Level-Dependent Quasi-Birth-and-Death Processes

Jeffrey P. Kharoufeh
Department of Industrial Engineering
University of Pittsburgh
1048 Benedum Hall
3700 O'Hara Street
Pittsburgh, PA 15261 USA
Ph: (412) 624-9832; Fax: (412) 624-9831
Email: jkharouf@pitt.edu

Abstract

This article defines and describes the level-dependent quasi-birth-and-death (LDQBD) process, a generalization of the homogeneous (or level-independent) quasi-birth-and-death (QBD) process. Like its level-independent counterpart, the LDQBD process is a bivariate Markov process whose transition probability matrix (or infinitesimal generator matrix) exhibits a block tri-diagonal structure. However, unlike the level-independent QBD, its transitions are explicitly dependent on the level. This article defines discrete- and continuous-time versions of the LDQBD process, discusses necessary and sufficient conditions for positive recurrence, and describes the limiting distribution for each case. Algorithmic approaches, extensions of the basic models, and suggestions for further reading are also provided.

Introduction

A quasi-birth-and-death (QBD) process is a bivariate Markov process with state space $S = \{(i,j) : i \geq 0,\ j = 1,2,\ldots,m\}$ where i is called the *level* of the process, j is called the *phase* of the process, and m is an integer that can be finite or infinite. The process is restricted in level jumps only to its nearest neighbors but is unrestricted in the phase dimension. More precisely, from state $(i,j) \in S$ the process may transition to states of the form (i,k), $(i-1,k)$ or $(i+1,k)$, but not to states of the form $(i \pm n, k)$ where $n \geq 2$. Clearly, the QBD process is an extension of the standard birth-and-death process whose state space consists only of the level i. When the transitions of a QBD process are independent of the level, it is termed a *homogeneous* or *level-independent* QBD process. Otherwise, it is termed an *inhomogeneous* or *level-dependent* QBD (LDQBD) process. This very general class of models finds wide applicability in the modeling of queueing systems, particularly those with complex arrival and/or service processes such as the $PH/M/\infty$ queue. Several queueing, and non-queueing, applications of LDQBD processes will be discussed in the last section (Further Reading and Applications). The level-independent version has been studied extensively by a number of researchers and is given a very lucid treatment in the excellent text by Latouche and Ramaswami [30]. As a prototypical example, the homogeneous QBD process can be used to model the $M/M/1$ queueing system in a random environment where the level i corresponds to the number of customers in the system, and the phase j is the state of an exogenous environment process that governs the arrival and/or service rates (see Neuts [36]). The class of QBD models, and their associated matrix-analytic solution techniques, provide a powerful set of tools for analyzing a wide variety of stochastic models arising in queueing theory, computer and communications systems, reliability modeling and analysis, inventory systems, and many others. For further information on the homogeneous QBD process, please see the article *Level-Independent Quasi-Birth-and-Death Processes.*

The main purpose of the present article is to formally define and describe level-dependent quasi-birth-and-death processes whose transitions depend explicitly on the level i. Although their structure, and many of their attributes, mirror those of the level-independent version, their analysis and computational aspects are significantly more complicated. Here, we review both discrete- and continuous-time versions of the LDQBD process. For each type, we discuss necessary and sufficient conditions to establish positive recurrence, describe the limiting distribution, and discuss computational methods proposed for analyzing these models. We begin by describing the discrete-time version in the next section.

Discrete-Time LDQBD Processes

Suppose $\{(X_n, J_n) : n \geq 0\}$ is a discrete-time Markov process on the state space $S = \{(i, j) : i \geq 0,\ j = 1, 2, \ldots, m\}$ where i and j are the level and phase of the process, respectively. For the purposes of this article, we will assume m is finite; however, m may be infinite and/or depend explicitly on the level i (as described by Bright and Taylor [13]). For the sake of notational brevity, let $\{Z_n : n \geq 0\} \equiv \{(X_n, J_n) : n \geq 0\}$. The one-step transition probability matrix of this process possesses the typical block tri-diagonal form of a QBD process and is given by

$$P = \begin{bmatrix} A_1^{(0)} & A_0^{(0)} & 0 & 0 & \cdots \\ A_2^{(1)} & A_1^{(1)} & A_0^{(1)} & 0 & \cdots \\ 0 & A_2^{(2)} & A_1^{(2)} & A_0^{(2)} & \cdots \\ 0 & 0 & A_2^{(3)} & A_1^{(3)} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}. \tag{1}$$

For $i \geq 0$ and $\ell = 0, 1, 2$, $A_\ell^{(i)}$ are m-order non-negative matrices, and the row sums are equal to unity, i.e., the row sums of $A^{(i)} \equiv A_2^{(i)} + A_1^{(i)} + A_0^{(i)}$ and $A_1^{(0)} + A_0^{(0)}$ are equal to 1. To interpret the entries of P, we note that for $i \geq 1$, the process transitions from (i, j) to $(i-1, k)$ with probability $[A_2^{(i)}]_{j,k}$, or it transitions to state $(i+1, k)$ with probability $[A_0^{(i)}]_{j,k}$. A transition from (i, j) to (i, k) occurs with probability $[A_1^{(i)}]_{j,k}$. Finally, when $i = 0$, from state $(0, j)$ the process transitions to state $(1, k)$ with probability $[A_0^{(0)}]_{j,k}$, or to state $(0, k)$ with probability $[A_1^{(0)}]_{j,k}$. It should be clear that the rows of P are blocks of order m, and the structure indicates that the process is skip-free in the level in both directions. For the remainder of this section, we assume that P is irreducible.

