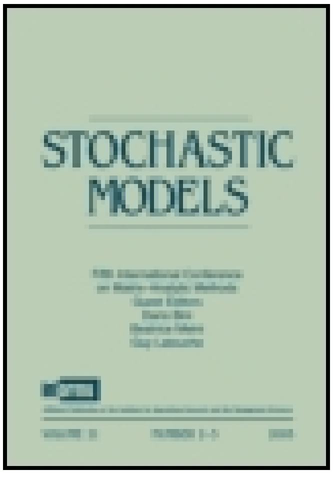

Communications in Statistics. Stochastic Models

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/lstm19

Matching moments to phase distributions: Mixtures of erlang distributions of common order

Mary A. Johnson [a] & Michael R. Taaffe [b]

[a] School of Industrial Engineering , Purdue University , West Lafayette, Indiana, 47907

[b] School of Industrial Engineering , Purdue University , West Lafayette, Indiana, 47907

Published online: 13 Dec 2007.

COMMUN. STATIST.-STOCHASTIC MODELS, 5(4), 711-743 (1989)

MATCHING MOMENTS TO PHASE DISTRIBUTIONS: MIXTURES OF ERLANG DISTRIBUTIONS OF COMMON ORDER

MARY A. JOHNSON

MICHAEL R. TAAFFE

School of Industrial Engineering
Purdue University
West Lafayette, Indiana 47907

ABSTRACT

One approach to the moment-matching problem for phase distributions to to restrict selection to an appropriate subset of phase distributions. We investigate the use of mixtures of Erlang distributions of common order to match moments feasible for distributions with support on $[0, \infty)$. We show that, except for special cases, the first k (finite) moments of any nondegenerate distribution with support on $[0, \infty)$ can be matched by a mixture of Erlang distributions of (sufficiently high) common order. Moreover, we show that any k-tuple of first k moments feasible for a mixture of n-stage Erlang distributions (E_n's) is feasible for a mixture of $\lfloor k/2 \rfloor + 1$ or fewer E_n's. The three-moment-matching problem is considered in detail. The set of pairs of second and third standardized moments feasible for mixtures of E_n's is characterized. An analytic expression is derived for the minimum order, n, such that a given set of first three moments is feasible for a mixture of E_n's. Expressions are also given for the parameters of the unique mixture of two E_n's that matches a feasible set of first three moments. Methods for implementation of these results are suggested and evaluated. In our evaluation, we consider distributional properties such as dimension, numerical stability, and density-function shape.

1. INTRODUCTION

The family of phase-type (PH) distributions is an important tool in algorithmic probability. The prominence of PH distributions in stochastic models is explained by its flexibility and its favorable analytic and computational properties. However, realization of the full utility of the PH family depends on the availability of methods for efficiently selecting PH distributions that are suitable for the intended computational analysis and adequately reflect the randomness being modelled.

In this paper, we focus on the method of moment matching and the subset of the PH family consisting of mixtures of Erlang distributions of common order. We prove the feasibility of matching the first k (finite) moments feasible for a distribution with support on $[0, \infty)$ to a mixture of Erlang distributions of (sufficiently high) common order. (We define *support* of a distribution to be any set over which the associated probability equals one.) Further, we show that a set of first k moments feasible for a mixture of Erlang distributions of order n is feasible for a mixture of $\lfloor k/2 \rfloor + 1$ or fewer Erlang distributions of order n. We investigate theoretical and empirical aspects of the $k = 3$ case in detail. The set of first three moments feasible for mixtures of Erlang distributions of order n is described. We then show that any triple of first three moments feasible for some PH distribution is feasible for a mixture of two Erlang distributions of (sufficiently large) common order. Also, for fixed first three moments, we derive the minimum n such that the moments are feasible for a mixture of two Erlang distributions of order n. Finally, we derive the parameters of the unique mixture of two Erlang distributions of order n that matches a feasible triple of first three moments. Implementation issues associated with procedures suggested by our theoretical results are investigated. We consider several properties of selected distributions, including dimension, numerical stability, and density-function shape.

The organization of this paper is as follows. Section 2 provides background material, which consists of an introduction to the PH family and our motivation for emphasizing three-moment results. In Section 3 we present theoretical results on matching k moments to mixtures of Erlang distributions. In Section 4 we develop a framework for the three-moment-matching problem. In Section 5 feasible pairs of second and third standardized moments of mixtures of Erlang distributions of common order are identified and characterized. Moment-matching formulas are presented in Section 6. Section 7 provides a discussion of issues that arise in implementation of the results in Sections 5 and 6. Section 8 provides a brief summary and conclusion.

2. BACKGROUND

2.1 Introduction to PH distributions. A PH distribution is defined as the distribution of the time until absorption in a finite-state Markov process with n ($n = 1, 2, \cdots$) transient states and one absorbing state, state $n+1$. The parameters of a PH distribution are its dimension n and the elements of an n-dimensional row vector $\vec{\alpha}$ and an $n \times n$ matrix $\mathbf{T}$. The elements of $\vec{\alpha}$ are the initial-state probabilities assigned to each of the transient states. The matrix $\mathbf{T}$ is obtained by deleting the last row and last column of the generator matrix of the Markov process associated with the PH distribution. The pair $(\vec{\alpha}, \mathbf{T})$ is called the *representation* of the PH distribution specified by it. An important property of the PH family is that PH distributions do not have unique representations. Special cases of PH distributions include the exponential distribution, Erlang distributions (convolutions of identical exponential distributions), and hyperexponential distributions (mixtures of exponential distributions).

The cumulative distribution function (cdf) associated with the representation $(\vec{\alpha}, \mathbf{T})$ is $F(x) = 1 - \vec{\alpha} \exp(\mathbf{T}\, x)\vec{e}$, where $\vec{e}$ is a $n{\times}1$ column vector of one's. The matrix $\mathbf{T}$ is nonsingular, since states $1, 2, \cdots, n$ are transient if and only if T is nonsingular. The kth noncentral moment of F is $\mu_k = (-1)^k\, k!\, \vec{\alpha}\, \mathbf{T}^{-k} \vec{e}$. Thus, the existence of $\mathbf{T}^{-1}$ implies that all noncentral moments of F are finite.

The denseness of the PH family in the set of distributions with support on $[0, \infty)$ is well known. See Johnson and Taaffe [8] for a precise statement and rigorous proof of the denseness property. The denseness proof is based on mixtures of Erlang distributions with common rate parameter. But, this family is such a restrictive subset of PH distributions that its exclusive use in a selection algorithm often leads to selected PH distributions of high dimension. See Bux and Herzog [4]. So, the proof of the theoretical versatility of PH distributions does not naturally lead to algorithmic methods for exploiting this versatility. Rather, the richness of the PH family, compounded with the nonuniqueness of PH-distribution representations, makes the PH-distribution-selection problem elusive. This suggests that a reasonable approach to the selection problem is to restrict selection to a PH-family subset that is not too restrictive nor unnecessarily general. Our results indicate that mixtures of Erlang distributions of common order may be an appropriate subset in the context of moment matching.

2.2 Three-moment approximations. Moment matching is a common method for approximating distributions, especially in the area of queueing approximations. Though two-moment queueing approximations are common, they may lead to serious error when the coefficient of variation, c, (the standard deviation divided by the mean) is high. See Whitt [18, 19], Klincewicz and Whitt [12], and Johnson and Taaffe [10].

Whitt [17] and Altiok [3] provide analytic three-moment-matching results for the $c > 1$ case. Whitt matches three moments to a two-stage hyperexponential distribution (H_2); Altiok matches moments to an alternative representation of the H_2 selected by Whitt. (This is easily shown using the results of Cumani [5].) However, the family of distributions used by Whitt and Altiok accommodates only some of the feasible combinations of three moments when $c > 1$ and none of the feasible combinations when $c \leq 1$. Our results on matching three moments to mixtures of two Erlang distributions generalize Whitt's results on matching three moments to H_2's and accommodates any triple of first three moments feasible for some PH distribution.

3. MATCHING MOMENTS TO MIXTURES OF ERLANG DISTRIBUTIONS OF COMMON ORDER

In this section, we derive results on matching k moments to mixtures of Erlang distributions of common order. We use three lemmas to prove our main result, Theorem 1. Lemma 1 describes the feasible region in k-dimensional Euclidean space for vectors of first k noncentral moments of distributions with support on $[0, \infty)$. Lemma 2 is the key to deriving results about mixtures of Erlang distributions from results about distributions with support on $[0, \infty)$. This lemma requires the assumption that all of the mixed Erlang distributions have common order. Lemma 3 motivates the second statment in Theorem 1 by showing that a discrete distribution with mass at $\lfloor k/2 \rfloor + 1$ or fewer points is sufficient to match the first k moments of any distribution with support on $[0, \infty)$.

To state our results precisely, we introduce the following notation. The k-tuple $\vec{\mu}_k = (\mu_1, \mu_2, \cdots, \mu_k)$ of first k noncentral moments of a distribution is a point in k-dimensional Euclidean space. The set of nondegenerate distributions with support on $[0, \infty)$ and finite moments is denoted by $\mathcal{F}$. (To avoid cumbersome details, we consider only nondegenerate distributions with finite moments. However, for results about $\vec{\mu}_k$, we need only assume that the first k or $k + 1$ moments are finite, depending on whether k is odd or even. See the proof of Lemma 3.) The set of finite mixtures of Erlang distributions of common order is denoted by $\mathcal{ME}$; the subset of $\mathcal{ME}$ that consists of mixtures of n-stage Erlang distributions is denoted by $\mathcal{ME}_n$; and the subset of $\mathcal{ME}_n$ that consists of mixtures of m n-stage Erlang distributions, each with positive mixing probability and distinct rate parameter, is denoted by $\mathcal{ME}_n(m)$. The set of points $\vec{\mu}_k$ that are feasible for distributions in a set $\mathcal{C}$ is denoted by $M_k(\mathcal{C})$. The closure of a set S is denoted by $\bar{S}$; the interior of S is S^o; and the boundary of S is $\delta(S)$.

The determinants of the two matrices defined below are used to specify $M_k(\mathcal{F})$ for $k \geq 2$. Let $\det(C)$ denote the determinant of square matrix C. For $s = 1, 2, \cdots, \lfloor k/2 \rfloor$, let

$$A(\vec{\mu}_{2s}) = \left[\mu_{i+j}\right]_{i,j=0}^{s} = \begin{bmatrix} \mu_0 & \mu_1 & \cdots & \mu_s \\ \mu_1 & \mu_2 & \cdots & \mu_{s+1} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \mu_s & \mu_{s+1} & \cdots & \mu_{2s} \end{bmatrix},$$

and for $s = 0, 1, \cdots, \lfloor (k-1)/2 \rfloor$, let $B(\vec{\mu}_{2s+1}) = \left[\mu_{i+j+1}\right]_{i,j=0}^{s}$.

Lemma 1 follows from the theory of Tchebycheff systems. Karlin and Studden [11] is a general treatment of this topic. Johnson and Taaffe [9] introduces Tchebycheff systems in the context of probability applications and moment matching. Lemma 1 is a restatement of Theorem 6 in [9], restricted to allow for nondegenerate distributions only. This restatement uses the result that a symmetric real matrix is positive definite (positive semi-definite) if and only if all of its leading principle minors are positive (nonnegative), Strang [16, pp. 238, 244].

