AN ABSTRACT OF THE DISSERTATION OF

Andreas Lang for the degree of Doctor of Philosophy in Statistics presented on August 11, 1994.

Title: An Empirical Evaluation of Parameter Approximation Methods for Phase-Type Distributions

Abstract approved: _____
Jeffrey L. Arthur

The stochastic modelling of real-world situations has gained widespread attention and has developed into a diversity of sub-areas, such as Markov processes and queueing theory. The widely accepted use of the exponential distribution in stochastic modelling is due (at least to a great extent) to simplifications which result from its lack-of-memory property, rather than based on empirical evidence. A significant development in providing greater modelling flexibility while retaining analytical tractability was the introduction of phase-type distributions, which constitute a relationship between Erlang's method of stages and absorption time distributions in finite-state Markov processes. A major advantage of phase-type distributions is their versatility and powerful applicability in solutions to a wide variety of stochastic models based on *matrix-analytic methods.*

The development of statistical fitting procedures and methods for the approximation of empirical distribution functions by phase-type distributions has been difficult due to several related aspects of these distributions, such as the highly non-linear fitting problem, the (often) large number of parameters to be approximated, the generally non-trivial relationship between the parameters and the shape of a phase-type distribution and the fact that representations (the sets of parameters) of phase-type distributions are typically not unique.

Although various methods for approximating the parameters of phase-type distributions have been discussed in the literature, the relative merits and weaknesses

of the proposed approximation methods have not been thoroughly studied. The presented analysis of two moment-matching and two maximum-likelihood based parameter approximation methods leads to several conclusions: *phase-type behaved* distribution functions can be adequately approximated by certain subsets of phase-type distributions, such as *mixture of Erlang* and *acyclic phase-type distributions.* Distributions exhibiting *non-phase-type behavior*, such as long non-exponential tails, low variability or sharp jumps are only approximated by high order phase-type distributions and might even lead to inappropriate results in some cases. The investigated parameter approximation methods involve widely varying performance characteristics and assume different levels of *a priori* knowledge about the properties of phase-type distributions. Based on the cases studied, we conclude that there does not yet exist a single superior parameter approximation method for phase-type distributions.

Key words and phrases: Phase-type distributions, stochastic models, matrix-analytic methods, Markov chains, parameter approximation, distribution fitting, moment matching, maximum likelihood, EM algorithm.

An Empirical Evaluation of Parameter Approximation Methods for Phase-Type Distributions

by
Andreas Lang

A DISSERTATION
submitted to
Oregon State University

in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Completed August 11, 1994
Commencement June 1995

Doctor of Philosophy dissertation of Andreas Lang presented on August 11, 1994

APPROVED:

Redacted for Privacy

Major Professor, representing Statistics

Redacted for Privacy

Chair of Department of Statistics

Redacted for Privacy

Dean of Graduate School

I understand that my dissertation will become part of the permanent collection of Oregon State University libraries. My signature below authorizes the release of my thesis to any reader upon request.

Redacted for Privacy

Andreas Lang, Author

For Katja

What one knows is, in youth, of little moment;
They know enough who know to learn.
The education of Henry Brooks Adams [1907], ch. 21.

ACKNOWLEDGEMENTS

The project presented here would neither have been started nor completed without the constant support and encouragement of numerous people.

Thanks is due first and foremost to my long-time advisor, Dr. Jeffrey Arthur, whose patience, goodwill and perpetual support I have certainly tried over time. His inspiration, guidance and critique throughout the research and writing of the dissertation was invaluable.

I am deeply indebted to my friend Katja for her (often long-distance) support and sacrifice. None of this would have occurred without her love, kindness and understanding patience. She has made this endeavor more meaningful.

Without the friendship of fellow students and people from outside academia, as well as the support and encouragement of my parents and family over the past five years, this project would have never gotten out of its infancy.

Special thanks go to my committee, Drs. Birkes, Butler, Ramsey, Rossignol and Waymire for their important suggestions, and to the Department of Statistics and its Chair, Dr. Seely, who provided financial support and the necessary work environment.

Several researchers have greatly contributed to the successful completion of this project: Drs. Asmussen (Aalborg University, Denmark), Bobbio (University of Brescia, Italy), Johnson (University of Illinois) and Schmickler (ANT Nachrichtentechnik GmbH, Germany). I also wish to express my appreciation to A. Pfening (Technical University of Budapest, Hungary) and M. Olsson (Chalmers University of Göteborg, Sweden) for always being available for fruitful discussions. A very warm thank you to Dr. Marcel Neuts (University of Arizona), who triggered my interest in the topic of this dissertation and who was always available for help of any sort.

Dr. Saunders' (Stanford University) kind permission to use the NPSOL software package involved in this analysis is gratefully acknowledged.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

LIST OF FIGURES

LIST OF FIGURES (Continued)

LIST OF FIGURES (Continued)

LIST OF TABLES

LIST OF TABLES (Continued)

An Empirical Evaluation of Parameter Approximation Methods for Phase-Type Distributions

Chapter 1

Introduction

The stochastic modelling of real-world situations has gained widespread attention and has developed into a diversity of sub-areas, such as Markov processes and queueing theory. Queueing problems arise in a variety of daily life situations; examples include queueing up in front of one of several cashiers at a grocery store, waiting lines at a post office, aircraft circling over an airport waiting for a runway to become available, broken machinery waiting to be fixed by a repair person, and so on. More recently, a number of queueing problems became apparent in connection with teletraffic theory, computer networks and data transmission applications (Asmussen [4]). Queueing problems present a great challenge to the probabilists trying to find mathematical models that are adequate in their representation of the considered real-world situation. Even in simple stochastic models, a significant hurdle in the search for explicit solutions is the always present need to find the conditional probability distributions that arise in their analysis. It is well known that the widely accepted use of the *exponential distribution* and the related *Poisson process* in stochastic modelling is due (at least to a great extent) to the significant simplification in conditioning which results from the remarkable lack-of-memory property of the exponential distribution, rather than based on empirical evidence (see Neuts [51]).

For that reason, it was necessary to develop generalizations of the family of exponential distributions, preserving most of their analytic tractability, yet providing more powerful tools to realistically model and represent situations arising in practice. One of the earliest results in this direction is due to Erlang [20], who suggested that many such random variables could be viewed as being composed of a finite number of *stages*, each being exponentially distributed with a common mean. The main

difference with the former viewpoint, where customers were being counted, was to count stages.

A further generalization of these familiar Erlang distributions is the class of *phase-type distributions.* Introduced by Neuts [53] as a natural probabilistic generalization of Erlang's approach, phase-type distributions constitute a relationship between Erlang's method of stages and absorption time distributions in finite-state Markov processes. They have an appealing algebraic representation in the form of vectors and matrices, and rely only on real arithmetic. Computations with phase-type distributions involve matrix multiplications and inversions in lieu of transform methods. The resulting expressions may often be given useful interpretations (see Neuts [51]). A more detailed introduction to phase-type distributions is given in chapter 2 of this dissertation. An extensive bibliography of articles, papers, theses and technical reports on phase-type distributions is available in Neuts [54].

One of the most useful features of phase-type distributions is that they allow for the use of *matrix-analytic methods* in stochastic models. Using these methods, cumbersome numerical integrations arising in the detailed study of many stochastic models are replaced by matrix operations that develop naturally in the analysis of structured Markov chains. These Markov models are typically generalizations of the chains embedded in the classical $GI/M/1$ and $M/G/1$ queues. Section 2.4 gives a brief introduction to these powerful methods. A comprehensive treatment of this topic is available in Neuts' two research monographs [51, 52]. A more informal description of the role of matrix-analytic methods in queueing theory is presented in Neuts [55].

One common argument against the use of phase-type approximations is that the exponential stages of the underlying Markov process often have no physical interpretation and are purely descriptive. However, approximating an unknown distribution with one that is analytically tractable provides the possibility to obtain solutions in at least a mathematically satisfying fashion. Moreover, many applications from areas such as epidemiology, drug kinetics and demography incorporate a nice prob-

abilistic interpretation of the scheme of exponential stages; examples can be found in Faddy [21, 22, 23].

The development of statistical fitting procedures and methods for the approximation of other probability distribution functions by phase-type distributions has been rather slow. The implementation of efficient numerical procedures for approximating the parameters of phase-type distributions remains an open problem which limits the use of this class of distributions in applications. The difficulty of the fitting problem is due to several related aspects of phase-type distributions, including:

- the fitting problem is highly nonlinear,
- the number of parameters to be approximated is often large,
- the relationship between the parameters and the shape of a phase-type distribution is generally non-trivial.
- representations (the set of parameters) of phase-type distributions are typically not unique. This is known as the *identifiability problem.* The question arises, under what conditions are two representations of a phase-type distribution equivalent? O'Cinneide [60] and Rydén [67] have studied this problem in detail. Their results leave unresolved the question of existence of necessary and sufficient conditions for determining the smallest number of parameters (*minimal order*) among all representations of a phase-type distribution. Lower bounds on the order of a phase-type distribution have been proposed by Aldous and Shepp [2], O'Cinneide [61, 62] and Maier [45]. Maier [45] also derived an upper bound on the order. Recently, the special class of *triangular* phase-type distributions has been studied by O'Cinneide [63], introducing a simpler quantity called *triangular order.* It is hoped that this development will suggest approaches to the study of order in the general phase-type setting.

Due to the lack of standardized test examples and comparative performance measures, an analysis of the merit of the various fitting methods has been difficult to establish. Notable exceptions are the results by Nielsen [59], Madsen and Nielsen [44]

and the development of a benchmark during a workshop in Aalborg, Denmark in 1991. This benchmark was first used in work done by Bobbio and Telek [12, 13]. An evaluation of moment-matching algorithms for fitting phase-type distributions on the basis of their performance in queueing models is presented in Johnson [34]. Attention is focused on the steady-state mean queue length of the $GI/M/1$ queue, where the approximating phase-type distributions are used as the interarrival time distributions.

Khoshgoftaar and Perros [38] present a comparison of three methods of parameter approximation (moment matching, minimum distance, maximum likelihood) for fitting general distributions with coefficient of variation greater than one by a Coxian distribution (see Cox [17]) of order two with real parameters, which can be represented by a hyperexponential distribution of order two. They conclude that in the case where the distribution to be approximated is known, the moment matching approach performed the best. However, in the case when the exact properties of the distribution to be approximated are not known (e.g. empirical data), the maximum likelihood and minimum distance criteria yielded better results.

This thesis evaluates several parameter approximation methods by defining appropriate performance measures and comparing the more promising fitting methods using a set of test examples. The experimental evaluations are based on samples from a variety of known probability distribution functions used as empirical data to be approximated. In addition, actual empirical data sets based on telecommunication applications are incorporated. Comparisons of several approximation methods are carried out using known properties of probability distribution functions, such as the moments and shape considerations. Several parameter approximation approaches, such as *moment matching* and *maximum likelihood based procedures*, in combination with *heuristics* and *nonlinear programming methods*, are involved in the study.

The remainder of this thesis is organized as follows: Chapter 2 introduces the fundamental properties of phase-type distributions, provides some examples, and discusses the related topics of phase-type renewal processes and matrix-analytic meth-

ods in stochastic modelling. Chapter 3 contains an overview of the most promising methods for approximating the parameters of phase-type distributions. Chapter 4 describes the data sets used in the computational experiments, and the performance measures used for comparisons. The results and their analysis are presented in chapter 5. Chapter 6 ends the study with a summary and discussion of the results, conclusions, and suggestions for future research.

Chapter 2
Probability Distributions of Phase-Type

This chapter is intended to provide a brief overview of phase-type distributions and their most important properties (section 2.1). Several important examples of phase-type distributions are given as special cases (section 2.2). A natural extension of the development of phase-type distributions are *renewal processes of phase-type*, which are introduced in section 2.3. A wide class of applications for phase-type distributions in stochastic models are generalizations called *matrix-analytic methods* of classical results in queueing theory. Mathematically less demanding sub-areas of these generalizations are known as *matrix-geometric methods*, and are introduced in section 2.4. Considered in this study are only continuous phase-type distributions with support on the nonnegative real numbers. There exists an entirely analogous development for phase-type distributions defined on the nonnegative integers.