Positive Recurrence and Limiting Distribution

A natural question is whether the Markov process with transition matrix P of (1) possesses a limiting distribution, and if so, under what conditions does the distribution exist and what is its form? Unfortunately, the extreme generality of the LDQBD process precludes the development of generic tools that can be used to establish such properties as irreducibility, recurrence, and positive recurrence; however, in practice, it is often possible to establish these properties for specific applications using the attributes of the model. Stochastic dominance arguments are also often employed to show that the LDQBD process is stochastically dominated by a simpler process (e.g., a homogeneous QBD process) whose irreducibility and recurrence properties can be more easily established. Our aim here is to review the theoretical conditions for positive recurrence of irreducible LDQBD processes.

Denote the limiting distribution of $\{Z_n : n \geq 0\}$ by a positive row vector $\boldsymbol{\pi}$ that uniquely solves the system $\boldsymbol{\pi} P = \boldsymbol{\pi}$ and $\boldsymbol{\pi} \boldsymbol{e} = 1$, where $\boldsymbol{e}$ is a column vector of ones. The vector $\boldsymbol{\pi}$ is partitioned by levels into subvectors so that

$$\boldsymbol{\pi} = (\boldsymbol{\pi}_0, \boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \ldots),$$

and $\boldsymbol{\pi}_i$ contains the limiting probabilities for states in level i, i.e., the states in the set $L_i \equiv \{(i,1),\, (i,2), \ldots, (i,m)\}$. Note that $\boldsymbol{\pi}_i$ has m components for each i. As shown in Chapter 6 of [30], the vector $\boldsymbol{\pi}_i$ assumes a *matrix-geometric* form when the process is level-independent, i.e., when for $\ell = 0, 1, 2$, $A_\ell^{(i)} = A_\ell$ for every $i \geq 0$. More specifically, for $i \geq 0$, the vector $\boldsymbol{\pi}_i$ satisfies

$$\boldsymbol{\pi}_i = \boldsymbol{\pi}_0 \, R^i,$$

where R is a fixed matrix (independent of the level) such that for any pair of phases j and k, $[R]_{j,k}$ is the expected number of visits to state $(1,k)$ before returning to the set L_0, provided that the process starts in state $(0,j)$. The result is a matrix analogue to the scalar geometric distribution. (For a more thorough treatment of the matrix-geometric distribution, please see the article *Matrix-Geometric Distributions.*) Before stating the necessary and sufficient conditions for positive recurrence of a LDQBD process, we first assume it is irreducible, aperiodic, and positive recurrent in order to characterize the limiting distribution $\boldsymbol{\pi}$. The following theorem mirrors Theorem 12.1.1. in [30] and is stated here without proof.

Theorem 1 *If the LDQBD process with transition probability matrix given by* (1) *is irreducible, aperiodic, and positive recurrent, then there exist matrices* $\{R_i : i \geq 1\}$*, such that*

$$\boldsymbol{\pi}_i = \boldsymbol{\pi}_{i-1} R_i, \quad i \geq 1, \tag{2}$$

where the matrix R_i *records the expected number of visits to the set* L_i *between any two visits to the set* L_{i-1}*, and the sequence* $\{R_i\}$ *is the minimal nonnegative solution of the set of equations given by*

$$R_i = A_0^{(i-1)} + R_i \, A_1^{(i)} + R_i \, R_{i+1} \, A_2^{(i+1)}, \quad i \geq 1.$$

Stated more precisely, $[R_i]_{j,k}$ is equal to the expected number of visits to $(i+1,k)$ before a return to the set $L_0 \cup L_1 \cup \cdots \cup L_i$, provided the process starts in state (i,j). Notice that the expected number of visits depends explicitly on the level i from which we start. To determine conditions for positive recurrence, we next discuss first passage times of $\{Z_n : n \geq 0\}$ and introduce some important matrices that are very similar to those typically used to characterize the behavior of level-independent QBD processes.

