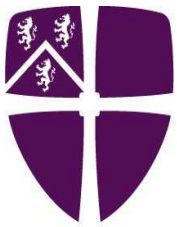
Limiting Conditional Distributions: Imprecision and Relation to the Hazard Rate

Richard J. Crossman

A Thesis presented for the degree of
Doctor of Philosophy



Statistics and Probability Research Group
Department of Mathematical Sciences
University of Durham
England

May 2009

Dedicated to

Pauline, Aenea, and Josh

Limiting Conditional Distributions: Imprecision and Relation to the Hazard Rate

Richard J Crossman

Submitted for the degree of Doctor of Philosophy
May 2009

Abstract

Many Markov chains with a single absorbing state have a unique *limiting conditional distribution* (LCD) to which they converge, conditioned on non-absorption, regardless of the initial distribution. If this limiting conditional distribution is used as the initial distribution over the non-absorbing states, then the probability distribution of the process at time n, conditioned on non-absorption, is equal for all values of $n > 0$. Such an initial distribution is known as the *quasi-stationary distribution* (QSD). Thus the LCD and QSD are equal. These distributions can be found in both the discrete-time and continuous-time case.

In this thesis we consider finite Markov chains which have one absorbing state, and for which all other states form a set which is a single communicating class. In addition, every state is aperiodic. These conditions ensure the existence of a unique LCD. We first consider continuous Markov chains in the context of *survival analysis*. We consider the *hazard rate*, a function which measures the risk of instantaneous failure of a system at time t conditioned on the system not having failed before t. It is well-known that the QSD leads to a constant hazard rate, and that the hazard rate generated by any other initial distribution tends to that constant rate. Claims have been made by Aalen [1], and Aalen and Gjessing [2] that it may be possible to predict the shape of hazard rates generated by *phase type distributions* (first passage time distributions generated by atomic initial distributions) by comparing these initial distributions with the QSD. In Chapter 2 we consider these claims, and demonstrate through the use of several examples that the behaviour considered by

those conjectures is more complex then previously believed.

In Chapters 3 and 4 we consider discrete Markov chains in the context of imprecise probability. In many situations it may be unrealistic to assume that the transition matrix of a Markov chain can be determined exactly. It may be more plausible to determine upper and lower bounds upon each element, or even determine closed sets of probability distributions to which the rows of the matrix may belong. Such methods have been discussed by Kozine and Utkin [42] and Škulj [62], [63], and in each of these papers results were given regarding the long-term behaviour of such processes. None of these papers considered Markov chains with an absorbing state. In Chapter 3 we demonstrate that, under the assumption that the transition matrix cannot change from time step to time step, there exists an imprecise generalisation to both the LCD and the QSD, and that these two generalisations are equal. In Chapter 4, we prove that this result holds even when we no longer assume that the transition matrix cannot change from time step to time step. In each chapter, examples are presented demonstrating the convergence of such processes, and Chapter 4 includes a comparison between the two methods.

Declaration

The work in this thesis is based on research carried out at the Statistics and Probability Research Group, Department of Mathematical Sciences, Durham University, England. No part of this thesis has been submitted elsewhere for any other degree or qualification and is all my own work unless referenced to the contrary in the text.

Acknowledgements

So many people to acknowledge, so little space. Thanks to Frank for supervising me, and to both Frank and Iain for their mathematical wisdom and help in general. My further gratitude to Erik and Phil, with whom I had several detailed discussions and exchanges of examples which informed Chapter 2. Thanks also to Matthias and Dave, for interesting discussions and the beer that accompanied them. Damjan gets thanks for that as well, but in addition has my endless gratitude for his contributions to Chapter 4, written partly during his two months in Durham and partly during my three-week stay in Slovenia (thanks also to the University of Ljubljana for letting me haunt their halls for that period), and for a fascinating and useful e-mail correspondence.

Further thanks to Anna, Kate, Chris and Tara for keeping me sane; my parents for so generously ensuring I couldn't bankrupt myself (and for proofreading duties); Becky, Becka, Jonathan and Joey for keeping me supplied with fuel (of one kind or another); Paul, James and Jamie for distracting me with bright lights and shifting colours; and Nathan, Rachel and Ben for picking up the torch in various different ways. There are a lot of other people who contributed to the process of keeping me on the rails, so suffice it to say, if you've accompanied me to the New Inn, Queen's Head, Elm Tree, Woodsman, Tap, or that restauraunt in Prague where someone threw a plate at Scott Ferson, then I appreciate the effort.

Above all else, though, I want to thank my first supervisor Pauline, who gave up so much of her time for me, for discussions and corrections and general support. None of us have as much time as we'd like, and Pauline had far less than most, and far, far less than she deserved. I am proud of this thesis, and I very much hope that she would be too.

Contents

Chapter 1

Introduction

1.1 Background and Motivation

Markov chains provide a flexible method with which to model many real-world situations. A great deal of work has been done on describing the long-term behaviour of Markov chains under various conditions. In general, a Markov process will tend toward a pattern of behaviour which does not depend on the initial distribution of the process. This long-term behaviour is referred to as the *limiting distribution*. The irreducibility and aperiodicity of a finite chain are sufficient conditions for this to occur. Sometimes it is alternatively called the *stationary distribution*, but in this thesis that term is reserved for using the limiting distribution as the *initial distribution* for the process. Taking the stationary distribution as the initial distribution ensures that the distribution of the process at time step n is identical for all values of $n > 0$.

The specific case considered in this thesis is the situation in which a Markov chain contains an *absorbing state*, which is a state that, once reached, can never be left. Markov chains of this type have for example been used to model animal populations, such as those of reindeer imported onto the island of St George in 1911 (Scheffer [55]). Other alternative applications include catalytic reactions (Parsons and Pollett [52]) and marriage lifetimes (see Aalen [2]).

It is well known that in general eventual absorption is inevitable in such cases. It is also known, however, that in certain situations the expected time to absorption

is so long that the process can settle into a pattern of behaviour before absorption occurs. By conditioning upon non-absorption then, a distribution can be found that describes the pattern of behaviour before absorption occurs. Such a distribution is called the *limiting conditional distribution* (LCD). Sufficient conditions for the existence of the LCD for a finite chain are that the non-absorbing states form a single communicating set, and each state is aperiodic. It is also sometimes referred to as the *quasi-stationary distribution* (QSD), but, in a similar manner to the stationary distribution, we reserve the term quasi-stationary distribution for the initial distribution over the non-absorbing states which is equal to the limiting conditional distribution. Taking the quasi-stationary distribution as the initial distribution ensures that the distribution of the process at time step n, conditioned on non-absorption, is identical for all values of n.

These ideas and terms are applicable to both discrete-time and continuous-time Markov chains. In this thesis both discrete and continuous Markov chains will be considered, although the context in which discrete chains are considered is very different to the one in which continuous chains are considered. In both cases, we assume a finite number of states, and that the non-absorbing states form a single communicating class, each state of which is aperiodic. Doing so guarantees the existence of a unique limiting conditional distribution, and hence a unique quasi-stationary distribution.

In the continuous case, we consider Markov chains in combination with the *survival function*, which describes the probability of absorption not having occurred by a specific time, and the *hazard rate*, which is calculated from the survival function and describes the rate of failure at any point, given that by that point failure has not occurred. Both these concepts can be found in e.g. Aalen *et al.* [3]; examples of the application of hazard rates are given by Olave and Salvador [48] (an example in finance), and Rogers *et al.* [53] (an example from the world of medicine). It is well known that by starting with the quasi-stationary distribution the hazard rate will be constant, and moreover that by starting with any other initial distribution over the non-absorbing states, the hazard rate will eventually converge to that same constant value. What is less well understood is the behaviour of the hazard rate *before* con-

vergence. Certain situations are well described, others are only partially described or almost entirely unknown. It has been suggested by Aalen and Gjessing [2] that there are circumstances under which the shape of a hazard rate can be predicted by comparing the corresponding initial distribution with the quasi-stationary distribution. Specifically, in situations in which the distance between a transient state and the absorbing state can be sensibly determined, and in which the only possible initial distributions are those which guarantee a given starting state, it may be possible to predict the shape of a hazard rate by comparing the corresponding initial distribution with the quasi-stationary distribution. In Chapter 2, the specific conjectures put forward in [2] will be considered at length, and attempts to mathematically rigorise them will be made. Chapter 2 will demonstrate that the behaviour these conjectures consider is in fact more complicated than initially believed.

Chapters 3 and 4 deal with the modelling of discrete-time Markov chains in the context of *imprecise probability* (Walley [66]). There are situations in which a discrete-time Markov chain is a reasonable model for a situation, but for which exact transition probabilities may be difficult or even impossible to find. In such situations imprecise probability may be very useful. Rather than each probability having an exact value, each one is assumed to lie within a closed set. Further, rather than using a single probability vector as the initial distribution, a set of probability vectors are used as possible initial distributions.

The challenge then is to discover what can still be said regarding the long-term behaviour of the process. The time-homogeneous case, in which each transition probability is assumed to be independent of time, has been considered by Kozine and Utkin [42]. The time-inhomogeneous case, in which transition probabilities are not only unknown but may change from step to step, has been considered by Škulj [62], [63]. In both these cases, however, absorption is still in general a certainty. In this thesis, then, it is shown how one can consider the long-term behaviour of imprecise Markov chains with an absorbing state conditioned upon non-absorption. If [42], [62] and [63] thus generalise the concept of the limiting and stationary distributions, this thesis generalises the limiting conditional and quasi-stationary distributions. Specifically, we prove that such a generalisation does indeed exist, and is independent

of the choice of initial distributions.

1.2 Literature and Notation

In this section a brief review is presented on previous work regarding the concepts of Markov chains (both discrete and continuous) with an absorbing state, and of imprecise probability. We also introduce notation and terminology which will be used throughout the thesis.

1.2.1 Markov Chains

In this subsection we introduce the notation that will be used for the state space of a Markov chain, and for the transition probabilities or intensities for the discrete and continuous cases, respectively.

Discrete Time Markov Chains

The notation we use for discrete-time Markov chains is as follows. Let $\mathcal{X} = \{X(n),\ n = 0, 1, \ldots\}$ be a discrete-time Markov chain on the state space $S = \{-1\} \cup C$ with $C = \{0, 1, \ldots, s\}$, where -1 is an absorbing state, and C is a set of transient states. In general, we assume that C is a single communicating class, and each state in C is aperiodic. We also assume the absorbing state is reachable from C. These conditions are sufficient to ensure that a unique LCD exists for the chain. We denote by $p_{ij}^{(n)}$ the probability that the chain is in state j at time $n+1$, given that it is state i at time n, i.e. $p_{ij}^{(n)} = P(X(n+1) = j | X(n) = i)$. We can thus define the transition matrix P_n as follows:

$$P_n = \begin{pmatrix} 1 & 0 & 0 & 0 & \ldots & 0 & 0 \\ p_{0,-1}^{(n)} & p_{00}^{(n)} & p_{01}^{(n)} & p_{02}^{(n)} & \cdots & p_{0s-1}^{(n)} & p_{0s}^{(n)} \\ p_{1,-1}^{(n)} & p_{10}^{(n)} & p_{11}^{(n)} & p_{12}^{(n)} & \cdots & p_{1s-1}^{(n)} & p_{1s}^{(n)} \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ p_{s-1,-1}^{(n)} & p_{s-1,0}^{(n)} & p_{s-1,1}^{(n)} & p_{s-1,2}^{(n)} & \cdots & p_{s-1,s-1}^{(n)} & p_{s-1,s}^{(n)} \\ p_{s,-1}^{(n)} & p_{s0}^{(n)} & p_{s1}^{(n)} & p_{s2}^{(n)} & \cdots & p_{ss-1}^{(n)} & p_{ss}^{(n)} \end{pmatrix}. \tag{1.1}$$

We also denote the initial distribution by $\boldsymbol{\pi}(0)$.

A discrete Markov chain is either *time-homogeneous* or *time-inhomogeneous*. In a time-homogeneous chain, the one-step transition probabilities are assumed to be independent of time, thus $P(X(n+1) = j \,|X(n) = i\,) = p_{ij}^{(n)} =: p_{ij}$, $i, j \in S$ for all $n \geq 0$. The one-step transition probability matrix $P = (p_{ij})_{i,j \in S}$ of the chain can therefore be written as

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 & \dots & 0 & 0 \\ p_{0,-1} & p_{00} & p_{01} & p_{02} & \cdots & p_{0s-1} & p_{0s} \\ p_{1,-1} & p_{10} & p_{11} & p_{12} & \cdots & p_{1s-1} & p_{1s} \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ p_{s-1,-1} & p_{s-1,0} & p_{s-1,1} & p_{s-1,2} & \cdots & p_{s-1,s-1} & p_{s-1,s} \\ p_{s,-1} & p_{s0} & p_{s1} & p_{s2} & \cdots & p_{ss-1} & p_{ss} \end{pmatrix}. \tag{1.2}$$

Let P^* be defined as follows

$$P^* = \begin{pmatrix} p_{00} & p_{01} & p_{02} & \cdots & p_{0s-1} & p_{0s} \\ p_{10} & p_{11} & p_{12} & \cdots & p_{1s-1} & p_{1s} \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ p_{s-1,0} & p_{s-1,1} & p_{s-1,2} & \cdots & p_{s-1,s-1} & p_{s-1,s} \\ p_{s0} & p_{s1} & p_{s2} & \cdots & p_{ss-1} & p_{ss} \end{pmatrix}. \tag{1.3}$$

A matrix is called *strictly substochastic* if each row sums to a value in the interval $[0, 1]$, and at least one row sums to a value in the interval $[0, 1)$. Since at least one value $p_{i,-1}$ must be strictly positive for $i \geq 0$, in order that the absorbing state be reachable from C, it follows that P^* is a strictly substochastic matrix. For a time-homogeneous Markov chain with one-step transition probability matrix P, it is known (see e.g. [41]) that

$$p_{ij}(n) := P(X(n) = j \,|X(0) = i\,) = [P^n]_{ij} \;\forall i, j \in S. \tag{1.4}$$

Let $p_j(n) := P(X(n) = j)$ be the probability that the chain is in state j at time n, then the probability distribution of the Markov chain at time n is given by $\boldsymbol{p}(n) = (p_{-1}(n), p_0(n), \dots, p_s(n))$.

For the *time-inhomogeneous* case, the transition matrix describing the chain is allowed to change from one time step to the next. Very little work on the long-term

behaviour of time-inhomogenous chains has been done, and such work frequently assumes an underlying time-homogeneous process which is perturbed by so called "mutations" which require either strong conditions upon them or tend to zero over time (see for example Bergin and Lipman [9], or Pak [50]); this assumption of mutation from an underlying transition matrix which is independent of time may be unrealistic in practice.

The origins of the theory of quasi-stationarity for discrete chains can be in found in the work of Yaglom [71][1]. In that paper the existence of what is now known as the limiting conditional distribution was proved. Perhaps the most important work on quasi-stationary distributions in the discrete-time case is Darroch and Seneta [21], a paper which deals with finite state spaces. In this paper it is demonstrated under which conditions the limiting conditional distribution exists. Moreover, it is proved that such a distribution, if it exists, is unique, and that it equals the quasi-stationary distribution. Various other results have followed since, involving multiple absorption states or an infinite state space. These will not be discussed here, but interested readers may like to consider Seneta and Vere-Jones [60], or Kesten [40].

We now define a specific kind of discrete Markov chain, known as the *birth-death process*. In a birth-death process, a transition from state i to state j over one time step is impossible unless $|i - j| \leq 1$. Continuing to assume that each chain has -1 as an absorbing state, the transition matrix at time step n therefore takes the following form:

$$P_n = \begin{pmatrix} 1 & 0 & 0 & 0 & \dots & 0 & 0 \\ p_{0,-1}^{(n)} & p_{00}^{(n)} & p_{01}^{(n)} & 0 & \dots & 0 & 0 \\ 0 & p_{10}^{(n)} & p_{11}^{(n)} & p_{12}^{(n)} & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & p_{s-1,s-1}^{(n)} & p^{(s-1,s)} \\ 0 & 0 & 0 & 0 & \dots & p_{ss-1}^{(n)} & p_{ss}^{(n)} \end{pmatrix}. \tag{1.5}$$

[1]This paper is in Russian, and thus has not been read directly by the author; the reference above was originally given by Schrijner [56].

Continuous Time Markov Chains

Darroch and Seneta [22] adapted many of the results from [21] to the continuous case. Specifically, the existence of a unique LCD that is equal to the QSD was once again proved. It is well known that the quasi-stationary distribution is intrinsically linked with the concept of a hazard rate, which is a measure of the risk of spontaneous absorption at time t, assuming absorption has not occurred by time t. Specifically, the QSD leads to a constant hazard rate, and any other initial distribution will generate a hazard rate that tends towards that constant value (see Aalen [3]). This will be explored in greater detail in Chapter 2.

The concept of phase type distributions (first passage time distributions generated by atomic initial distributions, that is those for which only one starting state is permitted) is described by Neuts [47] as having been initially based on ideas proposed by A.K. Erlang (no citation is given in the former paper). The theory was then expanded by Cox [16]. A phase type distribution is the distribution of the first-passage time between a given non-absorbing state and the absorbing state. In [1] these phase type distributions were considered in the context of survival analysis, in order to better understand the behaviour of the hazard rate. As an extension of this, the conjecture was put forward in [2] that by comparing an initial distribution to the QSD, predictions might be made regarding the hazard rate that initial distribution would generate. This comparison was vaguely described as some sort of comparative measure of "distance" from the absorbing state; initial distributions "closer" to the absorbing state than was the QSD would have a decreasing hazard rate, those "further" from the absorbing state than was the QSD would have increasing hazard rates, and those between the two extremes would display unimodal behaviour.

The notation used throughout our consideration of continuous-time Markov chains will now be given. A continuous-time Markov process $\mathcal{X} = \{X(t), t \geq 0\}$ upon the state space S, with $|S| < \infty$, will have the following properties. Denote the *transition probability functions* by

$$p_{ij}(t) = P(X(t) = j | X(0) = i) \ \forall i, j \in S. \tag{1.6}$$

The matrix of these functions will be denoted $P(t)$. The constant q_{ij} is a *transition*

intensity if

$$q_{ij} = \begin{cases} \lim_{h\to 0+} \frac{p_{ij}(h)}{h} \geq 0 & i \neq j \\ \lim_{h\to 0+} \frac{p_{ii}(h)-1}{h} \leq 0 & i = j \end{cases} . \tag{1.7}$$

We write the matrix of these constants as Q, and call Q the *generator* corresponding to the Markov process. The generator defines the process as follows:

$$P(t) = \exp(Qt). \tag{1.8}$$

This is the solution to the Kolmogorov equations (see Aalen *et al.* [3], and Darroch and Seneta [22]).

A generator $Q = (q_{ij})$ corresponding to a Markov chain with state space S is called *conservative* if $\sum_{j\in S} q_{ij} = 0 \; \forall i \in S$. Note that (1.7) gives

$$\sum_{j\in S} q_{ij} = \lim_{h\to 0+} \frac{\sum_{j\in S} p_{ij}(h) - 1}{h} = 0 \tag{1.9}$$

making the generator Q conservative. A state i in a Markov chain is *absorbing* if

$$p_{ii}(t) = 1 \tag{1.10}$$

holds for all t. Also, a state i in a Markov chain $\mathcal{X} = \{X(t), t \geq 0\}$ is *transient* (see e.g. [49]) if, assuming $X(0) = i$ and there exists $s = \inf_t\{X(t) \neq i\}$,

$$P(S_i < \infty) < 1 \tag{1.11}$$

where $S_i = \inf_t\{X(t+s) = i\}$ is the time between the process leaving i and the first return to i. In other words, i is a transient state if there exists a positive probability that once the process has left i it will never return there.

When considering a finite-state Markov chain the transience property is necessary in order to guarantee eventual absorption to occur with probability 1. Were a state $i \in C$ not transient, then we would expect to return to that state each time we left it, and therefore absorption could be postponed indefinitely.

The continuous Markov chain $\mathcal{X} = \{X(t), t \geq 0\}$ is *time homogeneous* if

$$P(X(t) = i|X(0) = j) = P(X(t+s) = i|X(s) = j) \tag{1.12}$$

holds for all states i and j and for all $t \geq 0$, $s \geq 0$. All continuous chains in this thesis will be assumed to be time-homogeneous.

Just as in the discrete case, there is a specific form of continuous Markov chain known as the birth-death processes. A continuous-time Markov process on the state space $S = \{-1, \ldots, s\}$ is called a birth-death process if $q_{ij} = 0$ for all $i, j \in S$ for which $|i - j| > 1$. Thus the process, when in state i, can move only to states $i - 1$ and $i + 1$. We refer to the value $q_{i,i+1}$ as the *birth rate* from state i, which we label λ_i, and similarly refer to $q_{i,i-1}$ as the *death rate* from state i, which will be denoted μ_i.

Each of the birth-death processes considered will be of the following type. The continuous Markov chain $\mathcal{X} = \{X(t), t \geq 0\}$ will have a state space $S = \{-1\} \cup \{0, \ldots, s\}$ as previously described. This results in a generator Q of the form shown below

$$Q = \begin{bmatrix} 0 & 0 & 0 & 0 & \ldots & 0 & 0 \\ \mu_0 & -(\lambda_0 + \mu_0) & \lambda_0 & 0 & \ldots & 0 & 0 \\ 0 & \mu_1 & -(\lambda_1 + \mu_1) & \lambda_1 & \ldots & 0 & 0 \\ 0 & 0 & \mu_2 & -(\lambda_2 + \mu_2) & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & \mu_s & -\mu_s \end{bmatrix}. \tag{1.13}$$

We also define the intensity matrix Q^* as follows

$$Q^* = \begin{bmatrix} -(\lambda_0 + \mu_0) & \lambda_0 & 0 & \ldots & 0 & 0 \\ \mu_1 & -(\lambda_1 + \mu_1) & \lambda_1 & \ldots & 0 & 0 \\ 0 & \mu_2 & -(\lambda_2 + \mu_2) & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & \mu_s & -\mu_s \end{bmatrix}. \tag{1.14}$$

The eigenvalues of this matrix will be of great importance in Chapter 2. More on continuous birth-death processes and their conditional limiting distributions can be found in work by Seneta [59], and van Doorn [24].

1.2.2 Imprecise Probability

Research on Markov chains where the one-step transition probability matrix is not completely known has been carried out from several perspectives. Schweitzer [58]

expressed imprecision in the form of perturbations in otherwise known transition probability matrices. Avrachenkov and Sanchez [6] introduced imprecision by means of fuzzy Markov chains.

A comprehensive approach to imprecision was written by Walley [66]. The approach to imprecise probability taken in this thesis is based on Weichselberger [68], which in turn is based upon [66]. Let S be a non-empty set and $\mathcal{A}$ the $\sum$-algebra of all subsets of S. The term *classical probability* is used to describe any set function $p\colon \mathcal{A} \to \mathbb{R}$ which satisfies Kolmogorov's axioms. An interval probability is now defined as follows. Let L and U be set functions on $\mathcal{A}$ such that $L \leq U$, $L(\emptyset) = U(\emptyset) = 0$, and $L(S) = U(S) = 1$. The interval-valued function $P(\cdot) = [L(\cdot), U(\cdot)]$ is called an *interval probability*. Each of these interval probabilities P generates a *structure*, which is the set $\mathcal{M}$ of classical probability measures on the measurable space $(S, \mathcal{A})$ that lie between L and U. For an interval probability with a non-empty structure, the quadruple $(S, \mathcal{A}, L(\cdot), U(\cdot))$ is described as an *R-field*. If the following properties holds for an R-field

$$L(A) = \inf_{p \in \mathcal{M}} p(A), \text{ and } U(A) = \sup_{p \in \mathcal{M}} p(A), \ \forall A \in \mathcal{A} \tag{1.15}$$

meaning that the lower bound L and upper bound U are strict, then it is described as an *F-field*. This is closely related to Walley's concept of coherence, and in fact the two coincide upon finite Markov chains. In brief, coherence requires that our bounds be as tight as possible, which is clearly the case with F-fields.

Kozine and Utkin [42] used the theory of interval-valued coherent previsions in order to generalise discrete-time, time-homogeneous irreducible Markov chains to interval-valued probabilities. A general procedure of interval-valued probability elicitation from heterogeneous and partial information is also analysed. Skulj [62, 63] obtained the relation between the sets of invariant distributions and limiting distributions for discrete-time, time-inhomogeneous irreducible Markov chains on a finite state space with interval probabilities based on theories of Weichselberger [68,69]. Recently de Cooman *et al.* [15] studied the time evolution of this same class of chains using upper and lower previsions (expectations).

Two notes are made here on terminology. First, once an attempt is made to extend our knowledge into the rich area of imprecision, care needs to be taken

regarding what the term time-homogeneity means. Without imprecision, a process is called time-homogeneous when the transition probabilities are constant, while in the imprecise situation the unknown transition probabilities are not necessarily constant. However, the bounds on the transition probabilities can be either time-dependent or constant and this latter case can also be considered as time-homogeneous within an imprecise framework. In this thesis, a Markov chain is called imprecise and time-homogeneous if the transition probabilities are assumed constant and known to exist within constant intervals, whereas an imprecise and time-inhomogeneous chain has transistion probabilities that might change from time step to time step, but always lie within constant intervals. Second, an attempt is made here to pre-empt any confusion as to the difference between probability intervals and interval probablility. The former implies a probability which has a single value, with that value only known to reside within a given interval. The latter case is the more general theory of mathematically expressing uncertainty across a range of situations.

1.3 Thesis Outline

In Chapter 2 continuous-time Markov chains are considered, in an attempt to investigate the claim made by Aalen and Gjessing [2] that the shape of the hazard rate generated by the initial distribution $\boldsymbol{\pi}(0)$ can be predicted by comparing $\boldsymbol{\pi}(0)$ with the QSD of the chain. It is demonstrated that the relationship postulated in [2] cannot occur without significant assumptions being made upon both the nature of the Markov chain and the set of possible initial distributions. Previously known results regarding the behaviour of hazard rates for birth-death processes are reviewed and expanded upon, and possible avenues for further study are discussed.

In Chapter 3 the long-term behaviour of discrete-time-homogeneous Markov chains with imprecision is considered. Thus it is assumed that there exists a single transition matrix, which describes the behaviour of the chain at every time step, but the elements of that matrix are known only to exist within a given closed set. The possible initial distributions over the non-absorbing states are also represented by a closed set. Thus all possible distributions over the state space can be found

by multiplying the set of initial distributions by the union of the set of transition matrices to the nth power. The results obtained in [42] are applied to the case of such chains with one absorbing state, as are those of [62] following the necessary adaptation, and it is proved that absorption remains certain. These results are then adapted to allow for conditioning on non-absorption, and it is shown that following these adaptations an invariant set of conditional distributions can be found that describes the long-term behaviour, conditioned on non-absorption, of the union of all chains described by the closed set of transition matrices. Moreover, it is proved that using this set as the set of possible initial distributions over the non-absorbing states ensures that the set of all possible distributions, conditioned on non-absorption, will be equal for all time steps, thus making this set the generalistion of the QSD in the precise time-homogeneous case. Approximations to this invariant set of conditional distributions are given in an example. The paper [17] contains a condensed form of the work done in this chapter.

The layout of Chapter 4 follows roughly that of Chapter 3. In Chapter 4 the assumption of time-homogeneity is removed. Thus not only is the transition matrix for each time step known only to exist within a given closed set, there is also no reason to believe the same element of that set will describe the Markov chain at the next time step. The results of [62] and [63] are applied, once again proving that absorption remains certain. We then adapt the method given in these papers to include conditioning upon non-absorption. It is then proved that there exists an invariant set of conditional distributions that describes the long-term behaviour of these chains, conditioned on non-absorption. Moreover, it is proved that using this set as the set of possible initial distributions over the non-absorbing states ensures that the set of all possible distributions, conditioned on non-absorption, will be independent of time. An example for the time-inhomogeneous case is then given. Finally, two further examples are given comparing the time-inhomogeneous case with the time-homogeneous case in Chapter 3.

Papers [18] and [19] combine to form a condensed version of this Chapter 4. The former concentrates on the theory which is presented in Chapter 4, whilst the latter compares the model presented in Chapter 4 with the model presented in Chapter 3.

Damjan Škulj is a named author on both these two papers. This reflects the fact that Section 4.2 and 4.6 were co-written with Škulj, and much of Chapter 4 is based upon various discussions and correspondence with him.

Chapter 2

Hazard Rates for Continuous Time Birth-Death Processes

2.1 Introduction

The research contained in this chapter investigates claims that have been made by Aalen [2] and Aalen and Gjessing [1]. The topic of those two papers is the study of *hazard rates*. The hazard rate is a function measuring the risk of absorption at time t for a Markov chain with one absorbing state, given that by t absorption has yet to occur. In this chapter, as in the papers cited above, time t is considered to be a continuous variable. The hazard rate exists in the discrete case also, but since Aalen considers only the continuous case, we shall do the same. We also consider only finite chains, so as to ensure each such chain has a unique *quasi-stationary distribution* (QSD). It is well-known (see e.g. [1]) that setting the initial distribution of the Markov process as equal to the QSD of that process will lead to a constant hazard rate. The thrust of [2] is the study of how such hazard rates change according to the initial distribution of the Markov process. For example, were one to look at graphs of the hazard rate $h(t)$, one might ask under what circumstances specific shapes occur. Monotonically increasing or decreasing, unimodal, and "bathtub" (decreasing then increasing) shapes are all possible (see [2]), along with more complex shapes.

In [1] and [2] it is suggested that, for a given Markov process with an absorbing state, there exists under certain conditions a method for comparing the chosen initial

distribution and the (family of) QSD(s) of the process that will provide information on the hazard rate's shape. To be specific, the following two conjectures were posed by Aalen and Gjessing [2]:

A1. The shape of the hazard rate is created by a balance between the attraction of the absorbing state, and general diffusion within the transient states;

A2. The hazard rate's shape is determined by how "close" the initial distribution is to the *absorbing state* compared to how close the quasi-stationary distribution is to that same state. Obviously this second statement only makes intuitive sense when distance from the absorbing state has some meaning (see below).

For the sake of brevity, these conjectures will be referred to from now on as "Aalen's conjectures", partially because he is the first named author of [2], and partially because similar comments were made in [1], for which Aalen is the only credited author. Conjectures A1 and A2 clearly lack specificity. Determining the relative "distance" of two initial distributions from the starting state requires two things. Firstly, since A2 only makes intuitive sense when distance from the absorbing state has some meaning, a method is needed by which two states can be compared in terms of their distance from the absorbing state. Second, a method is required by which two initial distributions over the state space can be compared.

In [1] and [2] this first problem is addressed by considering only chains for which the distance between states is intuitively obvious. One such chain is the birth-death process (see Section 1.2.1), for which transitions from state i to state j are impossible unless $|i - j| < 1$. This condition means that the minimum number of transitions from state $i \in \{0, 1, \ldots, s\}$ to the absorbing state, denoted -1, is $i + 1$. If this minimum number of transitions is considered as the distance to the absorbing state, comparing distances becomes simple.

The second problem is addressed only briefly in [1] and [2]. There are many different methods available by which two distributions can be compared (some of which will be discussed in Section 2.5), and determining which method is appropriate (if indeed any of them are) is a non-trivial task. Aalen and Gjessing assume that each initial distribution is atomic, meaning each distribution over the set of transient

states C takes the form $\boldsymbol{e}_i = (\delta_{0i}, \delta_{1i}, \ldots, \delta_{si})$, where δ_{ij} is the Kronecker delta. This initial distribution forces the process to start in state i, and leads to first-passage times that are *phase type* distributed. Comparing the "distance" between such distributions is not difficult, one can simply compare the distance between the two starting states. However, the QSD will not be an atomic distribution if $s > 0$, which makes comparing the QSD's distance to the absorbing state with that of $\boldsymbol{e}_i$ difficult.

In this chapter we too consider birth-death processes, in order to have a simple method for comparing the distance between states. In Section 2.2 the spectral representation of the probability function for such processes will be discussed. Section 2.3 will discuss the variable known as *time to absorption*, and will define the hazard rate in detail. Section 2.4 presents examples of the behaviour of the hazard rates of birth-death processes. Section 2.5 offers a method of comparing two initial distributions. Section 2.6 concerns itself with the nature of the hazard rate for birth-death processes with first-passage distributions of phase type (see Cox [16]). Finally, Section 2.7 contains conclusions, along with possible directions for future work.

2.2 The Spectral Representation

The continuous birth-death process was defined in Section 1.2.1. Recall that we assume a finite number of states $S = \{-1\} \cup \{-0, \ldots, s\}$, where -1 is an absorbing state and $C := \{0, 1, \ldots, s\}$ is an irreducible set of transient states. We now introduce notation which follows that used by Kijima [41]. The eigenvalues of the intensity matrix Q^* (see (1.14)) are denoted by the real values (see [41]) $-x_0, -x_1, \ldots, -x_s$, where $x_i > 0$ for all i and $x_0 \leq x_1 \leq x_2 \leq \ldots \leq x_s$. In Mandelson [46] and Kijima [41] it is shown that for a *regular* C, where regular denotes a set in which all states are transient and all states communicate, there is a unique eigenvalue $-x_0$ of Q^*, with maximal real part, which is real and less than zero. Therefore we have that $x_0 < x_1$.

In this section, previously known results will be utilised to show that each element i of a left (respectively right) eigenvector of Q^* is equal to a polynomial $L_i(x)$

(respectively $R_i(x)$), which will be defined later, evaluated at the absolute value of the corresponding eigenvalue. Before considering the polynomials themselves, however, it will be advantageous to consider various matrix properties which will be used in the rest of the chapter. This first definition can be found in e.g. Solomon *et al.* [64].

Definition 2.2.1 The square matrix $Q^* = [q_{ij}]_{ij}$ is said to be *weakly symmetric* if there exist strictly positive numbers $\{m_0, \ldots, m_s\}$ such that

$$m_i q_{ij} = m_j q_{ji}. \tag{2.1}$$

If this is the case, $\{m_0, \ldots, m_s\}$ is called a *symmetric measure*.

From [32] we have that the intensity matrix Q^* of a birth-death process is weakly symmetric with symmetric measure $\{\pi_0, \ldots, \pi_s\}$, where

$$\pi_0 = 1;\ \pi_i = \frac{\lambda_0 \lambda_1 \ldots \lambda_{i-1}}{\mu_1 \mu_2 \ldots \mu_i},\ i = 1, \ldots, s. \tag{2.2}$$

Denote by $\boldsymbol{R}(x) = (R_0(x), R_1(x), \ldots, R_s(x))'$ the vector such that $\boldsymbol{R}(x_i)$ is the right eigenvector of Q^* corresponding to eigenvalue $-x_i$. Similarly denote by $\boldsymbol{L}(x) = (L_0(x), L_1(x), \ldots, L_s(x))$ the polynomial vector such that $\boldsymbol{L}(x_i)$ is the left eigenvector of Q^* corresponding to eigenvalue $-x_i$. Since if $\boldsymbol{v}$ is a eigenvector of Q^* so too is $c\boldsymbol{v}$ for $c \in \mathbb{R}$, we can normalise both $\boldsymbol{R}(x)$ and $\boldsymbol{L}(x)$ so that $R_0(x) = L_0(x) = 1$.

Lemma 2.2.1 The polynomials $\{R_j(x)\}_{j=0}^s$ satisfy the following equations:

1. $-xR_0(x) = -(\lambda_0 + \mu_0)R_0(x) + \lambda_0 R_1(x)$
2. $-xR_i(x) = \mu_i R_{i-1}(x) - (\lambda_i + \mu_i)R_i(x) + \lambda_i R_{i+1}(x),\ i = 1, \ldots, s-1$
3. $-xR_s(x) = \mu_s R_{s-1}(x) - \mu_s R_s(x)$

Similarly the polynomials $\{L_j(x)\}_{j=0}^s$ satisfy the following equations:

1. $-xL_0(x) = -(\lambda_0 + \mu_0)L_0(x) + \mu_1 L_1(x)$
2. $-xL_i(x) = \lambda_{i-1} L_{i-1}(x) - (\lambda_i + \mu_i)L_i(x) + \mu_{i+1} L_{i+1}(x),\ i = 1, \ldots, s-1$
3. $-xL_s(x) = \lambda_{s-1} L_{s-1}(x) - \mu_s L_s(x)$

Proof See [32]. □

The relationship between the intensity matrix Q and the transition probabilities $p_{ij}(t)$ was given in (1.8). Combining that result with the vector $\boldsymbol{R}(x)$, transition probabilities can be expressed in terms of polynomials. We have from Abate [4] that

$$p_{ij}(t) = \pi_j \sum_{k=0}^{s} e^{-x_k t} R_i(x_k) R_j(x_k) c_k^2, \tag{2.3}$$

where

$$c_k = (\sum_{i=0}^{s} \pi_i R_i^2(x_k))^{-\frac{1}{2}}. \tag{2.4}$$

These equations allow us to differentiate and integrate the transition probabilities. This will prove useful in Section 2.3, in which we move between probability densities and survival functions.

2.3 Absorption Times and Hazard Rates

As stated in Section 2.1, our primary focus of consideration is the behaviour of phase type distributions, just as it was in [2]. In this section we consider the distribution of the time to absorption from a given state. This will then lead to an expression for the hazard rate for a birth-death process with atomic initial distributions.

