 – 103 (1969).

[35] S. Asmussen, H. Hering. Branching processes. Birkhäuser, Basel (1983).

[36] E. Pardoux. Probabilistic Models of Population Evolution. Springer International Publishing: Switzerland, (2016).

[37] N. M. Yanev, A. Yu. Yakovlev. On the distribution of marks over a proliferating cell population obeying the Bellman-Harris branching process. Mathematical Biosciences **75**, 159 – 173 (1985).

[38] M. Deistler, G. Feichtinger. The linear model formulation of a multitype branching process applied to population dynamics. Journal of the American Statistical Association, **69**, 662 – 664 (1974).

[39] N. M. Yanev, C. Jordan, S. Catlin, A. Yu. Yakovlev. Two-type Markov branching processes with immigration as a model of leukemia cell kinetics. C. R. Acad. Bulg. Sci. **59**, No. 9, 1025 – 1032 (2005).

[40] K. Danesh, R. Durrett, L. J. Havrilesky, E. Myers. A branching process model of ovarian cancer. Journal of Theoretical Biology, **314**, 10 – 15. (2012)

[41] I. V. Basawa, D. J. Scott. Asymptotic optimal inference for nonergodic models. Springer, New York (1983).

[42] V. A. Vatutin, A. M. Zubkov. Branching Processes-1. Journal of Soviet Mathematics **39** 2431 – 2475 (1987).

[43] P. Guttorp. Statistical inference for branching processes, Wiley, New York, (1991).

[44] G. Sankaranarayanan. Branching processes and its estimation theory. Wiley, New York, (1989).

[45] N. M. Yanev. Statistical inference for branching processes. Ch. 7 in: Records and branching processes (Eds M. Ahsanullah, G. P. Yanev), Nova Science Publishers, New York (2008), 143 – 168.

[46] J. Foster. A limit theorem for a branching process with state dependent immigration. Ann. Math. Statist. **42**, 1773 – 1776 (1971).

[47] A. G. Pakes. A branching process with a state-dependent immigration component. Adv. Appl. Prob. **3**, 301 – 314 (1971).

[48] K. V. Mitov, N. M. Yanev. Bellman-Harris branching processes with state-dependent immigration. J. Appl. Prob. **22**, 757 – 765 (1985).

[49] K. V. Mitov, N. M. Yanev. Bellman-Harris branching processes with a special type of state - dependent immigration. Adv. Appl. Prob. **21**, 270 – 283 (1989).

[50] N. M. Yanev, M. N. Slavtchova. Convergence in distribution of supercritical Bellman-Harris Branching processes with state-dependent immigration. Mathematica Balkanica **3** , 35 – 42 (1992).

[51] N. M. Yanev, M. Nikiforova-Slavtchova. Non-critical Bellman-Harris branching processes with state-dependent immigration. Serdica Bulg. Math. Publ. **17**, 67 – 79 (1991).

[52] K. V. Mitov. Limit theorems for state-dependent branching processes. C. R. Acad. Bulg. Sci. **36**, No. 2, 189 – 192 (1983) (in Russian).

[53] K. V. Mitov. Limit theorem for multitype Galton-Watson processes with decreasing state-dependent immigration. Math. and Education in Math. **17**, 293 – 299 (1988)(in Russian).

[54] K. V. Mitov. The maximal number of offspring of one particle in a branching process with state-dependent immigration, Math. and Education in Math. **27**, 92 – 97 (1998).

[55] K. V. Mitov, G. P. Yanev. Maximum family size in branching processes with state-dependent immigration. Math. and Education in Math. **28**, 142 – 149 (1999).

[56] K. V. Mitov. The maximal number of particles in a branching process with state dependent immigration. Pliska Stud. Math. Bulg. **13**, 161 – 168 (2000).

[57] K. Athreya, P. Ney. Branching Processes. Springer: New York, (1972).

[58] T. Harris. The Theory of Branching Processes. Springer: Berlin, (1963).

[59] C. J. Mode. Multitype branching processes: Theory and applications, American Elsevier Publishing Co. Inc, New York (1971).

[60] C. Mode. Stochastic Processes in Demography and Their Computer Implementation. Springer: Berlin/Heidelberg, (1985).

[61] Y. Iwasa, F. Michor, M. Nowak. Evolutionary dynamics of escape from biomedical intervention. Proc. Biol. Sci. B, **270**, 2573-2578, (2003).