Suppose $\{Z_n : n \geq 0\}$ starts in some state (i,j), i.e., in level $i > 0$ and phase j at time 0, and define S_1 as the first return time to level i. That is,

$$S_1 \equiv \inf\{n \geq 1 : Z_n \in L_i | Z_0 = (i,j)\}$$

Furthermore, adopting the notation in [30], let τ denote the first passage time of $\{Z_n : n \geq 0\}$ to level $i-1$. Define the following two probabilities:

$$[U_i]_{j,k} = \mathbb{P}\left(S_1 < \tau,\, Z_{S_1} = (i,k) | Z_0 = (i,j)\right), \quad (3)$$

$$[G_i]_{j,k} = \mathbb{P}\left(\tau < \infty,\, Z_\tau = (i-1,k) | Z_0 = (i,j)\right). \quad (4)$$

Note that equations (3) and (4) bear a striking resemblance to equations (6.5) and (6.6) of [30] with one major exception – the probabilities above depend explicitly on the level i whereas they are completely independent of i when the QBD is level-independent. We shall see that this dependence significantly complicates the process of computing the limiting distribution of a LDQBD process. The (j,k)th element of the matrix U_i is the probability that, given we start in level i, the process returns to state i before hitting level $i-1$, and the (j,k)th element of G_i denotes the probability that the process hits level $i-1$ in a finite duration of time, given that it starts somewhere in level i. It is not surprising that the sequences of matrices, $\{R_i\}$, $\{G_i\}$, and $\{U_i\}$ are interrelated. In fact, any one of the sequences determines the other two as noted by Theorem 12.1.2 in [30] which is restated here without proof.

Theorem 2 *Any one of the sequences $\{R_i\}$, $\{G_i\}$, and $\{U_i\}$ determines the other two via the following set of relations:*

$$\begin{aligned}
G_i &= (I - U_i)^{-1} A_2^{(i)}, \\
R_i &= A_0^{(i-1)} (I - U_i)^{-1}, \\
U_i &= A_1^{(i)} + A_0^{(i)} G_{i+1}, \\
U_i &= A_1^{(i)} + R_{i+1} A_2^{(i)}.
\end{aligned}$$

Note that the same relationships hold for R, G, and U in the level-independent case wherein none of the matrices depend on the level i. For computational purposes, the following result, Theorem 12.1.3 of [30], shows the explicit relationship between these matrices

Theorem 3 *For each $i \geq 1$, the matrices R_i, U_i, and G_i verify the following set of equations:*

$$\begin{aligned} G_i &= A_2^{(i)} + A_1^{(i)} G_i + A_0^{(i)} G_{i+1} G_i, \\ R_i &= A_0^{(i-1)} + R_i A_1^{(i)} + R_i\, R_{i+1}\, A_2^{(i+1)}, \\ U_i &= A_1^{(i)} + A_0^{(i)} \left(I - U_{i+1}\right)^{-1} A_2^{(i+1)}. \end{aligned}$$

In light of the above two theorems and equation (2), it is clear that the condition for positive recurrence can be restated in terms of the matrix G_1. The following theorem provides the necessary and sufficient condition that ensures positive recurrence of P (see also Theorem 12.1.4 of [30]).

Theorem 4 *The LDQBD process is positive recurrent if and only if there exists a strictly positive solution to the system of equations*

$$\boldsymbol{\pi}_0 = \boldsymbol{\pi}_0 \left(A_1^{(0)} + A_0^{(0)} G_1 \right) \tag{5}$$

subject to the normalization condition

$$\boldsymbol{\pi}_0 \left(\sum_{n=0}^{\infty} \prod_{i=1}^{n} R_i \right) \boldsymbol{e} = 1 \tag{6}$$

where the empty product (when $n = 0$) results in the identity matrix.

We pause here to remark that the matrix G_1 can only be obtained approximately in practice; therefore, it is difficult to assert positive recurrence using (5) and (6). However, the matrix equation (5) can be replaced by

$$\boldsymbol{\pi}_0 = \boldsymbol{\pi}_0 \left(A_1^{(0)} + R_1\, A_2^{(1)} \right) \tag{7}$$

due to the equivalence of $A_0^{(0)} G_1$ and $R_1\, A_2^{(1)}$. Note that, if $m = \infty$, then we also require that the embedded Markov chain on L_0 be positive recurrent. This condition is obviously met when $m < \infty$. Taken together, equations (2), (6), and (7) provide the limiting distribution $\boldsymbol{\pi}$. It is obvious that we must compute the family of matrices $\{R_i : i \geq 1\}$ to obtain this limiting distribution. By contrast, we need only to compute the matrix R (or G or U) in the level-independent case to completely characterize $\boldsymbol{\pi}$. We will review some algorithmic approaches for both the level-independent and level-dependent versions later. First, we describe and review the continuous-time version of the LDQBD process which parallels the discrete-time case.

Continuous-Time LDQBD Processes

The discrete-time LDQBD process has a natural analogue in continuous time which we now describe. Suppose $\{(X(t), J(t)) : t \geq 0\}$ is a continuous-time Markov process on the state space

$S = \{(i,j) : i \geq 0,\, j = 1, 2, \ldots, m\}$ where i denotes the level of the process and j denotes the phase. The infinitesimal generator matrix of the continuous-time LDQBD process is of the form

$$Q = \begin{bmatrix} A_1^{(0)} & A_0^{(0)} & 0 & 0 & \cdots \\ A_2^{(1)} & A_1^{(1)} & A_0^{(1)} & 0 & \cdots \\ 0 & A_2^{(2)} & A_1^{(2)} & A_0^{(2)} & \cdots \\ 0 & 0 & A_2^{(3)} & A_1^{(3)} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}. \tag{8}$$

