6th International Symposium on Imprecise Probability: Theories and Applications, Durham, United Kingdom, 2009

Imprecise Markov Chains with an Absorbing State

Richard J. Crossman
Durham University
r.j.crossman@durham.ac.uk

Pauline Coolen-Schrijner
Durham University

Damjan Škulj
University of Ljubljana
damjan.skulj@fdv.uni-lj.si

Frank P.A. Coolen
Durham University
frank.coolen@durham.ac.uk

Abstract

Several authors have presented methods for considering the behaviour of Markov chains in the generalised setting of imprecise probability. Some assume a constant transition matrix which is not known precisely, instead bounds are given for each element. Others consider a transition matrix which is neither known precisely nor assumed to be constant, though each element is known to exist within intervals that are constant over time. In both cases results have been published regarding the long-term behaviour of such chains. When a finite Markov chain is considered with a single absorbing state, however, eventual absorption is generally certain in both cases. Thus it is of interest to consider the long-term behaviour of the chain, conditioned on non-absorption, within the setting of imprecise probability. Methods have previously been presented for the case of a constant transition matrix, and submitted for the case of a non-constant transition matrix. In this paper the methods for the two cases are compared.

Keywords. Absorbing state, imprecise probability, Markov chains, time-inhomogeneity

1 Introduction

There are several papers in which the theory of interval probability has been applied to the consideration of Markov chains. Kozine and Utkin [10] consider the situation in which the individual elements of the transition matrix are assumed to be constant over time, but may not be known precisely (thus that paper can be thought of as generalising the time-homogeneous case). Instead, all that is known are the intervals in which each individual matrix element is contained. This property can be relaxed, as it was by Škulj [13, 14], by only requiring that the intervals to which those elements belong remain constant over time, and allowing the elements to vary with time (thus those papers can be thought of as generalising the time-inhomogeneous case). In those same papers the concept of the initial distribution is also generalised, so that rather than assume a specific initial distribution, an entire set of possible initial distributions is defined. The papers then considered the long-term behaviour of such chains, and proved that, subject to certain conditions, the possible distributions as time approaches infinity form a set that is independent of the set of initial distributions. An alternate method for considering the situation found in [13, 14] was offered by de Cooman *et al.* in [3]; we explain in Section 3 why we have not adopted their method in this paper.

It can be proved that for a finite Markov chain with one absorbing state eventual absorption is certain both in the case given by Kozine and Utkin [10], and also the case found in [14], assuming the conditions required in that paper (the respective proofs for these results can be found in Crossman *et al.* [4, 5]). In this situation, then, it is of more interest to consider the long-term behaviour of the chain when conditioning on non-absorption at each step.

What follows can be thought of as a generalisation of the *limiting conditional distribution* in the precise case. The limiting conditional distribution, if used as the initial distribution, is referred to as the *quasi-stationary distribution* (QSD). The QSD has many applications. For example, it is used by Pakes [11] to better understand population sizes, which are modelled in that paper as birth-death processes with catastrophes. In this case the QSD represents the long-term behaviour of a stable population, before the point at which it becomes extinct. Further, Parsons and Pollet [12] apply QSDs to describe the long-term behaviour of certain catalytic chemical reactions.

Crossman *et al.* [4] considered the long-term behaviour conditioned on non-absorption for the model given in [10], and the consideration of sets of initial distributions is introduced as in [13]. Crossman and Škulj [5] then applied this consideration to the model

given in [13], though the restrictions upon each row of the transition matrix is given as a closed probability set, rather than a group of intervals. In this paper the method found in [4] is similarly expanded to using closed probability sets (more on this can be found in Crossman [6]), and the two different approaches are compared.

1.1 Markov chains with imprecision

The following model is given in a slightly different form by Škulj [13]. Let $\mathcal{X} = \{X(n), n = 0, \ldots\}$ be a discrete-time Markov chain on the state space $S = \{-1\} \cup C$ with $C = \{0, \ldots, s\}$ where -1 is an absorbing state and C is a set of transient states. Imprecision is introduced by the assumption that the transition matrix for any given time step is not known precisely. Instead, limitations are imposed upon the possible values of each transition probability at each step.

Define $s + 2$ closed sets of probability distributions, $\mathcal{P}^{(i)},\ i = -1, 0, \ldots, s$.

Definition 1.1 All potential transition matrices for a given time step belong to the set

$$\mathcal{M}(P) := \left\{ \begin{pmatrix} \boldsymbol{p}^{(-1)} \\ \vdots \\ \boldsymbol{p}^{(s)} \end{pmatrix} \mid \boldsymbol{p}^{(i)} \in \mathcal{P}^{(i)},\ \forall i \in C \right\}$$

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

Thus, each row of the transition matrix for a given time step is chosen from a set of probability distributions, and each choice is made independently.

