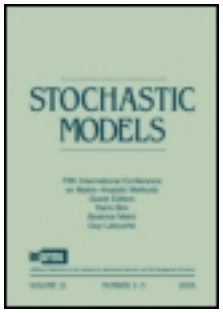

Communications in Statistics. Stochastic Models
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/lstm19

Matching moments to phase distributions: nonlinear programming approaches

Mary A. Johnson [a] & Michael R. Taaffe [b]
[a] Systems and Industrial Engineering , The University of Arizona , Arizona, 85721
[b] Department of Industrial and Manufacturing Engineering , The University of Rhode Island , Kingston, Rhode Island, 02881-0805
Published online: 21 Mar 2007.

To cite this article: Mary A. Johnson & Michael R. Taaffe (1990) Matching moments to phase distributions: nonlinear programming approaches, Communications in Statistics. Stochastic Models, 6:2, 259-281, DOI: 10.1080/15326349908807147

To link to this article: http://dx.doi.org/10.1080/15326349908807147

COMMUN. STATIST.-STOCHASTIC MODELS, 6(2), 259-281 (1990)

MATCHING MOMENTS TO PHASE DISTRIBUTIONS: NONLINEAR PROGRAMMING APPROACHES

MARY A. JOHNSON

Systems and Industrial Engineering
The University of Arizona
Tucson, Arizona 85721

MICHAEL R. TAAFFE

Department of Industrial and
Manufacturing Engineering
The University of Rhode Island
Kingston, Rhode Island 02881-0805

ABSTRACT

We present a nonlinear programming (NLP) approach to the problem of matching three moments to phase distributions. We first discuss the formulation and implementation of a general NLP problem and then consider NLP problems for searching over two families of phase distributions: mixtures of two Erlang distributions and real-parametered continuous Coxian distributions. Restricting the search to select from a subset of phase distributions allows us to greatly simplify the NLP problem, resulting in more efficient and predictable search procedures. Conversely, the restriction also reduces the variety of distributions the search algorithm can select. Tradeoffs between the formulations and possible refinements and extensions are discussed.

1. INTRODUCTION

The family of phase (PH) distributions has been central to the development of algorithmic probability. However, an obstacle to the use of existing algorithmic methods for analysis of stochastic systems has been the problem of PH-distribution selection. Though PH distributions are very versatile, they are not parsimonious, nor are representations of PH distributions unique. Thus, selecting a PH-distribution representation, given a non-PH distribution to be approximated or partial information such as three moments or data, is a key problem in algorithmic probability. In theory, the distribution of any nonnegative random variable can be approximated arbitrarily closely by a PH distribution, Asmussen [1, p. 76]. Thus, models based on PH distributions are, in principle, much more general than classical Markov-type models. By presenting practical methods of PH-distribution selection, we narrow the gap between the theoretical and practical utility of PH distributions.

In a collection of three papers, Johnson and Taaffe (J & T) [6] and [7] and this paper, we consider the problem of selecting PH distributions. Our emphasis is on matching three moments to PH distributions. Our decision to study methods for matching three moments is motivated by the appropriateness of matching three moments in some queueing-approximation applications. The reader is cautioned against blindly using this type of approximation without assessing its appropriateness for the intended application. We briefly describe the three papers of this collection below.

In J & T [6], we develop analytic results for matching moments to mixtures of Erlang distributions of common order (MECO's). We show that, except for a special case, the first k moments of a distribution with support on $[0, \infty)$, i.e., with no probability assigned to the interval $(-\infty, 0)$, can be matched to a mixture of $\lfloor k/2 \rfloor + 1$ or fewer Erlang distributions of common order. For the $k = 3$ case, we develop analytic expressions for the parameters of a mixture of two Erlang distributions of fixed common order that match a triple of first three moments and for the minimum feasible order of the mixed Erlang distributions.

In this paper, we discuss the formulation and implementation of NLP problems for matching three moments to distributions of three families of PH distributions with support on $(0, \infty)$: (general) continuous PH distributions with support on $(0, \infty)$, real-parametered Coxian distributions with support on $(0, \infty)$, and mixtures of two Erlang distributions (not necessarily of common order). Throughout this paper, we refer to the first two families simply as continuous PH distributions and Coxian distributions, respectively. Though our focus is on matching three moments, our NLP-problem formulations can be generalized to match any number of moments.

In J & T [7], we demonstrate the variety of density-function shapes obtainable with the methods of J & T [6] and this paper. We also show how to use those methods to effect desired changes in distribution characteristics for a given triple of moments.

The organization of the remaining sections of this paper is as follows. In Section 2 we formulate an NLP problem for matching three moments to a continuous PH distribution of fixed dimension. We also discuss implementation difficulties and the accompanying motivation for developing alternative, more restrictive formulations. In Section 3 we introduce an NLP problem for matching three moments to a mixture of two Erlang distributions, and in Section 4 we discuss an NLP problem that selects distributions from the Coxian family. In Section 5 we conclude by reviewing some ideas for further development of NLP methods for selecting PH distributions.

2. CONTINUOUS PH DISTRIBUTIONS

2.1 PH-distribution Definition. A PH distribution of dimension n is defined as the distribution of the time until absorption in Markov process with a finite number, n, of transient states and one absorbing state, state $n+1$. A PH distribution is *represented* by $(\vec{\alpha}, \mathbf{T})$, where $\vec{\alpha}$ is an n-dimensional row vector of initial-state probabilities assigned to the transient states and $\mathbf{T}$ is the $n \times n$ matrix obtained by deleting the last row and column of the generator matrix of an associated Markov process. Our restriction to PH distributions with support on $(0, \infty)$ is equivalent to requiring that $\sum_{i=1}^{n} \alpha_i = 1$. Let $\vec{e}$ denote an n-dimensional column vector of one's. The density function of the n-dimensional PH distribution with representation $(\vec{\alpha}, \mathbf{T})$ is $f(x) = -\vec{\alpha} \exp(\mathbf{T}x)\mathbf{T}\vec{e}$ for $x > 0$, and the ith noncentral moment of this distribution is $\mu_i = (-1)^i\, i!\, \vec{\alpha}\mathbf{T}^{-i}\vec{e}$. Note that the existence of $\mathbf{T}^{-1}$ implies that all noncentral moments of the associated PH distribution exist and are finite.

2.2 Formulation of NLP Problem. We refer to the moments to be matched as the *target moments*. Let μ_i denote the ith noncentral target moment, and let c and γ denote the corresponding second and third standardized moments. We define c to be the coefficient of variation (the standard deviation divided by the mean); γ is the coefficient of skewness (the third central moment divided by the cube of the standard deviation). As discussed in J & T [6], the mean of a distribution on $[0, \infty)$ is a scale parameter, and the standardized moments reflect the shape of the distribution function. Moreover, by multiplying the matrix $\mathbf{T}$ by a scalar, the mean of a PH distribution with representation $(\vec{\alpha}, \mathbf{T})$ is easily adjusted to any positive number without changing the standardized moments. So, the problem of matching (μ_1, μ_2, μ_3) to a PH distribution can be solved by matching c and γ to a PH distribution and then multiplying the selected matrix $\mathbf{T}$ by the appropriate scalar.

Our NLP problem for matching c and γ to a continuous PH distribution is as specified below. We refer to this problem as GNLP.

$$\text{minimize} \quad w_1(c - c')^2 + w_2(\gamma - \gamma')^2 \tag{2.1}$$

$$\text{subject to} \quad \vec{\alpha} \geq \vec{0} \tag{2.2}$$

$$\vec{\alpha}\,\vec{e} = 1 \tag{2.3}$$

$$t_{1,1} = -1 \tag{2.4}$$

$$t_{ii} < 0 \qquad \text{for} \quad i = 2, \cdots, n \tag{2.5}$$

$$t_{ij} \geq 0 \qquad \text{for} \quad i \neq j \tag{2.6}$$

$$\sum_{j \neq i} t_{ij} \leq -t_{ii} \qquad \text{for} \quad i = 1, \cdots, n \tag{2.7}$$

$$\det(\mathbf{T}) \neq 0 \tag{2.8}$$

In objective function (2.1), c' and γ' are the second and third standardized moments of the final solution, and w_1 and w_2 are positive weights assigned to the squared differences between target and final-solution moments. The target moments are matched when the value of the objective function is zero. We refer to a solution which matches the target moments as a *moment-matching solution.* The weights in (2.1) are in the spirit of multiple-objective optimization. That is, the weights may be chosen to reflect the relative importance the user places on matching the respective standardized moments. More important, we have found that modifying the weights is a helpful technique for guiding the search to a moment-matching solution. Constraints (2.2) and (2.3) force $\vec{\alpha}$ to be a probability vector. Constraint (2.4) is included to avoid changes in the current solution that result in scale changes only. Constraints (2.5), (2.6), and (2.7) are necessary and sufficient for phases 1, 2, $\cdots$, n to be nonabsorbing states of a Markov process. Constraint (2.8) is necessary and sufficient for phases 1, 2, $\cdots$, n to be transient.

