*Adv. Appl. Prob.* **17**, 514–530 (1985)
*Printed in N. Ireland*


# CONDITIONS FOR EXPONENTIAL ERGODICITY AND BOUNDS FOR THE DECAY PARAMETER OF A BIRTH–DEATH PROCESS

ERIK A. VAN DOORN, **Centre for Mathematics and Computer Science, Amsterdam*

**Abstract**

This paper is concerned with two problems in connection with exponential ergodicity for birth–death processes on a semi-infinite lattice of integers. The first is to determine from the birth and death rates whether exponential ergodicity prevails. We give some necessary and some sufficient conditions which suffice to settle the question for most processes encountered in practice. In particular, a complete solution is obtained for processes where, from some finite state $n$ onwards, the birth and death rates are rational functions of $n$. The second, more difficult, problem is to evaluate the decay parameter of an exponentially ergodic birth–death process. Our contribution to the solution of this problem consists of a number of upper and lower bounds.

DUAL BIRTH–DEATH PROCESSES; ORTHOGONAL POLYNOMIALS; SPECTRAL REPRESENTATION; TRANSITION PROBABILITIES

## 1. Introduction

Consider a standard, conservative Markov process in continuous time, whose state space $E \equiv \{0, 1, 2, \cdots\}$ constitutes an irreducible class. Let its (stationary) transition probabilities be denoted by $p_{ij}(t)$ $(i, j \in E,\ t \geqq 0)$. The transition $i \to j$ is then said to be *exponentially ergodic* if $p_{ij}(t)$ tends to its ergodic limit $p_j$ (independent of $i$ because of the irreducibility of $E$) exponentially fast, i.e., if there exists an $\alpha > 0$ such that

$$p_{ij}(t) - p_j = O(\exp(-\alpha t)) \tag{1.1}$$

as $t \to \infty$. We shall study the phenomenon of exponential ergodicity in the context of birth–death processes and thus continue the works of Callaert (1971), (1974) and Callaert and Keilson (1973a), (1973b). Our main tool will be Karlin and McGregor's (1957a) *spectral representation* for the transition probabilities of a birth–death process, which says that for this type of Markov

Received 26 November 1982; revision received 30 August 1984.
* Present address: Department of Applied Mathematics, Twente University of Technology, P.O. Box 217, 7500 AE Enschede, The Netherlands.

process

$$p_{ij}(t) = \pi_j \int_0^\infty \exp(-xt) Q_i(x) Q_j(x) \, d\psi(x) \tag{1.2}$$

($i, j \in E, t \geqq 0$). Here $\pi_j$ are constants and $\{Q_n\}$ is a system of polynomials properly normalized and orthogonal with respect to the mass distribution $d\psi$ which has its support $S(d\psi)$ on the non-negative real axis.

In view of (1.2), the basic results of Kingman (1963a), (1963b) on exponential ergodicity for Markov processes become transparent for birth–death processes. Let us first consider Kingman's (1963a) *solidarity theorem* for the transition probabilities of a transient or null-recurrent Markov process, which states that the supremum of the $\alpha$'s satisfying (1.1) is the same for each pair $i, j$, so that in particular either all or none of the $p_{ij}(t)$ go to their (zero) limits exponentially fast. This common supremal value is then called the *decay parameter* of the process; if it is positive the process itself is called *exponentially ergodic*. Now let $\gamma = \gamma(d\psi)$, where

$$\gamma(d\psi) \equiv \inf\{x \mid x > 0 \text{ and } x \in S(d\psi)\}, \tag{1.3}$$

and $d\psi$ the mass distribution associated with a transient or null-recurrent birth–death process. Then it is not difficult to show with (1.2) that for each pair $i, j$

$$p_{ij}(t) = O(\exp(-\gamma t)) \tag{1.4}$$

as $t \to \infty$. To establish that for any pair $i, j$ the factor $\gamma$ in (1.4) cannot be improved (i.e., enlarged) is somewhat more troublesome. Callaert (1971), (1974) uses a rather complicated argument involving theorems of Widder's on Stieltjes transforms and Laplace–Stieltjes transforms, but less sophisticated methods lead to the same conclusion. For it is clear that $d\psi$ cannot have an isolated point mass at 0, so that $\gamma$ must be the smallest point in $S(d\psi)$. A well-known theorem on zeros of orthogonal polynomials then tells us that $Q_n(\gamma) \neq 0$ for all $n$. Subsequently using a straightforward argument of the type on p. 105 of Van Doorn (1981) yields Callaert's result. Thus $\gamma(d\psi)$ is Kingman's decay parameter for a birth–death process with mass distribution $d\psi$ if the process is transient or null-recurrent.

If a birth–death process is positive recurrent, then the associated mass distribution $d\psi$ has positive mass at 0. Indeed, we have (Karlin and McGregor (1957b))

$$p_j = \pi_j \, d\psi(0) > 0 \tag{1.5}$$

($j \in E$). Since the $Q_n(x)$ are normalized such that $Q_n(0)=1$ for all $n$ in this case, it follows that instead of (1.2) we can write

$$p_{ij}(t) - p_j = \pi_j \int_{0+}^{\infty} \exp(-xt) Q_i(x) Q_j(x) \, d\psi(x). \tag{1.6}$$

A small complication now arises, which is also reflected in Kingman's (1963b) result for positive recurrent Markov processes. For again we have

$$p_{ij}(t) - p_j = O(\exp(-\gamma t)) \tag{1.7}$$

as $t \to \infty$, where $\gamma = \gamma(d\psi)$, but there may be pairs $i, j$ for which the factor $\gamma$ in (1.7) can be improved. This contingency is brought about when $d\psi$ has an isolated point mass at $\gamma$ and $Q_i(\gamma)=0$ or $Q_j(\gamma)=0$. This being an exceptional case (there is at most one $n$ such that $Q_n(\gamma)=0$), it is quite natural, indeed common practice, to call $\gamma$ the decay parameter of the process and the process exponentially ergodic if $\gamma > 0$. Proofs for the above statements (which are essentially Callaert's) may be given along the alternative lines sketched for the transient or null-recurrent case.

