Stochastic Processes and their Applications 1 (1973) 217–235. © *North-Holland Publ. Co.*

ON EXPONENTIAL ERGODICITY AND SPECTRAL STRUCTURE FOR BIRTH–DEATH PROCESSES, II

Herman CALLAERT and Julian KEILSON

Katholieke Universiteit, Leuven, Belgium
University of Rochester, Rochester, N.Y., USA

Received February 1972

Abstract. In Part I, Feller's boundary theory was described with simple conditions for process classification. The implications of this boundary classification scheme for spectral structure and exponential ergodicity are examined in Part II. Conditions under which the spectral span is finite or infinite are established. A time-dependent norm is exhibited describing the exponentiality of the convergence and its uniformity. Specific systems are discussed in detail.

Contents:

7. Spectral structure for the M/M/1 process

As an illustration of the spectral structure exhibited in the previous sections (see [1a]), we discuss the queue-length process $N(t)$ for the M/M/1 queue [10]. This process is a birth–death process with transition rates $\lambda_n = \lambda$, $n \geq 0$, $\mu_n = \mu$, $n \geq 1$, and $\mu_0 = 0$. We successively treat the passage-time densities (downward and upward), the transition probabilities and the classification of the process.

7.1. *The downward passage-time densities*

It has been shown in Section 4 that the half-plane of convergence of $\sigma_n^-(s)$, $n \geq 1$, terminates at a branch point β_n^-. This branch point is independent of n and has the value $-(\sqrt{\lambda} - \sqrt{\mu})^2$. We note that $|\beta_n^-| > 0$ if the queue is transient or positive recurrent, whereas $|\beta_n^-| = 0$ in the null-recurrent case.

7.2. *The upward passage time densities*

The abscissa of convergence β_n^+ of $\sigma_n^+(s)$ is a simple pole for all $n \geq 0$ and the sequence $\{\beta_n^+\}_{n=0}^{\infty}$ is strictly monotonic in the sense that $\beta_{n-1}^+ < \beta_n^+ < 0$ (see Theorem 4.3). By (2.2), $\beta_0^+ = -\lambda$, and hence the poles β_n^+ start at $-\lambda$ and move to the right as n increases from 0 to ∞. We now show that $\lim \{\beta_n^+ : n \to \infty\} = 0$ in the positive-recurrent case and that β_n^+ is uniformly bounded (upper bound) by $-(\sqrt{\lambda} - \sqrt{\mu})^2$ in the transient case.

Lemma 7.1. *For the queue-length process the abscissae of exponential convergence of the upward passage-time densities satisfy*:

$$\lim_{n\to\infty} \beta_n^+ = 0 \qquad \textit{if} \quad \lambda < \mu \,,$$

$$\lim_{n\to\infty} \beta_n^+ \leq -(\sqrt{\lambda} - \sqrt{\mu})^2 < 0 \qquad \textit{if} \quad \lambda > \mu \,.$$

Proof. If the process is positive recurrent, $\rho_n = \lambda_n/\mu_n = \lambda/\mu = \rho < 1$, so that we may apply [12, Theorem 3.1] which states that

$$\lim_{n\to\infty} \frac{\mathrm{E}[\tau_n^+]}{\Sigma_{j=0}^n \mathrm{E}[\tau_j^+]} = 1 - \rho \,. \tag{7.1}$$

From (2.4),

$$\lim_{n\to\infty} \sum_{j=0}^{n} \mathrm{E}[\tau_j^+] = \sum_{n=0}^{\infty} (\lambda_n \pi_n)^{-1} \sum_{j=0}^{n} \pi_j > \sum_{n=0}^{\infty} \lambda^{-1} = \infty \,,$$

and hence by (7.1), $\lim \{\zeta_n^+ : n \to \infty\} = \infty$. On the other hand, $\zeta_n^+ \leq |\beta_n^+|^{-1}$ by (3.4). Combining the last two results, one has $\lim \{\beta_n^+ : n \to \infty\} = 0$.

We now consider the transient case. In addition to the process $N(t)$ with state space $\{0, 1, 2, ...\}$, we consider the homogeneous birth–death process $H(t)$ defined on the state space $\{...-2, -1, 0, 1, 2, ...\}$ with transition rates $\lambda_n = \lambda$ and $\mu_n = \mu$ for all n. We now show that $S_{\mathrm{H}n}^+(\tau) \gg_{\mathrm{c}} S_{\mathrm{N}n}^+(\tau)$, where $S_n^+(\tau)$ is the c.d.f. of τ_n^+, and H and N stand for the homogeneous and the queue-length process, respectively.

Consider fig. 7.1 with B a set of states entered at $\tau = 0$. We have $B_{\mathrm{N}} = \{n-1, n-2, ..., 1, 0\}$ and $B_{\mathrm{H}} = \{n-1, n-2, ..., 1, 0, -1, -2, ...\}$. The random variables T_{N} and T_{H}, denoting the random dwell time in B, have

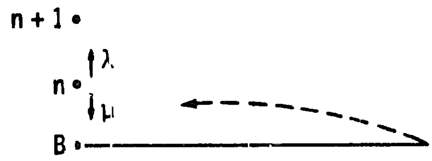

Fig. 7.1.

a non-defective c.d.f. $S_N(\tau)$ and $S_H(\tau)$ since $\lambda > \mu$. Moreover, from the nature of both processes it follows that $\bar{S}^+_{Hn}(\tau) \geq \bar{S}^+_{Nn}(\tau)$. To see this, we note that for $n = 0$ in Fig. 7.1, $S_N(\tau) = U_0(\tau)$, where $U_0(\tau)$ is the Heaviside function of unit jump at the point zero. Clearly $S_H(\tau) \geq \bar{S}_N(\tau)$, and the result then follows from Theorem 4.6 and induction on n. By P. 3.3, $\beta^+_{Hn} \geq \beta^+_{Nn}$, which means that the abscissa of convergence of $\sigma^+_{Nn}(s)$ for the queue-length process does not exceed the abscissa of convergence of $\sigma^+_{Hn}(s)$ for the homogeneous process.

