Available online at www.sciencedirect.com
SCIENCE DIRECT®

Performance Evaluation 63 (2006) 524–552

PERFORMANCE EVALUATION
An International Journal

www.elsevier.com/locate/peva

# Closed form solutions for mapping general distributions to quasi-minimal PH distributions

Takayuki Osogami*, Mor Harchol-Balter

*Computer Science Department, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, United States*

Available online 27 July 2005

**Abstract**

Approximating general distributions by phase-type (PH) distributions is a popular technique in stochastic analysis, since the Markovian property of PH distributions often allows analytical tractability. This paper proposes an algorithm for mapping a general distribution, $G$, to a PH distribution, which matches the first three moments of $G$. Efficiency of our algorithm hinges on narrowing the search space to a particular subset of the PH distributions, which we refer to as Erlang–Coxian (EC) distributions. The class of EC distributions has a small number of parameters, and we provide closed form solutions for these. Our solution applies to any distribution whose first three moments can be matched by a PH distribution. Also, our resulting EC distribution requires a nearly minimal number of phases, within one of the minimal number of phases required by any acyclic PH distribution.


*Keywords:* PH distribution; Moment matching; Closed form; Normalized moment

## 1. Motivation

There is a large body of literature on the topic of approximating general distributions by phase-type (PH) distributions, whose Markovian properties make them far more analytically tractable. Much of this

* Corresponding author.
*E-mail addresses:* osogami@cs.cmu.edu (T. Osogami), harchol@cs.cmu.edu (M. Harchol-Balter).
http://www.cs.cmu.edu/~osogami/
http://www.cs.cmu.edu/~harchol/

0166-5316/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.peva.2005.06.002

research has focused on the specific problem of finding an algorithm which maps a general distribution, $G$, to a PH distribution, $P$, where $P$ and $G$ agree on the first three moments. Throughout this paper we say that $G$ is *well represented* by $P$ if $P$ and $G$ agree on their first three moments. We choose to limit our discussion in this paper to three-moment matching, because matching the first three moments of an input distribution has been shown to be effective in predicting mean performance for variety of computer system models [5,17,23,29]. However, three moments might not always suffice for every problem, and we leave the problem of matching more moments to future work.

Moment matching algorithms can be evaluated along four different measures: (i) *The number of moments matched*: In general matching more moments is more desirable. (ii) *The computational efficiency of the algorithm*: It is desirable that the algorithm have short running time. Ideally, one would like a closed form solution for the parameters of the matching PH distribution. (iii) *The generality of the solution*: Ideally the algorithm should work for as broad a class of distributions as possible. (iv) *The minimality of the number of phases*: It is desirable that the matching PH distribution, $P$, have a small number of phases. Recall that the goal is to find a $P$ which can replace the input distribution $G$ in some stochastic process to model it as a Markov chain. Since it is desirable that the state space of this resulting Markov chain be kept small, we want to keep the number of phases in $P$ low.

This paper proposes moment matching algorithms which perform very well along all four of these measures. This constitutes the primary contribution of the paper. Our solution matches three moments, provides a closed form representation of the parameters of the matching PH distribution, applies to all distributions which can be well represented by a PH distribution, and is nearly minimal in the number of phases required.

The general approach in designing moment matching algorithms in the literature is to start by defining a subset $\mathcal{S}$ of the PH distributions, and then match each input distribution $G$ to a distribution in $\mathcal{S}$. The reason for limiting the solution to a distribution in $\mathcal{S}$ is that this narrows the search space and thus improves the computational efficiency of the algorithm. Observe that $n$-phase PH distributions have $\Theta(n^2)$ free parameters (see Fig. 1), while $\mathcal{S}$ can be defined to have far fewer free parameters. One has to be careful in defining the subset $\mathcal{S}$, however. If $\mathcal{S}$ is too small, it may limit the space of distributions that can be well represented. Also, if $\mathcal{S}$ is too small, it may exclude solutions with a minimal number of phases.

In this paper we define a subset of PH distributions, which we call Erlang–Coxian (EC) distributions. EC distributions have only six free parameters, which allows us to derive a closed form solution for these parameters in terms of the input distribution. The set of EC distributions is general enough, however, that for any distribution, $G$, that can be well represented by an $n$-phase acyclic PH distribution, there exists an EC distribution, with at most $n + 1$ phases, that well represents $G$.

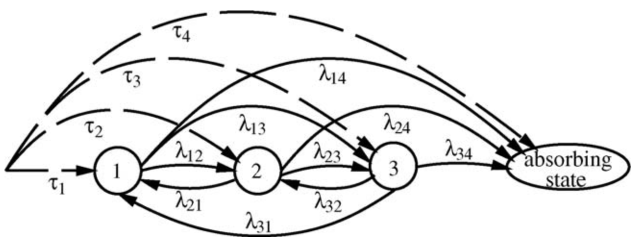



Fig. 1. The continuous time Markov chain whose absorption time defines a three-phase PH distribution.

It is not clear whether restricting our search space to the set of *acyclic* PH distributions (as is used throughout the literature) is limiting. While it is theoretically possible that the minimum phase solution is cyclic, in practice we have not been able to find a situation where the minimal solution requires cycles, and this question is left as an open problem. However, an acyclic PH distribution has a computational advantage over a cyclic one, since the generator matrix of the underlying Markov chain of an acyclic PH distribution is upper triangular. Therefore, in some applications, one might prefer an acyclic PH distribution with more phases to a cyclic PH distribution with less phases. Thus, in this paper, we limit our focus to the set of *acyclic* PH distributions.

To prove that our moment matching algorithm results in a nearly minimal number of phases, we need to know the minimal number of phases needed to well represent an input distribution by a PH distribution. As a secondary contribution, this paper provides a formal characterization of the set of distributions that are well represented by an $n$-phase acyclic PH distribution, for each $n = 1, 2, 3, \ldots$. This characterization is used to prove the minimality of the number of phases used in our moment matching algorithms.

## 2. Overview of key ideas and definitions

We start with some definitions that we use throughout the paper.

**Definition 1.** A PH distribution is the distribution of the absorption time in a continuous time Markov chain. A PH distribution, $F$, is specified by a generator matrix, $\mathbf{T}^F$, and an initial probability vector, $\vec{\tau}^F$.

Fig. 1 shows a three-phase PH distribution, $F$, with $\vec{\tau}^F = (\tau_1, \tau_2, \tau_3)$ and

$$\mathbf{T}^F = \begin{pmatrix} -(\lambda_{12} + \lambda_{13} + \lambda_{14}) & \lambda_{12} & \lambda_{13} \\ \lambda_{21} & -(\lambda_{21} + \lambda_{23} + \lambda_{24}) & \lambda_{23} \\ \lambda_{31} & \lambda_{32} & -(\lambda_{31} + \lambda_{32} + \lambda_{34}) \end{pmatrix}.$$

There are $n = 3$ internal states. With probability $\tau_i$ we start in the $i$th state. The absorption time is the sum of the times spent in each internal state before reaching the absorption state. Note that the absorbing state (state 4) and the associated initial probability and transition rates are not present in $\vec{\tau}^F$ and $\mathbf{T}^F$.

An important subset of PH distributions is the set of acyclic PH distributions and the set of Coxian PH distributions, which are defined as follows.

**Definition 2.** An acyclic PH distribution is a PH distribution with $\lambda_{ij} = 0$ for all $i > j$. An $n$-phase Coxian PH distribution is an $n$-phase acyclic PH distribution with $\tau_i = 0$ for $2 \leq i \leq n$ and $\lambda_{ij} = 0$ for $i + 1 < j \leq n$. An $n$-phase Coxian$^+$ PH distribution is an $n$-phase Coxian distribution with $\tau_1 = 1$.

Note that any acyclic PH distribution can be represented by a Coxian PH distribution, based on the result of Cumani [3].

In providing a *simple* representation and analysis of our closed form solution, it will be very helpful to start by defining an alternative to the standard moments, which we refer to as *normalized moments*.

**Definition 3.** Let $\mu_k^F$ be the $k$th moment of a distribution $F$ for $k = 1, 2, 3$. The *normalized kth moment* $m_k^F$ of $F$ for $k = 2, 3$ is defined to be $m_2^F = \mu_2^F/(\mu_1^F)^2$ and $m_3^F = \mu_3^F/\mu_1^F\mu_2^F$.

Notice the relationship between the normalized moments and the coefficient of variation $C_F$ and the skewness $\gamma_F$ of $F$: $m_2^F = C_F^2 + 1$ and $m_3^F = \nu_F \sqrt{m_2^F}$, where $\nu_F = \mu_3^F/(\mu_2^F)^{3/2}$. ($\nu_F$ and $\gamma_F$ are closely related, since $\gamma_F = \bar{\mu}_3^F/(\bar{\mu}_2^F)^{3/2}$, where $\bar{\mu}_k^F$ is the centralized $k$th moment of $F$ for $k = 2, 3$.)

**Definition 4.** A distribution $G$ is *well represented* by a distribution $F$ if $F$ and $G$ agree on their first three moments.

**Definition 5.** $\mathcal{PH}_3$ refers to the set of distributions that are well represented by a PH distribution.

It is known that a distribution $G$ is in $\mathcal{PH}_3$ iff its normalized moments satisfy $m_3^G > m_2^G > 1$ [10]. Since any nonnegative distribution $G$ satisfies $m_3^G \geq m_2^G \geq 1$ [13], $\mathcal{PH}_3$ contains almost all the nonnegative distributions.

**Proposition 1.** *The set of nonnegative distributions are dense in* $\mathcal{PH}_3$.

**Definition 6.** OPT($G$) is defined to be the minimum number of phases in an acyclic PH distribution, allowing a mass probability at zero, that well represents a distribution $G$.

### 2.1. Moment matching algorithms

*Previous work on moment matching algorithms*. Prior work has contributed a large number of moment matching algorithms. While all of these algorithms excel with respect to some of the four measures mentioned earlier, they all are deficient in at least one of these measures as explained below.

In cases where matching only two moments suffices, it is possible to achieve solutions that perform very well along all the other three measures. Sauer and Chandy [21] provide a closed form solution for matching two moments of a general distribution with squared coefficient of variation $C^2 > 0$ (i.e. any distribution in $\mathcal{PH}_3$). They use a two-phase hyper-exponential distribution for matching distributions with $C^2 > 1$ and a generalized Erlang distribution for matching distributions with $C^2 < 1$. Marie [15] provides a closed form solution for matching two moments of a general distribution with $C^2 > 0$. He uses a two-phase Coxian$^+$ PH distribution for distributions with $C^2 > 1$ and a generalized Erlang distribution for distributions with $C^2 < 1$.

If one is willing to match only a subset of distributions, then again it is possible to achieve solutions that perform very well along the remaining three measures. Whitt [28] and Altiok [2] focus on the set of distributions with $C^2 > 1$ and sufficiently high third moment. They obtain a closed form solution for matching three moments of any distribution in this set. Whitt matches to a two-phase hyper-exponential distribution, and Altiok matches to a two-phase Coxian$^+$ PH distribution. Telek and Heindl [25] focus on the set of distributions with $C^2 \geq 1/2$ and various constraints on the third moment. They obtain a closed form solution for matching three moments of any distribution in this set, by using a two-phase acyclic PH distribution with no mass probability at zero.

Johnson and Taaffe [9,10] come closest to achieving all four measures. They provide a closed form solution for matching the first three moments of any distribution $G \in \mathcal{PH}_3$. They use a mixed Erlang distribution with common order. Unfortunately, this mixed Erlang distribution requires $2\text{OPT}(G) + 2$ phases in the worst case.

In complementary work, Johnson and Taaffe [11,12] again look at the problem of matching the first three moments of any distribution $G \in \mathcal{PH}_3$, this time using three types of PH distributions: a mixture of Erlang distributions, a Coxian$^+$ PH distribution, and a general PH distribution. Their solution is nearly minimal in that it requires at most OPT($G$) + 2 phases. Unfortunately, their algorithm requires solving a nonlinear programming problem and hence is computationally inefficient, requiring time exponential in OPT($G$).

Above we have described the prior work focusing on moment matching algorithms, which is the focus of this paper. There is also a large body of work focusing on fitting the *shape* of an input distribution using a PH distribution. Recent research has looked at fitting heavy-tailed distributions to PH distributions [4,6,7,14,20,24]. There is also work which combines moment matching with the goal of fitting the shape of the distribution [8,22]. The work above is clearly broader in its goals than simply matching three moments. Unfortunately there is a tradeoff: obtaining a more precise fit requires more phases, and it can sometimes be computationally inefficient [8,22].

*The key idea behind our algorithm: The EC distribution*. In all the prior work on computationally efficient moment matching algorithms, the approach is to match a general input distribution $G$ to some subset, $\mathcal{S}$, of the acyclic PH distributions. In this paper, our subset $\mathcal{S}$ is the EC distribution:

**Definition 7.** An $n$-phase Erlang–Coxian (EC) distribution is a convolution of an $(n-2)$-phase Erlang distribution, $E_{n-2}$, and a two-phase Coxian$^+$ distribution possibly with mass probability at zero.

