UNIVERSITÀ DEGLI STUDI DI PISA

CORSO DI LAUREA MAGISTRALE IN MATEMATICA

Markovian binary trees subject to catastrophes: computation of the extinction probability

16 Settembre 2013

TESI DI LAUREA MAGISTRALE

Candidato

Fabio Cecchi

cecchi@mail.dm.unipi.it

Relatore
Prof. Beatrice Meini
Università di Pisa

Controrelatore
Prof. Dario A. Bini
Università di Pisa

ANNO ACCADEMICO 2012/2013

Contents

Chapter 1

Introduction

If we consider a set of things, is natural to consider the evolution in time of this set. Maybe some of these things will transform, other will vanish and new one will be created. A process of this kind is the argument of interest of the evolutionary problems.

Due to the generic definition of these processes, it is not difficult to imagine application of these in the more disparate fields, from the biology to the chemistry, from the physics to the demography, from the finance to the ecology.

Nevertheless, it was Sir Francis Galton who deserves the credit for formulating rigorously this problem for the first time. In fact around the 1870 he was asked to answer some questions concerning the evolution of the diffusion of the various surnames. He observed indeed how some surnames, once largely diffused, were at the moment close to extinction. In particular, in its famous Problem 4001, he formulated mathematically the problem, and posed the question "Would all surnames eventually become extinct?". That question was answered by the Rev. William Watson and, in 1874, Galton and Watson wrote together a paper entitled "On the probability of extinction of families". They deduced a simple mathematical conclusion, proving that if the average number of a man's sons is at most 1, then their surname will almost surely die out, while if it is more than 1, then there is a strictly positive probability that the surname will survive for any given number of

generations. A simple process of this kind, wherein an individual of the set generates offsprings with a certain probability, passing them an inherited property is called *Galton Watson process* and is the fundamental brick for what we actually know by the name of *branching processes*.

As a matter of fact, the term branching process was forged by Kolmogorov and Dmitriev in the work "Branching processes" of 1947. Nowadays a branching process is defined as a Markov process that models a population wherein each individual in generation n produces some random number of individuals in generation $n + 1$, according to a certain probability.

In this thesis we are interested in studying a particular evolutionary process. In such process the individuals may belong to a finite number of typologies and each individual, after an exponentially distributed interval of time, will either change type or die or generate offspring. Since we limit ourselves to the case where at most one descendant at a time is generated, such a process takes the name of *Markovian binary tree*.

After a brief introduction of notation and basic definitions, in Chapter 2 we will describe in detail the Markovian binary tree model, showing how to fit the properties of the branching processes to our model. In particular we will be interested in computing the probability that a certain population will become extinct, a well known problem in the setting of Markovian binary trees. We will conclude the chapter describing briefly the existing algorithm for the computation of the extinction probability in the case wherein all the individuals evolve independently one from the other.

In Chapter 3 we will introduce a parallel process of catastrophes running alongside the population process. When a catastrophe happens, all the individuals alive are affected and either survive or die according to a certain probability. This problem is much harder and the computation of the extinction probability is beyond our means. It will be introduced a parameter, denoted by ω, whose positivity or negativity plays a discerning role between populations having a positive probability to survive forever and those who will become extinct almost surely. However, even the computation of such a parameter is challenging, since it can be compared to the computation of the maximal Lyapunov exponent associated to a linear random dynamical

system.

Therefore, in Chapter 4, we will describe how to bound the parameter ω. We will cite some existing bounds and deduce a couple of new one exploiting the matrix properties of the problem. All these bounds will be tested and compared in Chapter 5, where the results of some experiments will be illustrated.

1.1   Notation and basic definitions

In order to make the whole text flow smoother, we use this section to recall some basic definitions and to fix some notation.

Numerical analysis and linear algebra

The matrices are denoted by capital letters, and the symbol I_n denotes the identity matrix of dimension n, where we omit the subscript if the dimension appears obvious from the context. The vectors are denoted by boldface letters and, if not specified differently, they are generally considered column vectors. Among them a special role is played by the vectors $\mathbf{0}$ and $\mathbf{e}$ having all the entries equal to 0 and to 1 respectively, and by the vectors $\mathbf{e}_i$ having all the entries null apart from the i-th which is equal to 1. The superscript ^{T} denotes the transpose operator. The symbol A_{ij} stands in for the i, j-th entry of the matrix A and v_i for the i-th entry of the vector $\mathbf{v}$.

Given a real matrix A, we say that $A \geq 0 (A > 0)$ if $A_{ij} \geq 0 (A_{ij} > 0)$ for each i, j. Given two real matrices A and B of the same dimension, we say that $A \geq B (A > B)$ if $A - B \geq 0 (A - B > 0)$. The same definition applies to vector, where the inequalities are applied component-wise.

Usually the vectors we consider belong to $\mathbb{R}^n$ or, sometimes, to the its subset $\mathcal{S}_n$, where

$$\mathcal{S}_n = \{\mathbf{x} \in \mathbb{R}^n \mid x_i \in \mathbb{N} \cup \{0\}, \ \forall i = 1, \ldots, n\} \tag{1.1}$$

is formed by the vectors having nonnegative integer entries.

Given a square matrix $A \in \mathbb{R}^{n \times n}$, we say that the n-dimensional non zero

vectors $\boldsymbol{v}$ and $\boldsymbol{u}$ are left and right *eigenvectors* for A, corresponding to the *eigenvalue* $\lambda \in \mathbb{C}$, if and only if

$$A\boldsymbol{u} = \lambda\boldsymbol{u}, \tag{1.2}$$

and

$$\boldsymbol{v}^T A = \lambda\boldsymbol{v}^T. \tag{1.3}$$

We define the *characteristic polynomial* of a matrix $A \in \mathbb{R}^{n \times n}$ by the formula

$$p_c(\lambda) = \det(A - \lambda I)$$

, where $\det(\cdot)$ denotes the determinant operator. The roots of the polynomial $p_c(\lambda)$, counted with their multiplicity, form the spectrum of the matrix A, which is denoted by

$$\sigma(A) = \{\lambda_1, \ldots, \lambda_n\}. \tag{1.4}$$

It is known that λ is an eigenvalue for the matrix A if and only if $\lambda \in \sigma(A)$. If the value 0 belongs to the spectrum of the matrix A, the matrix is said to be a *singular matrix*. An eigenvalue $\bar{\lambda}$ is said to be a *simple eigenvalue* if it is a simple root of the characteristic polynomial, i.e. $p_c(\bar{\lambda}) = 0$ and $p_c'(\bar{\lambda}) \neq 0$, where p_c' denotes the first derivative of the polynomial p_c.

The *spectral radius* of a matrix $A \in \mathbb{R}^{n \times n}$ is defined as the maximum of the absolute values of the eigenvalues of A and is denoted by $\rho(A)$, i.e.

$$\rho(A) = \max_{\lambda \in \sigma(A)} |\lambda|. \tag{1.5}$$

Another useful notion is the *norm*, denoted by $\|\cdot\|$. A *vector norm* is a function, $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}$ which associates a proper nonnegative value to each vector belonging to a vector space. Precisely every function such that, for

each $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$, fulfills the following properties

$$\|\alpha\boldsymbol{u}\| = |\alpha|\|\boldsymbol{u}\|,$$
$$\|\boldsymbol{u} + \boldsymbol{v}\| \leq \|\boldsymbol{u}\| + \|\boldsymbol{v}\|,$$
$$\|\boldsymbol{u}\| = 0 \Leftrightarrow \boldsymbol{u} = \boldsymbol{0},$$

is a vector norm on the vector space $\mathbb{R}^n$. It is possible to define a *matrix norm* by replacing $\boldsymbol{u}$ and $\boldsymbol{v} \in \mathbb{R}^n$ with A and $B \in \mathbb{R}^{n \times n}$ and requiring the same conditions to be satisfied. In particular, we observe that, with an abuse of notation, given $\|\cdot\|$ a vector norm we are allowed to define immediately an induced matrix norm by the defining the function

$$\|A\| = \max_{\|x\|=1}\{\|Ax\|\}. \tag{1.6}$$

Indeed the function just defined is a matrix norm and, moreover, it also fulfills a sub-multiplicative property, in fact

$$\|AB\| \leq \|A\|\|B\|, \tag{1.7}$$

for each $A, B \in \mathbb{R}^{n \times n}$. A rather special norm is the *maximum norm*, defined by

$$\|\boldsymbol{u}\|_\infty = \max\{|u_1| \ \ldots, |u_n|\}, \tag{1.8}$$

in the vector form and inducing the following matrix norm

$$\|A\|_\infty = \max_{i=1,\ldots,n} \sum_{j=1}^{n} |A_{ij}|. \tag{1.9}$$

Given a couple of matrices of generic dimension it is possible to perform on them the *Kronecker product*, an operation defined as follow,

Definition 1.1.1. *For each $A^{(1)} \in \mathbb{R}^{m_1 \times n_1}$ and $A^{(2)} \in \mathbb{R}^{m_2 \times n_2}$, the Kronecker product of the matrices $A^{(1)}$ and $A^{(2)}$ is defined by the matrix $A^{(1)} \otimes A^{(2)} \in$*

$\mathbb{R}^{(m_1+m_2)\times(n_1+n_2)}$ *such that*

$$A^{(1)} \otimes A^{(2)} = \begin{pmatrix} A_{11}^{(1)} A^{(2)} & \cdots & A_{n_11}^{(1)} A^{(2)} \\ \vdots & \ddots & \vdots \\ A_{1m_1}^{(1)} A^{(2)} & \cdots & A_{n_1m_1}^{(1)} A^{(2)} \end{pmatrix}. \tag{1.10}$$

This product possesses a lot of important properties, among these we mention the mixed product property, which states that given the matrices A, B, C, D, having dimensions allowing to define the standard products AC and BD, the equality

$$(A \otimes B)(C \otimes D) = (AC \otimes BD), \tag{1.11}$$

is verified.

The *matrix exponential* may be defined through the Taylor expansion, in parallel with what is usually done for the scalar exponential.

Definition 1.1.2. *For every matrix $A \in \mathbb{C}^{n \times n}$ the matrix exponential e^A is defined by*

$$e^A = I + A + \frac{A^2}{2!} + \frac{A^3}{3!} + \cdots = \sum_{i=0}^{\infty} \frac{A^i}{i!}. \tag{1.12}$$

It is known that such a series has an infinite radius of convergence, therefore we can differentiate the entries of the matrix obtained term by term yielding the formula

$$\frac{d}{dt} e^{At} = A e^{At} = e^{At} A.$$

There exist many other equivalent way to define the matrix exponential, see [19] for a detailed survey.

Although the definition of the exponential function is the same both in the scalar and in the matrix case, many of the properties that were taken for granted in the scalar case are no more true in the matrix one. This is due to the non commutativity of the matrix product, and the following property constitutes an example.

Theorem 1.1.1. *For $A, B \in \mathbb{C}^{n \times n}$, it holds*

$$e^{A+B} = e^A e^B \Leftrightarrow AB = BA. \tag{1.13}$$

The proof of this fact is a direct consequence of the formula (1.12) and can be found in [19].

Probability theory

Many of the concepts handled in the following chapters require some basic notions of probability theory. Such a field is huge and we aim to give just a brief introduction, which surely is not as complete and formal as the topic demands, but we hope may be of some help for the understanding of what follows.

Intuitively a probability function is a measure of how likely is an event to happen. This measure is represented by a value between 0 and 1, where the more an event is probable the higher is the measured value. In order to be more formal, we need to introduce the *probability space*, which is given by the triple

$$\mathcal{P} = (\Omega, \mathcal{F}, \mathrm{P}), \tag{1.14}$$

where

- Ω denotes the *sample space*, a collection of all the possible outcomes that may be attained by running a certain experiment;

- $\mathcal{F}$ is a fixed σ-algebra on Ω, which means that $\mathcal{F}$ is a nonempty collection of subsets of Ω such that is closed under the complement and the countable unions of its members and contains Ω itself;

- P is the *probability function* and associates a proper probability measure to a certain subset of the whole sample space. Thus P is a function $\mathrm{P} : \mathcal{F} \to [0, 1]$ such that $\mathrm{P}(\Omega) = 1$ and is a countably addictive function, i.e. if $\{A_i\} \subseteq \mathcal{F}$ is a countable collection of pairwise disjoint sets, then $\mathrm{P}(\bigcup_i A_i) = \sum_i \mathrm{P}(A_i)$.

9

A very important role in probability theory is played by the *random variables*, which are functions going from the space of the outcomes Ω into a certain subset $\mathcal{C}$ of $\mathbb{R}$, where $\mathcal{C}$ represents the feasible numerical outcomes of the experiment. In a certain sense a random variables perform as the connection between the real world experiment and the mathematical model we work on. A random variable may be either discrete or continuous depending on the structure of its co-domain $\mathcal{C}$ and is allowed to assume various random values depending on the outcome of the experiment and on the probability function defined in (1.14). In practice, denoting X our random variable, with an abuse of notation we say that the probability for the value of X to be $c \in \mathcal{C}$ is given by

$$\mathrm{P}[X = c] = \mathrm{P}[\{w \in \Omega \mid X(w) = c\}], \tag{1.15}$$

which is a known value if $\{w \in \Omega \mid X(w) = c\} \in \mathcal{F}$.

Given a random variable X, it is of a certain interest the understanding of what is the value we could expect X to possess. Such a quantity is denoted by $\mathrm{E}[X]$ and is called *expected value*. If X is a discrete random variable

$$E[X] = \sum_{x \in \mathcal{C}} \mathrm{P}[X = x] x, \tag{1.16}$$

while in presence of continuous random variable the summation is replaced by the integral on the whole set $\mathcal{C}$.

The *probability generating function* associated to a multivariate random variable $\mathbf{X} = (X_1, \ldots, X_n) \in \mathbb{R}^n$, is the complex valued function $F(\mathbf{s})$ defined by the following formula

$$F(\mathbf{s}) = E[\mathbf{s}^{\mathbf{X}}] = E[s_1^{X_1}, \ldots, s_k^{X_k}], \tag{1.17}$$

where $\mathbf{s}^{\mathbf{X}}$ is a multivariate random variable too and whose convergence is guaranteed as soon as $\|\boldsymbol{s}\|_\infty \leq 1$, where $\|\cdot\|_\infty$ is the maximum norm, see (1.8).

Another useful definition is that of *conditional probability*, that is to say the probability for a certain random variable X to take a certain value, say n, when it is known the value taken by another random value Y, say m. We

10

denote this probability

$$P[X = n | Y = m] = \frac{P[X = n \text{ and } Y = m]}{P[Y = m]}. \tag{1.18}$$

We avoid to discuss here more about probability theory, apart from briefly introduce a couple of useful probability distributions for a given random variable X.

- The *Poisson distribution* of parameter $\lambda > 0$ is a discrete probability distribution such that

$$P[X = n] = \frac{e^{-\lambda} \lambda^n}{n!} \tag{1.19}$$

- The *Exponential distribution* of parameter $\lambda > 0$ is a continuous probability distribution such that

$$P[a < X < b] = \int_a^b f(x, \lambda) \mathrm{d}x, \tag{1.20}$$

  where

$$f(x, \lambda) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0. \end{cases} \tag{1.21}$$

Stochastic processes

Given a probability space $(\Omega, \mathcal{F}, P)$, see (1.14), a *stochastic process* on a certain *state space* $D \subseteq \mathbb{R}^n$ is a collection of ordered random variables on Ω having values in D. We say that the process is a *discrete time process* if the collection is indexed by a numerable set of times $\mathcal{T}$, i.e.

$$\{X_n \mid n \in \mathbb{N}\}, \tag{1.22}$$

it is a *continuous time process* if the set $\mathcal{T}$ is continuous, i.e.

$$\{X(t) \mid t \geq 0\}. \tag{1.23}$$

The following definitions are presented for the discrete case, but may be easily generalized to the continuous case. The value of a certain entry of the process depends, in general, on the whole history of the process and on its own time. If the future of a process depends only on the current state, i.e.

$$P[X_n = d_n \mid X_{n-1} = d_{n-1}, \ldots, X_0 = d_0] = P[X_n = d_n \mid X_{n-1} = d_{n-1}], \quad (1.24)$$

the process is called a *Markov process.*

A process fulfilling the following time homogeneity condition

$$P[X_n \mid X_{n-1}] = P[X_1 \mid X_0], \quad (1.25)$$

for each $n \geq 1$, is said to be a *homogeneous process.* In the discrete time case a stationary Markov process takes the name of *Markov chain.*

Another useful property for a stochastic process is the ergodicity, an *ergodic processes* is a process whose statistical properties may be deduced from a single, sufficiently long, sample of the process. For a formal definition we refer to [20].

In order to fix some more concepts, we briefly show how a Markov chain works in practice. We suppose to have chosen the starting probability of the process, i.e. $P[X_0 = s]$ for every $s \in D$. The probability value assumed by X_1 depends on the probability value of the variable X_0 and on the probability distribution defining the transitioning phase, more precisely we define

$$p_{(s_i, s_j)} = P[X_1 = s_j \mid X_0 = s_i], \quad (1.26)$$

so that $p_{(s_i, s_j)}$ represents the probability that in a slot of time the stochastic process passes from the state $s_i \in D$ to the state $s_j \in D$. Such probabilities are called *transition probabilities* and may be considered the fundamental brick in the construction of a Markov chains. The chain evolves in this way obtaining progressively $P[X_2 = s], P[X_3 = s], \ldots$, for each $s \in D$.

In case we are allowed to provide with an index every element of the state space D, i.e. if D is a numerable space, we can define the *transition probability*

matrix P, such that the (i, j)-th entry is given by

$$P_{ij} = p_{(i,j)}, \tag{1.27}$$

where $p_{(i,j)}$ is defined in (1.26).

The matrix P is a *stochastic matrix*, which means that the sum over the entries of each row is equal to 1, due to the properties of the probability function. Furthermore we observe that the matrix P defines completely the whole Markov chain, in fact if we denote by $\boldsymbol{x}_{(n)}$ the vector whose entry $(x_{(n)})_i = \mathrm{P}[X_n = i]$, we have that

$$\boldsymbol{x}_{(1)}^T = \boldsymbol{x}_{(0)}^T P, \tag{1.28}$$

and

$$\boldsymbol{x}_{(n)}^T = \boldsymbol{x}_{(n-1)}^T P = \boldsymbol{x}_{(0)}^T P^n. \tag{1.29}$$

The matrix P is said to be *irreducible* if, for each $i, j \in S$, there exists $\tilde{n}$ such that $(P^{\tilde{n}})_{ij} > 0$. In the Markov chain context, an irreducible transition probability matrix means that there exists a strict positive probability of transitioning (even in more time slot) from every state to each other.

Given a state $s \in S$, we define T_s as the random variable representing the time used by the process for coming back to state s after starting in that state. Each state $s \in S$ may be classified through T_s, indeed s is either a *transient state*, if $\mathrm{P}[T_s < \infty] < 1$, or a *recurrent state* otherwise. Moreover if $\mathrm{E}[T_s] < \infty$ a recurrent state is said to be *positive recurrent state*.

An important characteristic of a Markov chain is the behavior that may be expected in the long term evolution of the process, this is displayed by the *stationary distribution*, denoted by $\boldsymbol{\pi}$, where

$$\pi_i = \lim_{n \to \infty} \mathrm{P}[X_n = i]. \tag{1.30}$$

This probability distribution must fulfill the following property

$$\boldsymbol{\pi}^T P = \boldsymbol{\pi}^T, \tag{1.31}$$

so that $\boldsymbol{\pi}$ is a left eigenvector for the matrix P relatively to the eigenvalue 1. Such a probability distribution is known to exist and to be unique if the matrix P is irreducible and all its states are positive recurrent.

The continuous time stationary Markov processes are the equivalent to the Markov chains in the continuous time case. In this setting, the role of the transition probability matrix is played by the *rate matrix*, or generator matrix. This matrix is characterized by having nonnegative off diagonal entries and the values on the diagonal such that the sum over each row is equal to 0. In such a case, denoted by R the rate matrix, the process remains in state i for a random time which follows an exponential probability distribution with parameter $-R_{ii}$, see (1.20), and then the process moves to another state j according to a probability $R_{ij}/(-R_{ii})$.

The stationary distribution $\boldsymbol{\pi}$ is defined like in (1.30) and must fulfill the condition

$$\boldsymbol{\pi}^T R = \mathbf{0}^T. \tag{1.32}$$

It is of a certain interest that if R is a rate matrix, $-R$ is a singular M-matrix, which is a class of matrices defined as follows

Definition 1.1.3. *A matrix M is an M-matrix if it is possible to express M in the form $M = \gamma I - B$ where $B \geq 0$ and $\gamma \geq \rho(B)$, where $\rho(B)$ is the spectral radius of B, see (1.5).*

An important feature of this class of matrices is the *inverse positive* property, that is to say that given A a not singular M-matrix we have that $A^{-1} \geq 0$.

Graph theory

A *graph* is a structure composed by a set of objects where some pairs of them are connected by links. The objects forming the graph are named *vertices*, while the connecting links take the name of *edges*. A graph G may be denoted in a compact form by

$$G = (V, E), \tag{1.33}$$

where V is the set of objects, i.e. the vertices, and $E \subseteq V \times V$ represents the set of the edges connecting the objects of the graph. If there exists an edge between the pair of vertices $v, w \in V$, we specifies that edge by $(v, w) \in E$. In our analysis we don't make difference between the edge (v, w) and (w, v), which means that the edges don't have a direction, that is to say that the graphs we consider are *undirected graph*.

If the number of vertices and edges is finite it is possible to define the *order* and the *size* of a graph by the number of vertices and edges respectively, i.e. $\text{order}(G) = |V|$ and $\text{size}(G) = |E|$.