On the other hand, $\sigma^+_{Hn}(s)$ is independent of n and satisfies

$$\sigma^+_H(s) = \lambda(s+\mu+\lambda-\mu\,\sigma^+_H(s))^{-1} .$$

Hence $\beta^+_{Hn} = -(\sqrt{\lambda}-\sqrt{\mu})^2$ for all n, by a reasoning similar to that for the downward passage-time densities (see Section 7.1). This means that for $\lambda > \mu$,

$$\lim_{n\to\infty} \beta^+_{Nn} \leq -(\sqrt{\lambda}-\sqrt{\mu})^2 < 0 ,$$

which completes the proof of the lemma.

Actually $\lim\{\beta^+_{Nn} : n \to \infty\} = -(\sqrt{\lambda}-\sqrt{\mu})^2$ in the transient case. This is so because $S^+_{Nn}(\tau)$ is stochastically monotonic increasing with n (as may be seen from Theorem 4.6 (i)) and converges in distribution by P.3.2f since

$$E[\tau^+_n] = (\lambda_n \pi_n)^{-1} \sum_{j=0}^{n} \pi_j = \frac{1-(\mu/\lambda)^{n+1}}{\lambda-\mu} ,$$

which is bounded from above by $(\lambda-\mu)^{-1} < \infty$.

It is worth noting that for the queue-length process $N(t)$,

(i) $-\nu_0 = -\lambda = \beta_0^+$, $-\nu_n = -(\lambda+\mu) < -\lambda < \beta_n^+$ for $n \geq 1$, and
(ii) $-\nu_n = -(\lambda+\mu) < -(\sqrt{\lambda} - \sqrt{\mu})^2 = \beta_n^-$ for $n \geq 1$,
illustrating Theorems 4.3 and 4.4.

7.3. *The index of convergence γ_n of the transition probabilities $p_{nn}(t)$*

It has been shown (Theorem 5.5) that $\beta_{nR} \leq \gamma_n$ in the transient case and $\beta_{nL} \leq \gamma_n \leq \beta_{nR}$ in the positive-recurrent case, the only possible exception being the case when $\beta_{nL} = \beta_{nR}$ and both are poles. β_{nL} and β_{nR} correspond to β_{n-1}^+ and β_{n+1}^- (or conversely).

For a transient queue-length process $\beta_n^- = -(\sqrt{\lambda}-\sqrt{\mu})^2$, whereas $\beta_n^+ < -(\sqrt{\lambda} - \sqrt{\mu})^2$ for all n (see Sections 7.1 and 7.2). Hence $\beta_{nR} = \beta_{n+1}^-$ and $-(\sqrt{\lambda} - \sqrt{\mu})^2 \leq \gamma_n$ for all n.

The positive-recurrent case is somewhat more complicated in that β_{n+1}^- and β_{n-1}^+ may satisfy either of the relations

(i) $\beta_{n+1}^- < \beta_{n-1}^+$,
(ii) $\beta_{n+1}^- = \beta_{n-1}^+$,
(iii) $\beta_{n+1}^- > \beta_{n-1}^+$.

In particular, (i) is satisfied for all n if $\lambda < \frac{1}{4}\mu$. In any of the cases (i), (ii) and (iii), we have $\beta_{nL} \leq \gamma_n \leq \beta_{nR}$. If $\beta_{nR} = \beta_{n+1}^-$ (which is a branch point), then $\gamma_n = \beta_{n+1}^-$.

The above results are in accordance with those in [1], where it is shown that $\gamma_n = -(\sqrt{\lambda} - \sqrt{\mu})^2$ and that even in the positive-recurrent case, γ_n cannot be improved for any state n.

7.4. *The classification of the process*

Since $\lim \{\rho_n : n \to \infty\} = \lambda/\mu$ and $\lim \{\rho_n^* : n \to \infty\} = \lambda/\mu$, the classification depends entirely on the value of λ/μ. It is shown in table 7.1. We note that, by a similar reasoning as above, the queue-length process M/M/s has a natural boundary at infinity for every s (even for $s = \infty$).

Table 7.1

	Classification	cf.
$\lambda/\mu < 1$	natural, positive-recurrent	(6.19)
$\lambda/\mu = 1$	natural, null-recurrent	(6.18)
$\lambda/\mu > 1$	natural, transient	(6.16)

8. Exponential ergodicity for processes with entrance, exit, and regular boundaries

The ideas developed in Sections 4 and 5 permit one to infer quickly that all birth–death processes with entrance, exit or regular boundaries are exponentially ergodic.

Theorem 8.1. *Every birth–death process with an entrance boundary is exponentially ergodic.*

Proof. We have seen in Section 6 that every such birth–death process is positive recurrent and that $E[\tau_{\infty 0}] = \lim \{E[\tau_{n0}] : n \to \infty\} < \infty$. Since $\tau_{\infty 0} = \Sigma_{n=1}^{\infty} \tau_n^-$, and τ_n^- is a positive random variable, it follows that $\tau_{\infty 0}$ is itself a positive random variable. Moreover, for the same process truncated by a reflecting boundary at L, the passage time τ_{Ln} will have a log-concave probability density function (cf. [13]) and τ_{Ln} converges in distribution to $\tau_{\infty n}$.[1] Hence $\tau_{\infty n}$ has a log-concave p.d.f., and by P.3.8 this is exponentially convergent with negative index. Hence $\sigma_n^-(s) = \sigma_{\infty,n+1}(s)/\sigma_{\infty,n}(s)$ has a negative abscissa of convergence β_n^- by P.3.5. The theorem then follows from Theorem 5.1 and P.5.2.

