

COMPUTABLE BOUNDS FOR THE DECAY PARAMETER OF A BIRTH–DEATH PROCESS

DAVID SIRL,^{* **}
HANJUN ZHANG^{* ***} AND
PHIL POLLETT,^{* ****} *The University of Queensland*

Abstract

We present bounds on the decay parameter for absorbing birth–death processes adapted from results of Chen (2000), (2001). We address numerical issues associated with computing these bounds, and assess their accuracy for several models, including the stochastic logistic model, for which estimates of the decay parameter have been obtained previously by Nåsell (2001).

Keywords: Approximation; exponential ergodicity; quasi-stationary distribution

2000 Mathematics Subject Classification: Primary 60J27
Secondary 65C20

1. Introduction

Determination of the decay parameter is critical to the analysis of absorbing countable state Markov processes, and there are several approximations and bounds for the particular case of the birth–death process; see, for example, Bordes and Roehner [2], Zeifman [29], and van Doorn [25], [27]. More accurate bounds have been obtained for specific models. For instance, Nåsell [19] has derived approximations for the expected time to absorption, starting with the quasi-stationary distribution (this being the reciprocal of the decay parameter) for the stochastic logistic model. Whilst Nåsell’s approximations are generally good, their accuracy for given parameter values is unknown. We use results of Chen [8], [9] (see also [10, Chapter 5]) to give upper and lower bounds that differ by less than a factor of four for a general absorbing birth–death process, to develop robust numerical methods for approximating the decay parameter of any finite-state absorbing birth–death process. We apply this to the stochastic logistic model, comparing our results with those of Nåsell [19]. We will see that the bounds frequently provide very good approximations to the decay parameter. We also demonstrate that both the upper and lower bounds are sharp in different senses which we will describe.

2. Preliminaries

Let $(X(t), t \geq 0)$ be a Markov process taking values in a countable state space $S = \{0\} \cup C$, where C is an irreducible class from which the sole absorbing state 0 is accessible. For simplicity we will take C to be either $\{1, 2, \dots, N\}$ or $\{1, 2, \dots\}$ according to whether we require the state

Received 8 December 2006; revision received 13 April 2007.

* Postal address: Department of Mathematics, The University of Queensland, Brisbane, QLD 4072, Australia.

** Email address: dsirl@maths.uq.edu.au

*** Email address: hjz@maths.uq.edu.au

**** Email address: pkp@maths.uq.edu.au

space to be finite or infinite. We denote by $P(t) = (p_{ij}(t), i, j \in S, t > 0)$ the transition function of the process, and by $Q = (q_{ij}, i, j \in S) = P'(0^+)$ its q-matrix of transition rates, assumed to be both stable and conservative. Furthermore, we shall assume that the process is absorbed with probability 1; if absorption occurs with probability less than 1, we may transform the rates to obtain a process which is absorbed with probability 1 and thereby study the original process conditioned on absorption occurring (see Waugh [28]).

We will now define the decay parameter of an absorbing Markov process and briefly discuss its importance, outlining the main previously known results. We then focus on birth–death processes and discuss what is known about the decay parameter in this more analytically tractable case. Following this we shall describe Chen’s results and explore some of their more important implications. It will be convenient to separate the cases of finite and infinite state space. In the infinite state case, we analyse our bounds in the context of the random walk on the nonnegative integers with an absorbing barrier at zero. We then investigate the numerical evaluation of these bounds for finite state processes, in particular for the stochastic logistic model.

The *decay parameter* of an absorbing Markov process is the nonnegative, finite quantity given by

$$\lambda_C := \lim_{t \rightarrow \infty} \left[-\frac{1}{t} \log p_{ij}(t) \right].$$

Kingman [15] proved that the limit exists and is the same for all $i, j \in C$. He also established that there are constants $M_{ij} \in (0, \infty)$ (with $M_{ii} = 1$) such that $p_{ij}(t) \leq M_{ij}e^{-\lambda_C t}$, so that when $\lambda_C > 0$ the transition probabilities decay to zero exponentially fast as t tends to infinity. However, it is important to realise that positivity of the decay parameter is equivalent to exponential ergodicity: recall that a process is called *exponentially ergodic* (with index α) if for all states i and j there are positive constants c_{ij} and α such that its transition function satisfies

$$|p_{ij}(t) - p_{ij}(\infty)| \leq c_{ij}e^{-\alpha t} \tag{1}$$

(for further details see Anderson [1, Section 6.6]). We note that, although the state space here is not irreducible, the limiting distribution is still unique when the class structure is as we have assumed. Although nearly all work on exponential ergodicity is presented in the context of convergence of the transition probabilities to the unique limiting distribution of a positive recurrent process, these results are just as applicable to the convergence of the transition probabilities to the degenerate limiting distribution with all of its mass at the absorbing state.

It is the close link with exponential ergodicity from which the decay parameter derives most of its usefulness. One of the most useful quantities associated with an absorbing process is the probability of absorption having occurred at or before an arbitrary time t . Depending upon what a particular model represents, this may have meaning in terms of, for example, the probability of extinction of an endangered species or the probability that an infection will die out. We can show that, for suitable constants $x_i, i \in C$, we have $p_{i0}(t) \approx x_i e^{-\lambda_C t}$ for $i \in C$ and t large. In addition, if we suppose that the process starts with the quasi-stationary distribution (see below) then it is easy to show that the time to absorption has a negative exponential distribution, with rate parameter λ_C , so that then the expected time to extinction is precisely $1/\lambda_C$. This quantity is of significant value in applications, as frequently it can be assumed that the process of interest has been running for sufficient time that it is reasonable to take the quasi-stationary distribution as the initial distribution.

Additionally, the decay parameter plays an important role in the theory of quasi-stationary distributions, which we now outline briefly.