Introduced by Neuts [53], phase-type distributions have gained widespread acceptance in algorithmic methods in queueing theory and related areas. A comprehensive treatment of probability distributions of phase-type, together with their properties and various applications, is available in Neuts [51].

2.1 Definitions and Properties

Consider a finite-state Markov process on the states $\{1, \ldots, m+1\}$, where the states labeled 1 through m are transient and state $m+1$ is absorbing. A random variable X has a phase-type distribution $F(\cdot)$, if X may be viewed as the time until absorption in such a Markov process.

The infinitesimal generator $\boldsymbol{Q}$ of such a Markov process is given by

$$\boldsymbol{Q} = \begin{bmatrix} \boldsymbol{T} & \boldsymbol{t}_0' \\ \mathbf{0} & 0 \end{bmatrix}, \tag{2.1}$$

where $\boldsymbol{T}$ is an $m \times m$ matrix satisfying $T_{ii} < 0$ for $1 \leq i \leq m$, and $T_{ij} \geq 0$ for $i \neq j$. The column vector $\boldsymbol{t}_0'$ corresponds to the absorbing state $m+1$ and is called the *exit rate vector*. Moreover, since $\boldsymbol{Q}$ is an infinitesimal generator, $\boldsymbol{T}\boldsymbol{e}' + \boldsymbol{t}_0' = \mathbf{0}'$, where $\mathbf{0}'$ is a $m \times 1$ column vector of zeroes and $\boldsymbol{e}'$ is a $m \times 1$ column vector with all components equal to one.

The initial probability vector of $\boldsymbol{Q}$ is given by $(\boldsymbol{\alpha}, \alpha_{m+1})$, where the $1 \times m$ row vector $\boldsymbol{\alpha}$ corresponds to the first m transient states of the Markov process, and $\boldsymbol{\alpha}\boldsymbol{e}' + \alpha_{m+1} = 1$.

The following lemmas and definition 1 are from Neuts [51].

Lemma 2.1 *The states $1, \ldots, m$ are transient if and only if the matrix $\boldsymbol{T}$ is non-singular.*

Lemma 2.2 *The probability distribution $F(\cdot)$ of the time until absorption in the state $m+1$ is given by*

$$F(x) = 1 - \boldsymbol{\alpha} \mathrm{e}^{\boldsymbol{T}x} \boldsymbol{e}', \qquad \text{for} \quad x \geq 0. \tag{2.2}$$

In lemma 2.2 above, note that $\mathrm{e}^{\boldsymbol{T}x} = \sum_{n=0}^{\infty} \frac{(\boldsymbol{T}x)^n}{n!}$ is a matrix, $\boldsymbol{\alpha}$ is a row vector and $\boldsymbol{e}'$ is a column vector, so that $\boldsymbol{\alpha}\mathrm{e}^{\boldsymbol{T}x}\boldsymbol{e}'$ is in fact a scalar.

Definition 1 *A probability distribution $F(\cdot)$ on $[0, \infty)$ is a* **distribution of phase-type (PH-distribution)** *if and only if it is the distribution of the time until absorption in a finite Markov process of the type defined in (2.1). The pair $(\boldsymbol{\alpha}, \boldsymbol{T})$ is called a* **representation** *of $F(\cdot)$.*

Definition 2 *The m transient states of the Markov process corresponding to a phase-type distribution are called* **phases**.

Definition 3 *The number of transient states of the Markov process corresponding to a phase-type distribution is called the* **order** *of the representation (or distribution).*

The following statements are readily verified (see Neuts [51]):

- The distribution $F(\cdot)$ has a jump of height α_{m+1} at $x = 0$, and its density portion $F'(x)$ on $(0, \infty)$ is given by

$$f(x) = F'(x) = \boldsymbol{\alpha} \mathrm{e}^{\boldsymbol{T}x} \boldsymbol{t}_0'. \tag{2.3}$$

- The Laplace-Stieltjes transform $\tilde{F}(s)$ of $F(\cdot)$ is given by

$$\tilde{F}(s) = \alpha_{m+1} + \boldsymbol{\alpha}(s\boldsymbol{I} - \boldsymbol{T})^{-1}\boldsymbol{t}_0', \qquad \text{for} \qquad \mathrm{Re}(s) \geq 0. \tag{2.4}$$

- The noncentral moments μ_i' of $F(\cdot)$ are all finite and given by

$$\mu_i' = \int_0^\infty x^i \, dF(x) = (-1)^i i! (\boldsymbol{\alpha} \boldsymbol{T}^{-i} \boldsymbol{e}'), \qquad \text{for} \quad i \geq 0. \tag{2.5}$$

Thus, one can see that computations involving phase-type distributions require essentially matrix calculations, i.e. matrix exponentials for distributions and densities, and matrix inverses for the moments. The matrix $\boldsymbol{T}$ may always be inverted, for the definition of phase-type distributions implies that $\boldsymbol{T}$ is nonsingular (see lemma 2.1).

The class of phase-type distributions is closed under finite mixtures and convolutions, under infinite mixtures with a discrete phase-type density as the mixing density, and under the formation of maxima and minima. In each case, knowing the representations for the distributions involved, it is possible to obtain a representation for the new distribution. For a thorough discussion of closure properties of phase-type distributions, see Neuts [51] and Assaf and Levikson [7].

A number of immediate applications result from these closure properties (see Neuts [51, 55]). For example, the stationary waiting time distribution of an $M/PH/1$ queue is itself of phase-type and has a simple representation. Another example is the $GI/G/1$ queue with interarrival distribution $A(\cdot)$ and service distribution $B(\cdot)$

(this example is described in Asmussen [5]). For such a model, no explicit or even computationally tractable solution for the steady-state waiting time distribution $W(A, B)$ can be found. Assuming, however, that the service distribution $B(\cdot)$ is of phase-type, $W(A, B)$ is again of phase-type. This implies that the $GI/G/1$ queue may be approximated by a $GI/PH/1$ queue, with the steady-state waiting time distribution of the latter model computed and used as an approximation to the waiting time distribution of the original queueing model.

A third important implication of the closure properties is that mixtures of phase-type distributions may be used to approximate probability distributions which reflect qualitatively different features of a queueing model. For instance, two different kinds of customers or servers (with different service times) may be modeled by using a mixture of two (or more) Erlang distributions with possibly different parameters. The resulting distribution will still be of phase-type, and can be used in computing performance measures for queueing models.

2.2 Examples

1. The **exponential distribution** with parameter $\lambda > 0$ and density

$$f(x) = \lambda \mathrm{e}^{-\lambda x} \qquad \text{for} \quad 0 < x < \infty. \tag{2.6}$$

It is easily seen that the corresponding Markov process has only one transient state, i.e. the phase-type representation of the exponential distribution is of order one. The representation is $\boldsymbol{\alpha} = \alpha_1 = 1$ and

$$\boldsymbol{T} = [-\lambda], \qquad \boldsymbol{t}_0' = [\lambda]. \tag{2.7}$$

The corresponding transition diagram is shown in figure 2.1.

2. The **Erlang distribution** E_k with k phases (of order k), i.e. a Gamma distribution with integral shape parameter k and density

$$f(x) = \frac{x^{k-1}}{(k-1)!}\lambda^k \mathrm{e}^{-\lambda x}, \qquad \text{for} \quad 0 < x < \infty. \tag{2.8}$$

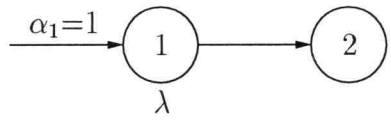

Figure 2.1. Transition diagram for an exponential distribution.

The classical Erlang distribution corresponds to a convolution of k exponential random variables, each with parameter $\lambda > 0$. The phase-type representation is $\boldsymbol{\alpha} = (1, 0, \ldots, 0)$ and

$$\boldsymbol{T} = \begin{bmatrix} -\lambda & \lambda & 0 & \cdots & 0 & 0 \\ 0 & -\lambda & \lambda & \cdots & 0 & 0 \\ \vdots & & & \ddots & & \vdots \\ 0 & 0 & 0 & \cdots & -\lambda & \lambda \\ 0 & 0 & 0 & \cdots & 0 & -\lambda \end{bmatrix}, \quad \boldsymbol{t}_0' = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ \lambda \end{bmatrix}. \tag{2.9}$$

The transition diagram for an order-k Erlang distribution is shown in figure 2.2.

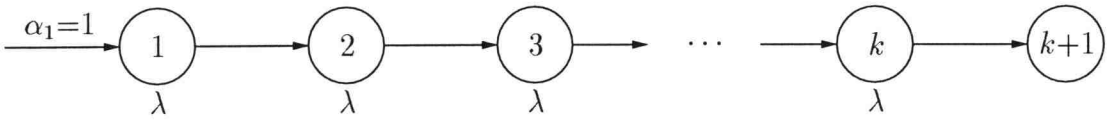

Figure 2.2. Transition diagram for an order-k Erlang distribution.

3. The **generalized Erlang distribution** GE_k of order k is the convolution of k exponential densities, each with a (possibly different) positive parameter $\lambda_1, \ldots, \lambda_k$. It has the phase-type representation $\boldsymbol{\alpha} = (1, 0, \ldots, 0)$ and

$$\boldsymbol{T} = \begin{bmatrix} -\lambda_1 & \lambda_1 & 0 & \cdots & 0 & 0 \\ 0 & -\lambda_2 & \lambda_2 & \cdots & 0 & 0 \\ \vdots & & & \ddots & & \vdots \\ 0 & 0 & 0 & \cdots & -\lambda_{k-1} & \lambda_{k-1} \\ 0 & 0 & 0 & \cdots & 0 & -\lambda_k \end{bmatrix}, \quad \boldsymbol{t}_0' = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ \lambda_k \end{bmatrix}. \tag{2.10}$$

The transition diagram for an order-k generalized Erlang distribution is shown in figure 2.3.

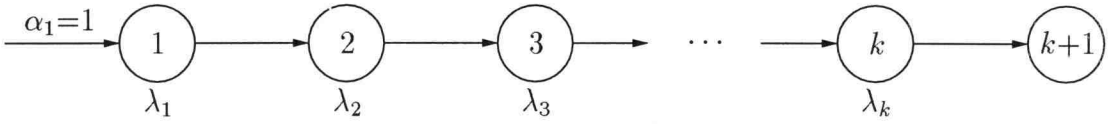

Figure 2.3. Transition diagram for an order-k generalized Erlang distribution.

4. The **Hyperexponential distribution** H_k of order k, i.e. a mixture of k exponential densities with positive parameters $\lambda_1, \lambda_2, \ldots, \lambda_k$, respectively, and density

$$f(x) = \sum_{i=1}^{k} \alpha_i \lambda_i \mathrm{e}^{-\lambda_i x} \qquad \text{for} \quad 0 < x < \infty. \tag{2.11}$$

The phase-type representation is $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_k)$ and

$$\boldsymbol{T} = \begin{bmatrix} -\lambda_1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & -\lambda_2 & 0 & \cdots & 0 & 0 \\ \vdots & & & \ddots & & \vdots \\ 0 & 0 & 0 & \cdots & -\lambda_{k-1} & 0 \\ 0 & 0 & 0 & \cdots & 0 & -\lambda_k \end{bmatrix}, \quad \boldsymbol{t}_0' = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_{k-1} \\ \lambda_k \end{bmatrix}. \tag{2.12}$$

The transition diagram for an order-k hyperexponential distribution is shown in figure 2.4.