Constraint (2.8) is the only nonlinear constraint of GNLP. Since nonlinear constraints typically induce greater algorithmic difficulties than linear constraints, elimination of constraint (2.8) is desirable. One approach is to use a barrier method to incorporate the constraint into the objective function. However, such a modification does not necessarily eliminate algorithmic difficulties, since the new objective function surface is probably more ill-behaved than the original surface. We have chosen to simply eliminate the constraint $\det(\mathbf{T}) \neq 0$ from the implemented NLP problem and deal with it interactively. If during execution of the NLP search the matrix inversion routine fails due to the singularity of $\mathbf{T}$, we modify the pair (w_1, w_2) or the initial solution and then re-execute the search. As explained below, user interaction is often necessary or desirable anyway. Also, our limited experience indicates that the search seldom encounters a singular matrix.

2.3 Implementation. We have implemented all of our NLP problems on a Gould NP-1 computer. Implementation procedures and insights for GNLP are discussed in this section.

2.3.1 NLP algorithm. In our implementation of GNLP we use the package GRG2 (release 1), which is a collection of algorithms for implementing the *generalized reduced gradient* method in double-precision arithmetic. See Lasdon and Waren [8] for details on the package or Luenberger [9] for an introduction to the generalized reduced gradient method. We estimate gradients of the objective and constraint functions by a central finite-difference approximation (fda). Other algorithm parameters are left at default settings.

Our overall strategy is to interactively manipulate (w_1, w_2) and the initial solution until a moment-matching solution is obtained. We always set w_1 and w_2 to values greater than or equal to one and often set w_1 initially to a value at least three times as great as w_2. If the search matches one moment and not the other, the weight associated with the unmatched moment is increased. If the moments of the final solution approximate the target moments but do not match them to the desired precision (usually three significant digits), we increase the magnitude of both w_1 and w_2 to force the search to continue until greater precision is achieved. If neither c' nor γ' is close to the target moments, adjustments of the ratio w_1/w_2 and/or the initial solution are used to redirect the search.

2.3.2 Dimensionality. A major difficulty with GNLP is its dimensionality. If GNLP searches over the set of n-dimensional continuous PH distributions, it has a total of $n(n+1) = O(n^2)$ decision variables: n elements of the vector $\vec{\alpha}$ and n^2 elements of the matrix T (counting $t_{1,1}$). When n is very small, say two or three, dimensionality is not an issue, but when $n = 7$, for example, there are fifty-six decision variables.

The large number of decision variables makes implementation of GNLP cumbersome and tedious. To implement GNLP the user must specify (preferably feasible) initial values and bounds for each of the $n(n+1)$ decision variables. Though much of the input does not change from one execution of the search to the next, initial values of the decision variables may have to be changed often, since a particular initial solution may not lead to a moment-matching solution. Thus, trial-and-error modifications of the initial solution can be time consuming in comparison to the alternatives discussed in Sections 3 and 4.

The numerical effort required by GNLP is also far greater than that of our alternative NLP problems. Evaluation of the objective function of GNLP is an $O(n^3)$ operation. Use of a central fda to approximate the objective-function gradient requires that the objective function be evaluated at the current solution and twice for each of the $n(n+1)$ decision variables. Thus, evaluation of the objective-function gradient via central fda is an $O(n^5)$ operation. (Analytic evaluation of the objective-function gradient is also $O(n^5)$.) Evaluation of constraints (2.2) - (2.7) is only an $O(n^2)$ operation, but approximation of the Jacobian of the left sides of constraints (2.3) and (2.7) via an fda is an $O(n^4)$ operation. (Since these constraints are linear, computation of an analytic Jacobian would be trivial; but GRG2 requires that the same method, fda or

analytic, be used to compute the gradients of all functions, whether objective or constraint.) The complexity of computing objective and constraint functions and their gradients typically results in nonnegligible run-times, often a few minutes long.

2.3.3 Objective-function surface. Because the number of decision variables in GNLP typically far exceeds the number of constraints, most instances of GNLP have multiple solutions and thus a nonconvex objective-function surface. This causes two types of problems. First, it implies that some combinations of initial solution and (w_1, w_2) may not lead to a moment-matching solution. Second, for a given dimension the user generally does not know what types of solutions exist or how to guide the search toward a preferable solution. Thus, without further sophistication, the flexibility of GNLP translates into unpredictability. With experience, a user may develop intuition about appropriate initial solutions, shortening but not eliminating the time-consuming task of interactively adjusting the initial solution and/or (w_1, w_2) until an acceptable solution is obtained.

2.3.4 Initial solution. The dimension of the PH distribution selected may be considered part of the initial solution. Our approach to estimating the minimum feasible dimension for a given pair (c, γ) is to use insight gained from matching moments to restricted families of PH distributions. For the families of PH distributions discussed in Sections 3 and 4, we provide results on the set of moments feasible for distributions of a fixed dimension. For each family, these results immediately lead to results on the minimum feasible dimension necessary to match a particular (c, γ) pair. Thus, we can obtain upper bounds for the minimum dimension necessary in the general case. Though we are uncertain of the exact minimum feasible value, the difficulties of an iterative search for a feasible dimension makes our approach attractive. An iterative search for a feasible dimension would begin with a low value of n, say two, and then successively increment n and solve the NLP problem until a moment-matching solution is obtained. The poor behavior of the objective-function surface makes such an approach potentially computationally expensive.

Some of our experience with initial solutions for a fixed value of n is reported below. Typically, we initially set $\vec{\alpha}$ to $(1/n, 1/n, \cdots, 1/n)$ or $(1, 0, \cdots, 0)$. An initial matrix $\mathbf{T}$ that we sometimes use consists of $t_{i,i} = -1$ and $t_{i,j} = 1/n$ for $i \neq j$ and $i = 1, 2, \cdots, n$. This solution sets all transition probabilities, $t_{i,j}/-t_{i,i}$, equal to $1/n$ and all mean sojourn times, $-t_{i,i}^{-1}$, equal to one. One reason for this choice of initial solution is that it does not bias the final solution toward a distribution representation that contains little or no feedback. We are especially interested in the effect of feedback, since all of our restricted families of PH-distribution representations do not allow feedback. However, this initial solution tends to lead to one of two problems. First, the reduced gradient at this solution is sometimes zero. Second, more often than for other initial solutions, this one leads to encountering a singular matrix $\mathbf{T}$ during the search. One modification that more often leads to a moment-matching solution consists of setting $t_{i,i}$ to $-i$ and $\lambda_{i,j}$ to i/n for $i \neq j$ and

$i = 1, 2, \cdots, n$. This modification retains the above transition probabilities, but systematically varies mean sojourn times over the values $1/n$, $1/(1-n)$, $\cdots$, 1. For some sets of target moments, neither of the above initial solutions leads to a moment-matching solution. Moreover, we consistently observed that most routing probabilities and initial-state probabilities were reduced to zero in the final moment-matching solutions. This motivates another alternative initial solution, a convolution of exponential distributions, for which $\alpha_1 = 1$, $t_{i,i+1} = -t_{i,i}$ for $i = 1, 2, \cdots, n$.

2.4 Effects of Restricting Selection to a Subset of PH Distributions. The difficulties discussed in Section 2.3 indicate that use of GNLP is a troublesome means of selecting PH distributions to match three moments. Since tedious user interaction is often necessary, and since the computation time may be high, the user may expend considerable effort in obtaining solutions and have little control over what solutions are obtained.

One way of guiding the search is to impose additional constraints on the selected distribution. Constraints can be used to guide the search toward a preferred solution or to obtain a restricted version of the general moment-matching problem that is easier to solve. In the spirit of the latter purpose, we have developed moment-matching methods that select from three subsets of PH distributions: mixtures of two Erlang distributions of common order (MECO-2's), mixtures two of Erlang distributions (not necessarily of common order), and Coxian distributions. These families are listed in order of increasing generality, i.e., each family contains the previous family. All three families accommodate all triples of first three moments that can be matched to some continuous PH distribution. Whether our restrictions are helpful in obtaining desirable characteristics or induce undesirable properties is partially addressed in J & T [7]. Algorithmic trade-offs among these families are summarized below.

