Markov Processes Relat. Fields **15**, 191–204 (2009)

Markov Processes and Related Fields

Quasi-Stationary Distributions for Reducible Absorbing Markov Chains in Discrete Time

E.A. van Doorn[1] **and P.K. Pollett**[2]

[1] Department of Applied Mathematics, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands. E-mail: e.a.vandoorn@utwente.nl
[2] Department of Mathematics, The University of Queensland, Qld 4072, Australia. E-mail: pkp@maths.uq.edu.au

Received November 17, 2008, revised April 6, 2009

Abstract. We consider discrete-time Markov chains with one coffin state and a finite set S of transient states, and are interested in the limiting behaviour of such a chain as time n tends to infinity, conditional on survival up to n. It is known that, when S is irreducible, the limiting conditional distribution of the chain equals the (unique) quasi-stationary distribution of the chain, while the latter is the (unique) ρ-invariant distribution for the one-step transition probability matrix of the (sub)Markov chain on S, ρ being the Perron – Frobenius eigenvalue of this matrix. Addressing similar issues in a setting in which S may be reducible, we identify all quasi-stationary distributions and obtain a necessary and sufficient condition for one of them to be the unique ρ-invariant distribution. We also reveal conditions under which the limiting conditional distribution equals the ρ-invariant distribution if it is unique. We conclude with some examples.

Keywords: absorbing Markov chain, ρ-invariant distribution, limiting conditional distribution, survival-time distribution

AMS Subject Classification: Primary 60J10, Secondary 15A18

1. Introduction

We consider discrete-time Markov chains with one coffin state and a finite set S of transient states, and are interested in the limiting behaviour of such a chain as time n tends to infinity, conditional on survival up to n. It is known that, when S is irreducible, the limiting conditional distribution of the chain equals the (unique) quasi-stationary distribution of the chain, while the latter is the (unique) ρ-invariant distribution for the one-step transition probability matrix of the (sub)Markov chain on S, ρ being the Perron – Frobenius eigenvalue of this matrix. Our aim in this paper is to investigate to what extent these results can be generalized if we allow S to be reducible.

Motivated by the modelling approaches in survival analysis of Aalen and Gjessing [1, 2] and Steinsalz and Evans [14], we addressed similar problems in a continuous-time setting in [7], which may be viewed as a companion paper to the paper at hand. It appears that the current discrete-time setting imposes additional problems as a consequence of the possible occurrence of periodicity. On the other hand, application of the discrete-time results of Lindqvist [8] allows a more succinct derivation of the main results.

The plan of the paper is as follows. In the next section we will introduce the relevant concepts and obtain a preliminary result (Theorem 2.1). In Section 3 we summarize what is known about quasi-stationary distributions and limiting conditional distributions when the state space of the Markov chain is irreducible. Subsequently, in the Sections 4 and 5, these results will be generalized in the setting of a state space that may be reducible. In particular, we will identify all quasi-stationary distributions and obtain a necessary and sufficient condition for one of them to be the unique ρ-invariant distribution in Section 4. In Section 5 we reveal conditions under which the limiting conditional distribution equals the ρ-invariant distribution if the latter is unique.

Finally, we discuss two examples in Section 6. The first example concerns a model for two competing species on a habitat patch. Our main concern is the question which of the species have survived given that the patch has been inhabited for a long time. The second example is a pure-death chain with killing as a model for describing the status of a patient suffering from a progressive disease. Application of the results of the Sections 4 and 5 enables us to obtain the distribution of the status of a long-term surviving patient.

2. Preliminaries

Let $\mathcal{X} \equiv \{X(n),\ n = 0, 1, \ldots\}$ denote a homogeneous discrete-time Markov chain on a state space $\{0\} \cup S$ consisting of an absorbing state 0 (the *coffin state*) and a *finite* set of transient states $S := \{1, 2, \ldots, s\}$. We let $P \equiv (P_{ij})$ be the matrix of one-step transition probabilities of the (sub)Markov chain on S and write

$$\kappa_i := 1 - \sum_{j \in S} P_{ij} \geq 0, \quad i \in S,$$

for the probabilities of absorption into state 0 (*killing probabilities*). Since all states in S are assumed to be transient, at least one of the killing probabilities must be positive and eventual killing is certain.

In what follows we identify a probability distribution $\{u_i\}$ over S with the row vector $\boldsymbol{u} \equiv (u_i, i \in S)$. We write $\mathbb{P}_i(\cdot)$ for the probability measure of the process when $X(0) = i$ and $\mathbb{E}_i(\cdot)$ for the expectation with respect to this measure. For any distribution $\boldsymbol{u}$ we let $\mathbb{P}_{\boldsymbol{u}}(\cdot) := \sum_i u_i \mathbb{P}_i(\cdot)$. The n-step transition probabilities of the process $\mathcal{X}$ are denoted by $P_{ij}(n) := \mathbb{P}_i(X(n) = j)$. Hence $P_{ij}(1) = P_{ij}$, and the matrix $P(n) := (P_{ij}(n),\ i, j \in S)$ of n-step transition probabilities satisfies $P(n) = P^n$.