5. The **Acyclic Phase-Type distribution** APH_k of order k. Only transitions $i \to j$ for $j \geq i + 1$ are allowed. Three equivalent canonical forms (i.e. unique representations) of acyclic phase-type distributions are possible (c.f. Bobbio and Cumani [11]):

   (a) A series model of exponential phases with parameters $\lambda_1, \lambda_2, \ldots, \lambda_k$ and

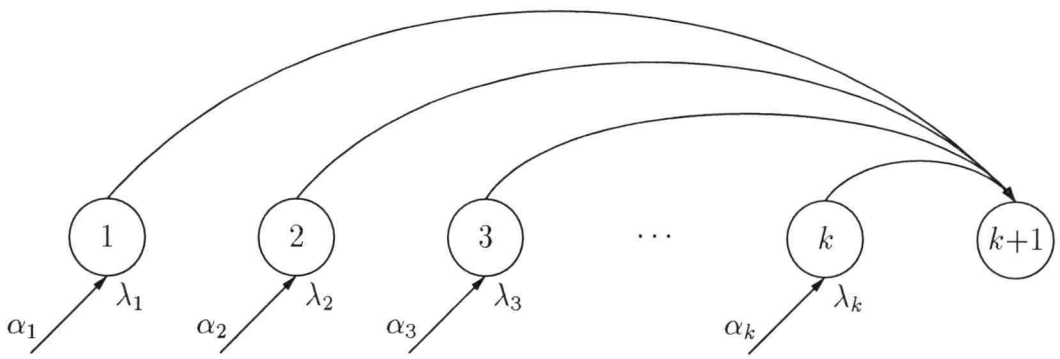

Figure 2.4. Transition diagram for an order-k hyperexponential distribution

initial probability vector $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_k)$:

$$\boldsymbol{T} = \begin{bmatrix} -\lambda_1 & \lambda_1 & 0 & \cdots & 0 & 0 \\ 0 & -\lambda_2 & \lambda_2 & \cdots & 0 & 0 \\ \vdots & & & \ddots & & \vdots \\ 0 & 0 & 0 & \cdots & -\lambda_{k-1} & \lambda_{k-1} \\ 0 & 0 & 0 & \cdots & 0 & -\lambda_k \end{bmatrix}, \quad \boldsymbol{t}_0' = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ \lambda_k \end{bmatrix}. \tag{2.13}$$

The transition diagram for this canonical form of an order-k acyclic phase-type distribution is shown in figure 2.5. The real parameters must satisfy the following conditions:

$$\begin{cases} 0 \leq \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_k \\ 0 \leq \alpha_i \leq 1 \qquad \text{for} \quad i = 1, 2, \ldots, k, \\ \sum_{j=1}^{k} \alpha_j = 1. \end{cases} \tag{2.14}$$

(b) A series model of exponential phases with parameters $\lambda_1, \lambda_2, \ldots, \lambda_k$ and transitions $i \rightarrow j$ with probabilities p_{ij} for $j \geq i + 1$. The phase-type

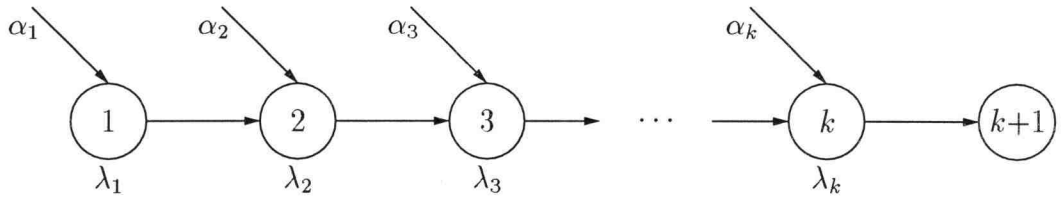

Figure 2.5. Canonical form 1 for an order-k acyclic phase-type distribution.

representation is $\boldsymbol{\alpha} = (1, 0, \ldots, 0)$ and

$$\boldsymbol{T} = \begin{bmatrix} -\lambda_1 & \gamma_{12} & \gamma_{13} & \cdots & \gamma_{1,k-1} & \gamma_{1k} \\ 0 & -\lambda_2 & \gamma_{23} & \cdots & \gamma_{2,k-1} & \gamma_{2k} \\ \vdots & & & \ddots & & \vdots \\ 0 & 0 & 0 & \cdots & -\lambda_{k-1} & \gamma_{k-1,k} \\ 0 & 0 & 0 & \cdots & 0 & -\lambda_k \end{bmatrix}, \boldsymbol{t}_0' = \begin{bmatrix} \gamma_{1,k+1} \\ \gamma_{2,k+1} \\ \\ \gamma_{k-1,k+1} \\ \lambda_k \end{bmatrix}, \tag{2.15}$$

where $\gamma_{ij} = p_{ij}\lambda_i$ and $\sum_{j=i+1}^{k+1} p_{ij} = 1$, for $i = 1, \ldots, k$.

The transition diagram for this canonical form of an order-k acyclic phase-type distribution is shown in figure 2.6. The following requirements on the real parameters must be satisfied:

$$\begin{cases} 0 \leq \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_k \\ 0 \leq p_{ij} \leq 1 \quad \text{for} \quad i = 1, 2, \ldots, k-1, \; j \geq i+1 \\ \sum_{j=i+1}^{k+1} p_{ij} = 1 \quad \text{for} \quad i = 1, 2, \ldots, k. \end{cases} \tag{2.16}$$

(c) The third canonical form of an acyclic phase-type distribution is depicted by the representation of the family of Coxian distributions below. However, the parameters of the exponential phases and the branching probabilities are real and must satisfy

$$\begin{cases} 0 \leq \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_k \\ 0 \leq p_i \leq 1 \quad \text{for} \quad i = 1, 2, \ldots, k-1, \; p_k = 1. \end{cases} \tag{2.17}$$

6. The **Coxian distribution** C_k of order k. Only transitions $i \rightarrow i+1$ and $i \rightarrow k+1$ (the absorbing state) are allowed, with (possibly complex) probabilities

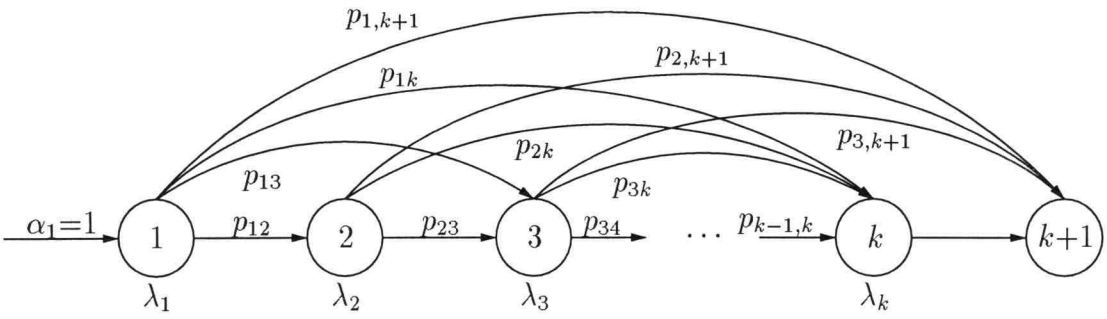

Figure 2.6. Canonical form 2 for an order-k acyclic phase-type distribution.

p_i and $q_i = 1 - p_i$, respectively, for $i = 1, \ldots, k-1$. The individual exponential phases, however, have different parameters $\lambda_1, \lambda_2, \ldots, \lambda_k$ (possibly complex), and the representation is $\boldsymbol{\alpha} = (1, 0, \ldots, 0)$ and

$$\boldsymbol{T} = \begin{bmatrix} -\lambda_1 & \gamma_1 & 0 & \cdots & 0 & 0 \\ 0 & -\lambda_2 & \gamma_2 & \cdots & 0 & 0 \\ \vdots & & & \ddots & & \vdots \\ 0 & 0 & 0 & \cdots & -\lambda_{k-1} & \gamma_{k-1} \\ 0 & 0 & 0 & \cdots & 0 & -\lambda_k \end{bmatrix}, \quad \boldsymbol{t}_0' = \begin{bmatrix} q_1\lambda_1 \\ q_2\lambda_2 \\ \\ q_{k-1}\lambda_{k-1} \\ \lambda_k \end{bmatrix}. \tag{2.18}$$

where $\gamma_i = p_i\lambda_i$. The transition diagram for an order-k Coxian distribution is shown in figure 2.7. Here the *a priori* probability of zero life-time is assumed to be zero, i.e. $\alpha_1 = 1$. The general development, however, allows for a positive probability of immediately going to the absorbing state (zero life-time), i.e. $\alpha_1 \leq 1$.

This family of distributions was introduced by Cox [17] and has been very popular in stochastic modeling. Cox showed that schemes different than the one above – such as allowing at each phase i the possibility of entering the absorbing state or any other phase $j \neq i$ (not only phase $i+1$) – do not lead to a class of distributions more general than (2.18). In fact, the entire family of phase-type distributions is a subset of the family of Coxian distributions.

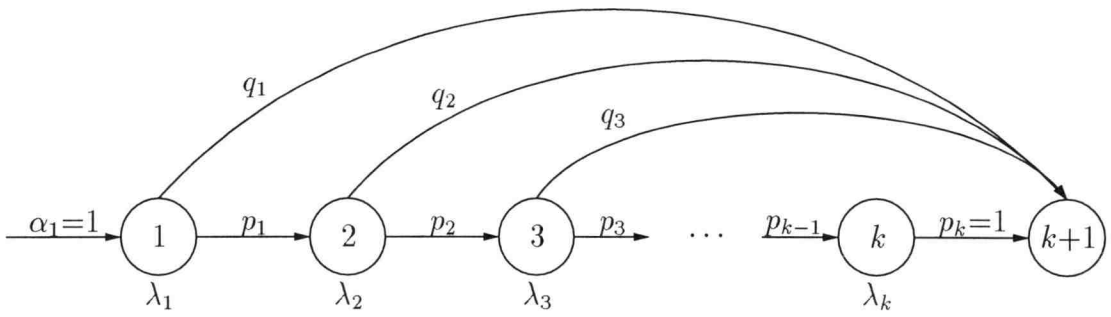

Figure 2.7. Transition diagram for an order-k Coxian distribution.

2.3 Renewal Processes of Phase-Type

The notion of a *renewal process of phase-type* was introduced by Neuts [57], where a thorough discussion may be found. Other references on this topic are Neuts [51] and Lucantoni and Ramaswami [43].

Definition 4 *A renewal process where the times between renewals are i.i.d. random variables distributed according to a phase-type distribution with representation* $(\boldsymbol{\alpha}, \boldsymbol{T})$ *is called a* **renewal process of phase-type**.

Based on the Markovian construction of a phase-type distribution, a phase-type renewal process may be viewed as follows. Suppose the Markov process $\boldsymbol{Q}$ (c.f. equation (2.1)) enters the absorbing state $m+1$. Instantaneously restart the process $\boldsymbol{Q}$ according to the initial probability vector $(\boldsymbol{\alpha}, \alpha_{m+1})$. For convenience, assume that $\alpha_{m+1} = 0$, so that the phase-type distribution according to this Markov process does not have an atom at 0. Repeating this procedure indefinitely, we construct a new Markov process on the state space $\{1, 2, \ldots, m\}$ (the state $m+1$ is an instantaneous state) with initial probability vector $\boldsymbol{\alpha}$ and infinitesimal generator

$$\boldsymbol{Q}^* = \boldsymbol{T} + \boldsymbol{t}_0' \boldsymbol{\alpha} \tag{2.19}$$

The point process obtained by this procedure is a *renewal process of phase-type.*

Let $\{N(t), t \geq 0\}$ denote the number of renewals in $(0,t]$, and $\{J(t), t \geq 0\}$ the Markov process with generator $\boldsymbol{Q}^*$. Define the $m \times m$ matrices $\boldsymbol{P}(n,t) = \{P_{ij}(n,t)\}$ by

$$P_{ij}(n,t) = P\{N(t) = n,\ J(t) = j \mid N(0) = 0,\ J(0) = i\}, \tag{2.20}$$

for $n \geq 0, t \geq 0, 1 \leq i \leq m, 1 \leq j \leq m$.

$P_{ij}(n,t)$ is the conditional probability, given the initial phase is i, that at time t the process $\boldsymbol{Q}^*$ is in state j, and n renewals have occured. As shown in Neuts [51], the matrices $\boldsymbol{P}(n,t)$ satisfy the Chapman-Kolmogorov differential equations

$$\begin{aligned} \frac{d}{dt}\boldsymbol{P}(0,t) &= \boldsymbol{T}\boldsymbol{P}(0,t) = \boldsymbol{P}(0,t)\boldsymbol{T}, \\ \frac{d}{dt}\boldsymbol{P}(n,t) &= \boldsymbol{P}(n,t)\boldsymbol{T} + \boldsymbol{P}(n-1,t)\boldsymbol{t}_0'\boldsymbol{\alpha}, \qquad \text{for} \quad n \geq 1. \end{aligned} \tag{2.21}$$

2.4 Matrix-Geometric Solutions

This section is a brief introduction to matrix-geometric solutions of certain stochastic models, and is intended to illustrate the power these methods provide regarding computationally tractable solutions for a large number of problems arising in practice.