Further conditions are now given. First, as -1 is an absorbing state $\mathcal{P}^{(-1)} = \{(1, 0, \ldots, 0)\}$ is required. Further, each of the possible transition matrices must guarantee that C is a single communicating class with each set in C aperiodic.[1]

Definition 1.2 The set of all possible initial distributions over S is denoted by

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

Furthermore, $\mathcal{D}_0$ is used to denote a strict subset of $\mathcal{M}_0$.

Thus, $\mathcal{D}_0$ can be thought of as the set of initial distributions deemed possible for a given process, where this conclusion is arrived at by some unspecified method. $\mathcal{M}_0$ would be used only when nothing whatsoever is known about the initial distribution.

[1] Note that this is a more general formulation than can be found in [4], the justification for this change can be found in [6].

2 Imprecise Markov chains with constant transition matrix

In this section it is assumed that there is a single element of $\mathcal{M}(P)$ that describes the transition probabilities at every time step, i.e. the transition matrix is unknown, but constant.

As mentioned, $\mathcal{D}_0$ represents the set of initial distributions over S that have been judged possible. Thus for a matrix $P \in \mathcal{M}(P)$ the set $\tilde{\mathcal{D}}_n(P)$ of all possible distributions over S at time $n \geq 1$ can be defined as follows.

Definition 2.1

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

where $\tilde{\mathcal{D}}_0(P) := \mathcal{D}_0$. Should every possible initial distribution be considered possible, the appropriate definition becomes

$$\tilde{\mathcal{M}}_n(P) := \{\boldsymbol{v}P | \boldsymbol{v} \in \tilde{\mathcal{M}}_{n-1}(P)\} = \{\boldsymbol{v}P^n | \boldsymbol{v} \in \tilde{\mathcal{M}}_0(P)\}$$

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

However, since it is unknown which element of the set $\mathcal{M}(P)$ actually describes the behaviour of the chain, it is of more pratical use to introduce the following definition.

Definition 2.2

$$\tilde{\mathcal{M}}_n := \bigcup_{P \in \mathcal{M}(P)} \tilde{\mathcal{M}}_n(P). \tag{2.1}$$

Thus $\tilde{\mathcal{M}}_n$ contains every distribution possible at time n.

Theorem 2.1 For each $P \in \mathcal{M}(P)$ and $n \geq 0$,

$$\tilde{\mathcal{M}}_{n+1}(P) \subseteq \tilde{\mathcal{M}}_n(P).$$

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

$$\begin{aligned} \tilde{\mathcal{M}}_{n+1}(P) &= \{\boldsymbol{v}P \mid \boldsymbol{v} \in \tilde{\mathcal{M}}_n(P)\} \\ &\subseteq \{\boldsymbol{v}P \mid \boldsymbol{v} \in \tilde{\mathcal{M}}_{n-1}(P)\} \\ &= \tilde{\mathcal{M}}_n(P). \end{aligned}$$

□

It is therefore appropriate to define

Definition 2.3

$$\tilde{\mathcal{M}}_\infty(P) := \bigcap_{n=0}^{\infty} \tilde{\mathcal{M}}_n(P)$$

This set $\tilde{\mathcal{M}}_\infty(P)$ describes the behaviour of the chain as time approaches infinity. Once again, though, since the correct matrix from $\mathcal{M}(P)$ is unknown, the following definition is of more practical use.

Definition 2.4

$$\tilde{\mathcal{M}}_\infty = \bigcup_{P\in\mathcal{M}(P)} \tilde{\mathcal{M}}_\infty(P).$$

It is proved in [4] that in our current case

$$\tilde{\mathcal{M}}_\infty = \bigcup_{P\in\mathcal{M}(P)} \{(1,0,\ldots,0)\} = \{(1,0,\ldots,0)\}.$$

Thus, since $\tilde{\mathcal{D}}_n(P) \subseteq \tilde{\mathcal{M}}_n(P)$ for all n, eventual absorption is certain irrespective of our choice of $\mathcal{D}_0$ or the actual element of the set $\mathcal{M}(P)$ that correctly describes the chain. We therefore consider the situation in which the chain is conditioned on non-absorption at each step.

We now define the following functions.

Definition 2.5 For

$$\boldsymbol{v} \in \mathcal{M}_0 \setminus \{(1,0,\ldots,0)\} \tag{2.2}$$

we have

$$f(\boldsymbol{v}) = f((v_{-1}, \boldsymbol{v}^*)) = \frac{1}{1-v_{-1}}\boldsymbol{v}^*,$$

and

$$\begin{aligned}\tilde{f}_\alpha(f(\boldsymbol{v})) &= \tilde{f}_\alpha(\frac{1}{1-v_{-1}}(v_0,\ldots,v_s)) \\ &:= (\alpha, (1-\alpha)(v_0,\ldots,v_s))\end{aligned}$$

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

Thus $f(\cdot)$ takes a distribution over S (for which absorption is not certain) and conditions it on non-absorption. $\tilde{f}_\alpha(\cdot)$ takes a distribution over C and maps it to a distribution in S for which the relative probabilities for being in any two states in C remain constant.