We allow S to be reducible, so we suppose that S consists of the classes (maximal subsets of communicating states) $S_1, S_2, \ldots, S_L$, and let P_k be the submatrix of P corresponding to the states in S_k. Since we are interested in the long-term behaviour of $\mathcal{X}$, we will exclude trivial cases by assuming that at least one of the classes S_k is (in the words of Seneta [13]) *self-communicating*, that is,

$$\mathbb{P}_i(X(1) \in S_k) > 0, \quad i \in S_k. \tag{2.1}$$

Note that $P_k \neq O$, the zero matrix, unless S_k is *not* self-communicating, in which case S_k consists of a single state j with $P_{jj} = 0$.

We define a partial order on $\{S_1, S_2, \ldots, S_L\}$ by writing $S_i \prec S_j$ (or $S_j \succ S_i$) when S_i is *accessible* from S_j, that is, when there exists a sequence of states $k_0, k_1, \ldots, k_\ell$, $\ell \geq 1$, such that $k_0 \in S_j$, $k_\ell \in S_i$, and $P_{k_m k_{m+1}} > 0$ for every m. Note that $S_k \prec S_k$ if and only if S_k is self-communicating. We will assume in what follows that the states are labelled such that P is in lower block-triangular form (*Frobenius normal form*), so that

$$S_i \prec S_j \implies i \leq j. \tag{2.2}$$

As is well known (see, for example, [10, Section 8.3]), the eigenvalue $\rho \equiv \rho(P)$ with maximal real part (the *Perron – Frobenius eigenvalue* of P) is real and nonnegative. Noting that the matrices P_k reside on the diagonal of P, it follows easily that the set of eigenvalues of P is precisely the union of the sets of eigenvalues of the individual P_k's. So, letting $\rho_k := \rho(P_k)$, the Perron – Frobenius eigenvalue of P_k (so that ρ_k is real and nonnegative), we have $\rho = \max_k \rho_k$. Since $\rho_k > 0$ unless S_k is *not* self-communicating (see, for example, [10, p. 673]), our assumptions imply that $\rho_k > 0$ for at least one k. Moreover, since all states in S are transient we must have $\rho_k < 1$ for all k (see, for example, [10, p. 696]). As a consequence,

$$0 < \rho < 1. \tag{2.3}$$

A (proper) probability distribution $\boldsymbol{u} \equiv (u_i, i \in S)$ will be called *x-invariant* for P (on S) if $\boldsymbol{u}$ is a left x-eigenvector of P, that is,

$$\sum_{i \in S} u_i P_{ij} = x u_j, \quad j \in S. \tag{2.4}$$

We observe that an x-invariant distribution $\boldsymbol{u}$ for P satisfies

$$\sum_{i\in S} u_i P_{ij}(n) = x^n u_j, \quad j \in S,\ n \geq 0, \tag{2.5}$$

so that $\boldsymbol{u}$ is actually x^n-invariant for P^n for all $n \geq 1$. We let $T := \inf\{n \geq 0\colon X(n) = 0\}$ denote the *survival time* (or *killing time*) – the random variable representing the time at which killing occurs – and define the *killing probability* corresponding to a probability distribution $\boldsymbol{u} \equiv (u_i, i \in S)$ by

$$\kappa_{\boldsymbol{u}} := \mathbb{P}_{\boldsymbol{u}}(T = 1) = \sum_{i\in S} u_i \kappa_i. \tag{2.6}$$

By summing (2.4) over all $j \in S$, we see that

$$\boldsymbol{u} \ \text{ is } \ x\text{-invariant} \implies x = 1 - \kappa_{\boldsymbol{u}}. \tag{2.7}$$

A (proper) probability distribution $\boldsymbol{u} \equiv (u_i, i \in S)$ is said to be a *quasi-stationary distribution* for $\mathcal{X}$ if the distribution of $X(n)$, conditional on absorption not yet having taken place, is constant over n when $\boldsymbol{u}$ is the initial distribution, that is, for all $n \geq 0$, one has $\mathbb{P}_{\boldsymbol{u}}(T > n) > 0$ and

$$\mathbb{P}_{\boldsymbol{u}}(X(n) = j \mid T > n) = u_j, \quad j \in S. \tag{2.8}$$

We can now formulate the following theorem.

