J. Appl. Prob. **51,** 756–768 (2014)
Printed in England

EXISTENCE AND UNIQUENESS OF A QUASISTATIONARY DISTRIBUTION FOR MARKOV PROCESSES WITH FAST RETURN FROM INFINITY

SERVET MARTÍNEZ * AND
JAIME SAN MARTÍN,* *Universidad de Chile*
DENIS VILLEMONAIS,** *Université de Lorraine*

Abstract

We study the long-time behaviour of a Markov process evolving in $\mathbb{N}$ and conditioned not to hit 0. Assuming that the process comes back quickly from ∞, we prove that the process admits a unique *quasistationary distribution* (in particular, the distribution of the conditioned process admits a limit when time goes to ∞). Moreover, we prove that the distribution of the process converges exponentially fast in the total variation norm to its quasistationary distribution and we provide a bound for the rate of convergence. As a first application of our result, we bring a new insight on the speed of convergence to the quasistationary distribution for birth-and-death processes: we prove that starting from any initial distribution the conditional probability converges in law to a unique distribution ρ supported in $\mathbb{N}^*$ if and only if the process has a unique quasistationary distribution. Moreover, ρ is this unique quasistationary distribution and the convergence is shown to be exponentially fast in the total variation norm. Also, considering the lack of results on quasistationary distributions for nonirreducible processes on countable spaces, we show, as a second application of our result, the existence and uniqueness of a quasistationary distribution for a class of possibly nonirreducible processes.

Keywords: Process with absorption; quasistationary distribution; Yaglom limit; mixing property; birth-and-death process

2010 Mathematics Subject Classification: Primary 37A25; 60B10; 60F99
Secondary 60J80

1. Introduction

Let X be a stable continuous-time Markov process evolving in $\mathbb{N} = \{0, 1, 2, \dots\}$ such that 0 is an absorbing point, so $X_t = 0$ for all $t \geq T_0$, and absorption occurs almost surely, that is, for all $x \in \mathbb{N}$, $\mathbb{P}_x(T_0 < +\infty) = 1$, where $T_0 = \inf\{s \geq 0,\ X_s = 0\}$. In this paper we provide a sufficient condition for the existence and uniqueness of a quasistationary distribution for X and for the conditional distribution of X to converge exponentially fast to it.

A *quasistationary distribution* (QSD) for X is a probability measure ρ on $\mathbb{N}^* = \{1, 2, 3, \dots\}$ such that, for all $t \geq 0$,

$$\rho(\cdot) = \mathbb{P}_\rho(X_t \in \cdot \mid t < T_0).$$

Received 23 February 2012; revision received 11 September 2013.
* Postal address: Departemento Ingeniería, Matemática and Centro de Modelamiento Matemático, Universidad de Chile, UMI 2807, CNRS, Universidad de Chile, Santiago, Chile.
** Postal address: Institut Élie Cartan de Lorraine, Université de Lorraine, Site de Nancy, B.P. 70239, F-54506 Vandoeuvre-lès- Nancy Cedex, France. Email address: denis.villemonais@inria.fr

Thus, a QSD is stationary for the process conditioned not to be absorbed. The notion of a QSD has always been closely related to the study of the long-time behaviour of a process conditioned not to be absorbed. Indeed, it is well known (see, for instance, [6] and [9]) that a probability measure ρ is a QSD if and only if it is a *quasilimiting distribution* (QLD), which means that there exists a probability measure μ on $\mathbb{N}^*$ such that

$$\rho(\cdot) = \lim_{t\to\infty} \mathbb{P}_\mu(X_t \in \cdot \mid t < T_0). \tag{1}$$

Existence and uniqueness of QSDs and QLDs have been extensively studied in the past decades. They were originally investigated by Yaglom [11], who stated their existence for subcritical Galton–Watson processes. In their seminal work [1], Darroch and Seneta proved that irreducible finite state space processes admit a unique QSD. In our case of a process X evolving in a countable state space, the question is more intricate since the existence or uniqueness of a QSD is not always true. In 1995, Ferrari *et al.* [4] proved a necessary and sufficient condition for the existence of a QSD for X under the assumption that it is irreducible and that the process does not come back from ∞ in finite time. More precisely, the authors proved that if $\mathbb{N}^*$ is an irreducible class for the process X and if $\lim_{x\to+\infty} \mathbb{P}_x(T_0 < t) = 0$ for any $t > 0$, then the existence of a QSD for X is equivalent to $\mathbb{E}_x(\mathrm{e}^{\lambda T_0}) < +\infty$ for some constants $x \in \mathbb{N}^*$ and $\lambda > 0$. The much-studied birth-and-death processes are of particular interest, since explicit sufficient and necessary conditions have been proved by van Doorn [7] characterising the three possible cases: there is no QSD, a unique QSD, or an infinite continuum of QSDs. (For more information on QSDs/QLDs, we refer the reader to the recent surveys [6] and [8].) In this paper we give a sufficient criterion for the existence and uniqueness of a QSD for countable state space processes. In the particular case of birth-and-death processes, we shall see that the criterion is in fact equivalent to the existence and uniqueness of a QSD.

While the existence of a QSD is interesting in itself, it is only the first step towards the understanding of a conditioned process' long-time behaviour. Indeed, it is of first practical importance to determine the initial distributions μ for which convergence (1) holds and, as stressed in [6], to determine the speed of convergence to the QSD. In the present paper, our aim is twofold: we give a criterion ensuring the existence and uniqueness of a QSD, and we prove that the conditional distribution of the process converges exponentially fast in the total variation norm to a unique QSD. Moreover, we provide an explicit upper bound for the rate of convergence, in terms of the constants c_1, c_2, and c_3 appearing in Hypotheses H1, H2, and H3 below. More precisely, we prove that there exists a unique QSD ρ and a constant $\gamma \in (0, 1)$ such that

$$\|\mathbb{P}_\mu(X_t \in \cdot \mid t < T_0) - \rho(\cdot)\|_{\mathrm{TV}} \le 2(1-\gamma)^{[t]} \quad \text{for all } \mu \in \mathcal{M}_1(\mathbb{N}^*) \text{ and all } t \ge 0,$$

where $\|\cdot\|_{\mathrm{TV}}$ denotes the total variation norm for signed measures, $[t]$ is the integer part of t, and $\mathcal{M}_1(\mathbb{N}^*)$ refers to the set of probability measures on $\mathbb{N}^*$. As we shall see, our proof uses a purely probabilistic approach, allowing us to answer the long standing question of the speed of convergence of a birth-and-death process to its unique QSD. Spectral tools give in this situation an exponential convergence when there is a second gap and when the initial distribution is finitely supported (see [7]).

