Bilateral Phase Type Distributions

Soohan Ahn
Department of Statistics, The University of Seoul
Jeonnong-dong, 90 Dong daeman-gu
Seoul 130-743, South Korea

V. Ramaswami
AT&T Labs-Research, 180 Park Avenue
Florham Park, NJ 07932, USA

Abstract

A new class of probability distributions called "bilateral phase type distributions (BPH)" on $(-\infty, \infty)$ is defined as a generalization of the versatile class of phase type (PH) distributions on $[0, \infty)$ introduced by Marcel F. Neuts. We derive the basic descriptors of such distributions in an algorithmically tractable manner and show that this class has many interesting closure properties and is dense in the class of all distributions on the real line. Based on the established versatility and tractability of phase type distributions, we believe that this class has high potential for general use in statistics, particularly to cover non-normal distributions, and also that its inherent connection to Markov chains may make it suitable for inference based on hidden Markov chain methods and MCMC type approaches.

1 Introduction

We begin with a quick review of phase type distributions on $[0, \infty)$ and an overview of this work which extends them to bilateral phase type distributions on $(-\infty, \infty)$.

1.1 Phase Type Distributions

M.F. Neuts [12] defined the interesting class of phase type (PH) distributions on $[0, \infty)$ as a generalization of the exponential distribution and many others derived from exponential distributions through convolutions and mixtures.

Such distributions are obtained as distributions of absorption times in finite state Markov chains with one absorbing state. More formally, we have the following definition in the continuous time case.

Definition 1 *Consider a continuous time Markov chain (CTMC) on the state space* $S = \{1, \dots, m, a\}$ *with infinitesimal generator*

$$Q = \begin{pmatrix} T & t^a \\ 0_{1m} & 0 \end{pmatrix} \tag{1}$$

where T *is an* $m \times m$ *matrix,* t^a *is a* $m \times 1$ *vector, and* $\boldsymbol{0}_{jk}$ *is a zero matrix of order* $j \times k$. *Let the initial probability vector of the CTMC be given by* $(\alpha, 1 - \alpha\mathbf{1})$ *where* $\alpha \geq 0$, $\alpha\boldsymbol{1} \leq 1$, *and* $\boldsymbol{1}$ *is a column vector of 1's of appropriate order. Then the distribution of the time until absorption in state* a *has the cdf* $F(\cdot)$ *and pdf* $f(\cdot)$ *given by*

$$F(x) = 1 - \alpha e^{Tx}\boldsymbol{1}, \quad f(x) = \alpha e^{Tx} t^a, \quad x > 0, \tag{2}$$

and is called the phase type (PH)-distribution with representation (α, T) *and is denoted by* $PH(\alpha, T)$.

A similar definition may be given in the discrete case yielding PH-distributions on the nonnegative integers. For a detailed discussion of this class of distributions and point processes defined using such distributions, we refer to Neuts [12], [13], [14], [15] and Latouche & Ramaswami [10].

PH-distributions have received much attention in the applied probability literature related to queues, dams, insurance risk, reliability, etc., and the reasons for that have been many:

- Denseness: The class of PH-distributions can be shown to be dense (in the Prohorov metric of weak convergence) in the set of all probability distributions on $[0, \infty)$. Although fitting phase type distributions and approximating other distributions with phase type distributions continue to be active areas of research and many problems do remain open, some success has been reported recently; see e.g., [5], [20].

- Closure: The class of PH-distributions is closed under finite convolutions and mixtures and under Boolean operations of taking the max or min of (independent) PH-random variables. In addition, many derived distributions in models involving phase type distributions are themselves phase type. For example, the stationary waiting time distributions in queues with renewal input and phase type service time distributions are known [4], [17] to be of phase type. Also, the stationary distribution of the fluid level in an ergodic stochastic fluid flow modulated by a finite state CTMC (used much in high speed network performance modeling, e.g.) is of phase type [6], [18], [1].

- Tractability: A very attractive feature of phase type distributions is their computational tractability. Due to the connection with an underlying Markov chain, in models involving phase type distributions, conditioning arguments become easier through the inclusion of the state of the Markov chain as an auxiliary variable. Indeed, much of matrix-analytic methods for queueing theory was developed in the context of such models; see [14], [15], [19], [10].

- Point Process Models: Using the phase type distribution's construction as a basis, one may define many tractable Markovian point processes, and these provide a versatile class of models for practical use and are heavily used due to their computational tractability; see [13], [15], [16], [11]. Generalizations of these to spatial point pattern models have also been made in the literature, see e.g., [9].

1.2 Extension to the Real Line

Our goal in this paper is to generalize the class of phase type distributions to a class we shall call "bilateral phase type distributions (BPH)" on the entire line $(-\infty, \infty)$ in a manner that retains the essential simplicity and versatility of phase type distributions. We achieve this through the consideration of a Markov reward (fluid flow) model defined on an absorbing continuous time Markov chain. The details of the construction are given in Section 2 in which we also present some properties of this class that can be derived easily. In Section 3, we briefly recall some known results on stochastic fluid flow models and use them in Section 4 to derive a transform free formula for the density of the bilateral phase type distribution. This is used in Section 5 to show that if X is BPH, then $X^+ = max(0, X)$ and $X^- = -min(0, X)$ are both phase type, a result that implies that BPH distributions are dense in the class of all probability distributions on the real line. Then in Section 6, we recall an algorithm that provides a fast method to compute one of the key matrices needed in the computation of the BPH distribution. In Section 7, we present some interesting examples of BPH densities computed using that algorithm and conclude in Section 8 with a few remarks.

There have been attempts in the past to generalize phase type distributions in several directions. Important among these are the generalizations to the class of matrix exponential distributions by Asmussen and Bladt [7] in which they demonstrate that distributions on $[0, \infty)$ with rational Laplace transforms can be handled by a matrix formalism similar to that of phase type distributions; see also [8]. In that formalism, the connection to Markov chains is lost as the matrices involved may not satisfy the conditions for being an infinitesimal generator of a Markov chain. Another generalization is that of J.G. Shanthikumar [21] who defined a bilateral phase type random variable as one whose positive and negative parts can be represented as a (possibly infinite) mixture of sums of *iid* exponentially distributed random variables. Our class of distributions can be shown to be a subset of those defined by Shanthikumar, but our representation is much more parsimonious and involves only finite state Markov chains; and yet, they form a dense set in the class of all distributions on the real line just as does the class introduced by Shanthikumar.

2 Bilateral Phase-type Distribution

In this section, we define the bilateral phase type distribution and obtain some of its interesting properties that are easily determined. A transform free formula for the density will be given in Section 4 after establishing some connections with stochastic fluid flow models.