Theorem 2.1. *Let $\boldsymbol{u} \equiv (u_i, i \in S)$ represent a proper probability distribution over S, then the following statements are equivalent:*

(i) *$\boldsymbol{u}$ is a quasi-stationary distribution for $\mathcal{X}$;*

(ii) *$\boldsymbol{u}$ is ρ_k-invariant for P for some $k \in \{1, 2, \ldots, L\}$;*

(iii) *$\boldsymbol{u}$ is x-invariant for P for some $x > 0$;*

(iv) *for some $k \in \{1, 2, \ldots, L\}$ one has $\mathbb{P}_{\boldsymbol{u}}(X(n) = j) = \rho_k^n u_j$ for all $j \in S$, $n \geq 0$;*

(v) *for some $x > 0$ one has $\mathbb{P}_{\boldsymbol{u}}(X(n) = j) = x^n u_j$ for all $j \in S,\ n \geq 0$.*

Proof. To prove (i) $\Rightarrow$ (iii), let $\boldsymbol{u}$ be a quasi-stationary distribution, so that $\mathbb{P}_{\boldsymbol{u}}(X(0) = j) = u_j$ and $\mathbb{P}_{\boldsymbol{u}}(X(n) = j) = \mathbb{P}_{\boldsymbol{u}}(T > n)u_j$ for all $n \geq 1$. Then,

$$\sum_{i\in S} u_i P_{ij} = \mathbb{P}_{\boldsymbol{u}}(X(1) = j) = \mathbb{P}_{\boldsymbol{u}}(T > 1)u_j = (1 - \kappa_{\boldsymbol{u}})u_j,$$

while, by definition of a quasi-stationary distribution, $1 - \kappa_{\boldsymbol{u}} = \mathbb{P}_{\boldsymbol{u}}(T > 1) > 0$. This establishes (iii).

The equivalence of (2.4) and (2.5) implies (ii) $\Leftrightarrow$ (iv) and (iii) $\Leftrightarrow$ (v). Moreover, a simple substitution shows (iv) $\Rightarrow$ (i).

Finally, we will show (iii) $\Rightarrow$ (ii). So let $x > 0$ and assume that $\boldsymbol{u}$ represents an x-invariant distribution. Recalling that P is in lower block-triangular form we decompose the vector $\boldsymbol{u} = (\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_L)$ accordingly, and note that

$$\boldsymbol{u}_L P_L = x\boldsymbol{u}_L.$$

If $\boldsymbol{u}_L \neq \mathbf{0}$ (the row vector of zeros) then S_L must be self-communicating and so, by the *Perron – Frobenius theorem* (see, for example, [10, p. 673]) applied to the matrix P_L, we must have $x = \rho_L$. On the other hand, if $\boldsymbol{u}_L = \mathbf{0}$ we must have

$$\boldsymbol{u}_{L-1} P_{L-1} = x\boldsymbol{u}_{L-1},$$

and we can repeat the argument. Thus proceeding we conclude that there must be a $k \in \{1, 2, \ldots, L\}$ such that $x = \rho_k$. This establishes (ii) and completes the proof of the theorem. □

Using (2.7) it follows in particular that if $\boldsymbol{u} \equiv (u_i, i \in S)$ is a quasi-stationary distribution, so that the distribution of $X(n)$ conditional on survival up to time n is constant over n, then

$$\mathbb{P}_{\boldsymbol{u}}(T > n) = \sum_{j \in S} \mathbb{P}_{\boldsymbol{u}}(X_n = j) = (1 - \kappa_{\boldsymbol{u}})^n, \quad n \geq 0.$$

So in this case the distribution of the residual survival time conditional on survival up to time n is also constant over n, and given by

$$\mathbb{P}_{\boldsymbol{u}}(T > n + m \mid T > n) = (1 - \kappa_{\boldsymbol{u}})^m, \quad m \geq 0. \tag{2.9}$$

In what follows we are interested in the *limiting distribution as* $n \to \infty$ of $X(n)$ conditional on survival up to time n, that is,

$$\lim_{n \to \infty} \mathbb{P}_{\boldsymbol{w}}(X(n) = j \mid T > n), \quad j \in S, \tag{2.10}$$

and in the *limiting distribution as* $n \to \infty$ of the residual survival time conditional on survival up to time n, that is,

$$\lim_{n \to \infty} \mathbb{P}_{\boldsymbol{w}}(T > n + m \mid T > n), \quad m \geq 0, \tag{2.11}$$

for any initial distribution $\boldsymbol{w} \equiv (w_i, i \in S)$ over S, provided these limiting conditional distributions exist. Theorem 2.1 plays a key role in the analysis.

