# Necessary and sufficient conditions for representing general distributions by Coxians

Takayuki Osogami[1] Mor Harchol-Balter[2]

September 2002

CMU-CS-02-178

School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

**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 stages in the representative Coxian distribution. In this paper, we formally characterize the set of distributions $G$ which are well-represented by an $n$-stage Coxian distribution, in the sense that the Coxian distribution matches the first three moments of $G$. We also discuss a few common, practical examples. Lastly, we derive a partial characterization of the set of busy period durations which are well-represented by an $n$-stage Coxian distribution.

[1]Carnegie Mellon University, Computer Science Department. Email: osogami@cs.cmu.edu

[2]Carnegie Mellon University, Computer Science Department. Email: harchol@cs.cmu.edu. This work was supported by NSF Career Grant CCR-0133077, by NSF ITR Grant 99-167 ANI-0081396 and by Spinnaker Networks via Pittsburgh Digital Greenhouse Grant 01-1.

**Keywords:** moment matching, Coxian distribution, busy period, phase-type distribution, normalized moment, queueing, matrix analytic

# 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 [19]. 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 [20, 16], 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.*

It has been shown that matching three moments is sufficient for accurate modeling of many computer systems [9, 22]. Matching fewer moments is less desirable since some queueing systems, e.g. the $H_2/M/1$ queue, have response times which are heavily dependent on the third moment of $H_2$ [30, 11].

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 them [29, 13, 12, 25, 18]. 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 stages. For example, for any acyclic PH distribution $P_n$, there exists a Coxian distribution $C_n$ with the same number of stages 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 stages 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

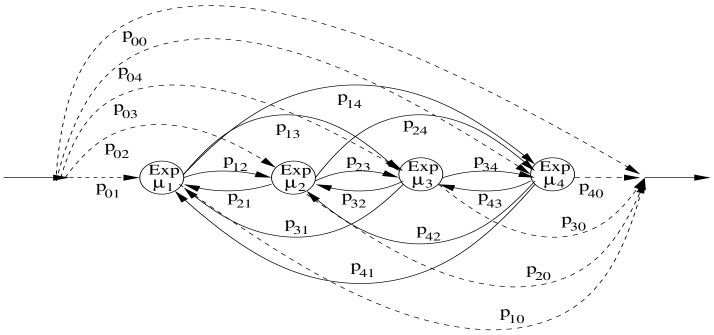


Figure 1: *A 4-stage PH distribution. There are $n = 4$ states, where the $i$th state has exponentially-distributed service time with rate $\mu_i$. With probability $p_{0i}$ we start in the $i$th state, and the next state is state $j$ with probability $p_{ij}$. Each state $i$ has probability $p_{i0}$ that it will be the last state. The value of the distribution is the sum of the times spent in each of the states.*

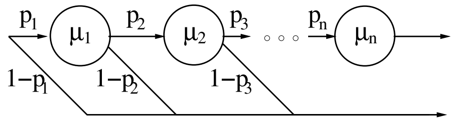


Figure 2: *An $n$-stage Coxian distribution. Observe the recursive definition: with probability $1 - p_1$, the value is zero, and with probability $p_1$, the value is an exponential random variable with rate $\mu_1$ followed by an $(n-1)$-stage Coxian distribution.*

is not known what is the minimal number of phases necessary to 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 $n$-stage Coxian distribution, for each $n = 1, 2, 3, \ldots$.

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

Our characterization of $\{S^{(n)}, n \geq 1\}$ will allow one to determine, for any distribution $G$, the minimal number of stages 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 [13], require knowing the number of stages $n$ in the minimal

[1]One might initially argue that $S^{(2)}$, the set of distributions well-represented by a two-stage Coxian distribution, should include all distributions, since a 2-stage Coxian distribution has four parameters ($p_1$, $p_2$, $\mu_1$, $\mu_2$), 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 gives two solutions for parameters ($p_2$, $\mu_1$, $\mu_2$) as functions of $p_1$. However, in both solutions, one of $\mu_1$ and $\mu_2$ is $p_1 - (1 + \sqrt{4(p_1-1)^2+1})/2$, which is not positive for all possible choices of $p_1$.

Coxian distribution. The current approach involves simply iterating over all choices for $n$ [13], whereas our characterization would immediately specify $n$.

A *secondary goal* of this paper is to specify the necessary and sufficient number of stages needed to well-represent busy period durations by Coxian distributions. Fitting busy period durations to Coxian distributions has become relevant recently in the solution of common computer systems problems involving cycle stealing, see [9, 22]. In [9, 22], transitions between states in a Markov chain represent busy period durations, which are modeled via Coxian distributions for tractability. In addition to standard busy periods, it is also common to model the busy period started by $N$ jobs. This paper will specify the number of stages needed to well-represent such busy periods by Coxian distributions.

Providing sufficient and necessary conditions for a distribution to be in $S^{(n)}$ 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 *third goal* of this paper is to provide examples of practical distributions which are included in $S^{(n)}$ for particular integers $n$.

In finding *simple* characterizations of $S^{(n)}$, 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 $X$ be a distribution and $E[X^k]$ be the $k$-th moment of $X$ for $k = 1, 2, 3$. The* normalized second moment $m_2$ *of $X$ and the* normalized third moment $m_3$ *of $X$ are defined to be*

$$m_2 = \frac{E[X^2]}{(E[X])^2} \quad \text{and} \quad m_3 = \frac{E[X^3]}{E[X]E[X^2]}.$$

Notice the correspondence to the squared coefficient of variability, $C^2$, and the skewness factor, $\gamma$: $m_2 = C^2 + 1$ and $m_3 = \gamma\sqrt{m_2}$.

### Relevant previous work

All prior work on characterizing $S^{(n)}$ has focused only on characterizing $S^{(2)^*}$, where $S^{(2)^*}$ is the set of distributions which are well-represented by a 2-stage Coxian distribution, where the definition of the two-stage Coxian distribution used is more restrictive than our definition – specifically, there is no mass probability at zero, i.e. $p_1 = 1$. Observe $S^{(2)^*} \subset S^{(2)}$. Altiok [2] showed a sufficient condition for a distribution $G$ to be in $S^{(2)^*}$. More recently, Telek and Heindl [28] expanded Altiok's condition and proved the necessary and sufficient condition for a distribution $G$ to be in $S^{(2)^*}$. While neither Altiok nor Telek and Heindl expressed these conditions in terms of normalized moments, the results can be expressed much simpler with our normalized moments, as shown in Theorem 1. In this paper, we extend these results to characterize $S^{(2)}$, as well as characterizing $S^{(n)}$, for all integers $n \geq 2$.

**Our results**

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

**Definition 4** *For integers $n \geq 2$, let $S_V^{(n)}$ denote the set of distributions with the following property on their normalized moments: $m_2$ and $m_3$:*

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

The main contribution of this paper is a derivation of the nested relationship between $S_V^{(n)}$ and $S^{(n)}$ for all $n \geq 2$. This relationship is illustrated in Figure 3 and proven in Section 3. There are three points to observe: (i) $S^{(n)}$ is a proper subset of $S^{(n+1)}$ for all integers $n \geq 2$, and likewise $S_V^{(n)}$ is a proper subset of $S_V^{(n+1)}$; (ii) $S_V^{(n)}$ is contained in $S^{(n)}$ and close to $S^{(n)}$ in size; providing a simple characterization for $S^{(n)}$; (iii) $S^{(n)}$ is almost contained in $S_V^{(n+1)}$ for all integers $n \geq 2$ (more precisely, we will show $S^{(n)} \subset S_V^{(n+1)} \cup E^{(n)}$, where $E^{(n)}$ is the set of distributions well-represented by an Erlang-n distribution). This result yields a necessary number and a sufficient number of stages 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 $S^{(2)}$, we derive the exact necessary and sufficient condition for a distribution $G$ to be in $S^{(2)}$ as a function of the normalized moments of $G$. This extends the results of Telek and Heindl, who analyzed $S^{(2)^*}$, which is a subset of $S^{(2)}$. (See Section 2).

We next investigate the fitting of M/G/1 busy periods by Coxian distributions. Let $B$ denote the duration of an M/G/1 busy period where $G$ is an arbitrary distribution with finite third moment and where the size of the job starting the busy period is in $S_V^{(n)}$. We prove that any such $B$ has distribution in $S_V^{(n)}$. This is surprising in that the number of stages which suffice to represent the busy period is determined solely by the *first job* starting the busy period, which may be a simple setup cost, and it is not required to consider the distributions of the other jobs in the busy period. Furthermore, let $B_N$ denote the duration of an M/G/1 busy period where $G$ is an arbitrary distribution with finite third moment and where the busy period is started by $N$ jobs with service time distribution in $S_V^{(n)}$ where $N$ is the number of Poisson arrivals during a random variable with distribution in $S_V^{(n)}$. We prove that any such $B_N$ is in $S_V^{(n)}$. (See Section 4).