2.1 Definition

On a probability space $(\Omega, \mathcal{A}, \mathcal{P})$, define $\mathcal{J} = \{J(t), t \geq 0\}$ to be an absorbing Markov process with state space S with $|S| = m+1$ and infinitesimal generator Q with the structure

$$Q = \begin{pmatrix} T & t^a \\ \mathbf{0} & 0 \end{pmatrix}, \tag{3}$$

where T is a $m \times m$ nonsingular matrix and t^a is a $m \times 1$ vector. The state of this CTMC at time u will be referred to as "the phase at time u." By the properties governing an infinitesimal generator, note that T has negative diagonal and nonnegative off-diagonal elements, and furthermore that $T\mathbf{1} + t^a = \mathbf{0}$.

We assume that S is partitioned as $S = S_1 \cup S_2 \cup S_3$, where $S_3 = \{a\}$ and a denotes the absorbing state. The set of transient states is $S_t = S_1 \cup S_2$. We partition T and t^a according to the sets S_1 and S_2 and write

$$T = \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix}, \quad t^a = \begin{pmatrix} t_1^a \\ t_2^a \end{pmatrix}. \tag{4}$$

Throughout, the (j,k)-th element of a matrix A will be denoted by $A(j,k)$ or by $[A]_{jk}$, whereas the symbol A_{jk} will always denote the submatrix of A formed by the elements $A(r,s)$ with $r \in S_j$ and $s \in S_k$. For later use, we also define the diagonal matrices

$$T^a = diag(t^a(j), j \in S_t);\ T^a_k = diag(t^a(j),\ j \in S_k),\ k = 1, 2. \tag{5}$$

On the same probability space, define a Markov modulated reward (fluid flow) process $\mathcal{F} = \{F(t) : t \geq 0\}$ with $F(0) = 0$ such that during each sojourn of $\mathcal{J}$ in state $j \in S_1$, the total accumulated net reward $F(\cdot)$ increases at rate $c_j > 0$ and during each sojourn of $\mathcal{J}$ in state $k \in S_2$, the total accumulated net reward decreases at rate $c_k > 0$. Assume further that once the CTMC reaches the state a, no further changes occur to the fluid level. We shall denote this model by $\mathcal{F}(T, C^*)$, where the diagonal matrices $C_j, j = 1, 2$, and C^* are such that

$$C_1 = diag(c_k, k \in S_1),\ C_2 = diag(c_k, k \in S_2),\ \text{and } C^* = diag(C_1, -C_2). \tag{6}$$

For later use, we also define the $(m+1) \times (m+1)$ diagonal matrix

$$C = diag(C_1, C_2, 1). \tag{7}$$

We denote the initial probability vector of the Markov chain by (α, α_a), where the m-vector α is such that for $k \in S_t$, $\alpha(k)$ is the probability that the CTMC starts in state $k \in S_t$. Also, we denote the absorbing time of the underlying Markov process $\mathcal{J}$ by τ.

Now, we will define a bilateral phase-type random variable.

Definition 2 *Let X denote the total accumulated reward until absorption, that is, $X = F(\tau)$. X is called a bilateral phase-type random variable or a bilateral phase-type distributed random variable with representation (α, T, C^*), and we denote this fact by*

$$X \sim BPH(\alpha, T, C^*).$$

It is clear from the construction that the BPH distribution has an atom at 0 iff $\alpha_a = 1 - \alpha\mathbf{1} > 0$; indeed, $\alpha_a = P[X = 0]$. Also, since T is invertible and therefore absorption occurs in a finite amount of time w.p. 1, we have that X is finite a.s. The assumptions about T also entail that there exists a constant $d < 0$ such that d is an eigenvalue of T, and furthermore that all other eigenvalues of T have negative real parts less than or equal to d.

2.2 Characteristic Function

Define the diagonal matrix

$$T_d = diag\{-T(1,1), \cdots, -T(m,m)\}.$$

Let $[f(x)]_k$ denote the probability density function of X given $J(0) = k \in S_t$, and let $[\phi(\cdot)]_k$ denote the associated characteristic function. That is,

$$[\phi(u)]_k = \mathcal{E}_k[e^{iuF(\tau)}] = \mathcal{E}_k[e^{iuX}], \quad -\infty < u < \infty, \tag{8}$$

where $\mathcal{E}_k$ denotes conditional expectation given $J(0) = k$, and $i = \sqrt{-1}$. Then for $k \in S_t$, $[\phi(u)]_k$ satisfies the equation

$$\begin{aligned}[\phi(u)]_k &= \int_0^\infty e^{iuc_k^* v} e^{T(k,k)v} t^a(k)\, dv \\ &\quad + \int_0^\infty e^{iuc_k^* v} e^{T(k,k)v}\, dv \sum_{j \neq k,a} T(k,j)[\phi(u)]_j,\end{aligned}$$

where $c_k^* = c_k$ for $k \in S_1$, and $c_k^* = -c_k$ for $k \in S_2$; this equation is obtained by conditioning on the first transition epoch v of the underlying CTMC. This may be written in matrix form as

$$\phi(u) = -(iuC^* - T_d)^{-1} t^a - (iuC^* - T_d)^{-1}(T + T_d)\phi(u), \tag{9}$$

where $\phi(u)$ is the m-vector of elements $[\phi(u)]_k$, $k \in S_t$. From this, we can get the following theorem.

Theorem 1 *The vector $\phi(\cdot)$ of characteristic functions is given by*

$$\phi(u) = -(iuC^* + T)^{-1} t^a, \quad -\infty < u < \infty. \tag{10}$$

Proof: The result follows immediately from equation (9) by noting that the inverse in equation (10) exists due to the fact that all eigenvalues of T have negative real parts. □

Theorem 1 immediately yields the characteristic function of the BPH r.v. X. We state the result as a theorem.

Theorem 2 *If we denote the characteristic function of the BPH r.v. $X \sim BPH(\alpha, T, C^*)$ by $\xi(u)$, then*

$$\xi(u) = \alpha_a - \alpha\,(iuC^* + T)^{-1} t^a, \quad -\infty < u < \infty. \tag{11}$$

Moment formulae for a BPH distribution can be derived from the above by routine differentiation at $u = 0$. The general formula for the k-th moment of X is given by

$$E[X^k] = k!\ \alpha(-T^{-1}C^*)^k\, \mathbf{1};$$

we omit the details.

2.3 Properties of BPH distributions

BPH distributions share many of the properties of PH distributions. Though easy and straightforward, we list them below to highlight the importance and versatility of the class BPH. Of these, we note that closure under convolutions and mixtures is proven in a manner very similar to those for PH-distributions using probabilistic constructions; see Chapter 2 in [10] for proofs in the case of PH-distributions which can be adapted easily to the present case.