3. Irreducible state space

To set the stage we will first assume that S is irreducible, that is $L = 1$, and so $S_1 = S$ and $P_1 = P \neq O$. As noted in the previous section the Perron–Frobenius eigenvalue ρ of P satisfies $0 < \rho < 1$. We can (and will) choose the associated left and right eigenvectors $\boldsymbol{u} = (u_i, i \in S)$ and $\boldsymbol{v} = (v_i, i \in S)$ strictly positive componentwise (see, for example, [10, p. 673]). It will also be convenient to normalize $\boldsymbol{u}$ and $\boldsymbol{v}$ such that

$$\sum_{i \in S} u_i = 1 \quad \text{and} \quad \sum_{i \in S} u_i v_i = 1. \tag{3.1}$$

Assuming S to be aperiodic the transition probabilities $P_{ij}(n)$ then satisfy

$$\lim_{n \to \infty} \rho^{-n} P_{ij}(n) = v_i u_j > 0, \quad i, j \in S \tag{3.2}$$

(see, for example, [6], or, for increasingly more general results, [9] and [8]).

Since $\boldsymbol{u}$ represents a ρ-invariant probability distribution for P we have, by Theorem 2.1,

$$\mathbb{P}_{\boldsymbol{u}}(X(n) = j) = \rho^n u_j, \quad j \in S, \quad n \geq 0. \tag{3.3}$$

Considering that

$$\mathbb{P}_{\boldsymbol{u}}(T > n) = \mathbb{P}_{\boldsymbol{u}}(X(n) \in S) = \rho^n,$$

it follows that for all $n \geq 0$

$$\mathbb{P}_{\boldsymbol{u}}(T > n + m \mid T > n) = \rho^m, \quad m \geq 0. \tag{3.4}$$

Moreover, also by Theorem 2.1, $\boldsymbol{u}$ is a quasi-stationary distribution of $\mathcal{X}$, so that (2.8) holds true for all n. Assuming S to be aperiodic, Darroch and Seneta [6] have shown that similar results hold true in the limit as $n \to \infty$ when the initial distribution differs from $\boldsymbol{u}$. Namely, for any initial distribution $\boldsymbol{w}$ one has

$$\lim_{n \to \infty} \mathbb{P}_{\boldsymbol{w}}(X(n) = j \mid T > n) = u_j, \quad j \in S, \tag{3.5}$$

and

$$\lim_{n \to \infty} \mathbb{P}_{\boldsymbol{w}}(T > n + m \mid T > n) = \rho^m, \quad m \geq 0. \tag{3.6}$$

So when S is aperiodic and all states in S communicate the limits (2.10) and (2.11) are determined by the Perron–Frobenius eigenvalue of P and the corresponding left eigenvector.

These results can be generalized to a setting in which S may consist of more than one class, as we will show in the next two sections.

4. General case: quasi-stationary distributions

We return to the setting of Section 2, so we will assume that S consists of the classes $S_1, S_2, \ldots, S_L$, and allow $L \geq 1$. In view of Theorem 2.1 we can identify all quasi-stationary distributions by identifying all x-invariant distributions for P such that $x = \rho_k$ for some k. That is, we must identify all nonnegative, nonzero left ρ_k-eigenvectors of P. Fortunately, the problem of identifying all nonnegative eigenvectors of a nonnegative matrix has been resolved completely in the literature. We summarize the results in Theorem 4.1 below, where a class S_k is called a *maximal class* if $\rho_j < \rho_k$ for all $j \neq k$ such that $S_j \prec S_k$ (which is vacuously true if no other class is accessible from S_k).

Theorem 4.1.

(i) *There exists a nonnegative left x-eigenvector of P if and only if there exists a maximal class S_k such that $x = \rho_k$.*

(ii) *If S_k is a maximal class, then there is a (up to scalar multiples) unique left ρ_k-eigenvector $\boldsymbol{u} \equiv (u_i, i \in S)$ of P such that $u_i > 0$ if i is accessible from S_k and $u_i = 0$ otherwise.*

(iii) *If $\boldsymbol{u} \equiv (u_i, i \in S)$ is a nonnegative left x-eigenvector of P, then $\boldsymbol{u}$ is a linear combination with nonnegative coefficients of the eigenvectors defined in* (ii) *corresponding to the maximal classes S_k with $x = \rho_k$.*

The above theorem combines Schneider [12, Theorems 3.1 and 3.7], which are based on earlier results of Schneider [11], Carlson [3] and Victory [15]. A mild generalization of these results is presented by Lindqvist [8, Theorem 6.1]. Recalling (2.7) and translating Theorem 4.1 in terms of quasi-stationary distributions with the help of Theorem 2.1, we obtain the following

Theorem 4.2.

