The Annals of Applied Probability
1997, Vol. 7, No. 1, 134–155

THE QUASI-STATIONARY BEHAVIOR OF QUASI-BIRTH-AND-DEATH PROCESSES[1]

BY N. G. BEAN, L. BRIGHT, G. LATOUCHE, C. E. M. PEARCE, P. K. POLLETT AND P. G. TAYLOR

University of Adelaide, University of Adelaide, Université Libre de Bruxelles, University of Adelaide, University of Queensland and University of Adelaide

For evanescent Markov processes with a single transient communicating class, it is often of interest to examine the limiting probabilities that the process resides in the various transient states, conditional on absorption not having taken place. Such distributions are known as quasi-stationary (or limiting-conditional) distributions. In this paper we consider the determination of the quasi-stationary distribution of a general level-independent quasi-birth-and-death process (QBD). This distribution is shown to have a form analogous to the matrix-geometric form possessed by the stationary distribution of a positive recurrent QBD. We provide an algorithm for the explicit computation of the quasi-stationary distribution.

1. Introduction. An elegant development in the literature on Markov processes over the last three decades has been the discovery of fine structure, distinguishing types of transient class, within the classification of the ergodic character of states. Consider the simplest case, that of a countable discrete-time Markov chain with a single absorbing state 0 accessible from an irreducible class $\mathscr{C} = \{1, 2, \dots\}$ with n-step transition probabilities $p_{ij}^{(n)}$ For $i, j \geq 1$, the power series $\sum_{n=0}^{\infty} p_{ij}^{(n)} z^n$ possess a common radius of convergence $\alpha \geq 1$, and the further classification follows the classical one, with $p_{ij}^{(n)} \alpha^n$ taking the place of $p_{ij}^{(n)}$. Thus the chain is classified as α-positive, α-null or α-transient depending on the convergence or divergence of $\sum_{n=0}^{\infty} p_{ij}^{(n)} \alpha^n$ and the behavior of $\limsup p_{ij}^{(n)} \alpha^n$. For a detailed account see [21] and [19].

Allied with these developments was an emergent theory of α-invariant (or quasi-stationary) distributions. These frequently give the limiting behavior of the chain, conditional on absorption not yet having taken place. If this is the case, the distribution is also termed a limiting-conditional distribution. In stark contrast to the theory of irreducible chains, limiting-conditional distributions may exist when the process is α-transient. Indeed, a homogeneous birth-and-death process is α-transient, yet there is always a limiting-conditional distribution. For further detail see [2], [5] and [1].

In order to treat concrete examples, one has to find the value of the critical parameter, α, which is often difficult. For this reason there are very few

Received September 1995; revised September 1996.
[1]Research supported by Australian Research Grant A69602762 and the University of Adelaide.
AMS 1991 *subject classification.* Primary 60K25.
Key words and phrases. Quasi-birth-and-death process, quasi-stationary distribution, limiting-conditional distribution.

substochastic chains for which a full analysis is available. Historical exceptions were finite-state processes, the Galton–Watson branching process, simple birth-and-death chains and the work of Kyprianou [8] on GI/M/1 queues (see also [9] for an analysis under conditions of heavy traffic). Recently, a notable advance was made by Kijima [6] who gave an algebraic equation for the convergence norm of PH/PH/1 queues (in fact, more generally, for processes of M/G/1 and GI/M/1 type). In the queueing context considered in [6], this equation can be solved by use of the Laplace–Stieltjes transform of the interarrival and service time distributions. Kijima also gave the form of the quasi-stationary distribution for the special cases of the M/PH/1 and PH/M/1 queues. This work was extended by Makimoto [13] who gave an explicit representation of the quasi-stationary distribution for PH/PH/c queues in terms of solutions to a matrix equation. Makimoto did not, however, discuss methods of solution for this equation in the general case.

In this paper we extend the results of [6] and [13] by examining the limiting-conditional behavior of general quasi-birth-and-death processes (QBD's), which includes PH/PH/c queues as a subclass. Neuts [16] introduced QBD's as a generalization of the ordinary birth-and-death process, replacing each state by a "level" of states. Processes of this type have proved to be a potent tool in modeling queueing and telecommunications systems. The authors believe that they also have great potential for modeling a variety of biological phenomena.

Our major result provides a method for the computation of the radius of convergence α and the unique quasi-stationary distribution, which, under the mild assumptions we shall make, admits a limiting-conditional interpretation. Thus, at the same time, we provide a means of studying the long-term behavior of a flexible class of stochastic models, as well as providing a significant new class of processes for which a complete limiting-conditional analysis is possible.

The paper is organized as follows. In Section 2 we briefly review some of the basic results on limiting-conditional and quasi-stationary distributions.

In Section 3 we generalize the familiar geometric matrix R of a QBD to a version $R(\beta)$. The scalar β acts as a time-discounting factor, so that $R(\beta)$ may be interpreted as a matrix of expected rewards. The matrix $R(\beta)$ is shown to have basic properties similar to those of R; in particular, when its entries are finite, it constitutes the minimal nonnegative solution to a certain matrix-quadratic equation.

Section 4 addresses the question of finding the radius of convergence α. The majority of the results in this section follow from [6]. Here, an algebraic expression is given for the largest value of β for which $R(\beta)$ has finite entries. This value is α. A consequence is that the process is α-transient. This constitutes a major difficulty for the conclusion that the process has a limiting-conditional distribution, since classical theory deals only with the α-recurrent case. However, recent results of Kesten [5] concerning α-transient processes are applicable in our situation. It follows from the structure of the model that the jumps are of bounded size and Kesten's uniform irreducibility condition is satisfied. With the mild extra assumption of aperiodicity, the existence of a limiting-conditional distribution follows from Kesten's Theorem 2.

In Section 5 we give an explicit representation for the unique limiting-conditional distribution. The form of this distribution extends that identified by Makimoto [13] in the special case of the PH/PH/c queue.

Section 6 contains the major contribution of the paper. In this section we develop an efficient numerical technique for computing the radius of convergence, α, the matrix $R(\alpha)$ and, ultimately, the limiting-conditional distribution. Our solution incorporates the logarithmic reduction algorithm developed by Latouche and Ramaswami [12] for QBD structures.

Finally, in Section 7 we report on numerical results for two examples.

2. Limiting-conditional distributions. Consider a discrete-time Markov chain $(X_n;\ n \in \mathbb{Z}_+)$ on a countable state space $\mathscr{S} = \{0, 1, \ldots\}$ with transition matrix P. Assume (X_n) has an absorbing state 0 and an irreducible and aperiodic communicating class $\mathscr{C} \equiv \mathscr{S} \setminus \{0\}$. Let $\widehat{P}$ denote the restriction of P to $\mathscr{C}$. We assume that the expected time until absorption is finite from one (and then all) states $i \in \mathscr{C}$. This is an important assumption which we emphasize for later reference.

ASSUMPTION 1. The expected time until absorption is finite from all states $i \in \mathscr{C}$.

Let T denote the time until absorption of the process. A distribution, $\boldsymbol{\pi} = (\pi_i,\ i \in \mathscr{C})$, over $\mathscr{C}$ is called a *quasi-stationary distribution* if, whenever $P(X_0 = j) = \pi_j$, $j \in \mathscr{C}$,

$$P(X_n = j | T > n) = \pi_j, \qquad j \in \mathscr{C},$$

for all $n \geq 1$, so that, conditional on the chain being in $\mathscr{C}$, the state probabilities do not vary with time.

A nontrivial, nonnegative row vector $\mathbf{m}(\beta)$ that satisfies

$$\mathbf{m}(\beta) = \beta \mathbf{m}(\beta) \widehat{P} \tag{2.1}$$

is called a β-invariant measure. It is elementary to show that $\boldsymbol{\pi}$ is a quasi-stationary distribution if and only if, for some $\beta > 1$, it is a normalized β-invariant measure, in which case

$$\beta^{-1} = \sum_{i \in \mathscr{C}} \sum_{j \in \mathscr{C}} \pi_i p_{ij} = 1 - \sum_{i \in \mathscr{C}} \pi_i p_{i0}.$$

Seneta and Vere-Jones [19] and Kesten [5] showed that under certain conditions the quasi-stationary distribution $\boldsymbol{\pi}$ is also a *limiting-conditional distribution* in that

$$\lim_{n \to \infty} P(X_n = j | X_0 = i,\ T > n) = \pi_j, \qquad j \in \mathscr{C},$$

no matter what the initial state i, in which case β must be the *convergence radius* associated with $\widehat{P}$.

