Bounds and asymptotics for the rate of convergence of birth-death processes

Erik A. van Doorn[a], Alexander I. Zeifman[b] and Tatyana L. Panfilova[c]

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

[b] Vologda State Pedagogical University
Institute of Informatics Problems RAS
and VSCC CEMI RAS
S. Orlova 6, Vologda, Russia
E-mail: zai@uni-vologda.ac.ru

[c] Vologda State Pedagogical University
S. Orlova 6, Vologda, Russia

September 22, 2008

Abstract. We survey a method initiated by one of us in the 1990's for finding bounds and representations for the rate of convergence of a birth-death process. We also present new results obtained by this method for some specific birth-death processes related to mean-field models and to the $M/M/N/N+R$ service system. The new findings pertain to the asymptotic behaviour of the rate of convergence as the number of states tends to infinity.

2000 Mathematics Subject Classification: Primary 60J80, Secondary 60J27

1   Introduction

The problem of relating the rate of convergence to stationarity of an ergodic birth-death process to the transition rates of the process – the birth and death rates – has been tackled in the literature by a variety of techniques. This range of solution methods is, at least partly, brought about by the varying degrees of generality with which one can approach the problem. One can, for example, formulate the problem in a operator-theoretic setting, and use methods appropriate for this setting to obtain general results that may subsequently be applied to birth-death processes. This approach is chosen, for example, in [4]. But one can also choose a setting tailored to birth-death processes by starting with Karlin and McGregor's [16], [17] spectral representation for the transition probabilities of a birth-death process, and dealing with the problem using methods from the theory of orthogonal polynomials. This approach is adopted, for example, in [8]. In the present paper we take an intermediate position by choosing Kolmogorov's forward differential equations for the transition probabilities of an ergodic Markov chain as our point of departure.

The purpose of this paper is twofold. In the first place we want to survey the method for studying the rate of convergence of a birth-death process that one of us, inspired by [11], initiated in the early 1990's (see [29], [30] and [31]), and developed further in cooperation with Granovsky (see [12], [13] and [14]). Secondly, we want to present some new results obtained by the method, related in particular to the asymptotic behaviour of the rate of convergence for some specific birth-death processes as the number of states tends to infinity. The method is based on two main ingredients. The first is the concept of *logarithmic norm* of a square matrix, developed independently by Dahlquist [6] and Lozinskiĭ [20] as a tool to derive error bounds in the numerical integration of initial-value problems for a system of ordinary differential equations (see also the survey papers [25] and [24]). The second ingredient is a suitable transformation of the Kolmogorov forward equations. We will consider the simple setting of a *finite* Markov chain in order to be able to emphasize methodology and not be distracted by technical details.

The plan of the paper is as follows. We start off in Section 2 in the setting of a system of differential equations of the form

$$\boldsymbol{x}'(t) = B\boldsymbol{x}(t), \quad t \geq 0, \tag{1}$$

where $B \equiv (b_{ij})$ is an *essentially positive* square matrix, that is, B is irreducible and $b_{ij} \geq 0$ if $i \neq j$. We show how a technique involving the logarithmic norm of B leads to representations and bounds for the eigenvalue of B with maximal real part and, accordingly, to representations and bounds for the (suitably defined) norm of $\boldsymbol{x}(t)$ (Theorems 1, 2 and 3). In Section 3 we apply the results of Section 2 to a system of differential equations obtained by a suitable transformation from the Kolmogorov forward equations for the transient state probabilities of a finite

1

Markov chain. Under the assumption that the matrix determining the system is essentially positive, this approach results in representations and bounds for the rate of convergence of the Markov chain to its stationary distribution, and in representations and bounds for the (suitably defined) distance between the distribution at time $t \geq 0$ and the stationary distribution (Theorems 5 and 6). Then, it is shown in Section 4 that the assumption underlying the results of Section 3 is justified if the Markov chain is a birth-death process, leading to unconditional results in this setting (Theorems 8, 9 and 10).

By substituting the birth and death rates and choosing values for a set of parameters, the results of Section 4 can be transformed into explicit bounds for the rate of convergence of a specific birth-death process. In exceptional cases an appropriate choice of the parameter values can even produce the exact value of the rate of convergence. More common is the situation in which the parameter values can be chosen such that the bounds produced are asymptotically (as the number of states tends to infinity) equal to the rate of convergence itself. In the Sections 5 and 6 we exhibit some birth-death processes for which such a choice is possible. Concretely, birth-death processes related to *mean field models* will be considered in Section 5, and the number of customers in an $M/M/N/N + R$ service system in Section 6.

2 Preliminary results

In what follows $\mathbf{0}$ and $\mathbf{1}$ denote column vectors of zeros and of ones, respectively, of appropriate length. Inequality for vectors and for matrices indicates elementwise inequality, and superscript ^{T} denotes transposition. We assume that the matrix norm $|| \cdot ||$ is induced by a vector norm, that is, $||A|| := \max\{||A\boldsymbol{a}|| : ||\boldsymbol{a}|| = 1\}$. As a consequence we have $||A\boldsymbol{a}|| \leq ||A|| \, ||\boldsymbol{a}||$ for any matrix A and vector $\boldsymbol{a}$ of suitable dimensions. Moreover, for any square matrix A with eigenvalue λ we have $|\lambda| \leq ||A||$.

Now let $B \equiv (b_{ij})$ be an $n \times n$ matrix, and suppose that $\boldsymbol{x}(.)$ is determined by the system (1), so that

$$\boldsymbol{x}(t) = e^{tB}\boldsymbol{x}(0), \quad t \geq 0.$$

It is well known (see, for example, [25, Lemma 1c] or [13, Assertion 3]) that

$$||e^{tB}|| \leq e^{t\gamma(B)}, \quad t \geq 0, \tag{2}$$

and hence

$$||\boldsymbol{x}(t)|| \leq ||e^{tB}|| \, ||\boldsymbol{x}(0)|| \leq e^{t\gamma(B)}||\boldsymbol{x}(0)||, \quad t \geq 0, \tag{3}$$

where

$$\gamma(B) := \lim_{h\downarrow 0} \frac{||I + hB|| - 1}{h} \tag{4}$$

2

is the *logarithmic norm* of B.

We denote the eigenvalues of B by $\lambda_1, \lambda_2, \ldots, \lambda_n$ and assume they are numbered such that $\Re(\lambda_n) \leq \Re(\lambda_{n-1}) \leq \ldots \leq \Re(\lambda_1)$. Since $e^{\lambda t}$ is an eigenvalue of e^{tB} if λ is an eigenvalue of B, we have

$$e^{t\Re(\lambda_1)} = |e^{\lambda_1 t}| \leq ||e^{tB}||, \quad t \geq 0,$$

and hence, in view of (2),

$$\Re(\lambda_1) \leq \gamma(B). \tag{5}$$

On the other hand, by Gelfand's spectral radius formula (see, for example, [23, Theorem 10.13] or [27, Theorem 3.4]),

$$||e^{kB}||^{1/k} \to e^{\Re(\lambda_1)} \quad \text{as} \quad k \to \infty,$$

so that

$$\Re(\lambda_1) = \inf\{b \in \mathbb{R} : \ ||e^{tB}|| = \mathcal{O}\left(e^{bt}\right) \quad \text{as} \quad t \to \infty\}.$$

Since $||e^{tB}|| = \max\{||\boldsymbol{x}(t)|| : \ ||\boldsymbol{x}(0)|| = 1\}$, we also have

$$\Re(\lambda_1) = \inf\{b \in \mathbb{R} : \ ||\boldsymbol{x}(t)|| = \mathcal{O}\left(e^{bt}\right) \quad \text{as} \quad t \to \infty \ \text{for all} \ \boldsymbol{x}(0)\}. \tag{6}$$