Lemma 1: Let $\mu_0 = 1$ and assume $k \geq 2$.

Part I:
$\vec{\mu}_k \in M_k^o(\mathcal{F})$ if and only if for $s = 1, 2, \cdots, \lfloor k/2 \rfloor$, $\det(A(\vec{\mu}_{2s})) > 0$, and for $s = 0, 1, \cdots, \lfloor (k-1)/2 \rfloor$, $\det(B(\vec{\mu}_{2s+1})) > 0$.

Part II:
$\vec{\mu}_k \in \bar{M}_k(\mathcal{F})$ if and only if for $s = 1, 2, \cdots, \lfloor k/2 \rfloor$, $\det(A(\vec{\mu}_{2s})) \geq 0$, and for $s = 0, 1, \cdots, \lfloor (k-1)/2 \rfloor$, $\det(B(\vec{\mu}_{2s+1})) \geq 0$.

The following notation is used to describe Erlang distributions and mixtures of Erlang distributions. Denote an n-stage Erlang distribution by E_n and an E_n with rate parameter λ by $E_n(\lambda)$. Let $\vec{\lambda}_m = (\lambda_1, \lambda_2, \cdots, \lambda_m)$ and $\vec{p}_m = (p_1, p_2, \cdots, p_m)$, where the λ_i's are distinct and $p_i > 0$ for $i = 1, 2, \cdots, m$. Let ME_n denote a mixture of E_n's; $ME_n(m)$ denote a mixture of m E_n's; and $ME_n(m, \vec{p}_m, \vec{\lambda}_m)$ denote a mixture of Erlang distributions $E_n(\lambda_1), E_n(\lambda_2), \cdots, E_n(\lambda_m)$, where $E_n(\lambda_i)$ is assigned mixing probability p_i, $i = 1, 2, \cdots, m$. Also, let $P_{n,t} = (n+t-1)!/(n-1)!$, the number of permutations of $n+t-1$ items taken t at a time.

All of the key theoretical results in this paper make use of Lemma 2. Lemma 2 states the relationship between the noncentral moments of an $ME_n(m, \vec{p}_m, \vec{\lambda}_m)$ and a related distribution G, which has the following interpretation. If the $ME_n(m, \vec{p}_m, \vec{\lambda}_m)$ is interpreted as an E_n with random mean Λ^{-1}, then G is the distribution of Λ^{-1}.

Lemma 2: Let G denote the discrete probability distribution that assigns probability p_i to λ_i^{-1}, where $\lambda_i > 0$, for $i = 1, 2, \cdots, m$. Let μ_t, $\mu_t(i)$ $(i = 1, 2, \cdots, m)$, and θ_t denote the tth noncentral moment of $ME_n(m, \vec{p}_m, \vec{\lambda}_m)$, $E_n(\lambda_i)$, and G, respectively. Then, the moments of $ME_n(m, \vec{p}_m, \vec{\lambda}_m)$ and G are related as follows.

$$\mu_t = P_{n,t}\,\theta_t, \qquad t = 1, 2, \cdots \tag{1}$$

Proof: From the Laplace transform of $E_n(\lambda_i)$, one can easily show that $\mu_t(i) = P_{n,t}\lambda_i^{-t}$. Hence,

$$\mu_t = \sum_{i=1}^{m} p_i \mu_t(i) = \sum_{i=1}^{m} p_i P_{n,t}\,\lambda_i^{-t} = P_{n,t} \sum_{i=1}^{m} p_i\,\lambda_i^{-t} = P_{n,t}\,\theta_t \square$$

Lemma 3: For $k \geq 2$ and any $\vec{\mu}_k \in M_k(\mathcal{F})$, there exists a discrete distribution in $\mathcal{F}$ with mass at $\lfloor k/2 \rfloor + 1$ or fewer points that corresponds to $\vec{\mu}_k$. If $\vec{\mu}_k \in M_k^o(\mathcal{F})$, then this distribution may be chosen so that it has no mass at zero.

Proof: See Appendix. □

Recall that the Erlang distribution with rate parameter $n\lambda$ converges to the degenerate distribution at $1/\lambda$ as $n \to \infty$. Thus, Lemma 3 suggests that for $\vec{\mu}_k \in M_k^o(\mathcal{F})$, $\vec{\mu}_k$ can at least be approximated by the moments of a mixture of $\lfloor k/2 \rfloor + 1$ or fewer Erlang distributions of high order.

The proof of Theorem 1 is tedious and has been placed in the Appendix. Briefly, the reasoning is as follows. Lemmas 1 and 2 imply $M_k^o(\mathcal{F}) = M_k^o(\mathcal{ME})$. Lemmas 2 and 3 imply that for any point in $M_k^o(\mathcal{ME}_n)$, there exists an $ME_n(m)$, where $m \leq \lfloor k/2 \rfloor + 1$, that matches it.

Theorem 1: For any $\vec{\mu}_k \in M_k^o(\mathcal{F})$, there exists n^* such that for any $n \geq n^*$, $\vec{\mu}_k \in M_k^o(\mathcal{ME}_n)$. Thus, $M_k^o(\mathcal{F}) = M_k^o(\mathcal{ME})$. Further, $M_k(\mathcal{ME}_n) = \bigcup_{m=1}^{\lfloor k/2 \rfloor + 1} M_k(\mathcal{ME}_n(m))$. That is, any k-tuple of moments feasible for a mixture of E_n's is feasible for a mixture of $\lfloor k/2 \rfloor + 1$ or fewer E_n's.

4. A FRAMEWORK FOR THE THREE-MOMENT-MATCHING PROBLEM

Our attention now turns to the three-moment-matching problem. We begin by introducing standardized moments and describing the feasible sets of first and second standardized moments for distributions in $\mathcal{F}$.

4.1 Standardized moments. For distributions in $\mathcal{F}$, we define standardized moments as follows. For $t = 2, 3, \cdots$, let $\bar{\mu}_t$ be the tth central moment of a distribution. The second standardized moment is the coefficient of variation, which is denoted by c and defined as $\bar{\mu}_2^{1/2}/\mu_1$. For $t = 3, 4, \cdots$, the tth standardized moment is $\bar{\mu}_t/\bar{\mu}_2^{t/2}$. The third standardized moment, the coefficient of skewness, is denoted by γ. Because standardized moments are independent of the mean (a scale parameter) and reflect distribution shape, they provide a convenient perspective for the moment-matching problem. Thus, for $k \geq 2$, the first k noncentral moments of a nondegenerate distribution in $\mathcal{F}$ can be obtained by matching the second through kth standardized moments and then adjusting the scale of the selected distribution to obtain the mean. In the case of PH distributions, the change in scale is effected by multiplying the matrix T by $\mu(1)/\mu(2)$, where $\mu(1)$ is the mean of the distribution before rescaling and $\mu(2)$ is the mean being matched. For $k = 3$, the focus can be confined to matching c and γ.

4.2 Feasible regions. The set of feasible second and third standardized moments for distributions in $\mathcal{F}$ is characterized in Proposition 1. In the following, $\mathcal{B}$ denotes the set of *generalized Bernoulli distributions*, i.e., the set of distributions of random variables of the form aX, where X is a Bernoulli random variable and $a > 0$.

Proposition 1: Inequalities (2) and (3) are necessary and sufficient conditions for c and γ to be feasible for a distribution in $\mathcal{F}$.

$$c > 0 \tag{2}$$

$$\gamma \geq c - 1/c \tag{3}$$

Moreover, (3) is tight if and only if (c, γ) corresponds to an element of $\mathcal{B}$.

Proof: By Part II of Lemma 1, $\vec{\mu}_3 \in \bar{M}_3(\mathcal{F})$ if and only if (4) - (6) hold.

$$\mu_1 \geq 0 \tag{4}$$

$$\mu_2 - \mu_1^2 \geq 0 \tag{5}$$

$$\mu_1 \mu_3 - \mu_2^2 \geq 0 \tag{6}$$

Since (5) is tight only for degenerate distributions, the points on the line $\mu_2 - \mu_1^2 = 0$ are not in $M_3(\mathcal{F})$. Restricting (5) to a strict inequality leads to (2). One can easily verify that there is a one-to-one correspondence between the elements of $\mathcal{B}$ and the points in $\{\vec{\mu}_3 \mid \gamma = c - 1/c\}$. So, $\{\vec{\mu}_3 \mid \gamma = c - 1/c\} = M_3(\mathcal{B})$. Since $\gamma = c - 1/c$ implies that $\mu_1 \mu_3 - \mu_2^2 = 0$, $\vec{\mu}_3 \in \delta(M_3(\mathcal{F}))$. Thus, the distribution that corresponds to $\vec{\mu}_3$ is unique. (See the last paragraph of the proof of Lemma 3.) So, the only distribution in $\mathcal{F}$ that corresponds to $\vec{\mu}_3$ is a generalized Bernoulli distribution. Since $\mathcal{B} \subset \mathcal{F}$, $\{\vec{\mu}_3 \mid \gamma = c - 1/c\} \subset M_3(\mathcal{F})$. So, (3) follows from (6). □

Since generalized Bernoulli distributions are not PH distributions, Proposition 1 shows that a point $\vec{\mu}_3$ that satisfies $\gamma = c - 1/c$ cannot be matched by any PH distribution. However, any generalized Bernoulli distribution can be arbitrarily closely approximated by a special case of the PH distribution, a mixture of the degenerate distribution at zero and an Erlang distribution.

Condition (3) of Proposition 1 suggests a graphical framework for the problem of matching c and γ to a distribution. When $c - 1/c$ is plotted on the horizontal axis and γ is plotted on the vertical axis, the feasible second and third moments of the elements of $\mathcal{F}$ are represented by the half plane that lies on and above the line $\gamma = c - 1/c$. For any set of distributions $\mathcal{C}$, let $R(\mathcal{C}) = \{(c - 1/c, \gamma): (c - 1/c, \gamma) \text{ corresponds to a distribution in } \mathcal{C}\}$. Then $R(\mathcal{F}) = \bar{R}(\mathcal{F}) = \{(c - 1/c, \gamma): c > 0, \gamma \geq c - 1/c\}$, and $R(\mathcal{B}) = \bar{R}(\mathcal{B}) = \{(c - 1/c, \gamma): c > 0, \gamma = c - 1/c\}$.

5. FEASIBLE REGIONS OF MIXTURES OF ERLANG DISTRIBUTIONS ON THE $(c - 1/c, \gamma)$ PLANE

In this section, the region $R(\mathcal{ME}_n)$ is characterized. In Theorem 2, we specify $R(\mathcal{ME}_n)$. We also show that any point in this region, i.e., any point feasible for an ME_n, is feasible for an E_n or an $ME_n(2)$. Moreover, any point in the *interior* of this region is feasible for an $ME_n(2)$.