The existence and uniqueness criterion is based on the three following hypotheses, where the positive constants c_1, c_2, c_3, and c_4 will appear in the expression of γ. Our first hypothesis states that there exists a subset of $\mathbb{N}^*$ where the probability of extinctions at any time t are balanced.

Hypothesis H1. *There exist $K \subset \mathbb{N}$ and a constant $c_1 > 0$ such that, for all $t \geq 0$,*

$$\frac{\inf_{x \in K} \mathbb{P}_x(t < T_0)}{\sup_{x \in K} \mathbb{P}_x(t < T_0)} \geq c_1.$$

Remark. One can check that when the process is irreducible, that is, when $\mathbb{P}_x(X_t = y) > 0$ for all $x, y \in \mathbb{N}^*$, this property is fulfilled for any finite subset $K \subset \mathbb{N}^*$. Note also that the smaller the subset K, the weaker the requirement on the constant c_1.

Let K satisfy Hypothesis H1. Our second hypothesis states that the process comes back quickly from any point to $K \cup \{0\}$ and, starting from some particular point in K, it has a relatively high probability to be in K afterwards. We denote by $T_K = \inf\{t \geq 0,\ X_t \in K\}$ the hitting time of K.

Hypothesis H2. *There exist some constants $\lambda_0 > 0$, $c_2 > 0$, and $c_3 > 0$, and a point $x_0 \in K$ such that, for all $t \geq 0$,*

$$\sup_{x \in \mathbb{N}^*} \mathbb{E}_x(\mathrm{e}^{\lambda_0 T_K \wedge T_0}) \leq c_2 \quad \textit{and} \quad \mathbb{P}_{x_0}(X_t \in K) \geq c_3 \mathrm{e}^{-\lambda_0 t}.$$

Remark. Usually, there exists an interval of values of λ_0 acceptable here, as it will clearly appear in the birth-and-death case (see the proof of Theorem 2 below). Note also that the larger the subset K, the weaker the requirements on the constants λ_0, c_2, and c_3. When K is finite and Hypothesis H2 holds for $\lambda_0 > 0$, then necessarily λ_0 is greater than or equal to Kingman's decay parameter.

Our last hypothesis states that the conditioned process comes back in time 1 to a point $x_0 \in K$ with a minimal probability.

Hypothesis H3. *There exists a constant $c_4 > 0$ and a point $x_0 \in K \subset E$ such that*

$$\inf_{x \in \mathbb{N}^*} \mathbb{P}_x(X_1 = x_0 \mid T_0 > 1) \geq c_4.$$

Remark. If the rate of absorption is uniformly bounded over $\mathbb{N}^*$ then $\inf_{x \in \mathbb{N}^*} \mathbb{P}_x(T_0 > 1) > 0$, and, thus, Hypothesis H3 is equivalent to the existence of x_0 and c_4 such that $\inf_{x \in \mathbb{N}^*} \mathbb{P}_x(X_1 = x_0) \geq c_4$. This is closely related to the existence of a *small set*, following the terminology of Down *et al.* [2] for processes without absorption, where $T_0 = +\infty$ happens $\mathbb{P}_x$-almost surely.

We are now able to state our main theorem, which is proved in Section 2. As an application, we also provide a corollary on birth-and-death processes, and show a generalization of the recent results of Ferrari and Mariĉ [3].

Theorem 1. *If Hypotheses H1, H2, and H3 are fulfilled, then there exists a unique QSD ρ for X. Moreover, for any probability measure μ on $\mathbb{N}^*$, we have*

$$\|\mathbb{P}_\mu(X_t \in \cdot \mid t < T_0) - \rho\|_{\mathrm{TV}} \leq 2\left(1 - \frac{c_1 c_3 c_4}{2c_2}\right)^{[t]} \quad \textit{for all } t \geq 0. \tag{2}$$

Remarks. (i) Inequality (2) implies that ρ is a QLD for X and any initial distribution, which means that, for any probability measure μ on $\mathbb{N}^*$,

$$\lim_{t \to \infty} \mathbb{P}_\mu(X_t \in \cdot \mid t < T_0) = \rho(\cdot).$$

(ii) Our approach is based on a strong mixing property inspired by Villemonais and Del Moral [10]. In particular, we prove that

$$\begin{aligned}&\|\mathbb{P}_\mu(X_t \in \cdot \mid t < T_0) - \mathbb{P}_\nu(X_t \in \cdot \mid t < T_0)\|_{\mathrm{TV}} \\ &\qquad \le 2\bigg(1 - \frac{c_1 c_3 c_4}{2c_2}\bigg)^{[t]} \quad \text{for all } \mu, \nu \in \mathcal{M}_1(\mathbb{N}^*) \text{ and all } t \ge 0.\end{aligned}$$

(iii) Note that c_1, c_2, c_3, and c_4 can be chosen in such a way that they satisfy $c_1c_3c_4/2c_2 < 1$. Nevertheless, as a consequence of the proof of Theorem 1 below, these constants always satisfy $c_1c_3c_4/2c_2 \le 1$ (see the argument after (5)).

We present two applications of our result. In Section 3 we develop the case of birth-and-death processes, and prove that such a process admits a unique QSD ρ if and only if Hypotheses H1, H2, and H3 are fulfilled, which implies that its conditional distribution converges exponentially fast to ρ, uniformly in its initial distribution. Note that this result provides new insight into the quasilimiting behaviour of birth-and-death processes. Moreover, its proof reveals that our criterion is optimal for a birth-and-death process: such a process satisfies Hypotheses H1, H2, and H3 if and only if it admits a unique QSD.