Lastly, we provide a few examples of common, practical distributions included in the set $S_V^{(n)} \subset S^{(n)}$. All distributions we consider have finite third moment. The Pareto distribution and the Bounded

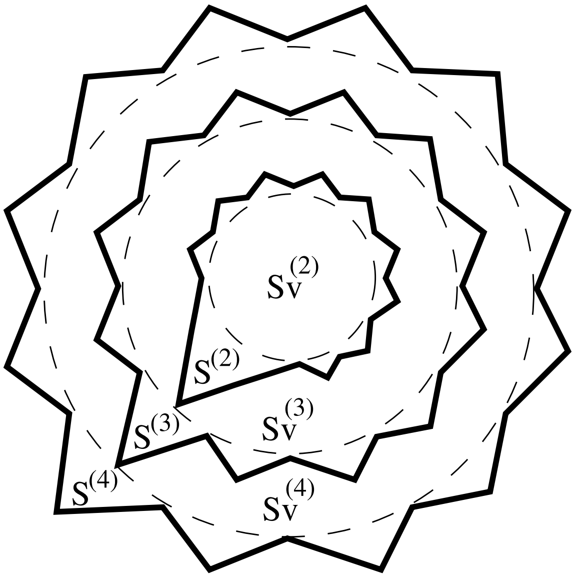


Figure 3: *The main contribution of this paper: a simple characterization of $S^{(n)}$ by $S_V^{(n)}$. Solid lines delineate $S^{(n)}$ (which is irregular) and dashed lines delineate $S_V^{(n)}$ (which is regular – has a simple specification). Observe the nested structure of $S^{(n)}$ and $S_V^{(n)}$.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

Pareto distribution (as defined in [7]) have been shown to fit many recent measurement of job service requirement in computing systems, including HTTP requests [3, 4], UNIX jobs [17, 8], and the duration of FTP transfers [23]. We show that the Bounded Pareto with high variability is in $S^{(2)}$. We also provide conditions under which the Pareto and uniform distributions are in $S_V^{(n)}$ for each $n \geq 2$. (See Section 5).[2]

## 2 Full characterization of $S^{(2)}$

The Telek and Heindl [28] result may be expressed in terms of normalized moments as follows:

**Theorem 1 (Telek, Heindl)** *$G \in S^{(2)^*}$ iff $G$ is in the following union of sets:*[3]

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

[2]Our results show that the *first three moments* of the Bounded Pareto distribution are matched by a two-stage Coxian distribution and the *first three moments* os the Pareto distribution with high variability are matched by a Coxian distribution with a small number of stages. Note however that this does not necessarily imply that the *shape* of these distributions is well-matched by a Coxian distribution with few stages, since the tail of these distributions is not exponential. Recently, fitting the *shape* of heavy-tailed distributions by phase-type distributions such as hyperexponential distributions has been studied [6, 27, 15].

[3]Throughout this paper, $\{$conditions on normalized moments$\}$ denotes the set of distributions that satisfy the conditions. For example, $\left\{\frac{3}{2}m_2 < m_3 \cap 2 < m_2\right\}$ denotes set $\left\{X | \frac{3}{2}m_2^X < m_3^X \text{ and } 2 < m_2^X\right\}$.

Since only an outline of a proof is given in [28], we derive our own proof of Theorem 1 in Appendix A for completeness. We now show a simpler characterization of $S^{(2)}$:

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

$$\left\{ \frac{4}{3}m_2 \le m_3 \le \frac{6(m_2-1)}{m_2} \quad \cap \quad \frac{3}{2} \le m_2 \le 2 \right\} \quad \bigcup \quad S_V^{(2)}, \tag{2}$$

*where recall $S_V^{(2)}$ is the set: $\left\{ \frac{4}{3}m_2 \le m_3 \;\cap\; 2 < m_2 \right\}$.*

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

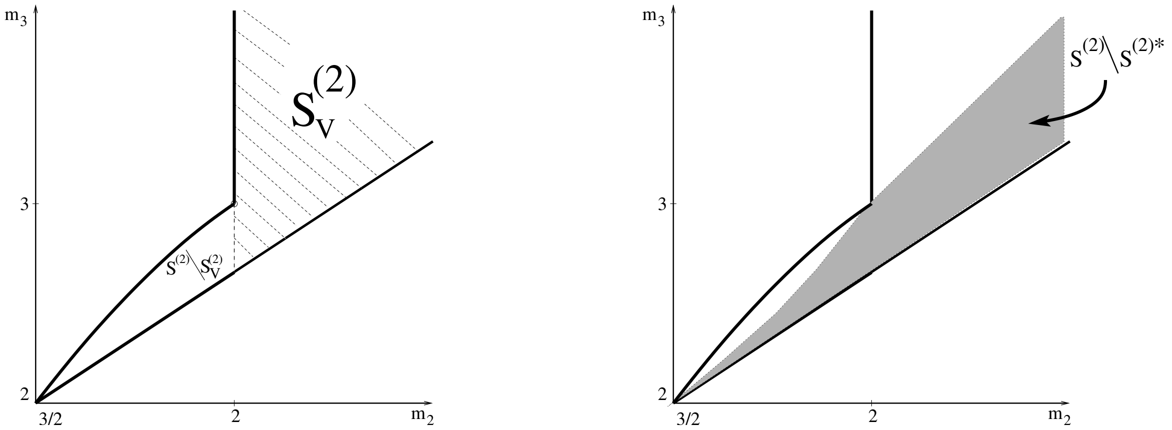


Figure 4: *(a) The thick solid lines delineate $S^{(2)}$. The dashed lines (striped region) show $S_V^{(2)} \subset S^{(2)}$. (b) Again, the thick solid lines delineate $S^{(2)}$. The shaded area shows the region $S^{(2)} \setminus S^{(2)^*}$.*

**Proof**:[Theorem 2] The theorem will be proved by reducing $S^{(2)}$ to $S^{(2)^*}$ and employing Theorem 1. The proof hinges on the following observation: An arbitrary distribution $G \in S^{(2)}$ iff $G$ is well-represented by some distribution $Z$, where

$$Z = \begin{cases} X & \text{with probability } p \\ 0 & \text{with probability } 1-p \end{cases}$$

for some $X \in S^{(2)^*}$. It therefore suffices to show that $Z$ is in set (2).

Let $m_i^Z$ be the normalized $i$-th moment of $Z$ and $m_i^X$ be the normalized $i$-th moment of $X$ for $i = 2, 3$. Observe that $E[Z^i] = pE[X^i]$ for $i = 1, 2, 3$ and $m_i^Z = \frac{m_i^X}{p}$ for $i = 2, 3$.

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

$$\left\{ \frac{9m_2 - 12 + 3\sqrt{2}(2-m_2)^{\frac{3}{2}}}{m_2} \le m_3 \le \frac{6(m_2-1)}{m_2} \;\cap\; \frac{3}{2} \le m_2 < 2 \right\} \bigcup \{m_3 = 3 \;\cap\; m_2 = 2\} \bigcup \left\{ \frac{3}{2}m_2 < m_3 \;\cap\; 2 < m_2 \right\}.$$

Thus, $Z$ is in the following union of sets:[4]

$$\left\{ \frac{9pm_2 - 12 + 3\sqrt{2}(2-pm_2)^{\frac{3}{2}}}{p^2 m_2} \le m_3 \le \frac{6(pm_2-1)}{p^2 m_2} \;\cap\; \frac{3}{2p} \le m_2 < \frac{2}{p} \right\} \bigcup \left\{ m_3^Z = \frac{3}{p} \;\cap\; m_2^Z = \frac{2}{p} \right\} \bigcup \left\{ \frac{3}{2} m_2^Z < m_3^Z \;\cap\; \frac{2}{p} < m_2^Z \right\} \tag{3}$$

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

$$\left\{ \frac{4}{3} m_2 \le m_3 \le \frac{6(m_2-1)}{m_2} \;\cap\; \frac{3}{2} \le m_2 \le 2 \right\} \bigcup \left\{ \frac{4}{3} m_2 \le m_3 \le \frac{3}{2} m_2 \;\cap\; 2 < m_2 \right\} \bigcup \left\{ \frac{3}{2} m_2 < m_3 \;\cap\; 2 < m_2 \right\} \tag{4}$$

Observe that (3) and (4) are now in similar forms. We now prove that set (3) is a subset of set (4), and set (4) is a subset of set (3). The technical details are postponed to Appendix C, Lemma 9.2. ■

# 3 A characterization of $S^{(n)}$

In this section, we prove that $S_V^{(n)}$ is contained in $S^{(n)}$, where $S_V^{(n)}$ is the set of distributions whose normalized moments satisfy (1) and that $S^{(n)}$ is almost contained in $S_V^{(n+1)}$. Figure 6 provides a graphical view of the $S_V^{(n)}$ sets with respect to the normalized moments. We prove the following theorem:

**Theorem 3** $S_V^{(n)} \subset S^{(n)} \subset S_V^{(n+1)} \cup E^{(n)}$, *where* $E^{(n)}$ *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 shown in Figure 5. Notice that the normalized moments

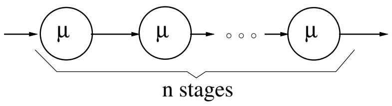


Figure 5: *An Erlang-*$n$ *distribution.*

of distributions in $E^{(n)}$, $m_2^{E^{(n)}}$ and $m_3^{E^{(n)}}$, satisfy the following conditions:

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

Theorem 3 tells us that $S^{(n)}$ is "sandwiched between" $S_V^{(n)}$ and $S_V^{(n+1)}$. From Figure 6, we see that $S_V^{(n)}$ and $S_V^{(n+1)}$ are quite close for higher $n$. Thus we have a very accurate representation of $S^{(n)}$. Theorem 3 follows from the next two lemmas:

[4] {conditions on normalized moments in terms of $p$} denotes the set of distributions that satisfy the conditions for some $p$. For example, $\left\{ \frac{3}{2} m_2 < m_3 \;\cap\; \frac{2}{p} < m_2 \right\}$ denotes set $\left\{ X \mid \exists p \text{ s.t. } 0 \le p \le 1 \text{ and } \frac{3}{2} m_2^X < m_3^X \text{ and } \frac{2}{p} < m_2^X \right\}$.