Fig. 2 shows the Markov chain whose absorption time defines an $n$-phase EC distribution. Below, an $N$-phase Erlang distribution, $E_N$, is also called an Erlang-$N$ distribution.

We now provide some intuition behind the creation of the EC distribution. Recall that a Coxian$^+$ PH distribution is very good for approximating a distribution with high variability. In particular, a two-phase Coxian$^+$ PH distribution is known to well represent any distribution that has high second and third moments (any distribution $G$ that satisfies $m_2^G > 2$ and $m_3^G > (3/2)m_2^G$) [19]. However a Coxian$^+$ PH distribution requires more phases for approximating distributions with lower second and third moments. For example, a Coxian$^+$ PH distribution requires at least $n$ phases to well represent a distribution $G$ with $m_2^G \leq (n+1)/n$ for $n \geq 1$ (see Section 3). The large number of phases needed implies that many free parameters must be determined, which implies that any algorithm that tries to well represent an arbitrary distribution using a minimal number of phases is likely to suffer from computational inefficiency.

By contrast, an $n$-phase Erlang distribution has only two free parameters and is also known to have the least normalized second moment among all the $n$-phase PH distributions [1,16]. However the Erlang distribution is obviously limited in the set of distributions which it can well represent.

By combining the Erlang distribution with the two-phase Coxian$^+$ PH distribution, we can represent distributions with all ranges of variability, while using only a small number of phases. Furthermore, the fact that the EC distribution has a small number of parameters $(n, p, \lambda_Y, \lambda_{X1}, \lambda_{X2}, p_X)$ allows us to obtain closed from expressions for these parameters that well represent any given distribution in $\mathcal{PH}_3$.

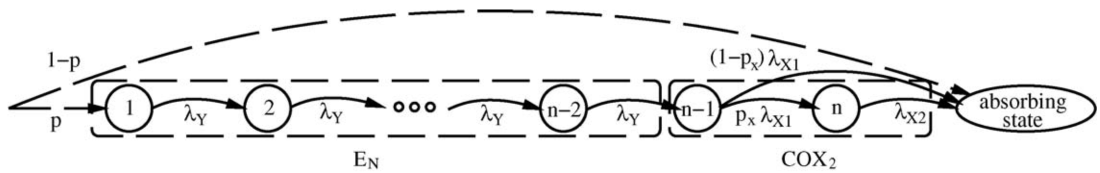



Fig. 2. The Markov chain whose absorption time defines an $n$-phase EC distribution. The first box above depicts the Markov chain whose absorption time defines an Erlang-$N$ distribution, where $N = n - 2$, and the second box depicts the Markov chain whose absorption time defines a two-phase Coxian$^+$ PH distribution. Notice that the rates in the first box are the same for all states.

### 2.2. Characterizing PH distributions

We now turn to our second goal of the paper, namely characterizing the set of distributions that are well represented by an $n$-phase acyclic PH distribution.

**Definition 8.** Let $\mathcal{S}^{(n)}$ denote the set of distributions that are well represented by an $n$-phase acyclic PH distribution for positive integer $n$.

All prior work on characterizing $\mathcal{S}^{(n)}$ has focused on characterizing $\mathcal{S}^{(2)^*}$, where $\mathcal{S}^{(2)^*}$ is the set of distributions which are well represented by a two-phase Coxian$^+$ PH distribution. Observe $\mathcal{S}^{(2)^*} \subset \mathcal{S}^{(2)}$. Altiok [2] showed a sufficient condition for a distribution to be in $\mathcal{S}^{(2)^*}$. More recently, Telek and Heindl [25] expanded Altiok's condition and proved the necessary and sufficient condition for a distribution to be in $\mathcal{S}^{(2)^*}$. While neither Altiok nor Telek and Heindl expressed these conditions in terms of normalized moments, the results can be expressed more simply with our normalized moments:

**Theorem 1** (Telek and Heindl [25]). *$G \in S^{(2)^*}$ iff $G$ satisfies exactly one of the following three conditions:* (i) $(9m_2^G - 12 + 3\sqrt{2}(2 - m_2^G)^{3/2})/m_2^G \le m_3^G \le 6(m_2^G - 1)/m_2^G$ *and* $3/2 \le m_2^G < 2$, (ii) $m_3^G = 3$ *and* $m_2^G = 2$, *or* (iii) $(3/2)m_2^G < m_3^G$ *and* $2 < m_2^G$.

In this paper, we will characterize $\mathcal{S}^{(n)}$, for all integers $n \ge 2$.

*Our Characterization of PH distributions*. While our goal is to characterize the set $\mathcal{S}^{(n)}$, this characterization turns out to be ugly. One of the key ideas is that there is a set $\mathcal{T}^{(n)} \supset \mathcal{S}^{(n)}$ which is very close to $\mathcal{S}^{(n)}$ in size, such that $\mathcal{T}^{(n)}$ has a simple specification via normalized moments. Thus, much of the proofs in our characterization revolve around $\mathcal{T}^{(n)}$.

**Definition 9.** For integers $n \ge 2$, let $\mathcal{T}^{(n)}$ denote the set of distributions, $F$, that satisfy exactly one of the following two conditions: (i) $m_2^F > (n+1)/n$ and $m_3^F \ge ((n+2)/(n+1))m_2^F$, or (ii) $m_2^F = (n+1)/n$ and $m_3^F = (n+2)/n$.

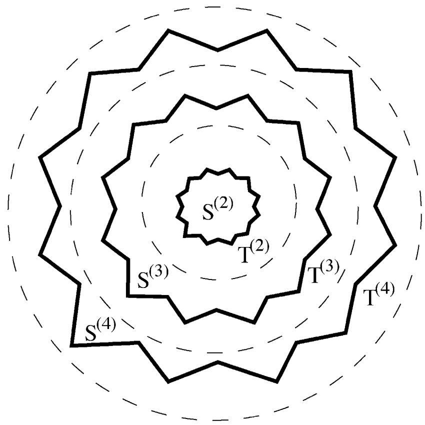
The main contribution of our characterization of acyclic PH distributions is a derivation of the nested relationship between $\mathcal{S}^{(n)}$ and $\mathcal{T}^{(n)}$ for all $n \ge 2$. This relationship is illustrated in Fig. 3. Observe that $\mathcal{S}^{(n)}$ is a proper subset of $\mathcal{S}^{(n+1)}$, and likewise $\mathcal{T}^{(n)}$ is a proper subset of $\mathcal{T}^{(n+1)}$ for all integers $n \ge 2$. More importantly, the nested relationship between $\mathcal{S}^{(n)}$ and $\mathcal{T}^{(n)}$ is formally characterized in the next theorem.

**Theorem 2.** *For all integers $n \ge 2$, $\mathcal{S}^{(n)} \subset \mathcal{T}^{(n)} \subset \mathcal{S}^{(n+1)}$.*

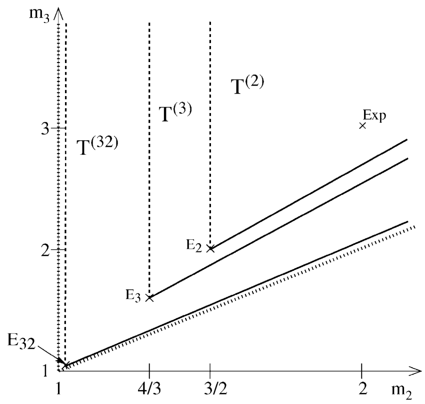
The property $\mathcal{S}^{(n)} \subset \mathcal{T}^{(n)}$ is important because it will allow us to prove that the EC distribution produced by our moment matching algorithm uses a nearly minimal number of phases. The property $\mathcal{T}^{(n)} \subset \mathcal{S}^{(n+1)}$ is important in completing our characterization of $\mathcal{S}^{(n)}$. This property will follow immediately from our construction of a moment matching algorithm.

### 2.3. Outline of paper

The first part of the paper will describe the characterization of $\mathcal{S}^{(n)}$, which is covered primarily in Section 3. This characterization will be used in the second part, which involves the construction of moment matching algorithms (Sections 4–7). Our moment matching algorithms depend on properties of EC distributions, which will be discussed in depth in Section 4. In Sections 5–7, we present three variants of closed form solutions, `Simple`, `Complete`, and `Positive`, each of which uses at most $\mathrm{OPT}(G) + 1$ phases but achieves slightly different goals. The `Simple` solution (see Section 5) has the




(a) Characterizing $\mathcal{S}^{(n)}$ via $\mathcal{T}^{(n)}$




(b) Set $\mathcal{T}^{(n)}$

Fig. 3. (a) The nested structure of $\mathcal{S}^{(n)}$ and $\mathcal{T}^{(n)}$: solid lines delineate $\mathcal{S}^{(n)}$ (which is irregular) and dashed lines delineate $\mathcal{T}^{(n)}$ (which is regular—has a simple specification). (b) Set $\mathcal{T}^{(n)}$ is depicted as a function of the normalized moments. $\mathcal{T}^{(n)}$ sets are delineated by solid lines, which include the border, and dashed lines, which do not include the border ($n = 2, 3, 32$). Observe that all possible nonnegative distributions lie within the region delineated by the two dotted lines: $m_2 \geq 1$ and $m_3 \geq m_2$.

advantage of simplicity and readability; however it does not work for all distributions in $\mathcal{PH}_3$ (although it works for almost all). The `Complete` solution (see Section 6) is defined for all the input distributions in $\mathcal{PH}_3$, and the number of phases used in the `Complete` solution is no more than those used in the `Simple` and `Positive` solutions for any distribution. In the `Simple` and `Complete` solution, the matching EC distribution can have mass probability at zero ($p < 1$). In some applications, however, it is desirable that the matching PH distribution has no mass probability at zero. The `Positive` solution (see Section 7) has no mass probability at zero ($p = 1$), and is defined for almost all distributions in $\mathcal{PH}_3$.

## 3. Characterizing phase type distributions

Set $\mathcal{S}^{(n)}$ is characterized by Theorem 2: $\mathcal{S}^{(n)} \subset \mathcal{T}^{(n)} \subset \mathcal{S}^{(n+1)}$ *for all* $n \geq 2$. In this section we prove the first part of the theorem.

**Lemma 1.** *For all integers* $n \geq 2$, $\mathcal{S}^{(n)} \subset \mathcal{T}^{(n)}$.

The second part, $\mathcal{T}^{(n)} \subset \mathcal{S}^{(n+1)}$, follows immediately from the construction of the `Complete` solution (see Corollary 3 in Section 6).

We begin by defining the ratio of the normalized moments.

**Definition 10.** The ratio of the normalized moments of a distribution $F$, $r^F$, is defined as $r^F = m_3^F / m_2^F$ and is also referred to as the $r$-value of $F$.

A nice property of the $r$-value is that it is insensitive to the mass probability at zero.

**Proposition 2.** *Let* $Z$ *be a mixture of two distributions*, $X$ *and* $O$, *where* $X$ *is a nonnegative distribution with* $\mu_1^X > 0$ *and* $O$ *is the distribution of the degenerate random variable* $V \equiv 0$. *Then*, $r^Z = r^X$.

**Proof.** Let $Z(\cdot)$, $X(\cdot)$, and $O(\cdot)$ be the (cumulative) distribution functions of $Z$, $X$, and $O$, respectively. Then, there exists $0 < p < 1$ such that $Z(\cdot) = pX(\cdot) + (1 - p)O(\cdot)$. Therefore, by definition,

$$r^Z = \frac{(p\mu_3^X)(p\mu_1^X)}{(p\mu_2^X)^2} = \frac{(\mu_3^X)(\mu_1^X)}{(\mu_2^X)^2} = r^X. \qquad \square$$

Below, unless otherwise stated, we denote the (cumulative) distribution function of a distribution, $X$, by $X(\cdot)$. Below, we use the notation $O$ repeatedly.

**Definition 11.** Let $O$ denote the distribution of the degenerate random variable $V \equiv 0$.

Note that, using the normalized second moment and the $r$-value, $\mathcal{T}^{(n)}$ can be redefined as the set of distributions, $F$, that satisfy exactly one of the following two conditions:

(i) $\quad m_2^F > \frac{n+1}{n} \quad \text{and} \quad r^F \geq \frac{n+2}{n+1}$, or

(ii) $\quad m_2^F = \frac{n+1}{n} \quad \text{and} \quad r^F = \frac{n+2}{n+1}$.

To show $\mathcal{S}^{(n)} \subset \mathcal{T}^{(n)}$, consider an arbitrary distribution, $X \in \mathcal{S}^{(n)}$. By definition of $\mathcal{S}^{(n)}$, there exists an $n$-phase acyclic PH distribution, $P$, that well represents $X$. Thus, $X \in \mathcal{T}^{(n)}$ iff $P \in \mathcal{T}^{(n)}$. Hence, it suffices to prove that all $n$-phase acyclic PH distributions are in $\mathcal{T}^{(n)}$. This can be shown by proving the two properties of the Erlang-$n$ distribution: (i) the set of Erlang-$n$ distributions has the least normalized second moment among all the $n$-phase (acyclic) PH distributions and (ii) the Erlang-$n$ distribution has the least $r$-value among all the $n$-phase acyclic PH distributions. Note that the Erlang-$n$ distribution, $E_n$, has

$$m_2^{E_n} = \frac{n+1}{n} \quad \text{and} \quad r^{E_n} = \frac{n+2}{n+1}.$$

Property (i) of the Erlang-$n$ distribution immediately follows from the prior work by Aldous and Shepp [1] and O'Cinneide [16], who prove that the set of Erlang-$n$ distributions is the unique class of $n$-phase PH distributions with the least second moment among all the $n$-phase PH distributions with a fixed first moment. Thus, all that remains is to prove property (ii).

