Quasi-stationary distributions

Erik A. van Doorn[a] and Philip K. Pollett[b]

[a] Department of Applied Mathematics, University of Twente
P.O. Box 217, 7500 AE Enschede, The Netherlands
E-mail: e.a.vandoorn@utwente.nl

[b] Department of Mathematics, The University of Queensland
Qld 4072, Australia
E-mail: pkp@maths.uq.edu.au

June 17, 2011

Abstract. This paper contains a survey of results related to quasi-stationary distributions, which arise in the setting of stochastic dynamical systems that eventually evanesce, and which may be useful in describing the long-term behaviour of such systems before evanescence. We are concerned mainly with continuous-time Markov chains over a finite or countably infinite state space, since these processes most often arise in applications, but will make reference to results for other processes where appropriate. Next to giving an historical account of the subject, we review the most important results on the existence and identification of quasi-stationary distributions for general Markov chains, and give special attention to birth-death processes and related models. Results on the question of whether a quasi-stationary distribution, given its existence, is indeed a good descriptor of the long-term behaviour of a system before evanescence, are reviewed as well. The paper is concluded with a summary of recent developments in numerical and approximation methods.

Keywords: Applied probability, Markov processes

Memorandum 1945 (June 2011). ISSN 1874-4850. Available from: http://www.math.utwente.nl/publications
Department of Applied Mathematics, University of Twente, Enschede, The Netherlands

1 Introduction

Many biological systems are certain to "die out" eventually, yet appear to be stationary over any reasonable time scale. This phenomenon, termed *quasi stationarity*, is illustrated in Figure 1. Here, a model for the number $X(t)$ of occupied habitat patches in an n-patch metapopulation (population network) is simulated. For the parameter values given, the expected time to total extinction starting with one patch occupied is 4.7287×10^7 years, yet over the period of 300 years simulated, the number of patches occupied stabilises near 16. The notion of a *quasi-stationary distribution* has proved to be a useful tool in modelling this kind of behaviour. Figure 2 shows the same simulation with the quasi-stationary distribution superimposed. Notice that while the limiting distribution necessarily assigns all its probability mass to the extinction state 0, the quasi-stationary distribution assigns mass to states in a way that mirrors the quasi stationarity observed.

To make this notion more precise, think of an observer who at some time t is aware of the occupancy of some patches, yet cannot tell exactly which patches are occupied. What is the chance of there being precisely i patches occupied? If we were equipped with the full set of state probabilities $p_i(t) = \Pr(X(t) = i)$, $i \in \{0, 1, \ldots, n\}$, we would evaluate the conditional probability $u_i(t) = \Pr(X(t) = i | X(t) \neq 0) = p_i(t)/(1-p_0(t))$, for i in the set $S = \{1, \ldots, n\}$ of transient states. Then, in view of the behaviour observed in our simulation, it would be natural for us to seek a distribution $\boldsymbol{u} = (u_i,\, i \in S)$ over S such that if $u_i(s) = u_i$ for a particular $s > 0$, then $u_i(t) = u_i$ for all $t > s$. Such a distribution is called a *stationary conditional distribution* or *quasi-stationary distribution*. Our key message is that $\boldsymbol{u}$ can usually be determined from the transition rates of the process and that $\boldsymbol{u}$ might then also be a *limiting conditional distribution* in that $u_i(t) \to u_i$ as $t \to \infty$, and thus be of use in modelling the long-term behaviour of the process. When the set S of transient states is finite, classical matrix theory is enough to establish the existence of a quasi-stationary distribution, which is unique when S is irreducible and admits limiting conditional interpretation that is independent of initial conditions. When S is infinite the question even of the existence and then uniqueness of a quasi-stationary distribution is both subtle and interesting, and not yet fully resolved.

We shall be concerned here with continuous-time Markov chains over a finite or countably-infinite state space, since these processes most often arise in applications, but we will make reference to results for other processes where appropriate. We will review theoretical results on quasi-stationary distributions and limiting conditional distributions in Sections 3 and 4, giving special attention to birth-death processes and related models in Section 5. Recent developments in numerical and approximation methods are summarised in Section 6. We start with some historical background in Section 2. Our review is by no means exhaustive. For additional references on quasi-stationary distributions and related work, we refer the reader to the annotated bibliography maintained here: `http://www.maths.uq.edu.au/~pkp/papers/qsds.html`

2 Modelling quasi stationarity

Yaglom [151] was the first to identify a limiting conditional distribution, establishing the existence of such for the subcritical Bienaymé-Galton-Watson branching process (a result refined later by Heathcote et al. [69]). However, this process does not exhibit quasi stationary behaviour of the kind depicted in Figure 1; rather, it reaches the extinction state quickly, and Yaglom's limit is a mathematical manifestation of the process being "forced" to stay positive. The idea of using a quasi-stationary distribution to account for apparent stationarity in evanescent stochastic processes was due to Bartlett [15, Page 24]:

> "It still may happen that the time to extinction is so long that it is still of more relevance to consider the effectively ultimate distribution (called a 'quasi-stationary' distribution) of [the process] N."

He gave details (in the context of an absorbing birth-death process) of one approach to modelling quasi stationarity whereby the process is imagined to be returned to state 1 (corresponding to one individual) at some small rate ϵ at the moment of extinction and the stationary distribution π^{ϵ} (if it exists)

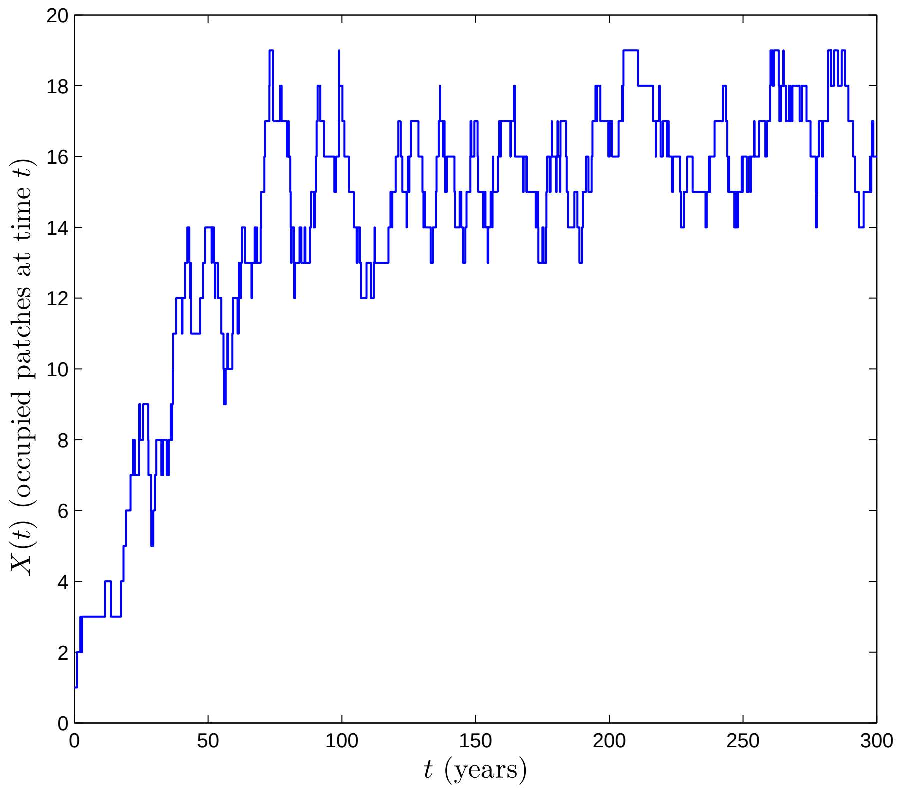

Fig 1. Simulation of a 20-patch metapopulation model with colonization rate $c = 0.1625$ and local extinction rate $e = 0.0325$, starting with one patch occupied.

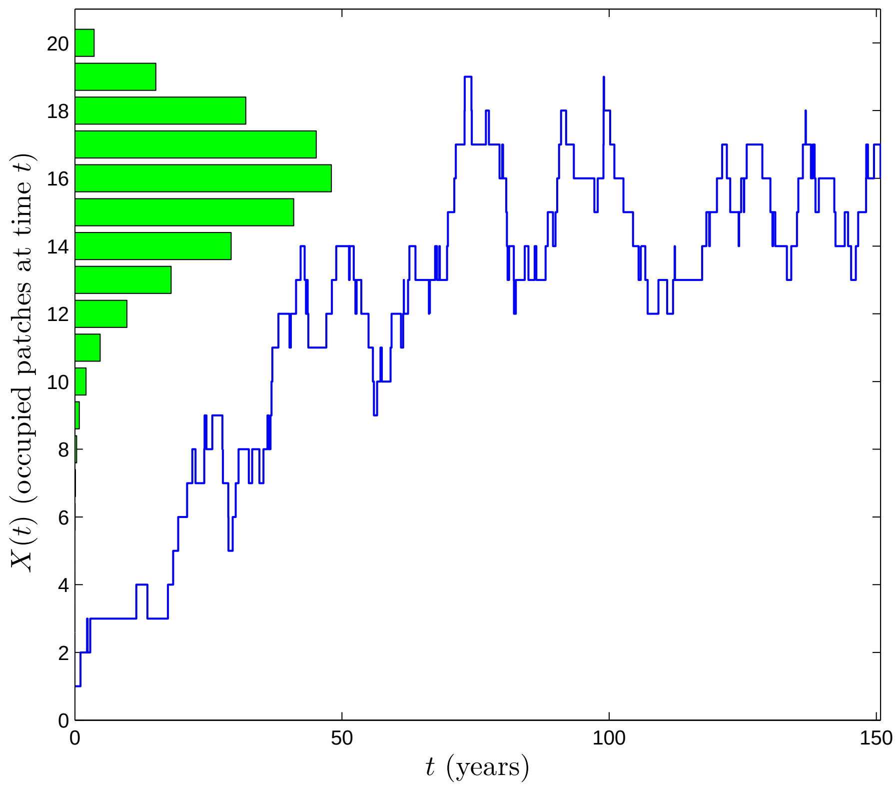

Fig 2. The same simulation as in Figure 1 with the quasi-stationary distribution superimposed.

of the resulting "returned process" is used to model the long-term behaviour of the original process. Bartlett then argued that, under natural conditions, the dependence of π^{ϵ} on ϵ would be weak. Ewens [54, 55] exploited the idea for certain diffusion models that arise in population genetics, as well as their discrete-state analogues. Ewens returned his processes to an arbitrary state, and he coined the term *pseudo-transient* for the stationary distribution of the returned process. More generally, one might consider returning the process to the set S of transient states according to a probability measure $\boldsymbol{m}$ over S, and then evaluating the stationary distribution $\pi^{\boldsymbol{m}}$ of the returned process. This was fleshed out by Darroch and Seneta [39, Section 2] in the context of discrete-time Markov chains, but they raised an objection that $\pi^{\boldsymbol{m}}$ depends on $\boldsymbol{m}$ "to such an extent that it can be made into almost any distribution" over S by a suitable choice of $\boldsymbol{m}$. They described several other possibilities for a "quasi-stationary" distribution, many that (importantly) *do not* depend on the initial distribution, the most natural being those described in Section 1: the *stationary conditional* distribution (now usually termed *quasi-stationary distribution*) and the *limiting conditional distribution* (studied earlier by Mandl [100, 101]). These notions gained prominence in the literature and, following treatments for finite-state Markov chains by Darroch and Seneta [39, 40], there were significant early developments by Seneta and Vere-Jones [136, 147, 148] for countable-state chains that built on important work by Vere-Jones [146] on x-invariant measures and geometric ergodicity (but see also Albert [3] and Kingman [86]).

The notions of quasi-stationary distribution and pseudo-transient distribution can been reconciled, at least in our present context of countable-state Markov chains. Under mild assumptions (to be detailed later), the quasi-stationary distribution $\boldsymbol{u}$ is unique and is a *fixed point* of the map $\boldsymbol{m} \mapsto \pi^{\boldsymbol{m}}$ (called the "return map"), that is, $\boldsymbol{u}$ satisfies $\boldsymbol{u} = \pi^{\boldsymbol{u}}$ uniquely. Under these assumptions also, the return map is contractive, and so iteration leads us to $\boldsymbol{u}$ (see, for example, Ferrari et al. [58]). Furthermore, $\pi^{\boldsymbol{m}}$ is expected to be "close" to $\boldsymbol{u}$ for a range of return distributions $\boldsymbol{m}$, a statement that is made precise in Barbour and Pollett [13], thus assuaging to some extent the concerns of Darroch and Seneta. This has practical implications, for $\pi^{\boldsymbol{m}}$, interpreted as a "ratio of means" (its j-th component is the ratio of the expected time spent in state j and the expected time to extinction) [39, Section 3], can often be evaluated explicitly (see, for example, Artalejo and Lopez-Herrero [8]). Furthermore, since $\pi^{\boldsymbol{m}}$ is a *stationary distribution*, there is a range of efficient computational methods and accurate approximation (truncation) methods that can be exploited.

Bartlett [15, Page 25] mentioned a further, final, approach to modelling quasi stationarity whereby an approximating distribution of the state variable is identified, there (and typically) a Gaussian distribution, the quality of this approximation improving under some limiting regime (typically the size of the system increasing). The idea of using a Gaussian approximation for a Markovian state variable dates back at least to Kac [71, Page 384]. It was made concrete by Van Kampen [72] and, since then, "Van Kampen's method" has become ubiquitous in biological and physical sciences literature. It was given a rigorous treatment as a diffusion approximation by Kurtz [90, 91] and Barbour [10, 11] (see also McNeil and Schach [102]) for the class of *density-dependent* Markovian models, the connection with quasi stationarity crystallized by Barbour [12]. Within this rigorous framework, one can not only identify the most appropriate approximating model, but delineate limiting conditions under which the approximation is faithful.

The nineteen sixties and seventies saw further developments in the theory of quasi-stationary distributions for countable-state Markov chains (for example Flaspohler [59], Tweedie [143]), as well as generalizations to semi-Markov processes (Arjas et al. [4], Cheong [28, 29], Flaspohler and Holmes [60], Nummelin [110]) and Markov chains on a general state space (Tweedie [144]), and detailed results for generic models, for example, birth-death processes (Cavender [23], Good [66], Kesten [78]), random walks (Daley [36], Pakes [115], Seneta [131]), queueing systems (Kyprianou [92, 93, 94]) and branching processes (Buiculescu [20], Evans [52], Green [64], Seneta and Vere-Jones [137]). Many of these early developments were influenced by ratio limit theory, which itself enjoyed significant growth during this period (see, for example, Foguel [61], Kingman and Orey [87], Orey [113], Papangelou [118], Port [127], Pruitt [128]).

Our review is concerned with the most recent theoretical developments, within the context of continuous-time Markov chains and related generic models. For diffusions and other continuous-state processes, a good starting point is Steinsaltz and Evans [140] (but see also Cattiaux et al. [22] and Pinsky [119]) and for branching processes there is an excellent recent review by Lambert [95, Section 3]. Whilst many issues remain unresolved, the theory has reached maturity, and the use of quasi-stationary distributions is now widespread, encompassing varied and contrasting areas of application, including cellular automata (Atman and Dickman [9]), complex systems (Collet et al. [34]), ecology (Day and Possingham [41], Gosselin [63], Gyllenberg and Sylvestrov [68], Kukhtin et al. [89], Pollett [122]) epidemics (Nåsell [106, 107, 108], Artalejo et al. [6, 7]), immunology (Stirk et al. [141]), medical decision making (Chan et al. [24]), physical chemistry (Dambrine and Moreau [37, 38], Oppenheim et al. [112], Pollett [121]), queues (Boucherie [17], Chen et al. [25], Kijima and Makimoto [84]), reliability (Kalpakam and Shahul-Hameed [73], Kalpakam [74], Li and Cao [98, 99]), survival analysis (Aalen and Gjessing [1, 2], Steinsaltz and Evans [139]) and telecommunications (Evans [53], Ziedins [152]).

A common feature of many physical systems is the concept of ageing: the individual, or the system as a whole, moves irreversibly through a series of states before reaching a stable regime. For example, individual patients suffering a progressive disease move from lower to higher risk states, while an ecosystem, consisting of species that affect one another's ability to survive, will shed some species before a state of coexistence is reached. This necessitates examining, and in some cases re-examining, the theory of quasi-stationarity within the context of a *reducible* state space.

3 Existence and identification

3.1 Introduction

This section contains an survey of results on the existence and identification of quasi-stationary and limiting conditional distributions. As announced we will focus on continuous-time Markov chains and consider finite and countably infinite state spaces separately in the Subsections 3.2 and 3.3, respectively. In the infinite setting we restrict ourselves to Markov chains for which the non-absorbing states constitute an irreducible class. We will briefly discuss quasi-stationarity for discrete-time Markov chains in Subsection 3.4. Some special cases of continuous-time Markov chains, to wit birth-death processes and birth-death processes with killing, are analysed in more detail in Section 5.

