Closed Form Solutions for Mapping General Distributions to Minimal PH Distributions

Takayuki Osogami, * Mor Harchol-Balter [1]

Department of Computer Science, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA

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 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.

Key words: PH distribution, moment matching, closed form, normalized moment
PACS:

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 research has focused on the

* Corresponding author

Email addresses: osogami@cs.cmu.edu (Takayuki Osogami,), harchol@cs.cmu.edu (Mor Harchol-Balter).

URLs: http://www.cs.cmu.edu/~osogami/ (Takayuki Osogami,), http://www.cs.cmu.edu/~harchol/ (Mor Harchol-Balter).

[1] This work was supported by NSF Career Grant CCR-0133077, by NSF ITR Grant 99-167 ANI-0081396, and by IBM via PDG Grant 2003.

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 [4,5,21,25,31]. Clearly, 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 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 [16] (see Figure 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 which 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 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 all distributions G that can be well-represented by a PH distribution, there exists an EC distribution that well-represents G. Furthermore, the class of EC distributions is broad enough such that for any distribution, G, that is well-represented by an 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.

It is not clear whether restricting our search space to the set of *acyclic* PH distributions (as is used throughout literature) is limiting. While it is theoretically possible that minimum phase solution is cyclic, in practice we have not been able to find a situation where minimal solution requires cycles, and this question is left as an open problem. However, note that 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. Unfortunately, the minimal number of phases is not known for general distributions. This makes it difficult to evaluate the effectiveness of different algorithms and also makes the design of moment matching algorithms open-ended. As a secondary contribution, this paper provides a formal characterization of the set of distributions that are well-represented by an n-phase PH distribution, for each $n = 1, 2, 3, ...$ 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, $\vec{\tau}^F$.*

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

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

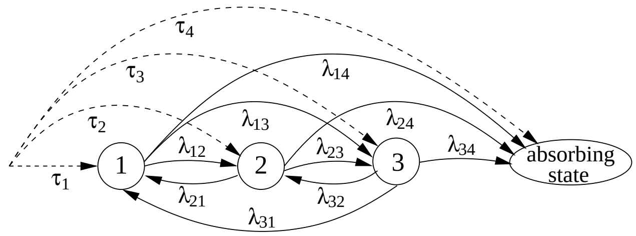


Fig. 1. *The continuous time Markov chain underlying a three-phase PH distribution.*

There are $n = 3$ internal states. With probability τ_i we start in the ith state. The absorption time is the sum of the times spent in each of the states before reaching the absorption state.

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

Observe that an acyclic PH distribution, F, has upper triangular $\mathbf{T^F}$. 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 μ_k^F be the k-th moment of a distribution F for $k = 1, 2, 3$. The* ***normalized k-th moment*** *m_k^F of F for $k = 2, 3$ is defined to be $m_2^F = \frac{\mu_2^F}{(\mu_1^F)^2}$ and $m_3^F = \frac{\mu_3^F}{\mu_1^F \mu_2^F}$.*

Notice the relationship between the normalized moments and the coefficient of variability C_F and the skewness γ_F of F: $m_2^F = C_F^2 + 1$ and $m_3^F = \nu_F \sqrt{m_2^F}$, where $\nu_F = \frac{\mu_3^F}{(\mu_2^F)^{3/2}}$. ($\nu_F$ and γ_F are closely related, since $\gamma_F = \frac{\bar{\mu}_3^F}{(\bar{\mu}_2^F)^{3/2}}$, 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 *Almost all the nonnegative distributions are in $\mathcal{PH}_3$.*

Definition 6 *$OPT(G)$ is defined to be the minimum number of phases in a PH distribution 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 (number of moments matched; generality of the solution; computational efficiency of the algorithm; and minimality of the number of phases), 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 which perform very well along all the other three measures. Sauer and Chandy [23] provide a closed-form solution for matching two moments of a general distribution in $\mathcal{PH}_3$. They use a two-branch hyper-exponential distribution for matching distributions with squared coefficient of variability $C^2 > 1$ 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 in $\mathcal{PH}_3$. 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 which perform very well along the remaining three measures. Whitt [30] 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-branch hyper-exponential distribution, and Altiok matches to a two-phase Coxian$^+$ PH distribution. Telek and Heindl [27] focus on the set of distributions with $C^2 \geq \frac{1}{2}$ 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 Coxian$^+$ PH distribution.

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 $2OPT(G) + 2$ 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

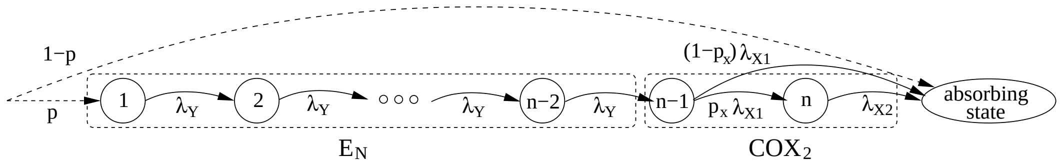


Fig. 2. *The Markov chain underlying an n-phase EC distribution. The first box above depicts the underlying continuous time Markov chain in an Erlang-N distribution, where $N = n-2$, and the second box depicts the underlying continuous time Markov chain in a two-phase Coxian$^+$ PH distribution. Notice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

nearly minimal in that it requires at most $OPT(G)+2$ phases. Unfortunately, their algorithm requires solving a nonlinear programing 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. Of particular recent interest has been work on fitting heavy-tailed distributions to PH distributions [3,6,7,14,22,26]. There is also work which combines the goals of moment matching with the goal of fitting the shape of the distribution [8,24]. The work above is clearly broader in its goals than simply matching three moments. Unfortunately there's a tradeoff: obtaining a more precise fit requires more phases. Additionally it can sometimes be computationally inefficient [8,24].

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 PH distributions. In this paper, we show that by using the set of EC distributions as our subset $\mathcal{S}$, we achieve a solution which excels in all four desirable measures mentioned earlier. We also use the EC distribution as a building block in designing variants of our closed form solution. We define the EC distributions as follows:

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

Figure 2 shows the underlying Markov chain of an n-phase EC 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 $m_3^G > \frac{3}{2}m_2^G$) [20]. 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 $m_2^G \leq \frac{n+1}{n}$ 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 [1,17]. However the Erlang distribution is obviously limited in the set of distributions which it can well-represent.

Our approach is therefore to combine the Erlang distribution with the two-phase Coxian$^+$ PH distribution, allowing us to 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 allows us to obtain closed-from expressions for the parameters (n, p, λ_Y, λ_{X1}, λ_{X2}, p_X) of the EC distribution that well-represents any given distribution in $\mathcal{PH}_3$.

2.2 Characterizing PH distributions

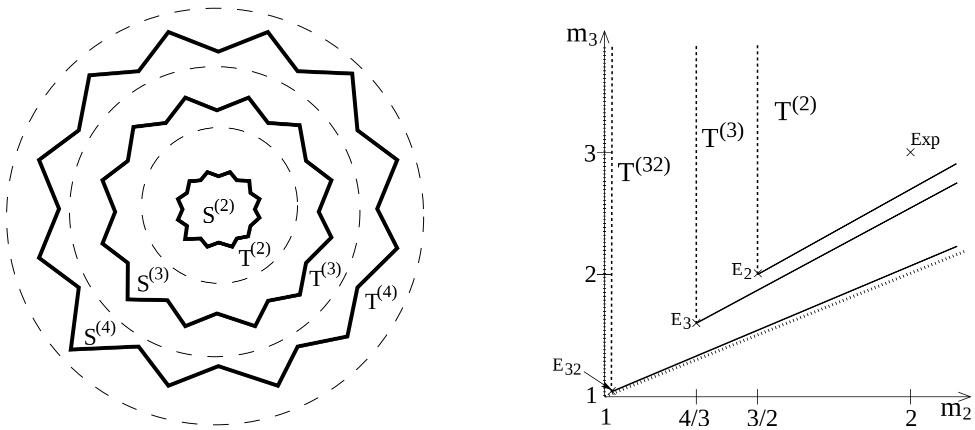
We now turn to our second goal of the paper, namely characterizing the set of distributions that are well-represented by an n-phase acyclic PH distribution.