P.8.2. *Every birth–death process with a regular reflecting boundary is exponentially ergodic.*

Proof. The proof parallels that above completely, since $E[\tau_{\infty 0}] < \infty$ and τ_{Ln} converges in distribution to $\tau_{\infty n}$.

Theorem 8.3. *Every birth–death process with a regular boundary is exponentially ergodic.*

Proof. The regular boundary will be either reflecting, the case covered in P.8.2, or will be completely or partially absorbing. If we denote the process with reflecting boundary by $N_R(t)$, and a process with absorbing boundary by $N_A(t)$, one has immediately

$$s_{An}^-(\tau) \leqslant s_{Rn}^-(\tau)$$

for the downward passage time densities from state n. Consequently,

[1] This is a direct consequence of Theorem 4.6 (i), a simple induction argument, and P.3.?

$\sigma_{An}^-(s)$ will be analytic for $\mathrm{Re}(s) > \beta_{Rn}^-$, and $\beta_{An}^- < \beta_{Rn}^- < 0$. The theo again follows from Theorem 5.1 and P.5.2.

Theorem 8.4. *Every birth–death process with an exit boundary is exponentially ergodic.*

It is known that all such processes are transient and that $\mathbf{E}[\tau_{0\infty}] = \lim\{\mathbf{E}[\tau_{0n}] : n \to \infty\} < \infty$. Again from P.3.2f and P.3.7, $\tau_{0\infty}$ will be a proper random variable with a log-concave probability density function, as will all $\tau_{m\infty}$. Clearly for such a process commencing at state $n = 0$, one has $\tau_{0\infty} > \tau_{m\infty}$ for all m, so that $\bar{A}_m(x) = \mathbf{P}\{\tau_{m\infty} > x\} \leq \mathbf{P}\{\tau_{0\infty} > x\} = \bar{A}_0(x)$. Also (cf. [4, XVII.4])

$$(8.1) \qquad \sum_{n=0}^{\infty} p_{0n}(t) = \mathbf{P}\{\tau_{0\infty} > t\} = \bar{A}_0(t)\,.$$

The log-concavity of the density of $\tau_{0\infty}$ implies that $\bar{A}_0(t)$ is exponentially convergent with negative index from P.4.7, so that $\Sigma\, p_{0n}(t)$ and $p_{0n}(t)$ are also such, implying exponential ergodicity. Note that

$$(8.2) \qquad \sum_{n=0}^{\infty} p_{mn}(t) = \mathbf{P}\{\tau_{m\infty} > t\} = \bar{A}_m(t) \leq \bar{A}_0(t)\,.$$

The structure exhibited in (8.2) was pointed out by Karlin and McGregor [9] via spectral theory.

9. Exponential ergodicity for processes with natural boundaries

The results of this section are based on the analytic theory of continued fractions, an exposition of which has been given by Wall [25]. The recursion relation (2.3) may be written in the form

$$(9.1) \qquad \sigma_n^-(s) = \alpha_n\{\beta_n(s) + \sigma_{n+1}^-(s)\}^{-1}\,,$$

where

$$(9.2) \qquad \alpha_n = -q_n/p_n\,,$$

$$(9.3) \qquad \beta_n(s) = -(s+\nu_n)/(\nu_n p_n)\,.$$

From (9.1), we then have a representation of $\sigma_1^-(s)$ as the continued fraction

(9.4) $$\sigma_1^-(s) = \cfrac{\alpha_1}{\beta_1(s) + \cfrac{\alpha_2}{\beta_2(s) + \dots}} \,.$$

The continued fraction is defined to be the limit of the sequence of approximants $\sigma_{1N}^-(s)$ obtained from (9.1) by setting $\sigma_N^-(s)$ equal to zero, provided the sequence converges as N goes to infinity. For the natural case here treated, the boundary at infinity is inaccessible. Such truncation at infinity then does not alter the process, and (9.4) is a valid expression for $\sigma_1^-(s)$, for both recurrent and transient cases. For a regular boundary, (9.4) would be valid only if the boundary is totally absorbing, and the classical definition of the continued fraction would not be suitable.

An equivalent continued fraction to (9.4) (the approximants coincide with those of (9.4)) is given by (cf. [25, p. 19])

(9.5) $$\sigma_1^-(s) = \cfrac{\alpha_1^*(s)}{1 - \cfrac{\alpha_2^*(s)}{1 - \dots}}$$

where

(9.6) $$\alpha_1^*(s) = q_1 \{\nu_1 (\nu_1 + s)^{-1}\} \,,$$

(9.7) $$\alpha_n^*(s) = p_{n-1}\, q_n \, \frac{\nu_{n-1}}{s+\nu_{n-1}} \, \frac{\nu_n}{s+\nu_n} \,, \qquad n \geqslant 2 \,.$$

When $\liminf \nu_n = 0$, we have seen in P.5.3 that the process cannot be exponentially ergodic. We may therefore restrict examination to the case $\nu_n > \nu^* > 0$. For this case we have the following sufficient condition for exponential ergodicity.

Theorem 9.1. *If for the natural birth–death process $N(t)$ one has*

(i) $\nu_n > \nu^* > 0, n \geqslant 0$,

(ii) $\limsup p_{n-1} q_n < \frac{1}{4}$,

then $N(t)$ is exponentially ergodic.

