Necessary and Sufficient Conditions for Representing General Distributions by Coxians

Takayuki Osogami[1] and Mor Harchol-Balter[1]

Department of Computer Science, Carnegie Mellon University
5000 Forbes Avenue, Pittsburgh, PA 15213, USA
{osogami, harchol}@cs.cmu.edu

Abstract. A common analytical technique involves using a Coxian distribution to model a general distribution G, where the Coxian distribution agrees with G on the first three moments. This technique is motivated by the analytical tractability of the Coxian distribution. Algorithms for mapping an input distribution G to a Coxian distribution largely hinge on knowing a priori the necessary and sufficient number of phases in the representative Coxian distribution. In this paper, we formally characterize the set of distributions G which are well-represented by an n-phase Coxian distribution, in the sense that the Coxian distribution matches the first three moments of G. We also discuss a few common, practical examples.

1 Introduction

Background Approximating general distributions by phase-type (PH) distributions has significant application in the analysis of stochastic processes. Many fundamental problems in queueing theory are hard to solve when general distributions are allowed as inputs. For example, the waiting time for an M/G/c queue has no nice closed formula when $c > 1$, while the waiting time for an M/M/c queue is trivially solved. Tractability of M/M/c queues is attributed to the memoryless property of the exponential distribution. A popular approach to analyzing queueing systems involving a general distribution G is to approximate G by a PH distribution. A PH distribution is a very general mixture of exponential distributions, as shown in Figure 1 [21]. The Markovian nature of the PH distribution frequently allows a Markov chain representation of the queueing system. Once the system is represented by a Markov chain, this chain can often be solved by matrix-analytic methods [18, 21], or other means.

When fitting a general distribution G to a PH distribution, it is common to look for a PH distribution which matches the first three moments of G. In this paper, we say that:

Definition 1. *A distribution G is* well-represented *by a distribution F if F 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

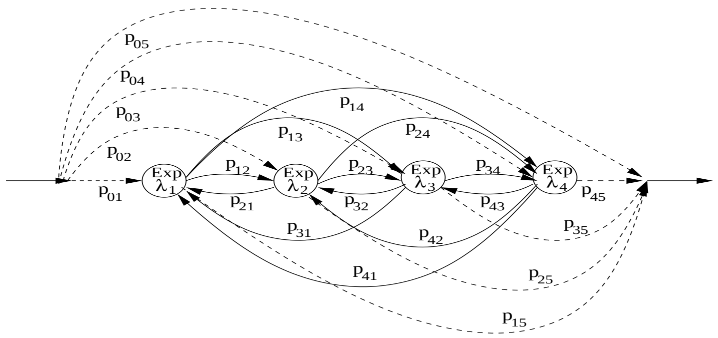

Fig. 1. A PH distribution is the distribution of the absorption time in a finite state continuous time Markov chain. The figure shows a 4-phase PH distribution. There are $n = 4$ states, where the ith state has exponentially-distributed sojourn time with rate λ_i. With probability p_{0i} we start in the ith state, and the next state is state j with probability p_{ij}. Each state i has probability p_{i5} that the next state will be the absorbing state. The absorption time is the sum of the times spent in each of the states.

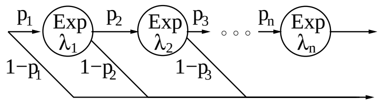

Fig. 2. An n-phase Coxian distribution is a particular n-phase PH distribution whose underlying Markov chain is of the form in the figure, where $0 \le p_i \le 1$ and $\lambda_i > 0$ for all $0 \le i \le n$.

to be effective in predicting mean performance for variety of many computer system models [7, 10, 23, 29, 33]. Clearly, however, three moments might not always suffice for every problem, and we leave the problem of matching more moments to future work.

Most existing algorithms for fitting a general distribution G to a PH distribution, restrict their attention to a subset of PH distributions, since general PH distributions have so many parameters that it is difficult to find time-efficient algorithms for fitting to the general PH distributions [14, 15, 20, 27, 32]. The most commonly chosen subset is the class of Coxian distributions, shown in Figure 2. Coxian distributions have the advantage of being much simpler than general PH distributions, while including a large subset of PH distributions without needing additional phases. For example, for any acyclic PH distribution P_n, there exists a Coxian distribution C_n with the same number of phases such that P_n and C_n have the same distribution function [5]. In this paper we will restrict our attention to Coxian distributions.

Motivation and Goal When finding a Coxian distribution C which well-represents a given distribution G, it is desirable that C be *minimal*, i.e., the number of phases in C be as small as possible. This is important because it minimizes the additional states necessary in the resulting Markov chain for the queueing system. Unfortunately, it is not known what is the minimal number of phases necessary to well-represent a given distribution G by a Coxian distribution. This makes it difficult to evaluate the effectiveness of different algorithms and also makes the design of fitting algorithms open-ended.

The *primary goal* of this paper is to characterize the set of distributions which are well-represented by an n-phase Coxian distribution, for each $n = 1, 2, 3, \ldots$.

Definition 2. *Let $\mathcal{S}^{(n)}$ denote the set of distributions that are well-represented by an n-phase Coxian distribution for positive integer n.*

Our characterization of $\{\mathcal{S}^{(n)}, n \geq 1\}$ will allow one to determine, for any distribution G, the minimal number of phases that are needed to well-represent G by a Coxian distribution.[1] Such a characterization will be a useful guideline for designing algorithms which fit general distributions to Coxian distributions. Another application of this characterization is that some existing fitting algorithms, such as Johnson and Taaffe's nonlinear programming approach [15], require knowing the number of phases n in the minimal Coxian distribution. The current approach involves simply iterating over all choices for n [15], whereas our characterization would immediately specify n.

Providing sufficient and necessary conditions for a distribution to be in $\mathcal{S}^{(n)}$ does not always immediately give one a sense of *which* distributions satisfy those conditions, or of the magnitude of the set of distributions which satisfy the condition. A *secondary goal* of this paper is to provide examples of common distributions which are included in $\mathcal{S}^{(n)}$ for particular integers n.

In finding *simple* characterizations of $\mathcal{S}^{(n)}$, 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*

$$m_2^F = \frac{\mu_2^F}{(\mu_1^F)^2} \quad \text{and} \quad m_3^F = \frac{\mu_3^F}{\mu_1^F \mu_2^F}.$$

Notice the correspondence to the coefficient of variability C_F and skewness γ_F of F: $m_2^F = C_F^2 + 1$ and $m_3^F = \nu_F \sqrt{m_2^F}$, where $\nu_F = \frac{\mu_3^F}{(\mu_2^F)^{3/2}}$. (Notice the correspondence between ν_F and the skewness of F, γ_F, where $\gamma_F = \frac{\bar{\mu}_3^F}{(\bar{\mu}_2^F)^{3/2}}$ and $\bar{\mu}_k^F$ is the centralized k-th moment of F for $k = 2, 3$.)

[1] One might initially argue that $\mathcal{S}^{(2)}$, the set of distributions well-represented by a two-phase Coxian distribution, should include all distributions, since a 2-phase Coxian distribution has four parameters (p_1, p_2, λ_1, λ_2), whereas we only need to match three moments of G. A simple counter example shows this argument to be false. Let G be a distribution whose first three moments are 1, 2, and 12. The system of equations for matching G to a 2-phase Coxian distribution with three parameters (λ_1, λ_2, p) results in either λ_1 or λ_2 being negative.