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 2-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 [27] 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, Heindl) *$G \in S^{(2)^*}$ iff G satisfies exactly one of the following three conditions: (i) $\frac{9m_2^G - 12 + 3\sqrt{2}(2-m_2^G)^{\frac{3}{2}}}{m_2^G} \leq m_3^G \leq \frac{6(m_2^G-1)}{m_2^G}$ and $\frac{3}{2} \leq m_2^G < 2$, (ii) $m_3^G = 3$ and $m_2^G = 2$, or (iii) $\frac{3}{2}m_2^G < m_3^G$ and $2 < m_2^G$.*

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



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

Fig. 3. *(a) Solid lines delineate $\mathcal{S}^{(n)}$ (which is irregular) and dashed lines delineate $\mathcal{T}^{(n)}$ (which is regular – has a simple specification). Observe the nested structure of $\mathcal{S}^{(n)}$ and $\mathcal{T}^{(n)}$: $\mathcal{S}^{(n)} \subset \mathcal{T}^{(n)} \subset \mathcal{S}^{(n+1)}$ for all integers $n \geq 2$. (b) Set $\mathcal{T}^{(n)}$ is depicted as a function of the normalized moments. $\mathcal{T}^{(n)}$ sets are delineated by solid lines, which includes the border, and dashed lines, which does 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 $m_3 \geq m_2$.*

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

The main contribution of our characterization of PH distributions is a derivation of the nested relationship between $\mathcal{S}^{(n)}$ and $\mathcal{T}^{(n)}$ for all $n \geq 2$. This relationship is illustrated in Figure 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 $n \geq 2$. 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 $n \geq 2$, $\mathcal{S}^{(n)} \subset \mathcal{T}^{(n)} \subset \mathcal{S}^{(n+1)}$.*

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 $\mathcal{S}^{(n)}$. This is covered primarily in Section 3, and will be used in the second part of the paper, which involves the construction of moment matching algorithms (Section 4-6).

Our moment matching algorithms depend on properties of EC distributions, which will be discussed in depth in Section 4 We find that for the purpose of moment matching it suffices to narrow down the set of EC distributions further from six free parameters to five free parameters, by optimally fixing one of the parameters.

In Section 5-6, we present three variants of closed form solutions for the remaining free parameters of the EC distribution, each of which achieves slightly different goals. The first closed-form solution provided, which we refer to as *the simple solution*, (see Section 5) has the advantage of simplicity and readability; however it does not work for all distributions in $\mathcal{PH}_3$ (although it works for almost all). This solution requires at most $OPT(G) + 2$ phases. The second closed-form solution provided, which we refer to as *the improved solution*, (see Section 6.1) is defined for all the input distributions in $\mathcal{PH}_3$ and uses at most $OPT(G) + 1$ phases. The improved solution is only lacking in numerical stability for a small subset of $\mathcal{PH}_3$. In practice, this is not a problem, since distributions lying in the small subset can be perturbed to move out of the subset. In [18,19], we also provide numerically stable solutions.

In the simple solution and the improved 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 third closed-form solution provided, which we refer to as *the positive solution*, (see Section 6.2) has no mass probability at zero ($p = 1$). This solution is defined for almost all distributions in $\mathcal{PH}_3$ and uses at most $OPT(G) + 1$ phases.

3 Characterizing PH 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, i.e. the following lemma:

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 a moment matching algorithm (see Corollary 3 in Section 6.1).

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 $r^F = \frac{m_3^F}{m_2^F}$ and is also referred to as the r-value of F.*

One of the nice properties of the r-value is that it is insensitive to the mass probability at zero, as shown in the proposition below:

Proposition 2 *Let $Z(\cdot) = pX(\cdot) + (1-p)O(\cdot)$, 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: By definition, $r^Z = \frac{(p\mu_3^X)(p\mu_1^X)}{(p\mu_2^X)^2} = \frac{(\mu_3^X)(\mu_1^X)}{(\mu_2^X)^2} = r^X$. ■

To shed light on the expression $Z(\cdot) = pX(\cdot) + (1-p)O(\cdot)$, consider a random variables V_1 whose distribution is X, Then, random variable

$$V_2 = \begin{cases} V_1 & \text{with probability } p \\ 0 & \text{with probability } 1-p. \end{cases}$$

has distribution Z, since $\Pr(V_2 < t) = p\Pr(V_1 < t) + (1-p)$.) 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) $m_2^F > \frac{n+1}{n}$ and $r^F \geq \frac{n+2}{n+1}$, or (ii) $m_2^F = \frac{n+1}{n}$ and $r^F = \frac{n+2}{n+1}$.

To show $\mathcal{S}^{(n)} \subset \mathcal{T}^{(n)}$, consider an arbitrary distribution, $X \in \mathcal{S}^{(n)}$. Let P be an n-phase acyclic PH distribution that well-represents X. Then $X \in \mathcal{T}^{(n)}$ iff $P \in \mathcal{T}^{(n)}$. Hence, it suffices to prove that all acyclic 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 is the unique class of n-phase PH distributions with the least normalized second moment among all the n-phase PH distributions and (ii) the Erlang-n distribution has the least r-value among all the n-phase acyclic PH distributions. Note that an Erlang-n distribution refers to the convolution of n i.i.d. exponential distributions (the distribution of the sum of n i.i.d. exponential random variables). Thus, the Erlang-n distribution, E_n, has $m_2^{E_n} = \frac{n+1}{n}$ and $r^{E_n} = \frac{n+2}{n+1}$.

Property (i) of the Erlang-n distribution immediately follows from the prior work by Aldous and Shepp [1] and O'Cinneide [17], 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 mean. Thus, all that remain is to prove property (ii).

Our approach is different from Aldous and Shepp [1] and O'Cinneide [17]. Aldous and Shepp prove the least variability of the Erlang-n distribution via quadratic variation (a property related to the second moment), and hence it is unlikely that their approach can be applied to prove property (ii), which relies on higher moments, in particular the r-value. O'Cinneide extends the work by Aldous and Shepp, considering a convex function, $f(\cdot)$, applied to a random variable with an n-phase PH distribution with a fixed mean. He proves that the expectation of $f(V)$ is minimized when the random variable V has an Erlang distribution. Unfortunately, the r-value of a distribution, G, is not an expectation of $f(V)$, where V is a random variable with distribution G, for a convex function $f(\cdot)$, and the theory of majorization does not directly apply to the r-value.

Our proof makes use of 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 in our proof is that any acyclic PH distribution, P, can be seen as a mixture of convolutions of exponential distributions, and one of the convolutions of exponential distributions has no greater r-value than P. This allows us to related 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 $i = 1, ...n$. Then, there exists $i \in [1, n]$ such that $r^Z \geq r^{X_i}$.*

Proof: We prove the lemma by induction on n. Without loss of generality, we let $X_1 \geq \cdots \geq X_n$.

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

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

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

Inductive case: Suppose that the lemma holds for $n \le k$. When $n = k+1$, Z can be seen as a mixture of two distributions, $Y(\cdot) = \frac{1}{1-p_{k+1}} \sum_{i=1}^{k} p_i X_i(\cdot)$ and $X_{k+1}(\cdot)$. When $r^{X_{k+1}} \le r^Z$, the lemma holds for $n = k+1$. When $r^{X_{k+1}} < r^Z$, we have $r^Y \le r^Z$ by the base case. By the inductive hypothesis, there exist $i \in [1,k]$ such that $r^Y \ge r^{X_i}$. Thus, the lemma holds for $n = k+1$, which completes the proof. ■

We are now ready to prove that an m-phase Erlang distribution has no greater r-value than any n-phase PH distribution. We prove the following lemma.