Summarizing, birth–death processes provide an illustrative example of Kingman's solidarity theorems for Markov processes in view of Karlin and McGregor's spectral representation (1.2) and Callaert's fundamental result (which can be given a relatively simple proof) that the decay parameter of a birth–death process equals $\gamma(d\psi)$, where $d\psi$ is the associated mass distribution.

Two obvious problems now arise in the context of birth–death processes, viz., (i) to give criteria for $\gamma(d\psi)$ to be positive in terms of the parameters which usually define a birth–death process (the birth and death rates), and more specifically (ii) to determine the value of $\gamma(d\psi)$ or at least bounds for $\gamma(d\psi)$ in terms of the rates. These are the problems to which this paper is addressed.

The plan of the paper is as follows. In Section 2 we formally introduce the necessary concepts and results related to birth–death processes and, in particular, to the spectral representation for their transition probabilities. In Section 3, which is the core of the paper, we give a characterization for the decay parameter of a birth–death process. Then, in Section 4, we shall obtain bounds for the decay parameter which are based on this characterization. Most of the preparatory work in this respect is done in a separate paper (Van Doorn (1984)) in which the more abstract terminology of orthogonal polynomials is used. Finally, problem (i) above will be tackled in Section 5. That is, we give conditions for a birth–death process to be exponentially ergodic. In particular we give the precise conditions for exponential ergodicity when, from some finite state $n$ onwards, the birth and death rates are rational functions of $n$.

## 2. Properties of birth–death processes

2.1. *Preliminaries.* A birth–death process on the set $E' \equiv \{-1, 0, 1, \cdots\}$, where $-1$ is an absorbing barrier and $E \equiv \{0, 1, \cdots\}$ constitutes an irreducible class, is faithfully represented by an array of functions $\{p_{ij}(t) \mid i, j \in E', t \geqq 0\}$ (the transition probabilities), satisfying the conditions

$$\sum_j p_{ij}(t) \leqq 1, \tag{2.1}$$

$$p_{ij}(t) \geqq 0, \tag{2.2}$$

$$p_{ij}(0) = \delta_{ij}, \tag{2.3}$$

$$p_{ij}(t+s) = \sum_k p_{ik}(t) p_{kj}(s), \tag{2.4}$$

$$p'_{ij}(t) = \sum_k a_{ik} p_{kj}(t), \tag{2.5}$$

$$p'_{ij}(t) = \sum_k p_{ik}(t) a_{kj}, \tag{2.6}$$

for $i, j \in E'$ and $t, s \geqq 0$. Here $a_{-1,j} = 0$ for all $j$, and, for $i \in E$,

$$a_{ij} = \begin{cases} \mu_i & \text{if } j = i-1 \\ -(\lambda_i + \mu_i) & \text{if } j = i \\ \lambda_i & \text{if } j = i+1 \\ 0 & \text{otherwise,} \end{cases} \tag{2.7}$$

where $\lambda_i$ and $\mu_i$, the *birth and death rates*, respectively, are positive with the exception $\mu_0 \geqq 0$. The *backward equations* (2.5) are equivalent to the more usual postulates

$$\begin{aligned} p_{i,i+1}(t) &= \lambda_i t + o(t) \\ p_{ii}(t) &= 1 - (\lambda_i + \mu_i)t + o(t) \\ p_{i,i-1}(t) &= \mu_i t + o(t) \end{aligned} \tag{2.8}$$

as $t \downarrow 0$, for $i \in E$. The *forward equations* (2.6) are not always encountered as a postulate. However, it has been shown by Karlin and McGregor (1959) that these equations must be satisfied in order that the sample paths of the process are continuous except for simple discontinuities with jump $\pm 1$, which we consider desirable.

It is well known that any set $\{\lambda_n, \mu_n\}_{n=0}^{\infty}$ of birth and death rates corresponds to at least one process $\{p_{ij}(t)\}$ satisfying (2.1)–(2.7). Karlin and McGregor (1957a) have shown that a set of rates $\{\lambda_n, \mu_n\}_n$ determines a process uniquely if and only if

$$\sum_{n=0}^{\infty} \{\pi_n + (\lambda_n \pi_n)^{-1}\} = \infty, \tag{2.9}$$

where

$$\pi_0 = 1; \qquad \pi_n = \frac{\lambda_0\lambda_1 \cdots \lambda_{n-1}}{\mu_1\mu_2 \cdots \mu_n}, \qquad n > 0. \tag{2.10}$$

Throughout this paper we shall only consider birth–death processes which are uniquely determined by their rates; it will be convenient to describe such processes as *simple.*

The initial condition (2.3) and the forward equations (2.6) imply

$$p_{i,-1}(t) = \mu_0 \int_0^t p_{i0}(\tau)\, d\tau \tag{2.11}$$

($i \in E$), while $p_{-1,j}(t) = \delta_{-1,j}$ for all $j$ and $t \geqq 0$ by (2.3) and the backward equations (2.5). Otherwise the transition probabilities involving the absorbing state $-1$ do not enter in an essential way in (2.2)–(2.7). Therefore, we might as well forget about $-1$ and represent a birth–death process by an array of functions $\{p_{ij}(t) \mid i, j \in E, t \geqq 0\}$ satisfying (2.2)–(2.7) and

$$\mu_0 \int_0^t p_{i0}(\tau)\, d\tau + \sum_j p_{ij}(t) \leqq 1 \tag{2.12}$$

for $i, j \in E$ and $t, s \geqq 0$, where all summations extend over $E$ instead of $E'$. This representation will be our starting point.

We should mention that Karlin and McGregor (1957a) postulate

$$\sum_j p_{ij}(t) \leqq 1 \tag{2.13}$$

instead of (2.12). However, by adapting Karlin and McGregor's Theorem 7, it can be shown that the set of postulates (2.2)–(2.7) and (2.12) is equivalent to the set (2.2)–(2.7) and (2.13), so that Karlin and McGregor's results carry over to the present context.