2.3.1 Absorption Times

Let T denote the random variable of *time to absorption* for a given finite birth-death process $\mathcal{X}$ with state space $S = \{-1\} \cup C$ and for a given initial distribution. The *lifetime distribution function* of $T \geq 0$, denoted $G_{\boldsymbol{\pi}(0)}(t)$ is the probability that absorption has occurred by time t given initial distribution $\boldsymbol{\pi}(0)$. Thus the lifetime distribution function is the cumulative distribution function of T. Let

$$g_{\boldsymbol{\pi}(0)}(t) = \frac{\mathrm{d}}{\mathrm{d}t} G_{\boldsymbol{\pi}(0)}(t). \tag{2.5}$$

For a finite birth-death process which is irreducible over C, it is known that absorption is certain (see e.g. Darroch and Seneta [22]), hence

$$\int_0^\infty \mathrm{d}G_{\boldsymbol{\pi}(0)}(t) = \int_0^\infty g_{\boldsymbol{\pi}(0)}(t)\mathrm{d}t = 1. \tag{2.6}$$

Lemma 2.3.1 The lifetime distribution function $G_{\boldsymbol{e}_i}(t)$ of a finite birth-death process $\mathcal{X}$ with guaranteed starting state i can be written as

$$G_{\boldsymbol{e}_i}(t) = \mu_0 \int_0^t [\sum_{k=0}^{s} e^{-x_k\tau} R_i(x_k) c_k^2] \mathrm{d}\tau \tag{2.7}$$

and is equal to

$$G_{\boldsymbol{e}_i}(t) = \mu_0 \left(\sum_{k=0}^{s} \frac{1 - e^{-x_k t}}{x_k} R_i(x_k) c_k^2 \right) \tag{2.8}$$

where c_k is as defined in (2.4).

Proof $G_{\boldsymbol{e}_i}(t) = P(T \leq t | X(0) = i)$, so

$$G_{\boldsymbol{e}_i}(t) = 1 - P(X(t) \neq -1 | X(0) = i) = 1 - \sum_{k=0}^{s} p_{ik}(t). \tag{2.9}$$

Differentiating the right-hand side and using Kolmogorov's forward equations $P'(t) = P(t)Q^*$ (see e.g Kijima [41]) yields

$$g_{\boldsymbol{e}_i}(t) = -\frac{\mathrm{d}}{\mathrm{d}t} \sum_{j=0}^{s} p_{ij}(t) = -\sum_{j=0}^{s} \sum_{k=0}^{s} p_{ik}(t) q_{kj}. \tag{2.10}$$

Then, recalling from (1.14) that the intensity matrix Q^* of a continuous birth-death process takes the form

$$Q^* = \begin{bmatrix} -(\lambda_0 + \mu_0) & \lambda_0 & 0 & \dots & 0 & 0 \\ \mu_1 & -(\lambda_1 + \mu_1) & \lambda_1 & \dots & 0 & 0 \\ 0 & \mu_2 & -(\lambda_2 + \mu_2) & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \mu_s & -\mu_s \end{bmatrix} \tag{2.11}$$

we have that

$$\begin{aligned} \sum_{j=0}^{s} \sum_{k=0}^{s} p_{ik}(t) q_{kj} &= -(\lambda_0 + \mu_0) p_{i0}(t) + \mu_1 p_{i1}(t) \\ &+ \sum_{j=1}^{s-1} (\lambda_{j-1} p_{ij-1}(t) - (\lambda_j + \mu_j) p_{ij}(t) + \mu_{j+1} p_{ij+1}(t)) \\ &+ \lambda_{s-1} p_{is-1}(t) - \mu_s p_{is}(t) \\ &= -\mu_0 p_{i0}(t) \end{aligned} \tag{2.12}$$

and hence, using (2.3) along with (2.10),

$$g_{\boldsymbol{e}_i}(t) = -\frac{\mathrm{d}}{\mathrm{d}t}\sum_{j\in C} p_{ij}(t) = \mu_0 p_{i0}(t) = \mu_0 \sum_{k=0}^{s} e^{-x_k t} R_i(x_k) c_k^2 \tag{2.13}$$

where C is the set of transient states. Therefore, using (2.5), the probability that absorption occurs before time t, given that $\boldsymbol{\pi}(0) = \boldsymbol{e}_i$, can be written as

$$\begin{aligned} G_{\boldsymbol{e}_i}(t) &= \mu_0 \int_0^t p_{i0}(\tau)\mathrm{d}\tau \\ &= \mu_0 \int_0^t \sum_{k=0}^{s} e^{-x_k \tau} R_i(x_k) c_k^2 \mathrm{d}\tau \\ &= \mu_0 \left(\sum_{k=0}^{s} \frac{1 - e^{-x_k t}}{x_k} R_i(x_k) c_k^2 \right). \end{aligned} \tag{2.14}$$

□

Lemma 2.3.2

$$\sum_{k=0}^{s} \frac{1}{x_k} R_i(x_k) c_k^2 = \mu_0^{-1} \tag{2.15}$$

Proof Through the use of (2.13) and (2.6) with $\boldsymbol{\pi}(0) = \boldsymbol{e}_i$ we have that

$$\begin{aligned} 1 &= \mu_0 \int_0^\infty \sum_{k=0}^{s} e^{-x_k t} R_i(x_k) c_k^2 \mathrm{d}t \\ &= \mu_0 \sum_{k=0}^{s} R_i(x_k) c_k^2 \int_0^\infty e^{-x_k t} \mathrm{d}t \\ &= \mu_0 \sum_{k=0}^{s} R_i(x_k) c_k^2 \frac{1}{x_k} \end{aligned} \tag{2.16}$$

where the interchange of summation and integration is justified by the fact that $R_i(x_k)$ and C_k^2 are finite constants, and the area under each $e^{-x_k t}$ for $0 \le t \le \infty$ is finite since all x_k are positive. □

Combining Lemmas 2.3.1 and 2.3.2 yields the following corollary.

Corollary 2.3.1 The lifetime distribution of a finite birth-death process with $\boldsymbol{\pi}(0) = \boldsymbol{e}_i$ equals

$$G_{\boldsymbol{e}_i}(t) = 1 - \mu_0 \sum_{k=0}^{s} \frac{1}{x_k} e^{-x_k t} R_i(x_k) c_k^2. \tag{2.17}$$

2.3.2 The Hazard Rate

The hazard rate is best thought of as a measure of the risk of instantaneous absorption undergone by a process at time t, conditional on the fact that by time t the process has not yet been absorbed. Such a measure is of use in a variety of fields, such as medicine and finance (see Section 1.1).

Definition 2.3.1 The hazard rate for a continuous Markov process $\mathcal{X}$ with initial distribution $\boldsymbol{\pi}(0)$ is (see e.g Keilson [36])

$$h_{\boldsymbol{\pi}(0)}(t) = \frac{g_{\boldsymbol{\pi}(0)}(t)}{1 - G_{\boldsymbol{\pi}(0)}(t)}. \tag{2.18}$$

For a finite time-homogeneous birth-death process $\mathcal{X}$ this becomes

$$h_{\boldsymbol{\pi}(0)}(t) = \frac{\mu_0 \sum_{k=0}^{s} \pi_k(0) p_{k0}(t)}{\sum_{k=0}^{s} \sum_{j=0}^{s} \pi_k(0) p_{kj}(t)} = \frac{\mu_0 \sum_{k=0}^{s} \pi_k(0) p_{k0}(t)}{1 - \sum_{k=0}^{s} \pi_k(0) p_{k,-1}(t)}. \tag{2.19}$$

For atomic initial distributions, we have

$$h_{\boldsymbol{e}_i}(t) = \frac{\mu_0 p_{i0}(t)}{\sum_{j=0}^{s} p_{ij}(t)}. \tag{2.20}$$

The latter two definitions may be obvious, or can be derived directly from (2.5) and (2.10). As has been mentioned, the shape of a hazard rate is closely connected with the concept of a QSD. The QSD was briefly described in Section 1.1, we now provide the definition for the continuous case. This definition can be found in many places, see for example van Doorn [24].

Definition 2.3.2 For a finite-state birth-death process $\mathcal{X}$ with state space $S = \{-1\} \cup C$ where C is a set of transient states, an initial distribution $\boldsymbol{\pi}(0)$ over C is referred to as the quasi-stationary distribution if

$$\frac{\pi_j(t)}{1 - \pi_{-1}(t)} = q_j, \forall j \in C \tag{2.21}$$

where $\pi_j(t) = P_{\boldsymbol{\pi}(0)}\{X(t) = j\}$ and q_j is independent of time. The quasi-stationary distribution is thus the vector $\boldsymbol{q}$.

As t tends to infinity the hazard rate always tends to the value of the constant hazard rate generated by using the QSD as the initial distribution. There are several

ways to prove this fact, one such method is given by Aalen *et al.* [3]. In that book it is proved that for a birth-death process $\mathcal{X}$ on the state space $-\{1\} \cup C$, where C is a finite single communicating class with all states aperiodic, and for which state i has birth rate λ_i and death rate μ_i, the hazard rate $h(t)$ can be described as follows:

$$h(t) = \mu_0 P(X(t) = 0 | X(t) \geq 0). \tag{2.22}$$

Since we know convergence to the QSD is certain in the limit, we have

$$\lim_{t \to \infty} h(t) = \mu_0 d_0. \tag{2.23}$$

Moreover,

$$h_{\boldsymbol{d}}(t) = \mu_0 d_0, \ \forall t > 0 \tag{2.24}$$

where $\boldsymbol{d}$ is the quasi-stationary distribution.

In the next section, we begin to explore Aalen's two conjectures, as described in Section 2.1. The main question we consider is if it can be shown, in fact, that there exists some method of comparing $\boldsymbol{e}_i$ and the QSD for a given birth-death process that will allow us to predict the hazard rate $h_{\boldsymbol{e}_i}(t)$?

2.4 Examples

In this section two examples are presented, in order to demonstrate the difficulties in attempting to first rigorise and then prove (or disprove) Aalen's conjectures. The first example is taken from [2], and expanded upon here. It was with this example that Aalen and Gjessing demonstrated the behaviour that they believe can be generalised into Conjectures A1 and A2.

Example 2.4.1

Consider a continuous-time Markov chain on the state space $S = \{-1, \ldots, 4\} = \{-1\} \cup C$, where -1 is the absorbing state and $C = \{0, \ldots, 4\}$ is a set of transient states. Let $\lambda_i = 1.5$ for all $0 \leq i < 4$ and $\mu_i = 1$ for all $0 \leq i \leq 4$.

The absolute value of the dominant eigenvalue is 0.037 (hence $x_0 = 0.037$), with corresponding (normalised) left eigenvector $(0.037, 0.090, 0.167, 0.276, 0.430)$, which

is the quasi-stationary distribution. This tells us both that 0.037 is the constant hazard rate under quasi-stationarity (see (2.23)). It is also easy to calculate that the expected value of starting state for the quasi-stationary distribution is 2.972. We need to be careful with the idea of an expected value of the starting state, since of course our numbering of the states is essentially arbitrary. However, in the examples considered in this thesis, the consistency of our labelling of states combined with the meaningful concept of distance between states means that the expected value of the starting state is a value that can be considered safely.

Three initial distributions are considered in [2], $\boldsymbol{e}_0$, $\boldsymbol{e}_2$, and $\boldsymbol{e}_4$; these are called Cases 0, 2 and 4 respectively. Two further cases are given here, employing initial distributions $\boldsymbol{e}_1$ and $\boldsymbol{e}_3$; these are called Cases 1 and 3 respectively. Adding these cases allows for a more complete overview.

Figure 2.1 shows the hazard rates for all five cases. Note that in all but Case 0 the hazard rate is zero at $t = 0$, and that $h_{\boldsymbol{e}_0}(0) = 1$. This effect can be explained by the fact that $p_{j0}(0) = 0$ for all $j \neq 0$ and by the fact that absorption is only possible from state 0. Also note that in each case the hazard rate converges towards 0.037, which follows from (2.23).

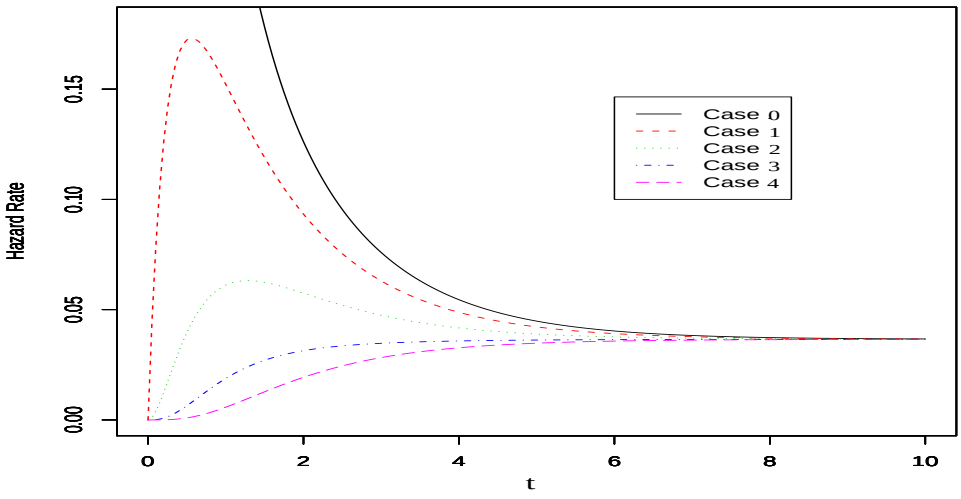


Figure 2.1: Hazard rates corresponding to Cases 0 - 4 in Example 2.4.1.

The suggested general trend given by Aalen's conjectures is demonstrated here. Case 0 produces a decreasing hazard rate, Cases 1 and 2 produce unimodal hazard rates, and Cases 3 and 4 produce increasing hazard rates. It is also interesting to note that $h_{\boldsymbol{e}_i}(t) \geq h_{\boldsymbol{e}_{i+1}}(t)$ for $i = 0, \ldots, 3$. The maximum value of the hazard rate in Case 2 is both smaller and occurs later than that of Case 1. Cases 3 and 4 have no maximum, but at any given point in time the hazard rate for Case 4 is further from the limit than is the hazard rate for Case 3.

The other important aspect to Example 2.4.1 is that the expected starting state for the quasi-stationary distribution is 2.972, and that Cases 1 and 2 have unimodal hazard rates, whereas Cases 3 and 4 have non-decreasing hazard rates. It appears that it is for this reason that Aalen's conjectures suggest it should be possible to compare an initial distribution $\boldsymbol{e}_i$ with the quasi-stationary distribution $\boldsymbol{q}$, in terms of their relative distances from absorption. It will be shown in Section 2.6 that it does not hold in general either that $i > E(X_q)$ implies an increasing hazard rate, or that $i < E(X_q)$ implies a unimodal hazard rate, where X_q is the discrete random variable with probability distribution $\boldsymbol{q}$ (thus $E(X_q)$ is the expected value of the starting state for the QSD).

Based on the above example, we present our own pair of conjectures:

B1. $h_{\boldsymbol{e}_{i+1}}(t_0) \leq h_{\boldsymbol{e}_i}(t_0)$ for all $t_0 \in [0, \infty)$ and for all $i = 0, \ldots, s-1$;

B2. $h_{\boldsymbol{e}_i}(t)$ has at most one turning point for all values of $i \in C$.

The first of these conjectures is equivalent to claiming there exists a hazard rate ordering for the hazard rates $h_{\boldsymbol{e}_i}(t)$, this is a term which will be discussed in Section 2.5. Were these conjectures proved then it would follow that there exists a value r^* for which $h_{\boldsymbol{e}_i}(t)$ is unimodal for $i \leq r^*$ and $h_{\boldsymbol{e}_i}(t)$ is strictly increasing for $i > r^*$. In other words there exists a set of states A for which starting in state $i \in A$ guarantees a unimodal distribution, and a set of states B for which starting in state $j \in B$ guarantees a non-decreasing function. Further, $\{0\} \cup A \cup B = C$, $A \cap B = \emptyset$ and $\max_{i \in A} i = \min_{j \in B} j - 1$. Let us now compare Conjectures B1 and B2 with Conjectures A1 and A2. A1 and A2 suggest that starting in a state far enough away from absorption will produce an increasing hazard rate, and starting

from a state closer to absorption will produce a unimodal hazard rate, and that the value that determines "far enough away" is related in some way to the QSD. Conjectures B1 and B2 also assume this value exists, and labels it r^*. However, we do not say anything about how r^* can be found. What we do claim, however, is that the reason such a value can be found is that $h_{\boldsymbol{e}_i}(t)$ has only one turning point, and that $h_{\boldsymbol{e}_{i+1}}(t_0) \leq h_{\boldsymbol{e}_i}(t_0)$, which combined with the fact that $h_{\boldsymbol{e}_i}(t)$ has the same limit for all values of i means that if $h_{\boldsymbol{e}_n}(t)$ is non-decreasing, then $h_{\boldsymbol{e}_{n+m}}(t)$ must be non-decreasing also. It will be shown in Section 2.6 that $h_{\boldsymbol{e}_s}(t)$ is in fact strictly increasing, so $r^* < s$, assuming Conjectures B1 and B2 are true.

In the following example we demonstrate that without the assumption that each initial distribution is atomic, the behaviour demonstrated in Example 2.4.1 may not occur.

Example 2.4.2

This example proves that, even whilst restricting attention to birth-death processes, the corresponding hazard rate does not have to be either increasing, decreasing, or unimodal for general initial distributions. It is highly likely that it is for this reason that Aalen and Gjessing restricted their attention to atomic initial distributions, a restriction which we also use in general.

Let $s = 5$, $\lambda_i = 0.6$ for $i = 0, 1, \ldots, 4$, and $\mu_i = 0.3$ for $i = 0, 1, \ldots, 5$, leading to the following transition intensity matrix for the transient states

$$Q^* = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.3 & -0.9 & 0.6 & 0 & 0 & 0 & 0 \\ 0 & 0.3 & -0.9 & 0.6 & 0 & 0 & 0 \\ 0 & 0 & 0.3 & -0.9 & 0.6 & 0 & 0 \\ 0 & 0 & 0 & 0.3 & -0.9 & 0.6 & 0 \\ 0 & 0 & 0 & 0 & 0.3 & -0.9 & 0.6 \\ 0 & 0 & 0 & 0 & 0 & 0.3 & -0.3 \end{pmatrix}. \tag{2.25}$$

The quasi-stationary distribution of this birth-death process is (0.0085, 0.0255, 0.0592, 0.1261, 0.2588, 0.5220). The initial distribution (0.086, 0.010, 0.010, 0.010,

0.010, 0.874) then produces the hazard rate illustrated in Figure 2.2. This hazard rate is not monotonically increasing or decreasing, or unimodal.

The above example makes it clear that we cannot expect to prove Aalen's conjectures without at least some restrictions on the possible initial distributions. This may well have been clear to Aalen at the time, and may have led to the restriction in [2] to only consider atomic initial distributions, though we have not come across any specific comments regarding this. Following Example 2.4.2, we will continue to consider Aalen's conjectures by exclusively using atomic initial distributions.

Example 2.4.1 demonstrated that $h_{\boldsymbol{e}_i}(t)$ is decreasing if and only if $i = 0$. We still require a method of comparison between the QSD and an initial distribution $\boldsymbol{e}_i$ with $i > 0$ that will allow us to determine whether the hazard rate $h_{\boldsymbol{e}_i}(t)$ is unimodal or increasing. Two very common methods for comparing probability distributions are stochastic domination of the first order, and of the second order. These methods will be defined in the next section, before being applied to our ongoing problem.

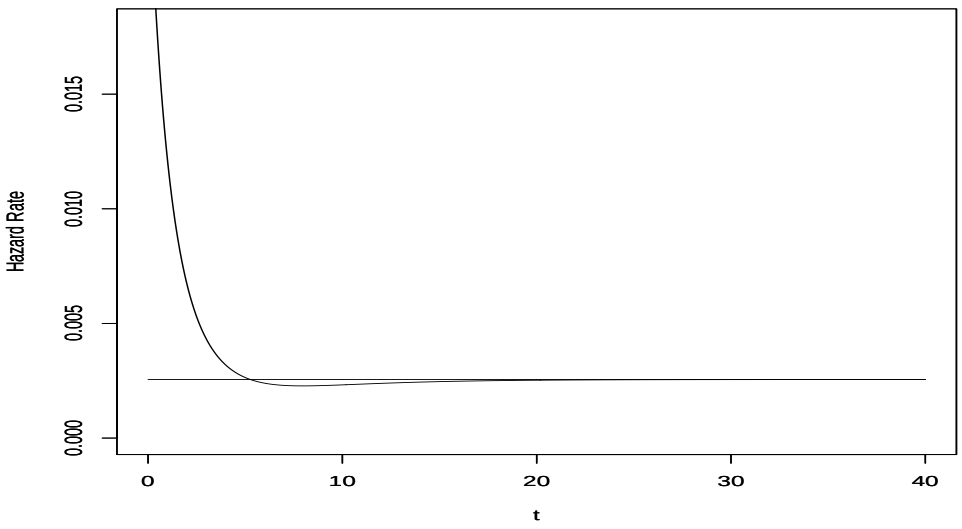


Figure 2.2: "Bath-tub" shaped hazard rate in Example 2.4.2.

2.5 Stochastic Orderings

Stochastic orderings can be thought of as comparisons between two probability distributions. In our case, we can consider stochastic orderings between two different initial distributions, or we can consider stochastic orderings between two continuous distributions which have been generated by two different initial distributions. In this thesis, we have concentrated exclusively on the former, but we will include in this section some information on the latter as well.

The hope is that there exists such methods that will allow us to determine the value r^*, described in the Section 2.4, through the comparison of the quasi-stationary distribution $\boldsymbol{q}$ with an atomic initial distribution $\boldsymbol{e}_i$. There are many methods available for comparing distributions. We will consider *first* and *second order stochastic dominance* (see [41] and [7], respectively). The reason for this choice comes from Aalen's conjectures. The claim made in those conjectures is that the shape of the hazard rate depends on the distance between the initial distribution and absorption, which in the case of an atomic initial distribution $\boldsymbol{e}_i$ means the distance $i+1$. This distance is then compared with the distance between the quasi-stationary distribution and absorption. The reason why this suggests the use of first order stochastic dominance is given below. First, however, we give the necessary definition.

Definition 2.5.1 For random variables $X = \{0, 1, \ldots, s\}$ and $Y = \{0, 1, \ldots, s\}$, with discrete distributions given by the vectors $\boldsymbol{a} = (a_0, \ldots, a_s)'$ and $\boldsymbol{b} = (b_0, \ldots, b_s)'$ respectively, X *is greater than* Y in the sense of *stochastic dominance of the first order* (also *stochastically greater*), denoted by $X \geq_{\text{st}} Y$ or $\boldsymbol{a} \geq_{\text{st}} \boldsymbol{b}$, if and only if

$$\sum_{i=0}^{n} a_i \leq \sum_{i=0}^{n} b_i, \ \forall \ n = 0, \ldots, s. \tag{2.26}$$

In words, if $\boldsymbol{a}$ and $\boldsymbol{b}$ are distributions over the set of transient states C, $\boldsymbol{a} \geq_{\text{st}} \boldsymbol{b}$ is true if and only if for every state i, the probability of being in state i or above is greater in distribution $\boldsymbol{a}$ than it is in $\boldsymbol{b}$. With -1 an absorbing state, it does not seem unreasonable to argue therefore that $\boldsymbol{a} \geq_{\text{st}} \boldsymbol{b}$ means that the initial distribution

$\boldsymbol{a}$ is further from absorption than $\boldsymbol{b}$ is. This makes it a logical choice of comparative method to apply to our problem, in which the distance from absorption is important.

In Example 2.1, $\boldsymbol{e}_4 \geq_{\text{st}} \boldsymbol{q} \geq_{\text{st}} \boldsymbol{e}_0$, where $\boldsymbol{q}$ is the QSD. Initial distributions $\boldsymbol{e}_1$, $\boldsymbol{e}_2$, and $\boldsymbol{e}_3$ cannot be ordered in such a way with relation to the quasi-stationary distribution. This also illustrates the obvious fact that for probability vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ it is *not* the case that either $\boldsymbol{a} \geq_{\text{st}} \boldsymbol{b}$ or $\boldsymbol{b} \geq_{\text{st}} \boldsymbol{a}$ must hold.

Alternatively, for continuous distributions the following definition (see e.g. [54]) is used instead.

Definition 2.5.2 For two continuous variables X and Y, $X \geq_{\text{st}} Y$ if

$$P(X > a) \geq P(Y > a) \tag{2.27}$$

for all a.

Thus the hazard rate itself could be compared. However, we will not make explicit use of this definition, save to discuss the concept of hazard rate ordering later in the section.

Definition 2.5.3 For the situation of Definition 2.5.1, X *is greater than* Y in the sense of *stochastic dominance of the second order*, denoted by $X \geq_{\text{2st}} Y$ or $\boldsymbol{a} \geq_{\text{2st}} \boldsymbol{b}$, if

$$\sum_{k=0}^{n}\sum_{j=0}^{k} a_j \leq \sum_{k=0}^{n}\sum_{j=0}^{k} b_j, \ \forall\ n = 0, \ldots, s. \tag{2.28}$$

Second order stochastic dominance is less easy to define in words, but can be thought of as a measure of risk. Note that first order stochastic dominance implies second order stochastic dominance, but that the reverse does not hold. Therefore we concentrate on applying second order stochastic dominance, since it follows that if second order stochastic dominance cannot provide sufficient conditions, then first order cannot either.

The following theorem is fairly simple, and therefore we were surprised to not find it anywhere in the literature. For the sake of completeness, we prove it here.

Theorem 2.5.1 $\boldsymbol{e}_i \geq_{\text{2st}} \boldsymbol{q} \Leftrightarrow \mathrm{E}(X_{e_i}) = i \geq \mathrm{E}(X_q)$, where $X_{\pi(0)}$ is a random variable with $\boldsymbol{\pi}(0)$ as its probability mass function.

Proof The following notation is introduced,

$$\sum_{k=0}^{n}\sum_{j=0}^{k}(e_i)_j = U_n^i \tag{2.29}$$

and

$$\sum_{k=0}^{n}\sum_{j=0}^{k} q_j = V_n. \tag{2.30}$$

We need to prove that

$$\sum_{k=0}^{n}\sum_{j=0}^{k}(e_i)_j \leq \sum_{k=0}^{n}\sum_{j=0}^{k} q_j \ \forall n = 0,\ldots,s \Leftrightarrow \mathrm{E}(X_{e_i}) \geq \mathrm{E}(X_q). \tag{2.31}$$

We first prove that (2.31) holds for the case where $n = s$.

$$\sum_{k=0}^{s}\sum_{j=0}^{k}(\boldsymbol{e}_i)_j \leq \sum_{k=0}^{s}\sum_{j=0}^{k} q_j \Leftrightarrow \mathrm{E}(X_{e_i}) \geq \mathrm{E}(X_q). \tag{2.32}$$

We have that

$$\begin{aligned}\sum_{k=0}^{s}\sum_{j=0}^{k}\pi_j(0) &= (s+1)\pi_0(0) + s\pi_1(0) + \ldots + 2\pi_{s-1}(0) + \pi_s(0) \\ &= s+1-\mathrm{E}(X_{\pi(0)}) \end{aligned} \tag{2.33}$$

Hence

$$\sum_{k=0}^{s}\sum_{j=0}^{k}(e_i)_j \leq \sum_{k=0}^{s}\sum_{j=0}^{k} q_j \Leftrightarrow s+1-\mathrm{E}(X_{e_i}) \leq s+1-\mathrm{E}(X_q) \Leftrightarrow \mathrm{E}(X_{e_i}) \geq \mathrm{E}(X_q) \tag{2.34}$$

as required. Next we prove that for vectors $\boldsymbol{e}_i$ and $\boldsymbol{q}$, if (2.28) holds for $n = s$, it holds for all $0 \leq n < s$. In other words, it must be proven that

$$\sum_{k=0}^{s}\sum_{j=0}^{k}(e_i)_j \leq \sum_{k=0}^{s}\sum_{j=0}^{k} q_j \Rightarrow \sum_{k=0}^{n}\sum_{j=0}^{k}(e_i)_j \leq \sum_{k=0}^{n}\sum_{j=0}^{k} q_j, \ \forall\ n = 0,\ldots,s. \tag{2.35}$$

The proof uses the special nature of $\boldsymbol{e}_i$ and induction. From (2.32) it follows that $E(X_{e_i}) \geq E(X_q) \Leftrightarrow U_s^i \leq V_s$. Assume now that $U_n^i \leq V_n$ holds for $n \geq j$; it remains to show that $U_{j-1}^i \leq V_{j-1}$. There are two possibilities, either $j \geq i$ or $j < i$. From (2.29) and (2.30)

$$U_n^i - U_{n-1}^i = \sum_{j=0}^{n}(e_i)_j = \begin{cases} 1 & \text{if } n \geq i \\ 0 & \text{if } n < i \end{cases} \tag{2.36}$$

and so $U_n^i - U_{n-1}^i = 0$ if $n < i$, that is if both U_n^i and U_{n-1}^i are equal to 0. In contrast, however,

$$V_n - V_{n-1} = \sum_{j=0}^{n} q_j \leq \sum_{j=0}^{s} q_j = 1. \tag{2.37}$$

Therefore

$$U_n^i - U_{n-1}^i \geq V_n - V_{n-1},\ \forall n = i, \ldots, s \tag{2.38}$$

and

$$U_n^i = 0 \leq V_n,\ \forall n = 0, \ldots, i-1. \tag{2.39}$$

It is demonstrated in (2.38) that U_n^i decreases by at least as much as V_n as n goes from j to $j-1$ as long as $j \geq i$. Hence the assumption that $U_j^i \leq V_j$ leads to $U_{j-1}^i \leq V_{j-1}$. By induction this gives us

$$U_n^i \leq V_n,\ \forall n = i, \ldots, s. \tag{2.40}$$

If $j < i$, proving $U_{j-1}^i \leq V_{j-1}$ is even easier, since (2.39) gives $U_{j-1}^i = 0$. Combining (2.39) and (2.40) therefore gives

$$U_n^i \leq V_n,\ \forall n = 0, \ldots, s. \tag{2.41}$$

Hence the condition $U_s^i \leq V_s \Rightarrow U_n^i \leq V_n$ holds for all $n = 0, \ldots, s$. □

An alternative to considering stochastic dominance would be to make use of hazard rate ordering. In [41] hazard rate ordering is defined as an ordering between two discrete distributions, hence this method can be used to compare initial distributions

Definition 2.5.4 *X is greater than Y* in the sense of *hazard rate order*, denoted by $X \geq_{\text{hr}} Y$ or $\boldsymbol{a} \geq_{\text{hr}} \boldsymbol{b}$, if

$$(\sum_{k\geq i}^{s} a_k)(\sum_{k\geq j}^{s} b_k) \geq (\sum_{k\geq j}^{s} a_k)(\sum_{k\geq i}^{s} b_k),\ \forall\ i > j. \tag{2.42}$$

It can be proven that $\boldsymbol{e}_0 \leq_{\text{hr}} \boldsymbol{q}$, $\boldsymbol{e}_s \geq_{\text{hr}} \boldsymbol{q}$, and that no other ordering is possible for atomic initial distributions.

Lemma 2.5.1 The equation

$$\boldsymbol{e}_r \geq_{\text{hr}} \boldsymbol{q} \tag{2.43}$$

can only hold when $r = s$. The same equation with the inequality reversed can only hold when $r = 0$.

Proof Assume $r = s$. This forces $\sum_{k\geq i}^{s} (\boldsymbol{e}_s)_k = 1$ for every $0 \leq i \leq s$. Thus (2.42) reduces to

$$\sum_{k\geq j}^{s} q_k \geq \sum_{k\geq i}^{s} q_k \tag{2.44}$$

which is obviously true since all elements of $\boldsymbol{q}$ are non-negative and $i > j$. Hence $\boldsymbol{e}_s \geq_{\text{hr}} \boldsymbol{q}$.

Now assume $r = 0$. This forces $\sum_{k\geq 0}^{s} (\boldsymbol{e}_s)_k = 1$, and $\sum_{k\geq j}^{s} (\boldsymbol{e}_s)_k = 0$ for all $j > 0$. Since $i > j$, i can never be zero, and so (2.42) becomes either

$$0 \leq \sum_{k\geq i}^{s} q_k \tag{2.45}$$

if $j = 0$, or has both sides equal to zero if $j > 0$. Thus $\boldsymbol{e}_0 \leq_{\text{hr}} \boldsymbol{q}$.

Finally, assume $0 < r < s$. If $i > r$, we have that $\sum_{k\geq i}^{s} (\boldsymbol{e}_r)_k = 0$, and if $i \leq r$, we have $\sum_{k\geq i}^{s} (\boldsymbol{e}_r)_k = 1$. Since $i > j$ there is at least one combination of i and j for which the left hand side and right hand side of (2.42) are re-written as 0 and $\sum_{k\geq i}^{s} q_k$ respectively. Clearly $0 \leq \sum_{k\geq i}^{s} q_k$. Thus it is impossible for $\boldsymbol{e}_r \geq_{\text{hr}} \boldsymbol{q}$ to hold for any $r < s$.

We now prove $\boldsymbol{e}_r \leq_{\text{hr}} \boldsymbol{q}$ is also impossible for $r > 0$. This comes from the fact that in the case where $j < i \leq r$, we have that $\sum_{k\geq i}^{s} (\boldsymbol{e}_r)_k = \sum_{k\geq j}^{s} (\boldsymbol{e}_r)_k = 1$. This reduces (2.42) to

$$\sum_{k\geq j}^{k} q_k \geq \sum_{k\geq i}^{k} q_k \tag{2.46}$$

for those values of i and j. This inequality rules out the possibility that $\boldsymbol{e}_r \leq_{\text{hr}} \boldsymbol{q}$, completing the proof. □

Therefore using hazard rate ordering to compare initial distributions is of little use to us, since there are only cases for which the ordering can be used, namely for initial distributions $\boldsymbol{e}_0$ and $\boldsymbol{e}_s$, and in both these cases the behaviour of the hazard rate is already known (see Section 2.6).

An alternate way to define hazard rate ordering is to compare the rates themselves directly. This is done in e.g. [54]. In this sense two hazard rates can be ordered only if one is greater than the other for all values of t , but it can be shown that (to immediately relate the result to our case),

$$\begin{aligned} & h_{e_i}(t_0) \geq h_{e_{i+1}}(t_0) \ \forall t_0 \in [0, \infty) \\ \Leftrightarrow\ & P(X_{e_i} > s + t | X_{e_i} > t) \leq P(X_{e_{i+1}} > s + t | X_{e_{i+1}} > t) \end{aligned} \tag{2.47}$$

where X_d is the random variable of time to absorption given initial distribution $\boldsymbol{d}$. Also equivalent to the above is that

$$(X_{e_i})_t \leq_{\text{st}} (X_{e_{i+1}})_t \tag{2.48}$$

where $\leq_{\text{st}}$ is as defined in (2.27), and $(X_d)_t$ is the time to absorption of a time given initial distribution $\boldsymbol{d}$, and conditioned on non-absorption by time t.

Considering this relation may be fruitful for future work. For now, however, we continue to compare initial distributions. What we want is for second order stochastic dominance to suffice as a measure of an initial distribution's "distance" from absorption, thus allowing us to prove Aalen's conjecture regarding predicting the shape of the hazard rate. We are therefore interested in whether either or both of the following statements hold:

C1. $\boldsymbol{a} \geq_{2\text{st}} \boldsymbol{q}$ or its converse is a sufficient condition for a specific shape of hazard rate;

C2. $\boldsymbol{a} \geq_{2\text{st}} \boldsymbol{q}$ or its converse is a necessary condition for a specific shape of hazard rate.

We will show in the next section that neither of these statements hold in general, and therefore that first or second order stochastic dominance cannot be used to find the value r^* which followed from Conjectures B1 and B2, which in turn means that these methods cannot be used to prove Conjectures A1 and A2.

2.6 Phase Type Distributions

As Aalen and Gjessing [2] suggested, and Section 2.4 demonstrated, restrictions need to be placed upon the possible initial distributions for a birth-death process in order for their conjectures to not be immediately disproved. In this section we consider the behaviour of hazard rates with initial distributions of the form $\boldsymbol{e}_i$ for $i = 0, \ldots, s$. In fact, the shape of the hazard rate $h_{\boldsymbol{e}_0}(t)$ and $h_{\boldsymbol{e}_s}(t)$ is known for all possible intensity matrices Q^* for a given birth-death process, and this will be demonstrated in this section. For initial distributions $\boldsymbol{e}_i$ for $i = 1, \ldots, s-1$, the shape of the hazard rate depends on the intensity matrix Q^*. In this section we attempt to predict this behaviour using second order stochastic dominance, as defined in Section 2.5, for the reasons given in that section. We will prove here that $\boldsymbol{e}_i \leq_{\text{2st}} \boldsymbol{q}$ guarantees a decreasing hazard rate, but that $\boldsymbol{e}_i \geq_{\text{2st}} \boldsymbol{q}$ in general implies nothing about the shape of the hazard rate unless $i = s$. These results in combination will demonstrate that in general second order stochastic dominance does not allow us to predict the shape of the hazard rate when dealing with phase type distributions in the manner suggested by Aalen's conjectures.

Lemma 2.6.1

$$\boldsymbol{e}_0 \leq_{\text{2st}} \boldsymbol{q} \leq_{\text{2st}} \boldsymbol{e}_s. \tag{2.49}$$

Further, $\boldsymbol{e}_i \leq_{\text{2st}} \boldsymbol{q}$ is impossible for $i > 0$.

Proof We have from (2.28) that $\boldsymbol{e}_0 \leq_{\text{2st}} \boldsymbol{q}$ requires that

$$\sum_{k=0}^{n} 1 \geq \sum_{k=0}^{n} \sum_{j=0}^{k} q_j, \ \forall\ 0 \leq n \leq s \tag{2.50}$$

where the left hand side follows from the nature of $\boldsymbol{e}_0$. Inequality (2.50) holds if $\sum_{j=0}^{k} q_j \leq 1$ for all k, which must be the case since $\boldsymbol{q}$ is a probability distribution.

