Adv. Appl. Prob. **39,** 271–292 (2007)
Printed in Northern Ireland

ON MATRIX EXPONENTIAL DISTRIBUTIONS

QI-MING HE,* *Dalhousie University*
HANQIN ZHANG,** *Chinese Academy of Sciences*

Abstract

In this paper we introduce certain Hankel matrices that can be used to study ME (matrix exponential) distributions, in particular to compute their ME orders. The Hankel matrices for a given ME probability distribution can be constructed if one of the following five types of information about the distribution is available: (i) an ME representation, (ii) its moments, (iii) the derivatives of its distribution function, (iv) its Laplace–Stieltjes transform, or (v) its distribution function. Using the Hankel matrices, a necessary and sufficient condition for a probability distribution to be an ME distribution is found and a method of computing the ME order of the ME distribution developed. Implications for the PH (phase-type) order of PH distributions are examined. The relationship between the ME order, the PH order, and a lower bound on the PH order given by Aldous and Shepp (1987) is discussed in numerical examples.

Keywords: Hankel matrix; matrix exponential distribution; phase-type distribution; matrix-analytic methods

2000 Mathematics Subject Classification: Primary 60A99
Secondary 15A18

1. Introduction

The ME (matrix exponential) distribution, as a generalization of the PH (phase-type) distribution [23], [24], was studied in [3] and [5]. For every ME distribution there exist many matrix representations (called ME representations). Of the ME representations of an ME distribution, naturally some will be of minimal order; this order is called the ME order of the ME distribution. In this paper we introduce certain Hankel matrices that can be used to study ME distributions and to compute their ME orders.

ME distributions and ME representations deserve attention from researchers for a number of reasons. First, ME distributions are useful in the analysis of stochastic models and, as was demonstrated in [3] and [21], can be used in the analysis of renewal processes and queueing systems. The study of ME representations of ME distributions makes it easier to use such distributions in stochastic modeling, financial engineering, and insurance and risk analysis. Second, the ME representation problem is closely related to the representation problem of positive linear systems [4], [22]. Some of the results obtained in this paper can be extended to that area. Third, the class of ME distributions includes all PH distributions and all Coxian distributions. To a certain extent, the study of PH distributions and Coxian distributions depends on results related to ME distributions, as demonstrated in [12] and [13] (and in this paper).

Received 16 November 2005; revision received 29 September 2006.
* Postal address: Department of Industrial Engineering, Dalhousie University, Halifax, NS B3J 2X4, Canada.
Email address: qi-ming.he@dal.ca
** Postal address: Institute of Applied Mathematics, Chinese Academy of Sciences, 100080 Beijing, P. R. China.
Email address: hanqin@amt.ac.cn

PH distributions have been widely used in stochastic modeling [2], [20], [24], [25]–[28]. Furthermore, it was shown in [7] that the minimal PH representation problem is equivalent to the positive realization problem in stochastic control theory. Thus, a better understanding of ME distributions can further the study of some stochastic models, the realization problem in stochastic control theory, and the two abovementioned special types of distribution (PH distributions and Coxian distributions).

The literature on ME distributions is limited. Asmussen and Bladt [3] studied ME distributions and applied them to the study of a class of queueing systems. They identified some necessary and sufficient conditions for an ME representation to be minimal and developed a method for computing a minimal ME representation from an ME representation. Bladt and Neuts [5] studied the class of ME distributions and their related ME renewal processes through a randomly stopped deterministic flow model. Fackrell [9], [10] studied the fitting of ME distributions and characteristics of ME distributions under the provision that the Laplace–Stieltjes transform of the original distribution is known. In [12] and [13], ME distributions were used in the study of PH distributions. Some results of [14] and [30] were related to the ME order of ME distributions. In [31], a method was developed for approximating ME distributions.

One of the important issues in the application of ME distributions is computational efficiency. In order to improve the efficiency of algorithms involving ME distributions (or PH distributions), minimal ME representations are desired. The ME order can play an important role in finding minimal ME representations. In this paper we introduce certain Hankel matrices that can be used to compute the ME order of ME distributions. The construction of these Hankel matrices is based on one of the following five types of information about the ME distribution: (i) an ME representation, (ii) its moments, (iii) derivatives of its distribution function, (iv) its Laplace–Stieltjes transform, or (v) its distribution function.

In [3], the ME order of an ME distribution was found by using its Laplace–Stieltjes transform or its ME representations. If the Laplace–Stieltjes transform, which is a rational function such that the numerator and denominator are coprime polynomials, is known, then the degree of the denominator gives the ME order. In comparison with that method, the method introduced here does not require one to find the form of the Laplace–Stieltjes transform. If an ME representation is known, the ME order can be found by computing the dimension of some subspace generated from the matrix and vectors of the representation [3], [14], which is a straightforward and efficient technique. In this case, for our method to work we need to calculate the moments or derivatives before using the Hankel matrices to find the ME order. To the authors' knowledge, only ME representations and Laplace–Stieltjes transforms have been used to find ME orders and minimal ME representations of ME distributions directly. Thus, the use of Hankel matrices (constructed from one of the five abovementioned types of information) to find the ME order is new.

Furthermore, using the Hankel matrices, in this paper we establish some relationships between the Laplace–Stieltjes transforms, the distribution functions, and the minimal ME representations of ME distributions. A necessary and sufficient condition for a probability distribution function to be an ME distribution function is found. Since a PH distribution is also an ME distribution, the ME order of a PH distribution calculated using the Hankel matrices provides a lower bound on its PH order. Thus, we demonstrate that the Hankel matrices can play an important role in the study of ME distributions and PH distributions.

The method used in this paper is related to identifiability of functions of finite Markov chains [11], [14], [30]. Although ME representations in general are not associated with hidden Markov chains, some of the ideas used in the study of the identifiability problem are useful. The method used in this paper is also related to the Chebyshev systems that have been widely used in function approximations [15], [16], [17], [18]. Johnson and Taaffe [16] used the Chebyshev systems to develop algorithms for computing PH distributions that fit the moments of the original distributions up to a fixed level. Our method gives the ME order (the minimal order) for a possible ME fit in distribution but does not actually compute an ME representation.

The remainder of the paper is organized as follows. In Section 2 the ME distribution and ME representation are mathematically defined, and two basic properties of ME distributions presented for later use. In Section 3 we introduce a certain Hankel matrix and establish a relationship between the Hankel matrix and the ME order. A method of computing the ME order is developed. In Section 4 the results in Section 3 are generalized to three other Hankel matrices. A necessary and sufficient condition for a probability distribution to be an ME distribution is obtained. In Section 5 we present an application to PH distributions. Some results on PH simplicity and PH fullness are obtained. In Section 6 some numerical examples are presented to provide insight into the relationship between the ME order, the PH order, and the lower bound on the PH order given in [1]. Finally, some technical details are collected in two appendices.

2. ME distributions and ME representations

The ME distribution was studied in [3]. A probability distribution function $F(t)$ is called an ME distribution if there exist a dimension-m row vector $\boldsymbol{\alpha}$, an $m \times m$ matrix $\boldsymbol{T}$, and a dimension-m column vector $\boldsymbol{u}$ such that

$$F(t) = 1 - \boldsymbol{\alpha} \exp\{\boldsymbol{T} t\}\boldsymbol{u}, \qquad t \geq 0. \tag{2.1}$$

The triple $(\boldsymbol{\alpha}, \boldsymbol{T}, \boldsymbol{u})$ is called an ME representation of the ME distribution $F(t)$. There is no restriction on the elements of $\boldsymbol{\alpha}$, $\boldsymbol{T}$, and $\boldsymbol{u}$ except that the right-hand side of (2.1) be a probability distribution. The integer m is the order of the ME representation $(\boldsymbol{\alpha}, \boldsymbol{T}, \boldsymbol{u})$.

The ME representation of an ME distribution is not unique. An ME representation of minimal order is called a minimal ME representation, and the ME order of an ME distribution is defined to be the order of (any of) its minimal ME representations. We point out that, since

$$F(t) = 1 - \boldsymbol{\alpha} \exp\{\boldsymbol{T} t\}\boldsymbol{u} = 1 - \boldsymbol{\alpha} \exp\{(-\lambda \boldsymbol{I} + \boldsymbol{T})t\}\boldsymbol{u}\mathrm{e}^{\lambda t}, \qquad t \geq 0,$$

where $\boldsymbol{I}$ is the $m \times m$ identity matrix and λ is a positive real number, finding a minimal version of $(\boldsymbol{\alpha}, \boldsymbol{T}, \boldsymbol{u})$ is equivalent to finding a minimal version of $(\boldsymbol{\alpha}, -\lambda\boldsymbol{I} + \boldsymbol{T}, \boldsymbol{u})$. Since we can always choose λ such that all eigenvalues of the matrix $-\lambda\boldsymbol{I} + \boldsymbol{T}$ are nonzero, we assume that the matrix $\boldsymbol{T}$ is invertible throughout this paper.

Next we present two basic properties of ME distributions which shall be repeatedly used in the paper.

Proposition 2.1. *Assume that a random variable X has an ME distribution with an ME representation $(\boldsymbol{\alpha}, \boldsymbol{T}, \boldsymbol{u})$ and a distribution function $F(t)$ of the form (2.1). Then*

$$\boldsymbol{\alpha}\boldsymbol{T}^k\boldsymbol{u} = \begin{cases} -F^{(k)}(0) & \text{for } k \geq 1, \\ 1 - F(0) & \text{for } k = 0, \\ (-1)^k \,\mathrm{E}[X^{-k}]/(-k)! & \text{for } k \leq -1, \end{cases} \tag{2.2}$$

where $F^{(k)}(t)$ denotes the kth derivative of the function $F(t)$.

https://doi.org/10.1239/aap/1175266478 Published online by Cambridge University Press

Proof. A proof of (2.2) can be found in [3]. For completeness, we include the following simple proof. By taking derivatives on both sides of (2.1) and letting $t = 0$, we obtain (2.2) for $k \geq 0$. By integrating on both sides of (2.1), we obtain

$$\mathrm{E}[X^k] = \int_0^\infty t^k \,\mathrm{d}F(t) = (-1)^k k!\,\boldsymbol{\alpha}\boldsymbol{T}^{-k}\boldsymbol{u}, \qquad k \geq 1,$$

which yields the last part of (2.2). This completes the proof.

Next we consider an ME distribution with two ME representations, $(\boldsymbol{\alpha}, \boldsymbol{T}, \boldsymbol{u})$ and $(\boldsymbol{\beta}, \boldsymbol{S}, \boldsymbol{v})$.

Proposition 2.2. *The ME representations* $(\boldsymbol{\alpha}, \boldsymbol{T}, \boldsymbol{u})$ *and* $(\boldsymbol{\beta}, \boldsymbol{S}, \boldsymbol{v})$ *represent the same probability distribution if and only if* $\boldsymbol{\alpha}\boldsymbol{T}^k\boldsymbol{u} = \boldsymbol{\beta}\boldsymbol{S}^k\boldsymbol{v}$, $-\infty < k < \infty$.

Proof. If the two ME representations $(\boldsymbol{\alpha}, \boldsymbol{T}, \boldsymbol{u})$ and $(\boldsymbol{\beta}, \boldsymbol{S}, \boldsymbol{v})$ represent the same probability distribution, then, by Proposition 2.1, we must have $\boldsymbol{\alpha}\boldsymbol{T}^k\boldsymbol{u} = \boldsymbol{\beta}\boldsymbol{S}^k\boldsymbol{v}$, $-\infty < k < \infty$.