A standard approach to the detailed study of many stochastic models is investigation of the presence of embedded Markov chains. The following results are from Neuts [51, 55] and Asmussen [4], where a thorough description may be found along with discussions of more complicated cases.

Two classic examples of stochastic models exhibiting embedded Markov chains are the $M/G/1$ and $GI/M/1$ queues. For the $M/G/1$ queue, having exponential interarrival times with parameter λ and service times following a continuous distribution function $B(x)$, $x \geq 0$, the idea is to observe that the queue length *just after*

departures is a Markov chain with transition probability matrix

$$\boldsymbol{P}_1 = \begin{bmatrix} a_0 & a_1 & a_2 & a_3 & a_4 & \cdots \\ a_0 & a_1 & a_2 & a_3 & a_4 & \cdots \\ 0 & a_0 & a_1 & a_2 & a_3 & \cdots \\ 0 & 0 & a_0 & a_1 & a_2 & \cdots \\ 0 & 0 & 0 & a_0 & a_1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \end{bmatrix}, \tag{2.22}$$

where

$$a_j = \int_0^\infty \mathrm{e}^{-\lambda t} \frac{(\lambda t)^j}{j!} dB(t) \qquad \text{for} \quad j = 0, 1, \ldots. \tag{2.23}$$

Generalizations of Markov chains $\boldsymbol{P}_1$ of $M/G/1$ type and their *matrix-analytic solutions* are extensively discussed in Neuts' monograph [52] and will not be covered here.

Considering the $GI/M/1$ queue, with times between successive arrivals distributed according to the continuous distribution function $A(x)$, $x \geq 0$, and service times having an exponential distribution with parameter μ, we note that the queue length *just prior to arrivals* is a Markov chain with transition probability matrix

$$\boldsymbol{P}_2 = \begin{bmatrix} b_0 & a_0 & 0 & 0 & 0 & \cdots \\ b_1 & a_1 & a_0 & 0 & 0 & \cdots \\ b_2 & a_2 & a_1 & a_0 & 0 & \cdots \\ b_3 & a_3 & a_2 & a_1 & a_0 & \cdots \\ b_4 & a_4 & a_3 & a_2 & a_1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \end{bmatrix}, \tag{2.24}$$

where, for $j = 0, 1, \ldots$

$$a_j = \int_0^\infty \mathrm{e}^{-\mu t} \frac{(\mu t)^j}{j!} dA(t), \tag{2.25}$$

and

$$b_j = 1 - \sum_{k=0}^{j} a_k. \tag{2.26}$$

The classic result due to Khinchine [37] states that the queue length at arrivals in the stationary $GI/M/1$ queue has a geometric distribution. A corresponding

result holds for widely differing stochastic models with embedded Markov chains which are generalizations of the paradigm $\boldsymbol{P}_2$ for the $GI/M/1$ queue. The transition probability matrix $\boldsymbol{P}$ of an embedded *Markov chain of the* $GI/M/1$ *type* is of the (canonical) block-partitioned form

$$\boldsymbol{P} = \begin{bmatrix} \boldsymbol{B}_0 & \boldsymbol{A}_0 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots \\ \boldsymbol{B}_1 & \boldsymbol{A}_1 & \boldsymbol{A}_0 & \mathbf{0} & \mathbf{0} & \cdots \\ \boldsymbol{B}_2 & \boldsymbol{A}_2 & \boldsymbol{A}_1 & \boldsymbol{A}_0 & \mathbf{0} & \cdots \\ \boldsymbol{B}_3 & \boldsymbol{A}_3 & \boldsymbol{A}_2 & \boldsymbol{A}_1 & \boldsymbol{A}_0 & \cdots \\ \boldsymbol{B}_4 & \boldsymbol{A}_4 & \boldsymbol{A}_3 & \boldsymbol{A}_2 & \boldsymbol{A}_1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \end{bmatrix}, \tag{2.27}$$

where the elements of the matrix $\boldsymbol{P}$ are now $m \times m$ nonnegative matrices. The matrix $\boldsymbol{P}$ in (2.27) is called *canonical* because the crucial feature is the particular structure of the overall transition probability matrix.

The following main result is stated only for the common case where the matrix $\boldsymbol{A}$, given by

$$\boldsymbol{A} = \sum_{j=0}^{\infty} \boldsymbol{A}_j, \tag{2.28}$$

is an irreducible stochastic matrix.

Denoting by $\boldsymbol{\pi}$ the invariant probability vector of the matrix $\boldsymbol{A}$, (i.e. $\boldsymbol{\pi}\boldsymbol{A} = \boldsymbol{\pi}$, $\boldsymbol{\pi}\boldsymbol{e}' = 1$) and defining the column vector $\boldsymbol{\beta}'$ by

$$\boldsymbol{\beta}' = \sum_{j=1}^{\infty} j\boldsymbol{A}_j\boldsymbol{e}', \tag{2.29}$$

then the Markov chain $\boldsymbol{P}$ is positive recurrent if and only if

$$\boldsymbol{\pi}\boldsymbol{\beta}' > 1. \tag{2.30}$$

Let $\boldsymbol{x}$ denote the invariant probability vector of $\boldsymbol{P}$, so that $\boldsymbol{x}\boldsymbol{P} = \boldsymbol{x}$, $\boldsymbol{x}\boldsymbol{e}' = 1$, and partition $\boldsymbol{x}$ into $1 \times m$ vectors as

$$\boldsymbol{x} = [\boldsymbol{x_0}, \boldsymbol{x_1}, \ldots]. \tag{2.31}$$

Then, for $i = 0, 1, \ldots$

$$\boldsymbol{x_i} = \boldsymbol{x_0} \boldsymbol{R}^i \tag{2.32}$$

for some $(m \times m)$ matrix $\boldsymbol{R}$.

A probability vector $\boldsymbol{x}$ which satisfies (2.32) is called a *matrix-geometric probability vector*. If the Markov chain $\boldsymbol{P}$ is positive recurrent (c.f. condition (2.30)), the matrix $\boldsymbol{R}$ is the minimal nonnegative solution to the nonlinear matrix equation

$$\boldsymbol{R} = \sum_{j=0}^{\infty} \boldsymbol{R}^j \boldsymbol{A}_j. \tag{2.33}$$

The vector $\boldsymbol{x_0}$ is obtained by solving the system of linear equations

$$\boldsymbol{x_0} = \boldsymbol{x_0} \sum_{j=0}^{\infty} \boldsymbol{R}^j \boldsymbol{B}_j, \qquad \boldsymbol{x_0}(\boldsymbol{I} - \boldsymbol{R})^{-1} \boldsymbol{e}' = 1. \tag{2.34}$$

We note that the first equation in (2.34) corresponds to the steady-state equations for the first m states. The second equation in (2.34) is the normalizing condition which guarantees that the vector $\boldsymbol{x}$ is a probability vector. It can be shown that the condition (2.30) implies that all m eigenvalues of the matrix $\boldsymbol{R}$ are less than one in modulus ($\boldsymbol{R}$ is of spectral radius less than one), which in turn implies that the matrix $(\boldsymbol{I} - \boldsymbol{R})$ is nonsingular.

The stochastic matrix $\boldsymbol{R}$, called the *rate matrix* of the Markov chain $\boldsymbol{P}$, has an interesting probabilistic interpretation. R_{jv} is the expected number of visits to the state $(i+1, v)$ before the first return to the set of states $(i, \cdot)$, given the chain $\boldsymbol{P}$ started in state (i, j). For a statement and proof see Neuts [51]).

Knowing the matrix $\boldsymbol{R}$ and the vector $\boldsymbol{x_0}$, many quantities of the underlying stochastic model may be computed. For instance, the marginal density $\{q_k,\ k \geq 0\}$, given by

$$q_k = \boldsymbol{x_k} \boldsymbol{e}' = \boldsymbol{x_0} \boldsymbol{R}^k \boldsymbol{e}', \tag{2.35}$$

is, in many queueing models, the steady-state density of the queue length.

The following example is intended to show the power of matrix-geometric solutions by presenting easily computable formulae for various quantities of interest,

using an application from queueing theory. This example is thoroughly discussed in Neuts [51], and the key results are also stated in Lucantoni and Ramaswami [43] and Neuts [58].

Example. Consider the $GI/PH/1$ queue, where the service times are random variables with common phase-type distribution function of order m with representation $(\boldsymbol{\gamma}, \boldsymbol{S})$. For simplicity, assume that $\gamma_{m+1} = 0$. The mean service time is given by $\mu_1' = -\boldsymbol{\gamma}\boldsymbol{S}^{-1}\boldsymbol{e}'$. The interarrival times have a probability distribution $G(\cdot)$ with mean λ_1'. Denote the queue length immediately *prior* to the r^{th} arrival by I_r, the phase state immediately *after* the r^{th} arrival by J_r and the interarrival time between the $(r-1)^{st}$ and the r^{th} arrival by τ_r. Then the sequence of triples $\{(I_r, J_r, \tau_r), r \geq 0\}$ is a *Markov renewal sequence* on the state space $\{(i, j, x), i \geq 0, 1 \leq j \leq m, x \geq 0\}$. The transition probability matrix $\widetilde{\boldsymbol{P}}(\cdot)$ of that Markov renewal sequence is given by

$$\widetilde{\boldsymbol{P}}(x) = \begin{bmatrix} \widetilde{\boldsymbol{B}}_0(x) & \widetilde{\boldsymbol{A}}_0(x) & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots \\ \widetilde{\boldsymbol{B}}_1(x) & \widetilde{\boldsymbol{A}}_1(x) & \widetilde{\boldsymbol{A}}_0(x) & \mathbf{0} & \mathbf{0} & \cdots \\ \widetilde{\boldsymbol{B}}_2(x) & \widetilde{\boldsymbol{A}}_2(x) & \widetilde{\boldsymbol{A}}_1(x) & \widetilde{\boldsymbol{A}}_0(x) & \mathbf{0} & \cdots \\ \widetilde{\boldsymbol{B}}_3(x) & \widetilde{\boldsymbol{A}}_3(x) & \widetilde{\boldsymbol{A}}_2(x) & \widetilde{\boldsymbol{A}}_1(x) & \widetilde{\boldsymbol{A}}_0(x) & \cdots \\ \widetilde{\boldsymbol{B}}_4(x) & \widetilde{\boldsymbol{A}}_4(x) & \widetilde{\boldsymbol{A}}_3(x) & \widetilde{\boldsymbol{A}}_2(x) & \widetilde{\boldsymbol{A}}_1(x) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \end{bmatrix}, \tag{2.36}$$

where

$$\widetilde{\boldsymbol{A}}_k(x) = \int_0^x \boldsymbol{P}(k, t)\, dG(t), \qquad \text{for} \quad k \geq 0, \tag{2.37}$$

and the $m \times m$ matrices $\boldsymbol{P}(k, t)$ are defined as in equation (2.20).

The following are key results:

- The $GI/PH/1$ queue is *stable* (i.e. exhibits stationary behavior) if and only if the Markov chain $\widetilde{\boldsymbol{P}}$ is positive recurrent, i.e. condition (2.30) is satisfied. Denoting $\boldsymbol{\pi}$ as the invariant probability vector of $\widetilde{\boldsymbol{A}} = \sum_{j=0}^{\infty} \widetilde{\boldsymbol{A}}_j$ and $\boldsymbol{\beta}' = \sum_{j=1}^{\infty} j\widetilde{\boldsymbol{A}}_j \boldsymbol{e}'$, it can be shown (c.f. Neuts [51], chapter 4) that the inner product $\boldsymbol{\pi\beta}'$ reduces to λ_1'/μ_1', where λ_1' and μ_1' are the (finite) mean interarrival and service times, respectively. This shows that the Markov chain $\widetilde{\boldsymbol{P}}$ is positive recurrent if and only if the classical equilibrium condition $\lambda_1' > \mu_1'$ holds.