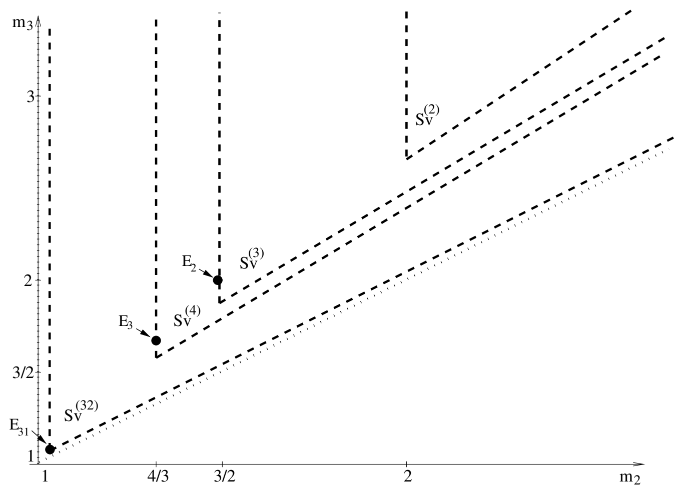


Figure 6: *Depiction of $S_V^{(n)}$ sets for $n = 2, 3, 4, 32$ as a function of the normalized moments. The outermost dotted lines ($m_2 \geq 1$ and $m_3 \geq m_2$) delineate the set of all the possible nonnegative distributions (that is, any nonnegative distribution $G$ satisfies $m_2^G \geq 1$ and $m_3^G \geq m_2^G$) [14]. $S_V^{(n)}$ for $n = 2, 3, 4, 32$ are delineated by dashed lines.*

**Lemma 3.1** $S^{(n)} \subset S_V^{(n+1)} \cup E^{(n)}$.

**Lemma 3.2** $S_V^{(n)} \subset S^{(n)}$.

**Proof**:[Lemma 3.1] The proof proceeds by induction. When $n = 2$, the lemma follows from (1), (5), and Theorem 2. Assume that $S^{(n)} \subset S_V^{(n+1)} \cup E^{(n)}$ for $n \leq k-1$. For any distribution $G \in S^{(k)}$, there exists a $k$-stage Coxian distribution $Z$ by which $G$ is well-represented, where $Z$ can be expressed as

$$Z = \begin{cases} X + Y & \text{with probability } p \\ 0 & \text{with probability } 1-p, \end{cases}$$

and where $X$ is an exponential distribution and $Y$ is a $(k-1)$-stage Coxian distribution. By the assumption of induction, $Y \in S_V^{(k)} \cup E^{(k-1)}$. We prove that (i) if $Y \in S_V^{(k)}$, then $Z \in S_V^{(k+1)}$ and (ii) if $Y \in E^{(k-1)}$, then $Z \in S_V^{(k+1)} \cup E^{(k)}$.

(i) Suppose $Y \in S_V^{(k)}$: We first prove that $m_2^Z > \frac{k+1}{k}$. First observe that

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

where the inequality follows from $Y \in S_V^{(k)}$. The derivative of the right hand side with respect to $E[Y]$ is

$$\frac{2(E[Y] - (k-1))}{p(k-1)(1+E[Y])^3},$$

which is minimized when $E[Y] = k-1$. 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 $p$:

$$\frac{m_3^Z}{m_2^Z} = \frac{(6+6E[Y]+3E[Y^2]+E[Y^3])(1+E[Y])}{(2+2E[Y]+E[Y^2])^2}.$$

Since $\frac{m_3^Z}{m_2^Z}$ is an increasing function of $E[Y^3]$, it is minimized at $E[Y^3] = \frac{k+2}{k+1}\frac{E[Y^2]^2}{E[Y]}$, since $Y \in S_V^{(k)}$. Thus,

$$\frac{m_3^Z}{m_2^Z} \geq \frac{(1+E[Y])(6(k+1)E[Y]+6(k+1)E[Y]^2+3(k+1)E[Y]E[Y^2]+(k+2)E[Y^2]^2)}{(k+1)E[Y](2+2E[Y]+E[Y^2])^2}. \tag{6}$$

Denoting the r.h.s. of (6) by $f(E[Y^2])$, we now find $E[Y^2]$ which minimizes $f(E[Y^2])$. Since

$$\frac{\partial f(E[Y^2])}{\partial E[Y^2]} = \frac{(1+E[Y])\left((4+4E[Y]+(k+1)(4+E[Y]))E[Y^2]-6(k+1)E[Y](1+E[Y])\right)}{(k+1)E[Y](2+2E[Y]+E[Y^2])^3},$$

the infimum of $f(E[Y^2])$ occurs at:

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

By evaluating $\frac{m_3^Z}{m_2^Z}$ at $E[Y^2] = \frac{k}{k-1}E[Y]^2$, we have

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

By Lemma 9.5 in Appendix C, $\frac{m_3^Z}{m_2^Z} \geq \frac{k+3}{k+2}$. By evaluating $\frac{m_3^Z}{m_2^Z}$ at

$$E[Y^2] = \frac{6(k+1)E[Y](1+E[Y])}{4+4E[Y]+(k+1)(4+E[Y])},$$

we have

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

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

$$\frac{6(k+1)E[Y](1+E[Y])}{4+4E[Y]+(k+1)(4+E[Y])} > \frac{k}{k-1}E[Y]^2,$$

since

$$\frac{6(k+1)E[Y](1+E[Y])}{4+4E[Y]+(k+1)(4+E[Y])} \geq \frac{k}{k-1}E[Y]^2$$
$$\iff g(E[Y]) \equiv (k^2+5k)(E[Y])^2 - 2(k^2-4k-3)E[Y] - 6(k+1)(k-1) \leq 0$$

and $g(0) \leq 0$ and

$$g\left(\frac{8k}{k+9}\right) = \frac{6(k+1)(k+3)(7k^2-6k+27)}{(k+9)^2} > 0$$

for $k \geq 2$.

(ii) Suppose $Y \in E^{(k-1)}$: We will prove that (a) if $E[Y] = (k-1)E[X]$ and $p = 1$, then $Z \in E^{(k)}$, and (b) if $E[Y] \neq (k-1)E[X]$ or $p < 1$, then $Z \in S_V^{(k+1)}$: For part (a), observe that if $Y \in E^{(k-1)}$, $E[Y] = (k-1)E[X]$, 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 $Z \in E^{(k)}$. For part (b), if $E[Y] \neq (k-1)E[X]$ or $p < 1$, then first notice that $m_2^Z > \frac{k+1}{k}$, since $m_2^Z$ is minimized when $E[Y] = (k-1)E[X]$ and $p = 1$. Also, since

$$E[Y^3] = \frac{k+1}{k} > \frac{k+2}{k+1},$$

$\frac{m_3^Z}{m_2^Z} \geq \frac{k+3}{k+2}$ by part (i), and hence $Z \in S_V^{(k+1)}$. ■

**Proof**:[Lemma 3.2] When $n = 2$, the lemma follows from Theorem 2. The remainder of the proof assumes $n \geq 3$. We prove that for any distribution $G \in S_V^{(n)}$, there exists an $n$-stage Coxian $Z$ such that the normalized moments of $G$ and $Z$ agree. Notice that the mean of $Z$ is easily matched to $G$ without changing the normalizing moments of $Z$ by multiplying a constant to the rates, $\mu_1$, ..., $\mu_n$, 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 $G \in S_V^{(n)}$ and $m_3^G > 2m_2^G - 1$: We need to show that $G$ is well-represented by some $n$-stage 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-stage 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 [21]). The normalized moments of $X$ are chosen as follows:

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

The mean of $Y$ is chosen as follows: $E[Y] = (n-2)(m_2^X - 1)E[X]$. 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 $y = \frac{E[Y]}{E[X]}$. Finally, we will show that there exists a two-stage 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 $G \in S_V^{(n)}$. 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: $G \in S_V^{(n)}$ and $m_3^G \leq 2m_2^G - 1$: We again must show that $G$ is well-represented by an $n$-stage Coxian distribution. We will show that $G$ is well-represented by a distribution $Z$:

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

where $p = \frac{1}{2m_2^G - m_3^G}$ and the normalized moments of $F$ satisfy $m_2^F = pm_2^G$ and $m_3^F = pm_3^G$. It is easy to see that the normalized moments of $G$ and $Z$ agree. Therefore, it suffices to show that $F$ is well-represented by an $n$-stage Coxian distribution $W$, since then $G$ is well represented by an $n$-stage Coxian distribution $Z$:

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

We will prove that $F$ is well-represented by an $n$-stage Coxian distribution $W = X + Y$, where $X$ is a two-stage 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:

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

the mean of $Y$ is chosen as follows: $E[Y] = (n-2)(m_2^X - 1)E[X]$. It is easy to see that the normalized moments of $F$ and $W$ agree:

$$
\begin{aligned}
m_2^W &= \frac{m_2^X + 2y + m_2^Y y^2}{(1+y)^2} = m_2^F; \\
m_3^W &= \frac{m_2^X m_3^X + 3m_2^X y + 3m_2^Y y^2 + m_2^Y m_3^Y y^3}{(m_2^X + 2y + m_2^Y y^2)(1+y)} = 2m_2^F - 1 = m_3^F,
\end{aligned}
$$

where $m_2^Y = \frac{n-1}{n-2}$ and $m_3^Y = \frac{n}{n-2}$ are the normalized moments of $Y$, and $y = \frac{E[Y]}{E[X]}$. Finally, we will show that there exists a two-stage 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