Now assume that $\boldsymbol{\alpha}\boldsymbol{T}^k\boldsymbol{u} = \boldsymbol{\beta}\boldsymbol{S}^k\boldsymbol{v}$, $-\infty < k < \infty$. By (2.1) and the Taylor expansion of the distribution function $F(t)$, we have

$$F(t) = 1 - \sum_{k=0}^{\infty} \frac{t^k \boldsymbol{\alpha}\boldsymbol{T}^k\boldsymbol{u}}{k!} = 1 - \sum_{k=0}^{\infty} \frac{t^k \boldsymbol{\beta}\boldsymbol{S}^k\boldsymbol{v}}{k!}, \qquad t \geq 0.$$

Therefore, $(\boldsymbol{\alpha}, \boldsymbol{T}, \boldsymbol{u})$ and $(\boldsymbol{\beta}, \boldsymbol{S}, \boldsymbol{v})$ represent the same probability distribution. This completes the proof.

Remark 2.1. The condition in Proposition 2.2 can be changed to either $\boldsymbol{\alpha}\boldsymbol{T}^k\boldsymbol{u} = \boldsymbol{\beta}\boldsymbol{S}^k\boldsymbol{v}$, $0 \leq k < \infty$, or $\boldsymbol{\alpha}\boldsymbol{T}^k\boldsymbol{u} = \boldsymbol{\beta}\boldsymbol{S}^k\boldsymbol{v}$, $-\infty < k \leq 0$.

3. ME order and Hankel matrices

Consider a random variable X associated with a distribution function $F(t)$ and an order-m ME representation $(\boldsymbol{\alpha}, \boldsymbol{T}, \boldsymbol{u})$. For $-\infty < k < \infty$ and $n \geq 1$, define an $n \times n$ matrix

$$\boldsymbol{\Lambda}_n^{[k]} = (a_{i,j}^{[k]}) \qquad \text{with} \qquad a_{i,j}^{[k]} = \boldsymbol{\alpha}\boldsymbol{T}^{i+j-2+k}\boldsymbol{u}, \quad 1 \leq i, j \leq n. \tag{3.1}$$

Since the matrix $\boldsymbol{T}$ is assumed to be invertible, the matrix $\boldsymbol{\Lambda}_n^{[k]}$ is well defined for all integers k. The matrix $\boldsymbol{\Lambda}_n^{[k]}$ is symmetric and is a Hankel matrix [19]. For $n = 3$ and $k = -2$, we have

$$\boldsymbol{\Lambda}_3^{[-2]} = \begin{pmatrix} \boldsymbol{\alpha}\boldsymbol{T}^{-2}\boldsymbol{u} & \boldsymbol{\alpha}\boldsymbol{T}^{-1}\boldsymbol{u} & \boldsymbol{\alpha}\boldsymbol{u} \\ \boldsymbol{\alpha}\boldsymbol{T}^{-1}\boldsymbol{u} & \boldsymbol{\alpha}\boldsymbol{u} & \boldsymbol{\alpha}\boldsymbol{T}\boldsymbol{u} \\ \boldsymbol{\alpha}\boldsymbol{u} & \boldsymbol{\alpha}\boldsymbol{T}\boldsymbol{u} & \boldsymbol{\alpha}\boldsymbol{T}^2\boldsymbol{u} \end{pmatrix}.$$

An important observation is that, by Proposition 2.1, the definition of the matrix $\boldsymbol{\Lambda}_n^{[k]}$ is independent of the specific ME representation used in (3.1) but depends on the moments and the derivatives of the distribution function at 0.

Next we show that the ME order of the ME distribution can be found by using the determinants of the matrices $\boldsymbol{\Lambda}_n^{[k]}$, $-\infty < k < \infty$, $n \geq 1$.

Lemma 3.1. *Let* $(\boldsymbol{\alpha}, \boldsymbol{T}, \boldsymbol{u})$ *be an ME representation of order* m*. For positive integers* n*, we have*

$$\boldsymbol{\Lambda}_n^{[k]} = \boldsymbol{L}(n)\boldsymbol{T}^k\boldsymbol{R}(n), \qquad -\infty < k < \infty,$$

where $\boldsymbol{L}(n)$ is an $n \times m$ matrix and $\boldsymbol{R}(n)$ an $m \times n$ matrix respectively given by

$$\boldsymbol{L}(n) = \begin{pmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\alpha T} \\ \vdots \\ \boldsymbol{\alpha T}^{n-1} \end{pmatrix} \quad \textit{and} \quad \boldsymbol{R}(n) = (\boldsymbol{u}, \boldsymbol{Tu}, \ldots, \boldsymbol{T}^{n-1}\boldsymbol{u}). \tag{3.2}$$

Proof. Lemma 3.1 follows from straightforward calculations using matrix multiplication.

From Lemma 3.1, the following result on the determinant of the matrix $\boldsymbol{\Lambda}_n^{[k]}$ can be obtained.

Lemma 3.2. *For an order-m ME representation $(\boldsymbol{\alpha}, \boldsymbol{T}, \boldsymbol{u})$, we have $\det(\boldsymbol{\Lambda}_n^{[k]}) = 0$ for $-\infty < k < \infty$ and $n \geq m+1$, where $\det(\boldsymbol{\Lambda}_n^{[k]})$ denotes the determinant of the matrix $\boldsymbol{\Lambda}_n^{[k]}$.*

Proof. By the Binet–Cauchy formula (see [19]), the rank of the matrix $\boldsymbol{\Lambda}_n^{[k]}$ is less than or equal to the smallest respective rank of the matrices $\boldsymbol{L}(n)$, $\boldsymbol{T}$, and $\boldsymbol{R}(n)$. Therefore, the rank of $\boldsymbol{\Lambda}_n^{[k]}$ is less than or equal to m, since the ranks of $\boldsymbol{L}(n)$, $\boldsymbol{T}$, and $\boldsymbol{R}(n)$ are all less than or equal to m. Consequently, $\det(\boldsymbol{\Lambda}_n^{[k]}) = 0$ for $n \geq m+1$. This completes the proof.

Lemma 3.3. *For an order-m ME representation $(\boldsymbol{\alpha}, \boldsymbol{T}, \boldsymbol{u})$, we have either $\det(\boldsymbol{\Lambda}_m^{[k]}) = 0$ for all k or $\det(\boldsymbol{\Lambda}_m^{[k]}) \neq 0$ for all k.*

Proof. In this case, $\boldsymbol{L}(m)$ and $\boldsymbol{R}(m)$ are square matrices of order m. Thus,

$$\det(\boldsymbol{\Lambda}_m^{[k]}) = \det(\boldsymbol{L}(m)) \det(\boldsymbol{T})^k \det(\boldsymbol{R}(m)) = \det(\boldsymbol{T})^k \det(\boldsymbol{\Lambda}_m^{[0]}). \tag{3.3}$$

Since $\det(\boldsymbol{T}) \neq 0$, $\det(\boldsymbol{\Lambda}_m^{[k]}) = 0$ if and only if $\det(\boldsymbol{L}(m))$ and/or $\det(\boldsymbol{R}(m))$ is 0, which leads to the conclusion. This completes the proof.

We are now ready to state and prove the first main result of this paper.

Theorem 3.1. *Consider a random variable X that has an ME distribution. The ME order of X can be obtained as*

$$\max\{n\colon\ \max_{-\infty<k<\infty}\{|\det(\boldsymbol{\Lambda}_n^{[k]})|\} \neq 0,\ n \geq 1\}, \tag{3.4}$$

$$\max\{n\colon\ \det(\boldsymbol{\Lambda}_n^{[2-2n]}) \neq 0,\ n \geq 1\}, \quad \textit{or} \tag{3.5}$$

$$\max\{n\colon\ \det(\boldsymbol{\Lambda}_n^{[k]}) \neq 0,\ n \geq 1\} \quad \textit{for any given integer } k. \tag{3.6}$$

Proof. Note that the ME order of an ME distribution is finite. First we prove (3.4). Suppose that the ME distribution of the random variable X has an order-m ME representation $(\boldsymbol{\alpha}, \boldsymbol{T}, \boldsymbol{u})$, and that this is minimal. By Theorem 2.1 of [3], $\{\boldsymbol{\alpha}, \boldsymbol{\alpha T}, \boldsymbol{\alpha T}^2, \ldots, \boldsymbol{\alpha T}^{m-1}\}$ and $\{\boldsymbol{u}, \boldsymbol{Tu}, \boldsymbol{T}^2\boldsymbol{u}, \ldots, \boldsymbol{T}^{m-1}\boldsymbol{u}\}$ are two sets of linearly independent vectors. This implies that the matrices $\boldsymbol{L}(m)$ and $\boldsymbol{R}(m)$ defined in (3.2) are invertible. By (3.3), the matrix $\boldsymbol{\Lambda}_m^{[0]}$ is invertible. Therefore, the quantity given in (3.4) is greater than or equal to m (the ME order of X). In fact, for any k, since T is invertible $\boldsymbol{\Lambda}_m^{[k]} = \boldsymbol{L}(m)\boldsymbol{T}^k\boldsymbol{R}(m)$ is invertible if m is the ME order of X. Thus, the ME order of X is less than or equal to the quantity given in (3.4). On the other hand, if m is the ME order of X then it follows from Lemma 3.2 that $\det(\boldsymbol{\Lambda}_n^{[k]}) = 0$ for $-\infty < k < \infty$ and $n \geq m+1$. Therefore, the quantity given in (3.4) is less than or equal to the ME order of X. Combining these two arguments, we conclude that the quantity given in (3.4) equals the ME order of X.

Equations (3.5) and (3.6) are consequences of Lemmas 3.2 and 3.3 and (3.4). This completes the proof.

An immediate consequence of Theorem 3.1 is that, for any k, $\det(\mathbf{\Lambda}_n^{[k]}) = 0$ if n is greater than the ME order of the ME distribution. Furthermore, Theorem 3.1 leads to the following method of computing the ME order of an ME distribution.

(a) If an ME representation of order m is available then we can use either (3.5) or (3.6), for $1 \le n \le m$, to find the ME order.

(b) If the moments of the ME distribution are available, (3.5) can be used to compute the ME order. An upper limit for n must be chosen, to decide where to terminate the computation process. If the Laplace–Stieltjes transform of the ME distribution is available, we first calculate the moments of the ME distribution and then use (3.5) to find the ME order.

(c) If the derivatives of the distribution function at 0 are available, (3.6) can be used to compute the ME order. Again, an upper limit for n must be chosen, to decide where to terminate the computation process.

Remark 3.1. The matrix $\mathbf{\Lambda}_n^{[k]}$ defined in (3.1) is a special case of some matrices introduced in [11] in relation to the identifiability problem for functions of finite Markov chains. Although ME representations are not directly related to hidden Markov chains, Theorem 3.1 shows that the Hankel matrix can still be useful.

Remark 3.2. In [15], a matrix similar to the one defined in (3.1) for $k = -2n + 2$ was introduced for parameter estimation and the fitting of PH distributions. It was assumed that the first n moments of a probability distribution are known, and a method for finding a PH distribution having the same first n moments as that of the original probability distribution was developed.

The matrix $\mathbf{\Lambda}_n^{[k]}$ defined in (3.1) can be ill conditioned: the determinant of $\mathbf{\Lambda}_n^{[k]}$ can be either very large or very small, which may lead to computation instability. To deal with this problem, we present an extension of $\mathbf{\Lambda}_n^{[k]}$. This extension does not have an explicit probabilistic interpretation, but it can be useful in computations. Let $\boldsymbol{c} = (c_1, c_2, \dots, c_n)$ and $\boldsymbol{d} = (d_1, d_2, \dots, d_n)$. For an ME representation $(\boldsymbol{\alpha}, \boldsymbol{T}, \boldsymbol{u})$, define, for $n \ge 1$,

$$\hat{\mathbf{\Lambda}}_n^{[k]}(g, \boldsymbol{c}, \boldsymbol{d}) = (\hat{a}_{i,j}^{[k]}) \qquad \text{with} \qquad \hat{a}_{i,j}^{[k]} = c_i d_j g^k \boldsymbol{\alpha} \boldsymbol{T}^{i+j-2+k} \boldsymbol{u}, \quad 1 \le i, j \le n, \tag{3.7}$$

where g is a constant. By choosing $\{g, \boldsymbol{c}, \boldsymbol{d}\}$ appropriately, it is possible to overcome the numerical difficulty associated with $\mathbf{\Lambda}_n^{[k]}$. It is easy to see that the definition in (3.7) is independent of the specific ME representation of the ME distribution used there.

