APPROXIMATIONS OF QUASISTATIONARY DISTRIBUTIONS FOR MARKOV CHAINS

L.A. Breyer and A.G. Hart

Department of Mathematics
The University of Queensland
Brisbane, Australia

ABSTRACT

We consider a simple and widely used method for evaluating quasistationary distributions of continuous time Markov chains. The infinite state space is replaced by a large, but finite approximation, which is used to evaluate a candidate distribution.

We give some conditions under which the method works, and describe some important pitfalls.

μ-subinvariant measures; conditioned processes; limiting conditional distributions

AMS 1991 SUBJECT CLASSIFICATION: PRIMARY 60J27
SECONDARY 60J35

1 INTRODUCTION

Various models used in applied probability feature a lifetime τ, after which their behaviour becomes 'uninteresting'. For example, epidemics usually end after a certain (perhaps long) time. Chemical reactions may stop, having exhausted one of the reactants. Market options expire. Endangered species become extinct.

When the model involves a Markov chain, it has proved useful to study the associated family of so-called quasistationary distributions. These probability distributions typically arise in the following generic way. Consider a Markov chain (X_t) on a state space S, together with a transient irreducible class $C \subseteq S$ for which the first exit time τ from C is almost surely finite. What happens to (X_t) after time τ is not of immediate interest; the states outside C are amalgamated into one single absorbing set $\{0\}$.

A *quasistationary distribution* (QSD) is a probability measure $m = (m_i)$ on C related to the process (X_t) by the equation

$$\Pr(X_t = j \mid \tau > t, X_0 \sim m) = m_j,$$

where the notation $X_0 \sim m$ means that X_0 has distribution m. QSDs exist and are unique whenever C is finite (see Darroch and Seneta (1967)). In the infinite case, it is natural to ask whether the class C may be replaced by a large but finite subset $C^{(n)}$, such that the corresponding QSD approximates one sought after on C. Indeed, such a technique is commonly used for the numerical evaluation of QSDs. A major aim of the present paper is to point out that this strategy does not always work.

Complications arise in many ways. The class C may admit zero, one, or a continuum of QSDs, the birth-death process being a case in point (van Doorn (1991)). In addition, the approximate QSDs may not converge as $C^{(n)}$ increases, or may converge to the wrong QSD on C.

We concentrate on continuous time Markov chains. The discrete time results which provided the impetus for this work are described in Seneta (1967). We comment further on the relationship between continuous time and discrete time in Section 7. A complete account of the early truncation literature may be found in Seneta (1981).

2 NOTATION AND ASSUMPTIONS

Let $P(t) = (p_{ij}(t))$ denote the transition probabilities of a continuous time Markov chain (X_t) with countable state space S, that is, $p_{ij}(t) = \Pr(X_t = j \mid X_0 = i)$, $i, j \in S$. The associated q-matrix, given by $q_{ij} = \lim_{t\to 0+}(p_{ij}(t) - \delta_{ij})/t$, is assumed stable: $-q_{ii} < \infty$. The state space is the union of an irreducible class C and a single absorbing state: $S = \{0\} \cup C$. The hitting time of $\{0\}$ (or first exit time from C) is denoted τ.

Henceforth, the transition matrix is assumed minimal (Anderson (1991)). The reason for this will become apparent following Lemma 2.

A quasistationary distribution is an example of a λ-*invariant measure*, that is, a measure (m_i) on C satisfying the equation

$$\sum_{i\in C} m_i p_{ij}(t) = e^{-\lambda t} m_j, \quad j \in C, t \geq 0 \tag{1}$$

for some real number $\lambda \geq 0$. In fact, it has been shown by Nair and Pollett (1993) that quasistationary distributions are characterised as the finite λ-invariant measures on C for P normalised so as to have probability mass 1. In contrast, a positive vector (x_j) is called a λ-*invariant vector* if it satisfies

$$\sum_{j\in C} p_{ij}(t) x_j = e^{-\lambda t} x_i, \quad i \in C.$$

Tweedie (1974) showed that the numbers (m_j) defined by (1) always satisfy

$$\sum_{i\in C} m_i q_{ij} = -\lambda m_j, \quad j \in C \tag{2}$$

in the case where P is the minimal process. Pollett (1986) gave necessary and sufficient conditions for the converse to hold.

In the remainder of this section, we recall some further results that we will need. All these facts may be found in Anderson (1991).