Similarly, $\boldsymbol{q} \leq_{\text{2st}} \boldsymbol{e}_s$ requires the following inequalities to hold

$$\sum_{k=0}^{n} \sum_{j=0}^{k} q_j \geq 0, \ \forall\ n = 0, \ldots, s-1 \tag{2.51}$$

and

$$\sum_{k=0}^{s} \sum_{j=0}^{k} q_j \geq 1 \tag{2.52}$$

where the right hand sides of (2.51) and (2.52) follow from the nature of $\boldsymbol{e}_s$. Clearly the first inequality is true. The second inequality also holds once it is realised that $\sum_{k=0}^{s}\sum_{j=0}^{k} q_j \geq \sum_{j=0}^{s} q_j = 1$.

It is now proved that $\boldsymbol{e}_i \leq_{\text{2st}} \boldsymbol{q}$ is only possible if $i = 0$. By the definition of $\boldsymbol{e}_i$

$$0 = \sum_{k=0}^{n}\sum_{j=0}^{k}(e_i)_j < \sum_{k=0}^{n}\sum_{j=0}^{k} q_j, \ \forall n < i. \tag{2.53}$$

Since it is already known that $q_0 > 0$, the above inequality is a contradiction of the conditions necessary for $\boldsymbol{e}_i \leq_{\text{2st}} \boldsymbol{q}$ if $i > 0$. □

In Conjectures C1 and C2 we suggested that $\boldsymbol{e}_i \leq_{\text{2st}} \boldsymbol{q}$ and $\boldsymbol{q} \leq_{\text{2st}} \boldsymbol{e}_i$ might be sufficient and/or necessary conditions for predicting the shape of the hazard rate. From Lemma 2.6.1 we now have that $\boldsymbol{e}_i \leq_{\text{2st}} \boldsymbol{q}$ can only hold if $i = 0$. In the following subsection we prove that, indeed, the hazard rate $h_{\boldsymbol{e}_0}(t)$ can be predicted.

It is worth noting that the results given in Subsections 2.6.1 and 2.6.2 seem to be well-accepted in the relevant literature. Despite this, however, we are aware of no specific proof, and thus include our own here for the sake of completeness.

2.6.1 Starting State 0

In this subsection it is proved that the initial distribution $\boldsymbol{e}_0$ will lead to a non-increasing hazard rate. This result follows immediately from Theorems 5.4 B and C and 5.8 B in Keilson [36], and is thus presented as a corollary. The proof of this corollary requires the definition of a *completely monotone* function, taken from Kijima [41].

Definition 2.6.1 An infinitely-differentiable function $g(t)$ is called *completely monotone* if $(-1)^n \frac{\mathrm{d}^n}{\mathrm{d}t^n} g(t) \geq 0$ for all t, and all n.

Definition 2.6.2 A twice-differentiable function g(t) is called *convex* if $\frac{\mathrm{d}^2}{\mathrm{d}t^2} g(t) \geq 0$, for all t. A function g(t) is called *log-convex* if $\log(g(t))$ is a convex function.

Corollary 2.6.1 For a finite birth-death process with absorbing state -1, the hazard rate corresponding to initial distribution $\boldsymbol{e}_0$, $h_{\boldsymbol{e}_0}(t)$ is a non-increasing function and bounded from below by x_0, where $-x_0$ is the dominating eigenvalue of the intensity matrix Q^*.

Proof From (2.13)

$$g_{\boldsymbol{e}_0}(t) = -\frac{\mathrm{d}}{\mathrm{d}t}\sum_{j\in C} p_{0j}(t) = \mu_0 p_{00}(t). \tag{2.54}$$

Using (2.9), (2.10), and (2.20)

$$h_{\boldsymbol{e}_0}(t) = \frac{\mu_0 p_{00}(t)}{\sum_{j\in C} p_{0j}(t)} = \frac{g_{\boldsymbol{e}_0}(t)}{\overline{G}_{\boldsymbol{e}_0}(t)} = -\frac{\mathrm{d}}{\mathrm{d}t}\log(\overline{G}_{\boldsymbol{e}_0}(t)). \tag{2.55}$$

From Theorem 5.2 in [41] it is known that for a birth-death process each of the transition probability functions $p_{ii}(t)$ are completely monotone, and consequently, using (2.54), $g_{\boldsymbol{e}_0}(t)$ is completely monotone. From Theorems 5.4 B and C and Theorem 5.8 in [36] any completely monotone density function is also log-convex and that if a density function $g_{\boldsymbol{e}_0}(t)$ is log-convex, then $\overline{G}_{\boldsymbol{e}_0}(t)$ is log-convex also. This means that $\frac{\mathrm{d}}{\mathrm{d}t}\log(\overline{G}_{\boldsymbol{e}_0}(t))$ is a non-decreasing function, and hence that $h_{\boldsymbol{e}_0}(t)$ is non-increasing, as required. □

We have now proved that $\boldsymbol{q} \geq_{2\mathrm{st}} \boldsymbol{e}_i$ is a sufficient condition for the shape of the hazard rate to be non-increasing. It also must be a necessary condition, since $h_{\boldsymbol{e}_i}(0) = 0$ for $i = 1, \ldots, s$ as was discussed in Section 2.4.

2.6.2 Starting State s

In this subsection it is proved that for a finite birth-death process with absorbing state, the hazard rate is a non-decreasing function when starting in the state furthest from absorption, and further that it is bounded from above by the constant hazard rate obtained when the quasi-stationary distribution is taken as initial distribution of the process. From Theorem 5.5 in Kijima [41] we have the following result. Suppose that $\lambda_s = 0$, $\mu_0 > 0$ and $\boldsymbol{\pi}(0) = \boldsymbol{e}_s$, then the time until absorption T is the sum of $s+1$ independent and exponentially distributed random variables with distinct parameters.

The proof of the theorem below also requires the following definition of Polya-frequency densities of infinite order (PF_∞), obtained from [36].

Definition 2.6.3 A probability density function $g(x)$ is a *Polya-frequency density of infinite order* ($g \in \mathrm{PF}_\infty$) if, possibly after translation, $g(x)$ is (the limit of a sequence of densities, each of which is) a convolution of a finite number of exponential densities.

Theorem 5.3 in Karlin [34] proves that if the probability density functions $f(x)$ and $g(x)$ lead to increasing hazard rates, then the convolution $f*g(x)$ is a probability density function (pdf) that also leads to an increasing hazard function. Therefore if an exponential distribution leads to an increasing hazard rate, then a convolution of such distributions must also lead to an increasing hazard rate. Since an exponential distribution does lead to an increasing hazard rate, we have that a hazard rate is non-decreasing if the associated pdf is a Polya-frequency density of infinite order. Proving that the pdf $f(x)$ of the first passage time from s to -1 is such that $f(x) \in \mathrm{PF}_\infty$ will therefore prove that the hazard rate $h_{\boldsymbol{e}_s}(t)$ is non-decreasing.

It can now be proved that for a finite birth-death process with absorbing state, the hazard rate, when starting in state s, is a non-decreasing function and bounded from above by the constant hazard rate obtained when the quasi-stationary distribution is taken as initial distribution of the process. It has already been shown that

$$h_{\boldsymbol{e}_s}(t) = \frac{\mu_0 p_{s0}(t)}{1 - p_{s,-1}(t)} = \frac{\mu_0 p_{s0}(t)}{\sum_{j \in \mathcal{C}} p_{sj}(t)}. \tag{2.56}$$

Also we know that the constant hazard rate obtained when the quasi-stationary distribution is taken as initial distribution of the process equals x_0, where $-x_0$ is the dominating eigenvalue of the intensity matrix Q^*.

Theorem 2.6.1 For a finite birth-death process $\mathcal{X}$ on the state space $\mathcal{S} = \{-1\} \cup C$, the hazard rate, when starting in the final state s, is a non-decreasing function and is bounded from above by x_0.

Proof From (2.18) with $\boldsymbol{\pi}(0) = \boldsymbol{e}_s$ it follows that

$$h_{\boldsymbol{e}_s}(t) = \frac{g_{\boldsymbol{e}_s}(t)}{\bar{G}_{\boldsymbol{e}_s}(t)} = -\frac{d}{dt}\log(\bar{G}_{\boldsymbol{e}_s}(t)). \tag{2.57}$$

From Theorem 5.5 in [41] it follows that, when $\boldsymbol{\pi}(0) = \boldsymbol{e}_s$, the time to absorption T is the sum of $s+1$ independent and exponential distributed random variables and consequently the pdf of T is a Polya-frequency density of infinite order. □

2.6.3 Starting States Between 0 and s

The shape of the hazard rate when the process has guaranteed starting state r, for $0 < r < s$, is now considered. It is first demonstrated that $\boldsymbol{e}_i \geq_{\text{2st}} \boldsymbol{q}$ is neither a necessary nor sufficient condition for the hazard rate to be increasing (since $\boldsymbol{e}_s \geq_{\text{2st}} \boldsymbol{q}$ and $h_{\boldsymbol{e}_s}(t)$ is non-decreasing it immediately follows that $\boldsymbol{e}_i \geq_{\text{2st}} \boldsymbol{q}$ cannot be a necessary or sufficient condition for the hazard rate to be unimodal). This is shown by the following two examples.

Example 2.6.1

Let $s = 5$, and define Q^* as follows

$$Q^* = \begin{pmatrix} -0.9 & 0.3 & 0 & 0 & 0 & 0 \\ 0.6 & -0.9 & 0.3 & 0 & 0 & 0 \\ 0 & 0.6 & -0.9 & 0.6 & 0 & 0 \\ 0 & 0 & 0.6 & -0.9 & 0.3 & 0 \\ 0 & 0 & 0 & 0.6 & -0.9 & 0.3 \\ 0 & 0 & 0 & 0 & 0.6 & -0.6 \end{pmatrix}. \tag{2.58}$$

This chain has the quasi-stationary distribution (0.176, 0.233, 0.220, 0.175, 0.122, 0.074). Further, $E(X_q) = 2.0581$, where X_q is the expected starting state for the quasi-stationary distribution. If taking $\boldsymbol{\pi}_0 = \boldsymbol{e}_3$ leads to a non-decreasing hazard rate, it would support Conjectures C1. As can be seen in Figure 2.3, however, that is not the case. Hence $\boldsymbol{e}_r \geq_{\text{2st}} \boldsymbol{q}$ is not a sufficient condition for a monotonically non-decreasing hazard rate.

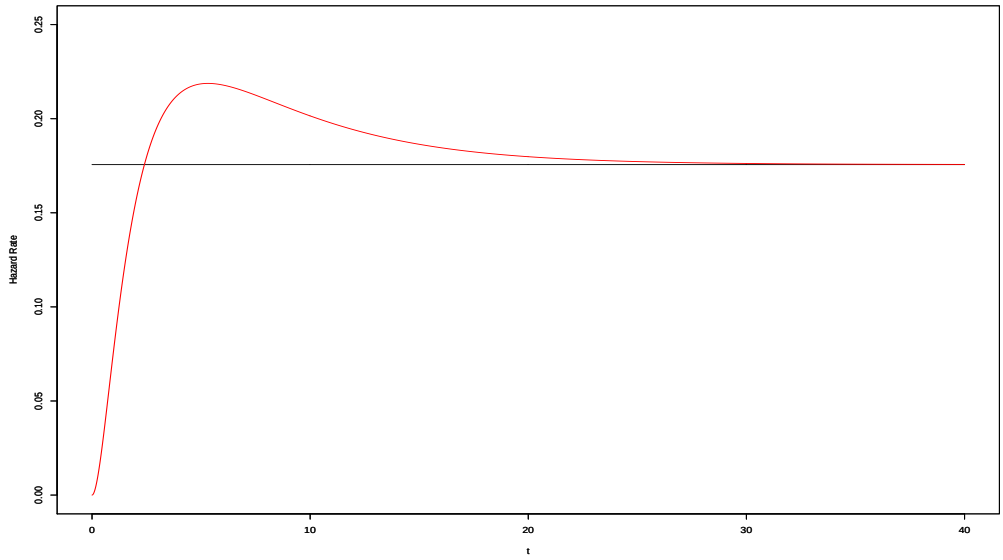


Figure 2.3: Hazard rate from state 3 in Example 2.6.1.

Example 2.6.2

Keep $s = 5$, and define Q^* as follows

$$Q^* = \begin{pmatrix} -0.9 & 0.6 & 0 & 0 & 0 & 0 \\ 0.3 & -0.9 & 0.6 & 0 & 0 & 0 \\ 0 & 0.3 & -0.9 & 0.6 & 0 & 0 \\ 0 & 0 & 0.3 & -0.9 & 0.6 & 0 \\ 0 & 0 & 0 & 0.3 & -0.9 & 0.6 \\ 0 & 0 & 0 & 0 & 0.3 & -0.3 \end{pmatrix}. \tag{2.59}$$

The quasi-stationary distribution for this process is (0.009, 0.026, 0.059, 0.126, 0.259, 0.522). Note that $E(X_q) = 4.1671$, and hence $\boldsymbol{q} \leq_{\text{2st}} \boldsymbol{e}_4$ does not hold. Were $h_{e_4}(t)$ to be unimodal, this would support Conjecture C2. As can be seen in Figure 2.4, however, the hazard rate is in fact non-decreasing. Thus $\boldsymbol{e}_r \geq_{\text{2st}} \boldsymbol{q}$ is neither a sufficient nor a necessary condition for a non-decreasing hazard rate.

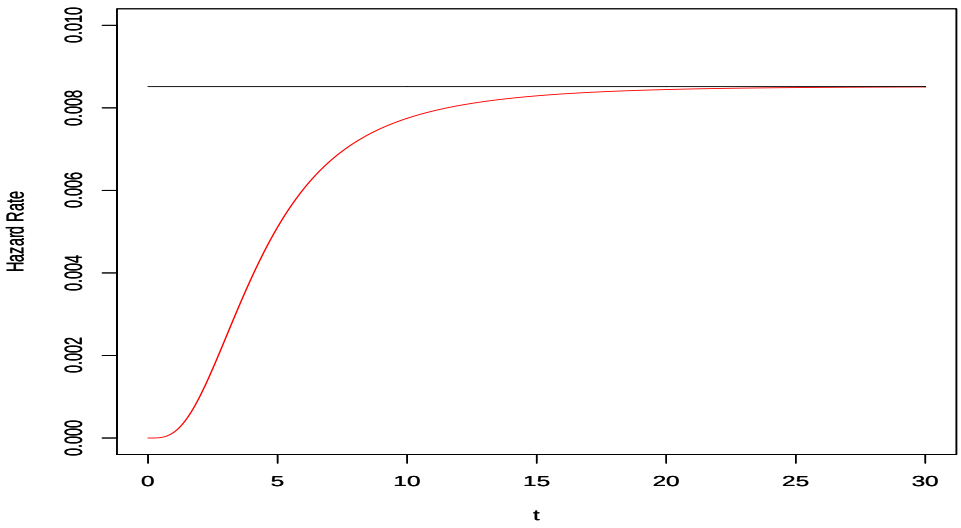


Figure 2.4: Hazard rate from state 5 in Example 2.6.2.

Although the two methods considered in Sections 2.5 and 2.6 are the most obvious ones, there are of course other possible stochastic orderings which might be employed here, in a continued effort to prove Aalen's conjectures. Rather than attempt to apply each one in turn in the hope of progress, however, other methods and ideas are now briefly discussed which might lead to advances in the future.

2.7 Alternative Approaches and Concluding Remarks

Although the comments made in [2] are somewhat vague, we have demonstrated in this chapter that if we attempt to assume logical reading of the evidence in that paper, the suggestion that the expecting value of the starting state for the QSD determines the value r^* does *not* hold. However, it is worth noting that during the research of which the results are summarised in this chapter many examples were calculated, and in every case r^* was found to exist and to often be close to the expected value of the starting state for the QSD. It is therefore our hope that a relation between the two can be found to exist, either directly, or through some third variable or property that affects the values of both r^* and the QSD. Potential

approaches which we might choose to take in order to progress from this point are given below, lack of time meant these methods could not be explored whilst preparing this thesis.

2.7.1 Alternative Approaches

Further progress may be made by considering under what conditions the hazard rate is non-decreasing. Recall that we claimed that a value $r^* \in \{1, 2, \ldots, s\}$ exists such that $h_{\boldsymbol{e}_r}$ is non-decreasing for $r^* \leq r \leq s$ and unimodal for $0 < r < r^*$. We focus now on what conditions are necessary for the hazard rate to be non-decreasing. Naturally one necessary and sufficient condition is that its first derivative is non-negative. We have that

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}h_{\boldsymbol{e}_r}(t) = &\frac{(-\mu_0 \sum_{k=0}^{s} x_k e^{-x_k t} R_r(x_k) c_k^2)(-\mu_0 \sum_{k=0}^{s} x_k^{-1} e^{-x_k t} R_r(x_k) c_k^2)}{(\mu_0 \sum_{k=0}^{s} x_k^{-1} e^{-x_k t} R_r(x_k) c_k^2)^2} \\ &- \frac{(\mu_0 \sum_{k=0}^{s} e^{-x_k t} R_r(x_k) c_k^2)(-\mu_0 \sum_{i=0}^{s} e^{-x_i t} R_r(x_k) c_k^2)}{(\mu_0 \sum_{k=0}^{s} x_k^{-1} e^{-x_k t} R_r(x_k) c_k^2)^2}\end{aligned} \tag{2.60}$$

from (2.20), (2.3) and (2.17). Thus in order for the hazard rate to be non-decreasing it is required that

$$\left(\mu_0 \sum_{k=0}^{s} e^{-x_k t} R_r(x_k) c_k^2\right)^2 - \left(\mu_0 \sum_{k=0}^{s} x_k e^{-x_k t} R_r(x_k) c_k^2\right)\left(\mu_0 \sum_{k=0}^{s} x_k^{-1} e^{-x_k t} R_r(x_k) c_k^2\right) \geq 0 \tag{2.61}$$

for all values of $t \geq 0$. By use of (2.3) and (2.17) it can be seen that (2.61) is equivalent to

$$\begin{aligned}&(\mu_0 p_{r0}(t))^2 + \mu_0 \left(\frac{\mathrm{d}}{\mathrm{d}t} p_{r0}(t)\right) \overline{G}_{\boldsymbol{e}_r}(t) \;\; \geq \;\; 0 \\ \Leftrightarrow\; &\mu_0 (p_{r0}(t))^2 + \left(\frac{\mathrm{d}}{\mathrm{d}t} p_{r0}(t)\right) \overline{G}_{\boldsymbol{e}_r}(t) \;\; \geq \;\; 0, \;\; \forall t \geq 0.\end{aligned} \tag{2.62}$$

Therefore $\mu_0 p_{r0}(t) = g_{\boldsymbol{e}_r}(t)$ and hence (2.61) is equivalent to

$$\begin{aligned}p_{r0}(t) g_{\boldsymbol{e}_r}(t) + \left(\frac{\mathrm{d}}{\mathrm{d}t} p_{r0}(t)\right) \overline{G}_{\boldsymbol{e}_r}(t) \geq 0, \;\; \forall t \geq \;\; &0 \\ \Leftrightarrow \frac{\left(\frac{\mathrm{d}}{\mathrm{d}t} p_{r0}(t)\right)}{p_{r0}(t)} \geq -\frac{g_{\boldsymbol{e}_r}(t)}{\overline{G}_{\boldsymbol{e}_r}(t)}, \;\; \forall t \geq \;\; &0 \\ \Leftrightarrow \frac{\mathrm{d}}{\mathrm{d}t} \log p_{r0}(t) \geq \frac{\mathrm{d}}{\mathrm{d}t} \log \overline{G}_{\boldsymbol{e}_r}(t), \;\; \forall t \geq \;\; &0 \quad .\end{aligned} \tag{2.63}$$

If we now define $y := \overline{G}_{\boldsymbol{e}_r}(t) = 1 - \mu_0 \int_0^t p_{r0}(\tau)\mathrm{d}\tau$ then clearly $y' = -g_{\boldsymbol{e}_r}(t)$ and $y'' = -\mu_0 \frac{\mathrm{d}}{\mathrm{d}t} p_{r0}(t)$. Thus finally the inequality (2.61) is equivalent to

$$\frac{y''}{y'} \geq \frac{y'}{y} \Leftrightarrow y''y \geq (y')^2. \tag{2.64}$$

If it is possible to find conditions under which this differential inequality holds true, then we would also have conditions for which the hazard rate is non-decreasing, but we lacked the necessary time to investigate further.

An alternative might be to use of the approach taken by Glaser [28], where sufficient conditions are given for non-increasing, non-decreasing, unimodal, and bathtub shaped hazard rates. To list these conditions, the following definitions are required.

Definition 2.7.1 The function $l(t)$ is the reciprocal of the hazard rate, that is

$$l(t) := \frac{1}{h(t)} = \frac{\overline{G}(t)}{g(t)}. \tag{2.65}$$

Definition 2.7.2

$$\eta(t) = -\frac{g'(t)}{g(t)} \tag{2.66}$$

The following results are given in [28]:

1. If $\eta'(t) > 0$ for all $t > 0$, then $h(t)$ is non-decreasing.

2. If $\eta'(t) < 0$ for all $t > 0$, then $h(t)$ is non-increasing.

3. If there exists t_0 such that $\eta'(t) < 0$ for all $t \in (0, t_0)$, $\eta'(t_0) = 0$, and $\eta'(t) > 0$ for all $t \in (t_0, \infty)$, and if $l'(t)$ has at least one zero, then $h(t)$ is bathtub shaped.

4. If there exists t_0 such that $\eta'(t) < 0$ for all $t \in (0, t_0)$, $\eta'(t_0) = 0$, and $\eta'(t) > 0$ for all $t \in (t_0, \infty)$, and if $l'(t)$ has no zeros, then $h(t)$ is non-decreasing.

5. If there exists t_0 such that $\eta'(t) > 0$ for all $t \in (0, t_0)$, $\eta'(t_0) = 0$, and $\eta'(t) < 0$ for all $t \in (t_0, \infty)$, and if $l'(t)$ has at least one zero, then $h(t)$ is unimodal.

6. If there exists t_0 such that $\eta'(t) > 0$ for all $t \in (0, t_0)$, $\eta'(t_0) = 0$, and $\eta'(t) < 0$ for all $t \in (t_0, \infty)$, and if $l'(t)$ has no zeros, then $h(t)$ is non-increasing.

It is a trivial task to apply these results to the special case considered in this chapter. Specifically, we have that a given hazard rate is increasing if either $\eta'(t)$ is everywhere positive, or that $\eta'(t)$ has one zero, $\eta'(t + c) > 0$ for any positive constant c, and $l'(t)$ has no zeros. Similarly, it can be proved that a given hazard rate is unimodal by showing that $\eta'(t)$ has one zero, $\eta'(t + c) < 0$ for any positive constant c, and $l'(t)$ has at least one zero.

Note that through the use of the chain rule

$$\eta'(t) = \frac{(g'(t))^2 - g(t)g''(t)}{(g(t))^2} \tag{2.67}$$

and hence considering the sign of $\eta'(t)$ is equivalent to considering at what values of t the inequality $(g'(t))^2 > g(t)g''(t)$ holds. Note further at this point that it is proven in [39] that $g(t)$ is unimodal.

The justification given in [28] for the use of this method is that $\eta(t)$ can often be easier to find than $h(t)$. Whether or not this is true for our current situation, we have not yet had the time to consider.

Finally, as mentioned in Section 2.5, it may be possible to consider either stochastic dominance or hazard rate ordering for the hazard rates themselves, rather than their initial distributions. Applying such comparisons, however, would be a non-trivial task, it is also not immediately clear as to whether such considerations would lead us to a method by which r^* can be calculated, which is the goal of our research in this area.

Whilst writing this thesis attempts were made to find a method by which r^* could be found in the case where all birth rates for states 0 to $s-1$ were equal to λ, and all death rates for states 0 to s were equal to μ. For this special case, the ratio $\frac{\lambda}{\mu}$ was considered. It was hoped that this ratio would at best allow the calculation of r^*, or at least that altering the ratio would affect the value r^* in a way that could be predicted. Whilst the former goal was not achieved, we acheived some possible partial success with the latter. We summarise the results of this attempt below; perhaps further study will bring more insight.

λ	μ	$\frac{\lambda}{\mu}$	$\lceil E(X_q) \rceil$	r^*
0.8	1.2	0.6667	7	11
0.9	1.1	0.8181	9	12
1	1	1	13	13
1.1	0.9	1.2222	16	15
1.2	0.8	1.5	18	16

Table 2.1: $\frac{\lambda}{\mu}$ ratio for $s = 19$.

λ	μ	$\frac{\lambda}{\mu}$	$\lceil E(X_q) \rceil$	r^*
0.3	0.7	0.4286	4	20
0.4	0.6	0.6667	6	21
0.5	0.5	1	25	26
0.6	0.4	1.5	37	35
0.7	0.3	2.3333	39	17

Table 2.2: $\frac{\lambda}{\mu}$ ratio for $s = 39$.

We began with the case where $s = 19$. Five situations were considered, $\lambda = 1.2$ and $\mu = 0.8$, $\lambda = 0.8$ and $\mu = 1.2$, $\lambda = 1.1$ and $\mu = 0.9$, $\lambda = 1.1$ and $\mu = 0.9$, and finally $\lambda = \mu = 1$. Results demonstrated that for the cases where $\frac{\lambda}{\mu} > 1$, $\lceil E(X_q) \rceil > r^*$, and for the cases where $\frac{\lambda}{\mu} < 1$, $\lceil E(X_q) \rceil < r^*$. When $\frac{\lambda}{\mu} = 1$, $\lceil E(X_q) \rceil = r^*$. Finally, there was a direct correspondence between the size of $\frac{\lambda}{\mu}$ and the value of $E(X_q)$. This information is summarised in Table 2.1.

Next, similar situations were considered in the case where $s = 39$. This time we considered the situations where $\lambda = 0.7$ and $\mu = 0.3$, $\lambda = 0.3$ and $\mu = 0.7$, $\lambda = 0.6$ and $\mu = 0.4$, $\lambda = 0.4$ and $\mu = 0.6$, and finally $\lambda = \mu = 0.5$. This time the case in which $\frac{\lambda}{\mu} = 1$ led to $\lceil E(X_q) \rceil < r^*$, but it still remains true that where $\frac{\lambda}{\mu} > 1$, $\lceil E(X_q) \rceil > r^*$ and where $\frac{\lambda}{\mu} < 1$, $\lceil E(X_q) \rceil < r^*$. This information is summarised in Table 2.2.

Overall, experimental results suggest that there is a critical region for the value $\frac{\lambda}{\mu}$ within which $\lceil E(X_q) \rceil = r^*$. If $\frac{\lambda}{\mu}$ lies above that interval, then it appears $\lceil E(X_q) \rceil > r^*$, and when it lies below this interval, $\lceil E(X_q) \rceil < r^*$. The size and location of this

theoretical interval is dependent upon the value of s, when $s = 29$ the ratio $\frac{1.03}{0.97}$ leads to $\lceil E(X_q) \rceil = r^*$ but when $s = 34$ this is no longer the case, and that ratio leads to $\lceil E(X_q) \rceil > r^*$ instead.

2.7.2 Concluding Remarks

In this chapter we have demonstrated that, if Aalen's conjectures do in fact hold, the method by which they can be defined in rigorous mathematical terms is less obvious than might originally have been believed. The expected value of the starting state for the initial distribution, $E(X_q)$, gives in general a good approximation of the value r^*, but the two are not always equal, and moreover examples can be constructed that make the estimation less impressive, and in addition so far no method can be found which allows the latter to be directly determined by the former. We are unaware of any other attempt to describe Aalen's conjectures in more detail, or to either prove or disprove them. Clearly this subject will benefit from more attention. Subsection 2.7.1 offers two alternate approaches to finding a method of comparing initial distributions, but time constraints prevented detailed consideration of either of them.

Another potential avenue of future research would be considering hazard rates for imprecise Markov chains, and what can be said about their shape. Discrete time imprecise Markov chains are defined in Chapters 3 and 4, but there are few results regarding such chains in continuous-time. It is not clear how one would define the hazard rate in such a setting (would one give upper and lower bounds upon the survival function, for instance?), though Coolen and Newby [13] give one suggestion. It may however be possible to adapt Aalen's conjectures by specifying two values, r_* and r^*, such that the imprecise hazard rate is known to be unimodal for initial distribution $\boldsymbol{e}_i$ with $i < r_*$, and known to be non-decreasing for initial distribution $\boldsymbol{e}_i$ with $i > r^*$.

Chapter 3

Time-Homogeneous Markov Chains with Imprecision

3.1 Introduction

In this chapter we discuss discrete-time Markov chains on a finite state space and with one absorbing state. For these chains it is not assumed that all one-step transition probabilities are precisely known, instead each individual row of the transition matrix is known to be an element of a given set of probability distributions. While the precise value of $p_{ij}^{(n)} := P(X(n+1) = j | X(n) = i)$ may be unknown, it is assumed that $p_{ij}^{(n)} = p_{ij}^{(m)}$, for all integers $n, m \geq 0$. In other words, it is known that we are in the time-homogeneous case (see Section 1.2.1).

Over the course of this chapter we will describe a generalisation of the idea of the *limiting conditional distribution* to the imprecise case. It will be proven, subject to mild conditions similar to those required in the precise case, that as time approaches infinity these imprecise chains, conditioned on non-absorption, tend towards a set of distributions which are conditionally invariant. Methods for approximating this set will be presented, and applied to examples.

3.2 Time Homogeneous Markov Chains with Imprecision

In this section imprecision is introduced to time-homogeneous Markov chains. The approach taken follows work by Kozine and Utkin [42], in which each element of the transition matrix is known to be within a given interval. They thus use interval probabilities, which we defined in Section 1.2.2. It was also explained in that section that each interval probability has an associated structure, which is made up of all probability measures on the measurable space $(S, \mathcal{A})$ that lie between the bounds of the interval. In this chapter we relax the model given in [42] by allowing elements of the transition matrix to take values, not just from a single interval probabilities, but from finite unions of interval probabilities. Each of these unions will also have an associated structure.

The reason for this generalisation of Kozine and Utkin's approach to imprecise Markov chains may not be obvious at first. It is difficult to imagine a particularly plausible situation in which a transition probability is considered to be independent of time, but also to lie within a known union of intervals rather than a single interval. What data or intuition could imply a value of $p_{ij}^{(n)}$ which was independent of n and belonged either to the interval $[\frac{1}{4}, \frac{1}{3}]$ or the interval $[\frac{2}{3}, \frac{3}{4}]$, for example? Situations can be considered for which this model *could* be applied, say one in which a coin is known to be biased such that $p \in [\frac{1}{4}, \frac{1}{3}]$ but with p either the probability of heads or the probability of tails. More importantly, the generalisation is consistent with the work in this chapter to follow along similar lines to Chapter 4, where the assumption that $p_{ij}^{(n)}$ is independent of n is no longer used. At such a point situations in which the possible values of transition probabilities do not include every value of a single interval become more plausible.

We begin by describing the transition matrix for the chain. Each row of the transition matrix will be taken from a closed set of probability distributions. In turn, each element of these probability distributions will be taken from a finite union of interval probabilities, as discussed above.

Definition 3.2.1 Define $s+2$ closed sets of probability distributions, $\mathcal{R}^{(i)}$, $i = -1, 0, \ldots, s$. The transition matrix for the chain takes the form

$$P = \begin{pmatrix} \boldsymbol{r}^{(-1)} \\ \boldsymbol{r}^{(0)} \\ \\ \cdots \\ \boldsymbol{r}^{(s)} \end{pmatrix} \tag{3.1}$$

where $\boldsymbol{r}^{(i)} \in \mathcal{R}^{(i)}$, for all $i \in S$. Clearly $\mathcal{R}^{(-1)} = \{(1, 0, \ldots, 0)\}$ in order to ensure that -1 is an absorbing state.

Definition 3.2.2

$$\mathcal{P} := \{P : \boldsymbol{r}^{(i)} \in \mathcal{R}^{(i)},\ \forall i \in S\} \tag{3.2}$$

Thus $\mathcal{P}$ denotes the set of all possible transition matrices for the chain. Let $\underline{c}_{ij} := \min_{\boldsymbol{r}^{(i)} \in \mathcal{R}^{(i)}} r_j^{(i)}$ and $\overline{c}_{ij} := \max_{\boldsymbol{r}^{(i)} \in \mathcal{R}^{(i)}} r_j^{(i)}$. These values are of great use in several results in this chapter, some of which require the existence of maxima and minima over each $\mathcal{R}^{(i)}$, and it is for this reason that it is assumed that the sets $\mathcal{R}^{(i)}$ are closed.

As was stated in Section 1.2.1, the limiting conditional distribution of a precise Markov chain only exists under certain mild conditions. Generalisations of these conditions are necessary in this chapter. Specifically, the sets $\mathcal{R}^{(i)}$ must be defined so that all transition matrices in $\mathcal{P}$ describe Markov chains for which C is a single communicating class, and for which each state in C is aperiodic. A specific method for ensuring that these conditions hold will not be given. However, both the periodicity of each state and the number of communicating classes are properties which depend only on which transitions are and are not possible. If we know with certainty whether or not a jump is possible, irrespective of how likely or not that jump is, then either every transition matrix in $\mathcal{P}$ has C a single communicating class with each state aperiodic, or none of them do. Therefore sufficient conditions for the aperiodicity of each state in C and that C be a single communicating class are given as follows: let there be no i, j for which $0 = \underline{c}_{ij} < \overline{c}_{ij}$, and let any transition matrix

in $\mathcal{P}$ have C a single communicating class with each state aperiodic. We assume these conditions from this point forward.

Since in the case considered in this chapter it is assumed that one of the matrices in $\mathcal{P}$ is in fact the actual matrix for all time steps, the two conditions given above ensure that the chain has a *quasi-stationary distribution* (QSD), even though our lack of knowledge regarding the chain means that the QSD is unknown.

Lemma 3.2.1 For a finite time-homogeneous Markov chain $\mathcal{X}$ on the state space $S = \{-1\} \cup C$ with one-step transition probability matrix $P \in \mathcal{P}$ and $\boldsymbol{v}$ a proper distribution over S,

$$\boldsymbol{v}P = (1, 0, \ldots, 0) \Longleftrightarrow \boldsymbol{v} = (1, 0, \ldots, 0). \tag{3.3}$$

Proof To prove that $\boldsymbol{v}P = (1, 0, \ldots, 0) \Rightarrow \boldsymbol{v} = (1, 0, \ldots, 0)$ we assume the contrary. If there exists $\boldsymbol{v} \neq (1, 0, \ldots, 0)$ and $P \in \mathcal{P}$ such that $\boldsymbol{v}P = (1, 0, \ldots, 0)$ then there is at least one element $v_i,\ i \in C$ such that $v_i > 0$. Since P is the transition matrix of a chain which is irreducible over C, there must be a strictly positive element P_{ij} for some $j \geq 0$. Hence we must have that $(\boldsymbol{v}P)_i > 0$, contradicting our assumption.

The fact that $\boldsymbol{v}P = (1, 0, \ldots, 0) \Leftarrow \boldsymbol{v} = (1, 0, \ldots, 0)$ is obvious. □

Lemma 3.2.1 demonstrates that at any finite time step absorption cannot be certain unless it was certain in the initial distribution. Therefore $(1, 0, \ldots, 0)$ can be excluded from any set of initial distributions without fear that it will re-appear at any finite time step.

Definition 3.2.3 Let

$$\mathcal{M}_0 := \{\boldsymbol{v} = (v_i)_{i \in S} \mid 0 \leq v_i \leq 1,\ \forall i \in S, \sum_{i \in S} v_i = 1\} \setminus \{(1, 0, \ldots, 0)\} \tag{3.4}$$

denote all initial distributions over the set S. Further, let $\mathcal{D}_0$ be the set of all possible initial distributions corresponding to $\mathcal{X}$ in a specific situation, where $\mathcal{D}_0$ is non-empty.

It is important to understand what is happening here, to avoid confusion. $\mathcal{M}_0$ describes all distributions over a set of $s+2$ states, with the exception of $(1, 0, \ldots, 0)$.

$\mathcal{D}_0$ describes all the distributions that have been determined as possible for the *specific situation* being described. Hence

$$\mathcal{D}_0 \subseteq \mathcal{M}_0. \tag{3.5}$$

If no restrictions are placed on the initial distribution, beyond the fact that the possibility of certain absorption at time 0 is excluded, then we have

$$\mathcal{M}_0 = \mathcal{D}_0. \tag{3.6}$$

This can be considered as the case of "maximum imprecision" with regard to the initial distribution. In fact, in Sections 3.3 and 3.4 it will be proved that the choice of $\mathcal{D}_0$ does not affect the long-term behaviour of the chain both with or without conditioning on non-absorption. From this point on, the phrase "possible initial distributions" will be used to refer to those distributions that have been determined to be possible for a given situation.

3.3 Long-Term Behaviour

In order to study the long-term behaviour of imprecise Markov chains it is necessary to consider the possible distributions at each finite time step n. Since we are in the time-homogeneous case, we know that there exists a single matrix in $\mathcal{P}$ which contains the actual transition probabilities for every time step. We thus make use of the following definition. For a one-step transition probability matrix $P \in \mathcal{P}$ and a given $\mathcal{D}_0$, the set $\mathcal{D}_n(P)$ of all possible state distributions of $\mathcal{X}$ at time $n \geq 1$ over the state space S can be defined inductively.

Definition 3.3.1

$$\mathcal{D}_n(P) = \{\boldsymbol{v}P \mid \boldsymbol{v} \in \mathcal{D}_{n-1}(P)\} = \{\boldsymbol{v}P^n \mid \boldsymbol{v} \in \mathcal{D}_0(P)\} \tag{3.7}$$

where $\mathcal{D}_0(P) := \mathcal{D}_0$. Further,

$$\mathcal{M}_n(P) = \{\boldsymbol{v}P \mid \boldsymbol{v} \in \mathcal{M}_{n-1}(P)\} = \{\boldsymbol{v}P^n \mid \boldsymbol{v} \in \mathcal{M}_0(P)\} \tag{3.8}$$

where $\mathcal{M}_0(P) := \mathcal{M}_0$.