- PH distributions on $[0, \infty)$ form a subset of BPH distributions and correspond to the special case where S_2 is empty and $c_j \equiv 1$ for all $j \in S_1$.
- If $X \sim BPH(\alpha, T, C^*)$, then $kX \sim BPH(\alpha, T, kC^*)$ for all real $k \neq 0$.
- The sum of a finite number of independent BPH random variables is itself BPH.
- A linear combination of a finite number of independent BPH random variables is itself a BPH random variable.
- A mixture of a finite number of BPH distributions is itself BPH.

In the definition of the BPH distribution, we may without loss of generality assume that $c_j \equiv 1$ for all j. This follows from a direct verification using characteristic functions, e.g., that

$$BPH(\alpha, T, C^*) = BPH(\alpha, C^{-1}T, I^*)$$

where I^* is a diagonal matrix with $I^*(j, j) = 1$ for $j \in S_1$ and $I^*(j, j) = -1$ for $j \in S_2$.

Note that in defining the BPH distribution, we could have allowed also for the presence of a set S_4 of (transient) states in the CTMC in which no accrual of reward occurs (i.e., $c_k^* = 0$ for $k \in S_4$). However, such a model is not parsimonious since the CTMC obtained from it by gluing together the paths restricted to the set $S_1 \cup S_2 \cup \{a\}$ would yield the same distribution for the total reward up to the absorption epoch, but with a smaller set of transient phases $S_1 \cup S_2$. This observation can help to reduce the dimensionality of the representation in some cases. However, just as with phase type distributions, the minimal representation problem – namely, that of obtaining a representation with the smallest dimension for T – remains open.

In Section 5, we shall derive transform free formulae for the density of the BPH distribution in terms of phase type densities. A consequence of that derivation will be a demonstration of the following important result; a proof is deferred to that section.

Theorem 3 *Let X be BPH. Then $X^+ = max(0, X)$ and $X^- = -min(0, X)$ are both phase type random variables.*

We also have the following theorem.

Theorem 4 *The set of BPH distributions is dense (in the weak convergence metric) in the class of all distributions on $(-\infty, \infty)$.*

Proof: It is easy to see that given any two PH distributions $PH(\alpha_1, T_1)$ and $PH(\alpha_2, T_2)$ without atoms at 0 and constants $0 \leq p, q \leq 1$ with $p + q \leq 1$, it is possible to construct a BPH random variable such that $P(X > 0) = p$, $P(X < 0) = q$ and $P(X = 0) = 1-p-q$, and the given PH distributions are the respective conditional distributions of $|X|$, given $X > 0$ and given $X < 0$. Indeed, the BPH distribution $BPH((p\alpha_1, q\alpha_2), diag(T_1, T_2), diag(I, -I))$ has this property. The asserted result is now trivial from this and the fact that in the weak convergence metric, PH distributions are dense in the class of all distributions on $[0, \infty)$.

3 Fluid flow models

Consider the distribution $BPH(\alpha, T, C^*)$ defined in Section 2. The evolution of the associated reward process over time is immediately seen to be the same as that of an unrestricted fluid process on the state space $S_1 \cup S_2 \cup S_3$, where $S_3 = \{a\}$ is a set after reaching which no further changes occur either to the fluid level or to the environmental Markov chain. This allows us to draw from the work on stochastic fluid flow models considered in [18], [1], [2]. In particular, we note that the BPH random variable X is indeed the fluid level $F(\infty)$ in this corresponding fluid flow model, whence the BPH distribution is also the steady state distribution of that fluid flow. Taking note of these, in this section we recall some important results from our prior work on fluid flow models which will be used repeatedly in what follows. For their proofs, we refer to [2].

3.1 Restricted Fluid Flow Models

Consider the CTMC $J(\cdot)$ with infinitesimal generator Q on a finite state space $S = S_1 \cup S_2 \cup S_3$ and an associated fluid flow process $F^+(t)$ starting with $F^+(0) = 0$ evolving in such a way that the fluid level increases at rate $c_j > 0$ while $J(t) = j \in S_1$, decreases at the rate $c_j > 0$ while $J(t) = j \in S_2$ and $F^+(t) > 0$, and remains constant while $J(t) \in S_3$. For later reference, we shall denote this model by the symbol $\mathcal{F}^+(Q, C^*)$, or simply as $\mathcal{F}^+$ when there is no confusion, and shall also carry over the earlier notation with respect to C and C^* and their submatrices. Note that in this model, the fluid level is restricted to nonnegative values; hence the name "restricted fluid flow." We characterized the transient behavior of such a fluid model in [2] and its stationary behavior in [1] using matrix-geometric methods proposed in [18]. Following are some key results obtained in those papers.

Busy Period

Assume that $F^+(0) = 0$ and $J(0) = j \in S_1$; we recall that in S_1, the fluid level increases linearly. Let Z denote the duration of the busy period of the restricted fluid flow model; i.e.,

$$Z = \inf\{t > 0 \,:\, F^+(t) = 0\},$$

with the standard convention that the infimum over the empty set is set to ∞. For $Re(s) > 0$, denote by $\Psi(s)$ the $|S_1| \times |S_2|$ matrix whose elements are the transforms $[\Psi(s)]_{jk}$ such that

$$[\Psi(s)]_{jk} = \int_0^\infty e^{-st} d_t \mathcal{P}_j(Z \le t, J(Z) = k), \quad j \in S_1, \ k \in S_2,$$

where $\mathcal{P}_j(A)$ denotes the conditional probability of the event A under the assumption $J(0) = j$.

Transforms $\Psi(s)$ for certain restricted fluid models will be used to determine the joint distribution of the absorption time of the CTMC and the fluid level thereat for the process defining a BPH random variable. However, the marginal distribution of a BPH random variable will require only these matrices evaluated at $s = 0$. Efficient algorithms for computing $\Psi(s)$ have been obtained in [3], and from among these a quadratically convergent scheme for the matrix $\Psi(\cdot)$ is given in Section 6 for completeness; for details of its derivation, refer to [3].

Having defined $\Psi(s)$, we introduce the following three kernels.

$$\tilde{\Psi}(s) = C_1 \Psi(s) C_2^{-1}, \tag{12}$$
$$\tilde{\Theta}(s) = [Q_{13} + \tilde{\Psi}(s) Q_{23}](sI - Q_{33})^{-1}, \tag{13}$$
$$\tilde{K}(s) = [(Q_{11} - sI) + \tilde{\Psi}(s) Q_{21} + \tilde{\Theta}(s) Q_{31}] C_1^{-1}. \tag{14}$$

In terms of the above kernels, the following result has been obtained as Theorem 13a in [2]. For the restricted fluid model $\mathcal{F}^+(Q, C^*)$, it provides the conditional joint distribution of the fluid level and phase at epochs within a busy period, given the phase at the start of the busy period.