Our proof makes use of the recursive structure of PH distributions and shows that an $n$-phase Erlang distribution has no greater $r$-value than any $n$-phase acyclic PH distribution. The key idea is that any acyclic PH distribution, $P$, can be seen as a mixture of the convolutions of exponential distributions, and one of the convolutions of exponential distributions has no greater $r$-value than $P$. This allows us to relate the minimal convolution to an Erlang distribution when all the rates of the exponential distributions are the same. The following lemma provides the key property of the $r$-value used in our proof.

**Lemma 2.** *Let $Z(\cdot) = \sum_{i=1}^n p_i X_i(\cdot)$, where $n \geq 2$ and $X_i$ are nonnegative distributions with $\mu_1^{X_i} > 0$ for $1 \leq i \leq n$. Then, there exists $i \in \{1, \ldots, n\}$ such that $r^Z \geq r^{X_i}$.*

**Proof.** We prove the lemma by induction on $n$. Without loss of generality, we let $r^{X_1} \geq \cdots \geq r^{X_n}$.

*Base case* ($n = 2$): Let $v = \mu_1^{X_2}/\mu_1^{X_1}$ and $w = \mu_2^{X_2}/\mu_2^{X_1}$. Then,

$$\begin{aligned} r^Z - r^{X_2} &= \frac{p_1^2 r^{X_1} + p_1 p_2 r^{X_1} v + p_1 p_2 r^{X_2} w^2/v + p_2^2 r^{X_2} w^2}{(p_1 + p_2 w)^2} - r^{X_2} \\ &\geq \frac{(p_1^2 + p_1 p_2 v + p_1 p_2 w^2/v + p_2^2 w^2 - (p_1 + p_2 w)^2) r^{X_2}}{(p_1 + p_2 w)^2} = \frac{p_1 p_2 (w - v)^2 r^{X_2}}{v (p_1 + p_2 w)^2} \geq 0, \end{aligned}$$

where the first inequality follows from $r^{X_1} \geq r^{X_2}$.

*Inductive case*: Suppose that the lemma holds for $n \leq k$. When $n = k + 1$, $Z$ can be seen as a mixture of two distributions, $Y(\cdot) = (1/(1 - p_{k+1})) \sum_{i=1}^{k} p_i X_i(\cdot)$ and $X_{k+1}(\cdot)$. When $r^{X_{k+1}} \leq r^Z$, the lemma holds for $n = k + 1$. When $r^{X_{k+1}} > r^Z$, we have $r^Y \leq r^Z$ by the base case. By the inductive hypothesis, there exist $i \in \{1, \ldots, k\}$ such that $r^Y \geq r^{X_i}$. Thus, the lemma holds for $n = k + 1$. □

We are now ready to prove that an $n$-phase Erlang distribution has no greater $r$-value than any $n$-phase PH distribution, which completes the proof of Lemma 1.

**Lemma 3.** *The Erlang distribution has the least r-value among all the acyclic PH distribution with a fixed number of phases, n, for all* $n \geq 1$.

**Proof.** We prove the lemma by induction on $n$.

*Base case* ($n = 1$): Any one-phase PH distribution is a mixture of $O$ and an exponential distribution, and the $r$-value is always $\frac{3}{2}$.

*Inductive case*: Suppose that the lemma holds for $n \leq k$. We show that the lemma holds for $n = k + 1$ as well.

Consider any $(k + 1)$-phase acyclic PH distribution, $G$, which is not an Erlang distribution. We first show that there exists a PH distribution, $F_1$, with $r^{F_1} \leq r^G$ such that $F_1$ is the convolution of an exponential distribution, $X$, and a $k$-phase PH distribution, $Y$. The key idea is to see any PH distribution as a mixture of PH distributions whose initial probability vectors, $\vec{\tau}$, are base vectors. For example, the three-phase PH distribution, $G$, in Fig. 1, can be seen as a mixture of $O$ and the three 3-phase PH distribution, $G_i$ ($i = 1, 2, 3$), whose parameters are $\vec{\tau}^{G_1} = (1, 0, 0)$, $\vec{\tau}^{G_2} = (0, 1, 0)$, $\vec{\tau}^{G_3} = (0, 0, 1)$, and $\mathbf{T}^{G_1} = \mathbf{T}^{G_2} = \mathbf{T}^{G_3} = \mathbf{T}^G$. Proposition 2 and Lemma 2 imply that there exists $i \in \{1, 2, 3\}$ such that $r^{G_i} \leq r^G$. Without loss of generality, let $r^{G_1} \leq r^G$ and let $F_1 = G_1$; thus, $r^{F_1} \leq r^G$. Note that $F_1$ is the convolution of an exponential distribution, $X$, and a $k$-phase PH distribution, $Y$.

Next we show that if $F_1$ is not an Erlang distribution, then there exists a PH distribution, $F_2$, with no greater $r$-value (i.e. $r^{F_2} \leq r^{F_1}$). Let $Z$ be a mixture of $O$ and an Erlang-$k$ distribution, $E_k$, (i.e. $Z(\cdot) = pO(\cdot) + (1 - p)E_k(\cdot)$), where $p$ is chosen such that $\mu_1^Z = \mu_1^Y$ and $m_2^Z = m_2^Y$. There always exists such a $Z$, since the Erlang-$k$ distribution has the least $m_2$ among all the PH distributions (in particular $m_2^{E_k} \leq m_2^Y$) and $m_2$ is an increasing function of $p$ ($m_2^Z = m_2^{E_k}/(1 - p)$). Also, observe that, by Proposition 2 and the inductive hypothesis, $r^Z \leq r^Y$. Let $F_2$ be the convolution of $X$ and $Z$, i.e. $F_2(\cdot) = X(\cdot) * Z(\cdot)$. We prove that $r^{F_2} \leq r^{F_1}$. Let $y = \mu_1^Y/\mu_1^X$. Then,

$$
\begin{aligned}
r^{F_1} &= \frac{(r^X(m_2^X)^2 + 3m_2^X y + 3m_2^Y y^2 + r^Y(m_2^Y)^2 y^3)(1 + y)}{(m_2^X + 2y + m_2^Y y^2)^2} \\
&\geq \frac{(r^X(m_2^X)^2 + 3m_2^X y + 3m_2^Z y^2 + r^Z(m_2^Z)^2 y^3)(1 + y)}{(m_2^X + 2y + m_2^Z y^2)^2} = r^{F_2},
\end{aligned}
$$

where the inequality follows from $\mu_1^Z = \mu_1^Y$, $m_2^Z = m_2^Y$, and $r^Z \leq r^Y$.

Finally, we show that an Erlang distribution has the least $r$-value. $F_2$ is the convolution of $X$ and $Z$, and it can also be seen as a mixture of $X$ and a distribution, $F_3$, where $F_3(\cdot) = X(\cdot) * E_k(\cdot)$. Thus, by Lemma 2, at least one of $r^X \leq r^{F_2}$ and $r^{F_3} \leq r^{F_2}$ holds. When $r^X \leq r^{F_2}$, the $r$-value of the Erlang-$(k + 1)$ distribution, $r^{E_{k+1}}$, is smaller than $r^{F_2}$, since $r^{E_{k+1}} < r^X \leq r^{F_2}$. When $r^X > r^{F_2}$ (and hence $r^{F_3} \leq r^{F_2}$), $r^{E_{k+1}} \leq r^{F_3} \leq r^{F_2}$ can be proved by showing that $r^{F_3}$ is minimized when $\mu_1^X = \mu_1^{E_k}/k$. Let $y = \mu_1^X/\mu_1^{E_k}$.

Then,

$$r^{F_3} = \frac{(r^{E_k}(m_2^{E_k})^2 + 3m_2^{E_k}y + 6y^2 + 6y^3)(1+y)}{m_2^{E_k} + 2y + 2y^2},$$

where $r^{E_k} = (k+2)/(k+1)$ and $m_2^{E_k} = (k+1)/k$. Therefore,

$$\frac{\partial r^{F_3}}{\partial y} = \frac{2k(k+1)(6ky^2 + 6ky + k - 1)}{\left(\frac{k+1}{k} + 2y + 2y^2\right)}\left(y - \frac{1}{k}\right).$$

Since $k > 1$, $r^{F_3}$ is minimized at $y = 1/k$. □

## 4. EC distribution

The purpose of this section is two-fold: to provide a detailed characterization of the EC distribution, and to discuss a narrowed-down subset of the EC distributions with only five free parameters ($\lambda_Y$ is fixed) which we will use in our moment matching algorithms. Both results are summarized in Theorem 3.

To motivate the theorem in this section, suppose one is trying to match the first three moments of a given distribution, $G$, to a distribution, $P$, which is the convolution of exponential distributions (possibly with different rates) and a two-phase Coxian$^+$ PH distribution. If $G$ has sufficiently high second and third moments, then a two-phase Coxian$^+$ PH distribution alone suffices and we need no exponential distributions (recall Theorem 1). If the variability of $G$ is lower, however, we might try appending an exponential distribution to the two-phase Coxian$^+$ PH distribution. If that does not suffice, we might append two exponential distributions to the two-phase Coxian$^+$ PH distribution. Thus, if $G$ has very low variability, we might be forced to use many phases to get the variability of $P$ to be low enough. Therefore, to minimize the number of phases in $P$, it seems desirable to choose the rates of the exponential distributions so that the overall variability of $P$ is minimized. One could express the appending of each exponential distribution as a "function" whose goal is to reduce the variability of $P$ yet further.

**Definition 12.** Let $X$ be an arbitrary distribution. **Function** $\phi$ maps $X$ to $\phi(X)$ such that $\phi(X) = Y * X$, the convolution of $Y$ and $X$, where $Y$ is an exponential distribution whose rate, $\lambda_Y$, is chosen so that the normalized second moment of $\phi(X)$ is minimized. Also, $\phi^i(X) = \phi(\phi^{i-1}(X))$ refers to the distribution obtained by applying function $\phi$ to $\phi^{i-1}(X)$ for integers $i \geq 1$, where $\phi^0(X) = X$.

Observe that, when $X$ is a $k$-phase PH distribution, $\phi^N(X)$ is a $(k+N)$-phase PH distribution.

In theory, function $\phi$ allows each successive exponential distribution which is appended to have a different rate. Surprisingly, however, the following theorem shows that if the exponential distribution $Y$ being appended by function $\phi$ is chosen so as to minimize the normalized second moment of $\phi(X)$ (as specified by the definition), then the rate of each successive $Y$ is always *the same* and is defined by the simple formula shown in the theorem below. The theorem also characterizes the normalized moments of $\phi^i(X)$.

**Theorem 3.** *Let* $\phi^i(X) = Y_i * \phi^{i-1}(X)$, *and let* $\lambda_{Y_i}$ *be the rate of the exponential distribution* $Y_i$ *for* $i = 1, \ldots, N$. *Then,*

$$\lambda_{Y_i} = \frac{1}{(m_2^X - 1)\mu_1^X} \tag{1}$$

*for* $i = 1, \ldots, N$. *The normalized moments of* $Z_N = \phi^N(X)$ *are*:

$$m_2^{Z_N} = \frac{(m_2^X - 1)(N+1) + 1}{(m_2^X - 1)N + 1},$$

$$m_3^{Z_N} = \frac{m_2^X m_3^X}{((m_2^X - 1)(N+1) + 1)((m_2^X - 1)N + 1)^2} + \frac{(m_2^X - 1)N(3m_2^X + (m_2^X - 1)(m_2^X + 2)(N+1) + (m_2^X - 1)^2(N+1)^2)}{((m_2^X - 1)(N+1) + 1)((m_2^X - 1)N + 1)^2}.$$

**Proof.** *We first characterize* $Z = \phi(X) = Y * X$, where $X$ is an arbitrary distribution with a finite third moment and $Y$ is an exponential distribution. The normalized second moment of $Z$ is

$$m_2^Z = \frac{m_2^X + 2y + 2y^2}{(1+y)^2},$$

where $y = \mu_1^Y / \mu_1^X$. Observe that $m_2^Z$ is minimized when $y = m_2^X - 1$, i.e. when

$$\mu_1^Y = (m_2^X - 1)\mu_1^X. \tag{2}$$

Observe that when $\mu_1^Y$ is set at this value, $m_2^Z$ and $m_3^Z$ satisfy:

$$m_2^Z = 2 - \frac{1}{m_2^X} \quad \text{and} \quad m_3^Z = \frac{1}{m_2^X(2m_2^X - 1)} m_3^X + \frac{3(m_2^X - 1)}{m_2^X}.$$

*We next characterize* $Z_i = \phi^i(X) = Y_i * \phi^{i-1}(X)$ *for* $2 \le i \le N$. By the above expression on $m_2^Z$ and $m_3^Z$, the second part of the theorem on the normalized moments of $Z_N$ follows from solving the following recurrence equations (where we use $b_i$ to denote $m_2^{\phi^i(X)}$ and $B_i$ to denote $m_3^{\phi^i(X)}$):

$$b_{i+1} = 2 - \frac{1}{b_i} \quad \text{and} \quad B_{i+1} = \frac{B_i}{b_i(2b_i - 1)} + \frac{3(b_i - 1)}{b_i}.$$

The solutions for these recurrence equations are

$$b_{i+1} = \frac{(b_1 - 1)(i+1) + 1}{(b_1 - 1)i + 1},$$

$$B_{i+1} = \frac{b_1 B_1 + (b_1 - 1)i(3b_1 + (b_1 - 1)(b_1 + 2)(i+1) + (b_1 - 1)^2(i+1)^2)}{((b_1 - 1)(i+1) + 1)((b_1 - 1)i + 1)^2}$$

for all $i \ge 0$. These solutions can be verified by substitution. This completes the proof of the second part of the theorem.

