The Annals of Probability
1995, Vol. 23, No. 2, 501–521

EXISTENCE OF QUASI-STATIONARY DISTRIBUTIONS. A RENEWAL DYNAMICAL APPROACH

BY P. A. FERRARI, H. KESTEN, S. MARTINEZ AND P. PICCO

Universidade de São Paulo, Cornell University, Universidad de Chile and CNRS

We consider Markov processes on the positive integers for which the origin is an absorbing state. Quasi-stationary distributions (qsd's) are described as fixed points of a transformation Φ in the space of probability measures. Under the assumption that the absorption time at the origin, R, of the process starting from state x goes to infinity in probability as $x \to \infty$, we show that the existence of a qsd is equivalent to $E_x e^{\lambda R} < \infty$ for some positive λ and x. We also prove that a subsequence of $\Phi^n \delta_x$ converges to a minimal qsd. For a birth and death process we prove that $\Phi^n \delta_x$ converges along the full sequence to the minimal qsd. The method is based on the study of the renewal process with interarrival times distributed as the absorption time of the Markov process with a given initial measure μ. The key tool is the fact that the residual time in that renewal process has as stationary distribution the distribution of the absorption time of $\Phi\mu$.

1. Introduction. Let $X(t)$ be a continuous time Markov process in $\mathbb{N} = \{0, 1, 2, \ldots\}$. Let Q be the corresponding transition rate matrix: $q(x, y)$ is the rate of jumping from x to $y \neq x$. We assume throughout that the state 0 is absorbing [i.e., $q(0, x) = 0$ for $x \in \mathbb{N}$] and that all states other than 0 form an irreducible class, and finally that Q is conservative and regular, that is,

$$-q(x, x) = \sum_{y \neq x} q(x, y) < \infty, \text{ and the minimal process} \quad (1.1)$$
$$\{x(t)\}_{t \geq 0} \text{ corresponding to } Q \text{ is an honest process.}$$

In particular, (1.1) says that

$$P_\mu(X(t) \in \mathbb{N}) = 1 \quad \text{for all } t \geq 0 \text{ and all initial distributions } \mu \text{ on } \mathbb{N},$$

and, consequently, any right continuous strong Markov process with transition rate matrix Q equals the minimal process with probability 1. In this paper, $X(t)$ always denotes the minimal process for Q. This process can then be constructed explicitly as a nice jump process [see, for instance, Asmussen (1987), Section 2.2 or Breiman (1968), Section 15.6].

We further define

$$R = \inf\{t \geq 0 \colon X(t) = 0\},$$

the absorption time at 0. We shall only be interested in processes for which $E_x R < \infty$ for all $x \geq 1$.

Received July 1993; revised March 1994.
AMS 1991 *subject classifications*. Primary 60J27, 60J80; secondary 60K05.
Key words and phrases. Quasi-stationary distributions, renewal processes, residual time.

A quasi-stationary distribution (qsd) μ is a probability measure on $\{1, 2, \dots\}$ with the property that, starting with μ, the conditional distribution, given the event that at time t the process has not been absorbed, is still μ. That is,

$$\frac{\sum_{y \geq 1} \mu(y) P_y(X(t) = x)}{\sum_{y \geq 1} \mu(y) P_y(X(t) \neq 0)} = \mu(x). \tag{1.2}$$

Quasi-stationary distributions for Markov processes and chains have been studied by several authors. Vere-Jones (1962), Seneta and Vere-Jones (1966) and Kingman (1963) studied the case of a general denumerable state space.

Intimately related with the notion of qsd is the so-called Yaglom limit:

$$\lim_{t \to \infty} \frac{P_y(X(t) = x)}{P_y(X(t) \neq 0)}. \tag{1.3}$$

Indeed, if the Yaglom limit exists and it is a probability measure, then it is a qsd [see Vere-Jones (1969), Theorems 1 and 2]. The existence of the limit in (1.3) was established for branching processes in a pioneer work by Yaglom (1947), for asymmetric random walks by Seneta (1966), for left continuous random walks by Pakes (1973) and for birth and death processes by Good (1968), Kijima and Seneta (1991), van Doorn (1991) and van Doorn and Schrijner (1994b). In the case of a finite Markov chain there is only one qsd and it is the Yaglom limit; see Darroch and Seneta (1965).

In this paper we study qsd's by means of the distribution of the absorption time R. Our main result is as follows.

THEOREM 1.1. *Assume that* (1.1) *holds and that* Q *restricted to* $\{1, 2, \dots\}$ *is irreducible. Assume further that*

$$\lim_{x \to \infty} P_x(R < t) = 0 \quad \text{for any } t \geq 0 \tag{1.4}$$

and that $P_x(R < \infty) = 1$ *for some (and hence all)* x. *Then a necessary and sufficient condition for the existence of a qsd is that*

$$E_x e^{\lambda R} < \infty \tag{1.5}$$

for some $\lambda > 0$ *and for some* $x \in \{1, 2, \dots\}$ *(and hence for all* x*).*

One can see immediately that condition (1.5) is necessary. In fact, if there exists a qsd μ, then $P_\mu(R > s + t) = P_\mu(R > s) P_\mu(R > t)$. So F^μ is exponential. Since $P_x(R < \infty) = 1$ for all x, also $P_\mu(R < \infty) = 1$ and consequently the exponentially distributed R must have $E_\mu R < \infty$. Then for $\lambda < 1/E_\mu R$, we have $E_\mu e^{\lambda R} < \infty$. Since the irreducibility of Q implies that μ charges every point and $E_x e^{\lambda R} \leq (1/\mu(x)) E_\mu e^{\lambda R}$, we deduce $E_x e^{\lambda R} < \infty$. Sufficiency of (1.5) will be proven in Section 4.

For a birth and death process, Theorem 1.1 was proven by van Doorn (1991), by Ferrari, Martinez and Picco (1992) and by van Doorn and Schrijner (1994a).

We notice that (1.5) is equivalent to exponential decay of the tail of the distribution of F^{δ_x}, that is,

(1.6) there exist positive constants C, γ such that $P_x(R > t) \leq Ce^{-\gamma t}$.

It is worth noting that if one defines

$$\lambda_1 = \lim_{t\to\infty} \frac{-1}{t} \log(P_x(X(t) = y)) \tag{1.7}$$

(this is independent of $x, y \geq 1$; see Lemma 5.2 for more details), then processes satisfying (1.4) and (1.5) can be λ_1-recurrent or λ_1-transient. That is, there are examples for which

$$\int_0^\infty e^{\lambda_1 t} P_x(X(t) = x)\, dt \tag{1.8}$$

diverges and other examples for which it converges. Thus, our results seem to have little connection with λ_1-recurrence, which is the basis for much of the previous work on qsd's. For instance, Seneta (1966) proved that (1.8) converges for the continuous time random walk on $\mathbb{N}$ which moves one step to the right and left at rates $p \in (0, \frac{1}{2})$ and $q = 1 - p$, respectively (with absorption at 0). An example for which (1.8) diverges is provided by a continuous time subcritical Markov branching process. If one starts with one particle, then

$$P_1(R > t) = P(\text{branching process survives until time } t) \sim (-A(0))e^{-\beta t}$$

for some $A(0) < 0$ and $\beta > 0$, by Athreya and Ney [(1972), Theorem III.8.1 and Corollary III.8.1]. Moreover, by Yaglom's theorem [see Athreya and Ney (1972), Theorem III.7.4] there exists some $b(1) > 0$ such that

$$P_1(X(t) = 1) \sim b(1)P_1(R > t).$$

Of course, $\lambda_1 = \beta$ in this case.

Condition (1.4) is easy to check directly from the transition rates. It is fulfilled in many examples. Pakes (1994) investigates what happens in a number of examples when (1.4) fails.