Given a vertex $v \in V$, the *degree* of v is given by the number of edges that connect to it. A graph G is said to be a *simple graph* if $(v, v) \notin E$ for each $v \in V$ and there is no more than one edge between two different vertices.

A *path* is a sequence of edges connecting a sequence of vertices, it is denoted by

$$P = \{(v_1, v_2), (v_2, v_3), \ldots, , (v_{n-1}, v_n)\}, \tag{1.34}$$

where $v_1, \ldots, v_n \in V$. We say that v_1 and v_n are the start and the end vertex of the path while $v_2, \ldots, v_{n-1}$ are its internal vertices. A path P is said to be a *simple path* if it doesn't cross any vertex more than once, i.e. $v_i \neq v_j$ for each $i, j = 1, \ldots, n \ i \neq j$. Among the not simple path we distinguish the *cycles* as those paths having $v_1 = v_n$.

Definition 1.1.4. *The length of a path is given by the the number of edges it employs, i.e. $length(P) = |P|$. The distance between two vertices is determined by the minimal length of a path connecting them, if there isn't any connecting path between them the distance is established to be ∞.*

If it is possible to find a path between v and w for each $v, w \in V$, $v \neq w$, we say that the graph is *connected*.

It is now possible to define the *tree*, a special kind of graph that will be largely employed in the following chapters.

Definition 1.1.5. *A tree is an undirected simple graph G such that G is connected and there exist no cycles in G.*

Given a tree G, a vertex of G having degree 1 is called *leaf*. On the other

hand a vertex with degree at least 2 is said to be an *internal vertex*.

Sometimes it is identified a special vertex in V, the *root*. If this is the case, the edges possess a natural orientation, indeed an edge belongs to one and only one path starting from the root, and this path can be crossed in a direction going towards or away from the root.

In a *rooted tree* it is defined a partial ordering on V, indeed we say that $v < w$ if the unique path going from the root towards w passes through v. In particular, if $v < w$ and $(v, w) \in E$, we say that v is the *parent* of w and w is a *child* of v. We observe that the parent is only one, while the children can be many.

A special class of the trees is composed by the *binary trees*. For this class of trees, the children associated to a parent are always 2, and take the names of *left child* and *right child*.

Chapter 2

Markovian Binary Trees

2.1 Continuous time Markovian multitype branching processes

A *multitype branching process* is a process which describes in a mathematical way the evolution of a population of individuals. These individuals may belong to a distinct, finite number of different typologies. For simplicity the various types are denoted by the elements of the set $\mathcal{M} = \{1, \ldots, m\}$.

Every individual is supposed to evolve in an independent way and during its life may generate a random number of offsprings following specific stochastic rules. Once an offspring is generated, this evolves following the rules inherited by its parent and possibly gives birth to its own offsprings. This process continues until there is at least an individual alive. Classical theory related to this topic can be found in [3, 13, 30], wherein are also illustrated many examples of practical application of this process.

- Demographic model: the main interest is the study of the aging structure of a certain population. Usually the different types represent the various age ranges, see [35].

- Environmental niches model: in biology sometimes some individual belonging to a certain niche may generate offspring belonging to different niches, for a detailed survey see [34].

- Population genetics model: the objective is the study of the evolution of a certain genotype or phenotype in a certain family considering that the offspring is affected by individuals which are external to the considered family, see [20].

- Physics: many examples can be found in physics, for instance in the setting of cosmic ray cascades the electrons and the photons generate each other.

In more detail, in a multitype branching process, at the moment of its death, an individual of type i eventually generates offsprings of the various types according to a certain probability distribution on $\mathcal{S}_m$, see (1.1). Indeed, given $\mathbf{j} \in \mathcal{S}_m$, the probability

$$p_i(\mathbf{j}) = p_i(j_1, \ldots, j_m), \tag{2.1}$$

denotes the probability that a type i individual generates at its death j_k descendants of type k for each $k \in \mathcal{M}$.

In order to properly investigate the features of a multitype branching process we need to employ the probability generating functions, introduced in (1.17), which regulate the probabilistic behavior of the reproduction. In fact a key role is performed by the functions $f_1, \ldots, f_m$, where f_i represents the probability generating function of the random variable indicating the number of offsprings of the various types generated by an individual of type i,

$$f_i(\mathbf{s}) = \sum_{\mathbf{j} \in \mathcal{S}_m} p_i(\mathbf{j}) \mathbf{s}^{\mathbf{j}}, \tag{2.2}$$

where $\mathbf{s} = (s_1, \ldots, s_m) \in [0, 1]^m$ and $\mathbf{s}^{\mathbf{j}} = s_1^{j_1} \cdots s_m^{j_m}$.

For any vector $\mathbf{i} \in \mathcal{S}_m$ it is well defined the continuous time Markov process

$$\{\mathbf{Z}^{\mathbf{i}}(t), \ t \geq 0\}, \tag{2.3}$$

on the state space $\mathcal{S}_m$, where the vector $\mathbf{Z}^{\mathbf{i}}(t) = (Z_1^{\mathbf{i}}(t), \ldots, Z_m^{\mathbf{i}}(t))$ is such that $Z_k^{\mathbf{i}}(t)$ represents the number of individuals of type k alive at time t

given that, the starting configuration of individuals was given by the entries of vector $\mathbf{i}$, i.e. $\mathbf{Z^i}(0) = \mathbf{i}$. We omit the superscript $\mathbf{i}$ whenever the context doesn't require it.

For each pair of vectors $\mathbf{i}, \mathbf{j} \in \mathcal{S}_m$, it is possible to define the probability $P^{\mathbf{i}}(\mathbf{j}; t)$ by

$$P^{\mathbf{i}}(\mathbf{j}; t) = \mathrm{P}[\mathbf{Z}(\tau + t) = \mathbf{j} \mid \mathbf{Z}(\tau) = \mathbf{i}] = \mathrm{P}[\mathbf{Z^i}(t) = \mathbf{j}], \qquad (2.4)$$

for each $\tau \geq 0$, where the last equality holds since the probability distribution of the offsprings is supposed to be homogeneous in time. The probability generating function of the variable $\mathbf{Z^i}(t)$ is denoted by $F^{\mathbf{i}}(\mathbf{s}; t)$ and may be expressed as

$$F^{\mathbf{i}}(\mathbf{s}; t) = \sum_{\mathbf{j} \in \mathcal{S}_m} P^{\mathbf{i}}(\mathbf{j}; t) \mathbf{s^j}, \qquad (2.5)$$

where $\mathbf{s} \in [0, 1]^m$. In the following we will be mostly interested in the cases wherein the starting population is given just by one individual. To this end we define the vector of functions $\mathbf{F}(\mathbf{s}, t)$ as

$$\mathbf{F}(\mathbf{s}, t) = (F^{\mathbf{e_1}}(\mathbf{s}; t), \dots, F^{\mathbf{e_m}}(\mathbf{s}; t)), \qquad (2.6)$$

whose entries may be interpreted, in some sense, as a canonical basis. Indeed these entries may generate any generic $F^{\mathbf{i}}(\mathbf{s}; t)$, as specified better after the following classical result.

Proposition 2.1.1. *Given $t, \tau > 0$, then*

$$\mathbf{F}(\mathbf{s}, t + \tau) = \mathbf{F}(\mathbf{F}(\mathbf{s}, \tau), t). \qquad (2.7)$$

Proof. Let's define $N_j^{k,i}(\tau)$ as the number of type j offsprings that are generated by the k-th individual of type i alive at time t. It is clear that

$$Z_j(t + \tau) = \sum_{i=1}^{m} \sum_{k=1}^{Z_i(t)} N_j^{k,i}(\tau). \qquad (2.8)$$

By passing to the generating functions the thesis is proven. □

In general if $\mathbf{Z}(t) = \mathbf{j} = (j_1, \ldots, j_m) \in \mathcal{S}_m$, then $\mathbf{Z}(t+\tau)$ may be obtained as the sum of $j_1 + \ldots + j_m$ independent random vectors, where j_k of them have generating function $F^{\mathbf{e}_k}(\mathbf{s}; \tau)$ for each $k \in \mathcal{M}$.

These generating functions $F^{\mathbf{e}_i}(\mathbf{s}; t)$ are often difficult to express for general multitype branching process. We need therefore to introduce the *multitype Markovian branching processes*. These processes have the peculiarity that the lifespan of an individual of type $i \in \mathcal{M}$ follows an exponential distribution of parameter α_i, see (1.20). In fact, with this restriction, the Kolmogorov forward and backward equations that govern the process and define $F^{\mathbf{e}_i}(\mathbf{s}; t)$ were completely deducted by Sevast'yanov in [38]. These formulas are quoted here for completeness.

Theorem 2.1.1 (Kolmogorov equations). *The following equations hold,*

- *Forward Kolmogorov equation:*

$$\frac{\partial}{\partial t} F^{\mathbf{e}_i}(\mathbf{s}; t) = \sum_{k=1}^{m} u^{(k)}(\mathbf{s}) \frac{\partial}{\partial s_k} F^{\mathbf{e}_i}(\mathbf{s}; t), \qquad (2.9)$$

  where $u^{(i)}(\mathbf{s}) = \alpha_i[f_i(\mathbf{s}) - s_i]$ and $\frac{1}{\alpha_i}$ is the average lifespan of a particle of type i.

- *Backward Kolmogorov equation:*

$$\frac{\partial}{\partial t} F^{\mathbf{e}_i}(\mathbf{s}; t) = u^{(i)}[\mathbf{F}(\mathbf{s}; t)], \qquad (2.10)$$

A central role in the theory of the branching processes is hold by the mean matrix, which may be defined in case of *not explosive* processes, that is in the case where

$$\frac{\partial}{\partial s_j} f_i(\mathbf{s})|_{\mathbf{s}=\mathbf{e}} < \infty, \qquad (2.11)$$

for each $i, j \in \mathcal{M}$, where the functions $f_i(\mathbf{s})$ are defined in (2.2). It is shown in [3] that such a condition is sufficient to verify

$$E[Z_j(t)|\mathbf{Z}(0) = \mathbf{e}_i] < \infty, \qquad (2.12)$$

for each $i, j \in \mathcal{M}$ and $t \geq 0$. It is now possible to define the mean matrix we mentioned before,

Definition 2.1.1 (Mean matrix). *The mean matrix is the $m \times m$ matrix*

$$M(t) = (m_{ij}(t))_{i,j=1,\ldots,m},$$

where,

$$m_{ij}(t) = E[Z_j(t) \mid \mathbf{Z}(0) = \mathbf{e}_i]$$

.

It can be shown by means of (2.7) that these matrices fulfill

$$M(t + \tau) = M(t)M(\tau), \tag{2.13}$$

so that the set of the mean matrices, $\{M(t), t \geq 0\}$, forms a semi-group, i.e. a set with an associative binary operation. Moreover, thanks to (2.10), it can be shown that the continuity condition

$$\lim_{t \to 0} M(t) = I_m, \tag{2.14}$$

is satisfied too. These properties together guarantee the semi-group $\{M(t), t \geq 0\}$ to be generated by a certain matrix $\Omega \in \mathbb{R}^{m \times m}$, such that for every $t > 0$, the equality

$$M(t) = \exp(\Omega t) \tag{2.15}$$

holds. An interesting interpretation may be given to the entries of the matrix Ω. In fact Ω_{ij} represents the rate at which a particle of type i influences the whole number of type j individual. This intuition leads us to express the entries of the matrix Ω by the formula

$$\Omega_{ij} = \alpha_i b_{ij}, \tag{2.16}$$

where α_i is the average lifespan of a type i individual and

$$b_{ij} = \frac{\partial f_i}{\partial s_j}(\mathbf{s})|_{\mathbf{s}=\mathbf{e}} - \delta_{ij}, \tag{2.17}$$

21

where δ_{ij} is the Kronecker's delta having value 1 if $i = j$ and 0 otherwise. We observe that b_{ij} represents the expected variation in the number of type j individuals alive just after the death of a type i individual. It is possible to give some other definitions in order to identify better the various typologies of branching processes,

Definition 2.1.2. *A multitype branching process is called positively regular if there exists $t_0 > 0$ such that $M(t_0) > 0$, where the matrices $M(t)$ are defined in 2.1.1.*

It is now important to quote the Perron-Frobenius theorem that we will largely employ later, for a proof we refer to [6],

Theorem 2.1.2 (Perron-Frobenius Theorem. Positive matrices case)**.** *A positive matrix M has a maximum modulus eigenvalue which is positive, real, simple and equal to the spectral radius of M. All the other eigenvalues are strictly smaller in modulus. Moreover, there exist positive right and left eigenvectors associated to such eigenvalue.*
The same statement holds even if we require the matrix to be nonnegative, irreducible and aperiodic.

By applying this theorem to the positive matrix $M(t_0)$, we find that this matrix has a maximum modulus real eigenvalue $\lambda(t_0)$ and that every other eigenvalue $\lambda(t_0)^{(i)}$ of $M(t_0)$, is such that $|\lambda(t_0)^{(i)}| < \lambda(t_0)$. We observe that it must exist λ eigenvalue of Ω such that $\lambda(t_0) = \exp(\lambda t_0)$, i.e.

$$\lambda = \frac{\log \lambda(t_0)}{t_0}, \tag{2.18}$$

moreover λ is the eigenvalue of maximal modulus of Ω, and due to the Perron-Frobenius theorem applied to $M(t_0)$, must also be simple and real.
Another important definition is the following,

Definition 2.1.3 (Singular process)**.** *A multitype branching process is said to be singular if the generating functions $f_1(\mathbf{s}), \ldots, f_m(\mathbf{s})$, defined in (2.2), are all linear in $s_1, \ldots, s_m$, with no constant term.*

It is now possible to quote another important result and its corollary, a proof can be found (for the discrete time case) in [13, 30],

Theorem 2.1.3. *Suppose that the process is positive regular and non singular. Then each vector $\mathbf{j} \in \mathcal{S}_m$, apart from $\mathbf{j} = \mathbf{0}$, is a transient state of the Markov chain $\{\mathbf{Z}(t),\ t \geq 0\}$.*

And in particular it leads to the important following result,

Corollary 2.1.1. *Under the hypotheses of theorem 2.1.3, for each $\mathbf{r} \in \mathcal{S}_m$ such that $\mathbf{r} \neq \mathbf{0}$ it holds that*

$$\lim_{t \to \infty} P[\mathbf{Z}(t) = \mathbf{r}] = 0. \tag{2.19}$$

This means that, under these simple hypotheses, the process will either die out or explode almost surely.

2.1.1 Extinction Probabilities

The population becomes extinct if for some $t \geq 0$ it happens that $\mathbf{Z}(t) = \mathbf{0}$, and it is of a certain interest determining the probability of this event given that $\mathbf{Z}(0) = \mathbf{e}_i$ for each $i \in \mathcal{M}$. Let

$$q_i = \mathrm{P}[\exists\, T \geq 0 \text{ such that } \mathbf{Z}(T) = \mathbf{0}|\mathbf{Z}(0) = \mathbf{e}_i]. \tag{2.20}$$

We observe that, due to the absorbing property of state $\mathbf{0}$, it is straightforward that

$$\mathbf{Z}(t) = \mathbf{0} \Rightarrow \mathbf{Z}(t + \tau) = \mathbf{0}, \quad \forall \tau \geq 0,$$

therefore if we define

$$q_i(t) = P[\mathbf{Z}(t) = \mathbf{0}|\mathbf{Z}(0) = \mathbf{e}_i], \tag{2.21}$$

then we have immediately that for every $\tau \geq 0$ the relation $q_i(t) \leq q_i(t + \tau)$ holds. So, it is equivalent to define the extinction probability q_i as

$$q_i = \lim_{t \to \infty} q_i(t), \tag{2.22}$$

for each $i \in \mathcal{M}$. By means of the backward Kolmogorov equation (2.10) it can be shown, see [3], that the extinction probability vector is the minimal nonnegative solution of

$$\mathbf{u}(\mathbf{s}) = 0, \tag{2.23}$$

where $\mathbf{u}(\mathbf{s}) = (u^{(1)}(\mathbf{s}), \dots, u^{(m)}(\mathbf{s}))$ is defined in (2.9). This is the same as requiring that $\mathbf{q}$ is the minimal nonnegative vector such that the equality

$$f_i(\mathbf{q}) - q_i = 0, \tag{2.24}$$

holds for each $i \in \mathcal{M}$, where the function $f_i(\mathbf{s})$ is the generating function defined in (2.2).

A fundamental theorem that has to be quoted is this one, see [30, Theorem 7.1],

Theorem 2.1.4. *Let Ω be the generator of the semi-group $\{M(t),\ t \geq 0\}$, and suppose the process to be positive regular, let λ be the eigenvalue of the matrix Ω defined in (2.18), then:*

1. if $\lambda \leq 0$, then $\mathbf{q} = \mathbf{e}$,

2. if $\lambda > 0$, then $\mathbf{q} < \mathbf{e}$.

In any case, $\mathbf{q}$ is equal to the minimal nonnegative solution of the vector equation

$$\mathbf{f}(\mathbf{s}) = \mathbf{s}, \tag{2.25}$$

where $\mathbf{f}(\mathbf{s}) = (f_1(\mathbf{s}), \dots, f_m(\mathbf{s}))$

Due to the key role played by the equation (2.25), we name it the *extinction equation*. We observe that this vector equation is simply a compact reformulation of the set of equations (2.24).

The Theorem 2.1.4 leads us to distinguish the various Markovian multitype branching processes on the basis of the value of the eigenvalue λ.

Definition 2.1.4. *Given λ the eigenvalue of the matrix Ω determined in (2.18) we say that*

- *if $\lambda < 0$ the process is subcritical,*

- *if $\lambda = 0$ the process is critical,*

- *if $\lambda > 0$ the process is supercritical.*

Obviously we will be mostly interested in the computation of the extinction probabilities in the non trivial cases, that is to say in presence of supercritical processes.

2.2   Transient Markovian arrival processes

A *Markovian arrival process* (MAP) is a two dimensional continuous time Markovian process $\{(N(t), \varphi(t)),\ t \geq 0\}$ in the space $(\{0\} \cup \mathbb{N}) \times \{1, .., m\}$, where m is finite. The process $\varphi(t)$ is called the *phase process*. The transition between the phases may be *observable* or *hidden*, and the distinction between these kind of transitions is determined by the happening of a certain event during the transition. The process $N(t)$ is the counter of the observable transitions that happen in the time interval $[0, t]$.

In a MAP the range of the possible transitions that can originate from a certain state (n, i) is restricted to those pointing towards the states belonging to the union of the sets $\{(n, j),\ 1 \leq j \leq n,\ j \neq i\}$ and $\{(n + 1, j),\ 1 \leq j \leq n\}$. The transfer towards one of these two sets determine respectively the transition that are said to be hidden or observable.

Therefore the process is governed by two distinct matrices

$$D_1, D_0 \in \mathbb{R}^{m \times m}, \tag{2.26}$$

which determine the transition rates in case respectively of occurrence of an event or not, i.e. the values of $(D_0)_{ij}$ and of $(D_1)_{ij}$ represent the rates of transition associated to moving from the state (n, i) to the state (n, j) and $(n + 1, j)$ respectively. So, the whole rate matrix which generates the MAP

may be expressed by

$$
Q = \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \\ \vdots \end{array}
\begin{pmatrix}
D_0 & D_1 & 0 & 0 & \cdots \\
0 & D_0 & D_1 & 0 & \\
0 & 0 & D_0 & D_1 & \ddots \\
0 & 0 & 0 & D_0 & \ddots \\
\vdots & & \ddots & \ddots & \ddots
\end{pmatrix},
\tag{2.27}
$$

where the levels define the different values of the process $N(t)$.

Since Q is a generator matrix for a continuous time Markov chain, it has to verify the condition $Q\mathbf{e} = 0$, which means that

$$
(D_0)_{ii} = -(\sum_{j=1, j \neq i}^{m} (D_0)_{ij} + \sum_{j=1}^{m} (D_1)_{ij}).
\tag{2.28}
$$

Given that $N(0) = 0$, we say that the process just described is a MAP$(\boldsymbol{\alpha}, D_0, D_1)$, where $\boldsymbol{\alpha}$ denotes the initial probability distribution on the phase space, i.e. $\alpha_i = \mathrm{P}[\varphi(0) = i]$.

We are now interested in describing the *transient MAP* [27]. We add to the phase space an additional *absorbing phase*, denoted by 0. This phase is absorbing in the sense that, when the phase process enters the phase 0, no more transition will occur. In other worlds, if $\varphi(t_0) = 0$, then for every $t \geq t_0$ we have that $\varphi(t) = 0$ and $N(t) = N(t_0)$, hence the process is said to be ended. We introduce the vector $\mathbf{d} \in \mathbb{R}^m$ such that, for each $i \in 1, \ldots, m$, the entry d_i represent the rate of transition from phase i to phase 0.

In a transient MAP the matrices D_0 and D_1 which form the (2.27) are extended respectively by D_0^* and $D_1^* \in \mathbb{R}^{(m+1) \times (m+1)}$ defined as

$$
D_0^* = \begin{pmatrix} 0 & \mathbf{0}^T \\ \mathbf{d} & D_0 \end{pmatrix}, \qquad
D_1^* = \begin{pmatrix} 0 & \mathbf{0}^T \\ \mathbf{0} & D_1 \end{pmatrix}.
\tag{2.29}
$$

All the phases different from 0 are supposed to be transient, therefore eventually the MAP will enter its absorbing phase and the process ends. A MAP

with m transient phases is thus defined by the tuple $(D_0, D_1, \mathbf{d}, \boldsymbol{\alpha})$, where $D_0, D_1 \in \mathbb{R}^{m \times m}$ and $\mathbf{d}, \boldsymbol{\alpha} \in \mathbb{R}^m$.