The analytic results available for MECO-2's lead to efficient, straightforward algorithms for finding the minimum feasible order of the mixed distributions and matching three moments to a unique MECO-2 of specified feasible order. (However, implementation issues arise even for this case; see J & T [6].) So, by forfeiting much of the flexibility of PH distributions, one may have a simple, efficient, predictable moment-matching algorithm that requires no user interaction and can easily be implemented as a part of a larger algorithm.

The family of mixtures of two Erlang distributions of (possibly) different orders requires use of NLP methods to match moments and sometimes leads to nonunique solutions. But our NLP problem for matching c and γ to a mixture of two Erlang distributions has only two (bounded) variables and no additional constraints. Moreover, minimum feasible orders can still be computed analytically for this family. Also, mixing Erlang distributions of different orders allows for selecting distributions of lower dimension than feasible for MECO-2's.

Matching c and γ to Coxian distributions involves an NLP problem with $2n-1$ (bounded) variables and $2n-2$ linear constraints. For this family, we can no longer state with certainty the minimum feasible dimension for a given point (c, γ), though we believe we can approximate it well.

3. MIXTURES OF TWO ERLANG DISTRIBUTIONS

3.1 Motivation and Notation. Our motivation for generalizing the MECO-2 family to mixtures of two Erlang distributions of possible different orders is threefold. The obvious motivation is to gain a wider variety of distributions. A second motivation is that, as shown below, for a given pair (c, γ), this generalization allows selection of representations of lower dimension than feasible for MECO-2's. The third motivation for this family is the favorable algorithmic properties of Erlang mixtures. For example, the density function of an Erlang mixture is easily evaluated, while the density function of an arbitrary PH distribution involves a matrix exponential, which in general is difficult to compute. Thus, using Erlang mixtures whenever they are adequate is appropriate.

We use the following notation to describe mixtures of Erlang distributions. An Erlang distribution of order n is denoted by E_n, and an E_n with rate parameter λ is denoted by $E_n(\lambda)$. A mixture of Erlang distributions of distinct orders $n_1, n_2, \cdots$, and n_m is denoted by $ME_{\vec{n}}$, where $\vec{n} = (n_1, n_2, \cdots, n_m)$. An $ME_{\vec{n}}$ that mixes k_i E_{n_i}'s for $i = 1, 2, \cdots, m$ is denoted by $ME_{\vec{n}}(\vec{k})$, where $\vec{k} = (k_1, k_2, \cdots, k_m)$. In particular, a mixture of two E_n's is an $ME_n(2)$, and a mixture of an E_{n_1} and an E_{n_2} is an $ME_{(n_1,n_2)}(1, 1)$. A mixture of $E_{n_1}(\lambda_1)$, $E_{n_2}(\lambda_2)$, $\cdots$, and $E_{n_k}(\lambda_k)$, with probability p_j assigned to $E_{n_j}(\lambda_j)$, is denoted by $\sum_{j=1}^{k} p_j E_{n_j}(\lambda_j)$. In this notation the n_j's are not necessarily distinct. For fixed $\vec{n}$, the set of all $ME_{\vec{n}}$'s is denoted by $m\mathcal{E}_{\vec{n}}$.

In the context of matching three moments, the $(c-1/c, \gamma)$-plane provides a convenient framework. We plot $c-1/c$ instead of c because $\gamma \geq c-1/c$ for any distribution with support on $[0, \infty)$. So on the $(c-1/c, \gamma)$-plane, the region feasible for distributions with support on $[0, \infty)$ is the half plane on and above the $\gamma = c-1/c$ line. For any set of distributions, say $\mathcal{V}$, the region on the $(c-1/c, \gamma)$-plane feasible for $\mathcal{V}$ is $R(\mathcal{V})$. See Sections 3 and 4 of J & T [6] for further explanation of the $(c-1/c, \gamma)$-plane and $R(m\mathcal{E}_n)$, the region on the $(c-1/c, \gamma)$-plane feasible for mixtures of Erlang distributions of order n.

The following theorem indicates that allowing mixed Erlang distributions to be of different orders, is a "phase-saving" relaxation of the common-order restriction of J & T [6]. See the appendix for a proof of Theorem 1.

Theorem 1: Let $n_2 \geq n_1 \geq 1$ and $\vec{n} = (n_1, n_2)$. Then, $R(m\mathcal{E}_{\vec{n}}) = R(m\mathcal{E}_{n_2})$. Moreover, for any $(c-1/c, \gamma) \in R(m\mathcal{E}_{n_2})$, there exists an $ME_{\vec{n}}(1, 1)$ that matches $(c-1/c, \gamma)$.

Theorem 1 can be restated as follows. The set of points on the $(c - 1/c, \gamma)$-plane feasible for mixtures of E_{n_1}'s and E_{n_2}'s is the set of points feasible for mixtures of Erlang distributions of order $\max\{n_1, n_2\}$; and any point in this set corresponds to a mixture of one E_{n_1} and one E_{n_2}. Thus, setting $n_1 = 1$, Theorem 1 implies that for any $(c - 1/c, \gamma) \in R(\mathcal{ME}_{n_2})$, $(c - 1/c, \gamma)$ can be matched to a mixture of an E_1, i.e., an exponential distribution, and an E_{n_2}. So, let n^* denote the minimum n such that $(c - 1/c, \gamma) \in R(\mathcal{ME}_n)$. (See Proposition 4 of J & T [6] for an expression for n^* in terms of c and γ.) By Theorem 1, a PH distribution of dimension $n^* + 1$ can be matched to $(c - 1/c, \gamma)$. In contrast, the minimum dimension of a MECO-2 that can be matched to $(c - 1/c, \gamma)$ is $2n^*$. Thus, for large n^*, $(c - 1/c, \gamma)$ is feasible for a PH distribution of approximately half the dimension necessary when distributions are selected from MECO-2's.

Figures 1 and 2 provide intuition lacking in the rigorous proof of Theorem 1. If scale is ignored and n_1 and n_2 are fixed, the parameters of $p\,E_{n_1}(\lambda_1) + (1-p)E_{n_2}(\lambda_2)$ can be expressed as p and r, where $r = (\lambda_2/n_2)/(\lambda_1/n_1)$, the ratio of the mean of $E_{n_1}(\lambda_1)$ to the mean of $E_{n_2}(\lambda_2)$. Constant-p contours and constant-r contours of $p\,E_1(\lambda_1) + (1-p)E_{10}(\lambda_2)$ are plotted in Figures 1 and 2. Figure 1 also shows the boundaries of $R(\mathcal{ME}_1)$ and $R(\mathcal{ME}_{10})$ (dotted lines), and Figure 2 shows the boundary of $R(\mathcal{ME}_{10})$. As shown in J & T [6], the lower bound on $c - 1/c$ in $R(\mathcal{ME}_n)$ is given by $c = 1/\sqrt{n}$; the lower bound on γ as a function of c corresponds to the DE_n curve, where DE_n denotes a mixture of an E_n and the degenerate distribution at zero. Notice how the contour curves sweep through $R(\mathcal{ME}_{10})$, approaching its boundaries for extreme values of p and r.

The endpoints of the contour curves can be explained analytically. Consider the constant-p curves first. Let $DE_n(p)$ denote a DE_n that assigns probability p to the E_n. For fixed p, $p\,E_1(\lambda_1) + (1-p)E_{10}(\lambda_2)$ converges weakly to a $DE_1(p)$ as $r \to \infty$ and to a $DE_{10}(1-p)$ as $r \to 0$. Hence, the points on the constant-p contour converge to the points that correspond to $DE_1(p)$ and $DE_{10}(1-p)$ as $r \to \infty$ and $r \to 0$, respectively. Now consider the constant-r curves. When $p = 1$, the mixture is simply an E_1. So the $p = 1$ end of each constant-r contour is at $(c - 1/c, \gamma) = (0, 2)$, the point that corresponds to an E_1. When $p = 0$, the mixture is an E_{10}. So, the $p = 0$ end of each contour is at $(c - 1/c, \gamma) = (-2.846, .632)$, which corresponds to an E_{10}.