3.2 Finite Markov chains

Preliminaries

We start off by introducing some notation and terminology, and deriving some basic results. Consider a continuous-time Markov chain $\mathcal{X} := \{X(t),\, t \geq 0\}$ on a state space $\{0\} \cup S$ consisting of an absorbing state 0 and a *finite* set of *transient* states $S := \{1, 2, \ldots, n\}$. The generator of $\mathcal{X}$ then takes the form

$$\begin{pmatrix} 0 & \mathbf{0} \\ \boldsymbol{a}^T & Q \end{pmatrix}, \tag{1}$$

where $Q = (q_{ij})$ is the generator of the (sub)Markov chain on S and the vector $\boldsymbol{a} = (a_i,\, i \in S)$ of absorption (or *killing*) rates satisfies

$$\boldsymbol{a} = -\mathbf{1}Q^T \geq \mathbf{0},\ \boldsymbol{a} \neq \mathbf{0}. \tag{2}$$

Here $\mathbf{0}$ and $\mathbf{1}$ are row vectors of zeros and ones, respectively, superscript ^{T} denotes transposition, and $\geq$ indicates componentwise inequality. Since all states in S are transient, state 0 is accessible from any state in S. Hence, whichever the initial state, the process will eventually escape from S into the absorbing state 0 with probability one.

We write $\mathbb{P}_i(\cdot)$ for the probability measure of the process when the initial state is i, and $\mathbb{E}_i(\cdot)$ for the expectation with respect to this measure. For any vector $\boldsymbol{u} = (u_i,\, i \in S)$ representing a distribution over S we let $a_{\boldsymbol{u}} := \sum_{i\in S} u_i a_i$ and $\mathbb{P}_{\boldsymbol{u}}(\cdot) := \sum_{i\in S} u_i \mathbb{P}_i(\cdot)$. We also write $P_{ij}(\cdot) := \mathbb{P}_i(X(\cdot) = j)$, and recall that the matrix $P(t) = (P_{ij}(t),\ i, j \in S)$ satisfies

$$P(t) = e^{Qt} := \sum_{k=0}^{\infty} \frac{Q^k}{k!} t^k, \quad t \geq 0 \tag{3}$$

(see, for example, Kijima [83, Section 4.6]).

We allow S to be reducible, so we suppose that S consists of communicating classes $S_1, S_2, \ldots, S_L$, with $L \geq 1$, and let Q_k be the submatrix of Q corresponding to the states in S_k. We define a partial order on $\{S_1, S_2, \ldots, S_L\}$ by writing $S_i \prec S_j$ when S_i is *accessible* from S_j, that is, when there exists a sequence of states $k_0, k_1, \ldots, k_\ell$, such that $k_0 \in S_j$, $k_\ell \in S_i$, and $q_{k_m k_{m+1}} > 0$ for every m. We assume in what follows that the states are labelled such that Q is in lower block-triangular form, so that we must have

$$S_i \prec S_j \implies i \leq j. \tag{4}$$

Considering that the matrices Q_k reside on the diagonal of Q, it is easily seen that the set of eigenvalues of Q is precisely the union of the sets of eigenvalues of the individual Q_k's. It is well known (see, for example, Seneta [134, Theorem 2.6]) that the eigenvalue with maximal real part of Q_k, denoted by $-\alpha_k$, is unique, simple, and negative. Hence $\alpha := \min_k \alpha_k > 0$, and $-\alpha$ is the (possibly degenerate) eigenvalue of Q with maximal real part. The quantity α plays an crucial role in what follows and will be referred to as the *decay parameter* of $\mathcal{X}$.

A vector $\boldsymbol{u} = (u_i,\, i \in S)$ and, if appropriate, the probability distribution over S represented by $\boldsymbol{u}$, are called *x-invariant for Q* if

$$\sum_{i\in S} u_i q_{ij} = -x u_j, \quad j \in S, \tag{5}$$

that is, in the finite setting at hand, if $\boldsymbol{u}$ is a left eigenvector of Q corresponding to the eigenvalue $-x$. The vector (or distribution) $\boldsymbol{u}$ is called *x-invariant for P* if

$$\sum_{i\in S} u_i P_{ij}(t) = e^{-xt} u_j, \quad j \in S,\ t \geq 0. \tag{6}$$

We recall that $\boldsymbol{u}$ is a *quasi-stationary distribution* for $\mathcal{X}$ if the distribution of $X(t)$, conditional on non-absorption up to time t, is the same for all $t \geq 0$ when $\boldsymbol{u}$ is the initial distribution. That is, $\boldsymbol{u}$ is a quasi-stationary distribution if, for all $t \geq 0$,

$$\mathbb{P}_{\boldsymbol{u}}(X(t) = j \,|\, T > t) = u_j, \quad j \in S, \tag{7}$$

where $T := \sup\{t \geq 0 : X(t) \in S\}$ is the *absorption time* (or *survival time*) of $\mathcal{X}$, the random variable representing the time at which escape from S occurs. The notions of x-invariant distribution and quasi-stationary distribution are intimately related, as the next theorem shows. The theorem seems to be stated in its entirety only in a discrete-time setting (in [48]), so for completeness' sake we also furnish a proof.

Theorem 1 Let $\boldsymbol{u} = (u_i,\, i \in S)$ represent a proper probability distribution over $S = \cup_{k=1}^{L} S_k$. Then the statements

(a) $:\Longleftrightarrow$ $\boldsymbol{u}$ is a quasi-stationary distribution for $\mathcal{X}$,

(b) $:\Longleftrightarrow$ $\boldsymbol{u}$ is x-invariant for Q for some $x > 0$,

(c) $:\Longleftrightarrow$ $\boldsymbol{u}$ is x-invariant for P for some $x > 0$,

(d) $:\Longleftrightarrow$ $\boldsymbol{u}$ is α_k-invariant for Q for some $k \in \{1, 2, \ldots, L\}$,

(e) $:\Longleftrightarrow$ $\boldsymbol{u}$ is α_k-invariant for P for some $k \in \{1, 2, \ldots, L\}$,

are equivalent. Moreover, if $\boldsymbol{u}$ is x-invariant for Q then $x = a_{\boldsymbol{u}} > 0$.

Proof The last claim is proven by summing (5) over $j \in S$ and noting that $x = 0$ would contradict the fact that all states in S are transient.

Since S is finite it follows readily from (3) that an x-invariant distribution for Q is also x-invariant for P. Conversely, taking derivatives in (6) and letting $t \to 0$ yields (5). So (5) and (6) are equivalent, and, as a consequence, (b) $\Longleftrightarrow$ (c) and (d) $\Longleftrightarrow$ (e). Moreover, a simple substitution shows (e) $\Longrightarrow$ (a). To prove (a) $\Longrightarrow$ (b), let $\boldsymbol{u}$ be a quasi-stationary distribution. Then, evidently,

$\mathbb{P}_{\boldsymbol{u}}(X(t) = j) = u_j \mathbb{P}_{\boldsymbol{u}}(T > t)$ for all $j \in S$ and $t \geq 0$, that is,

$$\sum_{i \in S} u_i P_{ij}(t) = u_j \left(1 - \sum_{i \in S} u_i P_{i0}(t) \right), \quad j \in S, \ t \geq 0.$$

Taking derivatives and letting $t \to 0$ subsequently shows that $\boldsymbol{u}$ is $a_{\boldsymbol{u}}$-invariant for Q. This establishes (b), since $a_{\boldsymbol{u}} > 0$ by the last claim.

Finally, we will show (b) $\Longrightarrow$ (d). So let $x > 0$ and assume that $\boldsymbol{u}$ represents an x-invariant distribution for Q, that is, $\boldsymbol{u}Q = -x\boldsymbol{u}$. Recalling that Q is assumed to be 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

$$\boldsymbol{u}_L Q_L = -x \boldsymbol{u}_L.$$

If $\boldsymbol{u}_L \neq \mathbf{0}$ then, by [134, Theorem 2.6] applied to the matrix Q_L, we have $x = \alpha_L$. On the other hand, if $\boldsymbol{u}_L = \mathbf{0}$ we must have

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

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

Theorem 1 identifies all quasi-stationary distributions for $\mathcal{X}$. We call a distribution on S a *limiting conditional distribution* for $\mathcal{X}$ if it is the limiting distribution as $t \to \infty$ of $X(t)$ conditional on survival up to time t, that is,

$$\lim_{t \to \infty} \mathbb{P}_{\boldsymbol{w}}(X(t) = j \,|\, T > t), \quad j \in S, \tag{8}$$

for some initial distribution $\boldsymbol{w} = (w_i, \ i \in S)$. Vere-Jones [148, Theorem 2] has shown (in a more general setting) that if the limits (8) constitute a proper distribution, then this distribution must be a quasi-stationary distribution. Conversely, any quasi-stationary distribution is of course a limiting conditional distribution, so Theorem 1 also identifies all limiting conditional distributions. Evidently, what remains to be solved is the problem of identifying the quasi-stationary distribution (if any) that is the limiting conditional contribution for any given initial distribution.

Noting that $\mathbb{P}_{\boldsymbol{u}}(T > t) = \mathbb{P}_{\boldsymbol{u}}(X(t) \in S) = \sum_{j \in S} \sum_{i \in S} u_i P_{ij}(t)$, the equivalence of the statements (a) and (e) in Theorem 1 immediately yields the following.

Corollary 2 If $\boldsymbol{u} = (u_i,\, i \in S)$ is a quasi-stationary distribution for $\mathcal{X}$ over $S = \cup_{k=1}^{L} S_k$, then

$$\mathbb{P}_{\boldsymbol{u}}(T > t) = e^{-\alpha_k t}, \quad t \geq 0,$$

for some $k \in \{1, 2, \ldots, L\}$.

So the residual survival time conditional on survival up to some time t is exponentially distributed if the initial distribution is a quasi-stationary distribution. In what follows we are also interested in

$$\lim_{t \to \infty} \mathbb{P}_{\boldsymbol{w}}(T > t + s \,|\, T > t), \quad s \geq 0, \tag{9}$$

that is, in the limiting distribution as $t \to \infty$ of the residual survival time conditional on survival up to time t, for *any* initial distribution $\boldsymbol{w} = (w_i,\ i \in S)$.

Irreducible state space

Let us first assume that S is irreducible, that is, $L = 1$, and so $S_1 = S$ and $Q_1 = Q$. Hence $-\alpha$, the eigenvalue of Q with maximal real part, is unique, simple, and negative. It is well known (see, for example, [134, Theorem 2.6] again) that the associated left and right eigenvectors $\boldsymbol{u} = (u_i,\, i \in S)$ and $\boldsymbol{v}^T = (v_i,\, i \in S)^T$ can be chosen strictly positive componentwise, and hence such that

$$\boldsymbol{u}\mathbf{1}^T = 1 \ \text{ and } \ \boldsymbol{u}\boldsymbol{v}^T = 1. \tag{10}$$

Classical Markov-chain theory (see, for example, [83, Theorem A.7]) then tells us that the transition probabilities $P_{ij}(t)$ satisfy

$$e^{\alpha t} P_{ij}(t) = v_i u_j + o(e^{-\varepsilon t}) \ \text{ as } \ t \to \infty, \quad i, j \in S, \tag{11}$$

for some $\varepsilon > 0$, which explains why α is called the decay parameter of $\mathcal{X}$.

Our definition of $\boldsymbol{u}$ implies that the distribution represented by $\boldsymbol{u}$ is α-invariant for Q, whence, by Theorem 1, $\boldsymbol{u}$ is the unique quasi-stationary distribution for $\mathcal{X}$. So if we let $\boldsymbol{u}$ be the initial distribution, then, conditional on survival up to time t, the distribution of $X(t)$ is constant over t, and, by Corollary 2, the remaining survival time has an exponential distribution with parameter α. Darroch and Seneta [40] have shown that similar results hold true in the limit as $t \to \infty$ when the initial distribution differs from $\boldsymbol{u}$. Namely, for any initial distribution $\boldsymbol{w}$ one has

$$\lim_{t\to\infty} \mathbb{P}_{\boldsymbol{w}}(X(t) = j \,|\, T > t) = u_j, \quad j \in S. \tag{12}$$

and

$$\lim_{t\to\infty} \mathbb{P}_{\boldsymbol{w}}(T > t + s \,|\, T > t) = e^{-\alpha s}, \quad s \geq 0. \tag{13}$$

We summarize these results in a theorem.

Theorem 3 [40, Section 3] When all states in S communicate the Markov chain $\mathcal{X}$ has a unique quasi-stationary distribution $\boldsymbol{u} = (u_i,\, i \in S)$, which is the (unique, positive) solution of the system $\boldsymbol{u}Q = -\alpha\boldsymbol{u}$ and $\boldsymbol{u}\mathbf{1}^T = 1$. Moreover, for any initial distribution $\boldsymbol{w} = (w_i,\, i \in S)$ the limits (8) and (9) exist, and are given by (12) and (13), respectively, where $\boldsymbol{u} = (u_i,\, i \in S)$ is the quasi-stationary distribution of $\mathcal{X}$.

General state space

The situation is more complicated when $L \geq 1$, and we must introduce some more notation and terminology before we can state the results. We recall that $-\alpha_k < 0$ is the unique and simple eigenvalue of Q_k with maximal real part, and that $-\alpha = -\min_k \alpha_k < 0$ is the eigenvalue of Q with maximal real part. We let $I_\alpha := \{k : \alpha_k = \alpha\}$, so that $\mathrm{card}(I_\alpha)$ is the algebraic multiplicity of the eigenvalue $-\alpha$, and write $S_{\min} := S_{\min I_\alpha}$. Class S_k will be called a *minimal class for* α if it is a minimal element in the set $\{S_j,\ j \in I_\alpha\}$ with respect to the partial order $\prec$, that is, for all $j \neq k$,

$$S_j \prec S_k \implies j \notin I_\alpha.$$

Letting m_α be the number of minimal classes for α, we have $m_\alpha \geq 1$, since $S_{\min}$ is always a minimal class for α. Moreover, it is shown in [47, Section 6] that there are precisely m_α linearly independent, *nonnegative* vectors $\boldsymbol{u}$ satisfying $\boldsymbol{u}Q = -\alpha\boldsymbol{u}$. (Hence, with g_α denoting the geometric multiplicity of the eigenvalue $-\alpha$, we have $m_\alpha \leq g_\alpha \leq \mathrm{card}(I_\alpha)$.) It is also shown in [47, Section 6] that if $\boldsymbol{u} = (u_i,\, i \in S)$ is a quasi-stationary distribution from which $S_{\min}$ is accessible (by which we mean that there is a state i such that $u_i > 0$ and S_k is accessible from i), then $\boldsymbol{u}$ must satisfy $\boldsymbol{u}Q = -\alpha\boldsymbol{u}$, and $u_i > 0$ if and only if state i is accessible from $S_{\min}$. So, in view of Theorem 1, we can generalize the first part of Theorem 3 as follows.

Theorem 4 [47, Theorem 10] The Markov chain $\mathcal{X}$ has a unique quasi-stationary distribution $\boldsymbol{u}$ from which $S_{\min}$ is accessible if and only if $S_{\min}$ is the only minimal class for α, in which case $\boldsymbol{u}$ is the (unique) nonnegative solution to the system $\boldsymbol{u}Q = -\alpha\boldsymbol{u}$ and $\boldsymbol{u}\mathbf{1}^T = 1$, and has a positive ith component if and only if state i is accessible from $S_{\min}$.

As shown in [47, Section 6] the second part of Theorem 3 can be generalized in the same spirit, leading to the next theorem.

Theorem 5 [47, Theorem 11] If the initial distribution $\boldsymbol{w}$ of the Markov chain $\mathcal{X}$ is such that $S_{\min}$ is accessible, and $S_{\min}$ is the only minimal class for α, then the limits (8) and (9) exist and are given by (12) and (13), respectively, where $\boldsymbol{u}$ is the unique quasi-stationary distribution of $\mathcal{X}$ from which $S_{\min}$ is accessible.

We illustrate the preceding results by an example. Suppose that the generator of $\mathcal{X}$ is given by

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 2 & -2 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 1-\gamma & \gamma & -1 & 0 \\ 0 & 0 & 1 & 1 & -2 \end{pmatrix}, \tag{14}$$

where $0 \leq \gamma \leq 1$. Evidently, we have $S_i = \{i\}$ for $i = 1, 2, 3$ and 4. Moreover, $\alpha = 1$ and $I_\alpha = \{2, 3\}$.

If $0 < \gamma \leq 1$ then S_2 is the only minimal class, so $m_\alpha = 1$. Hence, by Theorem 4 there is a unique quasi-stationary distribution from which state 2 is accessible, which is readily seen to be $(\frac{1}{2}, \frac{1}{2}, 0, 0)$. Observe that the restriction is relevant, since $(1, 0, 0, 0)$ is a quasi-stationary distribution too, but not one from which state 2 is accessible. Theorem 5 tells us that $(\frac{1}{2}, \frac{1}{2}, 0, 0)$ is the limiting conditional distribution for any initial distribution from which state 2 is accessible.

If $\gamma = 0$ then both S_2 and S_3 are minimal classes so $m_\alpha = 2$. In this case there is no unique quasi-stationary distribution from which state 2 is accessible. In fact, it is easy to see that there are infinitely many such distributions. Also the limiting conditional distribution is not unique in this case, but depends on the initial distribution.