Let's now consider a MAP in state (n, i), due to the properties of the continuous time Markov process, the time of permanence in such a state follows an exponential distribution of parameter $\nu_i = -(D_0)_{ii}$ independently of n, and at the end of this period it may occur:

- an hidden transition to state (n, j), where $j \neq 0, i$, with probability $\frac{(D_0)_{ij}}{\nu_i}$;

- an observable transition to state $(n+1, j)$, where $j \neq 0$, with probability $\frac{(D_1)_{ij}}{\nu_i}$;

- a transition to the absorbing state $(n, 0)$, with probability $\frac{d_i}{\nu_i}$.

It is possible to introduce the sequence $\{T_1, T_2, ...\}$ as the sequence of the intervals between two consecutive observable events.

Let's now define a particular type of probability distribution, which is
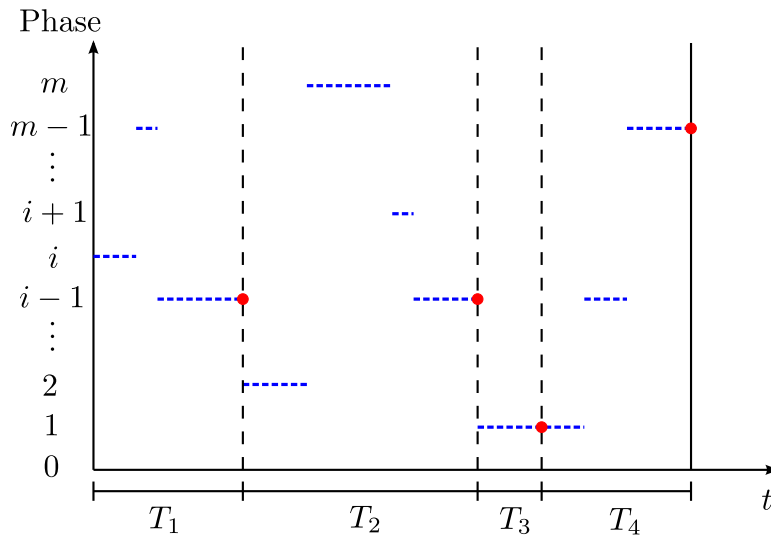

Figure 2.1: Phase process of a transient MAP that is absorbed after 3 observable transitions.

based on the transient MAP defined above. We consider a transient MAP with m transient phases, defined by $(D_0, D_1, \mathbf{d}, \boldsymbol{\alpha})$. Let's focus on the phase

only process $\varphi(t)$, see figure 2.2, which is a continuous time Markov process generated by the rate matrix $\tilde{Q}$ defined by

$$\tilde{Q} = \begin{pmatrix} 0 & \mathbf{0}^T \\ \mathbf{d} & D \end{pmatrix},$$

where $D = D_0 + D_1$. We are interested in the distribution of the random variable X representing the time necessary for the process to enter the absorbing state 0. Such a distribution is called *phase-type distribution* and we write $X \sim PH(\boldsymbol{\alpha}, D)$. This class of distribution is quite important for its versatility, indeed, as can be seen in [26], the phase-type distributions are dense into the family of positive valued distributions.

[SI POTREBBE FARE ESEMPI MAP (POISSON e PH-RENEWAL PROCESSES)]

2.3 Markovian binary trees

A *general Markovian tree* (GMT) is a mathematical model which allows to describe the life evolution of a set of individuals evolving independently one from the other. The term Markovian appears in the name because the life of each individual is governed by a transient Markovian arrival process as the one described in the previous section. When an observable transition occurs, in this setting, it means that offsprings are generated, with a distribution that is embedded in the parameters defining the MAP.

In the following analysis we will be mostly interested into a special kind of Markovian tree, the *Markovian Binary Tree* (MBT) introduced in [4]. The peculiarity of this tree is that the various individuals are allowed to generate only one descendant per observable transition.

2.3.1 Markovian Trees as continuous time Markov chain

A Markovian tree may be seen as a continuous time Markov process with states

$$X(t) = (N(t), \varphi_1(t), ..., \varphi_{N(t)}(t)), \tag{2.30}$$

in the space $\bigcup_{k=0}^{\infty}\{\{k\}\times\{1,...,m\}^k\}$, where $N(t)$ represents the total number of individuals alive at time t and $\varphi_k(t) \in \{1,...,m\}$ indicates the phase that the k-th individual possesses at that time.

In order to understand the dynamics that are admitted in an MBT model, here we present in detail all the possible transitions that may occur. Let's suppose that the process stands in the state defined by the configuration

$$
\begin{array}{ccccccccc}
(M, & a, & b, & \cdots & c, & r, & d, & \cdots & e), \\
 & \uparrow & \uparrow & & \uparrow & \uparrow & \uparrow & & \uparrow \\
 & 1 & 2 & \cdots & i-1 & i & i+1 & \cdots & M
\end{array}
$$

where the second line identifies the various individuals in the tree. We recall that the life of each individual is governed by the same transient MAP with parameters $(D_0, D_1, \mathbf{d}, \boldsymbol{\alpha})$. In particular, focusing on the individual i, it may occur:

1. An hidden transition: the phase of individual i changes from phase r to phase $s \neq 0$, this happens with rate $(D_0)_{rs}$, and the process moves to the state having configuration

$$
\begin{array}{ccccccccc}
(M, & a, & b, & \cdots & c, & s, & d, & \cdots & e), \\
 & \uparrow & \uparrow & & \uparrow & \uparrow & \uparrow & & \uparrow \\
 & 1 & 2 & \cdots & i-1 & i & i+1 & \cdots & M
\end{array}
$$

2. An absorption: the phase of individual i changes from phase r to the absorbing phase 0, causing the individual i to die. This happens with rate d_r, and the process moves to the state having configuration

$$
\begin{array}{cccccccc}
(M-1, & a, & b, & \cdots & c, & d, & \cdots & e), \\
 & \uparrow & \uparrow & & \uparrow & \uparrow & & \uparrow \\
 & 1 & 2 & \cdots & i-1 & i & \cdots & M-1
\end{array}
$$

3. An observable transition: the individual i generates a descendant. We suppose for instance that in the process of reproduction the phase of individual i changes from phase r to phase s and that the new born

29

individual begin its life in phase σ. We define the matrix $B \in \mathbf{R}^{m \times m^2}$, called the *birth rate* matrix, whose entries define the rates of occurrence of a transition of this kind. The described transition happens with rate $B_{r,(s,\sigma)}$ causing the process to move to the state having configuration

$$
\begin{array}{ccccccccc}
(M+1, & a, & b, & \cdots & c, & s, & \sigma & d, & \cdots & e). \\
 & \uparrow & \uparrow & & \uparrow & \uparrow & \uparrow & \uparrow & & \uparrow \\
 & 1 & 2 & \cdots & i-1 & i & i+1 & i+2 & \cdots & M+1
\end{array}
$$

Therefore, in general the whole process is ruled by the parameters $(D_0, B, \mathbf{d})$. The birth rate matrix B is given by

$$
B_{i,(j,k)} = (D_1)_{i,j} p_{i|(j,k)}, \tag{2.31}
$$

where $p_{i|(j,k)}$ is the conditional probability that a child starts its life in phase j, given that its parent was in phase i and has made a transition to phase k. In the case the probability distribution of the starting phase of the newborn individual does not depend on the type of the parent, B can be fully obtained by the parameters of the transient MAP $(D_0, D_1, \mathbf{d}, \boldsymbol{\alpha})$, indeed

$$
B_{i,(j,k)} = (D_1)_{i,j} \alpha_k. \tag{2.32}
$$

2.3.2 MBT as a multitype Markovian branching process

We would like to exploit the branched structure of the Markovian Tree in order to model it as branching process similar to the one described in Section 2.1. There are many ways to do it, here we present the two that are more intuitive.

In this first model we set as the types of the multitype branching process the different phases of the MAP, so that $\mathcal{M} = \{1, \ldots, m\}$. A branch, which starts with a type i individual, lives for an exponentially distributed time with parameter $\nu_i = -(D_0)_{ii}$ before running into a branching point, allowing the process evolve. The evolution may be described by means of the probabilities
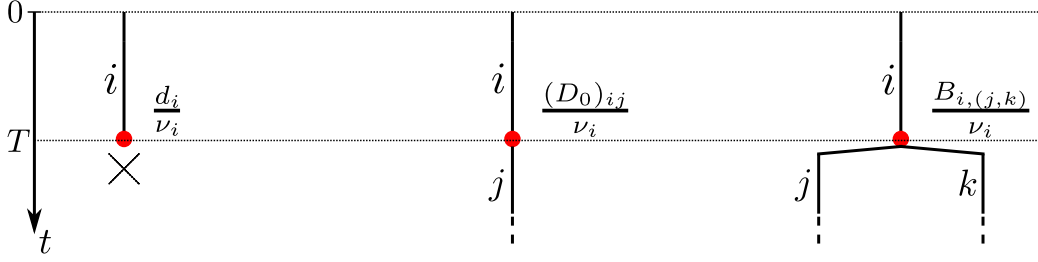
Figure 2.2: MBT as a Markovian branching process. The time T is a random variable following an exponential distribution of parameter $\nu_i = -(D_0)_{ii}$.

introduced in (2.1), indeed we have that the possible outcomes of a branching point are showed in figure 2.3.2 and occur with the following probabilities

$$
p_i(\mathbf{j}) = \begin{cases}
\frac{d_i}{\nu_i} & \text{if } \mathbf{j} = \mathbf{0} \\
\frac{(D_0)_{ij}}{\nu_i} & \text{if } \mathbf{j} = \mathbf{e}_j \quad \forall j = 1, \ldots, m \text{ and } j \neq i \\
\frac{B_{i,(j,k)}}{\nu_i} & \text{if } \mathbf{j} = \mathbf{e}_j + \mathbf{e}_k \quad \forall j, k = 1, \ldots, m \\
0 & \text{otherwise,}
\end{cases}
\tag{2.33}
$$

where B is the birth matrix defined in (2.31) and the process is completely determined by the parameters $(D_0, B, \mathbf{d})$. Since this is a multitype Markovian branching process, we can take advantage of the properties described in Section 2.1. In particular the probability generating functions $f_i(\mathbf{s})$ for $i \in \mathcal{M}$, see (2.2), take this form

$$
f_i(\mathbf{s}) = \frac{d_i}{\nu_i} + \sum_{j=1, j \neq i}^{m} \frac{(D_0)_{ij}}{\nu_i} s_j + \sum_{j,k=1}^{m} \frac{B_{i,(j,k)}}{\nu_i} s_j s_k,
\tag{2.34}
$$

or in vectorial form

$$
\mathbf{f}(\mathbf{s}) = (\tilde{D}_0)^{-1}\mathbf{d} + (\tilde{D}_0)^{-1}(D_0 + \tilde{D}_0)\mathbf{s} + (\tilde{D}_0)^{-1}B(\mathbf{s} \otimes \mathbf{s}),
\tag{2.35}
$$

where $\tilde{D}_0$ is the diagonal matrix with entries $\nu_1, \ldots, \nu_m$ and $\otimes$ denotes the Kronecker operator, see Definition 1.1.1. Moreover, since the process is not explosive, see (2.11), it is possible to define the mean matrix $M(t)$ as in Definition 2.1.1 and in particular the generator of the mean matrices semi-

group Ω, see (2.15). Indeed Ω can be expressed, according to (2.16), by the formula

$$\Omega_{ij} = \nu_i(\frac{\partial f_i}{\partial s_j}(\mathbf{s})|_{\mathbf{s}=\mathbf{e}} - \delta_{ij}) = (D_0)_{ij} + \sum_{k=1}^{m}(B_{i,(j,k)} + B_{i,(k,j)}), \qquad (2.36)$$

or in a vectorial form by

$$\Omega = D_0 + B(I_m \otimes \mathbf{e} + \mathbf{e} \otimes I_m). \qquad (2.37)$$

As we observed in Theorem 2.1.4, the extinction probabilities of a multitype Markovian branching process, defined in (2.20) may be obtained by solving the extinction equation. By modeling the MBT as just described, the extinction equation (2.25) assumes the following expression,

$$\begin{aligned}
0 = f_i(\mathbf{s}) - s_i &= \\
&= \frac{1}{\nu_i}(d_i + \sum_{k=1,k\neq i}^{m}(D_0)_{ik}s_k + \sum_{j,k=1}^{m}B_{i,jk}s_js_k + (D_0)_{ii}s_i) \\
&= \frac{1}{\nu_i}(d_i + \sum_{k=1}^{m}(D_0)_{ik}s_k + \sum_{j,k=1}^{m}B_{i,jk}s_js_k),
\end{aligned}$$

for each $i \in 1, \ldots, m$. So that the extinction equation (2.25) may be written in this simple vectorial form

$$\mathbf{0} = \mathbf{d} + D_0\mathbf{s} + B(\mathbf{s} \otimes \mathbf{s}). \qquad (2.38)$$

In order to find the minimal nonnegative solution of this equation we need numerical methods that we will present later.

2.3.3 MBT as a general branching process

Another way to interpret the MBT as a branching process is by taking the cue from the intrinsic branching structure of the relations developing between parents and descendants. In this model a branching point is created only if an individual either generates offsprings or dies, while in contrast with the

model suggested in Section 2.3.2, the simple changes of type happen along the edges and don't generate branching points. This is to say that, if the lifespan
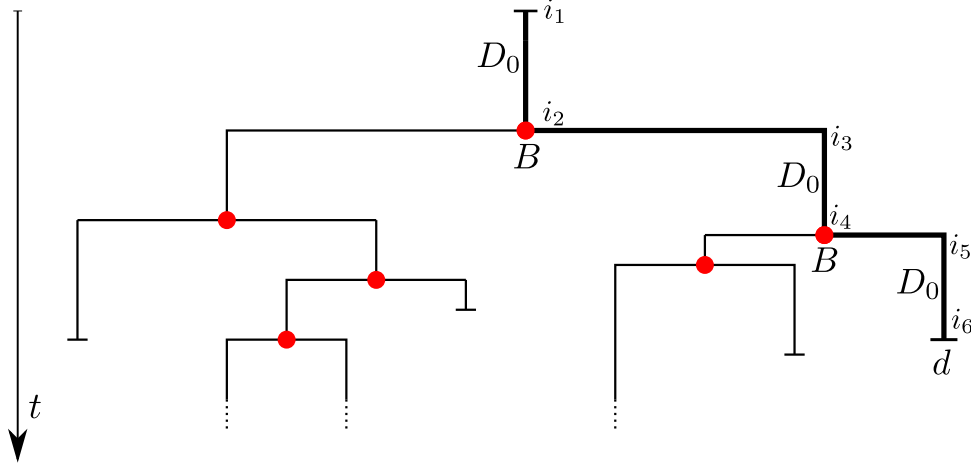

Figure 2.3: MBT as a general branching process. The time running between consecutive branching is a random variable following a general distribution given by the sum of various exponential distribution. With the thick line is pointed out the life span of the starting individual.

is controlled by a transient MAP $(D_0, B, \mathbf{d})$, a branching point coincides with either an observable transition or an absorption, while the hidden transitions are scattered within the branches and don't generate branching point. This behavior is shown in figure 2.3.3, where it is presented a possible evolution of the process in case it is ruled by a MAP.

In such a case the length of an edge is a random variable having a value given by the sum of many exponential probability distribution. We present here the probabilities defined in (2.1) hoping to get some hints regarding the evolution of the process.

$$p_i(\mathbf{j}) = \begin{cases} \theta_i & \text{if } \mathbf{j} = \mathbf{0} \\ \psi_{i,(j,k)} & \text{if } \mathbf{j} = \mathbf{e}_j + \mathbf{e}_k \quad \forall j, k = 1, \ldots, m \\ 0 & \text{otherwise,} \end{cases} \tag{2.39}$$

where we define $\boldsymbol{\theta} \in \mathbb{R}^m$ as the vector whose component θ_i represents the probability that a branch starting in type i dies without generating any off-

33

springs, conditioned that its branch has encountered a branching point. Furthermore, we denote $\Psi \in \mathbb{R}^{m \times m^2}$ as the matrix whose entry $\psi_{i,(j,k)}$ represents the probability that a branch starting in i individual generates a descendant starting in type k and change its own type to type j, conditioned that its branch has encountered a branching point.

The vector $\mathbf{f}(\mathbf{s})$, whose entries are the probability generating functions defined in (2.2), may be expressed in the following way

$$\mathbf{f}(\mathbf{s}) = \boldsymbol{\theta} + \Psi(\mathbf{s} \otimes \mathbf{s}). \tag{2.40}$$

Unfortunately, we are not allowed to use the results relative to the multitype Markovian branching processes obtained in section 2.1, indeed the length of an edge does not follow an exponential distribution so that the branching process is not Markovian.

2.4 Algorithmic approach for computing the extinction probabilities

In the previous section we observed how, by interpreting the MBT as a multitype Markovian branching process, see section 2.3.2, it is possible obtain the extinction probability vector $\boldsymbol{q}$, as the minimal nonnegative solution of the extinction formula expressed in (2.38), that we quote here for completeness

$$\mathbf{0} = \mathbf{d} + D_0 \mathbf{s} + B(\mathbf{s} \otimes \mathbf{s}). \tag{2.41}$$

Since D_0 is a non singular M-matrix, see Definition 1.1.3, this equation may be rewritten in the form

$$\mathbf{s} = \boldsymbol{\theta} + \Psi(\mathbf{s} \otimes \mathbf{s}), \tag{2.42}$$

where $\boldsymbol{\theta} = (-D_0)^{-1}\boldsymbol{d}$ and $\Psi = (-D_0)^{-1}B$ have the same meaning of the parameters defined in the equation (2.39).

The following algorithms provide general strategies for the computation of the minimal nonnegative solution of a vectorial quadratic equation having the

form of equation (2.42). However for some of these algorithms it is possible to give a probabilistic interpretation which may give us some intuitions about how the algorithms actually work in practice.
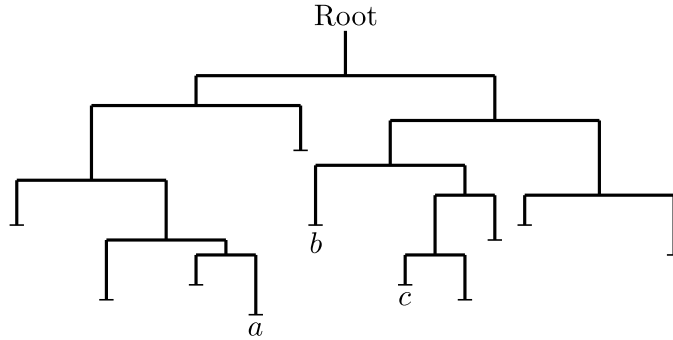

Figure 2.4: The evolution of a process depicted as described in Section 2.3.3. This process becomes extinct, in fact all the individuals dies after a while. The process can be interpreted as a rooted tree, see the Definition 1.1.5, wherein the vertices are identified by the birth of the starting individual (the root), the branching points (the internal vertices) and the deaths of any individual (the leaves)

2.4.1 Depth algorithm

The *depth algorithm* was proposed in the thesis of Nectarios Kontoleon [25] and further analyzed in [4] and [14], which focused respectively on the probabilistic and on the analytical aspects of the algorithm. From the algorithmic point of view, the depth algorithm is simply a fixed point iteration applied to equation (2.42). Indeed we are allowed to define the sequence

$$
\boldsymbol{q}_n^{(d)} = \begin{cases} \boldsymbol{\theta}, & \text{if } n = 0 \\ \boldsymbol{\theta} + \Psi(\boldsymbol{q}_{n-1}^{(d)} \otimes \boldsymbol{q}_{n-1}^{(d)}), & \text{if } n > 0, \end{cases} \tag{2.43}
$$

which is proved to be a monotonically increasing sequence converging to the extinction probability vector $\boldsymbol{q}$. In order to study the convergence, we define the approximation error by the formula

$$
\mathbf{E}_n^{(d)} = \boldsymbol{q} - \boldsymbol{q}_n^{(d)} \geq \mathbf{0}, \tag{2.44}
$$

and the following theorem is proved to be true, see [14, Theorem 2.3.4.]

Theorem 2.4.1. *If the MBT is not critical, then an upper bound for the approximation error for Depth algorithm is given by*

$$\mathbf{E}_n^{(d)} \leq \Psi(\boldsymbol{q} \otimes I + I \otimes \boldsymbol{q})\mathbf{E}_{n-1}^{(d)}. \tag{2.45}$$

The criticality of a process is defined in Definition 2.1.4.

The convergence is assured, in the not critical case, by assuming that Ω, defined by (2.37), is irreducible, in fact this hypothesis is sufficient for the fulfillment of the following inequality

$$\rho(\Psi(\boldsymbol{q} \otimes I + I \otimes \boldsymbol{q})) < 1, \tag{2.46}$$

where $\rho(\cdot)$ denotes the spectral radius, see (1.5). Therefore a linear convergence is guaranteed.

We consider a tree like the one represented in Figure 2.4, the depth of a vertex of the tree, is given by its distance from the root, see Definition 1.1.4, for instance the depths of the vertices a, b and c are respectively $5, 3$ and 5. The *depth* of a tree is given by the maximal depth of one of its vertices, it is possible to check that the tree in Figure 2.4 possesses a depth of 5.