[62] Y. Iwasa, F. Michor, M. Nowak. Evolutionary dynamics of invasion and escape. J. Theor. Biol., **226**, 2, 205-214, (2004).

[63] P. Haccou, V. Vatutin. Establishment success and extinction risk in autocorrelated environments. Theoretical Population Biology, **64**, 303–314, (2003).

[64] M. Serra, P. Haccou. Dynamics of escape mutants. Theor. Popul. Biol., **72**, 167-178, (2007).

[65] S. Sagitov, M. Serra. Multitype Bienaymé–Galton–Watson processes escaping extinction. Advances in Applied Probability, 41(1), 225-246 (2009).

[66] A. Ghosh, M. Serra, P. Haccou. Quantifying stochastic introgression processes in random environments with hazard rates. Theoretical Population Biology, **100**, 1-5 (2015).

[67] R. Bellman. On the theory of dynamic programming. Proceedings of the National Academy of Sciences **38** (8): 716–719 (1952).

[68] R. Bellman. Dynamic Programming. Princeton University Press, Princeton, NJ (1957).

[69] R. Howard. Dynamic programming and Markov processes. Cambridge, MA: MIT Press (1960).

[70] M. L. Puterman. Markov Decision Processes, 2nd ed. John Wiley & Sons, Hoboken, NJ, (2005).

[71] Bertsekas, D.P. (2017). Dynamic Programming and Optimal Control: Approximate Dynamic Programming, 4 e. Belmont, MA.: Athena Scientific.

[72] R. Bellman, S. Dreyfus. (1959). Functional approximations and dynamic programming. Mathematical Tables and Other Aids to Computation **13**: 247–251 (1959).

[73] R. Bellman, R. Kalaba, B. Kotkin. Polynomial approximation— a new computational technique in dynamic programming: Allocation processes. Mathematics of Computation **17**: 155–161 (1963).

[74] K. Judd. Numerical Methods in Economics. MIT Press (1998).

[75] P. Werbos. Approximate dynamic programming for real-time control and neural modelling. In: Handbook of Intelligent Control: Neural, Fuzzy, and Adaptive Approaches (eds. D.J. White and D.A. Sofge), 493–525. Van Nostrand (1992).

[76] J. Si, A. Barto, W. Powell, and D. Wunsch. Learning and Approximate Dynamic Programming: Scaling up to the Real World. New York: John Wiley and Sons (2004).

[77] S. R. Pliska. Optimization of multitype branching processes. Management Science, **23**, 2, 117-124, (1976), http://www.jstor.org/stable/2629818.

[78] W. B. Powell. Approximate Dynamic Programming: Solving the Curses of Dimensionality, Second Edition. John Wiley & Sons, (2011), DOI:10.1002/9781118029176.

[79] W. B. Powell. Clearing the Jungle of Stochastic Optimization. In INFORMS Tutorials in Operations Research, (2014), https://doi.org/10.1287/educ.2014.0128.

[80] W. B. Powell. A unified framework for stochastic optimization. European Journal of Operational Research, **275**, 3, 795-821, (2019), https://doi.org/10.1016/j.ejor.2018.07.014.

[81] W. B. Powell. From Reinforcement Learning to Optimal Control: A unified framework for sequential decisions. In: Handbook on Reinforcement Learning and Optimal Control, Studies in Systems, Decision and Control, 29–74 (2021).

[82] W. B. Powell. Reinforcement Learning and Stochastic Optimization. Wiley (2022).

[83] I. J. Bienayme. De la loi de multiplication et de la duree des familles. Soc. Philomath. Paris, Extraits, Ser. 5, 37 – 39 (1845).

[84] C. C. Heyde, E. Seneta. I. J. Bienaymé: Statistical theory anticipated. Springer, New York (1977).

[85] H. W. Watson, F. Galton. On the probability of the extinction of families. The Journal of the Anthropological Institute of Great Britain and Ireland **4**, 138 - 144 (1875).

[86] P. Jagers. Some notes on the history of branching processes, from my perspective. Oberwolfach Workshop on Random Trees (2009). http://www.cim.pt/magazines/bulletin/14/article/118/pdf

[87] D. G. Kendall. Branching processes since 1873. J. London Math. Soc **41**, 385 - 406 (1966).

[88] J. F. Steffenson. On Sandsynligheden for at Afkommet uddor. Matem. Tiddskr. B, 19-23 (1930).