In our second application, developed in Section 4, we show that the sufficient condition for existence and uniqueness of a QSD proved in [3] can be considerably relaxed. While the practical implications of this application are nowadays less manifest than the previous application, it is of much theoretical interest. Indeed, it demonstrates that our result applies to reducible Markov processes on a countable state space, which is an exciting area under development where most of the existing results on QSDs do not apply.

2. Proof of Theorem 1

The proof of Theorem 1 is divided into three parts. In the first step we show that, for all $t \ge 0$,

$$\frac{\mathbb{P}_{x_0}(t < T_0)}{\sup_{x\in\mathbb{N}^*} \mathbb{P}_x(t < T_0)} \ge \frac{c_1 c_3}{2c_2}. \tag{3}$$

In the second step, using the techniques developed in [10], we prove inequality (2) for all $t \ge 0$. In the third step, we conclude the proof by showing that (2) implies the existence and uniqueness of a QSD.

Step 1: Prove that (3) holds. For all $x \in E$, we have

$$\mathbb{P}_x(t < T_0) = \mathbb{E}_x(\mathbf{1}_{\{t<T_K\wedge T_0\}}) + \mathbb{E}_x(\mathbf{1}_{\{T_K\le t<T_0\}}).$$

On the one hand, we deduce from Hypothesis H2 that, for all $t \ge 0$,

$$\begin{aligned}\mathbb{E}_x(\mathbf{1}_{\{t<T_K\wedge T_0\}}) &\le \mathrm{e}^{-\lambda_0 t}\mathbb{E}_x(\mathrm{e}^{\lambda_0 T_K\wedge T_0}) \\ &\le \frac{\mathbb{P}_{x_0}(X_t \in K)}{c_3} c_2 \\ &\le \frac{\mathbb{P}_{x_0}(X_t \in \mathbb{N}^*)}{c_3} c_2 \\ &= \frac{\mathbb{P}_{x_0}(t < T_0)}{c_3} c_2.\end{aligned}$$

https://doi.org/10.1239/jap/1409932672 Published online by Cambridge University Press

On the other hand, the Markov property and Hypothesis H2 yields

$$
\begin{aligned}
\mathbb{E}_x(\mathbf{1}_{\{T_K \le t < T_0\}}) &= \mathbb{E}_x(\mathbf{1}_{\{T_K \le t \le T_0\}}\, \mathbb{P}_{X_{T_K}}(t - T_K \le T_0)) \\
&= \mathbb{E}_x(\mathbf{1}_{\{T_K \le t \le T_0\}}\, \mathrm{e}^{\lambda_0 T_K \wedge T_0} \mathrm{e}^{-\lambda_0 T_K \wedge T_0} \mathbb{P}_{X_{T_K}}(t - T_K < T_0)) \\
&\le \mathbb{E}_x(\mathrm{e}^{\lambda_0 T_K \wedge T_0}) \sup_{y \in K} \sup_{s \in [0,t]} \mathrm{e}^{-\lambda_0 s} \mathbb{P}_y(t - s < T_0) \\
&\le c_2 \sup_{y \in K} \sup_{s \in [0,t]} \mathrm{e}^{-\lambda_0 s} \mathbb{P}_y(t - s < T_0).
\end{aligned}
$$

Now using Hypotheses H1 and H2, and the Markov property, we have, for all $s \in [0, t]$ and any $y \in K$,

$$
\mathrm{e}^{-\lambda_0 s} \mathbb{P}_y(t - s < T_0) \le \frac{\mathbb{P}_{x_0}(X_s \in K)}{c_3} \frac{\inf_{z \in K} \mathbb{P}_z(t - s < T_0)}{c_1} \le \frac{\mathbb{P}_{x_0}(t < T_0)}{c_1 c_3}.
$$

We deduce that

$$
\mathbb{E}_x(\mathbf{1}_{\{T_K \le t < T_0\}}) \le \frac{c_2}{c_1 c_3} \mathbb{P}_{x_0}(t < T_0).
$$

Finally, we have

$$
\mathbb{P}_x(t < T_0) \le \left(\frac{c_2}{c_3} + \frac{c_2}{c_1 c_3} \right) \mathbb{P}_{x_0}(t < T_0)
$$

which implies (3), since c_1 is necessarily smaller than 1.

Step 2: Prove that (2) holds. Let us define, for all $0 \le s \le t \le T$, the linear operator $R^T_{s,t}$ by

$$
R^T_{s,t} f(x) = \mathbb{E}_x(f(X_{t-s}) \mid T - s < T_0) = \mathbb{E}(f(X_t) \mid X_s = x,\ T < T_0),
$$

using the Markov property. We begin by proving that, for any $T > 0$, the family of operators $(R^T_{s,t})_{0 \le s \le t \le T}$ is a Markov semigroup. We have, for all $0 \le u \le s \le t \le T$,

$$
R^T_{u,s}(R^T_{s,t} f)(x) = \mathbb{E}_x(\mathbb{E}_{X_{s-u}}(f(X_{t-s}) \mid T - s < T_0) \mid T - u < T_0).
$$

For any measurable function g, the Markov property implies that

$$
\begin{aligned}
\mathbb{E}_x(g(X_{s-u})\, \mathbf{1}_{\{T-u<T_0\}}) &= \mathbb{E}_x(g(X_{s-u}) \mathbb{P}_{X_{s-u}}(T - u - (s - u) < T_0)) \\
&= \mathbb{E}_x(g(X_{s-u}) \mathbb{P}_{X_{s-u}}(T - s < T_0)).
\end{aligned}
$$

Applying this equality to $g\colon y \mapsto \mathbb{E}_x(f(X_{t-s}) \mid T - s < T_0)$, we deduce that

$$
\begin{aligned}
R^T_{u,s}(R^T_{s,t} f)(x) &= \frac{\mathbb{E}_x(\mathbb{E}_{X_{s-u}}(f(X_{t-s})\, \mathbf{1}_{\{T-s<T_0\}}))}{\mathbb{P}_x(T - u < T_0)} \\
&= \frac{\mathbb{E}_x(f(X_{t-s+(s-u)})\, \mathbf{1}_{\{T-s+(s-u)<T_0\}})}{\mathbb{P}_x(T - u < T_0)} \\
&= R^T_{u,t} f(x),
\end{aligned}
$$

where we have used the Markov property a second time. Thus, the family $(R^T_{s,t})_{0 \le s \le t \le T}$ is a semigroup.