Relevant Previous Work 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$^+$ distribution, where a Coxian$^+$ distribution is simply a Coxian distribution with no mass probability at zero, i.e. $p_1 = 1$. Observe $\mathcal{S}^{(2)^*} \subset \mathcal{S}^{(2)}$. Altiok [2] showed a sufficient condition for a distribution G to be in $\mathcal{S}^{(2)^*}$. More recently, Telek and Heindl [31] expanded Altiok's condition and proved the necessary and sufficient condition for a distribution G 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, as shown in Theorem 1. In this paper, we will characterize $\mathcal{S}^{(2)}$, as well as characterizing $\mathcal{S}^{(n)}$, for all integers $n \geq 2$.

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

Definition 4. *For integers $n \geq 2$, let $\mathcal{S}_{\mathcal{V}}{}^{(n)}$ denote the set of distributions, F, with the following property on their normalized moments:*

$$m_2^F > \frac{n}{n-1} \quad \text{and} \quad m_3^F \geq \frac{n+2}{n+1} m_2^F. \tag{1}$$

The main contribution of this paper is a derivation of the nested relationship between $\mathcal{S}_{\mathcal{V}}{}^{(n)}$ and $\mathcal{S}^{(n)}$ for all $n \geq 2$. This relationship is illustrated in Figure 3 and proven in Section 3. There are three points to observe: (i) $\mathcal{S}^{(n)}$ is a proper subset of $\mathcal{S}^{(n+1)}$ for all integers $n \geq 2$, and likewise $\mathcal{S}_{\mathcal{V}}{}^{(n)}$ is a proper subset of $\mathcal{S}_{\mathcal{V}}{}^{(n+1)}$; (ii) $\mathcal{S}_{\mathcal{V}}{}^{(n)}$ is contained in $\mathcal{S}^{(n)}$ and close to $\mathcal{S}^{(n)}$ in size; providing a simple characterization for $\mathcal{S}^{(n)}$; (iii) $\mathcal{S}^{(n)}$ is almost contained in $\mathcal{S}_{\mathcal{V}}{}^{(n+1)}$ for all integers $n \geq 2$ (more precisely, we will show $\mathcal{S}^{(n)} \subset \mathcal{S}_{\mathcal{V}}{}^{(n+1)} \cup \mathcal{E}^{(n)}$, where $\mathcal{E}^{(n)}$ is the set of distributions well-represented by an Erlang-n distribution). This result yields a necessary number and a sufficient number of phases for a given distribution to be well-represented by a Coxian distribution. Additional contributions of the paper are described below.

With respect to the set $\mathcal{S}^{(2)}$, we derive the exact necessary and sufficient condition for a distribution G to be in $\mathcal{S}^{(2)}$ as a function of the normalized moments of G. This complements the results of Telek and Heindl, who analyzed $\mathcal{S}^{(2)^*}$, which is a subset of $\mathcal{S}^{(2)}$. (See Section 2).

Lastly, we provide a few examples of common, practical distributions included in the set $\mathcal{S}_{\mathcal{V}}{}^{(n)} \subset \mathcal{S}^{(n)}$. All distributions we consider have finite third moment. The Pareto distribution and the Bounded Pareto distribution (as defined in [8]) have been shown to fit many recent measurements of job service requirements in computing systems, including the file size requested by HTTP requests [3,4], the CPU requirements of UNIX jobs [9,19], and the duration of FTP transfers [24]. We show that a large subset of Bounded Pareto distributions is in $\mathcal{S}_{\mathcal{V}}{}^{(2)}$.

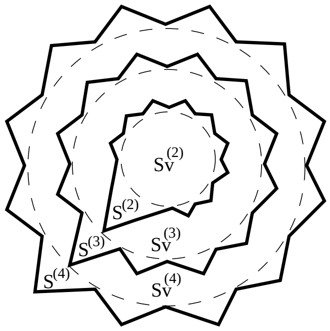

Fig. 3. The main contribution of this paper: a simple characterization of $\mathcal{S}^{(n)}$ by $\mathcal{S_V}^{(n)}$. Solid lines delineate $\mathcal{S}^{(n)}$ (which is irregular) and dashed lines delineate $\mathcal{S_V}^{(n)}$ (which is regular – has a simple specification). Observe the nested structure of $\mathcal{S}^{(n)}$ and $\mathcal{S_V}^{(n)}$. $\mathcal{S_V}^{(n)}$ is close to $\mathcal{S}^{(n)}$ in size and is contained in $\mathcal{S}^{(n)}$. $\mathcal{S}^{(n)}$ is almost contained in $\mathcal{S_V}^{(n+1)}$.

We also provide conditions under which the Pareto and uniform distributions are in $\mathcal{S_V}^{(n)}$ for each $n \geq 2$. (See Section 4).[2]

2 Full Characterization of $\mathcal{S}^{(2)}$

The Telek and Heindl [31] result may be expressed in terms of normalized moments as follows:

Theorem 1 (Telek, Heindl). *$G \in \mathcal{S}^{(2)^*}$ iff G is in the following union of sets:*

$$\left\{ F \middle| \frac{9m_2^F - 12 + 3\sqrt{2}(2-m_2^F)^{\frac{3}{2}}}{m_2^F} \leq m_3^F \leq \frac{6(m_2^F-1)}{m_2^F} \text{ and } \frac{3}{2} \leq m_2^F < 2 \right\}$$
$$\bigcup \left\{ F \middle| m_3^F = 3 \text{ and } m_2^F = 2 \right\} \bigcup \left\{ F \middle| \frac{3}{2} m_2^F < m_3^F \text{ and } 2 < m_2^F \right\}.$$

We now show a simple characterization for $\mathcal{S}^{(2)}$:

Theorem 2. *$G \in \mathcal{S}^{(2)}$ iff G is in the following union of sets:*

$$\left\{ F \middle| \frac{4}{3} m_2^F \leq m_3^F \leq \frac{6(m_2^F-1)}{m_2^F} \text{ and } \frac{3}{2} \leq m_2^F \leq 2 \right\} \quad \bigcup \quad \mathcal{S_V}^{(2)}, \tag{2}$$

where recall $\mathcal{S_V}^{(2)}$ is the set: $\left\{ F \middle| \frac{4}{3} m_2^F \leq m_3^F \text{ and } 2 < m_2^F \right\}$.

[2] Our results show that the *first three moments* of the Bounded Pareto distribution and the Pareto distribution are matched by a Coxian distribution with a small number of phases. Note however that this does not necessarily imply that the *shape* of these distributions is well-matched by a Coxian distribution with few phases, since the tail of these distributions is not exponential. Fitting the *shape* of heavy-tailed distributions by phase-type distributions such as PH distributions is studied in several recent papers [6, 11, 12, 17, 26, 30].

A summary of Theorems 1 and 2 is shown in Figure 4. Figure 4(a) illustrates how close $\mathcal{S}^{(2)}$ and $\mathcal{S_V}^{(2)}$ are in size. Figure 4(b) shows the distributions which are in $\mathcal{S}^{(2)}$ but not $\mathcal{S}^{(2)^*}$.