Lemma 3 *The Erlang distribution has the least r-value among all the acyclic PH distribution with a fixed number of phases, m, for all $m \ge 1$. An Erlang distribution, either mixed with O or by itself, constitutes the unique class of acyclic PH distributions with the least r-value.*

Proof: We prove the lemma by induction on m.

Base case ($m = 1$): Any PH distribution with one phase 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 $m \le k$. We show that the lemma holds for $m = k+1$ as well.

Consider any $(k+1)$-phase acyclic PH distribution, G, which is neither an Erlang distribution nor a mixture of O and an Erlang distribution. We first show that there exists a PH distribution, F_1, with no greater r-value (i.e. $r^{F_1} \le r^G$) such that F_1 is a convolution of an exponential distribution, X, and an k-phase PH distribution, H_1. The key idea is to see any PH distribution as a mixture of PH distributions whose τ vectors defined in Section 2 are base vectors. For example, the three-phase PH distribution, G, in Figure 1, can be seen as a mixture of O and the three 3-phase PH distribution, G_i ($i = 1, ..., 3$), where $\vec{\tau}^{G_1} = (1,0,0)$, $\vec{\tau}^{G_2} = (0,1,0)$, $\vec{\tau}^{G_3} = (0,0,1)$, and $\mathbf{T^{G_1}} = \cdots = \mathbf{T^{G_3}} = \mathbf{T^G}$. Proposition 2 and Lemma 2 imply that there exists $i \in [1,3]$ such that $r^{G_i} \le r^G$. Without loss of generality, let $r^{G_1} \le r^G$ and let $F_1 = G_1$. Note that F_1 has no greater r-value than G (i.e. $r^{F_1} \le r^G$), and F_1 is a convolution of an exponential distribution, X, and an k-phase PH distribution, H_1.

Next we show that if F_1 is neither an Erlang distribution nor a mixture of O and an Erlang distribution, then there exists a PH distribution, F_2, with no greater r-value (i.e. $r^{F_2} \le r^{F_1}$). Let H_2 be a mixture of O and an Erlang-k distribution, E_k, (i.e. $H_2(\cdot) = pO(\cdot) + (1-p)E_k(\cdot)$), where p is chosen such that $\mu_1^{H_2} = \mu_1^{H_1}$ and $m_2^{H_2} = m_2^{H_1}$. There always exists such an H_2, since the Erlang-k distribution has the least m_2 among all the PH distributions (in particular

$m_2^{E_k} \le m_2^{H_1}$) and m_2 is an increasing function of p ($m_2^{H_2} = \frac{m_2^{E_k}}{1-p}$). Also, observe that by the inductive hypothesis $r^{H_2} \le r^{H_1}$. Let F_2 be a convolution of X and H_2, i.e. $F_2(\cdot) = X(\cdot) * H_2(\cdot)$. We prove that $r^{F_2} \le r^{F_1}$. Let $y = \frac{\mu_1^{H_1}}{\mu_1^X}$. Then,

$$
\begin{aligned}
r^{F_1} &= \frac{\left(r^X (m_2^X)^2 + 3m_2^X y + 3m_2^{H_1} y^2 + r^{H_1} (m_2^{H_1})^2 y^3\right)(1+y)}{\left(m_2^X + 2y + m_2^{H_1} y^2\right)^2} \\
&\ge \frac{\left(r^X (m_2^X)^2 + 3m_2^X y + 3m_2^{H_2} y^2 + r^{H_2} (m_2^{H_2})^2 y^3\right)(1+y)}{\left(m_2^X + 2y + m_2^{H_2} y^2\right)^2} = r^{F_2},
\end{aligned}
$$

where the inequality follows from $\mu_1^{H_2} = \mu_1^{H_1}$, $m_2^{H_2} = m_2^{H_1}$, and $r^{H_2} \le r^{H_1}$.

Finally, we show that a mixture of O and an Erlang distribution has the least r-value. F_2 is a convolution of X and H_2, 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 $r^X \le r^{F_2}$ and $r^{F_3} \le r^{F_2}$ holds. When $r^X \le r^{F_2}$, we found an Erlang-1 distribution (exponential distribution), X, with $r^X \le r^{F_2} \le r^{F_1} \le r^F$. When $r^X > r^{F_2}$, $r^{F_3} \le r^{F_2}$ holds. Let F_4 be the Erlang-$(k+1)$ distribution. We prove that $r^{F_4} \le r^{F_3}$, which will complete the proof. It suffices to prove that r^{F_3} is minimized when $\mu_1^X = \frac{1}{k}\mu_1^{E_k}$. Let $y = \frac{\mu_1^X}{\mu_1^{E_k}}$. Then, $r^{F_3} = \frac{\left(r^{E_k}(m_2^{E_k})^2 + 3m_2^{E_k} y + 6y^2 + 6y^3\right)(1+y)}{\left(m_2^{E_k} + 2y + 2y^2\right)}$, where $r^{E_k} = \frac{k+2}{k+1}$ and $m_2^{E_k} = \frac{k+1}{k}$. 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 $y = \frac{1}{k}$. ■

4 EC distribution: Motivation and properties

The purpose of this section is twofold: 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 (λ_Y is fixed) which we will use in our moment matching algorithm. Both results are summarized in Theorem 3.

To motivate the theorem in this section, consider the following story. Suppose one is trying to match the first three moments of a given distribution, G, to a distribution, P, which is a 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 PH Coxian$^+$

distribution alone suffices and we need no exponential distributions. If the variability of G is lower, however, we might try appending an exponential distribution to the two-phase PH Coxian$^+$ distribution. If that doesn't 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.

Continuing with our story, one could express the appending of each exponential distribution as a "function" whose goal is to reduce the variability of P yet further. We call this "function ϕ."

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

Observe that, when X is a k-phase PH distribution, $\phi(X)$ is a $(k+1)$-phase PH distribution. In theory, function ϕ allows each successive exponential distribution which is appended to have a different first moment. Surprisingly, however, the following theorem shows that if the exponential distribution Y being appended by function ϕ is chosen so as to minimize the normalized second moment of $\phi(X)$ (as specified by the definition), then the first moment of each successive Y is always *the same* and is defined by the simple formula shown the theorem below, which also characterizes the normalized moments of $\phi^l(X)$.

Theorem 3 *Let $\phi^l(X) = Y_l * \phi^{l-1}(X)$, where Y_l is an exponential distribution with rate λ_{Y_l} for $l = 1, ..., N$. Then, $\lambda_{Y_l} = \frac{1}{(m_2^X - 1)\mu_1^X}$ for $l = 1, ..., N$. The normalized moments of $Z_N = \phi^N(X)$ are: $m_2^{Z_N} = \frac{(m_2^X-1)(N+1)+1}{(m_2^X-1)N+1}$ and*

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

Observe that, when X is a k-phase PH distribution, $\phi^N(X)$ is a $(k+N)$-phase PH distribution. The remainder of this section will prove the above theorem and a corollary.

Proof:[Theorem 3] 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 $y = \frac{\mu_1^Y}{\mu_1^X}$. Observe that m_2^Z is minimized when $y = m_2^X - 1$, namely when $\mu_1^Y = (m_2^X - 1)\mu_1^X$. Observe that when μ_1^Y is set at this value, the normalized moments of Z satisfies: $m_2^Z = 2 - \frac{1}{m_2^X}$ and $m_3^Z = \frac{1}{m_2^X(2m_2^X - 1)} m_3^X + \frac{3(m_2^X - 1)}{m_2^X}$.

