Imprecise Markov Chains with Absorption

Richard J. Crossman, Frank P. A. Coolen, and Pauline Coolen-Schrijner

Durham University, UK

Damjan Škulj

University of Ljubljana, Slovenia

1 The Quasi-Stationary Distribution

Consider a discrete time Markov chain $\mathcal{X} = \{X(n), n = 0, 1, \ldots\}$ with finite state space $S = \{-1\} \cup C$, where $C = \{0, \ldots, s\}$ is a single communicating class with all states aperiodic. Assuming -1 can be reached from C, absorption is certain, making the *limiting distribution* $(1, 0, \ldots, 0)$. Instead of considering the limiting distribution, then, we condition at non-absorption at each time step. The behaviour of the chain as time approaches infinity, conditioned on non-absorption, tends towards the *limiting conditional distribution* (LCD). If used as the initial distribution over C, then the distribution over C, conditioned on non-absorption, remains the same for all time steps. This initial distribution is known as the *quasi-stationary distribution* (QSD). This distribution is of interest with regard to population models, chemical reactions, and disease progression, amongst other things.

2 Imprecise Markov Chains

We introduce imprecision into the model by choosing each row $\boldsymbol{r}^{(i)}$ of the transition matrix at time n from a closed set of probability distributions $\boldsymbol{R}^{(i)}$. Hence

$$P^{(n)} = \begin{pmatrix} \boldsymbol{r}^{(-1)} \\ \boldsymbol{r}^{(0)} \\ \vdots \\ \boldsymbol{r}^{(s)} \end{pmatrix}$$

where $\boldsymbol{r}^{(i)} \in \boldsymbol{R}^{(i)}$. We force $\boldsymbol{R}^{(i)} = (1, 0, \ldots, 0)$ to ensure -1 is an absorbing state, and exclude $(\delta_{-1i}, \delta_{0i}, \ldots, \delta_{si})$ from $\boldsymbol{R}^{(i)}$, $i \geq 0$ to ensure -1 is the only absorbing state. The $\boldsymbol{R}^{(i)}$ are also chosen to fulfil the following criteria:

1. every possible combination must leave C a single communicating class;
2. every possible combination must leave all states in C aperiodic;
3. a jump from state i to state j must be possible for all combinations, or impossible for all combinations.

The first two conditions ensure a unique LCD exists for each possible transition matrix. The third condition prevents multiplications of possible matrices from not having a unique LCD.

The set of matrices fulfilling these conditions is denoted $\mathcal{M}(P)$.

We also define the functions $f(v_{-1}, v_0, \ldots, s) = \frac{1}{1-v_{-1}}(v_0, \ldots, v_s)$ and $\tilde{f}_\alpha(v_0, \ldots, v_s) = (\alpha, (1-\alpha)(v_0, \ldots, v_s))$, for $\alpha \in (0, 1]$. Thus $f(\cdot)$ conditions a vector on non-absorption, and $\tilde{f}_\alpha(\cdot)$ takes a distribution over C and converts it to a distribution over S with probability of absorption α. It is important to note that for the vector $\boldsymbol{w}$ over C

$$f(\tilde{f}_\alpha(\boldsymbol{w})P) = f(\tilde{f}_\beta(\boldsymbol{w})P)$$

for any $P \in \mathcal{M}(P)$ and any $\alpha, \beta \in (0, 1]$. This allows us to take a distribution conditioned on non-absorption, and find the equivalent distribution conditioned on non-absorption at the next time step.