Lemma 2.1 $f(\tilde{f}_\alpha(\boldsymbol{v})P) = f(\tilde{f}_\beta(\boldsymbol{v})P)$ for any $P \in \mathcal{M}(P)$, independently of the values of α and β.

Proof.

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

□

Using $f(\cdot)$ it becomes possible to define the set of all possible distributions over C, conditioned on non-absorption, given $\mathcal{D}_0$, in the following way

Definition 2.6

$$\tilde{\mathcal{M}}_n^C := \{f(\boldsymbol{v}) | \boldsymbol{v} \in \tilde{\mathcal{M}}_n \setminus \{(1,0,\ldots,0)\}\} \tag{2.3}$$

and

$$\tilde{\mathcal{D}}_n^C := \{f(\boldsymbol{v}) | \boldsymbol{v} \in \tilde{\mathcal{D}}_n \setminus \{(1,0,\ldots,0)\}\}.$$

Theorem 2.2 For each $P \in \mathcal{M}(P)$ and $n \geq 0$,

$$\tilde{\mathcal{M}}_{n+1}^C \subseteq \tilde{\mathcal{M}}_n^C.$$

Proof. For each $P \in \mathcal{M}(P)$, and for $\mathcal{M}_0^C(P) = \mathcal{M}_0^C$, we have from (2.3) and Theorem 2.1 that

$$\begin{aligned}\tilde{\mathcal{M}}_{n+1}^C(P) = &\quad \{f(\boldsymbol{v}) \mid \boldsymbol{v} \in \tilde{\mathcal{M}}_{n+1}(P)\backslash\{(1,0,\ldots,0)\}\} \\ \subseteq &\quad \{f(\boldsymbol{v}) \mid \boldsymbol{v} \in \mathcal{M}_n(P)\backslash\{(1,0,\ldots,0)\}\} \\ = &\quad \tilde{\mathcal{M}}_n^C(P).\end{aligned}$$

By taking the union of both sides over all $P \in \mathcal{M}(P)$ the proof is complete. □

The following definition is therefore appropriate.

Definition 2.7

$$\tilde{\mathcal{M}}_\infty^C := \bigcap_{i=0}^{\infty} \tilde{\mathcal{M}}_i^C.$$

Equivalently

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

Thus $\tilde{\mathcal{M}}_\infty^C$ contains the possible distributions, conditioned on non-absorption as time goes to infinity, for all possible matrices from $\mathcal{M}(P)$ assuming nothing is known about the initial distribution.

Associated with each element $P \in \mathcal{M}(P)$ is a unique limiting conditional distribution $\boldsymbol{\alpha}_P$. In [4] it is proved that

$$\tilde{\mathcal{M}}_\infty^C = \bigcup_{P\in\mathcal{M}(P)} \tilde{\mathcal{M}}_\infty^C(P) = \bigcup_{P\in\mathcal{M}(P)} \boldsymbol{\alpha}_P \tag{2.4}$$

That is, although the correct element of $\mathcal{M}(P)$ is unknown, we know that the only possible distributions that can occur as time approaches infinity are the limiting conditional distributions of the elements of $\mathcal{M}(P)$. Since we have from Darroch and Seneta [7] that $\boldsymbol{\alpha}_P$ is reached independently of the initial distribution, we have that the right hand side of (2.4) represents the long-term behaviour of the chain, conditioned on non-absorption, independent of the choice of $\mathcal{D}_0^C$.

3 Imprecise Markov chains with non-constant transition matrix

In the case considered in Section 2, the long-term behaviour conditioned on non-absorption is easy to define, since such behaviour in the time-homogeneous case is well-known. In this section it is no longer assumed that the unknown transition matrix for time step n equals that for time step $m \neq n$. This corresponds in the precise case to the concept of time-inhomogeneous chains, the long-term behaviour of which is far less well understood.

A further condition is required in this case, namely that if $[P]_{ij} = 0$ for any $P \in \mathcal{M}(P)$ then $[Q]_{ij} = 0$ for all $Q \in \mathcal{M}(P)$. Thus a jump from state i to state j is either possible at all time steps, or impossible at all time steps. This is to prevent situations in which two or more transition matrices, each of which has C as a single communicating class, can be chosen from $\mathcal{M}(P)$ which, when multiplied, form a matrix for which C is *not* a single communicating class. It is certainly true that such matrices exist, but it is possible that they may already be disqualified by the conditions in Section 1.1 (most critically the assumption of independence), making this new condition unnecessary. Work is currently being conducted into ascertaining whether or not the new condition is redundant.

Definition 3.1 The set of possible n step transition matrices $\mathcal{M}^n(P)$ is defined as follows:

$$\mathcal{M}^n(P) := \{P_1 P_2 \dots P_n |\ P_i \in \mathcal{M}(P)\}.$$

Definition 3.2 The set $\mathcal{M}(P)$ is referred to as *regular* if for some n every $P \in \mathcal{M}^n(P)$ has only strictly positive elements. Further, the set $\mathcal{M}(P)$ is referred to as *conditionally regular on* C if for some r every $P \in \mathcal{M}^r(P)$ has all elements $[P]_{ij}$ strictly positive, where $i \in C$, $j \in S$.