2.1. Quasi-stationary distributions and λ_C

When describing the long-term behaviour of Markov processes with an absorbing state we cannot use the usual notions of stationary and limiting distributions, for these distributions have mass one at the absorbing state and, thus, provide no information about the process before it evanesces. We need to consider the long-term behaviour of the process conditional on it not having been absorbed. A *quasi-stationary distribution* (QSD) is a probability measure $m = (m_i, i \in C)$ that, for some $\nu > 0$, satisfies

$$\sum_{i \in C} m_i p_{ij}(t) = e^{-\nu t} m_j,$$

for all $j \in C$ and $t > 0$. We can interpret this as meaning that if the Markov process has initial distribution m , then the state probabilities remain in the same proportions as in the initial distribution, but decay to zero like $e^{-\nu t}$. In fact, the state probabilities conditioned on nonabsorption are stationary and given by the distribution m . It is further known that, if it exists, the *limiting conditional distribution* (LCD) $b = (b_i, i \in C)$ given by

$$b_j = \lim_{t \rightarrow \infty} \frac{p_{ij}(t)}{1 - p_{i0}(t)},$$

independently of $i \in C$, is a QSD with $\nu = \lambda_C$. We can see that positivity of λ_C is necessary for the existence of an LCD. Though here we content ourselves with this brief introduction to the role of the decay parameter in the theory of QSDs and LCDs, which is but a scratch on the surface of this extremely rich field of interest in the study of countable state Markov processes. For further details and references see Anderson [1, Chapter 5] or Pollett [20].

2.2. Previous results

If the set C is finite, the process is always exponentially ergodic and the best (i.e. largest) constant α in (1) is effectively the Perron–Frobenius eigenvalue of the restricted q -matrix $Q_C = (q_{ij}, i, j \in C)$ (see Section 5). Although the problem of establishing exponential ergodicity for infinite-state processes was in principle solved by Tweedie [24], his necessary and sufficient conditions are usually impractical to check, and making any progress in this direction is dependent on the few useful sufficient conditions which are available and/or any special structure present in the particular model of interest.

3. Birth–death processes

Recall that a birth–death process (BDP) is a Markov process on (a subset of) the nonnegative integers with the property that direct transitions from any state i are possible only to the states $i - 1$ and $i + 1$; this is equivalent to specifying that the q -matrix has the tridiagonal form

$$q_{ij} = \begin{cases} \lambda_i & \text{if } j = i + 1, \\ \mu_i & \text{if } j = i - 1, \\ -(\lambda_i + \mu_i) & \text{if } j = i, \\ 0 & \text{otherwise.} \end{cases}$$

Here $(\lambda_i, i \geq 0)$ and $(\mu_i, i \geq 1)$ are respectively the birth and the death rates, which we assume are all strictly positive, except $\lambda_0 = 0$, in order to ensure that the state space has the desired class structure $S = \{0\} \cup C$. As usual we define the potential coefficients $\pi = (\pi_i, i \in C)$ by $\pi_1 = 1$ and

$$\pi_k = \frac{\lambda_1 \lambda_2 \cdots \lambda_{k-1}}{\mu_2 \mu_3 \cdots \mu_k}, \quad k \geq 2.$$

We recall here that the condition of absorption with probability 1 (for infinite-state BDPs) is equivalent to the divergence of the series

$$\mathcal{A} = \sum_{n=1}^{\infty} \frac{1}{\lambda_n \pi_n}.$$

We also see appreciable simplification of the theory of QSDs when we restrict our attention to birth–death processes. In fact a near complete picture of the existence of QSDs in BDPs is contained in the following result (an agglomeration of Theorems 3.2 and 4.1 of van Doorn [26]), for which we define

$$\mathcal{D} = \sum_{n=1}^{\infty} \frac{1}{\mu_n \pi_n} \sum_{m=n}^{\infty} \pi_m.$$

Theorem 1. (Van Doorn [26].) *For an absorbing birth–death process on $\{0\} \cup C$ which is absorbed with probability 1,*

- (a) *if $\mathcal{D} < \infty$ then there is a unique QSD corresponding to $v = \lambda_C > 0$;*
- (b) *if $\mathcal{D} = \infty$ then either*
 - (i) *$\lambda_C = 0$ and there are no QSDs, or*
 - (ii) *$\lambda_C > 0$ and there is a one-parameter family of QSDs indexed by $v \in (0, \lambda_C]$.*

We can see from this theorem that, for birth–death processes, $\lambda_C > 0$ is not only necessary but sufficient for the existence of a QSD; this dates back to Good [12]. We also observe, in light of the fact that $\mathcal{D} < \infty$ is equivalent to strong ergodicity (Zhang *et al.* [30]), the implication $\mathcal{D} < \infty$ implies that $\lambda_C > 0$ in the above result is simply restating the fact that strong ergodicity entails exponential ergodicity; it had earlier been shown by Tweedie [24] that the convergence of Reuter’s well-known S series (see [21, Section 8.4]), which is equivalent to the convergence of \mathcal{D} , implies exponential ergodicity. For further results on strong ergodicity, see Mao [16] and the references therein.

The search for practical conditions that ensure exponential ergodicity for birth–death processes has been long; indeed we would expect there to be no closed form expression for the decay parameter of a BDP in terms of the birth and death rates, (λ_i) and (μ_i) , respectively. However, there are many results giving computable bounds for a particular BDP, or class of BDPs. Bounds on the best constant of exponential ergodicity (that is, the largest α for which (1) holds) for BDPs have been obtained by Bordes and Roehner [2] (involving, and with conditions on, Reuter’s [21] series R , S , and T), Zeifman [29] (for processes satisfying $\sup_i q_i < \infty$), and van Doorn [25], [27] who considered BDPs with no additional structure. In particular, Theorem 3.3 of [27] gives three representations for the decay parameter; these representations are as (attained) suprema and infima over sequence spaces, for example

$$\lambda_C = \max_{v \in V} \left\{ \inf_{n \geq 1} \left\{ \lambda_n + \mu_n - \frac{\lambda_{n-1} \mu_n}{v_n} - v_{n+1} \right\} \right\},$$

where V is the family of all positive sequences $v = (v_i, i \in C)$. Whilst such representations are not in general amenable to explicit evaluation, judicious choice of the sequence v can provide a good (in this case lower) bound for the decay parameter.

In addition, the decay parameter has even been explicitly evaluated for some models, including the random walk on the nonnegative integers (Seneta [23]), the so-called ‘linear BDP’ (see, for example, Anderson [1, pp. 165–166]), ‘asymptotically symmetric quadratic BDPs’ (see Roehner and Valent [22]), and several queueing models (see, for example, Kartashov [13]). Also, there are some necessary and some sufficient conditions for the prevalence of exponential ergodicity (positivity of the decay parameter). Van Doorn [26, Theorem 3.2(ii)] (or our Theorem 1) is one such result; it is also known that $\inf_{i \in C} q_i = 0$ implies $\lambda_C = 0$ (this follows immediately from $\lambda_C \leq \inf_{i \in C} q_i$, one of the first properties of the decay parameter which Kingman [15] discovered). Callaert and Keilson [5] and Tweedie [24] (already mentioned above) contain the most general such results. Theorem 5.3 of van Doorn [25] determined the exponential ergodicity or otherwise of a wide range of BDPs encountered in practice, and Kijima [14, Theorems 3.2 and 3.3] provided some results concerning the comparison of the decay parameters of two birth–death processes.