Condition (1.6) can often be verified in examples by finding a suitable supermartingale. The use of supermartingales for establishing ergodicity properties of Markov processes with some kind of central drift is well known; see for instance Meyn and Tweedie (1993) (especially Theorem 7.1) for a very general treatment of such methods. For our case, assume that there exist a function $f(\cdot)$ on $\mathbb{N}$ and constants $D_1, D_4, D_5 > 0$, $D_2, D_3, D_6 < \infty$, D_6 integer, such that

$$f(x) \geq 0 \quad \text{and} \quad f(x) \to \infty \quad \text{as } x \to \infty, \tag{1.9}$$

$$\sum_{y\neq x} q(x, y)f(y) \leq f(x) - D_1 \quad \text{for } x \geq D_6, \tag{1.10}$$

$$|f(x) - f(y)| \leq D_2 \quad \text{for } x \geq D_6 \text{ and } q(x, y) > 0, \tag{1.11}$$

$$\sum_{\substack{y \neq x \\ f(y) \geq n}} q(x, y) \leq D_3 e^{-D_4 n} \quad \text{for } 1 \leq x \leq D_6 - 1 \text{ and } n \geq 1, \tag{1.12}$$

and

$$-q(x, x) \geq D_5 \quad \text{for } x \geq D_6. \tag{1.13}$$

We show in Lemma 4.3 below that (1.9)–(1.13) imply (1.5).

There is a large class of processes satisfying (1.9)–(1.13). In particular, these conditions are satisfied by any Markov process on the integers with bounded jumps and with a drift toward the origin bounded away from zero [and irreducible on $\{1, 2, \ldots\}$ and with jump rates bounded away from 0, i.e., satisfying (1.13)].

We also remark that in the present setup, (1.6) is trivially equivalent to geometric ergodicity of $\{X(t)\}$, that is, to the exponential decay (as $t \to \infty$) at a rate independent of x and y of

$$|P_x(X(t) = y) - \lim_{s \to \infty} P_x(X(s) = y)|. \tag{1.14}$$

Indeed, when absorption is certain,

$$\lim_{s \to \infty} P_x(X(s) = y) = \begin{cases} 1, & \text{if } y = 0, \\ 0, & \text{if } y \neq 0, \end{cases}$$

and, in addition, for $y \neq 0$,

$$P_x(X(t) = y) \leq P_x(R > t).$$

Therefore, (1.6) is equivalent to the expression in (1.14) being $\leq Ce^{-\gamma t}$ for all y.

In contrast to ergodic Markov processes, for which the stationary measure is unique, there can be infinitely many qsd's. This has been pointed out by Cavender (1978) for the birth and death process. For this process, the set of qsd's is empty, a singleton or a continuum indexed by a real parameter; see Theorem 3.2 in van Doorn (1991). Some qsd's can be characterized as "minimal"; those that have minimal expected absorption time at 0.

In Section 2, we sketch the proof of Theorem 1 and give an alternative characterization of qsd's. Under this characterization a qsd is a fixed point of a certain transformation Φ on the space of probability measures. In Section 3 we introduce a family of renewal processes associated with this transformation and study the dynamical system induced by Φ. In Section 4, we prove Theorem 1.1 by showing that the set of fixed points of Φ is nonempty. Under the hypotheses of Theorem 1.1 plus the additional hypothesis (5.1), we show in Section 5 that $\Phi^n \delta_x$ converges along some subsequences to a minimal qsd. For the birth and death process we show in Theorem 6.1 that $\Phi^n \delta_x$ converges to the (unique) minimal qsd.

2. Stationary Markov processes related to qsd. To motivate our approach, assume that we can pass to the infinitesimal form of (1.2). So we would have that if μ is a qsd, then it satisfies

$$\sum_{y\geq 1} \mu(y)(q(y,x) + q(y,0)\mu(x)) = 0 \quad \text{for all } x \geq 1. \tag{2.1}$$

This equation can be interpreted by saying that μ is the invariant measure for the process on $\{1,2,\dots\}$ with transition rate matrix Q^μ given by

$$q^\mu(x,y) = q(x,y) + q(x,0)\mu(y), \qquad x,y \geq 1. \tag{2.2}$$

Condition (2.1) can be read as $\mu Q^\mu = 0$. The equivalence between (1.2) and (2.1) will be proven in Proposition 2.1. This implies that μ is a qsd if and only if $\mu Q^\mu = 0$ and suggests the introduction of the transformation Φ: $\mu \mapsto \nu$, where ν is the unique solution of $\nu Q^\mu = 0$. In fact, a fixed point of Φ is a qsd. We link the transformation Φ with a renewal process. Let F^μ be the distribution of the absorption time of the process starting with μ. We show that $F^{\Phi\mu}$ is the stationary distribution of the residual time of the renewal process with interarrival times distributed as F^μ. This allows us to show that if F^μ has exponentially bounded tails, then $F^{\Phi^n\mu}$ converges along subsequences to an exponential distribution. This together with tightness of $\Phi^n\mu$ and the Schauder–Tychonov fixed point theorem are the key tools to prove Theorem 1.1.

For any probability measure μ on $\{1,2,\dots\}$, let $Y^\mu(t)$ on $\{1,2,\dots\}$ be the Markov process with transition rate matrix Q^μ given by (2.2). It is convenient to have a construction of $Y^\mu(t)$ related to the absorbing process $X(t)$ with initial distribution μ. Let $\{X_k(t)\colon k = 1,2,\dots\}$ be a sequence of independent copies of $X(t)$ with initial distribution μ and absorption times $t_k = \inf\{t\colon X_k(t) = 0\}$. Define $s_0 = 0$, $s_k = \sum_{i=1}^k t_i$ for $k \geq 1$ and

$$Y^\mu(t) = \sum_{k=1}^{\infty} X_k(t - s_{k-1})1\{t \in [s_{k-1}, s_k)\}. \tag{2.3}$$

In words, each time that a copy of the process $X(t)$ is absorbed, it is replaced immediately by a new copy. Such "resurrection processes" have been considered in the literature before, for example, in Pakes (1993) and some of its references. It is easy to see that the process constructed by means of (2.3) has transitions rates given by (2.2) and initial distribution μ. Observe that the s_k are the occurrence times of a renewal process with interarrival intervals distributed as F^μ, the absorption time of the process with initial measure μ. To have a similar construction of $Y^\mu(t)$ for another initial distribution ρ, it suffices to take ρ as the initial distribution of the first copy $X_1(t)$.

Since all states $\{1,2,\dots\}$ connect, the assumption $E_\mu R < \infty$ implies that $Y^\mu(t)$ is positive recurrent, so there exists a unique invariant measure, that is, a unique probability measure ν on $\{1,2,\dots\}$ satisfying $\nu Q^\mu = 0$ [see Asmussen (1987), Theorem 2.4.2]. Let Φ be the transformation Φ: $\mu \mapsto \nu$. Thus Φ is a map from the probability distributions on $\{1,2,\dots\}$ into itself.

The stationary distribution Φ^μ of Y^μ is given by [see Asmussen (1987), Theorem 2.4.2]

$$\Phi\mu(y) = \frac{1}{E_\mu R}\int_0^\infty P_\mu(X(t) = y)\,dt. \tag{2.4}$$

We show now that the fixed points of Φ are qsd's.

PROPOSITION 2.1. *Let μ be a probability measure on $\{1,2,\ldots\}$ such that $E_\mu R < \infty$. Then $\Phi\mu = \mu$ implies $\mu Q^\mu = 0$. Also $\mu Q^\mu = 0$ is equivalent to μ being a qsd.*

PROOF. Since $\Phi\mu$ is defined as the unique probability measure ν satisfying $\nu Q^\mu = 0$, $\mu = \Phi\mu$ implies $\mu Q^\mu = 0$.