Theorem 5 *For $x > 0$, $Re(s) > 0$, let $\tilde{v}(s,x)$ denote the transform matrix of order $\mid S_1 \mid \times \mid S \mid$, describing the behavior within a busy period, defined by the elements*

$$[\tilde{v}(s,x)]_{kj} = \int_0^\infty e^{-st} \frac{d}{dx} \left(\mathcal{P}_k[Z > t,\ 0 < F^+(t) \le x,\ J(t) = j] \right) dt, \quad k \in S_1,\ j \in S. \tag{15}$$

Then, we have the following partitioned formula:

$$\tilde{v}(s,x) = C_1^{-1} e^{\tilde{K}(s)x} [I \vdots \tilde{\Psi}(s) \vdots \tilde{\Theta}(s)]. \tag{16}$$

We introduce the following notational convention to be used in the rest of the paper. Given a transform $A(s)$, we shall denote by A the value of the transform $A(s)$ at $s = 0$; if we have to refer to the transform we shall always write either $A(s)$ or $A(\cdot)$ so that there is no confusion. This slight abuse of notation will obviate much clutter in the formulae we shall derive later.

3.2 Reflections

Let us denote by $\mathcal{F} = \mathcal{F}(Q, C^*)$ the unrestricted fluid flow model; here the fluid level is allowed to assume negative values also, increases at rate c_j for $j \in S_1$, decreases at rate c_k for $k \in S_2$ and remains constant in S_3. Associated with this fluid flow model is the flow model $\mathcal{F}_r = \mathcal{F}(Q, -C^*)$. Note that for this flow, the fluid level increases at rate c_j while $J(t) = j \in S_2$ and decreases at rate c_k while $J(t) = k \in S_1$; the fluid level once again remains constant in the set S_3. Thus, the roles of S_1 and S_2 have been reversed. Note that the paths of each of these unrestricted flow models is obtained by reflecting the paths of the other about the time axis. For this reason, the flow $\mathcal{F}_r$ will be called the reflected flow. Clearly, $\mathcal{F}$ and $\mathcal{F}_r$ are reflections of each other.

We now consider the fluid flow obtained by restricting the reflected fluid flow $\mathcal{F}_r$ to non-negative values and denote the resulting process by $\mathcal{F}_r^+$. The busy period kernel associated with this restricted flow $\mathcal{F}_r^+$ will be denoted by $\Psi_r(s)$, and $\tilde{\Psi}_r(s)$, $\tilde{\Theta}_r(s)$ and $\tilde{K}_r(s)$ will be defined in terms of it analogously as in (12)-(14) and will be subject to interpretations similar to those given in (15), (16) for the corresponding kernels of $\mathcal{F}^+$. These quantities characterize the behavior of $\mathcal{F}_r^+$ within its busy period.

We will see in the sequel that the quantities Ψ and Ψ_r play an important role.

4 Transform-free Formulae for BPH

We now return to the analysis of the BPH distribution and the related flow introduced in Section 2. First, we consider the first return times to zero for the flow of Section 2; such a return may or may not occur in a finite amount of time. However, when a return does occur, we have two types of returns to consider.

(a) If the initial phase is some $j \in S_1$, then the return to the empty fluid level can occur only with some phase $k \in S_2$, and $[\Psi(s)]_{jk},\ j \in S_1,\ k \in S_2$ gives the joint distribution of the return time and the phase at the epoch of return to level 0 given the phase at the start of the busy period; this is obvious from the definition of $\Psi(\cdot)$. The related quantity $[\Psi]_{jk} = [\Psi(0)]_{jk}$ gives the probability that a return occurs in a finite amount of time and that the phase at the epoch of return to 0 is $k \in S_2$ given that the initial phase is $j \in S_1$.

(b) If the initial phase is some $j \in S_2$, then the return to the empty fluid level can occur only with some phase $k \in S_1$. A consideration of the reflected flow yields that $[\Psi_r(s)]_{jk}$, $j \in S_2$, $k \in S_1$ yields the joint distribution of the return time and the phase at the epoch of return given the initial phase. Also $[\Psi_r]_{jk}$ is the probability that a return to level 0 occurs in a finite amount of time and occurs in the state $k \in S_1$ given that the return time started in phase $j \in S_2$.

These now lead to the following result.

Lemma 1 *Let $[w(x)]_{jk}$ denote the conditional joint density of the fluid level and the phase just prior to the absorption epoch. That is,*

$$[w(x)]_{jk} = \frac{d}{dx}\mathcal{P}_j[F(\tau-) \le x,\ J(\tau-) = k], \quad j,k \in S_1 \cup S_2.$$

With our standard convention with regard to partitioning according to the subsets of S, we have the following formulae for the submatrices $w_{ij}(\cdot)$.
(a) For $x > 0$,

$$\begin{aligned} w_{11}(x) &= [I - \Psi\Psi_r]^{-1} C_1^{-1} e^{\tilde{K}x} T_1^a, & (17)\\ w_{12}(x) &= [I - \Psi\Psi_r]^{-1} C_1^{-1} e^{\tilde{K}x} \tilde{\Psi} T_2^a, & (18)\\ w_{21}(x) &= \Psi_r [I - \Psi\Psi_r]^{-1} C_1^{-1} e^{\tilde{K}x} T_1^a, & (19)\\ w_{22}(x) &= \Psi_r [I - \Psi\Psi_r]^{-1} C_1^{-1} e^{\tilde{K}x} \tilde{\Psi} T_2^a. & (20) \end{aligned}$$

(b) For $x < 0$,

$$\begin{aligned} w_{11}(x) &= \Psi [I - \Psi_r\Psi]^{-1} C_2^{-1} e^{\tilde{K}_r|x|} \tilde{\Psi}_r T_1^a, & (21)\\ w_{12}(x) &= \Psi [I - \Psi_r\Psi]^{-1} C_2^{-1} e^{\tilde{K}_r|x|} T_2^a, & (22)\\ w_{21}(x) &= [I - \Psi_r\Psi]^{-1} C_2^{-1} e^{\tilde{K}_r|x|} \tilde{\Psi}_r T_1^a, & (23)\\ w_{22}(x) &= [I - \Psi_r\Psi]^{-1} C_2^{-1} e^{\tilde{K}_r|x|} T_2^a. & (24) \end{aligned}$$

Proof: For the fluid flow $\mathcal{F}$ starting in a state of S_1 at fluid level 0, define a busy cycle as the interval of time up to the epoch when the fluid level returns to level 0 with phase in S_1. Note that for any initial state $j \in S_1$, w.p. 1, we may have only finitely many busy cycles in the process $\mathcal{F}$; otherwise, with positive probability absorption into a may not occur, and that would contradict our assumptions. This also implies that the j-th element of the row sum $\sum_{n=1}^{\infty} (\Psi\Psi_r)^n \mathbf{1}$, which gives the expected number of busy cycles starting in $j \in S_1$, must be finite. Thus, the inverse $[I - \Psi\Psi_r]^{-1}$ exists. A similar argument for the reflected flow $\mathcal{F}_r$ will show that the inverse $[I - \Psi_r\Psi]^{-1}$ also exists.