3 Time-Homogeneous Case

When using imprecision in the time-homogeneous case we assume that there is one matrix $P \in \mathcal{M}(P)$ that describes the behaviour of the chain at all time steps. We denote by $\mathcal{M}_0$ all possible initial distributions over S except $(1, 0, \ldots, 0)$. The set defined below

$$\mathcal{M}_n^C(P) := \{f(\boldsymbol{v}P^n) : \boldsymbol{v} \in \mathcal{M}_0\}$$

describes all possible distributions at time n for a given element of $\mathcal{M}(P)$. Thus

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

includes all possible distributions at time n. It can be shown that

$$\tilde{\mathcal{M}}_n^C(P) \subseteq \tilde{\mathcal{M}}_{n-1}^C(P)$$

and hence

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

allowing us to define

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

and

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

Note that $\mathcal{M}_\infty^C(P)P = \mathcal{M}_\infty^C(P)$. Thus $\tilde{\mathcal{M}}_\infty^C$ is a superset of the actual behaviour of the chain, conditioned on non-absorption, as time approaches infinity, and all elements of the superset represent possible behaviour for a given element of $\mathcal{M}(P)$.

It is not difficult to show that for a smaller set of possible initial distributions $\mathcal{D}_0$ where

$$\mathcal{D}_n^C(P) = \{f(\boldsymbol{v}P^n) : \boldsymbol{v} \in \mathcal{D}_0\}$$

and

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

$\tilde{\mathcal{M}}_\infty^C$ is still a superset of the behaviour of the chain, conditioned on non-absorption, as time approaches infinity, and each element of this set will still represent the possible long-term behaviour of the chain, conditioned on non-absorption, for a given element of $\mathcal{M}(P)$. This is despite $\mathcal{D}_n^C(P) \subseteq \mathcal{D}_{n-1}^C(P)$ not necessarily holding.

4 Time-Inhomogeneous Case

In the time-inhomogeneous case we allow a new (not necessarily distinct) matrix to be chosen from $\mathcal{M}(P)$ at each time step. Once again denoting by $\mathcal{M}_0$ all possible initial distributions over S except $(1, 0, \ldots, 0)$. The set defined below

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

includes all possible distributions at time n. It can be shown that

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

allowing us to define

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

Note that

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

where $\cdot$ represents element-wise multiplication. Thus $\mathcal{M}_\infty^C$ represents the behaviour of the chain, conditioned on non-absorption, as time approaches infinity.

Once again, it can be shown that or a smaller set of possible initial distributions $\mathcal{D}_0$ where

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

and $\mathcal{D}_0^C := f(\mathcal{D}_0)$, then the behaviour of the chain, conditioned on non-absorption, as time approaches infinity is still $\mathcal{M}_\infty^C$, even though $\mathcal{D}_n^C \subseteq \mathcal{D}_{n-1}^C$ does not necessarily hold.

We therefore have that $\mathcal{M}_\infty^C$ is, in fact, the time-inhomogeneous imprecise generalisation of the LCD, because this set is reached irrespective of the choice of set of possible initial distributions. Moreover, if $\mathcal{D}_0^C = \mathcal{M}_\infty^C$, then at each time step the set of possible distributions, conditioned on non-absorption, is still $\mathcal{M}_\infty^C$. Thus $\mathcal{M}_\infty^C$ is also the generalisation of the QSD.

5 Comparing Approaches

In our paper we compare the effects of assuming time-homogeneity and time-inhomogeneity for the same set of possible transition matrices $\mathcal{M}(P)$. Two examples are given below, in which bounds upon the components of $\mathcal{M}_n^C$ have been calculated exactly for $n = 2, 3, 4$, and approximated for $n = 100$.

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 $c \in [0.67, 0.73]$.