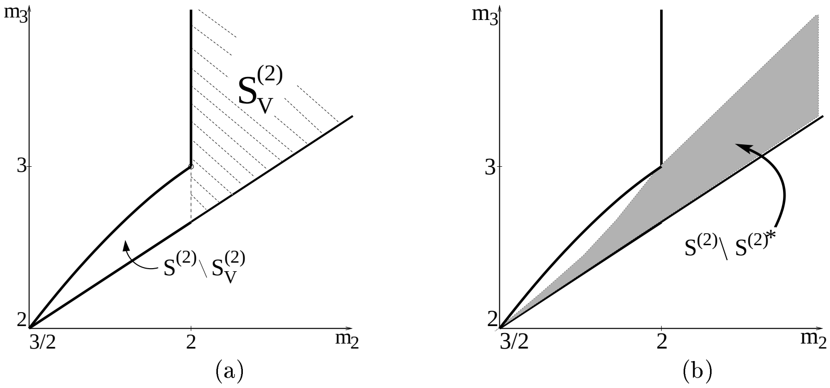

Fig. 4. (a) The thick solid lines delineate $\mathcal{S}^{(2)}$. The striped region shows $\mathcal{S_V}^{(2)} \subset \mathcal{S}^{(2)}$. (b) Again, the thick solid lines delineate $\mathcal{S}^{(2)}$. The shaded region shows $\mathcal{S}^{(2)} \setminus \mathcal{S}^{(2)^*}$.

Proof (Theorem 2). The theorem will be proved by reducing $\mathcal{S}^{(2)}$ to $\mathcal{S}^{(2)^*}$ and employing Theorem 1. The proof hinges on the following observation: an arbitrary distribution $G \in \mathcal{S}^{(2)}$ iff G is well-represented by some distribution[3] $Z(\cdot) = X(\cdot)p + 1 - p$ for some $X \in \mathcal{S}^{(2)^*}$. It therefore suffices to show that Z is in the set defined in (2).

By Theorem 1, since $X \in \mathcal{S}^{(2)^*}$, X is in the following union of sets:

$$\left\{F \,\middle|\, \frac{9m_2^F - 12 + 3\sqrt{2}(2 - m_2^F)^{\frac{3}{2}}}{m_2^F} \le m_3^F \le \frac{6(m_2^F - 1)}{m_2^F} \text{ and } \frac{3}{2} \le m_2^F < 2\right\}$$
$$\bigcup \left\{F \,\middle|\, m_3^F = 3 \text{ and } m_2^F = 2\right\} \bigcup \left\{F \,\middle|\, \frac{3}{2}m_2^F < m_3^F \text{ and } 2 < m_2^F\right\}.$$

Observe that $m_k^Z = \frac{m_k^X}{p}$ for $k = 2, 3$. Thus, Z is in the following union of sets:

$$\left\{F \,\middle|\, \exists p, \frac{9pm_2^F - 12 + 3\sqrt{2}(2 - pm_2^F)^{\frac{3}{2}}}{p^2 m_2^F} \le m_3^F \le \frac{6(pm_2^F - 1)}{p^2 m_2^F} \text{ and } \frac{3}{2p} \le m_2^F < \frac{2}{p}\right\}$$
$$\bigcup \left\{F \,\middle|\, \exists p, m_3^F = \frac{3}{p} \text{ and } m_2^F = \frac{2}{p}\right\} \bigcup \left\{F \,\middle|\, \exists p, \frac{3}{2}m_2^F < m_3^F \text{ and } \frac{2}{p} < m_2^F\right\} \tag{3}$$

[3] To shed light on this expression, consider random variables V_X whose distribution is X. Then random variable

$$V_Z = \begin{cases} V_X & \text{with probability } p \\ 0 & \text{with probability } 1 - p, \end{cases}$$

has distribution Z, since $\Pr(V_Z < t) = p\Pr(V_X < t) + (1 - p)$.

We want to show that Z is in the set defined in (2). To do this, we rewrite the set defined in (2) as:

$$\left\{F \middle| \frac{4}{3}m_2^F \le m_3^F \le \frac{6(m_2^F - 1)}{m_2^F} \text{ and } \frac{3}{2} \le m_2^F \le 2\right\}$$
$$\bigcup \left\{F \middle| \frac{4}{3}m_2^F \le m_3^F \le \frac{3}{2}m_2^F \text{ and } 2 < m_2^F\right\} \bigcup \left\{F \middle| \frac{3}{2}m_2^F < m_3^F \text{ and } 2 < m_2^F\right\}. \quad (4)$$

Observe that (3) and (4) are now in similar forms. We now prove that the set defined in (3) is a subset of the set defined in (4), and the set defined in (4) is a subset of the set defined in (3). The technical details are postponed to Appendix A, Lemma 3. □

3 A Characterization of $\mathcal{S}^{(n)}$

In this section, we prove that ${\mathcal{S}_\mathcal{V}}^{(n)}$ is contained in $\mathcal{S}^{(n)}$, where ${\mathcal{S}_\mathcal{V}}^{(n)}$ is the set of distributions whose normalized moments satisfy (1), and that $\mathcal{S}^{(n)}$ is almost contained in ${\mathcal{S}_\mathcal{V}}^{(n+1)}$. Figure 5 provides a graphical view of the ${\mathcal{S}_\mathcal{V}}^{(n)}$ sets with respect to the normalized moments. Figure 5 illuminates several points. First, there is a nested relationship between ${\mathcal{S}_\mathcal{V}}^{(n)}$ and ${\mathcal{S}_\mathcal{V}}^{(n-1)}$. This makes intuitive sense, since an n-phase Coxian can represent at least as many distributions as an $(n-1)$-phase Coxian. Next, observe that as either m_2^G or m_3^G decreases, more phases are needed to well-represent G. The intuition behind this is that the lower normalized moments, m_2 and m_3, imply moving towards a deterministic distribution (which has the minimum possible values of m_2 and m_3), and a deterministic distribution is well-known to require an infinite number of phases. On the flip side, for distributions with sufficiently high m_2 and m_3, two phases are all that is needed, since high m_2 and m_3 can be achieved by mixing two exponentials with very different rates. We prove the following theorem:

Theorem 3. *${\mathcal{S}_\mathcal{V}}^{(n)} \subset \mathcal{S}^{(n)} \subset {\mathcal{S}_\mathcal{V}}^{(n+1)} \cup \mathcal{E}^{(n)}$, where $\mathcal{E}^{(n)}$ is the set of distributions that are well-represented by an Erlang-n distribution for integers $n \ge 2$.*

An Erlang-n distribution refers to the distribution of a random variable, which is equal to the sum of n i.i.d. exponential random variables. Notice that the normalized moments of distributions in $\mathcal{E}^{(n)}$, $m_2^{\mathcal{E}^{(n)}}$ and $m_3^{\mathcal{E}^{(n)}}$, satisfy the following conditions:

$$m_2^{\mathcal{E}^{(n)}} = \frac{n+1}{n} \quad \text{and} \quad m_3^{\mathcal{E}^{(n)}} = \frac{n+2}{n}. \quad (5)$$

Theorem 3 tells us that $\mathcal{S}^{(n)}$ is "sandwiched between" ${\mathcal{S}_\mathcal{V}}^{(n)}$ and ${\mathcal{S}_\mathcal{V}}^{(n+1)}$. From Figure 5, we see that ${\mathcal{S}_\mathcal{V}}^{(n)}$ and ${\mathcal{S}_\mathcal{V}}^{(n+1)}$ are quite close for high n. Thus we have a very accurate representation of $\mathcal{S}^{(n)}$. Theorem 3 follows from the next two lemmas:

Lemma 1. $\mathcal{S}^{(n)} \subset {\mathcal{S}_\mathcal{V}}^{(n+1)} \cup \mathcal{E}^{(n)}$.