3.2 Formulation of NLP Problem. Our NLP problem matches c and γ to a mixture of two Erlang distributions by selecting parameters r and p of the mixture $p\,E_{n_1}(\lambda_1) + (1-p)E_{n_2}(\lambda_2)$, where $\lambda_1 = 1$ and $\lambda_2 = (n_2/n_1)r$. So r and p are as defined above. This formulation, denoted by ENLP, is given by (3.1) - (3.3). The elements (w_1, w_2) and (c', γ') of (3.1) are as defined for GNLP.

$$\text{minimize} \qquad w_1(c - c')^2 + w_2(\gamma - \gamma')^2 \tag{3.1}$$

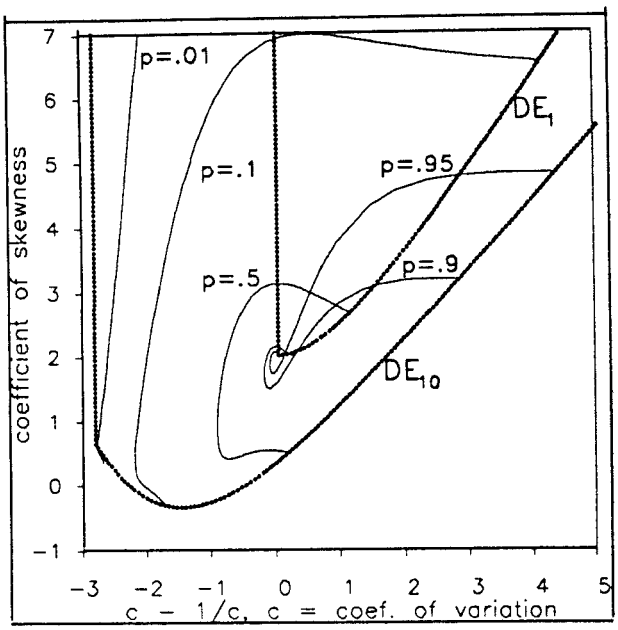

FIG. 1
Constant-p contours for $p\,E_1(\lambda_1) + (1-p)\,E_{10}(\lambda_2)$.

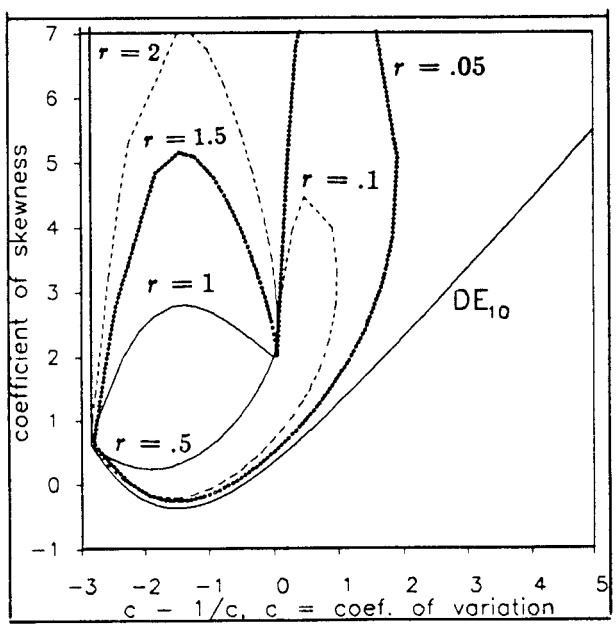

FIG. 2
Constant-r contours for $p\,E_1(\lambda_1) + (1-p)\,E_{10}(\lambda_2)$.

subject to $0 \le p \le 1$ (3.2)

$r \ge 0$ (3.3)

3.3 Implementation. Our description in Section 2.3.1 of the NLP algorithm for GNLP applies to ENLP also. Some implementation issues are discussed below.

3.3.1 Consistency with expected boundary behavior. Our experience in matching points near the boundary of $R(\mathcal{ME}_n)$ is consistent with boundary behavior of MECO-2's (discussed in J & T [6]) and with the contours in Figures 1 and 2. That is, we find that using an $ME_{(n_1,n_2)}(1, 1)$, where $n = n_2 \ge n_1$, to match points near the boundary of $R(\mathcal{ME}_n)$ requires extreme values of p and/or r. In some cases, we have not been able to match the target moments to three significant digits, as done for points farther inside the boundaries. Thus, numerical considerations indicate that using $ME_{(n_1,n_2)}(1, 1)$'s to match points only in $R(\mathcal{ME}_{n-1})$, where $n = \max\{n_1, n_2\}$, may be prudent.

3.3.2 Dimensionality. ENLP has only two decision variables, and the computational effort required to evaluate its objective and constraint functions is independent of n_1 and n_2. For the cases observed, the objective and constraint functions were typically evaluated 100 to 200 times (including evaluations for the central fda of the gradient). Our run-times are usually negligible and never more than several seconds.

3.3.3 Objective-function surface. As indicated by the crossing of constant-p contours in Figure 1, for at least some cases, objective-function (3.1) is nonconvex, and ENLP has more than one moment-matching solution. However, Theorem 2 and Conjecture 1 suggest that the objective-function surface of ENLP is still well behaved.

> **Theorem 2:** Let $n_2 \ge n_1$ and $\vec{n} = (n_1, n_2)$. Let c and γ be the second and third standardized moments, respectively, corresponding to (μ_1, μ_2, μ_3). Suppose $(c - 1/c, \gamma)$ is in the interior of $R(\mathcal{ME}_{n_1})$, i.e., $(c - 1/c, \gamma)$ is in $R(\mathcal{ME}_{n_1})$ but does not correspond to an E_{n_1}. Then there is an $ME_{\vec{n}}(1, 1)$ that matches (μ_1, μ_2, μ_3) and satisfies $\mu_{1(1)} < \mu_1 < \mu_{1(2)}$, where $\mu_{1(j)}$ is the mean of the mixed E_{n_j}; and there is another $ME_{\vec{n}}(1, 1)$ that matches (μ_1, μ_2, μ_3) and satisfies $\mu_{1(1)} < \mu_1 < \mu_{1(2)}$. (Note that the above statement does not exclude the possibility of more than two moment-matching solutions.)

In all of the constant-p and constant-r contour plots we have plotted, the crossings of contours consistently occur inside the feasible region of the lower-order mixed Erlang distribution. This suggests that for $n_2 > n_1$, nonunique solutions exist only for points in $R(\mathcal{ME}_{n_1})$. Our experience in using ENLP and the proofs of Theorems 1 and 2 motivate an even stronger statement, given in

Conjecture 1. See the appendix for an explanation of how the proofs of Theorems 1 and 2 motivate Conjecture 1.

Conjecture 1: Let $n_2 \geq n_1$. If $(c - 1/c, \gamma) \in R(\mathcal{ME}_{n_1})$, then there are *exactly* two $ME_{(n_1,n_2)}(1, 1)$'s, characterized in Theorem 2, that match $(c - 1/c, \gamma)$. If $(c - 1/c, \gamma) \in R(\mathcal{ME}_{n_2})$ but $(c - 1/c, \gamma) \notin R(\mathcal{ME}_{n_1})$, then there is *exactly* one $ME_{(n_1,n_2)}(1, 1)$ that matches $(c - 1/c, \gamma)$.

So, if Conjecture 1 is true, the number of moment-matching solutions to ENLP is easily determined, and this number is never more than two. Moreover, when ENLP has two moment-matching solutions, Theorem 2 implies that by replacing constraint (3.3) with $r \geq 1$ or $0 \leq r \leq 1$ one of the solutions can be excluded. Alternatively, as discussed below, adjustment of the initial solution is a convenient means of obtaining both moment-matching solutions. We have used this method exclusively to obtain alternative solutions.

3.3.4 Initial solution. Our experience indicates that ENLP requires fewer trial-and-error adjustments of the initial solution to obtain a moment-matching solution than are required by our other NLP problems. This good behavior is consistent with Theorem 2 and Conjecture 1. Some care is needed, however, in choosing an initial value of r. Initial values of r near one often result in the search ending with the exponential distribution, even when the target moments are far from those of the exponential. For fixed n_1 and n_2, an initial solution that often works for us is $(r, p) = (100, .10)$. Usually, if a particular initial solution does not lead to a moment-matching solution, setting r and p to their opposite extremes does result in a solution. For example, if $(r, p) = (100, .10)$ does not lead to a moment-matching solution, $(.01, .90)$ usually does. Moreover, for cases which are conjectured to have exactly two moment-matching solutions, those solutions can often be obtained with such complementary initial solutions

4. COXIAN DISTRIBUTIONS

4.1 Motivation and Notation. A Coxian distribution of order n is denoted by C_n and has a representation of the form depicted in Figure 3.

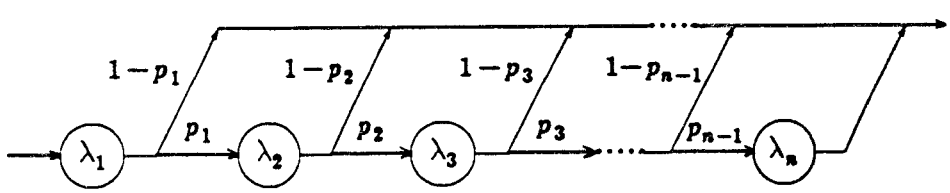

FIG. 3
Coxian distribution.