For each $i \geq 0$, the diagonal elements of $A_1^{(i)}$ are strictly negative, and the off-diagonal elements of $A_1^{(i)}$ are non-negative. The matrices $A_0^{(i)}$ and $A_2^{(i)}$ are non-negative. For $i \geq 1$, the matrix $A^{(i)} = A_0^{(i)} + A_1^{(i)} + A_2^{(i)}$ has zero row sums as does the matrix $A_0^{(0)} + A_1^{(0)}$. The structure of the generator matrix Q reveals that its transitions, like the discrete-time counterpart, are restricted to nearest neighbors in the levels and unrestricted (in general) across the phase dimension. For $i \geq 1$, the process transitions from (i,j) to $(i-1,k)$ at rate $[A_2^{(i)}]_{j,k}$, it transitions to state $(i+1,k)$ with rate $[A_0^{(i)}]_{j,k}$, or to state (i,k) at rate $[A_1^{(i)}]_{j,k}$. Finally, from state $(0,j)$, the process transitions to $(1,k)$ at rate $[A_0^{(0)}]_{j,k}$, or to state $(0,k)$ with rate $[A_1^{(0)}]_{j,k}$. If the process possesses a limiting distribution, $\boldsymbol{\pi} = (\boldsymbol{\pi}_0, \boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \ldots)$, then it must satisfy the system of equations $\boldsymbol{\pi} Q = \mathbf{0}$ and $\boldsymbol{\pi}\boldsymbol{e} = 1$, where $\mathbf{0}$ denotes the zero vector. We next review the necessary and sufficient condition for positive recurrence of the LDQBD process in continuous time and provide an expression for its limiting distribution.

Positive Recurrence and Limiting Distribution

As one might expect, the limiting distribution of the continuous-time version is analogous to its discrete-time counterpart. For the discussion that follows, we assume that $\{(X(t), J(t)) : t \geq 0\}$ with generator matrix Q is irreducible. The following main result shows that the limiting distribution restricted to any one of the levels depends explicitly on $\boldsymbol{\pi}_0$.

Theorem 5 *The LDQBD process with infinitesimal generator matrix given by* (8) *is positive recurrent if and only if there exists a strictly positive solution to the system of equations*

$$\boldsymbol{\pi}_0 \left(A_1^{(0)} + R_0\, A_2^{(1)} \right) = \mathbf{0} \tag{9}$$

subject to the normalization condition

$$\boldsymbol{\pi}_0 \left(\sum_{i=0}^{\infty} \prod_{n=0}^{i-1} R_i \right) \boldsymbol{e} = 1. \tag{10}$$

In such a case, the m-order row vector $\boldsymbol{\pi}_i$ is given by

$$\boldsymbol{\pi}_i = \boldsymbol{\pi}_0 \prod_{n=0}^{i-1} R_n, \quad i \geq 0. \tag{11}$$

In equations (10) and (11), when $i = 0$, the empty product results in the identify matrix I.

As for the discrete-time version, the limiting distribution is characterized by a level-dependent sequence of matrices $\{R_i : i \geq 0\}$ which, in general, can only be determined numerically. The elements of these matrices can be interpreted as follows. The (j, k)th element of R_i, denoted by $[R_i]_{j,k}$, is the expected sojourn time in state $(i+1, k)$ per unit sojourn in state (i, j), given that the process started in state (i, j). It is well known (cf. Bright and Taylor [13]) that the sequence $\{R_i\}$ is the minimal non-negative solution of the set of equations

$$A_0^{(i)} + R_i A_1^{(i+1)} + R_i \left(R_{i+1} A_2^{(i+2)} \right) = 0, \quad i \geq 0. \tag{12}$$

If the number of levels and/or phases is infinite, computing the limiting distribution has two major complications. First, one must truncate the infinite series of equation (10) in an appropriate manner, and second, the matrices $\{R_i : i \geq 0\}$ must be computed efficiently. The next section provides a brief overview of some of the algorithmic approaches to computing the limiting distribution of the discrete- and continuous-time LDQBD processes.

Numerical Algorithms

Numerical approaches for obtaining the limiting distribution of level-independent QBD processes have been studied fairly extensively in the literature. For both discrete- and continuous-time cases, Neuts [37] provides some methods for determining positive recurrence and computing the limiting distribution. Hajek [22] considered the case of homogeneous QBD processes with a finite number of levels and showed that the limiting distribution vector can be described by the sum of two matrix-geometric terms. For general, level-independent QBD processes, most noteworthy is the logarithmic reduction algorithm developed in 1993 by Latouche and Ramaswami [28] which guarantees quadratic convergence of the solution for the matrix R in level-independent QBD processes. A method for computing the stationary distribution of finite level-independent QBD processes was contributed by Elhasfi and Molle [19]. Some software tools that implement the most popular algorithms have been contributed by various researchers (see, for example, references [11, 23, 42, 44]).