There exists a number λ^* such that the integrals

$$\int_0^\infty e^{\lambda t} p_{ij}(t)\, dt, \quad i, j \in C \tag{3}$$

all converge for $\lambda < \lambda^*$ and diverge for $\lambda > \lambda^*$. It is given by

$$\lambda^* := -\lim_{t\to\infty} t^{-1} \log p_{ii}(t), \quad \text{(independently of } i \in C.)$$

Now suppose that $\lambda = \lambda^*$. If (3) diverges, the process (X_t) is called λ^*-recurrent and there exists an essentially unique measure (m_i) satisfying (1). An essentially unique λ^*-invariant vector (x_i) also exists.

Furthermore, the process is called λ^*-*positive recurrent* if $\sum_{i\in C} m_i x_i < \infty$. In that case, we have the limit

$$\lim_{t\to\infty} \Pr(X_t = j \mid \tau > t, X_0 = i) = m_j / \sum_{k\in C} m_k \tag{4}$$

which defines both a limiting conditional distribution and QSD (Vere-Jones (1969)) when the λ^*-invariant measure (m_j) is finite. In particular, this is true whenever the set C is finite, on account of the Perron-Frobenius theorem (Darroch and Seneta (1967)).

3 APPROXIMATING QUASISTATIONARY DISTRIBUTIONS

Let $(C^{(n)})$ be an increasing sequence of *finite* subsets of C such that

$$\emptyset \subset C^{(1)} \subseteq \cdots \subseteq C = \bigcup_n C^{(n)}. \tag{5}$$

The truncated q-matrix associated with $C^{(n)}$, $\left(q_{ij}^{(n)},\ i,j \in S\right)$, is defined by

$$q_{ij}^{(n)} = \begin{cases} q_{ij}, & \text{if } i,j \in C^{(n)}, \\ 0, & \text{otherwise.} \end{cases}$$

Associated with the matrix $(q_{ij}^{(n)})$ is a unique (and hence minimal) process with transition probabilities $p_{ij}^{(n)}(t) = \Pr(X_t = j, \tau^{(n)} > t \mid X_0 = i)$, $\quad i,j \in C^{(n)}$, where $\tau^{(n)}$ is the first exit time of X from $C^{(n)}$. Since $\lim_{n\to\infty} \uparrow \tau^{(n)} = \tau$, the monotone convergence theorem also implies (see Theorem 2.2.14 of Anderson (1991)) that

$$\lim_{n\to\infty} \uparrow p_{ij}^{(n)}(t) = p_{ij}(t), \quad i,j \in C,\ t \geq 0. \tag{6}$$

Lemma 1 *There exists a sequence $(C^{(n)})$ of finite sets satisfying (5), such that for all n, $C^{(n)}$ is irreducible for $(p_{ij}^{(n)}(t))$.*

Proof: If C is finite, set $C^{(n)} = C$ for all $n \geq 1$. Otherwise, enumerate the state space so that state i has a unique number n_i, set $C^{(1)} = \{a\}$ where $a \in C$, and recursively construct $C^{(n+1)}$ from $C^{(n)}$ as follows. Note that $C^{(1)}$ is trivially an irreducible class of $(p_{aa}^{(1)}(t)) > 0$.

Suppose that $C^{(n)}$ is finite and irreducible with respect to $(p_{ij}^{(n)}(t))$. Now, there exist states $i', j' \in C^{(n)}$ together with states $i'', j'' \in C \setminus C^{(n)}$ such that $q_{i'i''} > 0$ and $q_{j''j'} > 0$, for otherwise C would be reducible with $C^{(n)}$ closed. Choose $b \in C \setminus C^{(n)}$ with $n_b = \min\{n_k : k \in C \setminus C^{(n)}\}$. Then, since C is irreducible, there exist finite sequences $(i_k,\ k = 1, \ldots, l)$ and $(j_k,\ k = 1, \ldots, l')$ in C such that

$$q_{i''i_1} q_{i_1 i_2} \cdots q_{i_l b} > 0 \text{ and } q_{bj_1} q_{j_1 j_2} \cdots q_{j_{l'} j''} > 0.$$

The set $C^{(n+1)}$ is then taken to be

$$C^{(n)} \cup \{b, i', i'', i_1, i_2, \ldots, i_l, j_1, j_2, \ldots, j_{l'}, j'', j'\}.$$

$C^{(n+1)}$ is irreducible and finite.