Each circle in Figure 3 represents a phase. The mean sojourn time at phase i is λ_i^{-1}. Given that the current phase is phase i, the probability that the next transition will be to phase $i+1$ is p_i, and the probability that the next transition will be to the absorbing state is $1-p_i$. The more general definition of a Coxian distribution allows for complex parameters and mass at zero, Cox [3]. The Coxian representation depicted in Figure 3 is $(\vec{\alpha}, \mathbf{T})$, where $\vec{\alpha} = (1, 0, \ldots, 0)$ and T is upper bidiagonal with $t_{i,i} = -\lambda_i$ and $t_{i,i+i} = p_i \lambda_i$. As with representations of mixtures of Erlang distributions, algorithms can be designed to exploit the special structure of Coxian representations. Bux and Herzog [2] and Marie [10] provide examples of queueing models that use Coxian distributions.

The following theorem, which follows immediately from Cumani [4], shows that the set of Coxian distributions is more general than may be apparent. Note that an acyclic PH distribution is a PH distribution that can be represented by $(\vec{\beta}, \mathbf{S})$, where S is upper triangular.

> **Theorem 3**: Any n-phase acyclic continuous PH distribution has a unique representation of the form depicted in Figure 3 with $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \cdots \geq \lambda_n$.

The generality of Coxian distributions, along with the simple structure and few parameters of the Coxian representation (relative to unrestricted representations or general acyclic representations), makes the set of Coxian distributions an especially appealing subset of PH distributions. Moreover, Coxian representations are still sufficiently structured to allow one to obtain a "reasonable" estimate of the minimum order needed to match a fixed point $(c-1/c, \gamma)$. Let $\mathcal{C}_n$ denote the set of all Coxian distributions of order n. Theorem 4 follows immediately from and Theorems 1 and 3.

> **Theorem 4:** $R(\mathcal{ME}_n) \subset R(\mathcal{C}_{n+1})$.

Thus, if n^* is the minimum n such that $(c-1/c, \gamma) \in R(\mathcal{ME}_n)$, then for $m \geq n^*+1$, $(c-1/c, \gamma)$ can be matched to a C_m. But $R(\mathcal{ME}_n)$ is only a proper subset of $R(\mathcal{C}_{n+1})$. To see this, note that an $E_{n+1} \in \mathcal{C}_{n+1}$, though the point that corresponds to E_{n+1} on the $(c-1/c, \gamma)$-plane is left of $R(\mathcal{ME}_n)$. As explained in Section 4.3.1, however, we believe the approximate minimum feasible order implied by Theorem 4 is usually equal to the actual minimum.

4.2 Formulation of NLP Problem. Let $C_n(\vec{\lambda}, \vec{p})$ denote a C_n with parameters $\vec{\lambda}$ and $\vec{p}$, where $\vec{\lambda} = (\lambda_1, \lambda_2, \cdots, \lambda_n)$ and $\vec{p} = (p_1, p_2, \cdots, p_{n-1})$. Our NLP problem for matching c and γ to a $C_n(\vec{\lambda}, \vec{p})$ is expressed in terms of decision variables $\lambda_1, \lambda_2, \cdots, \lambda_n$ and $\lambda_{1,2}, \lambda_{2,3}, \cdots, \lambda_{n-1,n}$, where $\lambda_{i,i+1} = p_i \lambda_i$. This formulation, which we call CNLP, is as follows. Again, (w_1, w_2) and (c', γ') are as defined for GNLP.

$$\text{minimize} \quad w_1(c - c')^2 + w_2(\gamma - \gamma')^2 \tag{4.1}$$

$$\text{subject to} \quad \lambda_{i,i+1} \geq 0, \qquad i = 1, \cdots, n-1 \tag{4.2}$$

$$\lambda_i - \lambda_{i,i+1} \geq 0, \qquad i = 1, \cdots, n-1 \tag{4.3}$$

$$\lambda_i - \lambda_{i+1} \geq 0, \qquad i = 1, \cdots, n-1 \tag{4.4}$$

$$\lambda_n = 1 \tag{4.5}$$

Constraints (4.2) and (4.3) together ensure that $\lambda_i \geq 0$ and $0 \leq p_i \leq 1$ for $i = 1, \cdots, n-1$. Constraints (4.4) and (4.5) are intended to enhance convergence to a solution by avoiding alternative representations of the same solution. Theorem 3 indicates that ordering the rate parameters such that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ does not reduce the set of available distributions. Constraint (4.5) prevents simply scale changes and is analogous to fixing $t_{1,1}$ at negative one in GNLP and λ_1 at one in ENLP. Note that if in the final solution $p_i = 0$ for some $i = 1, \cdots, n-1$, then the solution is reducible, and phases $i+1, \cdots, n$ can be eliminated from the representation.

4.3 Implementation. The NLP algorithm used for GNLP and ENLP is also used for CNLP. Implementation insights and issues are as follows.

4.3.1 Behavior near boundary of feasible subset. For fixed n, points near the edges of the subset, $R(\mathcal{ME}_n)$, of $R(\mathcal{C}_{n+1})$ are either matched by the search or nearly matched. The difficulty experienced in matching points near the boundary of the subset suggests that this subset includes most of $R(\mathcal{C}_{n+1})$. More specifically, the representations of C_{n+1}'s selected for points near the DE_n curve suggest that the DE_n curve represents a lower bound on γ in $R(\mathcal{C}_{n+1})$ as well as $R(\mathcal{ME}_n)$.

First, consider the point $(c - 1/c, \gamma) = (2, 2.8)$ which is .25 above the DE_5 curve. When we matched (2, 2.8) to a C_6, the selected parameters were $\lambda_1 = 18.26$, $p_1 = 0.166$, $p_2 = 0.911$, and all other rates and probabilities equal one. Thus, this example illustrates a striking regularity we have observed repeatedly in C_{n+1} representations selected to match points just above the DE_n curve. When points farther above the DE_n curve are matched to a C_{n+1}, much less regularity is observed. For example, when we matched a C_6 to the point $(c - 1/c, \gamma) = (1.5, 4.19)$, which is well within $R(\mathcal{ME}_5)$, all the selected routing probabilities differed and only λ_5 and λ_6 were assigned the same value.

Now consider an experiment in which we attempted to match a C_6 to a point below the DE_5 curve. In $R(\mathcal{ME}_5)$ the lower bound on γ at $c - 1/c = 4$ is 4.8. When we attempted to match the point (4, 4.4) to a C_6, the NLP algorithm returned a solution with moments $(c - 1/c, \gamma) = (4.2, 4.8)$, which is almost on the DE_5 curve. Further, the NLP algorithm indicated the solution was unbounded and stopped at the solution $\lambda_1 = 423{,}000$, $\lambda_{1,2} = 270{,}000$, and all other parameters equal to one. Thus, by repeatedly increasing the rate parameter of the first phase, the search selected distributions that correspond to points closer and closer to the DE_5 curve at $c - 1/c = 4$ but never below it.

4.3.2 Dimensionality. The number of decision variables in CNLP is $2n - 1 = O(n)$. This is in contrast to two decision variables in ENLP and $O(n^2)$ in GNLP. Computation of objective function (4.1) involves an $O(n^2)$ sequence of upper-triangular-matrix multiplications and additions and one matrix inversion, which is only $O(n)$ when the bidiagonal structure of $\mathbf{T}$ is exploited. Approximating the gradient of the objective function via central fda is an $O(n^3)$ operation. Evaluation of the constraints of CNLP is an $O(n)$ operation. Approximation of the Jacobian of the left sides of (4.3) and (4.4) via central fda is $O(n^2)$. For the cases observed, the objective and constraint functions were typically evaluated at least 100 times and sometimes as high as 700 or more times. Still, run-times are typically only a few seconds.

4.3.3 Objective-function surface. Since the Coxian family is a generalization of the family of mixtures of two Erlang distributions, at least some points in $R(\mathcal{C}_n)$ also have nonunique solutions. Indeed, the increase in generality implies that nonuniqueness is probably more common for CNLP than for ENLP. We have made no attempt to further investigate the nonuniqueness of feasible solutions. However, for a fixed point on the $(c - 1/c, \gamma)$-plane, the number of feasible solutions clearly increases with n.

4.3.4 Initial solution. Most of the instances of CNLP that we have solved start with the following initial conditions: $\lambda_1 = n$, $\lambda_2 = n-1, \cdots, \lambda_n = 1$ and $\lambda_{i,i+1} = \lambda_i/2$ $(p_i = 1/2)$ for $i = 1, 2, \cdots, n-1$, with n typically set at $n^* + 1$ or $n^* + 2$, where n^* is the minimum n such that $(c - 1/c, \gamma) \in R(\mathcal{ME}_n)$. In cases where this set of initial conditions does not lead to a solution, the modification $\lambda_{i,i+1} = \lambda_i$ $(p_i = 1)$ for $i = 1, 2, \cdots, n-1$ usually leads to a solution. But at least in one instance, only after we made the additional modification, $\lambda_i = 1$ for all i, were we able to obtain a moment-matching solution. Overall, we have not experienced a great deal of difficulty in obtaining moment-matching solutions.