Lemma 3.1 All matrices which belongs to the set $\mathcal{M}^{s+1}(P)$ are conditionally regular on C.

Proof. Any matrix P_{s+1} contained in $\mathcal{M}^{s+1}(P)$ represents 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, and each state in C is aperiodic. There must therefore be a path of n states, denoted $\{a_k\}_{k=1,\dots,n}$, strictly 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 a jump from state i to state j is either possible or not at a given time step *independent of that time step*. 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 there remains no duplicated value in the path, which forces $n \leq s-1$. 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 as a single communicating class, and thus cannot contain a column of zeroes.

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 $P(X(s+1) = j | X(0) = i) > 0$. □

$\mathcal{M}_0$ and $\mathcal{D}_0$ are defined just as they were in Section 2. Since a constant transition matrix can no longer be assumed, the following definitions are required.

Definition 3.3

$$\mathcal{M}_n := \{\boldsymbol{v}P | \boldsymbol{v} \in \mathcal{M}_{n-1}, P \in \mathcal{M}(P)\}$$

and

$$\mathcal{D}_n := \{\boldsymbol{v}P | \boldsymbol{v} \in \mathcal{D}_{n-1}, P \in \mathcal{M}(P)\}.$$

Furthermore

$$\mathcal{M}_n^C := \{f(\boldsymbol{v}) | \boldsymbol{v} \in \mathcal{M}_n \setminus \{(1, 0, \dots, 0)\}\}$$

and

$$\mathcal{D}_n^C := \{f(\boldsymbol{v}) | \boldsymbol{v} \in \mathcal{D}_n \setminus \{(1, 0, \dots, 0)\}\}.$$

It should be clear that

$$\tilde{\mathcal{M}}_n^C \subseteq \mathcal{M}_n^C, \forall n > 0 \tag{3.1}$$

and moreover that

$$\tilde{\mathcal{M}}_1^C = \mathcal{M}_1^C$$

where $\tilde{\mathcal{M}}_n^C$ is as defined in (2.3).

Lemma 3.2

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

and

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

Proof. $\boldsymbol{v} \in \mathcal{D}_{n-1}^C \Rightarrow \tilde{f}_\alpha(\boldsymbol{v}) \in \mathcal{D}_{n-1}$ for some $\alpha \in [0,1)$ by definition. Thus $\tilde{f}_\alpha(\boldsymbol{v})P \in \mathcal{D}_n^C$. By Lemma 2.1, however $f(\tilde{f}_\alpha(\boldsymbol{v})P) = f(\tilde{f}_\beta(\boldsymbol{v})P)$ for any $\beta \in [0,1)$, and so in fact $f(\tilde{f}_\alpha(\boldsymbol{v})P) \in \mathcal{D}_n^C$ independently of our choice of α. □

It is proven in [13] that

$$\mathcal{M}_{n+1} \subseteq \mathcal{M}_n$$

making the following definition appropriate.

Definition 3.4

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

It is proven in [5] that

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

so absorption is certain even when the transition matrix is unknown and can change between time steps. Once again the long-term behaviour of the chain conditioned on non-absorption is considered.

It is proved in [5] (in an almost identical manner to Theorem 2.2) that

$$\mathcal{M}_{n+1}^C \subseteq \mathcal{M}_n^C \tag{3.2}$$

and hence the following definition is appropriate.

Definition 3.5

$$\mathcal{M}_\infty^C := \bigcap_{n=0}^{\infty} \mathcal{M}_n^C.$$

Definition 3.6 A set of distributions $\mathcal{M}$ is denoted a *conditionally invariant set of distributions*, henceforth known as CISD, if

$$f(\tilde{f}_\alpha(\mathcal{M}) \cdot \mathcal{M}(P)) = \mathcal{M}$$

for some α and therefore for every $\alpha \in [0,1)$, where $\cdot$ represents an element-wise product.

Thus if at any time-step the set of possible distributions over C is a CISD every subsequent time-step will have an identical set of possible distributions over C. Note that $\mathcal{M}_\infty^C$ must be a CISD By Lemma 2.1.

$\mathcal{M}_\infty^C$ describes the behaviour of the chain, conditioned on non-absorption, as time approaches infinity, assuming that there is nothing whatsoever that can be said regarding the initial distribution over C. An important property of the limiting conditional distribution in the precise case, however, is that the behaviour of the chain, conditioned on non-absorption, tends toward it independently of the choice of initial distribution over C. In what follows we outline the method by which the generalisation of this property can be proved.

Definition 3.7 Two sets of distributions over S, $\mathcal{M}$ and $\mathcal{N}$, are described as *conditionally equal* if $f(\mathcal{M}) = f(\mathcal{N})$, where $f(\mathcal{M}) := \{f(\boldsymbol{v}) | \boldsymbol{v} \in \mathcal{M}\}$.

