Representations for the rate of convergence of birth-death processes

Erik A. van Doorn

Faculty of Mathematical Sciences
University of Twente
P.O. Box 217, 7500 AE Enschede, The Netherlands
E-mail: doorn@math.utwente.nl

May 4, 2001

Abstract. We display some representations for the rate of convergence of a birth-death process, which are useful for obtaining upper and lower bounds. The expressions are brought to light by exploiting the spectral representation for the transition probabilities of a birth-death process and results from the theory of orthogonal polynomials.

Keywords and phrases: birth-death process, decay parameter, duality, orthogonal polynomials, rate of convergence, spectral measure

2000 Mathematics Subject Classification: Primary 60J80

1 Introduction

In a recent paper Kartashov [14] has shown that the rate of convergence to stationarity of an ergodic birth-death process on the nonnegative integers with birth rates λ_n and death rates μ_n is bounded below by the quantity

$$\sup_u \inf_{n\geq 0} \{\lambda_n + \mu_{n+1} - \frac{\lambda_n}{u_n} - \mu_{n+1}u_{n+1}\}, \tag{1.1}$$

where $u \equiv (u_0, u_1, \ldots)$ is any sequence of positive numbers. It is the purpose of this note to point out that, actually, the rate of convergence equals (1.1). Moreover, we will exhibit two other representations. Similar results will be presented for the rate of convergence to zero of the transition probabilities of a birth-death process which is not ergodic.

After introducing some notation and terminology, and some preliminary results in Section 2, the representations are given in Section 3. We show in Section 4 how bounds for the rate of convergence may be obtained from the representations, and we display some of these bounds. Section 5 contains some concluding remarks.

2 Preliminaries

In what follows $\mathcal{X} \equiv \{X(t),\ t \geq 0\}$ is a birth-death process taking values in $\mathcal{N} \equiv \{0, 1, \ldots\}$ with birth rates $\{\lambda_n,\ n \in \mathcal{N}\}$ and death rates $\{\mu_n,\ n \in \mathcal{N}\}$, all strictly positive except $\mu_0 \geq 0$. When $\mu_0 > 0$ the process may evanesce by escaping from $\mathcal{N}$, via state 0, to an absorbing state -1, say. We will assume that $\mathcal{X}$ is uniquely determined by its birth and death rates, which is equivalent to assuming

$$\sum_{n\in\mathcal{N}} \theta_n + \sum_{n\in\mathcal{N}} (\lambda_n\theta_n)^{-1} = \infty, \tag{2.1}$$

where

$$\theta_0 \equiv 1 \quad \text{and} \quad \theta_n \equiv \frac{\lambda_0\lambda_1\ldots\lambda_{n-1}}{\mu_1\mu_2\ldots\mu_n}, \quad n \geq 1.$$

(see Karlin and McGregor [12]).

It is well known that the transition probabilities

$$p_{ij}(t) \equiv \Pr\{X(t) = j \mid X(0) = i\}, \quad i, j \in \mathcal{N},\ t \geq 0,$$

have limits

$$\pi_j \equiv \lim_{t\to\infty} p_{ij}(t) = \begin{cases} \dfrac{\theta_j}{\sum_n \theta_n} & \text{if } \mu_0 = 0 \text{ and } \sum_{n=0}^{\infty} \theta_n < \infty \\ 0 & \text{otherwise,} \end{cases}$$

which are independent of the initial state i. Our interest focuses on the exponential rate of convergence of $p_{ij}(t)$ to its limit π_j, that is, on the quantities

$$\alpha_{ij} \equiv -\lim_{t\to\infty} \frac{1}{t} \log |p_{ij}(t) - \pi_j|, \quad i, j \in \mathcal{N}. \tag{2.2}$$