Proposition 3.1. *For the matrices defined respectively in (3.1) and (3.7), we have*

$$\hat{a}_{i,j}^{[k]} = c_i d_j g^k a_{i,j}^{[k]}, \qquad 1 \le i, j \le n, \tag{3.8}$$

$$\det(\hat{\mathbf{\Lambda}}_n^{[k]}(g, \boldsymbol{c}, \boldsymbol{d})) = g^{nk} \left(\prod_{i=1}^{n} c_i d_i \right) \det(\mathbf{\Lambda}_n^{[k]}), \qquad -\infty < k < \infty,\ n \ge 1. \tag{3.9}$$

If g and all elements of the vectors $\boldsymbol{c}$ and $\boldsymbol{d}$ are nonzero, then $\det(\mathbf{\Lambda}_n^{[k]}) = 0$ if and only if $\det(\hat{\mathbf{\Lambda}}_n^{[k]}(g, \boldsymbol{c}, \boldsymbol{d})) = 0$, for $n \ge 1$ and $-\infty < k < \infty$. Thus, the matrices $\hat{\mathbf{\Lambda}}_n^{[k]}(g, \boldsymbol{c}, \boldsymbol{d})$, $-\infty < k < \infty$, $n \ge 1$, can be used to determine the ME order.

Proof. Equation (3.8) is obvious from the definitions given in (3.1) and (3.7). To prove (3.9), we note that definition (3.7) is equivalent to

$$\hat{\boldsymbol{\Lambda}}_n^{[k]}(g, \boldsymbol{c}, \boldsymbol{d}) = \hat{\boldsymbol{L}}(\boldsymbol{c}, n)(g\boldsymbol{T})^k \hat{\boldsymbol{R}}(\boldsymbol{d}, n), \qquad -\infty < k < \infty, \tag{3.10}$$

where $\hat{\boldsymbol{L}}(\boldsymbol{c}, n)$ is an $n \times m$ matrix and $\hat{\boldsymbol{R}}(\boldsymbol{d}, n)$ an $m \times n$ matrix respectively given by

$$\hat{\boldsymbol{L}}(\boldsymbol{c}, n) = \begin{pmatrix} c_1\boldsymbol{\alpha} \\ c_2\boldsymbol{\alpha}\boldsymbol{T} \\ \vdots \\ c_n\boldsymbol{\alpha}\boldsymbol{T}^{n-1} \end{pmatrix} \quad \text{and} \quad \hat{\boldsymbol{R}}(\boldsymbol{d}, n) = (d_1\boldsymbol{u}, d_2\boldsymbol{T}\boldsymbol{u}, \dots, d_n\boldsymbol{T}^{n-1}\boldsymbol{u}). \tag{3.11}$$

Equation (3.9) is obtained from (3.10) and (3.11) using properties of the matrix determinant. This completes the proof.

Remark 3.3. If we choose $g = -1$ and $c_j = d_j = (-1)^j$, $1 \le j \le n$, then the matrix $\hat{\boldsymbol{\Lambda}}_n^{[k]}(g, \boldsymbol{c}, \boldsymbol{d})$ defined in (3.7) is nonnegative for $k \le -2n + 2$. This nonnegativity property can be useful in analysis.

4. Three generalizations

Theorem 3.1 established a relationship between the ME order, the moments, and the derivatives of the distribution function at 0. In this section we give three generalizations of Theorem 3.1.

4.1. ME order and Laplace–Stieltjes transforms

In this subsection Theorem 3.1 is generalized to depend on the Laplace–Stieltjes transform of the ME distribution. Assume that the ME distribution has an ME representation $(\boldsymbol{\alpha}, \boldsymbol{T}, \boldsymbol{u})$. Define an $n \times n$ matrix

$$\boldsymbol{\Lambda}_n^{*[k]}(s) = (a_{i,j}^{*[k]}(s)) \qquad \text{with} \qquad a_{i,j}^{*[k]}(s) = \boldsymbol{\alpha}(\boldsymbol{T} - s\boldsymbol{I})^{i+j-2+k}\boldsymbol{T}\boldsymbol{u}, \quad 1 \le i, j \le n, \tag{4.1}$$

where s is not an eigenvalue of $\boldsymbol{T}$. By definition, it is easy to see that $\boldsymbol{\Lambda}_n^{[k]} = \boldsymbol{\Lambda}_n^{*[k-1]}(0)$. Thus, the Hankel matrix defined in (3.1) is a special case of the one defined in (4.1).

If $k < -(i + j) + 1$ then

$$a_{i,j}^{*[k]}(s) = \frac{1}{(1 - k - i - j)!} f^{*(1-k-i-j)}(s),$$

where $f^*(s)$ is the Laplace–Stieltjes transform of the ME distribution, $f^{*(1-k-i-j)}(s)$ is the $(1 - k - i - j)$th derivative of $f^*(s)$, and s is not a pole of $f^*(s)$. If an ME representation $(\boldsymbol{\alpha}, \boldsymbol{T}, \boldsymbol{u})$ is known, then $f^*(s) = 1 - \boldsymbol{\alpha}\boldsymbol{u} - \boldsymbol{\alpha}(s\boldsymbol{I} - \boldsymbol{T})^{-1}\boldsymbol{T}\boldsymbol{u}$. If $k = -(i + j) + 1$ then $a_{i,j}^{*[k]}(s) = f^*(s) - f^*(\infty)$, and if $k > -(i + j) + 1$ then, by Proposition 2.1, $a_{i,j}^{*[k]}(s)$ can be expressed in terms of the derivatives of the distribution function at 0. Therefore, the matrix $\boldsymbol{\Lambda}_n^{*[k]}(s)$ is independent of the specific ME representation used in (4.1), but depends on the derivatives of the distribution function at 0 and the derivatives of the Laplace–Stieltjes transform of the ME distribution. Furthermore, if $k < -2n + 1$ then the matrix $\boldsymbol{\Lambda}_n^{*[k]}(s)$ is defined completely by the derivatives of the Laplace–Stieltjes transform of the ME distribution. From definition (4.1), if an ME representation $(\boldsymbol{\alpha}, \boldsymbol{T}, \boldsymbol{u})$ is available then it is easy to show that

$$\boldsymbol{\Lambda}_n^{*[k]}(s) = \boldsymbol{L}^*(n, s)(\boldsymbol{T} - s\boldsymbol{I})^k \boldsymbol{R}^*(n, s), \tag{4.2}$$

https://doi.org/10.1239/aap/1175266478 Published online by Cambridge University Press

where

$$L^*(n,s)=\begin{pmatrix}\boldsymbol{\alpha}\\ \boldsymbol{\alpha}(\boldsymbol{T}-s\boldsymbol{I})\\ \vdots\\ \boldsymbol{\alpha}(\boldsymbol{T}-s\boldsymbol{I})^{n-1}\end{pmatrix}\quad\text{and}\quad \boldsymbol{R}^*(n,s)=(\boldsymbol{T}\boldsymbol{u},(\boldsymbol{T}-s\boldsymbol{I})\boldsymbol{T}\boldsymbol{u},\dots,(\boldsymbol{T}-s\boldsymbol{I})^{n-1}\boldsymbol{T}\boldsymbol{u}). \tag{4.3}$$

Theorem 4.1. *The ME order of an ME distribution can be obtained as*

$$\max\{n\colon\ \max_{-\infty<k<\infty}\{|\det(\boldsymbol{\Lambda}_n^{*[k]}(s))|\}\neq 0,\ n\geq 1\}\quad \textit{or}$$
$$\max\{n\colon\ \det(\boldsymbol{\Lambda}_n^{*[-2n+2]}(s))\neq 0,\ n\geq 1\},$$

provided that s is not a pole of the Laplace–Stieltjes transform of the ME distribution.

Proof. First note that the matrix $\boldsymbol{T}-s\boldsymbol{I}$ is invertible if s is not an eigenvalue of $\boldsymbol{T}$. Also note that the vectors $\boldsymbol{\alpha},\boldsymbol{\alpha}(\boldsymbol{T}-s\boldsymbol{I}),\boldsymbol{\alpha}(\boldsymbol{T}-s\boldsymbol{I})^2,\dots,\boldsymbol{\alpha}(\boldsymbol{T}-s\boldsymbol{I})^{n-1}$ are linearly independent if and only if the vectors $\boldsymbol{\alpha},\boldsymbol{\alpha}\boldsymbol{T},\boldsymbol{\alpha}\boldsymbol{T}^2,\dots,\boldsymbol{\alpha}\boldsymbol{T}^{n-1}$ are linearly independent. By (4.2) and (4.3), Theorem 4.1 can be proved in a way similar to Theorem 3.1.

An interesting consequence of Theorem 4.1 is that the characteristic polynomial of the matrix $\boldsymbol{T}$ in a minimal ME representation can be obtained from the Hankel matrices, even when the matrix $\boldsymbol{T}$ is unknown.

Corollary 4.1. *Assume that the ME order of an ME distribution is equal to m and that the Laplace–Stieltjes transform of the ME distribution is known. Then, for any integer k, we have*

$$\frac{\det(\boldsymbol{\Lambda}_m^{*[k]}(s))}{\det(\boldsymbol{\Lambda}_m^{*[k-1]}(s))}=\det(\boldsymbol{T}-s\boldsymbol{I}), \tag{4.4}$$

where s is not a pole of the Laplace–Stieltjes transform and $\boldsymbol{T}$ is the $m\times m$ matrix of a minimal ME representation.

Proof. If s is not a pole of the Laplace–Stieltjes transform then, by (4.2), for any integer k the function $\det(\boldsymbol{\Lambda}_m^{*[k-1]}(s))$ is well defined. Equation (4.4) is obtained from (4.2) by direct calculation:

$$\frac{\det(\boldsymbol{\Lambda}_m^{*[k]}(s))}{\det(\boldsymbol{\Lambda}_m^{*[k-1]}(s))}=\frac{\det(\boldsymbol{L}^*(m,s))\det((\boldsymbol{T}-s\boldsymbol{I})^k)\det(\boldsymbol{R}^*(m,s))}{\det(\boldsymbol{L}^*(m,s))\det((\boldsymbol{T}-s\boldsymbol{I})^{k-1})\det(\boldsymbol{R}^*(m,s))}=\det(\boldsymbol{T}-s\boldsymbol{I}).$$

This completes the proof.

Corollary 4.1 implies that all the matrices of the minimal ME representations of an ME distribution have the same characteristic polynomial.

Remark 4.1. In [3], the ME order of an ME distribution was found from its Laplace–Stieltjes transform by (i) finding an ME representation and (ii) identifying the dimension of the minimal space related to the ME representation. In comparison with that method, our method is straightforward.