Note that we have added the transition matrix P to the notation, to reflect the fact that we are in the time-homogenous case, and thus that the transition matrix cannot change from time step to time step. This notation will be very useful when we begin to take the unions of these sets over all possible transition matrices for the chain.

Thus, if $\boldsymbol{v} \in \mathcal{D}_n(P)$ we have that $\boldsymbol{v} = (v_{-1}, v_0, \ldots, v_s)$ satisfies $v_j = \sum_{i \in S} \tilde{v}_i p_{ij}(n)$ for a $\tilde{\boldsymbol{v}} \in \mathcal{D}_0$ and $p_{ij}^{(n)} = [P^n]_{ij}$. The upcoming Lemma 3.3.1 shows that the sets $\mathcal{D}_n(P)$ are nested when $\mathcal{D}_0 = \mathcal{M}_0$. This is useful for calculating the set of *β-invariant distributions* of $\mathcal{X}$, defined below (see e.g. Li [44]).

Definition 3.3.2 Consider a finite time-homogeneous Markov chain $\mathcal{X}$ with one-step transition probability matrix P. A collection of nonnegative numbers $\boldsymbol{\mu} = \{\mu_i\}_{i \in S}$ with $\boldsymbol{\mu} \neq \mathbf{0}$ and $\sum_{i \in S} \mu_i = 1$ satisfying

$$\beta \sum_{i \in S} \mu_i p_{ij} = \mu_j, \quad j \in S \tag{3.9}$$

is called a *β-invariant distribution* of $\mathcal{X}$ (or P) over S. In the case where $\beta = 1$, $\boldsymbol{\mu}$ is also called an *invariant distribution*.

The stationary distribution (see Section 1.1) is an invariant distribution. A quasi-stationary distribution is a $\frac{1}{\lambda}$-invariant distribution where λ is the dominating eigenvalue of P^*, which is defined in (1.3). This follows from the fact that $\boldsymbol{\alpha}$ can be obtained by solving $\boldsymbol{\alpha} P^* = \lambda \boldsymbol{\alpha}$. The advantage of this terminology is that it allows us to describe a distribution as being in effect a quasi-stationary distribution and describing the equivalent dominant eigenvalue at the same time.

Lemma 3.3.1

$$\mathcal{M}_{n+1}(P) \subseteq \mathcal{M}_n(P) \tag{3.10}$$

Proof For each $P \in \mathcal{P}$, it follows from $\mathcal{D}_0 = \mathcal{M}_0$ and the fact that P is a stochastic matrix that $\mathcal{M}_1(P) = \{\boldsymbol{v}P \mid \boldsymbol{v} \in \mathcal{M}_0\} \subseteq \mathcal{M}_0$. Now assume that for a certain $n \geq 1$, $\mathcal{M}_n(P) \subseteq \mathcal{M}_{n-1}(P)$. Then

$$\mathcal{M}_{n+1}(P) = \{\boldsymbol{v}P \mid \boldsymbol{v} \in \mathcal{M}_n(P)\} \subseteq \{\boldsymbol{v}P \mid \boldsymbol{v} \in \mathcal{M}_{n-1}(P)\} = \mathcal{M}_n(P) \tag{3.11}$$

as needed, so the general property follows by induction. □

If we knew which matrix $P \in \mathcal{P}$ correctly described the chain, we would know that the set of possible distributions at time n was $\mathcal{D}_n(P)$. Since we do not have this information, however, we know only that the set of all possible distributions at time n must lie within the union of all $\mathcal{D}_n(P)$ for all $P \in \mathcal{P}$.

Definition 3.3.3 Let

$$\tilde{\mathcal{D}}_n := \bigcup_{P \in \mathcal{P}} \mathcal{D}_n(P) \tag{3.12}$$

and

$$\tilde{\mathcal{M}}_n := \bigcup_{P \in \mathcal{P}} \mathcal{M}_n(P). \tag{3.13}$$

We are interested in the set containing all possible limiting distributions of $\mathcal{X}$ corresponding to all $P \in \mathcal{P}$, that is, the set of all distributions of $\mathcal{X}$ at time n, with n tending to infinity.

Definition 3.3.4 Let

$$\mathcal{D}_\infty(P) := \{\lim_{n\to\infty}(\boldsymbol{v}P^n) : \boldsymbol{v} \in \mathcal{D}_0\}, \quad \forall P \in \mathcal{P} \tag{3.14}$$

and

$$\tilde{\mathcal{D}}_\infty := \bigcup_{P \in \mathcal{P}} \mathcal{D}_\infty(P). \tag{3.15}$$

Hence, $\boldsymbol{v} = (v_{-1}, v_0, \ldots, v_s) \in \mathcal{D}_\infty(P)$ satisfies

$$v_j = \lim_{n\to\infty} \sum_{i \in S} \tilde{v}_i p_{ij}(n) \tag{3.16}$$

for a $\tilde{\boldsymbol{v}} \in \mathcal{D}_0$, where $p_{ij}(n) = [P^n]_{ij}$. In the case $\mathcal{D}_0 = \mathcal{M}_0$ for all $P \in \mathcal{P}$, we define

$$\tilde{\mathcal{M}}_\infty = \bigcup_{P \in \mathcal{P}} \mathcal{M}_\infty(P) = \bigcup_{P \in \mathcal{P}} \{\lim_{n\to\infty}(\boldsymbol{v}P^n) : \boldsymbol{v} \in \mathcal{M}_0\} \tag{3.17}$$

Lemma 3.3.2

$$\tilde{\mathcal{M}}_\infty = \{\boldsymbol{\pi}\} \tag{3.18}$$

where $\boldsymbol{\pi} = (1, 0, \ldots, 0)$.

Proof It is known (see e.g. Kijima [41]) that absorption is certain for any $P \in \mathcal{P}$ and $\boldsymbol{v} \in \mathcal{M}_0$. Consequently $\mathcal{M}_\infty(P) = \{\boldsymbol{\pi}\}$ for all $P \in \mathcal{P}$ and all $\boldsymbol{v} \in \mathcal{M}_0$. Therefore $\bigcup_{P\in\mathcal{P}} \mathcal{M}_\infty(P) = \{\boldsymbol{\pi}\}$. □

Corollary 3.3.1

$$\tilde{\mathcal{D}}_\infty = \{\boldsymbol{\pi}\} \tag{3.19}$$

where $\boldsymbol{\pi} = (1, 0, \ldots, 0)$.

Proof Since we have from (3.5) that $\mathcal{D}_0 \subseteq \mathcal{M}_0$ for every $P \in \mathcal{P}$, we have that

$$\mathcal{D}_n(P) = \{\boldsymbol{v}P^n : \boldsymbol{v} \in \mathcal{D}_0\} \subseteq \{\boldsymbol{v}P^n : \boldsymbol{v} \in \mathcal{M}_0\} = \mathcal{M}_n(P) \tag{3.20}$$

for every $P \in \mathcal{P}$. It follows from (3.12) therefore that $\tilde{\mathcal{D}}_n \subseteq \tilde{\mathcal{M}}_n$ and from (3.15) that $\tilde{\mathcal{D}}_\infty \subseteq \tilde{\mathcal{M}}_\infty$. From Lemma 3.3.2, we have either that $\tilde{\mathcal{D}}_\infty = \{\boldsymbol{\pi}\}$, or that $\tilde{\mathcal{D}}_\infty = \emptyset$. Since $\mathcal{D}_0$ is non-empty, however, and every $\boldsymbol{v} \in \mathcal{D}_0$ is such that $\lim_{n\to\infty} \boldsymbol{v}P^n = \boldsymbol{\pi}$, $\tilde{\mathcal{D}}_\infty = \emptyset$ cannot hold. □

Let us consider what is learned from this lemma and its corollary. We now know that irrespective of how little information we have regarding a set of imprecise time-homogeneous Markov chains, each with one absorbing state and $s + 1$ transient, aperiodic states that form a single communicating class, it must be the case that each and every member of that set will become absorbed with certainty as time goes to infinity.

By the very definition of an absorbing state $\boldsymbol{\pi}$ must satisfy $\boldsymbol{\pi}P = \boldsymbol{\pi}$. Therefore $\boldsymbol{\pi}$ is an invariant distribution for each $P \in \mathcal{P}$ (see Definition 3.3.2). It is also known (see e.g. Kijima [41]) that under the conditions assumed in Section 3.2, each P has only one such invariant distribution. Therefore $\{\boldsymbol{\pi}\}$ can be referred to as an *invariant set of distributions*, in the sense that it is invariant under multiplication by a set of stochastic matrices. At this moment it may seem unnecessary to use such a term to describe a set with only one element, but results in Sections 3.4 and Chapter 4 will expand upon the concept. In the set notation used in Section 3.4 and Chapter 4, we have that $\{\boldsymbol{\pi}\}$ is the only set N for which

$$N\mathcal{P} = N \tag{3.21}$$

where $N\mathcal{P} = \{\boldsymbol{v}P : \boldsymbol{v} \in N, P \in \mathcal{P}\}$; this will be the standard method for multiplying sets from this point on. Any set N with this property is referred to as an invariant set of distributions.

Having described the long-term behaviour of the Markov chains under consideration, we move on to considering the effect of conditioning on non-absorption.

3.4 Conditional Distributions

For a precise Markov chain on a finite state space including an absorbing state, it is known (see e.g. Darroch and Seneta [21]) that the process may settle down to some kind of equilibrium over the non-absorbing states before absorption takes place. Here we consider similar properties on the long-term behaviour of the state probabilities of an imprecise finite time-homogeneous Markov chain $\mathcal{X}$ at time n conditioned on non-absorption, with set of possible one-step transition probability matrices $\mathcal{P}$. Let $\boldsymbol{d}(n) = (d_0(n), \ldots, d_s(n))$ be the distribution of $\mathcal{X}$ at time n under the condition that absorption has not occurred yet, then we have from [21] that the components of $\boldsymbol{d}(n)$ satisfy

$$d_j(n) = \frac{P(X(n) = j)}{P(X(n) \neq -1)} = \frac{p_j(n)}{1 - p_{-1}(n)}, \quad j \in C, \tag{3.22}$$

where, for $k \in S$, $p_k(n) = \sum_{i \in S} v_i p_{ik}(n)$ with $\boldsymbol{v} \in \mathcal{M}_0$ and $P \in \mathcal{P}$. We are interested in the set of all possible limiting conditional distributions $\boldsymbol{d} = (d_0, d_1, \ldots, d_s)$ of $\mathcal{X}$, where

$$d_j := \lim_{n \to \infty} d_j(n), \quad j \in C. \tag{3.23}$$

For an imprecise time-homogeneous Markov chain $\mathcal{X}$ there is still only one limiting conditional distribution, but its value is unknown, since the correct transition matrix is unknown, and the LCD is a $\frac{1}{\lambda}$-invariant distribution (see Definition 3.3.2) to that matrix. However, it is known to be a proper distribution over the transient states C. It is therefore of use to define a function that transforms probability distributions across S into probability distributions across C. At each time $n \geq 0$, the distribution $\boldsymbol{d}(n)$ of $\mathcal{X}$ conditioned on non-absorption is obtained from the unconditional distribution $\boldsymbol{p}(n)$ of $\mathcal{X}$ by applying the function $f : [0, 1) \times [0, 1]^{s+1} \to$

$[0,1]^{s+1}$ defined by

$$f(\boldsymbol{p}(n)) := f((v_{-1}, v_0, \ldots, v_s)) = \frac{1}{1-v_{-1}}(v_0, v_1, \ldots, v_s). \tag{3.24}$$

Note that $(1, 0, \ldots, 0)$ is not within the domain of f. We have

$$\boldsymbol{d}(n) = f(\boldsymbol{p}(n)). \tag{3.25}$$

From (3.24), the set of possible conditional distributions at time $n \geq 0$ for $\mathcal{X}$ with one-step transition probability matrix $P \in \mathcal{P}$ over C is

$$\mathcal{D}_n^C(P) = \{f(\boldsymbol{v}) \mid \boldsymbol{v} \in \mathcal{D}_n(P)\} \tag{3.26}$$

for the set of possible initial distributions $\mathcal{D}_0$. In the case where $\mathcal{D}_0 = \mathcal{M}_0$, i.e. all initial distributions other than $(1, 0, \ldots, 0)$ are deemed possible, we have

$$\mathcal{M}_n^C(P) = \{f(\boldsymbol{v}) \mid \boldsymbol{v} \in \mathcal{M}_n(P)\} \tag{3.27}$$

where

$$\mathcal{M}_0^C(P) = \{f(\boldsymbol{v}) \mid \boldsymbol{v} \in \mathcal{M}_0\} =: \mathcal{M}_0^C. \tag{3.28}$$

Hence, $\boldsymbol{v}^* = (v_0^*, v_1^*, \ldots, v_s^*) \in \mathcal{D}_n^C(P)$ satisfies

$$\begin{aligned} v_j^* &= \frac{v_j}{1-v_{-1}} \quad \text{for a } \boldsymbol{v} \in \mathcal{D}_n(P) \\ &= \frac{p_j(n)}{1-p_{-1}(n)}, \end{aligned} \tag{3.29}$$

where $p_k(n) = \sum_{i \in S} \tilde{v}_i p_{ik}(n)$ for a $\tilde{\boldsymbol{v}} \in \mathcal{D}_0(P)$ and $p_{ik}(n) = [P^n]_{ik}$. We therefore have from (3.22) that

$$v_j^* = d_j(n). \tag{3.30}$$

This allows us to consider the possible distributions, conditioned upon non-absorption, at any time step, which will be of great use in finding an imprecise equivalent to the limiting conditional distribution. The next theorem shows that the sets $\mathcal{M}_n^C(P)$ are nested in a similar way to the sets $\mathcal{M}_n(P)$ (see Lemma 3.3.1).

Theorem 3.4.1 For each $P \in \mathcal{P}$ and $n \geq 0$,

$$\mathcal{M}_{n+1}^C(P) \subseteq \mathcal{M}_n^C(P). \tag{3.31}$$

Proof For each $P \in \mathcal{P}$ we have from (3.26) and Lemma 3.3.1 that

$$\begin{aligned} \mathcal{M}_{n+1}^{C}(P) &= \{f(\boldsymbol{v}) \mid \boldsymbol{v} \in \mathcal{M}_{n+1}(P)\}\} \\ &\subseteq \{f(\boldsymbol{v}) \mid \boldsymbol{v} \in \mathcal{M}_{n}(P)\}\} \\ &= \mathcal{M}_{n}^{C}(P) \end{aligned} \tag{3.32}$$

as needed. □

We are interested in the set of all possible limiting conditional distributions of $\mathcal{X}$ corresponding to all possible one-step transition probability matrices $P \in \mathcal{P}$ and all possible initial distributions $\boldsymbol{v} \in \mathcal{M}_0$. Let the sets $\mathcal{M}_0^0$ and $\mathcal{M}_0^\epsilon$ be defined by

$$\mathcal{M}_0^0 = \{(0, v_0, v_1, \ldots, v_s) \mid 0 \leq v_i \leq 1,\ \forall i \in C,\ \sum_{i \in C} v_i = 1\} \tag{3.33}$$

and

$$\mathcal{M}_0^\epsilon = \{(\epsilon, v_0, \ldots, v_s) \mid 0 < \epsilon < 1,\ 0 \leq v_i \leq 1,\ \forall i \in C,\ \sum_{i \in C} v_i = 1 - \epsilon\}. \tag{3.34}$$

Then the set $\mathcal{M}_0$ can be written as

$$\mathcal{M}_0 = \mathcal{M}_0^0 \cup (\cup_{\epsilon > 0} \mathcal{M}_0^\epsilon). \tag{3.35}$$

Lemma 3.4.1

$$\mathcal{M}_0^C = \{f(\boldsymbol{v}) \mid \boldsymbol{v} \in \mathcal{M}_0^0\} \tag{3.36}$$

Proof From (3.26) $\mathcal{M}_0^C = \{f(\boldsymbol{v}) \mid \boldsymbol{v} \in \mathcal{M}_0\}$. Hence, we have to prove that for all $\tilde{\boldsymbol{v}} \in \mathcal{M}_0^\epsilon$, $f(\tilde{\boldsymbol{v}}) \in \{f(\boldsymbol{v}) \mid \boldsymbol{v} \in \mathcal{M}_0^0\}$. Suppose $\exists \tilde{\boldsymbol{v}} \in \mathcal{M}_0^\epsilon$ such that $\boldsymbol{v^*} = f(\tilde{\boldsymbol{v}}) \notin \{f(\boldsymbol{v}) \mid \boldsymbol{v} \in \mathcal{M}_0^0\}$. Then $(0, v_0^*, v_1^*, \ldots, v_s^*) \notin \mathcal{M}_0^0$, but this is in contradiction with the definition (3.33) of $\mathcal{M}_0^0$, and hence $f(\tilde{\boldsymbol{v}}) \in \{f(\boldsymbol{v}) \mid \boldsymbol{v} \in \mathcal{M}_0\}$ for all $\tilde{\boldsymbol{v}} \in \mathcal{M}_0^\epsilon$. □

As a result of Lemma 3.4.1 attention can be restricted in this section to the distributions belonging to $\mathcal{M}_0^0$. Once again, since it is unknown which element of $\mathcal{P}$ correctly describes the process, we consider unions of sets within which all possible distributions at time step n are included.

Definition 3.4.1

$$\tilde{\mathcal{D}}_n^C := \bigcup_{P\in\mathcal{P}} \mathcal{D}_n^C(P) \tag{3.37}$$

and

$$\tilde{\mathcal{M}}_n^C := \bigcup_{P\in\mathcal{P}} \mathcal{M}_n^C(P). \tag{3.38}$$

Theorem 3.4.1 tells us it is appropriate to define the set of all possible limiting conditional distributions of $\mathcal{X}$ corresponding to all possible one-step transition probability matrices $P \in \mathcal{P}$ and all possible initial distributions $\boldsymbol{v} \in \mathcal{M}_0$.

Definition 3.4.2

$$\tilde{\mathcal{M}}_\infty^C := \bigcup_{P\in\mathcal{P}} \mathcal{M}_\infty^C(P) \tag{3.39}$$

where

$$\mathcal{M}_\infty^C(P) = \lim_{n\to\infty} \mathcal{M}_n^C(P). \tag{3.40}$$

Hence, $\boldsymbol{v} \in \mathcal{M}_\infty^C(P)$ implies that $\boldsymbol{v} = (v_0, v_1, \ldots, v_s)$ satisfies

$$v_j = \lim_{n\to\infty} \frac{p_j(n)}{1 - p_{-1}(n)} = \lim_{n\to\infty} \frac{\sum\limits_{i\in C} \tilde{v}_i p_{ij}(n)}{1 - \sum\limits_{i\in C} \tilde{v}_i p_{i,-1}(n)}, \tag{3.41}$$

for a $\tilde{\boldsymbol{v}} \in \mathcal{M}_0^C$ and $p_{ij}(n) = [P^n]_{ij}$. The next result is a corollary to Theorem 3.4.1.

Corollary 3.4.1

$$\tilde{\mathcal{M}}_n^C \subseteq \tilde{\mathcal{M}}_{n-1}^C \tag{3.42}$$

Proof We have that

$$\tilde{\mathcal{M}}_n^C = \bigcup_{P\in\mathcal{P}} \mathcal{M}_n^C \subseteq \bigcup_{P\in\mathcal{P}} \mathcal{M}_{n-1}^C = \tilde{\mathcal{M}}_{n-1}^C \tag{3.43}$$

where the subset is justified by Theorem 3.4.1. □

As a final comment in this section, note that the set $\tilde{\mathcal{M}}_\infty^C$ has the following property

$$\tilde{\mathcal{M}}_\infty^C = \bigcup_{P\in\mathcal{P}} \mathcal{M}_\infty^C(P) = \bigcup_{P\in\mathcal{P}} (f(\tilde{f}_\alpha(\mathcal{M}_\infty^C(P))P)) \tag{3.44}$$

where $\tilde{f}_\alpha(\cdot)$ is defined in Definition 4.5.3, but for the moment can be considered a function that maps a distribution over C to a distribution over S such that $f(\tilde{f}_{\alpha_1}(\boldsymbol{v})) = f(\tilde{f}_{\alpha_2}(\boldsymbol{v})) = \boldsymbol{v}$. In Chapter 4 we will define the *conditionally invariant set of distributions* (see (4.50)), that is those sets $\mathcal{N}$ for which

$$f(\tilde{f}_\alpha(\mathcal{N})\mathcal{P}) = \mathcal{N} \tag{3.45}$$

(compare this to (3.21), which expressed a similar concept without conditioning upon non-absorption). The set $\tilde{\mathcal{M}}_\infty^C$ is not a true conditionally invariant distribution, because of the assumption of time-homogeneity. Nevertheless, the similarity between (3.44) and (3.45) should be noted.

Theorem 3.4.2 For an imprecise time-homogeneous Markov chain $\mathcal{X}$, with set of possible one-step transition probability matrices $\mathcal{P}$, and any given set of initial distributions, the set of all possible limiting conditional distributions of $\mathcal{X}$ is given by

$$\tilde{\mathcal{M}}_\infty^C = \bigcup_{P\in\mathcal{P}} \{\boldsymbol{\alpha}(P)\} \tag{3.46}$$

where $\boldsymbol{\alpha}(P)$ is the limiting conditional distribution of $\mathcal{X}$ corresponding to P.

Proof The proof follows from the fact that, as discussed from Section 1.2.1, each $P \in \mathcal{P}$ has a unique limiting conditional distribution, denoted by $\boldsymbol{\alpha}(P)$, which is independent of the initial distribution. Hence $\bigcup_{P\in\mathcal{P}}\{\boldsymbol{\alpha}(P)\}$ must be equal to the set of all limiting conditional distributions of $\mathcal{X}$ with set of possible one-step transition probability matrices $\mathcal{P}$. □

3.5 Calculations and Examples

At present we are not aware of a general method for finding $\tilde{\mathcal{M}}_n^C$, for finite values of $n > 0$, directly from the support of $\mathcal{P}$. Since $\mathcal{P}$ can be defined algebraically (see (3.53), for example), it is possible to define $\tilde{\mathcal{M}}_n^C$ as an algebraic vector. However, depending on $\mathcal{P}$, the elements of this vector may be n-degree polynomials in up to $s^2 + 3s + 2$ variables. Finding the set such an algebraic vector defines can be problematic even for fairly low values of n.

In this section, therefore, three methods are presented by which approximations can be derived for the sets $\tilde{\mathcal{M}}_n^C$ and $\tilde{\mathcal{M}}_\infty^C$. Methods 1 and 2 can be applied to any Markov chain of the type in this chapter, although again computational complexity may become an issue. Method 1 replaces $\mathcal{M}_0^C$ and $\mathcal{P}$ with discrete subsets, Method 2 involves calculating bounds upon the elements of the sets $\tilde{\mathcal{M}}_n^C$. Finally, a third method, adapted from the unconditional case used by Kozine and Utkin [42] will be discussed; it will then be explained why, once conditioning upon non-absorption is considered, the method no longer gives useful results.

The probability simplex representation [67] is used here in order to graphically represent three-element probability distributions with the two dimensional probability simplex. A probability simplex representation is an equilateral triangle with perpendicular height one unit, in which each vertex represents the probability distribution with all mass in one state of C. The probabilities assigned to the three elements of C are identified with perpendicular distances from the three sides of the triangle. From Lemma 3.4.1 it follows that the set $\mathcal{M}_0^C$ is represented by the whole simplex diagram.

It should be noted that in each example in this section the unknown elements of the transition matrix are taken from intervals, rather than unions of intervals. This is partially because, as stated in Section 3.2, situations in the time-homogeneous case for which elements are known to be within unions of intervals do not seem particularly realistic, and also because using Method 2 for finding bounds on $\tilde{\mathcal{M}}_n^C$ would become far more complicated were we to consider unions of intervals.

Method 1: Approximation of $\mathcal{M}_n^C(P)$.

Method 1 approximates the sets $\mathcal{M}_0^C$ and $\mathcal{P}$ with discrete subsets, and then uses these approximations to find an approximation to $\tilde{\mathcal{M}}_n^C$. The set of possible initial distributions $\mathcal{M}_0^C$ is approximated by the discrete finite set $\underline{\mathcal{M}}_0^C$ where

$$\underline{\mathcal{M}}_0^C = \{(\frac{i}{\gamma}, \frac{k}{\gamma}, 1 - \frac{i+k}{\gamma} | i, k \in \{0, 1 \ldots, \gamma\} \text{ with } i + k \leq \gamma\}, \tag{3.47}$$

where γ is a positive integer. Next, we consider each combination of integers i, j for which $\underline{c}_{ij} < \overline{c}_{ij}$. Each of these combinations corresponds to a variable a_{ij} that can

take any value in a known interval. Note that the fact that each matrix must be stochastic means for each i one a_{ij} is expressible in terms of a_{ik} for all $k \neq j$. We have that $1 \leq |\{a_{ij}\}| \leq s^2 + 3s + 2$. For each a_{ij} we define

$$A_{ij} = \{\underline{c}_{ij} + \frac{k(\overline{c}_{ij} - \underline{c}_{ij})}{\delta_{ij}} | k \in \{0, 1 \ldots, \delta_{ij}\}\} \tag{3.48}$$

for a positive integer δ_{ij}, and

$$\underline{\mathcal{P}} := \{P \in \mathcal{P} | p_{ij} \in A_{ij}\} \subseteq \mathcal{P}. \tag{3.49}$$

We can thus find a discrete subset of $\mathcal{M}_n^C(P)$ for $P \in \underline{\mathcal{P}}$ as follows

$$\underline{\mathcal{M}}_n^C(P) = \{\boldsymbol{v}P^n | \boldsymbol{v} \in \underline{\mathcal{M}}_0^C\} \tag{3.50}$$

and a discrete subset $\tilde{\mathcal{M}}_n^C$ as follows

$$\underline{\tilde{\mathcal{M}}}_n^C := \bigcup_{P \in \underline{\mathcal{P}}} \underline{\mathcal{M}}_n^C(P). \tag{3.51}$$

Similarly we can find a discrete subset of $\tilde{\mathcal{M}}_\infty^C$, which we will denote $\underline{\tilde{\mathcal{M}}}_\infty^C$, by calculating the quasi-stationary distribution for each matrix in $\underline{\mathcal{P}}$.

Method 2: Bounds for $\tilde{\mathcal{M}}_n^C$.

In this method, rather than using discrete subsets, we calculate bounds on each element in the set $\tilde{\mathcal{M}}_n^C$. As in Method 1, we define a_{ij} as a variable lying in the interval $\underline{c}_{ij}, \overline{c}_{ij}$. We then define the algebraic matrix P such that $P_{ij} = a_{ij}$ for all i, j for which $\underline{c}_{ij} < \overline{c}_{ij}$, and equals $P_{ij} := \underline{c}_{ij}$ for all elements where $\underline{c}_{ij} = \overline{c}_{ij}$. We also define the vector $\boldsymbol{v}$ as follows

$$\boldsymbol{v} = (0, v_0, v_1, v_2) \tag{3.52}$$

where $v_0 + v_1 + v_2 = 1, v_i \geq 0$ for all $i = 0, 1, 2$. Thus every possible distribution over C at time step n can be described by the vector $\tilde{\boldsymbol{v}} := f(\boldsymbol{v}P^n)$. We can then maximise and minimise each element of $\tilde{\boldsymbol{v}}$ over the region defined by the conditions $a_{ij} \in [\underline{c}_{ij}, \overline{c}_{ij}]$, $\sum p_{ij} = 1$, $v_0 + v_1 + v_2 = 1$, and $v_i \geq 0$ for all $i = 0, 1, 2$. This will give bounds upon the elements of $\tilde{\mathcal{M}}_n^C$.

Example 3.5.1

Consider a time-homogeneous Markov chain $\mathcal{X}$ with state space $S = \{-1\} \cup C$ where $C = \{0, 1, 2\}$, and let the set of all possible one-step transition probability matrices $\mathcal{P}$ be given by

$$\mathcal{P} = \{ \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & a_{10} & 0 & 1 - a_{10} \\ 0 & 0 & 0.75 & 0.25 \end{pmatrix} \mid a_{10} \in [0.1, 0.2] \} \tag{3.53}$$

For Method 1, we take $\gamma = 20$ and $\delta_{10} = 10$. This leads to

$$\underline{\mathcal{M}}_0^C = \{(0.05i, 0.05k, 1 - 0.05(i + k)),\ i, k \in \{0, 1 \ldots, 20\} \text{ with } i + k \leq 20\} \tag{3.54}$$

and

$$A_{10} = \{0.1 + \frac{0.1k}{10}, k \in \{0, 1 \ldots, 10\}\}. \tag{3.55}$$

As can be seen in Figure 3.1 a), b) and c), this choice of γ and δ_{10} appears to generate sets with good coverage, in that the shapes of the sets these figures are approximating seem very clear. More coverage could be achieved by increasing the values of γ or δ_{10}, or both, but it seems unlikely that this would lead to a substantial increase in comprehension of the shape of the sets.

Equations (3.54) and (3.55) lead to the following subset of the set of all possible conditional distributions at time 1,

$$\underline{\tilde{\mathcal{M}}}_1^C = \{f \left((0, v_0, v_1, v_2) \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & a_{10} & 0 & 1 - a_{10} \\ 0 & 0 & 0.75 & 0.25 \end{pmatrix} \right) | a_{10} \in A_{10}, \boldsymbol{v} \in \underline{\mathcal{M}}_0^C\} \tag{3.56}$$

and therefore

$$\underline{\tilde{\mathcal{M}}}_1^C = \left\{ \frac{1}{2(2 - v_0)} (4a_{10}v_1, 2v_0 + 3v_2, 4(1 - a_{10})v_1 + v_2) | a_{10} \in A_{10}, \boldsymbol{v} \in \underline{\mathcal{M}}_0^C \right\}. \tag{3.57}$$

This set is shown in Figure 3.1 a).

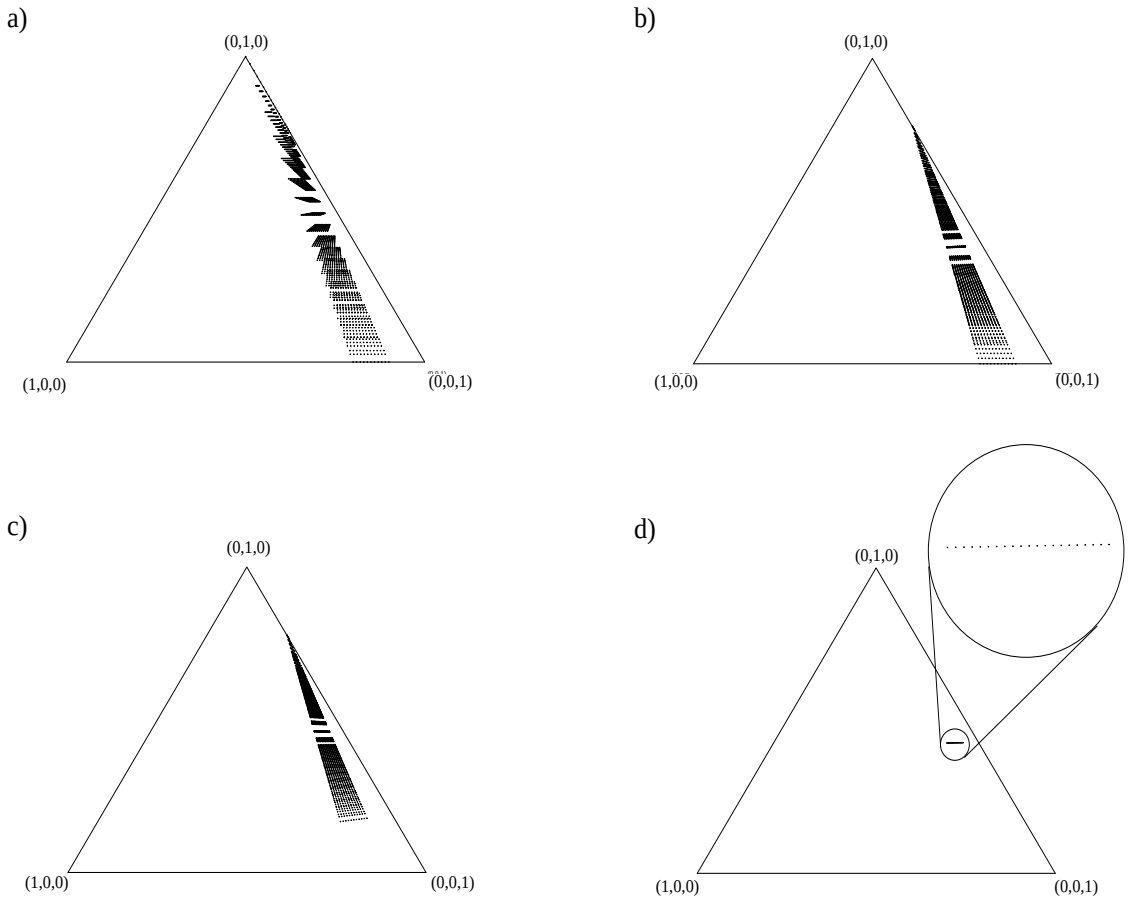


Figure 3.1: The sets a) $\underline{\tilde{\mathcal{M}}}_1^C$, b) $\underline{\tilde{\mathcal{M}}}_2^C$, c) $\underline{\tilde{\mathcal{M}}}_3^C$ and d) $\underline{\tilde{\mathcal{M}}}_\infty^C$.

Continuing in this way, subsets of all possible conditional distributions at time 2 and 3 are given as follows:

$$\underline{\tilde{\mathcal{M}}}_2^C = \{\frac{1}{8(2 - v_0 - a_{10}v_1)}(8a_{10}v_0 + 12a_{10}v_2, 4(3 - a_{10})v_1 + 3v_2,$$

$$8(1 - a_{10})v_0 + 4(1 - a_{10})v_1 + (13 - 12a_{10})v_2)$$

$$|a_{10} \in A_{10},\ \forall \boldsymbol{v} \in \underline{\mathcal{M}}_0^C\} \tag{3.58}$$

and

$$\underline{\tilde{\mathcal{M}}}_3^C = \{\frac{1}{8(8-(4+2a_{10})v_0-4a_{10}v_1-3a_{10}v_2)}(16a_{10}(3-a_{10})v_1+12a_{10}v_2,$$
$$8(3-a_{10})v_0+12(1-a_{10})v_1+3(13-4a_{10})v_2,$$
$$8(1-a_{10})v_0+4(13-17a_{10}+4a_{10}^2)v_1+(25-24a_{10})v_2)$$
$$|a_{10}\in A_{10},\ \forall \boldsymbol{v}\in \underline{\mathcal{M}}_0^C\}. \tag{3.59}$$

The simplex diagrams of the two sets (3.58) and (3.59) are given in Figure 3.1 b) and c), respectively.

Using Theorem 3.4.2 together with (3.49), an approximation of the set of all possible limiting conditional distributions of $\mathcal{X}$ is given by

$$\underline{\tilde{\mathcal{M}}}_\infty^C := \bigcup_{P\in\underline{\mathcal{P}}} \underline{\tilde{\mathcal{M}}}_\infty^C(P) = \bigcup_{P\in\underline{\mathcal{P}}} \boldsymbol{\alpha}(P) \tag{3.60}$$

where the quasi-stationary distribution $\boldsymbol{\alpha}(P)$ is obtained by solving

$$\boldsymbol{\alpha}(P)\, P^* = \lambda(P)\, \boldsymbol{\alpha}(P)$$

with $\lambda(P)$ the dominating eigenvalue of P^* (see (1.3)). Using Mathematica[1], we can find the set $\underline{\tilde{\mathcal{M}}}_\infty^C$. For a given value of a_{10}, the dominating eigenvalue of P^*, which we denote λ_{10} is found to be equal to

$$\lambda_{10} = \frac{(55-126a_{10}+6\sqrt{3}\sqrt{-441+328a_{10}-a_{10}^2+16a_{10}^3})^{\frac{1}{3}}+37-12a_{10}}{12(55-126a_{10}+6\sqrt{3}\sqrt{-441+328a_{10}-a_{10}^2+16a_{10}^3})^{\frac{1}{3}}}$$
$$+\frac{(55-126a_{10}+6\sqrt{3}\sqrt{-441+328a_{10}-a_{10}^2+16a_{10}^3})^{\frac{2}{3}}}{12(55-126a_{10}6\sqrt{3}\sqrt{-441+328a_{10}-a_{10}^2+16a_{10}^3})^{\frac{1}{3}}}. \tag{3.61}$$

Note that (3.61) contains complex values, all of these however cancel out across the expression. The associated quasi-stationary distribution is

$$\frac{3a_{10}}{a_{10}+3\lambda_{10}+4\lambda_{10}^2}\left(1, \frac{\lambda_{10}}{a_{10}}, \frac{2(2\lambda_{10}^2-a_{10})}{3a_{10}}\right). \tag{3.62}$$

Therefore

$$\tilde{\mathcal{M}}_\infty^C = \left\{\frac{3a_{10}}{a_{10}+3\lambda_{10}+4\lambda_{10}^2}(1, \frac{\lambda_{10}}{a_{10}}, \frac{2(2\lambda_{10}^2-a_{10})}{3a_{10}})|a_{10}\in A_{10}\right\} \tag{3.63}$$

where $\lambda=\lambda_{10}$ is given by (3.61). Table 3.1 gives the quasi-stationary distribution corresponding to each $a_{10}\in A_{10}$, while diagram d) in Figure 3.1 shows the simplex representation of the set (3.63).