[89] A. Kolmogorov, N. Dmitriev. Branching random processes. Dokl. Acad. Nauk SSSR, **56**, No. 1, 7 – 10 (See, English transl., Selected works of A. N. Kolmogorov, vol. II: Probability theory and Mathematical Statistics, Kluwer, Dordrecht, 1992), 1947.

[90] A. Kolmogorov, B. Sevastyanov. The calculation of final probabilities for the branching random processes. Dokl. Acad. Nauk SSSR **56**, No. 8, 783 – 786 (1947). (English transl., Selected works of A. N. Kolmogorov, vol. II: Probability theory and Mathematical Statistics, Kluwer, Dordrecht, 1992).

[91] V. Vatutin. Scientific and personal life of B. A. Sevastyanov. Pliska Stud. Math. **24**, 5 – 12 (2015).

[92] N. Yanev. In memory of N. A. Dmitriev. Pliska Stud. Math. Bulgar. **24**, 13 – 20 (2015).

[93] N. Dmitriev. Nikolai Aleksandrovich Dmitriev (obituary). Uspekhi Mat. Nauk, **56**, No.2, 403 – 408, (2001).

[94] R. Bellman, T.E. Harris. On the theory of age-dependent stochastic branching processes. Proc. Nat. Acad. Sci. USA , **34** , 601 – 604 (1948).

[95] K. Crump, C. Mode. A general age-dependent branching process I. Journal of Mathematical Analysis and Applications, **24**, No. 3, 494 – 508 (1968).

[96] K. Crump, C. Mode. A general age-dependent branching process II. Journal of Mathematical Analysis and Applications, **25**, No. 1, 8 – 17 (1969).

[97] N. Obreshkov. Probability theory. Nauka i Izkustvo, Sofia (1963) (in Bulgarian).

[98] A. Obretenov. Probability theory. Nauka i Izkustvo, Sofia (1974) (in Bulgarian).

[99] K. Mitov, G. Mitov, N. Yanev. Limit theorems for critical randomly indexed branching processes. Ch.7 in: Workshop on Branching Processes and Their Application, Lecture Notes in Statistics - Proceedings, M. Gonzales, M. Molina, I. del Puerto, M. Mota, R. Martinez, A. Ramos (Editors), Springer, Berlin 95 – 108, (2010).

[100] M. Gonzalez, R. Martinez, M. Slavtchova-Bojkova. Time to extinction of infectious diseases through age-dependent branching models. Ch.17 in: Workshop on Branching Processes and Their Application, Lecture Notes in Statistics - Proceedings, M. Gonzales, M. Molina, I. del Puerto, M. Mota, R. Martinez, A. Ramos (Editors), Springer, Berlin 241 – 256, (2010).

[101] C. C. Heyde (Editor). Branching processes. Proceedings of the First World Congress. Springer, New York (1995).

[102] K. B. Athreya, A. N. Vidyashankar. Branching processes. In D. M. Shanbag, C. R. Rao (Editors). Handbook of statistics **19**, 25 - 53 , Elsevier, Amsterdam (2001)

[103] K. V. Mitov, N. M. Yanev. : Regulation, regeneration, estimation, applications. Pliska. Stud. Math. Bulgar. **19** 5 – 58 (2009).

[104] R. Durrett. Branching process models of cancer. Springer, New York (2015).

[105] Z. Taib. Branching processes and neutral evolution. Springer, Berlin (1992).

[106] K. B. Athreya, P. Jagers (Editors). Classical and modern branching processes. Springer Science & Business Media, New York (2012).

[107] G. P. Yanev, N. M. Yanev. Limit theorems for branching processes with random migration stopped at zero. In K. B. Athreya, P. Jagers (Editors). Classical and modern branching processes. Springer Science & Business Media, New York (2012), pp. 323 – 336.

[108] J.-F. Le Gall. Spatial branching processes, random snakes and partial differential equations. Birkhäuser, Basel (2012).

[109] G. Kersting, V. A. Vatutin. Discrete time branching processes in random environment. Wiley, Hoboken, NJ (2017).

[110] I. Pazit, L. Pal. Neutron fluctuations. A treatise on the physica of branching processes. Elsevier, Amsterdam (2007).

[111] S. K. Srinivasan. Stochastic theory and cascade processes, American Elsevier, New York, 1969.

[112] I. I. Gihman, A. V. Skorohod. Stochastic processes, vol. 2. Nauka, Moscow, (1977) (in Russian).