For the level-dependent case, a few approaches are noteworthy. When the number of levels and the number of phases are both finite, Gaver et al. [20] provide stable algorithms for computing the limiting distribution and moments of first passage times in the continuous-time case. These algorithms are similar to the linear iteration schemes developed by Latouche and Ramaswami [29] who show how to use logarithmic reduction to compute the moments of passage times for level-independent processes. Ye and Li [45] developed the "folding algorithm" to compute the limiting distribution of a finite continuous-time LDQBD process. Their efficient algorithm directly solves $\boldsymbol{\pi} Q = \mathbf{0}$ by employing state-space reduction techniques that reduce the complexity induced by level dependence. Li and Cao [31] developed two types of RG-factorizations in order to analyze LDQBD processes with either finitely- or infinitely-many levels. They use their factorizations to derive the Laplace transforms of the conditional distributions of integral functionals of QBD processes.

When the number of levels and/or phases is infinite, the most widely accept approach for computing the limiting distribution is due to Bright and Taylor [13, 14] who provide an extension of the logarithmic reduction algorithm of Latouche and Ramaswami [28] for the level-dependent case. Define the continuous-time LDQBD process $\{Z(t) : t \geq 0\} \equiv \{(X(t), J(t)) : t \geq 0\}$ on the state space $\{(i, j) : i \geq 0,\ 1 \leq j \leq M_i\}$ where M_i is the number of phases in level i which can be infinite. Bright and Taylor [13] suggest truncating the infinite series of equation (10) at some level K and then re-normalizing to compute an approximate subvector of the form

$$\boldsymbol{\pi}_i(K) = \boldsymbol{\pi}_0(K) \prod_{n=0}^{i-1} R_n, \quad i \geq 1,$$

such that $\boldsymbol{\pi}_0(K)$ satisfies equation (9) and the normalization condition,

$$\boldsymbol{\pi}_0(K) \sum_{i=0}^{K} \left[\prod_{n=0}^{i-1} R_n \right] \boldsymbol{e} = 1.$$

The subvectors, $\{\boldsymbol{\pi}_i(K) : i \geq 0\}$, represent an invariant measure for the limiting distribution of all the states at or below level K. Therefore, for any $K \geq 0$, $\boldsymbol{\pi}_i(K)$ is an upper bound for $\boldsymbol{\pi}_i$ (componentwise), and $\boldsymbol{\pi}_i(K) \to \boldsymbol{\pi}_i$ as $K \to \infty$.

In order to compute $\boldsymbol{\pi}_i(K)$ for a chosen truncation level K, Bright and Taylor [13] examine the discrete-time Markov chain embedded at the jump epochs of $\{Z(t) : t \geq 0\}$ to compute the family of matrices $\{R_i : i \geq 0\}$ using a recursive scheme. The following result (Lemma 1 of [13]) is a continuous-time extension of a discrete-time result stated by Ramaswami and Taylor [40].

Lemma 1 *If $\{Z(t) : t \geq 0\}$ is positive recurrent, then for $i \geq 0$ the matrix R_i is given by*

$$R_i = \sum_{\ell=0}^{\infty} U_i^{\ell} \prod_{n=0}^{\ell-1} D_{k+2^{\ell-1}}^{\ell-1-n}, \quad i \geq 0 \tag{13}$$

where, for $i \geq 1$, U_i^{ℓ} and D_i^{ℓ} are $M_{i-1} \times M_{i-1+2^{\ell}}$ and $M_{i-1} \times M_{i-1-2^{\ell}}$ matrices, respectively, that are recursively defined by

$$\begin{aligned}
U_i^0 &= A_0^{(i)} \left(-A_1^{(i+1)} \right)^{-1}, \\
D_i^0 &= A_2^{(i)} \left(-A_2^{(i-1)} \right)^{-1}, \\
U_i^{\ell+1} &= U_i^{\ell} U_{i+2^{\ell}}^{\ell} \left[I - U_{i+2^{\ell+1}}^{\ell} D_{i+3\cdot 2^{\ell}}^{\ell} - D_{i+2^{\ell+1}}^{\ell} U_{i+2^{\ell}}^{\ell} \right]^{-1}, \\
D_i^{\ell+1} &= D_i^{\ell} D_{i-2^{\ell}}^{\ell} \left[I - U_{i-2^{\ell+1}}^{\ell} D_{i-2^{\ell}}^{\ell} - D_{i-2^{\ell+1}}^{\ell} U_{i-3\cdot 2^{\ell}}^{\ell} \right]^{-1}.
\end{aligned}$$

The infinite series of equation (13) can be truncated using a simple scheme (see Algorithms 2 and 3 of [13]). By rearranging the terms in equation (12), the family of matrices, $\{R_i : i \geq 0\}$, can be recursively computed by

$$R_i = A_0^{(i)} \left(-A_1^{(i+1)} - R_{i+1}\, A_2^{(i+2)} \right)^{-1} \tag{14}$$

(assuming the inverse exists) so that (13) need not be computed repeatedly. Algorithm 1 of [13] can be used to compute $\boldsymbol{\pi}_i(K)$ for a chosen value K. This algorithm is summarized in what follows.