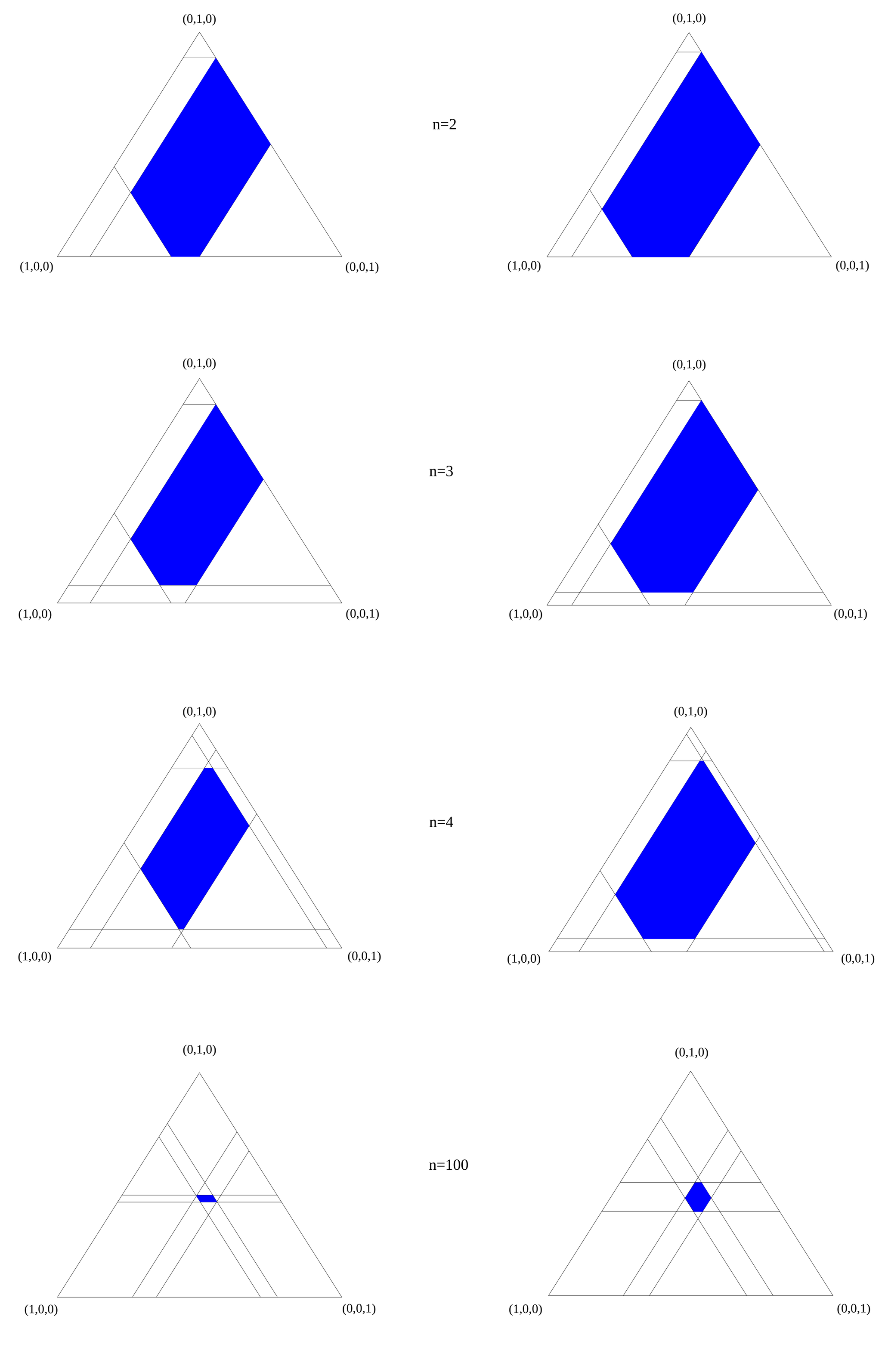

Figure 1: Simplex diagrams for Example 1

We know already that the size of the bounded areas are non-increasing from time step n to $n+1$. Figure 1 demonstrates this property 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 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.

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 $d \in [0.37, 0.73]$.