Now $\mu Q^\mu = 0$ is simply condition (2.1). The equivalence between (2.1) and μ being a qsd, under the condition (1.1), is just Theorems 3.4, 4.1 and the succeeding Remark (1) in Nair and Pollett (1993) [earlier discussions of the equivalence between (2.1) and (1.2) can be found in Vere-Jones (1969) and Pollett (1986)]. □

3. The associated renewal processes and the dynamics of the residual map. Let τ_i be the interarrival times of a renewal process. That is, the τ_i are independent and identically distributed as a random time τ with distribution F. Let $N(t)$ be the number of renewals up to time t and let $\sigma_i = \sum_{k=1}^i \tau_k$ be the time of the ith renewal. We assume that $E\tau < \infty$ and that F is not a lattice distribution. We introduce the residual map Ψ: $F \mapsto \Psi F$ acting on the set of time distributions where ΨF is the limit distribution of the residual time given by

$$1 - \Psi F(s) = \lim_{t\to\infty} P\big(\sigma_{N(t)+1} - t > s\big) = \frac{1}{E\tau}\int_s^\infty (1 - F(\omega))\,dw. \tag{3.1}$$

LEMMA 3.1. *If $E_\mu R < \infty$, then $F^{\Phi\mu} = \Psi F^\mu$.*

PROOF. The distribution of R under $\Phi\mu$ is given by

$$\begin{aligned} 1 - F^{\Phi\mu}(s) &= P_{\Phi\mu}(R > s) \\ &= \frac{1}{E_\mu R}\sum_{y\ge 1}\int_0^\infty P_\mu(X(t) = y)P_y(R > s)\,dt \\ &= \frac{1}{E_\mu R}\int_0^\infty P_\mu(R > t + s)\,dt \qquad (3.2) \\ &= \frac{1}{E_\mu R}\int_s^\infty P_\mu(R > t)\,dt. \\ &= 1 - \Psi F^\mu(s). \end{aligned}$$

□

We now analyze the residual map Ψ in the context of renewal processes. Let F be the interarrival distribution and write $m_k(F)$ for the kth moment of F. Assume that $m_k(F) < \infty$ for all k. Set $F_k = \Psi^k F$. These iterates of Ψ were investigated in Harkness and Shantaram (1969). They already derived the following relations (3.3)–(3.5):

$$1 - F_{k+1}(s) = \frac{1}{m_1(F_k)} \int_s^\infty (1 - F_k(w))\, dw. \tag{3.3}$$

Integration by parts and induction show that

$$m_1(F_k) = \frac{m_{k+1}(F)}{(k+1)m_k(F)}. \tag{3.4}$$

By induction on k this implies that

$$\frac{m_k(F_n)}{k!} = \prod_{i=n}^{k+n-1} m_1(F_i). \tag{3.5}$$

Harkness and Shantaram looked at limits (after normalization) of F_k along the full sequence of integers. Here we need conditions for subsequential convergence.

LEMMA 3.2. *Assume there exists a subsequence* $\mathcal{N} = \{n_1, n_2, \ldots\}$ *such that for all* $k \geq 1$

$$\lim_{j \to \infty} \frac{m_{k+n_j}(F)}{n_j m_{k+n_j-1}(F)} = \theta > 0.$$

Then F_{n_j} *converges to an exponential distribution with mean* θ.

PROOF.

$$\begin{aligned} m_k(F_n) &= k! \frac{1}{(n+1)\cdots(n+k)} \frac{m_{k+n}(F)}{m_n(F)} \\ &= k! \prod_{l=1}^{k} \frac{m_{l+n}(F)}{(n+l)m_{l+n-1}(F)} \to k!\theta^k \end{aligned}$$

as $n \to \infty$, $n \in \mathcal{N}$. Since

$$k!\theta^k = \frac{1}{\theta} \int_0^\infty t^k e^{-t/\theta}\, dt$$

and since

$$\sum_{k \geq 1} \left(\frac{1}{(2k)!\theta^{2k}} \right)^{1/2k} = \infty,$$

the result follows from the standard method of moments [cf. Chung (1974), Theorem 4.5.5]. □

PROPOSITION 3.3. *If for some* $C < \infty$, $\lambda > 0$, *it holds that* $1 - F(t) \le Ce^{-\lambda t}$ *for all* $t \ge 0$, *then there exists a* $\theta \le 1/\lambda$ *and a sequence* $\mathcal{N} = \{n_1 < n_2 < \cdots\}$ *such that*

$$F_n \text{ converges to an exponential distribution with mean } \theta$$

as $n \to \infty$, $n \in \mathcal{N}$ (*an exponential distribution with mean* 0 *is* δ_0).

PROOF. If $1 - F(t) \le Ce^{-\lambda t}$ for all $t \ge 0$, then

$$\begin{aligned} m_k(F) &= \int_0^\infty t^k \, dF(t) = \left[-t^k(1 - F(t))\right]_0^\infty + k\int_0^\infty t^{k-1}(1 - F(t)) \, dt \\ &= k\int_0^\infty t^{k-1}(1 - F(t)) \, dt \le Ck\int_0^\infty t^{k-1}e^{-\lambda t} \, dt \\ &= Ck!\lambda^{-k}. \end{aligned}$$

Thus

$$\prod_{j=1}^{k} \frac{m_j(F)}{jm_{j-1}(F)} = \frac{m_k(F)}{k!} \le C\lambda^{-k}$$

so that

$$\theta := \liminf \frac{m_j(F)}{jm_{j-1}(F)} \le \frac{1}{\lambda}.$$

Let $\mathcal{J} = \{j_1, j_2, \ldots\}$ be a subsequence for which

$$\frac{m_{j_l}(F)}{j_l m_{j_l-1}(F)} \to \theta \quad \text{for } l \to \infty. \tag{3.6}$$

If $\theta = 0$, then

$$m_1(F_{j_l-1}) = \frac{m_{j_l}(F)}{j_l m_{j_l-1}(F)} \to \theta = 0,$$

so that F_{j_l-1} converges to the Dirac distribution concentrated on $\{t = 0\}$ and we are done. If $\theta > 0$, then we use the fact that $m_{j+1}(F)/m_j(F)$ is increasing in j as a consequence of Schwarz's inequality

$$\left(m_j(F)\right)^2 \le m_{j-1}(F)m_{j+1}(F).$$

Therefore, for any $k \ge 0$,

$$\begin{aligned} \theta &\le \liminf_{l\to\infty} \frac{m_{j_l-k}(F)}{(j_l - k)m_{j_l-k-1}(F)} \le \limsup_{l\to\infty} \frac{m_{j_l-k}(F)}{(j_l - k)m_{j_l-k-1}(F)} \\ &= \limsup_{l\to\infty} \frac{m_{j_l-k}(F)}{j_l m_{j_l-k-1}(F)} \le \limsup_{l\to\infty} \frac{m_{j_l}(F)}{j_l m_{j_l-1}(F)} = \theta. \end{aligned}$$

It follows that for each fixed k,

$$\frac{m_{j_l-k}(F)}{(j_l-k)m_{j_l-k-1}(F)} \to \theta$$

and we can take $n_l = j_l - r_l$ with $r_l \to \infty$ so slowly that for each fixed $k \geq 0$,

$$\lim_{l\to\infty} \frac{m_{n_l+k}(F)}{(n_l+k)m_{n_l+k-1}(F)} = \theta.$$

The proposition now follows from Lemma 3.2. □

4. Existence of qsd.

THEOREM 4.1. *Assume that* (1.1), (1.4) *and* (1.5) *hold. Define*

$$\lambda_0 = \sup\{\lambda: E_x e^{\lambda R} < \infty\}. \tag{4.1}$$

Then λ_0 *is independent of the choice of* x *and for each* x,

$$F_n(t) := F^{\Phi^n\delta_x}(t) = P_{\Phi^n\delta_x}(R \leq t) \to 1 - e^{-t\lambda_0} \tag{4.2}$$

with

$$0 < \lambda_0 = \left(\lim_{n\to\infty} E_{\Phi^n\delta_x} R\right)^{-1} < \infty. \tag{4.3}$$

Finally, for fixed $x \geq 1$, *there exists a subsequence* $\mathcal{N}'$ *and a probability measure* μ_∞ *on* $\{1, 2, \ldots\}$ *such that* $\Phi^n\delta_x$ *converges weakly to* μ_∞ *along* $\mathcal{N}'$ *and such that*

$$P_{\mu_\infty}(R > t) = e^{-\lambda_0 t}. \tag{4.4}$$

REMARKS. (i) Condition (1.4) is unpleasant, but can probably not be dispensed with entirely. It is crucially used to prove tightness of the family (4.7) below and also at the end of the proof of Proposition 4.2.