4.2. ME order and derivatives of distribution functions

In this subsection Theorem 3.1 is generalized to depend on derivatives and integrals of distribution functions of ME distributions. Define

$$\mathbf{\Lambda}_n^{[k]}(t) = (a_{i,j}^{[k]}(t)) \qquad \text{with} \qquad a_{i,j}^{[k]}(t) = \boldsymbol{\alpha} \exp\{\boldsymbol{T}t\}\boldsymbol{T}^{i+j-2+k}\boldsymbol{u}, \quad 1 \le i, j \le n. \tag{4.5}$$

Since

$$\frac{\mathrm{d}}{\mathrm{d}t} a_{i,j}^{[k]}(t) = a_{i+1,j}^{[k]}(t) = a_{i,j+1}^{[k]}(t),$$

the determinant $\det(\mathbf{\Lambda}_n^{[k]}(t))$ is a Hankel determinant of the function $a_{1,1}^{[k]}(t)$ [17]. Since $\mathbf{\Lambda}_n^{[k]} = \mathbf{\Lambda}_n^{[k]}(0)$, the Hankel matrix $\mathbf{\Lambda}_n^{[k]}$ defined in (3.1) is a special case of the one defined in (4.5).

Note that $a_{i,j}^{[2-i-j]}(t) = 1 - F(t)$, where $F(t)$ is the distribution function. For $i+j+k-2 \ge 1$, we have

$$\frac{\mathrm{d}}{\mathrm{d}t} a_{i,j}^{[k-1]}(t) = a_{i,j}^{[k]}(t) = -F^{(i+j+k-2)}(t). \tag{4.6}$$

For $i+j+k-2 \le -1$, we have

$$\begin{aligned} a_{i,j}^{[k]}(t) &= -\int_t^\infty a_{i,j}^{[k+1]}(t_1)\,\mathrm{d}t_1 \\ &= (-1)^{2-i-j-k} \int_t^\infty \int_{t_1}^\infty \cdots \int_{t_{3-i-j-k}}^\infty (1 - F(t_{2-i-j-k}))\,\mathrm{d}t_1 \cdots \mathrm{d}t_{2-i-j-k}. \end{aligned} \tag{4.7}$$

In general, we have the following relationships between the Hankel matrices $\mathbf{\Lambda}_n^{[k]}(t)$:

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{\Lambda}_n^{[k]}(t) = \mathbf{\Lambda}_n^{[k+1]}(t) \quad \text{and} \quad \mathbf{\Lambda}_n^{[k]}(t) = -\int_t^\infty \mathbf{\Lambda}_n^{[k+1]}(t_1)\,\mathrm{d}t_1.$$

Equations (4.6) and (4.7) indicate that the matrix $\mathbf{\Lambda}_n^{(k)}(t)$ is independent of the specific ME representation used in (4.5), but depends on the derivatives and integrals of the distribution function $F(t)$. From the definition given in (4.5), if an ME representation $(\boldsymbol{\alpha}, \boldsymbol{T}, \boldsymbol{u})$ is available then it is easy to show that

$$\mathbf{\Lambda}_n^{[k]}(t) = \boldsymbol{L}(n)\exp\{\boldsymbol{T}t\}\boldsymbol{T}^k\boldsymbol{R}(n), \tag{4.8}$$

where $\boldsymbol{L}(n)$ and $\boldsymbol{R}(n)$ are as defined in (3.2).

Theorem 4.2. *The ME order of an ME distribution can be obtained, for any t, as*

$$\max\{n\colon \max_{-\infty<k<\infty}\{|\det(\mathbf{\Lambda}_n^{[k]}(t))|\} \ne 0,\ n \ge 1\} \quad \textit{or}$$
$$\max\{n\colon \det(\mathbf{\Lambda}_n^{[k]}(t)) \ne 0,\ n \ge 1\} \quad \textit{for any given integer } k.$$

Proof. Note that the matrix $\exp\{\boldsymbol{T}t\}$ is invertible for any real t. By (4.8), Theorem 4.2 can be proved in a way similar to Theorem 3.1.

Theorem 4.2 leads to a necessary and sufficient condition for a probability distribution function to be an ME distribution function, stated as follows.

Theorem 4.3. *A probability distribution function $F(t)$ is an ME distribution function if and only if the integer*

$$m = \max\{n\colon \max_{t\geq 0}\{|\det(\mathbf{\Lambda}_n^{[k]}(t))|\} \neq 0,\ n \geq 1\}$$

is finite for a given integer k, and m is then the ME order of the ME distribution. Furthermore, if $F(t)$ is an ME distribution function with ME order m then, for any integer k,

1. $\det(\mathbf{\Lambda}_n^{[k]}(t)) = 0$ *for all $t \geq 0$ if and only if $n \geq m+1$,*
2. *as a function of t, $\det(\mathbf{\Lambda}_m^{[k]}(t))$ is either always positive or always negative,*
3. $(1/t)\log(|\det(\mathbf{\Lambda}_n^{[k]}(t))/\det(\mathbf{\Lambda}_n^{[k]}(0))|)$ *is constant for $t \geq 0$ if and only if $n = m$ and, for $n = m$ and any $t > 0$,*

$$\sum_{i=1}^{m} \lambda_i = \sum_{i=1}^{m} t_{i,i} = \frac{1}{t}\log\left(\left|\frac{\det(\mathbf{\Lambda}_m^{[k]}(t))}{\det(\mathbf{\Lambda}_m^{[k]}(0))}\right|\right), \tag{4.9}$$

where $\{\lambda_1, \lambda_2, \ldots, \lambda_m\}$ and $\{t_{1,1}, t_{2,2}, \ldots, t_{m,m}\}$ are respectively the eigenvalues (counting multiplicities) and the diagonal elements of the matrix of any minimal ME representation.

Proof. Note that, implicitly, the function $F(t)$ is infinitely differentiable. Theorem 4.2 indicates that the integer

$$m = \max\{n\colon \max_{t\geq 0}\{|\det(\mathbf{\Lambda}_n^{[k]}(t))|\} \neq 0,\ n \geq 1\}$$

is finite for any ME distribution. Conversely, we shall show that if m is finite then $F(t)$ is an ME distribution with an ME representation of order m. First we show the result for the case in which $k = 0$. Here $\det(\mathbf{\Lambda}_{m+1}^{[0]}(t)) = 0$ for all $t \geq 0$. Recall that $\det(\mathbf{\Lambda}_{m+1}^{[0]}(t))$ is a Hankel determinant of $1-F(t)$. By [17, Theorem 7.1], $1 - F(t)$ is an exponential polynomial of order at most m (note that $a_{1,1}^{(0)}(t) = 1 - F(t)$), i.e.

$$1 - F(t) = \sum_{k=1}^{K} p_k(t)e^{\lambda_k t},$$

where $\lambda_i \neq \lambda_j$ if $i \neq j$, $p_1(t), p_2(t), \ldots, p_K(t)$ are polynomials of respective degrees $r_1 - 1$, $r_2 - 1, \ldots, r_K - 1$, and $r_1 + r_2 + \cdots + r_K = m$. Suppose that

$$p_k(t) = \sum_{j=0}^{r_k-1} p_{k,j}t^j, \qquad k = 1, 2, \ldots, K;$$

let $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_K)$, where, for $k = 1, 2, \ldots, K$,

$$\boldsymbol{\beta}_k = (p_{k,0} - p_{k,1}, p_{k,1} - 2!\, p_{k,2}, \ldots, \\ (r_k - 2)!\, p_{k,r_k-2} - (r_k - 1)!\, p_{k,r_k-1}, (r_k - 1)!\, p_{k,r_k-1});$$

and define the $m \times m$ matrix

$$\boldsymbol{S} = \begin{pmatrix} \boldsymbol{J}_1(\lambda_1) & & & \\ & \boldsymbol{J}_2(\lambda_2) & & \\ & & \ddots & \\ & & & \boldsymbol{J}_K(\lambda_K) \end{pmatrix},$$

https://doi.org/10.1239/aap/1175266478 Published online by Cambridge University Press

where the omitted entries all equal 0 and $\boldsymbol{J}_k(\lambda_k)$ are $r_k \times r_k$ matrices given by

$$\boldsymbol{J}_k(\lambda_k) = \begin{pmatrix} \lambda_k & 1 & & \\ & \lambda_k & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_k \end{pmatrix}, \qquad 1 \le k \le K.$$

Routine calculations verify that $1 - F(t) = \boldsymbol{\beta} \exp\{\boldsymbol{S}t\}\boldsymbol{e}$, where $\boldsymbol{e}$ is a column vector all of whose elements equal 1, which implies that $(\boldsymbol{\beta}, \boldsymbol{S}, \boldsymbol{e})$ is an ME representation of $F(t)$. By Theorem 4.2, the ME representation $(\boldsymbol{\beta}, \boldsymbol{S}, \boldsymbol{e})$ is a minimal ME representation.

If $k \ne 0$ then $\det(\boldsymbol{\Lambda}_{m+1}^{[k]}(t))$ is a Hankel determinant of $a_{1,1}^{[k]}(t)$. Using the above argument it can be shown that $a_{1,1}^{[k]}(t) = \boldsymbol{\beta} \exp\{\boldsymbol{S}t\}\boldsymbol{e}$. From (4.6) or (4.7), we find that $1 - F(t) = \boldsymbol{\beta} \exp\{\boldsymbol{S}t\}\boldsymbol{S}^{-k}\boldsymbol{e}$. Thus, the probability distribution has an ME representation $(\boldsymbol{\beta}, \boldsymbol{S}, \boldsymbol{S}^{-k}\boldsymbol{e})$. This proves the necessary and sufficient condition for a probability distribution to be an ME distribution.

Next we prove assertion 1. If $n \ge m + 1$ then, again by Theorem 4.2, $\det(\boldsymbol{\Lambda}_n^{(k)}(t)) = 0$ for all $t \ge 0$. If $n \le m$ and $\det(\boldsymbol{\Lambda}_n^{(k)}(t)) = 0$ for all $t \ge 0$ (and any fixed k), then, by [17, Theorem 7.1], and the above argument, the function $1 - F(t)$ is an exponential polynomial of order at most $n - 1$. This implies that $\det(\boldsymbol{\Lambda}_m^{(k)}(t)) = 0$ for all $t \ge 0$, which contradicts both the assumption that m is the ME order and Theorem 4.2. This proves assertion 1.

Since $\det(\boldsymbol{\Lambda}_m^{(k)}(t))$ is continuous in t and cannot equal 0 for all $t \ge 0$, then it is either always positive or always negative. This proves assertion 2.

Finally we prove assertion 3. If m is the ME order of the ME distribution then, by (4.8), we have

$$\det(\boldsymbol{\Lambda}_m^{[k]}(t)) = \det(\boldsymbol{\Lambda}_m^{[k]}(0)) \det(\exp\{\boldsymbol{T}t\}). \tag{4.10}$$

Since

$$\det(\exp\{\boldsymbol{T}t\}) = \exp\left\{\sum_{i=1}^{m} \lambda_i t\right\} = \exp\left\{\sum_{i=1}^{m} t_{i,i} t\right\},$$

(4.10) becomes

$$\det(\boldsymbol{\Lambda}_m^{[k]}(t)) = \det(\boldsymbol{\Lambda}_m^{[k]}(0)) \exp\left\{\sum_{i=1}^{m} \lambda_i t\right\},$$

which yields (4.9). If $n < m$ and $(1/t)\log(|\det(\boldsymbol{\Lambda}_n^{[k]}(t))/\det(\boldsymbol{\Lambda}_n^{[k]}(0))|)$ is constant in t, then, by [17, Corollary 7.1], the function $1 - F(t)$ is an exponential polynomial of order at most n, which contradicts the assumption that m is the ME order. This completes the proof of assertion 3 and, hence, the proof of Theorem 4.3.