The convergence radius can be characterized as follows. For $z \in \mathbb{R}$, let $N_{ij}(z)$ be defined by

$$N_{ij}(z) = \sum_{n=0}^{\infty} z^n \widehat{P}_{ij}^{(n)}, \tag{2.2}$$

where $\widehat{P}_{ij}^{(n)}$ is the (i, j)th entry of $\widehat{P}^n$. Theorem 6.1 of Seneta [18] states that, for a given value of z, either $N_{ij}(z)$ is finite for all (i, j) or $N_{ij}(z)$ is infinite for all (i, j). Thus we can define the convergence radius associated with $\widehat{P}$ as

$$\alpha = \sup\{z\colon N_{ij}(z) \text{ is finite}\}.$$

There are two possibilities for the behavior of $N_{ij}(z)$ at $z = \alpha$. In the case where (2.2) diverges for $z = \alpha$, $\widehat{P}$ is said to be α-recurrent (either positive or null), while in the case where (2.2) converges for $z = \alpha$, $\widehat{P}$ is said to be α-transient (see [21]).

Henceforth, assume that (X_n) is a quasi-birth-and-death process. This can be regarded as a two-dimensional Markov chain with $\mathscr{E} = \{(k, j)\colon k \geq 1,\ 1 \leq j \leq M\}$ and whose (stochastic) transition matrix is of the block-partitioned form

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & \cdots \\ A_2\mathbf{e} & A_1 & A_0 & 0 & 0 & \cdots \\ 0 & A_2 & A_1 & A_0 & 0 & \cdots \\ 0 & 0 & A_2 & A_1 & A_0 & \cdots \\ 0 & 0 & 0 & A_2 & A_1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \tag{2.3}$$

where A_0, A_1 and A_2 are $M \times M$ matrices. Here the partitioning corresponds to distinguishing subsets of states called levels. Level k is defined by $l(k) = \{(k, j)\colon 1 \leq j \leq M\}$ for $k \geq 1$ and level 0 is the absorbing state 0. In (2.3), and throughout, $\mathbf{e}$ denotes an $M \times 1$ vector of 1's.

We assume (X_n) is irreducible on $\mathscr{E}$ which implies the matrix $A = A_0 + A_1 + A_2$ is irreducible and that A_0 and A_2 are nonzero. We emphasize these properties for later reference.

PROPERTY 2. The matrix $A = A_0 + A_1 + A_2$ is irreducible and A_0 and A_2 are nonzero.

Consequently, the matrix

$$\widehat{P} = \begin{pmatrix} A_1 & A_0 & 0 & 0 & \cdots \\ A_2 & A_1 & A_0 & 0 & \cdots \\ 0 & A_2 & A_1 & A_0 & \cdots \\ 0 & 0 & A_2 & A_1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

is strictly substochastic.

Equation (2.1), which defines the β-invariant measures, can now be written as

$$\mathbf{m}_1(\beta) = \beta[\mathbf{m}_1(\beta)A_1 + \mathbf{m}_2(\beta)A_2], \tag{2.4}$$

$$\mathbf{m}_k(\beta) = \beta[\mathbf{m}_{k-1}(\beta)A_0 + \mathbf{m}_k(\beta)A_1 + \mathbf{m}_{k+1}(\beta)A_2], \qquad k \geq 2, \tag{2.5}$$

where the M-vector $\mathbf{m}_k(\beta)$ is the restriction of $\mathbf{m}(\beta)$ to level k.

In the rest of this paper, we describe how to find a limiting-conditional distribution for (X_n). We find the radius of convergence α and then solve (2.1) to determine an α-invariant measure. As a by-product, we show that (X_n) is α-transient. Theorem 2 of Kesten [5] is then invoked to show that, suitably normalized, the α-invariant measure gives the limiting-conditional distribution for (X_n).

The results of this paper can be applied to *continuous-time* QBD's which have generators of the form analogous to (2.3). This follows by applying our discrete-time results to the uniformized chain.

A continuous-time Markov chain, with generator $Q = [q_{ij}]$, is uniform if $q \equiv \sup_i \sum_j q_{ij} < \infty$ and then its uniformized chain is defined to be the discrete-time Markov chain with transition probability matrix $P_r \equiv Q/r + I$ for some $r \geq q$. If Q is the generator of a continuous-time QBD, then it must be uniform on account of its homogeneity. The matrix P_r is then the transition probability matrix of a discrete-time QBD. Thus our discrete-time results can be applied to the uniformized chain to find its limiting-conditional distribution. From [6], page 425, it is known that the limiting-conditional distribution for a uniform continuous-time Markov chain and its uniformized chain are identical. Thus we shall have calculated the limiting-conditional distribution for the original continuous-time QBD.

3. Absorbing QBD's. In this section we extend the matrix-geometric theory of QBD's, as developed by Neuts [16], to absorbing QBD's. We parallel the development of Neuts and for most results the proofs are very similar. Throughout, a matrix is termed finite if all its entries are finite.

Let $N_{11}(\beta)$ denote the $M \times M$ matrix whose (i, j)th entry is $N_{(1,i)(1,j)}(\beta)$ as defined in (2.2). Define

$$R(\beta) = \beta A_0 N_{11}(\beta), \tag{3.1}$$

and T to be the first passage time (greater than zero) to level 1. The entry $R_{ij}(\beta)$ has the interpretation

$$R_{ij}(\beta) = \mathbb{E}\left[\sum_{n=1}^{T} \beta^n \mathbb{I}[X_n = (2, j)] | X_0 = (1, i)\right], \tag{3.2}$$

and so can be interpreted as the expected total discounted reward for visits to state $(2, j)$ before returning to level 1, conditional on starting in state $(1, i)$ with a discount factor β. In the rest of this paper, we shall consider only the situation where β is greater than or equal to 1.

Since (X_n) is homogeneous on all levels greater than 1, the interpretation given in (3.2) also holds when levels 1 and 2 are replaced by levels k and $k+1$, respectively.

LEMMA 1. *The matrix $R(\beta)$ is finite for all $\beta < \alpha$ and finite for $\beta = \alpha$ if and only if the process is α-transient.*

PROOF. By the definition of the convergence radius, the matrix $N_{11}(\beta)$ is finite for $\beta < \alpha$ and for $\beta = \alpha$ in the α-transient case, and infinite otherwise. Therefore, (3.1) implies that so is the matrix $R(\beta)$. □

THEOREM 2. (i) *If there exists a finite nonnegative solution to*

$$S = \beta[A_0 + SA_1 + S^2 A_2], \tag{3.3}$$

then the matrix $R(\beta)$ defined by (3.1) *is finite.*

(ii) *If the matrix $R(\beta)$ defined by* (3.1) *is finite, then it is the minimal nonnegative solution to* (3.3). *Here, and throughout, a minimal solution is elementwise minimal.*

PROOF. The proof follows analogously to that of Lemmas 1.2.2 and 1.2.3 in [16].

Define the sequence of matrices $\{R^N(\beta)\}$ by

$$R_{ij}^N(\beta) = \mathbb{E}\left[\sum_{n=1}^{T\wedge N} \beta^n \mathbb{I}[X_n = (2, j)] | X_0 = (1, i)\right], \tag{3.4}$$

and the sequence of matrices $\{W^N(\beta)\}$ by the recursion

$$W^0(\beta) = 0 \tag{3.5}$$

and

$$W^{N+1}(\beta) = \beta[A_0 + W^N(\beta)A_1 + (W^N(\beta))^2 A_2]. \tag{3.6}$$

As in the proof of Lemma 1.2.3 of [16]:

1. $W^N(\beta) \le W^{N+1}(\beta)$;
2. $R^N(\beta) \le R^{N+1}(\beta)$;
3. $R^N(\beta) \le W^N(\beta)$.

To prove part (i), let $S(\beta)$ be a finite nonnegative solution to (3.3). Then, by induction, $W^N(\beta) \le S(\beta)$ for all n, and so $W(\beta) \equiv \lim_{N\to\infty} W^N(\beta)$ exists and $W(\beta) \le S(\beta)$. The matrix $W(\beta)$ satisfies (3.3) and therefore is the minimal nonnegative solution.

By inequality 3 above, $\lim_{N\to\infty} R^N(\beta)$ exists and is finite. Fatou's lemma then gives that $R(\beta)$, as defined by (3.1), is also finite.

We now turn to the proof of part (ii). As in Lemmas 1.2.2 and 1.2.3 of [16], we can show that if $R(\beta)$ defined in (3.1) is finite, then it is a solution to (3.3).

The assumption that $R(\beta)$ is finite replaces Neuts's assumption of positive recurrence to ensure convergence of all the series involved.

Since $R(\beta)$ is a finite nonnegative solution to (3.3), it can replace $S(\beta)$ in the argument to show that $W(\beta)$ is the minimal nonnegative solution to (3.3). Therefore, $W(\beta) \le R(\beta)$. By Lebesgue's dominated convergence theorem, $R(\beta) = \lim_{N\to\infty} R^N(\beta)$, and so, by inequality 3, $R(\beta) \le W(\beta)$. Thus $R(\beta) = W(\beta)$ and is the minimal nonnegative solution to (3.3). □

Theorem 2(i) shows that if there exists a finite nonnegative solution to (3.3), then $R(\beta)$ defined in (3.1) is finite. Further, Theorem 2(ii) shows that $R(\beta)$ is the minimal nonnegative solution to (3.3) whenever $R(\beta)$ is finite.

Finally, we prove a simple result that will be useful in Sections 4 and 5.