A non-symmetric distance measure $d(\cdot,\cdot)$ between two sets of distributions over S is defined in [5], where $d(\mathcal{M},\mathcal{N}) = 0$ if and only if for every $\boldsymbol{v} \in \mathcal{M}$ there is a $\boldsymbol{w} \in \mathcal{N}$ such that $f(\boldsymbol{v}) = f(\boldsymbol{w})$.

Corollary 3.1 Let $\mathcal{M}$ and $\mathcal{N}$ be closed sets of distributions. Then $f(\mathcal{M}) \subseteq f(\mathcal{N})$ if and only if $d(\mathcal{M},\mathcal{N}) = 0$.

Proof. $f(\mathcal{M}) \subseteq f(\mathcal{N})$ implies that for every $f(\boldsymbol{v}) \in \mathcal{M}$ there exists $\boldsymbol{w} \in \mathcal{N}$ such that $f(\boldsymbol{v}) = f(\boldsymbol{w})$. Thus $d(\mathcal{M},\mathcal{N}) = 0$.

Let $d(\mathcal{M},\mathcal{N}) = 0$. By the above assertion, for every $\boldsymbol{v} \in \mathcal{M}$ there exists $\boldsymbol{w} \in \mathcal{N}$ such that $d(\boldsymbol{v},\boldsymbol{w}) = 0$. Thus $f(\mathcal{M}) \subseteq f(\mathcal{N})$.

□

It is proven in [5] that, under the conditions given in this paper

$$d(\mathcal{M} \cdot \mathcal{M}(P), \mathcal{N} \cdot \mathcal{M}(P)) < d(\mathcal{M},\mathcal{N}) \tag{3.3}$$

and

$$f(\mathcal{M}) = f(\mathcal{M}') \Rightarrow d(\mathcal{M},\mathcal{N}) = d(\mathcal{M}',\mathcal{N}) \tag{3.4}$$

for any set of distributions $\mathcal{N}$.

Definition 3.8 Let $\mathcal{M}$ be a compact set of distributions and $\mathcal{M}(P)$ a set of transition matrices that are conditionally regular on C. Then $\mathcal{M}$ is a *fixed set of $\mathcal{M}(P)$ conditionally on C* if $f(\mathcal{M} \cdot \mathcal{M}(P)) = f(\mathcal{M})$, or equivalently, if $\mathcal{M}$ and $\mathcal{M} \cdot \mathcal{M}(P)$ are conditionally equal on C.

It is important to note that if $\mathcal{M}$ is a fixed set of $\mathcal{M}(P)$ conditionally on C, then $f(\mathcal{M})$ must be a conditionally invariant set of distributions.

Theorem 3.1 Let $\mathcal{M}$ and $\mathcal{N}$ be conditionally fixed sets of $\mathcal{M}(P)$ on C. Then they are conditionally equal on C.

Proof. It follows from Corollary 3.1 that the sets $\mathcal{M}$ and $\mathcal{N}$ are conditionally equal on C if and only if $d(\mathcal{M},\mathcal{N}) = d(\mathcal{N},\mathcal{M}) = 0$. Suppose that one of the distances is greater than 0, say $d(\mathcal{M},\mathcal{N}) > 0$. By the

assertion that both sets are conditionally fixed sets, we have that $f(\mathcal{M}) = f(\mathcal{M} \cdot \mathcal{M}(P))$ and $f(\mathcal{N}) = f(\mathcal{N} \cdot \mathcal{N}(P))$ Then, by Corollary 3.1, (3.3) and (3.4), $d(\mathcal{M},\mathcal{N}) = d(\mathcal{M} \cdot \mathcal{M}(P),\mathcal{N}) = d(\mathcal{M} \cdot \mathcal{M}(P),\mathcal{N} \times \mathcal{M}(P)) < d(\mathcal{M},\mathcal{N})$, which is a contradiction. □

Thus we have that there can be only one conditionally invariant set of distributions for a given imprecise Markov chain. Finally, [5] goes on to prove that convergence to this set is certain, conditioned upon non-absorption, irrespective of the choice of $\mathcal{D}_0^C$.

These results confirm the CISD as the imprecise analog to the QSD. Not only does setting $\mathcal{D}_0^C = \mathcal{M}_\infty^C$ ensure that $\mathcal{D}_n^C = \mathcal{M}_\infty^C$, for all n, but the set of possible distributions tends towards $\mathcal{M}_\infty^C$ no matter what initial distributions are allowed.

We now discuss the method offered in [3], and explain why we do not make use of it here. The results presented in Section 3 are based on the notion of regularity defined in Definition 3.2. In this sense the results on convergence directly generalise those found in [7], where the analogous notion of regularity is used in the precise case.