Computing $\boldsymbol{\pi}(K)$ for Given K:

Step 1: Calculate R_{K-1} using equation (13);

Step 2: Recursively calculate $R_{K-2}, R_{K-3}, \dots, R_0$ using equation (14);

Step 3: Solve the system of equations

$$\boldsymbol{\pi}_0(K)\left[A_1^{(0)} + R_0 A_2^{(1)}\right] = \mathbf{0}, \quad \boldsymbol{\pi}_0(K)\boldsymbol{e} = 1;$$

Step 4: For $i = 1$ to K, set

$$\boldsymbol{\pi}_i(K) = \boldsymbol{\pi}_{i-1}(K) R_{i-1},$$

and normalize $\boldsymbol{\pi}_n(K)$, $n = 0, 1, \dots, i$, so that

$$\sum_{n=0}^{i} \boldsymbol{\pi}_n(K)\boldsymbol{e} = 1.$$

Algorithms 2 and 3 of [13] provide the means by which to truncate the infinite series of (13) to effectively compute R_{K-1} in Step 1.

It is imperative to select an integer K that is large enough to ensure that the limiting probability of being in a state at or above level K is close to zero. Algorithm 4 of [13] describes a general method for choosing a truncation point that works well in practice. We next review the main steps of this algorithm.

Selecting the Integer K:

Step 1: Set $K_0 = 0$;

Step 2: Do the following:

(a) Choose K_1 such that $K_1 > K_0$;

(b) Calculate five terms of expression (13) for R_{K_1-1};

If the five terms are sufficient, go to Step 2(c),

else repeat Steps 2(a) and 2(b).

(c) Recursively calculate $R_{K_1-2}, R_{K_1-3}, \dots, R_{K_0}$ using equation (14).

(d) Simultaneously solve

$$\boldsymbol{\pi}_0(K_1)\left[A_1^{(0)} + R_0 A_2^{(1)}\right] = \mathbf{0}, \quad \boldsymbol{\pi}_0(K_1)\boldsymbol{e} = 1;$$

(e) For $i = K_0 + 1$ to K_1, do

$$\boldsymbol{\pi}_i(K_1) = \boldsymbol{\pi}_{i-1}(K_1)R_{i-1},$$

and normalize $\boldsymbol{\pi}_n(K)$, $n = 0, 1, \ldots, i$, such that

$$\sum_{n=0}^{i} \boldsymbol{\pi}_n(K_1)\boldsymbol{e} = 1.$$

(f) Set $K_0 = K_1$

until $\boldsymbol{\pi}_{K_1}(K_1)\boldsymbol{e} < \epsilon$.

Step 3: Set $K = K_1$.

Besides this general procedure for determining K and the subvectors $\{\boldsymbol{\pi}_i(K) : i \geq 0\}$, Bright and Taylor [13] also provide algorithms for some special cases including LDQBD processes with a large number of boundary states (Algorithm 5) and those which are stochastically dominated by another LDQBD process (Algorithms 6 and 7). The former is a type of LDQBD process that exhibits level-dependent behavior up to some fixed level $\bar{K}$, and level-independent behavior for each level i such that $i > \bar{K}$ (see Neuts [37]).

Further Reading and Applications

For extensive coverage of the level-independent QBD process in discrete and continuous time, it is best to first consult the two major texts due to Neuts [37, 38] and one due to Latouche and Ramaswami [30]. For a cogent summary of the level-independent case, please see the article *Level-Independent Quasi-Birth-and-Death processes* which provides the main results for limiting distributions and conditions for positive recurrence, as well as a more detailed discussion of algorithmic procedures for computing the limiting distribution. The best reference for understanding the rudimentary aspects of the level-dependent QBD process is Chapter 12 of the excellent text by Latouche and Ramaswami [30] which considers (primarily) the discrete-time version of the model. As noted above, the two papers by Bright and Taylor [13, 14] provide the details of an efficient algorithm for computing the limiting distribution of the infinite state space, continuous-time model.

Over the years, several extensions have been developed for both the level-independent and level-dependent QBD processes. One of the most important and interesting extensions of level-independent QBD processes is due to Ramaswami [41] who showed connections to the analysis of stochastic fluid queueing models. A stochastic fluid queueing model can be viewed as a QBD process wherein the level assumes values on the non-negative real line and the phase variable is discrete. The fluid level evolves in a piecewise linear manner depending on the current phase. Subsequent contributions to fluid queues include Ahn and Ramaswami [1] and Bean et al. [9, 10]. Of particular interest is the recent paper by da Silva Soares and Latouche [17] who analyze level-dependent fluid models within the QBD paradigm. Bean et al. [8] provide a complete quasistationary analysis of the discrete-time LDQBD process by assuming that level 0 behaves as an absorbing state. Choi et al. [16] provide a generalization of finite LDQBD processes that they call "nested QBD chains" which entail the addition of a third dimension, the period. They provide the fundamental matrices for these types of chains and prove other interesting properties as well.