We next characterize $Z_l = \phi^l(X) = Y_l * \phi^{l-1}(X)$ for $2 \le l \le N$: 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 follow from solving the following recurrence equations (where we use b_l to denote $m_2^{\phi^l(X)}$ and B_l to denote $m_3^{\phi^l(X)}$):

$$b_{l+1} = 2 - \frac{1}{b_l} \quad \text{and} \quad B_{l+1} = \frac{B_l}{b_l(2b_l - 1)} + \frac{3(b_l - 1)}{b_l}.$$

The solutions for these recurrence equations are

$$b_l = \frac{(b_1 - 1)l + 1}{(b_1 - 1)(l - 1) + 1} \text{ and } B_l = \frac{b_1 B_1 + (b_1 - 1)(l - 1)\left(3b_1 + (b_1 - 1)(b_1 + 2)l + (b_1 - 1)^2 l^2\right)}{((b_1 - 1)l + 1)\left((b_1 - 1)(l - 1) + 1\right)^2}$$

for all $l \ge 1$. 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 λ_{Y_l} is proved by induction. When $l = 1$, $\lambda_{Y_1} = \frac{1}{(m_2^X - 1)\mu_1^X}$ follows from the expression $\mu_1^Y = (m_2^X - 1)\mu_1^X$ derived above. Assume that $\lambda_{Y_l} = \frac{1}{(m_2^X - 1)\mu_1^X}$ holds when $l = 1, ..., t$. Let $Z_t = \phi^t(X)$. By the second part of the theorem, which is proved above, $m_2^{Z_t} = \frac{(m_2^X - 1)(t+1) + 1}{(m_2^X - 1)t + 1}$. Thus, by $\mu_1^Y = (m_2^X - 1)\mu_1^X$, $\mu_1^{Y_{t+1}} = (m_2^{Z_t} - 1)\mu_1^{Z_t} = (m_2^X - 1)\mu_1^X$. ■

Corollary 1 *Let $Z_N = \phi^N(X)$. If X is in set $\left\{F \mid 2 < m_2^F\right\}$, then Z_N is in set $\left\{F \mid \frac{N+2}{N+1} < m_2^F < \frac{N+1}{N}\right\}$.*

Corollary 1 suggests the number, N, of times that function ϕ must be applied to X to bring $m_2^{Z_N}$ into the desired range, given the value of m_2^X.

Proof:[Corollary 1] By Theorem 3, $m_2^{Z_N}$ is a continuous and monotonically increasing function of m_2^X. Thus, the infimum and the supremum of $m_2^{Z_N}$ are given by evaluating $m_2^{Z_N}$ at the infimum and the supremum, respectively, of m_2^X. When $m_2^X \to 2$, $m_2^{Z_N} \to \frac{N+2}{N+1}$. When $m_2^X \to \infty$, $m_2^{Z_N} \to \frac{N+1}{N}$. ■

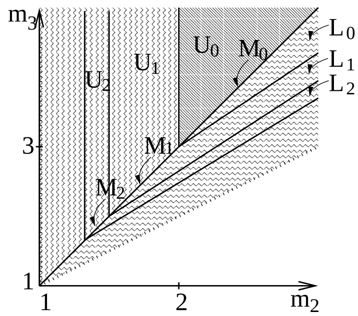


Fig. 4. *A classification of distributions. The dotted lines delineate the set of all nonnegative distributions G* ($m_3^G \geq m_2^G \geq 1$).

5 A simple closed form solution

Theorem 3 implies that the parameter λ_Y of the EC distribution can be fixed without excluding the distributions of lowest variability from the set of EC distributions. In the rest of the paper, we constrain λ_Y as follows:

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

and derive closed form representations of the remaining free parameters (n, p, λ_{X1}, λ_{X2}, p_X), where these free parameters will determine m_2^X and μ_1^X, which in turn gives λ_Y by (1). 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 use a smaller number of phases.

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 Figure 4). More formally,

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

$$\mathcal{U}_0 = \left\{ F \middle| m_2^F > 2 \text{ and } m_3^F > 2m_2^F - 1 \right\},$$
$$\mathcal{U}_i = \left\{ F \middle| \frac{i+2}{i+1} < m_2^F < \frac{i+1}{i} \text{ and } m_3^F > 2m_2^F - 1 \right\},$$
$$\mathcal{M}_0 = \left\{ F \middle| m_2^F > 2 \text{ and } m_3^F = 2m_2^F - 1 \right\},$$
$$\mathcal{M}_i = \left\{ F \middle| \frac{i+2}{i+1} < m_2^F < \frac{i+1}{i} \text{ and } m_3^F = 2m_2^F - 1 \right\},$$

$$\mathcal{L}_0 = \left\{ F \middle| \frac{3}{2} m_2^F < m_3^F < 2m_2^F - 1 \right\},$$
$$\mathcal{L}_i = \left\{ F \middle| \frac{i+3}{i+2} m_2^F < m_3^F \leq \frac{i+2}{i+1} m_2^F \text{ and } m_3^F < 2m_2^F - 1 \right\},$$

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 $\mathcal{L} = \mathcal{L}_0 \cup \mathcal{L}^+$.

The next theorem provides the intuition behind the sets $\mathcal{U}$, $\mathcal{M}$, and $\mathcal{L}$; namely, for any distribution X, X and $\phi(X)$ are in the same classification region (Figure 4).

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

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

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

where the last equality follows from $m_2^X = \frac{1}{2 - m_2^Z}$. ■

By Corollary 1 and Lemma 4, it follows that:

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 $F \in \mathcal{U}_0 \cup \mathcal{M}_0$, F can be well-represented by a two-phase Coxian$^+$ PH distribution, and $Z_N = \phi^N(X)$ can be well-represented by a $(2 + N)$-phase EC distribution. 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 ($p < 1$).

From these properties of $\phi^N(X)$, it is relatively easy to provide a closed form solution for the parameters (n, p, λ_{X1}, λ_{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. A little more effort is necessary to minimize the number of phases and to construct a solution with no mass probability at zero.

In this section, we give a simple solution, which assumes the following condition on the input distribution G: $G \in \mathcal{PH}_3^-$, 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 are not included. We also analyze the number of necessary phases and prove the following theorem:

Theorem 4 *Under the simple solution, the number of phases used to well-represent any distribution G by an EC distribution is at most $OPT(G) + 2$.*

The Closed Form Solution: 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 $p < 1$. For all cases, the parameters (n, p, λ_{X1}, λ_{X2}, p_X) are given by simple closed formulas.

(i) If $G \in \mathcal{U}_0 \cup \mathcal{M}_0$, 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 (λ_{X1}, λ_{X2}, p_X) of the two-phase Coxian$^+$ PH distribution are chosen as follows [27,20]: $\lambda_{X1} = \frac{u+\sqrt{u^2-4v}}{2\mu_1^G}$, $\lambda_{X2} = \frac{u-\sqrt{u^2-4v}}{2\mu_1^G}$, and $p_X = \frac{\lambda_{X2}\mu_1^G(\lambda_{X1}\mu_1^G-1)}{\lambda_{X1}\mu_1^G}$, where $u = \frac{6-2m_3^G}{3m_2^G-2m_3^G}$ and $v = \frac{12-6m_2^G}{m_2^G(3m_2^G-2m_3^G)}$.

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

$$n = \min\left\{k \middle| m_2^G > \frac{k}{k-1}\right\} = \left\lfloor \frac{m_2^G}{m_2^G - 1} + 1 \right\rfloor, \tag{2}$$

($N = \left\lfloor \frac{m_2^G}{m_2^G-1} - 1 \right\rfloor$). 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, where $Z(\cdot) = Y^{(n-2)*}(\cdot) * X(\cdot)$ and Y is an exponential distribution with rate given by (1), $Y^{(n-2)*}$ is the $(n-2)$-th convolution of Y, and $Y^{(n-2)*} * X$ is the convolution of $Y^{(n-2)*}$ and X.[2] By Theorem 3, this can be achieved by setting $m_2^X = \frac{(n-3)m_2^G-(n-2)}{(n-2)m_2^G-(n-1)}$, $m_3^X = \frac{\beta m_3^G - \alpha}{m_2^X}$, and $\mu_1^X = \frac{\mu_1^G}{(n-2)m_2^X-(n-3)}$, where

[2] To shed light on this expression, consider i.i.d. random variables V_1, ... V_k whose distributions are Y and a random variable V_{k+1}. Then random variable $\sum_{t=1}^{k+1} V_t$ has distribution Z.