Karlin and McGregor (1957a) (for $\mu_0 = 0$) and Kemperman (1962) (for $\mu_0 \geqq 0$) have shown that a simple birth–death process is *honest*, i.e., equality holds in (2.12) for all $t$, if and only if the series

$$\sum_{n=0}^{\infty} (\lambda_n \pi_n)^{-1} \sum_{i=0}^{n} \pi_i \tag{2.14}$$

diverges.

2.2. *The spectral representation.* We shall now properly introduce Karlin and McGregor's (1957a) representation formula for the transition probabilities of a birth–death process. For a set $\{\lambda_n, \mu_n\}_n$ of birth and death rates the polynomials

$Q_n(x)$ are defined by

$$\text{(2.15)}\qquad \begin{aligned} &\lambda_n Q_{n+1}(x) = (\lambda_n + \mu_n - x)Q_n(x) - \mu_n Q_{n-1}(x), \qquad n > 0 \\ &\lambda_0 Q_1(x) = \lambda_0 + \mu_0 - x, \qquad Q_0(x) = 1. \end{aligned}$$

A set of polynomials satisfying a recurrence relation of this type is orthogonal with respect to a mass distribution $d\psi$ on $[0, \infty)$, viz.,

$$\text{(2.16)}\qquad \int_0^\infty Q_i(x)Q_j(x)\, d\psi(x) = \pi_j^{-1}\delta_{ij}.$$

Here $\pi_j$ is as in (2.10), and the distribution $d\psi$ is of total mass 1, has infinite support, and satisfies

$$\text{(2.17)}\qquad \mu_0 > 0 \ \Rightarrow \int_0^\infty x^{-1}\, d\psi(x) \leqq \mu_0^{-1}.$$

Our standard assumption is that we are dealing with a simple birth–death process, i.e., (2.9) is fulfilled, which makes that $d\psi$ is uniquely determined by the above-mentioned properties. The transition probabilities $p_{ij}(t)$ of the birth–death process with rates $\lambda_n$ and $\mu_n$ can now be represented as

$$\text{(2.18)}\qquad p_{ij}(t) = \pi_j \int_0^\infty \exp(-xt)Q_i(x)Q_j(x)\, d\psi(x)$$

($i, j \in E$, $t \geqq 0$), which is Karlin and McGregor's representation formula.

The following can also be observed from Karlin and McGregor (1957a). Consider a pair $(\mu_0, d\psi)$, where $\mu_0 \geqq 0$ and $d\psi$ is a mass distribution on $[0, \infty)$ of total mass 1, with infinite support and finite moments of all orders, satisfying (2.17), and uniquely determined by the polynomials $Q_n(x)$ which are orthogonal with respect to $d\psi$ (or, equivalently, by its moments). Then the $Q_n(x)$ satisfy, if properly normalized, a recurrence relation of the type (2.15) with $\lambda_n, \mu_{n+1} > 0$ $(n \geqq 0)$. Since $\mu_0$ is fixed, all parameters $\lambda_n$ and $\mu_n$ are uniquely determined and satisfy, in fact, (2.9), so that they are the birth and death rates of a simple birth–death process. Conversely, the mass distribution arising in the representation formula (2.18) for this latter birth–death process is of course the original distribution $d\psi$. Summarizing, we see that a simple birth–death process can be represented by its rates as well as by the pair $(\mu_0, d\psi)$, where $\mu_0$ is the death rate in state 0, and $d\psi$ the mass distribution in the representation formula (2.18).

2.3. *The spectrum.* We consider a simple birth–death process with rates $\lambda_n$ and $\mu_n$. $\{Q_n(x)\}_n$ denotes the set of polynomials and $d\psi$ the mass distribution associated with $\{\lambda_n, \mu_n\}_n$ through (2.15)–(2.17). We will state some properties of the *spectrum* $S$ of the process, which is defined as the support of the distribution $d\psi$, i.e.,

(2.19) $$S = S(d\psi) \equiv \left\{ x \,\middle|\, \int_{x-\varepsilon}^{x+\varepsilon} d\psi(\xi) > 0 \text{ for all } \varepsilon > 0 \right\}.$$

Starting off with some preliminaries, we write $P_0(x) = Q_0(x) = 1$ and

(2.20) $$P_n(x) = (-1)^n \lambda_0 \lambda_1 \cdots \lambda_{n-1} Q_n(x), \quad n > 0,$$

and note that the sequence $\{P_n(x)\}$ satisfies the relations

(2.21) $$\begin{aligned} P_{n+1}(x) &= (x - \lambda_n - \mu_n) P_n(x) - \lambda_{n-1}\mu_n P_{n-1}(x), \qquad n > 0 \\ P_1(x) &= x - \lambda_0 - \mu_0, \qquad P_0(x) = 1. \end{aligned}$$

The following properties of the zeros of $Q_n(x)$ can now be obtained from Chihara's (1978) book on orthogonal polynomials. For all positive $n$, $Q_n(x)$ has $n$ positive, distinct zeros $x_{n1} < x_{n2} < \cdots < x_{nn}$. These quantities satisfy

(2.22) $$x_{n+1,i} < x_{ni} < x_{n+1,i+1}$$

$(i = 1, 2, \cdots, n;\ n = 1, 2, \cdots)$, whence

(2.23) $$\xi_i \equiv \lim_{n\to\infty} x_{ni} \quad \text{and} \quad \eta_j \equiv \lim_{n\to\infty} x_{n,n-j+1}$$

$(i, j = 1, 2, \cdots)$ exist (possibly, $\eta_j = \infty$). Also,

(2.24) $$0 \leqq \xi_i \leqq \xi_{i+1} < \eta_{j+1} \leqq \eta_j \leqq \infty,$$

so that both

(2.25) $$\sigma \equiv \lim_{i\to\infty} \xi_i \quad \text{and} \quad \tau \equiv \lim_{j\to\infty} \eta_j$$

exist and $0 \leqq \sigma \leqq \tau \leqq \infty$. Furthermore, we have

(2.26) $$\xi_{i+1} = \xi_i \Rightarrow \sigma = \xi_i$$

$(i = 1, 2, \ldots)$, and

(2.27) $$\eta_{j+1} = \eta_j \ \Rightarrow \ \tau = \eta_j$$

$(j = 0, 1, 2, \cdots)$, where $\eta_0 \equiv \infty$. From the preceding results and, e.g., Theorems IV.3.1 and IV.3.3 of Chihara (1978), it is easy to see that

(2.28) $$\eta_1 < \infty \Leftrightarrow \tau < \infty \Leftrightarrow \sup\{\lambda_n + \mu_n\} < \infty.$$

We next define

(2.29) $$\Xi \equiv \{\xi_1, \xi_2, \xi_3, \cdots\} \quad \text{and} \quad \mathrm{H} \equiv \{\eta_1, \eta_2, \eta_3, \cdots\}$$

and note that both sets may be finite.