- Let $\boldsymbol{x_k} = \{x_{kj},\ 1 \leq j \leq m\}$ for $k \geq 0$, where x_{kj} denotes the joint steady-state probability that an arrival finds k customers in the queue with the server in phase j of the service process. Then the invariant probability vector $\boldsymbol{x} = [\boldsymbol{x_0}, \boldsymbol{x_1}, \ldots]$ of $\widetilde{\boldsymbol{P}}$ is given by

$$[x_{k1}, x_{k2}, \ldots, x_{km}] = \boldsymbol{x_k} = C\boldsymbol{\gamma R}^k, \qquad \text{for} \quad k \geq 0, \tag{2.38}$$

  where

$$C = (\boldsymbol{\gamma}(\boldsymbol{I} - \boldsymbol{R})^{-1}\boldsymbol{e}')^{-1}. \tag{2.39}$$

  and the matrix $\boldsymbol{R}$ is the minimal nonnegative solution to the nonlinear matrix equation (2.33).

- The probability g_k that the queue length (prior to an arrival) equals k is

$$g_k = C\boldsymbol{\gamma R}^k\boldsymbol{e}', \qquad \text{for} \quad k \geq 0. \tag{2.40}$$

- The expected queue length L_1 (prior to arrivals) is given by

$$L_1 = C\boldsymbol{\gamma}(\boldsymbol{I} - \boldsymbol{R})^{-2}\boldsymbol{e}' - 1. \tag{2.41}$$

- The probability y_k that the queue length equals k at an arbitrary time is

$$\begin{aligned} y_0 &= 1 - \frac{\mu_1'}{\lambda_1'} = 1 - \rho, \\ y_k &= C\rho\boldsymbol{\gamma R}^k\boldsymbol{e}' + C(\lambda_1')^{-1}\boldsymbol{\gamma R}^{k-1}(\boldsymbol{R} - \boldsymbol{I})\boldsymbol{S}^{-1}\boldsymbol{e}', \text{ for } k \geq 1. \end{aligned} \tag{2.42}$$

- The mean L_2 of the density (2.42) above is

$$L_2 = \rho L_1 - C(\lambda_1')^{-1}\boldsymbol{\gamma}(\boldsymbol{I} - \boldsymbol{R})^{-1}\boldsymbol{S}^{-1}\boldsymbol{e}', \tag{2.43}$$

  where L_1 is given in formula (2.41).

- The expected number of customers Z served during a busy period is given by

$$Z = \boldsymbol{\gamma}(\boldsymbol{I} - \boldsymbol{R})^{-1}\boldsymbol{e}'. \tag{2.44}$$

Furthermore, an algorithm for computing the stationary waiting time distribution in terms of $\boldsymbol{R}$ is presented in Ramaswami and Lucantoni [65].

In order to recognize the connection of matrix-geometric solutions with classical results in queueing theory literature, we note that many examples of Markov chains of the type (2.27) have been discussed using methods based on complex analysis. The typical equation arising there is $\det[z\boldsymbol{I} - \sum_{k=0}^{\infty} \boldsymbol{A}_k z^k] = 0$, whose roots in the unit disk are usually solved by an argument based on *Rouché's Theorem.* There are exactly m such roots, which are the eigenvalues of the rate matrix $\boldsymbol{R}$ (c.f. Neuts [51, 58]). It should be pointed out that matrix-geometric solutions are straightforward matrix versions of the scalar equations mentioned above from classic queueuing theory.

Chapter 3

Parameter Approximation Methods

This chapter describes various parameter approximation methods for phase-type distributions and other related distributions that are subsets of the class of Coxian distributions. Phase-type as well as non-phase-type distributions are included in the description of these sub-families of Coxian distributions in section 3.1. The fitting methods discussed in this study, however, only encompass phase-type distributions as defined in chapter 2.

Most parameter approximation methods used for phase-type distributions can be separated into two general classes: *moment-matching* techniques and *maximum-likelihood based* approaches. An overview of the methods considered in this study is presented in sections 3.2 and 3.3.

It is important to note that the literature often uses terms such as *parameter estimation* and *maximum likelihood estimation.* However, based on the fact that the test distributions used for evaluating the different fitting methods are not of phase-type, nor were the data generated in the form of a random sample (see chapter 4), the use of such terms according to their proper statistical definition is prohibited. We shall therefore employ a different nomenclature, using the terms *parameter approximation* and *maximum likelihood based* methods.

3.1 Selection Subsets and Related Families of Distributions

The difficulties in approximating the parameters of phase-type distributions are attributable to several factors. As mentioned in chapter 1, the usually large number of

parameters, the non-linearity of the approximation problem, the complex relationship between the parameters and distribution properties and the non-uniqueness of the representation of phase-type distributions present non-trivial problems.

It is known that the class of phase-type distributions is *dense*, that is, for any cdf $G(\cdot)$ on $(0, \infty)$, there exists a sequence of order-n phase-type distributions $\{PH_n\}$ such that PH_n converges weakly to $G(\cdot)$ as $n \to \infty$ (a formal statement and proof is in Asmussen [4], p. 76). Thus, any cdf $G(\cdot)$ defined on $(0, \infty)$ can be approximated to an arbitrary precision by a phase-type distribution of appropriate order. A good approximation, however, may require a large number of phases. The large number of parameters presents problems, both in the approximation procedure as well as in applications. Matrix-analytic solutions to stochastic models are virtually useless if the dimensions of the matrices involved and their number of non-zero elements is large, since matrix operations require significant resources in terms of computing power.

Most approaches to fitting phase-type distributions (or other sub-families of Coxian distributions) that have been proposed in the literature rely on restricting the empirical distributions to be approximated by certain subsets of the family of phase-type distributions, called the *selection subset*. The benefit of such a restriction is avoiding over-parameterization by keeping the number of parameters to be approximated small and the structure of the underlying Markov processes simple and sparse. The numerical difficulties arising in the mostly non-linear approximation procedures may therefore be kept within certain bounds and the use of the approximating phase-type distribution in computations is less limited. Moreover, for certain subsets of phase-type distributions, unique representations may be found; such subfamilies are (generalized) Erlang, hyperexponential and acyclic phase-type distributions.

The main selection subsets based on phase-type distributions that have gained attention by several researchers in the area are *Erlang* and *hyperexponential* distributions. Aside from the fact that they exhibit a unique representation, their popularity

is based on the ability to distinguish them in terms of their variability. Erlang distributions have a coefficient of variation c less than one, whereas hyperexponential distributions have $c > 1$. The exponential distribution, which is both an Erlang and a hyperexponential distribution of order one, has $c = 1$. It is therefore possible to classify a distribution according to c and approximate this distribution by fitting the corresponding phase-type selection subset. Two of the parameter approximation methods compared in this study use mixtures of Erlang or hyperexponential distributions as the selection subset.

Several authors have raised the concern that, in fitting phase-type distributions to empirical data or approximating a given distribution, the phases of the resulting approximation are often fictitious and have no natural interpretation regarding the process being approximated (c.f. Asmussen [5], Lipsky [42] and van de Liefvoort [41]). This has led to the development of different classes of distributions which are not of phase-type. A close relative to the family of phase-type distributions are *generalized hyperexponential distributions* (see Botta, Harris and Marchal [14] and Harris et al. [29]), which are hyperexponential distributions with the mixing probabilities $(\alpha_1, \alpha_2, \ldots, \alpha_m)$ being replaced by weights which must sum to one, but are not restricted to be non-negative. The functions in the class of generalized hyperexponential distributions exhibit a unique representation. However, this class contains functions which are not probability distribution functions (see Harris et al. [29]) and is not closed under convolutions (see Botta, Harris and Marchal [14]). Maximum-likelihood procedures for generalized hyperexponential distributions are presented in Harris [27] and Harris and Sykes [28].

Another area of development related to Coxian distributions is the class of *matrix exponential distributions*, which consists of all distributions that have a rational Laplace-Stieltjes transform, but are not represented by a scheme of stages (or phases) as are Coxian distributions or phase-type distributions. The distribution functions in this family are expressed in terms of matrix exponentials. While phase-type distributions are a true subset of the family of matrix-exponential distributions (see

Lipsky [42]), the main difference is that the phase viewpoint is not being supported. Van de Liefvoort [41] presents an application of matrix exponential distributions in the moment problem for continuous distributions.

Finally, there are *mixtures of distributions*, largely motivated by applications involving mixed populations. Redner and Walker [66] present a survey of methods for fitting mixture densities, primarily using maximum-likelihood based approaches and the EM algorithm. Also mentioned are moment-matching methods. Fitting methods for finite mixtures of exponential and Weibull distributions are discussed in Kaylan and Harris [36] and in Mandelbaum and Harris [47], which is the basis for the maximum-likelihood based methods for generalized hyperexponential distributions discussed in Harris [27] and Harris and Sykes [28].

3.2 Moment Matching Techniques

A commonly used method in approximating the parameters of a probability distribution function $F(\cdot)$ of some random variable X, based on the *noncentral moments* $\mu'_n = \mathrm{E}(X^n) = \int_{-\infty}^{\infty} x^n \, dF(x)$, is called *moment matching*. The idea is to equate some set of empirical moments to their expected values, which results in a system of (generally nonlinear) equations for the parameters of the approximating distribution function. In our case, sample moments based on empirical data or the moments of a known probability distribution are matched to the corresponding moments of the approximating phase-type distribution (or a selection subset thereof), and the resulting equations yield values of the parameters of the approximating distribution. Only in some cases, however, is it possible to solve the moment-equations analytically. Such cases are typically based on two-phase hyperexponential distributions with balanced means (i.e. $\alpha_1/\lambda_1 = \alpha_2/\lambda_2$) or mixtures of two Erlang distributions of common order. In the case where only two moments are to be matched, closed-form expressions for the parameters of the approximating phase-type distribution can be obtained.

Several authors have proposed parameter approximation methods based on matching the first two moments of an empirical distribution to be approximated and the corresponding first two moments of an approximating phase-type distribution. Whitt [76, 77] uses the selection subset of Erlang distributions if the coefficient of variation c of the process to be approximated is less than one. The case where $c = 1$ forces the use of an exponential distribution. In the case where $c > 1$, the use of hyperexponential distributions is suggested. Sauer and Chandy [68, 69] also propose the use of a hyperexponential distribution for the case where $c > 1$. For distributions having $c < 1$, generalized Erlang distributions are being used as the selection subset. See also Tijms [75], Marie [48] and Sevcik, Levy, Tripathi and Zahorjan [70] for developments in this area.

The earliest standard references on fitting methods for phase-type distributions are Bux and Herzog [15, 16]. In addition to matching the first two moments, they also emphasize the shape constraints of the distribution to be fit, i.e. matching a number of cdf values of the approximating and the approximated distribution function. A numerical procedure using nonlinear programming (NLP) is developed to fit a Coxian distribution with common rate parameters at all phases (i.e. a mixture of pure Erlang distributions) to the first two sample moments of an empirical distribution and to user-specified values of the empirical cdf in the constraint set of the NLP. The objective is to minimize the number of phases. The primary disadvantage of this method, however, is the high order of the approximating Coxian distribution, resulting from the moment-matching requirements in the constraints.

Fitting approaches involving the matching of the first three moments are discussed in Altiok [3] and Khoshgoftaar and Perros [38], who present closed-form expressions for matching the first three empirical moments to a Coxian distribution of order two, and Whitt [76], who also presents closed-form expressions for a three-moment matching algorithm using two-phase hyperexponential distributions. More recent advances include Johnson and Taaffe [30], who use mixtures of Erlang distributions of common order to match the first three moments. Including also the

consideration of shape constraints of the distributions to be approximated, Johnson and Taaffe [31, 32] used NLP techniques and human interaction for the parameter approximation of mixtures of Erlang distributions (possibly of different order) and Coxian distributions. Recent work by Johnson [35] includes the development of a computer package MEFIT [33] for selecting phase-type distribution parameters using moment matching, NLP, heuristics and the use of Erlang distributions. This work is described in more detail in section 3.2.1.

Schmickler [71, 72, 73] and Schmickler and Südhofen [74] match three moments to mixtures of two or more Erlang distributions, not necessarily of the same order. Furthermore, they also numerically minimize the area between the approximated and approximating cdf's. Their work is discussed in more detail in section 3.2.2.

Additional methods emphasizing curve fitting are proposed by Bobbio, Cumani, Premoli and Saracco [9], who minimize the maximum difference between an empirical cdf and an acyclic phase-type distribution by iteratively solving a linearized version of the nonlinear optimization problem.