LEMMA 3. *For all β such that $R(\beta)$ is finite, the left kernel of A_0 is exactly the left kernel of $R(\beta)$.*

PROOF. Consider β such that $R(\beta)$ is finite. If $\mathbf{x}$ is in the left kernel of $R(\beta)$, then, by multiplying (3.3) by $\mathbf{x}$, it is clear that $\mathbf{x}$ must also be in the left kernel of A_0. The fact that the left kernel of A_0 is a subset of the left kernel of $R(\beta)$ follows from (3.1). □

4. Determination of the convergence radius. In this section we shall determine the largest value β^* of β for which (3.3) has a finite nonnegative solution. This will then be shown to be the convergence radius, α. The development in this section parallels that of Kijima ([6], Section 2), but is in terms of the matrix $R(\beta)$ rather than the matrix $\Phi(z)$ of Kijima's equation (2.7). As we shall show in Section 5, the quasi-stationary distribution depends on $R(\alpha)$. Our development leads directly to, and illuminates, the computational algorithm of Section 6.

First let us consider the solution of (3.3). For $0 < z \le 1$, let $\chi(z)$ be the maximal eigenvalue of the matrix

$$A(z) = A_0 + zA_1 + z^2A_2$$

and let $\mathbf{u}(z)$ and $\mathbf{v}(z)$ be the corresponding left and right eigenvectors normalized so that

$$\mathbf{u}(z)\mathbf{e} = 1 = \mathbf{u}(z)\mathbf{v}(z). \tag{4.1}$$

As is observed in [16], page 15, $\chi(z)$ is analytic on $(0, 1)$, continuous at $z = 1$ and may be defined by continuity at $z = 0$. Throughout this paper we shall be making use of Assumption 1 and Property 2. By Lemma 1.3.3 and Corollary 1.3.2 of [16], we have the following property.

PROPERTY 3a.

$$\chi'(1^-) > 1.$$

By further applying Corollary 1.3.3 and the last sentence of Lemma 1.3.4 of [16], we get the following property.

PROPERTY 3b. There exists $z_0 \in (0, 1)$ such that

$$z\chi'(z) < \chi(z) \quad \text{for all } z \in (0, z_0). \tag{4.2}$$

In the positive recurrent case, it is of interest to study the equation

$$z = \chi(z). \tag{4.3}$$

Since $A(1)$ is stochastic, this equation has one solution at $z = 1$. Because of Property 3 there is a second solution $z = \eta \in (0, 1)$ (see Lemma 1.3.4 of [16]).

In the transient case we are interested in the equation

$$\frac{z}{\beta} = \chi(z). \tag{4.4}$$

For all $0 < z < \eta$, it follows from Property 3 that $\chi(z) > z$ and so $\chi(z) > z/\beta$. Therefore, any set of solutions to (4.4) in $(0, 1]$ must be bounded away from 0. Thus, if there exists a solution to (4.4) in $(0, 1]$, then there exists a minimal nonnegative solution in $(0, 1]$.

THEOREM 4. (i) *If β is such that* (4.4) *has a solution in* $(0, 1]$, *then there exists a finite nonnegative solution to* (3.3) *and so $R(\beta)$ is finite.*

(ii) *If $R(\beta)$ is finite, then the maximal eigenvalue $\eta(\beta)$ of $R(\beta)$ is positive. It is the minimal nonnegative solution in* $(0, 1]$ *to* (4.4) *and the corresponding left eigenvector is $\mathbf{u}(\eta(\beta))$, which may be chosen to be positive.*

PROOF. (i) Consider the sequences of matrices $\{R^N(\beta)\}$ and $\{W^N(\beta)\}$ defined in the proof of Theorem 2. In that proof, under the assumption that there exists a finite nonnegative solution $S(\beta)$ to (3.3), we showed that $\{W^N(\beta)\}$ [and therefore $\{R^N(\beta)\}$] lies in the compact space $\{X\colon X \le S(\beta)\}$. In the hypothesis of this part of this theorem, we have not assumed the existence of the matrix $S(\beta)$ and so we take a different approach.

If β is such that (4.4) has a solution in $(0, 1]$, then let η^0 be the minimal such solution and $\mathbf{u}^0$ the corresponding left eigenvector of $A(\eta^0)$. By induction, it is easy to show that the sequence $\{W^N(\beta)\}$ [and therefore, by inequality 3, $\{R^N(\beta)\}$] satisfies

$$\mathbf{u}^0 R^N(\beta) \le \eta^0 \mathbf{u}^0. \tag{4.5}$$

The space of nonnegative matrices satisfying (4.5) is compact and so the sequence $\{R^N(\beta)\}$ converges and, by a similar argument to that in the proof of Theorem 2(ii), $R(\beta)$ is the minimal nonnegative solution to (3.3).

(ii) For $\beta \ge 1$, $R(\beta) \ge R(1)$ so that $\eta(\beta)$ is greater than or equal to the maximal eigenvalue of $R(1)$ which is strictly positive (see [16], Lemma 1.2.4).

This proof now follows similar lines to that of Lemma 1.3.2 of [16]. Let $\mathbf{s}$ be a nonnegative left eigenvector of $R(\beta)$ corresponding to the maximal eigenvalue $\eta(\beta)$. From (3.3) $\mathbf{s}$ is also a left eigenvector of the irreducible matrix $A(\eta(\beta))$ with eigenvalue $\eta(\beta)/\beta$. Since $\mathbf{u}(\eta(\beta))$ is positive and $\mathbf{s}$ is nonnegative, it follows from [3], pages 63–64, that $\mathbf{s} = \mathbf{u}(\eta(\beta))$ and $\eta(\beta)$ satisfies (4.4).

Since (4.4) has a solution, from the proof of part (i) we have that $\{R^N(\beta)\}$ satisfies (4.5) and therefore

$$\mathbf{u}^0 R(\beta) \leq \eta^0 \mathbf{u}^0. \tag{4.6}$$

Now Theorem 1.6 of [18] implies that $0 < \eta(\beta) \leq \eta^0$ and so $\eta(\beta) = \eta^0$. □

THEOREM 5. *For a process X_n such that Assumption* 1 *and Property* 2 *hold, the maximal value of β for which* (4.4) *has a solution $\eta \in (0, 1]$ is given by*

$$\beta^* = \left[\mathbf{u}(z_0)[A_1 + 2z_0 A_2]\mathbf{v}(z_0)\right]^{-1}, \tag{4.7}$$

where z_0 is the minimal solution to

$$\chi'(z)z = \chi(z) \tag{4.8}$$

in the interval $(0, 1)$, *and* $\mathbf{u}(z_0)$ *and* $\mathbf{v}(z_0)$ *are the Perron–Frobenius left and right eigenvectors of $A(z_0)$, respectively.*

PROOF. Kingman's theorem [7] shows that $\log \chi(e^{-s})$ is a convex function of s on the interval $[0, \infty)$. This result is a key ingredient in the proof and so we shall argue in terms of the function $\chi(e^{-s})$, $s \in [0, \infty)$. With this substitution, (4.4) becomes

$$\chi(e^{-s}) = e^{-s}/\beta. \tag{4.9}$$

We have already seen that when $\beta = 1$ there are two distinct solutions: one at $s = 0$ and the other at $s = -\log \eta$. Kingman's theorem then shows that the graphs of $\chi(e^{-s})$ and e^{-s}/β are as indicated in Figure 1. In particular, for β sufficiently large there are no solutions to (4.9).

Consider (4.8) which can be rewritten as

$$\left(\frac{d\chi(z)}{dz} \bigg|_{z=e^{-s}} \right) e^{-s} = \chi(e^{-s}). \tag{4.10}$$

Clearly, this is satisfied if the ratio of the left-hand side to the right-hand side is 1. Property 3a ensures that this ratio is greater than 1 in the limit as s approaches 0 from above, and Property 3b ensures that this ratio is less than 1 as s approaches ∞. The convexity of $\log \chi(e^{-s})$ implies that this ratio is decreasing. Therefore, there must exist a solution s_0 to (4.10) and accordingly a solution $z_0 = \exp(-s_0)$ to (4.8).

