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.

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 $C \subseteq S$. The first exit time τ from C must be 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; analogous results in discrete time are described in Seneta (1967), but their extension to continuous time is non-trivial.

2 NOTATION AND ASSUMPTIONS

Let $(p_{ij}(t))$ 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 3.

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 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

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

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

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

which defines a QSD (Vere-Jones (1969)) when the 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 $C^{(n)}$ 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 Thm 2.2.14 of Anderson (1991)) that

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

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

Proof: we sketch the proof. Take $C^{(1)}$ as a singleton, e.g. $C^{(1)} = \{a\}$ for some $a \in C$, which is always irreducible for $(p_{aa}^{(1)}(t))$. By induction, if $C^{(n)}$ is irreducible for $(p_{ij}^{(n)}(t))$ and $b \notin C^{(n)}$, the process (X_t) can go from $C^{(n)}$ to b and back in a finite time, with positive probability, passing through some sequence of states $a_1, \ldots, a_k, b, a_{k+1}, \ldots, a_r$. We obtain $C^{(n+1)}$ by adding these states to $C^{(n)}$. □

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 numbers $(m_j^{(n)} : j \in C^{(n)})$ satisfy

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

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 remark that this certainly holds if $\Pr(\tau = \infty) > 0$, since it is always possible to interchange monotone limits; but this is *not* usually the case, certainly not when considering quasistationary distributions.

One situation for which we *do* in fact have $\Pr(\tau = \infty) = 1$ is when $(p_{ij}(t))$ is positive recurrent ($\lambda^* = 0$) on C. In this case, the preceding remark yields a way of approximating the limiting distribution.

Theorem 2 *Let (X_t) be positive recurrent, with limiting distribution*

$$p_{ij}(t) = \Pr(X_t = j \mid X_0 = i) \to \pi_j, \quad i, j \in C \quad \textit{as } t \to \infty.$$

Then $m_j^{(n)} \to \pi_j$, where $(m_j^{(n)})$ is the unique positive vector satisfying (8).

A different way of estimating (π_j), which currently has the advantage of providing relative error bounds, was developed in Tweedie (1973). Writing $(c_{ij}^{(n)})$ for the cofactor matrix of $(q_{ij}^{(n)})$, he showed that

$$\lim_{n\to\infty} \uparrow \frac{c_{ij}^{(n)}}{c_{jj}^{(n)}} = \frac{\pi_i}{\pi_j} = \lim_{n\to\infty} \downarrow \frac{c_{ii}^{(n)}}{c_{ji}^{(n)}}.$$

4 CONVERGENCE

Lemma 3 $\lambda^* = \lim_{n\to\infty} \downarrow \lambda^{(n)}$.

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 actually equivalent (see Tweedie (1974)) to the q-matrix condition (2) in which the '=' sign is replaced by '$\leq$'.

Lemma 4 *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

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

holds for $j, a \in C^{(n)}$, it follows by (6) that for fixed $t > 0$ and all $k \geq 0$, $m_j^{(n+k)}/m_a^{(n+k)} \leq 1/p_{ja}^{(n)}(t) < \infty$. 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

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

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

(ii) Since $0 < m_j^{(n)} \leq 1$ for $n \geq 1, j \in C$, 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, then as above, Fatou's lemma shows that (9) holds, and hence $m'_j > 0$ for all j.

∎

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 $\lambda_* \leq \lambda^*$, 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.

Theorem 5 *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 satisfies

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

Proof: By general theory (Anderson (1991)), λ^*-recurrence guarantees that there exists precisely one λ^*-subinvariant measure (r_j) say, which is therefore, up to constant multiples, the one and only limit point (componentwise) of the set of measures $\{(m_j^{(n)}) \ : \ n \geq 1\}$, by Lemma 4. This proves the existence of the limit, and the second statement is well known (Anderson (1991), Prop. 5.2.11) with (r_j) in place of (m_j). □

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

Theorem 6 *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 $\lim_{n'\to\infty} m_j^{(n')} = m_j$. Fatou's lemma (see lemma 4) 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

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

As a result, we get 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_i) is λ^*-invariant, and (r_j) is a quasistationary distribution. □

The condition (FD) in the statement of Theorem 6 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,

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

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.

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 $m_j^{(n)} \to m_j$ as $n \to \infty$. 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 $m_j^{(n)} = \mu_1^{-1} \gamma_n \pi_j Q_j(\gamma_n)$, $(1 \leq j \leq n)$, where $-\gamma_n$ $(= -\lambda^{(n)})$ is the smallest eigenvalue of $(q_{ij}^{(n)})$. Now Lemma 3 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

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

6 EXAMPLE: BRANCHING PROCESSES

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

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

and C is an irreducible class. On account of the upper triangular form of the q-matrix, the components of a λ-invariant measure are given by a recurrence, which ensures essential uniqueness (per value of λ). Moreover, the minimal process is well known (Anderson (1991)) to satisfy (FD). Taking $C^{(n)} = \{1, \ldots, n\}$, it remains to check (10) for the approximations to converge. An expanded treatment of this will be discussed in Hart (1997).

Note that if we add the possibility of a catastrophe (see Pakes (1995) for details) so that $q_{i0} \geq \kappa > 0$ say, then we have $\lambda^* - \lambda_* \geq \kappa > 0$, and although a quasistationary distribution exists, the approximations $(m_j^{(n)})$ do *not* converge to it.

7 ACKNOWLEDGEMENTS

The authors are grateful to Phil Pollett and Gareth Roberts for many useful discussions. This work was carried out while they 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] Cavender, J.A. (1978) Quasi-stationary distributions of birth-and-death processes, *Adv. Appl. Prob.* **10**, 570–586.

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

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

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

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

[7] 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.

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

[9] Tweedie, R.L. (1973) The calculation of limit probabilities for denumerable Markov processes from infinitesimal properties, *J. Appl. Prob.* **10**, 84–99.

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

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

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