Denote by

$$u(x) = C_1^{-1} e^{\tilde{K}x} [I \vdots \tilde{\Psi}],$$

the matrix obtained from (16) by setting $s = 0$. For $x > 0$, by multiplying to the right in (15) by $t^a(j)$, note that for $k \in S_1$, $j \in S_1 \cup S_2$, the quantity $[u(x)]_{kj} t^a(j) = [u(x) T^a]_{kj}$ yields the density of absorption occurring into a before the first return to fluid level 0 and from the state $j \in S_1 \cup S_2$ and fluid level x, given that the initial phase $J(0) = k$; this is so, since $t^a(j)\,dt$ is the elementary probability that absorption occurs in $(t, t + dt)$ given that the state is j at time t. The equations (17)-(20) are now immediate and obtained by conditioning on the last return epoch to fluid level 0 before absorption occurs in the fluid flow model defining the BPH random variable, and upon noting that the premultipiers $[I - \Psi\Psi_r]^{-1}$ and $\Psi_r[I - \Psi\Psi_r]^{-1}$ in those formulae account for the distribution of the phase immediately after the last return to fluid level 0 before absorption occurs from the positive fluid level x. Equations (21)-(24) are obtained similarly by considering the reflected fluid flow $\mathcal{F}(Q, -C^*)$ and the associated restricted fluid flow $\mathcal{F}_r^+$.

Remark: Note that it is trivial to obtain from (15), (16) in an analogous manner the joint distribution of the time until absorption and the fluid level at the absorption epoch; our interest here is only on the marginal distribution of the fluid level at the absorption epoch, and we have stated the results in their simpler form as they relate to that marginal.

We now obtain the main result of this section.

Theorem 6 *Let $X \sim BPH(\alpha, T, C^*)$ and $f(x)$ denote its density function. Then*

$$f(x) = f_+(x)\chi(x > 0) + f_-(x)\chi(x < 0), \tag{25}$$

where χ denotes an indicator function and

$$f_+(x) = \beta e^{Kx}\kappa, \qquad f_-(x) = \beta_r e^{K_r|x|}\kappa_r, \tag{26}$$

$$K = C_1^{-1}\tilde{K}C_1, \qquad K_r = C_2^{-1}\tilde{K}_r C_2, \tag{27}$$

$$\beta = (\alpha_1 + \alpha_2\Psi_r)[I - \Psi\Psi_r]^{-1}, \quad \kappa = C_1^{-1}t_1^a + \Psi C_2^{-1}t_2^a, \tag{28}$$

$$\beta_r = (\alpha_1\Psi + \alpha_2)[I - \Psi_r\Psi]^{-1}, \quad \kappa_r = \Psi_r C_1^{-1}t_1^a + C_2^{-1}t_2^a. \tag{29}$$

Proof: First of all, note that since the fluid process $F(\cdot)$ is a.s. continuous at all time points t, we have $F(\tau-) = F(\tau+)$ a.s., whence the distribution of the fluid level just prior to the absorption epoch is the same as its distribution just after the absorption epoch. Now, the formula for $f_+(x)$ is obtained from pre-multiplying (17)-(18) by the initial probabilities α_1, pre-multiplying (19)-(20) by α_2, post-multiplying both by the appropriate vectors $\mathbf{1}$ (to reflect the fact that we do not care about the state from which absorption occurs), and summing the resulting expressions. The derivation of f_- is similar and uses (21)-(24).

For a BPH distribution, it is clear from the construction that the atom at 0, if it exists, is of size $1 - \alpha\mathbf{1}$; this is so since the fluid level at absorption is zero iff absorption occurs instantaneously, i.e., the CTMC starts in the absorbing state. We verify this directly from the formulae we have derived for the densities by proving the following result which, incidentally, is also a check on the correctness of the formulae in Theorem 6.

Theorem 7 *We have*

$$\int_{-\infty}^{\infty} f(x)\, dx = \alpha \mathbf{1}.$$

Proof: From equation (14), we have

$$K = C_1^{-1} Q_{11} + \Psi C_2^{-1} Q_{21}. \tag{30}$$

For the flow model $\mathcal{F}$, we have established in [2] (see equation (31) in [2]) the equation

$$Q_{12} + \tilde{\Psi}(s)(Q_{22} - sI) + \tilde{\Theta}(s) Q_{32} + \tilde{K}(s)\tilde{\Psi}(s) C_2 = 0,$$

from which it is easy to see that

$$C_1^{-1} Q_{12} + \Psi C_2^{-1} Q_{22} + K\Psi = 0. \tag{31}$$

From these, it follows that

$$\begin{aligned} K\mathbf{1} &= C_1^{-1}(Q_{11}\mathbf{1} + Q_{12}\mathbf{1}) + \Psi C_2^{-1}(Q_{21}\mathbf{1} + Q_{22}\mathbf{1}) + K\Psi\mathbf{1} \\ &= -C_1^{-1} t_1^a - \Psi C_2^{-1} t_2^a + K\Psi\mathbf{1} \\ &= -\kappa + K\Psi\mathbf{1} \end{aligned}$$

which yields

$$\kappa = -K(\mathbf{1} - \Psi\mathbf{1}). \tag{32}$$

A similar argument for the process $\mathcal{F}_r$ yields the equation

$$\kappa_r = -K_r(\mathbf{1} - \Psi_r\mathbf{1}). \tag{33}$$

Using these, it can be easily seen from (25), (26) that

$$\int_{-\infty}^{\infty} f(x)\, dx = \beta(\mathbf{1} - \Psi\mathbf{1}) + \beta_r(\mathbf{1} - \Psi_r\mathbf{1}).$$

Substituting the values for β and β_r from (28)-(29), and simplifying the expressions using the following equations (which are trivial from a series expansion of the inverses appearing in them)

$$\begin{aligned} \Psi_r[I - \Psi\Psi_r]^{-1} &= [I - \Psi_r\Psi]^{-1}\Psi_r, \\ \Psi[I - \Psi_r\Psi]^{-1} &= [I - \Psi\Psi_r]^{-1}\Psi, \end{aligned}$$

one can now routinely verify that

$$\int_{-\infty}^{\infty} f(x)\, dx = \alpha_1\mathbf{1} + \alpha_2\mathbf{1} = \alpha\mathbf{1};$$

we omit the details.