We will assume from now on that $|| \cdot || = || \cdot ||_1$, so that $||\boldsymbol{a}|| = \sum_i |a_i|$ for any (column) vector $\boldsymbol{a} \equiv (a_i)$, and $||A|| = \max_j \sum_i |a_{ij}|$ for any matrix $A \equiv (a_{ij})$. As a consequence,

$$\gamma(B) = \max_j \{b_{jj} + \sum_{i \neq j} |b_{ij}|\}.$$

If B is essentially positive then, for $r \in \mathbb{R}$ sufficiently large, the matrix $B + rI$ is non-negative, so that we can employ Perron-Frobenius theory (see, for example, [22] or [27]). It follows in particular that $\lambda_1 + r$, and hence λ_1, is real. Letting,

$$c_{min}(B) := \min_j \sum_i b_{ij} \quad \text{and} \quad c_{max}(B) := \max_j \sum_i b_{ij}, \tag{7}$$

we also have $\gamma(B) = c_{max}(B)$, so that (5) and (3) lead to $\lambda_1 \leq c_{max}(B)$ and

$$||\boldsymbol{x}(t)|| \leq e^{tc_{max}(B)}||\boldsymbol{x}(0)||, \quad t \geq 0,$$

respectively. Moreover, in the setting at hand e^{tB} is a positive operator, that is, $e^{tB}\boldsymbol{x} \geq \boldsymbol{0}$ if $\boldsymbol{x} \geq \boldsymbol{0}$ (see, for example, [13, Assertion 1] or [27, Theorem 8.2]). As a consequence

$$\boldsymbol{x}(0) \geq \boldsymbol{0} \implies ||\boldsymbol{x}(t)|| \geq e^{tc_{min}(B)}||\boldsymbol{x}(0)||, \quad t \geq 0 \tag{8}$$

(see [13, Assertion 2], where the condition $\sum_i b_{ij} \leq 0$ is not necessary). So, in view of (6), we must have $\lambda_1 \geq c_{min}(B)$. Clearly, replacing $\boldsymbol{x}(0) \geq \boldsymbol{0}$ by $\boldsymbol{x}(0) \leq \boldsymbol{0}$ in (8) does not alter the conclusion. We summarize the preceding results in the next theorem.

Theorem 1. Let B be essentially positive and $\boldsymbol{x}(0)$ any initial vector. Then

$$c_{min}(B) \leq \lambda_1 \leq c_{max}(B), \tag{9}$$

and

$$e^{tc_{min}(B)}||\boldsymbol{x}(0)|| \leq ||\boldsymbol{x}(t)|| \leq e^{tc_{max}(B)}||\boldsymbol{x}(0)||, \quad t \geq 0, \tag{10}$$

where the left inequality in (10) presupposes $\boldsymbol{x}(0) \geq \boldsymbol{0}$ or $\boldsymbol{x}(0) \leq \boldsymbol{0}$.

The system (1) does not change if we replace $\boldsymbol{x}(0)$ by $-\boldsymbol{x}(0)$ and B by $-B$, so we can try to repeat the arguments leading to Theorem 1 in terms of $-B$ rather than B. However, since $-B$ is *not* essentially positive if B is essentially positive, we can now use only those arguments that do not require essential positivity. Letting

$$s_{min}(B) := \min_{j}\{b_{jj} - \sum_{i \neq j} b_{ij}\} \quad \text{and} \quad s_{max}(B) := \max_{j}\{b_{jj} - \sum_{i \neq j} b_{ij}\}, \tag{11}$$

and noting that $-\gamma(-B) = \min_{j}\{b_{jj} - \sum_{i \neq j}|b_{ij}|\} = s_{min}(B)$ if B is essentially positive, this line of reasoning gives us the following additional results.

Theorem 2. Let B be essentially positive and $\boldsymbol{x}(0)$ any initial vector. Then

$$s_{min}(B) \leq \Re(\lambda_n), \tag{12}$$

and

$$e^{ts_{min}(B)}||\boldsymbol{x}(0)|| \leq ||\boldsymbol{x}(t)||, \quad t \geq 0. \tag{13}$$

Evidently, $s_{min}(B) \leq c_{min}(B)$ if B is essentially positive, so the bound (13) is relevant only if $\boldsymbol{x}(0)$ has negative as well as positive components.

Maintaining the assumption that B is essentially positive, we can do better than Theorem 1. First note that for nonsingular T we can write

$$T\boldsymbol{x}'(t) = (TBT^{-1})T\boldsymbol{x}(t), \quad t \geq 0,$$

so, applying Theorem 1 to TBT^{-1} rather than B, we conclude

$$e^{tc_{min}(TBT^{-1})}||\boldsymbol{x}(0)||_T \leq ||\boldsymbol{x}(t)||_T \leq e^{tc_{max}(TBT^{-1})}||\boldsymbol{x}(0)||_T, \quad t \geq 0, \tag{14}$$

where, for any $n \times n$ matrix T, the vector norm $||\cdot||_T$ is defined by

$$||\boldsymbol{a}||_T := ||T\boldsymbol{a}||,$$

and the left inequality in (14) presupposes $T\boldsymbol{x}(0) \geq \boldsymbol{0}$ or $T\boldsymbol{x}(0) \leq \boldsymbol{0}$.

Next choose $r \in \mathbb{R}$ again so large that the matrix $B + rI$ is nonnegative. Then the *Collatz-Wielandt theorem* (see, for example, [22, pp. 666-669] or [27, Theorem 2.9]) states that, for any diagonal matrix D with positive diagonal entries, the spectral radius $\rho(B + rI)$ of the matrix $B + rI$ satisfies either

$$c_{min}(D(B + rI)D^{-1}) < \rho(B + rI) < c_{max}(D(B + rI)D^{-1})$$

or

$$c_{min}(D(B + rI)D^{-1}) = \rho(B + rI) = c_{max}(D(B + rI)D^{-1}).$$

Moreover, the latter occurs if and only if $D\mathbf{1}$ is the positive eigenvector of $B + rI$ corresponding to the eigenvalue $\rho(B + rI)$ (which is unique up to a multiplicative constant). Noting that $\rho(B + rI) = \lambda_1 + r$, and replacing T by D in (14), we readily obtain the following generalization of Theorem 1.

Theorem 3. Let B essentially positive and $\mathbf{x}(0)$ any nonzero initial vector. Then, for any diagonal matrix D with positive diagonal entries,

$$c_{min}(DBD^{-1}) \leq \lambda_1 \leq c_{max}(DBD^{-1}), \tag{15}$$

and

$$e^{tc_{min}(DBD^{-1})}||\mathbf{x}(0)||_D \leq ||\mathbf{x}(t)||_D \leq e^{tc_{max}(DBD^{-1})}||\mathbf{x}(0)||_D, \quad t \geq 0, \tag{16}$$

where the left inequality in (16) presupposes $\mathbf{x}(0) \geq \mathbf{0}$ or $\mathbf{x}(0) \leq \mathbf{0}$. The extreme values in (15) (and hence in (16) if $\mathbf{x}(0) \geq \mathbf{0}$ or $\mathbf{x}(0) \leq \mathbf{0}$) are achieved if and only if $D\mathbf{1}$ is the positive eigenvector of B corresponding to λ_1.

The next corollary is immediate (see also [27, Exercise 8.2.6]).

Corollary 4. If B is essentially positive and D a diagonal matrix, then, for any nonzero initial vector $\mathbf{x}(0) \geq \mathbf{0}$,

$$||\mathbf{x}(t)||_D = e^{\lambda_1 t}||\mathbf{x}(0)||_D, \quad t \geq 0, \tag{17}$$

if and only if $D\mathbf{1}$ is the positive eigenvector of B corresponding to λ_1.