The results we have discussed, although in principle able to give significant information about the decay parameter of a birth–death process, are far from complete. Since the early 1990s, however, a series of papers by Mu-Fa Chen has appeared in the Chinese literature which addresses the problem of evaluating, or finding bounds for, the so-called *spectral gap* of various operators associated with not only countable-state Markov processes, but also a number of classes of differential operators with probabilistic applications. The main results with respect to birth–death processes are contained in the papers [6], [7], [8], [9] (see also Chen’s recent monograph [10]) but many of these results in fact apply to all countable-state Markov processes, or at least to those which are reversible. These results appear not to be widely known and indeed to the best of our knowledge have not previously been cited outside the Chinese literature.

In so far as these results are concerned with birth–death processes; in 1991, Chen [6] established the connection between the decay parameter, L^2 -exponential convergence, and the L^2 -spectral gap and established the validity of several previously known bounds under weaker assumptions than before. He then went on to use this representation to obtain variational formulae for the decay parameter, and from these obtained several bounds for the decay parameter, and then some sufficient, and some necessary and sufficient conditions for positivity of the decay parameter. Nearly all previously known results are encompassed by Chen’s work. Central to these results are the two operators defined by

$$I_i(f) = \frac{\sum_{j=i+1}^{\infty} \pi_j f_j}{\mu_{i+1} \pi_{i+1} (f_{i+1} - f_i)}, \quad i \geq 1,$$

and

$$H_i(f) = \frac{1}{f_i} \sum_{j=1}^i \frac{1}{\mu_j \pi_j} \sum_{m=j}^{\infty} \pi_m f_m, \quad i \geq 1,$$

which in general act on positive sequences $f = (f_i, i \in C)$ (though f must be π -integrable for $I_i(f)$ or $H_i(f)$ to be finite). Further restricting the class of sequences f upon which these operators act leads to Chen’s variational formulae (see [9, Section 2]). We now summarise Chen’s main results as applicable to our consideration of absorbing birth–death processes.

Theorem 2. Suppose that a birth–death process on $\{0\} \cup C$ is absorbed with probability 1 (i.e. $\mathcal{A} = \infty$) and write $\phi_i = \sum_{j=1}^i (\mu_j \pi_j)^{-1}$.

(a) (Chen [8].) If we define

$$R_n = \sum_{j=1}^n \frac{1}{\mu_j \pi_j} \sum_{j=n}^{\infty} \pi_j, \quad n \geq 1,$$

and put $R = \sup_{n \geq 1} R_n$, then $(4R)^{-1} \leq \lambda_C \leq R^{-1}$.

(b) (Chen [9].)

(i) Lower approximating sequence: put

$$f_i^{(1)} = \sqrt{\phi_i}, \quad f_i^{(n)} = f_i^{(n-1)} \Pi_i(f^{(n-1)}), \quad \delta_n'' = \sup_{i \geq 1} \frac{f_i^{(n+1)}}{f_i^{(n)}}.$$

Then $(\delta_n''^{-1}, n \geq 1)$ is increasing and $\delta_1''^{-1} \leq \dots \leq \delta_\infty''^{-1} \leq \lambda_C$.

(ii) Upper approximating sequence: put

$$f_i^{(1,k)} = \phi_{i \wedge k}, \quad f_i^{(n,k)} = f_i^{(n-1,k)} \Pi_i(f^{(n-1,k)}), \quad \delta_n' = \sup_{k \geq 1} \inf_{i \geq 1} \frac{f_i^{(n+1,k)}}{f_i^{(n,k)}}.$$

Then $(\delta_n'^{-1}, n \geq 1)$ is decreasing and $R^{-1} \geq \delta_1'^{-1} \geq \dots \geq \delta_\infty'^{-1} \geq \lambda_C$.

Analogous results hold if Π is replaced everywhere by I ; if we denote these approximating sequences by $\tilde{\delta}_n''$ and $\tilde{\delta}_n'$, respectively, then $\tilde{\delta}_n''^{-1} \leq \delta_n''^{-1} \leq \tilde{\delta}_{n+1}''^{-1}$ and $\tilde{\delta}_n' \geq \delta_n' \geq \tilde{\delta}_{n+1}'$.

Immediately obvious from Theorem 2(a) is the following necessary and sufficient condition for positivity of the decay parameter in terms of the transition rates of the process.

Corollary 1. We have $\lambda_C > 0$ if and only if $R < \infty$.

Remarks. Chen [8] also provides another upper bound δ^{-1} for λ_C , which is better (i.e. less) than R^{-1} , defined by

$$\delta = 2 \sup_{n \geq 1} \sum_{j=1}^n Q'_j v_j^{(n)},$$

where

$$Q'_j = \left(\sum_{i=1}^{j-1} (\mu_i \pi_i)^{-1} + (2\mu_j \pi_j)^{-1} \right) \sum_{i=j}^{\infty} \pi_i \quad \text{and} \quad v_j^{(n)} = \frac{(\mu_j \pi_j)^{-1}}{\sum_{k=1}^n (\mu_k \pi_k)^{-1}}.$$

Though this upper bound is better than R^{-1} , its more complicated form renders it more difficult to deal with analytically, and we shall see in Proposition 1 that it is in fact equal to $\delta_1'^{-1}$. We also note that the main use of Theorem 2(a) is to establish the positivity of the decay parameter; if we seek to approximate its value the better (but slightly more complicated) bounds given by the approximating sequences in Theorem 2(b) should be used.

We now proceed to establish the equality of two of the upper bounds we have described.

Proposition 1. *The bounds $\delta_1^{\prime-1}$ (defined in Theorem 2(b)(ii)) and δ^{-1} (defined in the above remarks) are equal.*

Proof. We firstly show that the quantity

$$\alpha_i^{(k)} = \frac{1}{\phi_{i \wedge k}} \sum_{j=1}^i (\mu_j \pi_j)^{-1} \sum_{m=j}^{\infty} \pi_m \phi_{m \wedge k}$$

achieves its infimum over $i \geq 1$ when $i = k$. To this end we note that it is not hard to show that $\alpha_i^{(k)}$ is decreasing for $i = 1, 2, \dots, k$, and trivial to see that it is increasing for $i \geq k$; thus establishing that $\arg \inf_{i \geq 1} \alpha_i^{(k)} = k$.