3.3 Infinite, irreducible Markov chains

The setting of this subsection is again that of a continuous-time Markov chain $\mathcal{X} := \{X(t),\, t \geq 0\}$ on a state space $\{0\} \cup S$ consisting of an absorbing state 0 and a set of transient states S. But now S is supposed to be countably infinite, so we set $S = \{1, 2, \dots\}$. We restrict ourselves to the case in which all states in S communicate. Note that absorption is not necessarily certain; we set $T = \infty$ if absorption does not occur, so that $\mathbb{P}(T > t) = 1 - \mathbb{P}(T \leq t)$. As before our aim is to identify quasi-stationary and limiting conditional distributions.

Preliminaries

We use the notation of the previous subsection insofar as it extends to the infinite setting at hand, so we let $Q = (q_{ij})$ be the generator of the (sub)Markov chain on S and $\boldsymbol{a} = (a_i,\, i \in S)$ the vector of absorption (or killing) rates. We will assume that Q is stable and conservative, that is,

$$-q_{ii} = a_i + \sum_{j \in S, j \neq i} q_{ij} < \infty, \quad i \in S,$$

and, in addition, that $\mathcal{X}$ is non-explosive and hence uniquely determined by Q.

The result (11) cannot be extended in full to the setting at hand. However, Kingman [86] has shown (see also Anderson [5, Theorem 5.1.9]) that under our assumptions there exist strictly positive constants c_{ij} (with $c_{ii} = 1$) and a parameter $\alpha \geq 0$ such that

$$P_{ij}(t) \leq c_{ij}e^{-\alpha t}, \quad t \geq 0, \ i, j \in S, \tag{15}$$

and

$$\alpha = -\lim_{t\to\infty} \frac{1}{t} \log P_{ij}(t), \quad i, j \in S. \tag{16}$$

Again we will refer to α as the *decay parameter* of $\mathcal{X}$. It is not difficult to see that α is the rate of convergence of the transition probabilities $P_{ij}(t)$ in the sense that

$$\alpha = \inf\left\{a \geq 0 : \int_0^\infty e^{at}P_{ij}(t)dt = \infty\right\}, \quad i, j \in S. \tag{17}$$

The definitions given in the previous section for x-invariant and quasi-stationary distributions remain valid, but the relationships between these notions given in Theorem 1 allow only a partial extension.

Theorem 6 Let $\boldsymbol{u} = (u_i,\, i \in S)$ represent a proper probability distribution over S. Between the statements (a), (b) and (c) defined in Theorem 1 the following relationships exist:

(i) (a) $\Longleftrightarrow$ (c),

(ii) (c) $\Longleftrightarrow$ (b) and $x = a_{\boldsymbol{u}}$.

Moreover,

(iii) (c) $\Longrightarrow$ $0 < x \leq \alpha$.

Proof Statement (i) is implied by Vere-Jones [148, Theorem 2] (see also Nair and Pollett [105, Proposition 3.1]), and statement (ii) combines [148, Theorem 5] (see also Pollett and Vere-Jones [126, Theorem 1]) and [126, Corollary 1]. Finally, [148, Theorem 4] and [126, Theorem 2] together yield statement (iii). $\square$

It is enlightening to point out some differences between this theorem and Theorem 1 with $L = 1$, the corresponding result in a finite setting.

First note that an x-invariant distribution (for Q or for P) in a finite setting must have $x = a_{\boldsymbol{u}} = \alpha$. In the infinite setting of Theorem 6 we have $0 < x = a_{\boldsymbol{u}} \leq \alpha$ if $\boldsymbol{u}$ is x-invariant for P, and we can even have $x < a_{\boldsymbol{u}}$ if $\boldsymbol{u}$ is only x-invariant for Q. Note that summing (5) over all $i \in S$ would yield $x = a_{\boldsymbol{u}}$ if the interchange of summation would be justified, but this is *not* the case in general. In Section 5 we will encounter examples in which x is any number in the interval $(0, \alpha]$.

Secondly, observe that $\alpha > 0$ in a finite setting, but we may have $\alpha = 0$ in an infinite setting. If $\alpha > 0$ the Markov chain $\mathcal{X}$ is called *exponentially transient*. Statements (i) and (iii) of Theorem 6 imply that exponential transience is necessary for the existence of a quasi-stationary distribution.

Thirdly, if $\boldsymbol{u}$ is a quasi-stationary distribution, then, for all $j \in S$,

$$u_j \mathbb{P}_{\boldsymbol{u}}(T > t) = \mathbb{P}_{\boldsymbol{u}}(X(t) = j \,|\, T > t) \mathbb{P}_{\boldsymbol{u}}(T > t) = \mathbb{P}_{\boldsymbol{u}}(X(t) = j). \tag{18}$$

Since, under our assumptions, the right hand side tends to zero as $t \to \infty$ (for any initial distribution $\boldsymbol{u}$), we must have $\mathbb{P}_{\boldsymbol{u}}(T > t) \to 0$ as $t \to \infty$, that is, absorption is certain. This is vacuously true in a finite, but not necessarily in an infinite setting, because there may be a drift to infinity. Moreover, if absorption is certain, then, for $\boldsymbol{u}$ to be a quasi-stationary distribution, it is also necessary for $\mathbb{P}_{\boldsymbol{u}}(X(t) = j)$ and $\mathbb{P}_{\boldsymbol{u}}(T > t)$ to have the same rate of convergence. Again, this is true (for any initial distribution $\boldsymbol{u}$) in a finite, but not necessarily in an infinite setting.

The preceding observation makes us define α_0 as the rate of convergence to zero of $\mathbb{P}_i(T > t) = 1 - P_{i0}(t)$, that is,

$$\alpha_0 := \inf\left\{ a \geq 0 : \int_0^\infty e^{at} \mathbb{P}_i(T > t) dt = \infty \right\}, \quad i \in S, \tag{19}$$

It follows easily from the irreducibility of S that α_0 is independent of i. Moreover, since $P_{ii}(t) \leq \mathbb{P}_i(T > t) \leq 1$, we have

$$0 \leq \alpha_0 \leq \alpha, \tag{20}$$

where each inequality can be strict (Jacka and Roberts [70, Remark 3.1.4]). It will be useful to note that

$$S_{\text{abs}} := \{i \in S : a_i > 0\} \text{ is finite } \implies \alpha_0 = \alpha \tag{21}$$

(see [70, Theorem 3.3.2 (iii)]).

We can now state the following necessary conditions for the existence of a quasi-stationary distribution.

Theorem 7 If there exists a quasi-stationary distribution $\boldsymbol{u}$ for the absorbing Markov chain $\mathcal{X}$, then absorption is certain and $0 < a_{\boldsymbol{u}} \leq \alpha_0 \leq \alpha$.

Proof We concluded already from (18) that absorption must be certain. If $\boldsymbol{u}$ is x-invariant for P, then, by summing (6) over all $j \in S$, we obtain

$$e^{-xt} = \mathbb{P}_{\boldsymbol{u}}(T > t) \geq u_i \mathbb{P}_i(T > t), \quad i \in S, \ t \geq 0.$$

whence $x \leq \alpha_0$. The theorem follows in view of (20) and the statements (i) and (iii) of Theorem 6. □

Thus the condition $\alpha_0 > 0$ is stronger than exponential transience, and necessary for a quasi-stationary distribution to exist. In all examples of quasi-stationary distributions known to us equality of α_0 and α prevails, but we do not know whether $\alpha_0 = \alpha$ is necessary for the existence of a quasi-stationary distribution. (The Markov chain of [70, Remark 3.1.4] satisfies $0 < \alpha_0 < \alpha$ but, in view of our Theorem 17, does not have a quasi-stationary distribution.)

Following [58, Page 515] we call a quasi-stationary $\boldsymbol{u}$ *minimal* if $a_{\boldsymbol{u}} = \alpha_0$. Clearly, the Theorems 6 and 7 lead to another sufficient condition for $\alpha_0 = \alpha$ besides (21).

Corollary 8 If $\boldsymbol{u}$ is an α-invariant quasi-stationary distribution for $\mathcal{X}$, then $\alpha_0 = \alpha$ and $\boldsymbol{u}$ is a minimal quasi-stationary distribution.

We continue this subsection with a survey of sufficient conditions for the existence of a quasi-stationary distribution. In some cases the Markov chain $\mathcal{X}$

is required to be *uniformizable*, which means that the matrix $Q = (q_{ij})$ of transition rates satisfies

$$-q_{ii} = a_i + \sum_{j \in S, j \neq i} q_{ij} \leq C,$$

for some constant C and all $i \in S$.

For *birth-death processes* it is known that exponential transience is necessary and sufficient for the existence of a quasi-stationary distribution when absorption at 0 is certain. Moreover, if the birth and death rates satisfy a certain condition that is weaker than uniformizability, then, for any number x in the interval $0 < x \leq \alpha$, there is a unique quasi-stationary distribution $\boldsymbol{u}$ such that $a_{\boldsymbol{u}} = x$. We postpone giving the details of these results to Subsection 5.1, but note at this point that Kijima [82, Theorem 3.3] has partly generalized these results to Markov chains that are uniformizable and *skip-free to the left,* that is, $q_{ij} = 0$ if $j < i - 1$. Namely, if $\alpha > 0$ and, for some x in the interval $0 < x \leq \alpha$, there is an x-invariant distribution for Q, then, for each y in the interval $x \leq y \leq \alpha$, there is a unique quasi-stationary distribution $\boldsymbol{u}$ such that $a_{\boldsymbol{u}} = y$. Of course, (21) implies that $\alpha_0 = \alpha$ in this case.

More concrete existence results are available in the setting of Markov chains in which *asymptotic remoteness* prevails, that is,

$$\lim_{i \to \infty} \mathbb{P}_i(T \leq t) = 0 \quad \text{for all } t > 0. \tag{22}$$

Then, by [58, Theorem 1.1], $\alpha_0 > 0$ is necessary and sufficient for the existence of a quasi-stationary distribution when absorption at 0 is certain. Moreover, [58, Theorem 4.1 and Proposition 5.1(a)] tell us that there exists an α_0-invariant quasi-stationary distribution if $\mathcal{X}$ is also uniformizable, while, by [58, Corollary 5.3], we must have $\alpha_0 = \alpha$ in that case. Summarizing we can state the following.

Theorem 9 [58] Let the absorbing Markov chain $\mathcal{X}$ be such that absorption is certain and $\alpha_0 > 0$. If asymptotic remoteness prevails then there exists a quasi-stationary distribution for $\mathcal{X}$. If, in addition, $\mathcal{X}$ is uniformizable, then $\alpha_0 = \alpha$ and there exists an α-invariant quasi-stationary distribution.

Since asymptotic remoteness prevails if the Markov chain $\mathcal{X}$ is uniformizable and skip-free to the left, Theorem 9 guarantees the existence of an α-invariant quasi-stationary distribution in the setting of Kijima's paper [82].

Given $\alpha_0 > 0$ and certain absorption, asymptotic remoteness is sufficient but not necessary for the existence of a quasi-stationary distribution. A counterexample is provided by certain birth-death processes (see Subsection 5.1). Other examples of Markov chains having a quasi-stationary distribution while (22) is violated are given by Pakes [117] and Bobrowski [16].

Another approach towards obtaining sufficient conditions for the existence of a quasi-stationary distribution is to confine attention to α-*recurrent* Markov chains, which are Markov chains satisfying

$$\int_0^\infty e^{\alpha t} P_{ii}(t) = \infty \tag{23}$$

for some state $i \in S$ (and then for all states $i \in S$). A chain is called α-*transient* if it is not α-recurrent. For later reference we also note at this point that an α-recurrent process is said to be α-*positive* if for some state $i \in S$ (and then for all states $i \in S$)

$$\lim_{t\to\infty} e^{\alpha t} P_{ii}(t)dt > 0, \tag{24}$$

and α-*null* otherwise. (Note that a *finite* absorbing Markov chain is always α-positive, in view of (11).) It is shown in [86, Theorem 4] that if $\mathcal{X}$ is α-recurrent then there exists, up to normalization, a unique positive solution to the system (5) with $x = \alpha$. However, besides $\alpha_0 > 0$ and certain absorption, additional restrictions on Q are required to ensure summability, and hence the existence of a (unique) α-invariant quasi-stationary distribution, even if $\mathcal{X}$ is α-positive (cf. [136, Page 414]). One such condition is given in the next theorem, which is inspired by the observations on Page 414 of [136] in a discrete-time setting (see also [70, Lemma 3.3.5]).

Theorem 10 Let the absorbing Markov chain $\mathcal{X}$ be such that absorption is certain and $\alpha > 0$. If $\mathcal{X}$ is α-recurrent and S_{abs} is finite, then $\alpha_0 = \alpha$ and there exists a unique α-invariant quasi-stationary distribution.

Proof Let $\mathcal{X}$ be α-recurrent and $\boldsymbol{u}$ the (up to a multiplicative constant) unique solution to the system (5) with $x = \alpha$. Then, by [120, Theorem 1 (iii)], $\boldsymbol{u}$ also solves (6) with $x = \alpha$. Summation over $j \in S_{\text{abs}}$ and integration yields

$$\sum_{i \in S} u_i \int_0^\infty \left(\sum_{j \in S_{\text{abs}}} P_{ij}(t) \right) dt = \alpha^{-1} \sum_{j \in S_{\text{abs}}} u_j,$$

since S_{abs} is finite and $\alpha > 0$. The integral represents the expected sojourn time in S_{abs} when the initial state is i, so is finite and bounded below by $(\max_{j \in S_{\text{abs}}} |q_{jj}|)^{-1} > 0$. It follows that $\boldsymbol{u}$ must be summable, and hence can be normalized to be a distribution, which, by Theorem 6, must then be a quasi-stationary distribution. From (21) we know already that $\alpha_0 = \alpha$ if S_{abs} is finite.

□

Since α-recurrence is usually difficult to verify, one might attempt to replace it by a condition which is stated directly in terms of Q. A result of this type is the continuous-time counterpart of Kesten's result [79, Theorem 2 and Page 657]. Insofar as it concerns quasi-stationary distributions, this result states that – besides finiteness of S_{abs} and certain absorption – uniformizability, certain restrictions on the number of nonzero downward rates, and a type of uniform irreducibility condition are sufficient for the existence of a unique α-invariant quasi-stationary distribution.

As observed already in a finite setting, any quasi-stationary distribution $\boldsymbol{u}$ for $\mathcal{X}$ is a *limiting conditional distribution*, in the sense that by a suitable choice of $\boldsymbol{w}$ (namely $\boldsymbol{w} = \boldsymbol{u}$) the limits (12) constitute a distribution represented by $\boldsymbol{u}$. Conversely, by [148, Theorem 2] (see also [117, Lemma 2.1]), any limiting conditional distribution must be a quasi-stationary distribution. So our quest for conditions on Q for a quasi-stationary distribution to exist may also be brought to bear on limiting conditional distributions. Evidently, if, for some initial distribution $\boldsymbol{w}$, the limits (12) exist and constitute a proper distribution (and, hence, a quasi-stationary distribution), then the rates of convergence of $\mathbb{P}_{\boldsymbol{w}}(X(t) = j)$ and $\mathbb{P}_{\boldsymbol{w}}(T > t)$ as $t \to \infty$ must be equal. Restricting ourselves to initial distributions that are concentrated on a single state, these observations lead to the following result.

Theorem 11 Let the absorbing Markov chain $\mathcal{X}$ be such that for some $i \in S$, the limits

$$u_j = \lim_{t\to\infty} \mathbb{P}_i(X(t) = j \,|\, T > t), \quad j \in S, \tag{25}$$

exist and constitute a proper distribution. Then absorption is certain, $\alpha = \alpha_0 > 0$, and $\boldsymbol{u} = (u_j,\ j \in S)$ is an α-invariant quasi-stationary distribution.

Proof If the limits (25) constitute a proper distribution then, as noted above, $\boldsymbol{u}$ is a quasi-stationary distribution, so that absorption must be certain and $a_{\boldsymbol{u}} > 0$. Moreover, since the rates of convergence of $\mathbb{P}_i(X(t) = j)$ and $\mathbb{P}_i(T > t)$ must be equal, we have $\alpha = \alpha_0$. Finally, by the argument given in the proof of [70, Lemma 4.1], $\boldsymbol{u}$ is an α-invariant distribution, so that $\alpha = a_{\boldsymbol{u}} > 0$. □

Remarks (i) In the statement of [70, Lemma 4.1] it is required that the limits (25) exist for *all* $i \in S$, but this is not used in the proof of the lemma.
(ii) Proposition 5.1(b) in [58] is similar to our Theorem 11, but contains the unnecessary requirement that $\mathcal{X}$ be uniformizable.

The existence of the limits in (25) has been proven in various settings, usually more restricted, however, than those required for the existence of a quasi-stationary distribution (see, for example, [131, 30]).