Thus, the process, (X_t), can move from $C^{(n)}$ to b and back again in a finite time with positive probability by traversing the sequence of states

$$i', i'', i_1, \ldots, i_l, b, j_1, \ldots, j_{l'}, j'', j'.$$

The enumeration (n_i) is needed to ensure that each state in C is accounted for in some $C^{(n)}$. $\square$

Lemma 1 is the direct analogue of Theorem 3 in Seneta (1968). There, Seneta proceeds essentially along the same lines as the proof above, except that no ordering is imposed upon C. Without this, the procedure for constructing the sequence $(C^{(n)})$ does not appear to guarantee $\bigcup_n C^{(n)} = C$. For example, suppose that C has a lattice-like structure, $C = C_1 \times C_2$ where $C_1 = \{1, 2, \ldots\}$ and $C_2 = \{1, 2\}$. If transitions are permitted from (i,j) to $(i, 3-j)$, from (i,j) to $(i \pm 1, j)$ for $i > 0$ and from $(0,j)$ to $(1,j)$, then, despite C being irreducible, it is possible to choose $(C^{(n)})$ to be the sequence given by $C^{(n)} = \{(i,1) \mid i = 1, 2, \ldots, n\}$. Whilst being strictly increasing and irreducible, $C^{(n)}$ only approximates half of C. This problem is easily rectified by the enumeration (n_i) which ensures each state's eventual inclusion in an approximating subclass.

Now set

$$\lambda^{(n)} = -\lim_{t\to\infty} t^{-1}\log p_{ii}^{(n)}(t). \tag{7}$$

From the previous section, we have the limit

$$\lim_{t\to\infty} \Pr(X_t = j \mid \tau^{(n)} > t) = m_j^{(n)}, \quad i,j \in C^{(n)},$$

where the collection $m^{(n)} := (m_j^{(n)},\ j \in C^{(n)})$ satisfies

$$\sum_{i\in C^{(n)}} m_i^{(n)} q_{ij}^{(n)} = -\lambda^{(n)} m_j^{(n)}, \quad j \in C^{(n)}, \qquad \sum_{i\in C^{(n)}} m_i^{(n)} = 1. \tag{8}$$

In essence, the remainder of this paper looks at the problem of what happens when we let n tend to infinity on both sides of (8). Alternatively, we are asking whether two different ways of approximating the set $\{\tau = \infty\}$ give the same result, at least when the limit (4) exists:

$$\lim_{n\to\infty}\lim_{t\to\infty} \Pr(X_t = j \mid \tau^{(n)} > t) \quad \stackrel{?}{=} \quad \lim_{t\to\infty}\lim_{n\to\infty} \Pr(X_t = j \mid \tau^{(n)} > t).$$

We shall see that convergence of the normalised $\lambda^{(n)}$-invariant measures of the truncated process to a λ-invariant probability distribution is not assured. In other words, it is not always possible to arbitrarily swap the order of limits in the above expression.

4 CONVERGENCE

If we are to expect any kind of convergence to occur as $n \to \infty$ in (8), we will first require $\lambda^{(n)}$ to converge.

Lemma 2 $\lambda^* = \lim_{n\to\infty} \downarrow \lambda^{(n)}$.

Proof: By (6), we have

$$\begin{aligned}
\lambda^{(n)} &= -\lim_{t\to\infty} t^{-1}\log p_{ii}^{(n)}(t) \\
&\geq -\lim_{t\to\infty} t^{-1}\log p_{ii}^{(n+1)}(t) = \lambda^{(n+1)} \\
&\geq -\lim_{t\to\infty} t^{-1}\log p_{ii}(t) = \lambda^*
\end{aligned}$$

and hence $\lambda^* \leq \lim_{n\to\infty} \lambda^{(n)}$. On the other hand, the function $t \mapsto -\log p_{ii}^{(n)}(t)$ is subadditive, so that the limit in (7) coincides with the infimum over $t > 0$. Therefore

$$-t^{-1}\log p_{ii}^{(n)}(t) \geq \inf_{t>0}\{-t^{-1}\log p_{ii}^{(n)}(t)\} = \lambda^{(n)} \geq \lim_{k\to\infty} \lambda^{(k)}.$$

Letting $n \to \infty$ on the left implies that $-t^{-1}\log p_{ii}(t) \geq \lim_{k\to\infty}\lambda^{(k)}$, and finally

$$\lambda^* = -\lim_{t\to\infty} t^{-1}\log p_{ii}(t) \geq \lim_{n\to\infty} \lambda^{(n)}.$$

□

Note that the minimality assumption on $(p_{ij}(t))$ is crucial here; otherwise, we can only say that $\lim_{n\to\infty} \downarrow \lambda^{(n)} \geq \lambda^*$.

For the next result, we recall that a non-trivial measure is *λ-subinvariant* if

$$\sum_{i\in C} m_i p_{ij}(t) \leq e^{-\lambda t} m_j, \quad j \in C \tag{9}$$

for some $\lambda \geq 0$. For the minimal process, the condition (9) is equivalent (see Tweedie (1974)) to the q-matrix condition (2) in which the '=' sign is replaced by '$\leq$'.

The following lemma demonstrates the existence of a subsequence of $(m^{(n)})$ whose elements approximate a λ^*-subinvariant measure on C for P.