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

Since $G \in S_V^{(n)}$, $m_3^G \leq \frac{n+2}{n+1}m_2^G$. Thus, $m_2^F \geq \frac{m_2^G}{2m_2^G - \frac{n+2}{n+1}m_2^G} = \frac{n+1}{n}$. Finally, $m_2^X \geq \frac{3}{2}$ follows from $m_2^F \geq \frac{n+1}{n}$. ■

# 4 A characterization of busy periods

In this section we characterize the set of $M/G/1$ busy period durations which are in $S_V^{(n)}$, and hence in $S^{(n)}$. As explained earlier, the tractability of queueing problems often relies not just on representing general distributions by Coxian distributions, but also on representing busy period durations by Coxian distributions [9, 22]. This includes busy periods started by $N$ jobs, where $N$ is the number of arrivals during a period of time. This section provides sufficient conditions on the number of stages needed to represent common types of busy period durations by Coxian distributions. Formally, we will prove the following theorems:

**Theorem 4** *Let $B$ denote the duration of an M/G/1 busy period where $G$ is an arbitrary distribution with finite third moment and where the job starting the busy period has size $K \in S_V^{(n)}$. Then, $B \in S_V^{(n)}$.*

The above theorem states that the number of stages which suffice for a busy period duration to be well-represented by a Coxian distribution is, surprisingly, determined solely by the distribution of the *first job* in the busy period.

**Lemma 4.1** *Let $\Psi_N(T, X)$ be the distribution of the sum of $N$ i.i.d. random variables with distribution $X \in S_V^{(n)}$ where $N$ is the number of Poisson arrivals with rate $\lambda$ during a random time $T \in S_V^{(n)}$. Then, $\Psi_N(T, X) \in S_V^{(n)}$.*

The following theorem follows from Theorem 4 and Lemma 4.1.

**Theorem 5** *Let $B_{\Psi_N}$ denote the duration of an M/G/1 busy period where $G$ is an arbitrary distribution with finite third moment and where the busy period is started by $\Psi_N(T, X)$ as defined in Lemma 4.1. Then, $B_{\Psi_N} \in S_V^{(n)}$.*

We will now prove Theorem 4 and Lemma 4.1.

**Proof**:[Theorem 4] When $\rho = 0$, $B = K \in S_V^{(n)}$, and hence the theorem is true. In the following we assume $0 < \rho < 1$. Let $b_2 \equiv \frac{E[B^2]}{(E[B])^2}$ and $b_3 \equiv \frac{E[B^3]}{(E[B])^3}$. We prove that $b_2 > \frac{n}{n-1}$ and $b_3 - \frac{n+1}{n} b_2^2 > 0$. Observe that together these imply the conditions in (1). Notice that the first three moments of $B$ are

$$\begin{aligned}
E[B] &= \frac{E[K]}{1-\rho}; \\
E[B^2] &= \frac{E[K^2]}{(1-\rho)^2} + \frac{\rho}{(1-\rho)^3}\frac{E[K]E[G^2]}{E[G]}; \\
E[B^3] &= \frac{E[K^3]}{(1-\rho)^3} + \frac{3\rho}{(1-\rho)^4}\frac{E[K^2]E[G^2]}{E[G]} + \frac{\rho}{(1-\rho)^4}\frac{E[K]E[G^3]}{E[G]} + \frac{3\rho^2}{(1-\rho)^5}\frac{E[K](E[G^2])^2}{(E[G])^2}.
\end{aligned}$$

It is easy to see that $b_2 > \frac{n}{n-1}$:

$$b_2 = k_2 + \frac{\rho}{1-\rho} g_2 \frac{E[G]}{E[K]} > k_2 > \frac{n}{n-1},$$

where $k_2 = \frac{E[K^2]}{(E[K])^2}$ and $g_2 = \frac{E[G^2]}{(E[G])^2}$. Next, we prove that $b_3 - \frac{n+1}{n} b_2^2 > 0$. Note that

$$b_3 = k_3 + \frac{3\rho}{1-\rho} g_2 k_2 \frac{E[G]}{E[K]} + \frac{\rho}{1-\rho} g_3 \left(\frac{E[G]}{E[K]}\right)^2 + \frac{3\rho^2}{(1-\rho)^2} g_2^2 \left(\frac{E[G]}{E[K]}\right)^2,$$

where $k_3 = \frac{E[K^3]}{(E[K])^3}$ and $g_3 = \frac{E[G^3]}{(E[G])^3}$. Thus,

$$\begin{aligned}
b_3 - \frac{n+1}{n} b_2^2 &= k_3 - \frac{n+1}{n} k_2^2 + \frac{n-2}{n}\frac{\rho}{1-\rho} g_2 k_2 \frac{E[G]}{E[K]} + \frac{\rho}{1-\rho} g_3 \left(\frac{E[G]}{E[K]}\right)^2 + \frac{2n-1}{n}\frac{\rho^2}{(1-\rho)^2} g_2^2 \left(\frac{E[G]}{E[K]}\right)^2 \\
&> k_3 - \frac{n+1}{n} k_2^2 > 0.
\end{aligned}$$

■

**Proof**:[Lemma 4.1] Let $p_2 \equiv \frac{E[\Psi_N(T,X)^2]}{(E[\Psi_N(T,X)])^2}$ and $p_3 \equiv \frac{E[\Psi_N(T,X)^3]}{(E[\Psi_N(T,X)])^3}$. We prove that $p_2 > \frac{n}{n-1}$ and $p_3 - \frac{n+1}{n} p_2^2 > 0$.

Notice that the Laplace transform of $\Psi_N(T,X)$ is $\tilde{\Psi}_N(T,X) = \tilde{T}(\lambda(1-\tilde{X}(s)))$. Thus, the first three moments of $\Psi_N(T,X)$ are

$$\begin{aligned}
E[\Psi_N(T,X)] &= \lambda E[T]E[X]; \\
E[\Psi_N(T,X)^2] &= \lambda^2 E[T^2]E[X]^2 + \lambda E[T]E[X^2]; \\
E[\Psi_N(T,X)^3] &= \lambda^3 E[T^3]E[X]^3 + 3\lambda^2 E[T^2]E[X]E[X^2] + \lambda E[T]E[X^3].
\end{aligned}$$

It is easy to see that $p_2 > \frac{n}{n-1}$:

$$p_2 = t_2 + \frac{x_2}{\lambda E[T]} > t_2 > \frac{n}{n-1},$$

where $t_2 = \frac{E[T^2]}{(E[T])^2}$ and $x_2 = \frac{E[X^2]}{(E[X])^2}$. Next, we prove that $p_3 - \frac{n+1}{n}p_2^2 > 0$. Note that

$$p_3 = t_3 + \frac{3t_2 x_2}{\lambda E[X]} + \frac{x_3}{(\lambda E[T])^2}.$$

where $t_3 = \frac{E[T^3]}{(E[T])^3}$ and $x_3 = \frac{E[X^3]}{(E[X])^3}$. Thus,

$$p_3 - \frac{n+1}{n}p_2^2 \quad = \quad \left(t_3 - \frac{n+1}{n}t_2^2\right) + \left(3 - 2\frac{n+1}{n}\right)\frac{t_2 x_2}{\lambda E[T]} + \frac{x_3 - \frac{n+1}{n}x_2^2}{(\lambda E[T])^2} > 0.$$

■

# 5 Examples of some common distributions in $S^{(n)}$

In this section, we give examples of distributions in $S_V^{(n)} \subset S^{(n)}$, and hence are well-represented by an $n$-stage Coxian distribution. A summary is shown in Figure 7.

## 5.1 Distributions in $S^{(2)}$

It is well-known that for all two-phase PH distributions $G$, there exists a two phase Coxian distribution $F$ such that $G$ and $F$ has the same distribution function, and hence $G \in S^{(2)}$. In the following, we show that the Bounded Pareto distributions with high variability are also in $S^{(2)}$.