The depth algorithm is named in this way since we can interpret the sequence $\{\boldsymbol{q}_0^{(d)}, \boldsymbol{q}_1^{(d)}, \boldsymbol{q}_2^{(d)} \ldots\}$ as follows: the n-th entry of the sequence represents the vector containing the probability that a tree is extinct and its depth is at most equal to n. We say that a tree is extinct if the process which is represented by it, becomes extinct. This interpretation provides us with an intuitive justification that the sequence $\{\boldsymbol{q}_0^{(d)}, \boldsymbol{q}_1^{(d)}, \boldsymbol{q}_2^{(d)} \ldots\}$ is monotonically increasing and converges to the extinction probability vector, i.e.

$$\lim_{n \to \infty} \boldsymbol{q}_n^{(d)} = \boldsymbol{q}. \tag{2.47}$$

In fact each element of the sequence is given by the probability for the process to produce a tree satisfying two conditions: the tree must be extinct and its depth must be bounded, see Figure 2.4.2. By increasing n, more and more extinct trees fulfill the condition about the bounded depth, so that the
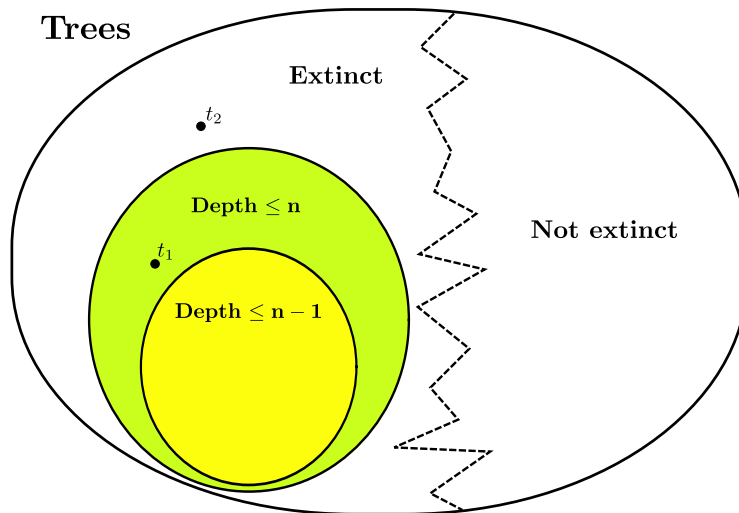
36

Figure 2.5: Graphic representation of what happens when we increase n in the depth algorithm.

sequence of probabilities $\{\boldsymbol{q}_n^{(d)}\}_{n\geq 0}$ is not decreasing in n. The convergence is assured by observing that a tree is extinct if and only if its depth is finite, therefore the tree t_2 in the figure 2.4.2 will be eventually absorbed by the subset of the depth bounded trees for a sufficiently large value of n.

Without delving in the details this algorithm may be also compared to the Neuts algorithm, which is an algorithm arising in the setting of the Quasi Birth and Death Markov chains, as described in [4].

2.4.2   Order algorithm

The *order algorithm* was proposed together with the depth algorithm in the thesis of Nectarios Kontoleon [25] and studied further in [4] and [14].

It works essentially in the same way the depth algorithm does, indeed, it is still a fixed point iteration. By employing (1.11), the equation (2.42) is rewritten in the equivalent form,

$$\mathbf{s} = [I - \Psi(\boldsymbol{s} \otimes I)]^{-1}\boldsymbol{\theta}, \tag{2.48}$$

37

and the fixed point iteration is applied to this version of the equation. It is observed that the matrix $I - \Psi(\boldsymbol{s} \otimes I)$ with $\boldsymbol{s} \in [0, 1]^m$ is a non singular M-matrix, see Definition 1.1.3, and therefore, thanks to the inverse positive property of the M-matrices $[I - \Psi(\boldsymbol{s} \otimes I)]^{-1} \geq 0$. We are now allowed to define the sequence

$$\boldsymbol{q}_n^{(o)} = \begin{cases} \boldsymbol{\theta}, & \text{if } n = 0 \\ [I - \Psi(\boldsymbol{q}_{n-1}^{(o)} \otimes I)]^{-1}\boldsymbol{\theta}, & \text{if } n > 0, \end{cases} \tag{2.49}$$

which is a monotonically increasing sequence. In order to study the convergence towards the extinction probability vector $\boldsymbol{q}$, we define the approximation error by the formula

$$\mathbf{E}_n^{(o)} = \boldsymbol{q} - \boldsymbol{q}_n^{(o)} \geq \mathbf{0}. \tag{2.50}$$

In parallel with Theorem 2.4.1, the following theorem is proved,

Theorem 2.4.2. *If the MBT is not critical, then an upper bound for the approximation error for Depth algorithm is given by*

$$\mathbf{E}_n^{(o)} \leq [I - \Psi(\boldsymbol{q} \otimes I]^{-1}\Psi(I \otimes \boldsymbol{q})\mathbf{E}_{n-1}^{(o)}. \tag{2.51}$$

The criticality of a process is defined in Definition 2.1.4 and for the proof of this theorem we refer to [14, Theorem 2.3.4.].

It is possible to prove through analytical arguments that, if the process is non critical the spectral radius of the matrix $[I - \Psi(\boldsymbol{q} \otimes I]^{-1}\Psi(I \otimes \boldsymbol{q})$ is smaller than 1, and in particular the following chain of inequalities holds

$$\rho([I - \Psi(\boldsymbol{q} \otimes I]^{-1}\Psi(I \otimes \boldsymbol{q})) \leq \rho(\Psi(\boldsymbol{q} \otimes I + I \otimes \boldsymbol{q})) < 1, \tag{2.52}$$

implying that the order algorithm converges faster than the depth algorithm, for a proof of this result see [14, Theorem 2.3.8.].

It is possible to prove the result of convergence even through probabilistic arguments, in parallel to what was done for the depth algorithm. We refer to Figure 2.4, the order of a certain vertex v is determined by number of left
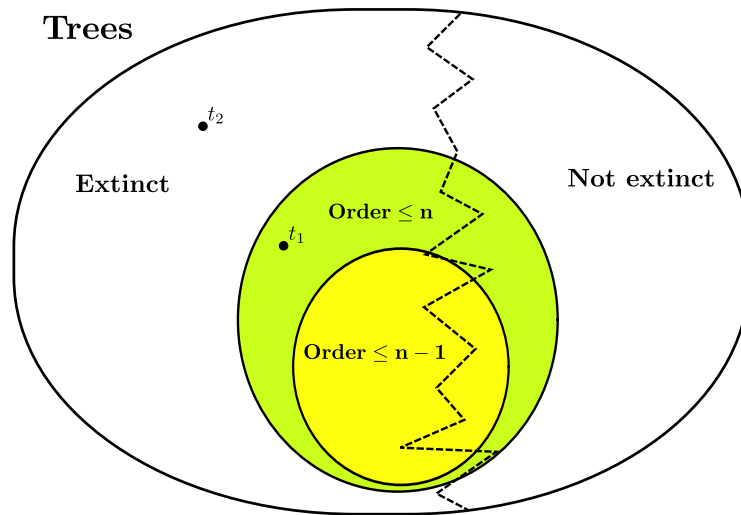
Figure 2.6: Graphic representation of what happens when we increase n in the order algorithm.

edges belonging to the unique path going between v and the root, where the branch connected to the root vertex is considered a left branch and must be counted.

In our setting, each vertex is associated to the individual whose life path passes through it, and the order of a vertex is given by the generation the associated individual belongs to, i.e. the root is said to be generation 0, the starting individual represents the first generation, the children of the starting individual represent the second generation, and so on. For instance, in Figure 2.4, the order of the leaves a, b and c are respectively $3, 3$ and 4. The *order* of a tree is given my the maximal order of one of its vertices, we observe that the tree in Figure 2.4 possesses a depth of 4.

The convergence to the effective extinction probability vector can be justified in an intuitive way with an argument similar to the one used for the depth algorithm, indeed an extinct tree necessarily have a finite order, even if it is not true the vice versa.

2.4.3   Thicknesses algorithm

The *thicknesses algorithm* is an immediate evolution of the order algorithm and was proposed by Hautphenne et al. in [17]. In fact, due to (1.11), equation (2.42) can be reformulated as

$$\mathbf{s} = [I - \Psi(I \otimes \boldsymbol{s})]^{-1}\boldsymbol{\theta}, \tag{2.53}$$

and the thicknesses algorithm works as a fixed point iteration applied in turn to (2.48) and (2.53). So that we can define the sequence

$$\boldsymbol{q}_n^{(t)} = \begin{cases} \boldsymbol{\theta}, & \text{if } n = 0, \\ [I - \Psi(\boldsymbol{q}_{n-1}^{(t)} \otimes I)]^{-1}\boldsymbol{\theta}, & \text{if } n > 0 \text{ and } n \text{ is even}, \\ [I - \Psi(I \otimes \boldsymbol{q}_{n-1}^{(t)})]^{-1}\boldsymbol{\theta}, & \text{if } n > 0 \text{ and } n \text{ is odd}, \end{cases} \tag{2.54}$$

which is a monotonic non decreasing sequence, due to the positive inverse property of the M-matrices, see Definition 1.1.3. Without delving into details, this algorithm is proved to converge at least linearly in the not critical case, and always faster than the depth algorithm, see [14]. A probabilistic interpretation exists also for this algorithm and is illustrated in [17]. We refer to figure 2.4 and observe that a path from the root to a vertex is given by a sequence of left and right edges, the thicknesses of a vertex is given by the number of times that this sequence of edges changes from right to left or left to right. The thicknesses of a tree is given by the maximum of the thicknesses of its vertex. For instance, in figure 2.4, the thicknesses of the vertices a, b and c is $3, 2$ and 4 respectively, while the tree possesses a thicknesses of 5.

2.4.4   Newton algorithm

The algorithms described until now were in practice fixed point iteration applied to various reformulation of (2.42), and all these converge linearly in the non critical case. It is possible to apply to the various reformulation of (2.42) also the classical Newton method for finding recursively better and better approximations of the extinction probability vector $\boldsymbol{q}$. For the details

concerning this approach we refer to the works [14, 16].

The equation (2.42), can be reformulated as $\mathscr{F}(\boldsymbol{q}) = \boldsymbol{0}$, where

$$\mathscr{F}(\boldsymbol{s}) = \boldsymbol{s} - \boldsymbol{\theta} - \Psi(\mathbf{s} \otimes \mathbf{s}). \tag{2.55}$$

The space $\mathbb{R}^m$ provided with any norm is a Banach space, and $\mathscr{F}$ is a mapping of $\mathbb{R}^m$ in itself. We say that the mapping $\mathscr{F} : \mathbb{R}^m \to \mathbb{R}^m$ is Fréchet differentiable at $\boldsymbol{x} \in \mathbb{R}^m$ if there exists a linear operator $A : \mathbb{R}^m \to \mathbb{R}^m$ such that

$$\lim_{\boldsymbol{h} \to \boldsymbol{0}} \frac{\|\mathscr{F}(\boldsymbol{x} + \boldsymbol{h}) - \mathscr{F}(\boldsymbol{x}) - A\boldsymbol{h}\|}{\|h\|} = 0 \tag{2.56}$$

the linear operator A is denoted by $\mathscr{F}'_{\boldsymbol{x}}$ and is called the Fréchet derivative of $\mathscr{F}$ at $\boldsymbol{x}$, for further details see [32]. It can be verified that the map defined in (2.55) is Frechét differentiable and that

$$\mathscr{F}'_x : \boldsymbol{s} \longmapsto \boldsymbol{s} - \Psi(\boldsymbol{s} \otimes \boldsymbol{x} + \boldsymbol{x} \otimes \boldsymbol{s}). \tag{2.57}$$

For a given starting point $\boldsymbol{q}_0^{(N)}$, it is now possible to define the sequence determined by the Newton method for the solution of $\mathscr{F}(\boldsymbol{x}) = \boldsymbol{0}$,

$$
\begin{aligned}
\boldsymbol{q}_{n+1}^{(N)} &= \boldsymbol{q}_n^{(N)} - (\mathscr{F}'_{\boldsymbol{q}_n^{(N)}})^{-1}(\mathscr{F}(\boldsymbol{q}_n^{(N)})) \\
&= \boldsymbol{q}_n^{(N)} - [I - \Psi(\boldsymbol{q}_n^{(N)} \otimes I + I \otimes \boldsymbol{q}_n^{(N)})]^{-1}[\boldsymbol{q}_n^{(N)} - \boldsymbol{\theta} - \Psi(\boldsymbol{q}_n^{(N)} \otimes \boldsymbol{q}_n^{(N)})] \\
&= [I - \Psi(\boldsymbol{q}_n^{(N)} \otimes I + I \otimes \boldsymbol{q}_n^{(N)})]^{-1}[\boldsymbol{\theta} - \Psi(\boldsymbol{q}_n^{(N)} \otimes \boldsymbol{q}_n^{(N)})]
\end{aligned}
$$

provided that $\mathscr{F}'_{\boldsymbol{x}}$ is not singular for each $\boldsymbol{x}$.

Theorem 2.4.3. *Given a not critical process and a starting point such that $\boldsymbol{0} \leq \boldsymbol{q}_0^{(N)} \leq \boldsymbol{\theta}$, the Newton sequence $\{\boldsymbol{q}_0^{(N)}, \boldsymbol{q}_1^{(N)}, \boldsymbol{q}_2^{(N)}, \ldots\}$ is such that:*

- *the sequence is well defined,*

- *the sequence is monotonically non decreasing, i.e. $\boldsymbol{q}_0^{(N)} \leq \boldsymbol{q}_1^{(N)} \leq \boldsymbol{q}_2^{(N)} \leq \ldots,$*

- *$\lim_{n \to \infty} \boldsymbol{q}_n^{(N)} = \boldsymbol{q},$*

where q is the minimal nonnegative solution of (2.42). Moreover there exists a positive constant C, such that

$$\|E_n^{(N)}\| \leq C \|E_{n-1}^{(N)}\|^2, \qquad (2.58)$$

where $E_n^{(N)} = q - q_n^{(N)}$, so that the Newton algorithm converges at least quadratically.

For the proof of this Theorem we refer to [16, Theorem 2.1.].
Here we briefly described only the Newton approach applied directly to equation (2.42), however similar methods arise by applying the Newton approach to (2.48) and (2.53), for further details see [14, Chapter 3].
A slightly different Newton approach based on an associate QBD process is described and analyzed in [18].
The Newton approaches proposed are proved to lead to methods converging quadratically if the process is not critical, see Definition 2.1.4. On the other hand, as we get close to critical processes the convergence rate decreases, becoming linear as soon as the process turns critical. We will see with the following algorithm how to partially bypass this inconvenience.

2.4.5 Optimistic algorithm

A completely different approach was proposed by Bini, Meini and Poloni in [7, 29]. Firstly, it is necessary to observe that the formula (2.42) may be rewritten in the following way

$$s = \theta + \mathscr{B}(s, s), \qquad (2.59)$$

where the notation $\mathscr{B}(s_1, s_2) = \Psi(s_1 \otimes s_2)$ shoots for pointing out that the map $(s_1, s_2) \longmapsto \Psi(s_1 \otimes s_2)$ is a bilinear vector valued map. It is also important to recall that, due to the definition of θ and Ψ, the vector e is always a solution of (2.59).
We set now $r = e - s$, and by substitution the equation (2.59) becomes

$$r = \mathscr{B}(r, s) + \mathscr{B}(s, r) - \mathscr{B}(r, r). \qquad (2.60)$$

In particular we know that $\boldsymbol{q}$, the the minimal nonnegative solution of (2.59), stands for extinction probability vector, so that the vector $\tilde{\boldsymbol{q}} = \boldsymbol{e} - \boldsymbol{q}$ may be interpreted as the survival probability vector, and must fulfill the equation (2.60).

Our objective becomes the computation of the probability that starting from a population of just one individual the process doesn't become extinct. This is the reason why this algorithm is called *optimistic algorithm*.

We define the linear application

$$H_{\boldsymbol{r}} : \boldsymbol{x} \longmapsto \mathscr{B}(\boldsymbol{x}, \boldsymbol{e}) + \mathscr{B}(\boldsymbol{e} - \boldsymbol{r}, \boldsymbol{x}), \tag{2.61}$$

so that (2.60) becomes

$$\boldsymbol{r} = H_{\boldsymbol{r}}\boldsymbol{r}. \tag{2.62}$$

We suppose Ω, defined in (2.37), to be irreducible, so that as soon as $\boldsymbol{r} \leq \boldsymbol{e}$, the matrix $H_{\boldsymbol{r}}$ is irreducible too. Furthermore $H_{\boldsymbol{r}}$ is even nonnegative so that by means of the Perron Frobenius theorem, see Theorem 2.1.2, it is possible to assure that $\rho(H_{\tilde{\boldsymbol{q}}}) = 1$ and $\tilde{\boldsymbol{q}}$ is a right positive eigenvector of matrix $H_{\tilde{\boldsymbol{q}}}$. Given a nonnegative irreducible matrix M, we define the operator $\mathrm{PV}(M)$ which compute the right eigenvector of M associated to the eigenvalue $\rho(M)$ conveniently normalized in a standard way, for the details we refer to [29]. We observe that the vector $\tilde{\boldsymbol{q}}$ must fulfill the following equality

$$\boldsymbol{r} = \mathrm{PV}(H_{\boldsymbol{r}}). \tag{2.63}$$

and in order to find $\tilde{\boldsymbol{q}}$ a fixed point iteration is applied to (2.63). So, given a starting approximation $\tilde{\boldsymbol{q}}_0$, it is possible to define the sequence $\{\tilde{\boldsymbol{q}}_0, \tilde{\boldsymbol{q}}_1, \tilde{\boldsymbol{q}}_2, \ldots\}$ wherein

$$\tilde{\boldsymbol{q}}_n = \mathrm{PV}(H_{\tilde{\boldsymbol{q}}_{n-1}}), \tag{2.64}$$

for each $n \geq 1$. In [7] a Newton approach is proposed for solving (2.63). Here it is sufficient to say that such an algorithm performs in some sense in a specular way compared to the standard Newton algorithms described in the previous section. Indeed the convergence of the algorithm is proved to be quadratic if the extinction probability vector is far from $\boldsymbol{0}$, and in

particular is quadratic if the process is critical. On the other hand, the rate of convergence decreases as the probability for the process to die out tends to 0, and become linear if the process survives forever with certain probability.

Chapter 3

Catastrophes

In the previous chapter we pointed out how the solution of the equation (2.42) gives us an easy expression for the characterization of the extinction probability vector. Such an equation holds for the simple MBT since the different branches of the tree evolve independently one from the other.

We would like now to surround the tree with an environment which influences at the same time all the individuals that belong to the branching process. In this case the equation (2.42) is no longer justified since there is no more independence in the evolution of the various individuals whose lives are now simultaneously conditioned by the external environment. Due to this reason the problem we would like to study now is a lot more complex and interesting than the analysis of the simple MBT.

3.1 An extended scenario: A process of catastrophes

In literature this problem has been studied mostly from a probabilistic point of view, on the other hand our approach means to consider a less general extension of the model with the purpose of being able to compute numerically at least some approximation of the parameters which influence the process. The external environment we consider is examined in detail by Latouche, Hautphenne and Nguyen in [15], it consists only of a temporal sequence of

times, $\{\tau_n, n \in \mathbb{N}\}$, such that each of these times coincide with the happening of a catastrophic event. When a catastrophe happens, it affects each individual alive which dies or survives according to a probability depending on its type, i.e. an individual of type i survives with probability $\delta_i \in [0, 1]$ and dies otherwise. It is possible to define the sequence $\{\xi_n, n \in \mathbb{N}\}$ of the times passing between two consecutive catastrophes, i.e. $\xi_1 = \tau_1$ and $\xi_n = \tau_n - \tau_{n-1}$ for each $n > 1$. Such a process is supposed to be stationary, ergodic and with a finite mean.

The life of each individual in the tree, apart from the catastrophes process, is still controlled by a transient MAP with parameters $(D_0, B, \mathbf{d})$ as the one described in the previous chapter.

In (2.3) we defined the population process $\{\mathbf{Z}(t), t \geq 0\}$, it is possible to extract from this continuous time process an embedded discrete process

$$\{\mathbf{Z}_n, n \in \mathbb{N}\}, \tag{3.1}$$

where $(\mathbf{Z}_n)_i$ denotes the number of individuals of type i alive just after the n-th catastrophe. In figure 3.1 we illustrate an example of a MBT which
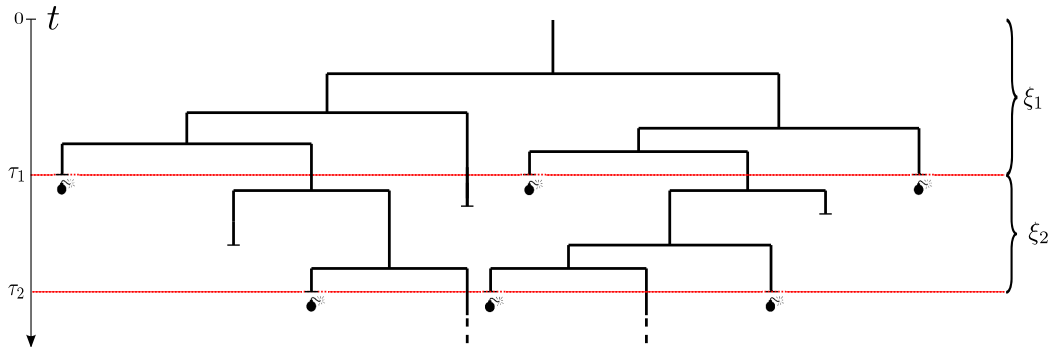

Figure 3.1: Markovian binary tree subject to catastrophes and generated by only an individual at time 0. In this case we have $\mathbf{Z}_0 = 1$, $\mathbf{Z}_1 = 3$ and $\mathbf{Z}_2 = 2$.

doesn't die out after two catastrophes.

We define the survival matrix Δ_δ as the diagonal matrix

$$\Delta_\delta = \begin{pmatrix} \delta_1 & 0 & \cdots & 0 \\ 0 & \delta_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \delta_m \end{pmatrix}, \tag{3.2}$$

and we use this matrix in the following proposition.

Proposition 3.1.1. *For every* $\mathbf{i} \in \mathcal{S}_m$*, given that* $\mathbf{Z}(0) = \mathbf{i}$ *and that the first catastrophe happens at time* ξ_1*, it holds that*

$$\mathbf{Z}_1 = (\mathbf{i}^T e^{\Omega \xi_1} \Delta_\delta)^T, \tag{3.3}$$

where Ω *is the generator of the mean matrices semi-group defined in* (2.15).