[113] C. Z. Wei, J. Winnicki. Estimation of means in the branching process with immigration. Ann. Statist. **18**, 1757 – 1773 (1990).

[114] J. Winnicki. Estimation of variances in the branching process with immigration. Probab. Theory Relat. Fields **88**, 77 – 106 (1991).

[115] B. A. Sevastyanov, A. M. Zubkov. Controlled branching processes. Theory Probab. Appl. **19**, 14-24 (1974).

[116] N. M. Yanev. Conditions of extinction of ϕ-branching processes with random ϕ. Theor. Probab. Appl. **20**, No 2, 433 – 440 (1975).

[117] N. M. Yanev, K. V. Mitov. Controlled branching processes with infinite mathematical means. Math. and Education in Math. **9** , 182 – 186 (1980) (in Russian).

[118] N. M. Yanev, G. P. Yanev. Conditions for extinction of controlled branching processes. Math. and Education in Math. **18**, 550 – 556 (1989).

[119] N. M. Yanev, K. V. Mitov. Controlled branching processes: the case of random migration. C. R. Acad. Bulg. Sci. **33**, No 4, 473 – 475 (1980).

[120] M. G. Velasco, I. García, G. P. Yanev. Controlled Branching Processes. John Wiley & Sons (2018).

[121] M. González, I. Del Puerto, N. Yanev, G. Yanev. Controlled branching processes with continuous time. Journal of Applied Probability, 58(3), 830-848 (2021). doi:10.1017/jpr.2021.8

[122] N. M. Yanev, K. V. Mitov. Limit theorems for controlled branching processes with decreasing emigration. Pliska Stud. Math. Bulg. **7**, 83 – 89 (1984) (in Russian).

[123] N. M. Yanev, K. V. Mitov. Limit theorems for controlled branching processes with non-homogeneous migration. C. R. Acad. Bulg. Sci. 35, No 3, 229 – 301 (1982).

[124] N. M. Yanev, K. V. Mitov. Controlled branching processes with non-homogeneous migration. Pliska Stud. Math. Bulg. **7**, 90 – 96 (1984) (in Russian).

[125] M. Drmota, G. Louchard, N. M. Yanev. Analysis of recurrence related to critical nonhomogeneous branching processes. Stoch. Analysis Appl. **24**, No. 1 , 37 – 59 (2006).

[126] K. V. Mitov, N. M. Yanev. Limit theorems for renewal, regeneration and branching processes. Math. and Education in Math. **30**, 32 – 41 (2001).

[127] K. V. Mitov, N. M. Yanev. Regenerative processes in the infinite mean cycle case. J. Appl. Prob. **38**, 165 – 179 (2001).

[128] P. I. Mayster. Alternating branching processes. J. Appl. Prob. **42**, 1095 – 1108 (2005).

[129] M. Slavtchova-Bojkova, N. M. Yanev. Non-critical branching processes with two types of state-dependent immigration. C. R. Acad. Bulg. Sci. **47**, No 6, 13 – 16 (1994).

[130] M. Slavtchova-Bojkova. On the subcritical age-dependent branching processes with two types of immigration. Math. and Education in Math. **31**, 187 – 191 (2002).

[131] G. Alsmeyer, M. Slavtchova-Bojkova. Limit theorems for subcritical age-dependent branching processes with two types of immigration. Stochas- tic Models **21**, No 1, 133 – 147 (2005).

[132] N. M. Yanev. Branching stochastic processes with immigration. Bulletin de l'Institut de Mathematiques (Bulg. Acad. Sci.) **5**, 71 – 88 (1972) (in Bulgarian).

[133] N. M. Yanev. On a class of decomposable age-dependent branching processes. Mathematica Balkanica **2**, 58 – 75 (1972) (in Russian).

[134] T. Harris. Branching processes. Ann. Math. Statist. **19**, 474 – 494 (1948).

[135] J.-P. Dion. Estimation of the mean and the initial probabilities of a branching process. J. Appl. Prob. **11**, 687 – 694 (1974).

[136] P. D. Feigin. A note on maximum likelihood estimation for simple branching processes. Austr. J. Statist. **19**, 152 – 154 (1977).

[137] N. Keiding, S. Lauritzen. Marginal maximum likelihood estimates and estimation of the offspring mean in branching process. Scand. J. Statist. **5**, 106 – 110 (1978).

[138] N. M. Yanev. On the statistics of branching processes. Theor. Probab. Appl. **20**, No 3, 623 – 633 (1975).