The last theorem of this section is a partial converse to Theorem 11 and gives a sufficient condition for the existence of the limits (8) and (9) when the initial distribution is concentrated on a single state. The theorem constitutes the continuous-time counterpart of (part of) [136, Theorem 3.1]. It can readily be proven with the help of [86, Theorem 4], but may also be established by combining the results of our Theorem 6 and [59, Theorem 1].

Theorem 12 Let the absorbing Markov chain $\mathcal{X}$ be such that absorption is certain and $\alpha > 0$. If $\mathcal{X}$ is α-positive and there exists a (unique) α-invariant quasi-stationary distribution $\boldsymbol{u} = (u_j,\ j \in S)$, then, for all $i \in S$,

$$\lim_{t\to\infty} \mathbb{P}_i(X(t) = j \,|\, T > t) = u_j, \quad j \in S, \tag{26}$$

and

$$\lim_{t\to\infty} \mathbb{P}_i(T > t + s \,|\, T > t) = e^{-\alpha s}, \quad s \geq 0. \tag{27}$$

Evidently, the statement of the theorem may be generalized to initial distributions with finite support.

We conclude this section with the observation that in the preceding theorems the condition of certain absorption can be relaxed, provided we work with the process "restricted to the event of absorption". Indeed, this allows us to deal with

$$\lim_{t\to\infty} \mathbb{P}_i(X(t) = j \,|\, t < T < \infty), \quad j \in S,$$

via the dual chain $\tilde{p}_{ij}(t) = p_{ij}(t)e_j/e_i$, $i, j \in S$, where $e_i = \mathbb{P}_i(T < \infty)$ is the probability of absorption starting in state i (see Waugh [149]). Our irreducibility assumption ensures that $e_i > 0$ for all i.

3.4 Discrete-time Markov chains

Most results for continuous-time Markov chains have more or less obvious analogues for discrete-time Markov chains. In one respect the discrete-time setting is simpler, since the requirement of uniformizability that we have encountered in several results of the previous section has no bearing on discrete-time Markov chains. On the other hand, the phenomenon of periodicity may pose problems in a discrete-time setting, in particular when considering limiting conditional distributions. In this subsection we will briefly describe how discrete-time results may be obtained from continuous-time results, and give appropriate references for further details.

So let $\mathcal{Y} := \{Y_n,\, n = 0, 1, \dots\}$ be a discrete-time Markov chain taking values in a state space consisting of an absorbing state 0 and a finite or countably infinite set of *transient* states S. We denote the matrix of 1-step transition probabilities *within* S by $P := (p_{ij},\, i, j \in S)$, and let p_{i0}, $i \in S$, be the 1-step absorption probabilities. The n-step transition probabilities are denoted by $P_{ij}(n)$, and the matrix of n-step transition probabilities *within* S by $P(n) := (P_{ij}(n),\ i, j \in S)$, so that $P_{ij}(1) = p_{ij}$, and $P(n) = P^n$. A vector $\boldsymbol{u} = (u_j,\ j \in$

S), or a proper probability distribution over S represented by $\boldsymbol{u}$, are called *x-invariant* for P if

$$\sum_{i \in S} u_i p_{ij} = x u_j, \quad j \in S. \tag{28}$$

By analogy with (7) and (8) the distribution $\boldsymbol{u}$ is said to be a *quasi-stationary distribution* for $\mathcal{Y}$ if, for all $n = 0, 1, \ldots,$

$$\mathbb{P}_{\boldsymbol{u}}(Y(n) = j \,|\, T > n) = u_j, \quad j \in S, \tag{29}$$

and *a limiting conditional distribution* for $\mathcal{Y}$ if, for some initial distribution $\boldsymbol{w}$,

$$\lim_{n \to \infty} \mathbb{P}_{\boldsymbol{w}}(Y(n) = j \,|\, T > n) = u_j, \quad j \in S. \tag{30}$$

As before $T := \inf\{n \geq 0 : Y(n) = 0\}$ denotes the *absorption time*, and $\mathbb{P}_{\boldsymbol{w}}(.)$ is the probability measure of the process when the initial distribution is $\boldsymbol{w}$.

The case in which S is finite, irreducible and aperiodic was analysed in the classic paper of Darroch and Seneta [39]. Their results have recently been generalized in [48] to a reducible setting, leading to discrete-time analogues of the Theorems 1, 4 and 5, with the restriction that the analogue of the latter requires an additional aperiodicity condition. We refer to [48] for details.

If S is countably infinite and irreducible there exists a real number ρ (the *decay parameter* of the Markov chain $\mathcal{Y}$ in S) such that $0 < \rho \leq 1$ and, for each $i,\ j \in S$,

$$(P_{ij}(n))^{1/n} \to \rho, \tag{31}$$

as $n \to \infty$ through the residue class modulo the period of P for which the sequence $\{P_{ij}(n)\}$ is not identically zero. (This result was stated for aperiodic chains in [146]; the generalization was observed in [86].) The chain is said to be *geometrically transient* if $\rho < 1$. A link with the results of Subsection 3.3 is established by observing that for any $q > 0$ we can associate with $\mathcal{Y}$ a continuous-time, uniformizable Markov chain $\mathcal{X}_q$ on $S \cup \{0\}$ by letting $Q = (q_{ij})$ and $\boldsymbol{a} = (a_i,\ i \in S)$ such that

$$q_{ij} = q(p_{ij} - \delta_{ij}), \quad a_i = q p_{i0}, \quad i, j \in S,$$

where δ_{ij} is Kronecker's delta. Namely, the decay parameter α_q of $\mathcal{X}_q$ is easily seen to satisfy $\alpha_q = q(1-\rho)$. Moreover, a vector $\boldsymbol{u} = (u_i,\, i \in S)$ is x-invariant for Q if and only if $\boldsymbol{u}$ is $(1 - x/q)$-invariant for P, and $\boldsymbol{u}$ is a quasi-stationary distribution for $\mathcal{X}_q$ if and only if it is a quasi-stationary distribution for $\mathcal{Y}$.

These observations enable us to translate all results for continuous-time, uniformizable Markov chains in Subsection 3.3 to the discrete-time setting at hand, with the restriction that we have to impose aperiodicity of S in statements involving limiting conditional distributions. For details we refer to Coolen-Schrijner and van Doorn [35].

We finally note that the existence of the limits (30) as a bona fide distribution – and hence the existence of a quasi-stationary distribution – has been proven in various settings, usually more restricted, however, than those required by results such as the discrete-time counterparts of the Theorems 9 and 10 (see, for example, Seneta and Vere-Jones [136], Daley [36], Pakes [116], Kijima [80, 81], Kesten [79], van Doorn and Schrijner [49, 50], Ferrari et al. [57] and Moler et al. [104]).

4 Speed of convergence to quasi stationarity

In the previous section we have focused on the existence of quasi-stationary distributions, given the parameters that determine the dynamics of the process, and on the identification of a quasi-stationary distribution as the limiting conditional distribution for a given initial distribution. However, as noted in the first paragraph of Section 2, the existence of a limiting conditional distribution does not necessarily imply that the process exhibits the type of "quasi stationary" behaviour depicted in Figure 1, characterized by relatively fast convergence to the limiting conditional distribution, and eventual evanescence after a much longer time. In the present section we will focus on the circumstances under which such a "quasi stationarity" scenario prevails, given the existence of a limiting conditional distribution.

As before our setting will be that of a continuous-time Markov chain $\mathcal{X}$ on a finite or countably infinite state space $\{0\} \cup S$. For convenience we will assume that Q, the generator of the (sub)Markov chain on S, is irreducible, and that, for some initial state $i \in S$, the limits (25) exist and constitute a proper distribution (and hence a quasi-stationary distribution) $\boldsymbol{u} = (u_j,\ j \in S)$. Theorem 11, which is obviously valid in a finite setting too, then tells us that absorption at 0 must be certain and $\alpha = \alpha_0 > 0$. Our task will be to characterize the time-scale on which the conditional probabilities $\mathbb{P}_i(X(t) = j \,|\, T > t)$ converge to their limits u_j, and to relate it to the time-scale on which absorption takes place.

We characterize the time-scale on which absorption takes place by α_0, the (common) rate of convergence to zero of the probabilities $\mathbb{P}_i(T > t) = 1 - P_{i0}(t)$, and the time-scale on which the conditional probabilities $\mathbb{P}_i(X(t) = j \,|\, T > t)$ converge to their limits $u_j,\ j \in S$, by

$$\beta := \inf\{\beta_{ij},\ i, j \in S\}. \tag{32}$$

Here β_{ij} is the rate of convergence of the conditional probability $\mathbb{P}_i(X(t) = j \,|\, T > t)$ to its limit u_j, that is,

$$\beta_{ij} := \inf\left\{a \geq 0 : \int_0^\infty e^{at} |\mathbb{P}_i(X(t) = j \,|\, T > t) - u_j| dt = \infty\right\}, \quad i, j \in S. \tag{33}$$

First assuming that the state space of $\mathcal{X}$ is finite, we know that $-\alpha$, the eigenvalue of Q with the largest real part, is real, simple and negative. It follows from classical matrix theory (see, for example, Schmidt [130, Theorem 1.3]) that if $-\alpha_2$ is the eigenvalue of Q with next largest real part (and largest multiplicity, if there is more than one such eigenvalue) then the result (11) can be stated more precisely as

$$e^{\alpha t} P_{ij}(t) = v_i u_j + \mathcal{O}(t^{m-1} e^{-\gamma t}) \ \text{ as } \ t \to \infty, \quad i, j \in S, \tag{34}$$

where m is the multiplicity of $-\alpha_2$ and

$$\gamma := \mathrm{Re}(\alpha_2) - \alpha > 0. \tag{35}$$

We will refer to γ as the *spectral gap* of Q (or $\mathcal{X}$). Since $\alpha_0 = \alpha$, it now follows readily that $\mathbb{P}_i(X(t) = j \,|\, T > t) - u_j = \mathcal{O}(t^{m-1} e^{-\gamma t})$, so that $\beta_{ij} \geq \gamma$ for all $i, j \in S$, and hence $\beta \geq \gamma > 0$. Perhaps examples can be constructed in which $\beta > \gamma$, but as a rule there will be some $i, j \in S$, such that the spectral expansion of $P_{ij}(t) - u_j \mathbb{P}_i(T > t)$ involves a term that is the product of $e^{-\alpha_2 t}$ and a nonzero polynomial in t, in which case $\beta = \gamma$.

The preceding observations lead to the conclusion that the type of "quasi stationary" behaviour depicted in Figure 1 occurs in the setting of a finite Markov chain only if the spectral gap γ is substantially larger than the decay parameter α. This fact was noted already in [39] in a discrete-time setting. (See [37, 38] for a similar conclusion in the setting of a finite birth-death process.)

The situation is more complicated when the state space $\{0\} \cup S$ of the Markov chain $\mathcal{X}$ is countably infinite, all other circumstances being unchanged. Again α_0 (as defined in (19)) may be used to characterize the time-scale on which absorption takes place, and we still have $\alpha_0 = \alpha$, but we must resort to operator theory to determine the rate of convergence of $\mathbb{P}_i(X(t) = j \mid T > t) - u_j$, by extending the notion of *spectral gap* to Markov chains with a countably infinite state space. While the spectral gap for ergodic Markov chains, and in particular reversible ergodic Markov chains, has received quite some attention in the literature (see, for example, Chen [27] and the references there), much less is known for absorbing Markov chains. However, if we restrict ourselves to birth-death processes (with killing), as we do in the next section, more detailed results can be obtained. Since the behaviour of birth-death processes is often indicative of what holds true in much greater generality, the results for birth-death processes given in the next section raise the expectation that for a wide class of Markov chains α-positivity is necessary (but by no means sufficient) for $\gamma > 0$, and hence for the type of quasi-stationary behaviour depicted in Figure 1.

In contrast, Pollett and Roberts [124] explain quasi stationarity using a dynamical systems approach. They proved, under mild conditions, that the Kolmogorov forward equations always admit a *centre manifold* consisting of points on a line in the unit $|S|$-simplex connecting the quasi-stationary distribution $\boldsymbol{u}$ with the degenerate limiting distribution $\boldsymbol{\pi}_0 = (1\ \mathbf{0})$, meaning that the state probability vector moves exponentially quickly to a region near that line, before moving slowly to $\boldsymbol{\pi}_0$.

5 Birth-death and related processes

In this section we will analyse in detail some special continuous-time Markov chains with a countably infinite, irreducible state space. So our setting is that of Subsection 3.3. We consider birth-death processes in Subsection 5.1 and birth-death processes *with killing* in Subsection 5.2.

5.1 Birth-death processes

The Markov chain $\mathcal{X}$ of Subsection 3.3 is now a *birth-death process*, that is, the generator (1) of $\mathcal{X}$ satisfies $q_{ij} = 0$ if $|i - j| > 1$. Since $S = \{1, 2, \ldots\}$ is supposed to be irreducible we must have

$$\lambda_i := q_{i,i+1} > 0 \text{ and } \mu_{i+1} := q_{i+1,i} > 0, \quad i \in S.$$

We also require that $a_1 > 0$ and $a_i = 0$ for $i > 1$, so that absorption at state 0 (*killing*) can only occur via state 1. (This assumption will be relaxed in the next subsection.) The parameters λ_i and μ_i are the *birth rate* and *death rate*, respectively, in state i. In the literature the killing rate a_1 is usually referred to as the death rate in state 1 (and denoted by μ_1). Throughout this subsection the birth and death rates are assumed to satisfy

$$\sum_{n=1}^{\infty} \frac{1}{\lambda_n \pi_n} = \infty, \tag{36}$$

where

$$\pi_1 := 1 \;\text{ and }\; \pi_n := \frac{\lambda_1 \lambda_2 \ldots \lambda_{n-1}}{\mu_2 \mu_3 \ldots \mu_n}, \quad n > 1, \tag{37}$$

which is necessary and sufficient for absorption to be certain and, hence, sufficient for $\mathcal{X}$ to be non-explosive (see, for example, [5, Section 8.1]).

Karlin and McGregor [75] have shown that the transition probabilities $P_{ij}(t)$ can be represented as

$$P_{ij}(t) = \pi_j \int_0^\infty e^{-xt} Q_{i-1}(x) Q_{j-1}(x) \psi(dx), \quad t \geq 0, \; i, j \in S. \tag{38}$$

Here $\{Q_n(.)\}$ is a sequence of polynomials satisfying the recurrence relation

$$\begin{aligned} \lambda_n Q_n(x) &= (\lambda_n + \mu_n - x) Q_{n-1}(x) - \mu_n Q_{n-2}(x), \quad n > 1, \\ \lambda_1 Q_1(x) &= \lambda_1 + a_1 - x, \quad Q_0(x) = 1, \end{aligned} \tag{39}$$

and ψ – the *spectral measure* of $\mathcal{X}$ – is the (unique) Borel measure of total mass 1 on the interval $(0, \infty)$ with respect to which the *birth-death polynomials* $Q_n(.)$ are orthogonal.

The next theorem shows how the decay parameters α, α_0 and β are related to $\text{supp}(\psi)$ – the support of the measure ψ, also known as the *spectrum* of $\mathcal{X}$ – and more specifically to

$$\xi_1 := \inf \text{supp}(\psi) \quad \text{and} \quad \xi_2 := \inf\{\text{supp}(\psi) \cap (\xi_1, \infty)\}. \tag{40}$$

The difference $\xi_2 - \xi_1$ is the *spectral gap* of $\mathcal{X}$. Not surprisingly, it can be interpreted as the limit as $n \to \infty$ of the spectral gap (as defined in (35)) of the suitably truncated birth-death process on $\{0, 1, 2, \ldots, n\}$. Interestingly, ξ_1 and ξ_2 are also the limits as $n \to \infty$ of the smallest zero and second smallest zero, respectively, of the polynomial $Q_n(x)$, all of whose zeros are distinct, real and positive (see, for example, Chihara [31]).

Theorem 13 The birth-death process $\mathcal{X}$ has $\alpha = \alpha_0 = \xi_1$ and $\beta = \xi_2 - \xi_1$.

Proof Evidently, (21) implies $\alpha = \alpha_0$. The representation for α was established by Callaert [21], and that for β in [43, Theorem 5]. □

In the setting at hand the existence of a quasi-stationary distribution can be established under much weaker conditions than those of the Theorems 9 or 10. In fact, certain absorption and exponential transience, which are necessary conditions by Theorem 7, happen to be sufficient as well. Moreover, all quasi-stationary distributions can actually be identified. A crucial role is played by the series

$$\sum_{n=1}^{\infty} \frac{1}{\lambda_n \pi_n} \sum_{j=n+1}^{\infty} \pi_j, \tag{41}$$

which, by a result of Keilson's [77] (see also [5, Section 8.1] or [83, Section 5.1]), can be interpreted as the limit as $n \to \infty$ of the expected first passage time from state n to state 1.

Theorem 14 [43, Theorem 3.2] Let the birth-death process $\mathcal{X}$ be such that absorption is certain and $\alpha > 0$. If the series (41) diverges then there is a one-parameter family of quasi-stationary distributions $\{\boldsymbol{u}(x) = (u_i(x),\, i \in S),\ 0 < x \leq \alpha\}$ for $\mathcal{X}$, where

$$u_i(x) = \frac{x\pi_i Q_{i-1}(x)}{a_1}, \quad i \in S. \tag{42}$$

If the series (41) converges then there is precisely one quasi-stationary distribution for $\mathcal{X}$, namely $\boldsymbol{u}(\alpha)$.