5 Phase Type Characterizations

The transform free expression for the BPH density helps us to immediately establish that both X^+ and X^- are phase type random variables. Indeed, the following result establishes Theorem 3 stated earlier.

Theorem 8 *Let $X \sim BPH(\alpha, T, C^*)$.*
a) The conditional distribution of X given $X > 0$ is $PH(\gamma, U)$, where

$$\begin{aligned}
\gamma &= \theta^{-1}\beta\Delta, \quad U = \Delta^{-1}K\Delta, \\
\Delta &= diag(\mathbf{1} - \Psi\mathbf{1}), \quad and \\
\theta &= \int_0^\infty f_+(x)\,dx = \beta(-K)^{-1}\kappa.
\end{aligned}$$

b) The conditional distribution of $-X$ given $X < 0$ is $PH(\gamma_r, U_r)$, where

$$\begin{aligned}
\gamma_r &= \theta_r^{-1}\beta_r\Delta_r, \quad U_r = \Delta_r^{-1}K_r\Delta_r, \\
\Delta_r &= diag(\mathbf{1} - \Psi_r\mathbf{1}), \quad and \\
\theta_r &= \int_0^\infty f_-(x)\,dx = \beta_r(-K_r)^{-1}\kappa_r.
\end{aligned}$$

Proof: Note that Ψ is nonnegative and has row sums less than one since the underlying Markov chain has an absorbing state a into which absorption occurs a.s. Thus, Δ is a nonnegative and invertible matrix. From this and the equations (30), (31), we have that that U has negative diagonal and nonnegative off-diagonal elements, and furthermore that

$$\begin{aligned}
K\mathbf{1} &= C_1^{-1}(Q_{11}\mathbf{1} + Q_{12}\mathbf{1}) + \Psi C_2^{-1}(Q_{21}\mathbf{1} + Q_{22}\mathbf{1}) + K\Psi\mathbf{1} \\
&= -C_1^{-1}t_1^a - \Psi C_2^{-1}t_2^a + K\Psi\mathbf{1} \\
&= -\kappa + K\Psi\mathbf{1}.
\end{aligned}$$

It follows that

$$\kappa = (-K)(\mathbf{1} - \Psi\mathbf{1}).$$

Now, using (26), we can calculate for $x > 0$ the density of the conditional distribution, say $g(x)$, of X given $X > 0$, and get

$$\begin{aligned}
g(x) &= \theta^{-1}\beta e^{Kx}\kappa \\
&= \theta^{-1}\beta e^{Kx}(-K)(\mathbf{1} - \Psi\mathbf{1}) \quad \text{by (32)} \\
&= \theta^{-1}\beta\Delta e^{\Delta^{-1}K\Delta x}(-\Delta^{-1}K\Delta)\Delta^{-1}(\mathbf{1} - \Psi\mathbf{1}) \\
&= \gamma e^{Ux}(-U\mathbf{1}).
\end{aligned}$$

We also see that

$$\begin{aligned}\theta &= \beta(-K^{-1})\kappa \\ &= \beta(-K^{-1})(-K)(\mathbf{1}-\Psi\mathbf{1}) \text{ by (32)} \\ &= \beta\Delta\mathbf{1},\end{aligned}$$

whence $\gamma\mathbf{1} = \theta^{-1}\beta\Delta\mathbf{1} = 1$. Comparing the formula for $g(x)$ with that of the density of a PH-distribution, we note that distribution to be $PH(\gamma, U)$. This proves (a). The proof of Part (b) is similar and omitted.

6 Algorithm

In this section, we briefly recall the algorithm developed in [3] for the computation of the matrix $\Psi(s)$ associated with the fluid flow $\mathcal{F} = \mathcal{F}(Q, C^*)$. The computation of $\Psi_r(s)$ is achieved by applying the same algorithm to the flow $\mathcal{F}_r$. While $\Psi(s)$ and $\Psi_r(s)$ are needed to determine the joint distribution of the absorption time and the value of the BPH random variable, note that the marginal distribution of the BPH random variable alone requires only the evaluation of these at $s = 0$.

Assume as given the fluid flow $\mathcal{F}(Q, C^*)$ and s with $Re(s) \geq 0$. We assume the conventions for the partitioning of the state space and various matrices as given in this paper. For $\lambda > 0$, let

$$P_\lambda = \frac{1}{\lambda}C^{-1}Q + I.$$

Choose (fixed) positive numbers λ and δ such that

$$\begin{aligned}\lambda \geq \max_{i\in S}\{-[C^{-1}Q]_{ii}\} \\ \max_{i\in S}\left[\frac{Re(s)}{\lambda}C^{-1}\right]_{ii} \leq \delta < 1, \text{ and} \\ \max_{i\in S}\left[P_\lambda - \frac{Re(s)}{\lambda}C^{-1}\right]_{ii} > 0.\end{aligned}$$

Define the matrices

$$A_2(s,\lambda) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \lambda C_2(sI + 2\lambda C_2)^{-1} & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

$$A_1(s,\lambda) = \Lambda C(sI + \Lambda C)^{-1}\begin{pmatrix} 0 & 0 & 0 \\ \frac{1}{2}P_{21} & \frac{1}{2}P_{22} & \frac{1}{2}P_{23} \\ P_{31} & P_{32} & P_{33} \end{pmatrix},$$

$$A_0(s,\lambda) = \begin{pmatrix} P_{11} - \frac{s}{\lambda}C_1^{-1} & P_{12} & P_{13} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

where $\Lambda = diag(\lambda I, 2\lambda I, \lambda I)$ and $P = P_\lambda$.

Consider now the following algorithm.

Algorithm

Fix $\epsilon > 0$ and set diff $= 100$;
$H^{**}(1,s,\lambda) = (I - A_1(s,\lambda))^{-1} A_0(s,\lambda)$;
$L^{**}(1,s,\lambda) = (I - A_1(s,\lambda))^{-1} A_2(s,\lambda)$;
$G^{**}(1,s,\lambda) = L^{**}(1,s,\lambda)$;
$T(1) = H^{**}(1,s,\lambda)$;
Do while (diff $> \epsilon$)
    $k = k + 1$;
    $U^{**}(k,s,\lambda) = H^{**}(k-1,s,\lambda)L^{**}(k-1,s,\lambda)$
        $+ L^{**}(k-1,s,\lambda)H^{**}(k-1,s,\lambda)$;
    $M = (H^{**}(k-1,s,\lambda))^2$;
    $H^{**}(k,s,\lambda) = (I - U^{**}(k,s,\lambda))^{-1} M$;
    $M = (L^{**}(k-1,s,\lambda))^2$;
    $L^{**}(k,s,\lambda) = (I - U^{**}(k,s,\lambda))^{-1} M$;
    $G^{**}(k,s,\lambda) = G^{**}(k-1,s,\lambda) + T(k-1)L^{**}(k,s,\lambda)$;
    $T(k) = T(k-1)H^{**}(k,s,\lambda)$;
    diff $= \max_{j,k \in S} \mid [G^{**}(k,s,\lambda)]_{j,k} - [G^{**}(k-1,s,\lambda)]_{j,k} \mid$;
end

$$\Psi(s) \approxeq G_{12}^{**}(k,s,\lambda)[G_{22}^{**}(k,s,\lambda)]^{-1}.$$

We have established in [3] that at its termination, the above algorithm yields the matrix $\Psi(s)$ with error at most ϵ in its entries, and furthermore that the error in the k-th iterate is $O[\{\eta(s)\}^{2^k}]$ for a constant $0 < \eta(s) < 1$ so much so that the iterates converge quadratically to the required limit. For details, refer to [3]. With this algorithm, the computation of the phase type representations for X^+ and X^- can be done efficiently.