A Bounded Pareto distribution has density function

$$f(x) \quad = \quad \alpha x^{-\alpha-1}\frac{k^\alpha}{1-\left(\frac{k}{p}\right)^\alpha} \qquad (k \leq x \leq p),$$

where $0 < \alpha < 2$. Bounded Pareto distributions have been empirically shown to fit many recent measurement of computing workloads. These include Unix process CPU requirements measured at Bellcore: $1 \leq \alpha \leq 1.25$ [17], Unix process CPU requirements measured at UC Berkeley: $\alpha \approx 1$ [8], sizes of files transferred through the Web: $1.1 \leq \alpha \leq 1.3$ [3, 4], sizes of files stored in Unix filesystems [10], I/O times [24], sizes of FTP transfers in the Internet: $.9 \leq \alpha \leq 1.1$ [23], and Pittsburgh Supercomputing

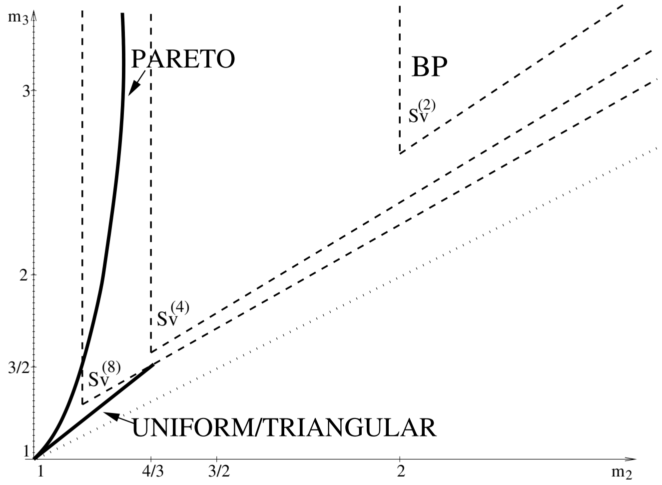


Figure 7: *A summary of the results in Section 5. A few particular distributions are shown in relation to $S_V^{(n)}$. BP refers to the class of bounded Pareto distributions with high variability described in Definition 5. All of these are contained in $S_V^{(2)}$. UNIFORM refers to the class of all uniform distributions described in Definition 6. We find that the larger the range of the UNIFORM distribution, the fewer the number of stages that suffices. TRIANGULAR refers to the set of symmetric triangular distributions, described in Definition 7. These interestingly have the same behavior as the UNIFORM distribution. Finally, PARETO refers to the class of Pareto($\alpha$) distributions with finite third moment, described in Definition 8. For this class, we find that the lower the value of the $\alpha$-parameter, the fewer the number of stages that are needed.*

Center workloads for distributed servers consisting of Cray C90 and Cray J90 machines [26]. In this section, we prove the necessary and sufficient condition for a Bounded Pareto distribution to be in $S_V^{(2)}$. Formally, we use the following definition:

**Definition 5** *$BP$ is a set of Bounded Pareto distributions satisfying $0 < \alpha < 2$ and $r \equiv \frac{p}{k}$ is greater than the maximum of the two lines shown in Figure 8.*

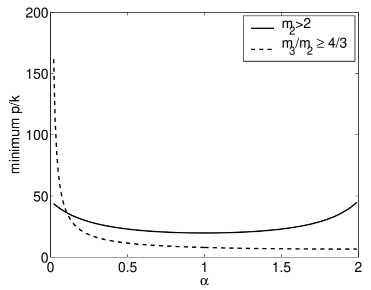


Figure 8: *The maximum of the two lines illustrates the lower bound needed on $r \equiv \frac{p}{k}$ in the definition of the BP distribution. These lines are derived from (7) and (8).*

With this definition, the theorem proven in this section is stated:

**Theorem 6** *$BP \subset S_V^{(2)}$ and $BP^C \cap S_V^{(2)} = \emptyset$, where $BP^C$ is the set of Bounded Pareto distributions not in $BP$ and $\emptyset$ is an empty set.*

**Proof**: Let $m_i^X$ be the normalized $i$-th moment of a distribution $X \in BP$ for $i = 2, 3$. When $\alpha = 1$, the moments of the Bounded Pareto are

$$E[X] = \frac{kp}{p-k}\log\frac{p}{k}, \quad E[X^2] = kp, \quad \text{and} \quad E[X^3] = \frac{1}{2}kp(k+p).$$

Thus,

$$m_2^X = \frac{(r-1)^2}{r(\log r)^2}, \quad m_3^X = \frac{(r-1)(r+1)}{2r\log r}, \quad \text{and} \quad \frac{m_3^X}{m_2^X} = \frac{(r+1)\log r}{2(r-1)}. \tag{7}$$

When $0 < \alpha < 1$ or $1 < \alpha < 2$, the moments of the Bounded Pareto are

$$E[X] = \frac{\alpha}{1-\alpha}\frac{p\left(\frac{k}{p}\right)^\alpha - k}{1-\left(\frac{k}{p}\right)^\alpha}, \quad E[X^2] = \frac{\alpha}{2-\alpha}\frac{p^2\left(\frac{k}{p}\right)^\alpha - k^2}{1-\left(\frac{k}{p}\right)^\alpha}, \quad \text{and} \quad E[X^3] = \frac{\alpha}{3-\alpha}\frac{p^3\left(\frac{k}{p}\right)^\alpha - k^3}{1-\left(\frac{k}{p}\right)^\alpha}.$$

Thus,

$$m_2^X = \frac{(1-\alpha)^2}{\alpha(2-\alpha)}\frac{(r^\alpha-1)(r^2-r^\alpha)}{(r-r^\alpha)^2} \quad \text{and} \quad m_3^X = \frac{(1-\alpha)(2-\alpha)}{\alpha(3-\alpha)}\frac{(r^\alpha-1)(r^3-r^\alpha)}{(r-r^\alpha)(r^2-r^\alpha)}, \tag{8}$$

and

$$\frac{m_3^X}{m_2^X} = \frac{(2-\alpha)^2}{(1-\alpha)(3-\alpha)}\frac{(r-r^\alpha)(r^3-r^\alpha)}{(r^2-r^\alpha)^2}.$$

By Lemmas 9.6-9.9 in Appendix C, both $m_2^X$ and $\frac{m_3^X}{m_2^X}$ are increasing functions of $r$. This makes intuitive sense since the higher moments are likely to increase as the upper bound $p$ (and thus $r$) increases.

Thus, the minimum $r$ such that $m_2^X > 2$ and $\frac{m_3^X}{m_2^X} > \frac{4}{3}$ can be obtained numerically for all $\alpha$ if there exists such a finite $r$. In the following we prove that there is such an $r$ for $0 < \alpha < 2$. When $\alpha = 1$, it is easy to see that as $r$ goes to infinity, both $m_2^X$ and $\frac{m_3^X}{m_2^X}$ go to infinity. Thus, there is a finite $r$ such that $m_2^X > 2$ and $\frac{m_3^X}{m_2^X} \geq \frac{4}{3}$. Next, consider the case with $0 < \alpha < 1$ or $1 < \alpha < 2$. By observing that $0 < \alpha < 2$, it is easy to see that $m_2^X$ goes to infinity as $r$ does. Thus, there is a finite $r$ such that $m_2^X > 2$. Next, we consider $\frac{m_3^X}{m_2^X}$. Observe that

$$\lim_{r\to\infty}\frac{m_3^X}{m_2^X} = \begin{cases} \frac{(2-\alpha)^2}{(1-\alpha)(3-\alpha)} & \text{when} \quad (0 < \alpha < 1) \\ \infty & \text{when} \quad (1 < \alpha < 2). \end{cases}$$

Thus, there is a finite $r$ such that $\frac{m_3^X}{m_2^X} \geq \frac{4}{3}$ for $1 < \alpha < 2$. When $0 < \alpha < 1$, a finite $r$ gives $\frac{m_3^X}{m_2^X} \geq \frac{4}{3}$ if and only if $\frac{(2-\alpha)^2}{(1-\alpha)(3-\alpha)} > \frac{4}{3}$. However, since

$$\frac{(2-\alpha)^2}{(1-\alpha)(3-\alpha)} > \frac{4}{3} \quad \Longleftrightarrow \quad 0 < \alpha(\alpha - 4)$$

There is a finite $r$ such that $\frac{m_3^X}{m_2^X} > \frac{4}{3}$ for $0 < \alpha < 1$, too. ∎

## 5.2 Distributions in $S^{(n)}$

In this section, we give examples of distributions in $S^{(n)}$. It is known that for all acyclic PH distributions $P_n$, there exists a Coxian distribution $C_n$ with the same number of phases as $P_n$ such that $C_n$ and $P_n$ have the same distribution function [5]. Therefore, all the $n$-phase acyclic PH distributions are in $S^{(n)}$.

In the rest of this section, we discuss uniform distributions, symmetric triangular distributions, and Pareto distributions. In particular, we derive the necessary and sufficient condition for these distributions to be in $S_V^{(n)}$. Formally, we use the following definitions:

**Definition 6** *$UNIFORM(l, u)$ refers to the distribution with lower bound $l$ and upper bound $u > 0$ having density function $f(x) = \frac{1}{u-l}$ in the region $l \le x \le u$ and zero otherwise.*

**Definition 7** *$TRIANGULAR(l, u)$ is the distribution with density functions of the form*

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

*where $0 \le l \le u$ and $u > 0$.*

**Definition 8** *$PARETO(\alpha)$ is the distribution with density functions of the form $f(x) = \alpha x^{-\alpha-1}$, where $\alpha > 3$.*

With these definitions, the three distributions are formally characterized as follows:

**Theorem 7** *The normalized moments of $UNIFORM(l, u)$ satisfy $1 \le m_2 \le \frac{4}{3}$ and $m_3 = 3 - \frac{2}{m_2}$ for all $0 \le l \le u$ and $u > 0$.*

**Theorem 8** *The normalized moments of $TRIANGULAR(l, u)$ satisfy $1 \le m_2 \le \frac{7}{6}$ and $m_3 = 3 - \frac{2}{m_2}$ for all $0 \le l \le u$ and $u > 0$.*

**Theorem 9** *The normalized moments of distributions in $PARETO(\alpha)$ satisfy*

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

*for all $\alpha > 3$.*

Simple consequences of the theorems are:

**Corollary 1** *$UNIFORM(l, u) \in S_V^{(n)}$ if and only if $n \geq \frac{8(1+2r+3r^2+2r^3+r^4)}{(1-r)^2(1+4r+r^2)}$, where $r = \frac{l}{u}$. In particular, $n = 8$ if $l = 0$ and $u > 0$, and $n > 8$ for all $0 < l \leq u$.*

**Corollary 2** *$TRIANGULAR(l, u) \in S_V^{(n)}$ if and only if $n \geq \frac{8(1+2r+3r^2+2r^3+r^4)}{(1-r)^2(1+4r+r^2)}$, where $r = \frac{l}{u}$. In particular, $n = 8$ if $l = 0$ and $u > 0$, and $n > 8$ for all $0 < l \leq u$.*

**Corollary 3** *$PARETO(\alpha) \in S_V^{(n)}$ if and only if $n > (\alpha - 1)^2$. In particular, $n > 4$ for all $\alpha > 3$.*

Below we prove Theorem 7. Theorems 8 and 9 can be proved very similarly (see Appendix B).