Proof. We employ the notations introduced in Definition (2.1.1), so we have that $M(\xi_1) = e^{\Omega \xi_1}$ is the matrix which describe the mean evolution of the population in an interval of time of length ξ_1 without catastrophes. Therefore, since $\mathbf{i} \in \mathcal{S}^m$ denotes the configuration of the population at time 0, we have that the expected population vector at time ξ_1, just before the catastrophe, is given by the transpose of the vector $\mathbf{i}^T e^{\Omega \xi_1}$. We conclude by observing that the entry $(\mathbf{Z}_1)_i$ is given by the expected number of individual alive just before the catastrophe, multiplied by δ_i. So, the following formula holds

$$(\mathbf{Z}_1)_i = (\mathbf{i}^T e^{\Omega \xi_1})_i \delta_i,$$

for each $i = 1, \ldots, m$, and the thesis is proved. $\qquad \square$

In parallel with the simple MBT, we are interested mainly into the study of the extinction probability vector. The presence of an external environment makes almost obsolete the analysis concerning the extinction of the process developed in the previous section. In fact, given λ the eigenvalue of Ω defined in (2.18), it is still true that $\lambda \leq 0$ implies $\mathbf{q} = \mathbf{e}$, but on the other hand the supercriticality of the process, i.e. the strict positiveness of λ, is no

more a sufficient condition for $\mathbf{q} < \mathbf{e}$. This is caused by the sequence of the catastrophes which may decrease, sometimes in a drastic way, the survival chances of a certain population.

In order to include the catastrophes in our analysis, we define the conditional probability of extinction given the successive times of catastrophes by the following formula,

$$q_i(\boldsymbol{\xi}) = \mathrm{P}[\lim_{n \to \infty} \mathbf{Z}_n = \mathbf{0} | \mathbf{Z}_0 = \mathbf{e}_i, \boldsymbol{\xi}], \qquad (3.4)$$

where $\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots)$ contains the sequence of time intervals between consecutive catastrophes. We present now a couple of results that give us some criteria for understanding better the behavior of the process, even in the supercritical case. The first one is an application of the theorem proved by Tanny in [39, Theorem 5.5, Corollary 6.3], which characterizes the scenario wherein we imagine to have only one type of user, i.e. $m = 1$.

Theorem 3.1.1. *If $m = 1$, then $\Omega = D_0 + 2B$ and $\Delta_\delta = \delta$ are both scalars and*

1. *If $e^{\Omega E[\xi]}\delta \leq 1$, then $P[q(\boldsymbol{\xi}) = 1] = 1$;*

2. *If $e^{\Omega E[\xi]}\delta > 1$, then $P[q(\boldsymbol{\xi}) < 1] = 1$ and*

$$\lim_{n \to \infty} 1/n \log(Z_n) = \Omega E[\xi] + \log(\delta),$$

where ξ is a random variable with the same distribution of any ξ_n and Ω is defined by formula (2.37).

This result is pretty intuitive, indeed it means that the extinction occurs certainly if, applying a catastrophe after a time $\mathrm{E}[\xi]$, the expected number of survivors is at most one. Otherwise the branching process has a strictly positive probability to survive forever.

The result we present now is more complex result and it is once again the application of a theorem proved by Tanny, see [40, Theorem 9.10]. We consider the case with a general number of types, i.e. $m \geq 1$, and the following statement is true.

Theorem 3.1.2. *The following limit exists*

$$\omega = \lim_{n \to \infty} \frac{1}{n} \log \|e^{\Omega \xi_1} \Delta_\delta \dots e^{\Omega \xi_n} \Delta_\delta\| \quad a.s., \tag{3.5}$$

and it is equal to ω independently of the matrix norm employed. Furthermore

1. *If $\omega \leq 0$, then $P[\mathbf{q}(\boldsymbol{\xi}) = \mathbf{1}] = 1$;*

2. *If $\omega > 0$, then $P[\mathbf{q}(\boldsymbol{\xi}) < \mathbf{1}] = 1$, and*

$$P[\lim_{n \to \infty} \frac{1}{n} \log \|\mathbf{Z}_n\| = \omega \mid \mathbf{Z}_0 = \mathbf{e}_i, \boldsymbol{\xi}] = 1 - q_i(\boldsymbol{\xi}) \quad a.s.,$$

for $i = 1, \dots, m$,

where Ω is defined by formula (2.37).

In the appendix of [15] it is verified explicitly that the kind of process we consider fulfills the hypotheses required by the theorems stated by Tanny in the papers [39, 40].

On the parameter controlling the extinction

Theorem 3.1.2 points out the key role that is played by the parameter ω for understanding if a process is doomed to extinguish or not.

The existence of the limit (3.5) is not obvious at all and has been an objective of study for a long time. Bellman, in the work [5] of 1954, proved a limit existence result for some infinite products of random matrices satisfying certain restrictive hypotheses. Those restrictive terms are fulfilled by the infinite product,

$$\lim_{n \to \infty} e^{\Omega \xi_1} \Delta_\delta \dots e^{\Omega \xi_n} \Delta_\delta, \tag{3.6}$$

so that, due to Bellman's result, the following limit is known to exist,

$$\omega_{\text{Bell}} = \lim_{n \to \infty} \frac{1}{n} \mathrm{E}[\log(e^{\Omega \xi_1} \Delta_\delta \dots e^{\Omega \xi_n} \Delta_\delta)_{ij}], \tag{3.7}$$

where $\mathrm{E}[\cdot]$ denotes the expected value relative to the sequence of random variables $\{\xi_1, \xi_2, \dots\}$ and the value of the limit does not depend on the matrix

entry chosen.

This result was extensively generalized by Furstenberg and Kesten in the outstanding work [11]. They firstly proved an almost sure convergence for the result obtained by Bellman, so that

$$\omega_{\text{Bell}} = \lim_{n \to \infty} \frac{1}{n} \log(e^{\Omega \xi_1} \Delta_\delta \ldots e^{\Omega \xi_n} \Delta_\delta)_{ij} \quad a.s.. \tag{3.8}$$

Then they observed the equivalence between choosing any one of the entries and working with the maximum norm, defined by (1.9), obtaining

$$\omega_{\text{Bell}} = \lim_{n \to \infty} \frac{1}{n} \log \|e^{\Omega \xi_1} \Delta_\delta \ldots e^{\Omega \xi_n} \Delta_\delta\|_\infty \quad a.s.. \tag{3.9}$$

Hence, the following simple step was the observation that on $\mathbb{R}^{n \times n}$ every norm is equivalent, yielding to

$$\omega = \omega_{\text{Bell}}, \tag{3.10}$$

where ω is the limit value defined in Theorem 3.1.2. In the following years, the result achieved by Kesten and Furstenberg was proved in a lot of different elegant ways, see the works by Kingman [24] and Oseledec [33]. Further studies gave rise to equivalent expressions for the parameter ω. For instance, we introduce a special class of function on the space of real valued matrices,

Definition 3.1.1 (Supermultiplicative function)**.** *A function $f : \mathbb{R}^{n \times n} \to [0, \infty)$ is called supermultiplicative if*

- *f is continuous;*

- *$f(\lambda A) = \lambda f(A)$, for each $\lambda \geq 0$;*

- *$f(AB) \geq f(A)f(B)$.*

Key, in the work [22, Theorem 1], employed the supermultiplicative functions as the ingredients for proving a result that, conveniently adjusted to our process, leads to the following equivalent expression for the parameter ω,

$$\omega = \lim_{n \to \infty} \frac{1}{n} \log f(e^{\Omega \xi_1} \Delta_\delta \ldots e^{\Omega \xi_n} \Delta_\delta) \quad a.s., \tag{3.11}$$

where f is a generic supermultiplicative function. Key's result may be applied only assuming certain restrictive conditions, but it can be shown that those terms are fulfilled by our process.

In the same paper Key provided an additional interesting formulation for ω, see [22, Corollary 1], in fact

$$\omega = \lim_{n \to \infty} \frac{1}{n} \log \rho(e^{\Omega \xi_1} \Delta_\delta \ldots e^{\Omega \xi_n} \Delta_\delta) \quad a.s., \tag{3.12}$$

where $\rho(\cdot)$ denotes the spectral radius, see (1.5). We highlight that this result is not an immediate consequence of (3.11) since the spectral radius is not a supermultiplicative function, but is based also on [22, Proposition 3] which states that $f(A) \leq \rho(A)$ for every f supermultiplicative function and $A \in \mathbb{R}^{n \times n}$.

The computation of the parameter ω, despite all its possible different formulation (3.5),(3.8),(3.11),(3.12), is a difficult matter. Indeed the catastrophe may have a stronger or weaker effect depending on the values of the components of the sequence $(\xi_1, \xi_2, \ldots)$ and on the types of the individuals that are alive when a catastrophe happens. Therefore, apart from particular simple cases, our objective will be the identification of incisive lower and upper bounds such that $\omega \in [\omega_l, \omega_u]$. We will use a combined numerical and probabilistic approach in order to narrow the interval $[\omega_l, \omega_u]$ as much as possible. Before proceeding with the analysis, it is interesting to open a parenthesis about the random linear dynamical systems and maximal Lyapunov exponent, so that it is more evident the hardness of the computation of the parameter ω.

3.1.1 Linear random dynamical system

In this subsection our goal is to briefly introduce the framework of the linear random dynamical system. In fact the computation of ω may be compared to the computation of the maximal Lyapunov exponent of a particular system of this kind.

Definition 3.1.2 (Random dynamical system). *A random dynamical system*

is defined by the tuple (S, Γ, Q), where

- *S denotes the space of the states where the system can stand, for simplicity $S \subseteq \mathbb{R}^k$ for some k;*

- *Γ is a set of maps from S in itself, a σ-algebra $\mathcal{G}$ is supposed to be defined on this set, see (1.14);*

- *Q is a probability measure defined on the σ-algebra $\mathcal{G}$, see (1.14).*

If the maps in Γ are linear, i.e. $\Gamma \subseteq \mathbb{R}^{k \times k}$, we talk about linear random dynamical system.

A system like this works intuitively in the following way: we suppose to be in state X_0, the first transition is defined by choosing a map $\gamma_1 \in \Gamma$ according to the probability Q, so that $X_1 = \gamma_1(X_0)$, then we choose in the same way $\gamma_2 \in \Gamma$ leading to $X_2 = \gamma_2(X_1) = \gamma_2\gamma_1(X_0)$ and so on. What we obtain is therefore a Markov chain $\{X_n, \ n \in \mathbb{N}\}$ on the state space S.

It is evident that if the set Γ is formed by only one linear map $A \in \mathbb{R}^{k \times k}$, we have $X_n = A^n X_0$ so that the behavior of the system depends on the spectrum of the matrix A. In particular if $A > 0$, the system will explode if and only if the spectral radius of A is greater than 1, i.e. $\rho(A) > 1$. In the case with a general number of linear maps, the role of the spectral radius is played, in some sense, by the maximal Lyapunov exponent of the system,

Definition 3.1.3 (Maximal Lyapunov exponent)**.** *The maximal Lyapunov exponent (MLE) of the random linear dynamical system defined in 3.1.2 is given by the formula,*

$$\lambda\{\Gamma, Q\} = \lim_{n \to \infty} \frac{1}{n} E \log \|A_1 A_2 \dots A_n\|, \qquad (3.13)$$

where $\{A_1, A_2, \dots\}$ is an iid sequence of maps chosen among the elements of Γ according to the probability Q.

The computation of $\lambda\{\Gamma, Q\}$ is a well known problem in literature, and in particular its hardness is witnessed in [24, 41].

We mentioned that in some sense the role of the spectral radius is played

by the MLE. This statement may be justified briefly by observing that if we have only one linear map, i.e. $\Gamma = \{A\}$, it is easy to compute the MLE in fact $\lambda\{\Gamma, Q\} = \log(\rho(A))$. Therefore the MLE may be seen as a sort of generalization of the spectral radius.

As far as we know, exact algorithms in the general case exist only in presence of maps with a certain structure [10, 21, 9] and mostly when the cardinality of the set Γ is finite [36]. More commonly, the objective is either to give bounds [22, 12] or to approximate the exact value of the MLE [31, 23, 28]. Going back to our problem, we recall that we are interested in computing

$$\omega = \lim_{n \to \infty} \frac{1}{n} \mathrm{E} \log \|e^{\Omega \xi_1} \Delta_\delta \dots e^{\Omega \xi_n} \Delta_\delta\|.$$

This problem can be interpreted as the computation of the MLE of the random Dynamical system having the following specific parameters,

- $S = \{\mathbf{x} \in \mathbb{R}^m \mid x_i \geq 0, \ i = 1, \dots, m\}$;

- $\Gamma = \{e^{\Omega t} \Delta_\delta \in \mathbb{R}^{m \times m} \mid t \geq 0\}$;

- Q is defined by the probability distribution which rule the time interval between two catastrophes.

So that the problem of computing ω inherits the whole set of difficulties that goes with the computation of the maximal Lyapunov exponent of a random dynamical system.

3.2 A duality approach

As explained in the previous section, in general computing the limit (3.5) is difficult. However it must be observed the special structure of the matrices which form the set of the maps of the linear random dynamical system we are interested into.

In this section our aim is to give a different expression for ω, in order to facilitate its analysis in the following chapter.

We assume that the matrix Ω defined in (2.37) is irreducible. We denote by λ the eigenvalue of Ω defined by (2.18). So we may define the matrix

$$\Omega^* = \Omega - \lambda I_m, \qquad (3.14)$$

which is a matrix with nonnegative off diagonal elements, and with all the eigenvalues with a strictly negative real part apart from one which is equal to 0, therefore $-\Omega^*$ is a singular M-matrix.

Lemma 3.2.1. *An equivalent expression for ω is given by the formula*

$$\omega = \lambda E[\xi] + \lim_{n \to \infty} \frac{1}{n} \log \|e^{\Omega^* \xi_1} \Delta_\delta \ldots e^{\Omega^* \xi_n} \Delta_\delta\|. \qquad (3.15)$$

Proof. It is known that an expression for ω is given by (3.5). Therefore the following equivalences holds,

$$\begin{aligned}
\omega &= \lim_{n \to \infty} \frac{1}{n} \log \|e^{(\Omega^* + \lambda I_m)\xi_1} \Delta_\delta \ldots e^{(\Omega^* + \lambda I_m)\xi_n} \Delta_\delta\| \\
&= \lim_{n \to \infty} \frac{1}{n} \log \|e^{\Omega^* \xi_1} e^{\lambda \xi_1} \Delta_\delta \ldots e^{\Omega^* \xi_n} e^{\lambda \xi_n} \Delta_\delta\| \\
&= \lim_{n \to \infty} \frac{1}{n} \log(e^{\lambda(\xi_1 + .. + \xi_n)} \|e^{\Omega^* \xi_1} \Delta_\delta \ldots e^{\Omega^* \xi_n} \Delta_\delta\|) \\
&= \lim_{n \to \infty} \lambda \frac{\xi_1 + \ldots + \xi_n}{n} + \lim_{n \to \infty} \frac{1}{n} \log \|e^{\Omega^* \xi_1} \Delta_\delta \ldots e^{\Omega^* \xi_n} \Delta_\delta\|.
\end{aligned}$$

The thesis follow from the strong law of large numbers. $\qquad \square$

There exists a corollary of the Perron-Frobenius theorem for singular M-matrices, we state it here.

Theorem 3.2.1 (Perron-Frobenius Theorem. Singular M-matrices case)**.** *The eigenvalue 0 for a singular irreducible M-matrix M is simple and there exist positive right and left eigenvectors associated to such eigenvalue.*

By applying this version of the Perron-Frobenius theorem to the matrix Ω^*, we know that there exist strictly positive left and right eigenvectors, denoted respectively $\boldsymbol{u}$ and $\boldsymbol{v}$, for such a matrix associated to the 0 eigenvalue.

Furthermore $\boldsymbol{u}$ and $\boldsymbol{v}$ are also the eigenvectors associated to the eigenvalue λ of the matrix Ω. We choose to normalize these eigenvectors in a canonical way, that is to say

$$\boldsymbol{u}^T\mathbf{e} = 1, \qquad \boldsymbol{u}^T\boldsymbol{v} = 1. \tag{3.16}$$

It is useful to define the diagonal matrices having these vectors as diagonals, to be more precise

$$\Delta_u = \mathrm{diag}(\boldsymbol{u}), \qquad \Delta_v = \mathrm{diag}(\boldsymbol{v}). \tag{3.17}$$

Indeed we need these matrices in order to express the fundamental matrix Θ, defined by this formula,

$$\Theta = \Delta_u^{-1}(\Omega^*)^T\Delta_u. \tag{3.18}$$

Proposition 3.2.1. *The matrix Θ is a generator matrix. Moreover its stationary probability vector $\boldsymbol{\pi}$ is given by*

$$\boldsymbol{\pi} = \Delta_u\boldsymbol{v} = \Delta_v\boldsymbol{u}.$$

Proof. In order to prove that Θ is a generator matrix it is sufficient to verify that the non diagonal entries of matrix Θ are nonnegative and that the row sums are equal to zero for each row. The first point is verified by the following chain of equalities,

$$\Theta_{ij} = \sum_{k,h=1}^{m} (\Delta_u^{-1})_{ik}((\Omega^*)^T)_{kh}(\Delta_u)_{hj} = \frac{1}{u_i}(\Omega^*)_{ji}u_j,$$

and by the non negativity of the off diagonal entries of matrix Ω^* and of $\boldsymbol{u}$. The row sums can be computed directly by multiplying Θ for the vector $\mathbf{e}$,

$$\Theta\mathbf{e} = \Delta_u^{-1}(\Omega^*)^T\Delta_u\mathbf{e} = \Delta_u^{-1}(\Omega^*)^T\boldsymbol{u} = \Delta_u^{-1}(\boldsymbol{u}^T\Omega^*)^T = \Delta_u^{-1}\mathbf{0} = \mathbf{0}.$$

The truth of the statement concerning the stationary probability vector may

be checked in a similar way

$$(\Delta_u \boldsymbol{v})^T \Theta = \boldsymbol{v}^T \Delta_u \Delta_u^{-1} (\Omega^*)^T \Delta_u = (\Omega^* \boldsymbol{v})^T \Delta_u = \boldsymbol{0}^T \Delta_u = \boldsymbol{0}^T.$$

$\square$

We are now interested in reformulating the expression for ω, see (3.15), so that the matrix Θ appears in it.

Lemma 3.2.2. *An equivalent expression the constant ω, defined by (3.15), is given by the formula $\omega = \lambda E[\xi] + \psi$, where*

$$\psi = \lim_{n \to \infty} \frac{1}{n} \log(e^{\Theta \eta_1} \Delta_\delta \ldots e^{\Theta \eta_n} \Delta_\delta)_{ij} \tag{3.19}$$

where we denote by $\eta_j = \xi_{-j}$ the intervals between events in the time reversed version of the catastrophe process.

Proof. Using (3.15) we can write $\omega = \lambda E[\xi] + \tilde{\psi}$, where

$$\tilde{\psi} = \lim_{n \to \infty} \frac{1}{n} \log(e^{\Omega^* \xi_1} \Delta_\delta \cdots e^{\Omega^* \xi_n} \Delta_\delta)_{ji}.$$

Therefore we have to prove that $\tilde{\psi} = \psi$. This is verified by the following

chain of equalities,

$$\tilde{\psi} = \lim_{n\to\infty} \frac{1}{n} \log((\Delta_u^{-1}e^{\Theta^T\xi_1}\Delta_u\Delta_\delta\Delta_u^{-1}e^{\Theta^T\xi_2}\Delta_u\Delta_\delta\cdots\Delta_u^{-1}e^{\Theta^T\xi_n}\Delta_u\Delta_\delta)^T)_{ij}$$

$$= \lim_{n\to\infty} \frac{1}{n} \log((\Delta_u^{-1}e^{\Theta^T\xi_1}\Delta_\delta e^{\Theta^T\xi_2}\Delta_\delta\cdots e^{\Theta^T\xi_n}\Delta_u\Delta_\delta)^T)_{ij}$$

$$= \lim_{n\to\infty} \frac{1}{n} \log(\Delta_\delta\Delta_u^{-1}e^{\Theta\xi_n}\Delta_\delta e^{\Theta\xi_{n-1}}\Delta_\delta\cdots\Delta_\delta e^{\Theta\xi_1}\Delta_u^{-1})_{ij}$$

$$= \lim_{n\to\infty} \frac{1}{n} \log(\Delta_\delta\Delta_u e^{\Theta\eta_1}\Delta_\delta e^{\Theta\eta_2}\Delta_\delta\cdots\Delta_\delta e^{\Theta\eta_n}\Delta_u^{-1})_{ij}$$

$$= \lim_{n\to\infty} \frac{1}{n} \log\rho(\Delta_\delta\Delta_u e^{\Theta\eta_1}\Delta_\delta e^{\Theta\eta_2}\Delta_\delta\cdots\Delta_\delta e^{\Theta\eta_n}\Delta_u^{-1})$$

$$= \lim_{n\to\infty} \frac{1}{n} \log\rho((\Delta_\delta\Delta_u)^{-1}\Delta_\delta\Delta_u e^{\Theta\eta_1}\Delta_\delta e^{\Theta\eta_2}\Delta_\delta\cdots e^{\Theta\eta_n}\Delta_u^{-1}(\Delta_\delta\Delta_u))$$

$$= \lim_{n\to\infty} \frac{1}{n} \log\rho(e^{\Theta\eta_1}\Delta_\delta\cdots e^{\Theta\eta_n}\Delta_\delta)$$

$$= \lim_{n\to\infty} \frac{1}{n} \log(e^{\Theta\eta_1}\Delta_\delta\cdots e^{\Theta\eta_n}\Delta_\delta)_{ij} = \psi,$$

where we used the equivalent formulations for the maximal Lyapunov exponent (3.12) and the stationarity of the process $\{\xi_n\}_n$ which imply that the tuple $(\eta_1,\ldots,\eta_n) = (\xi_{-1},..,\xi_{-n})$ has the same distribution as $(\xi_n,\ldots,\xi_1)$. $\quad\square$

After all this analysis, we observe that we can focus on the computation of ψ instead of ω, that means that the set of the linear maps of the random dynamical system we are interested into is now given by

$$\Gamma = \{e^{\Theta t}\Delta_\delta \,|\, t \geq 0\}. \tag{3.20}$$

The benefit of this replacement will be pointed out in the following section.