We now recall that

$$\delta_1' = \sup_{k \geq 1} \inf_{i \geq 1} \alpha_i^{(k)}$$

(see part (2) of Theorem 2.2 in [9]). Upon applying the above result and then interchanging the order of the summation this simplifies to

$$\sup_{k \geq 1} \frac{1}{\phi_k} \sum_{m=1}^{\infty} \pi_m \phi_{m \wedge k}^2,$$

an expression for δ which can be easily inferred from the proof of Theorem 3.5 of [8].

We shall see later that this result enables us to compute δ_1' for finite-state birth–death processes in time which is $O(N)$, as opposed to $O(N^2)$ without it.

4. Infinite state space

We can easily see that Theorem 2(a) encompasses two partial results which have been mentioned previously. Firstly, the fact that $\sum_{i \in \mathcal{C}} \pi_i = \infty$ implies $\lambda_{\mathcal{C}} = 0$ is immediately obvious from Corollary 1 because $\sum_{i \in \mathcal{C}} \pi_i = \infty$ implies that $R_n = \infty$ for all n . In addition, it is not difficult to show that $R_n < \mathcal{D}$ for all n and, hence, $\mathcal{D} < \infty$ implies that $\sup_n R_n < \infty$ and therefore $\lambda_{\mathcal{C}} > 0$ (in accordance with Theorem 1(a)).

The bounds given in Theorem 2(a) are a huge improvement on the previously available bounds; the error is strictly limited to a factor of four. Most significant is the universality of this factor across all birth–death processes; the accuracy of previous estimates of the decay parameter is both extremely variable and highly dependent on the particular birth–death process being studied. Having upper and lower bounds which are a constant multiple of each other ensures the triviality of Corollary 1 as a consequence of Theorem 2.

We now present an example for which we can explicitly evaluate some of the bounds and show that the upper bound for $\lambda_{\mathcal{C}}$ in Theorem 2(a) is sharp, in the sense that 1 is the smallest constant that can multiply R^{-1} , and the better lower bound $\delta_1^{\prime-1}$ is sharp, in the sense that it is sometimes equal to $\lambda_{\mathcal{C}}$ (in particular for the example we present). The example we use is the random walk on the nonnegative integers with an absorbing barrier at zero. This is a birth–death process with $\lambda_i = \lambda$ and $\mu_i = \mu$ for $i \geq 1$, and $\lambda_0 = 0$. In order that almost all sample paths are absorbed at zero, we must insist that $\lambda \leq \mu$. Seneta [23] showed that the decay parameter for this process is $\lambda_{\mathcal{C}} = (\sqrt{\lambda} - \sqrt{\mu})^2$. Clearly $\lambda_{\mathcal{C}} = 0$ if $\lambda = \mu$ (this also

follows from Corollary 1 as $\pi_i = 1$ for all i), so we restrict attention to the case in which $\lambda < \mu$. It is a relatively simple matter to evaluate R_n . We find that

$$R_n = \frac{\mu}{(\mu - \lambda)^2} \left(1 - \left(\frac{\lambda}{\mu} \right)^n \right),$$

which is (strictly) increasing, and so $R = \sup_{n \geq 1} R_n = \lim_{n \rightarrow \infty} R_n = \mu/(\mu - \lambda)^2$. Theorem 2(a) therefore provides the bounds

$$\frac{(\mu - \lambda)^2}{4\mu} \leq \lambda_C \leq \frac{(\mu - \lambda)^2}{\mu}.$$

Now, if we fix μ and let λ approach zero then both λ_C and the upper bound, R^{-1} , approach μ , which is the decay parameter of each state $i \geq 1$ of the pure death process obtained in the limit. In this way we can, given any $\varepsilon > 0$, find a BDP for which $R^{-1} - \lambda_C < \varepsilon$. For this example we can also evaluate the better upper bound, $\delta_1'^{-1}$. By firstly writing $\sum_{m=1}^{\infty} \pi_m \phi_m^2 \wedge k = \sum_{m=1}^k \pi_m \phi_m^2 + \phi_k^2 \sum_{m=k+1}^{\infty} \pi_m$, we find that

$$\delta_1' = \sup_{k \geq 1} \frac{\mu + \lambda}{(\mu - \lambda)^2} - \frac{2k}{(\mu - \lambda)((\mu/\lambda)^k - 1)}.$$

It is clear that the latter term of this expression decreases monotonically to zero, so

$$\lambda_C \leq \delta_1'^{-1} = \frac{(\mu - \lambda)^2}{\mu + \lambda},$$

which, it is plain to see, is strictly less than the previous upper bound of $(\mu - \lambda)^2/\mu$. We can also make analytical progress with the better lower bound, $\delta_1''^{-1}$. We can easily show that

$$\delta_{1,i}'' = \frac{1}{\sqrt{\phi_i}} \sum_{m=1}^{i-1} \pi_m \phi_m^{3/2} + \sqrt{\phi_i} \sum_{m=i}^{\infty} \pi_m \sqrt{\phi_m},$$

from which it follows that $\delta_{1,i}'' \leq (\sqrt{\lambda} - \sqrt{\mu})^{-2}$, and so $\lambda_C \geq \delta_{1,i}''^{-1} \geq (\sqrt{\lambda} - \sqrt{\mu})^2$. Therefore, we have shown that the lower bound is, in this instance, sharp, which is a clear improvement on the original lower bound of $(4R)^{-1}$.

In general it is not easy to explicitly evaluate R , δ_n' , or δ_n'' , so the bounds for the decay parameter given by Theorem 2 will only rarely yield a closed form expression. However, by using monotonicity properties and/or bounds for R_n it will frequently be possible to determine at least, the finiteness or otherwise of $R = \sup_n R_n$ and, hence, whether the decay parameter is zero or positive, which – at least in processes with infinite state space – is the most important conclusion to be made.

Even if no such analytical handle can be found, we can resort to numerical calculation of the R_n s and infer from a plot of R_n against n the behaviour of R_n as n tends to infinity and, thus, whether or not $\sup_n R_n$ can be expected to be finite. From this same plot one can also read off $n^* = \arg \sup_{n \geq 1} R_n$ and, by numerically calculating R_{n^*} , obtain (at least formal) bounds on λ_C . It is easy to show that the upper bound calculated in this way remains an upper bound for the birth–death process we are considering. If, however, not only $\sup_{n \geq 1} R_n$ but also $\sum_{i=n}^{\infty} \pi_i$ does not yield a closed form expression then we may resort to

approximating $\sum_{i=n}^{\infty} \pi_i$ by $\sum_{i=n}^M \pi_i$ for some large M ; this is the same as truncating the state space of the process to $\{0\} \cup \{1, 2, \dots, M\}$. It is again easy to show that the upper bound obtained in this manner is still a valid upper bound for the decay parameter of the original process and also that it is a worse upper bound than that obtained by considering the first M values of R_n . See also Breyer and Hart [3], which deals extensively with finite truncations.