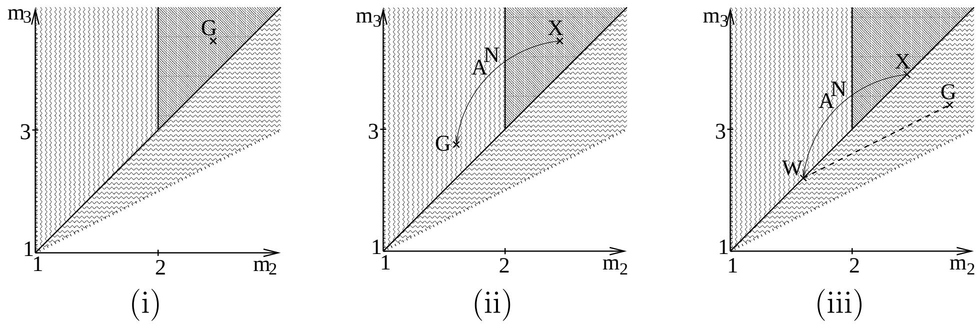


Fig. 5. *A graphical representation of the simple solution. Let G be the input distribution. (i) If $G \in \mathcal{U}_0 \cup \mathcal{M}_0$, G is well-represented by a two-phase Coxian$^+$ PH distribution X. (ii) If $G \in \mathcal{U}^+ \cup \mathcal{M}^+$, G is well-represented by $\phi^N(X)$, where X is a two-phase Coxian$^+$ PH distribution. (iii) If $G \in \mathcal{L}$, G is well-represented by Z, where Z is $W = \phi^N(X)$ with probability p and 0 with probability $1-p$ and X is a two-phase Coxian$^+$ PH distribution.*

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

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 μ_1^X, specified above.

(iii) <u>If $G \in \mathcal{L}$</u>, then let $p = \frac{1}{2m_2^G - m_3^G}$, $m_2^W = pm_2^G$, $m_3^W = pm_3^G$, and $\mu_1^W = \frac{\mu_1^G}{p}$. G is then well-represented by distribution Z, where $Z(\cdot) = pW(\cdot) + (1-p)O(\cdot)$.

Observe that p satisfies $0 \le p < 1$ and W satisfies $W \in \mathcal{M}$. If $W \in \mathcal{M}_0$, the parameters of W are provided by case (i), using m_2^W, m_3^W, and μ_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 μ_1^W, specified above.

Figure 5 shows a graphical representation of the simple solution.

Analyzing the number of phases required The proof of Theorem 4 relies on Lemma 1.

Proof:[Theorem 4] We will show that (i) if a distribution $G \in \mathcal{T}^{(l)} \cap (\mathcal{U} \cup \mathcal{M})$, then at most $l+1$ phases are used, and (ii) if a distribution $G \in \mathcal{T}^{(l)} \cap \mathcal{L}$, then at most $l+2$ phases are used. Since $\mathcal{S}^{(l)} \subset \mathcal{T}^{(l)}$ by Lemma 1, this will complete the proof. (i) Suppose $G \in \mathcal{U} \cup \mathcal{M}$. If $G \in \mathcal{T}^{(l)}$, then by (2) the EC distribution provided by the simple solution has at most $l+1$ phases. (ii) Suppose $G \in \mathcal{L}$. If $G \in \mathcal{T}^{(l)}$, then $m_2^W = \frac{m_2^G}{2m_2^G - m_3^G} > \frac{l+2}{l+1}$. By (2), the EC distribution provided by the simple solution has at most $l+2$ phases. ■

6 Variants of closed form solutions

In this section, we present two refinements of the simple solution (Section 5), which we refer to as the improved solution and the positive solution.

6.1 An improved closed form solution

The improved solution is defined for all the input distributions $G \in \mathcal{PH}_3$ and uses a smaller number of phases than the simple solution. Specifically, we prove that the number of phases required in the improved solution is characterized by the following theorem:

Theorem 5 *Under the improved solution, the number of phases used to well-represent any distribution G by an EC distribution is at most $OPT(G) + 1$.*

Figure 6 is an implementation of the improved solution. In this section, we first elaborate on the improved solution. Then, we prove Theorem 5. In fact, the improved solution can be improved upon yet further for many distributions (see [19] for details), but it does not improve the worst case performance.

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 (Section 5). If $G \notin \mathcal{PH}_3^-$, then to obtain the baseline EC distributing 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 $Z(\cdot) = pW(\cdot) + (1-p)O(\cdot)$. The parameters of the EC distribution that well-represents W are then obtained by the simple solution (Section 5).

Next, we describe an idea to reduce the number of phases used in the baseline EC distribution. The simple solution (Section 5) 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 $\left\{F \middle| \frac{3}{2} \leq m_2^X \leq 2 \text{ and } m_3^X = 2m_2^X - 1\right\}$, then X is well-represented by a two-phase Coxian$^+$ PH distribution.

The above ideas lead to the following solution:

(i) <u>If $G \in \mathcal{U} \cap \mathcal{PH}^-$</u>, then the simple solution (Section 5) provides the parameters $(n,\, p,\, \mu_{X1},\, \mu_{X2},\, p_X)$.

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

Input: the first three moments of a distribution G: μ_1^G, μ_2^G, and μ_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. $\mu_1^W = \frac{\mu_1^G}{p}; \quad m_2^W = pm_2^G; \quad m_3^W = pm_3^G.$

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

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

Fig. 6. *An implementation of the improved closed form solution.*

(ii) If $G \in \mathcal{U} \cap (PH^-)^c$, where $(PH^-)^c$ denotes the complement of $\mathcal{P}H^-$, then let $n = \frac{2m_2^G - 1}{m_2^G - 1}$, $m_2^W = \frac{1}{2}\left(\frac{n-1}{n-2} + \frac{n}{n-1}\right)$, $m_3^W = \frac{m_3^G}{m_2^G} m_2^W$, and $\mu_1^W = \frac{\mu_1^G}{p_W}$, where $p_W = \frac{m_2^W}{m_2^G}$. 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 μ_1^W. The parameters $(n,\, \mu_{X1},\, \mu_{X2},\, p_X)$ of W are provided by the simple solution (Section 5). Also, set $p = p_W$, since W has no mass probability at zero.

(iii) If $G \in \mathcal{M} \cup \mathcal{L}$, then the simple solution (Section 5) provides the parameters $(n,\, p,\, \mu_{X1},\, \mu_{X2},\, p_X)$, except that if the number n of phases calculated by (2) is $n > 2$, then n is decremented by one. The next theorem (Theorem 6) guarantees that parameters obtained with the reduced n are still feasible.

Theorem 6 *Let $Z = A^n(X)$. If $X \in \left\{F \mid \frac{3}{2} \leq m_2^F \leq 2 \text{ and } m_3^F = 2m_2^F - 1\right\}$, then $Z \in \left\{F \mid \frac{n+1}{n} \leq m_2^F \leq \frac{n}{n-1} \text{ and } m_3^F = 2m_2^F - 1\right\}$.*

Proof: By Theorem 3, m_2^Z is a continuous and monotonically increasing function of m_2^X, Thus, $\frac{n+1}{n} \leq m_2^Z \leq \frac{n}{n-1}$ 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 Theorem 4. ■

Now we prove Theorem 5.