Two important further insights for the case of unconditional convergence of imprecise Markov chains are found in [3], which suggest that the concept of regularity that we use might be too strong. First, de Cooman *et al.* show that even the concept of regularity itself can be transferred to the imprecise case in a weaker form, which suggests that there might be different types of convergence with different properties. However, their approach is substantially different from ours, where the main difference is that they represent imprecision in terms of lower and upper expectation operators instead of sets of probabilities and moreover, the calculations of the distributions at further time steps are done by the use of so called backwards recursion.

While in the case where sets of probabilities are convex, which certainly is the most important case, the representation with expectation operators coincides with the approach with sets of probabilities, our approach is more general if sets of probabilities are not assumed to be convex. Our stronger notion of regularity seems to be necessary in this case to assure convergence. The second problem with efficiently applying the approach taken in [3] to studying convergence under conditioning on non-absorption is that it is not obvious to us how the conditioning that must take place at every step would be combined with the backward recursion method. This effectively means that we do not see how the step performed in Lemma 2.1, which is shown to be easy using the forward calculations, could be done using the backwards recursion. Of course, while the chain is still finite, conditioning can be done at an arbitrary step n, but when convergence is in question as n tends to infinity, it is not clear how and where conditioning can be done, as it is clearly too late to condition at infinity where absorption takes place with certainty. Despite the above difficulties we believe that combining our results with those of de Cooman *et al.* is possible in some way, which is a possible path of our future research.

The second important insight given in [3] is that, in the case without conditioning and even in the precise case, instead of regularity a weaker condition called "regular absorption" is sufficient to assure unique convergence, which also seems to be possible to apply to the problem of unique convergence under conditioning.

4 Comparison between the models

In this section we consider two examples. In the first, movement from all three transient states exhibits imprecise behaviour, but the bounds on that behaviour are comparatively tight. In the second example, movement from only one transient state exhibits imprecise behaviour, but the bounds on that behaviour are comparatively much wider. In each example we consider the difference between applying the model given in Section 2 and that given in Section 3. Note that throughout this section $\mathcal{M}_0^C$ is used as the set of possible initial distributions over C.

In this section simplex diagrams (see e.g. Walley [16]) are used to graphically represent probability distributions with three elements. A simplex diagram is an equilateral triangle with 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. Thus the set $\mathcal{M}_0^C$ is represented by the whole simplex diagram.

Example 1

Consider a time-homogeneous birth-death process $\mathcal{X}$ with state space $\Omega = \{-1\} \cup C$ where $C = \{0, 1, 2\}$. The set of all possible one-step transition matrices $\mathcal{M}(P)$ is given as follows. Each $P \in \mathcal{M}(P)$ takes the form

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ a & 0 & 1-a & 0 \\ 0 & b & 0 & 1-b \\ 0 & 0 & c & 1-c \end{pmatrix}$$

where $a \in [0.1, 0.3]$, $b \in [0.5, 0.6]$, and $c \in [0.67, 0.73]$.

Generating either $\tilde{\mathcal{M}}_n^C$ or $\mathcal{M}_n^C$ in their entirety for this example (or any other) is a non-trivial task. There are

several alternative methods that can be used to gain sensible approximations. For instance, the maximum and minimum values of each element of the vectors contained in $\tilde{\mathcal{M}}_n^C$ and $\mathcal{M}_n^C$ can be calculated. The simplex diagrams in Figure 1 below show such approxiations for $\tilde{\mathcal{M}}_n^C$ for $n = 2, 3, 4$ (left column, from top to bottom), and $\mathcal{M}_n^C$, also for $n = 2, 3, 4$ (right column, from top to bottom). Bounds have also been approximated for the sets $\tilde{\mathcal{M}}_{100}^C$ and $\mathcal{M}_{100}^C$. These were found by randomly generating 1000 100-step transition matrices for each of the two cases, multiplying each one by $(1, 0, 0), (0, 1, 0)$ and $(0, 0, 1)$, and finding the overall maximum and minimum of each element. The 100th time step is an excellent approximation to the case as time approaches infinity.

Recall that it is known that the size of the bounded areas are non-increasing from time step n to $n+1$, from (2.3) and (3.2). Figure 1 demonstrates these properties very well. Note also that, as expected, for each times step the bounded areas on the right are larger than those on the left. This again is exactly what was expected given (3.1), and moreover is consistent with the idea that more can be said about the long term behaviour for the case where the transition matrix is constant than can be said for the case where the transition matrix is potentially non-constant between time steps. One could say that the second case allows for "more imprecision," in that less can be assumed about the underlying process.

It is important to note that the variable a does in fact play a role in the example, despite the fact that by conditioning on non-absorption we implicitly assume that every transition from state 0 must have been to state 1. This can be easily seen by noting that

$$f((0,x,y,z)\begin{pmatrix} 1 & 0 & 0 & 0 \\ a & 0 & 1-a & 0 \\ 0 & b & 0 & 1-b \\ 0 & 0 & c & 1-c \end{pmatrix})$$

$$= (\frac{1}{1-ax})(by, (1-a)x + cz, (1-b)y + (1-c)z)$$

and therefore conditioning on non-absorption does not prevent a from contributing to the distribution over C.