5. CONCLUSIONS

The NLP-problem formulations and implementations discussed here suggest that NLP methods have the potential to become a powerful tool for selecting approximating PH distributions. Limitations and potential areas of future development include the following.

One limitation of our NLP methods is that they are not well suited for applications in which the moment-matching procedure must be automated. One reason for this is that we are not able to assure convergence to a moment-matching solution. Also, some of our methods may be too computationally expensive. Finally, for at least some of our NLP problems, nonunique moment-matching solutions induce unpredictability.

Conversely, our NLP methods are well suited for interactive selection of PH distributions. Further, the nonuniqueness of solutions to our NLP problems implies that these problems may be a promising basis for developing more sophisticated algorithms. That is, the addition of new (preferably linear)

constraints or modification of the objective function may allow the user to specify in greater detail properties of selected distributions. For example, the numerical stability of the distributions selected by ENLP and CNLP may be enhanced by adding constraints of the form $a \le r \le b$ and $\lambda_1 - \lambda_n \le b$, respectively. Another interesting modification to CNLP is the addition of the following linear constraint to restrict selection to distributions with decreasing failure rates (DFR's).

$$\lambda_i + \lambda_{i+1,i+2} - \lambda_{i,i+1} - \lambda_{i+1} \ge 0 \quad \text{for} \quad i = 1, \cdots, n-1.$$

This constraint follows immediately from Ridder's result [12, p. 123] that the condition $(1-p_i)\,\lambda_i \ge (1-p_{i+1})\,\lambda_{i+1}$ for $i = 1, \cdots, n-1$ is sufficient for a C_n to be DFR. Unfortunately, analogous constraints to obtain IFR distributions are not apparent.

Use of alternative families of PH distributions may also enhance our ability to obtain desirable distribution properties. For example, use of a family that allows some feedback in the PH-distribution representation may be helpful in matching tail behavior. Also, our density-function observations in J & T [7] and our queueing-approximation results in Johnson [5] suggest that mixing more than two Erlang distributions may allow for improvements over the distributions selected by ENLP. Suppose a mixture of the form $\sum_{i=1}^{k} E_{n_i}(\lambda_i)$, where n_1, n_2, $\cdots$, and n_k are fixed, is to be selected. ENLP can be generalized to handle this problem. Let

$$r_i = (\lambda_i/n_i)/(\lambda_1/n_1) = \mu_{1(1)}/\mu_{1(i)},$$

where $\mu_{1(i)}$ is the mean of $E_{n_i}(\lambda_i)$. The constraints of the NLP can be formulated as follows.

$$0 \le p_i \le 1, \qquad i = 1, 2, \cdots, k \tag{5.3}$$

$$\sum_{i=1}^{k} p_i = 1 \tag{5.4}$$

$$r_i - r_{i-1} \ge 0, \qquad i = 2, 3, \cdots, k \tag{5.5}$$

$$r_1 = 1 \tag{5.6}$$

Constraint (5.6) prevents changes simply in scale. Constraint (5.5) is intended to further enhance convergence and is analogous to constraint (4.4) of CNLP. Here the ordering does restrict generality unless $n_1 = n_2 = \cdots = n_k$, but this loss in generality can be recovered interactively by rearranging the order of the mixed distributions.

We believe that the ability to modify an NLP-problem formulation to accommodate differences in users' priorities and objectives will be an important advantage of NLP methods and a key to obtaining more predictable and favorable solutions from NLP implementations. Development of these new formulations will be accompanied by a better understanding of available distributions from a given family and of methods for guiding the search toward a preferred solution. Characteristics that may be considered in future NLP-problem formulations include the value of the cdf at specified points, locations of quantiles, the value of the density function at zero, numerical stability, failure rate, and unimodality.

ACKNOWLEDGEMENTS

This paper is based on work performed at Purdue University. The first-named author is currently a postdoctoral fellow at the Laboratory for Algorithmic Research, University of Arizona, with research support from Bellcore. This support has contributed to the completion of the paper. We also acknowledge Colm A. O'Cinneide for his help with the proof of Theorem 1.

REFERENCES

1. S. Asmussen, *Applied Probability and Queues,* Wiley, New York, 1987.

2. 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.

3. D. R. Cox, "A Use of Complex Probabilities in the Theory of Stochastic Processes," *Proc. Camb. Phil. Soc.,* Vol. 51, 313-319, 1955.

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

5. M. A. Johnson, "*Phase Distributions: Selecting Parameters to Match Moments,*" Ph.D. dissertation, School of Industrial Engineering, Purdue University, August 1988.

6. M. A. Johnson, and M. R. Taaffe, "Matching Moments to Phase Distributions: Mixtures of Erlang Distributions of Common Order," *Commun. Statist.—Stochastic Models,* Vol. 5, No. 4, 1989.

7. M. A. Johnson, and M. R. Taaffe, "Matching Moments to Phase Distributions: Density Function Shapes," *Commun. Statist.—Stochastic Models,* to appear.

8. L. S. Lasdon, and A. D. Waren, *GRG2 User's Guide,* Department of Computer and Information Science, Cleveland State University, Cleveland, Ohio, 1982.

9. D. G. Luenberger, *Linear and Nonlinear Programming,* second edition, Addison-Wesley Publishing Co., Reading, Massachusetts,

10. R. Marie, "Calculating Equilibrium Probabilities for $\lambda(n)/C_k/1/N$ Queues," *Proceedings of Performance 80 International Symposium on Computer Performance Modelling,* 117-125, May 28-30, 1980.

11. A. A. N. Ridder, *Stochastic Inequalities for Queues* Ph.D. dissertation, Department of Mathematics and Computer Science, University of Leiden, The Netherlands,

APPENDIX

Proof of Theorem 1

We prove the last statement of the theorem first. The previous statement then follows easily. Notation in this proof is consistent with notation in the body of this paper. Though our proof uses noncentral moments only, we provide interpretations in terms of the $(c - 1/c, \gamma)$-plane also.

Let μ_j denote the jth noncentral moment of a probability distribution. Let $M_k(\mathcal{ME}_n)$ be the set of k-tuples of first k noncentral moments feasible for an ME_n, and let $\bar{M}_k(\mathcal{ME}_n)$ be the closure of this set. We review some facts proved in J & T [6]. Inequalities (A1) and (A2) are satisfied if and only if $(\mu_1, \mu_2) \in \bar{M}_2(\mathcal{ME}_n)$; inequalities (A1) - (A3) are satisfied if and only if $(\mu_1, \mu_2, \mu_3) \in \bar{M}_3(\mathcal{ME}_n)$. Moreover, if $(\mu_1, \mu_2, \mu_3) \in M_3(\mathcal{ME}_n)$, (A2) and (A3) are either both tight or both strict. If (A2) is tight for an ME_n, then the ME_n is an E_n and (A3) is also tight. Also, (A2) and (A3) are tight for every E_n. Inequality (A3) is tight if and only if (μ_1, μ_2, μ_3) can be matched by a DE_n. Thus, for fixed (μ_1, μ_2), the lower bound on μ_3 among ME_n's that match (μ_1, μ_2) is attained by the DE_n that matches (μ_1, μ_2). Note that the lower bounds on μ_2 and μ_3 in $M_3(\mathcal{ME}_n)$ decrease with increasing n.

$$\mu_1 \geq 0 \tag{A1}$$

$$\mu_2 \geq \left(\frac{n+1}{n}\right) \mu_1^2 \tag{A2}$$

$$\mu_3 \geq \left(\frac{n+2}{n+1}\right) \frac{\mu_2^2}{\mu_1} \tag{A3}$$

The dotted curves on the (μ_1, μ_2)-plane in Figure 4 show the E_1 and E_5 curves, i.e., plots of (A2) restricted to equality for $n = 1$ and 5. Note that $\bar{M}_2(\mathcal{ME}_1)$ is the region on or left of the μ_2 axis and on or above the E_1 curve; $\bar{M}_2(\mathcal{ME}_5)$ is on or above the E_5 curve. In terms of the $(c - 1/c, \gamma)$-plane, the region above the E_n curve of the (μ_1, μ_2)-plane corresponds to the half plane to