State	0.1	0.11	0.12	0.13	0.14	0.15	0.16	0.17	0.18	0.19	0.2
0	.044	.048	.053	.057	.062	.066	.071	.075	.080	.085	.089
1	.428	.428	.427	.427	.427	.427	.427	.427	.427	.427	.427
2	.528	.524	.520	.516	.511	.507	.502	.498	.493	.488	.484

Table 3.1: Quasi-stationary distribution for each a_{10}

We now discuss Figure 3.1. We note first that these diagrams reflect what is known from Corollary 3.4.1, namely that $\tilde{\mathcal{M}}_n^C \subseteq \tilde{\mathcal{M}}_{n-1}^C$, though the rate of convergence is difficult to judge. Further, d) demonstrates that even with imprecision, the long-term behaviour conditioned on non-absorption is known to be a single distribution belonging to what is a very small set compared to $\mathcal{M}_0^C$.

The other property of interest in Figure 3.1 is the stratification of the points in diagrams a) through c). The method by which $\underline{\mathcal{M}}_0^C$ is constructed ensures that the approximation of $\mathcal{M}_0^C$ is a regular lattice. The behaviour demonstrated in these diagrams, in which some areas of $\underline{\tilde{\mathcal{M}}}_n^C$ for $n > 0$ contain more elements than others, is worth commenting on. Note that throughout this chapter (and the next) we have very deliberately avoided assigning distributions to the intervals within which the elements of the transition matrix are known to lie. A value for a transition probability is either possible, or it is impossible, we do not consider whether one possible value is more or less likely than another. However, by approximating $\mathcal{M}_0^C$ as a regular lattice and then multiplying that lattice by a set of matrices which have been defined by taking regular points from the intervals in which the transition probabilities lie, we are in fact implicitly assuming uniform distributions on the set of initial distributions and the intervals used to define $\mathcal{P}$. In this sense, figures a) through c) are demonstrating that if the initial distributions are drawn from uniform distributions, and the matrices from $\mathcal{P}$ drawn likewise, there is no reason to believe the resulting distributions at time n with $n > 0$ will be uniformly distributed. In fact, it is for precisely this reason that we consider specific values of the transition probabilities as merely either possible or impossible. Were we to do otherwise we

[1]Version 6.

would have to consider combining distributions, and the run-on effects upon the theory would be considerable.

We begin Method 2 by noting that we have from (3.56) that every vector $\boldsymbol{v}^{(1)} \in \tilde{\mathcal{M}}_1^C(P)$ can be written as

$$\boldsymbol{v}^{(1)} = \frac{1}{2(2-v_0)}\left(4a_{10}v_1, 2v_0+3v_2, 4(1-a_{10})v_1+v_2\right) \tag{3.64}$$

with $a_{10} \in [0.1, 0.2]$ and $\boldsymbol{v} \in \mathcal{M}_0^0$. We now need to find

$$\min_{a_{10}\in[0.1,0.2],\, \boldsymbol{v}\in\mathcal{M}_0^0} v_i^{(1)} \quad \text{and} \quad \max_{a_{10}\in[0.1,0.2],\, \boldsymbol{v}\in\mathcal{M}_0^0} v_i^{(1)} \tag{3.65}$$

for $i = 0, 1, 2$. Since (v_0, v_1, v_2) is an honest probability distribution, we must have that $v_0 + v_1 = 1 - v_2$, and so

$$\boldsymbol{v}^{(1)} = \left(\frac{4a_{10}v_1}{2(2-v_0)}, \frac{3-v_0-3v_1}{2(2-v_0)}, \frac{1-v_0+(3-4a_{10})v_1}{2(2-v_0)}\right). \tag{3.66}$$

What is needed is the maximum and minimum values of each element of $\boldsymbol{v}^{(1)}$ within the region

$$R := \{(a_{10}, v_0, v_1) : a_{10} \in [0.1, 0.2], v_0 \in [0, 1], v_1 \in [0, 1], v_0 + v_1 \leq 1\}.$$

Note that R does not describe a probability space, it is simply a subset of $\mathbb{R}^3$. We find the minima and maxima over R by partially differentiating each element of $\boldsymbol{v}^{(1)}$

$$\frac{\partial}{\partial v_0}\boldsymbol{v}^{(1)} = \left(\frac{8a_{10}v_1}{(4-2v_0)^2}, \frac{2-6v_1}{(4-2v_0)^2}, \frac{(6-8a_{10})v_1-2}{(4-2v_0)^2}\right) \tag{3.67}$$

and

$$\frac{\partial}{\partial v_1}\boldsymbol{v}^{(1)} = \left(\frac{4a_{10}}{2(2-v_0)}, \frac{-3}{2(2-v_0)}, \frac{(3-4a_{10})}{2(2-v_0)}\right). \tag{3.68}$$

Note that for each element of $\boldsymbol{v}^{(1)}$ there is no co-ordinate within the region R at which its derivative with respect to v_0 and its derivative with respect to v_1 are both zero. Thus the maximum and minimum values of each element lie on the boundary of R. This leads us to $v_0^{(1)} \in [0, 0.2]$, $v_1^{(1)} \in [0, 1]$ and $v_2^{(1)} \in [0, 0.9]$. These bounds are shown in diagram a) of Figure 3.2.

Similarly, each vector $\boldsymbol{v}^{(2)} \in \bigcup_{P\in\mathcal{P}} \mathcal{M}_2^C(P)$ and $\boldsymbol{v}^{(3)} \in \bigcup_{P\in\mathcal{P}} \mathcal{M}_3^C(P)$ can be represented as

$$\boldsymbol{v}^{(2)} = \left(\frac{8a_{10}v_0+12a_{10}v_2}{8(2-v_0-a_{10}v_1)}, \frac{4(3-a_{10})v_1+3v_2}{8(2-v_0-av_1)}, \frac{8(1-a_{10})v_0+4(1-a_{10})v_1+(13-12a_{10})v_2}{8(2-v_0-a_{10}v_1)}\right)$$

and

$$\boldsymbol{v}^{(3)}=\Bigg(\frac{16a(3-a_{10})v_1+12a_{10}v_2}{8(8-(4+2a_{10})v_0-4a_{10}v_1-3a_{10}v_2)},\ \frac{8(3-a_{10})v_0+12(1-a_{10})v_1+3(13-4a_{10})v_2}{8(8-(4+2a_{10})v_0-4a_{10}v_1-3a_{10}v_2)},\ \frac{8(1-a_{10})v_0+4(13-17a_{10}+4a_{10}^2)v_1+(25-24a_{10})v_2}{8(8-(4+2a_{10})v_0-4a_{10}v_1-3a_{10}v_2)}\Bigg)$$

with $a_{10} \in [0.1, 0.2]$ and $v_0 + v_1 + v_2 = 1$. Finding the minimum and maximum of each component of $\boldsymbol{v}^{(2)}$ and $\boldsymbol{v}^{(3)}$ individually leads to $v_0^{(2)} \in [0, 0.2]$, $v_1^{(2)} \in [0, 0.778]$, $v_2^{(2)} \in [0.222, 0.9]$ and $v_0^{(3)} \in [0, 0.156]$, $v_1^{(3)} \in [0.167, 0.778]$ and $v_2^{(3)} \in [0.222, 0.746]$. These bounds are shown in Figure 3.2 in diagrams b) and c), respectively. By minimising and maximising the elements of (3.62) we can obtain bounds for the quasi-stationary distribution, yielding $v_0^{(\infty)} \in [0.044, 0.089]$, $v_1^{(\infty)} \in [0.427, 0.428]$, $v_2^{(\infty)} \in [0.484, 0.529]$, these bounds are shown in diagram d) in Figure 3.2. However, the minimum and maximum bounds for $v_1^{(\infty)}$ are so close that they are all but indistinguisable in the simplex diagram. We see that the set of possible conditional distributions at time n shrinks quickly with n.

We now compare Figure 3.1 and 3.2. The region in each diagram in Figure 3.2 is a strict superset of the equivalent region in Figure 3.1. Notice though that for every bound in Figure 3.2 there is a distribution in the equivalent diagram in Figure 3.1 that lies on that bound. In other words, the bounds on each element of $\tilde{\mathcal{M}}_n^C$ are the same as the bounds on each element of $\underline{\tilde{\mathcal{M}}}_n^C$ when $n = 1, 2, 3$ or $n = \infty$. Therefore Figure 3.1 provides the better approximation to $\tilde{\mathcal{M}}_n^C$, as the bounds can be calculated directly from the diagrams it contains, and those diagrams also give some insight into the shape of $\tilde{\mathcal{M}}_n^C$. Thus the only disadvantage we can see for Method 1 is the time required to calculate so many distributions, as has been previously mentioned. Of course, if possible, it would be most sensible to apply both methods, as this will generate both an idea of the shape of each set *and* precise bounds for which the set must lie within.

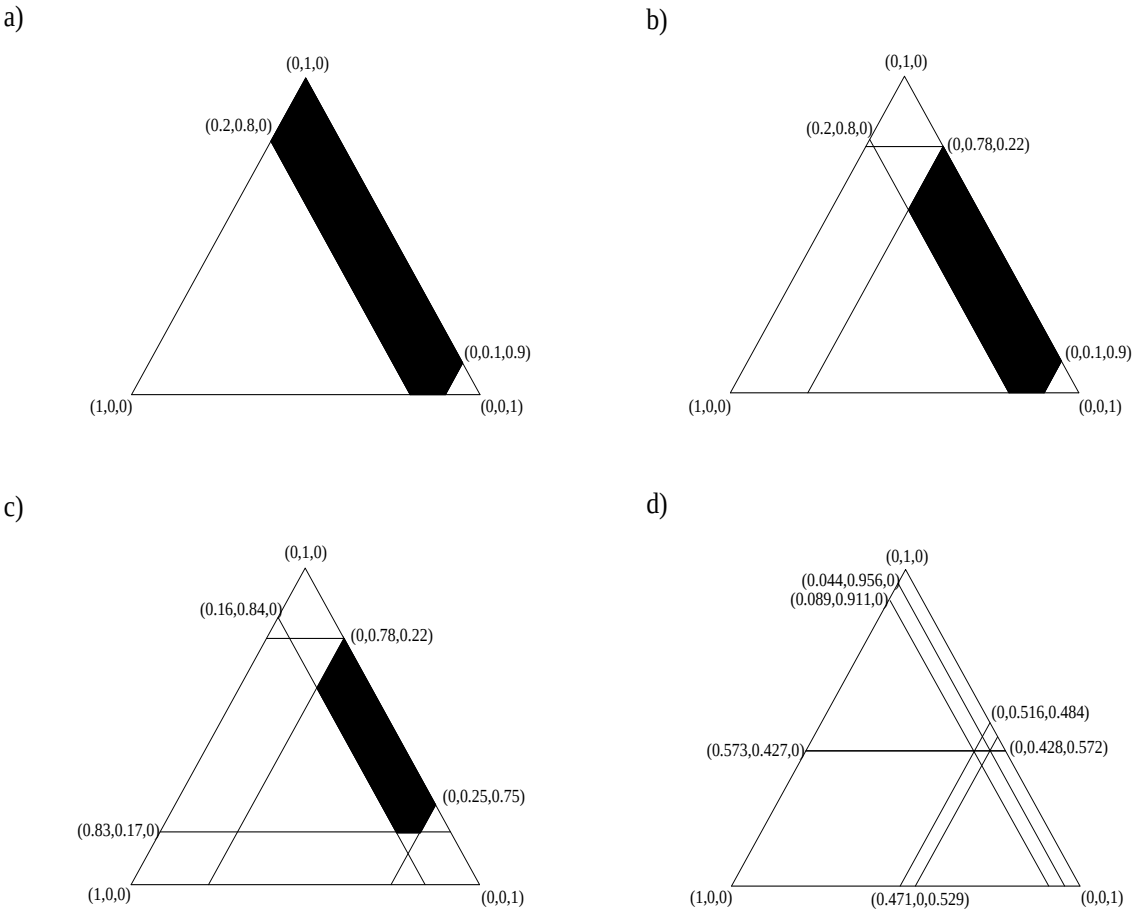


Figure 3.2: Bounds for the sets $\tilde{\mathcal{M}}_n^C$ for a) $n = 1$, b) $n = 2$, c $n = 3$ and d) $n = \infty$, when $a_{10} \in [0.1, 0.2]$.

Example 3.5.2

This example is similar to Example 3.5.1, the only change is a widening of the interval for the single unknown value. This will allow us to see how the bounds are affected by an increase in imprecision. This time $\mathcal{P}$ is

$$\mathcal{P} = \{ \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & a_{10} & 0 & 1 - a_{10} \\ 0 & 0 & 0.75 & 0.25 \end{pmatrix} \mid a_{10} \in [0.1, 0.4] \} \tag{3.69}$$

It is interesting to see how sensitive the bounds are to an increase in imprecision.

Figure 3.3 shows the diagrams corresponding to those in Figure 3.2, thus allowing direct comparison. Clearly, the bounds are wider in Example 3.5.2 than in Example 3.5.1, but once again the set of possible conditional distributions shrinks considerably with each time step shown. The bounds for the set of quasi-stationary distributions are $v_0^{(\infty)} \in [0.044, 0.187]$, $v_1^{(\infty)} \in [0.425, 0.428]$ and $v_2^{(\infty)} \in [0.388, 0.529]$. The interval in which v_1 is contained is so narrow that it is difficult to see the bounds upon the set, which demonstrates how much can be said about the long term behaviour of the chain, conditioned on non-absorption, even in a situation with imprecision.

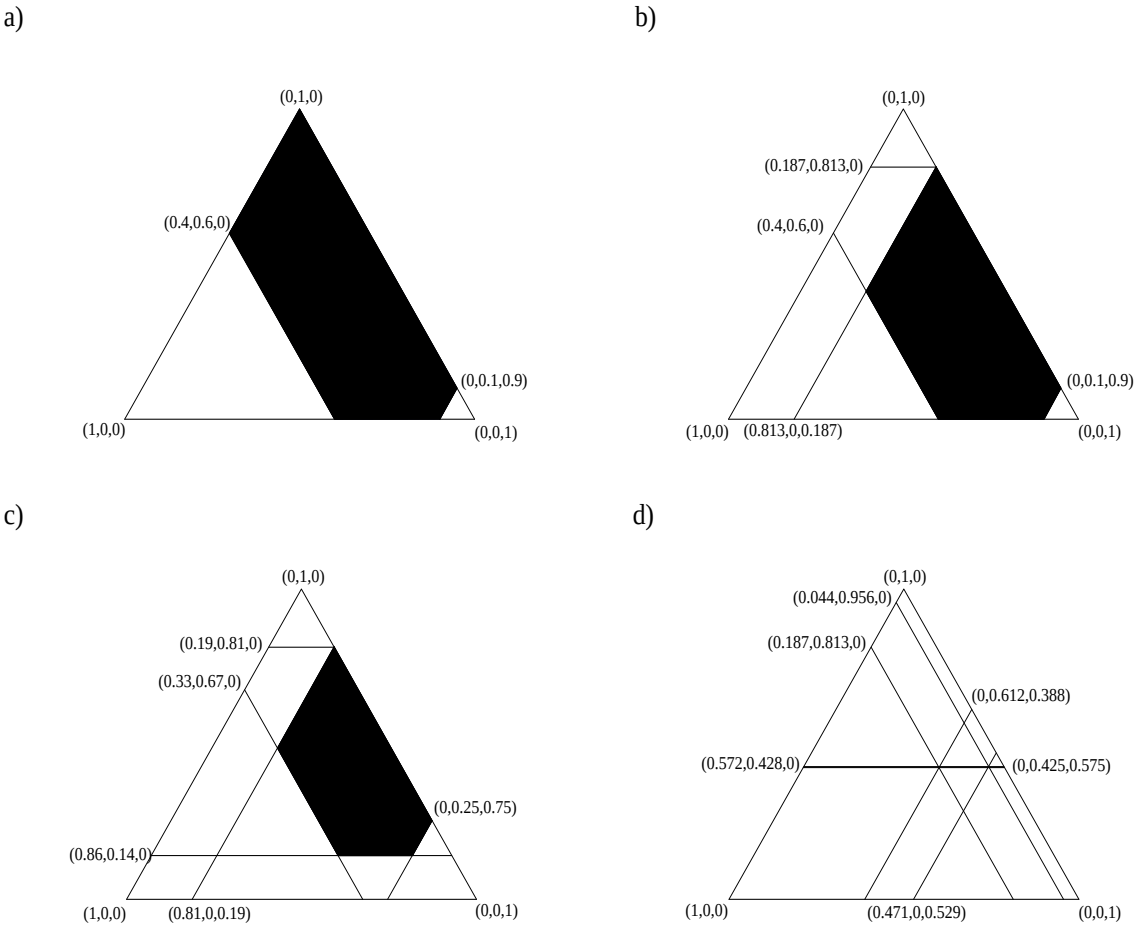


Figure 3.3: Bounds for the sets $\tilde{\mathcal{M}}_n^C$ for a) $n = 1$, b) $n = 2$, c) $n = 3$ and d) $n = \infty$, when $a_{10} \in [0.1, 0.4]$.

Note that, as one would expect, the region in each diagram in Figure 3.3 is a superset of the region of the corresponding diagram in Figure 3.2.

Example 3.5.3

In this third example there is no state $i \in C$ for which the transition probabilities p_{ij} are precisely known. This time the set of all possible one-step transition probability matrices $\mathcal{P}$ is given by

$$\mathcal{P}=\{ \begin{pmatrix} 1 & 0 & 0 & 0 \\ a_{0,-1} & 0 & 1-a_{0,-1} & 0 \\ 0 & a_{10} & 0 & 1-a_{10} \\ 0 & 0 & a_{21} & 1-a_{21} \end{pmatrix} | a_{0,-1} \in [0.1, 0.4]; a_{10} \in [0.4, 0.6]; a_{21} \in [0.65, 0.85] \} \tag{3.70}$$

Again, Method 2 is used to find bounds upon the elements of the possible distributions at time steps 1, 2, 3 and as time approaches infinity. The resulting simplex diagrams are shown in Figure 3.4. Note that in the first three time steps there is little difference between Examples 2 and 3 (most likely this is due to both chains being birth-death processes, limiting the paths that the process can take), but that the bounds upon $\tilde{\mathcal{M}}_\infty^C$ are indeed wider in Example 3.5.3 than those in Example 3.5.2, as would be expected due to the increased imprecision.

Kozine and Utkin [42] presented another method for finding bounds on the state distribution $\boldsymbol{p}(n)$ at time n for a finite time-homogeneous imprecise irreducible Markov chain. In the terminology of this thesis, this is equivalent to finding bounds on the elements of the set $\tilde{\mathcal{M}}_n = \cup_{P \in \mathcal{P}} \mathcal{M}_n(P)$ (note that these are sets of distributions over S and not C). This method can easily be adapted to obtain bounds on the elements of the conditional distribution $\boldsymbol{d}(n)$ at time n (see (3.22)), which of course are also bounds on the elements of the set $\tilde{\mathcal{M}}_n$. However, the resulting bounds may well prove to be of little use, as will be illustrated and discussed with an example.

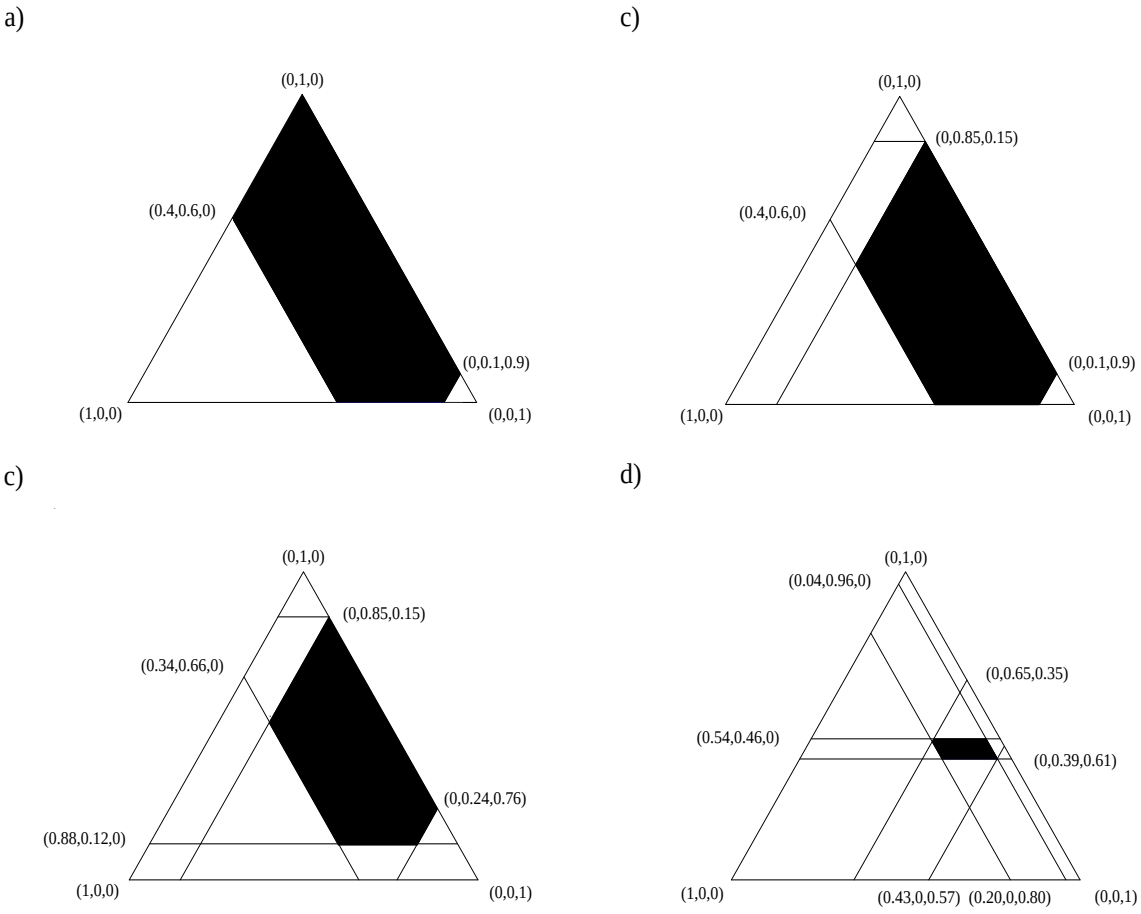


Figure 3.4: Bounds for the sets $\tilde{\mathcal{M}}_n^C$ for a) $n = 1$, b) $n = 2$, c) $n = 3$ and d) $n \to \infty$, when $a_{0,-1} \in [0.1, 0.4]$, $a_{10} \in [0.4, 0.6]$, and $a_{21} \in [0.65, 0.85]$.

Method 3: Alternate bounds for $\tilde{\mathcal{M}}_n^C$.

In [42] it is proved that for finite imprecise time-homogeneous Markov chains for which C is a single communicating class with all states aperiodic, and for bounds on the initial distribution, $\underline{p}_j(0) \leq p_j(0) \leq \overline{p}_j(0)$, bounds can be defined as follows

$$\underline{p}_j(n) := \sum_{i=-1}^{s} \underline{p}_i(n-1)\underline{c}_{ij} \tag{3.71}$$

and

$$\overline{p}_j(n) := \sum_{i=-1}^{s} \overline{p}_i(n-1)\overline{c}_{ij} \tag{3.72}$$

for all $j \in S$, so that $\underline{p}_j(n) \leq p_j(n) \leq \overline{p}_j(n)$ for all $n > 0$ and for all $j \in S$, where

$p_j(n) = P(X(n) = j)$. This result gives a method for calculating upper and lower bounds, $\overline{p}_j(n)$ and $\underline{p}_j(n)$, respectively, on the jth component of the elements of the set $\tilde{\mathcal{M}}_n$ for a given value of n, where $j \in S$.

A small adaptation to this result enables us to derive upper and lower bounds, which we shall denote $\overline{p}_j^C(n)$ and $\underline{p}_j^C(n)$ respectively, upon the jth element of the set $\tilde{\mathcal{M}}_n^C$ for a given value of n, where $j \in C$. The lower bounds are calculated first. First we set $\underline{p}_j(0) > 0$ for at least one value of j, thus ensuring $(1, 0, \ldots, 0)$ cannot be an initial distribution. By minimising over the probability that the process is in state j at time n, given absorption has not occurred, we define

$$\underline{p}_j^C(n) := \inf_{p_i, p_{ij}} \frac{\sum_{i=0}^s p_i(n-1)p_{ij}}{1 - \sum_{i=0}^s p_i(n-1)p_{i,-1}}, \ \forall j \in C \tag{3.73}$$

which holds as $p_0(n-1) < 1$ follows directly from $\overline{p}_j(0) < 1$ (see Lemma 3.2.1). Since all elements of the numerator are positive, as are all elements of the sum in the denominator, the numerator can be minimised and the denominator maximised simultaneously, leading to

$$\underline{p}_j^C(n) := \frac{\sum_{i=0}^s \underline{p}_i(n-1)\underline{c}_{ij}}{1 - \underline{p}_{-1}(n-1) - \sum_{i=0}^s \underline{p}_i(n-1)\underline{c}_{i,-1}}, \ \forall j \in C. \tag{3.74}$$

The upper bound is calculated in a similar way.

$$\overline{p}_j^C(n) := \sup_{p_i, p_{ij}} \frac{\sum_{i=0}^s p_i(n-1)p_{ij}}{1 - \sum_{i=0}^s p_i(n-1)p_{i,-1}}, \ \forall j \in C. \tag{3.75}$$

Note that by taking the supremum of $\sum_{i=0}^s p_i(n-1)p_{ij}$ over p_{ij} separately from the supremum of $\sum_{i=0}^s p_i(n-1)p_{i,-1}$ over p_i, there is no reason to believe that the resulting value represents a conditional probability. Therefore, there is also no reason that the resulting value cannot be greater than 1, and so we have

$$\overline{p}_j^C(n) = \min\Big\{\frac{\sum_{i=0}^s \overline{p}_i(n-1)\overline{c}_{ij}}{1 - \overline{p}_{-1}(n-1) - \sum_{i=0}^s \overline{p}_i(n-1)\overline{c}_{i,-1}}, 1\Big\}, \ \forall j \in C. \tag{3.76}$$

All together, this gives us $\underline{p}_j^C(n) \leq p_j^C(n) \leq \overline{p}_j^C(n)$, where $p_j^C(n) = P(X(n) = j | X(n) \geq 0)$. These bounds are illustrated in Example 3.5.4.

Example 3.5.4

The imprecise Markov chain $\mathcal{X} = \{X(n), n = 0, 1, \ldots\}$ has $\mathcal{P}$

$$\{\begin{pmatrix} 1 & 0 & 0 & 0 \\ a_{0,-1} & 0 & 1-a_{0,-1} & 0 \\ 0 & a_{10} & 0 & 1-a_{10} \\ 0 & 0 & a_{21} & 1-a_{21} \end{pmatrix} | a_{0,-1} \in [0.1, 0.2], a_{10} \in [0.45, 0.6], a_{21} \in [0.7, 0.8]\}. \tag{3.77}$$

By (3.74) and (3.76) the following equations are obtained

$$\begin{aligned} \underline{p}_0^C(n) &= \frac{0.45\underline{p}_1(n-1)}{1-0.1\underline{p}_0(n-1)-\underline{p}_{-1}(n-1)} \\ \underline{p}_1^C(n) &= \frac{0.8\underline{p}_0(n-1)+0.7\underline{p}_2(n-1)}{1-0.1\underline{p}_0(n-1)-\underline{p}_{-1}(n-1)} \\ \underline{p}_2^C(n) &= \frac{0.4\underline{p}_1(n-1)+0.2\underline{p}_2(n-1)}{1-0.1\underline{p}_0(n-1)-\underline{p}_{-1}(n-1)} \end{aligned} \tag{3.78}$$

and

$$\begin{aligned} \overline{p}_0^C(n) &= \frac{0.6\overline{p}_1(n-1)}{1-0.2\overline{p}_0(n-1)-\overline{p}_{-1}(n-1)} \\ \overline{p}_1^C(n) &= \frac{0.9\overline{p}_0(n-1)+0.8\overline{p}_2(n-1)}{1-0.2\overline{p}_0(n-1)-\overline{p}_{-1}(n-1)} \\ \overline{p}_2^C(n) &= \frac{0.55\overline{p}_1(n-1)+0.3\overline{p}_2(n-1)}{1-0.2\overline{p}_0(n-1)-\overline{p}_{-1}(n-1)}. \end{aligned} \tag{3.79}$$

Consider all possible initial distributions satisfying

$$\underline{p}_{-1}(0) = \overline{p}_{-1}(0) = 0 \tag{3.80}$$

$$\underline{p}_0(0) = \frac{1}{4},\ \underline{p}_1(0) = \frac{2}{5},\ \underline{p}_2(0) = \frac{1}{5} \tag{3.81}$$

$$\overline{p}_0(0) = \frac{7}{20},\ \overline{p}_1(0) = \frac{1}{2},\ \overline{p}_2(0) = \frac{2}{5}. \tag{3.82}$$

We have from (3.80) that there is zero probability of beginning in the absorbing state, immediately giving us $p_i^C(0) = p_i(0)$ for $i = 0, 1, 2$. From (3.81) and (3.82) we have lower and upper bounds for beginning in each of the three states in C. Hence $\underline{p}_j(n)$ and $\overline{p}_j(n)$ can be calculated for all $n > 0$ and for all $j \in S$ and $\underline{p}_j^C(n)$ and $\overline{p}_j^C(n)$ follow for all $n > 0$ and for all $j \in C$. Table 3.2 gives these values up to

n	$\underline{\boldsymbol{p}}(n)$	$\overline{\boldsymbol{p}}(n)$
1	$(0.025, 0.180, 0.340, 0.200)$	$(0.070, 0.300, 0.635, 0.395)$
2	$(0.043, 0.153, 0.284, 0.176)$	$(0.130, 0.381, 0.586, 0.468)$
3	$(0.058, 0.128, 0.246, 0.149)$	$(0.206, 0.352, 0.717, 0.463)$
4	$(0.071, 0.111, 0.206, 0.128)$	$(0.277, 0.430, 0.687, 0.533)$
5	$(0.082, 0.093, 0.178, 0.108)$	$(0.363, 0.412, 0.814, 0.538)$
6	$(0.091, 0.080, 0.150, 0.093)$	$(0.445, 0.488, 0.801, 0.609)$
7	$(0.099, 0.068, 0.129, 0.079)$	$(0.543, 0.480, 0.927, 0.623)$
8	$(0.106, 0.058, 0.109, 0.067)$	$(0.639, 0.556, 0.931, 0.697)$
n	$\underline{\boldsymbol{p}}^C(n)$	$\overline{\boldsymbol{p}}^C(n)$
1	$(0.185, 0.349, 0.205)$	$(0.323, 0.683, 0.425)$
2	$(0.160, 0.297, 0.184)$	$(0.438, 0.674, 0.538)$
3	$(0.136, 0.261, 0.158)$	$(0.443, 0.903, 0.583)$
4	$(0.119, 0.222, 0.138)$	$(0.595, 0.949, 0.737)$
5	$(0.101, 0.194, 0.118)$	$(0.646, 1.000, 0.843)$
6	$(0.088, 0.165, 0.102)$	$(0.880, 1.000, 1.000)$
7	$(0.075, 0.143, 0.087)$	$(1.000, 1.000, 1.000)$
8	$(0.065, 0.122, 0.075)$	$(1.000, 1.000, 1.000)$

Table 3.2: Bounds on $\boldsymbol{p}^C(n)$ for Example 3.5.4

$n = 8$, the values $\underline{p}_j(n)$, $\overline{p}_j(n)$, $\underline{p}_j^C(n)$ and $\overline{p}_j^C(n)$ are written as elements of vectors $\underline{\boldsymbol{p}}(n)$, $\overline{\boldsymbol{p}}(n)$, $\underline{\boldsymbol{p}}^C(n)$ and $\overline{\boldsymbol{p}}^C(n)$, respectively.

Note that for each $i = 0, 1, 2$, $\overline{p}_i^C(n)$ quickly reaches 1 as n increases. Further, the values of $\underline{p}_i^C(n)$ seem to tend to zero, by the 43rd time step all three elements of $\underline{\boldsymbol{p}}^C(n)$ are less than 0.0005. Considering that the actual amount of imprecision in this example is fairly small, it seems reasonable to conclude that for more general examples, this method will not prove to be very useful.

General concluding remarks on the long-term behaviour of imprecise Markov chains conditioned on non-absorption will be presented at the end of Chapter 4. For now, we simply note the different advantages and disadvantages to the first two methods. Method 1 seems to produce a far more accurate approximation to $\tilde{\mathcal{M}}_n^C$ than Method 2, since the 231 initial distributions that we use to approximate $\mathcal{M}_0^C$ are comparatively close together, thus providing a reasonable approximation, and thus the general shape of $\tilde{\mathcal{M}}_n^C$ is easy to make out from $\underline{\tilde{\mathcal{M}}}_n^C$, even though the latter is a strict subset of the former. However, the computational time required to find the nth power of eleven 3x3 matrices, and then multiplying each by 231 vectors before conditioning, will become problematic for large examples.

Method 2, on the other hand, guarantees bounds that lie on the set, and at each time step requires the calculation of only one algebraic vector by one algebraic matrix, for which maxima and minima can then be directly found using, for example, Mathematica. The obvious drawback is that the resulting supersets of $\tilde{\mathcal{M}}_n^C$ give no insight into the true shape of $\tilde{\mathcal{M}}_n^C$. Whenever possible, however, it would be best to use both methods, as each method has its own advantages, that combine well.

Chapter 4

Time-Inhomogeneous Markov Chains with Imprecision

4.1 Introduction

In this chapter we generalise the model defined in Chapter 3. Specifically, the property that transition probabilities are independent of time is no longer assumed. Thus, not only is the transition matrix unknown at each time step, it is allowed to change from one step to the next. In this sense, the work here can be considered as generalising the theory of finite time-inhomogeneous Markov chains in the classical case.

The motivation for this approach should be clear. There are many situations in which it would be unrealistic to assume that the transition matrix of a Markov chain will be independent of the time steps. This makes the time-inhomogeneous case a very important one, and it follows that describing an imprecise model for this case is worthwhile. Moreover, it is intuitively reasonable that a transition matrix that can change from time step to time step, potentially in unpredictable ways, might be sensibly modelled by placing bounds upon each element, thus the method of describing imprecision by interval probability lends itself well to time-inhomogeneous Markov chains. Lastly, there is not a great deal known about the long-term behaviour of time-inhomogeneous Markov chains in the precise case, and this is a situation that the method presented here can potentially change.

The long-term behaviour of the time-inhomogeneous case has been studied in detail by Škulj [62,63], who proved that, subject to mild conditions, there exists an invariant set of distributions that describes the long-term behaviour of such chains. The work in [62] is expanded upon in Sections 4.2 and 4.3.

Škulj [62, 63] did not however consider the effect of adding an absorbing state to the chain (doing so would violate the assumptions made in [63]). In Section 4.4 it is proven that with the inclusion of an absorbing state, and the imposition of mild conditions, absorption is certain. Thus a method for conditioning upon non-absorption in this case is of interest, and is presented in Section 4.5. The long-term behaviour of these chains conditioned upon non-absorption is then described. This satisfies one of the main goals of this thesis, namely to demonstrate that there exists a generalisation of the *limiting conditional distribution* in the imprecise time-inhomogeneous case. It is then demonstrated in Section 4.6 that this long-term behaviour conditioned upon non-absorption is independent of the choice of set of initial distributions, which strengthens the link between our results and those in the precise case. Section 4.7 contains examples illustrating our method, and Section 4.8 compares the model considered in this chapter with the one presented in Chapter 3. Finally, concluding remarks are given in Section 4.9.

4.2 Markov Chains with Interval Probabilities

Much of the set-up for the model to be introduced in this chapter follows that described in Chapter 3. Consider a Markov chain $\mathcal{X} = \{X(n), n = 0, \ldots\}$ with state space

$$S = \{-1, 0, \ldots, s\} = \{-1\} \cup C \tag{4.1}$$

where -1 is an absorbing state, and the finite set of states C is a single communicating class with each state aperiodic. C has the properties that $i, j \in C$ implies that i and j communicate, and that -1 is reachable from C. We will assume that known bounds exist on the possible values of each transition probability at each step. Škulj assumes in [62] and [63] that closed intervals can be used to define the possible values of transition probabilities. Just as in Chapter 3, this method is generalised

as follows.

Define $s+2$ closed sets of probability distributions, $\mathcal{R}^{(i)},\ i=-1,0,\ldots,s$. A transition matrix for this chain is defined by

$$P=\begin{pmatrix}\boldsymbol{r}^{(-1)}\\ \boldsymbol{r}^{(0)}\\ \cdots\\ \boldsymbol{r}^{(s)}\end{pmatrix}. \tag{4.2}$$

Thus the set of possible transition matrices for a given time step can be defined as follows.

Definition 4.2.1 All potential transition matrices for a given time step belong to the set

$$\mathcal{P}:=\{P|\ \boldsymbol{r}^{(i)}\in\mathcal{R}^{(i)},\ \forall i\in C\} \tag{4.3}$$

where the choice of the element from $\mathcal{R}^{(i)}$ has no effect on the choice of the element $\mathcal{R}^{(j)}$ if $i\neq j$.