(ii) Assume that the hypotheses of Theorem 4.1 hold, so that F_n converges to an exponential distribution with mean $\theta > 0$. Then is it true that

$$\lim_{t\to\infty} \frac{P_x(X(t) = y)}{P_x(X(t) \neq 0)} = \lim_{n\to\infty} (\Phi^n\delta_x)(y)? \tag{4.5}$$

In other words, does the Yaglom limit exist and does the limit in the right-hand side of (4.5) exist and are they equal? In the next section we shall prove that the limit in the right-hand side indeed exists under the additional assumption (5.1) and that it is a so-called minimal qsd. That (4.5) holds for a finite Markov chain was proven by Darroch and Seneta (1965). The relation (4.5) for the random walk follows from Section 3 of Seneta (1966) and Section 5 of Cavender (1978). van Doorn (1991) and van Doorn and Schrijner (1992) give conditions for the existence of the Yaglom limit for birth and death processes. Under these conditions, Theorem 4.1 of van Doorn (1991) and Theorem 6.1 in this paper imply that (4.5) holds for the birth and death case.

Kesten (1995) shows the Yaglom limit exists for a discrete time Markov chain on $\{0, 1, \dots\}$ with 0 as absorbing state, if the chain can only make jumps of bounded size and satisfies some further uniform irreducibility condition. He also gives a counterexample of a chain for which there exists a qsd but the Yaglom limit does not exist.

PROOF OF THEOREM 4.1. First we note that the independence of λ_0 from x follows easily from the irreducibility of Q. For the remainder of the proof we fix $x \geq 1$ and write F for $F_0 = F^{\delta_x}$, that is, $F(t) = P_x(R \leq t)$.

We next prove that

$$\liminf_{n\to\infty} \frac{m_{n+1}(F)}{(n+1)m_n(F)} \geq \frac{1}{\lambda_0}. \tag{4.6}$$

Indeed, if (4.6) fails, then as in the argument following (3.6) we can find a subsequence $\mathcal{N}$ such that along $\mathcal{N}$, F_n converges to an exponential distribution with mean $\theta < 1/\lambda_0$. We claim that this contradicts the maximality of λ_0. To see this, we set $\mu_n = \Phi^n \delta_x$. Then we note that the family

$$\{\mu_n : n \in \mathcal{N}\} \tag{4.7}$$

is tight. Indeed, for each $\varepsilon > 0$ we can find $t(\varepsilon)$ such that for all $n \in \mathcal{N}$,

$$\varepsilon \geq P_{\mu_n}(R > t(\varepsilon)) = 1 - F_n(t(\varepsilon)) = \sum_{y\geq 1} \mu_n(y) P_y(R > t(\varepsilon))$$

and consequently

$$\mu_n\{y : P_y(R \leq t(\varepsilon)) \leq 1/2\} \leq 2\varepsilon.$$

Thus, the family (4.7) is tight under (1.4). By going over to a further subsequence, if necessary, we may assume that

$$\mu_n \Rightarrow \mu_\infty \quad \text{as } n \to \infty \quad \text{in } \mathcal{N}$$

for some probability distribution μ_∞ on $\{1, 2, \dots\}$. Furthermore,

$$\begin{aligned} P_{\mu_\infty}(R > t) &= \sum_{y\geq 1} \mu_\infty(y) P_y(R > t) = \lim_{\substack{n\to\infty \\ n\in\mathcal{N}}} \sum_{y\geq 1} \mu_n(y) P_y(R > t) \\ &= \lim_{\substack{n\to\infty \\ n\in\mathcal{N}}} (1 - F_n(t)) = e^{-t/\theta}. \end{aligned}$$

Then μ_∞ charges some point y and

$$\mu_\infty(y) P_y(R > t) \leq P_{\mu_\infty}(R > t). \tag{4.8}$$

Thus

$$P_y(R > t) \leq \frac{1}{\mu_\infty(y)} e^{-t/\theta},$$

which indeed contradicts our choice of λ_0. So (4.6) must hold.

This same argument can be used to find a μ_∞ which satisfies (4.4). Indeed, by (4.1) there exists for any $\varepsilon > 0$ a constant $C = C(\varepsilon) > 0$ such that

$$P_x(R > t) \leq C \exp(-(\lambda_0 - \varepsilon)t)$$

for all $t \geq 0$. By Proposition 3.3 there exists a subsequence $\mathcal{N}'$ such that F_n converges along $\mathcal{N}'$ to an exponential distribution with mean less than or equal to $1/(\lambda_0 - \varepsilon)$. By doing this for different ε one can choose $\mathcal{N}'$ such that F_n converges along $\mathcal{N}'$ to an exponential distribution of mean at most $1/\lambda_0$. Then the preceding argument with $\mathcal{N}$ replaced by $\mathcal{N}'$ yields the desired μ_∞. Note also that the existence of this μ_∞ implies

$$P_x(R > t) \leq Ce^{-\lambda_0 t}.$$

Next we prove a converse of (4.6), namely,

$$\limsup_{n\to\infty} \frac{m_{n+1}(F)}{(n+1)m_n(F)} \leq \frac{1}{\lambda_0}. \tag{4.9}$$

Suppose that (4.9) does not hold. Then there exists an $\eta > 0$ and a subsequence $\mathcal{N}''$ such that

$$\frac{m_{n+1}(F)}{(n+1)m_n(F)} \geq \frac{1}{\lambda_0} + \eta \tag{4.10}$$

for any $n \in \mathcal{N}''$. Since $m_{k+1}(F)/m_k(F)$ is increasing in k,

$$\frac{m_{j+1}(F)}{(j+1)m_j(F)} \geq \frac{1}{\lambda_0} + \frac{\eta}{2} \tag{4.11}$$

for $n \leq j < j+1 \leq n(1 + C\eta)$, $n \in \mathcal{N}''$, and C satisfying

$$\frac{1}{1+C\eta}\left(\frac{1}{\lambda_0} + \eta\right) \geq \frac{1}{\lambda_0} + \frac{\eta}{2}.$$

For $n \in \mathcal{N}''$, take $i(n) =$ integer part of $n(1 + C\eta)$. Then

$$\frac{m_{i(n)}(F)}{i(n)!} = \frac{m_n(F)}{n!} \prod_{j=n}^{i(n)-1} \frac{m_{j+1}(F)}{(j+1)m_j(F)}$$

$$\geq \frac{m_n(F)}{n!}\left(\frac{1}{\lambda_0} + \frac{\eta}{2}\right)^{\lfloor C\eta n \rfloor},$$

by (4.11). Hence

$$\liminf_{\substack{n\to\infty \\ n\in\mathcal{N}''}} \left(\frac{m_{i(n)}(F)}{i(n)!}\right)^{1/i(n)} \geq \liminf_{\substack{n\to\infty \\ n\in\mathcal{N}''}} \left[\left(\frac{m_n(F)}{n!}\right)^{1/n}\left(\frac{1}{\lambda_0} + \frac{\eta}{2}\right)^{C\eta}\right]^{n/i(n)}$$

$$\geq \left[\frac{1}{\lambda_0}\left(\frac{1}{\lambda_0} + \frac{\eta}{2}\right)^{C\eta}\right]^{1/(1+C\eta)}.$$

However, this is impossible, since $P_x(R > t) \leq Ce^{-\lambda_0 t}$ implies $m_i(F) \leq Ci!(1/\lambda_0)^i$ as in the proof of Proposition 3.3. We have therefore proved

$$\lim_{n\to\infty} \frac{m_{n+1}(F)}{(n+1)m_n(F)} = \frac{1}{\lambda_0},$$

which, by Lemma 3.2, implies that F_n converges to an exponential distribution with mean $1/\lambda_0$. Moreover, by (3.4),

$$E_{\Phi^n\delta_x}R = \frac{m_{n+1}(F)}{(n+1)m_n(F)} \to \frac{1}{\lambda_0},$$

so that the equality in (4.3) holds.