For completeness' sake (and later reference) we also state the result one obtains by replacing B by DBD^{-1} and $\mathbf{x}(t)$ by $D\mathbf{x}(t)$ in (1) and (13), namely

$$e^{ts_{min}(DBD^{-1})}||\mathbf{x}(0)||_D \leq ||\mathbf{x}(t)||_D, \quad t \geq 0, \tag{18}$$

for any initial vector $\mathbf{x}(0)$ and diagonal matrix D with positive diagonal entries. Evidently, in view of Theorem 3, the inequality is relevant only if $\mathbf{x}(0)$ has negative as well as positive components, since $s_{min}(DBD^{-1}) \leq c_{min}(DBD^{-1})$.

Remark. A slight generalization of Theorem 3 and (18) may be obtained by considering, instead of (1), the system

$$\boldsymbol{y}'(t) = q(t)B\boldsymbol{y}(t), \quad t \geq 0, \tag{19}$$

where $q(\cdot)$ is a nonnegative and locally integrable (scalar) function on $[0, \infty)$. In this case we have

$$\boldsymbol{y}(t) = e^{\bar{q}(t)B}\boldsymbol{y}(0), \quad t \geq 0,$$

where $\bar{q}(t) := \int_0^t q(u)du$. Defining $\boldsymbol{x}(t) := \boldsymbol{y}(\bar{q}^{-1}(t))$, the system (19) reduces to the system (1), so we can immediately apply (16), (17) and (18) and translate the results in terms of $\boldsymbol{y}(\cdot)$. This amounts to replacing $\boldsymbol{x}(\cdot)$ by $\boldsymbol{y}(\cdot)$ in (16), (17) and (18) and writing $\bar{q}(t)$ instead of t in the exponents. $\qquad\square$

If B is *not* essentially positive one might look for a similarity transformation $C = TBT^{-1}$, such that C is essentially positive, and apply Theorem 3 to C instead of B to obtain bounds on λ_1.

3 Convergence of Markov chains

Consider a continuous-time Markov chain $\mathcal{X} \equiv \{X(t),\ t \geq 0\}$ taking values in $\{0, 1, \ldots, n\}$ with q-matrix $Q \equiv (q_{ij})$. The state probabilities $p_j(t) := \Pr\{X(t) = j\}$, $j = 0, 1, \ldots, n$, are determined by the initial distribution and the system of forward equations

$$\boldsymbol{p}'(t) = Q^T\boldsymbol{p}(t), \quad t \geq 0, \tag{20}$$

where $\boldsymbol{p}(t) \equiv (p_0(t), p_1(t), \ldots, p_n(t))^T$. We will assume that $\mathcal{X}$ is honest and irreducible, so that there exists a unique stationary distribution, represented by the vector $\boldsymbol{\pi} \equiv (\pi_0, \pi_1, \ldots, \pi_n)^T$. The distribution $\boldsymbol{\pi}$ satisfies $Q^T\boldsymbol{\pi} = \boldsymbol{0}$, and $\boldsymbol{p}(t)$ converges to $\boldsymbol{\pi}$ as $t \to \infty$, for any initial distribution $\boldsymbol{p}(0)$. In what follows we are interested in the *rate of convergence* (sometimes called the *decay parameter*) β, defined by

$$\beta := \sup\{a > 0 : ||\boldsymbol{p}(t) - \boldsymbol{\pi}|| = \mathcal{O}(e^{-at}) \ \text{ as } \ t \to \infty \ \text{ for all } \ \boldsymbol{p}(0)\}. \tag{21}$$

Since $\sum_{k=0}^n p_k(t) = \sum_{k=0}^n \pi_k = 1$, we evidently have

$$0 < |p_0(t) - \pi_0| \leq \sum_{k=1}^n |p_k(t) - \pi_k|, \quad t \geq 0.$$

So, letting $\boldsymbol{x}(t) := (p_1(t) - \pi_1, p_2(t) - \pi_2, \ldots, p_n(t) - \pi_n)^T$, we obtain

$$||\boldsymbol{x}(t)|| \leq ||\boldsymbol{p}(t) - \boldsymbol{\pi}|| \leq 2||\boldsymbol{x}(t)||, \quad t \geq 0, \tag{22}$$

and it follows that

$$\beta = \sup\{a > 0 : ||\boldsymbol{x}(t)|| = \mathcal{O}(e^{-at}) \text{ as } t \to \infty \text{ for all } \boldsymbol{x}(0)\}. \tag{23}$$

Also observe that the forward equations (20) imply

$$p_i'(t) = q_{0i}(p_0(t) - \pi_0) + \sum_{j=1}^{n} q_{ji}(p_j(t) - \pi_j)$$

$$= \sum_{j=1}^{n} (q_{ji} - q_{0i})(p_j(t) - \pi_j), \quad i = 1, 2, \ldots, n, \ t \geq 0,$$

so that

$$x_i'(t) = \sum_{j=1}^{n} (q_{ji} - q_{0i})x_j(t), \quad i = 1, 2, \ldots, n, \ t \geq 0.$$

That is, $\boldsymbol{x}(\cdot)$ actually satisfies the system of differential equations (1), and hence $\boldsymbol{x}(t) = e^{tB}\boldsymbol{x}(0)$, with $B \equiv (b_{ij})$ given by

$$b_{ij} := q_{ji} - q_{0i}, \quad i, \ j = 1, 2, \ldots, n. \tag{24}$$

As in the previous section we denote the eigenvalues of B by $\lambda_1, \lambda_2, \ldots, \lambda_n$ and assume they are numbered such that $\Re(\lambda_n) \leq \Re(\lambda_{n-1}) \leq \ldots \leq \Re(\lambda_1)$. Now suppose λ is an eigenvalue of Q and let $\boldsymbol{y}^* \equiv (y_0, y_1, \ldots, y_n)^T$ be a corresponding eigenvector. Since $Q\boldsymbol{1} = 0$, it follows readily that

$$B^T(\boldsymbol{y} - y_0\boldsymbol{1}) = \lambda(\boldsymbol{y} - y_0\boldsymbol{1}),$$

where $\boldsymbol{y} := (y_1, y_2, \ldots, y_n)^T$. So λ is also an eigenvalue of B, unless $\boldsymbol{y} - y_0\boldsymbol{1} = \boldsymbol{0}$, the latter occurring if and only if $\boldsymbol{y}^*$ is a constant vector, and hence $\lambda = 0$. Since $\mathcal{X}$ is an honest and irreducible Markov chain, the matrix Q has an eigenvalue 0 of multiplicity one, while all other eigenvalues have negative real parts. We thus conclude that $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the nonzero eigenvalues of Q, and satisfy

$$\Re(\lambda_n) \leq \Re(\lambda_{n-1}) \leq \ldots \leq \Re(\lambda_1) < 0.$$

Finally, (6) and (23) imply that

$$\beta = -\Re(\lambda_1), \tag{25}$$

so studying the rate of convergence of the Markov chain $\mathcal{X}$ with q-matrix $Q \equiv (q_{ij})$ amounts to studying the eigenvalue with largest real part of the matrix $B \equiv (b_{ij})$ defined by (24).

Although Q is essentially positive, the matrix B, in general, is *not*. So, as suggested at the end of the previous section, we must find a similarity transformation $C = TBT^{-1}$ such that C is essentially positive (if it exists at all), before

we can apply Theorem 3. In the next section we will display a suitable similarity transformation in the specific setting of birth-death processes, but in the remainder of this section we will only assume its existence. As a consequence of this assumption λ_1 must be real. Evidently, if B itself happens to be essentially positive we let $T = I$, the identity matrix. We note, parenthetically, that, by [13, Theorem 2], a suitable similarity transformation always exists when $\mathcal{X}$ is a *reversible* Markov chain (in which case *all* eigenvalues are real).

Under the assumption then that there exists a nonsingular matrix T such that $C = TBT^{-1}$ is essentially positive, we can apply the first part of Theorem 3 to C and, in view of (25), obtain the following information about β, the rate of convergence of the Markov chain $\mathcal{X}$.