7 Examples

In this section, we present a set of examples primarily to demonstrate the following facts: (a) a multitude of shapes for the BPH density can be generated with even a small number of phases; (b) interesting subfamilies of BPH

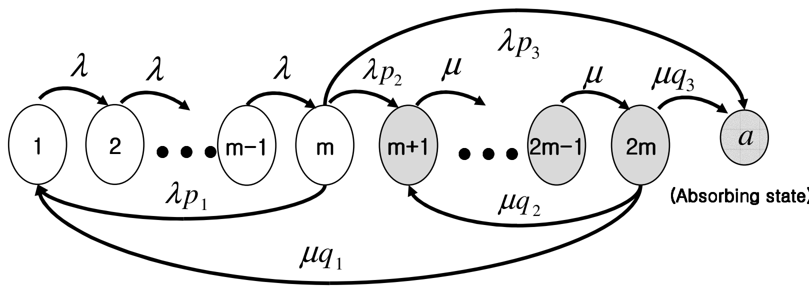


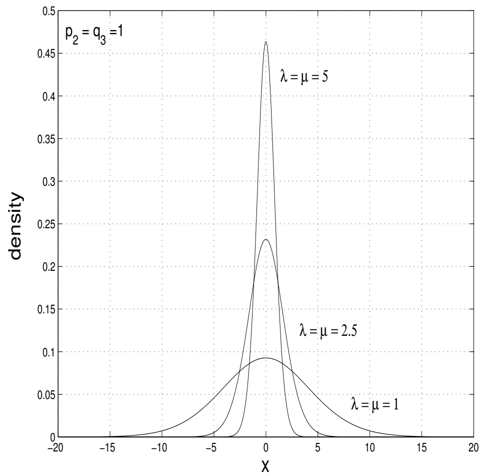
Figure 1: Markov Chain of Phases

distributions with a small number of parameters exist and can be used to model a variety of interesting characteristics.

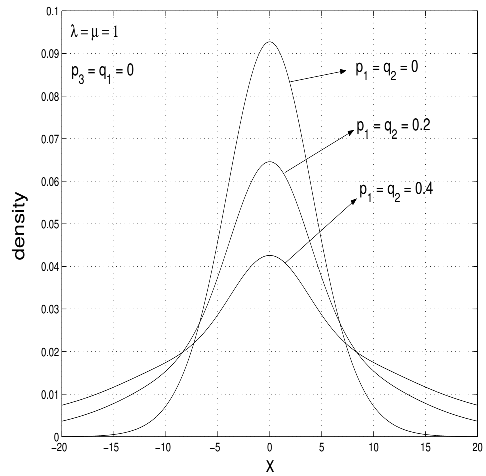
The canonical model of the CTMC we consider to generate the examples is one where there are 20 transient phases in the model. The first 10 have reward rates +1 and mean sojourn times $1/\lambda$, and the last ten have reward rates -1 and mean sojourn times $1/\mu$. Thus, $S_1 = \{1, \cdots, 10\}$ and $S_2 = \{11, \cdots, 20\}$. The phase process starts in phase i with probability α_i. Once it enters the set S_1, the process goes through the remaining phases in that set sequentially at the end of which it may go to phase 11 with probability p_2 or may return back to phase 1 with probability p_1. Similarly, once it enters the set S_2, it goes through the remaining phases in S_2 sequentially at the end of which it may loop back to either phase 1 with probability q_1 or to phase 11 with probability q_2 or may get absorbed with probability q_3. The general set up is shown in Figure 1.

The examples shown in Figure 2(a) correspond to the case $\alpha_1 = 1$, $\lambda = \mu$, $p_2 = q_3 = 1$. These examples correspond to the case where the BPH random variable is realized as the difference of two independent iid Erlang random variables. All distributions there have mean zero, and by choosing the value of m and λ appropriately, the variability can be controlled. Due to the Central Limit Theorem, by choosing the parameter m governing the number of phases, these distributions can be made arbitrarily close to a normal distribution.

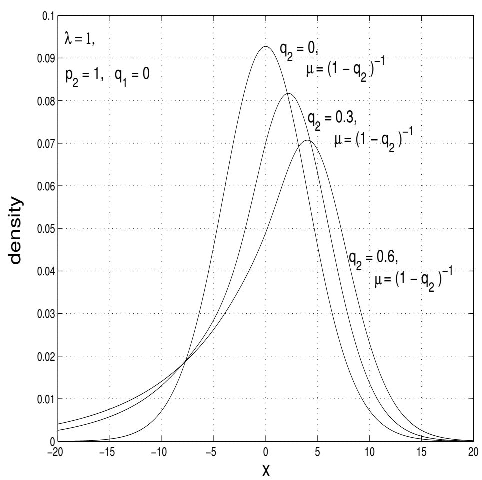
The examples shown in Figure 2(b) demonstrate a set of symmetric distributions with tails fatter than the ones considered in the first set. Here the return paths within each set of transient phases is used to model the positive and negative parts as a geometric mixture of successive convolutions



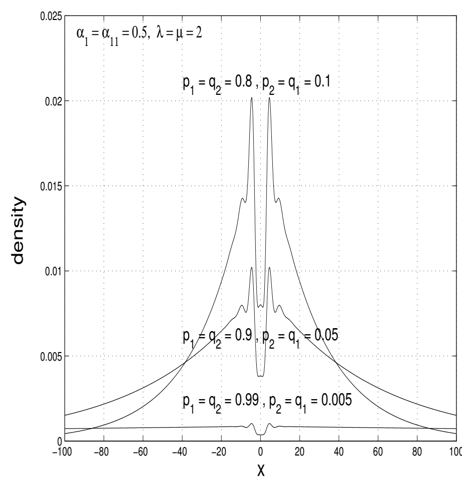
(a) Difference of Erlangs



(b) Difference of Mixtures of Erlangs



(c) Skewed Distributions with Mean Zero



(d) Some Multimodal Distributions

Figure 2: Examples of BPH distributions

of an Erlang distribution. As would be expected, with increasing feedback probabilities, the tails do get fatter. Here again, the intial probability vector is such that $\alpha_1 = 1$.