3.2.1 MEFIT

The following notation is used in describing Johnson's approach to selecting the parameters of a mixture of Erlang distributions. Denote the approximated cdf and pdf by $G(\cdot)$ and $g(\cdot)$, respectively. The cdf and pdf of the approximating mixed Erlang distribution is $F(\cdot)$ and $f(\cdot)$, respectively. The parameters of the Erlang mixture are r for the number of Erlang distributions, k_i, p_i and β_i for the order, mixing probability and mean of the i^{th} Erlang distribution in the mixture. The resulting Erlang mixture distribution is then

$$F(x) = \sum_{i=1}^{r} p_i F_i(x), \qquad \text{for} \quad x \geq 0, \tag{3.1}$$

where $F_i(\cdot)$ is the distribution of the i^{th} Erlang random variable used in the mixture.

The moments of both $F(\cdot)$ and $G(\cdot)$ are specified as standardized moments, i.e. the mean μ_1', (the noncentral first moment), the coefficient of variation $c = \mu_2^{1/2}/\mu_1'$, (the standard deviation divided by the mean), and the coefficient of skewness $\gamma = \mu_3/\mu_2^{3/2}$, (the third central moment divided by the cube of the standard deviation).

The following criteria are used for fitting:

1. matching the (standardized) moments of $F(\cdot)$ and $G(\cdot)$,
2. minimizing the sum of the absolute or squared differences between values of $F(\cdot)$ and $G(\cdot)$ (Δcdf) or $f(\cdot)$ and $g(\cdot)$ (Δpdf) at up to 200 points.

The following description of the moment matching algorithm was implemented in Fortran as the computer package MEFIT [33]. The user of the program may specify whether to include the above criteria in the objective function or the constraint set of the NLP used to determine the parameters of $F(\cdot)$. Typically the goodness-of-fit measures Δpdf or Δcdf are part of the objective function (which is to be minimized), if they are to be included in the approximation run at all. The moment-matching criteria may be selected to be included either in the objective function or as nonlinear constraints. Suppose ϕ is a moment criterion and $\phi(f)$ and $\phi(g)$ denote the values of that criterion for the approximating and the approximated distribution, respectively. Using criterion ϕ in the objective function means a term of the form

$$w(\phi(f) - \phi(g))^2, \qquad \text{for} \quad w > 0 \tag{3.2}$$

is added to the objective function. The purpose of the weight w is to reflect the relative importance of criterion ϕ. If criterion ϕ is to be included as a nonlinear constraint, a term of the form

$$-w\epsilon \leq w(\phi(f) - \phi(g)) \leq w\epsilon, \qquad \text{for} \quad w, \epsilon > 0, \tag{3.3}$$

is added to the constraint set of the NLP, where ϵ is the (unweighted) maximum tolerable difference between approximated and approximating moments.

The following scheme indicates how a parameter approximation is obtained. Input from the user is required for:

1. The number r ($r \leq 6$) of Erlang distributions to be mixed.
2. The orders $k_1, k_2, \ldots, k_r$ of the Erlang distributions to be mixed.
3. The fitting criteria, how to include them in the NLP (i.e. objective function or constraint set), and the assignment of weights.
4. The initial values of and bounds on (i) the means of the Erlang distributions to be mixed, and (ii) the mixing probabilities.

Output consists of the following:

1. The parameters of the selected mixed distribution.
2. For each criterion specified for matching, the value to be approximated and the approximating value.
3. (optional) Values of the approximating cdf $F(x_i)$ or pdf $f(x_i)$ at n ($n \leq 200$) equally spaced points $\{x_i,\ i = 1, \ldots, n\}$ along the interval $[0, b]$, where n and b are user-specified.

Heuristic arguments are used to select the number r of Erlang distributions to be mixed, as well as to determine their orders. Furthermore, the user also resorts to heuristics to provide initial solutions and bounds on the parameters. Johnson [35] discusses the experience gained throughout the development of this approach. As a more rigorous guideline in choosing the orders of the distributions, the following result is stated and proved in Johnson [35].

Theorem 3.1 (Johnson [35]) *Assume* $\mu_1' > 0$, $r \geq 2$ *and* $k_1, k_2, \ldots, k_r$ *are all distinct. Then the triple of moments* (μ_1', c, γ) *is feasible for a mixture of* r *Erlang distributions of orders* $k_1, k_2, \ldots, k_r$, *if and only if* $k_i \geq k^*$ *for some* $i = 1, 2, \ldots, r$, *where* k^* *is the minimum integer satisfying*

$$k^* > \frac{1}{c^2} \tag{3.4}$$

$$k^* > \frac{-\gamma + 1/c^3 + 1/c + 2c}{\gamma - (c - 1/c)}. \tag{3.5}$$

This result is used to choose at least one of the orders of the mixed Erlang distributions to be equal or greater than k^*.

The following example describes one of the ways in which Johnson's approach may be used to fit a mixture of Erlang distributions to a Weibull distribution.

Example. Suppose a Weibull distribution with standardized moments $\mu_1' = 1$, $c = 0.38$ and $\gamma = 0.21$ is to be approximated by a mixture of three Erlang distributions with orders $k_1 = 2$, $k_2 = 3$ and $k_3 = 10$ ($k^* = 9$). The objective function is used to match 6 values of the approximating and approximated pdf. Nonlinear constraints are used to match the first three moments, where the maximum tolerable difference between approximated and approximating moments is 0.02. Matching the mean μ_1' is associated with a weight of 10, resulting in the constraint $-0.20 \leq 10(\mu_1'(f) - 1.0) \leq 0.20$. Matching the coefficient of variation c is associated with a weight of 3, which yields the constraint $-0.06 \leq 3(c(f) - 0.38) \leq 0.06$ and matching the coefficient of skewness γ is associated with a weight of 1, resulting in the constraint $-0.02 \leq 1(\gamma(f) - 0.21) \leq 0.02$. Bounds are assigned to $\beta_2 - \beta_1$ and $\beta_3 - \beta_2$ to ensure that the initial ordering of the means $\beta_1, \beta_2, \beta_3$ of the individual

Erlang distributions is preserved. The resulting NLP problem is as follows:

$$
\begin{aligned}
\min \quad & z = \sum_{i=1}^{6} \left(f(x_i) - g(x_i)\right)^2 \\
\text{s.t.} \quad & \\
-0.20 \;&\leq\; 10(\mu_1'(f) - 1.0) \;\leq\; 0.20 \\
-0.06 \;&\leq\; 3(c(f) - 0.38) \;\leq\; 0.06 \\
-0.02 \;&\leq\; 1(\gamma(f) - 0.21) \;\leq\; 0.02 \\
0.01 \;&\leq\; \beta_{i+1} - \beta_i \leq 10^5, \qquad \text{for} \quad i = 1,2 \\
0 \;&\leq\; \beta_i \leq 10^5, \qquad \text{for} \quad i = 1,2,3 \\
0 \;&\leq\; p_i \leq 1, \qquad \text{for} \quad i = 1,2,3 \\
p_1 + p_2 + p_3 &= 1.
\end{aligned}
\tag{3.6}
$$

The parameters of the mixture of Erlang distributions are implicitly included in the constraints. The NLP is solved using the package NPSOL (version 4.0), which uses a sequential quadratic programming algorithm for dense NLP problems (see Gill, Murray, Saunders and Wright [24]).

3.2.2 MEDA

Schmickler's approach MEDA (Mixed Erlang Distributions for Approximation) to the parameter approximation of phase-type distributions as approximations of empirical distributions is similar to MEFIT, in that mixtures of Erlang distributions (not necessarily of the same order) are being used as the selection subset. Differences are the following:

1. The first three empirical moments are matched exactly. Higher moment information is taken into account by incorporating shape constraints of the empirical distribution based on the area between the approximating and the approximated cdf's. All measured data points are considered for fitting the shape of the empirical cdf.

2. Significant jumps in the empirical cdf can be taken into account by using weighted Dirac delta functions (or Erlang distributions with an infinite number of phases). This method yields remarkable improvements in the fit for the case of jumps, but inclusion of an E_∞ distribution in the approximation mixture makes this usable only if a phase-type distribution with a limited number of phases is not needed.

3. The orders of the Erlang distributions to be mixed are chosen automatically by the program. No user interaction is possible.

In the first version of MEDA [71], only two or three Erlang distributions were used in the mixture. The extension (Schmickler [72, 73] and Schmickler and Südhofen [74]) incorporates higher numbers of Erlang distributions, as well as a different nonlinear programming algorithm for the parameter approximation. As described in Schmickler [72, 73], a fast numerical algorithm for matching the first three empirical moments to the moments of a mixture of two Erlang distributions is implemented in MEDA. The moments of mixed (more than two) Erlang distributions can be reduced and normalized to the moments of a mixture of two Erlang distributions, which is incorporated in MEDA. This means that the parameters for the first two Erlang distributions to be mixed (i.e. the mixing probabilities p_1 and p_2 and the parameters λ_1 and λ_2) are chosen by a moment matching algorithm, whereas the remaining parameters of all other Erlang distributions in the mixture are approximated using the reduction and normalization step. The optimization of all parameters according to the objective function

$$\Delta = \frac{1}{\hat{\mu}_1} \int_0^\infty |F_e(x) - F_a(x)| \, dx, \tag{3.7}$$

which is the area between the empirical and the approximating cdf (normalized with respect to the first empirical moment $\hat{\mu}_1$) is done by using the *Flexible Polyhedron Search* (FPS) nonlinear programming algorithm by Nelder and Mead [50]. The FPS algorithm allows for the simultaneous optimization of several variables, but only continuous variables are permitted. Since the number of phases in each Erlang

distribution to be mixed is an integral parameter, mixed Gamma distributions (having continuous shape parameters) are used instead for the optimization run. After optimal values k_i^* for these parameters have been determined, a transformation is performed by searching for the best neighboring integer number k_i for k_i^*.

A typical approximation with MEDA is as follows:

1. The program MEDA reads the input file containing the empirical cdf and the first three empirical central moments. The cdf is then examined for jumps, according to a user-specified level of significance.

2. A setup for the optimization follows, where initial values for all parameters of the mixed Erlang distributions are determined, except the parameters of the first two Erlang distributions to be mixed, since these parameters are to be determined by the moment matching algorithm. A heuristic argument is used for these initial parameter values, based on a linear approximation of the empirical cdf in the lower range (see Schmickler [72, 73]).

3. The optimization of the parameter values is performed, using the FPS algorithm involving Gamma distributions.

4. The continuous shape parameters of the Gamma distributions are transformed into integral values to obtain Erlang distributions.

5. The objective function value Δ is computed.

Steps 3-5 above are performed iteratively until the objective function can no longer be improved. An approximation run as described above always starts with two or three Erlang distributions to be mixed. If the fit needs to be improved, more Erlang distributions can be included. Steps 2-5 above have to be carried out for each additional Erlang distribution included in the mixture; results of approximation runs for a mixture of three Erlang distributions are being used for the optimization run for the fourth Erlang distribution, and so on.

3.3 Maximum Likelihood Based Approximation

Maximum likelihood is one of the most popular methods in parametric point estimation. A vast body of literature has been published, both on theoretical properties as well as applications.

A general definition of maximum likelihood estimates is the following (see also Kotz and Johnson [39]). Consider a random sample $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$, whose joint distribution is given by the density $f_n(\boldsymbol{x}; \boldsymbol{\Theta})$ in the n-dimensional Euclidean space $\Re^n$. The unknown parameter vector $\boldsymbol{\Theta}$ is contained in the parameter space Ω.

Definition 5 *For fixed observations* $\boldsymbol{x}$, *the* **likelihood function** *of* $\boldsymbol{x}$ *is defined as* $L(\boldsymbol{x}; \boldsymbol{\Theta}) = f_n(\boldsymbol{x}; \boldsymbol{\Theta})$, *being a function of* $\boldsymbol{\Theta} \in \Omega$. *Any* $\hat{\boldsymbol{\Theta}} = \hat{\boldsymbol{\Theta}}(\boldsymbol{x}) \in \Omega$ *which maximizes* $L(\boldsymbol{x}; \boldsymbol{\Theta})$ *over* Ω *is called a* **maximum likelihood estimate** *(MLE) of the unknown true parameter vector* $\boldsymbol{\Theta}$.