Theorem 2: $R(\mathcal{ME}_n)$ is the set of points in $R(\mathcal{F})$ for which (7) and (8) are either both strict or both tight. Also, $R(\mathcal{ME}_n) = R(\mathcal{ME}_n(1)) \cup R(\mathcal{ME}_n(2))$, and $R^{\circ}(\mathcal{ME}_n) = R(\mathcal{ME}_n(2))$.

$$c \geq 1/\sqrt{n} \tag{7}$$

$$\gamma \geq \frac{1}{1+n}\left(\frac{1}{c^3} + (1-n)\frac{1}{c} + (2+n)\,c\right) \tag{8}$$

Proof: See Appendix. □

Figure 1 shows $R(\mathcal{ME}_n)$ for several values of n. Notice that for $n = 2, 3, \cdots$, $R(\mathcal{ME}_{n-1}) \subset R(\mathcal{ME}_n)$. By Theorem 1, $R^o(\mathcal{F}) = R^o(\mathcal{ME})$. As $n \to \infty$, the right side of (7) converges to zero and the right side of (8) converges to $c - 1/c$. Thus, Theorem 2 implies $R^o(\mathcal{ME}) = R(\mathcal{ME})$. Since points on the Bernoulli line are not feasible for any PH distribution, the feasible region for PH distributions is $R(\mathcal{ME}) = \{(c-1/c, \gamma)\colon\ \gamma > c - 1/c\}$.

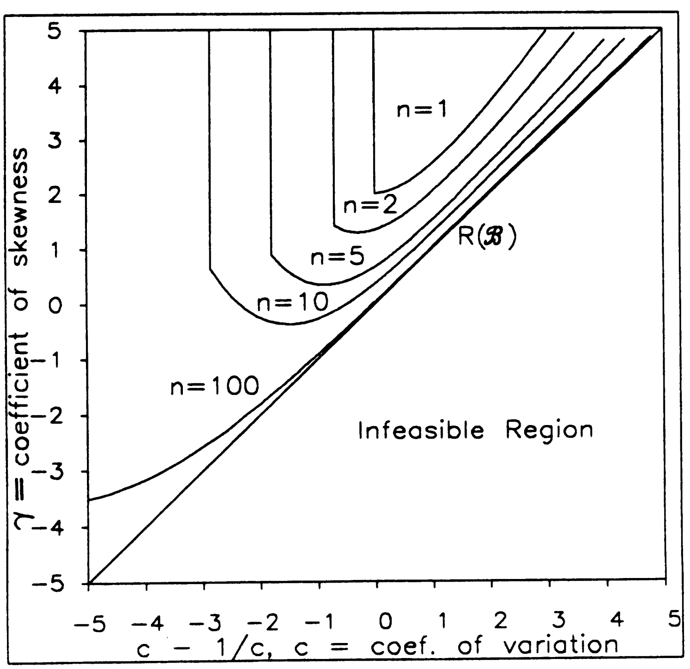

FIG. 1
$R(\mathcal{ME}_n)$ for $n = 1, 2, 5, 10, 100$.

Propositions 2 and 3 provide simple interpretations for part of the boundary of $R(\mathcal{ME}_n)$. Let DE_n denote a mixture of the degenerate distribution at zero and an E_n, and let $DE_n(p)$ denote a DE_n with mixing probability p assigned to the E_n. Proposition 2 shows that each point on the curve that defines the lower bound on γ in $R(\mathcal{ME}_n)$ corresponds to a DE_n. Proposition 3 shows that the n-stage Erlang distribution corresponds to the point at which the curves for the lower bounds on c and γ intersect. No simple interpretation for the lower bound on $c-1/c$ is apparent.

Proposition 2: Each point $(c-1/c, \gamma)$ for which (8) is tight corresponds to a DE_n. Moreover, for $p \in (0, 1]$ and $c \geq 1/\sqrt{n}$, there is a one-to-one correspondence between parameter p and the points for which (8) is tight.

Proof: Let λ be the rate parameter of the mixed E_n. Let μ_t be the tth noncentral moment of the mixture, and suppose the tth noncentral moment of the mixed E_n is η_t. Then, $\eta_t = P_{n,t}\lambda^{-t}$ (see equation (1)) and $\mu_t = p\,\eta_t$. The result follows from straightforward algebra. □

Proposition 3: An E_n has second and third standardized moments $c = 1/\sqrt{n}$ and $\gamma = 2/\sqrt{n}$. Inequalities (7) and (8) are tight for these moments.

Proof: The standardized moments of an E_n follow from equation (1). That (8) and (9) are tight for these moments is easily verified. (Also, this is the $p = 1$ case of Proposition 2.) □

By Theorem 1 and Proposition 1, every point in $R^o(\mathcal{F})$ can be matched to a mixture of two Erlang distributions of (sufficiently high) common order. More specifically, for any point $(c-1/c, \gamma)$ in $R^o(\mathcal{F})$, there exists n^* such that $(c-1/c, \gamma) \in R^o(\mathcal{ME}_n)$ for all $n \geq n^*$. Proposition 4 provides an expression for n^*.

Proposition 4: Let $(c-1/c, \gamma)$ be a point in $R^o(\mathcal{F})$. The minimum n^* such that $(c-1/c, \gamma) \in R^o(\mathcal{ME}_n)$ for all $n \geq n^*$ is the smallest integer that satisfies (9) and (10).

$$n^* > 1/c^2 \tag{9}$$

$$n^* > \frac{-\gamma + 1/c^3 + 1/c + 2c}{\gamma - (c - 1/c)} \tag{10}$$

Proof: The lower bounds on n^* are obtained by solving (7) and (8) for n. The result then follows from Theorem 2. □

6. MOMENT-MATCHING FORMULAS FOR MIXTURES OF TWO ERLANG DISTRIBUTIONS

For each point $\vec{\mu}_3$ that corresponds to a point in $R^o(\mathcal{ME}_n)$, the $ME_n(2)$ that exactly matches $\vec{\mu}_3$ is unique. Theorem 3 expresses the parameters of the $ME_n(2)$ in terms of μ_1, μ_2, and μ_3. The second and third noncentral moments are expressed in terms of the variables $y = \mu_2 - ((n+1)/n)\,\mu_1^2$ and $x = \mu_1\,\mu_3 - ((n+2)/(n+1))\,\mu_2^2$. By Lemma 1, $\vec{\mu}_3 \in \bar{M}_3(\mathcal{ME}_n)$ if and only if $y \geq 0$ and $x \geq 0$. (See inequalities (A8) and (A9) in the appendix.)

The proof of Theorem 3 is a generalization of the derivation in Abate and Whitt [1, Section 5.1] of the parameters of a two-stage hyperexponential distribution in terms of its first three noncentral moments. The steps of this proof follow corresponding steps of the proof given by Abate and Whitt. An alternative means of deriving the parameters of the $ME_n(2)$ that matches $\vec{\mu}_3$ is to note that the distribution of Λ^{-1}, as defined in Lemma 2, is the three-moment lower principle representation given in Section 4.1 of Johnson and Taaffe [9].

Theorem 3: Let μ_1, μ_2, μ_3 be the first three noncentral moments of F, a mixture of two Erlang cdf's, F_1 and F_2. The parameters of F are λ_1, λ_2, and p_1, where λ_1 and λ_2 are the rate parameters of F_1 and F_2, respectively, and p_1 is the mixing parameter assigned to F_1. (The mixing probability p_2 assigned to F_2 is $1-p_1$.) The parameters of F can be expressed as follows.

$$\lambda_i^{-1} = \left(-B \pm \sqrt{B^2 - 4AC}\right)/(2A) \tag{11}$$

and

$$p_1 = \left(\frac{\mu_1}{n} - \lambda_2^{-1}\right)/(\lambda_1^{-1} - \lambda_2^{-1}), \tag{12}$$

where

$$A = n(n+2)\mu_1\, y, \tag{13}$$

$$B = -\left(n\,x + \frac{n(n+2)}{n+1}\, y^2 + (n+2)\,\mu_1^2\, y\right), \tag{14}$$

$$C = \mu_1 x, \tag{15}$$

$$y = \mu_2 - \left(\frac{n+1}{n}\right)\mu_1^2, \tag{16}$$

and

$$x = \mu_1 \mu_3 - \left(\frac{n+2}{n+1}\right)\mu_2^2. \tag{17}$$

Proof: See Appendix. □

7. IMPLEMENTATION ISSUES

The results in Sections 5 and 6 suggest the following algorithm for matching a triple of first three moments, $\vec{\mu}_3$, in $M_3^o(\mathcal{F})$, which is the set feasible for PH distributions, to a mixture of two Erlang distributions of common order. First, use Proposition 4 to determine the minimum value, n^*, such that the moments are feasible for an $ME_{n^*}(2)$. (We ignore the possibility that for some value of n $\vec{\mu}_3$ may be feasible for an E_n, which corresponds to a single point in $\delta(R(\mathcal{ME}_n))$.) Then, use Theorem 3 to compute the parameters of the $ME_{n^*}(2)$ that matches $\vec{\mu}_3$. Call this procedure Algorithm A.

The simplicity of Algorithm A does not guarantee its algorithmic usefulness. Though the algorithm is straightforward, the acceptability of the properties of the selected distributions must be considered. Properties that may be important for the distributions selected by a particular moment-matching algorithm include dimension, numerical stability, kurtosis (fourth standardized moment), structure of the representation $(\vec{\alpha}, \mathbf{T})$, and density-function shape. For example, algorithm builders should determine when Algorithm A selects numerically unstable distributions and modify Algorithm A accordingly. Similarly, users should know what regions of the $(c-1/c, \gamma)$ plane correspond to selected distributions of unacceptable dimension. In the following subsections, we discuss the properties listed above for Algorithm A and consider modifications to Algorithm A. Highlights of our observations and conclusions are italicized.

7.1 Dimension. In many applications of PH distributions, low dimension is an important property for the selected PH distribution. For a given point $(c-1/c, \gamma)$, Algorithm A minimizes the dimension over the set of mixtures of Erlang distributions of common order but, in general, not over the set of PH distributions.

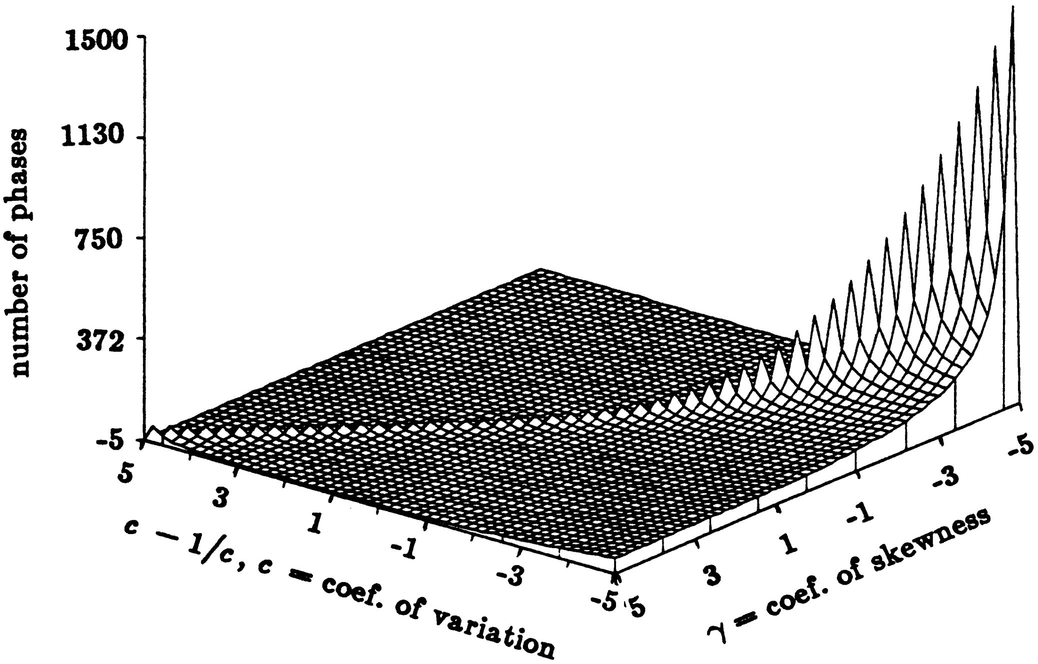

FIG. 2
Dimension of distributions selected by Algorithm A.