Proof. We may write

$$\alpha_n^*(s) = p_{n-1}\, q_n \, \frac{\nu_{n-1}}{\nu_{n-1} - \epsilon} \, \frac{\nu_n}{\nu_n - \epsilon} \, \frac{\nu_{n-1} - \epsilon}{\nu_{n-1} - \epsilon + (s+\epsilon)} \, \frac{\nu_n - \epsilon}{\nu_n - \epsilon + (s+\epsilon)} \,.$$

Clearly $\nu_n\,(\nu_n-\epsilon)^{-1}\leqslant\nu^*\,(\nu^*-\epsilon)^{-1}$ when $\epsilon<\nu^*$. Also

$$\left|\frac{\nu_n-\epsilon}{\nu_n-\epsilon+(s+\epsilon)}\right|\leqslant 1,\qquad n\geqslant 0,\quad\text{when}\quad \mathrm{Re}\,(s+\epsilon)\geqslant 0\,.$$

If we denote lim sup $p_{n-1}\,q_n$ by ζ, we then have

$$\limsup_{n\to\infty}|\alpha_n^*(s)|\leqslant\zeta\,\{\nu^*\,(\nu^*-\epsilon)^{-1}\}^2<\tfrac{1}{4},\quad\text{when}\quad\mathrm{Re}\,(s+\epsilon)\geqslant 0$$

if $0<\epsilon<\nu^*\,(1-\sqrt{4\zeta})$. This implies that for some positive integer K, $|\alpha_n^*(s)|<\frac{1}{4}$ for all $n\geqslant K$. We may then call on Worpitzky's Theorem [25, p. 42] which implies that the sequence of approximants for the continued fraction

$$\sigma_K^-(s)=\cfrac{\alpha_K^*(s)}{1-\cfrac{\alpha_{K+1}^*(s)}{1-\cfrac{\alpha_{K+2}^*(s)}{\cdots}}}$$

converges uniformly over the domain $D=\{s\colon \mathrm{Re}\,(s+\epsilon)\geqslant 0\}$. The approximants of $\sigma_K^-(s)$ are all analytic functions of s in D, and the uniform convergence guarantees that $\sigma_K^-(s)$ will be analytic in D. Hence the exponential ergodicity of the process $N(t)$ follows from P.5.2.

Corollary 9.2. *If for the natural birth–death process $N(t)$ one has*
(i) $\nu_n>\nu^*>0$,
(ii) $\lim\,\{\lambda_n/\mu_n : n\to\infty\}=\theta\neq 1,\ 0\leqslant\theta\leqslant\infty$,
then $N(t)$ is exponentially ergodic.

The proof is immediate since $\lim\,\{p_{n-1}\,q_n : n\to\infty\}=\theta\,(1+\theta)^{-2}<\frac{1}{4}$.

It is worth noting that all basic linear birth–death processes with $\lambda_n=n\alpha_1+\beta_1$, $\mu_n=n\alpha_2$, α_1, α_2 and β_1 positive, are natural processes as the reader will verify from (6.9), (6.11) and (6.12). Hence by Corollary 9.2, all such processes are exponentially ergodic when $\alpha_1\neq\alpha_2$. When $\theta=1$, i.e., $\alpha_1=\alpha_2$, one cannot have exponential ergodicity as has been shown analytically [1].

We remark that Theorem 9.1 provides a sufficient condition for exponential ergodicity which seems far from necessary. The condition does not seem able, for example, to handle sequences $\{\lambda_n,\mu_n\}$ for which λ_n/μ_n fluctuates.

10. Uniformity of exponential convergence

A simple identity plays a useful role in the discussion of convergence and its uniformity.

P.10.1. *Let $N(t)$ be any irreducible Markov chain in continuous time and let $e_n = \lim \{p_{nn}(t) : t \to \infty\}$. Then for any $m, n \geqslant 0$, one has*

$$(10.1) \qquad \sum_{n=0}^{\infty} \{p_{mn}(t) - e_n\} \{p_{nr}(t) - e_r\} = p_{mr}(2t) - e_r \ .$$

Proof. The statement is obtained trivially from the Chapman–Kolmogorov equation $\Sigma_n \, p_{mn}(s) \, p_{nr}(t) = p_{mr}(s+t)$ when one has transience or null recurrence. For positive recurrence, where $e_n \neq 0$, one also has $\Sigma_n \, e_n = 1$, $\Sigma_n \, p_{mn} = 1$, and $\Sigma_n \, e_n \, p_{nr} = e_r$, and the lemma follows.

P.10.2. *If $N(t)$ is reversible in time (cf.* [11, 16, 19]), *positive recurrent, and irreducible, then*

$$(10.2) \qquad \sum_{n=0}^{\infty} e_n^{-1} \{p_{mn}(t) - e_n\}^2 = e_m^{-1} \{p_{mm}(2t) - e_m\} \ .$$

To establish this, one employs P.10.1 with $r = m$. One then uses the relation $e_n \, p_{nm}(t) = e_m \, p_{mn}(t)$ for time reversible processes, and the positivity of e_m for all m implied by the irreducibility.

P.10.3. *If $N(t)$ is reversible in time, transient or null recurrent, and irreducible, then*

$$(10.3) \qquad \sum_{n=0}^{\infty} \pi_n^{-1} \, p_{mn}^2(t) = \pi_m^{-1} \, p_{mm}(2t) \, ,$$

where π_m is the potential coefficient for state m associated with the reversible process.

Again this follows from P.10.1 for $r = m$. One has $e_n = 0$ for all n, and $\pi_m \, p_{mn}(t) = \pi_n \, p_{nm}(t)$ for all m and n.

The positivity of the summands in (10.2) and (10.3) gives rise to a useful inequality:

P.10.4. *For any irreducible, time-reversible Markov chain in continuous time, and any $m,n \geqslant 0$,*

(10.4) $$\gamma_{mn} \leqslant \min\{\gamma_m, \gamma_n\}.$$

Proof. We note from (10.2) and (10.3) that for all m,n,

(10.5) $$\pi_n^{-1}\{p_{mn}(t) - e_n\}^2 \leqslant \pi_m^{-1}\{p_{mm}(2t) - e_m\}.$$

The proof then follows directly from P.3.1c (i).

The inequality (10.4) permits a quick derivation of an important result obtained previously by Callaert [1].