5. Finite state space

The chief usefulness of Theorem 2 when C is an infinite set is in the necessary and sufficient condition of Corollary 1 for $\lambda_C > 0$. This always holds when C is finite, again following from Theorem 2, because R_n is nonzero for only finitely many n and, hence, $R = \sup_n R_n$ is trivially finite. However, the bounds for λ_C offer the potential for analytical expressions that determine the order of the decay parameter and indeed describe the dependence of the decay parameter on model parameters. In general the only way to estimate λ_C is numerically: $-\lambda_C$ is the eigenvalue of the restricted q -matrix Q_C with largest real part (being real and negative) [11]. For some specific models there are estimates of λ_C . For example, Näsell [17], [18] provided estimates for λ_C in the stochastic logistic model (which we explore numerically below). We also remark that C being a finite set ensures that absorption at zero always occurs with probability 1.

Our main result, now stated for *finite-state* birth–death process, is as follows.

Theorem 3. *For a birth–death process on $\{0\} \cup \{1, 2, \dots, N\}$,*

(a) *if we define*

$$R_n = \sum_{j=1}^n \frac{1}{\mu_j \pi_j} \sum_{j=n}^N \pi_j, \quad n = 1, \dots, N,$$

and put $R = \max_{1 \leq n \leq N} R_n$, then $(4R)^{-1} \leq \lambda_C \leq R^{-1}$, and in addition

(b) *if we put*

$$\delta''_{1,i} = \frac{1}{\sqrt{\phi_i}} \sum_{j=1}^i (\mu_j \pi_j)^{-1} \sum_{k=j}^N \pi_k \sqrt{\phi_k},$$

and

$$\delta'_{1,k} = \frac{1}{\phi_k} \sum_{m=1}^N \pi_m \phi_{m \wedge k}^2,$$

then $\delta''_1 = \max_{1 \leq i \leq N} \delta''_{1,i}$ and $\delta'_1 = \max_{1 \leq k \leq N} \delta'_{1,k}$ satisfy $(4R)^{-1} \leq \delta''_1 \leq \lambda_C \leq \delta'_1 \leq R^{-1}$.

Remarks. We have simplified the statement of the above theorem by describing only the bounds which we will numerically evaluate later in this section. All four of the approximating sequences given in Theorem 2(b) have obvious analogues in finite-state models. In addition we have simplified the expression for $\delta'_{1,k}$ using Proposition 1.

Whilst we now lose the significance of being able to determine positivity of λ_C (this always being true), we can compute all of the finite number of R_n values (at least numerically) and therefore find the bounds exactly (to numerical precision), and of course the same applies to the better bounds. We compare the results of this procedure applied to the stochastic logistic model with the results of Näsell [17], [18], who developed methods of approximating λ_C specifically for this model.

The *stochastic logistic model* (SLM), or SIS epidemic model, is the birth–death process on $S = \{0\} \cup \{1, 2, \dots, N\}$ with birth and death rates respectively given by

$$\lambda_i = \frac{\lambda i}{N}(N - i) \quad \text{and} \quad \mu_i = \mu i, \quad i \in S.$$

This process is perhaps the simplest stochastic model of a population which incorporates retardation of population growth at higher population sizes (due to increased competition for resources, for example). It has also found application as a model for the transmission of infectious diseases which confer no long-term immunity and in the study of metapopulations – populations which inhabit discrete patches of habitat.

5.1. Numerical implementation

In evaluating the bounds given in Theorem 3, we must be careful to ensure that numerical errors are kept to a minimum; the errors of particular relevance here are those of numerical overflow and/or underflow. Even if the value we seek to calculate is of moderate magnitude, some of the intermediate quantities may be extreme, that is, too large or too close to zero to be handled within numerical precision. Typically such values would be rounded to infinity or zero, either of these occurring obviously renders any subsequent computations meaningless. For example, for the SLM with $(N, \lambda, \mu) = (500, 10, 9)$ we have $(\mu_{500}\pi_{500})^{-1} \approx 10^{196}$ even though the decay parameter in this instance is only about 0.05.

In the following paragraphs we briefly describe the techniques we use for avoiding such overflow or underflow in the computation of the two better bounds. We have also developed similar methods for calculating R , but these are superfluous as the other bounds are better – the significance of R for finite-state BDPs is its relative simplicity and the resultant possibility that it might yield explicit formulae describing the dependence of the decay parameter on the parameters of a given model (in a manner similar to Näsell [17, Section 8] for the SLM). We therefore include these bounds in our plots below so that we may see how they perform. We also stress that the expressions given in the rest of this subsection were developed with the SLM in mind: though they may very well (and appear to usually) work for other BDPs this is by no means guaranteed.

We firstly look at the better lower bound, $\delta''_{1,i}$, we find that it is useful to write $\delta''_{1,i}$ recursively as follows: $\delta''_{1,0} = 0$ and

$$\delta''_{1,i} = \delta''_{1,i-1} \sqrt{\rho_{i-1}} \sqrt{\frac{g_{i-1} + 1}{g_i + 1}} + \sum_{k=i}^N \frac{1}{\mu_k} \sqrt{\rho_i \cdots \rho_{k-1}} \sqrt{\frac{g_{i-1} + 1}{g_i + 1}}, \tag{2}$$

where $\rho_i = \lambda_i/\mu_i$ and $g_i = \sum_{j=1}^{i-1} (\mu_j \pi_j)^{-1} / (\mu_i \pi_i)^{-1}$, $i = 1, \dots, N$, is precalculated using $g_1 = 0$ and $g_{i+1} = \rho_i(1 + g_i)$, again to avoid overflow or underflow problems. Also, by observing that the sum over k in (2), call it s_i , can be written as $s_i = \sigma_i / \sqrt{g_i + 1}$, where $\sigma_N = \sqrt{g_N + 1}/N$ and

$$\sigma_i = \sigma_{i+1} \sqrt{\rho_i} + \mu_i^{-1} \sqrt{g_i + 1}, \quad i = N - 1, \dots, 1,$$

we reduce the computation time for $\delta''_{1,i}$ from $O(N^2)$ to $O(N)$.