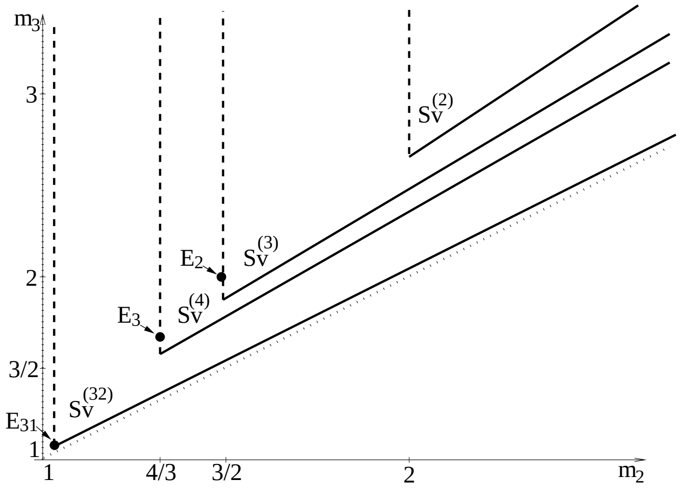

Fig. 5. Depiction of $\mathcal{S}_{\mathcal{V}}{}^{(n)}$ sets for $n = 2, 3, 4, 32$ as a function of the normalized moments. Observe that all possible nonnegative distributions lie within the region delineated by the two dotted lines: $m_2 \geq 1$ and $m_3 \geq m_2$ [16]. $\mathcal{S}_{\mathcal{V}}{}^{(n)}$ for $n = 2, 3, 4, 32$ are delineated by solid lines, which includes the border, and dashed lines, which does not include the border.

Lemma 2. $\mathcal{S}_{\mathcal{V}}{}^{(n)} \subset \mathcal{S}^{(n)}$.

Proof (Lemma 1). The proof proceeds by induction. When $n = 2$, the lemma follows from (1), (5), and Theorem 2. Next, assume that $\mathcal{S}^{(n)} \subset \mathcal{S}_{\mathcal{V}}{}^{(n+1)} \cup \mathcal{E}^{(n)}$ for $n \leq k-1$. Consider an arbitrary distribution $G \in \mathcal{S}^{(k)}$. Let $Z(\cdot) = (X(\cdot) \otimes Y(\cdot))p + 1 - p$, where X is an exponential distribution and Y is a $(k-1)$-phase Coxian distribution.[4] Observe that for any arbitrary distribution $G \in \mathcal{S}^{(k)}$, there exists some such Z which well-represents G. By the assumption of induction, $Y \in \mathcal{S}_{\mathcal{V}}{}^{(k)} \cup \mathcal{E}^{(k-1)}$. We prove that (i) if $Y \in \mathcal{S}_{\mathcal{V}}{}^{(k)}$, then $Z \in \mathcal{S}_{\mathcal{V}}{}^{(k+1)}$ and (ii) if $Y \in \mathcal{E}^{(k-1)}$, then $Z \in \mathcal{S}_{\mathcal{V}}{}^{(k+1)} \cup \mathcal{E}^{(k)}$. Without loss of generality, we can set the first moment of X to 1. To see why this is possible, observe that Z is comprised of k exponential phases, and the normalized second and third moments of Z, m_2^Z and m_3^Z are both invariant to multiplying all the rates of exponential phases in Z by the same constant. Thus, if the first moment of X equals $\mu_1^X \neq 1$, then the rates of all the phases in Z may be multiplied by μ_1^X to bring the first moment of X down to 1.

[4] To shed light on this expression, consider random variables V_X and V_Y whose distributions are X and Y, respectively. Then random variable

$$V_Z = \begin{cases} V_X + V_Y & \text{with probability } p \\ 0 & \text{with probability } 1-p, \end{cases}$$

has distribution Z, since $\Pr(V_Z < t) = p \Pr(V_X + V_Y < t) + (1-p)$.

(i) Suppose $Y \in \mathcal{S}_{\mathcal{V}}{}^{(k)}$: We first prove that $m_2^Z > \frac{k+1}{k}$. Observe that

$$m_2^Z = \frac{2 + 2\mu_1^Y + \mu_2^Y}{p(1+\mu_1^Y)^2} > \frac{2 + 2\mu_1^Y + \frac{k}{k-1}\mu_2^Y}{p(1+\mu_1^Y)^2},$$

where the inequality follows from $Y \in \mathcal{S}_{\mathcal{V}}{}^{(k)}$. The right hand side is minimized when $\mu_1^Y = k-1$. Thus, $m_2^Z > \frac{k+1}{pk} \geq \frac{k+1}{k}$. Next, we prove that $\frac{m_3^Z}{m_2^Z} \geq \frac{k+3}{k+2}$ for all $m_2^Z > \frac{k+1}{k}$. Notice that $\frac{m_3^Z}{m_2^Z}$ is independent of p:

$$\frac{m_3^Z}{m_2^Z} = \frac{(6 + 6\mu_1^Y + 3\mu_2^Y + \mu_3^Y)(1+\mu_1^Y)}{(2 + 2\mu_1^Y + \mu_2^Y)^2}.$$

Since $\frac{m_3^Z}{m_2^Z}$ is an increasing function of μ_3^Y, it is minimized at $\mu_3^Y = \frac{k+2}{k+1}\frac{(\mu_2^Y)^2}{\mu_1^Y}$, since $Y \in \mathcal{S}_{\mathcal{V}}{}^{(k)}$. Thus,

$$\frac{m_3^Z}{m_2^Z} \geq \frac{(1+\mu_1^Y)(6(k+1)\mu_1^Y + 6(k+1)(\mu_1^Y)^2 + 3(k+1)\mu_1^Y\mu_2^Y + (k+2)(\mu_2^Y)^2)}{(k+1)\mu_1^Y(2 + 2\mu_1^Y + \mu_2^Y)^2}.$$

The infimum of the right hand side occurs at:

$$\mu_2^Y = \max\left\{\frac{6(k+1)\mu_1^Y(1+\mu_1^Y)}{4 + 4\mu_1^Y + (k+1)(4+\mu_1^Y)}, \frac{k}{k-1}(\mu_1^Y)^2\right\}.$$

By evaluating $\frac{m_3^Z}{m_2^Z}$ at $\mu_2^Y = \frac{k}{k-1}(\mu_1^Y)^2$, we have

$$\frac{m_3^Z}{m_2^Z} \geq \frac{(1+\mu_1^Y)\left[6(k+1)(k-1)^2(1+\mu_1^Y) + 3k(k^2-1)(\mu_1^Y)^2 + k^2(k+2)(\mu_1^Y)^3\right]}{(k+1)\left[2(k-1) + 2(k-1)\mu_1^Y + k(\mu_1^Y)^2\right]^2}.$$

By Lemma 4 in Appendix A, $\frac{m_3^Z}{m_2^Z} \geq \frac{k+3}{k+2}$. By evaluating $\frac{m_3^Z}{m_2^Z}$ at

$$\mu_2^Y = \frac{6(k+1)\mu_1^Y(1+\mu_1^Y)}{4 + 4\mu_1^Y + (k+1)(4+\mu_1^Y)},$$

we have

$$\frac{m_3^Z}{m_2^Z} \geq \frac{3\left[8(1+\mu_1^Y) + (k+1)(8+5\mu_1^Y)\right]}{16(2+k)(1+\mu_1^Y)} \geq \frac{k+3}{k+2},$$

where the last inequality holds iff $\mu_1^Y \leq \frac{8k}{k+9}$. However, $\mu_1^Y \leq \frac{8k}{k+9}$ holds if

$$\frac{6(k+1)\mu_1^Y(1+\mu_1^Y)}{4 + 4\mu_1^Y + (k+1)(4+\mu_1^Y)} > \frac{k}{k-1}(\mu_1^Y)^2.$$