P.10.5. *Let $N(t)$ be any basic birth–death process. Then*

(10.5) $$\gamma_n \leqslant \gamma_0 \quad \textit{for all } n.$$

Proof. For any basic birth–death process, one has $s_{0m}(\tau) * p_{mm}(\tau) = p_{0m}(\tau)$, so that $\sigma_{0m}(s)\{s\pi_{mm}(s)\} = s\pi_{0m}(s)$. It is known that $\sigma_{0m}(s)$ is the reciprocal of a polynomial of degree m with all roots real and simple (see [13]). Hence $\gamma_{0m} = \gamma_m$. But (10.4) implies that $\gamma_{0m} \leqslant \gamma_0$, and the theorem follows.

For the transient case, a stronger result proven by Kingman [17] in a more general setting is available.

P.10.6. *If a basic birth–death process is transient and exponentially ergodic, then*

$$\gamma_{mn} = \gamma_{nn} = \gamma_n \qquad \textit{for all } m,n.$$

Proof. From (10.1) and the nonnegativity of $p_{mn}(t)$, one has for all m,n,

$$p_{mn}(t)\, p_{nn}(t) \leqslant p_{mn}(2t).$$

When one takes logarithms, divides by t, and permits t to go to $+\infty$, then from P.3.1c (i) one obtains $\gamma_{mn} + \gamma_n \leqslant 2\gamma_{mn}$, i.e., $\gamma_{mn} \geqslant \gamma_n$. From P.10.4, however, one has $\gamma_{mn} \leqslant \gamma_n$. Hence $\gamma_{mn} = \gamma_n$ for all m,n and the statement follows.

The uniformity of the convergence contained in P.10.5 may be presented for arbitrary initial distribution when the spectral representation of Karlin and McGregor [8] is available.

Theorem 10.7. *Let $N(t)$ be any entrance or natural positive recurrent basic birth–death process for which $\alpha_m = \mathbf{P}\{N(0)=m\}$ and $\Sigma_{m=0}^{\infty} e_m^{-1}\alpha_m^2 < \infty$. Let $p_k(t) = \mathbf{P}\{N(t)=k\}$, and let*

$$(10.7) \qquad H(t) = \left[\sum_{k=0}^{\infty} e_k^{-1}\{p_k(t)-e_k\}^2\right]^{1/2}.$$

Then $H(t)$ is finite for all t, monotonic decreasing and log-convex. Moreover, for any such initial distribution,

$$(10.8) \qquad H(t) \leqslant H(0)\exp[\gamma_0 t].$$

Proof. From (10.1), we infer as for (10.2) that

$$(10.9) \qquad \sum_{n=0}^{\infty} e_n^{-1} p_{mn}(t)\, p_{rn}(t) = e_r^{-1} p_{mr}(2t).$$

Let $p_{Kn}(t) = \Sigma_{m=0}^{K} \alpha_m p_{mn}(t)$. If we multiply (10.9) by $\alpha_m\alpha_r$ and sum over m, r from 0 to K, we obtain

$$(10.10) \qquad \sum_{n=0}^{\infty} e_n^{-1} p_{Kn}^2(t) = \sum_{m=0}^{K}\sum_{r=0}^{K} e_r^{-1}\alpha_m\, p_{mr}(t)\,\alpha_r.$$

The Karlin–McGregor [8] representation states that [2]

$$(10.11) \qquad p_{mr}(2t) = e_r \int_0^{\infty} Q_m(x)\, Q_r(x)\, \mathrm{e}^{-2xt}\rho(\mathrm{d}x),$$

where $Q_m(x)$ is a polynomial of degree m. Hence

$$(10.12) \qquad \sum_{n=0}^{\infty} e_n^{-1} p_{Kn}^2(t) = \int_0^{\infty}\left\{\sum_{m=0}^{K}\alpha_m Q_m(x)\right\}^2 \mathrm{e}^{-2xt}\rho(\mathrm{d}x)$$
$$\leqslant \int_0^{\infty}\left\{\sum_{m=0}^{K}\alpha_m Q_m(x)\right\}^2 \rho(\mathrm{d}x) = \sum_{n=0}^{\infty} e_n^{-1} p_{Kn}^2(0) = \sum_{n=0}^{K} e_n^{-1}\alpha_n^2$$

[2] We note that for the positive-recurrent case, $e_n/e_0 = \pi_n/\pi_0 = \pi_n$. Hence (10.11) coincides with [8, (1.7)] with $\rho(\mathrm{d}x) = \psi(\mathrm{d}x)/e_0$.

From (10.9), $p_{Kn}(t)$ is monotonic increasing with K. From (10.12), for $p_n(t) = \lim \{p_{Kn}(t) : K \to \infty\}$, we have $\Sigma_{n=0}^{\infty} e_n^{-1} p_n^2(t) \leq \Sigma_{n=0}^{\infty} e_n^{-1} \alpha_n^2$ when the latter is finite. Also from (10.12), and P.10.5,

$$(10.13) \quad H_K^2(t) \overset{\text{def}}{=} \sum_{n=0}^{\infty} e_n^{-1} p_{Kn}^2(t) - \sum_{n=0}^{K} e_n^{-1} \alpha_n^2$$

$$= \int_{0+}^{\infty} \left\{ \sum_{m=0}^{K} \alpha_m Q_m(x) \right\}^2 e^{-2xt} \rho(dx)$$

$$\leq \exp[2\gamma_0 t] \int_{0+}^{\infty} \left\{ \sum_{m=0}^{K} \alpha_m Q_m(x) \right\}^2 \rho(dx) = H_K^2(0) \exp[2\gamma_0 t] .$$

It follows from dominated convergence that $H_K^2(t) \to H^2(t)$, and

$$H^2(t) \leq H^2(0) \exp[2\gamma_0 t] .$$

From (10.13) we see that $H_K^2(t)$ is completely monotonic for every value of K. The limit of a convergent sequence of completely monotonic functions is completely monotonic, hence log-convex, and $H(t)$ will also be log-convex. This completes the proof.