Figures 2 and 3 show the dimension of the distribution selected by Algorithm A as a function of $(c-1/c, \gamma)$. In these figures, the point nearest the viewer is (−5, 5), the upper, left corner of Figure 1. Figure 2 shows the actual dimension, $2n^*$. In Figure 3, all dimensions above thirty are set equal to thirty so that the behavior of the lower portion of the surface becomes apparent. Since $R^o(\mathcal{ME})$ does not include its boundary, the lower bounds on $c-1/c$ and γ in $R^o(\mathcal{F})$ and $R^o(\mathcal{ME}_n)$, $n = 1, 2, \cdots$, were implemented as .01 inside the theoretical boundaries.

Figure 2 indicates that precise matching of the point $(c-1/c, \gamma)$ *via Algorithm A requires a PH distribution of unacceptably high dimension when* $c-1/c$ *is too low or* γ *is too close to* $c-1/c$. *(The value of the maximum acceptable dimension is application dependent.) Approaches to reducing the dimension of the selected distribution include matching two moments instead of three, adjusting the moments to be matched to reduce the dimension of the selected distribution, and using alternative three-moment matching methods.*

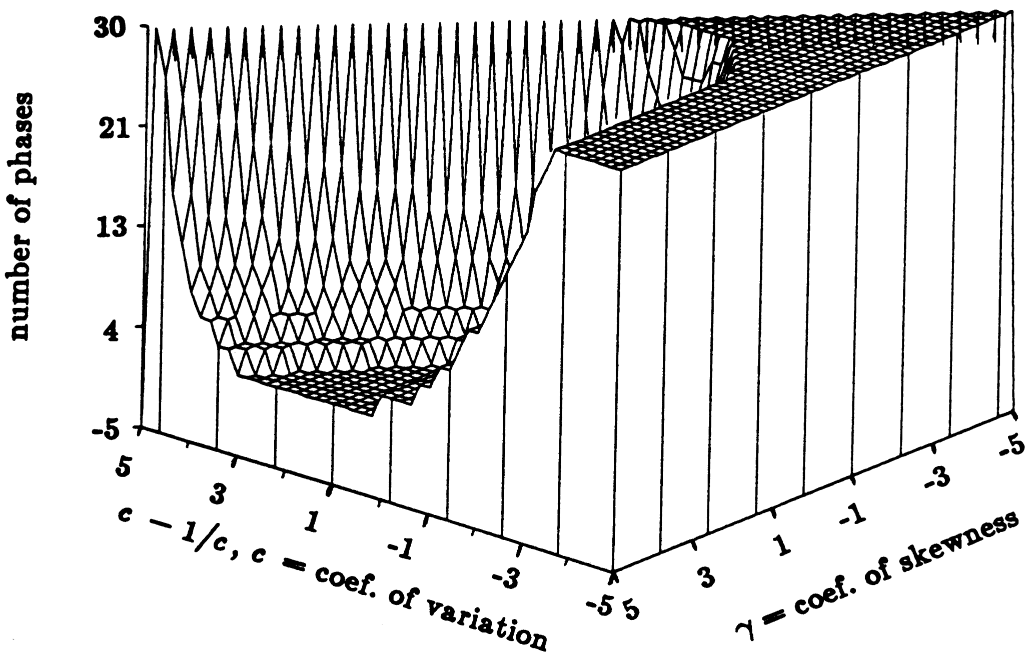

FIG. 3
Min(30, dimension of distribution selected by Algorithm A).

Matching only two moments can sometimes be appropriate. For example, two-moment queueing approximations are sometimes appropriate, particularly when the approximated distribution has $c < 1$. Matching two moments to an Erlang distribution (or some slight modification of an Erlang distribution) requires about half the dimension required by Algorithm A. Even this solution may yield an unacceptably high dimension for very low c. However, Aldous and Shepp [2] have shown that when $c < 1$, selecting an Erlang distribution minimizes dimension over all PH distributions that match the first two moments.

For points near the Bernoulli ($\gamma = c - 1/c$) line, dimension can be reduced substantially by imposing a small increase in γ. For example, at $c - 1/c = 3$, for $n = 100$, 20, and 10, the lower bounds on γ in $R(\mathcal{ME}_n)$ are 3.0390, 3.1874, and 3.3578, respectively. This suggests that adjusting γ to restrict the dimension of the selected distribution may not induce serious approximation errors.

In Johnson and Taaffe [6], we have demonstrated the use of nonlinear programming (NLP) methods to match moments to distributions of smaller dimension than feasible for mixtures of Erlang distributions of common order. The trade-offs between using a straightforward, efficient analytic method and a more computationally expensive method which may require user intervention are application dependent.

7.2 Structure of distribution representation. *Mixtures of Erlang distributions have representations with special structure that may be exploited in applications, possibly mitigating the effects of high dimension.* A mixture of two Erlang distributions (with or without common order) has a PH distribution representation that consists of a vector, $\vec{\alpha}$, with only two nonzero entries and an upper bidiagonal matrix $\mathbf{T}$. Further, absorption can occur from only two of the transient phases. See Neuts [14, pp. 127ff] and Ramaswami and Latouche [15] for examples where such properties have been exploited. Also see Johnson and Taaffe [6] for examples of moment-matching methods that select distributions with special structure *and* lower dimension than obtained from Algorithm A.

7.3 Numerical Stability. Numerical stability of selected distributions is another important implementation issue. Let $r = \lambda_2/\lambda_1$ and $p = p_1$, where p_1, λ_1, and λ_2 are defined as in Theorem 3. Without loss of generality, assume that $\lambda_2 > \lambda_1$. For fixed n, r and p can be considered functions of $(c - 1/c,\ \gamma)$. Further, the values of r and p can be used as measures of the numerical stability of selected distributions. Computation of very high values of r and values of p very close to zero or one require high machine precision to obtain meaningful results. Further, extreme values of r are likely to lead to ill-conditioned matrices and stiff sets of differential equations. Also, sharp changes in r and p on the $(c - 1/c,\ \gamma)$ plane indicate extreme sensitivity of the parameters to the precise location of the point $(c - 1/c,\ \gamma)$.

The behavior of the parameters of the distributions selected by Algorithm A is most easily understood by first studying, for fixed n, the parameters of $ME_n(2)$'s as a function of $(c - 1/c,\ \gamma)$. For example, Figures 4 and 5 show constant p contours and constant r contours, respectively, in $R(\mathcal{ME}_{10})$. These figures show that near the $c - 1/c = 1/\sqrt{n} - \sqrt{n}$ (i.e., $c = 1/\sqrt{n}$) boundary, p is very low and r is high, and near the DE_n curve (i.e., the γ boundary) r is high. This poor behavior near the boundaries is especially serious for Algorithm A, since in Algorithm A, points matched to an $ME_n(2)$ for $n \geq 2$ are near a boundary of $R(\mathcal{ME}_n)$. The following analytic result shows that r becomes arbitrarily large as $(c - 1/c,\ \gamma)$ moves closer to the DE_n curve.

Proposition 5: Let $ME_n(2)$ correspond to the point $(c - 1/c,\ \gamma)$, and let r be as defined above. As

$$\gamma \longrightarrow \frac{1}{1+n}\left(\frac{1}{c^3} + (1-n)\frac{1}{c} + (2+n)\,c\right), \tag{18}$$

the lower bound on γ in $R(\mathcal{ME}_n)$, $r \longrightarrow \infty$.

Proof: If the noncentral moments of the $ME_n(2)$ are denoted by μ_t, $t = 1, 2 \cdots$, then the convergence in (18) is equivalent to $\mu_1 \mu_3 - (n+2)\,\mu_2^2/(n+1) \longrightarrow 0$. (See (A9) in the appendix.) Since standardized moments are independent of the mean, we can assume without loss of generality that μ_1 is fixed. Then in terms of the parameters of an $ME_n(2)$ as defined in Theorem 3, $x \longrightarrow 0$, and y is fixed. Hence, $C \longrightarrow 0$, A remains fixed, and

$$B \longrightarrow -\left(\frac{n(n+2)}{n+1}\, y^2 + (n+2)\,\mu_1^2\, y\right). \tag{19}$$

Thus, $\lambda_i^{-1} \longrightarrow (-B \pm \sqrt{B^2})/(2A)$, where B equals the right side

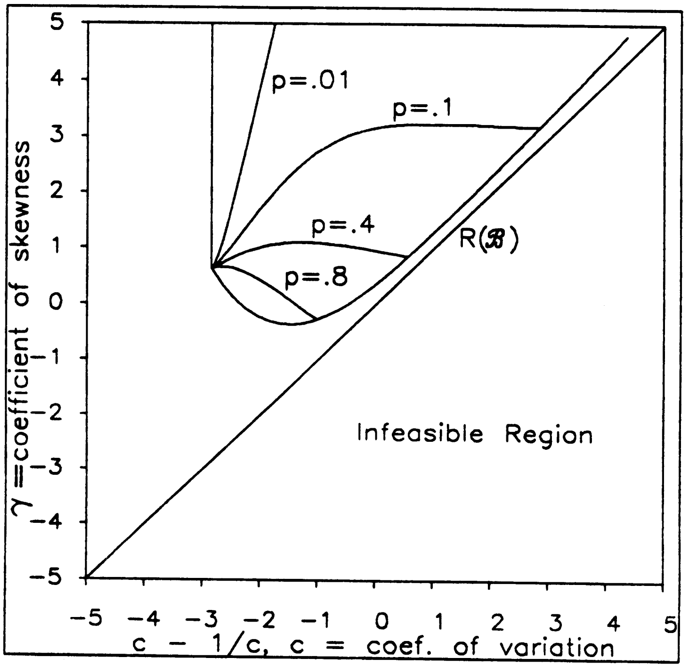

FIG. 4
Constant p contours for $ME_{10}(2)$'s.