Last, we observe that for any distribution μ on $\{1,2,\dots\}$ and any x in the support of μ,

$$E_\mu R \geq \mu(x)E_x R > 0.$$

In addition, $E_x R \to \infty$ as $x \to \infty$, by (1.4). Thus the second inequality in (4.3) follows, while the first inequality in (4.3) is obtained from (1.5). □

Let

$$\mathcal{M}_\theta = \{\mu\colon \mu \text{ a probability distribution on } \{1,2,\dots\} \text{ such that } P_\mu(R > t) = e^{-t/\theta}\}.$$

Then, from the last theorem, for $\theta = 1/\lambda_0$,

$$\mathcal{M}_\theta \neq \varnothing. \tag{4.12}$$

We provide $\mathcal{M}_\theta$ with the topology of pointwise convergence [i.e., $\mu_k \to \mu$ if and only if $\mu_k(x) \to \mu(x)$ for all $x \geq 1$]. Notice that

$$\Phi\colon \mathcal{M}_\theta \to \mathcal{M}_\theta. \tag{4.13}$$

In fact, if $\mu \in \mathcal{M}_\theta$, then by (3.2),

$$P_{\Phi\mu}(R > s) = \frac{1}{E_\mu R}\int_s^\infty P_\mu(R > t)\,dt$$

$$= \frac{1}{\theta}\int_s^\infty e^{-t/\theta}\,dt = e^{-s/\theta}.$$

PROPOSITION 4.2. *Under the hypotheses of Theorem* 4.1, *if* $\mathcal{M}_\theta \neq \varnothing$, *then the transformation* Φ *has a fixed point in* $\mathcal{M}_\theta$.

PROOF. We prove the proposition by showing that Φ is continuous on $\mathcal{M}_\theta$ and that $\mathcal{M}_\theta$ is convex and compact. The result then follows from an application of the Schauder–Tychonov fixed point theorem [see Dunford and Schwartz (1958), Theorem V.10.5].

Now we show the continuity of Φ. If $\mu_k(z) \to \mu(z)$ for all $z \geq 1$ and $\mu_k, \mu \in \mathcal{M}_\theta$, then

$$E_{\mu_k}R = E_\mu R = \theta.$$

Therefore, by (2.4) and Fatou's lemma, for any subset A of $\{1, 2, \ldots\}$,

$$\begin{aligned}
&\liminf_{k\to\infty} \sum_{y\in A} \Phi\mu_k(y) \\
&\quad = \liminf_{k\to\infty} \frac{1}{E_{\mu_k}R} \int_0^\infty \sum_{z\geq 1} \mu_k(z) \sum_{y\in A} P_z(X(t) = y)\, dt \qquad (4.14) \\
&\quad \geq \frac{1}{\theta} \int_0^\infty \sum_{z\geq 1} \mu(z) \sum_{y\in A} P_z(X(t) = y)\, dt = \sum_{y\in A} \Phi\mu(y).
\end{aligned}$$

Since this holds for A as well as A^c, we must have equality in (4.14). However, then also

$$\begin{aligned}
1 &= \lim_{k\to\infty} \sum_z \Phi\mu_k(z) \geq \limsup_{k\to\infty} \Phi\mu_k(y) + \liminf \sum_{z\neq y} \Phi\mu_k(z) \\
&\geq \limsup_{k\to\infty} \Phi\mu_k(y) + \sum_{z\neq y} \Phi\mu(z)
\end{aligned}$$

and it follows that

$$\limsup_{k\to\infty} \Phi\mu_k(y) = \Phi\mu(y),$$

which is the required continuity.

The convexity of $\mathcal{M}_\theta$ is trivial. For compactness it suffices to prove sequential compactness, since the topology on $\mathcal{M}_\theta$ is a metric one [e.g., $d(\mu, \mu') = \Sigma_{y\geq 1} |\mu(y) - \mu'(y)|/2^y$ is a metric on $\mathcal{M}_\theta$ whose induced topology is the topology of $\mathcal{M}_\theta$]. However, the sequential compactness of $\mathcal{M}_\theta$ can be proven in the same way as in Theorem 4.1. We merely have to show that any sequence $\{\mu_k\} \subset \mathcal{M}_\theta$ is tight, and this is done exactly as for the family (4.7). □

PROOF OF THEOREM 1.1. The theorem follows from Propositions 2.1 and 4.2 and (4.12). □

We finish this section by proving the following lemma announced in the Introduction.

LEMMA 4.3. *Statements* (1.9)–(1.13) *imply* (1.5).

PROOF. Define $\sigma_1, \sigma_2, \ldots$ as the successive jump times of $X(\cdot)$. Then, for $x \geq D_6$ we have by Neveu (1972), Lemma VII.2.8, or by Stout (1974), Lemma 5.4.1, that for $\lambda \geq 0$,

$$E_x \exp\bigl(\lambda[f(X(\sigma_1)) - f(x) + D_1]\bigr) \leq \exp\bigl(\Theta(\lambda)[D_1 + D_2]^2\bigr),$$

with

$$\Theta(\lambda) = [D_1 + D_2]^{-2}\bigl[\exp(\lambda(D_1 + D_2)) - 1 - \lambda(D_1 + D_2)\bigr],$$

because a.e. on $\{X(0) = x\}$ it holds that $|f(X(\sigma_1)) - f(X(0)) + D_1| \leq D_1 + D_2$.

However, then, for $0 < \varepsilon < D_5$,

$$\begin{aligned} E_x \exp\bigl(\lambda[f(X(\sigma_1)) - f(x)] + \varepsilon\sigma_1\bigr) &= E_x \exp\bigl(\lambda[f(X(\sigma_1)) - f(x)]\bigr) E_x \exp(\varepsilon\sigma_1) \\ &= E_x \exp\bigl(\lambda[f(X(\sigma_1)) - f(x)]\bigr) \frac{-q(x,x)}{-q(x,x) - \varepsilon} \\ &\le \exp\bigl(-\lambda D_1 + \Theta(\lambda)[D_1 + D_2]^2\bigr) \frac{D_5}{D_5 - \varepsilon}. \end{aligned}$$

Because $\Theta(\lambda) = O(\lambda^2)$ as $\lambda \downarrow 0$, we can fix $\lambda > 0$ such that

$$-\lambda D_1 + \Theta(\lambda)[D_1 + D_2]^2 \le -\frac{\lambda D_1}{2},$$

and then $\varepsilon > 0$ such that

$$\exp\bigl(-\lambda D_1 + \Theta(\lambda)[D_1 + D_2]^2\bigr) \frac{D_5}{D_5 - \varepsilon} \le 1.$$