(i) *There exists a quasi-stationary distribution $\boldsymbol{u} \equiv (u_i, i \in S)$ for $\mathcal{X}$ with killing probability $\kappa_{\boldsymbol{u}} = \kappa$ if and only if there exists a maximal class S_k such that $\rho_k = 1 - \kappa$.*

(ii) *If S_k is a maximal class, then there is a unique quasi-stationary distribution $\boldsymbol{u} \equiv (u_i, i \in S)$ for $\mathcal{X}$ with killing probability $\kappa_{\boldsymbol{u}} = 1 - \rho_k$ and such that $u_i > 0$ if and only if i is accessible from S_k.*

(iii) *If $\boldsymbol{u} \equiv (u_i, i \in S)$ is a quasi-stationary distribution for $\mathcal{X}$ with killing probability $\kappa_{\boldsymbol{u}} = \kappa$, then $\boldsymbol{u}$ is a linear combination with nonnegative coefficients of the quasi-stationary distributions defined in (ii) corresponding to the maximal classes S_k with $\rho_k = 1 - \kappa$.*

Evidently, there exists at least one quasi-stationary distribution since S_1 is a maximal class. Moreover, there is precisely one quasi-stationary distribution if and only if $S_1 \prec S_k$ for all k (which is vacuously true if $L = 1$) and $\rho_1 = \rho$. However, motivated by our interest in limiting conditional distributions, we shall be concerned in what follows with quasi-stationary distribution of a particular type rather than quasi-stationary distributions in general. We must introduce some further notation and terminology first.

We let $I(\rho) := \{k : \rho_k = \rho\}$, so that $\mathrm{card}(I(\rho))$ is the algebraic multiplicity of the Perron−Frobenius eigenvalue ρ, and define

$$a(\rho) := \min I(\rho). \tag{4.1}$$

Class S_k will be called a *ρ-maximal class* if S_k is a maximal class and $\rho_k = \rho$. (In the terminology of [8] the index k is *ρ-final*.) The number of ρ-maximal classes will be denoted by $m(\rho)$. Clearly, $m(\rho) \geq 1$ since $S_{a(\rho)}$ is a ρ-maximal class. By *accessibility* of a class S_k *from a distribution* $\boldsymbol{u} \equiv (u_i, i \in S)$ we mean that there is a state i such that $u_i > 0$ and S_k is accessible from i.

The next lemma provides the main ingredient for the proof of Theorem 4.3, which establishes under which circumstances there is precisely one quasi-stationary distribution from which $S_{a(\rho)}$ is accessible.

Lemma 4.1. *If $\boldsymbol{u} \equiv (u_i, i \in S)$ is a quasi-stationary distribution from which class $S_{a(\rho)}$ is accessible, then $\kappa_{\boldsymbol{u}} = 1 - \rho$.*

Proof. If the initial distribution $\boldsymbol{u}$ is a quasi-stationary distribution, then, by (2.7) and Theorem 2.1,

$$\mathbb{P}_{\boldsymbol{u}}(X(n) = j) = (1 - \kappa_{\boldsymbol{u}})^n u_j, \quad j \in S, \quad n \geq 0.$$

It follows that $u_j > 0$ for all states j that are accessible from $\boldsymbol{u}$. So, if $S_{a(\rho)}$ is accessible from $\boldsymbol{u}$, we have $u_j > 0$ for all $j \in S_{a(\rho)}$. Since

$$\mathbb{P}_{\boldsymbol{u}}(X(n) = j) \geq u_j P_{jj}(n),$$

it follows that

$$P_{jj}(n) \leq (1 - \kappa_{\boldsymbol{u}})^n, \quad j \in S_{a(\rho)}, \quad n \geq 0.$$

Assuming $S_{a(\rho)}$ to be aperiodic and in view of (3.2) applied to the process $\mathcal{X}$ restricted to $S_{a(\rho)}$, we therefore have $1 - \kappa_{\boldsymbol{u}} \geq \rho = \max_k \rho_k$. But since, by Theorem 2.1 again, $1 - \kappa_{\boldsymbol{u}} = \rho_k$ for some k, we must have $\kappa_{\boldsymbol{u}} = 1 - \rho$. Since $\rho(P^d) = (\rho(P))^d$, the same conclusion prevails if $S_{a(\rho)}$ has period $d > 1$, in view of (3.2) applied to the (aperiodic) process $\mathcal{X}^d \equiv \{X(dn),\ n = 0, 1, \ldots\}$ restricted to $S_{a(\rho)}$. □

Combining this lemma with Theorem 4.2 (and Theorem 2.1) leads to the following key result.