[139] N. M. Yanev, J.-P. Dion. A new transfer limit theorem. C. R. Acad. Bulg. Sci. **44**, No.1, 19 – 22 (1991).

[140] N. M. Yanev, J.-P. Dion. Limiting distributions of Galton-Watson branching processes with a random number of ancestors. C. R. Acad. Bulg. Sci. **44**, No. 3, 23 – 26 (1991).

[141] N. M. Yanev, J.-P. Dion. Estimation theory for branching processes with and without immigration. C. R. Acad. Bulg. Sci. 44, No. 4, 19 – 22 (1991).

[142] N. M. Yanev, J.-P. Dion. Statistical inference for branching processes with censored observations. C. R. Acad. Bulg. Sci. **45**, No. 12, 21 – 24 (1992).

[143] N. M. Yanev, J.-P. Dion. Statistical inference for branching processes with an increasing random number of ancestors. J. Stat. Planning and In- ference **39**, 329 – 352 (1994).

[144] N. M. Yanev, J.-P. Dion. Central limit theorem for martingales in BGWR branching processes with some statistical applications. Math. Methods of Statistics **4**, No. 3, 344 – 358 (1995).

[145] N. M. Yanev, J.-P. Dion. Limit theorems and estimation theory for branching processes with an increasing random number of ancestors. J. Appl. Prob. **34**, 309 – 327 (1997).

[146] C. Jacob, N. Lalam, N. M. Yanev. Statistical inference for processes depending on environments and application in regererenerative processes. Pliska Stud. Math. Bulg. **17**, 109 – 136 (2005).

[147] C. Jacob, N. Lalam. Estimation of the offspring mean in a general single-type size-dependent branching processes. Pliska Stud. Math. Bulg. **16**, 65 – 88 (2004).

[148] T. N. Sriram, I. V. Basawa, R. M. Huggins. Sequential estimation for branching processes with immigration. Annals of Statistics **19**, No. 4, 2232 – 2243 (1991).

[149] P. I. Mayster. Branching diffusion processes in a bounded domain with absorbing boundary. Theor. Probab. Appl. **19** , 589 - 596 (1974).

[150] P. I. Mayster. Mathematical expectation of a time-continuous branching diffusion process. Theor. Probab. Appl. **23**, 831 – 836 (1978).

[151] P. I. Mayster. Branching diffusion processes with Poisson initial distribution. Serdica Bulg. Math. Publ. **8**, 250 – 254 (1982) (in Russian).

[152] P. I. Mayster. Several examples of branching diffusion processes in an unbounded domain. Serdica Bulg. Math. Publ. **8**, 190 – 196 (1982) (in Russian).

[153] P. I. Mayster. Mathematical expectation of a branching process with small diffusion. Pliska Stud. Math. Bulg. **7**, 109 – 117 (1984) (in Russian).

[154] J. Kersten, K. Mattes, J. Mekke. Infinitely divisible point processes. Mir, Moscow, 1982.

[155] G. S. Tschobanov. Raumlich homogene kritische verzweigungsprozesse mit kontinuievlicher Zeit. Serdica Bulg. Math. Publ. **6**, 264 – 269 (1980)(in German).

[156] G. S. Tschobanov. Raumlich homogene kritische Verzweigungsprozesse mit beliebigem Markenraum. Pliska Stud. Math. Bulg. **7** (1984), 34 – 54 (1984) (in German).

[157] G. S. Tschobanov. Raumlich homogene kritische Bellman - Harris Prozesse. Pliska Stud. Math. Bulg. **7** (1984), 18 – 33 (1984) (in German).

[158] G. P. Yanev. Revisiting offspring maxima in branching processes, Pliska Stud. Math. Bulg. **18**, 401 – 426 (2007).

[159] G. P. Yanev, I. Rahimov. Maximal number of direct offspring in simple branching processes. Nonlinear Analysis: Theory, Methods and Applications **30**, No 4, 1115 – 1123 (1997).

[160] I. Rahimov, G. P. Yanev. On the maximum family size in branching processes. J. Appl. Prob. **36**, 632 – 643 (1999).

[161] K. V. Mitov. The maximal number of offspring of one particle in a branching process with state-dependent immigration, Math. and Education in Math. **27**, 92 – 97 (1998).

[162] K. V. Mitov, G. P. Yanev. Maximum family size in branching processes with state-dependent immigration. Math. and Education in Math. **28**, 142 – 149 (1999).