With these choices,

$$g(t) := \exp\bigl(\lambda[f(X(t)) - f(X(0))] + \varepsilon t\bigr)$$

satisfies

$$E_x g(\sigma_1) \le 1 = g(0) \quad \text{when } x \ge D_6. \tag{4.15}$$

If we now define

$$T = \inf\{t \ge \sigma_1 \colon X(t) < D_6\}$$

and

$$\nu = \inf\{k \colon \sigma_k \ge T\},$$

then $g(X(\sigma_{k \wedge \nu}))$, $k \ge 0$, is a positive supermartingale, by virtue of (4.15). Moreover, on the event $\{T \ge t\}$ we have $\sigma_\nu \ge t$ [note that $\sigma_k \to \infty$ as $k \to \infty$ by virtue of (1.1)] and hence

$$g(X(\sigma_\nu)) = \exp\bigl(\lambda[f(X(\sigma_\nu)) - f(X(0))] + \varepsilon\sigma_\nu\bigr) \ge \exp(-\lambda f(X(0)) + \varepsilon t).$$

Thus, for $x \ge D_6$,

$$P_x\{T \ge t\} \le \exp(\lambda f(x) - \varepsilon t) E_x g(X(\sigma_\nu)) \le \exp(\lambda f(x) - \varepsilon t). \tag{4.16}$$

(4.16) shows that T has an exponentially bounded tail, when $X(0) \ge D_6$. However, then, even for $X(0) = x < D_6$, one has

$$\begin{aligned} P_x\{T \ge t\} &\le E_x\bigl(\exp(\lambda f(X(\sigma_1)) + \varepsilon\sigma_1)\exp(-\varepsilon t); X(\sigma_1) \ge D_6\bigr) + P_x(\sigma_1 \ge t) \\ &\le D_7 \exp(-\varepsilon t) \end{aligned}$$

for some constant D_7 which is finite when λ and ε are taken sufficiently small, by virtue of (1.12) and (1.1). From this it is not hard to obtain that R also has an exponentially bounded tail, that is, that (1.6) holds. □

5. Minimal qsd. Define

$$\lambda(\mu) = \sup\{\lambda: E_\mu e^{\lambda R} < \infty\}.$$

Then, by (4.8), for any probability measure μ on $\{1, 2, \dots\}$,

$$\lambda(\mu) \le \lambda_0,$$

with λ_0 as in (4.1). If μ is a qsd, then F^μ is exponential with mean $E_\mu R = 1/\lambda(\mu)$. This implies that $\lambda(\mu) = 1/E_\mu R \le \lambda_0$. When there is equality we call μ a *minimal* qsd because the mean time of absorption is minimal.

Proposition 4.2, together with Proposition 2.1, shows that for any θ for which $\mathcal{M}_\theta$ is nonempty, $\mathcal{M}_\theta$ actually contains a qsd. We shall now show that minimal qsd's corresponding to $\theta = 1/\lambda_0$ can be obtained as subsequential limits of $\Phi^n \delta_x$, when (1.1) is strengthened to

$$-q(x,x) = \sum_{y \ne x} q(x,y) \le C_1 < \infty \tag{5.1}$$

for some constant C_1 and all $x \in \mathbb{N}$.

Specifically, we prove the following result.

PROPOSITION 5.1. *Assume that* (5.1) *and* (1.5) *hold.*

(a) *If* $\mathcal{N} = \{n_1 < n_2 < \cdots\}$ *is a sequence of integers such that for some* $x \ge 1$, $\Phi^{n_k}\delta_x$ *converges weakly to a probability measure* μ_∞ *on* $\{1, 2, \dots\}$ *(as* $k \to \infty$*), then* μ_∞ *is a minimal qsd.*

(b) *Similarly, if* $\mathcal{T} = \{t_k\}$ *is a sequence of real numbers with* $t_k \to \infty$ *and* μ_∞ *is a probability measure on* $\{1, 2, \dots\}$ *such that for some* $x \ge 1$,

$$\lim_{k\to\infty} \frac{P_x(X(t_k) = y)}{P_x(R > t_k)} = \mu_\infty(y), \qquad y \ge 1, \tag{5.2}$$

then μ_∞ *is a minimal qsd.*

Note that the last statement of Theorem 4.1 shows that (1.1), (1.4) and (1.5) together imply the existence of subsequences $\mathcal{N}$ for which Proposition 5.1(a) applies.

The proof of Proposition 5.1 rests on a simple lemma.

LEMMA 5.2. *If* (5.1) *holds, then for some* $\lambda_1 \in [\lambda_0, \infty)$ *we have for all* $x, y \ge 1$,

$$\lim_{t\to\infty} \left(P_x(X(t) = y)\right)^{1/t} = e^{-\lambda_1}. \tag{5.3}$$

Moreover, for all $x, y \ge 1$ *and all* $s \ge 0$,

$$\lim_{t\to\infty} \frac{P_x(X(t+s) = y)}{P_x(X(t) = y)} = e^{-\lambda_1 s}. \tag{5.4}$$

(λ_1 *is independent of* x *and* y.)

PROOF. For $h > 0$ and $kh \le t \le (k+1)h$ it holds that

$$\begin{aligned} &P_x(X(kh) = y)P_y(X(t-kh) = y) \\ &\quad \le P_x(X(t) = y) \\ &\quad \le P_x(X((k+1)h) = y)\big(P_y(X((k+1)h - t) = y)\big)^{-1}. \end{aligned}$$

Furthermore, for all $0 \le u \le h$,

$$1 \ge P_x(X(u) = x) \ge P_x(X(v) = x, \text{for } 0 \le v \le h) \ge e^{-C_1 h} \tag{5.5}$$

[by (5.1)].

From this one easily sees that it suffices to prove

$$\lim_{k\to\infty} \big(P_x(X(hk) = y)\big)^{1/k} = e^{-\lambda_1 h} \tag{5.6}$$

and

$$\lim_{k\to\infty} \frac{P_x(X(h(k+l)) = y)}{P_x(X(hk) = y)} = e^{-\lambda_1 lh} \tag{5.7}$$

for all $x, y \ge 1$, $l \ge 1$ and $h > 0$. However, (5.6) is a well known easy consequence of subadditivity [see Vere-Jones (1967), Theorem A], whereas the ratio theorem (5.7) for the discrete time Markov chain $\{X(kh)\}_{k\ge 0}$ is just Lemma 4 of Kesten (1995) [it is for this lemma that (5.1) is needed; the uniformity in x in the last inequality of (5.5) is not really needed].

The fact that $\lambda_1 \ge \lambda_0$ follows immediately from the definition (4.1), which shows that $P_x(R > t) = \sum_{y\ge 1} P_x(X(t) = y) \le e^{-(\lambda_0 - \varepsilon)t}$ for $\varepsilon > 0$ and t sufficiently large. □

PROOF OF PROPOSITION 5.1. (a) By (2.4), for any probability measure ν on $\{1, 2, \dots\}$ and $y \ge 1$,

$$\Phi\nu(y) = \frac{1}{E_\nu R}\int_0^\infty \sum_{x\ge 1} \nu(x)P_x(X(t) = y)\, dt.$$

Induction on n then shows

$$\begin{aligned} \Phi^n\nu(y) &= \prod_{k=0}^{n-1} (E_{\Phi^k\nu}R)^{-1} \int_0^\infty dt_1 \int_0^\infty dt_2 \cdots \int_0^\infty dt_n\, P_\nu(X(t_1 + t_2 + \cdots + t_n) = y) \\ &= \prod_{k=0}^{n-1} (E_{\Phi^k\nu}R)^{-1} \frac{1}{(n-1)!}\int_0^\infty t^{n-1} P_\nu(X(t) = y)\, dt. \end{aligned}$$