3.3 Matrix exponential of a generator matrix

In the previous section we have defined the set of the matrices which form the linear maps of our special linear random dynamical system, see (3.20). These matrices play a key role in the computation of ψ, so that it is important to spend some time studying their properties, hoping to obtain some useful information which could simplify the evaluation of the parameter ψ.

These matrices possesses a particular form, indeed the generic matrix is given by the multiplication of the exponential of a random matrix times a fixed diagonal one, i.e. $e^{\Theta t}\Delta_\delta$, where Θ and Δ_δ are defined respectively in (3.18) and (3.2), and the randomness is introduced by the scalar $t \geq 0$. In particular we are interested in computing the random product

$$V(t_1, \ldots, t_n) = e^{\Theta t_1}\Delta_\delta \ldots e^{\Theta t_n}\Delta_\delta. \tag{3.21}$$

Since Δ_δ is a diagonal matrix with positive diagonal entries, by applying the Definition 1.12 of the matrix exponential, it is possible to write this random product in the following equivalent form,

$$V(t_1, \ldots, t_n) = e^{\Theta t_1}e^{\tilde{\Delta}_\delta} \ldots e^{\Theta t_n}e^{\tilde{\Delta}_\delta}, \tag{3.22}$$

where $\tilde{\Delta}_\delta$ is the diagonal matrix such that $(\tilde{\Delta}_\delta)_{ii} = \log \delta_i$.

We observe that the evaluation of this product may be a lot easier if it could be possible to commute the factors, but the Theorem 1.1.1 states that this can be done if and only if $\Theta\tilde{\Delta}_\delta = \tilde{\Delta}_\delta\Theta$ which is in general not true. Therefore without further assumption

$$V(t_1, \ldots, t_n) \neq e^{\Theta \sum_{i=1}^n t_i}e^{n\tilde{\Delta}_\delta}. \tag{3.23}$$

We recall now that in Proposition 3.2.1 we proved that Θ is a rate matrix, and clearly Θt, for each $t \geq 0$, maintain such property. Moreover the probability stationary vector remains the same for each t. Therefore it could be interesting to examine the properties of the exponential of a rate matrix.

Proposition 3.3.1. *For any rate matrix $\Theta \in \mathbb{R}^{n \times n}$ having stationary probability vector $\boldsymbol{\pi} \in \mathbb{R}^n$ it holds that $e^{\Theta t}$ is a stochastic matrix with $\boldsymbol{\pi}$ as its stationary probability vector for each $t \geq 0$.*

Proof. We recall that from the Definition 1.12 we have that

$$e^{\Theta t} = \sum_{i=0}^{\infty} \frac{(t\Theta)^i}{i!}.$$

In order to prove the stochasticity of the matrix $e^{\Theta t}$ it must verified the positivity of its entries and that the sums over each row is equal to 1.

- The positivity of the entries is proved by considering $\Theta = H + \upsilon I$ for $\upsilon \leq 0$ such that $H \geq 0$, this decomposition exists since Θ is a rate matrix. Thanks to Theorem 1.1.1, we have that

$$e^{\Theta} = e^{H+\upsilon I} = e^{H} e^{\upsilon I},$$

  since $H\upsilon I = \upsilon I H = \upsilon H$ At this point is sufficient to observe that $e^{H} \geq 0$ due to the non negativity of H and that $e^{\upsilon I}$ is equal to $e^{\upsilon} I$ thanks to the definition of the matrix exponential. Therefore

$$e^{\Theta t} = e^{t} e^{\Theta} = e^{t} e^{\upsilon} e^{H} \geq 0.$$

- For the second part it is sufficient to observe that $e^{t\Theta} \mathbf{e} = \mathbf{e}$, which is verified by

$$e^{t\Theta}\mathbf{e} = \sum_{i=0}^{\infty} \frac{(t\Theta)^{i}}{i!}\mathbf{e} = I\mathbf{e} + \sum_{i=1}^{\infty} \frac{t^{i}\Theta^{i-1}}{i!}\Theta\mathbf{e} = \mathbf{e},$$

  where the last equality is true since Θ is a generator, i.e. $\Theta\mathbf{e} = \mathbf{0}$.

The argument concerning the fact about the stationary probability vector is quite similar, indeed

$$\boldsymbol{\pi}^{T} e^{t\Theta} = \boldsymbol{\pi}^{T} \sum_{i=0}^{\infty} \frac{(t\Theta)^{i}}{i!} = \boldsymbol{\pi}^{T} I + \sum_{i=1}^{\infty} \boldsymbol{\pi}^{T} \Theta \frac{t^{i}\Theta^{i-1}}{i!} = \boldsymbol{\pi}^{T},$$

where the last equality follow from $\boldsymbol{\pi}^{T}\Theta = \mathbf{0}^{T}$. □

Therefore the linear maps forming the set (3.20) have a very specific structure. Indeed, due to Proposition 3.3.1, $e^{\Theta t}$ is a stochastic matrix for each $t \geq 0$, and since $0 \leq \delta_{i} \leq 1$ fr each $i = 1, \ldots, m$, we have that all the matrices $e^{\Theta t}\Delta_{\delta}$ are subtochastic matrices.

Moreover, due to the definition of Θ, see (3.18), the irreducibility of Θ is

implied by the irreducibility of the matrix Ω, see (2.37), condition that we have already required at the beginning of the chapter. Therefore we can suppose Θ to be irreducible. Hence, if we also suppose the entries δ_i to be strictly greater than 0 for each $i = 1, \ldots, m$, we have that $e^{\Theta t} \Delta_\delta > 0$ for each $t > 0$.

Chapter 4

Estimate of the extinction probability and numerical results

In the previous chapter we introduced a process of catastrophes influencing the behavior of the branching process we are studying. Theorem 3.1.2 introduced the parameter ω whose value determines either a certain extinction or a strictly positive probability of infinite time survival of the process.

In Lemma 3.2.2 we managed to obtain a nicer expression for ω, introducing a new parameter ψ and observing that the computation of it may be likened to the computation of the maximal Lyapunov exponent of a particular linear random dynamical system, see definition 3.1.2, having a very specific set of linear maps

$$\Gamma = \{e^{\Theta t}\Delta_\delta \mid t \geq 0\}. \tag{4.1}$$

The properties of this set of maps are illustrated in the last part of the previous chapter. Our objective is now to exploit these properties for achieving some information concerning the extinction probability, which may be deduced through the estimation of the parameter ψ.

4.1 Particular cases: exact computation of the extinction probability

Despite the nice form of the set (4.1), the computation of the maximal Lyapunov exponent ψ, defined in (3.19), is still beyond our possibilities. An exact evaluation of such parameter may be obtained only in cases wherein some additional restrictive assumptions are made.

4.1.1 Type independent extinction probabilities

In this simpler scenario we suppose that the survival probabilities are independent from the type of the individuals, i.e. $\delta_i = \delta_j$ for each $i, j = 1, \ldots, m$. In this case ψ can be computed as stated in the following theorem, see [15, Corollary 5.1]

Theorem 4.1.1. *If for each $i = 1, \ldots, m$ we have $\delta = \delta_i$, then*

$$\psi = \log(\delta), \tag{4.2}$$

where ψ is the parameter defined in (3.19).

Proof. Since $\delta_i = \delta_j$ for every $i, j = 1, \ldots, m$, we have that $\Delta_\delta = \delta I_m$. By means of (3.19) and Theorem 1.1.1, it is possible to express ψ in this simple way

$$\begin{aligned}
\psi &= \lim_{n \to \infty} \frac{1}{n} \log \| e^{\Theta \eta_1} \Delta_\delta \ldots e^{\Theta \eta_n} \Delta_\delta \| \\
&= \lim_{n \to \infty} \frac{1}{n} \log \| \delta^n e^{\Theta \eta_1} \ldots e^{\Theta \eta_n} \| \\
&= \lim_{n \to \infty} \log(\delta^n)^{1/n} + \lim_{n \to \infty} \log \| e^{\Theta \tau_n} \| \\
&= \log(\delta) + \lim_{n \to \infty} \log \| e^{\Theta \tau_n} \|,
\end{aligned}$$

where $\tau_n = \sum_{i=1}^{n} \eta_i$ denotes the time of happening of the n-th catastrophic event. Since the maximal Lyapunov exponent is independent from the norm chosen, see definition (3.5), we employ the norm $\|\cdot\| = \|\cdot\|_\infty$ so that $\| e^{\Theta \tau_n} \| =$

1 because of the stochasticity of the matrix $e^{\Theta t}$, proved in Proposition 3.3.1. Hence we may conclude,

$$\psi = \log(\delta) + \lim_{n \to \infty} \log\|e^{\Theta \tau_n}\| = \log(\delta) + \log 1 = \log(\delta).$$

$\square$

4.1.2 Constant intervals between catastrophes

A different assumption that simplifies in a drastic way the problem is to consider the intervals between the consecutive catastrophes having a constant length. Even in this case it is possible to compute easily the parameter ψ, indeed the following theorem is proved.

Theorem 4.1.2. *If $\xi_i = \beta$ for each $i = 1, 2, \ldots$, then*

$$\psi = \log(\rho(e^{\Theta \beta} \Delta_\delta)), \tag{4.3}$$

where ψ is the parameter defined in (3.19).

Proof. In this case we have that the following chain of equalities is true,

$$
\begin{aligned}
\psi &= \lim_{n \to \infty} \frac{1}{n} \log\|e^{\Theta \eta_1} \Delta_\delta \ldots e^{\Theta \eta_n} \Delta_\delta\| \\
&= \lim_{n \to \infty} \log(\|(e^{\Theta \beta} \Delta_\delta)^n\|)^{1/n} \\
&= \log(\rho(e^{\Theta \beta} \Delta_\delta)),
\end{aligned}
$$

where the last equivalence is given by the Gelfand's formula, whose proof may be found in [37]. $\square$

4.2 General case: a probabilistic approach

As already stated, a computable formula for ψ is available only in presence of particular and too restrictive conditions. Therefore in the general case we must content by obtaining approximations or bounds.

In this section we engage the problem exploiting the probabilistic features of the process we are working on, in fact the generator Θ defines a continuous time Markov process

$$\{\phi(t) \mid t \geq 0\} \qquad (4.4)$$

on the state space $\mathbb{R}^m$.

So that the study of the whole population process may be reduced to the study of this single particle process. During its lifetime this particle faces a sequence of accidents, corresponding to the whole population process catastrophes, and survives with a probability according to its type when one of these happens, i.e. with probability δ_i for $i = 1, .., m$. We define the times of the successive occurrences of accidents as the sequence $\{\theta_0, \theta_1, \theta_2, \ldots\}$, whose entries are defined by $\theta_0 = 0$ and $\theta_n = \theta_{n-1} + \eta_n$ for $n \geq 1$, where the sequence of $(\eta_1, \eta_2, \ldots)$ is defined in Theorem 3.2.2. It is possible to extract from the process (4.4) an embedded discrete time process $\{\phi_n \mid n \in \mathbb{N}\}$, wherein $\phi_n = \phi(\theta_n)$. Moreover we denote by T the first epoch such that the particle does not survive an accident. It is now possible to give a probabilistic expression for the parameter ψ, in fact the formula (3.19) is equivalent to the following

$$\psi = \lim_{n \to \infty} \log \mathrm{P}[T > \theta_n, \phi_n = j \mid \phi_0 = i, \theta_0, \ldots, \theta_n], \qquad (4.5)$$

for any choice of $i, j = 1, \ldots, m$.

We refer to [15] for the proof of the following theorem,

Theorem 4.2.1. *If the sequence of times of catastrophes forms an ergodic stationary process, then $\psi_l \leq \psi \leq \psi_u$, with*

$$\psi_u = \lim_{n \to \infty} \log P[T > \theta_n, \phi_n = j \mid \phi_0 = i], \qquad (4.6)$$

and

$$\psi_l = \sum_{i=1}^{m} u_i v_i \log \delta_i, \qquad (4.7)$$

where $\boldsymbol{u}$ and $\boldsymbol{v}$ are respectively the left and the right eigenvector of Ω associated to the eigenvalue λ, and ψ is defined by formula (4.5). The vectors $\boldsymbol{u}$

and $\boldsymbol{v}$ are normalized in the canonical way as stated in (3.16).

It is possible to give an alternative expression for ψ_l which is based only on the matrix generator Θ, see (3.18), in fact

$$\psi_l = \sum_{i=1}^{m} \pi_i \log \delta_i, \tag{4.8}$$

where $\boldsymbol{\pi}$ is the stationary probability vector of the generator matrix Θ. This equality is a direct application of Proposition 3.2.1.

The bounds suggested in Theorem 4.2.1 are tight, indeed in the case with $\delta_i = \delta_j$ for each $i, j = 1, .., m$ we have that

$$\psi_l = \psi = \psi_u,$$

the details may be seen in [15].

The upper bound ψ_u depends from the probability distribution chosen for the process of the catastrophes, in [15] such an upper bound is evaluated in a couple of special cases.

Poisson process of catastrophes

We assume the process $\{\tau_n\}$ to be a Poisson process of rate $\beta = 1/\mathrm{E}[\xi]$. This means that the probability distribution of the time interval between two consecutive catastrophes follows an exponential distribution of parameter $1/\beta$. With these assumptions it is possible to compute the transition probabilities matrix $\tilde{P}$ associated to the discrete time phase process $\{\phi_n \,|\, n \in \mathbb{N}\}$. In fact without catastrophes the transition matrix would be

$$P = \int_0^{\infty} e^{\Theta t} \beta e^{-\beta t} \mathrm{d}t = \beta(\beta I - \Theta)^{-1},$$

therefore the transition matrix for the process just after the happening of a catastrophe is given by $\tilde{P} = P\Delta_\delta$. So, with this information, it is possible to

rewrite the upper bound given by Theorem 4.2.1 using the following equalities

$$\psi_u = \lim_{n \to \infty} \frac{1}{n} \log(\tilde{P})_{ij}^n = \log \rho(P \Delta_\delta) = \tag{4.9}$$

$$= \log \rho(\beta(\beta I - \Theta)^{-1} \Delta_\delta). \tag{4.10}$$

Markovian arrival process of catastrophes

A more general case is considered if we suppose that the epochs of the catastrophes are marked by the observable transitions of a MAP as the one described in section 2.2. This MAP generates a bidimensional process $\{M(t), \phi(t), \ t \geq 0\}$, so that the process $M(t)$ counts the catastrophes occurred before time t. We suppose that this MAP is governed by the matrices D_0 and D_1 defined by formulas (2.26) and that the initial distribution vector $\boldsymbol{\alpha}$ is equal to the stationary probability vector of $(-D_0)^{-1} D_1$. It is known from [1] that the time reversal version of a MAP is still a MAP wherein the governing matrices are given by

$$\tilde{D}_i = \Delta_\varepsilon^{-1} D_i^T \Delta_\varepsilon,$$

where $\boldsymbol{\varepsilon}$ is the stationary probability vector of $D_0 + D_1$ and Δ_ε is the diagonal matrix having the element of $\boldsymbol{\varepsilon}$ on the diagonal. The initial probability vector of the time reversed MAP is given by $\tilde{\boldsymbol{\alpha}}$ such that it is the stationary probability vector of $(-\tilde{D}_0)^{-1} \tilde{D}_1$.

In order to characterize the first epoch such that the particle does not survive an accident, it is necessary to keep track of the bidimensional process $\{(\varphi(t), \chi(t)), \ t \geq 0\}$, where $\varphi(t)$ is the phase of the population process at time t, see (4.4), and $\chi(t)$ is the phase of the time reversed MAP of the catastrophes. The infinitesimal generator of this process is given by the matrix $Q = Q_0 + Q_1$ where

$$Q_0 = \Theta \otimes I + I \otimes \tilde{A}_0, \qquad Q_1 = I \otimes \tilde{A}_1, \tag{4.11}$$

represent the transition rates without and in presence of catastrophes, respectively.

The transition probability matrix at the epoch of an accident is given by $P = (-Q_0)^{-1}Q_1$, therefore if we consider the moment just after the catastrophe the transition matrix becomes $\tilde{P} = P(\Delta_\delta \otimes I)$.

It is now possible to give an alternative expression for the upper bound formulated in Theorem 4.2.1,

$$\psi_u = \lim_{n \to \infty} \frac{1}{n} \log((I \otimes \tilde{\boldsymbol{\alpha}})\tilde{P}^n(I \otimes \mathbf{e}))_{ij}$$
$$= \log \rho(\tilde{P}) = \log \rho(-(\Theta \otimes I + I \otimes \tilde{A}_0)^{-1}(\Delta_\delta \otimes \tilde{A}_1)).$$

4.3   General case: a matrix approach

We are now interested in exploiting the properties of the matrix appearing in the various equivalent formulations available for the parameter ψ, in hopes of providing useful bounds.

We denote by $V(t)$ the substochastic matrix

$$V(t) = e^{\Theta t}\Delta_\delta, \quad t \geq 0. \tag{4.12}$$

The goal is to obtain informations concerning the parameter ψ, that we recall may be computed through one of the following expressions

$$\psi = \lim_{n \to \infty} \frac{1}{n} \log(V(\eta_1)\ldots V(\eta_n))_{ij} \quad a.s., \tag{4.13}$$

$$= \lim_{n \to \infty} \frac{1}{n} \log\|V(\eta_1)\ldots V(\eta_n)\| \quad a.s., \tag{4.14}$$

$$= \lim_{n \to \infty} \frac{1}{n} \log \rho(V(\eta_1)\ldots V(\eta_n)) \quad a.s., \tag{4.15}$$

for each entry $i, j = 1, \ldots, m$ and each matrix norm $\|\cdot\|$.

It is difficult to deduce bounds through formula (4.13), therefore we focus our attention to the other two expressions.

Norm formulation

Firstly, we analyze the formulation (4.14) and, in particular, we consider the maximum norm, see (1.8). We observe that it is possible to exploit some of

the properties of the matrix norm operator in order to prove the following proposition

Proposition 4.3.1. *For every $\eta_1, \ldots, \eta_n \geq 0$, the following chain of inequalities holds*

$$\|(\mathbf{e}\boldsymbol{\pi}^T \Delta_\delta)^n\|_\infty \leq \|V(\eta_1) \ldots V(\eta_n)\|_\infty \leq \|\Delta_\delta^n\|_\infty, \qquad (4.16)$$

where $\boldsymbol{\pi}$ is the steady state probability vector associated to the generator matrix Θ.

Proof. We prove separately the two inequalities.

- $\|V(\eta_1) \ldots V(\eta_n)\|_\infty \leq \|\Delta_\delta^n\|_\infty$:

  Since the maximum norm is submultiplicative, see (1.7), we have that

  $$\|V(\eta_1) \ldots V(\eta_n)\|_\infty = \|e^{\Theta\eta_1}\Delta_\delta \ldots e^{\Theta\eta_n}\Delta_\delta\|_\infty$$
  $$\leq \|e^{\Theta\eta_1}\|_\infty\|\Delta_\delta\|_\infty \ldots \|e^{\Theta\eta_n}\|_\infty\|\Delta_\delta\|_\infty.$$

  By recalling Proposition 3.3.1, we know that for each $i = 1, \ldots, n$ the matrix $e^{\Theta\eta_i}$ proves to be a stochastic matrix and therefore $\|e^{\Theta\eta_i}\|_\infty = 1$. Hence we have that

  $$\|V(\eta_1) \ldots V(\eta_n)\|_\infty \leq \|\Delta_\delta\|_\infty \ldots \|\Delta_\delta\|_\infty = \delta_{\max}^n.$$

  where $\delta_{\max} = \max_{i=1,\ldots,m} \delta_i$. The inequality is therefore proved since

  $$\delta_{\max}^n = \|\Delta_\delta^n\|_\infty.$$

- $\|(\mathbf{e}\boldsymbol{\pi}^T \Delta_\delta)^n\|_\infty \leq \|V(\eta_1) \ldots V(\eta_n)\|_\infty$:

  We prove this fact by induction.

  For $n = 1$ we have that for every $t \geq 0$,

  $$\|\mathbf{e}\boldsymbol{\pi}^T \Delta_\delta\|_\infty = \|\mathbf{e}\boldsymbol{\pi}^T e^{-\Theta t} e^{\Theta t}\Delta_\delta\|_\infty = \|\mathbf{e}\boldsymbol{\pi}^T e^{\Theta t}\Delta_\delta\|_\infty$$
  $$\leq \|\mathbf{e}\boldsymbol{\pi}^T\|_\infty\|e^{\Theta t}\Delta_\delta\|_\infty = \|e^{\Theta t}\Delta_\delta\|_\infty,$$

Therefore, by choosing $t = \eta_1$, we have that

$$\|\mathbf{e}\boldsymbol{\pi}^T\Delta_\delta\|_\infty \leq \|e^{\Theta\eta_1}\Delta_\delta\|_\infty = \|V(\eta_1)\|_\infty.$$

In order to prove the inductive step, we suppose the proposition to be true for each product of at most $n-1$ factors, in particular

$$\|(\mathbf{e}\boldsymbol{\pi}^T\Delta_\delta)^{n-1}\|_\infty \leq \|e^{\Theta t_1}\Delta_\delta \ldots e^{\Theta t_{n-1}}\Delta_\delta\|_\infty \qquad (4.17)$$

for each choice of $t_1, \ldots, t_{n-1} \geq 0$.