(ii) Suppose $Y \in \mathcal{E}^{(k-1)}$: We will prove that (a) if $\mu_1^Y = (k-1)$ and $p = 1$, then $Z \in \mathcal{E}^{(k)}$, and (b) if $\mu_1^Y \neq (k-1)$ or $p < 1$, then $Z \in \mathcal{S}_{\mathcal{V}}{}^{(k+1)}$. For part (a), observe that if $Y \in \mathcal{E}^{(k-1)}$, $\mu_1^Y = (k-1)$, and $p = 1$, then we have already seen that $m_2^Z = \frac{k+1}{k}$ in part (i). It is also easy to see that $m_3^Z = \frac{k+2}{k}$, and hence $Z \in \mathcal{E}^{(k)}$. For part (b), if $\mu_1^Y \neq (k-1)$ or $p < 1$, then first notice that $m_2^Z > \frac{k+1}{k}$, since m_2^Z is minimized when $\mu_1^Y = (k-1)$ and $p = 1$. Also, since $m_3^Y = \frac{k+1}{k-1} > \frac{k+2}{k-1}$, $\frac{m_3^Z}{m_2^Z} \geq \frac{k+3}{k+2}$ by part (i), and hence $Z \in \mathcal{S}_{\mathcal{V}}{}^{(k+1)}$. □

Proof (Lemma 2). When $n = 2$, the lemma follows from Theorem 2. The remainder of the proof assumes $n \geq 3$. We prove that for an arbitrary distribution $G \in \mathcal{S_V}^{(n)}$, there exists an n-phase Coxian Z such that the normalized moments of G and Z agree. Notice that the first moment of Z is easily matched to G by normalization without changing the normalizing moments of Z. The proof consists of two parts: (i) the case when the normalized moments of G satisfy $m_3^G > 2m_2^G - 1$; (ii) the case when the normalized moments of G satisfy $m_3^G \leq 2m_2^G - 1$.

(i) Suppose $G \in \mathcal{S_V}^{(n)}$ and $m_3^G > 2m_2^G - 1$: We need to show that G is well-represented by some n-phase Coxian distribution. We will prove something stronger that G is well-represented by a distribution Z where $Z = X + Y$, and X is a particular two-phase Coxian distribution with no mass probability at zero and Y is a particular Erlang-$(n-2)$ distribution. (For the intuition behind this particular way of representing G, please refer to [22]). The normalized moments of X are chosen as follows:

$$
\begin{aligned}
m_2^X &= \frac{m_2^G(n-3)-(n-2)}{m_2^G(n-2)-(n-1)};\\
m_3^X &= \big((n-1)m_2^X-(n-2)\big)\big((n-2)m_2^X-(n-3)\big)^2\,\frac{m_3^G}{m_2^X}\\
&\quad -\frac{(n-2)(m_2^X-1)\big(n(n-1)(m_2^X)^2-n(2n-5)m_2^X+(n-1)(n-3)\big)}{m_2^X}.
\end{aligned}
$$

The first moment of Y is chosen as follows: $\mu_1^Y = (n-2)(m_2^X - 1)\mu_1^X$. It is easy to see that the normalized moments of G and Z agree:

$$
\begin{aligned}
m_2^Z &= \frac{m_2^X + 2y + m_2^Y y^2}{(1+y)^2} = m_2^G;\\
m_3^Z &= \frac{m_2^X m_3^X + 3m_2^X y + 3m_2^Y y^2 + m_2^Y m_3^Y y^3}{(m_2^X + 2y + m_2^Y y^2)(1+y)} = m_3^G;
\end{aligned}
$$

where $m_2^Y = \frac{n-1}{n-2}$ and $m_3^Y = \frac{n}{n-2}$ are the normalized moments of Y, and $y = \frac{\mu_1^Y}{\mu_1^X}$. Finally, we will show that there exists a two-phase Coxian distribution with no mass probability at zero, with normalized moments m_2^X and m_3^X. By Theorem 1, it suffices to show that $m_2^X > 2$ and $m_3^X > \frac{3}{2}m_2^X$. The first condition, $m_2^X > 2$, can be shown using $\frac{n}{n-1} < m_2^G$, which follows from $G \in \mathcal{S_V}^{(n)}$. It can also be shown that $m_3^X > 2m_2^X - 1 \geq \frac{3}{2}m_2^X$ using $\frac{n}{n-1} < m_2^G$ and $m_3^G > 2m_2^G - 1$, which is the assumption that we made at the beginning of (i).

(ii) Suppose $G \in \mathcal{S_V}^{(n)}$ and $m_3^G \leq 2m_2^G - 1$: We again must show that G is well-represented by an n-phase Coxian distribution. We will show that G is well-represented by a distribution $Z(\cdot) = U(\cdot)p + 1 - p$ (See Section 2 for an explanation of Z), where $p = \frac{1}{2m_2^G - m_3^G}$ and the normalized moments of U satisfy $m_2^U = pm_2^G$ and $m_3^U = pm_3^G$. It is easy to see that the normalized moments of G and Z agree. Therefore, it suffices to show that U is well-represented by an n-phase Coxian distribution W, since then G is well represented by an n-phase Coxian distribution $Z(\cdot) = W(\cdot)p + 1 - p$ (See Section 2 for an explanation of

Z). We will prove that U is well-represented by an n-phase Coxian distribution W, where $W = X + Y$ and X is a two-phase Coxian distribution with no mass probability at zero and Y is an Erlang-$(n-2)$ distribution. The normalized moments of X are chosen as follows:

$$m_2^X = \frac{m_2^U(n-3)-(n-2)}{m_2^U(n-2)-(n-1)} \quad \text{and} \quad m_3^X = 2m_2^X - 1;$$

the first moment of Y is chosen as follows: $\mu_1^Y = (n-2)(m_2^X-1)\mu_1^X$. It is easy to see that the normalized moments of U and W agree:

$$m_2^W = \frac{m_2^X + 2y + m_2^Y y^2}{(1+y)^2} = m_2^U;$$
$$m_3^W = \frac{m_2^X m_3^X + 3m_2^X y + 3m_2^Y y^2 + m_2^Y m_3^Y y^3}{(m_2^X + 2y + m_2^Y y^2)(1+y)} = 2m_2^U - 1 = m_3^U,$$

where $m_2^Y = \frac{n-1}{n-2}$ and $m_3^Y = \frac{n}{n-2}$ are the normalized moments of Y, and $y = \frac{\mu_1^Y}{\mu_1^X}$. Finally, we will show that there exists a two-phase Coxian distribution with normalized moments m_2^X and m_3^X. By Theorem 2, it suffices to show that $\frac{3}{2} \leq m_2^X$, since

$$\frac{4}{3}m_2^X \leq m_3^X = 2m_2^X - 1 \leq \frac{6(m_2^X-1)}{m_2^X},$$

where the first inequality holds when $m_2^X \geq \frac{3}{2}$ and the second inequality holds when $\frac{3}{2} \leq m_2^X \leq 2$. Since $G \in \mathcal{S_V}^{(n)}$, $m_3^G \geq \frac{n+2}{n+1}m_2^G$. Thus, $m_2^U \geq \frac{m_2^G}{2m_2^G - \frac{n+2}{n+1}m_2^G} = \frac{n+1}{n}$. Finally, $m_2^X \geq \frac{3}{2}$ follows from $m_2^U \geq \frac{n+1}{n}$. □

4 Examples of Some Common Distributions in $\mathcal{S}^{(n)}$

In this section, we give examples of distributions that are well-represented by an n-phase Coxian distribution. In particular, we discuss Bounded Pareto distributions, uniform distributions, symmetric triangular distributions, and Pareto distributions, and derive the necessary and sufficient condition for these distributions to be in $\mathcal{S_V}^{(n)} \subset \mathcal{S}^{(n)}$. A summary is shown in Figure 6.