Let us now prove that, for any $s \le T - 1$, any $x \in \mathbb{N}^*$, and $f \ge 0$,

$$
R^T_{s,s+1} f(x) \ge \frac{c_4 c_1 c_3}{2 c_2} f(x_0). \tag{4}
$$

In other words, we prove that $c_4c_1c_3/2c_2$ is a Dobrushin coefficient, which will allow us to show that inequality (2) holds. We have

$$\begin{aligned}\mathbb{P}_x(T-s<T_0)R^T_{s,s+1}f(x) &= \mathbb{E}_x(f(X_1)\,\mathbf{1}_{\{T-s<T_0\}})\\ &\geq f(x_0)\mathbb{P}_x(X_1=x_0,\ T-s<T_0)\\ &\geq f(x_0)\mathbb{E}_x(\mathbf{1}_{\{X_1=x_0\}}\,\mathbb{P}_{x_0}(T-s-1<T_0)),\end{aligned}$$

by the Markov property. We infer from (3) that $\mathbb{P}_{x_0}(T-s-1<T_0)\geq(c_1c_3/2c_2)\sup_{y\in\mathbb{N}^*}\mathbb{P}_y(T-s-1<T_0)$. But Hypothesis H3 yields

$$\mathbb{P}_x(X_1=x_0)\geq c_4\mathbb{P}_x(1<T_0),$$

which implies that

$$\begin{aligned}\mathbb{P}_x(T-s<T_0)R^T_{s,s+1}f(x) &\geq f(x_0)c_4\mathbb{P}_x(1<T_0)\frac{c_1c_3}{2c_2}\sup_{y\in\mathbb{N}^*}\mathbb{P}_y(T-s-1<T_0)\\ &\geq \frac{c_4c_1c_3}{2c_2}f(x_0)\mathbb{P}_x(T-s<T_0);\end{aligned}$$

thus, (4) holds.

We are now able to prove inequality (2). For any pair of orthogonal probability measures μ_1 and μ_2 on $\mathbb{N}^*$ and any $f\geq 0$, we have, by (4),

$$\mu_i R^T_{s,s+1}f\geq\frac{c_4c_1c_3}{2c_2}f(x_0)\quad\text{for } i=1,2. \tag{5}$$

Thus, $\mu_iR^T_{s,s+1}-c_4c_1c_3\delta_{x_0}/2c_2$ is a positive measure whose weight is smaller than the constant $1-c_4c_1c_3/2c_2$. We deduce that

$$\begin{aligned}&\|\mu_1R^T_{s,s+1}-\mu_2R^T_{s,s+1}\|_{\mathrm{TV}}\\ &\quad\leq\left\|\left(\mu_1R^T_{s,s+1}-\frac{c_4c_1c_3}{2c_2}\delta_{x_0}\right)\right\|_{\mathrm{TV}}+\left\|\left(\mu_2R^T_{s,s+1}-\frac{c_4c_1c_3}{2c_2}\delta_{x_0}\right)\right\|_{\mathrm{TV}}\\ &\quad\leq 2\left(1-\frac{c_4c_1c_3}{2c_2}\right)\\ &\quad=\left(1-\frac{c_4c_1c_3}{2c_2}\right)\|\mu_1-\mu_2\|_{\mathrm{TV}}.\end{aligned}$$

If μ_1 and μ_2 are two different but not orthogonal probability measures, we can apply the previous result to the orthogonal probability measures $(\mu_1-\mu_2)_+/(\mu_1-\mu_2)_+(\mathbb{N}^*)$ and $(\mu_1-\mu_2)_-/(\mu_1-\mu_2)_-(\mathbb{N}^*)$. Then

$$\begin{aligned}&\left\|\frac{(\mu_1-\mu_2)_+}{(\mu_1-\mu_2)_+(\mathbb{N}^*)}R^T_{s,s+1}-\frac{(\mu_1-\mu_2)_-}{(\mu_1-\mu_2)_-(\mathbb{N}^*)}R^T_{s,s+1}\right\|_{\mathrm{TV}}\\ &\qquad\leq\left(1-\frac{c_4c_1c_3}{2c_2}\right)\left\|\frac{(\mu_1-\mu_2)_+}{(\mu_1-\mu_2)_+(\mathbb{N}^*)}-\frac{(\mu_1-\mu_2)_-}{(\mu_1-\mu_2)_-(\mathbb{N}^*)}\right\|_{\mathrm{TV}}.\end{aligned}$$

But $(\mu_1-\mu_2)_+(\mathbb{N}^*)=(\mu_1-\mu_2)_-(\mathbb{N}^*)$ since $\mu_1(\mathbb{N}^*)=\mu_2(\mathbb{N}^*)=1$. Then, multiplying this inequality by $(\mu_1-\mu_2)_+(\mathbb{N}^*)$, we deduce that

$$\begin{aligned}&\|(\mu_1-\mu_2)_+R^T_{s,s+1}-(\mu_1-\mu_2)_-R^T_{s,s+1}\|_{\mathrm{TV}}\\ &\qquad\leq\left(1-\frac{c_4c_1c_3}{2c_2}\right)\|(\mu_1-\mu_2)_+-(\mu_1-\mu_2)_-\|_{\mathrm{TV}}.\end{aligned}$$

https://doi.org/10.1239/jap/1409932672 Published online by Cambridge University Press

Since $(\mu_1 - \mu_2)_+ - (\mu_1 - \mu_2)_- = \mu_1 - \mu_2$, we obtain

$$\|\mu_1 R^T_{s,s+1} - \mu_2 R^T_{s,s+1}\|_{\mathrm{TV}} \le \left(1 - \frac{c_4 c_1 c_3}{2c_2}\right)\|\mu_1 - \mu_2\|_{\mathrm{TV}}.$$

In particular, using the semigroup property of $(R^T_{s,t})_{s,t}$, we deduce that, for any $x, y \in \mathbb{N}^*$,

$$\begin{aligned}
\|\delta_x R^T_{0,T} - \delta_y R^T_{0,T}\|_{\mathrm{TV}} &= \|\delta_x R^T_{0,T-1} R^T_{T-1,T} - \delta_y R^T_{0,T-1} R^T_{T-1,T}\|_{\mathrm{TV}} \\
&\le \left(1 - \frac{c_4 c_1 c_3}{2c_2}\right)\|\delta_x R^T_{0,T-1} - \delta_y R^T_{0,T-1}\|_{\mathrm{TV}} \\
&\le 2\left(1 - \frac{c_4 c_1 c_3}{2c_2}\right)^{[T]},
\end{aligned}$$

by induction, where $[T]$ denotes the integer part of T. Inequality (2) is thus proved for any pair of initial probability measures (δ_x, δ_y), with $(x, y) \in \mathbb{N}^* \times \mathbb{N}^*$.