*Theorem* 2.1. The spectrum $S$ of a simple birth–death process with rates $\lambda_n$

and $\mu_n$ has the following properties:

(i) If $\sigma=\infty$, then $S=\Xi$.

(ii) If $\sigma<\infty$ and $\sup\{\lambda_n+\mu_n\}=\infty$, then $S=\bar{\Xi}\cup S_l$ (a bar denoting closure), where $S_l$ is an unbounded subset of $(\sigma,\infty)$; also, $\sigma$ is the smallest limit point of $S$.

(iii) If $\sigma<\infty$ and $\sup\{\lambda_n+\mu_n\}<\infty$, then $S=\bar{\Xi}\cup S_l\cup \bar{\mathrm{H}}$, where $S_l$ is a (possibly empty) subset of $(\sigma,\tau)$; also, $\sigma$ is the smallest and $\tau$ is the largest limit point of $S$.

*Proof.* The problem of finding a mass distribution $d\psi$ of total mass 1, with infinite support and finite moments of all orders, and with respect to which the polynomials $Q_n(x)$ of (2.15) are orthogonal, may be formulated as a Hamburger moment problem (HMP). If this moment problem is determined, i.e., has a unique solution, then the statements in (i), (ii) and (iii) concerning the support of this solution are known (Chihara (1978), Section II.4). Now if $\sigma<\infty$, then (Chihara (1978), Van Doorn (1985)) the HMP is determined, which settles (ii) and (iii).

Next suppose $\sigma=\infty$ and the HMP is indeterminate, i.e., has more than one solution. By Theorem 5 of Chihara (1968), there exists a solution $d\phi$ of the HMP whose support equals $\Xi$. On the other hand, since we are dealing with a simple birth–death process, there is only one solution of the HMP with support on $[0,\infty)$ and satisfying the additional condition (2.17). Therefore, we are done if we can show that Chihara's 'natural' solution $d\phi$ fulfills condition (2.17). Indeed, from the results of Sections 3 and 5 of Chihara (1982a), it is readily verified that

$$\int_0^\infty x^{-1}\,d\phi(x)=\{(\lambda_0+\mu_0)M_0\}^{-1},$$

where $M_0$ is the *maximal initial parameter* for the *chain sequence*

$$\left\{\left(1-\frac{\mu_{n-1}}{\lambda_{n-1}+\mu_{n-1}}\right)\frac{\mu_n}{\lambda_n+\mu_n}\right\}_{n=1}^{\infty}$$

(for definitions, see Chihara (1978), (1982a)). It follows that $\mu_0/(\lambda_0+\mu_0)\leqq M_0$, so that $\int_0^\infty x^{-1}\,d\phi(x)\leqq\mu_0^{-1}$.

2.4. *Duality.* As a final and essential prerequisite we must mention the duality concept for birth–death processes that was introduced by Karlin and McGregor (1957a), (1957b) and some of its consequences. For a set $\mathscr{P}=\{\lambda_n,\mu_n\}_n$ of birth and death rates, the dual set $\mathscr{P}^d=\{\lambda_n^d,\mu_n^d\}_n$ is defined by

$$\begin{aligned}&\mu_0=0\ \Rightarrow\ \lambda_n^d=\mu_{n+1},\qquad \mu_n^d=\lambda_n\\ &\mu_0>0\ \Rightarrow\ \mu_0^d=0,\qquad \lambda_n^d=\mu_n,\qquad \mu_{n+1}^d=\lambda_n\end{aligned}\tag{2.30}$$

($n=0, 1, \cdots$). Clearly, this duality concept establishes a one-to-one correspondence between the sets of rates where $\mu_0=0$ and those where $\mu_0>0$. The relations between the polynomials corresponding to $\mathcal{P}$ and $\mathcal{P}^d$ are readily found to be

$$\begin{aligned}\mu_0=0 &\Rightarrow Q_n^d(x)=\lambda_n\pi_n(Q_{n+1}(x)-Q_n(x))/(-x)\\ \mu_0>0 &\Rightarrow Q_{n+1}^d(x)=\lambda_n\pi_n(Q_{n+1}(x)-Q_n(x))/\mu_0.\end{aligned} \tag{2.31}$$

For $\mu_0=0$, the polynomials $Q_n^d$ are known as the kernel polynomials with parameter 0 corresponding to the polynomials $Q_n$ (cf. Chihara (1978)). There exists a separation theorem for the zeros of kernel polynomials (Chihara (1978), Theorem I.7.2) which is easily seen to lead to

$$\begin{aligned}\mu_0=0 &\Rightarrow \xi_i\leqq\xi_i^d\leqq\xi_{i+1}\\ \mu_0>0 &\Rightarrow \xi_i^d\leqq\xi_i\leqq\xi_{i+1}^d\end{aligned} \tag{2.32}$$

(cf. (2.23); see Van Doorn (1985) for more detailed results).

Regarding the parameters $\pi_n$ and $\pi_n^d$ associated with $\mathcal{P}$ and $\mathcal{P}^d$, respectively, we clearly have

$$\begin{aligned}\mu_0=0 &\Rightarrow \pi_n^d=\lambda_0(\lambda_n\pi_n)^{-1}\\ \mu_0>0 &\Rightarrow \pi_n^d=\mu_0(\mu_n\pi_n)^{-1}.\end{aligned} \tag{2.33}$$

Since $\lambda_n\pi_n=\mu_{n+1}\pi_{n+1}$, it follows that

$$\sum_{n=0}^{\infty}\{\pi_n+(\lambda_n\pi_n)^{-1}\}=\infty \Leftrightarrow \sum_{n=0}^{\infty}\{\pi_n^d+(\lambda_n^d\pi_n^d)^{-1}\}=\infty, \tag{2.34}$$

i.e., a birth–death process is simple if and only if its dual process is simple.