Note that the probability sets themselves are independent of the current step, even though the transition matrix may change from one step to the next. Note also that $\mathcal{R}^{(-1)}=(1,0,\ldots,0)$ and that $\boldsymbol{e}_i\notin\mathcal{R}^{(i)}$ for $i>-1$, where $\boldsymbol{e}_i=(\delta_{-1i},\delta_{0i},\ldots,\delta si)$ as defined in Chapter 3. This ensures that our Markov chain has exactly one absorbing state. The assumption is also made that each of the possible transition matrices guarantee that C is a single communicating class with every state aperiodic. Finally, if $[P]_{ij}=0$ for any $P\in\mathcal{P}$ then it is assumed that $[P']_{ij}=0$ for all $P'\in\mathcal{P}$. Thus a jump from state i to state j is either possible independently of the time step, or impossible independently of the time step. This prevents situations in which the matrices $P_1,P_2,\ldots,P_r\in\mathcal{P}$ each represent chains which are irreducible over C but the chain represented by $\Pi_{i=1}^{r}P_i$ is *not* irreducible over C. While it is not difficult to prove that such a combination of matrices could exist, it is not clear that such combinations would be permitted by the earlier assumptions. Thus, it may be possible to prove that this final requirement is already covered by those assumptions previously made. We state it separately for now, however, due to its importance in proving Lemma 4.6.6 in Section 4.6.

Following on from Definition 4.2.1, the set of possible n-step transition matrices is now defined.

Definition 4.2.2

$$\mathcal{P}^n := \{P_1 P_2 \dots P_n,\ P_i \in \mathcal{P}\} \tag{4.4}$$

Note that

$$\mathcal{P}^1 = \mathcal{P}. \tag{4.5}$$

A property of sets of transition matrices is now defined that will become critical to later work.

Definition 4.2.3 The set $\mathcal{P}$ is called *regular* if for some n every $P \in \mathcal{P}^n$ has only strictly positive elements. Further, the set $\mathcal{P}$ is called *conditionally regular on* C if for some n every $P \in \mathcal{P}^n$ has all elements of the form P_{ij} strictly positive, where $i, j \in C$.

Lemma 4.2.1 All matrices which belongs to the set $\mathcal{P}^{s+1}$ have strictly positive elements beneath the first row, making $\mathcal{P}$ conditionally regular.

Proof Any matrix P that is contained in $\mathcal{P}^{s+1}$ will represent the behaviour of a time-inhomogeneous Markov chain over $s+1$ time steps. By assumption each of the time steps are described by transition matrices for which C is a single communicating class, with each element aperiodic. Therefore there is a path of n states, denoted $\{a_k\}_{k=1,\dots,n}$, between i and j, where $i, j \in C$, and no element of $\{a_k\}_{k=1,\dots,n}$ is equal to either i or j.

Assume $i \neq j$. By assumption the possibility of a jump from a given state to another given state is completely independent of which time step it is. Therefore if there exists $k_1 \neq k_2$ such that $a_{k_1} = a_{k_2}$, the elements $a_{k_1}, a_{k_1+1}, \dots, a_{k_2-1}$ can be removed from $\{a_k\}_{k=1,\dots,n}$ and the remainder still represents a viable path from i to j. This process can continue until no duplicated value in the path remains, forcing $n \leq s-1$. Thus j can be reached from i in s jumps, forcing $P(X(s) = j | X(0) = i) > 0$. $P(X(s+1) = j | X(0) = i) > 0$ follows immediately from the fact that each

possible transition matrix has C has each state aperiodic, and so if there exists a route from i to j in exactly s steps, there exists a route from i to j in exactly $s+t$ steps, $t > 0$, since at least one state in that route allows the process to remain in that state from one time step to the next.

Now assume $i = j$. The same process as above applies, except that without duplicated values in the path we have $n \leq s$, and hence we can return to i after $s+1$ jumps, and $P(X(s+1) = j | X(0) = i) > 0$. □

Lastly, we consider initial distributions. The set of all possible distributions for which absorption is not certain is denoted as $\mathcal{M}_0$, thus

$$\mathcal{M}_0 := \{\boldsymbol{v} = (v_{-1}, v_0, \ldots, v_s) | \sum_{i=-1}^{s} v_i = 1\} \setminus \{(1, 0, \ldots, 0)\}. \tag{4.6}$$

Therefore any set of initial distributions, which in general will be denoted as $\mathcal{D}_0$, will have the property that $\mathcal{D}_0 \subseteq \mathcal{M}_0$.

4.3 Distributions at Step n

In the precise case, one finds the distribution at step n by simply multiplying the distribution at step $n-1$ by the appropriate transition matrix. The following method simply generalises this concept. For the set of initial distributions $\mathcal{D}_0$ the set of distributions at time n is defined by

$$\mathcal{D}_n := \{\boldsymbol{v}P | \boldsymbol{v} \in \mathcal{D}_{n-1}; P \in \mathcal{P}\} \tag{4.7}$$

where $\mathcal{P}$ is as defined in (4.3). Obviously, in the special case where it is assumed that all distributions are possible at step 0,

$$\mathcal{M}_n := \{\boldsymbol{v}P | \boldsymbol{v} \in \mathcal{M}_{n-1}; P \in \mathcal{P}\}. \tag{4.8}$$

Lemma 4.3.1 For the sets of initial distributions $\mathcal{D}_0$

$$\mathcal{D}_0 \subseteq \mathcal{D}_1 \Rightarrow \mathcal{D}_n \subseteq \mathcal{D}_{n+1}, \ \forall n \tag{4.9}$$

and

$$\mathcal{D}_1 \subseteq \mathcal{D}_0 \Rightarrow \mathcal{D}_{n+1} \subseteq \mathcal{D}_n, \ \forall n. \tag{4.10}$$

Proof Both parts of the lemma are proved by induction and through the use of (4.7). For the first part, assume that $\mathcal{D}_{n-1} \subseteq \mathcal{D}_n$. Note that

$$\begin{aligned} \mathcal{D}_{n+1} &= \{\boldsymbol{v}P | \boldsymbol{v} \in \mathcal{D}_n; P \in \mathcal{P}\} \\ &\supseteq \{\boldsymbol{v}P | \boldsymbol{v} \in \mathcal{D}_{n-1}; P \in \mathcal{P}\} \\ &= \mathcal{D}_n. \end{aligned} \tag{4.11}$$

The argument is almost identical for the proof of the second part of the lemma. □

Lemma 4.3.2 For all $\boldsymbol{v} \in \mathcal{M}_n$ and all $P \in \mathcal{P}$,

$$\boldsymbol{v}P = (1, 0, \ldots, 0) \Leftrightarrow \boldsymbol{v} = (1, 0, \ldots, 0). \tag{4.12}$$

Proof Follows immediately from Lemma 3.2.1. □

Corollary 4.3.1

$$\mathcal{M}_{n+1} \subseteq \mathcal{M}_n, \ \forall n \tag{4.13}$$

Proof Since $\mathcal{M}_0$ contains every possible distribution for which absorption is not certain, it immediately follows that $\mathcal{M}_1 \subseteq \mathcal{M}_0$, so long as $(1, 0, \ldots, 0) \notin \mathcal{M}_1$. Lemma 4.3.2 proves that this cannot be the case. □

This corollary (proven in a slightly different way in [62]) immediately allows the following definition (also found in [62]) for the limiting set of distributions

$$\mathcal{M}_\infty = \bigcap_{n=0}^{\infty} \mathcal{M}_n. \tag{4.14}$$

The following definition is now introduced.

Definition 4.3.1 Any set of distributions N with the property

$$\{\boldsymbol{v}P | \boldsymbol{v} \in N; P \in \mathcal{P}\} = N \tag{4.15}$$

is described as an *invariant set of distributions*.

Clearly the set $\mathcal{M}_\infty$ is non-empty (it will contain the stationary distribution of every $P \in \mathcal{P}$, for example), and has the property that

$$\{\boldsymbol{v}P | \boldsymbol{v} \in \mathcal{M}_\infty; P \in \mathcal{P}\} = \mathcal{M}_\infty. \tag{4.16}$$

Thus it is an invariant set of distributions. This set is analogous to the stationary distribution in the precise case, in the sense that once the invariant set is arrived at, it can never be left in future time-steps. Methods for approximating this set for given situations are described in [62].

4.4 $\mathcal{M}_\infty$ and the Absorbing State

In this section it is shown that under the conditions assumed in this chapter, eventual absorption remains certain. The most important part of this process is to prove that the sequence $\boldsymbol{v}P_1P_2\dots P_n$, where $\boldsymbol{v} \in \mathcal{M}_0$ and each $P_i \in \mathcal{P}$, tends to $(1, 0, \dots, 0)$ as n tends to infinity. This result will follow from the theorem below.

Theorem 4.4.1 For $P_i \in \mathcal{P}^{s+1}$ (see Definition 4.2.2)

$$\lim_{n\to\infty} P_1P_2\dots P_n = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \dots & 0 \end{pmatrix}. \tag{4.17}$$

Proof Let

$$B^{(n)} := P_1P_2\dots P_n = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ b^{(n)}_{0,-1} & b^{(n)}_{00} & b^{(n)}_{01} & \dots & 1-\sum_{j=-1}^{s-1} b^{(n)}_{0j} \\ b^{(n)}_{1,-1} & b^{(n)}_{10} & b^{(n)}_{11} & \dots & 1-\sum_{j=-1}^{s-1} b^{(n)}_{1j} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ b^{(n)}_{s,-1} & b^{(n)}_{s0} & b^{(n)}_{s1} & \dots & 1-\sum_{j=-1}^{s-1} b^{(n)}_{sj} \end{pmatrix} \tag{4.18}$$

where $P_k \in \mathcal{P}^{s+1}$. As previously mentioned in the comments following Definition 4.2.3 this guarantees that each P_i will be strictly positive below the first row. It immediately follows that

$$b^{(n+1)}_{i,-1} = b^{(n)}_{i,-1} + \sum_{j=0}^{s} [P_{n+1}]_{j,-1} b^{(n)}_{ij} \tag{4.19}$$

and therefore that either $b_{i,-1}^{(n)} = 1$, or $b_{i,-1}^{(n)}$ is strictly increasing in n for each $i \in C$, since $[P_{n+1}]_{ji} > 0$, for all $j \in C$, which in turn means that $b_{ij}^{(n)} > 0$ for all $i \in C$ and $j \in S$.

Assume that for a given $i \in C$, $\lim_{n\to\infty} b_{i,-1}^{(n)} = \alpha_i < 1$. At each step n, therefore, we have that $\sum_{j=0}^{s} b_{ij}^{(n)} > 1 - \alpha_i > 0$. By the fact that each row of $\mathcal{P}$ (and hence of $\mathcal{P}^{s+1}$) is generated from a closed set of distributions, there exists a set of constants $\{\underline{c}_i\}$ such that $[P_k]_{ji} \geq \underline{c}_i > 0$ for all $j \in C$. Thus

$$\begin{aligned} b_{i,-1}^{(n+1)} &\geq b_{i,-1}^{(n)} + \sum_{j=0}^{s} \underline{c}_{-1} b_{ij}^{(n)} \\ &= b_{i,-1}^{(n)} + \underline{c}_{-1} \sum_{j=0}^{s} b_{ij}^{(n)} \\ &> b_{i,-1}^{(n)} + \underline{c}_{-1}(1 - \alpha_i). \end{aligned} \tag{4.20}$$

Since $\underline{c}_{-1}(1 - \alpha_i)$ is independent of n there must exist an n_0 such that $b_{i,-1}^{(n_0)} > 1$, which is clearly impossible. Thus no such $\alpha_i < 1$ can exist, and since $b_{i,-1}^{(n)}$ is known to be both increasing in n and never more than 1, it must be that $\lim_{n\to\infty} b_{i,-1}^{(n)} = 1$ as required. □

Corollary 4.4.1 Let $P_i \in \mathcal{P}^{s+1}$, $\forall i$. Further let $\boldsymbol{x} \in \mathcal{M}_0$. Then

$$\lim_{n\to\infty} \boldsymbol{x} P_1 P_2 \dots P_n = (1, 0, \dots, 0). \tag{4.21}$$

Proof

$$\begin{aligned} \lim_{n\to\infty} \boldsymbol{x} P_1 P_2 \dots P_n &= \boldsymbol{x}(\lim_{n\to\infty} P_1 P_2 \dots P_n) \\ &= \boldsymbol{x}(\mathbf{1}, \mathbf{0}, \dots, \mathbf{0}) \\ &= (\sum_{i=-1}^{s} x_i, 0, \dots, 0) = (1, 0, \dots, 0) \end{aligned} \tag{4.22}$$

where the first equality comes from the associativity of matrix multiplication, the second equality comes from Theorem 4.4.1, and the final equality comes from the fact that $\boldsymbol{x}$ is an honest probability distribution. □

Corollary 4.4.2 For the imprecise Markov chain $\mathcal{X} = \{X(n), n = 0, \ldots\}$ on the state space $S = \{-1\} \cup C$, where -1 is an absorbing state, the only probability distribution that can be an element of $\mathcal{M}_\infty$ (see [62]) is $(\pi_{-1}, \pi_0, \ldots, \pi_s) = (1, 0, \ldots, 0)$ where

$$\pi_i = \lim_{n\to\infty} (\boldsymbol{x} P_1 P_2 \ldots P_n)_i = (\boldsymbol{x} \Pi_{n=1}^{\infty} P_n)_i \tag{4.23}$$

with $P_i \in \mathcal{P}$. Hence,

$$\mathcal{M}_\infty = \{(1, 0, 0, \ldots, 0)\}. \tag{4.24}$$

Proof Define

$$\mathcal{M}_{i+1} = \{\boldsymbol{x} P : \boldsymbol{x} \in \mathcal{M}_i, P \in \mathcal{P}\} \tag{4.25}$$

and

$$\mathcal{M}_\infty = \cap_{i=0}^{\infty} \mathcal{M}_i. \tag{4.26}$$

This leads to

$$\mathcal{M}_\infty = \{\lim_{n\to\infty} \boldsymbol{x} P_1 P_2 \ldots P_n : \boldsymbol{x} \in \mathcal{M}_0; P_i \in \mathcal{P},\ \forall i\}. \tag{4.27}$$

From Theorem 4.4.1 and Corollary 4.4.1 we have that

$$\mathcal{M}_\infty = \{\lim_{n\to\infty} \boldsymbol{x} P_1' P_2' \ldots P_n' : \boldsymbol{x} \in \mathcal{M}_0; P_i' \in \mathcal{P}^{s+1},\ \forall i\}. \tag{4.28}$$

By elementary matrix algebra it follows that

$$a_{i,-1}^{(n+1)} = a_{i,-1}^{(n)} + \sum_{j=0}^{s} [P_{n+1}]_{j,-1} a_{ij}^{(n)} \geq a_{i,-1}^{(n)} \tag{4.29}$$

where $a_{ij}^{(n)} = (P_1 P_2 \ldots P_n)_{ij}$ with $P_i \in \mathcal{P}$. It is also known from (4.19) that

$$a_{i,-1}^{(n+s+1)} > a_{i,-1}^{(n)},\ \forall n \geq 0. \tag{4.30}$$

Thus the sequence $\{a_{i,-1}^{(n)}\}_{n\in\mathbb{Z}}$ is non-decreasing for each $i \in S$, and the subsequence $\{a_{i,-1}^{((s+1)n)}\}_{n\in\mathbb{Z}}$ is strictly increasing for each $i \in S$. This completes the proof. □

Therefore, even in this case where potentially very little is known about the behaviour of the Markov chain, so long as it is known that S is finite, that there is a single absorbing state, and that C is a single communicating class with each state aperiodic, absorption in the limit is certain. Clearly in such situations considering $\mathcal{M}_\infty$ is not particularly enlightening.

4.5 Conditioning Upon Non-Absorption

In this section it is proved that under the conditions assumed in this chapter each Markov chain has a *conditionally invariant set of distributions* to which it will tend as time goes to infinity, assuming no distribution over C is ruled out as a possible initial distribution. This conditionally invariant set is a generalisation of the limiting conditional distribution in the precise time-homogeneous case, as well as a generalisation of the set $\tilde{\mathcal{M}}_{\infty}^{C}$ discussed in Chapter 3 (see Definition 3.4.2). We begin by defining this initial distribution set.

Definition 4.5.1 Denote by $\mathcal{M}_0^C$ the set of all probability distributions over the set of transient states C,

$$\mathcal{M}_0^C = \{\boldsymbol{v} = (v_j)_{j=0}^s : v_j \in [0,1],\ \forall j; \sum_{j=0}^{s} v_j = 1\}. \tag{4.31}$$

Next, the set of strictly substochastic matrices (see (1.3)) that describe the behaviour over C is defined. Note that for all $P \in \mathcal{P}$

$$P := \begin{pmatrix} 1 & \mathbf{0} \\ \boldsymbol{p} & Q \end{pmatrix} \tag{4.32}$$

where $\mathbf{0}$ is a row vector with $s+1$ elements, all of which are zero; $\boldsymbol{p}$ is a column vector with $s+1$ elements; and Q is an $(s+1) \times (s+1)$ substochastic matrix, ensuring that transition to state -1 is possible.

Definition 4.5.2 Define the following set of substochastic matrices

$$\mathcal{P}^C = \left\{ Q : \begin{pmatrix} 1 & \mathbf{0} \\ \boldsymbol{p} & Q \end{pmatrix} \in \mathcal{P} \right\}. \tag{4.33}$$

Note that there is a one-to-one correspondence between the elements of $\mathcal{P}$ and the elements of $\mathcal{P}^C$. Hence the following matrix functions can be defined

$$g := P \to Q. \tag{4.34}$$

$$g^{-1} := Q \to P. \tag{4.35}$$

Since $\mathcal{M}_0$ is an infinite set of $(s+2)$-vectors, $\mathcal{M}_0^C$ is an infinite set of $(s+1)$-vectors. The vector function $f : \mathbb{R}^{s+2} \to \mathbb{R}^{s+1}$ is now introduced, where for $\boldsymbol{v} \in \mathcal{M}_n$,

$$f(\boldsymbol{v}) = f((v_{-1}, v_0, \ldots, v_s)) = \frac{1}{1 - v_{-1}}(v_0, v_1, \ldots, v_s). \tag{4.36}$$

In words, the function $f(\cdot)$ takes a distribution over S and conditions it upon the event that the process cannot be in the absorbing state, and hence gives a new distribution over the transient states. Therefore $f(\cdot)$ transforms a distribution over S into a distribution over C. Obviously the one exception to this is the distribution $(1, 0, \ldots, 0)$, but this distribution is not within the function's domain. Therefore

$$\boldsymbol{v} \in \mathcal{M}_0 \Rightarrow f(\boldsymbol{v}) \in \mathcal{M}_0^C. \tag{4.37}$$

Unlike the matrix function $g(\cdot)$, $f(\cdot)$ does not have a unique inverse. The following function is defined instead.

Definition 4.5.3

$$\tilde{f}_\alpha(\boldsymbol{v}) = \tilde{f}_\alpha(v_0, \ldots, v_s) := (\alpha, (1-\alpha)(v_0, \ldots, v_s)) \tag{4.38}$$

where $\alpha \in [0, 1)$.

Obviously this function has the following property

$$f(\tilde{f}_\alpha(\boldsymbol{v})) = \boldsymbol{v}. \tag{4.39}$$

Therefore $\tilde{f}_\alpha(\cdot)$ transforms a distribution over C into a distribution over S by assigning a value α to the probability of absorption.

Definition 4.5.4 For the set of possible initial distributions $\mathcal{D}_0$,

$$\mathcal{D}_n^C := \{f(\boldsymbol{v}) : \boldsymbol{v} \in \mathcal{D}_n\}. \tag{4.40}$$

When $\mathcal{D}_0 = \mathcal{M}_0$ we have

$$\mathcal{M}_n^C := \{f(\boldsymbol{v}) : \boldsymbol{v} \in \mathcal{M}_n\}. \tag{4.41}$$

Every element of the set $\mathcal{M}_n^C$ is a possible distribution at time n, conditioned upon non-absorption. This definition is going to prove problematic if there exists $n < \infty$ such that $(1, 0, \ldots, 0) \in \mathcal{M}_n$, but this is impossible due to Lemma 4.3.2. We therefore have that $\mathcal{M}_n^C$ can be calculated from $\mathcal{M}_n$. It is also possible to calculate $\mathcal{M}_n^C$ from $\mathcal{M}_{n-1}^C$. For each distribution $\boldsymbol{v}$ in $\mathcal{M}_{n-1}^C$ there exists a value α_v such that $\tilde{f}_{\alpha_v}(\boldsymbol{v}) \in \mathcal{M}_{n-1}$. Therefore $\tilde{f}_{\alpha_v}(\boldsymbol{v})P \in \mathcal{M}_n$ for any $P \in \mathcal{P}$, and hence $f(\tilde{f}_{\alpha_v}(\boldsymbol{v})P) \in \mathcal{M}_n^C$. On first consideration, it may appear that determining the value of α_v may be problematic. However, in the following lemma it is proved that, in fact, any value of α can be taken.

Lemma 4.5.1 $f((\alpha, (1-\alpha)\boldsymbol{v})P) = f((\beta, (1-\beta)\boldsymbol{v})P)$ for any $P \in \mathcal{P}$, independently of the values of α and β.

Proof

$$\begin{aligned} f(\tilde{f}_\alpha(\boldsymbol{v})P) &= f\left((\alpha, (1-\alpha)\boldsymbol{v}) \begin{pmatrix} 1 & \mathbf{0} \\ \boldsymbol{p} & Q \end{pmatrix}\right) \\ &= \frac{(1-\alpha)\boldsymbol{v}Q}{|(1-\alpha)\boldsymbol{v}Q|} \\ &= \frac{\boldsymbol{v}Q}{|\boldsymbol{v}Q|}. \end{aligned} \tag{4.42}$$

□

Therefore if two distributions at step n are equal after conditioning upon non-absorption, and are multiplied by the same transition matrix $P \in \mathcal{P}$, then they will be equal after conditioning upon non-absorption at step $n+1$. This leads to the alternative formulation given below

$$\mathcal{M}_n^C = \{f(\tilde{f}_\alpha(\boldsymbol{v})P) | \boldsymbol{v} \in \mathcal{M}_{n-1}^C, P \in \mathcal{P}\} \tag{4.43}$$

If it is decided that not every distribution in $\mathcal{M}_0^C$ is a possible initial distribution for the chain, the set of possible initial distributions over C can be denoted $\mathcal{D}_0^C \subset \mathcal{M}_0^C$.

The following lemma proves two results which will become useful when considering long-term behaviour.

Lemma 4.5.2

$$\mathcal{D}_1^C \subseteq \mathcal{D}_0^C \Rightarrow \mathcal{D}_{n+1}^C \subseteq \mathcal{D}_n^C \tag{4.44}$$

and

$$\mathcal{D}_1^C \supseteq \mathcal{D}_0^C \Rightarrow \mathcal{D}_{n+1}^C \supseteq \mathcal{D}_n^C. \tag{4.45}$$

Proof We prove (4.44) by induction. Assume $\mathcal{D}_n^C \subseteq \mathcal{D}_{n-1}^C$. Then

$$\tilde{f}_\alpha(\mathcal{D}_n^C) \subseteq \tilde{f}_\alpha(\mathcal{D}_{n-1}^C). \tag{4.46}$$

That means

$$\begin{aligned}
\mathcal{D}_{n+1}^C &= \{f(\boldsymbol{v}) : \boldsymbol{v} \in \mathcal{D}_{n+1} \setminus \{(1,0,\ldots,0)\}\} \\
&= \{f(\boldsymbol{v}P) : \boldsymbol{v} \in \mathcal{D}_n \setminus \{(1,0,\ldots,0)\}, P \in \mathcal{P}\} \\
&= \{f(\boldsymbol{v}P) : \boldsymbol{v} \in \tilde{f}_\alpha(\mathcal{D}_n^C) \setminus \{(1,0,\ldots,0)\}, P \in \mathcal{P}\} \\
&\subseteq \{f(\boldsymbol{v}P) : \boldsymbol{v} \in \tilde{f}_\alpha(\mathcal{D}_{n-1}^C) \setminus \{(1,0,\ldots,0)\}, P \in \mathcal{P}\} \\
&= \{f(\boldsymbol{v}) : \boldsymbol{v} \in \tilde{f}_\alpha(\mathcal{D}_n^C) \setminus \{(1,0,\ldots,0)\}\} \\
&= \{f(\boldsymbol{v}) : \boldsymbol{v} \in \mathcal{D}_n \setminus \{(1,0,\ldots,0)\}\} \\
&= \mathcal{D}_n^C
\end{aligned} \tag{4.47}$$

where the third and sixth lines follow from Lemma 4.5.1, and through the use of (4.39). The second part of the lemma is also proved by induction, and uses a very similar progression to that given in (4.47). □

Note that the result above does *not* follow immediately from Lemma 4.3.1, because of course, for example, $\mathcal{D}_1^C \subseteq \mathcal{D}_0^C$ does not imply $\mathcal{D}_1 \subseteq \mathcal{D}_0$.

Corollary 4.5.1

$$\mathcal{M}_{n+1}^C \subseteq \mathcal{M}_n^C \tag{4.48}$$

Proof We have that $\mathcal{M}_1^C \subseteq \mathcal{M}_0^C$. Lemma 4.5.2 completes the proof. □

This next definition is therefore appropriate, and describes the behaviour of the chain, conditioned on non-absorption, as time approaches infinity.

Definition 4.5.5

$$\mathcal{M}_{\infty}^{C} = \cap_{n=0}^{\infty} \mathcal{M}_{n}^{C} \tag{4.49}$$

Definition 4.5.6 $\mathcal{N}$ is described as a *conditionally invariant set of distributions over* C, henceforth known as CISD, if

$$f(\tilde{f}_{\alpha}(\mathcal{N})\mathcal{P}) = \mathcal{N} \tag{4.50}$$

for some $\alpha \in [0, 1)$.

Clearly by (4.49) and Lemma 4.5.1, $\mathcal{M}_{\infty}^{C}$ must be a CISD. This is crucial, because a CISD remains unchanged from time step to time step, once conditioning upon non-absorption has once more been applied. In this sense, it is the logical generalisation of the *quasi-stationary distribution* in the precise, time-homogeneous case. It is now demonstrated that neither $\mathcal{M}_{\infty}^{C}$ nor $\mathcal{M}_{0}^{C} \setminus \mathcal{M}_{\infty}^{C}$ are empty. There are several results that we will require in order to accomplish this. It is known that a quasi-stationary distribution exists for the homogeneous Markov chains generated by each P (Darroch and Seneta [21]). Moreover, this distribution has the property that $\boldsymbol{q}Q = \lambda \boldsymbol{q}$, where λ is the dominating eigenvalue of Q.

Lemma 4.5.3

$$\mathcal{Q} := \{\boldsymbol{q} : \exists Q \in \mathcal{P}^{C} \text{ for which } \boldsymbol{q} \text{ is a QSD to } Q\} \subseteq \mathcal{M}_{\infty}^{C} \tag{4.51}$$

Equivalently, every quasi-stationary distribution which corresponds to the time-homogenous Markov chain with generator $Q \in \mathcal{P}^{C}$ is contained in $\mathcal{M}_{\infty}^{C}$. Moreover, the left-hand set is not empty.

Proof Since $\mathcal{M}_{0}^{C}$ contains all distributions over C, it follows that $\mathcal{Q} \subseteq \mathcal{M}_{0}^{C}$. It is necessary to prove that

$$\mathcal{Q} \subseteq \mathcal{M}_{n}^{C} \Rightarrow \mathcal{Q} \subseteq \mathcal{M}_{n+1}^{C}. \tag{4.52}$$

By definition $\boldsymbol{q} \in \mathcal{Q} \subseteq \mathcal{M}_{n}^{C}$ implies that there exists $\alpha \in [0, 1)$ such that $\tilde{f}_{\alpha}(\boldsymbol{q}) \in \mathcal{M}_{n}$. It also follows that there exists a $Q \in \mathcal{P}^{C}$ such that $\boldsymbol{q}Q = \lambda \boldsymbol{q}$. Since $g^{-1}(Q) \in \mathcal{P}$, we have by Lemma 4.5.1 that

$$f(\tilde{f}_{\alpha}(\boldsymbol{q})g^{-1}(Q)) = \boldsymbol{q} \tag{4.53}$$

and so $\boldsymbol{q} \in \mathcal{M}_{n+1}^C$. Hence $\mathcal{Q} \subseteq \mathcal{M}_{n+1}^C$, as required. □

Thus $\mathcal{M}_\infty^C$ is non-empty. It is now proved that the set $\mathcal{M}_0^C \setminus \mathcal{M}_\infty^C$ is also non-empty.

Lemma 4.5.4 The set $\mathcal{M}_\infty^C$ is a strict subset of $\mathcal{M}_0^C$, that is

$$\mathcal{M}_\infty^C \subset \mathcal{M}_0^C \tag{4.54}$$

and therefore there exists at least one $\boldsymbol{v}$ for which $\boldsymbol{v} \in \mathcal{M}_0^C \setminus \mathcal{M}_\infty^C$ holds.

Proof From (4.4) the set of matrices

$$\{P_1 P_2 \dots P_{s+1} |\ P_i \in \mathcal{P}\} \tag{4.55}$$

is conditionally regular on C. Therefore

$$(P_1 P_2 \dots P_{s+1})_{ij} > 0,\ \forall i \in C, j \in S. \tag{4.56}$$

Consider the set of $(s+1)$-vectors $\{\boldsymbol{e}_i\}$ where

$$\boldsymbol{e}_i = (\delta_{i0}, \dots, \delta_{is}). \tag{4.57}$$

We prove that $\boldsymbol{e}_i$ $(i \geq 0)$ cannot lie within $\mathcal{M}_\infty^C$. Assume that in fact $\boldsymbol{e}_i \in \mathcal{M}_\infty^C$ for some $i \in C$. By the definition of $\mathcal{M}_\infty^C$ it must be the case that

$$(\alpha, (1-\alpha)\boldsymbol{e}_i) \in \mathcal{M}_\infty \tag{4.58}$$

for some $\alpha \in [0, 1)$. Thus by (4.13) and (4.14) for the same value of α

$$(\alpha, (1-\alpha)\boldsymbol{e}_i) \in \mathcal{M}_n,\ \forall n. \tag{4.59}$$

Combining (4.56) with the fact that

$$\mathcal{M}_{s+1} = \{\boldsymbol{v} P_1 P_2 \dots P_{s+1} : \boldsymbol{v} \in \mathcal{M}_0; P_i \in \mathcal{P}\} \tag{4.60}$$

implies that $(\alpha, (1-\alpha)\boldsymbol{e}_i)$ cannot belong to $\mathcal{M}_{s+1}$. This is a contradiction, which thus forces $\{\boldsymbol{e}_i\},\ i \in C$, to lie in the complement of $\mathcal{M}_\infty^C$. □

$\mathcal{M}_\infty^C$ has now been defined, and it has been demonstrated that it is neither empty, nor does it contain every element of $\mathcal{M}_0^C$. We therefore have that some distributions conditioned on non-absorption are possible as time approaches infinity, and that some are impossible, even though we have assumed nothing about the initial distribution over C.

In the following section it is proved that the assumption that the initial distribution could be any honest probability distribution over C is in fact redundant. Rather, it will be shown that as time approaches infinity the chain, conditioned upon non-absorption, will tend to $\mathcal{M}_\infty^C$ independently of the subset of $\mathcal{M}_0^C$ chosen as the set of initial distributions. This is another reason to claim that $\mathcal{M}_\infty^C$ is a generalisation of the concept of the limiting conditional distribution.

4.6 Convergence to Equilibrium

This section begins with a definition for a distance measure[1] which will eventually be employed to prove that for any set of initial distributions over C, denoted $\mathcal{D}_0^C \subseteq \mathcal{M}_0^C$, the sets $\mathcal{D}_n^C$ must converge to $\mathcal{M}_\infty^C$ as time approaches infinity.

4.6.1 Distances Between Sets

In order to demonstrate that as time approaches infinity, the sequence of sets $\mathcal{M}_n^C$ converges to $\mathcal{M}_\infty^C$, a method is required for judging how different two sets are after conditioning their elements upon non-absorption. This will be done by defining a distance measure between two sets of distributions that have been conditioned upon non-absorption. Let $\mathcal{N}$ and $\mathcal{N}'$ be sets of distributions on S. Recall from (4.36) that $f(v_{-1}, v_0, \ldots, v_s) = \frac{1}{1-v_{-1}}(v_0, \ldots, v_s)$.

[1] Technically this potentially an abuse of terminology, since we do not prove this measure obeys the triangle inquality. However, that particular property is not required anywhere in this thesis, and so the function can simply be considered as a "measure of dissimilarity", which when adapted to a function over sets of distributions at time n will be shown to tend to zero as n tends to infinity.

Definition 4.6.1 If the sets $\mathcal{N}$ and $\mathcal{N}'$ are such that

$$f(\mathcal{N}) = f(\mathcal{N}') \tag{4.61}$$

then they are referred to as *conditionally equal on* C, where C is the set of non-absorbing states.

Definition 4.6.2 Several steps are required in order to define a distance between two sets of distributions. We begin by defining a distance between two *individual* distributions. Let $\boldsymbol{v}$ and $\boldsymbol{w}$ be precise distributions on S and let $i \in S$ be such that $w_i > 0$. Define

$$\alpha_{\boldsymbol{v},\boldsymbol{w};i} := \frac{v_i}{w_i}. \tag{4.62}$$

The following equality holds for any distributions $\boldsymbol{v}, \boldsymbol{w}, \boldsymbol{x}$ where w_i and x_i are both non-zero.

$$\alpha_{\boldsymbol{v},\boldsymbol{x};i} = \alpha_{\boldsymbol{v},\boldsymbol{w};i}\alpha_{\boldsymbol{w},\boldsymbol{x};i}. \tag{4.63}$$

We now use the terms above to define a distance measure between two distributions.

Definition 4.6.3 If $w_i > 0$ for every $i \in C$, we let $\underline{\alpha}_{\boldsymbol{v},\boldsymbol{w}} := \min_{i\in C} \alpha_{\boldsymbol{v},\boldsymbol{w};i}$ and $\overline{\alpha}_{\boldsymbol{v},\boldsymbol{w}} := \max_{i\in C} \alpha_{\boldsymbol{v},\boldsymbol{w};i}$. A distance measure between distributions positive on every subset of C is then defined as follows

$$d(\boldsymbol{v}, \boldsymbol{w}) := \frac{\overline{\alpha}_{\boldsymbol{v},\boldsymbol{w}} - \underline{\alpha}_{\boldsymbol{v},\boldsymbol{w}}}{\underline{\alpha}_{\boldsymbol{v},\boldsymbol{w}}}. \tag{4.64}$$

Clearly, $d(\boldsymbol{v}, \boldsymbol{w}) \geq 0$ for any $\boldsymbol{v}$ and $\boldsymbol{w}$ for which it exists, but $d(\boldsymbol{v}, \boldsymbol{w}) = 0$ can hold for different $\boldsymbol{v}$ and $\boldsymbol{w}$, therefore this distance measure does not define a metric space, but rather a pseudometric space (see Steen [65]). The following lemmas demonstrate some important properties of the function $d(\cdot, \cdot)$.

Lemma 4.6.1 Let $\boldsymbol{v}$ and $\boldsymbol{w}$ be distributions that are positive on every subset of C. Then $d(\boldsymbol{v}, \boldsymbol{w}) = 0$ is equivalent to $f(\boldsymbol{v}) = f(\boldsymbol{w})$.

Proof We have that $d(\boldsymbol{v}, \boldsymbol{w}) = 0$ must be equivalent to $\overline{\alpha}_{\boldsymbol{v},\boldsymbol{w}} = \underline{\alpha}_{\boldsymbol{v},\boldsymbol{w}}$. Thus, $\alpha_{\boldsymbol{v},\boldsymbol{w};i} = \dfrac{v_i}{w_i} = \alpha$ for every $i \in C$ where α is a constant independent of i. Since $\boldsymbol{v}$

and $\boldsymbol{w}$ are probability measures, we can define $P_v(A) = \sum_{i\in A} v_i$. For every $A \subseteq S$, then,

$$P_v(A|C) = \frac{P_v(A \cap C)}{P_v(C)} = \frac{\sum_{iA\cap C} v_i}{\sum_{i\in C} v_i} = \frac{\sum_{iA\cap C} \alpha w_i}{\sum_{i\in C} \alpha w_i} = P_w(A|C). \tag{4.65}$$

To prove the implication in the opposite direction let $P_v(A|C) = P_w(A|C)$ for every $A \subseteq S$, or in particular $P_v(i|C) = P_w(i|C)$ for every $i \in C$. For every $i \in C$

$$\alpha_{\boldsymbol{v},\boldsymbol{w};i} = \frac{v_i}{w_i} = \frac{P_v(i|C)P_v(C)}{P_w(i|C)P_w(C)} = \frac{P_v(C)}{P_w(C)} \tag{4.66}$$

which is independent of i. Therefore $\underline{\alpha}_{\boldsymbol{v},\boldsymbol{w}} = \overline{\alpha}_{\boldsymbol{v},\boldsymbol{w}}$, which implies $d(\boldsymbol{v}, \boldsymbol{w}) = 0$. □

Note that $d(\boldsymbol{v}, \boldsymbol{w})$ is continuous in both terms. The following lemma shows that it is also symmetric.

Lemma 4.6.2 Let $\boldsymbol{v}$ and $\boldsymbol{w}$ be probability measures such that $v_i, w_i > 0$ for every $i \in C$. Then $d(\boldsymbol{v}, \boldsymbol{w}) = d(\boldsymbol{w}, \boldsymbol{v})$.