It is somewhat curious that birth–death processes are essentially ℓ_2 processes. One consequence is that not all initial distributions are finite in that they have a finite ℓ_2 norm $\Sigma_{k=0}^{\infty} e_k^{-1} \alpha_k^2$. The set of all distributions finite in this norm is a convex set which includes the distributions with $p_n(0) = \delta_{n,m}$ and the ergodic distribution. The ℓ_2 character of birth–death processes plays a key role in the work of Karlin and McGregor [8] and Kendall [15, 16].

It is also worth noting that for a transient or null-recurrent process a comparable theorem is readily available, stating that the norm

$$H(t) = \left\{ \sum_{k=0}^{\infty} \pi_k^{-1} p_k^2(t) \right\}^{1/2}$$

is monotonic decreasing and log-convex with

$$H(t) \leq H(0) \exp[\gamma_0 t] .$$

When $\gamma_0 = 0$, one does not have exponential decay of the norm. But one still has monotonic log-convex decay of the norm. This is a consequence of the reversibility of the process in time. Similar results may be expected for any time-reversible chain.

11. Finite and infinite spectral span

Definition 11.0. Let $g(t)$ be completely monotonic on $(0, t)$ and have the representation

$$(11.1) \qquad g(t) = \int_{|\beta|}^{|\alpha|} e^{-xt}\, \rho(dx)\,, \qquad 0 \leqslant |\beta| \leqslant |\alpha| \leqslant \infty\,,$$

where $\rho(dx)$ is a measure for all $0 < t < \infty$. The interval of support $(|\beta|, |\alpha|)$ will be said to have *spectral span* $|\alpha| - |\beta|$.

In P.3.13 and P.3.14, it has been seen that when $g(t)$ is bounded or integrable, $\gamma(s) = L\{g(t)\}$ has all its zeros and singularities on the negative interval (α, β) and is analytic elsewhere. For the passage time densities $s_n^+(\tau)$, $s_n^-(\tau)$ and $s_n(\tau)$, we designate the intervals by (α_n^+, β_n^+), (α_n^-, β_n^-) and $(\alpha_{on}, \beta_{on})$, respectively. For $p_{nn}(t)$, we designate the interval by $(\alpha_{\pi n}, \beta_{\pi n})$.

The key theorem for the discussion of the spectral span is the following.

Theorem 11.1. *Let $f(t)$ and $g(t)$ be bounded completely monotonic functions having Laplace transforms $\varphi(s)$ and $\gamma(s)$, respectively, such that*

$$(11.2) \qquad \varphi(s) = (s + \nu - \nu\gamma(s))^{-1}\,,$$

and $\gamma(0+) \leqslant 1$. Then the spectral spans of $f(t)$ and $g(t)$ are both finite or both infinite. When both are finite, $\alpha_f \leqslant \alpha_g$, i.e., $|\alpha_f| \geqslant |\alpha_g|$.

Proof. By P.3.13, $\varphi(s)$ and $\gamma(s)$ have the representation (3.6). We note (from [3, Theorem 1a, p. 416]) that a completely monotonic function always has a finite index of exponential convergence. Thus $|\beta_f| < \infty$ and $|\beta_g| < \infty$, and we need only consider α_f and α_g. From (3.6), we see that $\gamma(s) \to 0$ as $s \to +\infty$. Hence from (11.2), $s\varphi(s) \to 1$, and $f(0+) = 1$. We may rewrite (11.2) in either the form

$$(11.3)\qquad \nu\left\{\frac{\gamma(0)-\gamma(s)}{s}\right\}+\nu\left\{\frac{1-\gamma(0)}{s}\right\}=\frac{1-s\varphi(s)}{1-(1-s\varphi(s))}$$

or

$$(11.4)\qquad \varphi(s)=\frac{(s+\nu)^{-1}}{1-(\nu\,(s+\nu)^{-1})\,\gamma(s)}\,.$$

We note that $1-s\varphi(s)=L\{-\mathrm{d}f(t)/\mathrm{d}t\}$ and that α for $-\mathrm{d}f/\mathrm{d}t$ coincides with α_f. Similarly, $\{\gamma(0)-\gamma(s)\}/s=L\{\bar{G}(t)\}$, where $\bar{G}(t)=\int_t^\infty g(u)\,\mathrm{d}u$, and $\alpha_g=\alpha_{\bar{G}}$. If α_f is finite, then $1-s\varphi(s)<0$ on $(-\infty,\alpha_f)$ by P.3.13, and $1-s\varphi(s)\to 0$ as $s\to-\infty$ along the negative axis. It follows from (11.3) and P.3.13 that the left-hand side of (11.3), for which $\alpha=\alpha_g$, has $\alpha_g\geqslant\alpha_f$. Suppose that one knows instead of α_f finite that α_g is finite. Then one may consider form (11.4) and verify that as $s\to-\infty$, the denominator of (11.4) is ultimately positive, the numerator ultimately negative and both go to zero, in keeping with P.3.13. Hence α_g finite implies α_f finite, and the theorem is proven.

Theorem 11.1 has immediate consequences for birth–death processes.

P.11.2. *For any basic birth–death process, and any state n, $p_{nn}(t)$ and $s_n(t)$ both have finite spectral spans or both have infinite spectral spans. When both are finite, $\alpha_{\pi n}\leqslant\alpha_{on}$.*

Proof. The statement follows directly from (5.1) and Theorem 11.1. One need only observe that $s_n^+(\tau)$, $s_n^-(\tau)$ and $s_n(\tau)$ are all completely monotonic and bounded (see P.4.1a and P.4.1b), as is $p_{nn}(t)$.