In particular, if we fix $x \ge 1$ and write

$$c_n = \frac{1}{(n-1)!}\prod_{k=0}^{n-1} (E_{\Phi^k\delta_x}R)^{-1},$$

then

$$(\Phi^n\delta_x)(y) = c_n\int_0^\infty t^{n-1}P_x(X(t) = y)\, dt. \tag{5.8}$$

Assume now that along some subsequence $\mathcal{N}$, $\Phi^n\delta_x \Rightarrow \mu_\infty$ for some probability measure μ_∞ on $\{1, 2, \dots\}$. Then for $s \geq 0$, $y \geq 1$,

$$
\begin{aligned}
P_{\mu_\infty}(X(s) = y) &= \sum_{z \geq 1} \mu_\infty(z) P_z(X(s) = y) \\
&= \lim_{\substack{n \to \infty \\ n \in \mathcal{N}}} \sum_{z \geq 1} \Phi^n\delta_x(z) P_z(X(s) = y) \\
&= \lim_{\substack{n \to \infty \\ n \in \mathcal{N}}} \sum_{z \geq 1} c_n \int_0^\infty t^{n-1} P_x(X(t) = z) P_z(X(s) = y)\, dt \qquad (5.9) \\
&= \lim_{\substack{n \to \infty \\ n \in \mathcal{N}}} c_n \int_0^\infty t^{n-1} P_x(X(t+s) = y)\, dt.
\end{aligned}
$$

If we can show that for each fixed T the contribution of the integral over $[0, T]$ to the right-hand side of (5.9) is negligible, then it follows from (5.4) that the right-hand side of (5.9) equals

$$
\lim_{\substack{n \to \infty \\ n \in \mathcal{N}}} e^{-\lambda_1 s} c_n \int_0^\infty t^{n-1} P_x(X(t) = y)\, dt = e^{-\lambda_1 s} \mu_\infty(y).
$$

However, then

$$
P_{\mu_\infty}(X(s) = y) = e^{-\lambda_1 s} \mu_\infty(y) \qquad (5.10)
$$

and, by summing over y,

$$
P_{\mu_\infty}(R > s) = \sum_{y \geq 1} P_{\mu_\infty}(X(s) = y) = e^{-\lambda_1 s}. \qquad (5.11)
$$

Hence

$$
\frac{\sum_{x \geq 1} \mu_\infty(x) P_x(X(s) = y)}{\sum_{x \geq 1} \mu_\infty(x) P_x(X(s) \neq 0)} = \mu_\infty(y)
$$

and μ_∞ is a qsd. Moreover, (5.11) shows that for any x with $\mu_\infty(x) > 0$ and $\varepsilon > 0$,

$$
\mu_\infty(x) E_x\{e^{(\lambda_1 - \varepsilon)R}\} \leq E_{\mu_\infty}\{e^{(\lambda_1 - \varepsilon)R}\} < \infty,
$$

so that $\lambda_0 \geq \lambda_1$. Since we also have $\lambda_1 \geq \lambda_0$ (by Lemma 5.2), $\lambda_1 = \lambda_0$ and μ_∞ is a minimal qsd.

It remains to show that the integral over $[0, T]$ can be neglected in (5.9). However, on the one hand,

$$
\int_0^T t^{n-1} P_x(X(t) = y)\, dt \leq T^n/n, \qquad (5.12)
$$

while on the other hand, by (5.3), for $0 < \varepsilon < \lambda_1/2$ [recall that $\lambda_1 \geq \lambda_0 > 0$ by (1.5)] and n large,

$$\int_0^\infty t^{n-1} P_x(X(t) = y)\, dt \geq \int_{n/(2e\lambda_1)^m} t^{n-1} \exp(-(\lambda_1 + \varepsilon)t)\, dt \quad \text{[by (5.3)]}$$

$$\geq \int_0^\infty t^{n-1} \exp(-(\lambda_1 + \varepsilon)t)\, dt - \left(\frac{n}{2e\lambda_1}\right)^n \frac{1}{n} \tag{5.13}$$

$$= \frac{(n-1)!}{(\lambda_1 + \varepsilon)^n} - \frac{n^{n-1}}{(2e\lambda_1)^n} \geq \frac{(n-1)!}{2(\lambda_1 + \varepsilon)^n}.$$

Clearly (5.12) is small with respect to (5.13) as $n \to \infty$. This proves the desired negligibility of the integral over $[0, T]$, and hence proves Proposition 6.1(a).

(b) The proof that μ_∞ is a minimal qsd when (5.2) holds is even more direct. This time,

$$P_{\mu_\infty}(X(s) = y) = \sum_{z \geq 1} \mu_\infty(z) P_z(X(s) = y)$$

$$= \lim_{k \to \infty} \sum_{z \geq 1} c_k P_x(X(t_k) = z) P_z(X(s) = y), \tag{5.14}$$

where

$$c_k := (P_x(R > t_k))^{-1} = \left(\sum_{z \geq 1} P_x(X(t_k) = z)\right)^{-1}.$$

Note that we can take the limit outside the sum in the right-hand side of (5.14) by Scheffe's theorem [see Scheffé (1947)]. Continuing (5.14) we have

$$P_{\mu_\infty}(X(s) = y) = \lim_{k \to \infty} c_k P_x(X(t_k + s) = y)$$

$$= e^{-\lambda_1 s} \lim_{k \to \infty} c_k P_x(X(t_k) = y) \quad \text{[by (5.4)]}$$

$$= e^{-\lambda_1 s} \mu_\infty(y).$$

This is (5.10) again, and proves μ_∞ is a minimal qsd as before. □

COROLLARY 5.3. *If* (5.1), (1.4) *and* (1.5) *hold, then* $\lambda_1 = \lambda_0$.

PROOF. This is immediate because, with μ_∞ as in the last statement of Theorem 4.1, (4.4) as well as (5.11) must hold. □

REMARK. Other conditions for the equality $\lambda_0 = \lambda_1$ are given in Section 3.3 of Jacka and Roberts (1993).

6. The birth and death case. In this section we treat the birth and death process. The transition matrix has elements $q(0, y) = 0$, for all y, and $q(x, y) = 0$ if $|x - y| > 1$. We assume that the rates are uniformly bounded: there exists $C < \infty$ such that

$$-q(x, x) < C, \qquad x \in \mathbb{N}. \tag{6.1}$$

Notice that (6.1) implies (1.1) and (1.4). We further assume that all nonzero states connect:

(6.2) for all $x > 0$, $q(x, x+1) > 0$ and $q(x, x-1) > 0$.

The process evolves according to the (backward) equation

$$\frac{d}{dt} p_t(x, y) = \sum_{z=0}^{\infty} q(x, z) p_t(z, y), \tag{6.3}$$

where $p_t(x, y)$ is the probability that the process starting at x be at y at time t.

We assume that $E_x e^{\lambda R} < \infty$ for some $\lambda > 0$. Then by Theorem 1.1 there exist qsd's. The main result of this section is the following:

THEOREM 6.1. *For a birth and death process satisfying* (6.1), (6.2) *and* (1.5), $\lim_{n\to\infty} \Phi^n \delta_x$ *exists and is the unique minimal qsd.*

First we prove in the next lemma that the distribution of the absorption time determines the initial measure.

LEMMA 6.2. *The function* $\nu \mapsto F^\nu$ (*defined for* ν *a probability measure on* $\{1, 2, \ldots\}$) *is one to one: if* $\nu \neq \rho$, *then* $F^\nu \neq F^\rho$.