[163] K. V. Mitov, G. P. Yanev. Maximum individual score in critical two-type branching processes, C. R. Acad. Bulg. Sci. **55**, No 11, 17 – 22 (2002).

[164] K. V. Mitov, A. G. Pakes, G. P. Yanev. Extremes of geometric variables with applications to branching processes. Statistics and Probability Letters **65**, No 4, 379 – 388 (2003).

[165] H. Kesten, B. P. Stigum. A limit theorem for multidimensional Galton-Watson processes. Ann. Math. Statist. **37** , 1211 – 1223 (1966).

[166] F. M. Hoppe. Supercritical multitype branching processes. Ann. Probab. **4**, 393-401 (1976).

[167] S. Asmussen. Convergence rates for branching processes. Ann. Probab. **4**, 139 – 146 (1976).

[168] S. Asmussen. Almost sure behavior of linear functionals of supercritical branching processes. Trans. Amer. Math. Soc. **1**, 233 – 248 (1977).

[169] A. Joffe, F. Spitzer. On multitype branching processes with $\rho \leq 1$. J. Math. Anal. Appl. **19**, 409 – 430 (1967).

[170] J. A. C. Resing. Pokking systems and multiple branching processes. Queueing Systems, **13**, 409 – 426 (1993).

[171] F. Maaouia, A. Touati. Identification of multiple branching processes. Annals of Statistiics **33**, 2655 - 2694 (2005).

[172] S. Pennison. Conditional limit theorems for multitype branching processes and illustration in epidemiological risk analysis. Ph. D. thesis, University of Potsdam, Germany (2010).

[173] S. Singh, D. J. Schneider, C. R. Myers. Using multitype branching processes to quantify statistics of disease outbreaks in zoonotic epidemics. Phys. Rev. E 89, 032702 (2014).

[174] L. Demetrius, P. Schuster, K. Sigmund. Polynucleotide evolution and branching processes. Bulletin of Mathematical Miology, **47**, No.2 , 239 – 262 (1985).

[175] E. Crane, B. Rath, D. Yeo. Age evolution in the mean field forest fire model via multitype branching processes. Ann. Probab. **49**, No.4, 2031 – 2075 (2021)

[176] Ivanoff, B. G. The multitype branching diffusion. Journal of Multivariate Analysis **11**, No. 3, 289 – 318 (1981).

[177] Vazquez, A. Spreading dynamics on heterogeneous populations: multitype network approach. Physical Review E, **74**, No.6, 066114 (2006).

[178] K. B. Athreya, S. Karlin On branching processes with random environments, I Ex- tinction Probabilities. Ann. Math. Statist. **42**, 1499 - 1520 (1971).

[179] K. B. Athreya, S. Karlin On branching processes with random environments, II - Limit Theorems. Ann. Math. Statist. **42**, 1843 - 1858 (1971).

[180] D. Tanny. Limit theorems for branching processes in random environment. Ann. Probab. **5**, 100 - 116 (1977).

[181] I. M. Del Puerto, N. M. Yanev. Branching processes with multi-type random control functions. C. R. Acad. Bulg. Sci. **57**, No 6 , 29 - 36 (2004).

[182] I. M. Del Puerto, N. M. Yanev. Stationary distributions for branching processes with multi-type random control functions. J. Appl. Stat. Sci. **16**, 91 - 102 (2008).

[183] A. Yu. Yakovlev, N. M. Yanev. Distributions of continious labels in branching stochastic processes. C. R. Acad. Bulg. Sci. **59**, No. 11, 1123 – 1130 (2006).

[184] R. Garcia-Pelayo, I. Salazar, W. C. Schieve. A branching process model for sand avalanches. Journal of Statistical Physics **72**, 167 - 187 (1993).

[185] D. Aldous, L. Popovic. A critical branching process model for biodiversity. Adv. Appl. Prob. **37**, 1094 – 1115 (2005).

[186] L. Devroye. Branching processes in the analysis of the heights of trees. Acta Informatica **24**, 277 - 298 (1987).

[187] C. Gutierez, C. Minuesa. Predator-prey density-dependent branching processes. Stochastic Models (2022), doi: 10.1080/15326349.2022.2032755

[188] R. B. R. Azevedo, P. Olofsson. A branching process model of evolutionary rescue. Mathematical Biosciences **341**, 108708 (2021).