From Callaert [3] (see also [9]) we know that these limits exist, and that

$$\alpha \equiv \alpha_{00} \leq \alpha_{ij}, \quad i, j \in \mathcal{N}, \tag{2.3}$$

with equality whenever $\mu_0 > 0$, and inequality prevailing for at most one value of i or j when $\mu_0 = 0$. The quantity α is therefore indicative of the speed of convergence to stationarity of the process $\mathcal{X}$, and we refer to it as the *rate of convergence of* $\mathcal{X}$. Sometimes α is called the *decay parameter* of $\mathcal{X}$ (see, for example, [3], [9] and [16]), although this term was originally used by Kingman [17] to denote the quantity

$$-\lim_{t\to\infty} \frac{1}{t} \log p_{ij}(t),$$

which is independent of i and j, and differs from α when $\mu_0 = 0$ and $\sum_n \theta_n < \infty$.

It has been shown by Chen [4] (see also [5, Chapter 9]) that if $\mathcal{X}$ is ergodic, that is, if $\mu_0 = 0$ and $\sum_n \theta_n < \infty$, then α can be expressed as

$$\alpha = -\lim_{t\to\infty} \frac{1}{t} \log \|P(t) - \Pi\|, \tag{2.4}$$

where $(P(t)) \equiv (p_{ij}(t))_{i,j\in\mathcal{N}}$, $\Pi \equiv (\pi_j)_{i,j\in\mathcal{N}}$, and the matrix norm is the one induced by the vector norm

$$\|v\|^2 \equiv \sum_{n\in\mathcal{N}} \pi_n v_n^2, \quad v \equiv (v_n,\ n \in \mathcal{N}).$$

This representation has been the starting point of Kartashov's analysis in [14] leading to (1.1).

To justify our expressions for α in the next section we must recall Karlin and McGregor's [12] spectral representation for the transition probabilities $p_{ij}(t)$. Namely, we can write

$$p_{ij}(t) = \theta_j \int_0^\infty e^{-xt} Q_i(x) Q_j(x) \psi(dx), \quad t \geq 0, \; i, j \in \mathcal{N}, \tag{2.5}$$

where $\{Q_n(x)\}$ is a sequence of polynomials satisfying the recurrence relation

$$\begin{aligned} &\lambda_n Q_{n+1}(x) = (\lambda_n + \mu_n - x) Q_n(x) - \mu_n Q_{n-1}(x), \quad n > 1, \\ &Q_0(x) = 1, \quad \lambda_0 Q_1(x) = \lambda_0 + \mu_0 - x, \end{aligned} \tag{2.6}$$

and ψ – the *spectral measure* of $\mathcal{X}$ – is the (unique) measure of total mass 1 with respect to which the polynomials $\{Q_n(x)\}$ are orthogonal, and whose *support*,

$$\text{supp}(\psi) \equiv \{x \in \mathbb{R} \mid \psi\{(x - \epsilon, x + \epsilon)\} > 0 \;\text{ for all }\; \epsilon > 0\},$$

is a subset of the nonnegative real axis.

If $\mu_0 = 0$ we shall have use for a *dual* process $\mathcal{X}^d$, which is the birth-death process taking values in $\mathcal{N}$ with birth rates $\{\lambda_n^d \equiv \mu_{n+1}, \; n \in \mathcal{N}\}$ and death rates $\{\mu_n^d \equiv \lambda_n, \; n \in \mathcal{N}\}$ (so that $\mu_0^d > 0$). Associated with $\mathcal{X}^d$ are polynomials $\{Q_n^d(x)\}$ satisfying a recurrence relation similar to (2.6) and a spectral measure ψ^d. Of particular interest to us is the fact that ψ^d is related to the spectral measure ψ of $\mathcal{X}$ by

$$\lambda_0 \psi^d([0, x]) = \int_0^x u \psi(du), \quad x \geq 0, \tag{2.7}$$

(see [9]) so that the supports of both measures are identical, unless 0 is an isolated point in $\text{supp}(\psi)$, in which case 0 is *not* contained in $\text{supp}(\psi^d)$.