Given $\eta_1, \ldots, \eta_n \geq 0$, we observe therefore that the following chain of inequalities is verified

$$\|(\mathbf{e}\boldsymbol{\pi}^T\Delta_\delta)^n\|_\infty = |\boldsymbol{\pi}^T\Delta_\delta\mathbf{e}|\|(\mathbf{e}\boldsymbol{\pi}^T\Delta_\delta)^{n-1}\|_\infty$$
$$\leq |\boldsymbol{\pi}^T\Delta_\delta\mathbf{e}|\|e^{\Theta t_1}\Delta_\delta \ldots e^{\Theta t_{n-1}}\Delta_\delta\|_\infty$$

indeed we have just employed (4.17). Hence, by choosing $(t_1, \ldots, t_{n-1}) = (\eta_2, \ldots, \eta_n)$ we have that

$$\|(\mathbf{e}\boldsymbol{\pi}^T\Delta_\delta)^n\|_\infty \leq |\boldsymbol{\pi}^T\Delta_\delta\mathbf{e}|\|V(\eta_2) \ldots V(\eta_n)\|_\infty$$
$$= \|\boldsymbol{\pi}^T\Delta_\delta\mathbf{e}V(\eta_2) \ldots V(\eta_n)\|_\infty.$$

It is now possible to observe that there exists a constant $C \in \mathbb{R}$ such that

$$0 \leq \boldsymbol{\pi}^T\Delta_\delta\mathbf{e}I_m \leq \mathbf{e}\boldsymbol{\pi}^T\Delta_\delta + CI_m,$$

in particular this is verified as soon as

$$C \geq \max_{j=1,\ldots,m}\{\sum_{i=1}^m \pi_i\delta_i - \pi_j\delta_j\} > 0.$$

Therefore, by choosing C in this way, the proof can be completed by

observing that

$$\begin{aligned}
\|(\mathbf{e}\boldsymbol{\pi}^T\Delta_\delta)^n\|_\infty &\leq \|\boldsymbol{\pi}^T\Delta_\delta\mathbf{e}V(\eta_2)\ldots V(\eta_n)\|_\infty \\
&\leq \|(\mathbf{e}\boldsymbol{\pi}^T\Delta_\delta + CI_m)V(\eta_2)\ldots V(\eta_n)\|_\infty \\
&\leq \|\mathbf{e}\boldsymbol{\pi}^T\Delta_\delta V(\eta_2)\ldots V(\eta_n)\|_\infty + \|CV(\eta_2)\ldots V(\eta_n)\|_\infty \\
&\leq \|\mathbf{e}\boldsymbol{\pi}^T\Delta_\delta V(\eta_2)\ldots V(\eta_n)\|_\infty \\
&\leq \|V(\eta_1)V(\eta_2)\ldots V(\eta_n)\|_\infty,
\end{aligned}$$

where the last inequality is proved with an argument identical to the one used in the step with $n=1$ and by choosing $t = \eta_1$.

$\square$

This proposition leads directly to the following corollary which provides us with new bounds for the parameter ψ.

Corollary 4.3.1. *For each $\eta_1, \eta_2, \ldots$, we define ψ as in (4.14) and the following chain of inequalities holds,*

$$\log(\sum_{i=1}^{m}\pi_i\delta_i) \leq \psi \leq \log(\delta_{\max}), \tag{4.18}$$

where $\delta_{\max} = \max_{i=1,\ldots,m}\delta_i$.

Proof. The proof follows directly from Proposition 4.3.1 and from the monotonicity of the function logarithm, indeed it holds that

$$\lim_{n\to\infty}\frac{1}{n}\log\|(\mathbf{e}\boldsymbol{\pi}^T\Delta_\delta)^n\| \leq \psi \leq \lim_{n\to\infty}\frac{1}{n}\log\|(\Delta_\delta)^n\|.$$

So that, due to the Gelfand's formula, we have that

$$\lim_{n\to\infty}\frac{1}{n}\log\|(\Delta_\delta)^n\| = \log(\rho(\Delta_\delta)) = \log(\delta_{\max}),$$

and

$$\lim_{n\to\infty}\frac{1}{n}\log\|(\mathbf{e}\boldsymbol{\pi}^T\Delta_\delta)^n\| = \log(\rho(\mathbf{e}\boldsymbol{\pi}^T\Delta_\delta)) = \log(\sum_{i=1}^{m}\pi_i\delta_i).$$

$\square$

We would like to give now a probabilistic intuition concerning the meaning of the lower bound formula

$$\psi_l^{(norm)} = \log(\sum_{i=1}^{m} \pi_i \delta_i) \qquad (4.19)$$

we just found.

We observe that by letting the time flow, for big values of t, the Markov process of the phases $\{\phi(t), t \geq 0\}$, defined in (4.4), becomes stable. Once the stability is reached, the process $\phi(t)$ is in a certain phase according to the probability values given by the entries of the vector $\boldsymbol{\pi}$.

Therefore the expected value of the probability to die in a catastrophe may be approximated, for large values of t, by

$$\bar{\delta} = \sum_{i=1}^{m} \delta_i \mathrm{P}[\text{Phase of the process is } i] \simeq \sum_{i=1}^{m} \delta_i \pi_i.$$

So if we approximate Δ_δ with $\bar{\delta}I$, we obtain an approximation for ψ given by

$$\bar{\psi} = \log(\bar{\delta}) = \log(\sum_{i=1}^{m} \pi_i \delta_i) = \psi_l^{(norm)},$$

where the first equality holds thanks to Theorem 4.1.1.

It is very interesting to compare the newfound lower bound with ψ_l, defined in (4.7). We observe that

$$\psi_l^{(norm)} = \log(\sum_{i=1}^{m} \pi_i \delta_i) \geq \sum_{i=1}^{m} \pi_i \log \delta_i = \psi_l. \qquad (4.20)$$

This inequality follows directly from the concavity of the log function and from the fact that $\sum_{i=1}^{m} \pi_i = 1$.

Spectral radius formulation

We focus our attention now on the formulation for ψ involving the spectral radius, see (4.15). We define the function

$$\rho(t) = \rho(V(t)), \qquad (4.21)$$

as the value of the spectral radius associated the matrix $V(t) = e^{\Theta t}\Delta_\delta$. We observe that the matrix $e^{\Theta t}$ is nonnegative for $t \geq 0$ and turns out to be irreducible for $t > 0$. In fact the irreducibility is guaranteed by the definition of Θ, see (3.18), and by the assumed irreducibility of Ω, defined in (2.16). Moreover, we observe that the condition $\delta_i > 0$ for each $i = 1, \ldots, m$ is sufficient to assure the matrix $V(t)$ to be nonnegative and irreducible for every $t > 0$.

Under this mild hypotheses, it is possible to apply the Perron Frobenius Theorem, see Theorem 2.1.2, to each matrix $V(t)$. So that for every $t > 0$, $\rho(t)$ turns out to be the simple maximum modulus eigenvalue descending from the Perron Frobenius Theorem and moreover there must exist two positive vectors $\boldsymbol{u}(t)$ and $\boldsymbol{v}(t)$, such that

$$\boldsymbol{u}(t)^T V(t) = \rho(t)\boldsymbol{u}(t)^T, \qquad V(t)\boldsymbol{v}(t) = \rho(t)\boldsymbol{v}(t). \qquad (4.22)$$

Hence, we refer to Andrew et al., see [2, Theorem 2.1.], fitting their result to our scenario by stating the following theorem

Theorem 4.3.1. *If $V(t) = e^{\Theta t}\Delta_\delta$ is nonnegative and irreducible we have that the scalar function $\rho(t)$, defined in (4.21), is an analytic function on the domain $(0, \infty)$. Moreover, given the positive eigenvectors $\boldsymbol{u}(t)$ and $\boldsymbol{v}(t)$, defined in (4.22) and normalized as follow*

$$\boldsymbol{u}(t)^T \boldsymbol{v}(t) = 1 \qquad \boldsymbol{u}(t)^T \boldsymbol{e} = 1, \qquad (4.23)$$

we have that $\boldsymbol{u}(t)$ and $\boldsymbol{v}(t)$ are analytic vector functions on the domain $(0, \infty)$.

This theorem is a corollary of [2, Theorem 2.1.], and is proved just by ob-

serving that, under the hypotheses of Theorem 4.3.1, the maximum modulus eigenvalue of $V(t)$ is simple for $t > 0$.

We conjecture that the study of the function $\rho(t)$ is of a certain interest since we would like to compare the parameter ψ, defined in (4.15), to the function $\rho(t)$ computed in some particular point. We state now some remarks concerning the function $\rho(t)$ with the target of obtaining some information about its behavior in comparison with the parameter ψ.

Lemma 4.3.1. *The value of the function $\rho(t)$ is known in the cases wherein $t \to \infty$ and $t = 0$, in particular*

- *We have that $\rho(0) = \delta_{\max}$, where $\delta_{\max} = \max_{i=1,\ldots,m} \delta_i$.*
  Moreover, if we suppose $\delta_{\max} = \delta_i$ and $\delta_j < \delta_{\max}$ for each $j \neq i$ the functions $\boldsymbol{u}(t)$, $\boldsymbol{v}(t)$ and $\rho(t)$ are analytic on the domain $[0, \infty)$ and $\boldsymbol{u}(0) = \boldsymbol{v}(0) = \boldsymbol{e}_i$.

- *The limit for t going to infinity of the function $\rho(t)$ exists and we have that*

$$\lim_{t \to \infty} \rho(t) = \sum_{i=1}^{m} \pi_i \delta_i, \qquad (4.24)$$

where $\boldsymbol{\pi}$ is the steady state probability vector associated to the generator matrix Θ.

Moreover, it holds that

$$\lim_{t \to \infty} \boldsymbol{v}(t) = \boldsymbol{e}, \qquad \lim_{t \to \infty} \boldsymbol{u}(t)^T = \frac{1}{\rho_\infty} \boldsymbol{\pi}^T \Delta_\delta, \qquad (4.25)$$

where $\rho_\infty = \sum_{i=1}^{m} \pi_i \delta_i$.

Proof. • We observe that the function $\rho(t)$ can be easily evaluated in $t = 0$, in fact

$$\rho(0) = \rho(V(0)) = \rho(e^{\Theta 0} \Delta_\delta) = \rho(\Delta_\delta) = \delta_{\max} = \delta_i. \qquad (4.26)$$

Moreover, the hypothesis $\delta_j < \delta_{\max}$ for each $j \neq i$ guarantees us that the maximum modulus eigenvalue $\delta_{\max}$ of Δ_δ is simple. Therefore it is possible to extend the analyticity of the functions $\rho(t)$, $\boldsymbol{u}(t)$ and $\boldsymbol{v}(t)$

73

even to $t = 0$, thanks to [2, Theorem 2.1.].

This leads to the uniqueness of the normalized left and right eigen-vectors, and in particular it is possible to check trivially that $\boldsymbol{u}(0) = \boldsymbol{v}(0) = \boldsymbol{e}_i$.

- We recall that, since Θ is a generator matrix,

$$\lim_{t \to \infty} e^{\Theta t} = \boldsymbol{e}\boldsymbol{\pi}^T.$$

Therefore it is possible to evaluate $\lim_{t \to \infty} \rho(t)$, indeed

$$\lim_{t \to \infty} \rho(t) = \lim_{t \to \infty} \rho(V(t)) = \lim_{t \to \infty} \rho(e^{\Theta t} \Delta_\delta) = \rho(\boldsymbol{e}\boldsymbol{\pi}^T \Delta_\delta), \qquad (4.27)$$

So that $\lim_{t \to \infty} \rho(t)$ is given by the spectral radius of a rank 1 matrix, that can be easily computed

$$\lim_{t \to \infty} \rho(t) = \rho(\boldsymbol{e}\boldsymbol{\pi}^T \Delta_\delta) = \boldsymbol{\pi}^T \Delta_\delta \boldsymbol{e} = \sum_{i=1}^{m} \pi_i \delta_i. \qquad (4.28)$$

Since $\lim_{t \to \infty} V(t)$ is a rank 1 matrix, it is easy to check the statement concerning the eigenvectors.

$\square$

We focus now on the analysis of the first derivative of the function $\rho(t)$, which can be easily expressed as stated by the following lemma

Lemma 4.3.2. *Under the hypotheses of Theorem 4.3.1, it is possible to express the derivative of the function $\rho(t)$ for each $t > 0$ in the following way*

$$\rho(t)' = \rho(t)\boldsymbol{u}(t)^T \Theta \boldsymbol{v}(t), \qquad (4.29)$$

where $\boldsymbol{u}(t)$ and $\boldsymbol{v}(t)$ are the eigenvectors of $V(t)$ defined as in Theorem 4.3.1.

Proof. Firstly, we observe that due to Theorem 4.3.1, the first derivative of the function $\rho(t)$ is guaranteed to exist for $t > 0$ since an analytic function is infinitely differentiable.

We proceed by differentiating the equation

$$V(t)\boldsymbol{v}(t) = \rho(t)\boldsymbol{v}(t)$$

and by left multiplying both terms of the obtained equation by $\boldsymbol{u}(t)^T$, yielding the following equation

$$\boldsymbol{u}(t)^T(V(t)'\boldsymbol{v}(t) + V(t)\boldsymbol{v}(t)') = \boldsymbol{u}(t)^T(\rho(t)'\boldsymbol{v}(t) + \rho(t)\boldsymbol{v}(t)').$$

Hence, expanding the products, it is verified

$$\boldsymbol{u}(t)^T V(t)\Theta\boldsymbol{v}(t) + \rho(t)\boldsymbol{u}(t)^T\boldsymbol{v}(t)' = \rho(t)'\boldsymbol{u}(t)^T\boldsymbol{v}(t) + \rho(t)\boldsymbol{u}(t)^T\boldsymbol{v}(t)'$$
$$\boldsymbol{u}(t)^T V(t)\Theta\boldsymbol{v}(t) = \rho(t)'\boldsymbol{u}(t)^T\boldsymbol{v}(t)$$

and the proposition is proved since $\boldsymbol{u}(t)^T\boldsymbol{v}(t) = 1$ and $\boldsymbol{u}(t)^T V(t) = \rho(t)\boldsymbol{u}(t)^T$.

$\square$

Hence it is possible to evaluate the derivative of the function $\rho(t)$ in some special point. We recall that if there exist $\delta_{\max}$ such that $\delta_i = \delta_{\max}$ and $\delta_j < \delta_{\max}$ for each $j \neq i$, the function $\rho(t)$ becomes analytic on the domain $[0, \infty)$, and it is licit to to compute the derivative of $\rho(t)$ also in $t = 0$.

Corollary 4.3.2.   • *If the function $\rho(t)$ is analytic on the domain $[0, \infty)$, it is possible to compute $\rho(0)'$, in fact*

$$\rho(0)' = \delta_{\max}\Theta_{ii} < 0, \tag{4.30}$$

where Θ is defined in (3.18).

• *The limit for t going to infinity of the derivative of the function $\rho(t)$ exists and we have that*

$$\lim_{t \to \infty} \rho(t)' = 0. \tag{4.31}$$

Proof.   • If the hypothesis of analyticity holds, we have in particular that the function $\rho(t)'$ needs to be continuous on the domain $[0, \infty)$, therefore it is possible to compute $\rho(0)'$ through the limit $\rho(0)' =$

$\lim_{t \to 0^+} \rho(t)'$, so that

$$\rho(0)' = \lim_{t \to 0^+} \rho(t)' = \lim_{t \to 0^+} \rho(t)\boldsymbol{u}(t)^T \Theta \boldsymbol{v}(t) = \delta_{\max}\boldsymbol{e}_i^T \Theta \boldsymbol{e}_i = \delta_{\max}\Theta_{ii}.$$
(4.32)

thanks to Lemma 4.3.2 and Lemma 4.3.1.

- We exploit Lemma 4.3.2 and Lemma 4.3.1 even in this case, in fact

$$\lim_{t \to \infty} \rho(t)' = \lim_{t \to \infty} \rho(t)\boldsymbol{u}(t)^T \Theta \boldsymbol{v}(t) = \boldsymbol{\pi}^T \Delta_\delta \Theta \boldsymbol{e} = 0,$$
(4.33)

where the last equality descends from the fact that Θ is a rate matrix and therefore $\Theta \boldsymbol{e} = 0$.

$\square$

We observe that the function $\rho(t)$ is limited on the domain $[0, \infty)$, as it is stated by the following Lemma

Lemma 4.3.3. *For every $t \geq 0$, it is verified that*

$$\rho(0) \geq \rho(t) \geq \lim_{t \to \infty} \rho(t)$$
(4.34)

Proof. In order to prove this result, we need to recall Gelfand's formula, stating that for every matrix norm and square matrix A, the spectral radius of A can be computed as the following limit

$$\rho(A) = \lim_{N \to \infty} \|A^N\|^{1/N}.$$
(4.35)

In particular it is true that

$$\rho(t) = \rho(V(t)) = \lim_{N \to \infty} \|V(t)^N\|_\infty^{1/N},$$
(4.36)

and by means of Proposition 4.3.1 we can conclude, indeed

$$\begin{aligned}
\rho(t) &= \lim_{N\to\infty} \|V(t)^N\|_\infty^{1/N} \\
&\leq \lim_{N\to\infty} \|\Delta_\delta^N\|_\infty^{1/N} \\
&= \lim_{N\to\infty} \|V(0)^N\|_\infty^{1/N} \\
&= \rho(V(0)) = \rho(0),
\end{aligned}$$

and

$$\begin{aligned}
\rho(t) &= \lim_{N\to\infty} \|V(t)^N\|_\infty^{1/N} \\
&\geq \lim_{N\to\infty} \|(\mathbf{e}\boldsymbol{\pi}^T\Delta_\delta)^N\|_\infty^{1/N} \\
&= \lim_{N\to\infty} \lim_{t\to\infty} \|V(t)^N\|_\infty^{1/N} \\
&= \lim_{t\to\infty} \rho(V(t)) \\
&= \lim_{t\to\infty} \rho(t),
\end{aligned}$$

$\square$

This results leads to the following Corollary, which explains what we meant by comparing the function $\rho(t)$ with the parameter ψ.

Corollary 4.3.3. *There exists $\bar{t} \in [0,\infty)$ such that $\psi = \log\rho(\bar{t})$, where ψ is defined in* (4.15).

Proof. This corollary follows from the continuity of the function $\rho(t)$, proved in Theorem 4.3.1, by the monotony of the logarithmic function and by the fact that

$$\log\rho(0) = \log(\delta_{\max}) \geq \psi \geq \log(\sum_{i=1}^m \pi_i\delta_i) = \lim_{t\to\infty}\log\rho(t), \qquad (4.37)$$

which is guaranteed by Lemma 4.3.1 and by Corollary 4.3.1. $\square$

At this point, we strongly believe in the truth of the following conjecture,

Conjecture 4.3.1. *The function $\rho(t)$, defined in (4.21), is a decreasing convex function on the domain $[0, \infty)$.*

If this conjecture turns out to be true, it could be guaranteed the uniqueness of the $\bar{t} \in [0, \infty)$ such that $\psi = \log \rho(\bar{t})$, leading possibly to new methods for bounding and approximating such a parameter. We report here some examples, showing the behavior of $\log \rho(t)$ in these cases:

- We suppose Θ of this form

$$\Theta = \begin{bmatrix} -1 & 1 & & & \\ 1 & -2 & 1 & & \\ & 1 & -2 & 1 & \\ & & 1 & -2 & 1 \\ & & & 1 & -1 \end{bmatrix}$$

and a random Δ_δ. The behavior of $\log \rho(t)$ is shown in Figure 4.3.
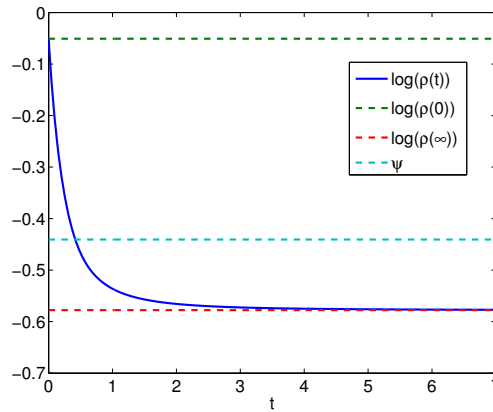
Figure 4.1: Behavior of $\log \rho(t)$, 1st example.

- We consider now two other examples, more similar to those we have to deal with if we consider Θ given by (3.18). In both cases, we consider Θ to be a full random generator matrix, and we generate $\delta_1, \ldots, \delta_m$ in a random way too. In these cases, the behaviors of $\log \rho(t)$ are illustrated in Figure 4.3, wherein the expected interval between two consecutive
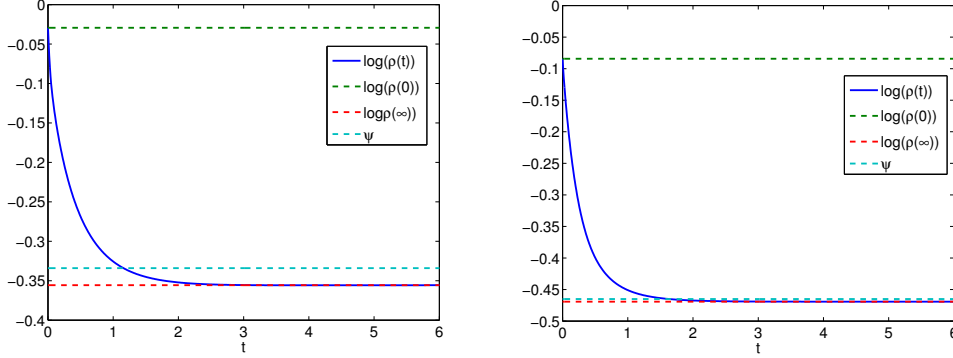
78

Figure 4.2: Behavior of $\log \rho(t)$, 2nd and 3rd examples.

catastrophes is given by $E[\xi] = 5$ and $E[\xi] = 15$. We observe that increasing $E[\xi]$, it seems that $\psi \to \lim_{t\to\infty} \log \rho(t)$, however this is just an intuitive observation without an appropriate theoretical foundation.