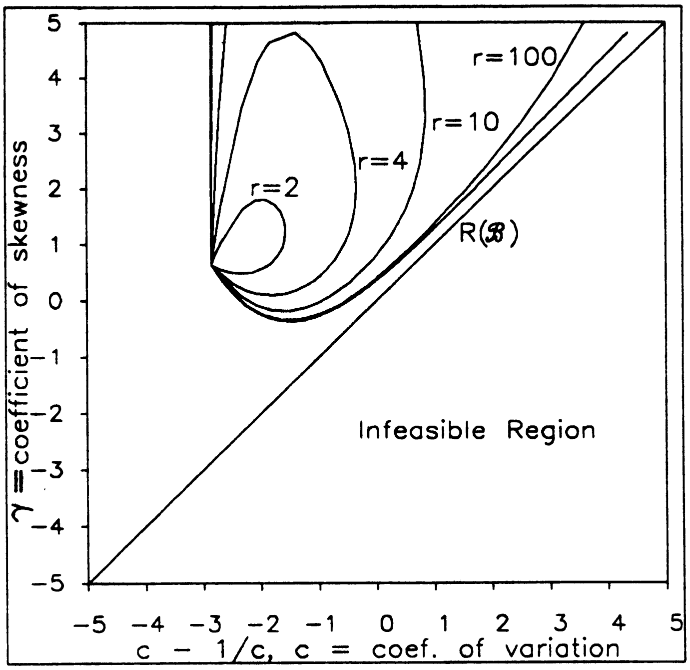

FIG. 5
Constant r contours for $ME_{10}(2)$'s.

of (19). So, $\lambda_1^{-1} \rightarrow -B/A$, $\lambda_2^{-1} \rightarrow 0$, and $p \rightarrow -\mu_1 A(nB)^{-1}$, where A is as defined in (13) and B equals the right side of (19). □

Because of the extreme behavior of the parameters of $ME_n(2)$'s near the boundaries of $R(\mathcal{ME}_n)$, any implementation of Algorithm A should set implemented boundaries of $R(\mathcal{ME}_n)$ inside the theoretical boundaries. One way to do this is to match points near a boundary of $R(\mathcal{ME}_n)$ to an $ME_{n'}(2)$, where $n' > n$. We have applied this idea to Algorithm A and have observed that small increases in dimension often result in noticeable improvement in r and p. Let n^* be the minimum order feasible for $(c-1/c, \gamma)$, i.e., the order selected by Algorithm A. Tables 1 and 2 illustrate the improvements achieved by incrementing, for all points on the $(c-1/c, \gamma)$ plane, the order of the mixed distributions. In each table, for points which correspond to several values of n^*, values of r and p are given for $n = n^*$, $n = n_1 \equiv n^* + 1$, and $n = n_2 \equiv \lceil 1.1 \rceil n^*$.

TABLE 1

Effect of small increase in dimension on r and p at $c-1/c=2.00$ and $c-1/c=4.00$.

	$c-1/c=2.00$			$c-1/c=4.00$		
n	γ	r	p	γ	r	p
$n^*=2$		36.2	.1717		118.0	.0624
$n_1=n_2=3$	$\gamma_1=3.657$	30.1	.1328	$\gamma_1=6.361$	95.5	.0486
$n^*=2$		2897.4	.2191		11805.2	.0790
$n_1=n_2=3$	$\gamma_2=3.110$	69.3	.1707	$\gamma_2=5.580$	222.5	.0617
$n^*=5$		173.7	.1660		547.6	.0599
$n_1=n_2=6$	$\gamma_1=2.663$	107.1	.1548	$\gamma_1=4.944$	336.1	.0558
$n^*=5$		2359.7	.1750		26245.0	.0633
$n_1=n_2=6$	$\gamma_2=2.560$	222.1	.1632	$\gamma_2=4.790$	742.6	.0590
$n^*=10$		635.4	.1584		1979.1	.0571
$n_1=n_2=11$	$\gamma_1=2.331$	352.5	.1549	$\gamma_1=4.472$	1096.7	.0559
$n^*=10$		2161.0	.1603		106393.8	.0580
$n_1=n_2=11$	$\gamma_2=2.310$	568.2	.1568	$\gamma_2=4.430$	2322.8	.0568
$n^*=20$		2431.1	.1531		7531.1	.0552
$n_1=21$	$\gamma_1=2.166$	1279.8	.1521	$\gamma_1=4.236$	3962.9	.0548
$n_2=22$		896.0	.1512		2773.4	.0545
$n^*=20$		8677.9	.1536		16327.2	.0553
$n_1=21$	$\gamma_2=2.160$	2049.7	.1526	$\gamma_2=4.230$	5512.2	.0550
$n_2=22$		1212.1	.1517		3445.9	.0546
$n^*=100$		58283.9	.1479		182386.6	.0533
$n_1=101$	$\gamma_1=2.033$	29558.2	.1478	$\gamma_1=4.047$	91805.3	.0533
$n_3=110$		5878.4	.1475		18144.6	.0532

TABLE 2

Effect of small increase in dimension on r and p at $\gamma = 2.00$ and $\gamma = 4.00$.

		$\gamma = 2.00$		$\gamma = 4.00$	
n	$c - 1/c$	r	p	r	p
$n^* = 2$		3.73	.500e00	3.73	.670e-01
$n_1 = n_2 = 3$	$c_1 - 1/c_1 = 0.000$	4.16	.364e00	4.30	.794e-01
$n^* = 2$		13.29	.360e-04	57.8	.168e-05
$n_1 = n_2 = 3$	$c_2 - 1/c_2 = -0.690$	2.10	.207e00	4.24	.139e-01
$n^* = 5$		2.67	.161e-01	6.20	.157e-02
$n_1 = n_2 = 6$	$c_1 - 1/c_1 = -1.500$	2.22	.590e-01	4.37	.663e-02
$n^* = 5$		26.00	.378-05	72.00	.469e-06
$n_1 = n_2 = 6$	$c_2 - 1/c_2 = -1.770$	2.95	.849e-02	6.72	.960e-03
$n^* = 10$		4.57	.797e-03	10.12	.122e-03
$n_1 = n_2 = 11$	$c_1 - 1/c_1 = -2.667$	2.99	.481e-02	6.08	.724e-03
$n^* = 10$		40.09	.553e-06	97.63	.904e-07
$n_1 = n_2 = 11$	$c_2 - 1/c_2 = -2.830$	4.57	.724e-03	9.99	.114e-03
$n^* = 20$		7.35	.623e-04	15.69	.116e-04
$n_1 = 21$	$c_1 - 1/c_1 = -4.130$	4.32	.436e-03	8.72	.805e-04
$n_2 = 22$		3.32	.129e-02	6.40	.236e-03
$n^* = 20$		41.102	.235e-06	92.93	.447e-07
$n_1 = 21$	$c_2 - 1/c_2 = -4.230$	6.69	.820e-04	14.13	.154e-04
$n_2 = 22$		4.19	.467e-03	8.38	.868e-04
$n^* = 100$		18.70	.319e-06	38.40	.715e-07
$n_1 = 101$	$c_1 - 1/c_1 = -9.849$	9.93	.248e-05	19.89	.555e-06
$n_3 = 110$		2.76	.323e-03	4.74	.718e-04
$n^* = 100$		45.69	.197e-07	95.39	.441e-08
$n_1 = 101$	$c_1 - 1/c_1 = -9.880$	13.78	.841e-06	28.00	.188e-06
$n_3 = 110$		2.76	.323e-03	4.94	.605e-04

Table 1 shows r and p at $c-1/c=2$ and 4, with γ chosen to demcnstrate the range of improvement in r that is attainable. For each case in Table 1, r and p are computed for $\gamma=\gamma_1$ and γ_2, where γ_1 is on the DE_{n^*-1} curve, and γ_2 is equal to the value of γ on the DE_{n^*} curve, rounded up to the nearest 100th. (Exception: For $n^*=100$, γ_2 is greater than γ_1, so r and p are not given for γ_2 in this case.) The change in r at γ_1 is the smallest improvement that can be attained by mixing Erlang distributions of order n_1 or n_2 instead of n^*. The change in r at γ_2 is representative of the improvement attained when $(c-1/c, \gamma)$ is very close to DE_{n^*}.

Inspection of Table 1 leads to the following conclusions. First, increasing n improves r much more when γ is very close to its lower bound in $R(\mathcal{ME}_{n^*})$ than when it is farther away. Second, a small increase in n induces only a small change (always a decrease) in p. Hence, the improvement in the numerical stability of r obtained via an increase in dimension does not induce numerical instability in p. Third, although the small increase in n often induces a substantial improvement in r, in many cases, r is still very large. Thus, additional increments in n may be warranted. As $n \rightarrow \infty$, the selected distribution converges to the discrete two-point distribution that matches the point. So, the limiting values of r and p correspond to this distribution. However, note that for points near the Bernoulli line, high values of r are inevitable, since the selected distributions reflect the behavior of Bernoulli distributions.

Table 2 shows r and p at $\gamma=2$ and 4, with $c-1/c$ chosen to demonstrate the range of improvement in p that is attainable. For each case in Table 2, r and p are computed for $c-1/c=c_1-1/c_1$ and $c-1/c=c_2-1/c_2$, which are defined as follows. The value c_1-1/c_1 is set to $1/(n^*-1)^{1/2}-(n^*-1)^{1/2}$, the lower bound on $c-1/c$ in $R(\mathcal{ME}_{n^*-1})$. The value c_2-1/c_2 is set to $(1/n^{*1/2}-n^{*1/2})+.01$ so that it is .01 inside the lower bound on $c-1/c$ in $R(\mathcal{ME}_{n^*})$. The change in p at c_1-1/c_1 is the minimum attainable improvement. The change at c_2-1/c_2 is representative of the improvement attained for a point very close to the boundary of the feasible region.

The following conclusions are based on Table 2. First, the parameter r is much better behaved near the $c-1/c$ boundaries than near the γ boundaries. So, although increasing n decreases r, these improvements are minor compared to those observed in Table 1. Second, as expected, the parameter p is very small when $n=n^*$ in Table 2. When n is increased to n_1 or n_2, p sometimes increases by orders of magnitude, but often remains very small even with this increase. For the points in Table 2, we have also observed improvements induced by higher values of n, up to $n=\lceil 5n^*\rceil$. At $n=\lceil 5n^*\rceil$, p is at least .03 when $\gamma=2$ and at least .10 when $\gamma=4$. Thus, these results imply that extremely low values of p can be avoided, but often at the cost of a substantial increase in dimension of the selected distribution.

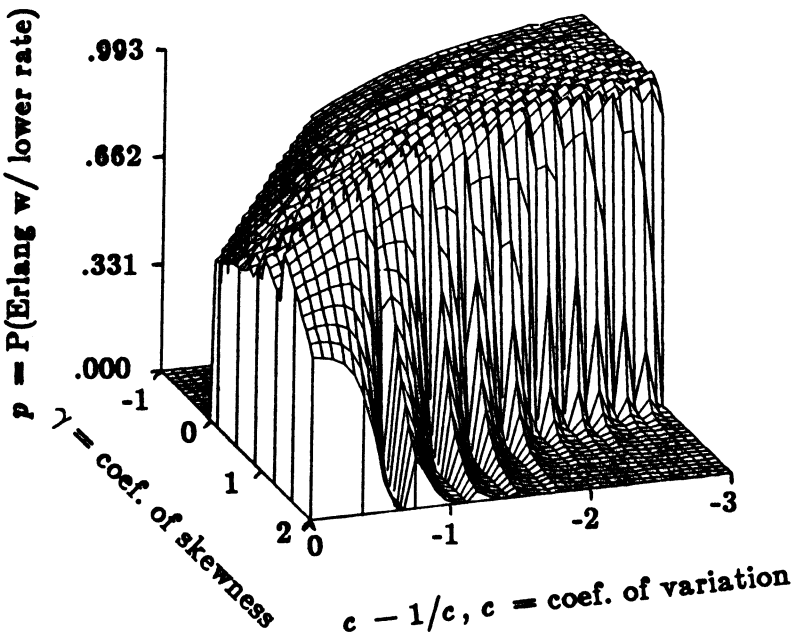

FIG. 6.
Parameter p over low $c - 1/c$, high γ region when two Erlang distributions of minimum feasible order are mixed.