Example 2

Consider a time-homogeneous birth-death process $\mathcal{X}$ with state space $\Omega = \{-1\} \cup C$ where $C = \{0, 1, 2\}$. The set of all possible one-step transition matrices $\mathcal{M}(P)$ is given as follows. Each $P \in \mathcal{M}(P)$ has the following form.

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0.6 & 0 & 0.4 & 0 \\ 0 & d & 0 & 1-d \\ 0 & 0 & 0.7 & 0.3 \end{pmatrix}$$

where $d \in [0.37, 0.73]$. The diagrams were created using identical methods to those used in the first example.

The same comments regarding Figure 1 also apply to Figure 2. It should also be noted that in the second example more can be said about the probability of being in state 1, conditioned on non-absorption. as time approaches infinity, but less can be said about the probabilities of being in states 1 or 3. This may be explained as follows. Note that in the method used in Section 2, the bounds upon $\tilde{\mathcal{M}}_\infty^C$ are simply the bounds upon the set $\bigcup_{P\in\mathcal{M}(P)} \boldsymbol{\alpha}_P$ (see (2.4)). Thus the bounds approximated in the bottom-left simplex of Figure 1 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 approximated in the bottom-left simplex of Figure 2 relate to the three elements of a vector function with one unknown, d, which has a comparatively large range. The elongated, thinner shape in Figure 2 is thus intuitively unsurprising, though the validity of this intuition can be questioned, as d will eventually appear in all transition proabilities given enough jumps.

The final point regarding Figures 1 and 2 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 completely 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 Section 3 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 comparatively little work has been done.

Note that is would also be 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 $\mathcal{M}(P)$. These can then be used to create sets of vectors to approximate $\tilde{\mathcal{M}}_n^C$ and $\mathcal{M}_n^C$. The drawback to this method is that it rapidly becomes computationally heavy. 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 $\mathcal{M}(P)$ 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.

5 Concluding remarks

In this paper we have summarised two methods in which imprecision can be applied to the theory of Markov chains, and discussed that in each case, given certain conditions and conditioned on non-absorption, convergence to a unique conditionally invariant set is guaranteed, and that using this set as the set of initial distributions, the possible behaviour of the chain is unchanging over time. It has also been demonstrated that, by considering this extension of the QSD, it is possible to say something regarding the long-term behaviour, conditioned on non-absorption, of finite Markov chains with an absorbing state, in situations in which the transition matrix at each time-step is not known precisely. Moreover, it has been shown that much can be said even in situations where the transition matrix is not assumed to be constant over time, and in which there is no transient state from which the transition probabilities are known precisely. This in turn means that the model presented in Section 3 could be applied when considering the long-term behaviour of certain precise time-inhomogeneous chains.

6 Acknowledgements

We are grateful to two referees for their detailed comments which helped to improve this paper.

References

[1] Beer, G. (1993) *Topologies On Closed And Closed Convex Sets*, Kluwer Academic Publishers, Dordrecht.

[2] Coolen-Schrijner, P. & van Doorn, E. (2006) Quasi-stationary distributions for a class of discrete-time Markov chains, *Methodology and Computing in Applied Probability*, **8**, 449-465.

[3] de Cooman, G., Quaeghebeur, E. & Miranda, E. (2009) Imprecise Markov chains and their limit behaviour. *Pre-print at arXiv.org*, arXiv:0801.0980.

[4] 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.

[5] Crossman, R.J. & Škulj D. (2009) Finite discrete-time Markov chains with absorbing state and imprecise probability, in submission, copy available upon request.

[6] Crossman, R.J. (2009) *Limiting Conditional Distributions: Imprecision and Relation to Hazard Rate*, PhD Thesis, Durham University, in submission, copy available upon request.

[7] Darroch, J.N. & Seneta, E. (1965) On quasi-stationary distributions in absorbing discrete-time finite Markov chains, *Journal of Applied Probability* **2**, 88-100.

[8] Hartfiel, D.J. (1998) *Markov Set-Chains*, Springer-Verlag, Berlin, Heidelberg, New York.

[9] Kijima, M. (1997) *Markov Processes for Stochastic Modeling*. Chapman & Hall, London.

[10] Kozine, I.O. & Utkin, L.V. (2002) Interval-valued finite Markov chains, *Reliable Computing 8*, **2**, 97-113.

[11] Pakes, A. (1987) Limit theorems for the population size of a birth and death process allowing catastrophes, *Journal of Mathematical Biology*, 25(3), 307-325.

[12] Parsons, R. & Pollett P. (1987) Quasistationary distributions for auto-catalytic reactions, *Journal of Statistical Physics*, 46(1-2), 249-254.

[13] Škulj, D. (2006) Finite discrete time Markov chains with interval probabilities, *Soft Methods for Integrated Uncertainty Modelling*, Springer, Berlin.