Theorem 4.3. *The process $\mathcal{X}$ with transition probability matrix P has a unique quasi-stationary distribution $\boldsymbol{u} \equiv (u_i, i \in S)$ from which $S_{a(\rho)}$ is accessible if and only if $m(\rho) = 1$ (that is, $S_{a(\rho)}$ is the only ρ-maximal class) in which case $\boldsymbol{u}$ is the (unique) ρ-invariant distribution for P and satisfies $u_i > 0$ if and only if i is accessible from $S_{a(\rho)}$.*

5. General case: limiting conditional distributions

We now turn to the question of whether for a given initial distribution $\boldsymbol{w}$ over S the limits (2.10) exist and constitute a proper distribution $\boldsymbol{u}$, say, and whether $\boldsymbol{u}$ can be identified with a (perhaps unique) quasi-stationary distribution. Since

$$\mathbb{P}_{\boldsymbol{w}}(X(n) = j \mid T > n) = 0$$

for all n if j is not accessible from $\boldsymbol{w}$, it will be no restriction of generality to assume that every class S_k is accessible from $\boldsymbol{w}$.

In order to study the asymptotic behaviour of $\mathbb{P}_{\boldsymbol{w}}(X(n) = j \mid T > n)$ as $n \to \infty$, we need information about the asymptotic behaviour of the individual transition probabilities $P_{ij}(n)$, which is given by Mandl [9] under the assumption that each class S_k is aperiodic (a summary is given in [6]), and by Lindqvist [8] in a general setting. Since the situation is complicated considerably by allowing each class S_k to have a period $d_k \geq 1$, we will make the simplifying assumption that class $S_{a(\rho)}$ (but not necessarily any other class) is aperiodic. Lindqvist's results enable us to present the generalization announced at the end of Section 3 (and already alluded to by Lindqvist [8, p. 586–587]) in the next theorem.

Theorem 5.1. *If $d_{a(\rho)} = m(\rho) = 1$ (that is, $S_{a(\rho)}$ is aperiodic and the only ρ-maximal class) and the initial distribution $\boldsymbol{w}$ is such that each class is accessible, then the limits* (2.10) *and* (2.11) *exist and are given by* (3.5) *and* (3.6), *respectively, where $\boldsymbol{u}$ is the unique quasi-stationary distribution of $\mathcal{X}$ from which $S_{a(\rho)}$ is accessible.*

Proof. Let τ denote the maximal number of classes S_k with $\rho_k = \rho$ that can be traversed in a path from one state in S to another, and let H denote the set of pairs of states (i, j) such that there exists a path from i to j traversing τ such classes. Observe that $S_{a(\rho)}$ must be accessible from i and j must be accessible from $S_{a(\rho)}$ for any pair $(i, j) \in H$, since $S_{a(\rho)}$ is the only ρ-maximal class. It now follows from [8, Theorems 5.4 and 5.8] (see also the proof of the latter theorem) that

$$\lim_{n\to\infty} \left(\frac{\rho}{n}\right)^{\tau-1} \rho^{-n} P_{ij}(n) = v_i u_j, \tag{5.1}$$

where $v_i > 0$ if $(i, j) \in H$ for some j and $v_i = 0$ otherwise, and $\boldsymbol{u} \equiv (u_i, i \in S)$ is the ρ-invariant distribution for P satisfying $u_i > 0$ if and only if i is accessible from $S_{a(\rho)}$. By Theorem 4.3 the latter distribution is in fact the unique ρ-invariant distribution for P, and also the unique quasi-stationary distribution of $\mathcal{X}$ from which $S_{a(\rho)}$ is accessible. As a consequence of (5.1) we have

$$\lim_{n\to\infty}\left(\frac{\rho}{n}\right)^{\tau-1}\rho^{-n}\sum_{j\in S}P_{ij}(n)=v_i,$$

which implies that

$$\lim_{n\to\infty}\left(\frac{\rho}{n}\right)^{\tau-1}\rho^{-n}\sum_{i\in S}w_i\sum_{j\in S}P_{ij}(n)=\sum_{i\in S}w_iv_i.$$

Since there must be a class S_k, say, such that $v_i>0$ for all states $i\in S_\ell\succ S_k$, while each class, in particular S_k, is supposed to be accessible from $\boldsymbol{w}$, we must have $w_iv_i>0$ for at least one $i\in S$. Hence, for all $j\in S$,

$$\lim_{n\to\infty}\mathbb{P}_{\boldsymbol{w}}(X(n)=j\mid T>n)=\lim_{n\to\infty}\frac{\sum_{i\in S}w_iP_{ij}(n)}{\sum_{i\in S}w_i\sum_{j\in S}P_{ij}(n)}=u_j,$$

and, for any $m\geq 0$,

$$\lim_{n\to\infty}\mathbb{P}_{\boldsymbol{w}}(T>n+m\mid T>n)=\lim_{n\to\infty}\frac{\sum_{i\in S}w_i\sum_{j\in S}P_{ij}(n+m)}{\sum_{i\in S}w_i\sum_{j\in S}P_{ij}(n)}=\rho^m,$$

as required. □

Under the conditions of this theorem the class $S_{a(\rho)}$ appears to be a "bottleneck" class in the sense that the limiting conditional distribution is supported by this class and those accessible from it, but not by any other class. This phenomenon is exemplified by the models discussed in the next section.