In order to compute the better upper bound, δ'_1 , without introducing overflow or underflow problems, we again proceed by deriving the following recurrence: writing

$$\delta'_{1,k} = \frac{\sum_{m=1}^N \pi_m \phi_{m \wedge (k-1)}^2 + (\mu_k \pi_k)^{-1} ((\mu_k \pi_k)^{-1} + 2\phi_{k-1}) \sum_{m=k}^N \pi_m}{\phi_{k-1} + (\mu_k \pi_k)^{-1}},$$

we find that

$$\delta'_{1,k} = \frac{\delta'_{1,k-1} + h_k(g_k^{-1} + 2)}{1 + g_k^{-1}},$$

where g_k is as in the previous paragraph and $h_k = (\mu_k \pi_k)^{-1} \sum_{m=k}^N \pi_m$ is calculated as $h_N = \mu_N^{-1}$ and $h_i = h_{i+1} \rho_i + \mu_i^{-1}$, $i = N - 1, \dots, 1$. Again, this yields an algorithm with execution times linear in N .

The above methods for calculating both δ'_1 and δ''_1 can be copied virtually straight into any programming environment (MATLAB[®], C[®], C++[®], etc.): calculate, in order, (ρ_i) , then (g_i) , (σ_i) and $(\delta''_{1,i})$, or (ρ_i) , then (g_i) , (h_i) and $(\delta'_{1,k})$.

We do not deal with further terms in the upper and lower approximating sequences, $(\delta_n'^{-1})$ and $(\delta_n''^{-1})$, respectively, on account of it being much more difficult to deal with the overflow or underflow issues which arise; we have not been able to satisfactorily resolve these difficulties. We also point out that the calculation of the upper bounds, $(\delta_n'^{-1})$, for $n > 1$ will take time which is quadratic in N ; for even though we can show that $\arg \inf_{i \geq 1} II_i(f^{(n,k)}) = k$ for all $n, k \geq 1$, $f^{(n,k)}$ depends on the whole sequence $f^{(n-1,k)}$, not just the k th component.

5.2. Other numerical approaches

We will compare the numerical methods obtained from the bounds of Theorem 3 (as described in Section 5.1) with two other methods of finding λ_C for the SLM. The first of these is an essentially exact method which relies on the fact that, for a finite-state absorbing Markov process, $-\lambda_C$ is the eigenvalue of Q_C with maximum real part.

We implement this approach using MATLAB's `eigs` routine to find the eigenvalue closest to zero. It yields the exact (to numerical precision) value of λ_C , but runs into trouble when the eigenvalue it seeks is of order similar to or smaller than machine epsilon. For these reasons, we take the decay parameter computed in this way to be the true value to which the approximations considered are compared. We emphasise that this method is applicable to all finite-state absorbing Markov processes, not just birth–death processes.

The other methods we consider are those developed by Nåsell in a series of results aiming to approximate (amongst other things) the expected time to extinction of the SLM starting from the quasi-stationary distribution. As noted in Section 2, this quantity is exactly $1/\lambda_C$, so Nåsell's results can also be used to approximate the decay parameter of the SLM. Nåsell developed these results in a series of papers culminating in [19] (this paper described such results for the Verhulst logistic model which is a generalisation of the SLM). Nåsell's methods are based on the identity $\lambda_C = \mu_1 m_1$ relating the decay parameter to the elements of the QSD, $m = (m_i, i \in C)$, corresponding to states from which direct transition to the absorbing state is possible. Nåsell approximates the QSD by the stationary distribution of a closely related process and then uses this identity to approximate λ_C ; see [19, Section 2] for further details.

Nåsell's approximations are delineated into three cases: where λ/μ is markedly less than one, approximately equal to one, and markedly greater than one. This is necessary because the approximation of the QSD used in each of these three situations is different. Many of Nåsell's results are not asymptotically exact, but nonetheless provide quite good estimates, particularly for large values of N . Their chief drawback is the lack of error bounds, there being no indication how accurate the approximation is, but they have the advantage of being simple and fast to evaluate.

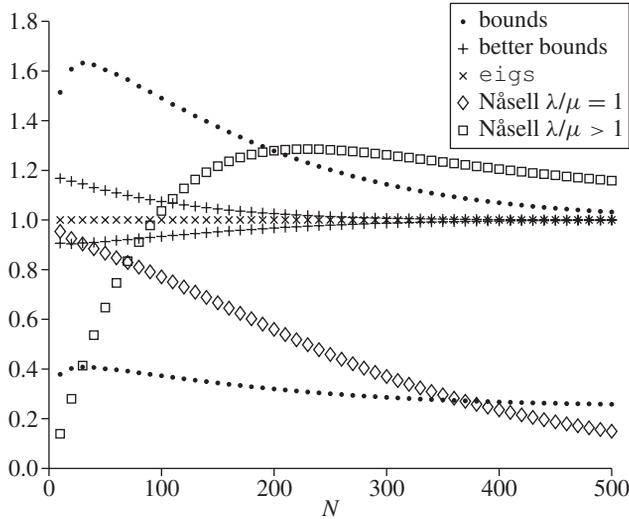


FIGURE 1: Behaviour of the relative bounds and approximations of λ_C for the SLM with varying N , where $\lambda = 1.2$, $\mu = 1$, and $\rho > 1$.

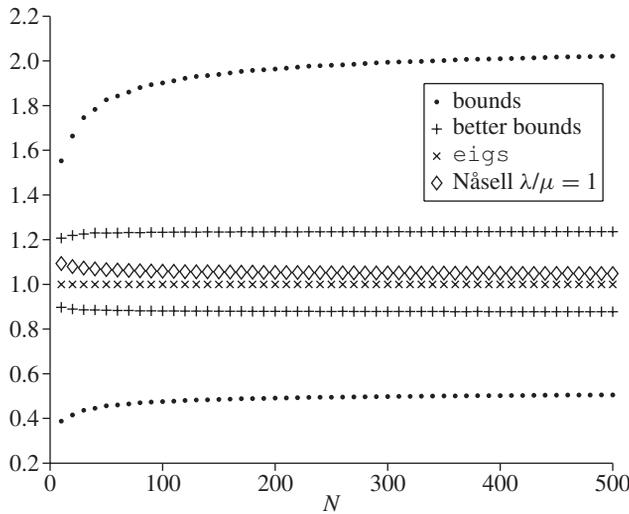


FIGURE 2: Behaviour of the relative bounds and approximations of λ_C for the SLM with varying N , where $\lambda = 1$, $\mu = 1$, and $\rho = 1$.