Figures 6 and 7 illustrate graphically the improvement in p attained when $n = n^* + 1$ instead of n^*. In these figures, the value of p is plotted over the region $c - 1/c \in [-3, 0]$ and $\gamma \in [-1, 2]$, a subregion of the plane shown in Figure 1. The point nearest the viewer is (0, 2). In Figure 6, n is set to n^*, the minimum feasible order; in Figure 7, n is set to $n^* + 1$. Although in Figure 7 there are still many points that map to very low values of p, some improvement is noticeable. The slope down to the low values of p is not as steep, so the parameter p is less sensitive to small changes in γ. Moreover, the high values of p are not as extreme, and some near-zero values are raised visibly above zero. The improvement is much more pronounced for $c - 1/c$ near zero than for lower values of $c - 1/c$. Thus, as $c - 1/c$ decreases (and n^* increases), noticeable improvement requires a greater increment in n.

A close look at the top of Table 2 reveals that for $n^* = 2$, at $c - 1/c = 0$, the increase in n actually resulted in a *lower* value of p at $\gamma = 2$ and a higher value of r at $\gamma = 2$ and 3. This is in contrast to the

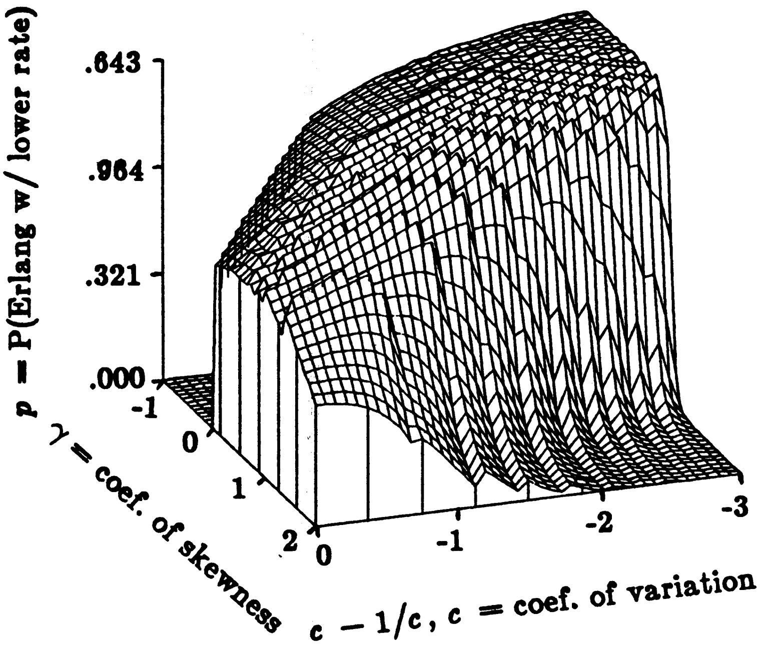

FIG. 7
Parameter p over low $c - 1/c$, high γ region when order of Erlang distributions mixed is one greater than minimum feasible order.

above generalizations. This exception occurred because the $c - 1/c = 0$ line is not close enough to the boundary of $R(\mathcal{ME}_2)$ to demonstrate boundary behavior. This and other examples that we have observed demonstrate that a trade-off between stability and dimension does *not* exist for points that are not near a $R(\mathcal{ME}_n)$ boundary. Conversely, the change in r and p for such points has negligible adverse effect on numerical stability.

We summarize our observations on numerical stability as follows. *The parameters of the $ME_n(2)$'s that correspond to points near the boundary of $R(\mathcal{ME}_n)$ often lead to numerically unstable distributions. For such points, mixing Erlang distributions of order one greater than the minimum feasible order often improves stability substantially. However, additional increments in the order of the mixed Erlang distributions may be needed to obtain adequate stability. Points near the Bernoulli line inevitably correspond to mixtures with high values of r. Finally, incrementing the order of the mixed Erlang distributions for points not near the boundary of a feasible region does not necessarily improve numerical stability of the selected distribution.*

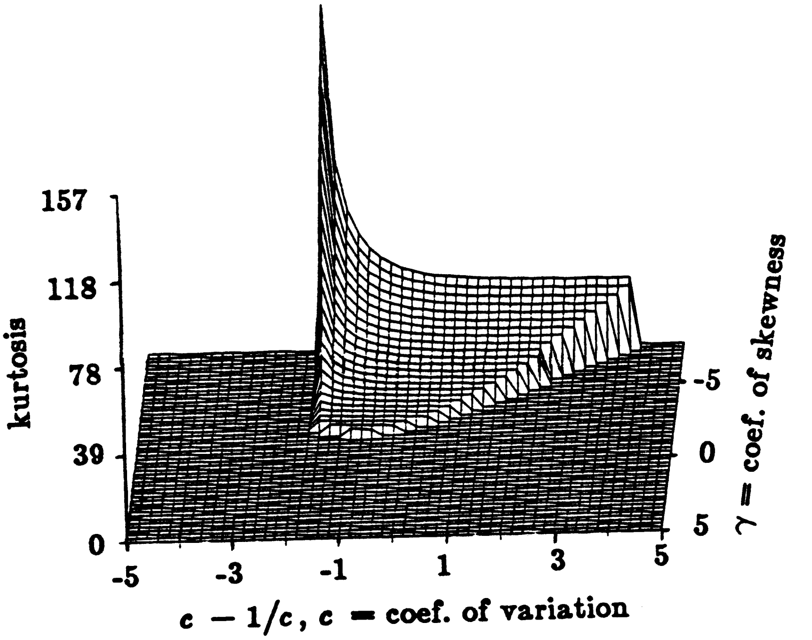

FIG. 8
Kurtosis surface over $R(\mathcal{ME}_5)$, when each point in $R(\mathcal{ME}_5)$ is matched to an $ME_5(2)$.

7.4 Kurtosis surface. When matching three moments, the behavior of the kurtosis of the selected distribution may be important for some applications. Even if the specific value of kurtosis is not important, an application may be sensitive to sudden or catastrophic changes in the height of the kurtosis surface over the $(c-1/c,\ \gamma)$ plane.

To investigate this aspect of selecting $ME_n(2)$'s, the kurtosis surface over the $(c-1/c,\ \gamma)$ plane has been plotted for several cases. Figure 8 shows the kurtosis surface over $R(\mathcal{ME}_5)$ when all points in the region are matched to an $ME_5(2)$. The implemented bounds of $R(\mathcal{ME}_5)$ are .01 inside the theoretical bounds, and kurtosis is set to zero outside $R(\mathcal{ME}_5)$. This surface shows that the kurtosis is almost flat over most of $R(\mathcal{ME}_5)$, but near the left edge of the region, kurtosis increases drastically. While no analytic results have been obtained, we have observed that kurtosis gets arbitrarily high as points closer to the left edge are plotted. Figure 9

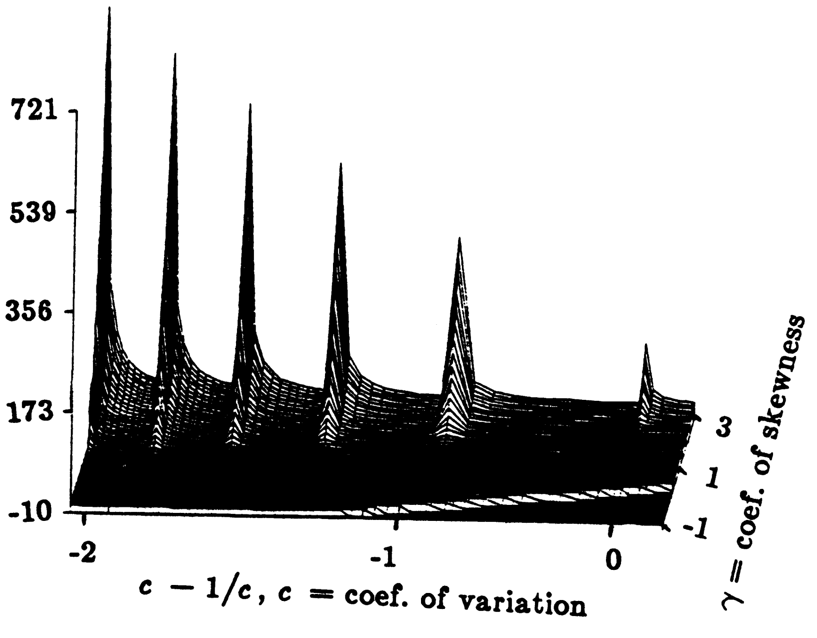

FIG. 9
Kurtosis over low $c-1/c$, high γ region when two Erlang distributions of minimum feasible order are mixed.

shows the kurtosis surface in the low $c-1/c$, high γ region of the $(c-1/c, \gamma)$ plane when Algorithm A is implemented with implemented bounds .01 inside theoretical bounds. For applications sensitive to kurtosis, the sharp changes shown in Figure 9 are undesirable. Figure 10 shows the kurtosis surface when Erlang distributions of order n^*+1 are mixed. ***The large smoothing effect is another indication in some regions that a small increment in n tends to improve significantly the properties of the selected distributions.***

7.5 Density-function shape. Density-function shapes of mixtures of Erlang distributions of common order are studied in detail in Johnson and Taaffe [7]. Important results in that paper include the following. ***First, shape properties of Erlang distributions are not necessarily preserved under finite mixing.*** In particular, bimodality is likely and is especially intuitive for mixtures of Erlang distributions of high order, since as order increases, the limiting distribution is a discrete two-point distribution. ***Second, use of the minimum feasible order, n^*, sometimes results in density functions with very sharp spikes in them; mixing two Erlang distributions of order n^* smooths the density function considerably.***

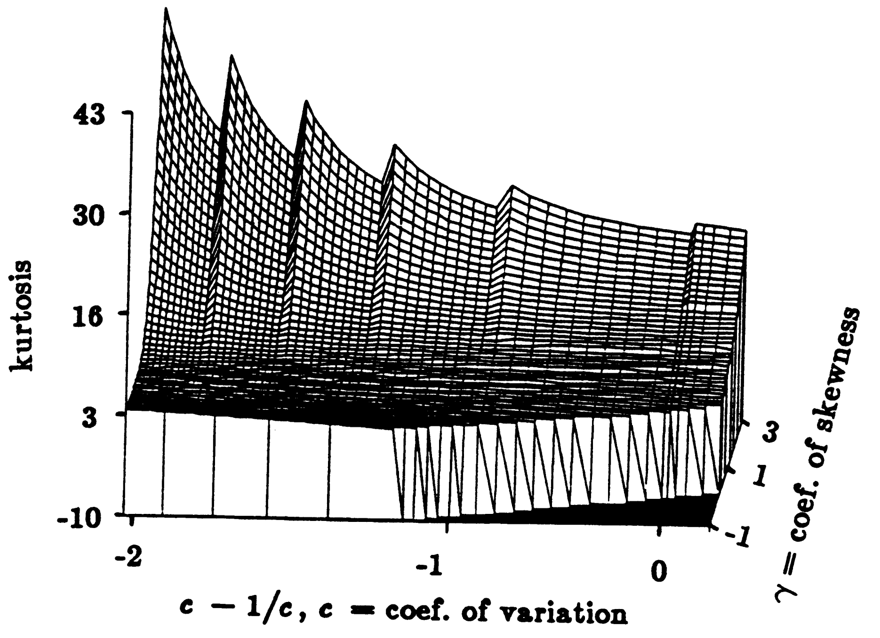

FIG. 10
Kurtosis over low $c - 1/c$, high γ region when order of Erlang distributions mixed is one greater than minimum feasible order.