Often it is advantageous to compute the MLEs by maximizing the logarithm of the likelihood function, denoted by $l(\boldsymbol{x}; \boldsymbol{\Theta}) = \log L(\boldsymbol{x}; \boldsymbol{\Theta})$. The logarithm is a monotone function; thus, any $\hat{\boldsymbol{\Theta}}(\boldsymbol{x})$ which maximizes $L(\boldsymbol{x}; \boldsymbol{\Theta})$ also maximizes $l(\boldsymbol{x}; \boldsymbol{\Theta})$.

In our setting, it is desired to estimate the parameters $(\boldsymbol{\alpha}, \boldsymbol{T})$ of a phase-type distribution based on observations $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$. The likelihood function is then given by

$$L(\boldsymbol{x}; \boldsymbol{\Theta}) = \prod_{i=1}^{n} \boldsymbol{\alpha} \mathrm{e}^{\boldsymbol{T} x_i} \boldsymbol{t}_0' = \prod_{i=1}^{n} \boldsymbol{\alpha} \mathrm{e}^{\boldsymbol{T} x_i} (-\boldsymbol{T}\boldsymbol{e}'). \tag{3.8}$$

Aside from the case where the order of the approximating phase-type distribution equals one (i.e. an exponential distribution), there is no known way to combine the exponents of equation (3.8) in order to obtain a sufficient statistic of reasonable dimension. Moreover, equation (3.8) cannot be differentiated in a straightforward manner (see Asmussen and Nerman [5]); no simple expression for the partial derivative of $\mathrm{e}^{\boldsymbol{T} x}$ w.r.t. T_{ij} is known (see Graham [25]). An example of a log-likelihood surface, based on a generalized Erlang distribution with two parameters and a data

set of 200 observations from one of the test distributions involved in this study is shown in figure 3.1. Difficulties such as extreme flatness of the log-likelihood function around maxima are evident.

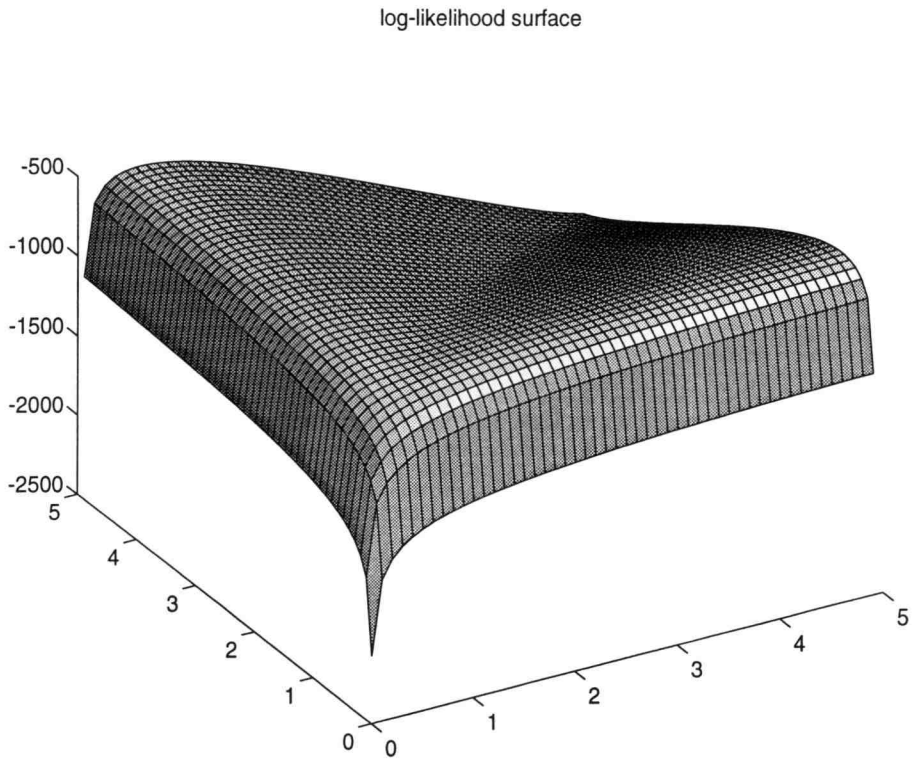

Figure 3.1. Log-likelihood surface.

Approaches on how to deal with these problems have been proposed by Bobbio and Cumani [10, 11], who define the restricted class of *acyclic phase-type distributions* (c.f. section 2.2) for which closed-form expressions of partial derivatives of equation (3.8) with respect to the parameters may be derived. A different approach by Asmussen and Nerman [5] suggests thinking of the observations $\boldsymbol{x}$ as partial information on the representation of the phase-type distribution. This interpretation motivates an iterative method, called the *EM method* (Expectation Maximization), which was implemented for the phase-type setting by Häggström, Asmussen and Nerman [26]. The standard reference on the EM algorithm is Dempster, Laird and

Rubin [19], who give a thorough introduction and describe a wide range of applications for the use of the EM algorithm in parameter estimation for distributions.

Both maximum likelihood estimation approaches, by Bobbio and Cumani [11] and by Asmussen and Nerman [5], will be described in more detail in the following sections. Since our computational experiments are based on test distributions which are not phase-type distributions and the samples from these distributions were not chosen randomly (see chapter 4), we shall use the term *maximum likelihood based parameter approximation* in the sequel. A third fitting method involving maximum likelihood was developed by Faddy [21]. He suggests using Coxian distributions with real parameters (see section 2.2, eqn. (2.18) and figure 2.7) as the selection subset in order to improve on the precision of the estimates, since the fitting of Coxian distributions involves fewer parameters than the fitting of general phase-type distributions. O'Cinneide [62] showed that every phase-type distribution whose Laplace transform has real poles has an equivalent Coxian representation. However, it is not guaranteed that such an equivalent Coxian representation is of the same order as the original phase-type distribution (see Botta, Harris and Marchal [14]). The fitting procedure by Faddy is implemented with a straightforward procedure in MATLAB [49], involving the computation of the matrix exponentials $e^{\boldsymbol{T}x_i}$ and numerically maximizing the log-likelihood function by using the Nelder-Mead simplex search method [50].

3.3.1 MLAPH

Bobbio and Cumani [10, 11] have developed a maximum likelihood based approach incorporating the canonical representation of the restricted class of *acyclic phase-type distributions* (see section 2.2, equations (2.13, 2.14, 2.15, 2.16, 2.17) and figures (2.5, 2.6, 2.7)). The term *acyclic phase-type distribution* was first used by Cumani [18], who also introduced the canonical representation, i.e. minimum order unique representations. Recently, O'Cinneide [63] investigated this family of distributions using the term *triangular phase-type distributions.*

Recalling the properties of the canonical representation of acyclic phase-type distributions, we note that they are a straightforward restriction of the family of Coxian distributions (see section 2.2), obtained by using only real-valued parameters. The necessary ordering of the transition rates of the exponential phases (i.e. the canonical representation, conditions (2.14, 2.16, 2.17)) ensures that the class of acyclic phase-type distributions exhibits unique representations. Moreover, they form a dense set for distributions with support on $[0, \infty)$ (c.f. Cumani [18]); that is, there exist acyclic phase-type distributions in canonical form that converge weakly to any given distribution on $[0, \infty)$. Any acyclic phase-type distribution in canonical form is a proper probability distribution function, and the class of acyclic phase-type distributions is closed under mixtures, convolutions and formations of coherent systems (c.f. Assaf and Levikson [7]).

These properties allow for the derivation of closed-form expressions of the cdf of acyclic phase-type distributions in canonical form, as well as the derivatives with respect to both the model parameters and time. Maximum likelihood based approximation is therefore possible.

The following results (Bobbio and Cumani [11]) are based on the canonical representation of an acyclic phase-type distribution of order k depicted in equations (2.13, 2.14) and figure (2.5). The vector of initial probabilities is denoted by $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_k)$, where α_{k+1} is assumed to be zero. The vector of transition

rates is $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_k)$. The cdf of the time t until absorption in state $k+1$ of the Markov chain (2.13) is denoted by $F(t, \boldsymbol{\alpha}, \boldsymbol{\lambda})$. The Laplace-Stieltjes transform of $F(t, \boldsymbol{\alpha}, \boldsymbol{\lambda})$ is

$$\tilde{F}(s, \boldsymbol{\alpha}, \boldsymbol{\lambda}) = \int_0^\infty \mathrm{e}^{-sx}\, dF(t, \boldsymbol{\alpha}, \boldsymbol{\lambda}) = \sum_{i=1}^{k} \alpha_i \tilde{F}^{(i)}(s, \boldsymbol{\alpha}, \boldsymbol{\lambda}), \tag{3.9}$$

where $\tilde{F}^{(i)}(s, \boldsymbol{\alpha}, \boldsymbol{\lambda})$ is the Laplace-Stieltjes transform of the cdf of an acyclic phase-type distribution comprised of the last i phases in figure (2.5). From the structure of the matrix $\boldsymbol{T}$ given by equation (2.13), it is clear that $(\lambda_1, \lambda_2, \ldots, \lambda_k)$ are its eigenvalues. For any $i = 1, 2, \ldots, k$, let r be the number of distinct eigenvalues in the set of parameters $(\lambda_{k-i+1}, \lambda_{k-i+2}, \ldots, \lambda_k)$, and let m_u be the multiplicity of the u^{th} distinct eigenvalue $\lambda_{(u)}$ (i.e. $\sum_{u=1}^{r} m_u = i$). It follows that

$$\tilde{F}^{(i)}(s, \boldsymbol{\alpha}, \boldsymbol{\lambda}) = \prod_{u=1}^{r} \left(\frac{\lambda_{(u)}}{s + \lambda_{(u)}} \right)^{m_u} \qquad \text{for} \quad i = 1, \ldots, k. \tag{3.10}$$

Partial fraction expansion of equation (3.10) results in

$$\tilde{F}^{(i)}(s, \boldsymbol{\alpha}, \boldsymbol{\lambda}) = \sum_{u=1}^{r} \sum_{l=1}^{m_u} \frac{c_{ul}^{(i)}}{(s + \lambda_{(u)})^l} \qquad \text{for} \quad i = 1, \ldots, k, \tag{3.11}$$

and inverting the Laplace-Stieltjes transform given in equation (3.11) yields

$$f^{(i)}(t, \boldsymbol{\alpha}, \boldsymbol{\lambda}) = \sum_{u=1}^{r} \mathrm{e}^{-\lambda_{(u)} t} \left(\sum_{l=1}^{m_u} \frac{c_{ul}^{(i)}}{(l-1)!}\, t^{l-1} \right) \qquad \text{for} \quad i = 1, \ldots, k. \tag{3.12}$$

Equation (3.12) is completely determined once the coefficients $c_{ul}^{(i)}$ are calculated. By integrating equation (3.12) term by term, the cdf $F^{(i)}(t, \boldsymbol{\alpha}, \boldsymbol{\lambda})$ can be obtained in closed form. Bobbio and Cumani [11] present an iterative algorithm for computing the coefficients $c_{ul}^{(i)}$.