It is enlightening to interpret $\boldsymbol{u}(x) = (u_i(x),\, i \in S)$ of (42) from the more general perspective of Theorem 6. Indeed, it is not difficult to see that $\boldsymbol{u}(x)$ is the unique x-invariant vector for Q satisfying $x = a_{\boldsymbol{u}(x)}(= a_1 u_1(x))$. But $\boldsymbol{u}(x)$ represents a proper distribution (if and) only if $x = \alpha$, or, $0 < x < \alpha$ and the series (41) diverges. If $0 < x < \alpha$ and the series (41) converges, it is possible to renormalize $\boldsymbol{u}(x)$ such that a proper distribution $\boldsymbol{v}$ results, but then $x < a_{\boldsymbol{v}}$ (see [43] for details).

We notice that asymptotic remoteness (see (22)) implies divergence of the series (41). Indeed, by Karlin and McGregor [76, Theorem 10] (or, for example, [95, Corollary 1.2.4.1]) we have

$$\mathbb{E}_i(T) = \frac{1}{a_1}\sum_{j=1}^{\infty}\pi_j + \sum_{n=1}^{i-1}\frac{1}{\lambda_n \pi_n}\sum_{j=n+1}^{\infty}\pi_j, \quad i \in S, \tag{43}$$

while $\mathbb{E}_i(T) \geq t\,\mathbb{P}_i(T > t)$ for all $t \geq 0$ by Markov's inequality. So if asymptotic remoteness prevails we must have $\mathbb{E}_i(T) \to \infty$ as $i \to \infty$, implying divergence of (41). (Actually, the converse holds true as well, cf. [117, Lemma 2.2]). We also remark that divergence of (41) is equivalent to the Kolmogorov forward equations having a unique solution (see, for example, [5]), the prevailing situation in most practical models.

As announced already we can conclude from Theorem 14 that certain absorption and exponential transience are necessary and sufficient conditions for the existence of a quasi-stationary distribution when we are dealing with a birth-death process. Given the birth and death rates verification of certain absorption is trivial through (36), but it is less straightforward to establish exponential transience. Some necessary and some sufficient conditions, which settle the problem for most processes encountered in practice, have been collected in [42]. One such result is

$$\alpha > 0 \implies \sum_{n=1}^{\infty} \pi_n < \infty, \tag{44}$$

which is implied by [76, Equation (9.19)] and the fact that $\alpha = \xi_1$. Also, a complete solution is given in [42] for processes having birth rates λ_i and death rates μ_i that are asymptotically rational functions of i.

Next turning to limiting conditional distributions the situation is again quite straightforward in the setting at hand, provided we restrict ourselves to initial distributions that are concentrated on a single state (or a finite set of states).

Theorem 15 [43, Theorem 4.1] Let the birth-death process $\mathcal{X}$ be such that absorption is certain. Then, for all $i \in S$,

$$\lim_{t\to\infty} \mathbb{P}_i(X(t) = j \mid T > t) = \frac{\alpha \pi_j Q_{j-1}(\alpha)}{a_1}, \quad j \in S, \tag{45}$$

so that these conditional limits constitute the minimal quasi-stationary distribution for $\mathcal{X}$ if $\alpha > 0$.

In Section 4 we have emphasized that one should use the (minimal) quasi-stationary distribution to approximate the unconditional time-dependent distribution of an absorbing Markov chain only if the decay parameter α of the unconditional process is substantially smaller than the decay parameter β of the process conditioned on nonabsorption. Unfortunately, determining $\beta = \gamma$ (the spectral gap) is usually at least as difficult as determining $\alpha = \xi_1$ (the first point in the spectrum). However, some information on β can be obtained if α is known explicitly. Namely, Miclo [103] and Chen [26, Theorem 3.5] have presented a necessary and sufficient condition for the spectral gap $\xi_2 - \xi_1$ to be positive in the setting of a non-absorbing, non-explosive, ergodic birth-death process (in which case $\xi_1 = 0$). With the transformation technique employed, for instance, in [44, Section 5] this result can be translated into the setting at hand, yielding

$$\beta > 0 \iff \sup_i \left(\sum_{n=1}^{i} \frac{1}{\lambda_n \pi_n Q_{n-1}(\alpha) Q_n(\alpha)} \right) \left(\sum_{n=i+1}^{\infty} \pi_n Q_{n-1}^2(\alpha) \right) < \infty. \tag{46}$$

It follows in particular that

$$\beta > 0 \implies \sum_{n=1}^{\infty} \pi_n Q_{n-1}^2(\alpha) < \infty. \tag{47}$$

As an aside we note that, by a classic result in the theory of orthogonal polynomials (see Shohat and Tamarkin [138, Corollary 2.6]), the conclusion in (47) is equivalent to the spectral measure ψ having a point mass at α, which is obviously necessary for the spectral gap to be positive. Interestingly, [46, Theorem 3.1] (which is a corrected version of [44, Theorem 5.1]) tells us that the conclusion is also equivalent to α-positivity of $\mathcal{X}$. In other words, an absorbing birth-death process with decay parameter α will have $\beta = 0$ if it is α-transient or α-null. As mentioned already in the previous section, we suspect this conclusion to be valid in much greater generality.

We conclude this subsection with a worked-out example. Consider then a birth-death process $\mathcal{X}$ with λ_1 and a_1 positive but otherwise unspecified, and constant rates $\lambda_i = \lambda$ and $\mu_i = \mu$ for $i > 1$. As a consequence the constants π_n of (37) are given by

$$\pi_1 = 1 \text{ and } \pi_n = \frac{\lambda_1}{\mu}\left(\frac{\lambda}{\mu}\right)^{n-2}, \quad n > 1.$$

Throughout we will assume that $\lambda \leq \mu$, so that (36) is satisfied and hence absorption at 0 is certain. From [45, Section 5] we learn that the smallest limit point in $\mathrm{supp}(\psi)$, the support of the spectral measure ψ of $\mathcal{X}$, is given by $\sigma := (\sqrt{\mu} - \sqrt{\lambda})^2$, and that ψ will have an isolated point mass to the left of σ if and only if

$$a_1 - \lambda_1(\sqrt{\mu/\lambda} - 1) < \sigma. \tag{48}$$

In this case the isolated smallest point in the support of ψ is given by the single root in the interval $(0, \sigma)$ of the equation $z - \lambda_1 - a_1 - \lambda_1\mu G(z) = 0$, where

$$G(z) := \frac{1}{2\lambda\mu}\left(z - \lambda - \mu + \sqrt{(z-\lambda-\mu)^2 - 4\lambda\mu}\right).$$

A little algebra reveals that this root is given by $\lambda_1(1-\nu) + a_1$, where

$$\nu := \frac{\lambda_1 - \lambda + a_1 - \mu + \sqrt{(\lambda_1 - \lambda + a_1 - \mu)^2 + 4\mu(\lambda_1 - \lambda)}}{2(\lambda_1 - \lambda)}. \tag{49}$$

We conclude that the decay parameter α of the process $\mathcal{X}$ is given by

$$\alpha = \xi_1 = \begin{cases} \lambda_1(1-\nu) + a_1 & \text{if } \; a_1 - \lambda_1(\sqrt{\mu/\lambda} - 1) < \sigma \\ \sigma & \text{otherwise,} \end{cases} \tag{50}$$

which, for constant λ_1, is seen to increase from 0 for $a_1 = 0$ to σ for $a_1 = \sigma + \lambda_1(\sqrt{\mu/\lambda} - 1)$, while, for constant a_1, it decreases to 0 as λ_1 increases to infinity. We also have

$$\beta = \xi_2 - \xi_1 = \begin{cases} \sigma - a_1 - \lambda_1(1-\nu) & \text{if } \; a_1 - \lambda_1(\sqrt{\mu/\lambda} - 1) < \sigma \\ 0 & \text{otherwise,} \end{cases} \tag{51}$$

which, for constant λ_1, decreases from σ to 0 as a_1 increases from 0 to $\sigma + \lambda_1(\sqrt{\mu/\lambda} - 1)$, while, for constant a_1, it increases to infinity as λ_1 increases to infinity.

To determine the minimal quasi-stationary distribution we note that after some algebraic manipulations the polynomials $Q_n(x)$ can be represented as

$$Q_n(x) = \frac{(\lambda_1 + a_1 - \lambda_1 z_1 - x)z_2^n - (\lambda_1 + a_1 - \lambda_1 z_2 - x)z_1^n}{\lambda(z_2 - z_1)}, \quad n \geq 0, \tag{52}$$

where $z_1 = z_1(x)$ and $z_2 = z_2(x)$ are the roots of the equation $\lambda z^2 - (\lambda + \mu - x)z + \mu = 0$ (with appropriate adaptations if the two roots are identical). First assuming that (48) is satisfied, so that $\alpha = \lambda_1(1-\nu) + a_1$, we obtain after some algebra the roots $z_1(\alpha) = \nu$ and $z_2(\alpha) = \mu/(\lambda\nu)$. It follows that in this case the minimal quasi-stationary distribution is given by

$$u_i(\alpha) = \frac{\alpha \pi_i Q_{i-1}(\alpha)}{a_1} = \begin{cases} 1 - \dfrac{\lambda_1}{a_1}(1-\nu), & i = 1 \\ \dfrac{\lambda_1}{\lambda}\left(1 - \dfrac{\lambda_1}{a_1}(1-\nu)\right)\left(\dfrac{\lambda\nu}{\mu}\right)^i, & i > 1. \end{cases} \tag{53}$$

If, however, (48) is not satisfied, then $\alpha = \sigma$ and $z_1(\alpha) = z_2(\alpha) = \sqrt{\mu/\lambda}$, and hence

$$Q_n(\alpha) = \left\{1 + \frac{1}{\lambda_1}\sqrt{\frac{\lambda}{\mu}}\left(a_1 - \lambda_1\left(\sqrt{\frac{\mu}{\lambda}} - 1\right) - \sigma\right)n\right\}\left(\frac{\mu}{\lambda}\right)^{n/2}, \quad n \geq 0.$$

It follows that the quasi-stationary distribution is given by

$$\begin{aligned} u_i(\alpha) = \frac{\alpha \pi_i Q_{i-1}(\alpha)}{a_1} &= \frac{\sigma}{\lambda a_1}\left((\lambda - \lambda_1)\mathbb{I}_{\{i=1\}} + \lambda_1\left(\frac{\lambda}{\mu}\right)^{i/2}\right) \\ &+ \frac{\sigma}{a_1\mu}\left(\lambda_1 + a_1 - \frac{\lambda_1}{\lambda}\sqrt{\lambda\mu} - \sigma\right) i \left(\frac{\lambda}{\mu}\right)^{(i-1)/2}, \quad i \in S. \end{aligned} \tag{54}$$

In the special case $\lambda_1 = \lambda$ this is a mixture of a geometric and a negative binomial distribution, namely,

$$u_i(\alpha) = p(1-r)r^i + (1-p)(1-r)^2 ir^{i-1}, \quad i \in S, \tag{55}$$

where $p := \mu(1-\sqrt{\lambda/\mu})/a_1$ and $r := \sqrt{\lambda/\mu}$.

¿From [44, Section 6] we know that $\mathcal{X}$ is α-positive if (48) is satisfied, and α-transient otherwise.

5.2 Birth-death processes with killing

As announced we generalize the setting of the previous subsection by allowing absorption from any state $i \in S$ rather than only state 1, that is, we allow $a_i \geq 0$ for all $i \in S$. We will assume that absorption is certain, which, by van Doorn and Zeifman [51, Theorem 1], is now equivalent to

$$\sum_{n=1}^{\infty} \frac{1}{\lambda_n \pi_n} \sum_{j=1}^{n} a_j \pi_j = \infty. \tag{56}$$

Evidently, we must have $a_i > 0$ for at least one state $i \in S$. The representation (38) remains valid provided we redefine the polynomials $Q_n(.)$ by means of the recurrence relation

$$\begin{aligned} \lambda_n Q_n(x) &= (\lambda_n + \mu_n + a_n - x)Q_{n-1}(x) - \mu_n Q_{n-2}(x), \quad n > 1, \\ \lambda_1 Q_1(x) &= \lambda_1 + a_1 - x, \quad Q_0(x) = 1, \end{aligned} \tag{57}$$

which reduces to (39) in the specific setting of the previous subsection. The quantities π_n are as in (37), and the measure ψ is again the (unique) Borel measure of total mass 1 on the interval $(0, \infty)$ with respect to which the polynomials $Q_n(.)$ are orthogonal. Defining ξ_1 and ξ_2 as in (40) we still have $\alpha = \xi_1$ and $\beta = \xi_2 - \xi_1$, but we do not necessarily have $\alpha_0 = \alpha$ any longer. (See [35] for proofs and developments.) However, by (21), we do have $\alpha_0 = \alpha$ if S_{abs} is finite, and under this condition Theorem 14 can be generalized as follows.

Theorem 16 [35, Theorems 6.5 and 6.6] Let the birth-death process with killing $\mathcal{X}$ be such that S_{abs} is finite, absorption is certain, and $\alpha > 0$. If the series (41) diverges then there is a one-parameter family of quasi-stationary distributions $\{\boldsymbol{u}(x) = (u_i(x),\, i \in S),\ 0 < x \leq \alpha\}$ for $\mathcal{X}$, where

$$u_i(x) = c(x)\pi_i Q_{i-1}(x), \quad i \in S, \tag{58}$$

and $c(x)$ is a normalizing constant. If the series (41) converges then there is precisely one quasi-stationary distribution for $\mathcal{X}$, namely $\boldsymbol{u}(\alpha)$.

We note that Li and Li [97, Theorem 6.2(i)] show that if, under the conditions of this theorem, S_{abs} is finite is replaced by the weaker condition $\lim_{i\to\infty} a_i = 0$, then divergence of (41) is sufficient for asymptotic remoteness, and hence, by Theorem 9, for the existence of a quasi-stationary distribution.

Without a restriction on S_{abs} the situation is less clear. Indeed, let $0 < x \leq \alpha$. It is easy to see that, up to a multiplicative constant, $(\pi_i Q_{i-1}(x),\ i \in S)$ is the unique x-invariant vector for Q. We also have $\pi_i Q_{i-1}(x) > 0$, as shown in [35]. So for $\boldsymbol{u}(x) = (u_i(x),\, i \in S)$ of (58) to be a quasi-stationary distribution it is, in view of Theorem 6, necessary and sufficient that $\sum_{i\in S} \pi_i Q_{i-1}(x)$ converges and $x = a_{\boldsymbol{u}}$, or equivalently,

$$x \sum_{i\in S} \pi_i Q_{i-1}(x) = \sum_{i\in S} a_i \pi_i Q_{i-1}(x) < \infty. \tag{59}$$

In view of the last statement of Theorem 16 the equality sign need not prevail, but even if it does the sums need not be bounded. Of course the latter can happen only if S_{abs} is infinite. Examples are given in [35] of birth-death processes with killing with certain absorption and $\alpha > 0$ for which

(i) no quasi-stationary distribution exists, and

(ii) an x-invariant quasi-stationary distributions exists if and only if $\gamma < x \leq \alpha$ for some $\gamma > 0$.

So the simple structure that prevails in the setting of pure birth-death processes is lost already in the setting of birth-death processes with killing as soon as we allow infinitely many positive killing rates. But we can state the following.

Theorem 17 Let the birth-death process with killing $\mathcal{X}$ be such that absorption is certain and $\alpha > 0$. If a quasi-stationary distribution for $\mathcal{X}$ exists then $\alpha_0 = \alpha$ and there exists an α-invariant quasi-stationary distribution.

Proof We have seen that a quasi-stationary distribution $\boldsymbol{u} = (u_i,\ i \in S)$ must satisfy $u_i = c\pi_i Q_{i-1}(x)$ for some $x \in (0, \alpha]$ and some constant c. Since, by [35, Equation (3.8)], $0 < Q_i(y) \leq Q_i(x)$ if $x \leq y \leq \alpha$, we must have $\sum_{i \in S} \pi_i Q_{i-1}(\alpha) < \infty$ if a quasi-stationary distribution exists, while, by [51, Theorem 2],

$$\alpha \sum_{i \in S} \pi_i Q_{i-1}(\alpha) = \sum_{i \in S} a_i \pi_i Q_{i-1}(\alpha).$$

So, up to a multiplicative constant, $(\pi_i Q_{i-1}(\alpha),\ i \in S)$ constitutes an α-invariant quasi-stationary distribution, and hence $\alpha_0 = \alpha$ by Corollary 8. $\Box$

Finally turning to limiting conditional distributions for birth-death processes with killing, the next result, generalizing Theorem 15, follows from statements in the proof of [51, Theorem 2].