Proof:[Theorem 5] Since $S^{(n)} \subset \mathcal{T}^{(n)}$ (Lemma 1), it suffices to prove that if a distribution $G \in \mathcal{T}^{(n)}$, then at most $n+1$ phases are needed. (i) Suppose $G \in \mathcal{U}$. If $G \in \mathcal{T}^{(n)}$, then the simple solution (Section 5) is used and at most $n+1$ phases are used. (ii) Suppose $G \in \mathcal{M}$. If $G \in \mathcal{T}^{(n)} \cap \mathcal{P}H^-$, then the number of phases used in the improved solution is one less than the simple solution. Therefore, at most n phases are used. If $G \in \mathcal{T}^{(n)} \cap (\mathcal{P}H^-)^c$, then exactly n phases are used. (iii) Suppose $G \in \mathcal{L}$. If $G \in \mathcal{T}^{(n)}$, then the number of phases used in the improved solution is one less than the simple solution. Therefore, at most $n+1$ phases are used. ■

Theorem 5 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 $n \geq 2$, $\mathcal{T}^{(n)} \subset \mathcal{S}^{(n+1)}$.*

6.2 A positive closed form solution

The simple solution and the improved solution can have a 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 queueing system analysis [5,21] and approximating shortfall distributions in inventory system analysis [28,29]. 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 the matrix analytic method when the PH distributions have no pass probability at zero; however, no simple analytical solution is known when the PH distributions have nonzero mass probability at zero.

The positive closed form solution is a variant of the improved solution, and it does not have mass probability at zero. The key idea in the design of the positive solution is to match the input distribution by a mixture of an EC distribution (with no mass probability at zero) and an exponential distribution.

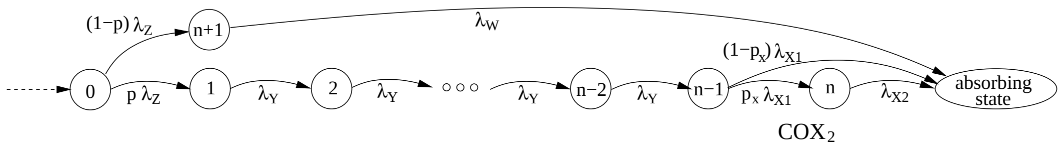


Fig. 7. *The Markov chain underlying an $(n+1)$-phase extended EC distribution.*

The use of this type of extended EC distribution makes intuitive sense, since it can approximate the EC distribution with mass probability at zero arbitrarily closely by letting the mean of the exponential distributions approach zero. It turns out, however, that it is not always easy (or even possible) to find a *closed form* expression for the parameters of the EC distribution and the exponential distribution. We find that in such cases a convolution of an EC distribution and an exponential distribution leads to tractability, and we can find the closed form expression for the parameter of the EC distribution and the exponential distribution. 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 $Z(\cdot) * (X(\cdot) + (1-p)W(\cdot)))$, where Z and W are exponential distributions, and X is an EC distribution with no mass probability at zero. (See Figure 7 for the underlying Markov chain of an extended EC distribution.)*

Note that the parameter n in an extended EC distribution denotes the number of phases in the EC portion of the extended EC distribution. Note also that in the positive solution exactly one of λ_W and λ_Z is set ∞, i.e. the distribution function of the extended EC distribution is $F(\cdot) = Z(\cdot) * X(\cdot)$ or $F(\cdot) = pX(\cdot) + (1-p)W(\cdot)$. Therefore, the total number of phases in an extended EC distribution is $n+1$.

Figure 8 shows an implementation of the positive solution. The extended EC distribution given by the positive solution has no mass probability at zero. The positive solution is defined for almost all the input distributions $\mathcal{PH}_3$; in particular it is defined for all distributions in

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

The number of phases required is characterized by the following theorem:

Theorem 7 *Under the positive solution, the number of phases used to well-represent any distribution G by an extended EC distribution is $\leq OPT(G)+1$.*

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

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

use `Improved`. Otherwise,

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

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

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

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

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

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

$\lambda_Z = \frac{1}{z\mu_1^X}$; $\quad \lambda_W = \infty$; 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. $z = \frac{m_2^G\left((m_3^G - 3) - 2\frac{k+3}{k+2}(m_2^G - 2)\right) + m_2^G\sqrt{(m_3^G-3)^2 + 8\frac{k+3}{k+2}(m_2^G-2)(\frac{3}{2} - r^G)}}{2\frac{k+3}{k+2}(m_2^G-2)^2}$;

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

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

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

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

Fig. 8. *An implementation of the positive closed form solution.*

Closed form solution

When the input distribution is in

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

the EC distribution produced by the improved solution does not have a mass probability at zero. Below, we focus on input distributions $G \in \mathcal{L}^+$ with an additional assumption that $r^G \neq \frac{3}{2}$.

We first consider the first approach of using a mixture of an EC distribution (with no mass probability at zero) and an exponential distribution (i.e. $\lambda_Z = \infty$). Given a distribution $G \in \mathcal{L}_k$ (see Definition 13), we seek m_2^X, m_3^X, $0 < p < 1$, and $w > 0$, where $w = \frac{\mu_1^W}{\mu_1^X}$, such that

$$\frac{k+2}{k+1} \le m_2^X < \frac{k+1}{k} \quad (3)$$

$$m_3^X = 2m_2^X - 1 \quad (4)$$

$$m_2^G = \frac{pm_2^X + 2(1-p)w^2}{(p+(1-p)w)^2} \quad (5)$$

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

Note that G is then well-represented by a distribution with $pX(\cdot)+(1-p)W(\cdot)$, where X is an EC distribution with no mass probability at zero and W is an exponential distribution.

The following lemma characterizes the parameters of distributions, X and W.

Lemma 5 *Suppose* $G \in \mathcal{L}_k$ *and* $\frac{(k+1)m_2^G+(k+4)}{2(k+2)} \le r^G < \frac{3}{2}$ *for* $k \ge 1$. *Let* $w = \frac{2-m_2^G}{4\left(\frac{3}{2}-r^G\right)}$, $p = \frac{(2-m_2^G)^2}{(2-m_2^G)^2+4(2m_2^G-1-m_3^G)}$, $m_2^X = 2w$, *and* $m_3^X = 2m_2^X - 1$. *Then,* $w > 0$, $0 < p < 1$, *and conditions (3)-(6) are satisfied.*

Proof: It is easy to check, by substitution, that conditions (4)-(6) are satisfied. It is easy to see $0 < p < 1$, since $m_3^G < 2m_2^G - 1$. Also, $m_2^X \ge \frac{k+2}{k+1}$ implies $w > 0$. Thus, it suffices to prove that condition (3) is satisfied.

We first consider the first inequality of condition (3). The assumption on r^G in the lemma gives $\frac{3}{2} - r^G \le \frac{3}{2} - \frac{(k+1)m_2^G+(k+4)}{2(k+2)} = \frac{k+1}{k+2} \cdot \frac{2-m_2^G}{2}$. Therefore, since $\frac{3}{2} > r^G$, it follows that $m_2^X = 2w = \frac{2-m_2^G}{2} \cdot \frac{1}{\frac{3}{2}-r^G} \ge \frac{k+2}{k+1}$.

We next consider the second inequality of condition (3). We begin by bounding the range of m_2^G for G considered in the lemma. Condition $G \in \mathcal{L}_k$ implies $m_2^G \ge \frac{k+2}{k+1}$. Also, if $m_2^G > 2$, then by the assumption on r^G in lemma,

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

This contradicts $r^G < \frac{3}{2}$. Thus, $m_2^G \le 2$. So far, we derived the range of m_2^G as $\frac{k+2}{k+1} < m_2^G \le 2$.