The first part of the theorem on $\lambda_{Y_i}$ is proved by induction. When $i = 1$, (1) follows from (2). Assume that (1) holds for $i = 1, \ldots, k$. Let $Z_k = \phi^k(X)$. By the second part of the theorem, which is proved above,

$$m_2^{Z_k} = \frac{(m_2^X - 1)(k+1) + 1}{(m_2^X - 1)k + 1}.$$

Thus, by (2),

$$\frac{1}{\lambda_{Y_{k+1}}} = \mu_1^{Y_{k+1}} = (m_2^{Z_k} - 1)\mu_1^{Z_k} = (m_2^X - 1)\mu_1^X. \quad \square$$

**Corollary 1.** *If* $X \in \{F | 2 < m_2^F\}$, *then* $Z = \phi^N(X) \in \{F | (N+2)/(N+1) < m_2^F < (N+1)/N\}$.

**Proof.** By Theorem 3, $m_2^Z$ is a continuous and monotonically increasing function of $m_2^X$. Thus, the infimum and the supremum of $m_2^Z$ are given by evaluating $m_2^Z$ at the infimum and the supremum, respectively, of $m_2^X$. When $m_2^X \to 2$, $m_2^Z \to (N+2)/(N+1)$. When $m_2^X \to \infty$, $m_2^Z \to (N+1)/N$. $\square$

Corollary 1 suggests the number, $N$, of times that function $\phi$ must be applied to $X$ to bring $m_2^Z$ into a desired range, given the value of $m_2^X$.

*Concluding remarks*: Theorem 3 implies that the parameter $\lambda_Y$ of the EC distribution can be fixed without excluding the distributions of lowest variability from the set of EC distributions. Below, we constrain $\lambda_Y$ as follows:

$$\lambda_Y = \frac{1}{(m_2^X - 1)\mu_1^X}, \tag{3}$$

and derive closed form representations of the remaining free parameters $(n, p, \lambda_{X1}, \lambda_{X2}, p_X)$, where these free parameters will determine $m_2^X$ and $\mu_1^X$, which in turn gives $\lambda_Y$ by (3). Obviously, at least three degrees of freedom are necessary to match three moments. As we will see, the additional degrees of freedom allow us to accept all input distributions in $\mathcal{PH}_3$ and to use a smaller number of phases.

## 5. Simple closed form solution

The `Simple` solution is the simplest among our three closed form solutions, and the `Complete` and `Positive` solutions will be built upon the `Simple` solution. Before, presenting the `Simple` solution, we first classify the input distributions. This classification is used, in particular, to determine the number of phases used in the `Simple` solution.

### 5.1. Preliminaries

Set $\mathcal{T}^{(n)}$, which is used to characterize set $\mathcal{S}^{(n)}$, gives us a sense of how many phases are necessary to well represent a given distribution. It turns out that it is useful to divide set $\mathcal{T}^{(n)}$ into smaller subsets to describe the closed form solutions compactly. Roughly speaking, we divide the set $\mathcal{T}^{(n)} \setminus \mathcal{T}^{(n-1)}$ into three subsets, $\mathcal{U}_{n-1}$, $\mathcal{M}_{n-1}$, and $\mathcal{L}_{n-1}$ (see Fig. 4). More formally,

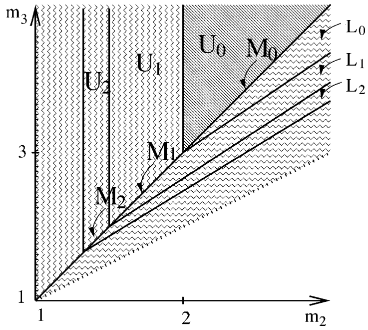



Fig. 4. A classification of distributions. The dotted lines delineate the set of all nonnegative distributions $G$ ($m_3^G \geq m_2^G \geq 1$).

**Definition 13.** We define $\mathcal{U}_i$, $\mathcal{M}_i$, and $\mathcal{L}_i$ as follows:

$$\mathcal{U}_0 = \{F | m_2^F > 2 \text{ and } m_3^F > 2m_2^F - 1\}, \qquad \mathcal{M}_0 = \{F | m_2^F > 2 \text{ and } m_3^F = 2m_2^F - 1\},$$

$$\mathcal{L}_0 = \left\{ F \left| \frac{3}{2} m_2^F < m_3^F < 2m_2^F - 1 \right. \right\},$$

and

$$\mathcal{U}_i = \left\{ F \left| \frac{i+2}{i+1} < m_2^F < \frac{i+1}{i} \text{ and } m_3^F > 2m_2^F - 1 \right. \right\},$$

$$\mathcal{M}_i = \left\{ F \left| \frac{i+2}{i+1} < m_2^F < \frac{i+1}{i} \text{ and } m_3^F = 2m_2^F - 1 \right. \right\},$$

$$\mathcal{L}_i = \left\{ F \left| \frac{i+3}{i+2} m_2^F < m_3^F < \frac{i+2}{i+1} m_2^F \text{ and } m_3^F < 2m_2^F - 1 \right. \right\}$$

for nonnegative integers $i$. Also, let $\mathcal{U}^+ = \cup_{i=1}^{\infty} \mathcal{U}_i$, $\mathcal{M}^+ = \cup_{i=1}^{\infty} \mathcal{M}_i$, $\mathcal{L}^+ = \cup_{i=1}^{\infty} \mathcal{L}_i$, $\mathcal{U} = \mathcal{U}_0 \cup \mathcal{U}^+$, $\mathcal{M} = \mathcal{M}_0 \cup \mathcal{M}^+$, and $\mathcal{L} = \mathcal{L}_0 \cup \mathcal{L}^+$. Further, let

$$\hat{\mathcal{U}} = \{F | m_3^F > 2m_2^F - 1\} \supset \mathcal{U}, \qquad \hat{\mathcal{M}} = \{F | m_3^F = 2m_2^F - 1\} \supset \mathcal{M},$$

$$\hat{\mathcal{L}} = \{F | m_3^F < 2m_2^F - 1\} \supset \mathcal{L}.$$

Observe that $\hat{\mathcal{U}}$ includes both $\mathcal{U}$ and borders between $\mathcal{U}_i$ and $\mathcal{U}_{i+1}$, for $i \geq 0$, that are not included in $\mathcal{U}$.

The sets $\hat{\mathcal{U}}$, $\hat{\mathcal{M}}$, and $\hat{\mathcal{L}}$ provide a classification of distributions into three categories such that, for any distribution $X$, $X$ and $\phi(X)$ lie in the same category.

**Lemma 4.** *Let $Z_N = \phi^N(X)$ for integers $N \geq 1$. If $X \in \hat{\mathcal{U}}$ (respectively, $X \in \hat{\mathcal{M}}$, $X \in \hat{\mathcal{L}}$), then $Z_N \in \hat{\mathcal{U}}$ (respectively, $Z_N \in \hat{\mathcal{M}}$, $Z_N \in \hat{\mathcal{L}}$).*

**Proof.** We prove the case when $N = 1$. The lemma then follows by induction. Let $Z = \phi(X)$. By Theorem 3, $m_2^X = 1/(2 - m_2^Z)$, and

$$m_3^Z > \text{(respectively, =, and <)} \ \frac{2m_2^X - 1}{m_2^X(2m_2^X - 1)} + 3\frac{m_2^X - 1}{m_2^X} > \text{(respectively, =, and <)} \ 2m_2^Z - 1,$$

where the last equality follows from $m_2^X = 1/(2 - m_2^Z)$. $\square$

By Corollary 1 and Lemma 4, we get the following corollary.

**Corollary 2.** *Let $Z_N = \phi^N(X)$ for $N \geq 0$. If $X \in \mathcal{U}_0$ (respectively, $X \in \mathcal{M}_0$), then $Z_N \in \mathcal{U}_N$ (respectively, $Z_N \in \mathcal{M}_N$).*

The corollary implies that for any $G \in \mathcal{U}_N \cup \mathcal{M}_N$, $G$ can be well represented by an $(N+2)$-phase EC distribution with no mass probability at zero ($p = 1$), because, for any $X \in \mathcal{U}_0 \cup \mathcal{M}_0$, $X$ can be well represented by a two-phase Coxian$^+$ PH distribution, and hence $Z_N = \phi^N(X)$ can be well represented by an $(N+2)$-phase EC distribution. (Recall $\mathcal{U}_N, \mathcal{M}_N, \mathcal{L}_N \subset \mathcal{T}_{N+1}$.) Below, it will also be shown that for any $G \in \mathcal{L}_N$, $G$ can be well represented by an $(N+2)$-phase EC distribution with positive mass probability at zero ($p < 1$).

By Corollary 2, it is relatively easy to provide a closed form solution for the parameters $(n, p, \lambda_{X1}, \lambda_{X2}, p_X)$ of an EC distribution, $Z$, so that a given distribution is well represented by $Z$. Essentially, one just needs to find an appropriate $N$ and solve $Z = \phi^N(X)$ for $X$ in terms of normalized moments, which is immediate since $N$ is given by Corollary 1 and the normalized moments of $X$ can be obtained from Theorem 3.

### 5.2. The `Simple` solution

We are now ready to present the `Simple` solution. The `Simple` solution assumes that $G \in \mathcal{PH}_3^-$, where $\mathcal{PH}_3^- = \mathcal{U} \cup \mathcal{M} \cup \mathcal{L}$. Observe $\mathcal{PH}_3^-$ includes almost all distributions in $\mathcal{PH}_3$. Only the borders between the $\mathcal{U}_i$'s, $\mathcal{M}_i$'s, and $\mathcal{L}_i$'s are not included. The solution differs according to the classification of the input distribution $G$. When $G \in \mathcal{U}_0 \cup \mathcal{M}_0$, a two-phase Coxian$^+$ PH distribution suffices to match the first three moments. When $G \in \mathcal{U}^+ \cup \mathcal{M}^+$, $G$ is well represented by an EC distribution with $p = 1$. When $G \in \mathcal{L}$, $G$ is well represented by an EC distribution with $p < 1$.

(i) *If $G \in \mathcal{U}_0 \cup \mathcal{M}_0$* (see Fig. 5(i)), then a two-phase Coxian$^+$ PH distribution suffices to match the first three moments, i.e., $p = 1$ and $n = 2$ ($N = 0$). The parameters $(\lambda_{X1}, \lambda_{X2}, p_X)$ of the two-phase

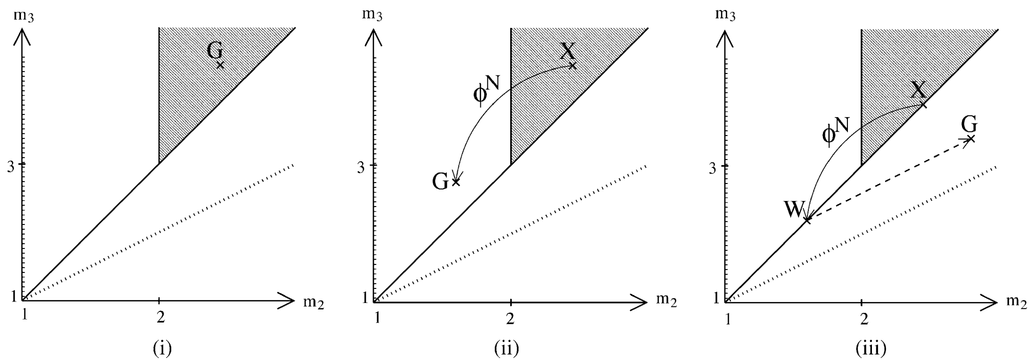



Fig. 5. Ideas in the `Simple` solution. Let $G$ be the input distribution. (i) If $G \in \mathcal{U}_0 \cup \mathcal{M}_0$, $G$ is well represented by a two-phase Coxian$^+$ PH distribution $X$. (ii) If $G \in \mathcal{U}^+ \cup \mathcal{M}^+$, $G$ is well represented by $\phi^N(X)$, where $X$ is a two-phase Coxian$^+$ PH distribution. (iii) If $G \in \mathcal{L}$, $G$ is well represented by $Z$, where $Z$ has a distribution function $p\phi^N(X)(\cdot) + (1-p)O(\cdot)$.