Finally turning to the mass distributions $d\psi$ and $d\psi^d$ associated with $\mathcal{P}$ and $\mathcal{P}^d$, respectively, we obtain from the preceding result and Lemmas 2 and 3 of Karlin and McGregor (1957a) that for a simple birth–death process

$$\begin{aligned}&\mu_0=0 \Rightarrow d\psi^d(x)=\lambda_0^{-1}x\,d\psi(x), \qquad x\geqq 0\\ &\mu_0>0 \Rightarrow d\psi^d(x)=\begin{cases}1-\mu_0\int_0^\infty x^{-1}\,d\psi(x) & x=0\\ \mu_0x^{-1}\,d\psi(x) & x>0.\end{cases}\end{aligned} \tag{2.35}$$

## 3. Representations for the decay parameter

Consider a simple birth–death process $\{p_{ij}(t)\mid i, j\in E, t\geqq 0\}$ and let $p_j$ denote the limit as $t\to\infty$ of $p_{ij}(t)$. The *decay parameter* $\alpha^*$ of this process is formally defined as

$$\alpha^* \equiv \sup \{\alpha \geqq 0 \mid p_{ij}(t) - p_j = O(e^{-\alpha t}) \text{ as } t \to \infty \text{ for all } i, j \in E\}, \tag{3.1}$$

and the process is said to be *exponentially ergodic* if $\alpha^* > 0$.

We have seen in Section 2.2 that the process $\{p_{ij}(t)\}$ can be represented by a pair $(\mu_0, d\psi)$, where $\mu_0$ is the death rate in state 0 and $d\psi$ the mass distribution appearing in Karlin and McGregor's representation formula for $p_{ij}(t)$. With $\gamma(d\psi)$ defined as in (1.3), we now have the following theorem, which, as mentioned in the introduction, is essentially due to Callaert (1971), (1974).

*Theorem* 3.1. The decay parameter of a simple birth–death process represented by $(\mu_0, d\psi)$ equals $\gamma(d\psi)$.

A birth–death process is usually not defined in terms of a pair $(\mu_0, d\psi)$, but in terms of a set of birth and death rates $\{\lambda_n, \mu_n\}_n$. What we seek, therefore, is to express $\gamma(d\psi)$ in the corresponding rates $\lambda_n$ and $\mu_n$. An intermediate step towards this goal is to relate $\gamma(d\psi)$ to the limit points $\xi_i$ of (2.23), which are uniquely determined by the birth and death rates via the polynomials $Q_n$ of (2.15).

*Lemma* 3.2. Let $d\psi$ and $\xi_i$, $i \geqq 1$, correspond to the same simple birth–death process, then $\gamma(d\psi) = \xi_2$ if $\xi_2 > \xi_1 = 0$, and $\gamma(d\psi) = \xi_1$ otherwise.

*Proof.* The cases $\xi_1 > 0$ and $\xi_2 > \xi_1 = 0$ follow immediately from Theorem 2.1. If $\xi_1 = \xi_2 = 0$, then, by (2.26), $\sigma = 0$. Since, by Theorem 2.1, $\sigma$ is the first limit point of the support of $d\psi$, it follows that $\gamma(d\psi) = 0$.

Although characterizations for $\xi_2$ can be given, it is much easier to work with $\xi_1$. This consideration leads us to bringing dual processes into our analysis as follows. First note that

$$\mu_0 > 0 \;\Rightarrow\; \xi_1 > 0 \quad \text{or} \quad \sigma = 0. \tag{3.2}$$

For, by (2.17), $d\psi$ has a finite moment of order $-1$ if $\mu_0 > 0$, so that there cannot be positive mass at 0; (3.2) then follows by Theorem 2.1. Consequently, by Lemma 3.2,

$$\mu_0 > 0 \Rightarrow \gamma(d\psi) = \xi_1. \tag{3.3}$$

On the other hand, from (2.35) we have

$$\gamma(d\psi) = \gamma(d\psi^d), \tag{3.4}$$

so that, by combining (3.3) in terms of the dual process and (3.4), we get

$$\mu_0 = 0 \Rightarrow \gamma(d\psi) = \xi_1^d. \tag{3.5}$$

Thus from the preceding results and Theorem 3.1 we conclude the following.

*Theorem* 3.3. The decay parameter of a simple birth–death process equals $\xi_1$ if $\mu_0>0$ and $\xi_1^d$ if $\mu_0=0$, where a superscript $d$ indicates the dual process.

We have now transformed the problem of finding the decay parameter of a simple birth–death process, given the birth and death rates, into that of finding the first limit point of the set of zeros of a sequence of orthogonal polynomials, given their recurrence relations. For the latter problem a substantial body of results (representations and bounds) is available (see Chihara (1978), (1982b), Van Doorn (1984) and references therein). In the next sections we shall apply some of these results.

## 4. Bounds for the decay parameter

We shall collect some bounds for the decay parameter $\alpha^*$ of a simple birth–death process with rates $\lambda_n$ and $\mu_n$. Throughout this section we shall assume $\mu_0=0$, which covers the most interesting case. It will be clear from the results of the previous section that bounds for the decay parameter of a process where $\mu_0>0$ may be obtained from the expressions given below by considering the dual process.

From Theorem 3.3 we see that $\alpha^*$ equals $\xi_1^d$, the first limit point of the set of zeros of the sequence of orthogonal polynomials $Q_n^d(x)$. By analogy with (2.20)–(2.21), we write $P_0^d(x)=Q_0^d(x)=1$ and

$$P_n^d(x)=(-1)^n\mu_1\mu_2\cdots\mu_nQ_n^d(x), \tag{4.1}$$

$n>0$, and find that

$$\begin{aligned}P_{n+1}^d(x)&=(x-\lambda_n-\mu_{n+1})P_n^d(x)-\lambda_n\mu_nP_{n-1}^d(x), \qquad n>0,\\ P_1^d(x)&=x-\lambda_0-\mu_1, \qquad P_0^d(x)=1,\end{aligned} \tag{4.2}$$

which brings us in a position to apply the results of Van Doorn (1984). The most general bounds for $\xi_1^d$ given in that paper involve an infinite number of free parameters. Here we shall mention the bounds that result from some obvious choices for these parameter sequences. Thus from Theorem 7 ($\chi_n=\sqrt{\lambda_n\mu_n}$) and Theorem 8 ($\beta_n=\frac{1}{4}$) of the aforementioned paper we obtain the following lower bounds.