We prove $m_2^X < \frac{k+1}{k}$ in two cases: (i) $\frac{k+1}{k} \le m_2^G \le 2$ and (ii) $\frac{k+2}{k+1} \le m_2^G < \frac{k+1}{k}$. Note that $m_2^X = \frac{2-m_2^G}{2\left(\frac{3}{2}-r^G\right)}$. (i) When $\frac{k+1}{k} \le m_2^G \le 2$,

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

The inequality follows from $m_2^G < \frac{k+1}{k}$ and $r^G \le \frac{k+2}{k+1}$. (ii) When $\frac{k+2}{k+1} \le m_2^G <$

$\frac{k+1}{k}$,

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

The inequality follows from $r^G = \frac{m_3^G}{m_2^G} < \frac{2m_2^G - 1}{m_2^G}$, which follows from $G \in \mathcal{L}_k$. This completes the proof. ■

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, and the solutions are not unique. By fixing some of the parameters, the system of equations can be reduced to have a unique solution; however, the system of equations is not necessarily simplified enough to provide simple closed form solutions. 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, r^G is significantly simplified: $r^G = \frac{2pr^X + 3(1-p)w}{2(p+(1-p)w)}$. Now, finding p and w is a relatively easy task. 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 a 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 lest of the input distribution, in fact, lead to simpler closed form expressions for solutions by both approaches.

Next, we consider the second approach of using a convolution of an EC distribution (with mass probability at zero) and an exponential distribution (i.e. $\lambda_W = \infty$). Given a distribution $G \in \mathcal{L}_k$ (we assume $\frac{m_3^G}{m_2^G} \neq \frac{3}{2}$), we seek m_2^X, m_3^X, and $z > 0$, where $z = \frac{\mu_1^Z}{\mu_1^X}$ such that

$$m_2^X \geq \frac{k+2}{k+1} \quad (7) \qquad m_2^G = \frac{m_2^X + 2z + 2z^2}{(1+z)^2} \quad (9)$$

$$m_3^X = \frac{k+3}{k+2} m_2^X \quad (8) \qquad m_3^G = \frac{m_2^X m_3^X + 3m_2^X z + 6z^2 + 6z^3}{(m_2^X + 2z + 2z^2)(1+z)} \quad (10)$$

Note that G is then well-represented by a distribution with $X(\cdot) * Z(\cdot)$, where X is an EC distribution and Z is an exponential distribution.

The following lemma characterizes the parameters of distributions, X and Z.

Lemma 6 *Suppose $G \in \mathcal{L}_k$, $r^G < \frac{3}{2}$, and $r^G < \frac{(k+1)m_2^G + (k+4)}{2(k+2)}$ for $k \geq 1$. If $m_2^G = 2$, we choose $z = \frac{m_3^G - 2\frac{k+3}{k+2}}{3 - m_3^G}$, $m_2^X = 2(1+z)$, and $m_3^X = \frac{k+3}{k+2} m_2^X$. If*

$m_2^G \neq 2$, we choose $m_2^X = (1+z)\Big(m_2^G(1+z) - 2z\Big)$ and $m_3^X = \frac{k+3}{k+2} m_2^X$, where

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

Then, $z > 0$ and conditions (7)-(10) are satisfied.

Proof: For each case, it is easy to check, by substitution, that conditions (8)-(10) are satisfied. Below, we prove condition (7) and $z > 0$.

We begin with the first case, where $m_2^G = 2$. It is easy to see (7) is true if $z > 0$, since $m_2^X = 2(1+z) > 2 > \frac{k+2}{k+1}$. However, $z > 0$ if $2\frac{k+3}{k+2} < m_3^G < 3$, which is true by $G \in \mathcal{L}_k$, $r^G < \frac{3}{2}$, and $m_2^G = 2$.

Below, we consider the second 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

$$\begin{aligned}
& m_3^G(m_2^X + 2z + 2z^2)(1+z) = m_2^X m_3^X + 3m_2^X z + 6z^2 + 6z^3 \quad \text{(by (10))} \\
\iff\ & m_3^G m_2^G (1+z)^3 = \frac{k+3}{k+2}(m_2^X)^2 + 3z(m_2^X + 2z + 2z^2) \quad \text{(by (8) and (9))} \\
\iff\ & m_3^G m_2^G (1+z)^3 = \frac{k+3}{k+2}(1+z)^2\left(m_2^G(1+z) - 2z\right)^2 + 3z(1+z)^2 m_2^G \quad \text{(by (9))} \\
\iff\ & m_3^G m_2^G (1+z) = \frac{k+3}{k+2}\left(m_2^G(1+z) - 2z\right)^2 + 3z m_2^G
\end{aligned}$$

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

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

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

Second, we show $m_2^X \leq \frac{k+2}{k+1}$. 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 (9), $m_2^X = z\Big((m_2^G - 2)z + 2(m_2^G - 1)\Big) + m_2^G$. Thus, if $m_2^G \geq 2$, then $m_2^X \geq m_2^G \geq \frac{k+2}{k+1}$. 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 (9). Thus, $m_2^X \geq \frac{k+2}{k+1}$ iff $(0 < z) \leq z^*$, where z^* is a larger solution, x, of the following quadratic equation: $\chi(x) = 0$, where

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

That is, $z^* \equiv \frac{m_2^G - 1 + \sqrt{\frac{k+2}{k+1} m_2^G - \frac{k+3}{k+1}}}{2 - m_2^G}$. Thus, it suffices to show $f(z^*) \geq 0$. Since $\chi(z^*) = 0$, we get $(z^*)^2$ as a linear function of z^*: $(z^*)^2 = \frac{2(m_2^G-1)z^* + m_2^G - \frac{k+2}{k+1}}{2-m_2^G}$. By substituting this $(z^*)^2$ into the expression for $f(z^*)$, we get

$$
\begin{aligned}
f(z^*) = & \frac{k+3}{k+2}(2-m_2^G)^2 \left(\frac{2(m_2^G-1)z^* + m_2^G - \frac{k+2}{k+1}}{2-m_2^G} \right) \\
& - m_2^G \left(2\frac{k+3}{k+2}(2-m_2^G) - (3-m_3^G) \right) z - (m_2^G)^2 \left(r^G - \frac{k+3}{k+2} \right) = 0 \\
= & \left(3m_2^G - 2\frac{k+3}{k+2}(2-m_2^G) - m_2^G m_3^G \right) z^* + 2\frac{k+3}{k+2} m_2^G - \frac{k+3}{k+1}(2-m_2^G) - m_2^G m_3^G \\
> & \left(3m_2^G - 2\frac{k+3}{k+2}(2-m_2^G) - (m_2^G)^2 \left(\frac{(k+1)m_2^G + (k+4)}{2(k+2)} \right) \right) z^* \\
& + 2\frac{k+3}{k+2} m_2^G - \frac{k+3}{k+1}(2-m_2^G) - (m_2^G)^2 \left(\frac{(k+1)m_2^G + (k+4)}{2(k+2)} \right) \\
= & \frac{2-m_2^G}{2(k+2)} \left((k+1)(m_2^G)^2 + 3(k+2)m_2^G - 4(k+3) \right) z^* \\
& - \frac{(k+1)m_2^G - (k+2)}{2(k+1)(k+2)} \left((k+1)(m_2^G)^2 + (2k+6)m_2^G - 4(k+3) \right)
\end{aligned}
$$

where the inequality follows from the assumption on r^G in the lemma. By substituting $z^* = \frac{m_2^G - 1 + \sqrt{\frac{k+2}{k+1} m_2^G - \frac{k+3}{k+1}}}{2 - m_2^G}$ into the last expression, we obtain $f(z^*) = g(m_2^G) + h(m_2^G)\sqrt{\frac{(k+2)m_2^G - (k+3)}{k+1}}$, where $g(m_2^G) \equiv \frac{(k+1)^2(m_2^G)^2 - (k-3)(k+2)m_2^G - 4(k+3)}{2(k+1)(k+2)}$ and $h(m_2^G) \equiv \frac{(k+1)(m_2^G)^2 + 3(k+2)m_2^G - 4(k+3)}{2(k+2)}$. Since

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

for $\frac{k+2}{k+1} \leq m_2^G < 2$, $g(m_2^G)$ and $h(m_2^G)$ are increasing functions of m_2^G in the range of $\frac{k+2}{k+1} \leq m_2^G < 2$. Since $g(\frac{k+2}{k+1}) = \frac{2}{(k+1)^2(k+2)} > 0$ and $h(\frac{k+2}{k+1}) = \frac{2}{k^2+3k+2} > 0$, we have $g(m_2^G) \geq 0$ and $h(m_2^G) \geq 0$ for all $\frac{k+2}{k+1} \leq m_2^G < 2$. This implies $f(z^*) \geq 0$. This completes the proof. ■

Finally, we prove Theorem 7.