Coxian$^+$ PH distribution are chosen as follows [25]:

$$\lambda_{X1} = \frac{u + \sqrt{u^2 - 4v}}{2\mu_1^G}, \qquad \lambda_{X2} = \frac{u - \sqrt{u^2 - 4v}}{2\mu_1^G}, \quad \text{and} \quad p_X = \frac{\lambda_{X2}(\lambda_{X1}\mu_1^G - 1)}{\lambda_{X1}}, \tag{4}$$

where

$$u = \frac{6 - 2m_3^G}{3m_2^G - 2m_3^G} \quad \text{and} \quad v = \frac{12 - 6m_2^G}{m_2^G(3m_2^G - 2m_3^G)}. \tag{5}$$

(ii) *If* $G \in \mathcal{U}^+ \cup \mathcal{M}^+$ (see Fig. 5(ii)), Corollary 1 specifies the number of phases needed:

$$n = \min\left\{k \,\middle|\, m_2^G > \frac{k}{k-1}\right\} = \left\lfloor \frac{m_2^G}{m_2^G - 1} + 1 \right\rfloor. \tag{6}$$

Let $N = n - 2$. Next, we find the two-phase Coxian$^+$ PH distribution $X \in \mathcal{U}_0 \cup \mathcal{M}_0$ such that $G$ is well represented by $Z = \phi^N(X)$. By Theorem 3, this can be achieved by setting

$$m_2^X = \frac{(n-3)m_2^G - (n-2)}{(n-2)m_2^G - (n-1)}, \qquad m_3^X = \frac{\beta m_3^G - \alpha}{m_2^X}, \quad \text{and} \quad \mu_1^X = \frac{\mu_1^G}{(n-2)m_2^X - (n-3)},$$

where

$$\alpha = (n-2)(m_2^X - 1)(n(n-1)(m_2^X)^2 - n(2n-5)m_2^X + (n-1)(n-3)),$$
$$\beta = ((n-1)m_2^X - (n-2))((n-2)m_2^X - (n-3))^2.$$

Thus, we set $p = 1$, and the parameters $(\lambda_{X1}, \lambda_{X2}, p_X)$ of $X$ are given by case (i), using $m_2^X$, $m_3^X$, and $\mu_1^X$, specified above.

(iii) *If* $G \in \mathcal{L}$ (see Fig. 5(iii)), then let

$$p = \frac{1}{2m_2^G - m_3^G}, \qquad m_2^W = pm_2^G, \qquad m_3^W = pm_3^G, \quad \text{and} \quad \mu_1^W = \frac{\mu_1^G}{p}.$$

Observe that $p$ satisfies $0 \le p < 1$. Also, since $W$ is in $\mathcal{M}$, $W$ can be chosen as an EC distribution with no mass probability at zero. If $W \in \mathcal{M}_0$, the parameters of $W$ are provided by case (i), using $m_2^W$, $m_3^W$, and $\mu_1^W$, specified above. If $W \in \mathcal{M}^+$, the parameters of $W$ are provided by case (ii), using $m_2^W$, $m_3^W$, and $\mu_1^W$, specified above. $G$ is then well represented by distribution $Z$, where $Z(\cdot) = pW(\cdot) + (1-p)O(\cdot)$.

### *5.3. Analyzing the number of phases required*

The number of phases used in the `Simple` solution is characterized by the following theorem.

**Theorem 4.** *The* `Simple` *solution uses at most* $\mathrm{OPT}(G) + 1$ *phases to well represent a distribution* $G \in \mathcal{PH}_3^-$.

**Proof.** Since $S^{(n)} \subset \mathcal{T}^{(n)}$ (by Lemma 1), it suffices to prove that if a distribution $G \in \mathcal{T}^{(n)}$, then at most $n+1$ phases are needed. If $G \in \mathcal{T}^{(n)} \cap (\mathcal{U} \cup \mathcal{M})$, then $m_2^G > (n+1)/n$. Also, if $G \in \mathcal{T}^{(n)} \cap \mathcal{L}$, then

$$m_2^W = \frac{m_2^G}{2m_2^G - 1} > \frac{n+1}{n}.$$

Thus, by (6), the EC distribution provided by the `Simple` solution has at most $n+1$ phases. □

## 6. Complete closed form solution

The `Complete` solution improves upon the `Simple` solution in the sense that it is defined for all the input distributions $G \in \mathcal{PH}_3$. Fig. 6 shows an implementation of the `Complete` solution. Below, we elabo-

$(n, p, \lambda_Y, \lambda_{X1}, \lambda_{X2}, p_X) = \texttt{Complete}(\mu_1^G, \mu_2^G, \mu_3^G)$

Input: the first three moments of a distribution $G$: $\mu_1^G$, $\mu_2^G$, and $\mu_3^G$.

Output: parameters of the EC distribution, $(n, p, \lambda_Y, \lambda_{X1}, \lambda_{X2}, p_X)$

1. $m_2^G = \frac{\mu_2^G}{(\mu_1^G)^2}$; $\quad m_3^G = \frac{\mu_3^G}{\mu_1^G \mu_2^G}$.

2. $p = \begin{cases} \frac{(m_2^G)^2 + 2m_2^G - 1}{2(m_2^G)^2} & \text{if } m_3^G > 2m_2^G - 1, \text{ and } \frac{1}{m_2^G - 1} \text{ is an integer,} \\ \frac{1}{2m_2^G - m_3^G} & \text{if } m_3^G < 2m_2^G - 1, \\ 1 & \text{otherwise.} \end{cases}$

3. $\mu_1^W = \frac{\mu_1^G}{p}$; $\quad m_2^W = p m_2^G$; $\quad m_3^W = p m_3^G$.

4. $n = \begin{cases} \left\lceil \frac{m_2^W}{m_2^W - 1} \right\rceil - 1 & \text{if } m_3^W = 2m_2^W - 1, \text{ and } m_2^W \le 2, \\ \left\lfloor \frac{m_2^W}{m_2^W - 1} + 1 \right\rfloor & \text{otherwise.} \end{cases}$

(When $n = 1$, set $p_X = 0$ and $\lambda_{X1} = \frac{1}{\mu_1^W}$, and exit. Below, we assume $n \ge 2$.)

5. $m_2^X = \frac{(n-3)m_2^W - (n-2)}{(n-2)m_2^W - (n-1)}$; $\quad \mu_1^X = \frac{\mu_1^W}{(n-2)m_2^X - (n-3)}$.

6. $\alpha = (n-2)(m_2^X - 1)\left(n(n-1)(m_2^X)^2 - n(2n-5)m_2^X + (n-1)(n-3)\right)$.

7. $\beta = \left((n-1)m_2^X - (n-2)\right)\left((n-2)m_2^X - (n-3)\right)^2$.

8. $m_3^X = \frac{\beta m_3^W - \alpha}{m_2^X}$.

9. $u = \begin{cases} 1 & \text{if } 3m_2^X = 2m_3^X \\ \frac{6 - 2m_3^X}{3m_2^X - 2m_3^X} & \text{otherwise} \end{cases}$; $\quad v = \begin{cases} 0 & \text{if } 3m_2^X = 2m_3^X \\ \frac{12 - 6m_2^X}{m_2^X(3m_2^X - 2m_3^X)} & \text{otherwise} \end{cases}$.

10. $\lambda_{X1} = \frac{u + \sqrt{u^2 - 4v}}{2\mu_1^X}$; $\lambda_{X2} = \frac{u - \sqrt{u^2 - 4v}}{2\mu_1^X}$; $p_X = \frac{\lambda_{X2}(\lambda_{X1}\mu_1^X - 1)}{\lambda_{X1}}$; $\lambda_Y = \frac{1}{(m_2^X - 1)\mu_1^X}$.

Fig. 6. An implementation of the `Complete` solution, defined for $G \in \mathcal{PH}_3$.

rate on the `Complete` solution, and prove an upper bound on the number of phases used in the `Complete` solution.

### 6.1. The `Complete` solution

Consider an arbitrary distribution $G \in \mathcal{PH}_3$. Our approach consists of two steps, the first of which involves constructing a baseline EC distribution, and the second of which involves reducing the number of phases in this baseline solution. If $G \in \mathcal{PH}_3^-$, then the baseline solution used is simply given by the `Simple` solution. Also, if $G \notin \mathcal{PH}_3^-$ but $G \in \hat{\mathcal{L}} \cup \hat{\mathcal{M}}$, then it turns out that the `Simple` solution could be defined for this $G$, and this gives the baseline solution. If $G \notin \mathcal{PH}_3^-$ but $G \in \hat{\mathcal{U}}$, then to obtain the baseline EC distribution we first find a distribution $W \in \mathcal{PH}_3^-$ such that $r^W = r^G$ and $m_2^W < m_2^G$ and then set $p$ such that $G$ is well represented by distribution $Z$, where $Z(\cdot) = pW(\cdot) + (1-p)O(\cdot)$ (see Fig. 7(a)). The parameters of the EC distribution that well represents $W$ are then obtained by the `Simple` solution.

To reduce the number of phases used in the baseline EC distribution, we exploit the subset of two-phase Coxian$^+$ PH distributions that are not used in the `Simple` solution. The `Simple` solution is based on the fact that a distribution $X$ is well represented by a two-phase Coxian$^+$ PH distribution when $X \in \mathcal{U}_0 \cup \mathcal{M}_0$. In fact, a wider range of distributions are well represented by the set of two-phase Coxian$^+$ PH distributions. In particular, if $X$ is in set $\mathcal{S} = \{F | 3/2 \le m_2^X \le 2 \text{ and } m_3^X = 2m_2^X - 1\}$, then $X$ is well represented by a two-phase Coxian$^+$ PH distribution (see Fig. 7(a)). By exploiting two-phase Coxian$^+$ PH distributions in $\mathcal{S} \cup \mathcal{U}_0 \cup \mathcal{M}_0$, the `Complete` solution reduces the number of phases used. The above ideas lead to the following solution:

(i) *If* $G \in \hat{\mathcal{U}} \cap \mathcal{PH}^-$, then the `Simple` solution provides the parameters $(n, p, \mu_{X1}, \mu_{X2}, p_X)$.
(ii) *If* $G \in \hat{\mathcal{U}} \cap (\mathcal{PH}^-)^c$ (see Fig. 7(a)), where $(\mathcal{PH}^-)^c$ denotes the complement of $\mathcal{PH}^-$, then let

$$n = \frac{2m_2^G - 1}{m_2^G - 1}, \tag{7}$$

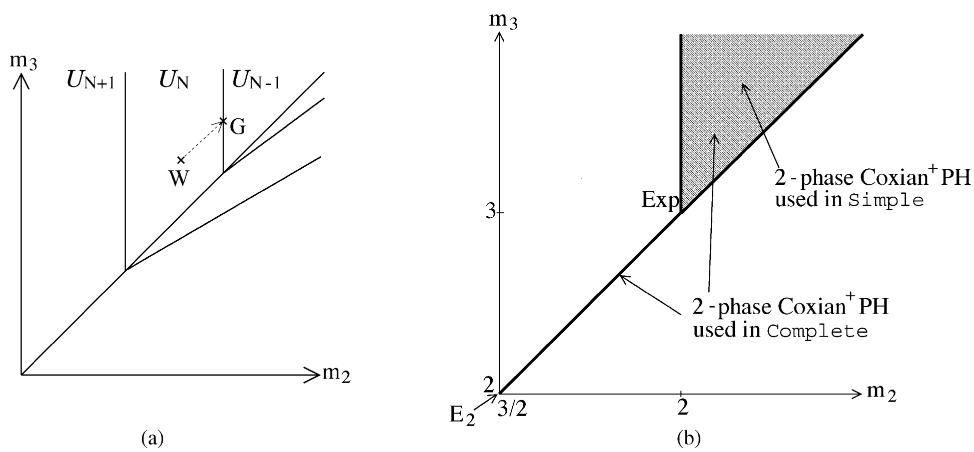



Fig. 7. Ideas in the `Complete` solution. (a) A distribution $G$ not in $\mathcal{PH}_3^-$ is well represented by an EC distribution $W$ mixed with $O$. (b) The set of two-phase Coxian$^+$ PH distributions used is extended.

and

$$m_2^W = \frac{1}{2}\left(\frac{n-1}{n-2} + \frac{n}{n-1}\right), \qquad m_3^W = \frac{m_3^G}{m_2^G} m_2^W, \quad \text{and} \quad \mu_1^W = \frac{\mu_1^G}{p_W},$$

where $p_W = m_2^W / m_2^G$. $G$ is then well represented by $Z$, where $Z(\cdot) = p_W W(\cdot) + (1 - p_W)O(\cdot)$, where $W$ is an EC distribution with normalized moments $m_2^W$ and $m_3^W$ and mean $\mu_1^W$. The parameters $(n, \mu_{X1}, \mu_{X2}, p_X)$ of $W$ are provided by the `Simple` solution. Also, set $p = p_W$, since $W$ has no mass probability at zero.