**Proof**:[Theorem 7] The first three moments of $X = UNIFORM(l, u)$ are $E[X] = \frac{u^2-l^2}{2(u-l)}$, $E[X^2] = \frac{u^3-l^3}{3(u-l)}$, and $E[X^3] = \frac{u^4-l^4}{4(u-l)}$. and the normalized second and third moments of $X$ are

$$m_2^X = \frac{4}{3}\frac{1 + r + r^2}{(1 + r)^2} \quad \text{and} \quad m_3^X = \frac{3}{2}\frac{1 + r + r^2 + r^3}{(1 + r)(1 + r + r^2)},$$

where $r = \frac{l}{u}$. Since $\frac{\partial}{\partial r} m_2^X = \frac{4}{3}\frac{r-1}{(r+1)^3} \leq 0$ for $0 \leq r \leq 1$, $m_2^X$ is a nonincreasing function of $r$. So, the minimum of $m_2^X$ is given by evaluating it at $r = 1$ and the maximum is given by evaluating it at $r = 0$. Thus, $1 < m_2^X < \frac{4}{3}$. Also, it is easy to see that $m_2^X$ and $m_3^X$ satisfy $m_3^X = 3 - \frac{2}{m_2^X}$. ■

# 6 Conclusion

The contribution of this paper is a characterization of the set $S^{(n)}$ of distributions $G$ which are well-represented by an $n$-stage Coxian distribution. Prior work has only analyzed $S^{(2)^*} \subset S^{(2)}$, and this characterization was messy. We introduce several ideas which help in creating a simple formulation of $S^{(n)}$. The first is the concept of normalized moments. The second is the notion of $S_V^{(n)}$, a nearly complete subset of $S^{(n)}$ 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 $S^{(n)}$ provides the minimum number of necessary phases and the 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), the goal is to come up with a *minimal* (fewest number of stages) 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 [21], we develop an algorithm for finding a minimal Coxian distribution that well-represents a given distribution. We find that the simple characterization of $S^{(n)}$ 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 [13], which require knowing a priori the number of stages $n$ in the minimal Coxian distribution.

In addition to characterizing those distributions in $S^{(n)}$, we also consider which M/G/1 busy periods have durations in $S^{(n)}$. We find that the number of stages which suffice for a busy period duration to be well-represented by a Coxian distribution is, surprisingly, determined solely by the distribution of the *first job* in the busy period. Furthermore we classify a few examples of common and practical distributions as being subsets of $S^{(n)}$ for some $n$.

Future work includes a simple characterization of the set of distributions that are well-represented by general $n$-phase PH distributions. It is known that the Erlang distribution has the lowest normalized second moment among all the $n$-phase PH distributions [1]. However, a lower bound on the normalized third moment of $n$-phase PH distributions is not known.

# 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 modelling 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] M. Harchol-Balter. Task assignment with unknown duration. *Journal of the ACM*, 49(2), 2002.

[8] M. Harchol-Balter and A. Downey. Exploiting process lifetime distributions for dynamic load balancing. In *Proceedings of SIGMETRICS '96*, pages 13–24, 1996.

[9] M. Harchol-Balter, C. Li, T. Osogami, A. Scheller-Wolf, and M. Squillante. Task assignment with cycle stealing under central queue. In *To appear in Proceedings of 23rd International Conference on Distributed Computing Systems (*ICDCS '03*).*, May 2003. For a copy see http://www.cs.cmu.edu/∼harchol/.

[10] G. Irlam. Unix file size survey - 1993. Available at `http://www.base.com/gordoni/ufs93.html`, September 1994.

[11] M. A. Johnson and M. F. Taaffe. A graphical investigation of error bounds for moment-based queueing approximations. *Queueing Systems*, 8:295–312, 1991.

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

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

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

[15] R. E. A. Khayari, R. Sadre, and B. Haverkort. Fitting world-wide web request traces with the EM-algorithm. In *Proceedings of SPIE: Internet Performance and Control of Network Systems II*, volume 4523, pages 211–220, 2001.

[16] G. Latouche and V. Ramaswami. *Introduction to Matrix Analytic Methods in Stochastic Modeling*. ASA-SIAM, Philadelphia, 1999.

[17] W. E. Leland and T. J. Ott. Load-balancing heuristics and process behavior. In *Proceedings of Performance and ACM Sigmetrics*, pages 54–69, 1986.

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

[19] M. F. Neuts. *Matrix-Geometric Solutions in Stochastic Models*. Johns Hopkins University Press, 1981.

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

[21] 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, February 2003.

[22] T. Osogami, M. Harchol-Balter, and A. Scheller-Wolf. Analysis of cycle stealing with switching cost. In *To appear in ACM Sigmetrics 2003 Conference on Measurement and Modeling of Computer Systems (*SIGMETRICS '03*).*, June 2003. For a copy see http://www.cs.cmu.edu/∼harchol/.

[23] V. Paxson and S. Floyd. Wide-are traffic: The failure of poisson modeling. *IEEE/ACM Transactions on Networking*, pages 226 – 244, June 1995.

[24] D. L. Peterson and D. B. Adams. Fractal patterns in DASD I/O traffic. In *CMG Proceedings*, December 1995.

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

[26] 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.

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

[28] M. Telek and A. Heindl. Moment bounds for acyclic discrete and continuous phase type distributions of second order. In *Proceedings of UK Performance Evaluation Workshop, UKPEW 2002*, 2002.

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

[30] W. Whitt. On approximations for queues, iii: Mixtures of exponential distributions. *AT&T Bell Laboratories Technical Journal*, 63:163 – 175, 1984.

# A A proof of the Telek and Heindl theorem

In this section, we give a proof of Theorem 1.

Consider a two-stage Coxian distribution $X$, where an exponential distribution with rate $\mu_1$ is followed by an exponential distribution with rate $\mu_2$ with probability $p$. Let $m_i^X$ be the normalized $i$-th moment of $X$ for $i = 2, 3$. Without loss of generality, we assume $E[X] = 1$. Then,

$$\begin{aligned} 1 &= \frac{1}{\mu_1} + p\frac{1}{\mu_2}, \\ E[X^2] &= p\left(\frac{2}{\mu_1^2} + \frac{2}{\mu_1\mu_2} + \frac{2}{\mu_2^2}\right) + (1-p)\left(\frac{2}{\mu_1^2}\right), \\ E[X^3] &= p\left(\frac{6}{\mu_1^3} + \frac{6}{\mu_1^2\mu_2} + \frac{6}{\mu_1\mu_2^2} + \frac{6}{\mu_2^3}\right) + (1-p)\left(\frac{6}{\mu_1^3}\right). \end{aligned} \tag{9}$$

Let $u = \mu_1 + \mu_2$ and $v = \mu_1\mu_2$. Then, two solutions of (9) can be expressed as

$$\mu_1 = \frac{u + \sqrt{u^2 - 4v}}{2}, \quad \mu_2 = \frac{u - \sqrt{u^2 - 4v}}{2}, \quad \text{and} \quad p = \frac{\mu_2(\mu_1 - 1)}{\mu_1},$$

and

$$\mu_1 = \frac{u - \sqrt{u^2 - 4v}}{2}, \quad \mu_2 = \frac{u + \sqrt{u^2 - 4v}}{2}, \quad \text{and} \quad p = \frac{\mu_2(\mu_1 - 1)}{\mu_1},$$

where $u = \frac{6 - 2m_3^X}{3m_2^X - 2m_3^X}$ and $v = \frac{12 - 6m_2^X}{m_2^X(3m_2^X - 2m_3^X)}$. The solution is feasible if and only if the following conditions are satisfied: (i) $\mu_1$ and $\mu_2$ are real numbers, (ii) $\mu_1, \mu_2 > 0$, (iii) $p \le 1$, and (iv) $p \ge 0$. In the following, we will first express the necessary and sufficient conditions in terms of $u$ and $v$, and then the theorem is proven.

The next lemma gives the necessary and sufficient conditions in terms of $u$ and $v$:

**Lemma 9.1** *For at least one solution to be feasible, the following conditions are necessary: $v \le \frac{1}{4}u^2$, $u > 0$, $v > 0$, and $v \le u$. Provided that these conditions are satisfied, the first solution is feasible if and only if $u \ge 2$ or $v \le u - 1$, and the second solution is feasible if and only if $v \ge u - 1$ and $u \ge 2$.*

Before proving the lemma, we state a few corollaries:

**Corollary 4** *The second solution is feasible only when the first solution is feasible.*

Thus, we only need to consider the first solution.

**Corollary 5** *The first solution is feasible if and only if*

$$0 < v \le \begin{cases} u - 1 & \text{when } 1 < u \le 2 \\ \frac{u^2}{4} & \text{when } 2 < u \le 4 \\ u & \text{when } 4 < u. \end{cases}$$

**Proof**:[Lemma 9.1] We prove that each of the four conditions for the feasibility of the solution corresponds to one of the conditions in the lemma. Namely, we show (i) $v \le \frac{1}{4}u^2$ iff $\mu_1$ and $\mu_2$ are real numbers; (ii) $u > 0$ and $v > 0$ iff $\mu_1, \mu_2 > 0$; (iii) $v \le u$ iff $p \le 1$. Also, provided that these conditions are satisfied, we show (iv) $u \ge 2$ or $v \le u - 1$ iff $p \ge 0$; (v) $v \ge u - 1$ and $u \ge 2$ iff $p \ge 0$.