*Theorem* 4.1. The decay parameter $\alpha^*$ of a simple birth–death process with rates $\lambda_n$ and $\mu_n$ ($\mu_0=0$) satisfies

(i) $$\alpha^*\geqq\inf_{n\geqq 1}\{\lambda_{n-1}+\mu_n-\sqrt{\lambda_{n-1}\mu_{n-1}}-\sqrt{\lambda_n\mu_n}\},$$

(ii) $$\alpha^*\geqq\inf_{n\geqq 1}\tfrac{1}{2}\{\lambda_{n-1}+\lambda_n+\mu_n+\mu_{n+1}-\sqrt{(\lambda_n+\mu_{n+1}-\lambda_{n-1}-\mu_n)^2+16\lambda_n\mu_n}\}.$$

We also mention two upper bounds for $\alpha^*$, which follow from Theorem 4 ($\chi_n = 1$) and Corollary 4.3 of the said paper.

*Theorem* 4.2. The decay parameter $\alpha^*$ of a simple birth–death process with rates $\lambda_n$ and $\mu_n$ ($\mu_0 = 0$) satisfies

$$\text{(i)} \qquad \alpha^* < \left\{1 + \sum_{i=n+1}^{n+k} \left(1 - 2\left(\frac{\lambda_i\mu_i}{(\lambda_{i-1} + \mu_i)(\lambda_i + \mu_{i+1})}\right)^{\frac{1}{2}}\right)\right\}\left\{\sum_{i=n}^{n+k} \frac{1}{\lambda_i + \mu_{i+1}}\right\}^{-1}, \qquad n, k = 0, 1, \cdots,$$

$$\text{(ii)} \qquad \alpha^* < \tfrac{1}{2}\{\lambda_{n-1} + \lambda_n + \mu_n + \mu_{n+1} - \sqrt{(\lambda_n + \mu_{n+1} - \lambda_{n-1} - \mu_n)^2 + 4\lambda_n\mu_n}\}, \qquad n = 1, 2, \cdots.$$

We illustrate the potency of these theorems with two examples.

*Example* 1. Lindvall (1979) considers a birth–death process with rates satisfying

$$\inf_{n \geqq 0} \{\lambda_n + \mu_n\} = c_0 > 0, \qquad \sup_{n \geqq 1} \{\lambda_n/(\lambda_n + \mu_n)\} = c_1 > \tfrac{1}{2} \tag{4.3}$$

and states that the decay parameter of this process is at least as large as that of the birth–death process with rates

$$\lambda_0 = c_0; \qquad \lambda_n = c_0c_1, \qquad \mu_n = c_0(1 - c_1), \qquad n > 0. \tag{4.4}$$

To obtain a lower bound for the decay parameter of the original process he then calculates the decay parameter $\alpha^*$ of the process with rates (4.4). With the previous results this calculation becomes very simple, for by Theorem 4.1(i) and Theorem 4.2(i) ($n = 0$, $k \to \infty$) it follows that $c \equiv c_0(1 - 2\sqrt{c_1(1 - c_1)})$ is both a lower bound and an upper bound for $\alpha^*$, whence $\alpha^* = c$.

*Example* 2. Consider the queue-length process of the *M/M/s* queue, which is a birth–death process with rates

$$\lambda_n = \lambda, \qquad \mu_n = \mu \min\{n, s\}, \qquad n \geqq 0, \tag{4.5}$$

where $\lambda, \mu > 0$ and $s$ is a positive integer. The decay parameter $\alpha^*$ of this process can be calculated as follows (Van Doorn (1981), Theorem 6.2.13). Let $\rho \equiv \lambda/s\mu$ denote the traffic intensity and define

$$C(x) = \tfrac{1}{2}(1 - x + \rho^{-1} - \sqrt{(1 - x + \rho^{-1})^2 - 4\rho^{-1}}), \tag{4.6}$$

and

$$\begin{aligned} R_{n+1}(x, y) &= 1 - x + n(sy)^{-1} - n(syR_n(x, y))^{-1}, \qquad n = 1, 2, \cdots, s-1 \\ R_1(x, y) &= 1 - x. \end{aligned} \tag{4.7}$$

Also, let $\rho^*$ be the largest root $<1$ of the equation

$$R_s(1 - y^{-\frac{1}{2}}, y) = y^{-\frac{1}{2}} \tag{4.8}$$

if $s>1$, and $\rho^*=0$ if $s=1$. Then, if $\rho \geqq \rho^*$,

$$\alpha^* = (\sqrt{\lambda} - \sqrt{s\mu})^2, \tag{4.9}$$

whereas for $\rho<\rho^*$, $\alpha^*$ equals $\lambda$ times the smallest positive root of the equation

$$R_s(x, \rho) = C(x). \tag{4.10}$$

If, instead of executing this complicated scheme, one chooses the relatively simple approach of calculating lower and upper bounds as given in Theorems 4.1 and 4.2, one can often find satisfactory and even exact results. This can be seen from Table 1, where we have given the results of all calculations for $s=10$, $\mu=1$ and various values for the traffic intensity.

We finally notice that for $\rho=0\cdot05$ we can apply Lindvall's argument of the previous example. The lower bound thus obtained is, however, inferior to those given in Table 1.