Lemma 3 *For each $a \in C$, there exists a subsequence (n') such that*

(i) $m'_j := \lim_{n'\to\infty} m_j^{(n')}/m_a^{(n')}$ *$(j \in C)$ is λ^*-subinvariant;*

(ii) either $m_j := \lim_{n'\to\infty} m_j^{(n')}$ *$(j \in C)$ is identically zero, or it is λ^*-subinvariant (and the measure is finite).*

Proof:

(i) Since

$$m_j^{(n)} p_{ja}^{(n)}(t) \le \sum_i m_i^{(n)} p_{ia}^{(n)}(t) = e^{-\lambda^{(n)}t} m_a^{(n)} \le m_a^{(n)}$$

holds for $j, a \in C^{(n)}$, it follows by (6) that for fixed $t > 0$ and all $k \ge 0$, $m_j^{(n+k)}/m_a^{(n+k)} \le 1/p_{ja}^{(n)}(t) < \infty$. Thus, there are bounds, (U_j), such that $0 < u_j^{(k)} := m_j^{(k)}/m_a^{(k)} < U_j < \infty$ for all $j \in C$. By Cantor's diagonal argument, there exists a subsequence (n') of (n) such that the numbers m'_j defined by (i) exist, simultaneously for all $j \in C$. Finally, Fatou's lemma gives

$$\begin{aligned} \sum_i m'_i p_{ij}(t) &= \sum_i \left(\lim_{n'\to\infty} u_i^{(n')} p_{ij}^{(n')}(t) \right) \\ &\le \lim_{n'\to\infty} \sum_i u_i^{(n')} p_{ij}^{(n')}(t) \\ &= \lim_{n'\to\infty} \left(e^{-\lambda^{(n')}t} u_j^{(n')} \right) \\ &= e^{-\lambda^* t} m'_j. \end{aligned}$$

Since $m'_a = 1$, we must have $m'_j \ge e^{\lambda^* t} m'_a p_{aj}(t) > 0$.

(ii) Since $0 < m_j^{(n)} \le 1$ for $n \ge 1, j \in C$, a subsequence (n') can be found such that $m_j := \lim_{n'\to\infty} m_j^{(n')}$ exists simultaneously for all $j \in C$. If the resulting measure (m_j) is not identically zero, Fatou's lemma shows that it satisfies (9), and as above $m_j > 0$ for all j.

□

Lemma 3 provides an alternative proof of the existence of a λ^*-invariant measure to that given by Theorem 2 in Kingman (1963). In principle, we now have the means to approximate ratios of λ^*-subinvariant quantities, e.g.,

$$\lim_{n'\to\infty} \frac{m_j^{(n)}}{m_i^{(n)}} = \lim_{n'\to\infty} \frac{m_j^{(n)}/m_a^{(n)}}{m_i^{(n)}/m_a^{(n)}} = \frac{m'_j}{m'_i}, \qquad i, j \in C,$$

which allows us to approximate a quasistationary distribution of a process up to a constant multiple.

While part (i) of the lemma always works, though it might give an infinite measure, part (ii) seems to be closely connected to the existence of finite λ^*-subinvariant measures. Unfortunately, these do not always exist; this has to do with a second parameter $\lambda_* \le \lambda^*$, studied in Jacka and Roberts (1996), and defined by

$$\lambda_* := -\lim_{t\to\infty} t^{-1} \log \Pr(\tau > t | X_0 = i) \quad \text{independently of } i \in C$$

when the limit exists. This number happens to be the supremum of those λ for which a finite λ-subinvariant measure exists. It follows that a necessary condition for the measure (m'_j) in part (ii) of the lemma to be nonzero is that $\lambda_* = \lambda^*$. The paper (Jacka and Roberts (1996)) has some sufficient conditions which guarantee this, the most important being that the limiting conditional distribution (4) exist.

If, in addition, we are able to determine that the λ^*-subinvariant measure is unique, then we need not be concerned with finding convergent subsequences of $(m^{(n)})$, for then all subsequences, and hence the entire sequence, will converge to the desired limit. The practical consequence of this is that any irreducible truncation $C^{(n)}$, for n large enough, will produce an adequate estimate. A sufficient condition for this is λ^*-recurrence of (X_t).

Theorem 4 *Suppose that $(p_{ij}(t))$ is λ^*-positive recurrent. Then for any $a \in C$,*

$$m'_j := \lim_{n\to\infty} m_j^{(n)}/m_a^{(n)}$$

exists and is the essentially unique λ^*-invariant measure on C for $P(t)$. Moreover, it satisfies

$$\lim_{t\to\infty} \Pr(X_t = j \mid \tau > t, X_0 = i) = m'_j / \sum_{i\in C} m'_i$$

where the righthand side is interpreted as zero if $\sum_{i\in C} m'_i = \infty$.