6. Examples

6.1. A model for two competing species on a habitat patch

Consider two species A and B that affect one another's ability to survive on a habitat patch. Let $X_A(n)$ and $X_B(n)$ denote the number of individuals of species A and B, respectively, at time n, and assume that

$$\mathcal{X}\equiv\{(X_A(n),X_B(n)),\ n=0,1,\ldots\}$$

is a Markov chain on a *finite* state space $S\cup\{(0,0)\}$, and with a matrix P of one-step transition probabilities of the (sub)Markov chain on S. We let S_{AB}, S_A and S_B denote the subsets of S consisting of states that correspond to the presence of individuals of both species, just species A, and just species B, respectively. Assuming irreducibility of these subsets, and excluding the possibility of immigration, $(0,0)$ is an absorbing state and S comprises the three classes S_{AB}, S_A, and S_B. We have $S_{AB} \succ S_A$ and $S_{AB} \succ S_B$, and will suppose for convenience that each class is aperiodic.

We denote the Perron−Frobenius eigenvalues of the submatrices of P corresponding to S_{AB} S_A, S_B, by ρ_{AB}, ρ_A, and ρ_B, respectively. Then there are essentially four cases to consider:

(1) $\rho_{AB} > \rho_A > \rho_B$ (same as $\rho_{AB} > \rho_B > \rho_A$),

(2) $\rho_{AB} > \rho_A = \rho_B$,

(3) ρ_A is largest or $\rho_A = \rho_{AB} > \rho_B$ (same as ρ_B is largest or $\rho_B = \rho_{AB} > \rho_A$),

(4) $\rho_A = \rho_B \geq \rho_{AB}$.

Theorem 4.3 tells us that one can associate a quasi-stationary distribution with each maximal class (let us call such a quasi-stationary distribution *basic*), and that linear combinations of basic quasi-stationary distributions yield quasi-stationary distributions again if the corresponding ρ's are equal. So in the current setting we have three basic quasi-stationary distributions in the cases (1) and (2), and two basic quasi-stationary distributions otherwise. The basic quasi-stationary distributions are the only quasi-stationary distributions in the cases (1) and (3), but there are infinitely many quasi-stationary distributions in the cases (2) and (4).

Assuming that the initial state is in S_{AB} and that there is only one ρ-maximal class, Theorem 5.1 tells us that there is a unique limiting conditional distribution, which equals the basic quasi-stationary distribution associated with the ρ-maximal class. This covers the cases (1), (2), and (3) of the present example. In case (4) there are two ρ-maximal classes, namely S_A and S_B, so then the conditions of Theorem 5.1 are not satisfied. Of course there exists a unique limiting conditional distribution, but it will be a linear combination of the two basic quasi-stationary distributions associated with S_A and S_B, with weights depending on the initial distribution.

In summary, given that the patch has been inhabited for a very long time the question which of the species have survived can be answered in terms of the eigenvalues ρ_{AB}, ρ_A, and ρ_B. To obtain the precise limiting conditional distribution the model must be specified in detail.

6.2. A model for the course of a progressive disease

Consider the setting of patients suffering from a progressive disease. If patients live, they can only remain in the same state or move to a higher-risk state but not to a lower-risk state. Suppose that there are s "alive states" $1, 2, \ldots, s$, listed in decreasing order of risk, and a "dead state" 0. Assume that patients are assessed periodically, and that a patient in state i (> 1) will, on the next assessment, have moved to state $i-1$ with probability q_i, have died with probability κ_i, or have remained at i with probability

$$r_i := 1 - q_i - \kappa_i.$$

The probability of death from state 1 is κ_1 and $r_1 := 1 - \kappa_1$ is the probability of remaining in state 1. Thus patients are assumed to be assessed sufficiently often that it is not possible for them to have skipped an alive state between assessments. We also assume that $0 \leq \kappa_i < 1$ and $0 < q_i < 1$ for all i. Such a model has been used by Chan et al. [4] in a study of patients with congestive heart failure. The process is thus a pure-death chain with killing on $\{1, 2, \ldots, s\}$, with death probability q_i and killing probability κ_i in state i, and coffin state 0.