Finally, the NLP methods of Johnson and Taaffe [6] provide density-function shapes not feasible for mixtures of two Erlang distributions of common order.

8. SUMMARY AND CONCLUSIONS

In this paper, we show that mixtures of Erlang distributions of common order is a convenient family for matching moments to PH distributions. We provide detailed analytic results for matching three moments to mixtures of two Erlang distributions of common order. The moment-matching results presented here have the advantage that they accommodate all triples of first three moments that can be accommodated by PH distributions. This set of feasible triples is considerably larger than those accommodated by previously available three-moment-matching results. Methods for implementing our three-moment-matching results are suggested and evaluated in this paper. Although the theoretical results of Sections 5 and 6 are straightforward, judicious use of these results will take into account the implementation issues discussed in Section 7.

We conclude with a few words of caution about moment matching in general. First, while moment matching is appropriate for many applications, it is not adequate for all applications. For example, some reliability applications may require that the selected distribution have monotone failure rate. This property is not necessarily obtained if only moments are matched. Second, the user should determine an appropriate number of moments to match. We have cited evidence indicating two- and three-moment approximations are appropriate for some queueing approximations. These conclusions are not intended to be valid for all applications. Finally, if moments are estimated from data, the statistical properties of the estimators should be considered. One issue is the effect of the poor statistical properties of higher-order moments on the parameters of the selected distribution. We have not yet examined the statistical issues associated with the methods suggested in this paper.

REFERENCES

[1] J. Abate and W. Whitt, "Transient Behavior of Regulated Brownian Motion, I: Starting at the Origin," *Advances in Applied Probability,* Vol. 19, 1987, 560-598.

[2] D. Aldous and L. Shepp, "The Least Variable Phase Type Distribution is Erlang," *Commun. Statist.—Stochastic Models,* Vol. 3, No. 3, 1987, 467-473.

[3] T. Altiok, "On the Phase-Type Approximations of General Distributions," *IIE Transactions,* Vol. 17, No. 2, 1985, 110-116.

[4] W. Bux, and U. Herzog, "The Phase Concept: Approximation of Measured Data and Performance Analysis," *Computer Performance,* Chandy, K. M., and Reiser, M., Editors, North Holland, 1977.

[5] A. Cumani, "On the Canonical Representation of Homogeneous Markov Processes Modelling Failure-Time Distributions," *Microelectronics and Reliability,* Vol. 22, 1982, 583-602.

[6] M. A. Johnson, and M. R., Taaffe, "Matching Moments to Phase Distributions: Nonlinear Programming Approaches," Research Memorandum No. 88-14, School of Industrial Engineering, Purdue University, 1988.

[7] M. A. Johnson, and M. R., Taaffe, "Matching Moments to Phase Distributions: Density Function Shapes," Research Memorandum No. 88-11, School of Industrial Engineering, Purdue University, 1988.

[8] M. A. Johnson, and M. R., Taaffe, "The Denseness of Phase Distributions," Research Memorandum No. 88-20, School of Industrial Engineering, Purdue University, 1988.

[9] M. A. Johnson, and M. R., Taaffe, "Tchebycheff Systems and Moment Spaces in a Probability Context," Research Memorandum No. 88-21, School of Industrial Engineering, Purdue University, 1988.

[10] M. A. Johnson, and M. R., Taaffe, "Error Bounds for Queueing Performance Measure Approximations Based on Moment Matching," Research Memorandum No. 88-22, School of Industrial Engineering, Purdue University, 1988.

[11] S. Karlin and W. J. Studden, *Tchebycheff Systems: With Applications in Analysis and Statistics,* John Wiley & Sons, New York, 1966.

[12] J. G. Klincewicz, and W. Whitt, "On Approximations for Queues, II: Shape Constraints," *AT&T Bell Labs Tech Journal,* Vol. 63, No. 1, 139-161, January, 1984.

[13] B. Kolman, *Elementary Linear Algebra,* Second Edition, MacMillan Publishing Co., New York, 1977.

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

[15] V. Ramaswami and G. Latouche, "An Experimental Evaluation of the Matrix-Geometric Method for the $GI/PH/1$ Queue," *Commun. Statist.—Stochastic Models,* to appear.

[16] G. Strang, *Linear Algebra and Its Applications,* Academic Press, New York, 1976.

[17] W. Whitt, "Approximating a Point Process by a Renewal Process, I: Two Basic Methods," *Operations Research,* Vol. 30, No. 1, Jan.-Feb., 1982, 125-147.

[18] W. Whitt, "On Approximations for Queues, I: Extremal Distributions," *AT&T Bell Labs Technical Journal,* Vol. 63, No. 1, January, 1984, 115-138.

[19] W. Whitt, "On Approximations for Queues, III: Mixtures of Exponential Distributions," *AT&T Bell Labs Tech Journal,* Vol. 63, No. 1, January 1984, 163-175.

APPENDIX

PROOF OF LEMMA 3

This proof is based on results in Johnson and Taaffe [9]. First, assume $\vec{\mu}_k \in M_k^o(\mathcal{F})$. If k is odd, then let F be the lower principle representation of $\vec{\mu}_k$ as specified on page 12 of [9]. If k is even, then let F be the upper principle representation of $\vec{\mu}_{k+1}$ as specified on page 13 of [9].

Now suppose $\vec{\mu}_k \in \delta(M_k(\mathcal{F}))$. As stated in Section 3.2 of [9], the distribution that corresponds to a point in $\delta(M_k(\mathcal{F}))$ is unique. This is because any point in $M_k(\mathcal{F})$ is also in the set of moments feasible for distributions with support on $[0, b]$, for sufficiently large b, [9, Section 3.2]. Call this set $M_k(\mathcal{F}[0, b])$. Note that if $\vec{\mu}_k \in \delta(M_k(\mathcal{F}))$, then $\vec{\mu}_k \in \delta(M_k(\mathcal{F}[0, b]))$. Thus, by Theorem 1 of [9], $\vec{\mu}_k \in \delta(M_k(\mathcal{F}))$ implies that the distribution associated with $\vec{\mu}_k$ is unique. Theorem 1 of [9] also says that this distribution has mass at $\lfloor k/2 \rfloor + 1$ or fewer points. Note that one of these points may be zero. □

PROOF OF THEOREM 1

If $k = 1$, the $E_n(\lambda)$ with $\lambda = n/\mu_1$ has mean μ_1. So, Theorem 1 is true for $k = 1$.

Now let $k \geq 2$ and assume $\vec{\mu}_k \in M_k^o(\mathcal{F})$. For some integer $n \geq k$ and for $t = 0, 1, \cdots, k$, let $\theta_t = \mu_t / P_{n,t}$. (Recall that $P_{n,t} = (n+t-1)!/(n-1)!$.) Let $\vec{\theta}_k = (\theta_1, \theta_2, \cdots, \theta_k)$. Then, for $s = 1, 2, \cdots, \lfloor k/2 \rfloor$,

$$A(\vec{\mu}_{2s}) = \left[P_{n,\, i+j}\ \theta_{i+j} \right]_{i,j=0}^{s} \tag{A1}$$

by definition of $A(\vec{\mu}_{2s})$ and θ_t, and similarly for $s = 0, 1, \cdots, \lfloor (k-1)/2 \rfloor$,

$$B(\vec{\mu}_{2s+1}) = \left[P_{n,\, i+j+1}\ \theta_{i+j+1} \right]_{i,j=0}^{s}. \tag{A2}$$

For any $r \times r$ matrix, $C = \left[c_{i,j} \right]_{i,j=1}^{r}$, the determinant of C is defined as

$$\det(C) = \sum_{\mathcal{P}(r)} (-1)^{\sigma} \prod_{i=1}^{r} c_{i,j(i)},$$

where $\mathcal{P}(r)$ is the set of all permutations $(j(1), j(2), \cdots, j(r))$ of the r-tuple $(1, 2, \cdots,)$ and σ is the number of reversals in order in $(j(1), j(2), \cdots, j(r))$ relative to $(1, 2, \cdots,)$, Kolman [13, p. 191]. In the expressions below, let $\mathcal{P}$ denote $\mathcal{P}(s+1)$. Then,

$$\det(A(\vec{\mu}_{2s})) = \sum_{\mathcal{P}} (-1)^{\sigma} \left(\prod_{i=0}^{s} P_{n,i+j(i)} \right) \left(\prod_{i=0}^{s} \theta_{i+j(i)} \right),$$

using (A1), and

$$\det(B(\vec{\mu}_{2s+1})) = \sum_{\mathcal{P}} (-1)^{\sigma} \left(\prod_{i=0}^{s} P_{n,i+j(i)+1}\right) \left(\prod_{i=0}^{s} \theta_{i+j(i)+1}\right)$$

using (A2). Note that $P_{n,t}$ is $O(n^t)$, i.e., $\lim_{n\to\infty} P_{n,t}/n^t = 1$. Thus, for any permutation $(j(0), j(1), \cdots, j(s))$, $\prod_{i=0}^{s} P_{n,i+j(i)}$ is $O(n^{q_1})$, where $q_1 = \sum_{i=0}^{s}(i+j(i)) = s^2+s$, and $\prod_{i=0}^{s} P_{n,i+j(i)+1}$ is $O(n^{q_2})$, where $q_2 = s^2 + 2s + 1$. Thus, for $l = 0, 1$, since $P_{n,i+j(i)+l}$ and $P_{n,2i+l}$ have the same order,

$$\lim_{n\to\infty} \prod_{i=0}^{s} (P_{n,i+j(i)+l}/P_{n,2i+l}) = 1. \tag{A3}$$

Since $\vec{\mu}_k \in M_k^o(\mathcal{F})$, $\det(A(\vec{\mu}_{2s})) > 0$ for $s = 0, 1, \cdots, \lfloor k/2 \rfloor$. (See Lemma 2.) Hence, for $s = 0, 1, \cdots, \lfloor k/2 \rfloor$, by definition of $\det(A(\vec{\mu}_{2s}))$,

$$\sum_{\mathcal{P}} (-1)^{\sigma} \prod_{i=0}^{s} \mu_{i+j(i)} > 0,$$

which implies, since $\prod_{i=0}^{s} \mu_{2i} > 0$,

$$\sum_{\mathcal{P}} (-1)^{\sigma} \prod_{i=0}^{s} \frac{\mu_{i+j(i)}}{\mu_{2i}} > 0,$$

which implies by definition of θ_t

$$\sum_{\mathcal{P}} (-1)^{\sigma} \left(\prod_{i=0}^{s} \frac{P_{n,i+j(i)}}{P_{n,2i}}\right) \left(\prod_{i=0}^{s} \frac{\theta_{i+j(i)}}{\theta_{2i}}\right) > 0,$$

which implies by (A3) that there exists $n_s{}^*$ such that for $n \geq n_s{}^*$,

$$\sum_{\mathcal{P}} (-1)^{\sigma} \prod_{i=0}^{s} \frac{\theta_{i+j(i)}}{\theta_{2i}} > 0,$$

which implies, since $\prod_{i=0}^{s} \theta_{2i} > 0$, that for $n \geq n_s{}^*$,

$$\det(A(\vec{\theta}_{2s})) > 0.$$

Similarly, for $s = 0, 1, \cdots, \lfloor (k-1)/2 \rfloor$, Lemma 1 says that $\det(B(\vec{\mu}_{2s+1})) > 0$, which implies that there exists n_s^{**} such that for $n \geq n_s^{**}$, $\det(B(\vec{\theta}_{2s+1})) > 0$. So, let

$$n^* = \max(n_1^*, \cdots, n_{\lfloor k/2 \rfloor}^*, n_0^{**}, \cdots, n_{\lfloor (k-1)/2 \rfloor}^{**}).$$