P.11.3. *All the upward passage time densities $s_n^+(\tau)$ have finite spectral spans. Moreover, α_n^+ is monotonic decreasing as n increases.*

Proof. One observes that (2.1) has a form to which Theorem 11.1 is immediately applicable. Since $s_0^+(\tau)=\lambda_0\exp[-\lambda_0\tau]$, with $\alpha_0^+=\beta_0^+=-\lambda_0$, and $\alpha_{n+1}^+\leqslant\alpha_n^+$, the result follows by induction. (It is known, in fact, that $\alpha_{n+1}^+<\alpha_n^+$ since $\sigma_n^+(s)$ is a rational function whose poles lie between those of $\sigma_{n+1}^+(s)$ (see [13]).)

P.11.4. *The downward passage time densities $s_n^-(\tau)$ either have infinite spectral spans for all $n\geqslant 1$ or have finite spectral spans for all $n\geqslant 1$. In the latter case, α_n^- is monotonic increasing as n increases.*

Proof. As for P.11.3, the statement is immediate from (2.3) and Theorem 11.1, and induction.

P.11.5. *If for any basic birth–death process $N(t)$, $\mathbf{E}[\tau_n^-] \to 0$ for $n \to \infty$, then $s_n(\tau)$ and $p_{nn}(\tau)$ have infinite spectral span for all n.*

Proof. We have seen in P.3.10 and the accompanying footnote 1 that for any passage time τ with completely monotonic probability density function

$$(11.5) \qquad |\beta| \leqslant (\mathbf{E}[\tau])^{-1} \leqslant |\alpha| \, .$$

When $\mathbf{E}[\tau_n^-] \to 0$, one must have $|\alpha_n^-|$ unbounded. From P.11.4, however, one must then have $|\alpha_n^-| = \infty$ for all n. Consequently for $s_n(\tau)$, $|\alpha_{on}| = \infty$ for all n, and by P.11.2, $|\alpha_{\pi n}| = \infty$.

For entrance and regular reflecting processes, one has immediately from $\mathbf{E}[\tau_{\infty 0}] < \infty$:

P.11.6. *All entrance processes, regular processes with a reflecting boundary at infinity, and those natural processes for which $\mathbf{E}[\tau_n^-] \to 0$ as $n \to \infty$ have infinite spectral span for all n.*

When the Karlin–McGregor representation is available, a similar conclusion is reached for other processes of interest.

P.11.7. *Let $N(t)$ be any exit, entrance or natural process. Then $\alpha_{\pi n} \geqslant \alpha_{\pi 0}$.*

Proof. For any such process, the Karlin–McGregor representation (10.11) is available; i.e.,

$$p_{nn}(t) = \pi_n \int_0^\infty Q_n^2(x) \, \mathrm{e}^{-xt} \, \rho(\mathrm{d}x) \, .$$

Since a spectral value for the support of $p_{00}(t)$ can only be removed from that for $p_{nn}(t)$ by a zero of $Q_n(x)$, the statement follows.

One then has:

Theorem 11.8. *Let $N(t)$ be a basic birth–death process in any of the following categories*: (a) *entrance*, (b) *exit*, (c) *natural, with* $\mathrm{E}[\tau_n^+] \to 0$ *as* $n \to \infty$, *or* $\mathrm{E}[\tau_n^-] \to 0$ *as* $n \to \infty$. *All such processes have infinite spectral span.*

Proof. The Karlin–McGregor representation is available for all of these processes. When $\mathrm{E}[\tau_n^+] \to 0$ or $\mathrm{E}[\tau_n^-] \to 0$, then $|\alpha_n^+| \to \infty$ or $|\alpha_n^-| \to \infty$, and hence $|\alpha_{\sigma n}| \to \infty$. Consequently, $|\alpha_{\pi n}| \to \infty$. From P.11.7, one then infers that $\rho(\mathrm{d}x)$ has infinite span.

It should be noted that a broad class of processes exists for which the spectral span is finite. We have in mind certain natural processes such as M/M/s, for which it is known analytically that the spectral span is finite. Details may be found, for example, in [10].

The knowledge that ν_n is unbounded is enough to insure that the spectral span is infinite. This conclusion is made evident from the following theorem.

Theorem 11.9. *Let $N(t)$ be any basic birth–death process for which $p_{nn}(t)$ has a spectral support interval bounded by $|\beta_{\pi n}|$ and $|\alpha_{\pi n}|$. Then*

$$(11.6)\qquad |\beta_{\pi n}| \leqslant \nu_n \leqslant |\alpha_{\pi n}| \,.$$

Proof. Complete monotonicity assures the representation

$$(11.7)\qquad p_{nn}(\tau) = \int_{|\beta_{\pi n}|}^{|\alpha_{\pi n}|} \mathrm{e}^{-x\tau}\,\rho_n(\mathrm{d}x)\,,$$

where $\rho_n(\mathrm{d}x)$ is a finite measure of mass unity, since $p_{nn}(0+) = 1$. Hence

$$(11.8)\qquad -p_{nn}'(0+) = \int_{|\beta_{\pi n}|}^{|\alpha_{\pi n}|} x\,\rho_n(\mathrm{d}x) \left\{ \int_{|\beta_{\pi n}|}^{|\alpha_{\pi n}|} \rho_n(\mathrm{d}x) \right\}^{-1} .$$

It is known, however, that (cf. P.5.2)

$$L\{-p_{nn}'(\tau)\} = 1 - s\,\pi_n(s) = \frac{\nu_n(1-\sigma_n(s))}{s+\nu_n(1-\sigma_n(s))}\,.$$

Hence $-p_{nn}'(0+) = \nu_n$ by the Tauberian argument. When this result is combined with (11.8), one obtains (11.6) immediately.

Corollary 11.10. *If for any basic birth–death process, ν_n is unbounded, the process has infinite spectral span (in the sense that* $\sup_n\{|\alpha_{\pi n}| - |\beta_{\pi n}|\} = \infty$.

This corollary provides a convenient alternative to Theorem 11.8 for the natural processes. It also conveys the sense in which regular processes, for all of which $\sup \nu_n = \infty$, have infinite spectral span even though the Karlin–McGregor representation is not applicable.