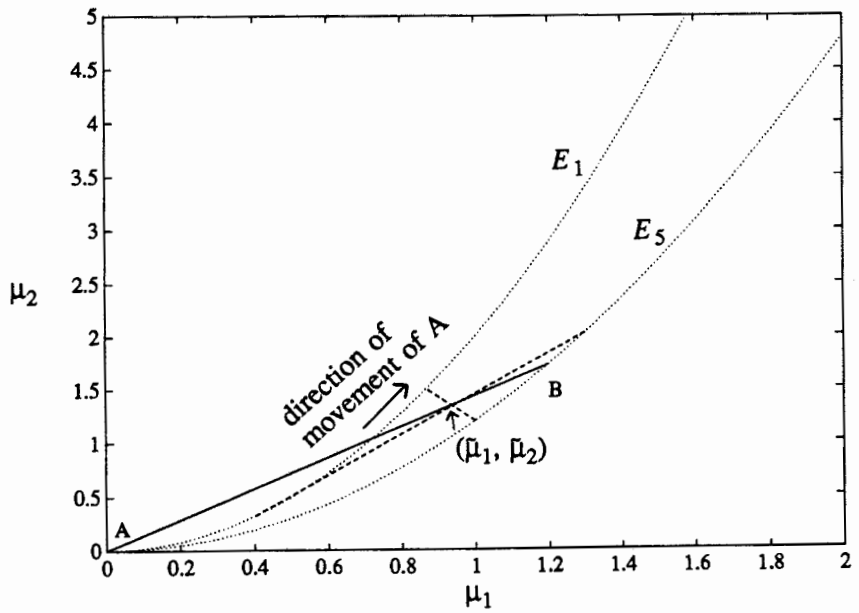

FIG. 4
Case I: $(\tilde{\mu}_1, \tilde{\mu}_2)$ between E_1 and E_5 curves.

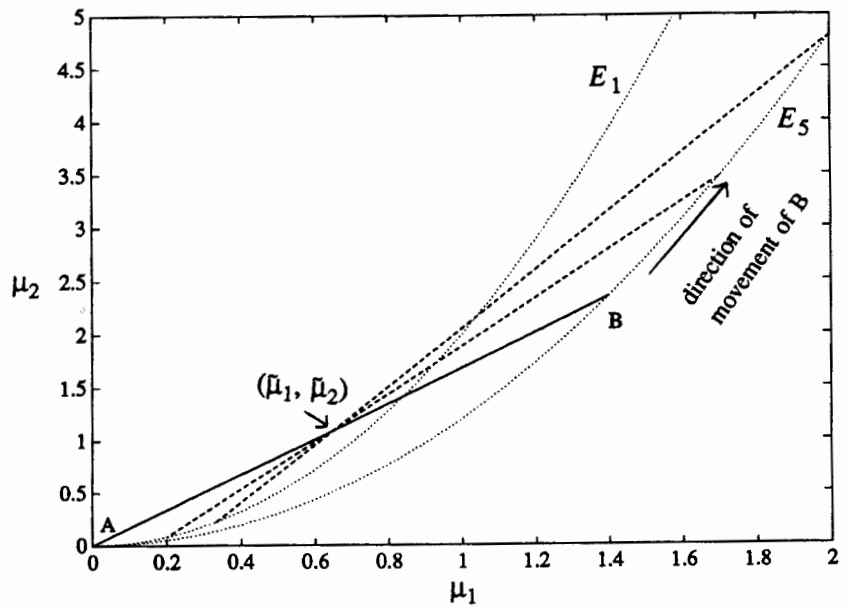

FIG. 5
Case II: $(\tilde{\mu}_1, \tilde{\mu}_2)$ above E_1 curve.

right of the $c - 1/c$ boundary of $R(\mathcal{ME}_n)$. The lower bound on μ_3 is a surface over $M_2(\mathcal{ME}_n)$. This surface corresponds to the DE_n curve on the $(c - 1/c, \gamma)$-plane.

Now let $n_2 > n_1$, $\vec{n} = (n_1, n_2)$, and $(\tilde{\mu}_1, \tilde{\mu}_2) \in M_2(\mathcal{ME}_{n_2})$. If (A2) is tight when $(\mu_1, \mu_2) = (\tilde{\mu}_1, \tilde{\mu}_2)$ and $n = n_2$, then $(\tilde{\mu}_1, \tilde{\mu}_2)$ corresponds to an E_{n_2} and hence an $ME_{\vec{n}}(1, 1)$ with probability one assigned to the E_{n_2}. So, for this case the second statement of Theorem 1 is true. For the rest of our proof of the second statement of Theorem 1, we assume $(\tilde{\mu}_1, \tilde{\mu}_2)$ does not correspond to an E_{n_2}. Let $p_1 E_{n_1}(\lambda_1) + p_2 E_{n_2}(\lambda_2)$ denote an $ME_{\vec{n}}(1, 1)$ that matches $(\tilde{\mu}_1, \tilde{\mu}_2)$. (That such a distribution exists will become obvious below.) Let $\tilde{\mu}_3$ denote the value of μ_3 for this $ME_{\vec{n}}(1, 1)$, and let $\tilde{c}$ and $\tilde{\gamma}$ respectively denote the values of c and γ associated with $(\tilde{\mu}_1, \tilde{\mu}_2, \tilde{\mu}_3)$. We vary p_1, λ_1, and λ_2 so that $\tilde{\mu}_3$ takes on all values for which $(\tilde{\mu}_1, \tilde{\mu}_2, \tilde{\mu}_3) \in M_3(\mathcal{ME}_{n_2})$, i.e., such that $(\tilde{\mu}_1, \tilde{\mu}_2, \tilde{\mu}_3)$ satisfies (A3) with strict inequality. Equivalently, $\tilde{\gamma}$ takes on all values of γ above the DE_{n_2} curve at $c - 1/c = \tilde{c} - 1/\tilde{c}$. Let $\mu_{j(i)}$ denote the jth noncentral moment of E_{n_i}.

Case I: $(\tilde{\mu}_1, \tilde{\mu}_2)$ lies between the E_{n_1} and E_{n_2} curves. (See Figure 4.)

In terms of the $(c - 1/c, \gamma)$-plane, this case corresponds to $\tilde{c} - 1/\tilde{c}$ being between the $c - 1/c$ boundaries of $R(\mathcal{ME}_{n_1})$ and $R(\mathcal{ME}_{n_2})$, i.e, $1/\sqrt{n_2} < \tilde{c} < 1/\sqrt{n_1}$. In Figure 4, the DE_{n_2} that matches $(\tilde{\mu}_1, \tilde{\mu}_2)$ is represented by the line segment that passes through $(\tilde{\mu}_1, \tilde{\mu}_2)$ and has left endpoint, A, at (0, 0) and right endpoint, B, on the E_{n_2} curve. Now suppose line segment $\overline{AB}$ is rotated so that endpoint A moves to the right (and up) on the E_{n_1} curve and B is the point on the E_{n_2} curve such that $\overline{AB}$ passes through $(\tilde{\mu}_1, \tilde{\mu}_2)$. Each new position of $\overline{AB}$ represents new values of p_1, λ_1, and λ_2. Since

$$\frac{\mu_{2(1)}}{\mu_{1(1)}} \to \infty \tag{A4}$$

as $\mu_{1(1)} \to \infty$, $\overline{AB}$ is asymptotically vertical as $\mu_{1(1)} \to \infty$. Thus, as $\mu_{1(1)} \to \infty$,

$$\mu_{1(2)} \to \tilde{\mu}_1, \tag{A5}$$

$$\mu_{2(2)} \to \left(\frac{n_2 + 1}{n_2}\right) \tilde{\mu}_1^2, \tag{A6}$$

and

$$p_1 \to 0 \quad \text{and} \quad p_2 \to 1. \tag{A7}$$

Also, since $\tilde{\mu}_2 = p_1 \mu_{2(1)} + p_2 \mu_{2(2)}$, by (A6) and (A7),

$$p_1 \mu_{2(1)} \to \tilde{\mu}_2 - \left(\frac{n_2 + 1}{n_2}\right) \tilde{\mu}_1^2, \tag{A8}$$

a finite, positive constant. Now consider the limiting behavior of $\tilde{\mu}_3$.

$$\begin{aligned} \tilde{\mu}_3 &= p_1 \mu_{3(1)} + p_2 \mu_{3(2)} \\ &= p_1 \left(\frac{n_1+2}{n_1+1}\right) \frac{\mu_{2(1)}^2}{\mu_{1(1)}} + p_2 \left(\frac{n_2+2}{n_2+1}\right) \frac{\mu_{2(2)}^2}{\mu_{1(2)}}. \end{aligned} \tag{A9}$$

By (A4) and (A8), the first term in (A9) approaches infinity as $\mu_{1(1)} \to \infty$; by (A5), (A6), and (A7), the second term converges to a finite constant. So, $\tilde{\mu}_3 \to \infty$ as $\mu_{1(1)} \to \infty$. Thus, by continuity, all values of $\tilde{\mu}_3$ such that $(\tilde{\mu}_1, \tilde{\mu}_2, \tilde{\mu}_3) \in M_3(\mathcal{ME}_{n_2})$ are obtained by moving endpoint A from $(0^+, 0^+)$ to the right along the E_{n_1} curve.