Another interesting result is provided by adapting the *mixing inequality* proved by Cohen et al. in [8, Theorem 4]. We introduce the diagonal matrix $\Delta_{\log \delta}$ such that $(\Delta_{\log \delta})_{ii} = \log \delta_i \leq 0$, hence we have that

$$\Delta_\delta = e^{\Delta_{\log \delta}}.$$

Lemma 4.3.4. *Given the function* $\rho(t) = \rho(V(t)) = \rho(e^{\Theta t}\Delta_\delta)$, *for every* $t_1, \ldots, t_n \geq 0$, *the following chain of inequalities holds,*

$$\rho(e^{T\Theta + n\Delta_{\log \delta}}) \leq \rho(e^{t_1\Theta}\Delta_\delta \ldots e^{t_n\Theta}\Delta) \leq \rho(e^{T\Theta}\Delta_\delta^n), \qquad (4.38)$$

where $T = t_1 + \ldots + t_n$

This inequality can be exploited immediately in order to obtain another couple of bounds for the parameter ψ defined by (4.15).

Corollary 4.3.4. *Given an ergodic sequence* $\eta_1, \eta_2, \ldots > 0$, *we have that*

$$\lim_{n\to\infty} \log \rho(e^{n(E[\eta]\Theta + \Delta_{\log \delta})})^{1/n} \leq \psi \leq \lim_{n\to\infty} \log \rho(e^{nE[\eta]\Theta}\Delta_\delta^n)^{1/n} \qquad (4.39)$$

where $E[\eta] = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \eta_i$.

Unfortunately, numerical experiments seem to assert that these last bounds are not as effective as the ones that we already have.

Chapter 5

Numerical experimentation

In the previous chapter, we explained how the parameter ψ, defined in (3.19) may be bounded. Although an exact computation for ψ is not available, we hope these bounds to be sufficient for the deduction of the positivity or negativity of the parameter ω, which can be computed as stated in Lemma 3.2.2. In fact, we recall that the sign of ω plays a discriminating role in determining if a MBT subject to catastrophes is doomed to extinction or not, see Theorem 3.1.2.

We briefly report here the bounds that we have at our disposal for the parameter ψ.

- In the work of Hautphenne et al., see [15], were proposed a couple of bounds, that we quoted in Theorem 4.2.1. In particular we have

    1. The upper bound

$$\psi_u^{\mathrm{P}} = \lim_{n \to \infty} \log \mathrm{P}[T > \theta_n, \phi_n = j \mid \phi_0 = i], \qquad (5.1)$$

    2. The lower bound

$$\psi_l^{\mathrm{P}} = \sum_{i=1}^{m} u_i v_i \log \delta_i. \qquad (5.2)$$

    The superscript $^{\mathrm{P}}$ reminds that these bounds are obtained through a probabilistic approach.

- It is of a certain interest to compare the function $\log \rho(t)$, where $\rho(t)$ is defined in (4.21), with the other bounds we possess. A particular attention is given to $\rho_{\mathrm{E}[\xi]} = \log \rho(\mathrm{E}[\xi])$, where $\mathrm{E}[\xi]$ represents the expected value of the time passing between two consecutive catastrophes. We don't have theoretical results concerning $\rho_{\mathrm{E}[\xi]}$, but perhaps its value can favor the identification of $\bar{t}$ defined in Corollary 4.3.3.

- By means of the function $\rho(t)$, it is possible to rewrite the bounds obtained through Proposition 4.3.1, indeed we have

  1. The upper bound

  $$\psi_u^{\mathrm{M}} = \log \rho(0) = \log(\delta_{\max}), \qquad (5.3)$$

  2. The lower bound

  $$\psi_l^{\mathrm{M}} = \lim_{t \to \infty} \log \rho(t) = \log\left(\sum_{i=1}^{m} \pi_i \delta_i\right), \qquad (5.4)$$

  where $\delta_{\max} = \max_{i=1,\dots,m} \delta_i$ and $\boldsymbol{\pi}$ is the steady state probability vector associated to the generator matrix Θ. The superscript $^{\mathrm{M}}$ reminds that these bounds are obtained through a matrix approach.

In order to compare this bounds to the parameter ψ, we need to compute it through simulation. This is possible when the data of the problem are sufficiently small. Indeed we simulate 1000 processes of catastrophes generating a sequence $\boldsymbol{\xi}^{(i)} = (\xi_1^{(i)}, \dots, \xi_{50}^{(i)})$ for $i = 1, \dots, 1000$, then we compute

$$\psi(i) = \frac{1}{50} \log \|V(\xi_1^{(i)}) \dots V(\xi_{50}^{(i)})\|_\infty. \qquad (5.5)$$

The parameter ψ we try to obtain is somehow approximated by the mean value of $\psi(1), \dots, \psi(1000)$.

5.1 Right whale model

This example is taken from [15] and is inspired by real data concerning North Atlantic right whales in the years 1980-81. In this model the different typologies of the individuals, i.e. the whales, are given by the different life stages. In particular the whales are differentiated among calf, immature, mature, reproducing females and postbreeding female which are associated to types $1, \ldots, 5$ in the same order. We consider a time unit of 1 year yielding the following matrices governing the MBT and defined in (2.26),

$$D_0 = \begin{bmatrix} -1 & 0.93 & \cdot & \cdot & \cdot \\ \cdot & -0.15 & 0.12 & \cdot & \cdot \\ \cdot & \cdot & -0.41 & \cdot & \cdot \\ \cdot & \cdot & \cdot & -1 & 0.97 \\ \cdot & \cdot & 0.99 & \cdot & -1 \end{bmatrix} \tag{5.6}$$

and

$$D_1 = \begin{bmatrix} \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & 0.40 & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix}. \tag{5.7}$$

Therefore the birth matrix, defined in (2.31), is given by $B = D_1 \otimes \boldsymbol{e}_1^T$ and the vector containing the death rates, see (2.29), is given by $\boldsymbol{d} = -D_0\boldsymbol{e} - B\boldsymbol{e}$. In this first example, the catastrophes are supposed to happen according to a Poisson process, with an expected interval between two catastrophes of $\mathrm{E}[\xi] = 25$, so that the upper bound ψ_u^{P} can be computed through formula (4.9). The survival probabilities are set to be $\delta_1 = \delta_3 = 0.2$ and $\delta_2 = \delta_4 = \delta_5 = 0.8$.

As can be seen in Table 5.1 the upper bound ψ_u^{P} turns out to be the best upper bound we possess, on the other hand the lower bound ψ_l^{M} is very close to the real solution and way better than the lower bound ψ_l^{P}. The approximation $\rho_{\mathrm{E}[\xi]}$ results lower than ψ but approximates ψ not a lot better that ψ_l^{M}.

	Value	Difference
ψ_u^{M}	- 0.223143551314210	+ 0.42679
ψ_u^{P}	- 0.638442689254770	+ 0.01149
ψ	- 0.649935440688771	
$\rho_{\mathrm{E}[\xi]}$	- 0.653684273135562	- 0.00374
ψ_l^{M}	- 0.653888127769856	- 0.00395
ψ_l^{P}	- 0.870034855601116	- 0.22009

Table 5.1: Results for the right whale experiment

For the sake of completeness, we observe that, due to Lemma 3.2.2, we have that $\omega = \lambda \mathrm{E}[\xi] + \psi$, therefore

$$\lambda \mathrm{E}[\xi] + \psi_l^{\mathrm{M}} \leq \omega \leq \lambda \mathrm{E}[\xi] + \psi_u^{\mathrm{P}}. \tag{5.8}$$

Since $\lambda \mathrm{E}[\xi] = 2.18879984044839$, we obtain that

$$\omega \in [1.53511, 1.55036]. \tag{5.9}$$

Therefore, since $\omega > 0$, there exist a positive probability for the right whales to not become extinct.

We reproduce now the same experiment changing the survival matrix. We consider in our second experiment, the case wherein the chances to survive are roughly halved compared with the first example. In the third experiment, we level the chances leading to a survival matrix having for diagonal elements $\mathrm{diag}(\Delta_\delta) = (0.2, 0.3, 0.35, 0.35, 0.3)$, and moreover we double the frequency of catastrophes, fixing $\mathrm{E}[\xi] = 12.5$.

By means of (5.8) and observing the results illustrated in Table 5.1, we deduce by computing $\lambda \mathrm{E}[\xi]$ that in the second case

$$\omega \in [0.84177, 0.85702], \tag{5.10}$$

	2nd experiment		3rd experiment	
	Value	Difference	Value	Difference
ψ_u^{M}	- 0.916290731874155	+ 0.42679	- 1.04982212449868	+ 0.09030
ψ_u^{P}	- 1.33178175544091	+ 0.01130	- 1.13864252213307	+ 0.00148
ψ	- 1.34308275648623		- 1.14012759360271	
$\rho_{\mathrm{E}[\xi]}$	- 1.34702755033161	- 0.00394	- 1.14017629343684	- 0.00005
ψ_l^{M}	- 1.34703530832980	- 0.00395	- 1.14022317192362	- 0.00009
ψ_l^{P}	- 1.56318203616106	- 0.22009	- 1.14930646648308	- 0.00918

Table 5.2: Results for the right whale experiment, second and third experiment.

while in the third

$$\omega \in [-0.04578, -0.04424]. \tag{5.11}$$

Therefore in the second case the population of whales won't become extinct with a positive probability, despite the parameter ω in this case is smaller than in the first because of the reduced survival probabilities. On the other hand, in the third example we observe that the parameter ω is negative implying the certain extinction of the whale population.

5.1.1 Insect model

Like the previous example, also this one is taken from [15]. We deal with a particular species of insect, which doesn't reproduce until its death. On the other hand, there are various occasion for an insect of this kind to die before reaching the reproductive stage. Once it reproduces, a geometrically distributed number of offsprings is produced.

We consider the following matrices governing the MBT and defined in (2.26),

$$
D_0 =
\begin{bmatrix}
-\alpha_2 & \gamma_0 & & & \\
& -\alpha_2 & \gamma_0 & & \\
& & \ddots & \ddots & \\
& & & -\alpha_2 & \gamma_0 \\
& & & & -\alpha_1
\end{bmatrix}
\tag{5.12}
$$

and

$$
D_1 =
\begin{bmatrix}
0 & & & & \\
& 0 & & & \\
& & \ddots & & \\
& & & 0 & \\
& & & & \gamma_1
\end{bmatrix}.
\tag{5.13}
$$

As in the previous example, the birth matrix, defined in (2.31), can be computed through $B = D_1 \otimes e_1^T$ and the vector containing the death rates, see (2.29), is given by $d = -D_0 e - Be = (\gamma_2, \ldots, \gamma_2, \gamma_0)$, where $\alpha_2 = \gamma_0 + \gamma_2$ and $\alpha_1 = \gamma_0 + \gamma_1$. The stages of life of the insects are denoted by m.

These parameters can be chosen in order for the insects to have a certain expected life $L = m/\gamma_0$ and to produce an expected number of eggs $E = \gamma_1/\gamma_0$ once reached its last stage of life. The number $C = E(\gamma_0/(\gamma_0 + \gamma_2))^{m-1}$ represents the expected number of eggs that an insect produces conditioned that is still in its first stage of life.

In this batch of examples we fix $m = 5$, $L = 12$, $E = 100$, $C = 2$ and consider three different Markovian arrival processes determining the process of catastrophes:

- A renewal process with Erlang distributed intervals between renewals, with order 6 and parameter equal to 0.5. So that

$$
A_0 =
\begin{bmatrix}
-0.5 & 0.5 & & \\
& \ddots & \ddots & \\
& & \ddots & 0.5 \\
& & & -0.5
\end{bmatrix}
, \quad
A_1 =
\begin{bmatrix}
0 & \cdots & \cdots & 0 \\
\vdots & \ddots & & \vdots \\
0 & & \ddots & \vdots \\
0.5 & 0 & \cdots & 0
\end{bmatrix}
, \tag{5.14}
$$

where $A_0, A_1 \in \mathbb{R}^{6 \times 6}$.

- An intermittent process, for which catastrophes are much more frequent when the MAP is in its first phase. In particular

$$A_0 = \begin{bmatrix} -21a & a & & & \\ & -2a & \ddots & & \\ & & \ddots & a & \\ & & & -2a \end{bmatrix}, \quad A_1 = \begin{bmatrix} 20a & & & \\ & a & & \\ & & \ddots & \\ & & & a \end{bmatrix}, \quad (5.15)$$

where $A_0, A_1 \in \mathbb{R}^{6 \times 6}$ and $a = 0.02$.

- A seesaw process, where the MAP is governed by the matrices

$$A_0 = \begin{bmatrix} -2b & b & & & & & \\ & -4b & b & & & & \\ & & -4b & b & & & \\ & & & -8b & b & & \\ & & & & -4b & b & \\ & & & & & -4b & b \\ b & & & & & & -2b \end{bmatrix},$$

$$A_1 = \begin{bmatrix} b & & & & & & \\ & 3b & & & & & \\ & & 3b & & & & \\ & & & 7b & & & \\ & & & & 3b & & \\ & & & & & 3b & \\ & & & & & & b \end{bmatrix}, \quad (5.16)$$

where $b = 1/36$.

Each of these catastrophic processes have the same expected value, more precisely $E[\xi] = 12$ for all of them. The results are illustrated in Table 5.1.1. It is clear that, also in this case, the best bounds that we can find for the parameter ψ are given by ψ_l^M and ψ_u^P. A point that needs to be stressed

	1st MAP		2nd MAP		3rd MAP	
	Value	Difference	Value	Difference	Value	Difference
ψ_u^{M}	- 0.369164550813516	+ 0.15549	- 0.369164550813516	+ 0.14511	- 0.369164550813516	+ 0.15154
ψ_u^{P}	- 0.521050123191361	+ 0.00360	- 0.510735670678260	+ 0.00353	-0.517680367464709	+ 0.00302
ψ	- 0.524654893539577		- 0.514271614865039		- 0.520703216296654	
$\rho_{\mathrm{E}[\xi]}$	- 0.524679369265143	- 0.00002	- 0.524679369265143	- 0.01041	- 0.524679369265143	- 0.00398
ψ_l^{M}	- 0.524679320975098	- 0.00002	- 0.524679320975098	- 0.01041	- 0.524679320975098	- 0.00398
ψ_l^{P}	- 0.552714098622790	- 0.02806	- 0.552714098622790	- 0.03844	- 0.552714098622790	- 0.03201

Table 5.3: Results for the insect experiments.

out is that the catastrophic process influences only the bound ψ_u^{P}, this fact is evident if we observe the definitions of the various bounds. However this remark tells us that other lower bounds, sensitive to the catastrophic process need to be found. Moreover, we observe that $\rho_{\mathrm{E}[\xi]}$ is very similar to ψ_l^{M}, implying in these cases that $\bar{t} \in [0, \mathrm{E}[\xi]]$, where $\bar{t}$ is defined in Corollary 4.3.3.

Bibliography

[1] A. T. Andersen, M. F. Neuts, and B. Friis Nielsen. On the time reversal of markovian arrival processes. *Stochastic Models*, 20:237–260, 2004.

[2] A. L. Andrew, K.-w. E. Chu, and P. Lancaster. Derivatives of eigenvalues and eigenvectors of matrix functions. *SIAM Journal on Matrix Analysis and Applications*, 14(4):903–926, 1993.

[3] K. B. Athreya and P. E. Ney. *Branching processes*. Springer-Verlag, 1972.

[4] N. G. Bean, N. Kontoleon, and P. G. Taylor. Markovian trees: properties and algorithms. *Annals of Operations Research*, 160(1):31–50, 2008.

[5] R. Bellman. Limit theorems for non-commutative operations. i. *Duke Mathematical Journal*, 21(3):491–500, 1954.

[6] A. Berman and R. J. Plemmons. *Nonnegative Matrices in the Mathematical Sciences*. Classics in Applied Mathematics. Society for Industrial and Applied Mathematics, 1994.

[7] D. A. Bini, B. Meini, and F. Poloni. On the solution of a quadratic vector equation arising in markovian binary trees. *Numerical Linear Algebra with Applications*, 18:981–991, 2011.

[8] J. E. Cohen, S. Friedland, T. Kato, and F. P. Kelly. Eigenvalue inequalities for products of matrix exponentials. *Linear Algebra and its Applications*, 45:55–95, 1982.

[9] J. E. Cohen and C. M. Newman. The stability of large random matrices and their products. *The Annals of Probability*, 12(2):283–310, 1984.

[10] P. J. Forrester. Lyapunov exponents for products of complex gaussian random matrices. *Journal of Statistical Physics*, 151(5):796–808, 2013.

[11] H. Furstenberg and H. Kesten. Products of random matrices. *The Annals of Mathematical Statistics*, 31(2):457–469, 1960.

[12] R. Gharavi and V. Anantharam. An upper bound for the largest lyapunov exponent of a markovian product of nonnegative matrices. *Theoretical Computer Science*, 332(1–3):543–557, 2005.

[13] T. E. Harris. *The theory of branching processes*. Springer-Verlag, 1963.

[14] S. Hautphenne. *An Algorithmic Look at Phase-Controlled Branching Processes*. PhD thesis, Université Libre de Bruxelles, Faculté des sciences, 2009.

[15] S. Hautphenne, G. Latouche, and G. T. Nguyen. Markovian trees subject to catastrophes: Would they survive forever? *Springer Proceedings in Mathematics & Statistics*, 27:87–106, 2013.

[16] S. Hautphenne, G. Latouche, and M.-A. Remiche. Newton's iteration for the extinction probability of a markovian binary tree. *Linear algebra and its applications*, 428(11):2791–2804, 2008.

[17] S. Hautphenne, G. Latouche, and M.-A. Remiche. Algorithmic approach to the extinction probability of branching processes. *Methodology and Computing in Applied Probability*, 13(1):171–192, 2011.

[18] S. Hautphenne and B. Van Houdt. On the link between markovian trees and tree-structured markov chains. *European journal of operational research*, 201(3):791–798, 2010.

[19] N. J. Higham. *Functions of matrices: theory and computation.* Society for Industrial and Applied Mathematics, 2008.

[20] S. Karlin and H. Taylor. *A First Course in Stochastic Processes.* Elsevier Science, 1975.

[21] E. Key. Computable examples of the maximal Lyapunov exponent. *Probab. Theory Related Fields*, 75(1):97–107, 1987.

[22] E. S. Key. Lower bounds for the maximal lyapunov exponent. *Journal of Theoretical Probability*, 3(3):477–488, 1990.

[23] B. J. Kim and G. H. Choe. High precision numerical estimation of the largest lyapunov exponent. *Communications in Nonlinear Science and Numerical Simulation*, 15(5):1378–1384, 2010.

[24] J. F. C. Kingman. Subadditive ergodic theory. *The annals of Probability*, 1:883–899, 1973.

[25] N. Kontoleon. *The Markovian Binary Tree: a model of the macroevolutionary process.* PhD thesis, University of Adelaide, School of Mathematical Sciences, Discipline of Applied Mathematics, 2006.

[26] G. Latouche and V. Ramaswami. *Introduction to matrix analytic methods in stochastic modeling*, volume 5. Society for Industrial and Applied Mathematics, 1987.

[27] G. Latouche, M.-A. Remiche, and P. Taylor. Transient markov arrival processes. *Annals of Applied Probability*, 13(2):628–640, 2003.

[28] R. Mainieri. Zeta function for the lyapunov exponent of a product of random matrices. *Phys. Rev. Lett.*, 68(13):1965–1968, 1992.

[29] B. Meini and F. Poloni. A perron iteration for the solution of a quadratic vector equation arising in markovian binary trees. *SIAM Journal on Matrix Analysis and Applications*, 32(1):248–261, 2011.

[30] C. Mode. *Multitype branching processes: theory and applications.* American Elsevier Pub. Co., 1971.

[31] J. Nielsen. Lyapunov exponents for products of random matrices, 1997.

[32] J. M. Ortega and W. C. Rheinboldt. *Iterative solution of nonlinear equations in several variables.* Number 30. Siam, 2000.

[33] V. I. Oseledec. A multiplicative ergodic theorem. lyapunov characteristic numbers for dynamical systems. *Trans. Moscow Math. Soc*, 19(2):197–231, 1968.

[34] E. Pollak. *Stochastic Theory of Gene Frequencies in Subdivided Populations.* Columbia University, 1964.

[35] J. H. Pollard. On the use of the direct matrix product in analysing certain stochastic population models. *Biometrika*, 53(3-4):397–415, 1966.

[36] M. Pollicott. Maximal lyapunov exponents for random matrix products. *Inventiones mathematicae*, 181(1):209–226, 2010.

[37] F. Riesz and B. Nagy. *Functional Analysis.* DOVER PUBN Incorporated, 1990.

[38] B. Sevast'yanov. The theory of branching random processes. *Uspekhi Mat. Nauk*, 6:47–99, 1951.

[39] D. Tanny. Limit theorems for branching processes in a random environment. *The Annals of Probability*, 5:100–116, 1977.

[40] D. Tanny. On multitype branching processes in a random environment. *Advances in Applied Probability*, 13:464–497, 1981.

[41] J. N. Tsitsiklis and V. D. Blondel. The lyapunov exponent and joint spectral radius of pairs of matrices are hard—when not impossible—to compute and to approximate. *Mathematics of Control, Signals and Systems*, 10:31–40, 1997.