12. Skip-free processes on the full lattice

The methods and results obtained for birth–death processes permit direct extension to their analogues on the full lattice of integers. In this class is any process $N(t)$ on a state space $N = \{n\colon -\infty < n < \infty\}$ which is temporally homogeneous, Markovian, and skip-free or lattice-continuous in both directions. Such processes are characterized by a set of transition rates $\{\lambda_n, \mu_n\}$ for which λ_n and μ_n are positive for all real n. They too are reversible in time. The notation and theory of our early sections is still applicable with only minor revision. For every set of state probabilities $p_n(t) = \mathbf{P}\{N(t) = n\}$, one has

$$(12.1) \qquad \mathrm{d}p_n/\mathrm{d}t = -(\lambda_n+\mu_n)p_n(t) + \lambda_{n-1}\, p_{n-1}(t) + \mu_{n+1}\, p_{n+1}(t)\,.$$

Again one finds

$$\mathrm{d}p_{nn}(t)/\mathrm{d}t = -\nu_n\, p_{nn}(t) + \nu_n\, p_{nn}(t) * s_n(t)\,,$$

where $s_n(t)$ is given by (1.5). Consequently, one still has

$$(12.2) \qquad \pi_{nn}(s) = (s+\nu_n - \nu_n\, \sigma_n(s))^{-1}\,,$$

where $\sigma_n(s) = p_{n-1}\, \sigma^+_{n-1}(s) + q_n\, \sigma^-_{n+1}(s)$. Clearly, exponential ergodicity for the full lattice process is equivalent to exponential convergence for $s^+_{n-1}(\tau)$ and $s^-_{n+1}(\tau)$. To establish the exponential ergodicity, one needs only establish the exponential ergodicity of the two component birth–death processes, obtained by setting up a boundary at $n = 0$, reflecting in both directions.

Acknowledgment

The authors wish to thank A.W. Marshall for his interest and helpful comments.

References

[1] H. Callaert, Exponential ergodicity for birth-and-death processes, Ph.D. Thesis, Katholieke Universiteit, Leuven (1971).
[1a] H. Callaert and J. Keilson, On exponential ergodicity and spectral structure for birth–death processes I, Stoch. Processes Appl 1 (1973) 187–216.
[2] W. Feller, The birth-and-death processes as diffusion processes, J. Math. Pures Appl. 38 (1959) 301–345.
[3] W. Feller, An introduction to probability theory and its applications, Vol. 2 (Wiley, New York, 1966).
[4] W. Feller, An introduction to probability theory and its applications, Vol. 1, 3rd ed. (Wiley, New York, 1968).
[5] C.R. Heathcote, Complete exponential convergence and some related topics, J. Appl. Probability 4 (1967) 217–256.
[6] I A. Ibragimov, On the composition of unimodal distributions, Theor. Probability Appl. 1 (1956) 255–260.
[7] S. Karlin, Total positivity (Stanford Univ. Press, Stanford, Calif., 1968).
[8] S. Karlin and J.L. McGregor, The differential equations of birth-and-death processes, and the Stieltjes moment problem, Trans. Am. Math. Soc. 85 (1957) 489–546.
[9] S. Karlin and J.L. McGregor, The classification of birth-and-death processes, Trans. Am. Math. Soc. 86 (1957) 366–400.
[10] S. Karlin and J.L. McGregor, Many server queueing processes with Poisson input and exponential service times, Pacific J. Math. 8 (1958) 87–118.
[11] J. Keilson, A review of transient behavior in regular diffusion and birth–death processes – Part II, J. Appl. Probability 2 (1965) 405–428.
[12] J. Keilson, A technique for discussing the passage time distribution for stable systems, J. Roy. Statist. Soc. B28 (1966) 477–486.
[13] J. Keilson, Log-concavity and log-convexity in passage time densities of diffusion and birth–death processes, J. Appl. Probability 8 (1971) 391–398.
[14] J.H.B. Kemperman, An analytical approach to the differential equations of the birth-and-death process, Michigan Math. J. 9 (4) (1962).
[15] D.G. Kendall, Unitary dilations of Markov transition operators and the corresponding integral representations for transition-probability matrices, in: U. Grenander, ed., Probability and statistics (Almqvist and Wiksell, Stockholm, 1959).
[16] D.G. Kendall, Unitary dilations of one-parameter semigroups of Markov transition operators, and the corresponding integral representations for Markov processes with a countable infinity of states, Proc. London Math. Soc. (3) 9 (1959) 417–431.
[17] J.F.C. Kingman, The exponential decay of Markov transition probabilities, Proc. London Math. Soc. (3) 13 (1963) 337–358.
[18] J.F.C. Kingman Ergodic properties of continuous-time Markov processes and their discrete skeletons, Proc. London Math. Soc. (3) 13 (1963) 593–604.
[19] J F.C. Kingman, Markov population processes, J. Appl. Probability 6 (1969) 1–18.
[20] W. Lederman and G.E.H. Reuter, Spectral theory for the differential equations of simple birth-and-death processes, Philos. Trans. Roy. Soc. London Ser. A 246 (1954) 321–369.

[21] A.W. Marshall and F. Proschan, Mean life of series and parallel systems, J. Appl. Probability 7 (1970) 165–174.
[22] D.S. Mitrinović, Analytic inequalities (Springer, Berlin, 1970).
[23] G.E.H. Reuter, Denumerable Markov processes and the associated contraction semigroups on ℓ, Acta Math. 97 (1957) 1–46.
[24] D. Vere-Jones, Geometric ergodicity in denumerable Markov chains, Quart. J. Math. Oxford 13 (1962) 7–28.
[25] H.S. Wall, Analytic theory of continued fractions (Van Nostrand, New York, 1948).
[26] D.V. Widder, The Laplace transform (Princeton Univ. Press, Princeton, N.J., 1946).