The level-dependent QBD process finds wide applicability as a model for systems with complex dynamics. It is most prevalent in the modeling of queueing systems with Markovian arrival process (MAP) input, or for those with phase-type distributions. For example, the LDQBD process is ideal for modeling the $PH/M/\infty$ queue. Some other interesting applications in queueing theory include models described in references [2, 6, 21, 24, 33]. The emergence of the service sector in many economies has sparked a renewed interest in retrial queueing models. These models can help assess the impact of retrial customers who are initially denied access to service but who retry the system after some random amount of time. Such models are particularly important in large-scale service systems, such as large customer contact centers. Consequently, many new single- and multi-server retrial queueing models have emerged over the past five years. The infinitesimal generator matrix of retrial queues is usually of the LDQBD-type because the retrial rate varies with the level (which corresponds to the number of customers in orbit). Some notable LDQBD retrial queueing models include [3, 4, 5, 7, 12, 27, 32, 34, 43]. In addition to queueing models, LDQBD processes are now emerging in reliability modeling and analysis and inventory systems modeling. Some interesting reliability applications are due to Pérez-Ocón and Montoro-Cazorla [15, 39]. Recent inventory models that exploit the LDQBD structure include [18, 35]. The homogeneous QBD process has found wide applicability in the modeling and analysis of computer and communications systems, and Bright and Taylor [13] provide numerous examples of such. Recently, LDQBD processes have been applied as models of real-time wireless code division multiple access (WCDMA) communications systems to estimate performance parameters such as blocking probability and expected delay (see for example references [25, 26]).

References

[1] S. Ahn and V. Ramaswami. Efficient algorithms for transient analysis of stochastic fluid flow models. *J. Appl. Probab.*, 42:531–549, 2005.

[2] A.S. Alfa. Vacation models in discrete time. *Queueing Systems*, 44:5–30, 2003.

[3] A.S. Alfa. Discrete-time analysis of $GI/G/1$ system with Bernoulli retrials: An algorithmic approach. *Annals of Operations Research*,141:51–66, 2006.

[4] V.V. Anisimov and J.R. Artalejo. Approximation of multiserver retrial queues by means of generalized truncated models. *Sociedad de Estadística e Investigacíon Operativa Top*, 10(1):51–66, 2002.

[5] J.R. Artalejo and S.R. Chakravarthy. Algorithmic analysis of the $MAP/PH/1$ retrial queue. *Sociedad de Estadística e Investigacíon Operativa Top*, 14(2):293–332, 2006.

[6] S. Asmussen and M. Pihlsgård. Transient properties of many-server queues and related QBDs. *Queueing Systems*, 46:249–270, 2004.

[7] F. Avram and A. Gómez-Corral. On bulk-service $MAP/PH^{L,N}/1/N$ G-queues with repeated attempts. *Annals of Operations Research*,141:109–137, 2006.

[8] N.G. Bean, P.K. Pollett, and P.G. Taylor. Quasistationary distributions for level-dependent quasi-birth-and-death processes. *Communications in Statistics – Stochastic Models*, 16(5):511–541, 2000.

[9] N.G. Bean, M.M. O'Reilly, and P.G. Taylor. Hitting probabilities and hitting times for stochastic fluid flows. *Stochastic Processes and Their Applications*, 115:1530–1556, 2005.

[10] N.G. Bean and M. M. O'Reilly. Performance measures of a multi-layer Markovian fluid model. *Annals of Operations Research*, 160:99–120, 2008.

[11] D.A. Bini, B. Meini, S. Steffé, and B. V. Houdt. Structured Markov chains solver: software tools. In *SMCtools '06: Proceeding from the 2006 workshop on Tools for solving structured Markov chains*, page 14, New York, NY, USA, 2006. ACM.

[12] L. Breuer, A. Dudin, and V. Klimenok. A retrial $BMAP/PH/N$ system. *Queueing Systems*, 40:433–457, 2002.

[13] L.W. Bright and P.G. Taylor. Calculating the equilibrium distribution in level dependent quasi-birth-and-death processes. *Communications in Statistics – Stochastic Models*, 11:497–525, 1995.

[14] L.W. Bright and P.G. Taylor. (1997). "Equilibrium Distributions for Level-Dependent Quasi-Birth-and-Death Processes." In S.R. Chakravarthy and A.S. Alfa (eds.), Matrix Analytic Methods in Stochastic Models: Proc. of the 1st Int. Conf., Marcel Dekker, NJ, pp. 359-375.

[15] D.M. Cazorla and R. Pérez-Ocón. An LDQBD process under degradation, inspection, and two types of repair. *European Journal of Operational Research*, 190:494–508, 2008.

[16] S.H. Choi, B. Kim, K. Sohraby, and B.D. Choi. On matrix-geometric solutions of nested QBD chains. *Queueing Systems*, 43:5–28, 2003.

[17] A. da Silva Soares and G. Latouche. Fluid queues with level dependent evolution. *European Journal of Operational Research*, 196:1041–1048, 2009.

[18] T.G Deepak, A. Krishnamoorthy, V.C. Narayanan, K. Vineetha. Inventory with service time and transfer of customers. *Annals of Operations Research*,160:191–213, 2008.

[19] E.H. Elhafsi and M. Molle. On the solution to QBD processes with finite state space. *Stochastic Analysis and Applications*, 25:763–779, 2007.

[20] D.P. Gaver, P.A. Jacobs, G. Latouche. Finite birth-and-death models in randomly changing environments. *Advances in Applied Probability*, 16:715–731, 1984.