Proof: By general theory (Anderson (1991)), λ^*-recurrence guarantees that there exists precisely one λ^*-subinvariant measure which is, in fact, λ^*-invariant. It is therefore, up to constant multiples, the one and only limit point (componentwise) of the set of measures $((m_j^{(n)}/m_a^{(n)}),\ n \geq 1)$, by Lemma 3. This proves the existence of the limit, (m'_j), and the second statement is well known (Anderson (1991), Proposition 5.2.10). □

Note that this theorem does *not* imply that $m_j^{(n)} \to m'_j/\sum_{i\in C} m'_i$ as $n \to \infty$. This is ideally what we would like to have happen. However, if $\lim_{n\to\infty} m_j^{(n)}$ exists, it will be a multiple of $m'_j/\sum_{i\in C} m'_i$. More precisely, Fatou's lemma together with (ii) of Lemma 3 allows us to deduce that $m_j^{(n)} \to \alpha \cdot (m'_j/\sum_{i\in C} m'_i)$ where $0 \leq \alpha \leq 1$.

As commonly encountered processes are not always λ^*-positive recurrent, the next result may be more useful in some circumstances.

Theorem 5 *Suppose that $(p_{ij}(t))$ satisfies the Feller-Dynkin condition,*

(FD) $\lim_{i\to\infty} p_{ij}(t) = 0$ *for all $j \in C$, $t > 0$.*

If, for some (and then all) $j \in C$,

$$m_j := \limsup_{n\to\infty} m_j^{(n)} > 0, \tag{10}$$

then $r_j := m_j/\sum_{i\in C} m_i$ is a quasistationary distribution associated with λ^*.

Proof: Take a subsequence (n') such that $\lim_{n'\to\infty} m_j^{(n')} = m_j$. Such a sequence always exists. An application of Fatou's lemma yields

$$\begin{aligned} \lim_{n'\to\infty} \sum_{i\in C} m_i^{(n')} &= \liminf_{n'\to\infty} \sum_{i\in C} m_i^{(n')} \\ &= 1 \\ &\geq \sum_{i\in C} \liminf_{n'\to\infty} m_i^{(n)} \\ &= \sum_{i\in C} \lim_{n'\to\infty} m_i^{(n)} \\ &= \sum_{i\in C} m_i > 0, \end{aligned}$$

whilst Lemma 3 shows that the (subprobability) measure (m_j) is λ^*-subinvariant. Also,

$$\begin{aligned} e^{-\lambda^{(n')}t} m_j^{(n')} &= \sum_i m_i^{(n')} p_{ij}^{(n')}(t) \\ &\leq \sum_i m_i^{(n')} p_{ij}(t) \\ &= \sum_i \left(m_i^{(n')} - m_i\right) p_{ij}(t) + \sum_i m_i p_{ij}(t). \end{aligned}$$

By (FD), the first sum on the right can be made arbitrarily small for large n'. Take a finite set K such that $p_{ij}(t) < \epsilon/4$ whenever $i \notin K$. Then for n' large enough, $|m_i^{(n')} - m_i| < \epsilon/2$ for all $i \in K$, so that

$$\begin{aligned}\left|\sum_{i\in C}\left(m_i^{(n')} - m_i\right)p_{ij}(t)\right| &\leq \sum_{i\in K}\left|m_i^{(n')} - m_i\right|p_{ij}(t) \\ &\qquad + \sum_{i\notin K}\left|m_i^{(n')} - m_i\right|p_{ij}(t) \\ &< \epsilon/2 + 2\cdot\epsilon/4 = \epsilon.\end{aligned}$$

As a result, we find as $n' \to \infty$,

$$e^{-\lambda^* t} m_j \leq \sum_{i\in C} m_i p_{ij}(t),$$

which implies that the measure (m_j) is λ^*-invariant, and (r_j) is a quasistationary distribution. □

Note that the problem of finding a suitable convergent subsequence of $(m^{(n)})$ remains. Also, the caveat discussed immediately following the proof of Theorem 4 also applies to Theorem 5, with n' substituted for n there.

The condition (FD) in the statement of Theorem 5 could be replaced (though we won't prove it here) by the tightness condition,

(T) For each $\epsilon > 0$, there is a finite $K \subset C$ such that, for all n large enough,

$$\lim_{t\to\infty} \Pr(X_t \in K \mid \tau^{(n)} > t) = \sum_{j\in K} m_j^{(n)} \geq 1 - \epsilon,$$

but this seems more difficult to check, unless one has good error bounds on the differences $(m_j^{(n)} - m_j)$, or the behaviour of the process (X_t) is well known. However, tightness does give rise to the situation where a subsequence $(m^{(n')})$ which converges to a proper probability distribution on C exists.