3 Representations

The representation formula (2.5) shows that

$$p_{00}(t) - \pi_0 = \int_{0+}^\infty e^{-xt} \psi(dx), \quad t \geq 0, \tag{3.1}$$

since we must have $\psi(\{0\}) = \pi_0$. It follows (see [3], [9]) that

$$\alpha \equiv \alpha_{00} = \inf\{x \mid x > 0 \;\text{ and }\; x \in \text{supp}(\psi)\}, \tag{3.2}$$

so that $\alpha = \min \text{supp}(\psi)$, unless 0 is an isolated point in $\text{supp}(\psi)$. The latter situation can occur only if $\psi(\{0\}) = \pi_0 > 0$, which, in turn, can occur only if $\mu_0 = 0$. Then, however, we can bring the dual process into play. Namely, if $\mu_0 = 0$, then (2.7) allows us to conclude that $\alpha = \min \text{supp}(\psi^d)$. Summarizing we have the following result, noted earlier in [9].

Theorem 3.1 *The rate of convergence α of the birth-death process $\mathcal{X}$ satisfies*

$$\alpha = \begin{cases} \min \text{supp}(\psi^d) & \text{if} \quad \mu_0 = 0 \\ \min \text{supp}(\psi) & \text{if} \quad \mu_0 > 0. \end{cases} \tag{3.3}$$

It is well known (see, for example, Chihara [7]) that zeros of orthogonal polynomials are real and that the smallest point in the support of the orthogonalizing measure for a sequence of orthogonal polynomials equals the limit as $n \to \infty$ of the smallest zero of the nth degree polynomial. In [10] several expressions for this limit are given in terms of the parameters in the three-terms recurrence relation satisfied by the polynomials. To apply these results in the present setting we observe that the *monic* polynomials

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

satisfy the recurrence

$$\begin{aligned} &P_n(x) = (x - \lambda_{n-1} - \mu_{n-1}) P_{n-1}(x) - \lambda_{n-2}\mu_{n-1} P_{n-2}(x), \quad n > 1, \\ &P_0(x) = 1, \quad P_1(x) = x - \lambda_0 - \mu_0, \end{aligned} \tag{3.4}$$

while, if $\mu_0 = 0$, the monic polynomials corresponding to the dual process

$$P_n^d(x) \equiv (-1)^n \lambda_0^d \lambda_1^d \ldots \lambda_{n-1}^d Q_n^d(x) = (-1)^n \mu_1 \mu_2 \ldots \mu_n Q_n^d(x), \quad n > 0,$$

satisfy the recurrence

$$\begin{aligned} &P_n^d(x) = (x - \lambda_{n-1} - \mu_n) P_{n-1}^d(x) - \lambda_{n-1}\mu_{n-1} P_{n-2}^d(x), \quad n > 1, \\ &P_0^d(x) = 1, \quad P_1^d(x) = x - \lambda_0 - \mu_1. \end{aligned} \tag{3.5}$$

We can now substitute the coefficients in the recurrence relations (3.5) and (3.4) in the expressions given in the Theorems 2, 6 and 9 of [10], and obtain, in view of Theorem 3.1, expressions for α in the settings $\mu_0 = 0$ and $\mu_0 > 0$, respectively. Before stating the results we recall that a sequence $v \equiv (v_1, v_2, \ldots)$

is a *chain sequence* if each v_n can be written as $v_n = (1-g_{n-1})g_n$, with $0 \le g_0 < 1$ and $0 < g_n < 1$ for $n > 0$ (see Chihara [7]). For example, $v = (\frac{1}{4}, \frac{1}{4}, \ldots)$ is a chain sequence, since we can choose $g_n = \frac{1}{2},\ n \ge 0$. By convention, $\lambda_{-1} \equiv 0$.