We first discuss the set of Bounded Pareto distributions. A Bounded Pareto distribution has a density function

$$f(x) = \alpha x^{-\alpha-1}\frac{l^\alpha}{1-\left(\frac{l}{u}\right)^\alpha}$$

for $l \leq x \leq u$ and 0 elsewhere, where $0 < \alpha < 2$ [8]. Bounded Pareto distributions have been empirically shown to fit many recent measurements of computing workloads. These include Unix process CPU requirements measured at Bellcore: $1 \leq \alpha \leq 1.25$ [19], Unix process CPU requirements measured at UC Berkeley: $\alpha \approx 1$ [9], sizes of files transferred through the Web: $1.1 \leq \alpha \leq 1.3$ [3,4], sizes of files stored in Unix filesystems [13], I/O times [25], sizes of FTP transfers in the

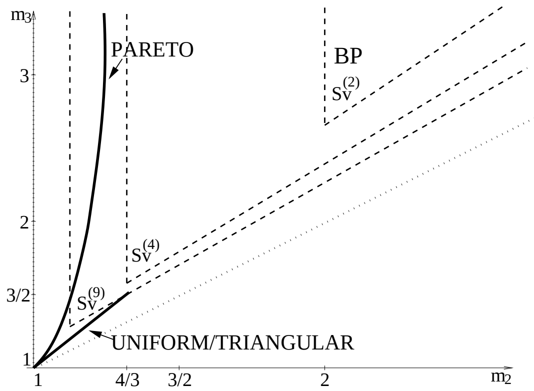

Fig. 6. A summary of the results in Section 4. A few particular classes of distributions are shown in relation to $\mathcal{S_V}^{(n)}$. $\mathcal{BP}^*$ refers to the subset of Bounded Pareto distributions contained in $\mathcal{S_V}^{(2)}$. $\mathcal{UNIFORM}$ refers to the class of all uniform distributions described in Definition 5. We find that the larger the support of the uniform distribution, the fewer the number of phases that suffices. $\mathcal{TRIANGULAR}$ refers to the set of symmetric triangular distributions, described in Definition 5. These interestingly have the same behavior as the uniform distribution. Finally, $\mathcal{PARETO}$ refers to the class of Pareto distributions with finite third moment, described in Definition 5. For this class, we find that the lower the value of the α-parameter, the fewer the number of phases that are needed.

Internet: $.9 \leq \alpha \leq 1.1$ [24], and Pittsburgh Supercomputing Center workloads for distributed servers consisting of Cray C90 and Cray J90 machines [28].

The normalized moments of a Bounded Pareto distribution, F, are

$$m_2^F = \frac{(r-1)^2}{r(\log r)^2}; \quad m_3^F = \frac{(r-1)(r+1)}{2r\log r}$$

when $\alpha = 1$, and

$$m_2^F = \frac{(1-\alpha)^2}{\alpha(2-\alpha)}\frac{(r^\alpha-1)(r^2-r^\alpha)}{(r-r^\alpha)^2}; \quad m_3^F = \frac{(1-\alpha)(2-\alpha)}{\alpha(3-\alpha)}\frac{(r^\alpha-1)(r^3-r^\alpha)}{(r-r^\alpha)(r^2-r^\alpha)},$$

when $0 < \alpha < 1$ or $1 < \alpha < 2$, where $r = \frac{u}{l}$. Not all Bounded Pareto distribution are in $\mathcal{S_V}^{(2)}$. However, a large subset of the Bounded Pareto distributions reside in $\mathcal{S_V}^{(2)}$. Figure 7 shows the necessary and sufficient condition on r as a function of α for a Bounded Pareto distribution to be in $\mathcal{S_V}^{(2)}$. Specifically, a Bounded Pareto distribution is in $\mathcal{S_V}^{(2)}$ if and only if $r = \frac{u}{l}$ is above the two lines shown in Figure 7. We use $\mathcal{BP}^*$ to denote the subset of Bounded Pareto distributions which are contained in $\mathcal{S_V}^{(2)}$.

Next, we discuss uniform distributions, symmetric triangular distributions, and Pareto distributions, and derive the necessary and sufficient condition for these distributions to be in $\mathcal{S_V}^{(n)}$. We use the following definitions:

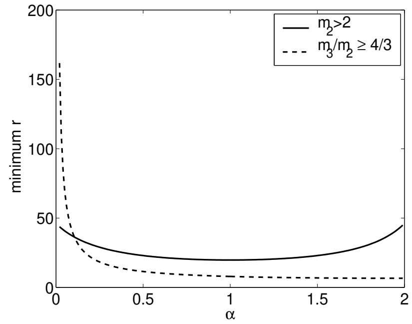

Fig. 7. The maximum of the two lines illustrates the lower bound needed on $r \equiv \frac{u}{l}$ in the definition of the $\mathcal{BP}^*$ distribution. These lines are derived from the conditions $m_2^F > 2$ and $m_3^F \geq \frac{4}{3} m_2^F$.

Definition 5. *$\mathcal{UNIFORM}$ refers to the set of distributions having density function $f(x) = \frac{1}{u-l}$ for $l \leq x \leq u$ and 0 elsewhere, for some $0 \leq l < u$.*

$\mathcal{TRIANGULAR}$ is the set of distributions with density function

$$f(x) = \begin{cases} \left(\frac{2}{u-l}\right)^2 (x-l) & \text{if } l \leq x \leq \frac{u+l}{2} \\ -\left(\frac{2}{u-l}\right)^2 (x-u) & \text{if } \frac{u+l}{2} \leq x \leq u \\ 0 & \text{otherwise,} \end{cases}$$

for some $0 \leq l < u$.

$\mathcal{PARETO}$ is the set of distributions with density function $f(x) = \alpha k^\alpha x^{-\alpha-1}$ for $x \geq k$ and 0 elsewhere, for some $\alpha > 3$ and $k > 0$.

Let $F_U \in \mathcal{UNIFORM}$ and $F_T \in \mathcal{TRIANGULAR}$ with parameters l and u, and let $F_P \in \mathcal{PARETO}$ with parameters α and k. The normalized moments of F_U, F_T, and F_P are:

$$m_2^{F_U} = \frac{4}{3} \frac{1+r+r^2}{(1+r)^2}; \quad m_2^{F_T} = \frac{7+10r+7r^2}{6(1+r)^2}; \quad m_2^{F_P} = \frac{(\alpha-1)^2}{\alpha(\alpha-2)};$$

$$m_3^{F_U} = \frac{3}{2} \frac{1+r^2}{1+r+r^2}; \quad m_3^{F_T} = \frac{3(3+2r+3r^2)}{7+10r+7r^2}; \quad m_3^{F_P} = \frac{(\alpha-1)(\alpha-2)}{\alpha(\alpha-3)},$$

where $r = \frac{u}{l}$. Note that $m_2^{F_P}$ and $m_3^{F_P}$ are independent of k.

Therefore, the three distribution classes are formally characterized as follows:

Theorem 4. *For all $F \in \mathcal{UNIFORM}$, $1 \leq m_2^F \leq \frac{4}{3}$ and $m_3^F = 3 - \frac{2}{m_2^F}$ for all $0 \leq l < u$.*

For all $F \in \mathcal{TRIANGULAR}$, $1 \leq m_2^F \leq \frac{7}{6}$ and $m_3^F = 3 - \frac{2}{m_2^F}$ for all $0 \leq l < u$.