[21] A. Gómez-Corral. On a finite-buffer bulk-service queue with disasters. *Mathematical Methods of Operations Research*, 61:57-84, 2005.

[22] B. Hajek. Birth and death processes on the integers with phases and general boundaries. *Journal of Applied Probability*, 19:488–499, 1982.

[23] B. Haverkort, A. Van Moorsel, and A. Dijkstra. MGMtool: A performance analysis tool based on matrix geometric methods. In R. Pooley and J. Hillston, editors, *Modelling Techniques and Tools*, pages 312–316. Edinburgh University Press, 1993.

[24] Q. He and M.F. Neuts. Two $M/M/1$ queues with transfers of customers. *Queueing Systems*, 42:377–400, 2002.

[25] N. Hedge and E. Altman. Capacity of multiservice WCDMA networks with variable GoS. *Wireless Networks*, 12:241–253, 2006.

[26] I. Koukoutsidis, E. Altman, and J.M. Kelif. A non-homogeneous QBD approach for the admission and GoS control in a multiservice WCDMA system. Proc. IWQoS 2005, LNCS 3552: 327–340.

[27] B.K. Kumar, R. Rukmani, and V. Thangaraj. An $M/M/c$ retrial queueing system with Bernoulli vacations. *J. Syst Sci Syst Eng*, 18(2):222–242, 2009.

[28] G. Latouche and V. Ramaswami. A logarithmic reduction algorithm for quasi-birth-and-death processes. *Journal of Applied Probabiliy*, 30:650–674, 1993.

[29] G. Latouche and V. Ramaswami. Expected passage times in homogeneous quasi-birth-and-death processes. *Communications in Statistics: Stochastic Models*, 11:103–122, 1995.

[30] G. Latouche and V. Ramaswami. *Introduction to Matrix Analytic Methods in Stochastic Modeling*. ASA-SIAM Series on Statistics and Applied Probability. SIAM, Philadelphia PA, 1999.

[31] Q. Li and J. Cao. Two types of RG-factorizations of quasi-birth-and-death processes and their applications to stochastic integral functionals. *Stochastic Models*, 20(3):299–340, 2004.

[32] Q. Li, Y. Ying, and Y.Q. Zhao. A $BMAP/G/1$ retrial queue with server subject to breakdowns and repairs. *Annals of Operations Research*, 141:233–270, 2006.

[33] Z. Lian, X. Liu, and L. Liu. Discriminatory processor sharing queues and the DREB method. *Stochastic Models*, 24:19–40, 2008.

[34] V.V. Mushko, M.J. Jacob, K.O. Ramakrishnan, A. Krishnanmoorthy, and A.N. Dudin. Multiserver queue with addressed retrials. *Annals of Operations Research*, 141:283–301, 2006.

[35] V.C. Narayanan, T.G. Deepak, A. Krishnamoorthy, B. Krishnakumar. On an (s, S) inventory policy with service time, vacation to server and correlated lead time. *Quality Technology & Quantitative Management*, 5(2):129–143, 2008.

[36] M.F. Neuts. The M/M/1 queue with randomly varying arrival and service rates. *Opsearch*, 15:139–157, 1978.

[37] M.F. Neuts. *Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach.* The Johns Hopkins University Press, Baltimore, MD, 1981.

[38] M.F. Neuts. *Structured Stochastic Matrices of M/G/1 Type and Their Applications.* Marcel Dekker, New York, 1989.

[39] R. Pérez-Ocón and D. Montoro-Cazorla A multiple warm standby system with operational and repair times following phaste-type distributions. *European Journal of Operational Research*, 169:178–188, 2006.

[40] V. Ramaswami and P.G. Taylor. Some properties of the rate operators in level dependent quasi-birth-and-death processes with countable number of phases. *Communications in Statistics – Stochastic Models*, 12:143–164, 1996.

[41] V. Ramaswami. Matrix analytic methods for stochastic fluid flows. In D. Smith and P. Hey, editors, *Teletraffic Engineering in a Competitive World (Proceedings of the 16th International Teletraffic Congress)*, pages 1019–1030. Elsevier Science B.V., Edinburgh, UK, 1999.

[42] A. Riska and E. Smirni. MAMSolver: A matrix-analytic methods tools. In T. Field, P. G. Harrison, J. Bradley, and U. Harder, editors, *Modelling Techniques and Tools. 12th International Conference, TOOLS 2002*, volume 2324 of *Lecture Notes in Computer Science*, pages 205–211, Berlin, 2002. Springer-Verlag.

[43] Y.W. Shin. Multi-server retrial queue with negative customers and disasters. *Queueing Systems*, 55:223–237, 2007.

[44] M.S. Squillante. MAGIC: A computer performance modeling tool based on matrix-geometric techniques. In G. Balbo and G. Serazzi, editors, *Computer Performance Evaluation: Modeling Techniques and Tools*, pages 411–425. North–Holland, Amsterdam, 1992.

[45] J. Ye and S. Li. Folding algorithm: A computational method for finite QBD processes with level-dependent transitions. *IEEE Transactions on Communications*, 42(2-4):625–639, 1994.