Theorem 3.2 *If $\mu_0 = 0$, then the rate of convergence α of the birth-death process $\mathcal{X}$ can be represented as*

(i) $\alpha = \max_v \left\{ \inf_{n\ge 0} \left\{ \lambda_n + \mu_{n+1} - \frac{\lambda_n \mu_n}{v_n} - v_{n+1} \right\} \right\},$

where $v \equiv (v_0, v_1, \ldots)$ is any sequence of positive numbers;

(ii) $$\alpha = \max_v \left\{ \inf_{n\ge 1} \frac{1}{2} \left\{ \lambda_{n-1} + \mu_n + \lambda_n + \mu_{n+1} - \sqrt{(\lambda_n + \mu_{n+1} - \lambda_{n-1} - \mu_n)^2 + 4v_n^{-1}\lambda_n\mu_n} \right\} \right\},$$

where $v \equiv (v_1, v_2, \ldots)$ is any chain sequence;

(iii) $\alpha = \inf_v \left\{ \liminf_{n\to\infty} \sum_{i=0}^{n} \left(v_i(\lambda_i + \mu_{i+1}) - 2\sqrt{v_{i-1}v_i\lambda_i\mu_i} \right) \right\},$

where $v_{-1} \equiv 0$ and $v \equiv (v_0, v_1, \ldots)$ is any sequence of nonnegative numbers satisfying $\sum_i v_i = 1$.

Theorem 3.3 *If $\mu_0 > 0$, then the rate of convergence α of the birth-death process $\mathcal{X}$ can be represented as*

(i) $\alpha = \max_v \left\{ \inf_{n\ge 0} \left\{ \lambda_n + \mu_n - \frac{\lambda_{n-1} \mu_n}{v_n} - v_{n+1} \right\} \right\},$

where $v \equiv (v_0, v_1, \ldots)$ is any sequence of positive numbers;

(ii) $$\alpha = \max_v \left\{ \inf_{n\ge 1} \frac{1}{2} \left\{ \lambda_{n-1} + \mu_{n-1} + \lambda_n + \mu_n - \sqrt{(\lambda_n + \mu_n - \lambda_{n-1} - \mu_{n-1})^2 + 4v_n^{-1}\lambda_{n-1}\mu_n} \right\} \right\},$$

where $v \equiv (v_1, v_2, \ldots)$ is any chain sequence;

(iii) $\alpha = \inf_v \left\{ \liminf_{n\to\infty} \sum_{i=0}^{n} \left(v_i(\lambda_i + \mu_i) - 2\sqrt{v_{i-1}v_i\lambda_{i-1}\mu_i} \right) \right\},$

where $v_{-1} \equiv 0$ and $v \equiv (v_0, v_1, \ldots)$ is any sequence of nonnegative numbers satisfying $\sum_i v_i = 1$.

Remarks. (i) Letting $u_n \equiv v_n/\mu_n$ for $n > 0$, we see that Kartashov's lower bound (1.1) can actually be identified with representation (i) in Theorem 3.2, as announced.

(ii) The representations (i) in both theorems have been observed earlier by Kijima [15] (see also [16, Section 5.7]). Representation (i) in Theorem 3.2 has also been derived by Chen [6] using a coupling technique.

(iii) The representations in Theorem 3.3 are valid also when $\mu_0 = 0$ and $\sum_n \theta_n = \infty$, since $\psi(\{0\}) = \pi_0 = 0$ in this case, so that $\alpha = \min \operatorname{supp}(\psi^d) = \min \operatorname{supp}(\psi)$, as a consequence of (2.7).