Theorem 5. Let T be a nonsingular matrix such that $C = TBT^{-1}$ is essentially positive, and let D be a diagonal matrix with positive diagonal entries. Then

$$c_{min}(DT(-B)(DT)^{-1}) \leq \beta \leq c_{max}(DT(-B)(DT)^{-1}), \tag{26}$$

where the extreme values are achieved if and only if $D\mathbf{1}$ is the positive eigenvector of C corresponding to the eigenvalue $\lambda_1 = -\beta$.

Defining $\boldsymbol{z}(\cdot) := T\boldsymbol{x}(\cdot)$, we have

$$\boldsymbol{z}'(t) = TBT^{-1}\boldsymbol{z}(t) = C\boldsymbol{z}(t), \quad t \geq 0,$$

since $\boldsymbol{x}(\cdot)$ satisfies the system of differential equations (1). Note further that

$$||\boldsymbol{z}(t)||_D = ||DT\boldsymbol{x}(t)|| \leq ||DT|| \, ||\boldsymbol{x}(t)||, \quad t \geq 0,$$

and

$$||\boldsymbol{x}(t)|| = ||(DT)^{-1}D\boldsymbol{z}(t)|| \leq ||(DT)^{-1}|| \, ||\boldsymbol{z}(t)||_D, \quad t \geq 0,$$

so that, in view of (22),

$$||DT||^{-1} \, ||\boldsymbol{z}(t)||_D \leq ||\boldsymbol{p}(t) - \boldsymbol{\pi}|| \leq 2 \, ||(DT)^{-1}|| \, ||\boldsymbol{z}(t)||_D, \quad t \geq 0. \tag{27}$$

We are now ready to apply to $\boldsymbol{z}(\cdot)$ the second part of Theorem 3 as well as (18). Together with (27) this readily leads to the following bounds for $||\boldsymbol{p}(t) - \boldsymbol{\pi}||$.

Theorem 6. Let T be a nonsingular matrix such that $C = TBT^{-1}$ is essentially positive, and let D be a diagonal matrix with positive diagonal entries. Then, for any initial distribution $\boldsymbol{p}(0)$,

$$||\boldsymbol{p}(t) - \boldsymbol{\pi}|| \leq \kappa \, e^{-c_{min}t} ||\boldsymbol{p}(0) - \boldsymbol{\pi}||, \quad t \geq 0, \tag{28}$$

and

$$||\boldsymbol{p}(t) - \boldsymbol{\pi}|| \geq \frac{1}{\kappa} e^{-s_{max}t} ||\boldsymbol{p}(0) - \boldsymbol{\pi}||, \quad t \geq 0, \tag{29}$$

where $\kappa := 2 \, ||DT|| \, ||(DT)^{-1}||$, c_{min} and c_{max} denote the left-hand and right-hand side, respectively, of (26), and $s_{max} := s_{max}(DT(-B)(DT)^{-1})$. Moreover, if $p_i(0) \geq \pi_i$ for $i = 1, 2, \ldots, n$, or $p_i(0) \leq \pi_i$ for $i = 1, 2, \ldots, n$, then

$$||\boldsymbol{p}(t) - \boldsymbol{\pi}|| \geq \frac{1}{\kappa} e^{-c_{max}t} ||\boldsymbol{p}(0) - \boldsymbol{\pi}||, \quad t \geq 0. \tag{30}$$

We note that the condition preceding (30) is satisfied, for instance, when the initial distribution concentrates all probability mass in state 0. Of course, (29) becomes irrelevant if (30) holds true, since $s_{max} < c_{max}$.

Since B and $DTB(DT)^{-1}$ have the same eigenvalues, Theorem 2 tells us that $-s_{max} \leq \Re(\lambda_n)$. We shall see in the next section that we can obtain more precise results for λ_n, the eigenvalue of Q with the smallest real part, in the specific setting of birth-death processes.

Our final observation in this section is the following.

Theorem 7. If there exists a nonsingular matrix T such that $C = TBT^{-1}$ is essentially positive, then there exists a nonnegative matrix Δ such that, for any initial distribution $\boldsymbol{p}(0)$ satisfying $p_i(0) \geq \pi_i$ for $i = 1, 2, \ldots, n$, or $p_i(0) \leq \pi_i$ for $i = 1, 2, \ldots, n$,

$$||\boldsymbol{p}(t) - \boldsymbol{\pi}||_\Delta = e^{-\beta t} ||\boldsymbol{p}(0) - \boldsymbol{\pi}||_\Delta, \quad t \geq 0.$$

Proof. Choose the diagonal matrix D such that $D\boldsymbol{1}$ is the positive eigenvector of C corresponding to λ_1. By Corollary 3, applied to C and $\boldsymbol{z}(\cdot)$, we then have

$$||\boldsymbol{z}(t)||_D = e^{-\beta t} ||\boldsymbol{z}(0)||_D, \quad t \geq 0.$$

Letting Δ be the matrix with zeros in the first row and column, and the remaining block equal to DT, we also have

$$||\boldsymbol{p}(t) - \boldsymbol{\pi}||_\Delta = ||\boldsymbol{x}(t)||_{DT} = ||\boldsymbol{z}(t)||_D, \quad t \geq 0,$$

so that the result follows. $\qquad \square$

In the next section we will show that the technique expounded in the present section for the analysis of the rate of convergence of a finite, continuous-time Markov chain is applicable, and thus yields bounds and representations, in the more restricted setting of birth-death processes.

4 Birth-death processes

We will assume in this section that the Markov chain $\mathcal{X}$ of the previous section is a birth-death process, so that its q-matrix has the form

$$
Q = \begin{pmatrix}
-a_0 & a_0 & 0 & \cdots & 0 & 0 \\
b_1 & -(a_1 + b_1) & a_1 & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & -(a_{n-1} + b_{n-1}) & a_{n-1} \\
0 & 0 & 0 & \cdots & b_n & -b_n
\end{pmatrix},
\tag{31}
$$

with positive *birth rates* a_i, $0 \le i < n$, and *death rates* b_i, $0 < i \le n$. It is well known that the nonzero eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$ of Q are real, distinct and negative (see, for example, [17]). We have seen in the previous section that these eigenvalues are precisely the eigenvalues of the matrix B of (24), which is now given by

$$
B = \begin{pmatrix}
-(a_0 + a_1 + b_1) & b_2 - a_0 & -a_0 & \cdots & -a_0 & -a_0 \\
a_1 & -(a_2 + b_2) & b_3 & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & -(a_{n-1} + b_{n-1}) & b_n \\
0 & 0 & 0 & \cdots & a_{n-1} & -b_n
\end{pmatrix}.
$$

Choosing T to be the upper triangular matrix with 1's on and above the diagonal, that is,

$$
T = \begin{pmatrix}
1 & 1 & 1 & \cdots & 1 & 1 \\
0 & 1 & 1 & \cdots & 1 & 1 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & 1 & 1 \\
0 & 0 & 0 & \cdots & 0 & 1
\end{pmatrix}
\quad \text{and} \quad
T^{-1} = \begin{pmatrix}
1 & -1 & 0 & \cdots & 0 & 0 \\
0 & 1 & -1 & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & 1 & -1 \\
0 & 0 & 0 & \cdots & 0 & 1
\end{pmatrix},
\tag{32}
$$

the similarity transformation $C = TBT^{-1}$ yields the essentially positive matrix

$$
C = \begin{pmatrix}
-(a_0 + b_1) & b_1 & 0 & \cdots & 0 & 0 \\
a_1 & -(a_1 + b_2) & b_2 & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & -(a_{n-2} + b_{n-1}) & b_{n-1} \\
0 & 0 & 0 & \cdots & a_{n-1} & -(a_{n-1} + b_n)
\end{pmatrix}.
\tag{33}
$$

So we can apply Theorem 5 and obtain after a little algebra the next theorem, which, in its full generality, was first stated in [31].