Let us now prove that the inequality extends to any couple of initial probability measures. Let μ be a probability measure on $\mathbb{N}^*$ and $x \in \mathbb{N}^*$. We have

$$\begin{aligned}
&\|\mathbb{P}_\mu(X_T \in \cdot \mid T < T_0) - \mathbb{P}_x(X_T \in \cdot \mid T < T_0)\|_{\mathrm{TV}} \\
&\quad = \frac{1}{\mathbb{P}_\mu(T < T_0)}\|\mathbb{P}_\mu(X_T \in \cdot) - \mathbb{P}_\mu(T < T_0)\mathbb{P}_x(X_T \in \cdot \mid T < T_0)\|_{\mathrm{TV}} \\
&\quad \le \frac{1}{\mathbb{P}_\mu(T < T_0)} \sum_{y\in\mathbb{N}^*} \mu(y)\|\mathbb{P}_y(X_T \in \cdot) - \mathbb{P}_y(T < T_0)\mathbb{P}_x(X_T \in \cdot \mid T < T_0)\|_{\mathrm{TV}} \\
&\quad \le \frac{1}{\mathbb{P}_\mu(T < T_0)} \sum_{y\in\mathbb{N}^*} \mu(y)\mathbb{P}_y(T < T_0)\|\mathbb{P}_y(X_T \in \cdot \mid T < T_0) \\
&\qquad\qquad - \mathbb{P}_x(X_T \in \cdot \mid T < T_0)\|_{\mathrm{TV}} \\
&\quad \le \frac{1}{\mathbb{P}_\mu(T < T_0)} \sum_{y\in\mathbb{N}^*} \mu(y)\mathbb{P}_y(T < T_0) 2\left(1 - \frac{c_4 c_1 c_3}{2c_2}\right)^{[T]} \\
&\quad \le 2\left(1 - \frac{c_4 c_1 c_3}{2c_2}\right)^{[T]}.
\end{aligned}$$

The same procedure, replacing δ_x by any probability measure, leads to inequality (2).

Step 3: Prove that (2) *implies the existence and uniqueness of a QSD for X*. Let us first prove the uniqueness of the QSD. If ρ_1 and ρ_2 are two QSDs, then we have $\mathbb{P}_{\rho_i}(X_t \in \cdot \mid t < T_0) = \rho_i$ for $i = 1, 2$ and any $t \ge 0$. Thus, we deduce from inequality (2) that

$$\|\rho_1 - \rho_2\|_{\mathrm{TV}} \le 2\left(1 - \frac{c_4 c_1 c_3}{2c_2}\right)^{[t]} \quad \text{for all } t \ge 0,$$

which yields $\rho_1 = \rho_2$.

Let us now prove the existence of a QSD. By [6, Proposition 1], this is equivalent to proving the existence of a QLD for X (see the introduction). Hence, it is sufficient to prove the existence of a point $x \in \mathbb{N}^*$ such that $\mathbb{P}_x(X_t \in \cdot \mid t < T_0)$ converges when t goes to ∞.

Let $x \in \mathbb{N}^*$ be any point in $\mathbb{N}^*$. We have, for all $s, t \geq 0$,

$$\begin{aligned}
\|\mathbb{P}_x(X_t \in \cdot \mid t < T_0) &- \mathbb{P}_x(X_{t+s} \in \cdot \mid t+s < T_0)\|_{\mathrm{TV}} \\
&= \|\mathbb{P}_x(X_t \in \cdot \mid t < T_0) - \mathbb{P}_{\delta_x R^{t+s}_{s,t+s}}(X_t \in \cdot \mid t < T_0)\|_{\mathrm{TV}} \\
&\leq 2\bigg(1 - \frac{c_4 c_1 c_3}{2c_2}\bigg)^{[t]} \\
&\to 0 \quad \text{as } s, t \to +\infty.
\end{aligned}$$

Thus, any sequence $(\mathbb{P}_x(X_t \in \cdot \mid t < T_0))_{t\geq 0}$ is a Cauchy sequence for the total variation norm. But the space of probability measures on $\mathbb{N}^*$ equipped with the total variation norm is complete, so $\mathbb{P}_x(X_t \in \cdot \mid t < T_0)$ converges when t goes to ∞.

Finally, we have proved that there exists a unique QSD ρ for X. The last assertion of Theorem 1 is proved as follows: for any probability measure μ on $\mathbb{N}^*$, we have

$$\begin{aligned}
\|\mathbb{P}_\mu(X_t \in \cdot | t < T_0) - \rho\|_{\mathrm{TV}} &= \|\mathbb{P}_\mu(X_t \in \cdot | t < T_0) - \mathbb{P}_\rho(X_t \in \cdot | t < T_0)\|_{\mathrm{TV}} \\
&\leq 2\bigg(1 - \frac{c_4 c_1 c_3}{2c_2}\bigg)^{[t]} \\
&\to 0 \quad \text{as } t \to +\infty.
\end{aligned}$$

This concludes the proof of Theorem 1.

3. The birth-and-death process case

In this section we consider *birth-and-death processes*, which are widely used to describe the stochastic evolution of a population whose individuals are reproducing and dying at a rate depending on the population size. A process X on $\mathbb{N}$ is said to be a *birth-and-death process with absorption* if there exist two families of positive constants $(b_x)_{x\geq 1}$ and $(d_x)_{x\geq 1}$ such that the transition rate matrix $(Q(x, y))_{x,y\in\mathbb{N}}$ of X is given by

$$Q(x, y) = \begin{cases} b_x & \text{if } x \geq 1 \text{ and } y = x + 1, \\ d_x & \text{if } x \geq 1 \text{ and } y = x - 1, \\ 0 & \text{otherwise.} \end{cases}$$

The families $(b_x)_{x\geq 1}$ and $(d_x)_{x\geq 1}$ are respectively referred to as the family of birth rates and the family of death rates. Also, it is easy to check that 0 is an absorbing point for X.