In closing this section we remark that Bordes and Roehner (1983) give a lower bound for the decay parameter of a birth–death process in the case that either the series (2.14) or the series

$$\sum_{n=0}^{\infty} (\lambda_n \pi_n)^{-1} \sum_{i=n+1}^{\infty} \pi_i \tag{4.11}$$

converges. (Note that convergence of *both* series is equivalent to convergence

TABLE 1

The exact value and bounds for the decay parameter $\alpha^*$ of the queue-length process of the $M/M/10$ queue with service intensity 1 and arrival intensity $\lambda$ ($\rho^*=0\cdot498$)

| | | lower bounds (Theorem 4.1) | | upper bounds (Theorem 4.2) | |
|---|---|---|---|---|---|
| $\rho=\lambda/10$ | $\alpha^*$ | (i) | (ii) | (i) | (ii) |
| 0·05 | 1·000 | 0·793 | 0·500 | 1·178 | 1·134 |
| 0·1 | 1·000 | 0·586 | 0·438 | 1·241 | 1·382 |
| 0·2 | 1·000 | 0·551 | 0·469 | 1·377 | 2·000 |
| 0·3 | 0·998 | 0·536 | 0·479 | 1·515 | 2·697 |
| 0·4 | 0·984 | 0·528 | 0·484 | 1·351 | 3·438 |
| 0·5 | 0·858 | 0·523 | 0·488 | 0·858 | 4·209 |
| 0·6 | 0·508 | 0·508 | 0·490 | 0·508 | 5·000 |
| 0·7 | 0·267 | 0·267 | 0·267 | 0·267 | 5·752 |
| 0·8 | 0·111 | 0·111 | 0·111 | 0·111 | 6·469 |
| 0·9 | 0·026 | 0·026 | 0·026 | 0·026 | 7·228 |
| 1·0 | 0·000 | <0 | 0·000 | 0·000 | 8·000 |
| 1·1 | 0·024 | <0 | 0·024 | 0·024 | 8·734 |
| 1·2 | 0·091 | <0 | 0·091 | 0·091 | 9·497 |
| 1·3 | 0·196 | <0 | 0·196 | 0·196 | 10·235 |
| 1·4 | 0·336 | <0 | 0·336 | 0·336 | 11·000 |
| 1·5 | 0·505 | 0·134 | 0·505 | 0·505 | 11·738 |

of the series in (2.9).) It is easy to see that the simple birth–death processes to which Bordes and Roehner's bounds apply can be identified as those which are not honest themselves or have dishonest duals.

## 5. Conditions for exponential ergodicity

A birth–death process is called exponentially ergodic if and only if its decay parameter is positive. Hence, by Theorem 3.1, a process represented by the pair $(\mu_0, d\psi)$ is exponentially ergodic if and only if $\gamma(d\psi) > 0$. An equivalent definition is given in the following simple but useful theorem, where $\sigma$ is the quantity defined in (2.25).

*Theorem* 5.1. A simple birth–death process with rates $\lambda_n$ and $\mu_n$ is exponentially ergodic if and only if $\sigma > 0$.

*Proof.* Follows immediately from Theorem 2.1.

Of course, the question of whether a process is exponentially ergodic can be phrased in terms of $\xi_1$ and $\xi_1^d$, cf. Theorem 3.3. However, formulating it in terms of $\sigma$ has certain advantages. First, we need not distinguish between the cases $\mu_0 = 0$ and $\mu_0 > 0$. Secondly, representations and bounds for $\sigma$ are generally simpler and more powerful than those for $\xi_1$, the reason being the following. Consider a set of rates $\{\lambda_n, \mu_n\}_n$ and define new sets $\{\lambda_n^{(k)}, \mu_n^{(k)}\}_n$, $k = 1, 2, \cdots$, by

$$\lambda_n^{(k)} = \lambda_{n+k}, \quad \mu_n^{(k)} = \mu_{n+k}. \tag{5.1}$$

From Theorem III.4.2 of Chihara (1978) we then obtain

$$\sigma^{(k)} = \sigma, \tag{5.2}$$

$k = 1, 2, \cdots$, where the notation should be clear. It follows that any finite number of changes in the rates of a simple birth–death process does not affect the value of $\sigma$, i.e., $\sigma$ (and hence the prevalence of exponential ergodicity) is determined only by the limiting behaviour of the birth and death rates.

In view of Theorem 5.1, necessary and/or sufficient conditions for exponential ergodicity of a process, given its rates, can be obtained from representations and bounds for $\sigma$. By using (2.20)–(2.21) a large number of such results follow from Chihara (1978), (1982b) and Van Doorn (1984). Here we shall just mention two bounds from the latter paper that will be of use later on.

*Theorem* 5.2. Let $\lambda_n$ and $\mu_n$ be the birth and death rates for a simple birth–death process. Then we have

(i) $$\sigma \geqq \liminf_{n\to\infty} \{\lambda_n + \mu_n - \sqrt{\lambda_{n-1}\mu_n} - \sqrt{\lambda_n\mu_{n+1}}\},$$

(ii) $$\sigma \leqq \liminf_{k\to\infty} \left\{\frac{1}{k}\sum_{n=0}^{k}(\lambda_n+\mu_n-2\sqrt{\lambda_{n-1}\mu_n})\right\}.$$

We next show that by using the above bounds and some other results from the literature, it is possible to calculate $\sigma$ exactly for a large class of processes, which includes the processes which have asymptotically rational birth and death rates. Thus, in view of Theorem 5.1, we can decide on the prevalence of exponential ergodicity for the vast majority of processes encountered in practice.

*Theorem* 5.3. Let $\lambda_n$ and $\mu_n$ be the birth and death rates for a simple birth–death process.