[14] Škulj, D. (2007) Finite discrete time Markov chains with interval probabilities. In J. Lawry, E.Miranda, A. Bugarin, S.Li, M.A. Gil, P.Grzegorzewski, O.Hryniewicz (esd.), *Soft Methods for Integrated Uncertainty Modelling*, 299-306. Springer, Berlin.

[15] Walley, P. (1981) Coherent lower (and upper) probabilities. Department of Statistics, University of Warwick.

[16] Walley, P. (1991) *Statistical Reasoning With Imprecise Probabilities*, Chapman & Hall, London.

[17] Weichselberger, K. (2000) The theory of interval-probability as a unifying concept for uncertainty, *International Journal of Approximate Reasoning*, **24** 149-170.

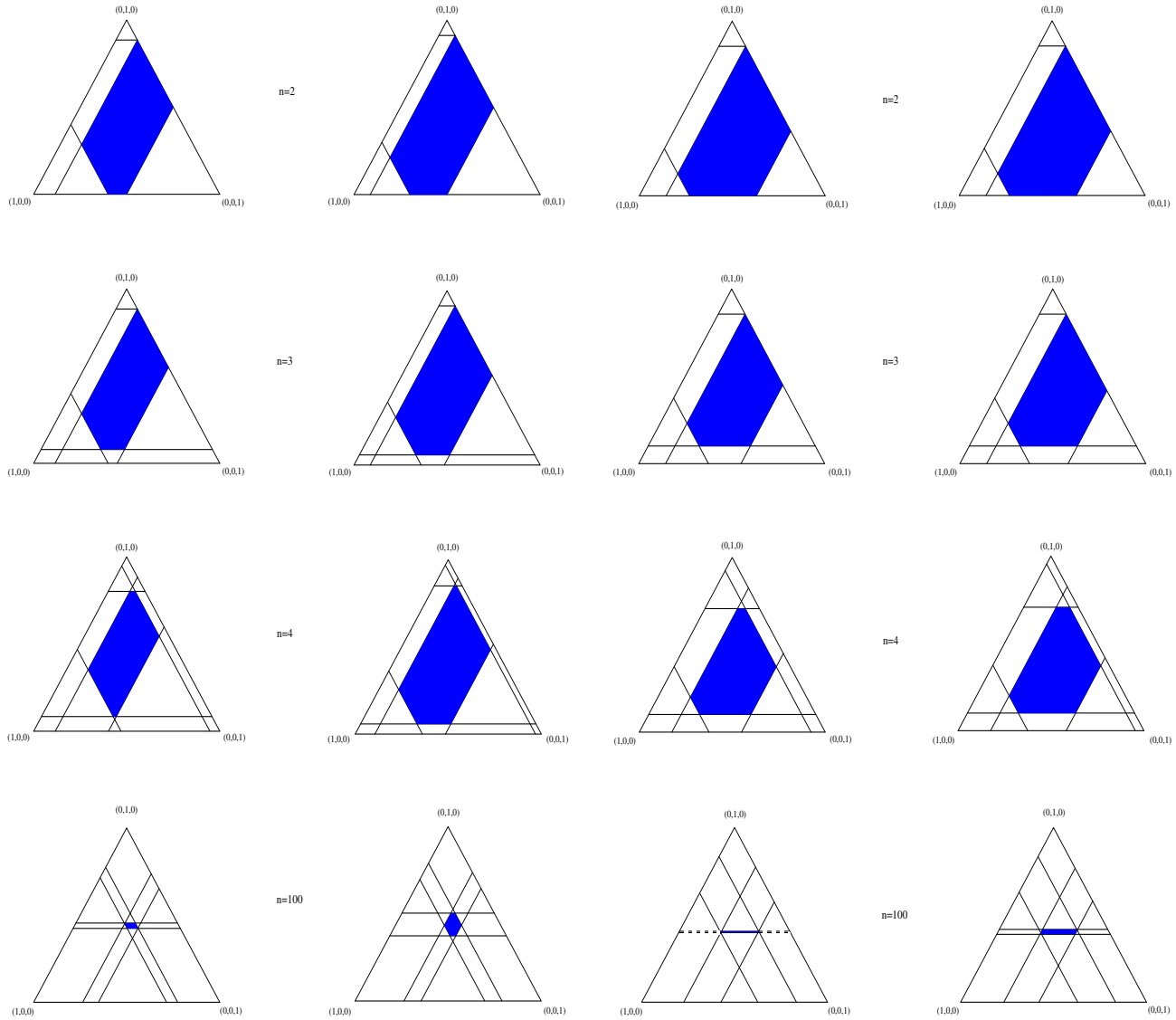

Figure 1: Bounds for the sets $\tilde{\mathcal{M}}_n^C$ and $\mathcal{M}_n^C$, all for $n = 2, 3, 4$ and 100.

Figure 2: Bounds for the sets $\tilde{\mathcal{M}}_n^C$ and $\mathcal{M}_n^C$, all for $n = 2, 3, 4$ and 100.