Applying Theorem 1, we show that the conditional distribution of a birth-and-death process converges exponentially fast to a uniquely determined distribution (which is then a QSD) for any initial distribution if and only if it admits a unique QSD. Also, as shall be seen in the proof, Hypotheses H1, H2, and H3 are equivalent to the uniqueness of a QSD in the birth-and-death case.

We note that existence and uniqueness criteria for birth-and-death processes have been well known since the works of van Doorn [7] (also see Hart and Pollett [5]). Indeed, setting $T_z = \inf\{t \geq 0,\ X_t = z\}$, the author proved that a birth-and-death process has a unique QSD if and only if

$$S := \sup_{x\geq 1} \mathbb{E}_x(T_1) < +\infty,$$

https://doi.org/10.1239/jap/1409932672 Published online by Cambridge University Press

where S can be easily computed, since, for any $z \geq 1$,

$$\sup_{x \geq z} \mathbb{E}_x(T_z) = \sum_{k \geq z+1} \frac{1}{d_k \alpha_k} \sum_{l \geq k} \alpha_l,$$

with $\alpha_k = (\prod_{i=1}^{k-1} b_i)/(\prod_{i=1}^{k} d_i)$. However, the spectral theory tools used to prove this result are not well suited to study the speed at which the conditional distribution converges to the QSD. In particular, existing results do not provide speed of convergence to the QLD nor the set of initial distributions such that limit (1) holds. As illustrated by the numerical numerical computations in [6], the speed of convergence and its dependence on the initial distribution are of first practical importance to know whether or not the existence of a QSD is relevant for the dynamic of the process. As a consequence, the following result provides new insight into the quasilimiting behaviour of birth-and-death processes, completing the picture offered in [7].

Theorem 2. *For a birth-and-death process X, the following statements are equivalent.*

(i) *There exists a unique QSD.*

(ii) *There exists a distribution $\rho \in \mathcal{M}_1(\mathbb{N}^*)$ such that, for any $\mu \in \mathcal{M}_1(\mathbb{N}^*)$,*

$$\mathbb{P}_\mu(X_t \in \cdot \mid t < T_0) \to \rho(\cdot) \quad \textit{as } t \to \infty. \tag{6}$$

(iii) *There exist a distribution $\rho \in \mathcal{M}_1(\mathbb{N}^*)$ and $\gamma \in (0, 1)$ such that, for any $\mu \in \mathcal{M}_1(\mathbb{N}^*)$,*

$$\|\mathbb{P}_\mu(X_t \in \cdot \mid t < T_0) - \rho\|_{\mathrm{TV}} \leq 2(1-\gamma)^{[t]} \quad \textit{for all } t \geq 0. \tag{7}$$

Moreover, in (ii) and (iii) the distribution ρ is the unique QSD.

We emphasize that our proof also provides a purely probabilistic argument to the already known fact that $S < +\infty$ implies existence and uniqueness of a QSD, while earlier proofs rely on much more complex arguments based on the spectral decomposition of the rate matrix Q.

Proof of Thereom 2. Let X be a birth-and-death process. If (6) holds then ρ is a QLD for X starting from any initial distribution and is thus the unique QSD for X.

Let us now prove that the existence and uniqueness of a QSD for X implies that Hypotheses H1, H2, and H3 hold. This will imply (7) by Theorem 1 and, thus, conclude the proof of Theorem 2.

Since X is irreducible, Hypothesis H1 is satisfied for any finite subset $K \subset \mathbb{N}^*$.

Setting $x_0 = 1$ and $\lambda_0 = b_1 + d_1$, we have, for any subset $K \subset \mathbb{N}^*$ containing x_0 and any $t \geq 0$,

$$\mathbb{P}_{x_0}(X_t \in K) \geq \mathbb{P}_{x_0}(X_s = x_0 \text{ for all } s \in [0, t]) = \mathrm{e}^{-\lambda_0 t}.$$

Since the birth-and-death process X has a unique QSD, we have $S < +\infty$ (see, for instance, [7]). In particular, we deduce that, for any $\varepsilon > 0$, there exists $z_\varepsilon \geq 1$ such that

$$\sup_{x \geq z_\varepsilon} \mathbb{E}_x(T_{z_\varepsilon}) = \sum_{k \geq z_\varepsilon+1} \frac{1}{\partial_k \alpha_k} \sum_{l \geq k} \alpha_l \leq \varepsilon.$$

The Markov inequality thus implies that

$$\sup_{x \geq z_\varepsilon} \mathbb{P}_x(T_{z_\varepsilon} \geq 1) \leq \varepsilon,$$

https://doi.org/10.1239/jap/1409932672 Published online by Cambridge University Press

which together with a renewal argument yields

$$\sup_{x\geq z_\varepsilon} \mathbb{P}_x(T_{z_\varepsilon} \geq n) \leq \varepsilon^n \text{ for all } n \geq 1.$$

As a consequence, we deduce that there exists $z_0 \geq 1$ such that

$$\sup_{x\geq z_0} \mathbb{E}_x(\mathrm{e}^{\lambda_0 T_{z_0}}) < +\infty.$$

In particular, setting $K = \{1, 2, \dots, z_0\}$, we deduce that

$$\sup_{x\in\mathbb{N}^*} \mathbb{E}_x(\mathrm{e}^{\lambda_0 T_K \wedge T_0}) < +\infty.$$

Hence, Hypothesis H2 is satisfied.