Theorem 8. The rate of convergence β of a birth-death process with q-matrix (31) satisfies

$$
\max_{\boldsymbol{d} > 0} \left\{ \min_{1 \le i \le n} \alpha_i \right\} = \beta = \min_{\boldsymbol{d} > 0} \left\{ \max_{1 \le i \le n} \alpha_i \right\},
\tag{34}
$$

10

where $\boldsymbol{d} \equiv (d_1, d_2, \ldots, d_n)$, and

$$\alpha_i := a_{i-1} + b_i - \frac{d_{i+1}}{d_i}a_i - \frac{d_{i-1}}{d_i}b_{i-1}, \quad i = 1, 2, \ldots, n, \tag{35}$$

with $a_n = b_0 = d_0 = d_{n+1} = 0$.

There is a vast literature on results related to Theorem 8, derived by a variety of methods, and often pertaining to ergodic birth-death processes with an infinite state space. We refer to [2], [3], [4], [7], [8], [9], [12], [13], [15], [18], [29], [30], [31], [32], and the references there for further information.

Information about λ_n, the smallest eigenvalue of Q, and hence of C, may be obtained by observing that $-\lambda_n$ is the largest eigenvalue of $-C$. Since the characteristic polynomial $P_r(\lambda)$ for the leading principal submatrix of order r of the matrix $-C$ satisfies the recursion

$$P_r(\lambda) = (a_{r-1} + b_r - \lambda)P_{r-1}(\lambda) - a_{r-1}b_{r-1}P_{r-2}(\lambda), \quad r = 2, 3 \ldots, n,$$
$$P_0(\lambda) = 1, \quad P_1(\lambda) = a_0 + b_1 - \lambda,$$

we see that $P_n(\lambda)$, and hence the spectrum of $-C$, does not change if we change the signs of its nondiagonal elements. In other words, $-\lambda_1 = \beta < -\lambda_2 < \ldots < -\lambda_n$ are the eigenvalues of the nonnegative (and irreducible, and hence essentially positive) matrix C^*, which is obtained from C by changing the signs of its diagonal elements. Subsequently applying Theorem 3 to C^* gives us the next result.

Theorem 9. Let $\chi = -\lambda_n$, with λ_n the smallest eigenvalue of the q-matrix (31). Then χ satisfies

$$\max_{\boldsymbol{d} > \boldsymbol{0}} \left\{ \min_{1 \leq i \leq n} \zeta_i \right\} = \chi = \min_{\boldsymbol{d} > \boldsymbol{0}} \left\{ \max_{1 \leq i \leq n} \zeta_i \right\}, \tag{36}$$

where $\boldsymbol{d} \equiv (d_1, d_2, \ldots, d_n)$, and

$$\zeta_i := a_{i-1} + b_i + \frac{d_{i+1}}{d_i}a_i + \frac{d_{i-1}}{d_i}b_{i-1}, \quad i = 1, 2, \ldots, n, \tag{37}$$

with $a_n = b_0 = d_0 = d_{n+1} = 0$.

We next give exponential bounds for $||\boldsymbol{p}(t) - \boldsymbol{\pi}||$ in the setting at hand. The results have essentially been derived earlier in [29] and [31].

Theorem 10. Let $\mathcal{X}$ be a birth-death process with q-matrix (31). Then, for any initial distribution $\boldsymbol{p}(0)$ and vector $\boldsymbol{d} \equiv (d_1, d_2, \ldots, d_n)$ such that $d := \min_i\{d_i\} > 0$, we have

$$||\boldsymbol{p}(t) - \boldsymbol{\pi}|| \leq \kappa \, e^{-\min_i\{\alpha_i\}t}||\boldsymbol{p}(0) - \boldsymbol{\pi}||, \quad t \geq 0, \tag{38}$$

and

$$||\boldsymbol{p}(t) - \boldsymbol{\pi}|| \geq \frac{1}{\kappa} e^{-\max_i\{\zeta_i\}t}||\boldsymbol{p}(0) - \boldsymbol{\pi}||, \quad t \geq 0. \tag{39}$$

where $\kappa = 4\sum_{i=1}^{n}(d_i/d)$, and α_i and ζ_i are given by (35) and (37), respectively. Moreover, if $p_i(0) \geq \pi_i$ for $i = 1, 2, \ldots, n$, or $p_i(0) \leq \pi_i$ for $i = 1, 2, \ldots, n$, then

$$||\boldsymbol{p}(t) - \boldsymbol{\pi}|| \geq \frac{1}{\kappa} e^{-\max_i\{\alpha_i\}t}||\boldsymbol{p}(0) - \boldsymbol{\pi}||, \quad t \geq 0. \tag{40}$$

Proof. Letting $D = \mathrm{diag}(d_1, d_2, \ldots, d_n)$ and T as in (32), we have

$$||DT|| = \sum_{i=1}^{n} d_i \quad \text{and} \quad ||(DT)^{-1}|| = \frac{2}{d}.$$

The results (38) and (40) now follow from Theorem 6, while (39) is implied by (18). $\qquad\square$

We note that, by Theorem 9, the exponent in the lower bound (39) is at best equal to $-\chi t$. In the setting of an arbitrary initial distribution (and unspecified $n > 1$) this cannot be improved, in the sense that examples can be constructed for which $||\boldsymbol{p}(t) - \boldsymbol{\pi}|| = \mathcal{O}(e^{-\chi t})$ as $t \to \infty$. One should bear in mind, however, that for large t the lower bound (39) is (almost surely) very conservative, since the spectral expansion of the quantities $p_j(t) - \pi_j$, and hence of $||\boldsymbol{p}(t) - \boldsymbol{\pi}||$, will involve a term of the order $e^{-\beta t}$, unless the birth and death rates and the initial distribution satisfy a very specific constraint (which happens with probability zero if, say, the initial distribution is chosen randomly).

We conclude this section with the simple example of a birth-death process with constant birth rates $a_i = a$, $0 \leq i < n$, and death rates $b_i = b$, $0 < i \leq n$. We will indicate dependence on n by writing β_n and χ_n instead of β and χ, respectively. Choosing

$$d_i = \left(-\sqrt{\frac{b}{a}}\right)^{i-1} \frac{\sin\frac{in\pi}{n+1}}{\sin\frac{n\pi}{n+1}}, \quad i = 1, \ldots, n,$$

(which is positive since $(i-1)\pi < \frac{in\pi}{n+1} < i\pi$) and exploiting the trigonometric identity

$$\sin\frac{(i+1)n\pi}{n+1} + \sin\frac{(i-1)n\pi}{n+1} = 2\sin\frac{in\pi}{n+1}\cos\frac{n\pi}{n+1}, \quad i = 1, \ldots, n,$$

it follows readily that the quantities α_i of (35) and ζ_i of (37) are constant and given by

$$\alpha_i = a + b - 2\sqrt{ab}\cos\frac{\pi}{n+1}, \quad \zeta_i = a + b + 2\sqrt{ab}\cos\frac{\pi}{n+1}, \quad i = 1, \ldots, n.$$

12

Hence, by the Theorems 8 and 9,

$$\beta_n = a + b - 2\sqrt{ab}\cos\frac{\pi}{n+1}, \ \ \chi_n = a + b + 2\sqrt{ab}\cos\frac{\pi}{n+1}, \quad n \geq 1. \quad (41)$$

These results are, in fact, well known (see, for example, [26, p. 13]) and [12, Example 2.3]), but the proof seems to be new. It follows in particular that

$$\lim_{n\to\infty}\beta_n = (\sqrt{a} - \sqrt{b})^2 \ \ \text{and} \ \ \lim_{n\to\infty}\chi_n = (\sqrt{a} + \sqrt{b})^2, \quad (42)$$

which are examples of the type of asymptotic results we will be interested in in the following sections.