For all $F \in \mathcal{PARETO}$,

$$1 < m_2^F < \frac{4}{3} \quad \text{and} \quad m_3^F = \frac{-2(m_2^F)^2 + 3m_2^F + 2(m_2^F - 1)\sqrt{m_2^F(m_2^F - 1)}}{4 - 3m_2^F}$$

for all $\alpha > 3$.

Simple consequences of the theorem are:

Corollary 1. *Let $F \in \mathcal{UNIFORM}$ with parameters l and u. Then, $F \in \mathcal{S_V}^{(n)}$ if and only if $n \geq \frac{7+14r+30r^2+14r^3+7r^4}{(1-r)^2(1+4r+r^2)}$, where $r = \frac{u}{l}$. In particular, for all values of u, $n = 7$ if $l = 0$, and $n > 7$ whenever $l > 0$.*

Let $F \in \mathcal{TRIANGULAR}$ with parameters l and u. Then, $F \in \mathcal{S_V}^{(n)}$ if and only if $n \geq \frac{4(11+34r+54r^2+34r^3+11r^4)}{(1-r)^2(5+14r+5r^2)}$, where $r = \frac{u}{l}$. In particular, for all values of l and u, $n \geq 9$.

Let $F \in \mathcal{PARETO}$ with parameters α and k. Then, $F \in \mathcal{S_V}^{(n)}$ if and only if $n > (\alpha - 1)^2$ for all values of k. In particular, $n > 4$ for all $\alpha > 3$ and k.

5 Conclusion

The contribution of this paper is a characterization of the set $\mathcal{S}^{(n)}$ of distributions G which are well-represented by an n-phase Coxian distribution. We introduce several ideas which help in creating a simple formulation of $\mathcal{S}^{(n)}$. The first is the concept of normalized moments. The second is the notion of $\mathcal{S_V}^{(n)}$, a nearly complete subset of $\mathcal{S}^{(n)}$ with an extremely simple representation. The arguments required in proving the above results have an elegant structure which repeatedly makes use of the recursive nature of the Coxian distributions.

Our characterization of $\mathcal{S}^{(n)}$ provides a necessary number of phases and a sufficient number of phases for a given distribution to be well-represented by a Coxian distribution, and these bounds are nearly tight. This result has several practical uses. First, in designing algorithms which fit general distributions to Coxian distributions (fitting algorithms), it is desirable to find a *minimal* (fewest number of phases) Coxian distribution. Our characterization allows algorithm designers to determine how close their Coxian distribution is to the minimal Coxian distribution, and provides intuition for coming up with improved algorithms. We have ourselves benefitted from exactly this point. In a companion paper [22], we develop an algorithm for finding a minimal Coxian distribution that well-represents a given distribution. We find that the simple characterization of $\mathcal{S}^{(n)}$ provided herein is very useful in this task. Our results are also useful as an input to some existing fitting algorithms, such as Johnson and Taaffe's nonlinear programming approach [15], which require knowing a priori the number of phases n in the minimal Coxian distribution. Furthermore we classify a few examples of common and practical distributions as being subsets of $\mathcal{S}^{(n)}$ for some n.

Future work includes a simple characterization of the set of distributions that are well-represented by general n-phase PH distributions. If we were to follow the approach in this paper, we would start by specifying the lower bounds for the second and third normalized moments of general n-phase PH distributions. However, this seems to be nontrivial: although the lower bound on the normalized second moment is known [1], the lower bound on the normalized third moment of n-phase PH distributions is not known.

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 - Stochastic Models*, 3:467 – 473, 1987.
2. T. Altiok. On the phase-type approximations of general distributions. *IIE Transactions*, 17:110 – 116, 1985.
3. M. E. Crovella and A. Bestavros. Self-similarity in World Wide Web traffic: Evidence and possible causes. *IEEE/ACM Transactions on Networking*, 5(6):835 – 846, December 1997.
4. M. E. Crovella, M. S. Taqqu, and A. Bestavros. Heavy-tailed probability distributions in the world wide web. In *A Practical Guide To Heavy Tails*, chapter 1, pages 1 – 23. Chapman & Hall, New York, 1998.
5. A. Cumani. On the canonical representation of homogeneous Markov processes modeling failure-time distributions. *Microelectronics and Reliability*, 22:583 – 602, 1982.
6. A. Feldmann and W. Whitt. Fitting mixtures of exponentials to long-tail distributions to analyze network performance models. *Performance Evaluation*, 32:245 – 279, 1998.
7. 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.
8. M. Harchol-Balter. Task assignment with unknown duration. *Journal of the ACM*, 49(2), 2002.
9. M. Harchol-Balter and A. Downey. Exploiting process lifetime distributions for dynamic load balancing. In *Proceedings of SIGMETRICS '96*, pages 13 – 24, 1996.
10. M. Harchol-Balter, C. Li, T. Osogami, A. Scheller-Wolf, and M. Squillante. Task assignment with cycle stealing under central queue. In *Proceedings of ICDCS '03*, pages 628–637, May 2003.
11. 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.
12. A. Horváth and M. Telek. Phfit: A general phase-type fitting tool. In *Proceedings of Performance TOOLS 2002*, pages 82 – 91, April 2002.
13. G. Irlam. Unix file size survey - 1993. Available at `http://www.base.com/gordoni-/ufs93.html`, September 1994.

14. M. A. Johnson and M. R. Taaffe. Matching moments to phase distributions: Density function shapes. *Communications in Statistics — Stochastic Models*, 6:283 – 306, 1990.
15. M. A. Johnson and M. R. Taaffe. Matching moments to phase distributions: Nonlinear programming approaches. *Communications in Statistics — Stochastic Models*, 6:259 – 281, 1990.
16. S. Karlin and W. Studden. *Tchebycheff Systems: With Applications in Analysis and Statistics*. John Wiley and Sons, 1966.
17. 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.
18. G. Latouche and V. Ramaswami. *Introduction to Matrix Analytic Methods in Stochastic Modeling*. ASA-SIAM, Philadelphia, 1999.
19. W. E. Leland and T. J. Ott. Load-balancing heuristics and process behavior. In *Proceedings of Performance and ACM Sigmetrics*, pages 54 – 69, 1986.
20. R. Marie. Calculating equilibrium probabilities for $\lambda(n)/c_k/1/n$ queues. In *Proceedings of* Performance 1980, pages 117 – 125, 1980.
21. M. F. Neuts. *Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach*. The Johns Hopkins University Press, 1981.
22. T. Osogami and M. Harchol-Balter. A closed-form solution for mapping general distributions to minimal PH distributions. In *Proceedings of TOOLS 2003*, September 2003.
23. 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.
24. V. Paxson and S. Floyd. Wide-are traffic: The failure of Poisson modeling. *IEEE/ACM Transactions on Networking*, pages 226 – 244, June 1995.
25. D. L. Peterson and D. B. Adams. Fractal patterns in DASD I/O traffic. In *CMG Proceedings*, December 1995.
26. A. Riska, V. Diev, and E. Smirni. Efficient fitting of long-tailed data sets into PH distributions. *Performance Evaluation*, 2003 (to appear).
27. C. Sauer and K. Chandy. Approximate analysis of central server models. *IBM Journal of Research and Development*, 19:301 – 313, 1975.
28. B. Schroeder and M. Harchol-Balter. Evaluation of task assignment policies for supercomputing servers: The case for load unbalancing and fairness. In *Proceedings of* HPDC 2000, pages 211–219, 2000.
29. M. Squillante. Matrix-analytic methods in stochastic parallel-server scheduling models. In *Advances in Matrix-Analytic Methods for Stochastic Models*. Notable Publications, July 1998.
30. D. Starobinski and M. Sidi. Modeling and analysis of power-tail distributions via classical teletraffic methods. *Queueing Systems*, 36:243 – 267, 2000.
31. 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.
32. W. Whitt. Approximating a point process by a renewal process: Two basic methods. *Operations Research*, 30:125 – 147, 1982.
33. 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.