Proof:[Theorem 7] When an input distribution, G, satisfies $m_3^G \geq 2m_2^G - 1$, the positive solution is the same as the improved solution (Section 6.1), and hence requires the same number of phases, which is at most $OPT(G) + 1$. When G satisfies $m_3^G < 2m_2^G - 1$, it is immediate, from the construction of the solution, that the positive solution requires at most one more phase than the improved solution. For this G, the improved solution requires $OPT(G)$

phases, and hence the positive solution requires $OPT(G)+1$ phases. ■

7 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 P, (iii) the number of phases used in P is at most $OPT(G)+c$, where c is a small constant, (iv) the solution is expressed in closed form.

The key idea is the definition and use of EC distributions, a subset of PH distributions. The set of EC distributions is defined so that it 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), (ii), and (iii). Also, the EC distribution is defined so that it has a small number (six) of 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 characterization is enabled by the simple definition of the EC distribution. 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.

One *take-home lesson* from this paper is that the moment-matching problem is better solved with respect to the above four goals by sewing together two or more types of distributions, so that one can gain the best properties of both. The EC distribution sews the two-phase Coxian PH distribution and the Erlang distribution. The point is that these two distributions provide several different and complementary desirable properties.

The second 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, 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 benefitted from exactly this point in this paper. Another benefit of characterizing $\mathcal{S}^{(n)}$ is that some existing moment matching algorithms, such as Johnson and Taaffe's nonlinear programming approach [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.

The closed form solutions proposed in this paper have been largely implemented and tested. Latest implementation of the solutions is available at an online code repository, http://www.cs.cmu.edu/~osogami/code/.

Acknowledgement

We would like to thank Miklos Telek for his help in improving the presentation and quality of this paper.

References

[1] D. Aldous and L. Shepp. The least variable phase type distribution is Erlang. *Communications in Statistics - Stotchastic Models*, 3:467 – 473, 1987.

[2] T. Altiok. On the phase-type approximations of general distributions. *IIE Transactions*, 17:110 – 116, 1985.

[3] A. Feldmann and W. Whitt. Fitting mixtures of exponentials to long-tail distributions to analyze network performance models. *Performance Evaluation*, 32:245 – 279, 1998.

[4] H. Franke, J. Jann, J. Moreira, P. Pattnaik, and M. Jette. An evaluation of parallel job scheduling for ASCI blue-pacific. In *Proceedings of Supercomputing '99*, pages 679 – 691, November 1999.

[5] M. Harchol-Balter, C. Li, T. Osogami, A. Scheller-Wolf, and M. S. Squillante. Analysis of task assignment with cycle stealing under central queue. In *Proceedings of ICDCS '03*, pages 628–637, May 2003.

[6] A. Horváth and M. Telek. Approximating heavy tailed behavior with phase type distributions. In *Advances in Matrix-Analytic Methods for Stochastic Models*, pages 191 – 214. Notable Publications, July 2000.

[7] A. Horváth and M. Telek. Phfit: A general phase-type fitting tool. In *Proceedings of Performance TOOLS 2002*, pages 82 – 91, April 2002.

[8] M. A. Johnson. Selecting parameters of phase distributions: Combining nonlinear programming, heuristics, and Erlang distributions. *ORSA Journal on Computing*, 5:69 – 83, 1993.

[9] M. A. Johnson and M. F. Taaffe. An investigation of phase-distribution moment-matching algorithms for use in queueing models. *Queueing Systems*, 8:129 – 147, 1991.

[10] M. A. Johnson and M. R. Taaffe. Matching moments to phase distributions: Mixtures of Erlang distributions of common order. *Communications in Statistics — Stochastic Models*, 5:711 – 743, 1989.

[11] M. A. Johnson and M. R. Taaffe. Matching moments to phase distributions: Density function shapes. *Communications in Statistics — Stochastic Models*, 6:283 – 306, 1990.

[12] M. A. Johnson and M. R. Taaffe. Matching moments to phase distributions: Nonlinear programming approaches. *Communications in Statistics — Stochastic Models*, 6:259 – 281, 1990.

[13] S. Karlin and W. Studden. *Tchebycheff Systems: With Applications in Analysis and Statistics*. John Wiley and Sons, 1966.

[14] R. E. A. Khayari, R. Sadre, and B. Haverkort. Fitting world-wide web request traces with the EM-algorithm. *Performance Evalutation*, 52:175 – 191, 2003.

[15] R. Marie. Calculating equilibrium probabilities for $\lambda(n)/c_k/1/n$ queues. In *Proceedings of* Performance 1980, pages 117 – 125, 1980.

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

[17] C. A. O'Cinneide. Phase-type distributions and majorization. *Annals of Applied Probability*, 1(3):219 – 227, 1991.

[18] T. Osogami and M. Harchol-Balter. A closed-form solution for mapping general distributions to minimal PH distributions. In *Proceedings of TOOLS 2003*, pages 200–217, September 2003.

[19] T. Osogami and M. Harchol-Balter. A closed-form solution for mapping general distributions to minimal PH distributions. Technical Report CMU-CS-03-114, School of Computer Science, Carnegie Mellon University, 2003.

[20] T. Osogami and M. Harchol-Balter. Necessary and sufficient conditions for representing general distributions by Coxians. In *Proceedings of TOOLS '03*, pages 182–199, September 2003.

[21] T. Osogami, M. Harchol-Balter, and A. Scheller-Wolf. Analysis of cycle stealing with switching cost. In *Proceedings of Sigmetrics '03*, pages 184–195, June 2003.

[22] A. Riska, V. Diev, and E. Smirni. Efficient fitting of long-tailed data sets into PH distributions. *Performance Evaluation*, 2003 (to appear).

[23] C. Sauer and K. Chandy. Approximate analysis of central server models. *IBM Journal of Research and Development*, 19:301 – 313, 1975.

[24] L. Schmickler. Meda: Mixed Erlang distributions as phase-type representations of empirical distribution functions. *Communications in Statistics — Stochastic Models*, 8:131 – 156, 1992.

[25] M. Squillante. Matrix-analytic methods in stochastic parallel-server scheduling models. In *Advances in Matrix-Analytic Methods for Stochastic Models*. Notable Publications, July 1998.

[26] D. Starobinski and M. Sidi. Modeling and analysis of power-tail distributions via classical teletraffic methods. *Queueing Systems*, 36:243 – 267, 2000.

[27] M. Telek and A. Heindl. Matching moments for acyclic discrete and continuous phase-type distributions of second order. *International Journal of Simulation*, 3:47 – 57, 2003.

[28] G. van Houtum and W. Zijm. Computational procedures for stochastic multi-echelon production systems. *International Journal of Production Economics*, 23:223 – 237, 1991.

[29] G. van Houtum and W. Zijm. Incomplete convolutions in production and inventory models. *OR Spektrum*, 10:97 – 107, 1997.

[30] W. Whitt. Approximating a point process by a renewal process: Two basic methods. *Operations Research*, 30:125 – 147, 1982.

[31] Y. Zhang, H. Franke, J. Moreira, and A. Sivasubramaniam. An integrated approach to parallel scheduling using gang-scheduling, backfilling, and migration. *IEEE Transactions on Parallel and Distributed Systems*, 14:236 – 247, 2003.