5 Mean-field models

In this section we consider a spin system $\Phi \equiv (\boldsymbol{\phi}(t),\ t \geq 0)$ on a complete graph with n vertices (or *sites*). The state space for this process is the set of all 2^n binary vectors of length n, where the ith component of a state vector indicates whether the ith site is empty (0) or occupied (1), say. The process is described by $2n$ nonnegative *flip rates* λ_i and μ_i, $i = 0, 1, \ldots, n-1$, where λ_i and μ_i are the rates with which an empty site and an occupied site, respectively, swap states if the site has i occupied neighbouring sites. Such spin systems are known in statistical physics as *mean-field models*.

Denoting the total number of ones in a binary vector $\boldsymbol{a}$ by $|\boldsymbol{a}|$ (so that $|\boldsymbol{\phi}(t)|$ represents the total number of occupied sites at time t), and defining

$$X(t) := |\boldsymbol{\phi}(t)|, \quad t \geq 0,$$

it is clear that $\mathcal{X} := \{X(t),\ t \geq 0\}$ is a birth-death process with state space $\{0, 1, \ldots, n\}$, and birth and death rates

$$a_i = (n - i)\lambda_i, \ \ 0 \leq i < n, \ \ \text{and} \ \ b_i = i\mu_{i-1}, \ \ 0 < i \leq n,$$

respectively. So the speed of convergence of the mean-field model Φ to its stationary regime may be characterized by the rate of convergence of the corresponding birth-death process $\mathcal{X}$, and hence the techniques of the previous section may be applied to study convergence of mean-field models. This approach was adopted in [12], [13] and [15].

For instance, by using Theorem 8 it was shown in [12], that if $\lambda_i + \mu_i = c$ for $0 \leq i \leq n - 1$, then

$$\beta_n \geq c - (n-1)\max_i\{|\lambda_i - \lambda_{i-1}|, |\mu_i - \mu_{i-1}|\}, \quad n \geq 1,$$

with equality subsisting if and only if

$$\lambda_i = \lambda_0 + ih, \quad \mu_i = \mu_0 - ih, \quad 0 \leq i \leq n - 1,$$

13

for some real number h, in which case $\beta_n = c - (n-1)h$, $n \geq 1$. In a similar fashion it was shown in [15] that, given $\lambda_i + \mu_i = c$, the value of χ_n can be determined exactly, namely

$$\chi_n = nc, \quad n \geq 1.$$

As before we have indicated dependence on n by writing β_n and χ_n instead of β and χ, respectively.

A special class of mean-field models, characterized by a uniform stationary distribution, arises when $\lambda_i = \mu_i$, $0 \leq i \leq n-1$. The case

$$\lambda_i = \mu_i = c(i+1)^s, \quad 0 \leq i \leq n-1,$$

for some $c > 0$ and $s \geq 0$, was considered in [13] and [15]. Specifically, by using Theorem 8 again, it was shown in [13] that

$$\lim_{n \to \infty} \frac{\beta_n}{n^s} = 2^{1-s}c, \quad 0 \leq s < 1,$$

and

$$\lim_{n \to \infty} \frac{\beta_n}{n} = c, \quad s \geq 1, \tag{43}$$

while an asymptotic result for χ_n was obtained in [15]. The result (43) can actually be generalized, as shown in the next theorem.

Theorem 11. Let $\lambda_0 = \mu_0 = c > 0$, and suppose $\lambda_i = \mu_i \geq c(i+1)$ for $0 < i \leq n-1$. Then

$$\lim_{n \to \infty} \frac{\beta_n}{n} = c.$$

Proof. Writing $d_1 = d/\lambda_0$, and choosing $d_{i+1} := 1/(i\lambda_i)$ for $i = 1, 2, \ldots, n$, it follows after some algebra that the quantities α_i of (35) are given by

$$\alpha_i = \begin{cases} \lambda_0 \left(\left(1 - \dfrac{1}{d}\right) n + 1 + \dfrac{1}{d} \right), & i = 1, \\[2ex] \dfrac{\lambda_1}{2}\left(n + 4 - 2d\right), & i = 2, \\[2ex] \dfrac{\lambda_{i-1}}{i}\left(n - \dfrac{i}{i-2}\right), & i = 3, 4, \ldots, n. \end{cases}$$

So, choosing $d := \sqrt{n}$, say, it follows that

$$\frac{\beta_n}{n} \geq \min_{1 \leq i \leq n}\left\{\frac{\alpha_i}{n}\right\} \geq c(1 + o(1)) \quad \text{as} \ n \to \infty.$$

On the other hand, we have

$$\min_{1 \leq i \leq n} \alpha_i \leq \alpha_1 = c(n + 1 - (n-1)/d_1) \leq c(n + o(n)) \quad \text{as} \ n \to \infty$$

for any $d_1 > 0$, so that the result follows from the max-min representation in Theorem 8. $\qquad \square$

6 The $M/M/N/N+R$ service system

In this section we will consider the $M/M/N/N+R$ service system, where $N \geq 1$ is the number of servers and $R \geq 0$ the number of waiting places. We allow the arrival rate $\lambda(N) > 0$ and $R \equiv R(N)$ to be functions of N. With μ denoting the service rate per server, the number of customers in this system is a birth-death process with rates

$$a_i = \lambda(N), \;\; 0 \leq i < n(N), \;\; \text{and} \;\; b_i = \min\{i, N\}\mu, \;\; 0 < i \leq n(N),$$

where $n(N) := N + R(N)$. We wish to study the asymptotic behaviour as $N \to \infty$ of $\beta \equiv \beta_{n(N)}$, the rate of convergence, and of $\chi \equiv \chi_{n(N)}$. Our first result concerns the case in which $\lambda(N)$ is constant.

Theorem 12. Let $\lambda(N) = \lambda > 0$, then, irrespective of the behaviour of $R(N)$,

$$\lim_{N \to \infty} \beta_{n(N)} = \mu.$$

Proof. Choosing

$$d_i = \begin{cases} 1, & 1 \leq i \leq N, \\ \left(\dfrac{N}{N-1}\right)^{i-N}, & N < i < N + R(N), \\ \dfrac{(N-1)\mu}{(N-1)\mu + \lambda}\left(\dfrac{N}{N-1}\right)^{R(N)} & i = N + R(N) \;\; (\text{if} \;\; R(N) > 0), \end{cases}$$

it follows after some algebra that the quantities α_i of (35) are given by

$$\alpha_i = \begin{cases} \mu, & 1 \leq i < N \;\; \text{and} \;\; i = N + R(N), \\ \mu - \dfrac{\lambda}{N-1}, & N \leq i < N + R(N) - 1, \\ \mu - \dfrac{\lambda(\lambda-\mu)}{(N-1)\mu + \lambda}, & i = N + R(N) - 1 \;\; (\text{if} \;\; R(N) > 1). \end{cases}$$

Consequently, $\alpha_i = \mu + o(1)$ as $N \to \infty$ for all i, and so, by Theorem 8, the statement of the theorem follows. $\square$

It is not difficult to see that under the conditions of Theorem 12 we also have

$$\lim_{N \to \infty} \frac{\chi_{n(N)}}{N} = \mu. \tag{44}$$

Indeed, by choosing $d_i = N^{i/2}$ for $1 \leq i \leq n(N)$, the quantities ζ_i of (37) satisfy $\zeta_i \leq N\mu + o(N)$, so that $\chi_{n(N)} \leq N\mu + o(N)$ as $N \to \infty$ by the min-max representation in (36). On the other hand, $\zeta_N > N\mu$ for any choice of d_i's so that $\chi_{n(N)} > N\mu$.