It has been observed in practice (see for example [4]) that patients surviving for a long time tend to dwell in the various states with constant probabilities. This phenomenon is reflected by our model since the results of the previous sections imply that a limiting conditional distribution exists. Indeed, our model fits in the setting of Section 2 with a matrix

$$P = \begin{pmatrix} r_1 & 0 & 0 & \cdots & 0 & 0 \\ q_2 & r_2 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & r_{s-1} & 0 \\ 0 & 0 & 0 & \cdots & q_s & r_s \end{pmatrix} \tag{6.1}$$

of one-step transition probabilities. The classes now consist of single states, so, maintaining the notation of the previous sections, we have $S_k = \{k\}$ and $S_1 \prec S_2 \prec \cdots \prec S_s$. Assuming $r := \max r_i > 0$, we also find that $\rho_k = r_k$ for all k, and hence

$$\rho = \max \rho_k = r > 0.$$

As in Section 4 we let $a(\rho) = \min\{k : \rho_k = \rho\}$. Evidently, $S_{a(\rho)}$ is the only ρ-maximal class, so we can apply the Theorems 4.3 and 5.1 and readily obtain the following result.

Proposition 6.1. *Let the initial distribution $\boldsymbol{w}$ be supported by at least one state $i \geq a(\rho)$. Then*

$$\lim_{n\to\infty} \mathbb{P}_{\boldsymbol{w}}(T > n + m \mid T > n) = r^m, \quad m \geq 0, \tag{6.2}$$

and

$$\lim_{n\to\infty} \mathbb{P}_{\boldsymbol{w}}(X(n) = j \mid T > n) = u_j, \quad j \in S, \tag{6.3}$$

where $\boldsymbol{u} = (u_1, u_2, \ldots, u_s)$ *is the (unique) quasi-stationary distribution from which state* $a(\rho)$ *is accessible, and given by*

$$u_j = \begin{cases} u_1 \displaystyle\prod_{i=1}^{j-1} \frac{r - r_i}{q_{i+1}}, & j \le a(\rho), \\ 0, & j > a(\rho), \end{cases} \tag{6.4}$$

where u_1 *is chosen so that* $\sum_{j=1}^{a(\rho)} u_j = 1$, *and an empty product should be interpreted as being equal to* 1.

Acknowledgement

The work of Phil Pollett is supported by the Australian Research Council Centre of Excellence for Mathematics and Statistics of Complex Systems.

References

[1] O.O. Aalen and H.K. Gjessing (2001) Understanding the shape of the hazard rate: a process point of view. *Statist. Sci.* **16**, 1–22.

[2] O.O. Aalen and H.K. Gjessing (2003) A look behind survival data: underlying processes and quasi-stationarity. In: *Mathematical and Statistical Methods in Reliability*, B.H. Lindqvist, K.A. Doksum (eds.), World Scientific Publishing, Singapore, 221–234.

[3] D. Carlson (1963) A note on M-matrix equations. *J. Soc. Indust. Appl. Math.* **11**, 1027–1033.

[4] D.C. Chan, P.A. Heidenreich, M.C. Weinstein and G. Fonarow (2008) Heart failure disease management programs: A cost-effectiveness analysis. *Amer. Heart J.* **155**, 332–338.

[5] C.D.H. Cooper (1973) On the maximum eigenvalue of a reducible non-negative real matrix. *Math. Z.* **131**, 213–217.

[6] J.N. Darroch and E. Seneta (1965) On quasi-stationary distributions in absorbing discrete-time finite Markov chains. *J. Appl. Probab.* **2**, 88–100.

[7] E.A. van Doorn and P.K. Pollett (2008) Survival in a quasi-death process. *Linear Algebra Appl.* **429**, 776–791.

[8] B.H. Lindqvist (1989) Asymptotic properties of powers of nonnegative matrices, with applications. *Linear Algebra Appl.* **114/115**, 555–588.

[9] P. Mandl (1959) On the asymptotic behaviour of probabilities within groups of states of a homogeneous Markov chain. *Časopis Pěst. Mat.* **84**, 140–149.

[10] C.D. Meyer (2001) *Matrix Analysis and Applied Linear Algebra.* SIAM, Philadelphia. Updates on `http://www.matrixanalysis.com`.

[11] H. Schneider (1956) The elementary divisors associated with 0 of a singular M-matrix. *Proc. Edinburgh Math. Soc.* **10**, 108–122.

[12] H. Schneider (1986) The influence of the marked reduced graph of a nonnegative matrix on the Jordan form and on related properties: a survey. *Linear Algebra Appl.* **84**, 161–189.

[13] E. Seneta (1981) *Non-negative Matrices and Markov Chains.* Springer, New York.

[14] D. Steinsaltz and S.N. Evans (2004) Markov mortality models: implications of quasistationarity and varying initial distributions. *Theoret. Population Biol.* **65**, 319–337.

[15] H.D. Victory (1985) On nonnegative solutions to matrix equations. *SIAM J. Algebraic Discrete Methods* **6**, 406–412.