Example 4.1. Consider a distribution with density function

$$F^{(1)}(t) = \left(1 + \frac{1}{4\pi^2}\right)(1 - \cos(2\pi t))\mathrm{e}^{-t}.$$

Note that Asmussen and Bladt [3], using a different method, proved that the ME order of this distribution equals 3 and found some minimal ME representations. By routine but tedious

https://doi.org/10.1239/aap/1175266478 Published online by Cambridge University Press

calculations (see Appendix A), we obtain

$$\begin{aligned} \det(\mathbf{\Lambda}_1^{[1]}(t)) &= a_{1,1}^{[1]}(t) = -F^{(1)}(t), \\ \det(\mathbf{\Lambda}_2^{[1]}(t)) &= (4\pi^2+1)^2(1-\cos(2\pi t))\mathrm{e}^{-2t}, \\ \det(\mathbf{\Lambda}_3^{[1]}(t)) &= (4\pi^2+1)^3\mathrm{e}^{-3t}, \\ \det(\mathbf{\Lambda}_n^{[1]}(t)) &= 0, \qquad n \ge 4. \end{aligned} \tag{4.11}$$

By Theorem 4.3, the distribution is an ME distribution with ME order 3. Since

$$\frac{1}{t}\log\left(\left|\frac{\det(\mathbf{\Lambda}_3^{[1]}(t))}{\det(\mathbf{\Lambda}_3^{[1]}(0))}\right|\right) = -3,$$

the sum of the diagonal elements of the matrix of any minimal ME representation is -3.

4.3. ME order and distribution functions

In this subsection we present a generalization of Theorem 3.1 that depends on $F(t)$, the distribution function of the ME distribution. For $\boldsymbol{x} = (x_1, x_2, \dots, x_n)$ and $\boldsymbol{y} = (y_1, y_2, \dots, y_n)$, define

$$\begin{aligned} &\mathbf{\Lambda}_n(\boldsymbol{x}, \boldsymbol{y}, t) = (a_{i,j}(x_i, y_j, t)) \qquad \text{with} \\ &\qquad a_{i,j}(x_i, y_j, t) = \boldsymbol{\alpha}\exp\{\boldsymbol{T}(x_i + y_j + t)\}\boldsymbol{u} = 1 - F(x_i + y_j + t), \quad 1 \le i, j \le n. \end{aligned}$$

The definition of the Hankel matrix $\mathbf{\Lambda}_n(\boldsymbol{x}, \boldsymbol{y}, t)$ is independent of the specific ME representation, but depends (only) on the distribution function. It is easy to show that

$$\mathbf{\Lambda}_n(\boldsymbol{x}, \boldsymbol{y}, t) = \boldsymbol{L}(n, \boldsymbol{x})\exp\{\boldsymbol{T}t\}\boldsymbol{R}(n, \boldsymbol{y}),$$

where

$$\boldsymbol{L}(n, \boldsymbol{x}) = \begin{pmatrix} \boldsymbol{\alpha}\exp\{\boldsymbol{T}x_1\} \\ \boldsymbol{\alpha}\exp\{\boldsymbol{T}x_2\} \\ \vdots \\ \boldsymbol{\alpha}\exp\{\boldsymbol{T}x_n\} \end{pmatrix} \quad \text{and} \quad \boldsymbol{R}(n, \boldsymbol{y}) = (\exp\{\boldsymbol{T}y_1\}\boldsymbol{u}, \dots, \exp\{\boldsymbol{T}y_n\}\boldsymbol{u}).$$

Theorem 4.4. *The ME order of an ME distribution* $(\boldsymbol{\alpha}, \boldsymbol{T}, \boldsymbol{u})$ *can be obtained, for any* t*, as*

$$\begin{aligned} \max\{n\colon \max\{|\det(\mathbf{\Lambda}_n(\boldsymbol{x}, \boldsymbol{y}, t))|\colon \boldsymbol{x} \in \mathbb{R}_+^n,\ \boldsymbol{y} \in \mathbb{R}_+^n, \\ x_1 < x_2 < \cdots < x_n,\ y_1 < y_2 < \cdots < y_n\} \ne 0,\ n \ge 1\}, \end{aligned} \tag{4.12}$$

where $\mathbb{R}_+^n$ *is the Cartesian product of* n *copies of the set of nonnegative real numbers. Consequently, if* n *is greater than the ME order then* $\det(\mathbf{\Lambda}_n(\boldsymbol{x}, \boldsymbol{y}, t)) = 0$ *for any* $\boldsymbol{x}$ *satisfying* $x_1 < x_2 < \cdots < x_n$*, any* $\boldsymbol{y}$ *satisfying* $y_1 < y_2 < \cdots < y_n$*, and any* t*.*

Proof. Suppose that m is the ME order of the ME distribution of interest and that there exists an order-m ME representation $(\boldsymbol{\alpha}, \boldsymbol{T}, \boldsymbol{u})$. By [3, Lemma 2.2], the dimensions of the space generated by $\{\boldsymbol{\alpha}\exp\{\boldsymbol{T}t\}\colon t \ge 0\}$ and the space generated by $\{\exp\{\boldsymbol{T}t\}\boldsymbol{u}\colon t \ge 0\}$ both equal m. Thus, there exist $\boldsymbol{x}$ and $\boldsymbol{y}$ such that $\{\boldsymbol{\alpha}\exp\{\boldsymbol{T}x_i\}\colon 1 \le i \le m\}$ and $\{\exp\{\boldsymbol{T}y_i\}\boldsymbol{u}\colon 1 \le i \le m\}$ are two sets of linearly independent vectors, and the matrices $\boldsymbol{L}(m, \boldsymbol{x})$ and $\boldsymbol{R}(m, \boldsymbol{y})$ are invertible. The quantity defined in (4.12) is then greater than or equal to the ME order, m. On the other hand, if $n > m$ then the vectors $\boldsymbol{\alpha}\exp\{\boldsymbol{T}x_i\}$, $1 \le i \le n$, are linearly dependent for any $\boldsymbol{x}$ and the vectors $\exp\{\boldsymbol{T}y_i\}\boldsymbol{u}$, $1 \le i \le n$, are linearly dependent for any $\boldsymbol{y}$, which implies that the quantity defined in (4.12) is less than or equal to m. This completes the proof.

5. An application to PH distributions

In this section we present an application of Theorem 3.1 to PH distributions. We begin with a general property related to the ME order of an ME distribution with an ME representation $(\boldsymbol{\alpha}, \boldsymbol{T}, \boldsymbol{u})$. We consider two subspaces, respectively generated by the vectors $\boldsymbol{\alpha}, \boldsymbol{\alpha T}, \boldsymbol{\alpha T}^2, \ldots$ and $\boldsymbol{u}, \boldsymbol{T u}, \boldsymbol{T}^2\boldsymbol{u}, \ldots$. Suppose that the dimensions of the two subspaces are n_1 and n_2, respectively. Since $\boldsymbol{T}$ is invertible, it can be proved (see [3, Lemma 2.2]) that the vectors $\boldsymbol{\alpha}, \boldsymbol{\alpha T}, \ldots, \boldsymbol{\alpha T}^{n_1-1}$ are linearly independent and the vectors $\boldsymbol{u}, \boldsymbol{T u}, \ldots, \boldsymbol{T}^{n_2-1}\boldsymbol{u}$ are linearly independent. Since the matrix $\boldsymbol{T}$ is of order m, we must have $\max\{n_1, n_2\} \le m$. Furthermore, we have the following proposition regarding the matrix $\boldsymbol{\Lambda}_n^{[k]}$.

Proposition 5.1. *For $n \ge \min\{n_1, n_2\} + 1$ and $-\infty < k < \infty$, we have $\det(\boldsymbol{\Lambda}_n^{[k]}) = 0$. Consequently, the ME order of the ME distribution $(\boldsymbol{\alpha}, \boldsymbol{T}, \boldsymbol{u})$ is less than or equal to $\min\{n_1, n_2\}$.*

Proof. By definition (see (3.2)), the rank of the matrix $\boldsymbol{L}(n)$ is at most n_1 and the rank of the matrix $\boldsymbol{R}(n)$ is at most n_2. By the Binet–Cauchy formula (see [19]), the rank of the matrix $\boldsymbol{\Lambda}_n^{[k]}$ is at most $\min\{n_1, n_2\}$. Therefore, we must have $\det(\boldsymbol{\Lambda}_n^{[k]}) = 0$ for $n \ge \min\{n_1, n_2\} + 1$ and $-\infty < k < \infty$. This completes the proof.

Next we present an application to PH distributions. The PH distribution was introduced by Neuts [23] as the probability distribution of the absorption time of a finite-state Markov process. For a given PH distribution $F(t)$, if

$$F(t) = 1 - \boldsymbol{\alpha} \exp\{\boldsymbol{T} t\}\boldsymbol{e}, \qquad t \ge 0, \tag{5.1}$$

$(\boldsymbol{\alpha}, 1 - \boldsymbol{\alpha e})$ is a probability vector, and T is a PH generator of order m (i.e. an $m \times m$ invertible matrix with negative diagonal elements, nonnegative off-diagonal elements, and nonnegative row sums), then we call the pair $(\boldsymbol{\alpha}, \boldsymbol{T})$ a *PH representation* of that PH distribution. The integer m is the order of the PH representation $(\boldsymbol{\alpha}, \boldsymbol{T})$. We refer the reader to [24, Chapter 2], for basic properties of PH distributions.

It is well known that the PH representation of a PH distribution is not unique. The *PH order* of a PH distribution is defined to be the minimal order of all its PH representations. By definitions (2.1) and (5.1), it is readily seen that PH distributions are special ME distributions. Thus, each PH distribution has a PH order and an ME order, and the PH order is greater than or equal to the ME order. Consequently, for PH distributions, Theorem 3.1 provides a lower bound on their PH orders. Next, using the ME order, we prove a result on the PH orders of PH representations with a PH-simple generator.

The following definition was first given in [25]. A PH generator T is *PH simple* if any two PH representations $(\boldsymbol{\alpha}, \boldsymbol{T})$ and $(\boldsymbol{\beta}, \boldsymbol{T})$, where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are different, represent two different probability distributions. According to [25, Theorem 2], a PH generator T of order m is PH simple if and only if $\boldsymbol{e}, \boldsymbol{T e}, \boldsymbol{T}^2\boldsymbol{e}, \ldots, \boldsymbol{T}^{m-1}\boldsymbol{e}$ are linearly independent. Related results on PH simplicity can be found in [6]. Under the PH simplicity condition, by Theorem 3.1 it is straightforward to obtain the following result.

Proposition 5.2. *Consider an order-m PH representation $(\boldsymbol{\alpha}, \boldsymbol{T})$ where $\boldsymbol{T}$ is PH simple. If $\boldsymbol{\alpha}, \boldsymbol{\alpha T}, \boldsymbol{\alpha T}^2, \ldots, \boldsymbol{\alpha} T^{m-1}$ are linearly independent then the PH order of the PH distribution $(\boldsymbol{\alpha}, \boldsymbol{T})$ is m.*

As an application of the above proposition, we consider an order-m Erlang distribution with parameter λ, where λ is a positive number. A PH representation of such an Erlang distribution

https://doi.org/10.1239/aap/1175266478 Published online by Cambridge University Press

has components

$$\boldsymbol{\alpha} = (0, 0, \ldots, 0, 1) \quad \text{and} \quad T = \begin{pmatrix} -\lambda & & & & \\ \lambda & -\lambda & & & \\ & \lambda & -\lambda & & \\ & & \ddots & \ddots & \\ & & & \lambda & -\lambda \end{pmatrix},$$

where $\boldsymbol{\alpha}$ is a dimension-m vector and $\boldsymbol{T}$ is an $m \times m$ matrix. We use Theorem 3.1 to prove the following well-known result, which implies that the ME order can be a tight lower bound of the PH order.

Proposition 5.3. *For an order-m Erlang distribution with parameter* λ*, we have* $\det(\boldsymbol{\Lambda}_m^{(k)}) \neq 0$ *for* $-\infty < k < \infty$*. Consequently, the ME order and the PH order of the Erlang distribution both equal* m.