(i) and (ii) are immediate from the expression of the solutions.

(iii) First, we consider the first solution. $p \le 1$ if and only if

$$\frac{\mu_2(\mu_1-1)}{\mu_1} \le 1 \iff \frac{u-\sqrt{u^2-4v}}{2}\left(\frac{u+\sqrt{u^2-4v}}{2}-1\right) \le \frac{u+\sqrt{u^2-4v}}{2} \iff v \le u.$$

Next, we consider the second solution. $p \le 1$ if and only if

$$\frac{\mu_2(\mu_1-1)}{\mu_1} \le 1 \iff \frac{u+\sqrt{u^2-4v}}{2}\left(\frac{u-\sqrt{u^2-4v}}{2}-1\right) \le \frac{u-\sqrt{u^2-4v}}{2} \iff v \le u.$$

(iv) Notice that $\mu_1 = \frac{u+\sqrt{u^2-4v}}{2}$.

$$p \ge 0 \iff \mu_1 \ge 1 \iff \sqrt{u^2-4v} \ge 2-u.$$

When $u \ge 2$, $\mu_1 \ge 1$ is always true, and when $u < 2$,

$$\mu_1 \ge 1 \iff u^2-4v \ge (2-u)^2 \iff u-v \ge 1.$$

(v) Notice that $\mu_1 = \frac{u-\sqrt{u^2-4v}}{2}$.

$$p \ge 0 \iff \mu_1 \ge 1 \iff \sqrt{u^2-4v} \le u-2 \iff u^2-4v \le (u-2)^2, u \ge 2 \iff u-v \le 1, u \ge 2.$$

■

Next, we state the condition in terms of $m_2^X$ and $m_3^X$. This gives a proof of the theorem.

**Proof**:[Theorem 1] We prove that the condition in Corollary 5 corresponds to the conditions in the theorem. First, we translate the three conditions $u \ge 1$, $u \ge 2$, and $u \ge 4$ into the conditions with respect to $m_2^X$ and $m_3^X$.

$$\begin{aligned}
u \ge 1 \quad &\iff \quad \frac{6-2m_3^X}{3m_2^X-2m_3^X} \ge 1 \\
&\iff \quad \left(\frac{3}{2}m_2^X > m_3^X \quad \text{and} \quad 0 \le m_2^X \le 2\right) \quad \text{or} \quad \left(\frac{3}{2}m_2^X < m_3^X \quad \text{and} \quad 2 \le m_2^X\right); \\
u \ge 2 \quad &\iff \quad \frac{6-2m_3^X}{3m_2^X-2m_3^X} \ge 2 \\
&\iff \quad \left(\frac{3}{2}m_2^X > m_3^X \quad \text{and} \quad m_3^X \ge 3m_2^X-3\right) \quad \text{or} \quad \left(\frac{3}{2}m_2^X < m_3^X \quad \text{and} \quad m_3^X \le 3m_2^X-3\right); \\
u \ge 4 \quad &\iff \quad \frac{6-2m_3^X}{3m_2^X-2m_3^X} \ge 4 \\
&\iff \quad \left(\frac{3}{2}m_2^X > m_3^X \quad \text{and} \quad m_3^X \ge 2m_2^X-1\right) \quad \text{or} \quad \left(\frac{3}{2}m_2^X < m_3^X \quad \text{and} \quad m_3^X \le 2m_2^X-1\right).
\end{aligned}$$

Next, we translate the conditions $v \le u-1$, $v \le \frac{u^2}{4}$, $v \le u$, and $v > 0$ into the conditions with respect to $m_2^X$ and $m_3^X$.

$$v \le u-1 \quad \iff \quad 1 \le \frac{12m_2^X-2m_2^Xm_3^X-12}{m_2^X(3m_2^X-2m_3^X)} \iff (3 > m_3^X \quad \text{and} \quad m_2^X = 2) \quad \text{or} \quad \left(\frac{3}{2}m_2^X < m_3^X\right);$$

$$\begin{aligned}
v \le \frac{u^2}{4} \quad &\Longleftrightarrow \quad u^2 - 4v \ge 0 \\
&\Longleftrightarrow \quad \frac{4}{m_2^X(3m_2^X - 2m_3^X)^2}\left((3 - m_3^X)^2 m_2^X - (12 - 6m_2^X)(3m_2^X - 2m_3^X)\right) \ge 0 \\
&\Longleftrightarrow \quad \left(m_2^X \ge 2\right) \quad \text{or} \quad \left(m_2^X < 2 \quad \text{and} \quad (m_3^X \ge \frac{3\left(3m_2^X - 4 + \sqrt{2}(2 - m_2^X)^{\frac{3}{2}}\right)}{m_2^X}\right); \\
v \le u \quad &\Longleftrightarrow \quad \frac{12 - 6m_2^X}{m_2^X(3m_2^X - 2m_3^X)} \le \frac{6m_2^X - 2m_2^X m_3^X}{m_2^X(3m_2^X - 2m_3^X)} \\
&\Longleftrightarrow \quad \left(6 - \frac{6}{m_2^X} \ge m_3^X \quad \text{and} \quad \frac{3}{2}m_2^X \ge m_3^X\right) \text{or} \left(6 - \frac{6}{m_2^X} \le m_3^X \quad \text{and} \quad \frac{3}{2}m_2^X \le m_3^X\right); \\
v > 0 \quad &\Longleftrightarrow \quad \left(m_2^X \ge 2 \quad \text{and} \quad m_3^X > \frac{3}{2}m_2^X\right) \quad \text{or} \quad \left(0 \le m_2^X < 2 \quad \text{and} \quad m_3^X < \frac{3}{2}m_2^X\right).
\end{aligned}$$

Summarizing these conditions gives the conditions in the theorem.

When $m_2^X = 2$ and $m_3^X = 3$, the distribution is exponential and a one-stage Coxian distribution can match the first three moments. ■

# B Proof of Theorems 8 and 9

**Proof**:[Theorem 8] The first three moments of $X = TRIANGULAR(l, u)$ are

$$E[X] = \frac{u+l}{2}, \quad E[X^2] = \frac{7u^2 + 10ul + 7l^2}{24}, \quad \text{and} \quad E[X^3] = \frac{3u^3 + 5u^2 l + 5ul^2 + 3l^3}{16}.$$

and the normalized second and third moments are

$$m_2^X = \frac{7 + 10r + 7r^2}{6(1+r)^2} \quad \text{and} \quad m_3^X = \frac{3(3 + 5r + 5r^2 + 3r^3)}{(1+r)(7 + 10r + 7r^2)},$$

where $r = \frac{l}{u}$.

Since $\frac{\partial}{\partial r} m_2^X = \frac{2}{3}\frac{r-1}{(r+1)^3} \le 0$ for $0 \le r \le 1$, $m_2^X$ is a nonincreasing function of $r$. So, the minimum of $m_2^X$ is given by evaluating it at $r = 1$ and the supremum is given by evaluating at $r = 0$. Thus, $1 \le m_2^X \le \frac{7}{6}$. Also, it is easy to see that $m_2^X$ and $m_3^X$ satisfy $m_3^X = 3 - \frac{2}{m_2^X}$. ■

**Proof**:[Theorem 9] The first three moments of $X = PARETO(\alpha)$ are

$$E[X] = \frac{\alpha}{\alpha - 1}, \quad E[X^2] = \frac{\alpha}{\alpha - 2}, \quad \text{and} \quad E[X^3] = \frac{\alpha}{\alpha - 3},$$

and the normalized second and third moments are

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

Since $\frac{\partial}{\partial r} m_2^X = -\frac{2(\alpha-1)}{(\alpha-2)^2\alpha^2} < 0$ for $\alpha > 3$, $m_2^X$ is a decreasing function of $\alpha$. So, the supremum of $m_2^X$ is given by evaluating it at $\alpha = 3$ and the infimum is given by letting $\alpha \to \infty$. Thus, $1 < m_2^X < \frac{4}{3}$.