[189] J. P. Gleeson, T. Onaga, P. Fennell, J. Cotter, R. Burke, D. J. P. O'Sullivan. Branching process descriptions of information cascades on Twitter. Journal of Complex Networks **8**, cnab002 (2021).

[190] M. C. Teich, B. E. A, Saleh. Branching processes in quantum electronics. IEEE J. on Selected Topics in Quantum Electronics **6**, No. 6, 1450 - 1457 (2000).

[191] K. Dusek, M. Demjanenko. Application of the theory of branching processes (cascade theory) to polymer degradation and crosslinking: Postgel stage. International Journal of Radiation Applications and Instrumentation. Part C. Radiation Physics and Chemistry **28**, No. 5 – 6, Pages 479 – 486 (1986).

[192] T. Epps. Stock prices as branching processes. Communications in statistics: Stochastic Models **12**, 529 – 558 (1996).

[193] G. K. Mitov, K. V. Mitov. Randomly indexed Galton-Watson branching processes, Math. and Education in Math. **35**, 275 – 281 (2006).

[194] G. K. Mitov, K. V. Mitov, N. M. Yanev. Statistics & Probability Letters **79**, No. 13, 1512 – 1521 (2009).

[195] G. K. Mitov, K. V. Mitov. Option pricing by branching processes, Pliska Stud. Math. Bulg. **18**, 213 - 224 (2007).

[196] Y. Jiao, C. Ma, S. Scotti. Alpha-CIR model with branching processes in sovereign interest rate modeling. Finance and Stochastics **21**, 789 – 813 (2017).

[197] H. Shapiro. "Linear Algebra and Matrices. Topics for a Second Course". American Mathematical Society, Providence, RI, USA, (2015).

[198] Jacob C., (2010). Branching processes: their role in epidemiology, International Journal of Environmental Research and Public Health, **7,** 1186-1204.

[199] K. Etessami, M. Yannakakis. Recursive markov decision processes and recursive stochastic games. J. ACM **62**(2), 11:1-11:69 (2015), https://doi.org/10.1145/2699431

[200] K. Etessami, A. Stewart, M. Yannakakis. Polynomial time algorithms for branching markov decision processes and probabilistic min(max) polynomial bellman equations. Math. Oper. Res. **45**(1), 34-62 (2020). https://doi.org/10.1287/moor.2018.0970

[201] U. Rothblum, P. Whittle. Growth optimality for branching Markov decision chains. Math. Oper. Res. **7**(4):582–601 (1982).

[202] E. M. Hahn, M. Perez, S. Schewe, F. Somenzi, A. Trivedi, D. Wojtczak. Model-Free Reinforcement Learning for Branching Markov Decision Processes. In: Silva, A., Leino, K.R.M. (eds) Computer Aided Verification. CAV 2021. Lecture Notes in Computer Science, **12760**, Springer, Cham. (2021) https://doi.org/10.1007/978-3-030-81688-9_30

[203] R. Sutton, A. Barto. Reinforcement Learning: An Introduction, Second Edition. MIT Press, Cambridge, MA, (2018).

[204] R. W. H. Sargent. Optimal control. Journal of Computational and Applied Mathematics, **124**, 361-371 (2000).

[205] R. Stengel. Stochastic Optimal Control: Theory and Application. Hoboken, NJ: John Wiley & Sons (1986).

[206] D. Kirk. Optimal Control Theory: An introduction. New York: Dover (2012).

[207] F. Lewis, D. Vrabie, V. Syrmos. Optimal Control, 3rd Edition. John Wiley & Sons (2012).

[208] S. Sethi. Optimal Control Theory: Applications to Management Science and Economics, 3 e. Boston: SpringerVerlag (2019).

[209] Python Software Foundation. Python Language Reference, version 3.8.13, https://docs.python.org/3.8/reference/index.html

[210] C. R. Harris, K. J. Millman, S. J. van der Walt, et al. Array programming with NumPy. Nature, **585**, 357–362 (2020), https://doi.org/10.1038/s41586-020-2649-2

[211] P. Virtanen, R. Gommers, T. E. Oliphant, et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. Nature Methods, **17**, 261-272 (2020), https://doi.org/10.1038/s41592-019-0686-2

[212] J. D. Hunter. Matplotlib: A 2D Graphics Environment. Computing in Science & Engineering, **9**, 90-95 (2007).

[213] yWorks GmbH. (2019). yEd. Retrieved from https://www.yworks.com/products/yed