Proof Since $\alpha_{\boldsymbol{v},\boldsymbol{w};i} = \alpha^{-1}_{\boldsymbol{w},\boldsymbol{v};i}$, $\underline{\alpha}_{\boldsymbol{v},\boldsymbol{w}} = \overline{\alpha}^{-1}_{\boldsymbol{w},\boldsymbol{v}}$ and $\overline{\alpha}_{\boldsymbol{v},\boldsymbol{w}} = \underline{\alpha}^{-1}_{\boldsymbol{w},\boldsymbol{v}}$. By substituting these identities into (4.64) the following is obtained

$$d(\boldsymbol{v}, \boldsymbol{w}) = \frac{\overline{\alpha}_{\boldsymbol{v},\boldsymbol{w}} - \underline{\alpha}_{\boldsymbol{v},\boldsymbol{w}}}{\underline{\alpha}_{\boldsymbol{v},\boldsymbol{w}}} = \frac{\underline{\alpha}^{-1}_{\boldsymbol{w},\boldsymbol{v}} - \overline{\alpha}^{-1}_{\boldsymbol{w},\boldsymbol{v}}}{\overline{\alpha}^{-1}_{\boldsymbol{w},\boldsymbol{v}}} = \frac{\overline{\alpha}_{\boldsymbol{w},\boldsymbol{v}} - \underline{\alpha}_{\boldsymbol{w},\boldsymbol{v}}}{\underline{\alpha}_{\boldsymbol{w},\boldsymbol{v}}} = d(\boldsymbol{w}, \boldsymbol{v}). \tag{4.67}$$

□

Having defined a measure for the distance between two distributions, measures are next defined for the distance between a single distribution and a set of distributions, and for the distance between two sets of distributions.

Definition 4.6.4 For a distribution $\boldsymbol{v}$ and a set of distributions $\mathcal{N}$, where all elements in $\mathcal{N}$ are positive on every subset of C, define

$$d(\boldsymbol{v}, \mathcal{N}) = \inf_{\boldsymbol{w}\in\mathcal{N}} d(\boldsymbol{v}, \boldsymbol{w}). \tag{4.68}$$

Furthermore, if $\mathcal{N}'$ is another set of probabilities, define

$$d_{\tilde{H}}(\mathcal{N}', \mathcal{N}) = \sup_{\boldsymbol{v}\in\mathcal{N}'} d(\boldsymbol{v}, \mathcal{N}). \tag{4.69}$$

Note that this is not a symmetric measure.

The above construction of the distance measures between sets of probabilities is analogous to the construction of the Hausdorff measure (see e.g. [63])

$$d_H(X,Y) = \max\left\{\sup_{x\in X}\inf_{y\in Y} d(x,y), \sup_{y\in Y}\inf_{x\in X} d(x,y)\right\} \tag{4.70}$$

where $(\mathcal{N}, d)$ is a non-empty complete metric space and $X, Y \subset \mathcal{N}$ (hence our use of $\tilde{H}$). This is usually used to extend a distance function from a complete metric space to the space of its compact subsets. However, $\mathcal{M}_0^C$ is not a complete space, so the two are not identical.

Lemma 4.6.3 Let $\boldsymbol{v}$ be a distribution and $\mathcal{N}$ a closed set of distributions. Then $d(\boldsymbol{v}, \mathcal{N}) = 0$ if and only if there is a distribution $\boldsymbol{w} \in \mathcal{N}$ such that $f(\boldsymbol{v}) = f(\boldsymbol{w})$. Consequently, $d_{\tilde{H}}(\mathcal{N}', \mathcal{N}) = 0$ if and only if for every $\boldsymbol{v} \in \mathcal{N}'$ there is a $\boldsymbol{w} \in \mathcal{N}$ such that $f(\boldsymbol{v}) = f(\boldsymbol{w})$.

Proof Since $\mathcal{N}$ is a closed subset of the compact set of distributions on a finite space, it is compact. Furthermore, if $\boldsymbol{v}$ is a distribution such that $d(\boldsymbol{v}, \mathcal{N}) = 0$, then there exists a sequence $\{\boldsymbol{w}_n\} \subset \mathcal{N}$ such that $\lim_{n\to\infty} d(\boldsymbol{v}, \boldsymbol{w}_n) = 0$. Because of the compactness of $\mathcal{N}$, this sequence has a subsequence which converges to some $\boldsymbol{w} \in \mathcal{N}$. Lastly, continuity of d implies that $d(\boldsymbol{v}, \boldsymbol{w}) = 0$, and hence $f(\boldsymbol{v}) = f(\boldsymbol{w})$.

Now assume that for every $\boldsymbol{v} \in \mathcal{N}'$ there is a $\boldsymbol{w} \in \mathcal{N}$ such that $f(\boldsymbol{v}) = f(\boldsymbol{w})$. It follows from Lemma 4.6.1 and (4.69) that $d(\boldsymbol{v}, \mathcal{N}) = 0$, for all $\boldsymbol{v} \in \mathcal{N}'$. Hence from (4.69) we have $d_{\tilde{H}}(\mathcal{N}', \mathcal{N}) = 0$. □

Corollary 4.6.1 Let $\mathcal{N}$ and $\mathcal{N}'$ be closed sets of distributions. Then $f(\mathcal{N}) \subseteq f(\mathcal{N}')$ if and only if $d_{\tilde{H}}(\mathcal{N}, \mathcal{N}') = 0$.

Proof Assume first that $f(\mathcal{N}) \subseteq f(\mathcal{N}')$. Then for every $\boldsymbol{v} \in \mathcal{N}$ there exists a $\boldsymbol{w} \in \mathcal{N}'$ such that $f(\boldsymbol{v}) = f(\boldsymbol{w})$. Thus $d(\boldsymbol{v}, \boldsymbol{w}) = 0$, which implies $d(\boldsymbol{v}, \mathcal{N}') = 0$ and hence that $d_{\tilde{H}}(\mathcal{N}, \mathcal{N}') = 0$. To prove the result in the opposite direction, let $d_{\tilde{H}}(\mathcal{N}, \mathcal{N}') = 0$. By Lemma 4.6.3, for every $\boldsymbol{v} \in \mathcal{N}$ there exists $\boldsymbol{w} \in \mathcal{N}'$ such that $d(\boldsymbol{v}, \boldsymbol{w}) = 0$. This gives $f(\boldsymbol{v}) = f(\boldsymbol{w})$, which consequently implies $f(\mathcal{N}) \subseteq f(\mathcal{N}')$. □

The value of this measure for the distance between sets is now becoming apparent. From Lemma 4.6.2, if every element of $\mathcal{N}$ and of $\mathcal{N}'$ contains no zeros, then Corollary 4.6.1 gives us $f(\mathcal{N}) = f(\mathcal{N}') \Leftrightarrow d_{\tilde{H}}(\mathcal{N}, \mathcal{N}') = d_{\tilde{H}}(\mathcal{N}', \mathcal{N}) = 0$. Note that by the condition of regularity given earlier, there exists n_0 so that every element of every set $\mathcal{M}_n^C$ is non-zero for $n \geq n_0$. It is now shown that two closed sets of distributions that are conditionally equal are always equally far away from any other closed set of distributions.

Lemma 4.6.4 For closed sets of distributions $\mathcal{N}$ and $\mathcal{N}'$, which are conditionally equal on C, and any other closed set of distributions $\mathcal{L}$,

$$d_{\tilde{H}}(\mathcal{N}, \mathcal{L}) = d_{\tilde{H}}(\mathcal{N}', \mathcal{L}) \tag{4.71}$$

and

$$d_{\tilde{H}}(\mathcal{L}, \mathcal{N}) = d_{\tilde{H}}(\mathcal{L}, \mathcal{N}'). \tag{4.72}$$

Proof The first part of the lemma is proved by showing that $d(\boldsymbol{v}, \boldsymbol{w}) = 0$ implies $d(\boldsymbol{v}, \boldsymbol{x}) = d(\boldsymbol{w}, \boldsymbol{x})$ for all probability distributions $\boldsymbol{w}$ and $\boldsymbol{x}$ which have no zero elements (thus allowing the distances to be well-defined). In other words, if $\boldsymbol{v}$ and $\boldsymbol{w}$ have no distance between them, then the distance from $\boldsymbol{v}$ to $\boldsymbol{x}$ must be the same as the distance from $\boldsymbol{w}$ to $\boldsymbol{x}$. Note that $d(\boldsymbol{v}, \boldsymbol{w}) = 0$ implies that

$$\alpha_{\boldsymbol{w},\boldsymbol{x};i} = \alpha_{\boldsymbol{w},\boldsymbol{v};i}\alpha_{\boldsymbol{v},\boldsymbol{x};i} = \underline{\alpha}_{\boldsymbol{w},\boldsymbol{v}}\alpha_{\boldsymbol{v},\boldsymbol{x};i} \tag{4.73}$$

since $\alpha_{\boldsymbol{w},\boldsymbol{v};i} = \underline{\alpha}_{\boldsymbol{w},\boldsymbol{v}}$ for every $i \in C$. This implies that $\underline{\alpha}_{\boldsymbol{w},\boldsymbol{x}} = \underline{\alpha}_{\boldsymbol{w},\boldsymbol{v}}\underline{\alpha}_{\boldsymbol{v},\boldsymbol{x}}$ and $\overline{\alpha}_{\boldsymbol{w},\boldsymbol{x}} = \underline{\alpha}_{\boldsymbol{w},\boldsymbol{v}}\overline{\alpha}_{\boldsymbol{v},\boldsymbol{x}}$. Substituting these identities into (4.64) produces

$$d(\boldsymbol{w}, \boldsymbol{x}) = \frac{\overline{\alpha}_{\boldsymbol{w},\boldsymbol{x}} - \underline{\alpha}_{\boldsymbol{w},\boldsymbol{x}}}{\underline{\alpha}_{\boldsymbol{w},\boldsymbol{x}}} = \frac{\underline{\alpha}_{\boldsymbol{w},\boldsymbol{v}}\overline{\alpha}_{\boldsymbol{v},\boldsymbol{x}} - \underline{\alpha}_{\boldsymbol{w},\boldsymbol{v}}\underline{\alpha}_{\boldsymbol{v},\boldsymbol{x}}}{\underline{\alpha}_{\boldsymbol{w},\boldsymbol{v}}\underline{\alpha}_{\boldsymbol{v},\boldsymbol{x}}} = \frac{\overline{\alpha}_{\boldsymbol{v},\boldsymbol{x}} - \underline{\alpha}_{\boldsymbol{v},\boldsymbol{x}}}{\underline{\alpha}_{\boldsymbol{v},\boldsymbol{x}}} = d(\boldsymbol{v}, \boldsymbol{x}). \tag{4.74}$$

It is now shown that for all closed sets of distributions $\mathcal{L}$ and any distributions $\boldsymbol{v}$ and $\boldsymbol{w}$ such that $d(\boldsymbol{v}, \boldsymbol{w}) = 0$, $d(\boldsymbol{v}, \mathcal{L}) = d(\boldsymbol{w}, \mathcal{L})$. This follows from

$$d(\boldsymbol{v}, \mathcal{L}) = \inf_{\boldsymbol{x} \in \mathcal{L}} d(\boldsymbol{v}, \boldsymbol{x}) = \inf_{\boldsymbol{x} \in \mathcal{L}} d(\boldsymbol{w}, \boldsymbol{x}) = d(\boldsymbol{w}, \mathcal{L}) \tag{4.75}$$

where the second equality follows from (4.74) and the first and third follow from Definition 4.6.4. Finally, let $\mathcal{N}, \mathcal{N}'$ and $\mathcal{L}$ be closed sets of distributions such that

all elements of each distribution in $\mathcal{L}$ are non-zero. It follows that

$$d_{\tilde{H}}(\mathcal{N},\mathcal{L}) = \sup_{\boldsymbol{v}\in\mathcal{N}} d(\boldsymbol{v},\mathcal{L}) \leq \sup_{\boldsymbol{w}\in\mathcal{N}'} d(\boldsymbol{w},\mathcal{L}) = d_{\tilde{H}}(\mathcal{N}',\mathcal{L}) \tag{4.76}$$

where the last inequality follows from Lemma 4.6.3, which for every $\boldsymbol{v} \in \mathcal{N}$ shows the existence of a distribution $\boldsymbol{w} \in \mathcal{N}'$ such that $d(\boldsymbol{v},\boldsymbol{w}) = 0$. Because of the symmetry between $\mathcal{N}$ and $\mathcal{N}'$ the opposite inequality also holds, and therefore the first equality in the lemma is proved.

The second equality $d(\mathcal{L},\mathcal{N}) = d(\mathcal{L},\mathcal{N}')$ will be proved using a method very similar to that of the first part. The crucial step is once more to show that $d(\boldsymbol{v},\boldsymbol{w}) = 0$ implies $d(\boldsymbol{x},\boldsymbol{v}) = d(\boldsymbol{x},\boldsymbol{w})$. Note that

$$\alpha_{\boldsymbol{x},\boldsymbol{w};i} = \alpha_{\boldsymbol{x},\boldsymbol{v};i}\alpha_{\boldsymbol{v},\boldsymbol{w};i} = \alpha_{\boldsymbol{x},\boldsymbol{v};i}\underline{\alpha}_{\boldsymbol{v},\boldsymbol{w}}. \tag{4.77}$$

Thus it can be derived that $d(\boldsymbol{x},\boldsymbol{w}) = d(\boldsymbol{x},\boldsymbol{v})$.

We have therefore that $d(\boldsymbol{x},\mathcal{N}) = \inf_{\boldsymbol{v}\in\mathcal{N}} d(\boldsymbol{x},\boldsymbol{v})$ for every $\boldsymbol{x}$, and since for every $\boldsymbol{v} \in \mathcal{N}$ there exists $\boldsymbol{w} \in \mathcal{N}'$ such that $d(\boldsymbol{v},\boldsymbol{w}) = 0$, this infimum must be smaller or equal to $\inf_{\boldsymbol{w}\in\mathcal{N}'} d(\boldsymbol{x},\boldsymbol{w}) = d(\boldsymbol{x},\mathcal{N}')$. Therefore, $d(\boldsymbol{x},\mathcal{N}) \leq d(\boldsymbol{x},\mathcal{N}')$ and, by symmetry, also $d(\boldsymbol{x},\mathcal{N}') \leq d(\boldsymbol{x},\mathcal{N})$. Thus $d(\boldsymbol{x},\mathcal{N}) = d(\boldsymbol{x},\mathcal{N}')$. Finally, note

$$d_{\tilde{H}}(\mathcal{L},\mathcal{N}) = \sup_{\boldsymbol{x}\in\mathcal{L}} d(\boldsymbol{x},\mathcal{N}) = \sup_{\boldsymbol{x}\in\mathcal{L}} d(\boldsymbol{x},\mathcal{N}') = d_{\tilde{H}}(\mathcal{L},\mathcal{N}') \tag{4.78}$$

which completes the proof. □

Lemma 4.6.5 Let $\boldsymbol{v}$ be a distribution and $\mathcal{N}$ a closed set of distributions. Then there exists a distribution $\boldsymbol{w} \in \mathcal{N}$ such that $d(\boldsymbol{v},\boldsymbol{w}) = d(\boldsymbol{v},\mathcal{N})$.

Proof A similar argument as in the proof of Lemma 4.6.3 is applied. Since there is a sequence of distributions $\{\boldsymbol{w}_n\} \subset \mathcal{N}$ such that $\lim_{n\to\infty} d(\boldsymbol{v},\boldsymbol{w}_n) = d$, the limit $\boldsymbol{w}$ of a convergent subsequence must be the probability required. □

We now have a comprehensive method of judging the similarity of two closed sets of conditional distributions. In Subsection 4.6.2, this method will be applied to demonstrate that for a given Markov chain there is in fact only one corresponding conditionally invariant set of distributions.

4.6.2 Fixed Sets

It will be proved in this subsection that $\mathcal{M}_\infty^C$ is the only set $\mathcal{N}$ with the property that

$$f(\tilde{f}_\alpha(\mathcal{N})\mathcal{P}) = \mathcal{N} \tag{4.79}$$

where, as always, $\mathcal{P}$ is the set of transition matrices. In other words, $\mathcal{M}_\infty^C$ is the only set that is conditionally invariant.

What follows will frequently require that no element of any transition matrix below the first row is equal to zero. Obviously, that is not necessarily the case from one time step to the next. This dilemma is solved by once again making use of $\mathcal{P}^{s+1}$ (see Definition 4.2.2). Later in this section it shall be shown that the long-term behaviour of $\mathcal{P}$ and of $\mathcal{P}^{s+1}$ is identical.

The smallest possible element P_{ij}, where $i, j \in C$, of any such matrix will usually be denoted by $\underline{m} > 0$. It is first proved that this constant does in fact exist.

Lemma 4.6.6 There exists a constant $\underline{m} > 0$ such that

$$P \in \mathcal{P}^{s+1} \Rightarrow P_{ij} \geq \underline{m} \tag{4.80}$$

for all $i, j \in C$.

Proof Define $\underline{P}$ such that $\underline{P}_{ij} = \min_j\{r_j^{(i)} : \boldsymbol{r}^{(i)} \in \mathcal{R}^{(i)}\}$ (this minimum exists since $\mathcal{R}^{(i)}$ is a closed set for all i). It follows that $\underline{P} \leq P$ for any $P \in \mathcal{P}$, and hence that $(\underline{P})^{s+1} \leq P'$ for any $P' \in \mathcal{P}^{s+1}$. Set $\underline{m} := \min_{i,j\in C}(\underline{P})_{ij}^{s+1}$. This proves that (4.80) holds. It remains to be proved that $\underline{m} > 0$, this is done as follows. Note that for each $P \in \mathcal{P}$

$$P = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ \underline{c}_{0,-1} + \alpha_{0,-1} & \underline{c}_{00} + \alpha_{00} & \underline{c}_{01} + \alpha_{01} & \dots & \underline{c}_{0s} + \alpha_{0s} \\ \underline{c}_{1,-1} + \alpha_{1,-1} & \underline{c}_{10} + \alpha_{10} & \underline{c}_{11} + \alpha_{11} & \dots & \underline{c}_{1s} + \alpha_{1s} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \underline{c}_{s,-1} + \alpha_{s,-1} & \underline{c}_{s0} + \alpha_{s0} & \underline{c}_{s1} + \alpha_{s1} & \dots & \underline{c}_{ss} + \alpha_{ss} \end{pmatrix} \tag{4.81}$$

where $0 \leq \alpha_{ij} \leq \overline{c}_{ij} - \underline{c}_{ij}$. By the assumption that each P must represent a Markov chain for which C is a single communicating class, we have that for any $P^{(s+1)} =$

$P_1 P_2 \dots P_{s+1}$, where $P_i \in \mathcal{P}$, every element of $P^{(s+1)}$ below the first row is strictly positive. It is also the case, from the assumptions stated in Section 4.2, that if it is possible for $\alpha_{ij} > 0$ to hold then $\underline{c}_{ij} > 0$, since all jumps must be either always possible, or always impossible. It therefore follows that every element of P^{s+1} is strictly positive beneath the first row, since for any element sum involved in matrix multiplication, α_{ij} can only make a positive contribution if $\underline{c}_{ij}$ makes one too. □

The following results guarantee that all probabilities in sets $\mathcal{M}_1^C, \mathcal{M}_2^C, \dots$ are bounded from below by a constant $\underline{n}$.

Corollary 4.6.2 Let $P \in \mathcal{P}$ be denoted by

$$P = \begin{pmatrix} \boldsymbol{P}_{-1} \\ \boldsymbol{P}_0 \\ \vdots \\ \boldsymbol{P}_s \end{pmatrix}. \tag{4.82}$$

Let $\underline{n} > 0$ be a constant such that $(f(\boldsymbol{v}))_i \geq \underline{n}$ for all $i \in C$ and for every $\boldsymbol{v} \in \boldsymbol{P}_k,\ k \geq 0$. Let $\boldsymbol{w}^+ \in \mathcal{M}_n$ for some $n > 0$. Then $(f(\boldsymbol{w}^+))_i \geq \underline{n}$ for all $i \in C$.

Proof The assumption of the corollary is that $\frac{P_{ij}}{\sum_{j \in C} P_{ij}} \geq \underline{n}$. Let $\boldsymbol{w}^+ = \boldsymbol{w}P$. Then

$$\begin{aligned} (f(\boldsymbol{w}^+))_j &= \frac{\sum_{i \in C} w_i P_{ij}}{\sum_{k \in C} \sum_{i \in C} w_i P_{ik}} \\ &= \frac{\sum_{i \in C} w_i P_{ij}}{\sum_{i \in C} w_i \sum_{k \in C} P_{ik}} \\ &\geq \frac{\sum_{i \in C} w_i \underline{n} \sum_{k \in C} P_{ik}}{\sum_{i \in C} w_i \sum_{k \in C} P_{ik}} \\ &= \underline{n}. \end{aligned} \tag{4.83}$$

□

It follows from the above corollary that conditional probabilities in every fixed set must be greater or equal to $\underline{n}$ on every non-empty subset of C.

The following lemma shows that multiplication with a conditionally regular set of transition matrices is a contraction with respect to the distance function d.

Lemma 4.6.7 Let $\boldsymbol{v}$ and $\boldsymbol{w}$ be distributions on S and let the set $\mathcal{K}$ contain matrices whose entries of the form P_{ij}, $i, j \in C$, are greater or equal to $< 0\underline{m} < \frac{1}{2}$. Assume also that $(f(\boldsymbol{w}))_i \geq \underline{m}$ for every $i \in C$. Take an arbitrary $P \in \mathcal{K}$ and denote $\boldsymbol{v}^+ = \boldsymbol{v}P$ and $\boldsymbol{w}^+ = \boldsymbol{w}P$. Then

$$d(\boldsymbol{v}^+, \boldsymbol{w}^+) \leq (1 - 2\underline{m}^2) d(\boldsymbol{v}, \boldsymbol{w}) \tag{4.84}$$

for $d(\cdot, \cdot)$ as given by Definition 4.6.3.

Note that since each distribution $\boldsymbol{w}$ has at least two elements (assuming we are not in the trivial case with only one non-absorbing state), it is impossible for $w_i \geq \underline{m}$ for all i if $\underline{m} > \frac{1}{2}$. Thus under the conditions of the corollary, $(1 - 2\underline{m}^2)$ is always positive.

Proof Take any $j \in C$ and calculate

$$\begin{aligned} \alpha_{\boldsymbol{v}^+, \boldsymbol{w}^+; j} &= \frac{\sum_{i \in C} v_i P_{ij}}{\sum_{i \in C} w_i P_{ij}} \\ &= \frac{\sum_{i \in C} \alpha_{\boldsymbol{v}, \boldsymbol{w}; i} w_i P_{ij}}{\sum_{i \in C} w_i P_{ij}} \\ &= \sum_{i \in C} \alpha_{\boldsymbol{v}, \boldsymbol{w}; i} \frac{w_i P_{ij}}{\sum_{k \in C} w_k P_{kj}}. \end{aligned} \tag{4.85}$$

Note that (4.85) is a convex combination of elements $\alpha_{\boldsymbol{v}, \boldsymbol{w}; i}$ with coefficients $\frac{w_i P_{ij}}{\sum_{k \in C} w_k P_{kj}} \geq \underline{m}^2$. Thus, $\underline{m}^2 \overline{\alpha}_{\boldsymbol{v}, \boldsymbol{w}} + (1 - \underline{m}^2) \underline{\alpha}_{\boldsymbol{v}, \boldsymbol{w}} \leq \alpha_{\boldsymbol{v}^+, \boldsymbol{w}^+; j} \leq (1 - \underline{m}^2) \overline{\alpha}_{\boldsymbol{v}, \boldsymbol{w}} + \underline{m}^2 \underline{\alpha}_{\boldsymbol{v}, \boldsymbol{w}}$. Using this and $\underline{\alpha}_{\boldsymbol{v}^+, \boldsymbol{w}^+} \geq \underline{\alpha}_{\boldsymbol{v}, \boldsymbol{w}}$ it follows that

$$\begin{aligned} d(\boldsymbol{v}^+, \boldsymbol{w}^+) &= \frac{\overline{\alpha}_{\boldsymbol{v}^+, \boldsymbol{w}^+} - \underline{\alpha}_{\boldsymbol{v}^+, \boldsymbol{w}^+}}{\underline{\alpha}_{\boldsymbol{v}^+, \boldsymbol{w}^+}} \\ &\leq \frac{(1 - \underline{m}^2) \overline{\alpha}_{\boldsymbol{v}, \boldsymbol{w}} + \underline{m}^2 \underline{\alpha}_{\boldsymbol{v}, \boldsymbol{w}} - \underline{m}^2 \overline{\alpha}_{\boldsymbol{v}, \boldsymbol{w}} - (1 - \underline{m}^2) \underline{\alpha}_{\boldsymbol{v}, \boldsymbol{w}}}{\underline{\alpha}_{\boldsymbol{v}, \boldsymbol{w}}} \\ &= (1 - 2\underline{m}^2) \frac{\overline{\alpha}_{\boldsymbol{v}, \boldsymbol{w}} - \underline{\alpha}_{\boldsymbol{v}, \boldsymbol{w}}}{\underline{\alpha}_{\boldsymbol{v}, \boldsymbol{w}}} \\ &= (1 - 2\underline{m}^2) d(\boldsymbol{v}, \boldsymbol{w}). \end{aligned} \tag{4.86}$$

□

Corollary 4.6.3 Let $\mathcal{N}$ and $\mathcal{N}'$ be closed sets of distributions such that for every $\boldsymbol{w} \in \mathcal{N}$, $w_i \geq \underline{m}$ for every $i \in C$; and let $\mathcal{K}$ be as in Lemma 4.6.7. Denote $\mathcal{N}'^+ = \mathcal{N}'\mathcal{K}$ and $\mathcal{N}^+ = \mathcal{N}\mathcal{K}$. Then

$$d_{\tilde{H}}(\mathcal{N}'^+, \mathcal{N}^+) \leq (1 - 2\underline{m}^2) d_{\tilde{H}}(\mathcal{N}', \mathcal{N}). \tag{4.87}$$

Proof Denote $d = d_{\tilde{H}}(\mathcal{N}', \mathcal{N})$ and take any $\boldsymbol{v}^+ = \boldsymbol{v}P$ where $\boldsymbol{v} \in \mathcal{N}'$ and $P \in \mathcal{K}$. Hence $\boldsymbol{v}^+ \in \mathcal{N}'^+$. Then there is a $\boldsymbol{w} \in \mathcal{N}$ such that $d(\boldsymbol{v}, \boldsymbol{w}) \leq d$. Denote $\boldsymbol{w}^+ = \boldsymbol{w}P$. By Lemma 4.6.7, $d(\boldsymbol{v}^+, \boldsymbol{w}^+) \leq (1 - 2\underline{m}^2)d(\boldsymbol{v}, \boldsymbol{w}) \leq (1 - 2\underline{m}^2)d$. Therefore, for every $\boldsymbol{v}^+ \in \mathcal{N}'^+$ we have $d(\boldsymbol{v}^+, \mathcal{N}^+) \leq (1 - 2\underline{m}^2)d$ and consequently $d_{\tilde{H}}(\mathcal{N}'^+, \mathcal{N}^+) = \sup_{\boldsymbol{v}^+ \in \mathcal{N}'^+} d(\boldsymbol{v}, \mathcal{N}^+) \leq (1 - 2\underline{m}^2)d$, which proves the corollary. □

This corollary demonstrates that, as long as $\underline{m} > 0$, the distance between two closed sets of distributions decreases with every successive multiplication by $\mathcal{P}$. Of course, the value of $\underline{m}$ is determined by $\mathcal{P}$, and it has already been proved that a value of $\underline{m} > 0$ can be found after at most $s + 1$ time steps.

Definition 4.6.5 Let $\mathcal{N}$ be a compact set of distributions on S and $\mathcal{K}$ a set of transition matrices that are conditionally regular on C. Then $\mathcal{N}$ is a *conditionally fixed set of* $\mathcal{K}$ if $f(\mathcal{N}\mathcal{K}) = f(\mathcal{N})$, or equivalently, if $\mathcal{N}$ and $\mathcal{N}\mathcal{K}$ are conditionally equal on C.

It immediately follows that if $\mathcal{N}$ is a conditionally fixed set of $\mathcal{P}$, then $f(\mathcal{N})$ is a conditionally invariant set of distributions (note the difference between the two terms). Therefore, if it can be shown that all conditionally fixed sets of $\mathcal{P}$ are conditionally equal, it follows that there can be only one conditionally invariant set of distributions. The first stage in proving this result is the theorem below.

Theorem 4.6.1 Assume every matrix in $\mathcal{P}$ has every element below the first row strictly positive. Let $\mathcal{N}$ and $\mathcal{N}'$ be conditionally fixed sets of $\mathcal{P}$. Then they are conditionally equal on C.

Proof It follows from Corollary 4.6.1 that the sets $\mathcal{N}$ and $\mathcal{N}'$ are conditionally equal on C if and only if $d_{\tilde{H}}(\mathcal{N}, \mathcal{N}') = d_{\tilde{H}}(\mathcal{N}', \mathcal{N}) = 0$. Suppose that one of the distances is greater than 0, say $d_{\tilde{H}}(\mathcal{N}, \mathcal{N}') > 0$. Then, by Lemma 4.6.4 and Corollary 4.6.3, $d_{\tilde{H}}(\mathcal{N}, \mathcal{N}') = d_{\tilde{H}}(\mathcal{N}\mathcal{P}, \mathcal{N}') = d_{\tilde{H}}(\mathcal{N}\mathcal{P}, \mathcal{N}'\mathcal{P}) \leq (1 - 2\underline{m}^2)d_{\tilde{H}}(\mathcal{N}, \mathcal{N}')$, which is a contradiction, since $\underline{m} > 0$. □

From Lemma 4.6.6 we have that each element of any matrix in $\mathcal{P}^{s+1}$ is positive below the first row. Therefore all conditionally fixed sets of $\mathcal{P}^{s+1}$ are conditionally

equal. The relationship between the conditionally fixed sets of $\mathcal{P}$ and the conditionally fixed sets of $\mathcal{P}^{s+1}$ is now given.

Lemma 4.6.8 If $\mathcal{N}$ is a conditionally fixed set of $\mathcal{P}$, it is also a conditionally fixed set of $\mathcal{P}^{s+1}$.

Proof It is proved first that $\mathcal{N}$ is a conditionally fixed set of $\mathcal{P}^2$ (see 4.2.2). Note that by Lemma 4.5.1 it follows that for $P_1, P_2 \in \mathcal{P}$,

$$f(\tilde{f}_\alpha(\boldsymbol{v})P_1P_2) = f(\tilde{f}_\beta(\boldsymbol{v})P_1P_2), \tag{4.88}$$

since $\mathcal{P}^2$ is a set of transition matrices corresponding to a Markov chain with a single absorbing state, and a set of transient states C which is a single communicating class with all states aperiodic. Let $\tilde{f}_\alpha(\boldsymbol{v}P_1) = \boldsymbol{v}_\alpha$ and $\tilde{f}_\beta(\boldsymbol{v}P_1) = \boldsymbol{v}_\beta$. By the fact that $\mathcal{N}$ is a conditionally fixed set of $\mathcal{P}$ we have that $f(\boldsymbol{v}_\alpha) \in \mathcal{N}$ and $f(\boldsymbol{v}_\beta) \in \mathcal{N}$. Hence, once again making use of Lemma 4.5.1, it follows that

$$f(\boldsymbol{v}_\alpha P_2) = f(\boldsymbol{v}_\beta P_2). \tag{4.89}$$

Hence $\mathcal{N}$ is a conditionally fixed set of $\mathcal{P}^2$. The proof follows by induction. Assume $\mathcal{N}$ is a conditionally fixed set of both $\mathcal{P}$ and $\mathcal{P}^n$. For $P_1 \in \mathcal{P}$ and $P_n \in \mathcal{P}^n$ it follows from Lemma 4.5.1 that

$$f(\tilde{f}_\alpha(\boldsymbol{v})P_1P_n) = f(\tilde{f}_\beta(\boldsymbol{v})P_1P_n). \tag{4.90}$$

Using Lemma 4.5.1 one last time gives us

$$f(\boldsymbol{v}_\alpha P_n) = f(\boldsymbol{v}_\beta P_n). \tag{4.91}$$

Hence $\mathcal{N}$ is a conditionally fixed set of $\mathcal{P}^{n+1}$, as needed. □

Thus any conditionally fixed set of $\mathcal{P}$ is also a conditionally fixed set of $\mathcal{P}^{s+1}$. Since we already know that all conditionally fixed sets of $\mathcal{P}^{s+1}$ are conditionally equal, the corollary below quickly follows.

Corollary 4.6.4 All conditionally fixed sets of $\mathcal{P}$ are conditionally equal.

Proof Let N and L be conditionally fixed sets of $\mathcal{P}$. By Lemma 4.6.8 they must also be conditionally fixed sets of $\mathcal{P}^{s+1}$, and therefore by Theorem 4.6.1 they must be conditionally equal. □

Corollary 4.6.4 proves that all conditionally fixed sets for $\mathcal{P}$ are conditionally equal, which of course means that for every such set $\mathcal{N}$, it holds that $f(\mathcal{N}) = \mathcal{M}_\infty^C$. Therefore $\mathcal{M}_\infty^C$ is the only set that a set of initial distributions over C can converge to as time approaches infinity, conditioned upon non-absorption. All that remains to be proved is that all such sets of initial distributions do indeed converge. This is done with the results below. To prove this convergence one more definition is introduced.

Definition 4.6.6 Denote by T_n the union of all possible distributions, conditioned upon non-absorption over steps 0 to n, that is

$$T_n = \overline{\bigcup_{i=0}^{n} D_i^C}. \tag{4.92}$$

Further

$$T_\infty = \overline{\bigcup_{i=0}^{\infty} D_i^C} \tag{4.93}$$

for the set of initial distributions D_0^C.

Note that

$$\overline{T}_\infty \supseteq \mathcal{Q} \tag{4.94}$$

where $\mathcal{Q}$ is the set of quasi-stationary distributions for the elements in $\mathcal{P}$. The theory in this chapter concludes with a final theorem, which proves that setting $\mathcal{D}_0^C$ equal to the set of all quasi-stationary distributions for the matrices in $\mathcal{P}$ will result in convergence to $\mathcal{M}_\infty^C$, followed by a corollary which proves that convergence to $\mathcal{M}_\infty^C$ will occur irrespective of $\mathcal{D}_0^C$.

Theorem 4.6.2 Let $\mathcal{Q}$ be the set of quasi-stationary distributions for the elements in $\mathcal{P}$. Set $D_0^C = \mathcal{Q}$. Then $D_\infty^C = M_\infty^C$.

Proof $D_0^C = \mathcal{Q}$ implies that $D_0^C \subseteq D_1^C$, since for every element $\boldsymbol{q} \in \mathcal{Q}$ there is a $P \in \mathcal{P}$ such that $f(\tilde{f}_\alpha(\boldsymbol{q}P)) = \boldsymbol{q}$. Thus by Lemma 4.5.2, $D_n^C \subseteq D_{n+1}^C$, $\forall n \geq 0$, and hence $\bigcup_{i=0}^n D_i^C = D_n^C$. Lemma 4.5.3 gives us that $\mathcal{Q} \subseteq \mathcal{M}_\infty^C$. Therefore, since $\mathcal{M}_\infty^C$ is a conditionally invariant set, $D_n^C \subseteq \mathcal{M}_\infty^C$ holds for all n. Thus the sequence of increasing sets $\mathcal{D}_n^C$ is bounded from above. Next, we demonstrate that the union of $D_\infty^C = \bigcup_{i=0}^\infty D_i^C$, and that its closure, denoted by $\mathcal{D}$, is a conditionally invariant set of distributions, and therefore that $D_\infty^C = \mathcal{M}_\infty^C$ by Corollary 4.6.4.

Since $\boldsymbol{v} \in D_\infty^C \Rightarrow \boldsymbol{v} \in D_n^C$ for some n, we have that for any $P \in \mathcal{P}$,

$$\boldsymbol{v} \in D_\infty^C \Rightarrow f(\tilde{f}_\alpha(\boldsymbol{v})P) \in D_\infty^C \tag{4.95}$$

since $f(\tilde{f}_\alpha(\boldsymbol{v})P) \subseteq D_{n+1}^C \subseteq D_\infty^C$. Hence for any $\boldsymbol{v}' \in D_n^C$ there exists $P \in \mathcal{P}$ and $\boldsymbol{w} \in D_{n-1}^C$ such that $f(\tilde{f}_\alpha(\boldsymbol{w})P) = \boldsymbol{v}'$. Since $D_{n-1}^C \subseteq D_\infty^C$ and $D_n^C \subseteq D_\infty^C$, it follows that

$$f(\tilde{f}_\alpha(D_\infty^C)\mathcal{P}) = D_\infty^C. \tag{4.96}$$

Therefore $\mathcal{D}_\infty^C$ has the property of a conditionally invariant set of distributions.

We now consider $\mathcal{D}$. Since both it and $\mathcal{P}$ are closed sets, proving that $\mathcal{D}$ is a conditionally invariant set of distributions requires only that

$$f(\tilde{f}_\alpha(\mathcal{D})\mathcal{P}) \subseteq D_\infty^C \cup \mathcal{D} \tag{4.97}$$

holds. This in conjunction with (4.96) will force

$$f(\tilde{f}_\alpha(D_\infty^C \cup \mathcal{D})\mathcal{P}) = D_\infty^C \cup \mathcal{D} \tag{4.98}$$

as needed.

We now prove that (4.97) holds. By the definition of the closure of an open set, it must be the case that for every distribution $\boldsymbol{v} \in \mathcal{D}$ there exists a sequence of vectors $\{\boldsymbol{v}_n\}_{n\in\mathbb{N}}$ such that $\boldsymbol{v}_n \in D_\infty^C$, $\forall n \geq 0$ and $\lim_{n\to\infty} \boldsymbol{v}_n = \boldsymbol{v}$. This means in particular that for any $\epsilon > 0$ there exists an $n_0 \geq 0$ such that $|v_i - (v_{n_0})_i| < \epsilon$ for all $n \geq n_0$ and for all $i \in C$.