PROOF. We have

$$\frac{d^n}{dt^n} p_t(x, y) = \sum_{z=0}^{\infty} q^{(n)}(x, z) p_t(z, y),$$

where $q^{(n)}(x, z)$ is the (x, z)-term of the matrix Q^n. So

$$\begin{aligned} \frac{d^n}{dt^n} p_t(x,0)\Big|_{t=0} &= q^{(n)}(x,0) \\ &= \sum_{z_1, \ldots, z_{n-2}} q(x, z_1) q(z_1, z_2) \cdots q(z_{n-2}, 1) q(1, 0). \end{aligned}$$

Then

$$\frac{d^n}{dt^n} p_t(n,0)\Big|_{t=0} = q(n, n-1) q(n-1, n-2) \cdots q(1,0)$$

and

$$\frac{d^n}{dt^n} p_t(z,0)\Big|_{t=0} = 0 \quad \text{if } z > n.$$

The facts that the $q(x, y)$ are uniformly bounded and that for any x only a finite number of z's satisfy $q(z, x) > 0$ imply that the nth derivative $(d^n/dt^n)F^\nu(t)$ exists and is given by

$$\frac{d^n}{dt^n} F^\nu(t) = \sum_{z=0}^{\infty} \nu(z) \frac{d^n}{dt^n} p_t(z, 0).$$

Since $(d^n/dt^n)p_t(z,0)|_{t=0} = 0$ if $z > n$, the above sum evaluated at $t = 0$ only contains contributions for $z = 1, \ldots, n$. From $(d^k/dt^k)p_t(k,0)|_{t=0} = q(k, k-1)\cdots q(1,0) > 0$ we deduce that $\nu(n)$ is a linear function of $(d^k/dt^k)F^\nu(t)|_{t=0}$, $k = 1, \ldots, n$. □

COROLLARY 6.3. *A probability measure μ is a qsd if and only if the absorption time distribution F^μ is exponential. For any θ there is at most one qsd μ with $E_\mu R = \theta$.*

PROOF. We already proved in Section 1 that F^μ is exponential when μ is a qsd. Conversely, if F^μ is exponential, let $\theta = E_\mu R$. By the last lemma, $\mathscr{M}_\theta$ must be the singleton $\{\mu\}$. By Proposition 4.2, μ is a qsd. □

PROOF OF THEOREM 6.1. As in (4.7), $\{\Phi^n \delta_x\}$ is right. From Proposition 5.1(a) we see that any limit point $\lim_{n\to\infty} \Phi^{n_l}\delta_x = \mu_\infty$ is a minimal qsd. Lemma 6.2 implies that all the limit points are the same. □

Acknowledgments. The authors thank the referees for suggesting some improvements in Theorem 1.1 and Section 5. They also thank A. G. Pakes for catching an error in Section 6. H.K., S.M. and P.P. acknowledge the kind hospitality of Instituto de Mathemática e Estatística da Universidade de São Paulo, where part of this work was realized. They also acknowledge FAPESP support during their visits to IME-USP. H.K. further thanks the American Mathematical Society for travel support to Brazil. P.A.F. is partially supported by CNPq and Vitae, H.K. by the NSF and S.M. by FONDECYT and Vitae.

REFERENCES

ASMUSSEN, S. (1987). *Applied Probability and Queues*. Wiley, New York.

ATHREYA, K. and NEY, P. (1972). *Branching Processes*. Springer, Berlin.

BREIMAN, L. (1968). *Probability*. Addison-Wesley, Reading, MA.

CAVENDER, J. A. (1978). Quasi-stationary distributions of birth and death processes. *Adv. in Appl. Probab.* **10** 570–586.

CHUNG, K. L. (1974). *A Course in Probability Theory*, 2nd ed. Academic Press, New York.

DARROCH, J. N. and SENETA, E. (1965). On quasi-stationary distributions in absorbing discrete-time finite Markov chains. *J. Appl. Probab.* **2** 88–100.

DUNFORD, N. and SCHWARTZ, J. T. (1958). *Linear Operators*. Interscience, New York.

FERRARI, P. A., MARTINEZ, S. and PICCO, P. (1992). Existence of non trivial quasi stationary distributions in the birth and death chain. *Adv. in Appl. Probab.* **24** 795–813.

GOOD, P. (1968). The limiting behaviour of transient birth and death processes conditioned on survival. *J. Austral. Math. Soc.* **8** 716–722.

HARKNESS W. L. and SHANTARAM, R. (1969). Convergence of a sequence of transformations of distribution functions. *Pacific J. Math.* **31** 403–415.

JACKA, S. D. and ROBERTS, G. O. (1993). Weak convergence of conditional processes on a countable state space. Preprint.

KESTEN, H. (1995). A ratio limit theorem for (sub) Markov chains on $\{1, 2, \ldots\}$ with bounded jumps. *Adv. in Appl. Probab.* **27**.

KIJIMA, M. and SENETA, E. (1991). Some results for quasi-stationary distributions of birth–death processes. *J. Appl. Probab.* **28** 503–511.

KINGMAN, J. F. C. (1963). The exponential decay of Markov transition probabilities. *Proc. London Math. Soc. (3)* **13** 337–358.

MEYN, S. P. and TWEEDIE, R. L. (1993). Stability of Markovian processes III: Foster–Lyapunov criteria for continuous time processes. *Adv. in Appl. Probab.* **25** 518–548.

NAIR, M. and POLLETT, P. (1993). On the relationship between μ-invariant measures and quasistationary distributions for continuous time Markov chains. *Adv. in Appl. Probab.* **25** 82–102; 717–719.

NEVEU, J. (1972). *Martingales a Temps Discrèt*. Masson and Cie, Paris.

PAKES, A. G. (1973). Conditional limit theorems for a left continuous random walk. *J. Appl. Probab.* **10** 317–329.

PAKES, A. G. (1993). Absorbing Markov and branching processes with instantaneous resurrection. *Stochastic Process. Appl.* **48** 85–106.

PAKES, A. G. (1994). Quasi-stationary laws for Markov processes: examples of an always proximate absorbing state. Preprint.

POLLETT, P. K. (1986). On the equivalence of μ-invariant measures for the minimal process and its q-matrix. *Stochastic Process. Appl.* **22** 203–221.

SCHEFFÉ, H. (1947). A useful convergence theorem for probability distributions. *Ann. Math. Statist.* **18** 434–438.

SENETA, E. (1966). Quasi-stationary behaviour in the random walk with continuous time. *Austral. J. Statist.* **8** 92–98.

SENETA, E. and VERE-JONES, D. (1966). On quasi stationary distributions in discrete-time Markov chains with a denumerable infinity of states. *J. Appl. Probab.* **3** 403–434.

STOUT, W. F. (1974). *Almost Sure Convergence*. Academic Press, New York.

VAN DOORN, E. A. (1991). Quasi-stationary distributions and convergence to quasi-stationarity of birth–death processes *Adv. in Appl. Probab.* **23** 683–700.

VAN DOORN, E. A. and SCHRIJNER, P. (1994a). Ratio limits and limiting conditional distributions for discrete-time birth-death processes. *J. Math. Anal. Appl.*

VAN DOORN, E. A. and SCHRYNER, P. (1994b). Geometric ergodicity and quasi-stationary in discrete-time birth-death processes. Preprint.

VERE-JONES, D. (1962). Geometric ergodicity in denumerable Markov chains. *Quart. J. Math. Oxford Ser. 2* **13** 7–28.

VERE-JONES, D. (1967). Ergodic properties of nonnegative matrices I. *Pacific J. Math.* **22** 361–396.

VERE-JONES, D. (1969). Some limit theorems for evanescent processes. *Austral. J. Statist.* **11** 67–78.

YAGLOM, A. M. (1947). Certain limit theorems of the theory of branching stochastic processes *Dokl. Akad. Nauk SSSR (N.S.)* **56** 795–798 (in Russian).

P. A. FERRARI
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
UNIVERSIDADE DE SÃO PAULO
CX. POSTAL 20570
01498-970 SÃO PAULO
BRASIL

H. KESTEN
DEPARTMENT OF MATHEMATICS
WHITE HALL
CORNELL UNIVERSITY
ITHACA, NEW YORK 14853-7901

S. MARTINEZ
DEPARTAMENTO DE INGENIERÍA MATEMÁTICA
UNIVERSIDAD DE CHILE
CASILLA 170
CORREO 3
SANTIAGO 3
CHILE

P. PICCO
LABORATOIRE PROPRE, LP7061
CENTRE DE PHYSIQUE THÉORIQUE CNRS
LUMINY, CASE 907F
13288 MARSEILLE CEDEX 9
FRANCE