Define β^* by

$$\chi'(z_0) = \frac{1}{\beta^*} \tag{4.11}$$

and so (4.8) and (4.10) are equivalent to

$$\frac{z_0}{\beta^*} = \chi(z_0). \tag{4.12}$$

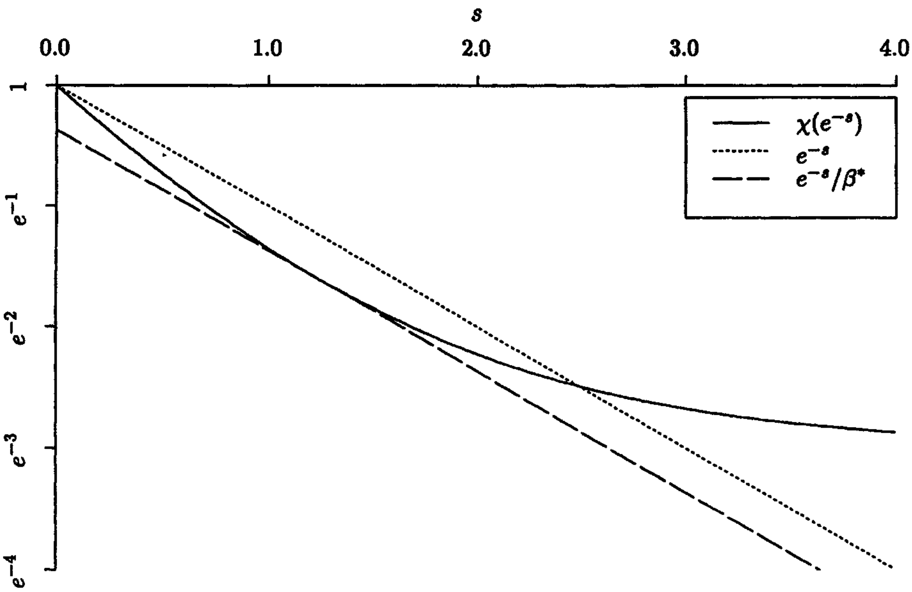

FIG. 1. *Graph of* $\chi(e^{-s})$ *and* e^{-s}/β*, on a log scale, against s.*

Equation (4.12) says that (4.4) is satisfied with $\beta = \beta^*$, and (4.11) says that the slope of the tangent to $\chi(z)$ at $z = z_0$ is $1/\beta^*$. Therefore, β^* is the maximal value of β for which (4.4) has a solution.

Returning to (4.10), Kingman's theorem shows only that, as a function of z, the ratio of the left-hand side to the right-hand side is increasing, not necessarily *strictly* increasing. Therefore, we cannot exclude the possibility that this ratio is 1 on an entire interval $[z_1, z_2]$. However, if this is the case, it is easy to see that $\chi(z)$ is of the form Cz on this interval, for some fixed value C. By (4.12), the value of C must be $1/\beta^*$. Thus β^* is uniquely determined even in this case.

For any β for which (4.4) has a solution, Theorem 4 shows that the maximal eigenvalue of the matrix $R(\beta)$ is given by the minimal such solution. When $\beta = \beta^*$, the minimal solution z_0 to (4.4) is also the minimal solution to (4.8). Thus the maximal eigenvalue of $R(\beta^*)$ is given by z_0.

To conclude the proof, it suffices to observe that (4.7) follows by substitution of (see [16], Lemma 1.3.3)

$$\chi'(z) = \mathbf{u}(z)[A_1 + 2zA_2]\mathbf{v}(z) \tag{4.13}$$

into (4.11). □

THEOREM 6. *The convergence radius* α *associated with* (X_n) *is equal to* β^* *defined in* (4.7).

PROOF. Lemma 1 shows that if $\beta < \alpha$, then $R(\beta)$ is finite. Therefore, Theorem 4(ii) implies that (4.4) has a solution and so $\beta \le \beta^*$. This proves that $\alpha \le \beta^*$. Anderson [1], Lemma 5.2.4(3), states that if there exists a nonnegative nonzero vector $\mathbf{m}$ such that $\beta \mathbf{m} \widehat{P} \le \mathbf{m}$ for some $\beta > 0$, then $\beta \le \alpha$. Therefore, if we can find such a vector for $\beta = \beta^*$, the proof is concluded.

Theorem 4(i) proves that the matrix $R(\beta^*)$ exists and so satisfies (3.3) with $\beta = \beta^*$. Choose $\mathbf{x}$ to be strictly positive and not in the left kernel of A_0. Then consider the vector $\mathbf{m}_j = \mathbf{x}R(\beta^*)^j$, $j \ge 1$, which is nonnegative and nonzero by Lemma 3. From (3.3) it follows that $\{\mathbf{m}_j\}$ satisfies (2.5) (with $\beta = \beta^*$). However, $\{\mathbf{m}_j\}$ does not satisfy (2.4) (with $\beta = \beta^*$); in fact, the right-hand side is $\mathbf{m}_1 - \beta^* \mathbf{x} A_0$. Since A_0 is a nonnegative matrix and $\mathbf{x}$ is a positive vector not in the left kernel of A_0, it follows that this is strictly less than $\mathbf{m}_1$. Hence we have that $\beta^* \mathbf{m} \hat{P} \le \mathbf{m}$. □

REMARK 1. In Kijima [6] the fact that $\beta^* \le \alpha$ is proved using the relation corresponding to (3.1). Our proof of this fact has the feature of exhibiting a class of (matrix-geometric) β-subinvariant measures for all $\beta \le \alpha$.

Since we have shown that $\beta^* = \alpha$, we shall henceforth use α to denote the quantity given by (4.7).

COROLLARY 7. *(X_n) is α-transient.*

PROOF. In the proof of Theorem 6, we observed that $R(\alpha)$ is finite. Lemma 1 then shows that (X_n) is α-transient. □

5. Determination of the limiting-conditional distribution. In this section we shall solve (2.1) for $\mathbf{m}(\alpha)$. Application of Theorem 2 from [5] determines the limiting-conditional distribution to be the normalized form of $\mathbf{m}(\alpha)$.

LEMMA 8. *Let $\mathbf{a} = \mathbf{u}(z_0)$ and $\mathbf{b} = \mathbf{u}'(z_0)$. Then the $1 \times M$ vector $\mathbf{b}$ is the unique solution to*

$$\mathbf{b}\left[z_0^2 A_2 + z_0\left(A_1 - \frac{1}{\alpha} I\right) + A_0\right] = -\mathbf{a}\left[2z_0 A_2 + \left(A_1 - \frac{1}{\alpha} I\right)\right] \tag{5.1}$$

subject to $\mathbf{be} = 0$.

PROOF. Consider the equation

$$\chi(z)\mathbf{u}(z) = \mathbf{u}(z)[A_0 + zA_1 + z^2 A_2]$$

that follows from the definition of $\chi(z)$ and $\mathbf{u}(z)$ as the eigenvalue–eigenvector pair of $A(z)$. Neuts ([16], page 17), shows that $\chi(\cdot)$ and $\mathbf{u}(\cdot)$ are differentiable functions of z. Hence (5.1) can be derived by differentiating the above equation with respect to z and evaluating at $z = z_0$ using (4.11) and (4.12). Thus $\mathbf{u}'(z_0)$

must exist and satisfy (5.1). By (4.1) the vector $\mathbf{u}(z)$ is normalized so that $\mathbf{u}(z)\mathbf{e} = 1$ and hence $\mathbf{be} = 0$.

To show that there is a unique solution to (5.1) subject to $\mathbf{be} = 0$, observe that any solution to (5.1) must be of the form $\mathbf{b} = \mathbf{b}_0 + \mathbf{c}$, where $\mathbf{b}_0$ is a particular solution to (5.1) and $\mathbf{c}$ satisfies $\mathbf{c}A(z_0) = z_0\mathbf{c}/\alpha$. The Perron–Frobenius theorem shows that the maximal eigenvalue of the irreducible matrix $A(z_0)$ is simple and so $\mathbf{c}$ must be some multiple of $\mathbf{u}(z_0)$. Since $\mathbf{u}(z_0)$ is normalized according to (4.1), the scalar multiple is uniquely defined by the condition $\mathbf{be} = 0$. □

THEOREM 9. *For a process (X_n) such that Assumption 1 and Property 2 hold, the limiting-conditional distribution is given by $c^{-1}\mathbf{m}_j(\alpha)$, where*

$$\mathbf{m}_j(\alpha) = z_0^j\mathbf{b} + jz_0^{j-1}\mathbf{a} - \mathbf{b}R(\alpha)^j \tag{5.2}$$

and

$$c = \frac{1}{(1-z_0)^2} - \mathbf{b}R(\alpha)(I - R(\alpha))^{-1}\mathbf{e}. \tag{5.3}$$

REMARK 2. Makimoto [13] derived the form of the limiting-conditional distribution for the PH/PH/c queue. In the PH/PH/1 case, this form is identical to (5.2).