Proof. By definition (see (3.2)), it is easy to verify that

$$\boldsymbol{L}(m) = \begin{pmatrix} 0 & 0 & \cdots & 0 & 0 & 1 \\ 0 & 0 & \cdots & 0 & \lambda & -\lambda \\ 0 & 0 & \iddots & \lambda^2 & -2\lambda^2 & \lambda^2 \\ \vdots & \iddots & \iddots & \iddots & \vdots & \vdots \\ 0 & \lambda^{m-2} & \iddots & (-1)^{m-4}\binom{m-2}{m-4}\lambda^{m-2} & (-1)^{m-3}\binom{m-2}{m-3}\lambda^{m-2} & (-\lambda)^{m-2} \\ \lambda^{m-1} & -\binom{m-1}{1}\lambda^{m-1} & \cdots & (-1)^{m-3}\binom{m-1}{m-3}\lambda^{m-1} & (-1)^{m-2}\binom{m-1}{m-2}\lambda^{m-1} & (-\lambda)^{m-1} \end{pmatrix},$$

$$\boldsymbol{R}(m) = \begin{pmatrix} 1 & -\lambda & \lambda^2 & \cdots & (-1)^{m-2}\binom{m-3}{0}\lambda^{m-2} & (-1)^{m-1}\binom{m-2}{0}\lambda^{m-1} \\ 1 & 0 & -\lambda^2 & \cdots & (-1)^{m-3}\binom{m-3}{1}\lambda^{m-2} & (-1)^{m-2}\binom{m-2}{1}\lambda^{m-1} \\ 1 & 0 & 0 & \ddots & (-1)^{m-4}\binom{m-3}{2}\lambda^{m-2} & (-1)^{m-3}\binom{m-2}{2}\lambda^{m-1} \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 1 & 0 & 0 & \ddots & 0 & -\binom{m-2}{m-2}\lambda^{m-1} \\ 1 & 0 & 0 & \cdots & 0 & 0 \end{pmatrix}.$$

Thus, both $\boldsymbol{L}(m)$ and $\boldsymbol{R}(m)$ have full rank (i.e. m). By (3.3), we have $\det(\boldsymbol{\Lambda}_m^{(k)}) \neq 0$ for $-\infty < k < \infty$. Therefore, the ME order and the PH order of the Erlang distribution both equal m. This completes the proof.

Next we introduce a concept related to PH generators. A PH generator T is *PH full* if there exists a probability vector $\boldsymbol{\alpha}$ such that the PH representation $(\boldsymbol{\alpha}, \boldsymbol{T})$ is a minimal PH representation. For a PH-full generator $\boldsymbol{T}$, the PH representation $(\boldsymbol{\alpha}, \boldsymbol{T})$ may have an equivalent PH representation of smaller order for some, but not all, probability vectors $\boldsymbol{\alpha}$. Using Proposition 5.2, we obtain the following sufficient condition for a PH generator to be PH full.

Theorem 5.1. *If $\boldsymbol{T}$ is PH simple then there exists a probability vector $\boldsymbol{\alpha}$ such that the PH representation $(\boldsymbol{\alpha}, \boldsymbol{T})$ is a minimal PH representation. Consequently, if $\boldsymbol{T}$ is PH simple then $\boldsymbol{T}$ is PH full.*

Proof. By definition, it is sufficient to show that there exists a probability vector $\boldsymbol{\alpha}$ such that the PH representation $(\boldsymbol{\alpha}, \boldsymbol{T})$ is a minimal PH representation. By the definition of PH simplicity and Proposition 5.2, it is sufficient to find a probability vector $\boldsymbol{\alpha}$ such that $\boldsymbol{\alpha}, \boldsymbol{\alpha T}, \boldsymbol{\alpha T}^2, \ldots, \boldsymbol{\alpha T}^{m-1}$ are linearly independent, where m is the order of the matrix $\boldsymbol{T}$. For this purpose, we utilize the Jordan canonical form of the matrix $\boldsymbol{T}$. According to [25, Theorem 2], $\boldsymbol{T}$ is PH simple if and only if no row eigenvector of $\boldsymbol{T}$ is orthogonal to the vector $\boldsymbol{e}$. If two (row) eigenvectors $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ correspond to the same eigenvalue, then $\boldsymbol{u}_1 - \boldsymbol{u}_2$ is also an eigenvector of that eigenvalue. After normalization (if possible), one of the vectors $\boldsymbol{u}_1, \boldsymbol{u}_2$, and $\boldsymbol{u}_1 - \boldsymbol{u}_2$ must have a zero summation. Therefore, [25, Theorem 2] implies that the algebraic multiplicities of all eigenvalues of $\boldsymbol{T}$ equal 1 (i.e. every eigenvalue is related to exactly one Jordan block or, equivalently, the eigenvalues of the Jordan blocks of $\boldsymbol{T}$ are different). Suppose that $\boldsymbol{T} = \boldsymbol{VJU}$, where $\boldsymbol{V} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_m) = \boldsymbol{U}^{-1}$, $\boldsymbol{U} = (\boldsymbol{u}_1^\top, \ldots, \boldsymbol{u}_m^\top)^\top$, and

$$J = \begin{pmatrix} \boldsymbol{J}_1(\rho_1) & & \\ & \ddots & \\ & & \boldsymbol{J}_K(\rho_K) \end{pmatrix} \quad \text{with} \quad \boldsymbol{J}_k(\rho_k) = \begin{pmatrix} \rho_k & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & \rho_k \end{pmatrix}, \quad 1 \le k \le K, \tag{5.2}$$

is the Jordan canonical form of $\boldsymbol{T}$. Here $\rho_1, \rho_2, \ldots, \rho_K$ are the K different eigenvalues of $\boldsymbol{T}$ and $\boldsymbol{J}_k(\rho_k)$ is of order m_k. If $\boldsymbol{u}_n$ is an eigenvector of $\boldsymbol{T}$ and all nonzero elements of $\boldsymbol{u}_n$ have the same sign, then the vector $\boldsymbol{u}_n$ is chosen to be nonnegative. We now define

$$\boldsymbol{\alpha} = \boldsymbol{\varepsilon U}/(\boldsymbol{\varepsilon U e}), \tag{5.3}$$

where $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \ldots, \boldsymbol{\varepsilon}_K)$ and $\boldsymbol{\varepsilon}_k$ is a vector of dimension m_k specified as follows. According to (5.2), $\boldsymbol{\varepsilon}_k$ corresponds to the Jordan block $\boldsymbol{J}_k(\rho_k)$. The vector $\boldsymbol{\varepsilon}_k$ equals either $(1, \ldots, 1)$ or $(1, \ldots, 1, c)$, where c is a very large, positive constant. If the corresponding eigenvector of ρ_k is not nonnegative then $\boldsymbol{\varepsilon}_k$ equals $(1, \ldots, 1)$; otherwise $\boldsymbol{\varepsilon}_k$ equals $(1, \ldots, 1, c)$. We choose c large enough that the vector $\boldsymbol{\alpha}$ defined in (5.3) is stochastic and every element of $\boldsymbol{\alpha}$ is positive.

If $\boldsymbol{\alpha}, \boldsymbol{\alpha T}, \boldsymbol{\alpha T}^2, \ldots, \boldsymbol{\alpha T}^{m-1}$ are linearly dependent then there exist $x_1, x_2, \ldots, x_m$, not all of which equal 0, such that

$$\begin{aligned} \boldsymbol{0} &= \sum_{i=1}^{m} x_i \boldsymbol{\alpha T}^{i-1} = \frac{1}{\boldsymbol{\varepsilon U e}} \boldsymbol{\varepsilon} \bigg(\sum_{i=1}^{m} x_i \boldsymbol{J}^{i-1} \bigg) \boldsymbol{V} \\ &= \frac{\boldsymbol{\varepsilon}}{\boldsymbol{\varepsilon U e}} \begin{pmatrix} \sum_{i=1}^{m} x_i \boldsymbol{J}_1(\rho_1)^{i-1} & & \\ & \ddots & \\ & & \sum_{i=1}^{m} x_i \boldsymbol{J}_K(\rho_K)^{i-1} \end{pmatrix} \boldsymbol{V} \\ &= \frac{1}{\boldsymbol{\varepsilon U e}} \bigg(\boldsymbol{\varepsilon}_1 \sum_{i=1}^{m} x_i \boldsymbol{J}_1(\rho_1)^{i-1}, \boldsymbol{\varepsilon}_2 \sum_{i=1}^{m} x_i \boldsymbol{J}_2(\rho_2)^{i-1}, \ldots, \boldsymbol{\varepsilon}_K \sum_{i=1}^{m} x_i \boldsymbol{J}_K(\rho_K)^{i-1} \bigg) \boldsymbol{V}. \end{aligned} \tag{5.4}$$

https://doi.org/10.1239/aap/1175266478 Published online by Cambridge University Press

By routine calculations, we have

$$\sum_{i=1}^{m} x_i \boldsymbol{J}_k(\rho_k)^{i-1} = \begin{pmatrix} \sum_{i=1}^{m} x_i \binom{i-1}{0}\rho_k^{i-1} & \sum_{i=1}^{m} x_i \binom{i-1}{1}\rho_k^{i-2} & \cdots & \sum_{i=1}^{m} x_i \binom{i-1}{m_k-1}\rho_k^{i-m_k} \\ & \sum_{i=1}^{m} x_i \binom{i-1}{0}\rho_k^{i-1} & \ddots & \vdots \\ & & \ddots & \sum_{i=1}^{m} x_i \binom{i-1}{1}\rho_k^{i-2} \\ & & & \sum_{i=1}^{m} x_i \binom{i-1}{0}\rho_k^{i-1} \end{pmatrix}. \tag{5.5}$$

Since the matrix $\boldsymbol{V}$ is invertible, (5.5) implies that (5.4) is equivalent to

$$0 = \sum_{i=1}^{m} x_i \binom{i-1}{j}\rho_k^{i-1-j}, \qquad 0 \le j \le m_k - 1,\ 1 \le k \le K. \tag{5.6}$$

Equation (5.6) can be rewritten as

$$0 = (x_1, \dots, x_m)(\boldsymbol{B}_1, \boldsymbol{B}_2, \dots, \boldsymbol{B}_K), \tag{5.7}$$

where

$$\boldsymbol{B}_k = \begin{pmatrix} 1 & & & \\ \binom{1}{0}\rho_k & 1 & & \\ \binom{2}{0}\rho_k^2 & \binom{2}{1}\rho_k^{2-1} & \ddots & \\ \vdots & \vdots & \ddots & 1 \\ \binom{m_k}{0}\rho_k^{m_k} & \vdots & \vdots & \binom{m_k}{m_k-1}\rho_k^{m_k-(m_1-1)} \\ \vdots & \vdots & \vdots & \vdots \\ \binom{m-1}{0}\rho_k^{m-1} & \binom{m-1}{1}\rho_k^{m-1-1} & \cdots & \binom{m-1}{m_k-1}\rho_k^{m-1-(m_k-1)} \end{pmatrix}.$$

The matrix $(\boldsymbol{B}_1, \boldsymbol{B}_2, \dots, \boldsymbol{B}_K)$ is a confluent, or generalized, Vandermonde matrix (see [19]). In Appendix B we show that

$$\det(\boldsymbol{B}_1, \boldsymbol{B}_2, \dots, \boldsymbol{B}_K) = \prod_{1 \le i < j \le K} (\rho_j - \rho_i)^{m_i m_j}. \tag{5.8}$$

Since $\rho_1, \rho_2, \dots, \rho_K$ are all different, the determinant of $(\boldsymbol{B}_1, \boldsymbol{B}_2, \dots, \boldsymbol{B}_K)$ is nonzero and the matrix $(\boldsymbol{B}_1, \boldsymbol{B}_2, \dots, \boldsymbol{B}_K)$ is invertible. Equation (5.7) then implies that $x_1, x_2, \dots, x_m$ all equal 0, which contradicts the assumption made about them. Therefore, $\boldsymbol{\alpha}, \boldsymbol{\alpha T}, \boldsymbol{\alpha T}^2, \dots, \boldsymbol{\alpha T}^{m-1}$ are linearly independent.