(iii) *If* $G \in \hat{\mathcal{M}} \cup \hat{\mathcal{L}}$, then the `Simple` solution provides the parameters $(n, p, \mu_{X1}, \mu_{X2}, p_X)$, except that (6) is replaced by

$$n = \begin{cases} \left\lceil \frac{m_2^G}{m_2^G - 1} \right\rceil - 1 & \text{if } m_2^G \le 2, \\ 2 & \text{otherwise.} \end{cases} \tag{8}$$

The next theorem guarantees that parameters obtained with the reduced $n$ are still feasible.

**Theorem 5.** *If $X$ is in set $\{F | 3/2 \le m_2^F \le 2$ and $m_3^F = 2m_2^F - 1\}$, then $Z = \phi^N(X)$ is in set $\{F | (N+1)/N \le m_2^F \le N/(N-1)$ and $m_3^F = 2m_2^F - 1\}$.*

**Proof.** By Theorem 3, $m_2^Z$ is a continuous and monotonically increasing function of $m_2^X$, Thus,

$$\frac{N+1}{N} \le m_2^Z \le \frac{N}{N-1}$$

follows by simply evaluating $m_2^Z$ at the lower and upper bound of $m_2^X$. $m_3^Z = 2m_2^Z - 1$ follows from Lemma 4. □

*6.2. Analyzing the number of phases required*

The number of phases used in the `Complete` solution is characterized by the following theorem.

**Theorem 6.** *The* `Complete` *solution uses at most* $\text{OPT}(G) + 1$ *phases to well represent any distribution* $G \in \mathcal{PH}_3$.

**Proof.** If $G \in \mathcal{PH}_3^-$, the number of phases used in the `Complete` solution is at most that used in the `Simple` solution, i.e. $\le \text{OPT}(G) + 1$. Thus, it suffices to prove that if a distribution $G \in (\mathcal{T}^{(n)} \setminus \mathcal{T}^{(n-1)}) \cap (\mathcal{PH}_3^-)^c$, then at most $n+1$ phases are needed (recall $\mathcal{S}^{(n)} \subset \mathcal{T}^{(n)}$ by Lemma 1). If $G \in (\mathcal{T}^{(n)} \setminus \mathcal{T}^{(n-1)}) \cap (\mathcal{PH}_3^-)^c \cap \hat{\mathcal{U}}$, then $m_2^G = n/(n-1)$. Thus, by (7), the number of phases used is $n+1$. If $G \in (\mathcal{T}^{(n)} \setminus \mathcal{T}^{(n-1)}) \cap (\mathcal{PH}_3^-)^c \cap (\hat{\mathcal{M}} \cup \hat{\mathcal{L}})$, then the number of phases given by (8) is exactly $n$, except for input distribution $G$ with $m_2^G \ge 2$ and $r^G = 3/2$ (i.e. exponential distribution possibly mixed with $O$) for which (8) gives 1. Thus, the number of phases used in the `Complete` solution for $G \in (\mathcal{PH}_3^-)^c \cap (\hat{\mathcal{M}} \cup \hat{\mathcal{L}})$ is $\text{OPT}(G)$. □

Theorem 6 implies that any distribution in $\mathcal{S}^{(n)}$ is well represented by an EC distribution of $n+1$ phases. In the proof, we prove a stronger property that any distribution in $\mathcal{T}^{(n)}$, which is a superset of $\mathcal{S}^{(n)}$, is well represented by an EC distribution of $n+1$ phases, which implies the following corollary.

**Corollary 3.** *For all integers $n \ge 2$, $\mathcal{T}^{(n)} \subset \mathcal{S}^{(n+1)}$.*

## 7. Positive closed form solution

The `Simple` and `Complete` solutions can have mass probability at zero (i.e. $p < 1$). In some applications, mass probability at zero is not an issue. Such applications include approximating busy period distributions in the analysis of multiserver systems [17] and approximating shortfall distributions in inventory system analysis [26,27]. However, there are also applications where a mass probability at zero increases the computational complexity or even makes the analysis intractable. For example, a PH/PH/1/FCFS queue can be analyzed efficiently via matrix analytic methods when the PH distributions have no mass probability at zero; however, no simple analytical solution is known when the PH distributions have nonzero mass probability at zero.

The `Positive` solution is built upon the `Complete` solution, but does not have mass probability at zero. The key idea in the design of the `Positive` solution is to match the input distribution either by a mixture of an EC distribution (with no mass probability at zero) and an exponential distribution, or by the convolution of an EC distribution (with positive mass probability at zero) and an exponential distribution. The use of these types of distributions makes intuitive sense, since they can approximate the EC distribution with mass probability at zero arbitrarily closely by letting the rate of the exponential distributions approach infinity. Therefore, in this section, we extend the definition of the EC distribution and use the extended EC distribution to well represent the input distribution.

**Definition 14.** An extended EC distribution has a distribution function either of the form $pX(\cdot) + (1 - p)W(\cdot)$ or of the form $Z(\cdot) * X(\cdot)$, where $X$ is an EC distribution with no mass probability at zero; $Z$ and $W$ are exponential distributions.

Note that the parameter $n$ in an extended EC distribution denotes the number of phases in the EC portion of the extended EC distribution. Therefore, the total number of phases in the extended EC distribution is $n + 1$.

### *7.1. The* `Positive` *solution*

The `Positive` solution is defined for almost all the input distributions in $\mathcal{PH}_3$. Specifically, it is defined for all the distributions in

$$\mathcal{U} \cup \hat{\mathcal{M}} \cup \left\{ F \,\middle|\, r^F \neq \frac{3}{2} \text{ and } m_3^F < 2m_2^F - 1 \right\}.$$

Although the `Positive` solution is not defined for a very small set (with measure 0) of input distributions, this is not a problem in practice, since distributions lying in the very small subset can be perturbed to move out of the subset. Fig. 8 shows an implementation of the `Positive` solution. Below, we elaborate on the `Positive` solution, and prove an upper bound on the number of phases used in the `Positive` solution.

When the input distribution $G$ is in $\mathcal{U} \cup \hat{\mathcal{M}}$, the EC distribution produced by the `Complete` solution does not have a mass probability at zero. When $G$ is in $\{F | m_3^F < 2m_2^F - 1 \text{ and } r^F > \frac{3}{2}\}$, $G$ can be well represented by a two-phase Coxian$^+$ PH distribution, whose parameters are given by (4) and (5). Below, we focus on input distributions $G \in \{F | m_3^F < 2m_2^F - 1 \text{ and } r^F < \frac{3}{2}\}$.

$(n, p, \lambda_Y, \lambda_{X1}, \lambda_{X2}, p_X, \lambda_W, \lambda_Z) = \texttt{Positive}(\mu_1^G, \mu_2^G, \mu_3^G)$

If $G \in \mathcal{U} \cup \left\{F \mid m_3^F = 2m_2^F - 1\right\} \cup \left\{F \mid m_3^F < 2m_2^F - 1 \text{ and } m_2^F > 2 \text{ and } r^F > 3/2\right\}$,

use $\texttt{Complete}$. Otherwise,

1. $m_2^G = \frac{\mu_2^G}{(\mu_1^G)^2}$; $m_3^G = \frac{\mu_3^G}{\mu_1^G \mu_2^G}$; $r^G = \frac{m_3^G}{m_2^G}$; $k = \lfloor \frac{2m_2^G - m_3^G}{m_3^G - m_2^G} \rfloor$.

If $m_3^G \geq \frac{(k+1)m_2^G + (k+4)}{2(k+2)} m_2^G$,

2. $w = \frac{2 - m_2^G}{4(3/2 - r^G)}$; $p = \frac{(2 - m_2^G)^2}{(2 - m_2^G)^2 + 4(2m_2^G - 1 - m_3^G)}$; $m_2^X = 2w$;

$m_3^X = 2m_2^X - 1$; $\mu_1^X = \frac{\mu_1^G}{p + (1-p)w}$; $\lambda_W = \frac{1}{w\mu_1^X}$; $\lambda_Z = \infty$; go to 3.

If $r^G < \frac{(k+1)m_2^G + (k+4)}{2(k+2)}$ and $m_2^G = 2$,

2. $z = \frac{m_3^G - 2\frac{k+3}{k+2}}{3 - m_3^G}$; $m_2^X = 2(1+z)$; $m_3^X = \frac{k+3}{k+2} m_2^X$; $\mu_1^X = \frac{\mu_1^G}{1+z}$.

$\lambda_Z = \frac{1}{z\mu_1^X}$; $\lambda_W = \infty$; go to 3.

If $m_3^G < \frac{(k+1)m_2^G + (k+4)}{2(k+2)} m_2^G$ and $m_2^G \neq 2$,

2. $z = \frac{m_2^G\left((m_3^G - 3) - 2\frac{k+3}{k+2}(m_2^G - 2)\right) + m_2^G\sqrt{(m_3^G - 3)^2 + 8\frac{k+3}{k+2}(m_2^G - 2)(3/2 - r^G)}}{2\frac{k+3}{k+2}(m_2^G - 2)^2}$;

$m_2^X = (1+z)\left(m_2^G(1+z) - 2\right)$; $m_3^X = \frac{k+3}{k+2} m_2^X$; $\mu_1^X = \frac{\mu_1^G}{1+z}$.

$\lambda_Z = \frac{1}{z\mu_1^X}$; $\lambda_W = \infty$; go to 3.

3. $\mu_2^X = m_2^X(\mu_1^X)^2$; $\mu_3^X = m_3^X \mu_1^X \mu_2^X$.

4. $(n, p, \lambda_Y, \lambda_{X1}, \lambda_{X2}, p_X) = \texttt{Complete}(\mu_1^X, \mu_2^X, \mu_3^X)$

Fig. 8. An implementation of the `Positive` solution, defined for a distribution $G \in \mathcal{U} \cup \hat{\mathcal{M}} \cup \{F | r^F \neq 3/2 \text{ and } m_3^F < 2m_2^F - 1\}$.

We first consider the first approach of using a mixture of an EC distribution (with no mass probability at zero), *X*, and an exponential distribution, *W*, (i.e. $\lambda_Z = \infty$). Let

$$\hat{\mathcal{L}}_N = \left\{ F \,\middle|\, \frac{N+3}{N+2}\, m_2^F < m_3^F \leq \frac{N+2}{N+1} m_2^F \text{ and } m_3^F < 2m_2^F - 1 \right\}.$$

Given a distribution $G \in \hat{\mathcal{L}}_N$, we seek $m_2^X$, $m_3^X$, $0 < p < 1$, and $w > 0$ such that

$$\frac{N+2}{N+1} \leq m_2^X < \frac{N+1}{N}, \tag{9}$$

$$m_3^X = 2m_2^X - 1, \tag{10}$$

$$m_2^G = \frac{pm_2^X + 2(1-p)w^2}{(p + (1-p)w)^2}, \tag{11}$$

$$m_3^G = \frac{pm_2^X m_3^X + 6(1-p)w^3}{(p + (1-p)w)(pm_2^X + 2(1-p)w^2)}. \tag{12}$$

Note that (9) and (10) guarantee that we can find, via the Complete solution, an $(N+1)$-phase EC distribution, $X$, such that $X$ has no mass probability at zero and has normalized moments $m_2^X$ and $m_3^X$. Let $W$ be the exponential distribution with $\mu_1^W = w/\mu_1^X$. (11) and (12) guarantee that, by choosing $\mu_1^X$ appropriately, $G$ is well represented by a distribution $Y$, where $Y(\cdot) = pX(\cdot) + (1-p)W(\cdot)$.

The following lemma provides conditions on the input distribution for which the first approach is defined.

**Lemma 5.** *Suppose*

$$G \in \hat{\mathcal{L}}_N \quad \textit{and} \quad \frac{(N+1)m_2^G + (N+4)}{2(N+2)} \leq r^G$$

*for* $N \geq 1$ (*see* Fig. 9(*a*)). *Let*

$$w = \frac{2 - m_2^G}{4\left(\frac{3}{2} - r^G\right)}, \qquad p = \frac{(2 - m_2^G)^2}{(2 - m_2^G)^2 + 4(2m_2^G - 1 - m_3^G)},$$

$m_2^X = 2w$, *and* $m_3^X = 2m_2^X - 1$. *Then*, $w > 0$, $0 < p < 1$, *and conditions* (9)–(12) *are satisfied.*

**Proof.** It is easy to check, by substitution, that conditions (10)–(12) are satisfied. It is easy to see $0 < p < 1$, since $m_3^G < 2m_2^G - 1$. Also, $m_2^X \geq (N+2)/(N+1)$ implies $w > 0$. Thus, it suffices to prove that condition (9) is satisfied.