PROOF. (i) We first show that

$$\mathbf{n}_j(\alpha) = \mathbf{y}\begin{pmatrix} z_0^j\mathbf{b} + jz_0^{j-1}\mathbf{a} \\ R(\alpha)^j \end{pmatrix} \tag{5.4}$$

satisfies (2.4) and (2.5), where $\mathbf{y}$ is any $1 \times (M+1)$ vector that satisfies

$$\mathbf{y}\begin{pmatrix} \mathbf{b} \\ I \end{pmatrix} A_0 = \mathbf{0}. \tag{5.5}$$

Substitution of (5.4) into the right-hand side of (2.4), with $\beta = \alpha$, gives

$$\begin{aligned}
\alpha[\mathbf{n}_1(\alpha)A_1 + \mathbf{n}_2(\alpha)A_2] &= \alpha\mathbf{y}\left[\begin{pmatrix} z_0\mathbf{b} + \mathbf{a} \\ R(\alpha) \end{pmatrix} A_1 + \begin{pmatrix} z_0^2\mathbf{b} + 2z_0\mathbf{a} \\ R(\alpha)^2 \end{pmatrix} A_2\right] \\
&= \alpha\mathbf{y}\begin{pmatrix} \mathbf{b}[z_0A_1 + z_0^2A_2] + \mathbf{a}[A_1 + 2z_0A_2] \\ R(\alpha)A_1 + R(\alpha)^2A_2 \end{pmatrix} \\
&= \mathbf{y}\begin{pmatrix} [z_0\mathbf{b} + \mathbf{a}] - \alpha\mathbf{b}A_0 \\ R(\alpha) - \alpha A_0 \end{pmatrix} \\
&= \mathbf{y}\begin{pmatrix} z_0\mathbf{b} + \mathbf{a} \\ R(\alpha) \end{pmatrix} \\
&= \mathbf{n}_1(\alpha),
\end{aligned}$$

by the definition of **a** and (5.1), (3.3) and (5.5). Similarly, substitution of (5.4) into the right-hand side of (2.5) when $\beta = \alpha$ gives, for $k \geq 2$,

$$\begin{aligned}
&\alpha[\mathbf{n}_{k-1}(\alpha)A_0 + \mathbf{n}_k(\alpha)A_1 + \mathbf{n}_{k+1}(\alpha)A_2] \\
&\quad = \alpha\mathbf{y}\left[\begin{pmatrix} z_0^{k-1}\mathbf{b} + (k-1)z_0^{k-2}\mathbf{a} \\ R(\alpha)^{k-1} \end{pmatrix} A_0 + \begin{pmatrix} z_0^{k}\mathbf{b} + kz_0^{k-1}\mathbf{a} \\ R(\alpha)^{k} \end{pmatrix} A_1 \right. \\
&\qquad\qquad \left. + \begin{pmatrix} z_0^{k+1}\mathbf{b} + (k+1)z_0^{k}\mathbf{a} \\ R(\alpha)^{k+1} \end{pmatrix} A_2 \right] \\
&\quad = \mathbf{y}\begin{pmatrix} kz_0^{k-1}\mathbf{a} + z_0^{k}\mathbf{b} \\ R(\alpha)^{k} \end{pmatrix} \\
&\quad = \mathbf{n}_k(\alpha).
\end{aligned}$$

(ii) Now we prove that the family $\{\mathbf{n}_j(\alpha)\}$, defined in (5.4) and (5.5), is equivalent (up to a scale factor) to $\{\mathbf{m}_j(\alpha)\}$ as given in (5.2).

Let the $1 \times (M+1)$ vector $\mathbf{y}$ be written as $(y_0, \mathbf{y}')$, where $\mathbf{y}'$ is a $1 \times M$ vector. Equation (5.5) then states that $y_0\mathbf{b} + \mathbf{y}'$ is in the left kernel of the matrix A_0. If A_0 has full rank, then $\mathbf{y}' = -y_0\mathbf{b}$ and $\mathbf{n}_j(\alpha)$ trivially has the form (5.2). Therefore, assume that A_0 has rank $k < M$. Let $\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_k$ be a basis for the left kernel of A_0. By Lemma 3 it follows that

$$\mathbf{q}_i R(\alpha) = 0 \quad \text{for all } i = 1, 2, \ldots, k. \tag{5.6}$$

Equation (5.5) can now be rewritten as

$$y_0\mathbf{b} + \mathbf{y}' = \sum_{i=1}^{k} w_i\mathbf{q}_i \tag{5.7}$$

for some $w_i \in \mathbb{R}$, $i = 1, 2, \ldots, k$.

Consider $\mathbf{n}_j(\alpha)$ as defined in (5.4):

$$\begin{aligned}
\mathbf{n}_j(\alpha) &= \mathbf{y}\begin{pmatrix} z_0^{j}\mathbf{b} + jz_0^{j-1}\mathbf{a} \\ R(\alpha)^{j} \end{pmatrix} \\
&= y_0(z_0^j\mathbf{b} + jz_0^{j-1}\mathbf{a}) + \mathbf{y}'R(\alpha)^j \\
&= y_0(z_0^j\mathbf{b} + jz_0^{j-1}\mathbf{a}) + \left(\sum_{i=1}^{k} w_i\mathbf{q}_i - y_0\mathbf{b}\right)R(\alpha)^j \\
&= y_0(z_0^j\mathbf{b} + jz_0^{j-1}\mathbf{a}) - y_0\mathbf{b}R(\alpha)^j \\
&= y_0\mathbf{m}_j(\alpha)
\end{aligned}$$

by (5.7) and (5.6).

(iii) Next we show that the family $\{\mathbf{m}_j(\alpha)\}$, defined in (5.2), is nonnegative.

By differentiating (4.4) with respect to β, we can show that $\eta'(\beta) > 0$ for all $\beta < \alpha$. Now consider, for $\beta < \alpha$, the equation

$$\eta(\beta)\mathbf{u}(\eta(\beta)) = \mathbf{u}(\eta(\beta))R(\beta) \tag{5.8}$$

that follows from the fact that $\eta(\beta)$ and $\mathbf{u}(\eta(\beta))$ are the eigenvalue–eigenvector pair of $R(\beta)$ by Lemma 1 and Theorem 4(ii). Differentiating with respect to β gives

$$\text{(5.9)}\qquad \mathbf{u}(\eta(\beta)) + \eta(\beta)\mathbf{u}'(\eta(\beta)) - \mathbf{u}'(\eta(\beta))R(\beta) = \mathbf{u}(\eta(\beta))\frac{R'(\beta)}{\eta'(\beta)}.$$

The vector $\mathbf{u}(\eta(\beta))$ is positive and, by definition, $R(\beta)$ is an increasing function. Therefore, the right-hand side of (5.9) is nonnegative for all $\beta < \alpha$ and hence also in the limit as $\beta \uparrow \alpha$. Observe that $\eta(\alpha) = z_0$ and from Lemma 8 that $\mathbf{u}(z_0) = \mathbf{a}$ and $\mathbf{u}'(z_0) = \mathbf{b}$. Therefore, the left-hand side of (5.9) is equal to $\mathbf{a} + z_0\mathbf{b} - \mathbf{b}R(\alpha)$ and so $\mathbf{m}_1(\alpha)$ is nonnegative. Finally, $\mathbf{m}_j(\alpha)$ can be rewritten as

$$\text{(5.10)}\qquad \mathbf{m}_j(\alpha) = \mathbf{m}_{j-1}(\alpha)R(\alpha) + z_0^{j-1}\mathbf{m}_1(\alpha).$$

Hence, by induction, $\{\mathbf{m}_j(\alpha)\}$ is nonnegative since $R(\alpha)$ is a nonnegative matrix.

(iv) Next we show that the sum $c = \sum_{j=1}^{\infty} \mathbf{m}_j(\alpha)\mathbf{e}$ is finite.

This follows simply by summing (5.2) to yield

$$c = \sum_{j=1}^{\infty} \mathbf{m}_j(\alpha)\mathbf{e} = \left(\frac{\mathbf{a}}{(1-z_0)^2} + \frac{\mathbf{b}z_0}{1-z_0} - \mathbf{b}R(\alpha)(I - R(\alpha))^{-1}\right)\mathbf{e}.$$

Since $\mathbf{a}$ and $\mathbf{b}$ are normalized according to $\mathbf{ae} = 1$ and $\mathbf{be} = 0$, c is given by (5.3). It is finite since, by Theorem 5, z_0 [the spectral radius of $R(\alpha)$] is strictly less than 1, so $(I - R(\alpha))^{-1}$ is finite.

(v) We have shown above that $c^{-1}\mathbf{m}(\alpha)$ is a nonnegative solution to (2.1) summing to 1. To show that it is the limiting-conditional distribution, we invoke Theorem 2 of Kesten [5].

We have assumed throughout that (X_n) is irreducible and aperiodic. As the QBD is spatially homogeneous in its levels and the matrices A_0, A_1 and A_2 are finite dimensional, it follows that (X_n) is uniformly irreducible and uniformly aperiodic, in the sense of Kesten [5], and that the jumps are of bounded size. Further, we have assumed that P is stochastic and that absorption is certain, which imply conditions (1.10) and (1.11) of [5]. Therefore, (1.13) of [5] is satisfied (in particular with $k = i$ and $m = 0$) and so the limiting-conditional distribution of (X_n) is $c^{-1}\mathbf{m}(\alpha)$. □

6. Computation of the limiting-conditional distribution. In this section we shall discuss the numerical computation of the limiting-conditional distribution $c^{-1}\mathbf{m}(\alpha)$. The steps involved are:

(i) Solution of (4.7) and (4.8) to find α, z_0 and $\mathbf{a}$.
(ii) Solution of (5.1) to find $\mathbf{b}$.
(iii) Solution of (3.3) to find $R(\alpha)$.
(iv) Computation of c from (5.3).
(v) Calculation of the limiting-conditional distribution from (5.2) or the recursion (5.10).

Below we discuss these steps in more detail.