Let us assume that there exist $\boldsymbol{v} \in \mathcal{D}$ and $P \in \mathcal{P}$ such that $f(\tilde{f}_\alpha(\boldsymbol{v})P)$ does not lie within $D_\infty^C \cup \mathcal{D}$. It follows that there exists a perpendicular distance between the

distribution $f(\tilde{f}_\alpha(\boldsymbol{v})P)$ and the set $D_\infty^C \cup \mathcal{D}$. This distance is denoted a. However, from (4.95) it is known that $f(\tilde{f}_\alpha(\boldsymbol{v}_n)P) \in D_\infty^C \cup \mathcal{D}$, for all n. Thus

$$\| f(\tilde{f}_\alpha(\boldsymbol{v})P) - f(\tilde{f}_\alpha(\boldsymbol{v}_n)P) \|^2 \geq a^2, \tag{4.99}$$

where $\| \cdot \|$ denotes the standard Euclidean distance. This contradicts continuity.

Thus, as claimed, $D_\infty^C \cup \mathcal{D}$ is a conditionally invariant set, as claimed, which completes the proof. □

Corollary 4.6.5 For any set of initial distributions D_0^C, $D_\infty^C = M_\infty^C$.

Proof We have from (4.94) that $\overline{T}_\infty \supseteq \mathcal{Q}$. This combined with Theorem 4.6.1 gives the result. □

From combining results from this section, then, it can be proved that for a given set $\mathcal{P}$ there is only one conditionally invariant set of distributions, and that the long-term behaviour of the process, conditioned upon non-absorption, tends to this set independently of the choice of D_0^C.

This is the set that we wanted to find. In the precise, time-homogeneous case, the limiting conditional distribution $\boldsymbol{d}$ has the property that if the distribution over C at time n is $\boldsymbol{d}$, then the distribution over C at time $n+1$, conditioned on non-absorption, must also be $\boldsymbol{d}$. Similarly, in the imprecise time-inhomogeneous case, if the set of possible distributions over C at time n is $\mathcal{M}_\infty^C$, then the set of possible distributions over C, conditioned on non-absorption, at time $n+1$ must also be $\mathcal{M}_\infty^C$. Moreover, just as convergence to the limiting conditional distribution $\boldsymbol{d}$ in the limit is certain in the precise time-homogeneous case, so too is convergence to $\mathcal{M}_\infty^C$ in the limit in the imprecise time-inhomogeneous case. This justifies our earlier claim that a generalisation exists in the imprecise, time-inhomogeneous case for both the limiting conditional distribution and the quasi-stationary distribution, and that these two concepts are generalised by sets of distributions that are in fact equal.

We do not at present have a method for directly calculating $\mathcal{M}_\infty^C$. Škulj presents in [62] a method for approximating $\mathcal{M}_\infty$ (so without conditioning) using Choquet integrals. De Cooman offers a precise method for calculating $\mathcal{M}_\infty$ in [15]. Whether

either of these methods can be adapted to our situation remains an open question. In the next section, we present an example in which $\mathcal{M}_\infty^C$ is approximated.

4.7 An Example

We now present an example which will demonstrate the convergence to the conditionally invariant set of distributions $\mathcal{M}_\infty^C$.

Example 4.7.1

Consider the Markov chain $\mathcal{X} := \{X(n), n = 0, 1, \ldots\}$ with state space $S = \{-1, 0, 1, 2\}$, where -1 is an absorbing state. Define each possible transition matrix as

$$P = \begin{pmatrix} \boldsymbol{r}^{(-1)} \\ \boldsymbol{r}^{(0)} \\ \boldsymbol{r}^{(1)} \\ \boldsymbol{r}^{(2)} \end{pmatrix}$$

where $\boldsymbol{r}^{(i)} \in \mathcal{R}^{(i)}$, and let

$$\begin{aligned} \mathcal{R}^{(-1)} &= \{(1, 0, 0, 0)\} \\ \mathcal{R}^{(0)} &= \{(0.5, 0, 0.5, 0)\} \\ \mathcal{R}^{(1)} &= \{(0, a, 0, 1-a) | a \in [0.1, 0.2])\} \\ \mathcal{R}^{(2)} &= \{(0, 0, 0.75, 0.25)\}. \end{aligned}$$

Therefore only the transition probabilities from state 1 are not known precisely. All possible initial distributions over C are allowed, thus the set of initial distributions is $\mathcal{M}_0^C$. Rather than attempt to find each $\mathcal{M}_n^C$ precisely, exact bounds are found upon each element of the vectors in the sets $\mathcal{M}_n^C$ for $n = 1 \ldots, 4$, in the same manner as in Method 2 in Section 3.5. Approximate bounds are also offered for the elements of the vectors in sets $\mathcal{M}_{50}^C$ and $\mathcal{M}_{100}^C$. These approximate bounds were found by randomly calculating 1000 50-step and 100-step matrices, multiplying each one by the distributions $(1, 0, 0)$, $(0, 1, 0)$, and $(0, 0, 1)$, conditioning each resulting distribution

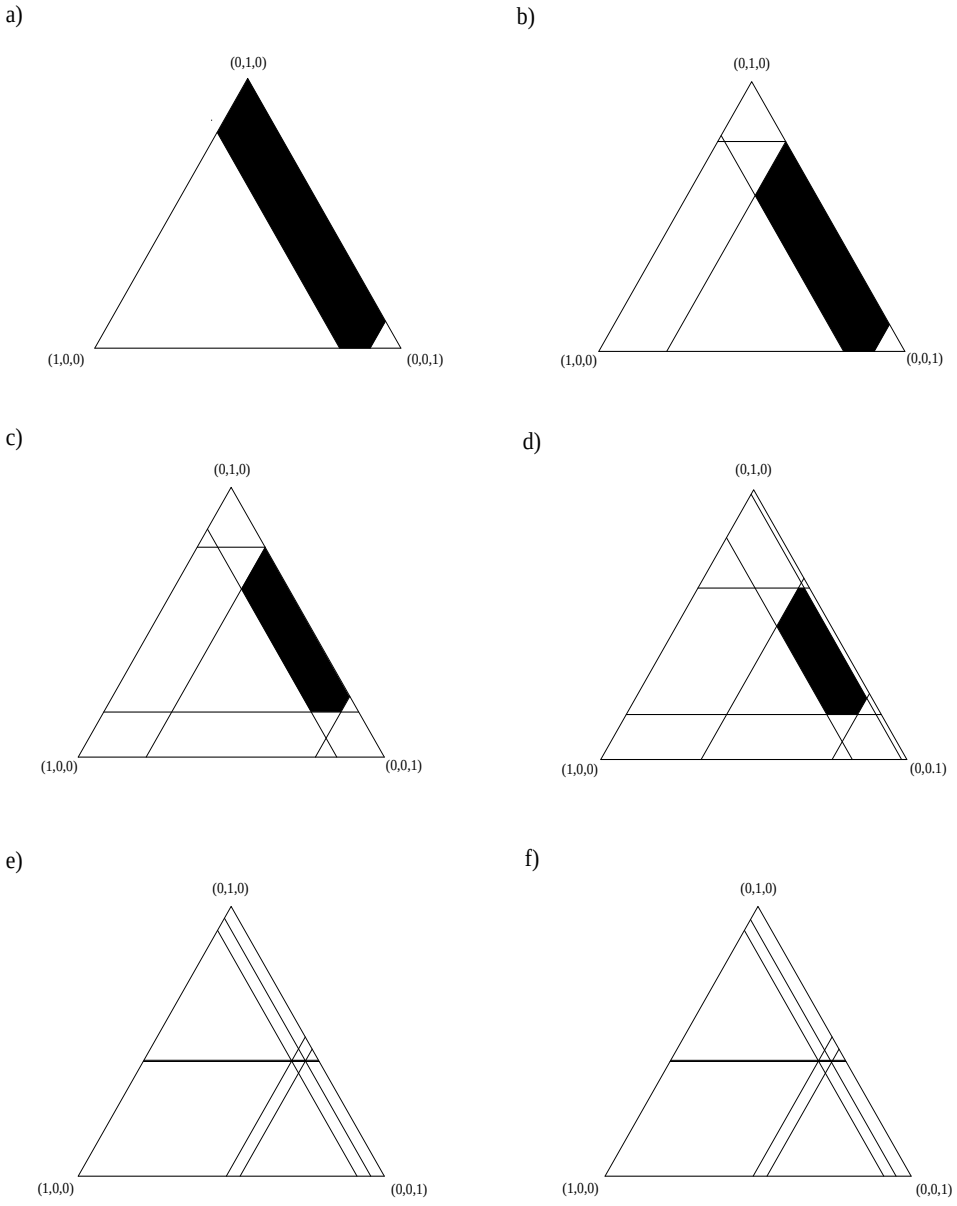


Figure 4.1: Precise bounds upon sets $\mathcal{M}_n^C$ for $n = 1, \ldots, 4$, (diagrams a) to d), respectively) and approximate bounds for $n = 50, 100$ (diagrams e) and f), respectively) for Example 4.7.1.

upon non-absorption, and then taking the minimum and maximum of each element over the resulting 3 000 distributions. The justification for using those three vectors is the fact that for any $\boldsymbol{v} \in \mathcal{M}_0^C$, $\boldsymbol{v} = \alpha(1,0,0) + \beta(0,1,0) + (1-\alpha-\beta)(0,0,1)$. Thus $\boldsymbol{v}P = \alpha(1,0,0)P + \beta(0,1,0)P + (1-\alpha-\beta)(0,0,1)P$, and so the vectors $(1,0,0)P$, $(0,1,0)P$ and $(0,0,1)P$ describe the set $\boldsymbol{v}P$. It was discovered during calculation

that these three vectors were equal to at least 4 decimal places. R[2] was used to perform these calculations.

The resulting diagrams demonstrate that even in a situation in which the transition matrix is free to change between time steps, the possible behaviour as time approaches infinity, conditioned on non-absorption, can be shown to lie within a comparatively small set of distributions. This is very important, and not only in the field of imprecise probability. Determining the long-term behaviour of precise time-inhomogeneous chains is often very difficult. Assuming that the range each element of the transition matrix must lie within over all time steps is known, however, $\mathcal{M}_{\infty}^{C}$ will be a superset of all possible distributions, conditioned upon non-absorption, that the process can display as time approaches infinity.

4.8 Comparison of Methods

In this section comparisons are made between the model described in this chapter, and that described in Chapter 3, in which the transition matrix, whilst containing unknown elements, was assumed to remain constant.

Two examples are presented. In the first, there is no state in C for which the transition probabilities from that state are known precisely. In the second example, there is only one state for which the transition probabilities from that state are not known precisely, but the interval describing that imprecision is much wider than any of those in the first example. This will allow us to compare two different forms of imprecision: imprecision in the sense of number of states for which we cannot exactly calculate their associated transition probablities, and imprecision in the sense of the level of our uncertainty regarding the transition probabilities associated with a specific state. In the simplest terms possible, what qualitative difference is there between a case for which the transition probabilities $p_{0,-1}$, p_{01}, p_{10} and p_{12} are known to lie in $[0.4, 0.6]$, and the case in which the transition probabilities $p_{0,-1}$ and p_{01} are known to lie in $[0.3, 0.7]$? In each example we consider the difference between applying the model presented in Chapter 3 and that presented in this chapter by

[2]Version 2.8.1

comparing the sizes of the bounded sets generated at equivalent time steps for each case. We will also compare the degree of convergence both within and between each method.

Note that throughout this section $\mathcal{M}_0^C$ is used as the set of all possible initial distributions over C. In each diagram, the method used in Chapter 3 is displayed on the left, and the method used in this chapter is displayed on the right.

Example 4.8.1

Consider the birth-death process $\mathcal{X}$ with state space $S = \{-1\} \cup C$ where $C = \{0, 1, 2\}$. The set of all possible one-step transition matrices $\mathcal{P}$ is given by

$$\mathcal{P} = \left\{ \begin{pmatrix} 1 & 0 & 0 & 0 \\ a & 0 & 1-a & 0 \\ 0 & b & 0 & 1-b \\ 0 & 0 & c & 1-c \end{pmatrix} \right\} \tag{4.100}$$

where $a \in [0.1, 0.3]$, $b \in [0.5, 0.6]$, and $c \in [0.67, 0.73]$.

From Definition 3.4.1 we have

$$\tilde{\mathcal{M}}_n^C = \bigcup_{P \in \mathcal{P}} \mathcal{M}_n^C(P) = \bigcup_{P \in \mathcal{P}} \{f(\boldsymbol{v}P^n) | \boldsymbol{v} \in \mathcal{M}_0\} \tag{4.101}$$

and from (4.43) we have

$$\mathcal{M}_n^C = \left\{ f(\tilde{f}_\alpha(\boldsymbol{v})P) | \boldsymbol{v} \in \mathcal{M}_{n-1}^C, P \in \mathcal{P} \right\}, \tag{4.102}$$

respectively. Deriving either of these sets for $n > 0$ in their entirety for this example is a non-trivial task. Instead, the maximum and minimum values of each element of the vectors contained in $\tilde{\mathcal{M}}_n^C$ and $\mathcal{M}_n^C$ will be calculated.

The simplex diagrams in Figure 4.2 show $\tilde{\mathcal{M}}_n^C$ for $n = 2, 3, 4$ (diagrams a), c), and e), respectively), and $\mathcal{M}_n^C$, also for $n = 2, 3, 4$ (diagrams b), d), and f), respectively). Bounds have also been approximated for the sets $\tilde{\mathcal{M}}_{100}^C$ and $\mathcal{M}_{100}^C$, in the same manner as those found in Section 4.7. We argue that these sets are close approximations to $\tilde{\mathcal{M}}_\infty^C$ and $\mathcal{M}_\infty^C$, respectively, as follows. A number of matrices $P \in \mathcal{P}^{100}$ were calculated, and for any vectors $\boldsymbol{v}, \boldsymbol{v}' \in \mathcal{M}_0^C$, it was calculated that

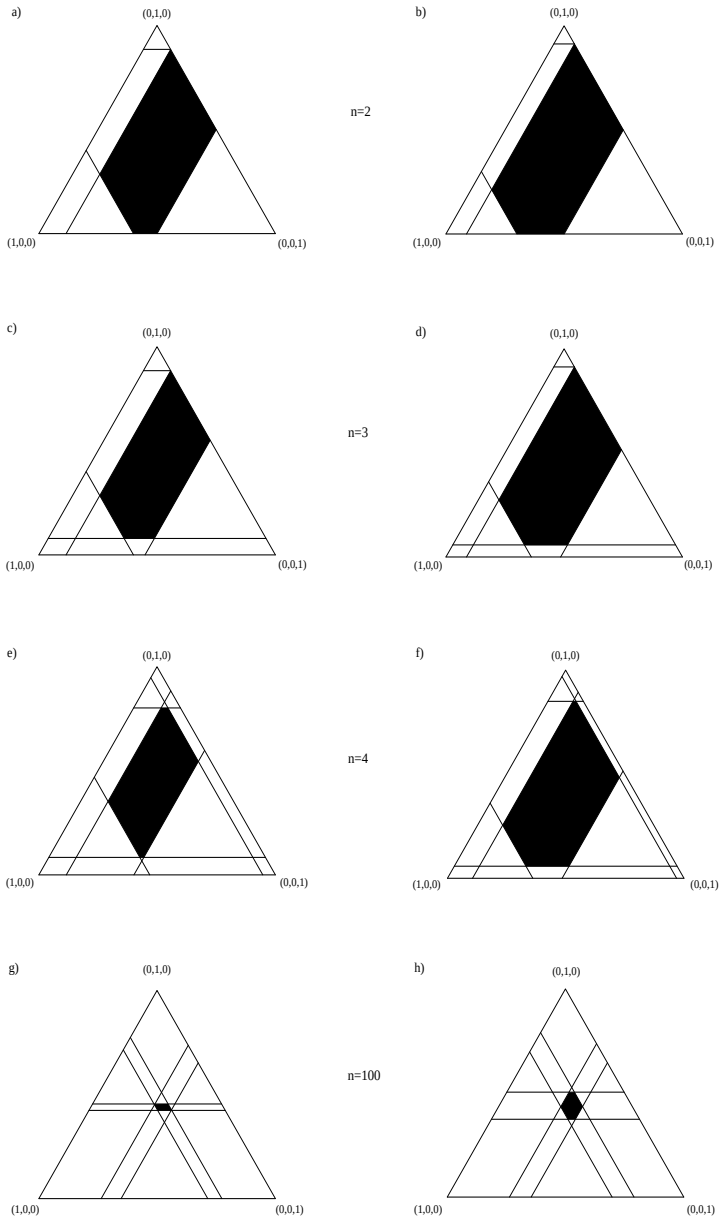


Figure 4.2: Bounds for the sets $\tilde{\mathcal{M}}_n^C$ and $\mathcal{M}_n^C$ in Example 4.8.1.

$\| (\boldsymbol{v} - \boldsymbol{v}')P \| < 0.000001$, where $\| \cdot \|$ indicates the supremum norm. It is thus argued that by time step 100, the convergence to the set $\mathcal{M}_\infty^C$ is all but complete. Given this, then, it seems entirely reasonable that by $n = 100$ the convergence will be almost complete in the time-homogeneous case also.

Recall that it is known that the size of the bounded areas are non-increasing from time step n to $n + 1$, according to (3.10) and (4.13). Figure 4.2 demonstrates

these properties very well. Note also that, as expected, for every time step the bounded areas on the right are larger than those on the left. This should come as no surprise considering what the two different models represent. For the sets $\mathcal{D}_0^C$ and $\mathcal{P}$ any long-term behaviour observed using the model described in Chapter 3 could theoretically also be observed in the model described in this chapter. The reverse, however, is not true. Thus it is entirely consistent that the sets shown in the simplex diagrams on the left hand side should be as small as or smaller than those on the right.

Example 4.8.2

Consider an imprecise birth-death process $\mathcal{X}$ with state space $S = \{-1\} \cup C$ where $C = \{0, 1, 2\}$. The set of all possible one-step transition matrices $\mathcal{P}$ is given by

$$\mathcal{P} = \left\{ \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0.6 & 0 & 0.4 & 0 \\ 0 & d & 0 & 1-d \\ 0 & 0 & 0.7 & 0.3 \end{pmatrix} \right\} \tag{4.103}$$

where $d \in [0.37, 0.73]$. The diagrams were created using identical methods to those used in the first example.

The same observations regarding Figure 4.2 can also be made about Figure 4.3. If the two figures are compared, it can be seen that in the second example more can be said about the probability of being in state 1 (in the sense that the distance between the bounds on the probability of being in state 1 are closer together in Figure 4.3), conditioned on non-absorption as time approaches infinity, but less can be said about the probabilities of being in states 0 or 2. This may be explained as follows. In the method used in Chapter 3, the bounds upon $\tilde{\mathcal{M}}_\infty^C$ are simply the bounds upon the set $\bigcup_{P \in \mathcal{P}} \boldsymbol{\alpha}_P$. Thus the bounds approximated in the bottom-left simplex of Figure 2 relate to the three elements of a vector function with three unknowns, a, b and c, all with comparatively small ranges. In comparison, the bounds which are approximated in the bottom-left simplex diagram of Figure 4.2 relate to the three elements of a vector function with one unknown, d, which has

a comparatively large range. The fact that the set in the bottom-left diagram of Figure 4.3 then is more elongated and thinner than the equivalent set in Figure 4.2 is therefore unsurprising. Moreover, the knowledge that $\tilde{\mathcal{M}}_{\infty}^{C} \subseteq \mathcal{M}_{\infty}^{C}$ makes it also unsurprising that the bottom-right diagrams in Figures 4.2 and 4.3 exhibit similar characteristics.

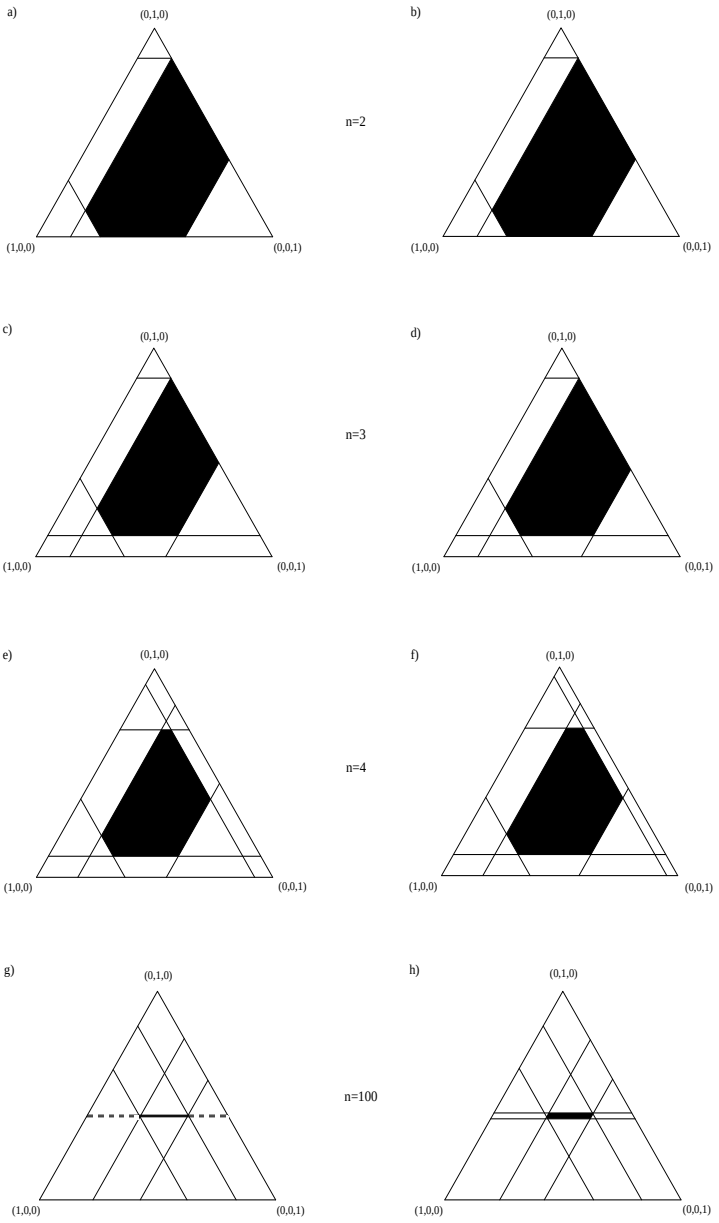


Figure 4.3: Bounds for the sets $\tilde{\mathcal{M}}_{n}^{C}$ and $\mathcal{M}_{n}^{C}$ in Example 4.8.2.

A final point regarding Figure 4.2 and 4.3 is the fact that in both the situation in which little is known about one state's behaviour, and in that where no state's behaviour is entirely known, there is much that can be said about the long-term behaviour conditioned on non-absorption. It is *not* the case, as may have been feared, that the imprecision grows with each new iteration until there is nothing to be said about a given time-step. Moreover, this is true even when the transition matrix is not assumed to be constant. This is particularly important because it suggests that the model used in this chapter can be applied to approximating the long-term behaviour of precise time-inhomogeneous chains with an absorbing state, conditioned upon non-absorption, an area in which very few results have been published, compared to what is known regarding the precise time-homogeneous case.

It is also possible to compare the two models by creating a set of r initial distributions to approximate $\mathcal{M}_0^C$ and a set of s transition matrices to approximate $\mathcal{P}$. These can then be used to create sets of vectors to approximate $\overline{\mathcal{M}}_n^C$ and $\mathcal{M}_n^C$.

The drawback to this method is that the number of calculations required rapidly becomes very large. In the example above, allowing $\mathcal{M}_0^C$ to be approximated by the 231 vectors $\{\frac{i}{20}, \frac{j}{20}, \frac{k}{20}\}$, where i, j, k are the set of non-negative integers for which $i + j + k = 20$, and allowing $\mathcal{P}$ to be approximated by the 264 matrices for which $a \in [0.1, 012, \ldots, 0.3]$, $b \in [0.5, 0.52, \ldots, 0.7]$, and $c \in [0.67, 0.69, 0.71, 0.73]$, then by the time $n = 4$ there are over a thousand *billion* vectors to calculate.

4.9 Concluding Remarks

In Chapters 3 and 4 we have demonstrated that in many cases an imprecise Markov chain will converge towards an analogue to the limiting conditional distribution of a precise Markov chains. Moreover, this result holds in conditions that are only slightly more restrictive than the conditions required in the precise case to guarantee a limiting conditional distribution. It has also been shown that taking the set of distributions to which an imprecise Markov chain tends, conditioned on non-absorption, and using it as the set of possible initial distributions over C leads to the set of possible distributions, conditioned on non-absorption, being equal at all

time steps. Therefore, our analogue to the limiting conditional distribution is also an analogue to the quasi-stationary distribution.

Furthermore, it has been shown that these results apply even in situations in which the transition matrix is allowed to change from time step to time step. This is a powerful result, both on its own terms, and because the set $\mathcal{M}_\infty^C$ must contain all possible limiting conditional distributions for a precise time-inhomogeneous Markov chain which draws its transition matrices from some subset of $\mathcal{P}$. As the long-term behaviour of time-inhomogeneous Markov chains is not well-known, the resulting approximation that our model gives may be of great value.

There are multiple possibilities for further research. The most obvious one, perhaps, is to find a method by which the conditionally invariant set of distributions can be calculated, or at least bounded using a method that offers more insight than taking the maximum and minimum of each element. For the situation in which no absorbing state is present, Škulj [62] presents a method for bounding the set of invariant distributions using Choquet integrals, and de Cooman *et al.* [15] provided a method of direct calculation by the application of a concept known as *lower previsions*. Adapting either of these methods to the case where conditioning is required would allow bounds on $\mathcal{M}_\infty^C$ to be calculated using iterative methods. Lack of time prevented further consideration of such methods in this thesis.

A second obvious extension to the work here would be to consider the situation in which there exists an infinite number of states. In the precise case this situation leads to each chain having families of limiting conditional distributions, rather than a single unique one (see Seneta and Vere-Jones, [60]). Whether this means that in our case, there will exist families of conditional sets of invariant distributions, is an open question.

The third potential avenue of progress is the most daunting. Would it be possible to achieve similar results in the case of *continuous* Markov chains? To our knowledge, imprecise Markov chains are an area in which almost no research has been carried out. Unlike the discrete case, we are not aware of any results on the long term behaviour of imprecise Markov chains whatsoever. Even the method by which such chains would be modelled is unclear; would we (for example) use interval

transition rates in Q, or would we attempt to introduce intervals into the matrix $P(t)$? These two approaches might lead to different results (see (1.8)). Ultimately, our aim is to describe the long-term behaviour of a continuous-time Markov chain with an absorbing state, conditioned on non-absorption, and then consider the effect imprecision has on the behaviour of the hazard rate, as defined in Chapter 2. This, however, is clearly something that lies some way in the future, if indeed it is possible at all.

Bibliography

[1] Aalen, O.O. (1995). Phase type distributions in survival analysis. *Scandinavian Journal of Statistics* **22**, 447-463.

[2] Aalen, O.O. & Gjessing, K. (2001). Understanding the shape of the hazard rate: a process point of view. *Statistical Sciences* **16**, 1-22.

[3] Aalen O.O., Borgan, Ø., & Gjessing, H.K. (2008). *Survival and Event History Analysis: A Process Point of View*, Springer.

[4] Abate, J. & Whitt, W. (1989). Spectral theory for skip-free Markov chains. *Probability in the Engineering and Informational Sciences* **3**, 77-88.

[5] Anderson, W.J. (1991). *Continuous-Time Markov Chains: an Applications-Oriented Approach.* Springer-Verlag.

[6] Avrachenkov, K. & Sanchez, E. (2002). Fuzzy Markov chains. *Fuzzy Optimisation and Decision Making* **1**, 143-159.

[7] Bawa, V. S. (1975). Optimal rules for ordering uncertain prospects. *Journal of Financial Economics* **2**, 95-121.

[8] Beer, G. (1993). *Topologies On Closed And Closed Convex Sets.* Kluwer Academic Publishers.

[9] Bergin, J. & Lipman, B. (1996). Evolution with state-dependent mutations. *Econometrica* **64**, 943-956.

[10] Billingsley, P. (1995). *Probability and Measures.* Wiley.

[11] Chihara, T.S. (1978). *An Introduction to Orthogonal Polynomials.* Gordon and Breach.

[12] Chikara, R.S. & Folks, J.L. (1989). *The Inverse Gaussian Distribution: Theory, Methodology and Applications.* Dekker.

[13] Coolen, F.P.A. & Newby, M.J. (1994) Bayesian reliability analysis with imprecise prior probabilities. *Reliability Engineering and System Safety*, **43**, 75-85.

[14] Coolen-Schrijner, P. & van Doorn, E.A. (2006). Quasi-stationary distributions for a class of discrete-time Markov chains, *Methodology and Computing in Applied Probability* **8**, 449-465.

[15] de Cooman, G., Quaeghebeur, E. & Miranda, E. (2009). Imprecise Markov chains and their limit behaviour. *Pre-print at arXiv.org*, arXiv:0801.0980.

[16] Cox, D.R. (1962). *Renewal Theory.* Methuen & Co Ltd.

[17] Crossman, R.J., Coolen-Schrijner, P. & Coolen, F.P.A. (2009). Time homogeneous birth-death processes with probability intervals and absorbing state. *Journal of Statistical Theory & Practice* **3**, 103-118.

[18] Crossman, R.J. & Škulj D. (2009). *Imprecise Markov chains with absorption.* In submission.

[19] Crossman, R.J., Coolen-Schrijner P., Škulj D., & Coolen, F.P.A. (2009). *Imprecise Markov chains with an absorbing state*, In T. Augustin, F.P.A. Coolen, S. Moral, M. Troffaes (eds.)., *Proceedings of the Sixth International Symposium on Imprecise Probabilities: Theory and Applications*, 119-128. Durham, UK.

[20] Dambrine, S. & Moreau, M. (1981). Note on the theory of a self-catalytic chemical reaction, I. *Physica A* **106**, 559-573.

[21] Darroch, J.N. & Seneta, E. (1965). On quasistationary distributions in absorbing discrete-time finite Markov chains. *Journal of Applied Probability* **2**, 88-100.

[22] Darroch, J.N. & Seneta, E. (1967). On quasistationary distributions in absorbing continuous-time finite Markov chains. *Journal of Applied Probability* **4**, 192-196.

[23] Doob, J.L. (1953). *Stochastic Processes.* Wiley.

[24] van Doorn, E.A. (1991). Quasi-stationary distributions and convergence to quasi-stationarity of birth-death processes. *Advances in Applied Probability* **23**, 683-700.

[25] van Doorn, E.A. & Schrijner, P. (1995a). Geometric ergodicity and quasi-stationarity in discrete-time birth-death processes. *Journal of the Australian Mathematical Society Series B* **37**, 121-144.

[26] van Doorn, E.A. & Schrijner, P. (1995b). Ratio limits and limiting conditional distributions for discrete-time birth-death processes. *Journal of Mathematical Analysis and Applications* **190**, 263-284.

[27] Feller, W. (1950). *An Introduction to Probability Theory and its Applications, Volume 1*, 3rd Edition. Wiley.

[28] Glaser, R.E. (1980). Bathtub and related failure rate characterizations, *Journal of the American Statistical Association* **75**, 667-672.

[29] Grimmett, G. & Stirzaker, D. (1992). *Probability and Random Processes.* Oxford.

[30] Hartfiel, D.J. (1998). *Markov Set Chains.* Springer-Verlag.

[31] Ibragimov, I.A. (1956). On the composition of unimodal distributions. *Theory of Probability and its Applications*, **1**.

[32] Karlin, S. & McGregor, J.L. (1957). The differential equations of birth-death processes, and the Stieltjes Moment Problem. *Transactions of the American Mathematical Society* **85**, 489-546.

[33] Karlin, S. & McGregor, J.L. (1957). The classification of birth and death processes. *Transactions of the American Mathematical Society* **86**, 366-400.

[34] Karlin, S. (1968). *Total Positivity 1*, Stanford University Press.

[35] Karlin, S & Taylor, H.M. (1981). *A Second Course in Stochastic Processes.* Academic Press.

[36] Keilson, J. (1979). *Markov chain models - rarity and exponentiality.* Springer-Verlag.

[37] Keilson, J. (1972). Log-concavity and log-convexity in passage time densities of diffusion and birth-death processes. *Journal of Applied Probability* **8**, 391-398.

[38] Keilson, J. & Styan, G.P.H. (1973). Markov chains and M-matrices: inequalities and equalities. *Journal of Mathematical Analysis and Applications* **41**, 439-459.

[39] Keilson, J. (1981). On the unimodality of passage time densities in birth-death processes. *Statistica Neerlandica* **35**, 49-56.

[40] Kesten, H. (1995). A ratio limit theorem for (sub) Markov chains on $1, 2, \ldots$ with bounded jumps. *Advances in Applied Probability* **27**, 652-691.

[41] Kijima, M. (1997). *Markov Processes for Stochastic Modeling.* Chapman & Hall.

[42] Kozine, I. & Utkin, L. (2002). Interval-valued finite Markov chains. *Reliable Computing* **8**, 97-113.

[43] Ledermann, W. & Reuter, G.E.H. (1954). Spectral theory for the differential equations of simple birth-and-death processes. *Philosophical Transactions of the Royal Society of London, Series A* **246**, 321-369.

[44] Li, Q.L. & Zhao, Y.Q. (2003). Beta-invariant measures for transition matrices of G1/M/1 type. *Stochastic Models* **319**, 201-233.

[45] Maller, R. and Zhou, X. (1996). *Survival Analysis With Long-Term Survivors.* John Wiley.

[46] Mandl, P. (1960). On the asymptotic behaviour of probabilities within groups of states of a homogeneous Markov process. *Časopis pro pestovani matematiky a fyziky* **85**, 448-456.

[47] Neuts, M.F. (1975). Probability distributions of phase type. *Liber Amicorum Prof. Emeritus H. Florin.* University of Louvain, Belgium, 173-206.

[48] Olave, P. & Salvador, M. (2007). Semi-parametric Bayesian analysis of the proportional hazard rate model: an application to the effect of training programs on graduate unemployment. *Journal of Applied Statistics* **34**, 1185-1205.

[49] Oloffson, P. (2005). *Probability, Statistics, and Stochasitc Processes.* Wiley.

[50] Pak, M. (2008). Stochastic stability and time-dependent mutations. *Games and Economic Behaviour* **64**, 650-665.

[51] Pakes, A. (1987). Limit theorems for the population size of a birth and death process allowing catastrophes. *Journal of Mathematical Biology* **25**, 307-325.

[52] Parsons, R. & Pollett, P. (1987). Quasistationary distributions for autocatalytic reactions. *Journal of Statistical Physics* **46**, 249-254.

[53] Rogers, G.S., Kopf, A.W., Rigel, D.S. *et al.* (1986). Hazard-rate analysis in stage I malignant melanoma. *Archives of Dermatology* **122**, 999-1002.

[54] Ross, S.M. (1983). *Stochastic Processes.* Wiley.

[55] Scheffer, V.B. (1951). The rise and fall of a reindeer herd, *The Scientific Monthly* **73**, 356-362.

[56] Schrijner, P. (1995). *Quasi-stationarity od Discrete-Time Markov Chains.* PhD thesis, University of Twente.

[57] Schrijner, P. (1997). Weak convergence of conditioned birth-death processes in discrete time. *Journal of Applied Probability* **34**, 46-53.

[58] Schweitzer, P. (1968). Perturbation theory and finite Markov chains. *Journal of Applied Probability* **5**, 401-413.

[59] Seneta, E. (1966). Quasi-stationary behaviour in the random walk with continuous time. *Australian Journal of Statistics* **8**, 92-98.

[60] Seneta, E. & Vere-Jones, D. (1966). On quasi-stationary distributions in discrete-time Markov chains with a denumerable infinity of states. *Journal of Applied Probability* **3**, 403-434.

[61] Seshadri, V. (1998). *The Inverse Gaussian Distribution: Statistical Theory and Applications.* Springer.

[62] Škulj, D. (2006). Finite discrete time Markov chains with interval probabilities. In J. Lawry, E. Miranda, A. Bugarin, S. Li, M.A. Gil, P. Grzegorzewski, O. Hryniewicz (eds.). *Soft Methods for Integrated Uncertainty Modelling*, 299-306. Springer.

[63] Škulj, D. (2007). Regular finite Markov chains with interval probabilities. In G. de Cooman, J. Vejnarová, M.Zaffalon (eds.)., *Proceedings of the Fifth International Symposium on Imprecise Probabilities and Their Applications*, 406-413. Prague, Czech Republic.

[64] Solomon, A., Barnes, J.W., Dokob, S.P. & Acevedo, R. (2003) Weakly Symmetric Graphs, Elementary Landscapes and the TSP. *Applied Mathematical Letters* **16**, 401-407.

[65] Steen, L.A. (1970). *Counterexamples in Topology.* Springer-Verlag.

[66] Walley, P. (1981). Coherent lower (and upper) probabilities. Department of Statistics, University of Warwick.

[67] Walley, P. (1991). *Statistical Reasoning with Imprecise Probabilities.* Chapman & Hall.

[68] Weichselberger, K. (2000). The theory of interval-probability as a unifying concept for uncertainty. *International Journal of Approximate Reasoning* **24**, 149-170.

[69] Weichselberger, K. (2001). *Elementare Grundbegriffe einer allgemeineren Wahrscheinlichkeitsrechnung I. Intervallwahrscheinlichkeit als umfassendes Konzept*. Physika: Heidelberg.

[70] Wise, J. (1962). The relationship between the mean and variance of a stationary birth-death process, and its economic application. *Biometrika* **1/2**, 253-255.

[71] Yaglom, A. (1947). Certain limit theorems of the theory of branching stochastic processes (in Russian). *Doklady Akademii Nauk SSSR* **56**, 795-798.