We first consider the first inequality of condition (9). The assumption on $r^G$ in the lemma gives

$$\frac{3}{2} - r^G \leq \frac{3}{2} - \frac{(N+1)m_2^G + (N+4)}{2(N+2)} = \frac{N+1}{N+2}\frac{2 - m_2^G}{2}.$$

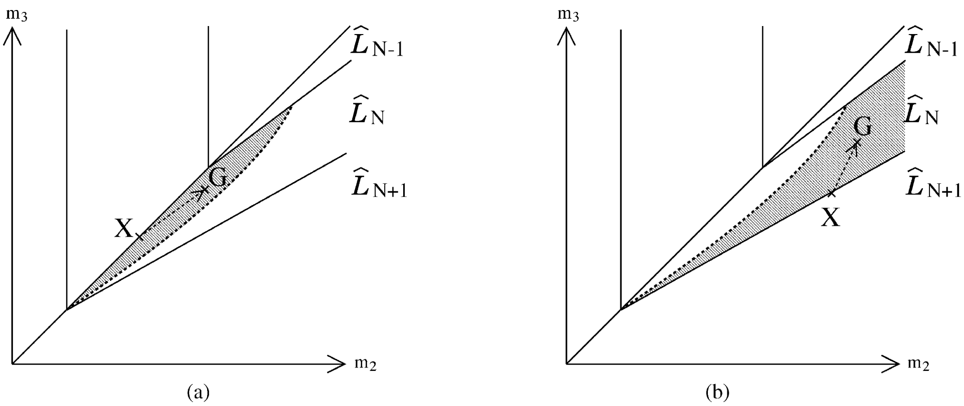



Fig. 9. Two regions in $\mathcal{L}_N$ that input $G$ can fall into under the Positive solution: (a) $G$ is well represented by a mixture of an EC distribution $X$ and an exponential distribution $W$; (b) $G$ is well represented by the convolution of an EC distribution $X$ and an exponential distribution $Z$.

Therefore, since $r^G < 3/2$, it follows that

$$m_2^X = 2w = \frac{2 - m_2^G}{2} \frac{1}{\frac{3}{2} - r^G} \geq \frac{N+2}{N+1}.$$

We next consider the second inequality of condition (9). We begin by bounding the range of $m_2^G$ for $G$ considered in the lemma. Condition $G \in \hat{\mathcal{L}}_N$ implies $m_2^G \geq (N+2)/(N+1)$. Also, if $m_2^G > 2$, then by the assumption on $r^G$ in the lemma,

$$r^G \geq \frac{(N+1)m_2^G + (N+4)}{2(N+2)} > \frac{2(N+1) + (N+4)}{2(N+2)} = \frac{3}{2}.$$

This contradicts $r^G < 3/2$. Thus, $m_2^G \leq 2$. So far, we derived the range of $m_2^G$ as $(N+2)/(N+1) < m_2^G \leq 2$.

We prove $m_2^X < (N+1)/N$ in two cases: (i) $(N+1)/N \leq m_2^G \leq 2$ and (ii) $(N+2)/(N+1) \leq m_2^G < (N+1)/N$. (i) When $(N+1)/N \leq m_2^G \leq 2$,

$$m_2^X = \frac{2 - m_2^G}{2\left(\frac{3}{2} - r^G\right)} < \frac{2 - \frac{N+1}{N}}{2\left(\frac{3}{2} - \frac{N+2}{N+1}\right)} = \frac{N+1}{N}.$$

The inequality follows from $m_2^G < (N+1)/N$ and $r^G \leq (N+2)/(N+1)$. (ii) When $(N+2)/(N+1) \leq m_2^G < (N+1)/N$,

$$m_2^X = \frac{2 - m_2^G}{2\left(\frac{3}{2} - r^G\right)} < \frac{2 - m_2^G}{2\left(\frac{3}{2} - \frac{2m_2^G - 1}{m_2^G}\right)} = m_2^G < \frac{N+1}{N}.$$

The inequality follows from

$$r^G = \frac{m_3^G}{m_2^G} < \frac{2m_2^G - 1}{m_2^G},$$

which follows from $G \in \hat{\mathcal{L}}_N$. □

The key idea behind Lemma 5 is to fix some of the parameters so that the set of equations becomes simpler and yet there exists a unique solution. The difficulty in finding closed form solutions is that we are given a system of nonlinear equations with high degree (10)–(12), and the solutions are not unique. By fixing some of the parameters, the system of equations can be reduced to have a unique solution. We find that $w$ given by Lemma 5 has nice characteristics. First, $m_2^X$ leads to a very simple expression: $m_2^X = 2w$. Second, with this expression of $m_2^X$, the expression involving $r^G$ is significantly simplified:

$$r^G = \frac{2pr^X + 3(1-p)w}{2(p + (1-p)w)}.$$

Now, solving (10)–(12) for $p$ and $w$ is a relatively easy task, and $p$ and $w$ immediately yield $m_2^X$ and $m_3^X$. Although Lemma 5 allows us to find a simple closed form solution, the set of input distributions defined for Lemma 5 is rather small. This necessitates the second approach of using the convolution of an EC distribution and an exponential distribution. Note that the second approach alone does not suffice, either. Applying the first approach to a small set of input distributions and applying the second approach to the rest of the input distribution lead to simpler closed form solutions by both approaches.

Next, we consider the second approach of using the convolution of an EC distribution (with mass probability at zero) and an exponential distribution (i.e. $\lambda_W = \infty$). Given a distribution $G \in \hat{\mathcal{L}}_N$ (we assume $r^G \neq 3/2$), we seek $m_2^X$, $m_3^X$, and $z > 0$ such that

$$m_2^X \geq \frac{N+2}{N+1}, \tag{13}$$

$$m_3^X = \frac{N+3}{N+2} m_2^X, \tag{14}$$

$$m_2^G = \frac{m_2^X + 2z + 2z^2}{(1+z)^2}, \tag{15}$$

$$m_3^G = \frac{m_2^X m_3^X + 3m_2^X z + 6z^2 + 6z^3}{(m_2^X + 2z + 2z^2)(1+z)}. \tag{16}$$

Note that (13) and (14) guarantee that we can find, via the `Complete` solution, an $(N+1)$-phase EC distribution, $X$, such that $X$ has no mass probability at zero and has normalized moments $m_2^X$ and $m_3^X$. Let $Z$ be the exponential distribution whose first moment is $\mu_1^Z = z\mu_1^X$, where $\mu_1^X$ is the first moment of $X$. (15) and (16) guarantee that, by choosing $\mu_1^X$ appropriately, $G$ is well represented by the convolution of $X$ and $Z$.

The following lemma provides conditions on the input distribution for which the second approach is defined.

**Lemma 6.** *Suppose*

$$G \in \hat{\mathcal{L}}_N \quad \text{and} \quad r^G < \frac{(N+1)m_2^G + (N+4)}{2(N+2)}$$

*for* $N \geq 1$ (*see* Fig. 9(*b*)). *If* $m_2^G = 2$, *we choose*

$$z = \frac{m_3^G - 2\frac{N+3}{N+2}}{3 - m_3^G}, \qquad m_2^X = 2(1+z), \qquad m_3^X = \frac{N+3}{N+2} m_2^X.$$

*If* $m_2^G \neq 2$, *we choose*

$$z = \frac{m_2^G\left((m_3^G - 3) - 2\frac{N+3}{N+2}(m_2^G - 2)\right) + m_2^G\sqrt{(m_3^G - 3)^2 + 8\frac{N+3}{N+2}(m_2^G - 2)\left(\frac{3}{2} - \frac{m_3^G}{m_2^G}\right)}}{2\frac{N+3}{N+2}(m_2^G - 2)^2},$$

$$m_2^X = (1+z)(m_2^G(1+z) - 2z), \qquad m_3^X = \frac{N+3}{N+2} m_2^X.$$

*Then*, $z > 0$ *and conditions* (13)–(16) *are satisfied.*

**Proof.** For each case, it is easy to check, by substitution, that conditions (14)–(16) are satisfied. Below, we prove condition (13) and $z > 0$.

We begin with the first case, where $m_2^G = 2$. It is easy to see (13) is true if $z > 0$, since

$$m_2^X = 2(1+z) > 2 > \frac{N+2}{N+1}.$$

Further, $z > 0$ if $2(N+3)/(N+2) < m_3^G < 3$, which is true by $G \in \mathcal{L}_N$, $r^G < 3/2$, and $m_2^G = 2$.

Below, we consider the case where $m_2^G \neq 2$. We first prove $z > 0$ by showing that $z$ is the larger solution of the two solutions of a quadratic equation that has a unique positive solution. Observe that

$$\begin{aligned}
& m_3^G(m_2^X + 2z + 2z^2)(1+z) = m_2^X m_3^X + 3m_2^X z + 6z^2 + 6z^3 && \text{(by (16))}\\
\iff\ & m_3^G m_2^G (1+z)^3 = \tfrac{N+3}{N+2}(m_2^X)^2 + 3z(m_2^X + 2z + 2z^2) && \text{(by (14) and (15))}\\
\iff\ & m_3^G m_2^G (1+z)^3 = \tfrac{N+3}{N+2}(1+z)^2(m_2^G(1+z) - 2z)^2 + 3z(1+z)^2 m_2^G && \text{(by (15))}\\
\iff\ & m_3^G m_2^G (1+z) = \tfrac{N+3}{N+2}(m_2^G(1+z) - 2z)^2 + 3z m_2^G
\end{aligned}$$

Thus, $z$ is a solution of the following quadratic equation: $f(z) = 0$, where

$$f(z) \equiv \frac{N+3}{N+2}(m_2^G - 2)^2 z^2 - m_2^G\left((m_3^G - 3) - 2\frac{N+3}{N+2}(m_2^G - 2)\right)z - (m_2^G)^2\left(r^G - \frac{N+3}{N+2}\right).$$

Since the coefficient of the leading term, $((N+3)/(N+2))(m_2^G - 2)^2$, is positive and $f(0) < 0$, there exists a unique positive solution of $f(z) = 0$.

Second, we show $m_2^X \geq (N+2)/(N+1)$. We consider two cases: (i) $m_2^G \geq 2$ and (ii) $m_2^G < 2$. Case (i) is easy to show. Suppose $m_2^G \geq 2$. Observe that by (15),

$$m_2^X = z((m_2^G - 2)z + 2(m_2^G - 1)) + m_2^G.$$

Thus, if $m_2^G \geq 2$, then $m_2^X \geq m_2^G \geq (N+2)/(N+1)$. Below, we consider case (ii).

Suppose $m_2^G < 2$. Observe that

$$m_2^X = -(2 - m_2^G)z^2 + 2(m_2^G - 1)z + m_2^G,$$

again by (15). Thus, $m_2^X \geq (N+2)/(N+1)$ iff $0 < z \leq z^*$, where $z^*$ is a larger solution, $x$, of the following quadratic equation: $\chi(x) = 0$, where

$$\chi(x) = -(2 - m_2^G)x^2 + 2(m_2^G - 1)x + m_2^G - \frac{N+2}{N+1}.$$

That is,

$$z^* \equiv \frac{m_2^G - 1 + \sqrt{\frac{N+2}{N+1}m_2^G - \frac{N+3}{N+1}}}{2 - m_2^G}. \tag{17}$$

Thus, it suffices to show $f(z^*) \geq 0$. Since $\chi(z^*) = 0$, we obtain $(z^*)^2$ as a linear function of $z^*$:

$$(z^*)^2 = \frac{2(m_2^G - 1)z^* + m_2^G - \frac{N+2}{N+1}}{2 - m_2^G}.$$

By substituting this $(z^*)^2$ into the expression for $f(z^*)$, we obtain

$$\begin{aligned}
f(z^*) &= \frac{N+3}{N+2}(2-m_2^G)^2 \left( \frac{2(m_2^G - 1)z^* + m_2^G - \frac{N+2}{N+1}}{2 - m_2^G} \right) \\
&\quad - m_2^G \left( 2\frac{N+3}{N+2}(2-m_2^G) - (3 - m_3^G) \right) z^* - (m_2^G)^2 \left( r^G - \frac{N+3}{N+2} \right) \\
&= \left( 3m_2^G - 2\frac{N+3}{N+2}(2-m_2^G) - m_2^G m_3^G \right) z^* + 2\frac{N+3}{N+2} m_2^G - \frac{N+3}{N+1}(2 - m_2^G) \\
&\quad - m_2^G m_3^G > \left( 3m_2^G - 2\frac{N+3}{N+2}(2-m_2^G) - (m_2^G)^2 \left( \frac{(N+1)m_2^G + (N+4)}{2(N+2)} \right) \right) z^* \\
&\quad + 2\frac{N+3}{N+2} m_2^G - \frac{N+3}{N+1}(2-m_2^G) - (m_2^G)^2 \left( \frac{(N+1)m_2^G + (N+4)}{2(N+2)} \right) \\
&= \frac{2 - m_2^G}{2(N+2)}((N+1)(m_2^G)^2 + 3(N+2)m_2^G - 4(N+3))z^* \\
&\quad - \frac{(N+1)m_2^G - (N+2)}{2(N+1)(N+2)}((N+1)(m_2^G)^2 + (2N+6)m_2^G - 4(N+3)),
\end{aligned}$$

where the inequality follows from the assumption on $r^G$ in the lemma. By substituting (17) into the last expression, we obtain

$$f(z^*) = g(m_2^G) + h(m_2^G)\sqrt{\frac{(N+2)m_2^G - (N+3)}{N+1}},$$

where

$$g(m_2^G) \equiv \frac{(N+1)^2(m_2^G)^2 - (N-3)(N+2)m_2^G - 4(N+3)}{2(N+1)(N+2)},$$

$$h(m_2^G) \equiv \frac{(N+1)(m_2^G)^2 + 3(N+2)m_2^G - 4(N+3)}{2(N+2)}.$$

Since

$$g'(m_2^G) = 2(N+1)^2 \left( m_2^G - \frac{N+2}{N+1} \right) + (N+2)(N+5) > 0$$

for $(N+2)/(N+1) \le m_2^G < 2$, $g(m_2^G)$ and $h(m_2^G)$ are increasing functions of $m_2^G$ in the range of $(N+2)/(N+1) \le m_2^G < 2$. Since

$$g\left(\frac{N+2}{N+1}\right) = \frac{2}{(N+1)^2(N+2)} > 0 \quad \text{and} \quad h\left(\frac{N+2}{N+1}\right) = \frac{2}{N^2+3N+2} > 0,$$

we have $g(m_2^G) \ge 0$ and $h(m_2^G) \ge 0$ for $(N+2)/(N+1) \le m_2^G < 2$. This implies $f(z^*) \ge 0$. □

### 7.2. Analyzing the number of phases required

The number of phases used in the `Positive` solution is characterized by the following theorem.