Theorem 18 Let the birth-death process with killing $\mathcal{X}$ be such that absorption is certain. Then, for all $i \in S$,

$$\lim_{t \to \infty} \mathbb{P}_i(X(t) = j \,|\, T > t) = \frac{\alpha \pi_j Q_{j-1}(\alpha)}{\sum_{k \in S} a_k \pi_k Q_{k-1}(\alpha)}, \quad j \in S, \tag{60}$$

which is to be interpreted as 0 if the sum diverges. If the sum converges and $\alpha > 0$ then these conditional limits constitute the minimal quasi-stationary distribution for $\mathcal{X}$.

6 Computational aspects

We consider here the numerical evaluation of quasi-stationary distributions, discussing a range of methods and giving guidance on how to implement these in MATLAB®, perhaps the most widely used package for scientific computing. MATLAB's numerical linear algebra features are built on LAPACK, a Fortran library developed for high-performance computers, and it is worth pointing out that other interfaces exist, including the open source Scilab[1] and Octave[2], which

[1] http://www.scilab.org/

[2] http://www.gnu.org/software/octave/

are gaining in popularity. LAPACK itself is in the public domain, and available to aficionados from netlib[3].

Suppose that $S = \{1, 2, \ldots, n\}$, so that Q is an $n \times n$ matrix. Restricting our attention to the case where S is irreducible, we seek to evaluate the left eigenvector $\boldsymbol{u} = (u_i,\, i \in S)$ of Q corresponding to the eigenvalue, $-\alpha$, with maximal real part (being simple, real and strictly negative). Once normalized so that $\boldsymbol{u}\mathbf{1}^T = 1$, $\boldsymbol{u}$ is the unique quasi-stationary (and then limiting conditional) distribution. If S is reducible then we would address an eigenvector problem within a restricted set of states (typically, when $-\alpha$ has geometric multiplicity one, we would determine $\boldsymbol{u}$ over states that are accessible from the minimal class $S_{\min}$, and then put $u_j = 0$ whenever j is not accessible from $S_{\min}$).

If the state space is infinite, then we would employ a truncation procedure whereby the infinite Q is approximated by a sequence $\{Q^{(n)}\}$ of irreducible finite square matrices (see for example Gibson and Seneta [62], Seneta [132, 133], Tweedie [142, 145]), in the hope that the corresponding sequence $\{\boldsymbol{u}^{(n)}\}$ of normalized left eigenvectors approximates the desired quasi-stationary distribution $\boldsymbol{u}$. For example, when S is irreducible it is always possible to construct an increasing sequence $\{S^{(n)}\}$ of *irreducible* finite subsets of S with limit S (see Breyer and Hart [19, Lemma 1]). Successive $\boldsymbol{u}^{(n)}$ would be evaluated using the methods described below. Ideally we would want truncations large enough to capture quasi stationarity of the process, and thus in choosing $\{S^{(n)}\}$ we might be led by the results of a simulation study or an analytical approximation. If the state space is multi-dimensional care would be needed to find an appropriate state-space enumeration that facilitates appending states as n increases; Brent's algorithm [18] provides one such method. If $Q^{(n)}$ is simply Q restricted to $S^{(n)}$ (and $\{S^{(n)}\}$ is an increasing sequence of irreducible finite subsets with limit S), then the corresponding sequence $\{\alpha^{(n)}\}$ of decay parameters will converge to α, the decay parameter of S [19, Lemma 2]. However, convergence of the corresponding sequence $\{\boldsymbol{u}^{(n)}\}$ of left eigenvectors is not guaranteed, let alone convergence to $\boldsymbol{u}$; see [134, Section 7.3] and [19, Section 4] for further de-

[3] `http://www.netlib.org/lapack/`

tails. Whilst technical subtleties abound in the most general setting, truncation is known to work under natural conditions including, for example, α-positivity, and is valid for a variety of standard models, including birth-death processes (Kijima and Seneta [85]) and branching processes [82].

Most standard mathematical software permits evaluation of all, or some, of the eigenvalues and/or eigenvectors of a square matrix. MATLAB provides two routines, `eig` and `eigs`. The first is, in principle, suitable for any square matrix, its utility limited by the availability of memory and processing power. The second is suitable for large sparse matrices (matrices populated primarily with zeros). We will explain how both routines are used to evaluate quasi-stationary distributions.

According to Cleve Moler[4] there are 16 different "code paths" for the `eig` function. The one trod in our case would be the QR algorithm preceded by a reduction of Q to Hessenberg form (for details see Golub and van Loan [65]), unless, exceptionally, `eig` identified some special structure that it could exploit. The following sequence of commands will usually suffice if Q is not too large:

```
[V,D]=eig(transpose(Q));
[mu,position]=max(real(diag(D)));                (61)
u=V(:,position); u=u/sum(u);
alpha=-mu;
```

The first step includes, importantly, transposing Q (MATLAB evaluates *right* eigenvectors). The result is a diagonal matrix D of all eigenvalues of Q and a matrix V whose columns are the corresponding eigenvectors ($Q^T V = Q^T D$, equivalently, $V^T Q = DQ$). The second identifies the eigenvalue with maximum real part (which, for our Q, is real) and records its position. The third step evaluates the quasi-stationary distribution by first extracting the relevant eigenvector and then normalizing it. The final step evaluates the decay parameter as the negative of the dominant eigenvalue. We mention here, and for later reference, that were Q to be *conservative* over S (and hence positive recurrent) the

[4]MATLAB News & Notes, Winter 2000.

above algorithm would return the unique *stationary* distribution, namely the unique solution to the system ($\boldsymbol{\pi} Q = \mathbf{0}\,;\ \boldsymbol{\pi}\mathbf{1}^T = 1$), and α would be returned as 0 (or very close to 0). However, this is not how one would evaluate a stationary distribution. Rather, since we are solving a system of linear equations, a standard factorize-and-solve method such as Gaussian elimination should be used; in MATLAB, the matrix right-divide command

```
u=[zeros(1,length(Q)) 1]/[Q ones(length(Q),1)];
```

will achieve this. Better still, we might use the GTH algorithm (Grassmann et al. [67], but see also [135]), a version of Gaussian elimination, which is regarded as the gold standard for Markov chains; its superior properties are detailed in O'Cinneide [111].

The above method assumes, of course, that Q is in the MATLAB workspace. But, setting up Q might not be a trivial matter, particularly when the state space is multi-dimensional. In these cases a bijection $f : S \to S'$, where $S' = \{1, 2, \ldots, |S|\}$, is needed to render Q as a square matrix over S'; the rate of transition from $\boldsymbol{x}$ to $\boldsymbol{y}$ in S is assigned to $q_{f(\boldsymbol{x}),f(\boldsymbol{y})}$. For example, if $X(t)$ in the metapopulation model referred to in Section 1 were constrained by another stochastic variable $Y(t)$, being the number of patches *suitable* for occupancy, then the extant states would form a triangular array $S = \{(x, y) : 1 \le x \le y \le n\}$. An appropriate bijection from S to $S' = \{1, 2, \ldots, \frac{1}{2}n(n+1)\}$ would be $f(x, y) := y + \frac{1}{2}(x-1)(2n-x)$. For more complex state spaces, a hash table might provide a more efficient implementation of f, although there will be some setup costs. Perhaps surprisingly, the inverse map is seldom required; typically we would be estimating quantities such as $\Pr(X(t) \in A | X(t) \in S)$ for $A \subset S$ (summing $u_{f(\boldsymbol{x})}$ over $\boldsymbol{x} \in A$), but even when identifying quantities such as the mode of $\boldsymbol{u}$, a simple search may suffice.

Notice that if, for example, $n = 1000$ in our metapopulation model, Q would have $250,500,250,000$ elements, and thus, stored as dense matrix, would require at least $2,000$ gigabytes of main memory. Yet, with only nearest neighbour transitions, typical of this sort of model, only 3 million (0.0003%) of these entries will be non-zero. For such problems, sparse matrix technology should be used. MATLAB provides the full range of sparse matrix operations. The `eigs` command implements Arnoldi's algorithm, which evaluates (typically a selection of) eigenvalues and eigenvectors of a sparse matrix. The algorithm is iterative. On each iteration, the "basic" Arnoldi method is used. Starting with a "seed vector" $\boldsymbol{v}$, an $m \times m$ upper-Hessenberg matrix H and an $n \times m$ matrix V is constructed in such a way that $V^T Q V = H$, with m fixed to be much smaller than n. The eigenvectors of H are determined by some efficient dense-matrix method and these are used to provide estimates of the extremal eigenvectors of Q. The idea is that if z is an eigenvector of H, then Vz should be close to an eigenvector of Q. Implementations differ in the way $\boldsymbol{v}$ is updated ready for the next iteration. MATLAB's `eigs` implements (through LAPACK) a (random) restart method due to Lehoucq and Sorensen [96]. An alternative restart method, one that is particularly suited to the present problem, is described in Pollett and Stewart [125], but presently not available in MATLAB. For further details, see [65, Chapter 9].

For large problems with sparse transition structure, we must first set up Q as a sparse matrix. The simplest way is to begin with the command `Q=sparse([])`, which initializes Q as an empty sparse matrix (replacing the usual step of setting Q to be the zero matrix: `Q=zeros(n,n)`). Then, we simply enter the non-zero elements as we would normally. (A more complicated method, but one which can markedly reduce execution time, involves setting up vectors of row indices and column indices of non-zero entries and a vector of their values.) Replacing the first step in the earlier procedure by

```
[V,D]=eigs(transpose(Q));
```

will achieve the desired effect, but, as we require the eigenvector corresponding to the eigenvalue with maximum real part, it is significantly more efficient proceed as follows:

```
[u,mu]=eigs(transpose(Q),1,'lr');
u=u/sum(u);                                    (62)
alpha=-mu;
```

(The incantation `[u,mu]=eigs(A,k,'lr')` yields the k eigenvalues of A with largest real part and the corresponding right eigenvectors.) It is quite remarkable that our dense-matrix code can be tweaked so simply.

MATLAB permits us a great deal of control over the way `eigs` is used. For example, the value of the Arnoldi parameter m can be changed from the default $m = 20$. If m is chosen too large or too small, the algorithm will be slow; if too large the time taken to evaluate the eigenvectors of H will be predominant, while if too small the number of outer iterations might be prohibitively large. Another useful feature of `eigs`, which is facilitated by LAPACK's remarkable "reverse communication" interface, is the ability to pass Q to `eigs` as a function (function handle) that evaluates $Q^T x$:

```
[u,mu]=eigs(@Qfun,1,'lr');
```

where

```
function y = Qfun(x)
  .....
end
```

declared elsewhere in our code, effects the operation $Q^T x$ as an efficient elementwise calculation. So, Q *does not need to be stored at all*, and in principle very much larger problems can be tackled. However, evaluation of $Q^T x$ can be time consuming and, because evaluation is frequent, the resulting code can be very slow.

We mention a final approach to evaluating the quasi-stationary distribution $\boldsymbol{u}$, which exploits the return map $\boldsymbol{m} \mapsto \pi^{\boldsymbol{m}}$ (recall that $\pi^{\boldsymbol{m}}$ is the stationary distribution of the process instantaneously returned to S, on departure, according to the measure $\boldsymbol{m}$, that is, $\pi^{\boldsymbol{m}}(Q + \boldsymbol{a}^T\boldsymbol{m}) = \mathbf{0}$). In the present finite state-space setting, the return map is contractive, and thus iteration leads us to $\boldsymbol{u}$. So, we start with an estimate of $\boldsymbol{u}$, which might for example be suggested by an analytical approximation such as a diffusion approximation, and then iterate until the desired accuracy is achieved, at each step using a Gaussian elimination algorithm to evaluate $\pi^{\boldsymbol{m}}$. If MATLAB's matrix right-divide command is used, sparsity or bandedness in the modified Q will be detected automatically and exploited. However, after at most two iterations the modified Q will have n more non-zero entries than Q.

To illustrate the methods presented above, we will first analyse the metapopulation model referred to in Section 1, and then analyse an elaboration which accounts for a dynamic landscape.

Example. Let $X(t)$ be the number of occupied patches in a network consisting of a fixed number of patches n. Each occupied patch becomes empty at rate e and colonization of empty patches occurs at rate c/n for each occupied-unoccupied pair. Thus we have a birth-death process over a finite space $\{0\} \cup S$, where $S = \{1, 2, \dots, n\}$, with birth rates $\lambda_i = (c/n)i(n-i)$ and death rates $\mu_i = ei$. It sometimes called the SIS (susceptible-infectious-susceptible) model, for it is often used to model the number of infectives in a population of fixed size n, with per-capita recovery rate e and per-proximate encounter transmission rate c. It is an example of Feller's [56] *stochastic logistic model*, and one of the earliest stochastic models for the spread of infections that do not confer any long lasting immunity, and where individuals become susceptible again after infection (Weiss and Dishon [150]). It appears, not only in ecology and epidemiology, but also in the propagation of rumours (Bartholomew [14]) and in chemical reaction kinetics (Oppenheim et al. [112]).

Clearly S is irreducible, and so there is a unique quasi-stationary distribution $\boldsymbol{u} = (u_i,\, i \in S)$ and this has a limiting conditional interpretation. Whilst it can be written down in terms of the coefficients (37) and the birth-death polynomials (39), by way of Theorems 14 and 15, neither these quantities, nor indeed the decay parameter α, can be exhibited explicitly. However, the quasi-stationary distribution is easily evaluated using code sequence (61) after setting up Q as follows:

```
Q=zeros(n,n);
i=1; lambda=(c/n)*i*(n-i); mu=e*i;
Q(i,i+1)=lambda; Q(i,i)=-(lambda + mu);
for i=2:n-1
```

```
    lambda=(c/n)*i*(n-i); mu=e*i;
    Q(i,i+1)=lambda; Q(i,i-1)=mu; Q(i,i)=-(lambda + mu);
end
i=n; mu=e*i; Q(i,i-1)=mu; Q(i,i)=-mu;
```

This quasi-stationary distribution for the 20-patch model with $c = 0.1625$ and $e = 0.0325$ is depicted in Figures 2 and 3. Figure 3 also depicts the pseudo-transient distribution when the initial distribution assigns all its mass to state 1, that is,

$$\pi_i = \pi_1 \frac{(n-1)!}{i(n-i)!} \left(\frac{c}{en}\right)^{i-1} \qquad (i = 1, 2, \ldots, n)$$

(see Clancy and Pollett [33], Kryscio and Lefèvre [88], Nåsell [108]) and the Gaussian approximation (for large n, $X(t)/n$ has an approximate normal $\mathrm{N}(1-e/c, e/c)$ distribution; see for example [123]). The remarkable closeness of the pseudo-transient distribution to the quasi-stationary distribution is explained in [13]. Further approximations to quasi-stationary distribution can be found in Nåsell [107, 108, 109], Ovaskainen [114] and Clancy and Mendy [32].

Next we consider an elaboration of this model, alluded to earlier, in which number of patches available for occupancy is a stochastic variable (Ross [129]).

Example. The network consists of a fixed number of patches n, but only $Y(t)$ of these are suitable for occupancy. As before, $X(t)$ is the number of occupied patches, but now $0 \le X(t) \le Y(t)$. The process $((X(t), Y(t))$ is assumed to have non-zero transition rates

$$q((x,y);(x,y+1)) = r(n-y),$$

$$q((x,y);(x,y-1)) = d(y-x),$$

$$q((x,y);(x-1,y-1) = dx,$$

$$q((x,y);(x+1,y)) = (c/n)x(y-x),$$

$$q((x,y);(x-1,y)) = ex.$$

The additional parameters, d and r, are, respectively, the disturbance rate (the per-capita rate at which patches suitable for occupancy become unsuitable) and the recovery rate (the per-capita rate at which unsuitable patches become

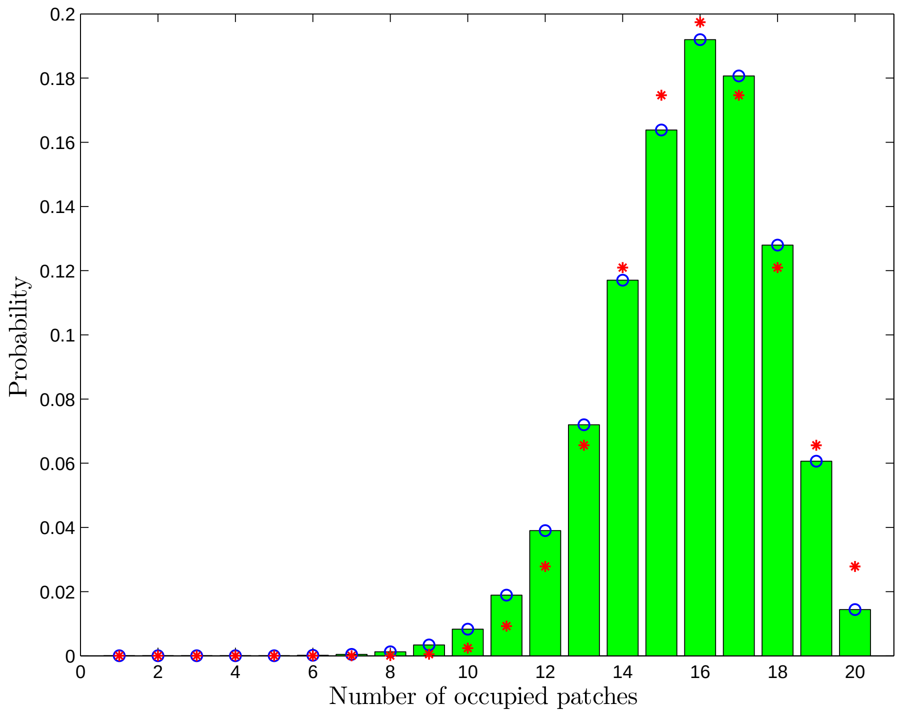

Fig 3. The quasi-stationary distribution (bars), the pseudo-transient distribution when the initial distribution assigns all its mass to state 1 (circles) and the Gaussian approximation (asterisks), for the 20-patch (SIS) metapopulation model with colonization rate $c = 0.1625$ and local extinction rate $e = 0.0325$.

suitable); notice that if an occupied patch is disturbed it also suffers local extinction. The state space consists of a set $S = \{(x, y) : 1 \le x \le y \le n\}$ (irreducible) of transient states that correspond to at least one patch being occupied ("extant" states), and a set $A = \{(0, y) : 0 \le y \le n\}$ (irreducible) in which the process is eventually trapped.