A Technical Lemmas

Lemma 3. *The set defined in (3) and the set defined in (4) are equivalent sets.*

Proof. Recall that the set defined in (3) is the union of the following three sets:

$$A_1 = \left\{ F \middle| \exists p, \frac{9pm_2^F - 12 + 3\sqrt{2}(2 - pm_2^F)^{\frac{3}{2}}}{p^2 m_2^F} \le m_3^F \le \frac{6(pm_2^F - 1)}{p^2 m_2^F} \text{ and } \frac{3}{2p} \le m_2^F < \frac{2}{p} \right\},$$

$$A_2 = \left\{ F \middle| \exists p, m_3^F = \frac{3}{p} \text{ and } m_2^F = \frac{2}{p} \right\}, \quad A_3 = \left\{ F \middle| \exists p, \frac{3}{2} m_2^F < m_3^F \text{ and } \frac{2}{p} < m_2^F \right\};$$

the set defined in (4) is the union of the following three sets:

$$B_1 = \left\{ F \middle| \frac{4}{3} m_2^F \le m_3^F \le \frac{6(m_2^F - 1)}{m_2^F} \text{ and } \frac{3}{2} \le m_2^F \le 2 \right\},$$

$$B_2 = \left\{ F \middle| \frac{4}{3} m_2^F \le m_3^F \le \frac{3}{2} m_2^F \text{ and } 2 < m_2^F \right\}, \quad B_3 = \left\{ F \middle| \frac{3}{2} m_2^F < m_3 \text{ and } 2 < m_2^F \right\}.$$

It suffices to prove that (i) $A_1 = B_1 \cup B_2$, (ii) $A_2 \subset B_1 \cup B_2$, and (iii) $A_3 = B_3$. (ii) and (iii) are immediate from the definition. To prove (i), we prove that $A_1 \subset B_1 \cup B_2$ and $B_1 \cup B_2 \subset A_1$.

Consider a distribution $F \in A_1$. We first show that $F \in B_1 \cup B_2$. Let $u(p)$ and $l(p)$ be the upper and lower bound of m_3^F, respectively:

$$l(p) = \frac{3\left(3pm_2^F - 4 + \sqrt{2}(2 - pm_2^F)^{\frac{3}{2}}\right)}{p^2 m_2^F}; \quad u(p) = \frac{6(pm_2^F - 1)}{p^2 m_2^F}.$$

Then, $u(p)$ and $l(p)$ are both continuous and increasing functions of p for $\frac{3}{2m_2^F} \le p \le \frac{2}{m_2^F}$. When $m_2^F \le 2$, the range of p is $\frac{3}{2m_2^F} \le p \le 1$. Thus,

$$\frac{4}{3} m_2^F = l\left(\frac{3}{2m_2^F}\right) \le m_3^F \le u(1) = \frac{6(m_2^F - 1)}{m_2^F},$$

and hence $F \in B_1$. When $2 < m_2^F$, the range of p is $\frac{3}{2m_2^F} \le p \le \frac{2}{m_2^F}$. Thus,

$$\frac{4}{3} m_2^F = l\left(\frac{3}{2m_2^F}\right) \le m_3^F \le u\left(\frac{2}{m_2^F}\right) = \frac{3}{2} m_2^F,$$

and hence $F \in B_2$. Therefore, $A_1 \subset B_1 \cup B_2$. However, since $u(p)$ and $l(p)$ are continuous functions of p, m_3^F can take any value between the lower and upper bounds. Therefore, $B_1 \cup B_2 \subset A_1$. $\square$

Lemma 4. *Let* $y \ge 0$ *and* $k \ge 1$. *Then,*

$$\frac{(1+y)\left[6(k+1)(k-1)^2(1+y) + 3k(k^2-1)y^2 + k^2(k+2)y^3\right]}{(k+1)\left[2(k-1) + 2(k-1)y + ky^2\right]^2} \ge \frac{k+3}{k+2}.$$

Proof. Let

$$\begin{aligned}
g(y,k) &= (1+y)\left[6(k+1)(k-1)^2(1+y)+3k(k^2-1)y^2+k^2(k+2)y^3\right](k+2) \\
&\quad -(k+1)\left[2(k-1)+2(k-1)y+ky^2\right]^2(k+3) \\
&= (2+4y+y^2)k^4-2(1+2y+4y^2+y^3)k^3-(2+4y+y^2-5y^3-y^4)k^2 \\
&\quad +2(1+y)(1+y+3y^2)k.
\end{aligned}$$

We prove that $g(y,k)\geq 0$. Let $h(y,k)=\frac{g(y,k)}{k}$. It suffices to prove $h(y,k)\geq 0$. Observe that $\frac{\partial h(y,k)}{\partial k}=0$ iff $k=\frac{2+4y+8y^2+2y^3\pm\sqrt{d(y)}}{3(2+4y+y^2)}$, where

$$d(y)=16+64y+108y^2+66y^3+17y^4+5y^5+y^6.$$

Notice that $d(y)\geq(4+8y+y^2+y^3)^2$. Thus,

$$\frac{2+4y+8y^2+2y^3+\sqrt{d(y)}}{3(2+4y+y^2)}\geq\frac{2+4y+8y^2+2y^3+(4+8y+y^2+y^3)}{3(2+4y+y^2)}\geq 1$$

for $y\geq 0$. Therefore, $h(y,k)$ is minimized when $k=\frac{2+4y+8y^2+2y^3+\sqrt{d(y)}}{3(2+4y+y^2)}$. Let

$$\begin{aligned}
s(y) &= h\left(y,\frac{2+4y+8y^2+2y^3+\sqrt{d(y)}}{3(2+4y+y^2)}\right) \\
&= \frac{2((28+83y+16y^2+y^3)d(y)-d(y)^{\frac{3}{2}})}{27(2+4y+y^2)^2} \\
&\quad -\frac{12(64+456y+1260y^2+1655y^3+889y^4+147y^5)}{27(2+4y+y^2)^2}.
\end{aligned}$$

It suffices to prove $s(y)\geq 0$. Let $t(y)=27(2+4y+y^2)^2s(y)$. It suffices to prove $t(y)\geq 0$. Notice that $t(0)=0$. Thus, it suffices to prove $t'(y)\geq 0$ for $y\geq 0$. However, $t'(y)=\frac{3}{\sqrt{d(y)}}v(y)$, where

$$\begin{aligned}
&v(y) \\
&= 2(128+688y+1922y^2+3216y^3+3055y^4+1562y^5+420y^6+56y^7+3y^8)\sqrt{d(y)} \\
&\quad -(64+216y+198y^2+68y^3+25y^4+6y^5)d(y) \\
&\geq 2(128+688y+1922y^2+3216y^3+3055y^4+1562y^5+420y^6+56y^7+3y^8)\cdot \\
&\quad (4+8y+y^2+y^3)-(64+216y+198y^2+68y^3+25y^4+6y^5)d(y) \\
&= 3y^2(912+5600+13212y^2+15184y^3+9604y^4+3914y^5+1175y^6+235y^7+21y^8) \\
&\geq 0.
\end{aligned}$$

$\square$