Since both $\{\boldsymbol{\alpha}, \boldsymbol{\alpha T}, \boldsymbol{\alpha T}^2, \dots, \boldsymbol{\alpha T}^{m-1}\}$ and $\{\boldsymbol{e}, \boldsymbol{Te}, \boldsymbol{T}^2\boldsymbol{e}, \dots, \boldsymbol{T}^{m-1}\boldsymbol{e}\}$ are sets of linearly independent vectors, by Proposition 5.2 the PH order of $(\boldsymbol{\alpha}, \boldsymbol{T})$ is m. This completes the proof.

6. Numerical examples

In [1], it was shown that the PH order of a PH random variable X is greater than or equal to $\mathrm{E}[X^2]/\operatorname{var}(X)$, where $\operatorname{var}(X)$ is the variance of X. The two are equal if and only if the PH distribution is an Erlang distribution. We shall call the smallest integer that is greater than or equal to $\mathrm{E}[X^2]/\operatorname{var}(X)$ the Aldous–Shepp lower bound. In this section we present a number of numerical examples to provide insight into the relationship between the ME order, the PH order, and the Aldous–Shepp lower bound.

Remark 6.1. In [27] a lower bound on the PH order was given. This lower bound is useful if the PH generator $\boldsymbol{T}$ has complex eigenvalues. However, we do not consider this lower bound here.

Example 6.1. Consider the PH generator

$$\boldsymbol{T} = \begin{pmatrix} -1.5 & 0 & 0 & 0 & 0.2 \\ 1 & -1.2 & 0 & 0.1 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0.2 & 0 & 1 & -1.5 & 0 \\ 0 & 1 & 0 & 1 & -2 \end{pmatrix}. \tag{6.1}$$

We consider the following two associated PH representations.

1. The PH representation $(\boldsymbol{\alpha}, \boldsymbol{T})$ with $\boldsymbol{\alpha} = (0.5, 0.1, 0.3, 0, 0.1)$ has ME order 5. Thus, its PH order also equals 5. Since $\mathrm{E}[X^2]/\operatorname{var}(X) = 0.948$, the Aldous–Shepp lower bound equals 1, which is far below the PH order.
2. The PH representation $(\boldsymbol{\alpha}, \boldsymbol{T})$ with $\boldsymbol{\alpha} = (0, 0, 0.8, 0, 0.2)$ has an ME order and a PH order that both equal 5, while $\mathrm{E}[X^2]/\operatorname{var}(X) = 2.051$. In this case, the Aldous–Shepp lower bound is 3, which is not far from the PH order.

For most of the PH distributions $(\boldsymbol{\alpha}, \boldsymbol{T})$ with $\boldsymbol{T}$ as in (6.1), we find that the ME order and the PH order both equal 5, while the corresponding Aldous–Shepp lower bound equals either 1 or 2. This example, as well as many other examples, shows that the Aldous–Shepp lower bound can be significantly smaller than the PH order. Furthermore, numerical examples also show that in many cases the Aldous–Shepp lower bound is smaller than the ME order of PH distributions. However, as the following example shows, the Aldous–Shepp lower bound can be greater than the ME order of a PH distribution.

Example 6.2. Consider the PH generator

$$\boldsymbol{T} = \begin{pmatrix} -2.5 & 0 & 0.3 \\ 1.5 & -1.5 & 0 \\ 0 & 0.1 & -1 \end{pmatrix}.$$

This matrix has three real eigenvalues, $-\lambda_1 = -2.4683$, $-\lambda_2 = -1.5841$, and $-\lambda_3 = -0.9475$, with corresponding eigenvectors

$$\begin{aligned} \boldsymbol{\alpha}_1 &= (1.2243, 0.0258, -0.2501), \\ \boldsymbol{\alpha}_2 &= (0.9115, 0.5566, -0.4681), \\ \boldsymbol{\alpha}_3 &= (0.1290, 0.1335, 0.7375). \end{aligned}$$

https://doi.org/10.1239/aap/1175266478 Published online by Cambridge University Press

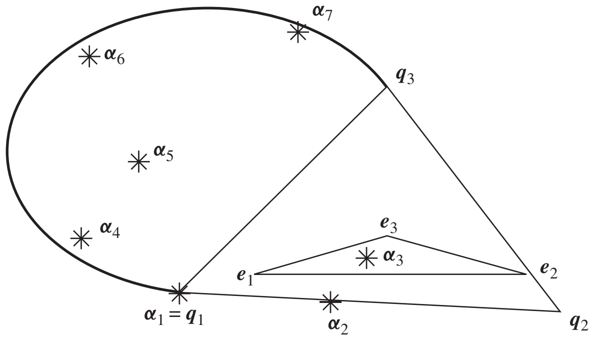

FIGURE 1: The convex set C_T.

Thus, $(\boldsymbol{\alpha}_i, \boldsymbol{T})$, $1 \leq i \leq 3$, represent three exponential distributions with respective parameters λ_1, λ_2, and λ_3. Using the theory developed in [8], a convex set C_T (plotted in Figure 1) can be constructed. In the figure, $\boldsymbol{q}_1$, $\boldsymbol{q}_2$, and $\boldsymbol{q}_3$ are defined as

$$\boldsymbol{q}_1 = \boldsymbol{\alpha}_1,$$
$$\boldsymbol{q}_2 = \frac{\lambda_2}{\lambda_2 - \lambda_1}\boldsymbol{\alpha}_1 + \frac{\lambda_1}{\lambda_1 - \lambda_2}\boldsymbol{\alpha}_2,$$
$$\boldsymbol{q}_3 = \frac{\lambda_3\lambda_2}{(\lambda_3 - \lambda_1)(\lambda_2 - \lambda_1)}\boldsymbol{\alpha}_1 + \frac{\lambda_1\lambda_3}{(\lambda_1 - \lambda_2)(\lambda_3 - \lambda_2)}\boldsymbol{\alpha}_2 + \frac{\lambda_1\lambda_2}{(\lambda_1 - \lambda_3)(\lambda_2 - \lambda_3)}\boldsymbol{\alpha}_3,$$

and thus calculated to be

$$\boldsymbol{q}_1 = (1.2243, 0.0258, -0.2501),$$
$$\boldsymbol{q}_2 = (0.3514, 1.5074, -0.8588),$$
$$\boldsymbol{q}_3 = (-1.9002, -1.7447, 4.6449).$$

All probability distributions whose Laplace–Stieltjes transforms have poles $-\lambda_1$, $-\lambda_2$, and $-\lambda_3$ are contained in the set C_T, i.e. for every such distribution (except those on the curve between $\boldsymbol{q}_1$ and $\boldsymbol{q}_3$), there exists an $\boldsymbol{\alpha}$ in C_T such that $(\boldsymbol{\alpha}, \boldsymbol{T}, \boldsymbol{e})$ is an ME representation of that distribution. The plotted polytope conv $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3\}$ [29], where $\boldsymbol{e}_1 = (1, 0, 0)$, $\boldsymbol{e}_2 = (0, 1, 0)$, and $\boldsymbol{e}_3 = (0, 0, 1)$, consists of all the probability vectors.

For all $\boldsymbol{\alpha}$ contained in the convex set conv $\{\boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{q}_3\}$, $(\boldsymbol{\alpha}, \boldsymbol{T}, \boldsymbol{e})$ represents a Coxian distribution. According to [27, Theorem 4.1], every $(\boldsymbol{\alpha}, \boldsymbol{T}, \boldsymbol{e})$ for which $\boldsymbol{\alpha}$ is contained in the interior of C_T also represents a Coxian distribution. For such $\boldsymbol{\alpha}$, these distributions have PH orders and ME orders. The ME orders are less than or equal to 3, but the PH orders can be very large. The Aldous–Shepp lower bound is computed for the following examples.

1. $(\boldsymbol{q}_1, \boldsymbol{T}, \boldsymbol{e})$ represents an exponential distribution. $\mathrm{E}[X^2]/\operatorname{var}(X) = 1$, so the Aldous–Shepp lower bound equals 1, which is also the PH order.

2. $(\boldsymbol{q}_2, \boldsymbol{T}, \boldsymbol{e})$ represents a Coxian distribution of order 2. $\mathrm{E}[X^2]/\operatorname{var}(X) = 1.901$, so the Aldous–Shepp lower bound equals 2, which is also the PH order.

3. $(\boldsymbol{q}_3, \boldsymbol{T}, \boldsymbol{e})$ represents a Coxian distribution of order 3. $\mathrm{E}[X^2]/\operatorname{var}(X) = 2.610$, so the Aldous–Shepp lower bound equals 3, which is also the PH order.

4. For $\boldsymbol{e}_1$, $\mathrm{E}[X^2]/\operatorname{var}(X) = 0.6510$, so the Aldous–Shepp lower bound equals 1. For $\boldsymbol{e}_2$, $\mathrm{E}[X^2]/\operatorname{var}(X) = 1.6335$, so the Aldous–Shepp lower bound equals 2. For $\boldsymbol{e}_3$, $\mathrm{E}[X^2]/\operatorname{var}(X) = 1.0299$, so the Aldous–Shepp lower bound equals 2. In these three cases both the PH order and the ME order equal 3.

5. $(\boldsymbol{\alpha}_i, \boldsymbol{T}, \boldsymbol{e})$, $i = 1, 2, 3$, represent exponential distributions. For each, $\mathrm{E}[X^2]/\operatorname{var}(X) = 1$, so the Aldous–Shepp lower bound equals 1, which is also the PH order.

6. For $\boldsymbol{\alpha}_4 = (1, -1, 1)$, $\mathrm{E}[X^2]/\operatorname{var}(X) = 0.3088$, so the Aldous–Shepp lower bound equals 1. The ME order equals 3.

7. For $\boldsymbol{\alpha}_5 = (-0.5, -2, 3.5)$, $\mathrm{E}[X^2]/\operatorname{var}(X) = 0.7538$, so the Aldous–Shepp lower bound equals 1. The ME order equals 3.

8. For $\boldsymbol{\alpha}_6 = (-1.5, -3.5, 6)$, $\mathrm{E}[X^2]/\operatorname{var}(X) = 1.1026$, so the Aldous–Shepp lower bound equals 2. The ME order equals 3.

9. For $\boldsymbol{\alpha}_7 = (-2.49, -3, 6.49)$, $\mathrm{E}[X^2]/\operatorname{var}(X) = 3.008$, so the Aldous–Shepp lower bound equals 4, which is greater than the ME order, 3.

In the last case, the PH order is greater than the ME order. Note that $\boldsymbol{\alpha}_7$ is close to the boundary of C_T. The PH order of $(\boldsymbol{\alpha}_7, \boldsymbol{T}, \boldsymbol{e})$ can be very large.