Since S is irreducible and finite, there is a unique quasi-stationary distribution $\boldsymbol{u} = (u_{(x,y)}, (x, y) \in S)$ and this has a limiting conditional interpretation. To evaluate this distribution we would first set up Q rendered as a square matrix over $S' = \{1, 2, \dots, \frac{1}{2}n(n+1)\}$ using the bijection $f(x, y) := y + \frac{1}{2}(x-1)(2n-x)$. For example, for the colonization transition we would have

```
for y=2:n
  for x=1:(y-1)
    i=index([x,y,n]); j=index([x+1,y,n]);
    Q(i,j)=(c/n)*x*(y-x);
  end
end
```

with the bijection defined elsewhere as

```
function i=index(state)
  x=state(1); y=state(2); n=state(3);
  i=x+(y-1)*(2*n-y)/2;
end
```

We could then use code sequence (61), as before. If n is large we would use sparse matrix code, preceding the above with `Q=sparse([])` and then using code sequence (62), or, better, setting up the rows and columns "manually":

```
Q_size=n*(n+1)/2;
Qnz=0;
   .....
for y=2:n
    for x=1:(y-1)
```

```
    Qnz=Qnz+1;
    Q_row(Qnz)=index([x,y,n]);
    Q_col(Qnz)=index([x+1,y,n]);
    Q_val(Qnz)=(c/n)*x*(y-x);
  end
end
   .....
Q=sparse(Q_row,Q_col,Q_val,Q_size,Q_size,Qnz);
```

We performed numerical experiments evaluating the quasi-stationary distribution on a PC equipped with an Intel® Xeon® 6-core 3.33 GHz processor, using parameters $e = 0.1$, $c = 0.6$, $d = 0.1$ and $r = 0.5$. Table 1 compares the execution time of `eig` versus `eigs` (with the default of $m = 20$ for the Arnoldi parameter) for the n-patch metapopulation model with different values of n. It is clear that sparse methods are considerably more efficient when the number of states is large. Table 2 compares the time needed to set up Q as a sparse matrix and the time to evaluate the quasi-stationary distribution using `eigs` ($m = 20$) for the n-patch metapopulation model with various values of n. Listed also is the corresponding size of the state space and the number of non-zero elements of Q $(= n(3n-2))$. It is clear that the execution time is dominated by the transition matrix set-up time. Figure 4 displays the average time, each average taken over 10 runs, needed to evaluate the quasi-stationary distribution of the 100-patch metapopulation model using `eigs` for values of m. Recall that when m is large the time taken to evaluate the eigenvectors of H will be predominant, while if too small the number of outer iterations might be prohibitively large. Notice that MATLAB's default value of $m = 20$ is close to optimal for our problem. Finally, Figure 5 plots the quasi-stationary distribution of the 100-patch metapopulation model. We also estimated the quasi-stationary distribution using *one* iteration of the return map starting from the diffusion approximation of Ross [129, Page 797] (the stationary distribution of the approximating Ornstein-Uhlenbeck process) and obtained an excellent approximation.

n	$\|S\|$	Execution time `eig`	`eigs`
20	400	0.056	0.024
30	900	0.281	0.042
50	2500	2.618	0.094
100	10000	118.294	0.300
150	22500	1120.634	0.702

Table 1. Time (in seconds) to evaluate the quasi-stationary distribution using `eig` and `eigs` (with $m = 20$) in the n-patch metapopulation model for various values of n. Listed also is the corresponding size of the state space.

n	$\|S\|$	nnz(Q)	Execution time Q setup	qsd $\boldsymbol{u}$
20	400	1,160	0.01163	0.02383
30	900	2,640	0.03275	0.04162
50	2500	7,400	0.15736	0.09393
100	10,000	29,800	1.97740	0.30042
150	22,500	67,200	16.01290	0.70218
200	40,000	119,600	59.80960	1.71450
300	90,000	269,400	340.28470	5.72000
500	250,000	749,000	2687.98260	8.58440

Table 2. Time (in seconds) to set up Q as a sparse matrix and the time to evaluate the quasi-stationary distribution using `eigs` (with $m = 20$) in the n-patch metapopulation model for various values of n. Listed also is the corresponding size of the state space and the number of non-zero elements of Q.

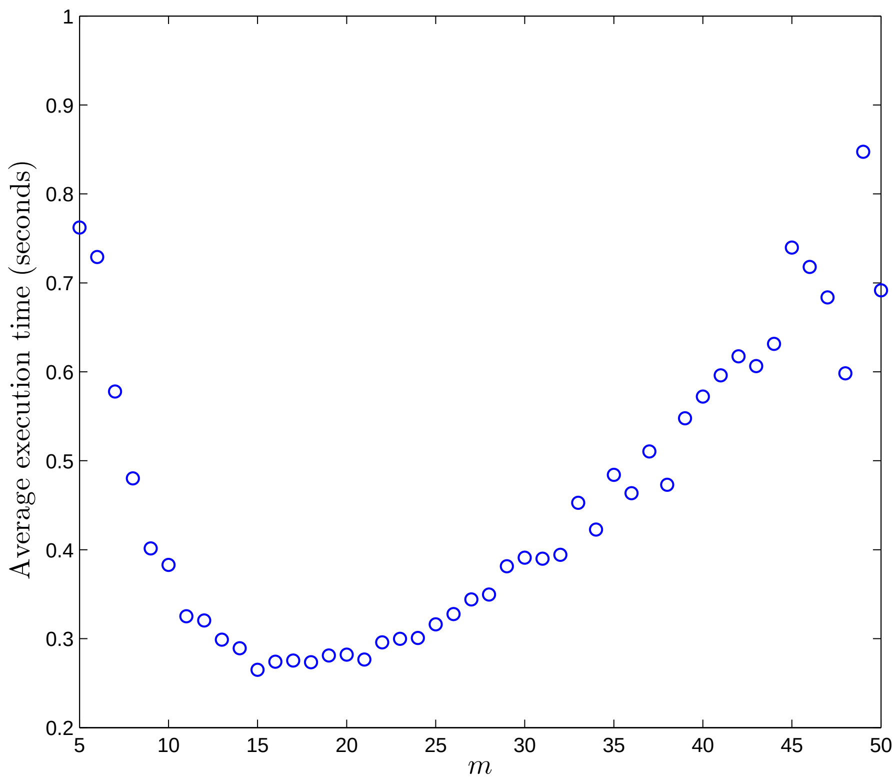

Fig 4. Execution times (each averaged over 10 runs) for evaluating the quasi-stationary distribution of the 100-patch metapopulation model using `eigs` for values of m.

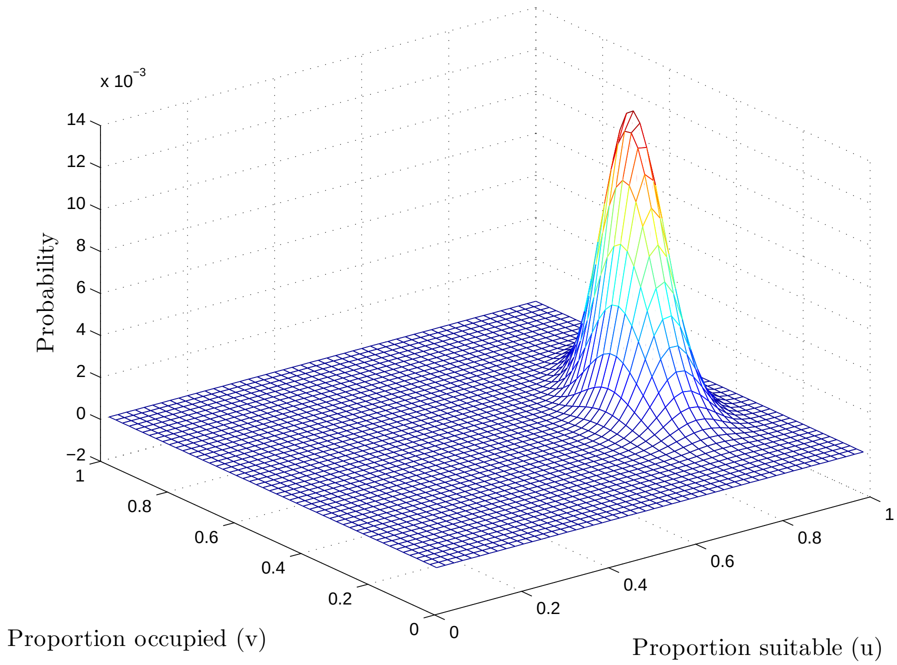

Fig 5. The quasi-stationary distribution of the 100-patch metapopulation model evaluated using `eigs` with $m = 20$.

Acknowledgement

The work of Phil Pollett is supported by the Australian Research Council (Discovery grant DP110101929 and the Centre of Excellence for Mathematics and Statistics of Complex Systems).

References

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

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

[3] Albert, E. (1963). Markov chains and λ-invariant measures. *J. Math. Anal. Appl.* **6**, 404-418.

[4] Arjas, E., Nummelin, E. and Tweedie, R.L. (1980). Semi-Markov processes on a general state space: α-theory and quasistationarity. *J. Austral. Math. Soc. Ser. A* **30**, 187-200.

[5] Anderson, W.J. (1991). *Continuous-time Markov Chains*. New York: Springer.

[6] Artalejo, J.R., Economou, A. and Lopez-Herrero, M.J. (2010). On the number of recovered individuals in the *SIS* and *SIR* stochastic epidemic models. *Math. Biosc.* **228**, 45-55.

[7] Artalejo, J.R., Economou, A. and Lopez-Herrero, M.J. (2010). The maximum number of infected individuals in SIS epidemic models: Computational techniques and quasi-stationary distributions. *J. Comput. Appl. Math.* **233**, 2563-2574.

[8] Artalejo, J.R. and Lopez-Herrero, M.J. (2010). Quasi-stationary and ratio of expectations distributions: A comparative study. *J. Theor. Biology* **266**, 264-274.

[9] Atman, A.P.F. and Dickman, R. (2002). Quasistationary distributions for the Domany-Kinzel stochastic cellular automaton. *Phys. Rev. E (3)*, 66(4):046135, 9.

[10] Barbour, A.D. (1972). The principle of the diffusion of arbitrary constants. *J. Appl. Probab.* **9**, 519-541.

[11] Barbour, A.D. (1974). On a functional central limit theorem for Markov population processes. *Adv. Appl. Probab.* **6**, 21-39.

[12] Barbour, A.D. (1976). Quasi-stationary distributions in Markov population processes. *Adv. Appl. Probab.* **8**, 296-314.

[13] Barbour, A.D. and Pollett, P.K. (2010). Total variation approximation for quasi-equilibrium distributions. *J. Appl. Probab.* **47**, 934-946.

[14] Bartholomew, D.J. (1976). Continuous time diffusion models with random duration of interest. *J. Math. Sociol.* **4**, 187-199.

[15] Bartlett, M.S. (1960). *Stochastic Population Models in Ecology and Epidemiology*. London: Methuen.

[16] Bobrowski, A. (2004). Quasi-stationary distributions of a pair of Markov chains related to time evolution of a DNA locus. *Adv. Appl. Probab.* **36**, 56-77.

[17] Boucherie, R.J. (1997). On the quasi-stationary distribution for queueing networks with defective routing. *J. Austral. Math. Soc. B* **38**, 454-463.

[18] Brent, R.P. (1973). Reducing the retrieval time of scatter storage techniques. *Communications of the ACM* **16**, 105-109.

[19] Breyer, L.A. and Hart, A.G. (2000). Approximations of quasi-stationary distributions for Markov chains. *Math. Computer Modelling* **31**, 69-79.

[20] Buiculescu, M. (1975). On quasi-stationary distributions for multi-type Galton-Watson processes. *J. Appl. Probab.* **12**, 60-68.

[21] Callaert, H. (1974). On the rate of convergence in birth-and-death processes. *Bull. Soc. Math. Belg.* **26**, 173-184.

[22] Cattiaux, P., Collet, P., Lambert, A., Martínez, S., Méléard, S. and San Martín, J. (2009). Quasi-stationary distributions and diffusion models in population dynamics. *Ann. Probab.* **37**, 1926-1969.

[23] Cavender, J.A. (1978). Quasistationary distributions of birth-death processes. *Adv. Appl. Probab.* **10**, 570-586.

[24] Chan, T., Pollett, P.K. and Weinstein, M.C. (2009). Quantitative risk stratification in Markov chains with limiting conditional distributions. *Medical Decision Making* **29**, 532-540.

[25] Chen, A., Li, J, Hou, Z. and Ng, K.W. (2010). Decay properties and quasi-stationary distributions for stopped Markovian bulk-arrival and bulk-service queues. *Queueing Syst.* **66**, 275-311.

[26] Chen, M.-F. (2000). Explicit bounds of the first eigenvalue. *Sci. China Ser. A* (English ed.) **43**, 1051-1059.

[27] Chen, M.-F. (2005). *Eigenvalues, Inequalities, and Ergodic Theory*. Springer, London, 2005.

[28] Cheong, C.K. (1970). Quasi-stationary distributions in semi-Markov processes. *J. Appl. Probab.* **7**, 388-399.

[29] Cheong, C.K. (1970). Correction: Quasi-stationary distributions in semi-Markov processes. *J. Appl. Probab.* **7**, 788.

[30] Cheong, C.K. (1972). Quasi-stationary distributions for the continuous-time Galton-Watson process. *Bull. Soc. Math. Belgique* **24**, 343-350.

[31] Chihara, T.S. (1978). *An Introduction to Orthogonal Polynomials*. New York: Gordon and Breach.

[32] Clancy, D. and Mendy, S.T. (2011). Approximating the quasi-stationary distribution of the SIS model for endemic infection. *Methodol. Comput. Appl. Probab.*, to appear.

[33] Clancy, D. and Pollett, P.K. (2003). A note on quasi-stationary distributions of birth-death processes and the SIS logistic epidemic. *J. Appl. Probab.* **40**, 821-825.

[34] Collet, P., Martínez, S. and Schmitt, B. (1996). Quasi-stationary distribution and Gibbs measure of expanding systems. *Nonlinear Phenomena and Complex Systems* **1**, 205-219.

[35] Coolen-Schrijner, P. and van Doorn, E.A. (2006). Quasi-stationary distributions for birth-death processes with killing. *J. Appl. Math. Stochastic Anal.* **2006**, Article ID 84640, 15 pages.

[36] Daley, D.J. (1969). Quasi-stationary behaviour of a left-continuous random walk. *Ann. Math. Statist.* **40**, 532-539.

[37] Dambrine, S. and Moreau, M. (1981). Note on the stochastic theory of a self-catalytic chemical reaction, I. *Physica A* **106**, 559-573.

[38] Dambrine, S. and Moreau, M. (1981). Note on the stochastic theory of a self-catalytic chemical reaction, II. *Physica A* **106**, 574-588.

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

[40] Darroch, J.N. and Seneta, E. (1967). On quasi-stationary distributions in absorbing continuous-time finite Markov chains. *J. Appl. Probab.* **4**, 192-196.

[41] Day, J.R. Possingham, H.P. (1995). A stochastic metapopulation model with variable patch size and position. *Theoret. Pop. Biol.* **48**, 333-360.

[42] van Doorn, E.A. (1985). Conditions for exponential ergodicity and bounds for the decay parameter of a birth-death process. *Adv. Appl. Probab.* **17**, 514-530.

[43] van Doorn, E.A. (1991). Quasi-stationary distributions and convergence to quasi-stationarity of birth-death processes. *Adv. Appl. Probab.* **23**, 683-700.

[44] van Doorn, E.A. (2003). Birth-death processes and associated polynomials. *J. Comput. Appl. Math.* **153**, 497-506.

[45] van Doorn, E.A. (2003). On associated polynomials and decay rates for birth-death processes. *J. Math. Anal. Appl.* **278**, 500-511.

[46] van Doorn, E.A. (2006). On the α-classification of birth-death and quasi-birth-death processes. *Stochastic Models* **22**, 411-421.