5.3. Numerical results

We now compare the two approximation methods, namely Chen’s bounds and Nåsell’s approximations, to the true value of the decay parameter for the SLM (remembering the shortcomings of MATLAB’s `eigs` routine when λ_C is too close to zero). We pay particular attention to the more interesting situation where $\lambda \geq \mu$, when extinction occurs over moderate to long timescales (see Nåsell [17], [18], or [19] for further discussion).

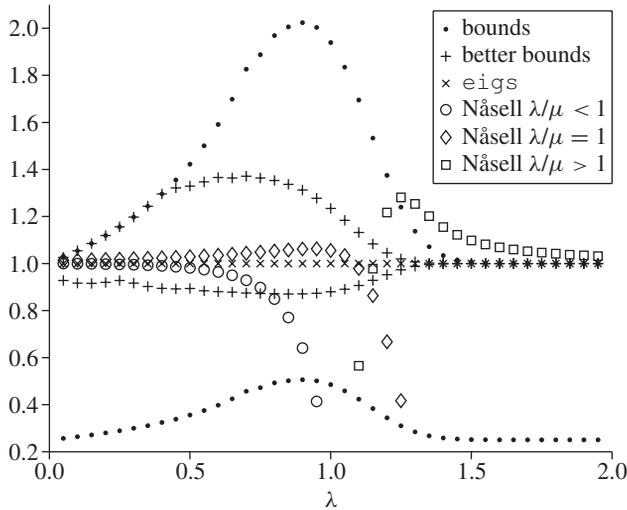


FIGURE 3: Behaviour of the relative bounds and approximations of λ_C for the SLM with varying $\rho (= \lambda)$, where $N = 150$ and $\mu = 1$.

Because of the vastly different orders of magnitude the decay parameter can take, we will look at the estimates relative to the true value calculated by the eigenproblem solver. Perhaps the first thing we observe from our numerical experimentation is that the behaviour of the bounds, and approximations relative to the true value of λ_C , depends on λ and μ almost exclusively through their ratio $\rho = \lambda/\mu$. For this reason we fix $\mu = 1$ for our analysis.

Figures 1 and 2 show the relative bounds and approximations of λ_C for the SLM with varying N and $(\lambda, \mu) = (1.2, 1)$ and $(1, 1)$, respectively. Figure 3 shows the behaviour when N is fixed and $\lambda (= \rho)$ varies.

We observe that when ρ is near one, the two upper bounds are quite different, with the better upper bound approximating λ_C reasonably well. When ρ is not near one the two upper bounds differ only marginally and as ρ deviates further away from one, these upper bounds get closer to the true value of λ_C (quite quickly for ρ increasing and relatively slowly for ρ decreasing). This qualitative behaviour is the same for different values of N , but the larger the value of N the more quickly these changes occur. In particular we note that as ρ increases above one the better upper bound very quickly becomes a very good approximation for λ_C . As ρ decreases below one the upper bounds become closer and then, rather slowly, become good approximations to λ_C . We also observe that for $\rho > 1$ the better upper bound is a very good approximation to λ_C , and that as N increases it very quickly becomes an excellent approximation. The better lower bound can be seen to be quite a good approximation of the decay parameter for all values of ρ and N , but again particularly when either ρ or N are large. There are some values of N where Nåsell's approximations are better but their error, and importantly, the sign of the error, are unknown. Even at their worst, the better bounds are still reasonable approximations, as can clearly be seen in Figures 4 and 5. Note that for large parts of both contour plots (where λ and N are large) λ_C is so small that `eigs` becomes unreliable. The boundary of this area is marked by an additional dashed 'contour' on the plots.

In the less interesting cases when the SLM typically sees absorption within a relatively short period of time ($\rho < 1$), Nåsell's approximations are better than our bounds. However, it should be noted that in this parameter region Nåsell's $\rho = 1$ approximation always

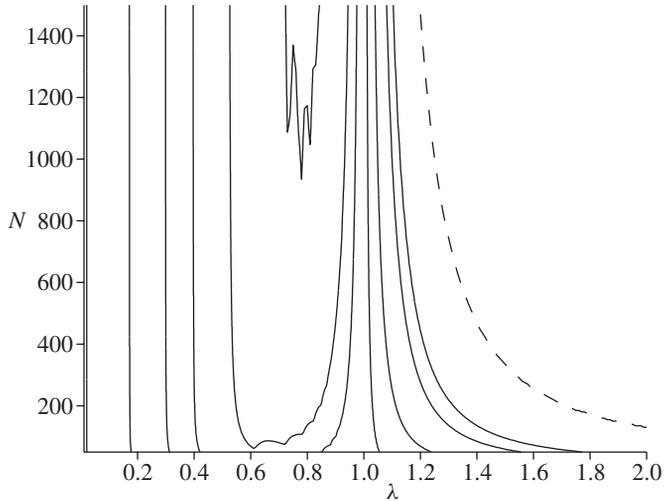


FIGURE 4: Contour plot of the ratio of the upper bound, $\delta_1'^{-1}$, and λ_C for various N and λ with $\mu = 1$. Contours are at 1.4, 1.35, 1.3, 1.2, 1.1, 1.01, and 1.001, with the innermost contour being at 1.4. The dashed ‘contour’ marks where λ_C becomes so small that *eigs* is unreliable.

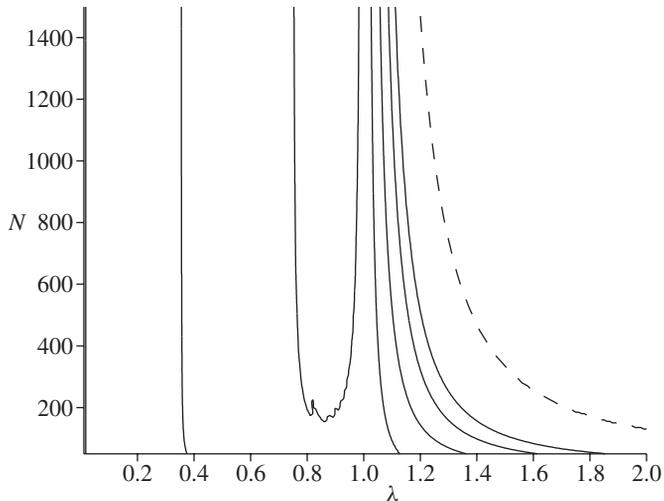


FIGURE 5: Contour plot of the ratio of the lower bound, $\delta_1''^{-1}$, and λ_C for various N and λ with $\mu = 1$. Contours are at 0.87, 0.9, 0.95, 0.99, and 0.999, with the innermost contour being at 0.87. The dashed ‘contour’ marks where λ_C becomes so small that *eig*s is unreliable.

overestimates λ_C , whereas his $\rho < 1$ approximation, although usually closer to the true value, underestimates it.