Case II: $(\tilde{\mu}_1, \tilde{\mu}_2)$ lies on or above the E_{n_1} curve of the (μ_1, μ_2)-plane. (See Figure 5.)

In terms of the $(c - 1/c, \gamma)$-plane, this case corresponds to $\tilde{c} - 1/\tilde{c}$ being on or to the right of the $c - 1/c$ boundary of $R(\mathcal{ME}_{n_1})$. Again, the DE_{n_2} that matches $(\tilde{\mu}_1, \tilde{\mu}_2)$ is represented in Figure 5 by the line segment, $\overline{AB}$, that passes through $(\tilde{\mu}_1, \tilde{\mu}_2)$ and has endpoint, A, at (0, 0) and endpoint, B, on the E_{n_2} curve. Now suppose $\overline{AB}$ is rotated so that B moves to the right on the E_{n_2} curve and A is the point on the E_{n_1} curve such that $\overline{AB}$ passes through $(\tilde{\mu}_1, \tilde{\mu}_2)$. The analysis proceeds as in Case I, except the roles of $(\mu_{1(1)}, \mu_{2(1)}, \mu_{3(1)})$ and $(\mu_{1(2)}, \mu_{2(2)}, \mu_{3(2)})$ are reversed.

The analysis of Cases I and II proves the second statement of the theorem. This result implies that $R(\mathcal{ME}_{n_2}) \subset R(\mathcal{ME}_{\vec{n}})$. To see that $R(\mathcal{ME}_{\vec{n}}) \subset R(\mathcal{ME}_{n_2})$, suppose $(c - 1/c, \gamma)$ corresponds to a distribution, F, in $\mathcal{ME}_{\vec{n}}$. If F is an ME_{n_2}, then trivially $(c - 1/c, \gamma) \in R(\mathcal{ME}_{n_2})$. If F is an ME_{n_1}, then again $(c - 1/c, \gamma) \in R(\mathcal{ME}_{n_2})$, since $R(\mathcal{ME}_{n_1}) \subset R(\mathcal{ME}_{n_2})$. Finally, suppose F assigns positive probability to at least one E_{n_1} and at least one E_{n_2}. This distribution can be written as $F = p_1 F_1 + p_2 F_2$, where F_1 is an ME_{n_1} and F_2 is an ME_{n_2}. Since $R(\mathcal{ME}_{n_1}) \subset R(\mathcal{ME}_{n_2})$, there exists an ME_{n_2}, say F_3, with the same first three moments as F_1. So, $p_1 F_3 + p_2 F_2$, which is an ME_{n_2}, has the same first three moments as F. Thus, any point $(c - 1/c, \gamma)$ that corresponds to an $ME_{\vec{n}}$ also corresponds to an ME_{n_2}, i.e., $R(\mathcal{ME}_{\vec{n}}) \subset R(\mathcal{ME}_{n_2})$. □

Proof of Theorem 2

Notation in this proof is as defined in the proof of Theorem 1. In Case II of the proof of Theorem 1, $(\tilde{\mu}_1, \tilde{\mu}_2)$ can be matched by a DE_{n_1} as well as the DE_{n_2} represented by the first position of $\overline{AB}$ in Figure 5. The DE_{n_1} can be represented on the (μ_1, μ_2)-plane by a segment, $\overline{CD}$, through $(\tilde{\mu}_1, \tilde{\mu}_2)$ with left endpoint, C, at (0, 0) and right endpoint, D, on the E_{n_1} curve. Suppose CD is rotated so that D moves to the right on the E_{n_1} curve and C is the point on the E_{n_2} curve such that $\overline{CD}$ passes through $(\tilde{\mu}_1, \tilde{\mu}_2)$. As $\mu_{1(1)} \to \infty$, the $ME_{\vec{n}}(1, 1)$'s represented by $\overline{CD}$ attain all values of $\tilde{\mu}_3$ above the μ_3 lower bound in $M_3(\mathcal{ME}_{n_1})$. In terms of the $(c - 1/c, \gamma)$-plane, $\overline{CD}$ attains all values of $\tilde{\gamma}$ above the DE_{n_1} curve at $c - 1/c = \tilde{c} - 1/\tilde{c}$. Thus, if $(\tilde{c} - 1/\tilde{c}, \tilde{\gamma})$ is in the interior of$R(\mathcal{ME}_{n_1})$, a distribution represented by a position of $\overline{AB}$ and another represented by a position of $\overline{CD}$ both match $(\tilde{\mu}_1, \tilde{\mu}_2, \tilde{\mu}_3)$. The distributions represented by $\overline{AB}$ satisfy $\mu_{1(1)} \le \tilde{\mu}_1 \le \mu_{1(2)}$, and the distributions represented by $\overline{CD}$ satisfy $\mu_{1(2)} \le \tilde{\mu}_1 \le \mu_{1(1)}$. (This can be shown analytically or observed from the rotation of $\overline{AB}$ and $\overline{CD}$ on the (μ_1, μ_2)-plane.) □

Motivation for Conjecture 1

Here we explain how the proofs of Theorems 1 and 2 motivate Conjecture 1. We show that the validity of Conjecture 1 rests on monotonicity assumptions which are plausible though not yet proved. Notation is as defined in the proof of Theorem 1.

In Case I of the proof of Theorem 1, each $ME_{\vec{n}}(1, 1)$ that corresponds to $(\tilde{\mu}_1, \tilde{\mu}_2, \tilde{\mu}_3)$ is represented by some position of $\overline{AB}$. So, *if* the change in $\tilde{\mu}_3$ is strictly monotone as $\overline{AB}$ is rotated, each $(\tilde{\mu}_1, \tilde{\mu}_2, \tilde{\mu}_3)$ such that $(c - 1/c, \gamma) \in R(\mathcal{ME}_{n_2})$ and $(c - 1/c, \gamma)$ is left of $R(\mathcal{ME}_{n_1})$ corresponds to exactly one $ME_{\vec{n}}(1, 1)$.

We divide Case II into two sub-cases.

Case IIa: Let $\tilde{\mu}_3$ be on or below the μ_3 bound in $M_3(\mathcal{ME}_{n_1})$. In terms of the $(c - 1/c, \gamma)$-plane, $(\tilde{\mu}_1, \tilde{\mu}_2, \tilde{\mu}_3)$ corresponds to a point $(\tilde{c} - 1/\tilde{c}, \tilde{\gamma})$ on or below the DE_{n_1} curve and above the DE_{n_2} curve and to the right of the $c - 1/c$ boundary of $R(\mathcal{ME}_{n_1})$. Each $ME_{\vec{n}}(1, 1)$ that matches a triple $(\tilde{\mu}_1, \tilde{\mu}_2, \tilde{\mu}_3)$ in this subcase is represented by a position of $\overline{AB}$ as defined for Case II. Thus, *if* the change in $\tilde{\mu}_3$ is monotone as $\overline{AB}$ is rotated, there is exactly one $ME_{\vec{n}}(1, 1)$, represented by a position of $\overline{AB}$, that matches it.

Case IIb: Let $\tilde{\mu}_3$ be above the μ_3 bound in $M_3(\mathcal{ME}_{n_1})$. In terms of the $(c - 1/c, \gamma)$-plane, $(\tilde{\mu}_1, \tilde{\mu}_2, \tilde{\mu}_3)$ corresponds to a point $(\tilde{c} - 1/\tilde{c}, \tilde{\gamma})$ in the interior

of $R(\mathcal{ME}_{n_1})$. Each $ME_{\vec{n}}(1, 1)$ that matches a triple $(\tilde{\mu}_1, \tilde{\mu}_2, \tilde{\mu}_3)$ in this subcase is represented by a position of $\overline{AB}$ or $\overline{CD}$ as defined for Case II. Thus, *if* the change in $\tilde{\mu}_3$ is monotone as $\overline{AB}$ or $\overline{CD}$ is rotated, there are exactly two $ME_{\vec{n}}(1, 1)$'s, one represented by a position of $\overline{AB}$ and the other by a position of $\overline{CD}$, that match it. □

Mary A. Johnson and Michael R. Taaffe

Matching moments with a class of phase distributions: Nonlinear programming approaches

Received: 10/10/1988

Revised: 9/27/1989

Accepted: 10/20/1989

Recommended by Brad Makrucki, Editor