(i) The minimal z_0 where (4.8) is satisfied can be found using a simple bisection search together with (4.13), testing whether the left-hand side of (4.8) is less than the right-hand side or vice versa. If $\chi'(z)z \geq \chi(z)$, then $z \geq z_0$ and if $\chi'(z)z < \chi(z)$, then $z < z_0$. A similar technique for a related problem [i.e., solving (4.3)] is mentioned in [16], page 40. The value of α can be calculated from (4.7) by first computing $\mathbf{a} = \mathbf{u}(z_0)$ and $\mathbf{v}(z_0)$.

Of course, the bisection search can find z_0 and hence $\alpha = \chi'(z_0)$ only to within some arbitrary accuracy. When the bisection search terminates, the estimate $\bar{\alpha}$ of α should be taken so that $\bar{\alpha} \leq \alpha$; otherwise, $R(\bar{\alpha})$ will not be finite. In our calculations we chose $\bar{\alpha}$ so that $0 \leq \alpha - \bar{\alpha} \leq 10^{-15}$. The implications which flow from the fact that $\bar{\alpha}$ is not exactly equal to α are quite involved and make for an interesting topic for further research. For notational convenience, in the rest of this paper we will suppress the distinction between α and $\bar{\alpha}$.

(ii) Since $\chi(z_0) = z_0/\alpha$ is the maximal eigenvalue of $A(z_0)$ and $A(z_0)$ is irreducible, the matrix on the left-hand side of (5.1) has rank $M - 1$. If we replace the last column of this matrix by the column $\mathbf{e}$ and the last entry of the vector on the right-hand side by 0, then, by the uniqueness part of Lemma 8, the resulting equation has a unique solution which is the required vector $\mathbf{b}$.

(iii) It might be expected that $R(\alpha)$ could be calculated using an iterative scheme similar to those traditionally used in the evaluation of $R(1)$. An example of such a scheme is the recursion (3.6). When this is done, $W^N(\alpha)$ has a probabilistic interpretation as the expected total discounted reward measured on sample paths, the maximum length of which increases with N (for more details, see [11]). However, the fact that we are trying to calculate $R(\beta)$ precisely at the supremum of the values of β for which it is finite greatly increases the influence of very long sample paths. Traditional algorithms thus converge extremely slowly.

In order to take long paths into account, we use a generalization of the logarithmic reduction algorithm of Latouche and Ramaswami [12], Theorem 6.1, for which the maximum length of the sample paths taken into account at step n very rapidly increases with n. We present this in the following theorem.

THEOREM 10. *The matrix $R(\beta)$ for $\beta \leq \alpha$ is given by*

$$R(\beta) = \sum_{k=0}^{\infty} B_0^{(k)} \left(\prod_{r=0}^{k-1} B_2^{(k-1-r)} \right), \tag{6.1}$$

where $B_m^{(k)}$ is defined for $m = 0, 2$ as

$$B_m^{(0)} = \beta A_m (I - \beta A_1)^{-1},$$
$$B_m^{(k+1)} = \left(B_m^{(k)}\right)^2 \left(I - B_0^{(k)} B_2^{(k)} - B_2^{(k)} B_0^{(k)}\right)^{-1}, \qquad k \geq 0.$$

PROOF. Latouche and Ramaswami [12] proved the result for the case $\beta = 1$. Define, for $m = 0, 2$,

$$\tilde{B}_m^{(0)} = A_m\left(\sum_{n=0}^{\infty} A_1^n\right), \tag{6.2}$$

$$\tilde{B}_m^{(k+1)} = (\tilde{B}_m^{(k)})^2\left(\sum_{n=0}^{\infty}(\tilde{B}_0^{(k)}\tilde{B}_2^{(k)} + \tilde{B}_2^{(k)}\tilde{B}_0^{(k)})^n\right), \qquad k \geq 0. \tag{6.3}$$

Theorem 6.1 of [12] proves that

$$R(1) = \sum_{k=0}^{\infty} \tilde{B}_0^{(k)}\left(\prod_{r=0}^{k-1} \tilde{B}_2^{(k-1-r)}\right). \tag{6.4}$$

This theorem may be interpreted as an algebraic statement: if $R(1)$ is finite, then all the relevant series converge and $R(1)$ is given by (6.4).

The same argument with A_0, A_1 and A_2 replaced by βA_0, βA_1 and βA_2 shows that if $R(\beta)$ is finite, then all relevant series converge and (6.1) holds.

Since (X_n) has been shown to be α-transient in Corollary 7, $R(\beta)$ is finite for all $\beta \leq \alpha$, and so the proof is complete. □

To use (6.1) to evaluate $R(\alpha)$, it is necessary to truncate the infinite sum. A simple way to decide where to truncate is to define $R^{(K)}(\alpha)$ as the sum of the first K terms and then to truncate at $K = K^*$, where K^* is the smallest value of K such that $||\mathbf{a}R^{(K)}(\alpha) - z_0\mathbf{a}||_\infty < \varepsilon$ for some tolerance ε.

For given β and k, the matrices $B_0^{(k)}$ and $B_2^{(k)}$ have the following probabilistic interpretation. Define the first-passage time (greater than 0) to level $1+2^k$ to be τ_k, for $k \geq 0$, and let T be the first-passage time (greater than 0) to level 1. Then

$$(B_0^{(k)})_{ij} = \mathbb{E}\left[\sum_{n=\tau_k}^{T\wedge\tau_{k+1}} \beta^n \mathbb{I}[X_n = (1+2^k, j)] | X_0 = (1, i)\right]$$

and

$$(B_2^{(k)})_{ij} = \mathbb{E}\left[\sum_{n=\tau_k}^{T\wedge\tau_{k+1}} \beta^n \mathbb{I}[X_n = (1+2^k, j)] | X_0 = (1+2^{k+1}, i)\right].$$

After some tedious manipulation, it can be shown that the (i, j)th entry of the kth term of the sum in (6.1) has the interpretation

$$\mathbb{E}\left[\sum_{n=\tau_k}^{T\wedge\tau_{k+1}} \beta^n \mathbb{I}[X_n = (2, j)] | X_0 = (1, i)\right].$$

That is, only the trajectories starting from $(1, i)$ that visit $(2, j)$ and reach the level $1+2^k$ but do not go beyond the level 2^{k+1} are included. Hence as k tends to ∞, the lengths of the sample paths whose influence is included in the sum grow at an exponential rate. Latouche and Ramaswami [12] noted that because of this structure, the probabilities of the sample paths will tend to 0 very quickly as k tends to ∞. Hence, in the special case where $\beta = 1$, only a small value of K^* will usually be necessary. Typically, K^* of the order of 5 or 6 will suffice. In the case where $\beta > 1$, as mentioned above, the weighting β^n tends to ∞ so that very long sample paths can make significant contributions. Thus the sum in (6.1) converges more slowly when $\beta = \alpha$ than when $\beta = 1$. We have found that K^* is usually between 30 and 50, when α is taken within 10^{-15} of its true value. Note that if $K^* = 30$, then we have taken into account the influence of sample paths that reach as high as level 2^{30} before returning to level 1.

REMARK 3. From our experience in calculating $R(\alpha)$, we found that as k gets large, the matrix $B_0^{(k)}$ becomes 0 to machine accuracy and the matrix $B_2^{(k)}$ becomes infinite to machine accuracy. Hence, to calculate the product $B_0^{(k)}(\prod_{r=0}^{k-1} B_2^{(k-1-r)})$ for large values of k, it is necessary to use an exponential scaling technique. A simple way to do this is to write all matrices $B_i^{(k)}$ in the form $\bar{B}_i^{(k)} \times 10^{b_i}$, where the maximum element of $\bar{B}_i^{(k)}$ is between 1 and 10, and then the product $B_i^{(k)} B_j^{(l)}$ is given by $\bar{B}_i^{(k)} \bar{B}_j^{(l)} \times 10^{(b_i+b_j)}$.

(iv) It is now a very simple matter to compute c from (5.3).

(v) The limiting-conditional distribution $\{c^{-1}\mathbf{m}_j(\alpha)\}$ can now be computed from (5.2) on division by c or via the recursion (5.10) after determining $c^{-1}\mathbf{m}_1(\alpha)$ from (5.2).

7. Numerical examples. In this section we present two numerical examples. The first example considers a continuous-time PH/PH/1 queue, while the second example considers a model of an animal population.