Also, it is easy to see that $m_2^X$ and $m_3^X$ satisfy

$$m_3^X = \frac{-2(m_2^X)^2 + 3m_2^X + 2(m_2^X - 1)\sqrt{m_2^X(m_2^X - 1)}}{4 - 3m_2^X}.$$

■

## C Technical lemmas

**Lemma 9.2** *This lemma proves that (3) and (4) are equivalent sets.*

**Proof**: Recall that set (3) is the union of the following three sets:

$$\begin{aligned}
A_1 &= \left\{ \frac{9pm_2 - 12 + 3\sqrt{2}(2 - pm_2)^{\frac{3}{2}}}{p^2 m_2} \le m_3 \le \frac{6(pm_2 - 1)}{p^2 m_2} \cap \frac{3}{2p} \le m_2 < \frac{2}{p} \right\}, \\
A_2 &= \left\{ m_3 = \frac{3}{p} \cap m_2 = \frac{2}{p} \right\}, \\
A_3 &= \left\{ \frac{3}{2} m_2 < m_3 \cap \frac{2}{p} < m_2 \right\};
\end{aligned}$$

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

$$\begin{aligned}
B_1 &= \left\{ \frac{4}{3} m_2 \le m_3 \le \frac{6(m_2 - 1)}{m_2} \cap \frac{3}{2} \le m_2 \le 2 \right\}, \\
B_2 &= \left\{ \frac{4}{3} m_2 \le m_3 \le \frac{3}{2} m_2 \cap 2 < m_2 \right\}, \\
B_3 &= \left\{ \frac{3}{2} m_2 < m_3 \cap 2 < m_2 \right\}
\end{aligned}$$

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 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 $Z \in A_1$. We first show that $Z \in B_1 \cup B_2$. Let $u(p)$ be the upper bound of $m_3^Z$:

$$u(p) = \frac{6(pm_2^Z - 1)}{p^2 m_2^Z},$$

and let $l(p)$ be the lower bound of $m_3^Z$:

$$l(p) = \frac{3\left(3pm_2^Z - 4 + \sqrt{2}(2 - pm_2^Z)^{\frac{3}{2}}\right)}{p^2 m_2^Z}.$$

Then, $u(p)$ and $l(p)$ are both continuous and increasing functions of $p$ for $\frac{3}{2m_2^Z} \le p \le \frac{2}{m_2^Z}$ by Lemmas 9.3 and 9.4. When $m_2^Z \le 2$, the range of $p$ is $\frac{3}{2m_2^Z} \le p \le 1$. Thus,

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

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

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

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

**Lemma 9.3** *Let $a > 0$. Then, $f(p) = \frac{ap-1}{p^2}$ is an increasing function of p for $0 < p < \frac{2}{a}$.*

**Proof**: Note that $f'(p) = \frac{-ap+2}{p^3} > 0$. The inequality holds when $0 < p < \frac{2}{a}$. ■

**Lemma 9.4** *Let $a > 0$. Then, $f(p) = \frac{3ap-4+\sqrt{2}(2-ap)^{\frac{3}{2}}}{p^2}$ is an increasing function of p for $\frac{3}{2a} \leq p \leq \frac{2}{a}$.*

**Proof**:

$$f'(p) \quad = \quad \frac{8 - 3ap - \frac{\sqrt{2}}{2}(8-ap)(2-ap)^{\frac{1}{2}}}{p^3}$$

Let $g(p) = 8 - 3ap - \frac{\sqrt{2}}{2}(8-ap)(2-ap)^{\frac{1}{2}}$. Then,

$$g'(p) = 3a\left(\frac{\sqrt{2}}{4}\frac{(4-ap)}{(2-ap)^{\frac{1}{2}}} - 1\right) \geq 3a\left(\frac{\sqrt{2}}{4}\frac{(4-ap)}{(2-\frac{3}{2})^{\frac{1}{2}}} - 1\right) \geq 3a\left(\frac{\sqrt{2}}{4}\frac{(4-2)}{(2-\frac{3}{2})^{\frac{1}{2}}} - 1\right) = 0$$

The first inequality follows from $\frac{3}{2a} \leq p$ and the second inequality follows from $p \leq \frac{2}{a}$. So, $g(p)$ is a non-decreasing function of $p$. Thus, $f'(p) \geq f'(\frac{3}{2}a) = \frac{\frac{1}{4}}{p^3} > 0$. ■

**Lemma 9.5** *Let $y \geq 0$ and $k \geq 1$. Then,*

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

**Proof**: Let

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

We prove that $g(y,k) \geq 0$. Let

$$h(y,k) = (2+4y+y^2)k^3 - 2(1+2y+4y^2+y^3)k^2 - (2+4y+y^2-5y^3-y^4)k + 2(1+y)(1+y+3y^2).$$

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

$$d(y) \geq (4 + 8y + y^2 + y^3)^2. \tag{10}$$

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

$$k = \frac{2 + 4y + 8y^2 + 2y^3 + \sqrt{d(y)}}{3(2 + 4y + y^2)}. \tag{11}$$

Let

$$\begin{aligned}
H(y) &= h\left(y, \frac{2 + 4y + 8y^2 + 2y^3 + \sqrt{d(y)}}{3(2 + 4y + y^2)}\right) \\
&= \frac{2((28 + 83y + 16y^2 + y^3)d(y) - d(y)^{\frac{3}{2}} - 6(64 + 456y + 1260y^2 + 1655y^3 + 889y^4 + 147y^5))}{27(2 + 4y + y^2)^2}.
\end{aligned}$$

It suffices to prove $H(y) \geq 0$. Let $G(y)$ be the numerator of $H(y)$. It suffices to prove $G(y) \geq 0$. Notice that $G(0) = 0$. Thus, it suffices to prove $G'(y) \geq 0$ for $y \geq 0$.

$$G'(y) = \frac{3}{\sqrt{d(y)}} F(y),$$

where

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

■

**Lemma 9.6** $f(r) = \frac{(r-1)^2}{r(\log r)^2}$ *is an increasing function for* $r > 1$.

**Proof**: Note that

$$f'(r) = \frac{r-1}{r^2(\log r)^3}\left(2 - 2r + (1+r)\log r\right).$$

Note that $\frac{r-1}{r^2(\log r)^3} > 0$ for $r > 1$, and let $g(r) = 2 - 2r + (1+r)\log r$. Then, $g(r)$ is positive for $r > 1$, since $g(1) = 0$, $g'(r) = \frac{1}{r} + \log r - 1$, $g'(1) = 0$, and $g''(r) = \frac{r-1}{r^2} > 0$. ■

**Lemma 9.7** $f(r) = \frac{r+1}{r-1}\log r$ *is an increasing function for* $r > 1$.

**Proof**: Note that

$$f'(r) \quad = \quad \frac{r^2 - 2r\log r - 1}{(r-1)^2 r}.$$

Note that $\frac{1}{(r-1)^2 r} > 0$ for $r > 1$, and let $g(r) = r^2 - 2r\log r - 1$. Then, $g(r)$ is positive for $r > 1$, since $g(1) = 0$, $g'(r) = 2r - 2\log r - 2$, $g'(1) = 0$, and $g''(r) = \frac{2(r-1)}{r} > 0$. ■

**Lemma 9.8** *Let $0 < \alpha < 1$ or $1 < \alpha < 2$. Then, $f(r) = \frac{(r^\alpha - 1)(r^2 - r^\alpha)}{(r - r^\alpha)^2}$ is an increasing function for $r > 1$.*

**Proof**: Note that

$$f'(r) = \frac{(r-1)r^\alpha}{(r-r^\alpha)^3}\left(\alpha r + (2-\alpha) + (\alpha-2)r^\alpha - \alpha r^{\alpha-1}\right).$$

Note that

$$\frac{(r-1)r^\alpha}{(r-r^\alpha)^3}\begin{cases} > 0 & (\text{when} \quad 0 < \alpha < 1) \\ < 0 & (\text{when} \quad 1 < \alpha < 2) \end{cases}$$

for $r > 1$. Let $g(r) = \alpha r + (2-\alpha) + (\alpha-2)r^\alpha - \alpha r^{\alpha-1}$. Then,

$$g(r)\begin{cases} > 0 & (\text{when} \quad 0 < \alpha < 1) \\ < 0 & (\text{when} \quad 1 < \alpha < 2) \end{cases}$$

for $r > 1$, since $g(1) = 0$, $g'(r) = \alpha + (\alpha-2)(1+\alpha)r^{\alpha-1} + (2-\alpha) - \alpha^2 r^{\alpha-2}$, $g'(1) = 0$, and

$$\begin{aligned} g''(r) \quad &= \quad \alpha(\alpha-1)(\alpha-2)r^{\alpha-3}(r-1), \\ &\begin{cases} > 0 & (\text{when} \quad 0 < \alpha < 1) \\ < 0 & (\text{when} \quad 1 < \alpha < 2). \end{cases} \end{aligned}$$

■

**Lemma 9.9** *Let $f(r) = \frac{(r-r^\alpha)(r^3-r^\alpha)}{(r^2-r^\alpha)^2}$. If $0 < \alpha < 1$, $f(r)$ is an increasing function for $r > 1$. If $1 < \alpha < 2$, $f(r)$ is a decreasing function for $r > 1$.*

**Proof**: Note that

$$f'(r) = \frac{(r-1)r^{\alpha+2}}{(r^2-r^\alpha)^3}\left(-(3-\alpha)(1-r^{\alpha-1}) - (\alpha-1)(r-r^{\alpha-2})\right).$$

Since $\frac{(r-1)r^{\alpha+2}}{(r^2-r^\alpha)^3} > 0$ for $r > 1$, it is easy to see that $f'(r) < 0$ for $r > 1$ when $1 < \alpha < 2$. To see that $f'(r) > 0$ for $r > 1$ when $0 < \alpha < 1$, we let

$$g(r) = -(3-\alpha)(1-r^{\alpha-1}) - (\alpha-1)(r-r^{\alpha-2}).$$

Then, $g(r) > 0$ for $r > 1$, since $g(1) = 0$, $g'(r) = (3-\alpha)(\alpha-1)(r^{\alpha-2}) - (\alpha-1)(1-(\alpha-2)r^{\alpha-3})$, $g'(1) = 0$, and

$$g''(r) \quad = \quad -(\alpha-1)(\alpha-2)(\alpha-3)(r^{\alpha-3})r^{\alpha-4}(r-1)$$

$$\begin{cases} > 0 & (\text{when} \quad 0 < \alpha < 1) \\ < 0 & (\text{when} \quad 1 < \alpha < 2). \end{cases}$$

■