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

[48] van Doorn, E.A. and Pollett, P.K. (2009). Quasi-stationary distributions for reducible absorbing Markov chains in discrete time. *Markov Process. Related Fields* **15**, 191-204.

[49] van Doorn, E.A. and Schrijner, P. (1995). Geometric ergodicity and quasi-stationarity in discrete-time birth-death processes. *J. Austral. Math. Soc. (B)* **37**, 121-144.

[50] van Doorn, E.A. and Schrijner, P. (1995). Ratio limits and limiting conditional distributions for discrete-time birth-death processes. *J. Math. Anal. Appl.* **190**, 263-284.

[51] van Doorn, E.A. and Zeifman, A.I. (2005). Extinction probability in a birth-death process with killing. *J. Appl. Probab.* **42**, 185-198.

[52] Evans, L.S. (1978). An upper bound for the mean of Yaglom's limit. *J. Appl. Probab.* **15**, 199-201.

[53] Evans, S.P. (1990). A quasi-stationary analysis of a virtual path in a B-ISDN network shared by services with very different characteristics. *Comput. Networks ISDN Systems* **20**, 391-399.

[54] Ewens, W.J. (1963). The diffusion equation and a pseudo-distribution in genetics. *J. Roy. Statist. Soc., Ser B* **25**, 405-412.

[55] Ewens, W.J. (1964). The pseudo-transient distribution and its uses in genetics. *J. Appl. Probab.* **1**, 141-156.

[56] Feller, W. (1939). Die Grundlagen der Volterraschen Theorie des Kampfes ums Dasein in wahrscheinlichkeitstheoretischer Behandlung, *Acta Biotheoretica* **5**, 11-40.

[57] Ferrari, P.A., Kesten, H. and Martínez, S. (1996). R-positivity, quasi-stationary distributions and ratio limit theorems for a class of probabilistic automata. *Ann. Appl. Probab.* **6**, 577-616.

[58] Ferrari, P.A., Kesten, H., Martínez, S. and Picco, S. (1995). Existence of quasistationary distributions. A renewal dynamical approach. *Ann. Probab.* **23**, 501-521.

[59] Flaspohler, D.C. (1974). Quasi-stationary distributions for absorbing continuous-time denumerable Markov chains. *Ann. Inst. Statist. Math.* **26**, 351-356.

[60] Flaspohler, D.C. and Holmes, P.T. (1972). Additional quasi-stationary distributions for semi-Markov processes. *J. Appl. Probab.* **9**, 671-676.

[61] Foguel, S.R. (1969). Ratio limit theorems for Markov processes. *Israel J. Math.* **7**, 384-392.

[62] Gibson, D. and Seneta, E. (1987). Augmented truncations of infinite stochastic matrices. *J. Appl. Probab.* **24**, 600-608.

[63] Gosselin, F. (1998). Reconciling theoretical approaches to stochastic patch-occupancy metapopulation models. *Bull. Math. Biol.* **60**, 955-971.

[64] Green, P.J. (1977). Generalizing the Yaglom limit theorems. In *Recent developments in statistics (Proc. European Meeting Statisticians, Grenoble, 1976)*, (pp. 441-444). Amsterdam: North-Holland.

[65] Golub, H.G. and van Loan, C.F. (1996). *Matrix Computations*. (3rd ed.). Baltimore: Johns Hopkins University Press.

[66] Good, P. (1968). The limiting behaviour of transient birth and death processes conditioned on survival. *J. Austral. Math. Soc. Ser. B* **8**, 716-722.

[67] Grassmann, W.K. Taksar, M.I. and Heyman, D.P. (1985). Regenerative analysis and steady state distributions for Markov chains. *Oper. Res.* **33**, 1107-1116.

[68] Gyllenberg, M. and Silvestrov, D.S. (1994). Quasi-stationary distributions of a stochastic metapopulation model. *J. Math. Biol.* **33**, 35-70.

[69] Heathcote, C.R., Seneta, E. and Vere-Jones, D. (1967). A refinement of two theorems in the theory of branching processes. *Teor. Verojatnost. i Primenen.* **12**, 341-346.

[70] Jacka, S.D. and Roberts, G.O. (1995). Weak convergence of conditioned processes on a countable state space. *J. Appl. Probab.* **32**, 902-916.

[71] Kac, M. (1947). Random walk and the theory of Brownian motion. *Amer. Math. Monthly* **54**, 369-391.

[72] Van Kampen, N.G. (1961). A power series expansion of the master equation. *Canad. J. Phys.* **39**, 551-567.

[73] Kalpakam, S. and Shahul Hameed, M.A. (1983). Quasi-stationary distribution of a two-unit warm-standby redundant system. *J. Appl. Probab.* **20**, 429-435.

[74] Kalpakam, S. (1993). On the quasi-stationary distribution of the residual lifetime. *IEEE Trans. Reliab.* **42**, 623-624.

[75] Karlin, S. and McGregor, J.L. (1957). The differential equations of birth-and-death processes, and the Stieltjes moment problem. *Trans. Amer. Math. Soc.* **85**, 489-546.

[76] Karlin, S. and McGregor, J.L. (1957). The classification of birth and death processes. *Trans. Amer. Math. Soc.* **86**, 366-400.

[77] Keilson, J. (1965). A review of transient behaviour in regular diffusion and birth-death processes. Part II. *J. Appl. Probab.* **2**, 405-428.

[78] Kesten, H. (1970). Review of Good (1968). *Math. Rev.* **39**, 410.

[79] Kesten, H. (1995). A ratio limit theorem for (sub)Markov chains on $\{0, 1, 2, \ldots\}$ with bounded jumps. *Adv. Appl. Probab.* **27**, 652-691.

[80] Kijima, M. (1992). On the existence of quasi-stationary distributions in denumerable R-transient Markov chains. *J. Appl. Probab.* **29**, 21-36.

[81] Kijima, M. (1993). Correction: On the existence of quasi-stationary distributions in denumerable R-transient Markov chains. *J. Appl. Probab.* **30**, 496.

[82] Kijima, M. (1993). Quasi-limiting distributions of Markov chains that are skip-free to the left in continuous time. *J. Appl. Probab.* **30**, 509-517.

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

[84] Kijima, M. and Makimoto, N. (1998). Quasi-stationary distributions of Markov chains arising from queueing processes: A survey. In J.G. Shantikumar and U. Sumita (Eds.), *Applied Probability and Stochastic Processes*, pp. 277-311. Boston: Kluwer.

[85] Kijima, M. and Seneta, E. (1991). Some results for quasistationary distributions of birth-death processes. *J. Appl. Probab.* **28**, 503-511.

[86] Kingman, J.F.C. (1963). The exponential decay of Markov transition probabilities. *Proc. London Math. Soc.* **13**, 337-358.

[87] Kingman, J.F.C. and Orey, S. (1964). Ratio limit theorems for Markov chains. *Proc. Amer. Math. Soc.* **15**, 907-910.

[88] Kryscio, R. and Lefèvre, C. (1989). On the extinction of the S-I-S stochastic logistic epidemic. *J. Appl. Probab.* **27**, 685-694.

[89] Kukhtin, V.V., Kuzmenko, N.V. and Shramko, O.V. (1997). Green light as a possible pressing factor for oceanic phytoplankton near the base of the euphotic zone. *J. Theor. Biol.* **188**, 319-322.

[90] Kurtz, T.G. (1970). Solutions of ordinary differential equations as limits of pure jump Markov processes. *J. Appl. Probab.* **7**, 49-58.

[91] Kurtz, T.G. (1971). Limit theorems for sequences of jump Markov processes approximating ordinary differential processes. *J. Appl. Probab.* **8**, 344-356.

[92] Kyprianou, E.K. (1971). On the quasi-stationary distribution of the virtual waiting time in queues with poisson arrivals. *J. Appl. Probab.* **8**, 494-507.

[93] Kyprianou, E.K. (1972). On the quasi-stationary distributions of the GI/M/1 queues. *J. Appl. Probab.* **9**, 117-128.

[94] Kyprianou, E.K. (1972). The quasi-stationary distributions of queues in heavy traffic. *J. Appl. Probab.* **9**, 821-831.

[95] Lambert, A. (2008). Population dynamics and random genealogies. *Stochastic Models* **24**, 45-163.

[96] Lehoucq, R.B. and Sorensen, D.C. (1996). Deflation techniques for an implicitly restarted Arnoldi iteration *SIAM. J. Matrix Anal. & Appl.* **17**, 789-821.

[97] Li, Y.R. and Li, J. (2009). Criteria for Feller transition functions. *J. Math. Anal. Appl.* **359**, 653-665.

[98] Li, W. and Cao, J. (1993). Limiting distributions of the residual lifetimes of several repairable systems. *Microelectron. Reliab.* **33**, 1069-1072.

[99] Li, W. and Cao, J. (1993). The limiting distributions of the residual lifetimes of a Markov repairable system. *Reliab. Eng. System Safety* **41**, 103-105.

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

[101] Mandl, P. (1960). On the asymptotic behaviour of probabilities within groups of states of a homogeneous Markov process. *Časopis Pěst. Mat.* **85**, 448-455 (in Russian).

[102] McNeil, D.R. and Schach, S. (1973). Central limit analogues for Markov population processes. *J. Roy. Statist. Soc. Ser. B* **35**, 1-23.

[103] Miclo, L. (1999). An example of application of discrete Hardy's inequalities. *Markov Process. Related Fields* **5**, 319-330.

[104] Moler, J.A., Plo, F. and San Miguel, M. (2000). Minimal quasi-stationary distribution under null R-recurrence. *Test* **9**, 455-470.

[105] Nair, N.G. and Pollett, P.K. (1993). On the relationship between μ-invariant measures and quasi-stationary distributions for continuous-time Markov chains. *Adv. Appl. Probab.* **25**, 82-102.

[106] Nåsell, I. (1991). On the quasistationary distribution of the Ross Malaria model. *Math. Biosci.* **107**,187-208.

[107] Nåsell, I. (1996). The quasi-stationary distribution of the closed endemic SIS model. *Adv. Appl. Probab.* **28**, 895-932.

[108] Nåsell, I. (1999). On the quasi-stationary distribution of the stochastic logistic epidemic. *Math. Biosci.* **156**, 21-40.

[109] Nåsell, I. (2003). Moment closure and the stochastic logistic model. *Theoret. Pop. Biol.* **63**, 159-168.

[110] Nummelin, E. (1976). Limit theorems for α-recurrent semi-Markov processes. *Adv. Appl. Probab.* **8**, 531-547.

[111] O'Cinneide, C.A. (1993). Entrywise perturbation theory and error analysis for Markov chains. *Numer. Math.* **65**, 109-120.

[112] Oppenheim, I., Schuler, K.K. and Weiss, G.H. (1977). Stochastic theory of nonlinear rate processes with multiple stationary states. *Physica A*, **88**, 191-214.

[113] Orey, S. (1961). Strong ratio limit property. *Bull. Amer. Math. Soc.* **67**, 571-574.

[114] Ovaskainen, O. (2001). The quasistationary distribution of the stochastic logistic model. *J. Appl. Probab.* **38**, 898-907.

[115] Pakes, A.G. (1973). Conditional limit theorems for a left continuous random walk. *J. Appl. Probab.* **10**, 317-329.

[116] Pakes, A.G. (1984). The Galton-Watson process with killing. *Math. Biosci.* **69**, 171-188.

[117] Pakes, A.G. (1995). Quasi-stationary laws for Markov processes: examples of an always proximate absorbing state. *Adv. Appl. Probab.* **27**, 120-145.

[118] Papangelou, F. (1967). Strong ratio limits, R-recurrence and mixing properties of discrete parameter Markov processes. *Z. Wahrsch. Verw. Gebiete* **8**, 259-297.

[119] Pinsky, R.G. (1985). On the convergence of diffusion processes conditioned to remain in a bounded region for a large time to limiting positive recurrent diffusion processes. *Ann. Probab.* **13**, 363-378.

[120] Pollett, P.K. (1986). On the equivalence of μ-invariant measures for the minimal process and its q-matrix. *Stochastic Process. Appl.* **22**, 203-221.

[121] Pollett, P.K. (1988). On the problem of evaluating quasistationary distributions for open reaction schemes. *J. Statist. Phys.* **53**, 1207-1215.

[122] Pollett, P.K. (1989). Modelling quasi-stationary behaviour in metapopulations. *Math. Computers Simulat.* **48**, 393-405.

[123] Pollett, P.K. (2001). Diffusion approximations for ecological models. In F. Ghassemi (Ed.), *Proceedings of the International Congress on Modelling and Simulation*, Vol. 2, Canberra, Australia, 2001, pp. 843-848. Modelling and Simulation Society of Australia and New Zealand.

[124] Pollett, P.K. and Roberts, A.J. (1990) A description of the long-term behaviour of absorbing continuous-time Markov chains using a centre manifold. *Adv. Appl. Probab.* **22**, 111-128.

[125] Pollett, P.K. and Stewart, D.E. (1994). An efficient procedure for computing quasistationary distributions of Markov chains with sparse transition structure. *Adv. Appl. Probab.* **26**, 68-79.

[126] Pollett, P.K. and Vere-Jones, D. (1992). A note on evanescent processes. *Austral. J. Statist.* **34**, 531-536.

[127] Port, S.C. (1965). Ratio limit theorems for Markov chains. *Pacific J. Math.* **15**, 989-1017.

[128] Pruitt, W.E. (1965). Strong ratio limit property for R-recurrent Markov chains. *Proc. Amer. Math. Soc.* **16**, 196-200.

[129] Ross, J.V. (2006). A stochastic metapopulation model accounting for habitat dynamics. *J. Math. Biol.* **52**, 788-806.

[130] Schmidt, E.J.P.G. (1986). An alternative approach to canonical forms of matrices. *Amer. Math. Monthly* **93**, 176-184.

[131] Seneta, E. (1966). Quasi-stationary behaviour in the random walk with continuous time. *Austral. J. Statist.* **8**, 92-98.

[132] Seneta, E. (1967). Finite approximations to infinite non-negative matrices I. *Proc. Camb. Phil. Soc.* **63**, 983-992.

[133] Seneta, E. (1968). Finite approximations to infinite non-negative matrices II: refinements and applications. *Proc. Camb. Phil. Soc.* **64**, 465-470.

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

[135] Seneta, E. (1998). Complementation in stochastic matrices and the GTH algorithm. *SIAM J. Matrix Anal. Appl.* **19**, 556-563.

[136] Seneta, E. and Vere-Jones, D. (1966). On quasi-stationary distributions in discrete-time Markov chains with a denumerable infinity of states. *J. Appl. Probab.* **3**, 403-434.

[137] Seneta, E. and Vere-Jones, D. (1968). On the asymptotic behaviour of subcritical branching processes with continuous state space. *Z. Wahrsch. Verw. Gebiete* **10**, 212-225.

[138] Shohat, J.A. and Tamarkin, J.D. (1963). *The Problem of Moments,* Math. Surveys I. (Rev. ed.). Providence, R.I.: American Mathematical Society.

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

[140] Steinsaltz, D. and Evans, S.N. (2007). Quasistationary distributions for one-dimensional diffusions with killing. *Trans. Amer. Math. Soc.* **359**, 1285-1324.

[141] Stirk, E.R., Lythe, G., van den Berg, H.A., Hurst, G.A.D. and Molina-París, C. (2010). The limiting conditional probability distribution in a stochastic model of T cell repertoire maintenance. *Math. Biosc.* **224**, 74-86.

[142] Tweedie, R.L. (1971). Truncation procedures for non-negative matrices. *J. Appl. Probab.* **8**, 311-320.

[143] Tweedie, R.L. (1974). Some ergodic properties of the Feller minimal process. *Quart. J. Math. Oxford* **25**, 485-495.

[144] Tweedie, R.L. (1974). Quasi-stationary distributions for Markov chains on a general state-space. *J. Appl. Probab.* **11**, 726-741.

[145] Tweedie, R.L. (1998). Truncation approximations of invariant measures for Markov chains. *J. Appl. Probab.* **35**, 517-536.

[146] Vere-Jones, D. (1962). Geometric ergodicity in denumerable Markov chains. *Quart. J. Math. Oxford* **13**, 7-28.

[147] Vere-Jones, D. (1967). Ergodic properties of nonnegative matrices, I. *Pacific J. Math.* **22**, 361-386.

[148] Vere-Jones, D. (1969). Some limit theorems for evanescent processes. *Austral. J. Statist.* **11**, 67-78.

[149] Waugh, W.A.O'N. (1958). Conditioned Markov processes. *Biometrika* **45**, 241-249.

[150] Weiss, G.H. and Dishon, M. (1971). On the asymptotic behaviour of the stochastic and deterministic models of an epidemic. *Math. Biosci.* **11**, 261-265.

[151] Yaglom, A.M. (1947). Certain limit theorems of the theory of branching processes. *Dokl. Acad. Nauk SSSR* **56**, 795-798 (in Russian).

[152] Ziedins, I.B. (1987). Quasi-stationary distributions and one-dimensional circuit-switched networks. *J. Appl. Probab.* **24**, 965-977.