It was shown in [1] that in the special case $R(N) = 0$ (the *Erlang loss system*) $\beta_{n(N)} \equiv \beta_N$ can be identified with the smallest zero of the polynomial

$$S(x) := \frac{\lambda(N)}{x} \left\{ c_{N+1}\left(\frac{x}{\mu}, \frac{\lambda(N)}{\mu}\right) - c_N\left(\frac{x}{\mu}, \frac{\lambda(N)}{\mu}\right) \right\},$$

where

$$c_n(x, a) := \sum_{m=0}^n (-1)^m \binom{n}{m}\binom{x}{m}\frac{m!}{a^m}, \quad n \geq 0, \tag{45}$$

are the *Charlier polynomials* (see, for example, [5, Section VI.1]). Since Charlier polynomials satisfy the recurrence relation

$$a c_{n+1}(x, a) - a c_n(x, a) + x c_n(x - 1, a) = 0, \quad n \geq 0,$$

we can write

$$S(x) = -c_N\left(\frac{x}{\mu} - 1, \frac{\lambda(N)}{\mu}\right),$$

and hence

$$\beta_N = \mu + \mu \xi_{N,1}\left(\frac{\lambda(N)}{\mu}\right), \tag{46}$$

where $\xi_{N,1}(a)$ denotes the smallest zero of the Charlier polynomial $c_N(x, a)$. It is known that $\xi_{N,1}(a) > 0$, so that

$$\beta_N > \mu, \tag{47}$$

but no explicit expression for $\xi_{N,1}(a)$ seems to be available (for general a). However, we can say more about the asymptotic behaviour of β_N as $N \to \infty$. The case $\lambda(N)$ is constant is covered by Theorem 12, and we will now consider the more interesting case $\lambda(N) = \lambda N$, studied, for example, in [10] and [28].

When $\lambda = \mu$, we can actually establish the exact value of β_N. Indeed, since $c_N(x, a) = c_x(N, a)$ for natural x, we have $c_N(1, N) = c_1(N, N) = 0$. But Charlier polynomials are orthogonal with respect to a measure consisting of point masses at the points $0, 1, \ldots$, so it follows from the theory of orthogonal polynomials (see [5, Ch. 2]) that the ith smallest zero of $c_N(x, a)$ is larger than $i - 1$, for $i = 1, 2, \ldots$. As a consequence we must have $\xi_{N,1}(N) = 1$, and hence, for all $N \geq 1$,

$$\beta_N = 2\mu \quad \text{if} \quad \lambda = \mu. \tag{48}$$

We also observe that $c_N(1, a) = c_1(N, a) = 1 - N/a < 0$ if $a < N$. Since $c_0(0, a) = 1$, it follows that $\xi_{N,1}(a) < 1$ if $a < N$, and hence

$$\beta_N < 2\mu \quad \text{if} \quad \lambda < \mu. \tag{49}$$

Moreover, by exploiting our Theorem 8 we can give the following asymptotic results, which are consistent with (unproven) statements in [10].

Theorem 13. Let $\lambda(N) = \lambda N$ with $\lambda > 0$, then

$$\lim_{N \to \infty} \beta_N = \mu \quad \text{if} \ \ \lambda < \mu,$$

and

$$\lim_{N \to \infty} \frac{\beta_N}{N} = (\sqrt{\lambda} - \sqrt{\mu})^2 \ \ \text{if} \ \ \lambda > \mu.$$

Proof. First assume $\lambda < \mu$. Let $L := [N\lambda/\mu]$ and choose $d_{N+1} := 0$, $d_N := (N-1)\mu$,

$$d_i := d_{i+1} + (d_{i+1} - d_{i+2})\frac{N\lambda}{i\mu}, \quad i = N-1, N-2, \ldots, L,$$

and $d_i := d_L$ for $i < L$. Clearly, $d_i > d_{i+1}$ for $i = L, L+1, \ldots, N$, and hence $d_i > 0$ for $1 \le i \le N$. Substitution of these d_i's in (35) readily yields that $\alpha_i = \mu$ for $i \ne L$. Moreover,

$$0 < 1 - \frac{d_{i+1}}{d_i} = \frac{d_{i+1}}{d_i}\left(1 - \frac{d_{i+2}}{d_{i+1}}\right)\frac{N\lambda}{i\mu} < \left(1 - \frac{d_{i+2}}{d_{i+1}}\right)\frac{N\lambda}{i\mu}, \quad L \le i < N,$$

so that

$$1 - \frac{d_{L+1}}{d_L} < \frac{N\lambda}{L\mu} \cdots \frac{N\lambda}{(N-1)\mu} < \left(\frac{\lambda}{\mu}\right)^{N-L}.$$

As a consequence

$$\alpha_L = \mu + \left(1 - \frac{d_{L+1}}{d_L}\right)N\lambda < \mu + N\lambda\left(\frac{\lambda}{\mu}\right)^{N-L} = \mu + o(1) \ \ \text{as} \ \ N \to \infty,$$

which, in view of (47) and the min-max representation in Theorem 8, proves our first claim.

Next assume $\lambda > \mu$. By choosing $d_i := (\mu/\lambda)^{i/2}$, $i = 1, 2, \ldots, N$, the quantities α_i of (35) satisfy

$$\alpha_i = \begin{cases} (\lambda - \sqrt{\lambda\mu})N - (\sqrt{\lambda\mu} - \mu)i + \sqrt{\lambda\mu}, & 1 \le i < N, \\ \lambda N - (\sqrt{\lambda\mu} - \mu)N + \sqrt{\lambda\mu}, & i = N, \end{cases}$$

so that $\alpha_{N-1} = \min_i\{\alpha_i\}$. By Theorem 8 we therefore have

$$\beta_N \ge \alpha_{N-1} = (\sqrt{\lambda} - \sqrt{\mu})^2 N + 2\sqrt{\lambda\mu} - \mu. \tag{50}$$

On the other hand, letting $M_1 := [N - \sqrt{N}]$ and $M_2 := [N - \frac{1}{2}\sqrt{N}]$, and choosing

$$d_i := \left(2\sqrt{\frac{\mu}{\lambda}}\right)^i, \quad 1 \le i < M_1, \tag{51}$$

and

$$
d_{i+1} := \begin{cases} \dfrac{i - M_1 + 3}{i - M_1 + 2}\sqrt{\dfrac{\mu}{\lambda}}d_i, & M_1 \le i < M_2, \\[2ex] \sqrt{\dfrac{\mu}{\lambda}}d_i, & i = M_2, \\[2ex] \dfrac{N - i}{N - i + 1}\sqrt{\dfrac{\mu}{\lambda}}d_i, & M_2 < i < N, \end{cases}
\tag{52}
$$

it follows after some algebra that the quantities α_i of (35) satisfy

$$
\alpha_i \le (\sqrt{\lambda} - \sqrt{\mu})^2 + o(N) \quad \text{as} \quad N \to \infty, \quad 1 \le i \le N.
$$

Hence, by the min-max representation in Theorem 8,

$$
\beta_N \le (\sqrt{\lambda} - \sqrt{\mu})^2 N + o(N) \quad \text{as} \quad N \to \infty,
\tag{53}
$$

and so, in view of (50), the second claim of the theorem follows. □

Remarks. 1. Using an entirely different technique Krasikov [19, Theorem 3] has shown that

$$
\xi_{N,1}(aN) > (\sqrt{a} - 1)^2 N + c\sqrt[3]{N} \quad \text{if} \quad a > 1,
$$

for some positive constant c, so that (50) can actually be strengthened to

$$
\beta_N > (\sqrt{\lambda} - \sqrt{\mu})^2 N + c\mu\sqrt[3]{N}.
$$

2. An alternative proof of (53) may be given along the lines of the proof of Lemma 36 in [21, Ch. 17], by exploiting the relation

$$
ac_n(x+1, a) + (n - a - x)c_n(x, a) + xc_n(x - 1, a) = 0, \quad n \ge 0.
$$

(This approach was suggested to us by Ilia Krasikov.) □

We can use the $\boldsymbol{d}$'s from the proof of the previous theorem to obtain upper bounds for $\|\boldsymbol{p}(t) - \boldsymbol{\pi}\|$ by means of (38). This results in

$$
\|\boldsymbol{p}(t) - \boldsymbol{\pi}\| \le \begin{cases} C_1 e^{-\{N(\sqrt{\lambda} - \sqrt{\mu})^2 + 2\sqrt{\lambda\mu} - \mu\}t}, & \text{if} \ \lambda > \mu \\ C_2 e^{-\mu t}, & \text{if} \ \lambda < \mu \end{cases}, \quad t \ge 0,
\tag{54}
$$

where C_1 and C_2 are positive constants that are independent of N. These bounds slightly improve upon the bounds given in [10, Propositions 6 and 10].