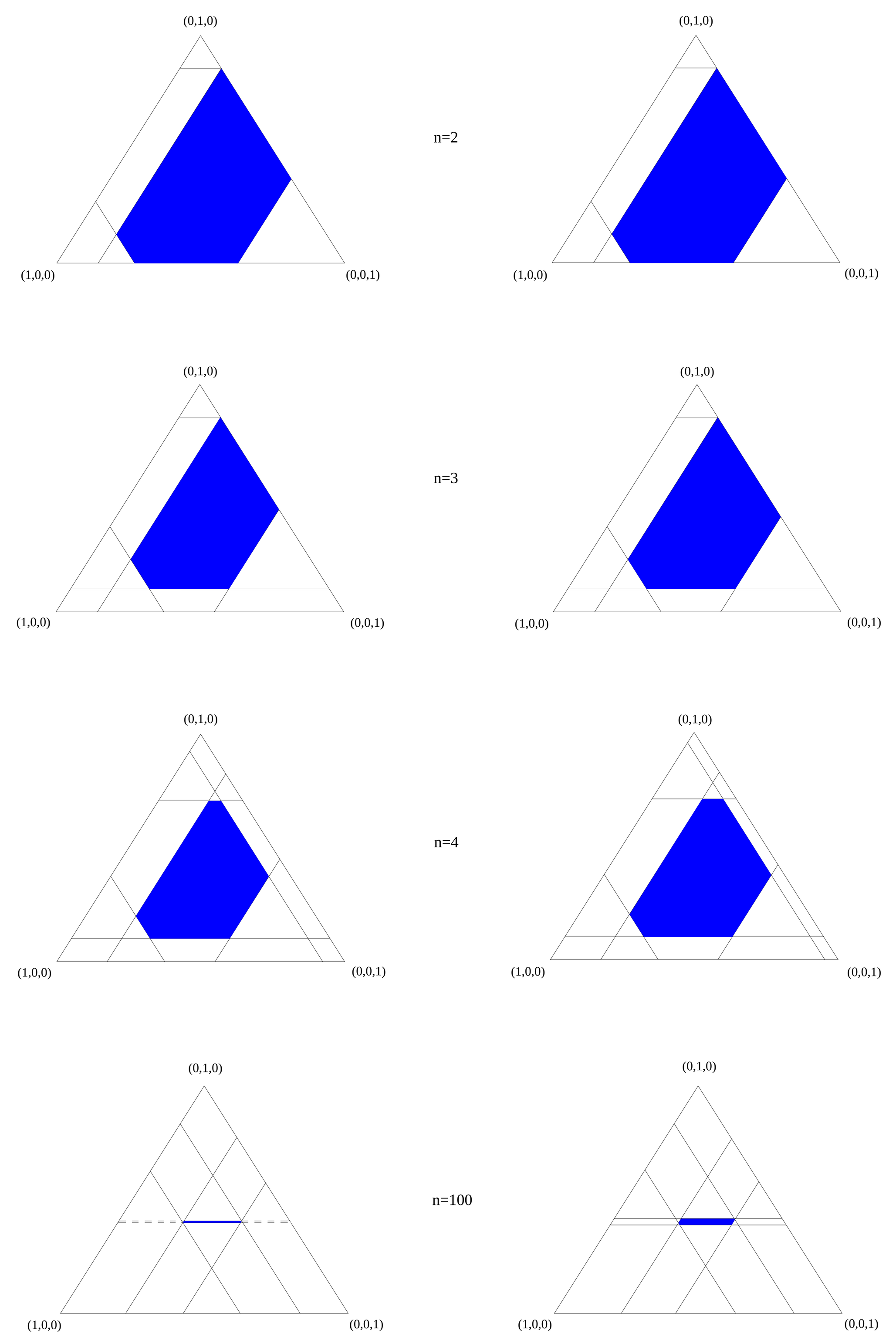

Figure 2: Simplex diagrams for Example 2

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 2, conditioned on non-absorption, as time approaches infinity, but less can be said about the probabilities of being in states 1 or 3. Further, 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.

References

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

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

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

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

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

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

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

[8] 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.

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

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

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