**Theorem 7.** *The* `Positive` *solution uses at most* $\mathrm{OPT}(G)+1$ *phases to well represent any distribution* $G \in \mathcal{U} \cup \hat{\mathcal{M}} \cup \{F | r^F \neq 3/2$ *and* $m_3^F < 2m_2^F - 1\}$.

**Proof.** Since $S^{(n)} \subset \mathcal{T}^{(n)}$ (by Lemma 1), it suffices to prove that if a distribution $G \in \mathcal{T}^{(n)}$, then at most $n+1$ phases are needed. When an input distribution $G \in \mathcal{U} \cup \hat{\mathcal{M}}$, the `Positive` solution is the same as the `Complete` solution, and hence requires the same number of phases, which is at most $\mathrm{OPT}(G)+1$. When $G \in \{F|r^F > 3/2$ and $m_3^F < 2m_2^F - 1\}$, the number of phase used in the `Positive` solution is $2 = \mathrm{OPT}(G)$. When $G \in \{F|r^F < 3/2$ and $m_3^F < 2m_2^F - 1\}$, it is immediate, from the construction of the solution, that the `Positive` solution requires at most one more phase than the `Complete` solution. For this $G$, the `Complete` solution requires $\mathrm{OPT}(G)$ phases, and hence the `Positive` solution requires $\mathrm{OPT}(G)+1$ phases. □

Finally, we remark that the `Positive` solution can be used as a building block for yet another solution, `ZeroMatching`, that not only matches the first three moments of the input distribution but also matches the mass probability at zero. Consider a distribution $G$ whose mass probability at zero is $q$. Then, $G$ can be expressed as a mixture of $O$ and a distribution $F$ that does not have mass probability at zero. The `Positive` solution can be used to match the first three moments of $F$ by an extended EC distribution, $E$. Now, a mixture of $O$ and $E$, whose distribution function is $qO(\cdot) + (1-q)E(\cdot)$, matches the first three moments and mass probability at zero of $G$. Observe that the `ZeroMatching` solution uses at most $\mathrm{OPT}(G)+1$ phases.

## 8. Conclusion

In this paper, we propose a closed form solution for the parameters of a PH distribution, $P$, that well represents a given distribution $G$. Our solution is the first that achieves all of the following goals: (i) the first three moments of $G$ and $P$ agree, (ii) any distribution $G$ that is well represented by a PH distribution (i.e., $G \in \mathcal{PH}_3$) can be well represented by some $P$, (iii) the number of phases used in $P$ is at most $\mathrm{OPT}(G)+1$, (iv) the solution is expressed in closed form.

The key idea is the definition and use of EC distributions, which sew together a two-phase Coxian$^+$ PH distribution and an Erlang distribution, each of which provides complementary properties. The set of EC distributions includes minimal PH distributions, in the sense that for any distribution, $G$, that is well represented by $n$-phase acyclic PH distribution, there exists an EC distribution, $E$, with at most

$n+1$ phases such that $G$ is well represented by $E$. This property of the set of EC distributions is the key to achieving the above goals (i)–(iii). Also, the EC distribution is defined so that it has only six free parameters. This property of the EC distribution is the key to achieving the above goal (iv).

We provide a complete characterization of the EC distribution with respect to the normalized moments. The analysis is an elegant induction based on the recursive definition of the EC distribution; the inductive analysis is enabled by a solution to a nontrivial recursive formula. Based on the characterization, we provide three variants of closed form solutions for the parameters of the EC distribution that well represents the input distribution. The closed form solutions proposed in this paper have been largely implemented and tested. The most recent implementation of the solutions is available at an online code repository: http://www.cs.cmu.edu/~osogami/code/.

Another contribution is a characterization of the set, $\mathcal{S}^{(n)}$, of distributions that are well represented by an $n$-phase acyclic PH distribution. We introduce two ideas that help in creating a simple formulation of $\mathcal{S}^{(n)}$. The first is the concept of normalized moments and their ratio, the $r$-value. The second is the notion of $\mathcal{T}^{(n)}$, which is a superset of $\mathcal{S}^{(n)}$, is close to $\mathcal{S}^{(n)}$ in size, and has a simple characterization via normalized moments. The characterization of $\mathcal{S}^{(n)}$ is used to prove the minimality of the number of phases used in our moment matching solutions. This characterization also has practical use in its own right, as it allows algorithm designers to determine how close their PH distribution is to the minimal PH distribution, and provides intuition for coming up with improved algorithms. We have ourselves benefited from exactly this point in this paper. Another benefit of characterizing $\mathcal{S}^{(n)}$ is that some existing moment matching algorithms, such as the nonlinear programming approach in [12], require knowing the number of phases, $n$, in the minimal PH distribution. The current approach involves simply iterating over all choices for $n$ [12], whereas our characterization would immediately specify $n$.

## Acknowledgements

This paper combines two papers [18,19] which appeared in TOOLS 2003, and also includes some new results beyond what was included in [18,19]. In particular, the `Positive` solution proposed in Section 7 of this paper is entirely new. The `Positive` solution is motivated by a discussion with Miklos Telek and Armin Heindl at the TOOLS conference. Also, Section 3 and Appendix A in [19] have been replaced by a simpler proof of the same result which is contained in Section 3 of this paper. Unfortunately, many interesting results in [18,19] are omitted in this paper due to space constraints. In particular, Sections 2 and 4 in [19] as well as Section 4.2 in [18] are entirely removed.

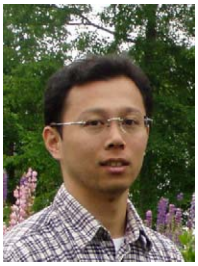
We thank Miklos Telek and anonymous reviewers of TOOLS 2003 and this special edition of Performance Evaluation for their very detailed and valuable comments and suggestions for this paper and [18,19]. This work was supported by NSF Career Grant CCR-0133077, NSF ITR Grant 99-167 ANI-0081396, and IBM via PDG Grant 2003.

## References

[1] D. Aldous, L. Shepp, The least variable phase type distribution is Erlang, Commu. Statist.—Stochast. Models 3 (1987) 467–473.

[2] T. Altiok, On the phase-type approximations of general distributions, IIE Trans. 17 (1985) 110–116.

[3] A. Cumani, On the canonical representation of homogeneous Markov processes modeling failure-time distributions, Microelectron. Reliab. 22 (1982) 583–602.
[4] A. Feldmann, W. Whitt, Fitting mixtures of exponentials to long-tail distributions to analyze network performance models, Perform. Eval. 32 (1998) 245–279.
[5] H. Franke, J. Jann, J. Moreira, P. Pattnaik, M. Jette, An evaluation of parallel job scheduling for ASCI blue-pacific, in: Proceedings of the Supercomputing'99, November 1999, pp. 679–691.
[6] A. Horváth, M. Telek, Approximating heavy tailed behavior with phase type distributions, in: Advances in Algorithmic Methods for Stochastic Models, Notable Publications, New Jersey, 2000, pp. 191–214.
[7] A. Horváth, M. Telek, Phfit: a general phase-type fitting tool, in: Proceedings of the TOOLS 2002, April 2002, pp. 82–91.
[8] M.A. Johnson, Selecting parameters of phase distributions: combining nonlinear programming, heuristics, and Erlang distributions, ORSA J. Comput. 5 (1993) 69–83.
[9] M.A. Johnson, M.F. Taaffe, An investigation of phase-distribution moment-matching algorithms for use in queueing models, Queueing Syst. 8 (1991) 129–147.
[10] M.A. Johnson, M.R. Taaffe, Matching moments to phase distributions: mixtures of Erlang distributions of common order, Commun. Statist.—Stochast. Models 5 (1989) 711–743.
[11] M.A. Johnson, M.R. Taaffe, Matching moments to phase distributions: density function shapes, Commun. Statist.—Stochast. Models 6 (1990) 283–306.
[12] M.A. Johnson, M.R. Taaffe, Matching moments to phase distributions: nonlinear programming approaches, Commun. Statist.—Stochast. Models 6 (1990) 259–281.
[13] S. Karlin, W. Studden, Tchebycheff Systems: With Applications in Analysis and Statistics, Wiley, New Yrok, 1966.
[14] R.E.A. Khayari, R. Sadre, B. Haverkort, Fitting world-wide web request traces with the EM-algorithm, Perform. Eval. 52 (2003) 175–191.
[15] R. Marie, Calculating equilibrium probabilities for $\lambda(n)/c_k/1/n$ queues, in: Proceedings of the Performance 1980, 1980, pp. 117–125.
[16] C.A. O'Cinneide, Phase-type distributions and majorization, Ann. Appl. Probab. 1 (3) (1991) 219–227.
[17] T. Osogami, Analysis of multi-server systems via dimensionality reduction of Markov chains, PhD Thesis, School of Computer Science, Carnegie Mellon University, 2005.
[18] T. Osogami, M. Harchol-Balter, A closed-form solution for mapping general distributions to minimal PH distributions, in: Proceedings of the TOOLS 2003, September 2003, pp. 200–217.
[19] T. Osogami, M. Harchol-Balter, Necessary and sufficient conditions for representing general distributions by Coxians, in: Proceedings of the TOOLS 2003, September 2003, pp. 182–199.
[20] A. Riska, V. Diev, E. Smirni, An EM-based technique for approximating long-tailed data sets with PH distributions, Perform. Eval. 55 (1/2) (2004) 147–164.
[21] C. Sauer, K. Chandy, Approximate analysis of central server models, IBM J. Res. Dev. 19 301–313 (1975).
[22] L. Schmickler, MEDA: mixed Erlang distributions as phase-type representations of empirical distribution functions, Commun. Statist.—Stochast. Models 8 (1992) 131–156.
[23] M. Squillante, Matrix-analytic methods in stochastic parallel-server scheduling models, in: Advances in Matrix-Analytic Methods for Stochastic Models, Notable Publications, New Jersey, 1998.
[24] D. Starobinski, M. Sidi, Modeling and analysis of power-tail distributions via classical teletraffic methods, Queueing Syst. 36 (2000) 243–267.
[25] M. Telek, A. Heindl, Matching moments for acyclic discrete and continuous phase-type distributions of second order, Int. J. Simulat. 3 (2003) 47–57.
[26] G. van Houtum, W. Zijm, Computational procedures for stochastic multi-echelon production systems, Int. J. Prod. Econ. 23 (1991) 223–237.
[27] G. van Houtum, W. Zijm, Incomplete convolutions in production and inventory models, OR Spektrum 10 (1997) 97–107.
[28] W. Whitt, Approximating a point process by a renewal process: two basic methods, Operat. Res. 30 (1982) 125–147.
[29] Y. Zhang, H. Franke, J. Moreira, A. Sivasubramaniam, An integrated approach to parallel scheduling using gang-scheduling, backfilling, and migration, IEEE Trans. Parallel Distrib. Syst. 14 (2003) 236–247.



**Takayuki Osogami** is a Ph.D. candidate at the Department of Computer Science, Carnegie Mellon University. He received a B.Eng. degree in electronic engineering from the University of Tokyo, Japan, in 1998. In 1998–2001, he was at IBM Tokyo Research Laboratory, where the principal project was development of optimization algorithms. His current research interest includes performance analysis of resource allocation policies for multi-server systems.

**Mor Harchol-Balter** is an associate professor of Computer Science at Carnegie Mellon University. She received her doctorate from the Computer Science Department at the University of California at Berkeley under the direction of Manuel Blum. She is a recipient of the McCandless Chair, the NSF CAREER award, the NSF Postdoctoral Fellowship in the Mathematical Sciences, multiple best paper awards, and several teaching awards, including the Herbert A. Simon Award for Teaching Excellence. She is heavily involved in the ACM SIGMETRICS research community. Her work focuses on designing new scheduling/resource allocation policies for various distributed computer systems including Web servers, distributed supercomputing servers, networks of workstations, and database systems. Her work spans both queueing analysis and implementation and emphasizes integrating measured workload distributions into the problem solution.