Our final result concerns the asymptotic behaviour of χ_N as $N \to \infty$ in the setting of Theorem 13.

Theorem 14. Let $\lambda(N) = \lambda N$ with $\lambda > 0$, then

$$\lim_{N \to \infty} \frac{\chi_N}{N} = \left(\sqrt{\lambda} + \sqrt{\mu} \right)^2.$$

Proof. By choosing $d_i := (\mu/\lambda)^{i/2}$, $i = 1, 2, \ldots, N$, the quantities ζ_i of (37) satisfy

$$\zeta_i = \begin{cases} (\lambda + \sqrt{\lambda\mu})N + (\sqrt{\lambda\mu} + \mu)i - \sqrt{\lambda\mu}, & 1 \le i < N, \\ \lambda N + (\sqrt{\lambda\mu} + \mu)N - \sqrt{\lambda\mu}, & i = N, \end{cases}$$

so that, for N sufficiently large, $\zeta_{N-1} = \max_i\{\zeta_i\}$. By (36) we therefore have

$$\chi_N \le \zeta_{N-1} = (\sqrt{\lambda} + \sqrt{\mu})^2 N - 2\sqrt{\lambda\mu} - \mu. \tag{55}$$

On the other hand, choosing d_i as in (51) and (52), it follows after some algebra that the quantities ζ_i of (37) satisfy

$$\zeta_i \ge (\sqrt{\lambda} + \sqrt{\mu})^2 + o(N) \quad \text{as} \quad N \to \infty, \quad 1 \le i \le N.$$

so that, by the max-min representation in (36),

$$\chi_N \ge (\sqrt{\lambda} + \sqrt{\mu})^2 N + o(N) \quad \text{as} \quad N \to \infty. \tag{56}$$

In view of (55) and (56) the theorem follows. $\quad\square$

Acknowledgements

References

[1] J.P.C. Blanc and E.A. van Doorn, Relaxation times for queueing systems. pp. 139-162 in: *Mathematics and Computer Science,* Proceedings of the CWI symposium, 1983, J.W. de Bakker, M. Hazewinkel and J.K. Lenstra, eds. CWI Monograph 1, North-Holland, Amsterdam, 1986.

[2] M.F. Chen, Estimation of spectral gap for Markov chains. *Acta Math. Sinica (N.S.)* **12** (1996) 337–360.

[3] M.F. Chen, Variational formulas and approximation theorems for the first eigenvalue in dimension one. *Sci. China Ser. A* **44** (2001) 409–418.

[4] M.F. Chen, *Eigenvalues, Inequalities and Ergodic Theory.* Springer-Verlag, London, 2005.

[5] T.S. Chihara, *An Introduction to Orthogonal Polynomials.* Gordon and Breach, New York, 1978.

[6] G. Dahlquist, *Stability and Error Bounds in the Numerical Integration of Ordinary Differential Equations.* Inaugural dissertation, University of Stockholm, Almqvist & Wiksells Boktryckeri AB, Uppsala 1958. Reprinted in: Transactions of the Royal Institute of Technology, No. 130, Stockholm, 1959.

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

[8] E.A. van Doorn, Representations for the rate of convergence of birth-death processes. *Theory Probab. Math. Statist.* **65** (2002) 37–43.

[9] E.A. van Doorn, N.D. van Foreest and A.I. Zeifman, Representations for the extreme zeros of orthogonal polynomials. *J. Comput. Appl. Math.*, to appear.

[10] C. Fricker, P. Robert and D. Tibi, On the rates of convergence of Erlang's model. *J. Appl. Probab.* **36** (1999) 1167-1184.

[11] B.V. Gnedenko and I.P. Makarov, Properties of the solutions of a problem with losses in the case of periodic intensities. *Differ. Equ.* **7** (1971) 1696-1698. (In Russian.)

[12] B.L. Granovsky and A.I. Zeifman, The decay function of nonhomogeneous birth and death processes, with application to mean-field models. *Stochastic Process. Appl.* **72** (1997) 105–120.

[13] B.L. Granovsky and A.I. Zeifman, The N-limit of spectral gap of a class of birth-death Markov chains. *Appl. Stoch. Models Bus. Ind.* **16** (2000) 235–248.

[14] B.L. Granovsky and A. Zeifman, Nonstationary queues: estimation of the rate of convergence. *Queueing Syst.* **46** (2004) 363-388.

[15] B.L. Granovsky and A.I. Zeifman, On the lower bound of the spectrum of some mean-field models. *Theory Probab. Appl.* **49** (2005) 148–155.

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

[17] S. Karlin and J.L. McGregor, Ehrenfest urn models. *J. Appl. Probab.* **2** (1965) 352–376.

[18] M. Kijima, Evaluation of the decay parameter for some specialized birth-death processes. *J. Appl. Probab.* **29** (1992) 781–791.

[19] I. Krasikov, Bounds for zeros of the Charlier polynomials. *Methods Appl. Anal.* **9** (2002) 599-610.

[20] S.M. Lozinskiĭ, Error estimate for numerical integration of ordinary differential equations, I. *Izv. Vysš. Učebn. Zaved. Matematika* **5** (1958) 52-90. Errata, **5** (1959) 222. (In Russian.)

[21] F.J. MacWilliams and N.J.A. Sloane, *The Theory of Error-Correcting Codes, Part II.* North-Holland, Amsterdam, 1977.

[22] C.D. Meyer, *Matrix Analysis and Applied Linear Algebra.* SIAM: Society for Industrial and Applied Mathematics, 2001. (updates available on `http://www.matrixanalysis.com`).

[23] W. Rudin, *Functional Analysis.* McGraw-Hill, New York, 1973.

[24] G. Söderlind, The logarithmic norm. History and modern theory. *BIT. Numerical Mathematics* **46** (2006) 631-652.

[25] T. Ström, On logarithmic norms. *SIAM J. Numer. Anal.* **12** (1975) 741–753.

[26] L. Takács, *Introduction to the Theory of Queues.* Oxford University Press, New York, 1962.

[27] R.S. Varga, *Matrix Iterative Analysis,* 2nd rev. ed. Springer-Verlag, Berlin, 2000.

[28] M. Voit, A note on the rate of convergence to equilibrium for Erlang's model in the subcritical case. *J. Appl. Probab.* **37** (2000) 918-923.

[29] A.I. Zeifman, Some estimates of the rate of convergence for birth and death processes. *J. Appl. Probab.* **28** (1991) 268-277.

[30] A.I. Zeifman, On the estimation of probabilities for birth and death processes. *J. Appl. Probab.* **32** (1995) 623-634.

[31] A.I. Zeifman, Upper and lower bounds on the rate of convergence for non-homogeneous birth and death processes. *Stochastic Process. Appl.* **59** (1995) 157-173.

[32] A.I. Zeifman, V.E. Bening and I.A. Sokolov, *Markov Chains and Models in Continuous Time.* Elex-KM, Moscow, 2008. (In Russian.)