4 Bounds

The expressions for α in the Theorems 3.2 and 3.3 involve an infinite number of unknown parameters v_n, so they do not readily yield explicit values. However, the expressions are particularly well suited for obtaining bounds. For instance, representation (i) in Theorem 3.2 (as well as Kartashov's bound (1.1), of course) tells us that if $\mu_0 = 0$, then

$$\alpha \geq \inf_{n\geq 0}\{\lambda_n + \mu_{n+1} - \frac{\lambda_n \mu_n}{v_n} - v_{n+1}\}, \tag{4.1}$$

for any positive sequence $v \equiv (v_0, v_1, \ldots)$. A clever choice for v can therefore lead to a good lower bound for α. In the same way the representation (i) in Theorem 3.3 and the representations (ii) in both theorems lead to lower bounds for α, while the representations (iii) generate upper bounds, all involving an infinite number of parameters v_n. Most (but not all) bounds for α that have appeared in the literature were derived or may be regained by an appropriate choice of these parameters. For a variety of bounds for α we refer to Blanc and van Doorn [2], Chen [4], [5, Chapter 9], [6], van Doorn [9], Kartashov [13], [14], Zeifman [19], [20], and the references there. By way of illustration we now give some examples of bounds – old and new – which are implied by the representations of the previous section.

Assuming $\mu_0 = 0$ we conclude from Theorem 3.2 (iii) that

$$\alpha \leq \lim_{N\to\infty} \inf \sum_{i=0}^{N} \left(v_i(\lambda_i + \mu_{i+1}) - 2\sqrt{v_{i-1}v_i\lambda_i\mu_i} \right), \tag{4.2}$$

for any sequence $v \equiv (v_0, v_1, \ldots)$ such that $\sum_i v_i = 1$. Choosing $v_i = 0$ for $i \neq n, n+1$, and minimizing the right hand side of (4.2) under this condition, yields after some algebra

$$\alpha \leq \frac{1}{2}\Big\{ \lambda_{n-1} + \mu_n + \lambda_n + \mu_{n+1} - \sqrt{(\lambda_n + \mu_{n+1} - \lambda_{n-1} - \mu_n)^2 + 4\lambda_n\mu_n} \Big\}. \tag{4.3}$$

for any value of $n \geq 1$. These upper bounds should be compared with the lower bound

$$\alpha \geq \inf_{n\geq 1} \frac{1}{2}\Big\{ \lambda_{n-1} + \mu_n + \lambda_n + \mu_{n+1} - \sqrt{(\lambda_n + \mu_{n+1} - \lambda_{n-1} - \mu_n)^2 + 16\lambda_n\mu_n} \Big\}, \tag{4.4}$$

which follows immediately by choosing $v_n = \frac{1}{4}$ for all $n \geq 1$ in the bound generated by Theorem 3.2 (ii). Slightly better lower bounds are obtained by letting $v_1 = \frac{1}{2}$ and $v_n = \frac{1}{4}$ for all $n > 1$, or by choosing $v_n = \frac{1}{4} + \frac{1}{16n(n+1)}$ for $n \geq 1$ (see Chihara [7, p. 98]). Both (4.3) and (4.4) were given earlier in [9].

Another interesting upper bound is obtained from (4.2) by choosing $v_i = \frac{1}{n}$ for $i = 1, 2, \ldots, n$ and $v_i = 0$ otherwise, and subsequently letting $n \to \infty$. This leads to

$$\alpha \leq \lim_{n\to\infty} \sup \Big\{ \lambda_n + \mu_{n+1} - 2\sqrt{\lambda_n\mu_n}) \Big\}, \tag{4.5}$$

which seems to be new. This upper bound may be compared with the lower bound

$$\alpha \geq \inf_{n\geq 0} \Big\{ \lambda_n + \mu_{n+1} - \sqrt{\lambda_n\mu_n} - \sqrt{\lambda_{n+1}\mu_{n+1}} \Big\}, \tag{4.6}$$

which follows from (4.1) by choosing $v_n = \sqrt{\lambda_n\mu_n}$ for $n \geq 1$, and which was given earlier in [9].