Appendix A. Proof of (4.11)

Define $w(t) = 1 - \cos(2\pi t)$. Then we have $F^{(1)}(t) = c\mathrm{e}^{-t}w(t)$, where $c = 1 + 1/(4\pi^2)$. It is easy to verify that

$$\begin{aligned} w^{(2n+1)}(t) &= (-1)^n(2\pi)^{2n+1}\sin(2\pi t), \qquad && n \ge 0, \\ w^{(2n)}(t) &= (-1)^{n-1}(2\pi)^{2n}\cos(2\pi t), \qquad && n \ge 1. \end{aligned} \tag{A.1}$$

From (A.1), by induction we obtain, for $n \ge 0$,

$$\begin{aligned} F^{(1+n)}(t) &= c\mathrm{e}^{-t}\left(\sum_{k=0}^{n}(-1)^k\binom{n}{k}w^{(n-k)}(t)\right) \\ &= c\mathrm{e}^{-t}\left(\sum_{k=0}^{n-1}(-1)^k\binom{n-1}{k}w^{(1+n-1-k)}(t)\right) - F^{(n)}(t). \end{aligned} \tag{A.2}$$

Using (A.2), it can be shown that

$$\det(\boldsymbol{\Lambda}_n^{[1]}(t)) = c^n\mathrm{e}^{-nt}\det\begin{pmatrix} w(t) & w^{(1)}(t) & \cdots & w^{(n-1)}(t) \\ w^{(1)}(t) & w^{(2)}(t) & \cdots & w^{(n)}(t) \\ \vdots & \vdots & \ddots & \vdots \\ w^{(n-1)}(t) & w^{(n)}(t) & \cdots & w^{(2n-2)}(t) \end{pmatrix}. \tag{A.3}$$

The first three equalities in (4.11) can be shown by setting n in (A.3) to equal 1, 2, and 3, respectively. The last equality in (4.11) is obtained from the fact that elements in the second column and the fourth column of the matrix in (A.3) satisfy $w^{(n)}(t) = -4\pi^2 w^{(n+2)}(t)$, $n \ge 1$. This completes the proof.

Appendix B. Proof of (5.8)

Equation (5.8) should be a consequence of the theory of total positivity developed in [17], but the authors have not found an appropriate reference. Thus, for completeness, we include its proof, in which we use the induction method.

If $K = 1$ then, by definition of $\boldsymbol{B}_1$, $\det(\boldsymbol{B}_1) = 1$. For $K \geq 2$, after multiplying the rth row of $\boldsymbol{B}_1$ by $-\rho_1$ and adding the result to its $(r+1)$th row for $r = m-1, m-2, \dots, 2, 1$, the matrix $\boldsymbol{B}_1$ becomes

$$\begin{pmatrix} 1 & & & & \\ 0 & 1 & & & \\ 0 & \binom{2}{1}\rho_1 - \rho_1 & \ddots & & \\ \vdots & \vdots & \ddots & 1 & \\ \vdots & \vdots & \vdots & \binom{m_k}{m_k-1}\rho_1 - \rho_1 & \\ \vdots & \vdots & \vdots & \vdots & \\ 0 & \binom{m-1}{1}\rho_1^{m-2} - \binom{m-2}{1}\rho_1^{m-2} & \cdots & \binom{m-1}{m_k-1}\rho_1^{m-m_k} - \binom{m-2}{m_k-1}\rho_1^{m-m_k} & \end{pmatrix}$$

$$= \begin{pmatrix} 1 & & & \\ 0 & 1 & & \\ 0 & \rho_1 & \ddots & \\ \vdots & \vdots & \ddots & 1 \\ \vdots & \vdots & \vdots & \binom{m_k-1}{m_k-2}\rho_1 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \rho_1^{m-2} & \cdots & \binom{m-2}{m_k-2}\rho_1^{m-m_k} \end{pmatrix} =: \begin{pmatrix} 1 & 0 \\ 0 & B_1' \end{pmatrix}, \tag{B.1}$$

where the identity $\binom{n}{i} = \binom{n-1}{i} + \binom{n-1}{i-1}$ has been used. For $2 \leq k \leq K$, the same operations yield for $\boldsymbol{B}_k$ the matrix

$$\begin{pmatrix} 1 & & & \\ \rho_k - \rho_1 & 1 & & \\ \rho_k(\rho_k - \rho_1) & \binom{2}{1}\rho_k - \rho_1 & \ddots & \\ \vdots & \vdots & \ddots & 1 \\ \rho_k^{m_k-1}(\rho_k - \rho_1) & \vdots & \vdots & \binom{m_k}{m_k-1}\rho_k - \rho_1 \\ \vdots & \vdots & \vdots & \vdots \\ \rho_k^{m-2}(\rho_k - \rho_1) & \binom{m-1}{1}\rho_k^{m-2} - \binom{m-2}{1}\rho_k^{m-3}\rho_1 & \cdots & \binom{m-1}{m_k-1}\rho_k^{m-m_k} - \binom{m-2}{m_k-1}\rho_k^{m-1-m_k}\rho_1 \end{pmatrix}.$$

https://doi.org/10.1239/aap/1175266478 Published online by Cambridge University Press

When the rth column of the above matrix is divided by $-1/(\rho_k - \rho_1)$ and the result added to the $(r+1)$th column for $r = 1, 2, \dots,$ the matrix becomes

$$\begin{pmatrix} 1 & \dfrac{-1}{\rho_k - \rho} & \cdots & \left(-\dfrac{1}{\rho_k - \rho}\right)^{m_k - 1} \\ \rho_k - \rho_1 & 0 & \cdots & 0 \\ \rho_k(\rho_k - \rho_1) & \rho_k - \rho_1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ \rho_k^{m_k-1}(\rho_k - \rho_1) & \vdots & \vdots & \rho_k - \rho_1 \\ \vdots & \vdots & \vdots & \vdots \\ \rho_k^{m-2}(\rho_k - \rho_1) & \dbinom{m-2}{1}\rho_k^{m-2-1}(\rho_k - \rho_1) & \cdots & \dbinom{m-2}{m_k - 1}\rho_k^{m-2-(m_k-1)}(\rho_k - \rho_1) \end{pmatrix}$$
$$=: \begin{pmatrix} \left(1, -\dfrac{1}{\rho_k - \rho}, \dots, \left(-\dfrac{1}{\rho_k - \rho}\right)^{m_k-1}\right) \\ (\rho_k - \rho_1)\boldsymbol{B}_k' \end{pmatrix}.$$

Note that the first column of the matrix in (B.1) has only one nonzero element in the first row. Then, by induction, we have

$$\begin{aligned} \det(\boldsymbol{B}_1, \boldsymbol{B}_2, \dots, \boldsymbol{B}_K) &= \left(\prod_{2 \le j \le K} (\rho_j - \rho_1)^{m_j}\right) \det(\boldsymbol{B}_1', \boldsymbol{B}_2', \dots, \boldsymbol{B}_K') \\ &= \left(\prod_{2 \le j \le K} (\rho_j - \rho_1)^{m_j}\right)\left(\prod_{2 \le j \le K} (\rho_j - \rho_1)^{m_j(m_1 - 1)}\right) \\ &\quad \times \left(\prod_{2 \le i < j \le K} (\rho_j - \rho_i)^{m_j m_i}\right), \end{aligned}$$

which leads to (5.8). This completes the proof.

Acknowledgements

The authors would like to thank an anonymous referee for his/her valuable comments and suggestions. This research was supported by an NSERC research grant (Canada) and a research grant from the Chinese Academy of Sciences, P. R. China.

References

[1] Aldous, D. and Shepp, L. (1987). The least variable phase type distribution is Erlang. *Stoch. Models* **3,** 467–473.
[2] Asmussen, S. (1987). *Applied Probability and Queues*. John Wiley, New York.
[3] Asmussen, S. and Bladt, M. (1996). Renewal theory and queueing algorithms for matrix-exponential distributions. In *Proc. First Internat. Conf. Matrix Analytic Methods Stoch. Models*, eds A. S. Alfa and S. Chakravarthy, Marcel Dekker, New York, pp. 313–341.
[4] Benvenuti, L. and Farina, L. (2004). A tutorial on the positive realization problem. *IEEE Trans. Automatic Control* **49,** 651–664.
[5] Bladt, M. and Neuts, M. F. (2003). Matrix-exponential distributions: calculus and interpretations via flows. *Stoch. Models* **19,** 113–124.

[6] Commault, C. and Chelma, J. P. (1993). On dual and minimal phase-type representations. *Stoch. Models* **9,** 421–434.
[7] Commault, C. and Mocanu, S. (2003). Phase-type distributions and representations: some results and open problems for system theory. *Internat. J. Control* **76,** 566–580.
[8] Dehon, M. and Latouche, G. (1982). A geometric interpretation of the relations between the exponential and the generalized Erlang distributions. *Adv. Appl. Prob.* **14,** 885–897.
[9] Fackrell, M. (2003). Characterization of matrix-exponential distributions. Doctoral Thesis, University of Adelaide.
[10] Fackrell, M. (2005). Fitting with matrix-exponential distributions. *Stoch. Models* **21,** 377–400.
[11] Gilbert, E. J. (1959). On the identifiability problem for functions of finite Markov chains. *Ann. Math. Statist.* **30,** 688–697.
[12] He, Q. and Zhang, H. (2005). An algorithm for computing minimal Coxian representations. To appear in *INFORMS J. Computing*.
[13] He, Q. and Zhang, H. (2005). Spectral polynomial algorithms for computing bi-diagonal representations for phase type distributions and matrix exponential distributions. *Stoch. Models* **22,** 289–317.
[14] Ito, H., Amari, S.-I. and Kobayashi, K. (1992). Identifiability of hidden Markov information sources and their minimum degrees of freedom. *IEEE Trans. Inf. Theory* **38,** 324–333.
[15] Johnson, M. A. and Taaffe, M. R. (1989). Matching moments to phase distributions: mixtures of Erlang distributions of common order. *Stoch. Models* **5,** 711–743.
[16] Johnson, M. A. and Taaffe, M. R. (1993). Tchebycheff systems for probabilistic analysis. *Amer. J. Math. Manag. Sci.* **13,** 83–111.
[17] Karlin, S. (1968). *Total Positivity,* Vol. I. Stanford University Press.
[18] Karlin, S. and Studden, W. J. (1966). *Tchebycheff Systems: With Applications in Analysis and Statistics*. John Wiley, New York.
[19] Lancaster, P. and Tismenetsky, M. (1985). *The Theory of Matrices*. Academic Press, New York.
[20] Latouche, G. and Ramaswami, V. (1999). *Introduction to Matrix Analytic Methods in Stochastic Modelling*. Society for Industrial and Applied Mathematics, Philadelphia, PA.
[21] Lipsky, L. (1992). *Queueing Theory: A Linear Algebraic Approach*. McMillan, New York.
[22] Luenberger, D. G. (1979). Positive linear systems. In *Introduction to Dynamic Systems: Theory, Models, and Applications*, John Wiley, New York.
[23] Neuts, M. F. (1975). Probability distributions of phase type. In *Liber Amicorum Prof. Emeritus H. Florin*, University of Louvain, pp. 173–206.
[24] Neuts, M. F. (1981). *Matrix-Geometric Solutions in Stochastic Models. An Algorithmic Approach*. The Johns Hopkins University Press, Baltimore, MD.
[25] O'Cinneide, C. A. (1989). On non-uniqueness of representations of phase-type distributions. *Stoch. Models* **5,** 247–259.
[26] O'Cinneide, C. A. (1990). Characterization of phase-type distributions. *Stoch. Models* **6,** 1–57.
[27] O'Cinneide, C. A. (1991). Phase-type distributions and invariant polytopes. *Adv. Appl. Prob.* **23,** 515–535.
[28] O'Cinneide, C. A. (1993). Triangular order of triangular phase-type distributions. *Stoch. Models* **9,** 507–529.
[29] Rockafellar, R. T. (1970). *Convex Analysis*. Princeton University Press.
[30] Ryden, T. (1996). On identifiability and order of continuous-time aggregated Markov chains, Markov-modulated Poisson processes and phase-type distributions. *J. Appl. Prob.* **33,** 640–653.
[31] Van de Liefvoort, A. and Heindl, A. (2005). Approximating matrix-exponential distributions by global randomization. *Stoch. Models* **21,** 669–693.