Let us now prove that Hypothesis H3 is fulfilled. We have, for any fixed $z \in \mathbb{N}^*$,

$$\begin{aligned}
\inf_{x\in\mathbb{N}^*} \mathbb{P}_x(X_1 = x_0 \mid T_0 > 1) &\geq \inf_{x\in\mathbb{N}^*} \mathbb{P}_x(X_1 = x_0) \\
&\geq \inf_{x\in\mathbb{N}^*} \mathbb{P}_x(T_{x_0} \leq 1)\mathbb{P}_{x_0}(X_t = 1 \text{ for all } t \in [0, 1]) \\
&\geq \mathrm{e}^{-\lambda_0} \inf_{x\in\mathbb{N}^*} \mathbb{P}_x(T_{x_0} \leq 1) \\
&\geq \mathrm{e}^{-\lambda_0} \inf_{x\in\mathbb{N}^*} \mathbb{P}_x\big(T_z \leq \tfrac{1}{2}\big)\mathbb{P}_z\big(T_{x_0} \leq \tfrac{1}{2}\big),
\end{aligned}$$

where we have used the strong then the weak Markov properties. Now, by the Markov inequality, we have, for any $\varepsilon > 0$,

$$\sup_{x\geq z_\varepsilon} \mathbb{P}_x\big(T_{z_\varepsilon} \geq \tfrac{1}{2}\big) \leq 2 \sup_{x\geq z_\varepsilon} \mathbb{E}_x(T_{z_\varepsilon}) \leq 2\varepsilon.$$

Choosing, for instance, $\varepsilon = \frac{1}{4}$, we deduce for $z = z_{1/4}$ that

$$\inf_{x\in\mathbb{N}^*} \mathbb{P}_x(X_1 = x_0 \mid T_0 > 1) \geq \mathrm{e}^{-\lambda_0}\Big(\tfrac{1}{2} \wedge \inf_{x<z} \mathbb{P}_x\big(T_z \leq \tfrac{1}{2}\big)\Big)\mathbb{P}_z\big(T_{x_0} \leq \tfrac{1}{2}\big).$$

Since X is irreducible, we immediately deduce that both $\inf_{x<z} \mathbb{P}_x(T_z \leq \frac{1}{2})$ and $\mathbb{P}_z(T_{x_0} \leq \frac{1}{2})$ are positive. In particular, we have

$$\inf_{x\in\mathbb{N}^*} \mathbb{P}_x(X_1 = x_0 \mid T_0 > 1) > 0,$$

so Hypothesis H3 is fulfilled.

Finally, Hypotheses H1, H2, and H3 are satisfied and, thus, applying Theorem 1, there exists a constant $\gamma > 0$ and a probability measure ρ on $\mathbb{N}^*$ such that, for any initial distribution μ on $\mathbb{N}^*$,

$$\|\mathbb{P}_\mu(X_t \in \cdot \mid t < T_0) - \rho\|_{\mathrm{TV}} \leq 2(1-\gamma)^{[t]} \quad \text{for all } t \geq 0.$$

This concludes the proof of Theorem 2.

Remark. Theorem 2 contains as a special case the recent result of [12, Theorem 4.1].

4. A criterion on the transition rate matrix of the process

In this section we consider a stable nonexplosive process X, conservative in $\mathbb{N}^*$, and we give a sufficient criterion on its transition rate matrix Q for the existence and uniqueness of a QSD for X. This result is of first theoretical importance since it applies to nonirreducible Markov processes, for which the lack of QSD-related results is patent (see, for instance, [8]). Moreover, the assumptions of the following theorem can be checked directly on the transition matrix expression.

Theorem 3. *Let $(Q(x,y))_{x,y\in\mathbb{N}}$ be the transition rate matrix of X, and assume that there exists a finite subset $K \subset \mathbb{N}^*$ such that*

$$\inf_{y\in\mathbb{N}^*\setminus K}\left(Q(y,0)+\sum_{x\in K}Q(y,x)\right) > \sup_{y\in\mathbb{N}^*} Q(y,0) \tag{8}$$

and that $P_t(x,y) > 0$ for all $x, y \in K$. Then there exists a positive constant $\gamma \in (0,1)$ and $\rho \in \mathcal{M}_1(\mathbb{N}^*)$ such that, for any initial distribution μ on $\mathbb{N}^*$ and all $t \geq 0$,

$$\|\mathbb{P}_\mu(X_t \in \cdot \mid t < T_0) - \rho\|_{\mathrm{TV}} \leq 2(1-\gamma)^{[t]}.$$

In particular, ρ is the unique QSD associated to X.

Remark. Our result is a generalization of Theorem 1.1 of [3], in which X is assumed to be irreducible,

$$\overline{q} := \sup_x \sum_{y\in\mathbb{N}\setminus\{x\}} Q(x,y) < +\infty,$$

and

$$\alpha := \sum_{x\in\mathbb{N}^*}\inf_{y\in\mathbb{N}^*\setminus x} Q(y,x) > C =: \sup_{y\in\mathbb{N}^*} Q(y,0).$$

Indeed, these assumptions imply that there exists a finite subset $K \subset \mathbb{N}^*$ such that

$$\inf_{y\in\mathbb{N}^*}\sum_{x\in K\setminus\{y\}} Q(y,x) \geq \sum_{x\in K}\inf_{y\in\mathbb{N}^*\setminus\{x\}} Q(y,x) > C,$$

and, thus, imply inequality (8). Moreover, we implicitly allow $\overline{q} = +\infty$ and remove the irreducibility assumption.

Proof of Theorem 3. Let us prove that Hypotheses H1, H2, and H3 hold under the assumptions of Theorem 3.

By assumption, we have

$$\inf_{y\in\mathbb{N}^*\setminus K}\sum_{x\in K} Q(y,x) > 0.$$

It follows that $\inf_{y\in\mathbb{N}^*\setminus K}\mathbb{P}_y(T_K \leq \frac{1}{2}) > 0$ and then

$$\inf_{y\in\mathbb{N}^*}\mathbb{P}_y\big(T_K \leq \tfrac{1}{2}\big) > 0.$$

Fix $x_0 \in K$. Since K is finite, we have, by assumption,

$$\min_{x\in K} P_x(X_{1/2} = x_0) > 0.$$

https://doi.org/10.1239/jap/1409932672 Published online by Cambridge University Press

Using the strong Markov property, we deduce from the last two inequalities that

$$\inf_{y\in\mathbb{N}^*} \mathbb{P}_y\big(T_{x_0} \in \big[\tfrac{1}{2}, 1\big]\big) > 0.$$

But the process is assumed to be stable, so that it remains in x_0 during at least a time $\frac{1}{2}$ with positive probability. We finally deduce that

$$\inf_{y\in\mathbb{N}^*} \mathbb{P}_y(X_1 = x_0) > 0, \tag{9}$$

which implies Hypothesis H3.