(i) If $\lambda_n \to \lambda \leqq \infty$, $\mu_n \to \mu \leqq \infty$ as $n\to\infty$, then

$$\sigma = (\sqrt{\lambda}-\sqrt{\mu})^2 \quad \text{if} \quad \lambda<\infty \quad \text{and} \quad \mu<\infty$$

and

$$\sigma=\infty \quad \text{if} \quad \lambda<\infty, \mu=\infty \quad \text{or} \quad \lambda=\infty, \mu<\infty.$$

(ii) If $\lambda_n = an^p + o(n^p)$, $\mu_n = dn^q + o(n^q)$ as $n\to\infty$, where $p, q>0$, then

$$\sigma=\infty \quad \text{if} \quad p\neq q \quad \text{or} \quad a\neq d.$$

(iii) If $\lambda_n = an^p + bn^{p-1} + o(n^{p-1})$, $\mu_n = an^p + en^{p-1} + o(n^{p-1})$ as $n\to\infty$, where $p$ is a natural number, then

$$\sigma = 0 \quad \text{if} \quad p=1$$

and

$$\sigma=\infty \quad \text{if} \quad p\geqq 3.$$

(iv) If $\lambda_n = an^2+bn+c+o(1)$, $\mu_n = an^2+en+f+o(1)$ as $n\to\infty$, then

$$\sigma = \frac{1}{4a}(a+e-b)^2.$$

*Proof.* The bounds given in the previous theorem readily yield (i) and (ii). Using (2.20)–(2.21) and after some algebra, the remaining cases can be reduced to situations analysed by Chihara (1982b).

*Remarks.* The first assertion in (i) follows also from a classical result known as Blumenthal's theorem (Chihara (1978)). Statement (ii) was proven earlier by Maki (1976) for natural $p$ and $q$.

Processes of the type described in (iv) have recently been studied in some detail by Roehner and Valent (1982) and Letessier and Valent (1984).

We conclude with some miscellaneous conditions for exponential ergodicity. First we cite Callaert and Keilson's (1973b) result to the effect that a birth–death process with rates $\lambda_n$ and $\mu_n$ is exponentially ergodic if

(5.3) $$\liminf_{n\to\infty}\{\lambda_n+\mu_n\}>0 \quad \text{and} \quad \limsup_{n\to\infty}\left\{\frac{\lambda_{n-1}\mu_n}{(\lambda_{n-1}+\mu_{n-1})(\lambda_n+\mu_n)}\right\}<\frac{1}{4}.$$

This criterion provides an alternative proof for the fact that the processes of Theorem 5.3(i) (with $\lambda \neq \mu$) and (ii) (with $p \neq q$ or $a \neq d$) are exponentially ergodic.

Another result of Callaert and Keilson's (1973b) (see also Tweedie (1981)) is that convergence of one of the series (2.14) and (4.11) is sufficient for exponential ergodicity. Indeed, it can be shown (cf. Van Doorn (1985)) that convergence of one (or both) of these series is equivalent to convergence of the series $\sum_{i=2}^{\infty} \xi_i^{-1}$, which, of course, implies that $\sigma = \infty$.

Finally, it can be shown from results of Karlin and McGregor (1957a), (1957b) and Theorem 2.1 that

$$\sum_{n=0}^{\infty} \pi_n = \sum_{n=0}^{\infty} (\lambda_n \pi_n)^{-1} = \infty \Rightarrow \sigma = 0, \tag{5.4}$$

while the first statement in (5.4) is equivalent to null-recurrence of the pertinent process. It follows that a necessary condition for exponential ergodicity of a birth–death process is that it is either transient or positive recurrent.

## References


BORDES, G. AND ROEHNER, B. (1983) Application of Stieltjes theory for $S$-fractions to birth and death processes. *Adv. Appl. Prob.* **15,** 507–530.

CALLAERT, H. (1971) Exponential Ergodicity for Birth–Death Processes (in Dutch). Ph.D. Thesis, University of Louvain.

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

CALLAERT, H. AND KEILSON, J. (1973a) On exponential ergodicity and spectral structure for birth–death processes. I. *Stoch. Proc. Appl.* **1,** 187–216.

CALLAERT, H. AND KEILSON, J. (1973b) On exponential ergodicity and spectral structure for birth–death processes II. *Stoch. Proc. Appl.* **1,** 217–235.

CHIHARA, T. S. (1968) On indeterminate Hamburger moment problems. *Pacific J. Math.* **27,** 475–484.

CHIHARA, T. S. (1978) *An Introduction to Orthogonal Polynomials.* Gordon and Breach, New York.

CHIHARA, T. S. (1982a) Indeterminate symmetric moment problems. *J. Math. Anal. Appl.* **85,** 331–346.

CHIHARA, T. S. (1982b) Spectral properties of orthogonal polynomials on unbounded sets. *Trans. Amer. Math. Soc.* **270,** 623–639.

KARLIN, S. AND MCGREGOR, J. L. (1957a) The differential equations of birth-and-death processes and the Stieltjes moment problem. *Trans. Amer. Math. Soc.* **85,** 489–546.

KARLIN, S. AND MCGREGOR, J. L. (1957b) The classification of birth and death processes. *Trans. Amer. Math. Soc.* **86,** 366–400.

KARLIN, S. AND MCGREGOR, J. L. (1959) A characterization of birth and death processes. *Proc. Nat. Acad. Sci. U.S.A.* **45,** 375–379.

KEMPERMAN, J. H. B. (1962) An analytical approach to the differential equations of the birth-and-death process. *Michigan Math. J.* **9,** 321–361.

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

KINGMAN, J. F. C. (1963b) Ergodic properties of continuous-time Markov processes and their discrete skeletons. *Proc. London Math. Soc.* **13,** 593–604.

LETESSIER, J. AND VALENT, G. (1984) The generating function method for quadratic asymptotically symmetric birth and death processes. *SIAM J. Appl. Math.* **44,** 773–783.

LINDVALL, T. (1979) A note on coupling of birth and death processes. *J. Appl. Prob.* **16,** 505–512.

MAKI, D. P. (1976) On birth-death processes with rational growth rates. *SIAM J. Math. Anal.* **7,** 29–36.

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.

TWEEDIE, R.L. (1981) Criteria for ergodicity, exponential ergodicity and strong ergodicity of Markov processes. *J. Appl. Prob.* **18,** 122–130.

VAN DOORN, E. A. (1981) *Stochastic Monotonicity and Queueing Applications of Birth–Death Processes.* Lecture Notes in Statistics **4,** Springer-Verlag, New York.

VAN DOORN, E. A. (1984) On oscillation properties and the interval of orthogonality of orthogonal polynomials. *SIAM J. Math. Anal.* **15,** 1031–1042.

VAN DOORN, E. A. (1985) On orthogonal polynomials with positive zeros and the associated kernel polynomials. *J. Math. Anal. Appl.*