In the set of Figure 2(c) are shown a set of skewed distributions with mean zero. Although there are many ways of generating increasingly skewed distributions, the method we have chosen here is to hold λ constant and to vary the values of μ and the feedback probability within the set S_2. The initial probability vector is once gain such that $\alpha_1 = 1$.

The denseness property demonstrates that essentially any shape can be approached by a BPH density. Even the simple special class considered by us can generate quite complex multimodal distributions. We have illustrated this in the last set of examples in Figure 2(d).

8 Concluding Remarks

A naive approach to fitting a BPH distribution based on an independent sample on X would be to consider the positive and negative observations in the sample separately and to fit PH distributions to each of them; for the latter, we could use, for instance, the EM algorithm given in [5]. After obtaining PH representations for X^+ and X^-, one could use the construction in the proof of Theorem 4 to get a BPH representation for the entire sample. It is, however, not clear that this would be an efficient approach. The matrices U and U_r may not be independent parameters in that they arise from two fluid processes that are reflections of each other and are therefore highly dependent, and this type of dependency is not captured in this approach. (Compare this approach, for instance, to fitting a normal distribution to data by separating positive and negative values in the sample.) Also, while BPH distributions can have densities that are continuous at the origin, the type of fit one obtains by considering positive and negative values separately will in all likelihood result in a discontinuity at zero; see the construction in the proof of Theorem 4 when $q = 1 - p$. Thus, the problem of fitting BPH distributions to data appears to be legitimately new and offering some interesting challenges.

In certain applications related to Markov additive processes to which our setup applies, we may be interested in the joint distribution of X, τ and $J(\tau-)$, and the successive observations in the sample may not even be assumed independent. That setup is even more complicated for statistical inference.

Note also that it is trivial to analyze the case where, in the construction of the BPH random variable, the initial fluid level is nonzero. One merit in considering this is that the resulting class of distributions is also closed under translations.

Finally, the introduction of general reward rates may have some computational advantages in that it allows one to choose the speeds in the CTMC conveniently.

All these deserve much further study.

Acknowledgment: The authors thank Soren Asmussen for his comments on a draft raising some interesting questions that led to the discussion in the Concluding Remarks. Soohan Ahn thanks AT&T where this work was initiated during his tenure as a post-doctoral researcher, and both authors thank the Institute of Mathematical Analysis, University of Minnesota during a visit to which some of the major results here were obtained. Soohan Ahn was supported partially also by Grant R01-2004-000-10284-0 in the Basic Research Program of the Korea Science & Engineering Foundation.

References

[1] Ahn, S & Ramaswami, V (2003): Fluid flow models & queues - A connection by stochastic coupling, *Stochastic Models,* vol. 19, pp. 325-348.

[2] Ahn, S & Ramaswami, V (2004) : Transient analysis of fluid flow models via stochastic coupling to a queue, *Stochastic Models,* vol. 20, pp. 71-102.

[3] Ahn, S & Ramaswami, V (2005): Efficient algorithms for transient analysis of stochastic fluid flow models, *J. Appl. Prob.*, to appear.

[4] Asmussen, S (1991): Phase type representation of waiting times, in *Queueing, Performance & Control in ATM*, J.W. Cohen & C.D. Pack (Eds.), pp. 157-159, North Holland, Amsterdam.

[5] Asmussen, S., Nerman, M. & Olsson, M (1996): Fitting phase-type distributions via the EM algorithm, *Scand. J. Stat.,* vol. 23, pp. 419-441.

[6] Asmussen, S (1995): Stationary distributions for fluid flow models with or without Brownian noise, *Stochastic Models,* vol. 11, pp. 1-20.

[7] Asmussen, S & Bladt, M (1996): Renewal theory and queueing algorithms for matrix exponential distributions, in *Matrix-Analytic Methods in Stochastic Models,* Chakravarthy, S & Alfa, A (Eds.), Marcel Dekker, NY.

[8] Bladt, M & Neuts, M.F (2003): Matrix-exponential distributions: calculus and interpretations via flows, *Stochastic Models*, vol. 19, pp. 113-124.

[9] Latouche, G & Ramaswami, V (1997): Spatial point patterns of phase type, in *Teletraffic Contributions for the Information Age*, Ramaswami, V & Wirth, P (Eds.), Elsevier Science B.V.

[10] Latouche, G & Ramaswami, V (1999): *Introduction to matrix analytic methods in Stochastic Modeling,* SIAM & ASA, Philadelphia.

[11] Lucantoni, D.M (1993): The BMAP/G/1 queue: A tutorial, in *Models and Techniques for Performance Evaluation of Computer and Communication Systems*, Donatiello, L & Nelson, R (Eds.), Springer Verlag, pp. 330-358.

[12] Neuts, M.F (1975): Probability distributions of phase type, in *Liber Amicorum Prof. Emeritus H. Florin*, Dept. of Math., University of Louvain, Belgium, 173-206.

[13] Neuts, M.F (1979): A versatile Markovian point process, *J. Appl. Prob.*, vol. 16, pp. 764-779.

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

[15] Neuts, M.F (1989): *Structured Stochastic Matrices of M/G/1 Type and Their Applications,* Marcel Dekker, NY.

[16] Ramaswami, V (1980): The N/G/1 queue and its detailed analysis, *Adv. Appl. Prob.,* vol. 12, pp. 222-261.

[17] Ramaswami, V (1990): From the matrix-geometric to the matrix-exponential, *QUESTA*, vol. 6, 229-260.

[18] Ramaswami, V (1999): Matrix analytic methods for stochastic fluid flows, in *Teletraffic Engineering in a Competitive World - Proc. of the 16th International Teletraffic Congress,* D. Smith and P. Key (eds.), pp. 1019–1030, Elsevier.

[19] Ramaswami, V (2000): Algorithmic analysis of stochastic models - The changing face of mathematics, Ramanujan Memorial Lecture in *Mathematics for the New Millennium*, Vijayakumar, A & Krishnakumar, K (Eds.), Anna University, Chennai, India.

[20] Ramaswami, V., Poole, D., Ahn, S., Byers, S., & Kaplan, A.E (2005): Assuring emergency services access: providing dial tone in the presence of long holding time internet dial-up calls, Informs Wagner Prize Finalist Paper, to appear in *Interfaces*.

[21] Shanthikumar, J.G (1985): Bilateral phase type distributions, *Naval Res. Log. Quarterly*, vol. 32, pp. 119-136.