Since K is finite, for any $t \geq 0$, there exists $x_t^{\max} \in K$ such that

$$\mathbb{P}_{x_t^{\max}}(t < T_0) = \max_{x\in K} \mathbb{P}_x(t < T_0).$$

For $t \geq 1$, the Markov property yields

$$\begin{aligned}
\mathbb{P}_x(t < T_0) &\geq \mathbb{P}_x(X_1 = x_t^{\max})\mathbb{P}_{x_{t-1}^{\max}}(t-1 < T_0)\\
&\geq \mathbb{P}_x(X_1 = x_t^{\max})\mathbb{P}_{x_t^{\max}}(t < T_0)\\
&\geq \min_{x',x''\in K} \mathbb{P}_{x'}(X_1 = x'')\, \mathbb{P}_{x_t^{\max}}(t < T_0).
\end{aligned}$$

But K is finite; thus, we have, by assumption, $\min_{x',x''\in K} \mathbb{P}_{x'}(X_1 = x'') > 0$. Finally, we deduce that

$$\inf_{t\geq 1} \frac{\min_{x\in K} \mathbb{P}_x(t < T_0)}{\max_{x\in K} \mathbb{P}_x(t < T_0)} \geq \min_{x',x''\in K} \mathbb{P}_{x'}(X_1 = x'') > 0.$$

Now, for $t \in [0, 1]$, we have

$$\inf_{t\in[0,1]} \frac{\min_{x\in K} \mathbb{P}_x(t < T_0)}{\sup_{x\in K} \mathbb{P}_x(t < T_0)} \geq \min_{x\in K} \mathbb{P}_x(X_s = x \text{ for all } s \in [0, 1]) > 0.$$

Thus, we deduce that Hypothesis H1 is fulfilled.

Since the absorption rate of the process is uniformly bounded by C, we have

$$\mathbb{P}_{x_0}(X_{t-1} \in \mathbb{N}^*) \geq \mathrm{e}^{-C(t-1)}.$$

By the Markov property, we deduce that

$$\mathbb{P}_{x_0}(X_t = x_0) \geq \inf_{y\in\mathbb{N}^*} \mathbb{P}_y(X_1 = x_0)\mathrm{e}^{-C(t-1)}.$$

In particular, setting $\lambda_0 = C$ and using (9), we deduce that the second point of Hypothesis H2 is fulfilled.

Let us now set

$$\alpha_K := \inf_{y\in\mathbb{N}^*\setminus K} \bigg(Q(y, 0) + \sum_{x\in K} Q(y, x)\bigg).$$

The process jumps into $K \cup \{0\}$ from any point $x \notin K \cup \{0\}$ with a rate bigger than α_K. This implies that $T_K \wedge T_0$ is uniformly bounded above by an exponential time of rate α_K. In particular, we have

$$\sup_{x\in\mathbb{N}^*} \mathbb{E}_x(\mathrm{e}^{CT_K\wedge T_0}) \leq \frac{\alpha_K}{\alpha_K - C} < \infty,$$

since $\alpha_K > C$ by assumption. As a consequence, the first part of Hypothesis H2 is also fulfilled with $\lambda_0 = C$.

This and Theorem 1 allows us to complete the proof of Theorem 3.

Acknowledgements

This work was partially done during the visit of D. Villemonais to the CMM. The authors acknowledge the support from Basal-Conicyt, MATH-Amsud, and the TOSCA team (INRIA Nancy – Grand Est). We thank an anonymous referee for useful comments and suggestions, and for pointing out reference [12].

References

[1] Darroch, J. N. and Seneta, E. (1967). On quasi-stationary distributions in absorbing continuous-time finite Markov chains. *J. Appl. Prob.* **4,** 192–196.
[2] Down, D., Meyn, S. P. and Tweedie, R. L. (1995). Exponential and uniform ergodicity of Markov processes. *Ann. Prob.* **23,** 1671–1691.
[3] Ferrari, P. A. and Marić, N. (2007). Quasi stationary distributions and Fleming–Viot processes in countable spaces. *Electron. J. Prob.* **12,** 684–702.
[4] Ferrari, P. A., Kesten, H., Martínez, S. and Picco, P. (1995). Existence of quasi-stationary distributions. A renewal dynamical approach. *Ann. Prob.* **23,** 501–521.
[5] Hart, A. G. and Pollett, P. K. (1996). Direct analytical methods for determining quasistationary distributions for continuous-time Markov chains. In *Athens Conference on Applied Probability and Time Series Analysis* (Lecture Notes Statist. **114**), Vol. I, Springer, New York, pp. 116–126.
[6] Méléard, S. and Villemonais, D. (2012). Quasi-stationary distributions and population processes. *Prob. Surveys* **9,** 340–410.
[7] Van Doorn, E. A. (1991). Quasi-stationary distributions and convergence to quasi-stationarity of birth–death processes. *Adv. Appl. Prob.* **23,** 683–700.
[8] Van Doorn, E. A. and Pollett, P. K. (2013). Quasi-stationary distributions for discrete state models. *Europ. J. Operat. Res.* **230,** 1–14.
[9] Vere-Jones, D. (1969). Some limit theorems for evanescent processes. *Austral. J. Statist.* **11,** 67–78.
[10] Villemonais, D. and Del Moral, P. (2011). Strong mixing properties for time inhomogeneous diffusion processes with killing. *Distributions quasi-stationnaires et méthodes particulaires pour l'approximation de processus conditionnés*, Doctoral Thesis (D. Villemonais), pp. 149–166.
[11] Yaglom, A. M. (1947). Certain limit theorems of the theory of branching random processes. *Doklady Akad. Nauk SSSR (N.S.)* **56,** 795–798 (in Russian).
[12] Zhang, H. and Zhu, Y. (2013). Domain of attraction of the quasistationary distribution for birth-and-death processes. *J. Appl. Prob.* **50,** 114–126.