Ramaswami and Latouche [17] modeled a PH/PH/1 queue in continuous time where the arrival time distribution was a five-stage Erlang distribution and the service time distribution was the mixture $0.7E_4(6.67778512) + 0.3E_2(1.03323700)$, where $E_i(\lambda)$ represents an i-stage Erlang distribution with parameter λ. If we uniformize this continuous-time process, then we obtain a discrete-time analogue of the PH/PH/1 queue with A_0, A_1 and A_2 given by

$$\begin{aligned} A_0 &= (\mathbf{T}^0\boldsymbol{\xi}) \otimes I_6, \\ A_1 &= I_{30} + T \otimes I_6 + I_5 \otimes S, \\ A_2 &= I_5 \otimes (\mathbf{S}^0\boldsymbol{\eta}). \end{aligned}$$

In the above, $\otimes$ is the Kronecker product, I_n is the $n \times n$ identity matrix, the vectors $\boldsymbol{\xi}$, $\boldsymbol{\eta}$, $\mathbf{T}^0$ and $\mathbf{S}^0$ are given by

$$\begin{aligned} \boldsymbol{\xi} &= (1, 0, 0, 0, 0), \\ \boldsymbol{\eta} &= (0.7, 0, 0, 0, 0.3, 0), \\ \mathbf{T}^0 &= (0, 0, 0, 0, \mu/\Delta)', \\ \mathbf{S}^0 &= (0, 0, 0, \lambda_1/\Delta, 0, \lambda_2/\Delta)' \end{aligned}$$

and the matrices T and S are given by

$$T = \begin{pmatrix} \frac{-\mu}{\Delta} & \frac{\mu}{\Delta} & 0 & 0 & 0 \\ 0 & \frac{-\mu}{\Delta} & \frac{\mu}{\Delta} & 0 & 0 \\ 0 & 0 & \frac{-\mu}{\Delta} & \frac{\mu}{\Delta} & 0 \\ 0 & 0 & 0 & \frac{-\mu}{\Delta} & \frac{\mu}{\Delta} \\ 0 & 0 & 0 & 0 & \frac{-\mu}{\Delta} \end{pmatrix},$$

$$S = \begin{pmatrix} \frac{-\lambda_1}{\Delta} & \frac{\lambda_1}{\Delta} & 0 & 0 & 0 & 0 \\ 0 & \frac{-\lambda_1}{\Delta} & \frac{\lambda_1}{\Delta} & 0 & 0 & 0 \\ 0 & 0 & \frac{-\lambda_1}{\Delta} & \frac{\lambda_1}{\Delta} & 0 & 0 \\ 0 & 0 & 0 & \frac{-\lambda_1}{\Delta} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{-\lambda_2}{\Delta} & \frac{\lambda_2}{\Delta} \\ 0 & 0 & 0 & 0 & 0 & \frac{-\lambda_2}{\Delta} \end{pmatrix},$$

with $\Delta = \lambda_1 + \lambda_2 + \mu$, $\lambda_1 = 6.67778512$ and $\lambda_2 = 1.03323700$. It is easy to show that the matrix $A = A_0 + A_1 + A_2$ is stochastic and irreducible.

The traffic intensity ρ is given by $\rho = \boldsymbol{\pi} A_0 \mathbf{e}/\boldsymbol{\pi} A_2 \mathbf{e}$, where $\boldsymbol{\pi}$ satisfies $\boldsymbol{\pi} A = \boldsymbol{\pi}$. By varying the value of μ, we can consider a range of values of ρ. In our computations we chose K^* to be the smallest value of K such that $||\mathbf{a}R^{(K)}(\alpha) - z_0\mathbf{a}||_\infty < 10^{-7}$. In Figure 2 we present a graph of α versus ρ for high values of ρ.

For all the values of ρ considered in Figure 2, the value of K^* was 19. The marginal quasi-stationary distribution for the levels $\{\mathbf{m}_k\mathbf{e}\}$ is given in Table 1 for $k = 1$ to 60 and $\rho = 0.95$.

In the mathematical modeling of ecological systems, there has been considerable interest in the incorporation of environmental stochasticity; see, for example, [4], [10] and [14]. These authors have recognized that birth-and-death process models and birth–death-and-catastrophe process models are inappropriate for modeling the evolution of many populations because they do not capture the effect of environmental variation. In each of [4], [10] and [14], methods were developed to include the environmental variation. However, none of the papers takes what seems to be the obvious way to incorporate environmental stochasticity, by adding an auxiliary variable to the model which reflects the state of the environment. With this auxiliary variable, a birth-and-death process model turns into a quasi-birth-and-death process model (QBD).

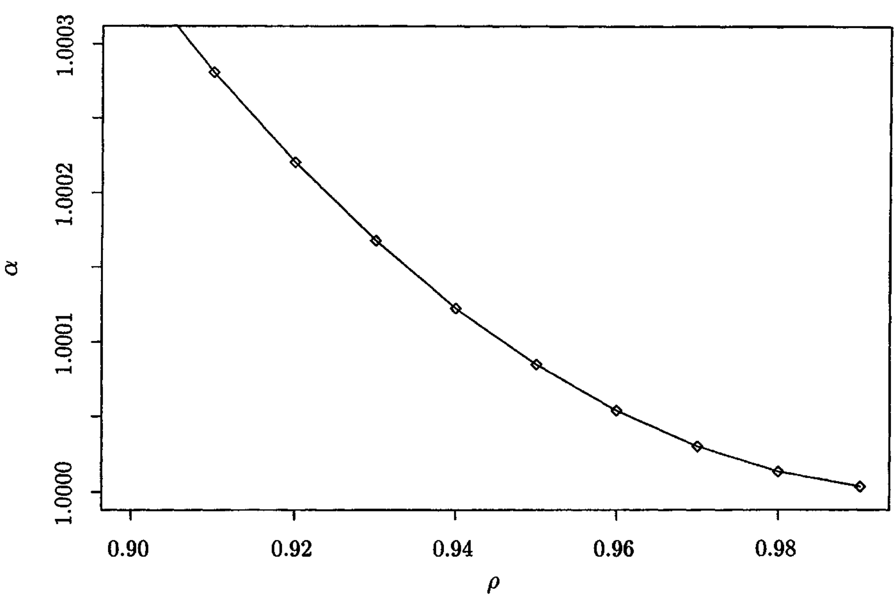

FIG. 2. *Graph of α against ρ.*

TABLE 1
Marginal quasi-stationary distribution for the PH / PH / 1 model with $\rho = 0.95$

k	$\mathbf{m}_k\mathbf{e}$	k	$\mathbf{m}_k\mathbf{e}$	k	$\mathbf{m}_k\mathbf{e}$
1	1.25550076e – 03	21	1.53793897e – 02	41	1.29890119e – 02
2	2.96270172e – 03	22	1.54540790e – 02	42	1.27581921e – 02
3	4.40736429e – 03	23	1.54963951e – 02	43	1.25242950e – 02
4	5.73225888e – 03	24	1.55088834e – 02	44	1.22879682e – 02
5	6.94587984e – 03	25	1.54939344e – 02	45	1.20498107e – 02
6	8.05132099e – 03	26	1.54537921e – 02	46	1.18103764e – 02
7	9.05415217e – 03	27	1.53905618e – 02	47	1.15701765e – 02
8	9.96060161e – 03	28	1.53062181e – 02	48	1.13296822e – 02
9	1.07768213e – 02	29	1.52026118e – 02	49	1.10893272e – 02
10	1.15086902e – 02	30	1.50814771e – 02	50	1.08495098e – 02
11	1.21617747e – 02	31	1.49444375e – 02	51	1.06105951e – 02
12	1.27413303e – 02	32	1.47930126e – 02	52	1.03729173e – 02
13	1.32523141e – 02	33	1.46286236e – 02	53	1.01367813e – 02
14	1.36993997e – 02	34	1.44525988e – 02	54	9.90246494e – 03
15	1.40869917e – 02	35	1.42661790e – 02	55	9.67022032e – 03
16	1.44192401e – 02	36	1.40705224e – 02	56	9.44027568e – 03
17	1.47000536e – 02	37	1.38667092e – 02	57	9.21283689e – 03
18	1.49331124e – 02	38	1.36557461e – 02	58	8.98808890e – 03
19	1.51218806e – 02	39	1.34385709e – 02	59	8.76619712e – 03
20	1.52696178e – 02	40	1.32160558e – 02	60	8.54730872e – 03

In the example below we illustrate how a population can be modeled as a QBD. Let the level represent the number in the population (or some function of this) and the phase the state of the environment. In a realistic model, there are many issues that need to be resolved, such as what a level should represent, what time scales should be considered and how the environment should be described. Any model will be highly dependent on the particular population under consideration. Here we develop a simple model for a population of the greater bilby (*Macrotis lagotis*) [20], commonly known as the bilby. The bilby is a small Australian marsupial that is currently an endangered species.

It has been observed that the breeding patterns of bilbies depend on the "quality" of the preceding seasons. An important measure of the quality of a season is how much rainfall there was during that season. In our model we specify that a season is either of good or of poor quality. Bilbies can sustain bad seasons and indeed breed during bad seasons, but, when there are consecutive bad seasons, they find it increasingly difficult, or impossible, to breed. We account for this by letting the phase represent how many consecutive bad seasons there have been.