5 EXAMPLE: BIRTH-DEATH PROCESSES

Consider a birth-death process (BDP) on $S = \{0\} \cup \{1, 2, \ldots\}$, with birth rates $\lambda_i > \lambda_0 = 0$ and death rates $\mu_i > \mu_0 = 0$. Suppose that the hitting time of $\{0\}$ is a.s. finite for the minimal process; in other words, we suppose that

$$A := \sum_{k=1}^{\infty} \frac{1}{\lambda_k \pi_k} = \infty,$$

with potential coefficients $\pi_1 = 1$ and $\pi_k = \pi_{k-1}(\lambda_{k-1}/\mu_k)$ if $k > 1$. In this situation, we can take $C^{(n)} = \{1, \ldots, n\}$, and $(q_{ij}^{(n)})$ represents the $n \times n$ north-west truncation of the original q-matrix.

Cavender (1978) showed that the sequence $(m_j^{(n)})$ converges as $n \to \infty$. Kijima and Seneta (1991) refined this work by proving that $m_j^{(n)} \to m_j$ as $n \to \infty$ where $m = (m_j,\ j \in C)$ is the normalised left eigenvector corresponding to the largest real eigenvalue of (q_{ij}) on C. Here, we consider the BDP in light of the preceding results.

In terms of the birth-death polynomials $Q_i(x)$ defined by $Q_0(x) = 0$, $Q_1(x) = 1$, and for $i > 1$,

$$\mu_i Q_{i-1}(x) - (\lambda_i + \mu_i) Q_i(x) + \lambda_i Q_{i+1}(x) = -x Q_i(x),$$

we can write $m_j^{(n)} = \mu_1^{-1}\gamma_n\pi_j Q_j(\gamma_n), \quad (1 \le j \le n)$, where $-\gamma_n$ $(= -\lambda^{(n)})$ is the largest real eigenvalue of $(q_{ij}^{(n)})$. Now Lemma 2 and the continuity in x of the polynomials immediately shows that

$$m_j := \lim_{n\to\infty} m_j^{(n)} = \mu_1^{-1}\gamma\pi_j Q_j(\gamma). \quad (\text{Here, } \gamma = \lambda^*).$$

When $\gamma > 0$, we are in the situation where the measures $(m_j^{(n)})$ satisfy condition (T): since $\sum_{i\in C} m_i = 1$ (van Doorn (1991)), we take K such that $\sum_{i\in K} m_i > 1-\epsilon/2$ and for n large enough, $\sup_{n>N(K)} \max_{j\in K} |m_j^{(n)} - m_j| < \epsilon/2|K|$. Thus

$$\sum_{j\in K} m_j^{(n)} \ge \sum_{j\in K} m_j - \sum_{j\in K} |m_j^{(n)} - m_j| > 1-\epsilon/2 - |K|\cdot\epsilon/2|K| = 1-\epsilon, \quad \text{when } n > N(K).$$

6 EXAMPLE: BRANCHING PROCESSES

Again, let $S = \{0,1,2,\ldots\}$ and consider a subcritical, Markov branching process on S, with offspring law $(p_i,\ i\ge 0)$ such that $p_0 > 0$, $p_1 = 0$ and $\sum_{i>1} p_i > 0$. The q-matrix is given by

$$q_{ij} = \begin{cases} 0, & \text{if } 0 \le j < i-1 \text{ or } i = 0, \\ -\rho i, & \text{if } j = i > 0, \\ \rho i p_{j-i+1}, & \text{if } 0 \le j = i-1 \text{ or } j > i > 0, \end{cases}$$

where $\rho > 0$ is a parameter determining the rate of process activity. The set $C = \{1,2,\ldots\}$ is a transient, irreducible class. On account of the upper triangular form of the q-matrix, the components of a λ-invariant measure, $u(\lambda) = (u_i(\lambda), i \in C)$, form a sequence of polynomials given by the recurrence,

$$\begin{aligned} u_1(\lambda) &\equiv 1 \text{ and} \\ u_{j+1}(\lambda) &= \frac{(\rho j - \lambda)u_j(\lambda) - \rho\sum_{i=1}^{j-1} i p_{j-i+1} u_i(\lambda)}{\rho(j+1)p_0}, \qquad j > 0, \end{aligned}$$