In different situations it may be desirable to have an estimate which we are certain either overestimates or underestimates λ_C . In more theoretical work regarding the existence of QSDs a lower bound is more important. However, if we are, for example, modelling the abundance of a threatened species, an overestimate of the decay parameter is much more desirable; as an

underestimate of λ_C leads to an overestimate of the viability of the population concerned, an error with potentially grave consequences for the threatened species. We have seen that Chen's bounds are frequently very good approximations for the decay parameter and, although other approximations may be more accurate, these bounds have the advantage of the certainty as to whether we have overestimated or underestimated the decay parameter, which is often an important consideration.

Acknowledgements

The support of the Australian Research Council Centre of Excellence for Mathematics and Statistics of Complex Systems is gratefully acknowledged. We also thank the referee for refocusing our attention on the overall significance of Chen's work, in particular reference [9], and for providing the calculation of the upper bound, $\delta_1''^{-1}$, in Section 4.

References

- [1] ANDERSON, W. J. (1991). *Continuous-Time Markov Chains: An Applications Oriented Approach*. Springer, New York.
- [2] BORDES, G. AND ROEHNER, B. (1983). Application of Stieltjes theory for S -fractions to birth and death processes. *Adv. Appl. Prob.* **15**, 507–530.
- [3] BREYER, L. A. AND HART, A. G. (2000). Approximations of quasi-stationary distributions for Markov chains. *Math. Comput. Modelling* **31**, 69–79.
- [4] CALLAERT, H. AND KEILSON, J. (1973). On exponential ergodicity and spectral structure for birth–death processes. I. *Stoch. Process. Appl.* **1**, 187–216.
- [5] CALLAERT, H. AND KEILSON, J. (1973). On exponential ergodicity and spectral structure for birth–death processes. II. *Stoch. Process. Appl.* **1**, 217–235.
- [6] CHEN, M.-F. (1991). Exponential L^2 -convergence and L^2 -spectral gap for Markov processes. *Acta. Math. Sinica New Ser.* **7**, 19–37.
- [7] CHEN, M.-F. (1996). Estimation of spectral gap for Markov chains. *Acta. Math. Sinica New Ser.* **12**, 337–360.
- [8] CHEN, M.-F. (2000). Explicit bounds of the first eigenvalue. *Sci. China Ser. A* **43**, 1051–1059.
- [9] CHEN, M.-F. (2001). Variational formulas and approximation theorems for the first eigenvalue in dimension one. *Sci. China Ser. A* **44**, 409–418.
- [10] CHEN, M.-F. (2005). *Eigenvalues, Inequalities, and Ergodic Theory*. Springer, London.
- [11] DARROCH, J. N. AND SENETA, E. (1967). On quasi-stationary distributions in absorbing continuous-time finite Markov chains. *J. Appl. Prob.* **4**, 192–196.
- [12] GOOD, P. (1968). The limiting behavior of transient birth and death processes conditioned on survival. *J. Austral. Math. Soc.* **8**, 716–722.
- [13] KARTASHOV, N. V. (2000). Calculation of the spectral ergodicity exponent for the birth and death process. *Ukrainian Math. J.* **52**, 1018–1028.
- [14] KIJIMA, M. (1992). Evaluation of the decay parameter for some specialized birth–death processes. *J. Appl. Prob.* **29**, 781–791.
- [15] KINGMAN, J. F. C. (1963). The exponential decay of Markov transition probabilities. *Proc. London Math. Soc.* (3) **13**, 337–358.
- [16] MAO, Y.-H. (2006). Convergence rates in strong ergodicity for Markov processes. *Stoch. Process. Appl.* **116**, 1964–1976.
- [17] NÅSELL, I. (1996). The quasi-stationary distribution of the closed endemic SIS model. *Adv. Appl. Prob.* **28**, 895–932.
- [18] NÅSELL, I. (1999). On the quasi-stationary distribution of the stochastic logistic epidemic. *Math. Biosci.* **156**, 21–40.
- [19] NÅSELL, I. (2001). Extinction and quasi-stationarity in the Verhulst logistic model. *J. Theoret. Biol.* **211**, 11–27.
- [20] POLLETT, P. K. (1995). The determination of quasistationary distributions directly from the transition rates of an absorbing Markov chain. *Math. Comput. Modelling* **22**, 279–287.
- [21] REUTER, G. E. H. (1957). Denumerable Markov processes and the associated contraction semigroups on l . *Acta Math.* **97**, 1–46.
- [22] ROEHNER, B. AND VALENT, G. (1982). Solving the birth and death processes with quadratic asymptotically symmetric transition rates. *SIAM J. Appl. Math.* **42**, 1020–1046.

- [23] SENETA, E. (1966). Quasi-stationary behaviour in the random walk with continuous time. *Austral. J. Statist.* **8**, 92–98.
- [24] TWEEDIE, R. L. (1981). Criteria for ergodicity, exponential ergodicity and strong ergodicity of Markov processes. *J. Appl. Prob.* **18**, 122–130.
- [25] VAN DOORN, E. A. (1985). Conditions for exponential ergodicity and bounds for the decay parameter of a birth–death process. *Adv. Appl. Prob.* **17**, 514–530.
- [26] VAN DOORN, E. A. (1991). Quasi-stationary distributions and convergence to quasi-stationarity of birth–death processes. *Adv. Appl. Prob.* **23**, 683–700.
- [27] VAN DOORN, E. A. (2002). Representations for the rate of convergence of birth–death processes. *Theory Prob. Math. Statist.* **65**, 37–43.
- [28] WAUGH, W. A. O’N. (1958). Conditioned Markov processes. *Biometrika* **45**, 241–249.
- [29] ZEĪFMAN, A. I. (1991). Some estimates of the rate of convergence for birth and death processes. *J. Appl. Prob.* **28**, 268–277.
- [30] ZHANG, H., CHEN, A., LIN, X. AND HOU, Z. (2001). Strong ergodicity of monotone transition functions. *Statist. Prob. Lett.* **55**, 63–69.