In our model we observe the population at time points that correspond to ends of seasons. Assume that in the state (k, j), k gives a measure of the number in the population and the phase $j = 1, 2, 3, 4, 5$ indicates there have been $j-1$ consecutive bad seasons, where $j = 1$ means that the last season was a good season. Assume the qualities of the seasons are independent random variables and that the probability of a season being good is g. Let b_j be the probability that the population moves up a level given the phase is j and let d_j be the probability that the population moves down a level given the phase is j. Define c_j by $c_j = 1 - b_j - d_j$. The modeling assumptions detailed above indicate that b_j is decreasing in j and d_j is increasing in j. The matrices A_0, A_1 and A_2 are given by

$$A_0 = \begin{pmatrix} gb_1 & (1-g)b_1 & 0 & 0 & 0 \\ gb_2 & 0 & (1-g)b_2 & 0 & 0 \\ gb_3 & 0 & 0 & (1-g)b_3 & 0 \\ gb_4 & 0 & 0 & 0 & (1-g)b_4 \\ gb_5 & 0 & 0 & 0 & (1-g)b_5 \end{pmatrix},$$

$$A_1 = \begin{pmatrix} gc_1 & (1-g)c_1 & 0 & 0 & 0 \\ gc_2 & 0 & (1-g)c_2 & 0 & 0 \\ gc_3 & 0 & 0 & (1-g)c_3 & 0 \\ gc_4 & 0 & 0 & 0 & (1-g)c_4 \\ gc_5 & 0 & 0 & 0 & (1-g)c_5 \end{pmatrix}$$

and

$$A_2 = \begin{pmatrix} gd_1 & (1-g)d_1 & 0 & 0 & 0 \\ gd_2 & 0 & (1-g)d_2 & 0 & 0 \\ gd_3 & 0 & 0 & (1-g)d_3 & 0 \\ gd_4 & 0 & 0 & 0 & (1-g)d_4 \\ gd_5 & 0 & 0 & 0 & (1-g)d_5 \end{pmatrix}.$$

TABLE 2
Marginal quasi-stationary distribution for the bilby model

k	$\mathbf{m}_k\mathbf{e}$	k	$\mathbf{m}_k\mathbf{e}$	k	$\mathbf{m}_k\mathbf{e}$
1	4.47898698e − 02	21	1.12862463e − 02	41	3.53678358e − 04
2	6.16374346e − 02	22	9.61375492e − 03	42	2.94731678e − 04
3	7.18438293e − 02	23	8.17266163e − 03	43	2.45472322e − 04
4	7.61207462e − 02	24	6.93480561e − 03	44	2.04336854e − 04
5	7.62851461e − 02	25	5.87448567e − 03	45	1.70008121e − 04
6	7.37472224e − 02	26	4.96852149e − 03	46	1.41377755e − 04
7	6.95059750e − 02	27	4.19620665e − 03	47	1.17514101e − 04
8	6.42841711e − 02	28	3.53919164e − 03	48	9.76348463e − 05
9	5.85946841e − 02	29	2.98132450e − 03	49	8.10837120e − 05
10	5.27931756e − 02	30	2.50846906e − 03	50	6.73106495e − 05
11	4.71189967e − 02	31	2.10831456e − 03	51	5.58550383e − 05
12	4.17261784e − 02	32	1.77018623e − 03	52	4.63314578e − 05
13	3.67069182e − 02	33	1.48486309e − 03	53	3.84176558e − 05
14	3.21092779e − 02	34	1.24440683e − 03	54	3.18443906e − 05
15	2.79504316e − 02	35	1.04200397e − 03	55	2.63868643e − 05
16	2.42265095e − 02	36	8.71822327e − 04	56	2.18575089e − 05
17	2.09198442e − 02	37	7.28881878e − 04	57	1.80999167e − 05
18	1.80042484e − 02	38	6.08939715e − 04	58	1.49837412e − 05
19	1.54488040e − 02	39	5.08388253e − 04	59	1.24004150e − 05
20	1.32205344e − 02	40	4.24165795e − 04	60	1.02595596e − 05

Note that this model does not fit into the class of PH/PH/c queues as defined, for example, in [16], Section 3.7. This is easy to see, since there are several values of $i \neq j$ for which both $(A_0)_{i,j}$ and $(A_2)_{i,j}$ are positive. When calculating α and $R(\alpha)$ we used the same tolerances as in the previous example. For the parameter values $g = 0.2$, $b_1 = 1$, $b_2 = 0.4$, $b_3 = 0.25$, $b_4 = 0.1$, $b_5 = 0$, $c_1 = 0$, $c_2 = 0.1$, $c_3 = 0.2$, $c_4 = 0.1$, $c_5 = 0$, $d_1 = 0$, $d_2 = 0.5$, $d_3 = 0.55$, $d_4 = 0.8$, $d_5 = 1$, we found K^* was equal to 23. The marginal quasi-stationary distribution for levels k ranging from 1 to 60 is given in Table 2.

Acknowledgments. The authors would like to thank Hugh Possingham for drawing our attention to the bilby example. The authors would also like to thank the referees for carefully reading an earlier draft of the paper and for making many constructive comments.

REFERENCES

[1] ANDERSON, W. J. (1991). *Continuous-Time Markov Chains: An Applications Oriented Approach*. Springer, New York.
[2] DARROCH, J. N. and SENETA, E. (1965). On quasi-stationary distributions in absorbing discrete-time Markov chains. *J. Appl. Probab.* **2** 88–100.
[3] GANTMACHER, F. R. (1959). *The Theory of Matrices*. Chelsea, New York.
[4] GOODMAN, D. (1987). Consideration of stochastic demography in the design and managment of biological reserves. *Natur. Resource Modeling* **1** 205–233.

[5] KESTEN, H. (1995). A ratio limit theorem for (sub) Markov chains on $\{1, 2, 3\ldots\}$ with bounded jumps. *Adv. in Appl. Probab.* **27** 652–691.

[6] KIJIMA, M. (1993). Quasi-stationary distributions of phase-type queues. *Math. Oper. Res.* **18** 423–437.

[7] KINGMAN, J. F. C. (1961). A convexity property of positive matrices. *Quart. J. Math. Oxford Ser. (2)* **12** 283–284.

[8] KYPRIANOU, E. K. (1972). On the quasi-stationary distributions of the GI/M/1 queues. *J. Appl. Probab.* **9** 117–128.

[9] KYPRIANOU, E. K. (1972). The quasi-stationary distributions of queues in heavy traffic. *J. Appl. Probab.* **9** 821–831.

[10] LANDE, R. (1993). Risks of population extinction from demographic and environmental stochasticity and random catastrophes. *American Naturalist* **142** 911–927.

[11] LATOUCHE, G. (1993). Algorithms for infinite Markov chains with repeating columns. In *Linear Algebra, Markov Chains and Queueing Models* (C. D. Meyer and R. J. Plemmons, eds.) 231–265. Springer, New York.

[12] LATOUCHE, G. and RAMASWAMI, V. (1993). A logarithmic reduction algorithm for quasi-birth-death processes. *J. Appl. Probab.* **30** 650–674.

[13] MAKIMOTO, N. (1993). Quasi-stationary distributions in a PH/PH/c queue. *Stochastic Models* **9** 195–212.

[14] MANGEL, M. and TIER, C. (1993). Dynamics of metapopulation with demographic stochasticity and environmental catastrophes. *Theoret. Population Biol.* **44** 1–31.

[15] NEUTS, M. F. (1978). Markov chains with applications in queueing theory, which have a matrix-geometric invariant probability vector. *Adv. in Appl. Probab.* **10** 185–212.

[16] NEUTS, M. F. (1981). *Matrix Geometric Solutions in Stochastic Models*. Johns Hopkins Univ. Press.

[17] RAMASWAMI, V. and LATOUCHE, G. (1981). An experimental evaluation of the matrix-geometric method for the GI/PH/1 queue. *Stochastic Models* **5** 629–667.

[18] SENETA, E. (1981). *Non-negative Matrices and Markov Chains*. Springer, New York.

[19] 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.

[20] SOUTHGATE, R. and POSSINGHAM, H. P. (1995). Population viability analysis of the bilby, *lagotis macrotis*. *Biological Conservation* **73** 143–150.

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

N. G. BEAN
DEPARTMENT OF APPLIED MATHEMATICS
UNIVERSITY OF ADELAIDE
S. A. 5005
AUSTRALIA
E-MAIL: nbean@maths.adelaide.edu.au

L. BRIGHT
DEPARTMENT OF APPLIED MATHEMATICS
UNIVERSITY OF ADELAIDE
S. A. 5005
AUSTRALIA
E-MAIL: lesb@nabaus.com.au

G. LATOUCHE
DÉPARTEMENT D'INFORMATIQUE
UNIVERSITÉ LIBRE DE BRUXELLES
BRUXELLES
BELGIUM
E-MAIL: latouche@ulb.ac.be

C. E. M. PEARCE
DEPARTMENT OF APPLIED MATHEMATICS
UNIVERSITY OF ADELAIDE
S. A. 5005
AUSTRALIA
E-MAIL: cpearce@maths.adelaide.edu.au

P. K. POLLETT
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF QUEENSLAND
QLD. 4072
AUSTRALIA
E-MAIL: pkp@maths.uq.oz.au

P. G. TAYLOR
DEPARTMENT OF APPLIED MATHEMATICS
UNIVERSITY OF ADELAIDE
S. A. 5005
AUSTRALIA
E-MAIL: ptaylor@maths.adelaide.edu.au