which ensures essential uniqueness (per value of λ). Pakes (1994) has shown that a λ-invariant measure exists for each $\lambda \in (0,\lambda^*]$ and that each such measure is finite. Thus, $u_i(\lambda) > 0$ for all $\lambda \in (0,\lambda^*]$ and $i \ge 1$. Also, by using Theorem 1.5.7 and Theorem 3.3.1 of Anderson (1991), the minimal process can be shown to satisfy (FD). Taking $C^{(n)} = \{1,\ldots,n\}$, this means that $m_j^{(n)}/m_1^{(n)}$ converges to a λ^*-invariant measure for some subsequence (n'). In turn, uniqueness of the λ^*-invariant measure shows that $m_j^{(n')}/m_1^{(n')}$ must converge to $u_j(\lambda^*)$ regardless of the subsequence chosen. Thus,

$$\lim_{n\to\infty} \frac{m_j^{(n)}}{m_1^{(n)}} = u_j(\lambda^*).$$

It remains to check the condition (10). First, note that $m^{(n)} = (m_j^{(n)},\ j = 1,2,\ldots,n)$ given by

$$m_j^{(n)} = \frac{u_j(\lambda^{(n)})}{\sum_{i=1}^n u_i(\lambda^{(n)})}, \quad j \in C$$

is the solution of (8). The characteristic polynomial of $Q^{(n)} = \left(q_{ij},\ i,j \in C^{(n)}\right)$ is given by

$$\mathrm{ch}_{Q^{(n)}}(x) := \det(Q^{(n)} - xI_n) = (-p_0)^n (n+1)! u_{n+1}(-x)$$

where I_n is the $n\times n$ identity matrix. Thus, the roots of $u_{n+1}(-x) = 0$ are the eigenvalues of $Q^{(n)}$. The $n\times n$ matrix, $T^{(n)} = I + (Q^{(n)}/q^{(n)})$, where $q^{(n)} = \max\{-q_{ii}^{(n)} \ :\ i = 1,2,\ldots,n\} = \rho n$, is substochastic and has eigenvalues $\kappa_1^{(n)},\ldots,\kappa_n^{(n)}$, ordered so that

$$0 < |\kappa_1^{(n)}| \le \ldots \le |\kappa_n^{(n)}| < 1.$$

Also,

$$\begin{aligned}
\mathrm{ch}_{T^{(n)}}(\kappa) &:= \det(T^{(n)} - \kappa I_n) \\
&= \det\left(\frac{1}{q^{(n)}}\big(Q^{(n)} + q^{(n)}(1-\kappa)I_n\big)\right) \\
&= \frac{1}{(q^{(n)})^n}\det\big(Q^{(n)} + q^{(n)}(1-\kappa)I_n\big) \\
&= \frac{1}{(q^{(n)})^n}\,\mathrm{ch}_{Q^{(n)}}\big(q^{(n)}(\kappa-1)\big) \\
&= \big(\frac{-p_0}{q^{(n)}}\big)^n (n+1)!u_{n+1}(\lambda)
\end{aligned}$$

where

$$\lambda = q^{(n)}(1-\kappa). \tag{11}$$

Once again, we have a one-to-one correspondence through (11) between the eigenvalues of $T^{(n)}$ and those of $Q^{(n)}$.

From the Perron-Frobenius theorem, $\kappa_n^{(n)}$ is real, simple and, in modulus, the largest eigenvalue of $T^{(n)}$. Therefore, $q^{(n)}(\kappa_n^{(n)} - 1)$ is, in its real part, the largest eigenvalue of $Q^{(n)}$ which ensures that $q^{(n)}(1-\kappa_n^{(n)})$ is the smallest, real root of $u_{n+1}(\lambda) = 0$. Hence, $\lambda^{(n)} = q^{(n)}(1-\kappa_n^{(n)})$.

Next, since $u_j(\lambda^*) > 0$ for all $j \geq 2$ and $u_j(\lambda)$ is continuous, it follows that $u_j(\lambda) \geq 0$ is monotonically decreasing in the interval $[\lambda^*, \lambda^{(j-1)}]$. The same is trivially true of $u_1(\lambda)$ which maintains the constant value 1. By Lemma 2, $\lambda^{(n)} \searrow \lambda^*$ as $n \to \infty$, and so $u_j(\lambda^{(n)})$ is monotonically increasing up to $u_j(\lambda^*)$ as $n \to \infty$ for all $j \geq 1$. Furthermore,

$$\sum_{i=1}^{n} u_i(\lambda^{(n-1)}) \leq \sum_{i=1}^{n+1} u_i(\lambda^{(n)}) \leq \ldots \leq \sum_{i=1}^{\infty} u_i(\lambda^*) < \infty. \tag{12}$$