Then, for $n \geq n^*$, $\det(A(\vec{\theta}_{2s})) > 0$ for $s = 1, 2, \cdots, \lfloor k/2 \rfloor$, and $\det(B(\vec{\theta}_{2s+1})) > 0$ for $s = 0, 1, \cdots, \lfloor (k-1)/2 \rfloor$. Hence, by Lemma 1, $\vec{\theta}_k \in M_k^o(\mathcal{F})$. Further, there exists a discrete cdf $G \in \mathcal{F}$ with finite support and first k moments $\vec{\theta}_k$. (See the second paragraph of the proof of Lemma 3.) Let $\lambda_1^{-1}, \cdots, \lambda_m^{-1}$, be the points at which G has mass, and for $i = 1, 2, \cdots, m$, let p_i be the probability assigned to λ_i^{-1}. By Lemma 2, the mixture of Erlang distributions, $ME_n(m, \vec{p}_m, \vec{\lambda}_m)$, has moments $\vec{\mu}_k$. Thus, $\vec{\mu}_k \in M_k(\mathcal{ME}_n)$. Moreover, since $\vec{\theta}_k$ is in the *interior* of $M_k(\mathcal{F})$, the cdf G is not unique. Hence, the corresponding ME_n is not unique, which implies that $\vec{\mu}_k \in M_k^o(\mathcal{ME}_n)$. (See the second paragraph of the proof of Lemma 3.) So, $M_k^o(\mathcal{F}) \subset M_k^o(\mathcal{ME})$. Since $\mathcal{ME} \subset \mathcal{F}$, $M_k^o(\mathcal{F}) = M_k^o(\mathcal{ME})$. By Lemma 3, there exists a cdf G as defined above such that $m \leq \lfloor k/2 \rfloor + 1$. Thus, $M_k(\mathcal{ME}_n) = \bigcup_{m=1}^{\lfloor k/2 \rfloor + 1} M_k(\mathcal{ME}_n)$. □

PROOF OF THEOREM 2

Let F be a nondegenerate cdf with first three noncentral moments $\vec{\mu}_3$, and second and third standardized moments c and γ. For $t = 1, 2, \cdots$, let $\theta_t = \mu_t / P_{n,t}$. By Lemma 1, $\vec{\theta}_3 \in M_3^o(\mathcal{F}))$ if and only if (A4) - (A6) hold.

$$\theta_1 > 0 \tag{A4}$$

$$\theta_2 - \theta_1^2 > 0 \tag{A5}$$

$$\theta_1 \theta_3 - \theta_2^2 > 0 \tag{A6}$$

Moreover, as mentioned in the proof of Theorem 1, any point in $M_3^o(\mathcal{F})$ can be matched by a discrete distribution with finite support. Thus, by Lemma 2, $\vec{\theta}_3 \in M_3^o(\mathcal{F})$ if and only if $\vec{\mu}_3 \in M_3^o(\mathcal{ME}_n)$. Inequalities (A4) - (A6) are equivalent to (A7) - (A9).

$$\mu_1 > 0 \tag{A7}$$

$$\mu_2 - \left(\frac{n+1}{n}\right) \mu_1^2 > 0 \tag{A8}$$

$$\mu_1 \, \mu_3 - \left(\frac{n+2}{n+1}\right) \mu_2^2 > 0 \tag{A9}$$

The left sides of (7) and (8) follow from the right sides of (A8) and (A9). Hence, for any point $(c-1/c, \gamma)$ for which (7) and (8) are strict, $\vec{\theta}_3 \in M_3^o(\mathcal{F})$, $\vec{\mu}_3 \in M_3^o(\mathcal{ME}_n)$, and $(c-1/c, \gamma) \in R^o(\mathcal{ME}_n)$.

Now consider $\delta(R(\mathcal{ME}_n))$, i.e., the points for which (7) or (8) is tight. If $c = 1/\sqrt{n}$, then $\theta_2 - \theta_1^2 = 0$, which implies that the only distribution that corresponds to $\vec{\theta}_3$ is degenerate. Hence, by Lemma 2, if $\vec{\mu}_3 \in M_3(\mathcal{ME}_n)$, $\vec{\mu}_3$ corresponds to an E_n. One can easily verify that (8) is also tight for an E_n. Now suppose (8) is tight. Then $\theta_1 \, \theta_3 - \theta_2^2 = 0$, which, by Proposition 1, implies that $\vec{\theta}_3$ corresponds to a generalized Bernoulli distribution. So again by Lemma 2, if $\vec{\mu}_3 \in M_3(\mathcal{ME}_n)$, $\vec{\mu}_3$ corresponds to an E_n. Hence, (7) is tight also.

Thus, $R(\mathcal{ME}_n)$ contains only one boundary point, that for which (7) and (8) are both tight, and this point corresponds to an E_n.) By Theorem 1, $R(\mathcal{ME}_n) = R(\mathcal{ME}_n(1)) \cup R(\mathcal{ME}(2))$. Since $R(\mathcal{ME}_n(1))$ consists of the points for which (7) and (8) are tight, $R^o(\mathcal{ME}_n) = R(\mathcal{ME}_n(2))$. □

PROOF OF THEOREM 3

The first three noncentral moments of F are

$$\mu_t = P_{n,t} \left(p_1 \, \lambda_1^{-t} + (1-p_1) \lambda_2^{-t} \right) \quad t = 1, 2, 3. \tag{A10}$$

The three equations, (A10) for $t = 1$, 2, and 3, are solved below in terms of λ_1^{-1}, λ_2^{-1}, and p_1.

Let $z = \lambda_2 / \lambda_1$. From for $t = 1$, 2 we have (A11) and (A12).

$$\frac{\lambda_2}{n} \mu_1 - 1 = (z-1) \, p_1 \tag{A11}$$

$$\frac{\lambda_2^2}{n\,(n+1)} \, \mu_2 - 1 = (z^2 - 1) \, p_1 \tag{A12}$$

Divide (A12) by (A11) and solve for z.

$$z = \frac{\lambda_2 \mu_1 - \dfrac{\lambda_2^2}{n+1} \mu_2}{n - \lambda_2 \mu_1}. \tag{A13}$$

For $t = 2$ and 3, express (A10) so that it has the same form as (A12) for $t = 1$ and 2. The result is (A15), and intermediate steps are as follows. Let $w = (\frac{n}{\lambda_1 \mu_1}) p_1$ and $\nu_t = (\frac{n}{n+t})(\frac{\mu_{t+1}}{\mu_1})$, $t = 1$ and 2. For $t = 2$ and 3, substitute $p_1 = \frac{w \lambda_1 \mu_1}{n}$ into (A10):

$$\frac{(n-1)!}{(n+t-1)!} \mu_t = (\frac{w \mu_1}{n}) \lambda_1^{-(t-1)} + (\frac{n - w \lambda_1 \mu_1}{n \lambda_2}) \lambda_2^{-(t-1)}, \qquad t = 2,\ 3$$

Multiply through by n/μ_1 and rearrange slightly:

$$\frac{(n-1)!}{(n+t-2)!} (\frac{n \mu_t}{(n+t-1) \mu_1}) = w \lambda_1^{-(t-1)} + (\frac{n - w \lambda_1 \mu_1}{\mu_1 \lambda_2}) \lambda_2^{-(t-1)}, \qquad t = 2,\ 3$$

Shift the t index by one and substitute $\nu_t = (\frac{n}{n+t})(\frac{\mu_{t+1}}{\mu_1})$:

$$\frac{(n-1)!}{(n+t-1)!} \nu_t = w \lambda_1^{-t} + (\frac{n - w \lambda_1 \mu_1}{\mu_1 \lambda_2}) \lambda_2^{-t}, \qquad t = 1,\ 2 \tag{A14}$$

By substituting $(w \lambda_1 \mu_1)/n$ for p_1 in (A10) with $t = 1$, one can verify that $(n - w \lambda_1 \mu_1)/(\mu_1 \lambda_2) = 1 - w$. So, (A14) is equivalent to (A15).

$$\nu_t = P_{n,t} \{w \lambda_1^{-t} + (1-w) \lambda_2^{-t}\}, \qquad t = 1,\ 2 \tag{A15}$$

Now, solve (A15), $t = 1$ and 2, for $z = \lambda_2/\lambda_1$, as done previously with equation (A10), $t = 1$ and 2, to obtain (A13). The result is

$$z = \frac{\lambda_2 \nu_1 - (\dfrac{\lambda_2^2}{n+1}) \nu_2}{n - \lambda_2 \nu_1}. \tag{A16}$$

Equate the right sides of (A13) and (A16) to eliminate z. Tedious algebra leads to the following quadratic equation in λ_2^{-1}.

$$n(\mu_1 - \nu_1) \lambda_2^{-2} - (\frac{n}{n+1})(\mu_2 - \nu_2) \lambda_2^{-1} + (\frac{1}{n+1})(\nu_1 \mu_2 - \mu_1 \nu_2) = 0 \tag{A17}$$

Hence,

$$\lambda_2^{-1} = \frac{-B' \pm \sqrt{B'^2 - 4A'C'}}{2A'} \tag{A18}$$

with A', B', and C' defined in the obvious manner.

Now, reverse the roles of p_1 and p_2 and of λ_1 and λ_2 in the above derivation to obtain an expression for λ_1^{-1}. The expression obtained is again the right side of (A18). (This can also be concluded by a symmetry argument.) So, λ_1^{-1} and λ_2^{-2} are the two roots of (A17).
To obtain an expression for p_1 in terms of λ_1 and λ_2, solve (A10), $k = 1$, for p_1. The result is (12).

Finally, to express λ_1^{-1} and λ_2^{-1} in terms of y and x, multiply equation (A17) by $-(n+1)(n+2)\mu_1^2/n$. The resulting equation can be expressed as $A\lambda_2^{-2} + B\lambda_2^{-1} + C = 0$, where A, B, and C are as defined in (13) - (15). Equation (11) follows immediately. □