As an aside we note that the bound

$$\alpha \leq \frac{\lambda_0}{1 - \pi_0}, \tag{4.7}$$

obtained for the ergodic case from a more general result by Chen [5, Corollary 9.19 (1)], does not seem to be recoverable from representation (*iii*) in Theorem 3.2. However, duality does help here, for (2.7) and (3.2) tell us

$$\lambda_0 = \int_0^\infty x\psi(dx) \geq \alpha \int_0^\infty \psi(dx) = \alpha(1 - \psi(\{0\}),$$

which gives (4.7) since $\psi(\{0\}) = \pi_0$. We note that (4.7) improves upon the simple bound

$$\alpha \leq \lambda_0 + \mu_1,$$

which follows from (4.2) by choosing $v_0 = 1$, but not necessarily upon the subtler bound (4.3).

5 Concluding remarks

If the initial distribution $\{p_i(0),\ i \in \mathcal{N}\}$ of the birth-death proces $\mathcal{X}$ does not concentrate all mass on a single state (or on a finite number of states) the rates of convergence of the state probabilities

$$p_j(t) \equiv \Pr\{X(t) = j\} = \sum_{i \in \mathcal{N}} p_i(0)p_{ij}(t), \quad j \in \mathcal{N},\ t \geq 0,$$

may be smaller than α. Actually, an example is given in [2] (see also [1]) in which the rate of convergence of $p_j(t)$ may be any number between 0 and α by an appropriate choice of the initial distribution. The results of Zeifman [19] indicate that this phenomenon can occur only in the ergodic case. Parenthetically, we note that Zeifman gives bounds of the type

$$\sum_{j \in \mathcal{N}} |p_j(t) - \pi_j| \leq ce^{-rt}, \quad t \geq 0,$$

where r does not depend on the initial distribution, but c does. Zeifman's bound applied to the example in [2] referred to above gives $c = \infty$ and is therefore consistent with the results in [2].

Concluding, we refer to [11] for results similar to the Theorems 3.1 - 3.3 in a discrete-time setting, and to [18] and [8] for a recent proposal of an alternative to the rate of convergence α as a measure of the speed of convergence.

References

[1] J. Abate, M. Kijima and W. Whitt, Decompositions of the $M/M/1$ transition function. *Queueing Syst. Theory Appl.* **9** (1991) 323-336.

[2] J.P.C. Blanc and E.A. van Doorn, Relaxation times for queueing systems. In *Mathematics and Computer Science – CWI Monographs 1*, J.W. de Bakker, M. Hazewinkel and J.K. Lenstra, eds. (North-Holland, Amsterdam, 1986) 139-162.

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

[4] M.F. Chen, Exponential L^2-convergence and L^2-spectral gap for Markov processes. *Acta Math. Sinica (N.S.)* **7** (1991) 19-37.

[5] M.F. Chen, *From Markov Chains to Non-Equilibrium Particle Systems* (World Scientific, Singapore, 1992).

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

[7] T.S. Chihara, *An Introduction to Orthogonal Polynomials* (Gordon & Breach, New York, 1978).

[8] P. Coolen-Schrijner and E.A. van Doorn, On the convergence to stationarity of birth-death processes, submitted.

[9] 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.

[10] E.A. van Doorn, Representations and bounds for zeros of orthogonal polynomials and eigenvalues of sign-symmetric tri-diagonal matrices. *J. Approx. Th.* **51** (1987) 254-266.

[11] E.A. van Doorn and P. Schrijner, Geometric ergodicity and quasi-stationarity in discrete-time birth-death processes. *J. Austral. Math. Soc. (B)* **37** (1995) 121-144.

[12] 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) 589-646.

[13] N.V. Kartashov, Uniformly ergodic jump Markov processes with bounded intensities. *Theory Probab. Math. Statist.* **52** (1996) 91-103.

[14] N.V. Kartashov, Calculation of the exponential ergodicity exponent for birth-death processes. *Theory Probab. Math. Statist.* **57** (1998) 53-60.

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

[16] M. Kijima, *Markov Processes for Stochastic Modeling* (Chapman & Hall, London, 1997).

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

[18] W. Stadje and P.R. Parthasarathy. On the convergence to stationarity of the many-server Poisson queue. *J. Appl. Probab.* **36** (1999) 546-557.

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

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