Therefore,

$$\begin{aligned}
m_1^{(n)} &= \frac{u_1(\lambda^{(n-1)})}{\sum_{i=1}^{n} u_i(\lambda^{(n-1)})} \\
&= \frac{1}{\sum_{i=1}^{n} u_i(\lambda^{(n-1)})} \\
&\geq \frac{1}{\sum_{i=1}^{\infty} u_i(\lambda^*)} > 0.
\end{aligned}$$

It follows that condition (10) is satisfied and the conclusions of Theorem 5 are valid. In other words, $m_j^{(n)} \to \alpha(m_j / \sum_{i\in C} m_i)$ as $n \to \infty$ for some $\alpha \in (0,1]$. We can improve upon this by applying dominated convergence to (12) which shows that $\alpha = 1$ and

$$\lim_{n\to\infty} m_j^{(n)} = \frac{m_j}{\sum_{i\in C} m_i}.$$

Thus, we may use $m^{(n)}$, for large enough n, to approximate the quasistationary distribution of a subcritical branching process. See Kijima (1993) for a more general discussion.

7 SOME FINAL REMARKS

In order to derive the results for the two examples presented, it is necessary to take the structure of the q-matrix into account. Without this, Theorems 4 and 5 lose some of their practical appeal. They do, however, provide insight into the continuous time case. We finish with some comments concerning the discrete time case.

Parallel results to those presented here exist for discrete time Markov chains. The reader is referred to Lemma 3 and Theorem 4 which form the basis for the analogy. Compared to discrete time, matters are more complicated here. Indeed, there is the possibility of multiple transition functions with the same q-matrix, which has no parallel in discrete time. According to Lemma 2, if the transition function $(p_{ij}(t))$ is non-minimal, the sequence of measures $(m^{(n)})$ (suitably normalized) will still only converge to a λ^*-subinvariant measure for the minimal transition function $(f_{ij}(t))$.

We also remark that the methods employed here, being based on ideas of compactness of measures related to weak and vague convergence, generalize very easily to more general state spaces. As an example, we state a continuous state space version of Theorem 5, whose proof proceeds exactly as for Theorem 5, *mutatis mutandis*.

Theorem 6 *Let X be a Markov process on a locally compact metric space C whose transition function $P_t(x,dy)$ maps the space of continuous functions vanishing at infinity into itself (Feller-Dynkin property). Suppose that for some sequence of subsets $C^{(n)} \uparrow C$, the n-th killed process has a quasistationary distribution $m^{(n)}$ with minimal support $\bar{C}^{(n)}$, the closure of $C^{(n)}$. If there exists a positive continuous function f with compact support in C such that*

$$\limsup_{n\to\infty} \int_{C^{(n)}} f dm^{(n)} > 0,$$

then $m^{(n)}$ converges vaguely along a subsequence to a multiple of some quasistationary distribution for P_t.

For weak and vague convergence, see Billingsley (1968).

8 ACKNOWLEDGEMENTS

The authors are grateful to Phil Pollett and Gareth Roberts for many useful discussions, and the referees whose comments helped to improve this paper. The first author also thanks Ruth Williams for pointing out an error in a previous version of this paper. This work was carried out while the authors were Ph.D. students in the Department of Mathematics at the University of Queensland.

References

[1] Anderson, W.J. (1991) *Continuous-Time Markov Chains*, Springer-Verlag.

[2] Billingsley, P. (1968) *Convergence of Probability Measures*, Wiley, New York.

[3] Cavender, J.A. (1978) Quasi-stationary distributions of birth-and-death processes, *Adv. Appl. Prob.* **10**, 570–586.

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

[5] Hart, A.G. (1997) *Quasistationary distributions for continuous-time Markov chains*, Ph.D. Thesis, The University of Queensland.

[6] Jacka, S.D. and Roberts, G.O. (1996) Weak convergence of conditioned processes on a countable state space, *J. Appl. Probab.* **32**, 902–916.

[7] 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.

[8] Kijima, Masaaki and Seneta, E. (1991) Some results for quasi-stationary distributions of birth-death processes. *J. Appl. Probab.*, **28**, 503–511.

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

[10] Pakes, A.G. (1994) On the recognition & structure of probability generating functions. *Research Report*, University of Western Australia.

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

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

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

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

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

[16] Seneta, E. (1981) *Non-Negative Matrices and Markov Chains*, Springer Series in Statistics. Springer-Verlag, New York, Berlin, 2nd edition.

[17] Tweedie, R.L. (1974) Some ergodic properties of the Feller minimal process, *Quart. J. Math. Oxford (2)* **25**, 485–495.

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

[19] Vere-Jones, D. (1969) Some limit theorems for evanescent processes, *Aust. J. Stats.* **11**, 67–78.