The next step is to derive the $(2m-1)$ first partial derivatives of $\tilde{F}(s, \boldsymbol{\alpha}, \boldsymbol{\lambda})$ with respect to the $(m-1)$ independent parameters $\boldsymbol{\alpha}$ and the m parameters $\boldsymbol{\lambda}$. The partial derivatives with respect to the α_i's are immediate, since equation (3.9) is linear in the α_i's. In order to compute the partial derivatives of $\tilde{F}(s, \boldsymbol{\alpha}, \boldsymbol{\lambda})$ with respect to the λ_i's, separate the contribution of λ_i and write

$$\tilde{F}(s, \boldsymbol{\alpha}, \boldsymbol{\lambda}) = \frac{\lambda_i}{s + \lambda_i} \tilde{\phi}(s) + \tilde{\psi}(s), \qquad \text{for} \quad i = 1, \ldots, k \tag{3.13}$$

where $\tilde{\psi}(s)$ contains the stages before the occurrence of the λ_i^{th} stage in the chain. Writing

$$\tilde{F}_i(s, \boldsymbol{\alpha}, \boldsymbol{\lambda}) = \frac{\lambda_i}{s + \lambda_i} \tilde{\phi}(s) = \sum_{j=i}^{k} \alpha_j F^{(j)}(s, \boldsymbol{\alpha}, \boldsymbol{\lambda}) \qquad \text{for} \quad i = 1, \ldots, k \tag{3.14}$$

and differentiating equation (3.13) with respect to λ_i then yields

$$\frac{\partial \tilde{F}(s, \boldsymbol{\alpha}, \boldsymbol{\lambda})}{\partial \lambda_i} = \frac{1}{\lambda_i} \left[\tilde{F}_i(s, \boldsymbol{\alpha}, \boldsymbol{\lambda}) - \frac{\lambda_i}{s + \lambda_i} \tilde{F}_i(s, \boldsymbol{\alpha}, \boldsymbol{\lambda}) \right] \qquad \text{for} \quad i = 1, \ldots, k. \tag{3.15}$$

To describe the maximum likelihood based approximation procedure, consider a data sample $\boldsymbol{x} = (x_1, x_2, \ldots, x_\eta, x_{\eta+1}, \ldots, x_\nu)$, such that $\eta \leq \nu$ are non-censored (e.g. observed life-times of items), and $\nu - \eta$ are censored (e.g. life-times of items that were still 'alive' when the data collection was terminated). Without loss of generality, arrange the order of the sample so that the η non-censored points are in the first η positions. Denote the maximum likelihood based estimators of $\boldsymbol{\alpha}$ and $\boldsymbol{\lambda}$ by $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\lambda}}$. The likelihood equation is

$$L(\boldsymbol{x}; \boldsymbol{\alpha}, \boldsymbol{\lambda}) = \prod_{i=1}^{\eta} f(x_i, \boldsymbol{\alpha}, \boldsymbol{\lambda}) \prod_{j=\eta+1}^{\nu} (1 - F(x_j, \boldsymbol{\alpha}, \boldsymbol{\lambda})). \tag{3.16}$$

The approximation problem is then to find $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\lambda}}$ such that $L(\boldsymbol{x}; \boldsymbol{\alpha}, \boldsymbol{\lambda})$ (or $l(\boldsymbol{x}; \boldsymbol{\alpha}, \boldsymbol{\lambda})$) is maximized under the canonical constraints $0 \leq \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_k$, $\alpha_i \geq 0$ for $i = 1, 2, \ldots, k$ and $\sum_{i=1}^{k} \alpha_i = 1$.

This nonlinear constrained optimization problem is then solved by resorting to an iterative linearization procedure. The first order derivative information described above is taken into account. Given initial values $\boldsymbol{\alpha}^{(0)}$ and $\boldsymbol{\lambda}^{(0)}$ for the parameters, the log-likelihood function $l(\boldsymbol{x}; \boldsymbol{\alpha}, \boldsymbol{\lambda})$ is linearized in an ϵ-neighborhood of $\boldsymbol{\alpha}^{(0)}$ and $\boldsymbol{\lambda}^{(0)}$. The solution of a modified simplex procedure along with a line search determines the next estimate $\boldsymbol{\alpha}^{(1)}$ and $\boldsymbol{\lambda}^{(1)}$, which is used as the initial solution for another step of the iterative procedure. The algorithm stops after a certain user-defined number of solutions have been computed, or if the rate of convergence becomes smaller than a user-specified level.

Results illustrating this parameter approximation procedure are presented in Bobbio and Cumani [10], and Bobbio and Telek [12, 13]. The major concerns mentioned by the authors are numerical instabilities. Running the same procedure on different computers yields significantly different results. Moreover, the algorithm is numerically very intensive and sensitive to the choice of the initial solution $\boldsymbol{\alpha}^{(0)}$ and $\boldsymbol{\lambda}^{(0)}$, because local maxima are present in the objective function. Bobbio and Telek [12, 13] suggest randomly generating the starting values $\boldsymbol{\alpha}^{(0)}$ and $\boldsymbol{\lambda}^{(0)}$. The rate of convergence of the algorithm is very slow and decreases with the order of the approximating acyclic phase-type distribution, due to the flatness of the objective function.

We note that this approach does not accommodate the approximation of the *order* of the fitted acyclic phase-type distribution.

3.3.2 EMPHT

A different approach in maximum likelihood based approximation for phase-type distributions incorporates the *EM algorithm* (see Dempster, Laird and Rubin [19]). Asmussen and Nerman [5] and Asmussen, Nerman and Olsson [6] describe the use of the EM algorithm for approximating the parameters $(\boldsymbol{\alpha}, \boldsymbol{T})$ of a phase-type distribution of order-k from data $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$, as well as fitting a phase-type distribution of order-k to a given distribution $G(\cdot)$ with support on $[0, \infty)$. Another similar approach by Olsson [64] presents the approximation of phase-type distributions based on censored samples.

The idea is to view the usual interpretation for a phase-type distribution (being the time x until absorption in a Markov process $\{J(t),\ 0 \leq t < x\}$ with a finite number of transient states) as a multi-parameter exponential family, provided the entire process $\{J(t),\ 0 \leq t < x\}$ is observed. This is equivalent to observing the underlying Markov chain J_0 (the initial state), $J_1, J_2, \ldots, J_{M-1}$ (where M is the number of transitions until absorption in state $k+1$, i.e. $J_M = k+1$ and J_i, $i = 1, \ldots, M-1$,

represents the state of the Markov chain after the i^{th} transition), and the sojourn times $S_0, S_1, \ldots, S_{M-1}$ ($S_M = \infty$). Thus, a *complete observation* y of a Markov process $\{J(t),\ 0 \leq t < x\}$ may be represented by $y = (j_0, j_1, \ldots, j_{m-1}, s_0, s_1, \ldots, s_{m-1})$. An *incomplete observation* x is then represented by $x = s_0 + s_1 + \ldots + s_{m-1}$. In practice, the data consists of n i.i.d. realizations of x, denoted by $x_1, x_2, \ldots, x_n$, each of which is regarded as an incomplete observation of the Markov process $J(t)$. It is an incomplete observation in the sense that no information about the initial state, the transient states visited and the sojourn times at the individual transient states is given. Denoting the n independent Markov processes by $J(t)^{[1]}, \ldots, J(t)^{[n]}$ and letting $J_0^{[r]}, J_1^{[r]}, \ldots, J_{M^{[r]}-1}^{[r]}$ and $S_0^{[r]}, S_1^{[r]}, \ldots, S_{M^{[r]}-1}^{[r]}$ be the underlying Markov chains and the sojourn times for the r^{th} process ($r = 1, \ldots, n$), the *complete*, yet unobservable, data set is then written as

$$\boldsymbol{y} = (j_0^{[1]}, \ldots, j_{m^{[1]}-1}^{[1]}, s_0^{[1]}, \ldots, s_{m^{[1]}-1}^{[1]}, \ldots, j_0^{[n]}, \ldots, j_{m^{[n]}-1}^{[n]}, s_0^{[n]}, \ldots, s_{m^{[n]}-1}^{[n]}). \quad (3.17)$$

The EM algorithm makes use of this complete sample in order to find the maximum likelihood estimate of $(\boldsymbol{\alpha}, \boldsymbol{T})$, based on the observed sample

$$\begin{aligned} \boldsymbol{x} &= (x_1, \ldots, x_n) \\ &= (s_0^{[1]} + \ldots + s_{m^{[1]}-1}^{[1]}, \ldots, s_0^{[n]} + \ldots + s_{m^{[n]}-1}^{[n]}). \end{aligned} \quad (3.18)$$

Based on the standard estimation theory for continuous-time Markov processes (see Albert [1] or Basawa and Prakasa [8]), the sufficient statistic for $\{J(t),\ 0 \leq t < x\}$ is

$$S = \left\{ \begin{array}{ll} (B_i), & i = 1, \ldots, k \\ (Z_i), & i = 1, \ldots, k \\ (N_{ij}), & i = 1, \ldots, k,\ j = 1, \ldots, k+1,\ i \neq j \end{array} \right\}, \quad (3.19)$$

where

$$\begin{aligned} B_i &= \sum_{r=1}^{n} I_{\{J_0^{[r]}=i\}}, \\ &= \text{the number of processes starting at state } i, \\ &\quad \text{for } i = 1, \ldots, k, \end{aligned}$$

$$
\begin{aligned}
Z_i &= \sum_{r=1}^{n} \sum_{u=0}^{m^{[r]}-1} I_{\{J_u^{[r]}=i\}} S_u^{[r]}, \\
&= \text{the total time spent in state } i, \\
&\quad \text{for } i = 1, \ldots, k, \\
N_{ij} &= \sum_{r=1}^{n} \sum_{u=0}^{m^{[r]}-1} I_{\{J_u^{[r]}=i, J_{u+1}^{[r]}=j\}}, \\
&= \text{the total number of transitions from state } i \text{ to state } j, \\
&\quad \text{for } i = 1 \ldots, k,\ j = 1, \ldots, k+1,\ i \neq j.
\end{aligned} \tag{3.20}
$$

The joint density of the complete sample $\boldsymbol{y}$ is given by

$$
f(\boldsymbol{y}; \boldsymbol{\alpha}, \boldsymbol{T}) = \prod_{i=1}^{k} \alpha_i^{B_i} \prod_{i=1}^{k} e^{T_{ii} Z_i} \prod_{i=1}^{k} \prod_{\substack{j=1 \\ j \neq i}}^{k+1} T_{ij}^{N_{ij}} \tag{3.21}
$$

with log-likelihood function

$$
l(\boldsymbol{y}; \boldsymbol{\alpha}, \boldsymbol{T}) = \sum_{i=1}^{k} B_i \ln(\alpha_i) + \sum_{i=1}^{k} T_{ii} Z_i + \sum_{i=1}^{k} \sum_{\substack{j=1 \\ j \neq i}}^{k+1} N_{ij} \ln(T_{ij}). \tag{3.22}
$$

The maximum likelihood estimators for the parameters $(\boldsymbol{\alpha}, \boldsymbol{T})$ are obtained by maximizing (3.22) subject to

$$
-\sum_{\substack{j=1 \\ j \neq i}}^{k+1} T_{ij} = T_{ii} \quad \text{for} \quad i = 1, \ldots, k \quad \text{and} \quad \sum_{i=1}^{k} \alpha_i = 1, \tag{3.23}
$$

which yields

$$
\begin{aligned}
\hat{\alpha}_i &= \frac{B_i}{n}, && \text{for} \quad i = 1 \ldots, k, \\
\hat{T}_{ij} &= \frac{N_{ij}}{Z_i}, && \text{for} \quad i = 1, \ldots, k,\ j = 1, \ldots, k+1,\ j \neq i, \\
\hat{T}_{ii} &= -\sum_{\substack{j=1 \\ j \neq i}}^{k+1} \hat{T}_{ij}, && \text{for} \quad i = 1 \ldots, k.
\end{aligned} \tag{3.24}
$$

Based on these estimates, the EM algorithm is applied to obtain maximum likelihood estimates for $(\boldsymbol{\alpha}, \boldsymbol{T})$ using the observed sample $\boldsymbol{x}$. The EM algorithm is

iterative in nature; starting from initial values $(\boldsymbol{\alpha}, \boldsymbol{T})^{(0)}$ for the parameters to be approximated, the values $(\boldsymbol{\alpha}, \boldsymbol{T})^{(m+1)}$ are derived from $(\boldsymbol{\alpha}, \boldsymbol{T})^{(m)}$ in two steps, the Expectation step and the Maximization step.

Expectation Step: Compute

$$
\begin{aligned}
E_{(\boldsymbol{\alpha},\boldsymbol{T})^{(m)}}(S|\boldsymbol{x}) &= \left\{ \begin{array}{ll} B_i^{(m+1)}, & i = 1\ldots, k, \\ Z_i^{(m+1)}, & i = 1\ldots, k, \\ N_{ij}^{(m+1)}, & i = 1\ldots, k, \\ & j = 1, \ldots, k+1,\ j \neq i \end{array} \right\} \\
&= \left\{ \begin{array}{ll} \left(E_{(\boldsymbol{\alpha},\boldsymbol{T})^{(m)}}(B_i|\boldsymbol{x})\right), & i = 1\ldots, k, \\ \left(E_{(\boldsymbol{\alpha},\boldsymbol{T})^{(m)}}(Z_i|\boldsymbol{x})\right), & i = 1\ldots, k, \\ \left(E_{(\boldsymbol{\alpha},\boldsymbol{T})^{(m)}}(N_{ij}|\boldsymbol{x})\right), & i = 1\ldots, k, \\ & j = 1, \ldots, k+1,\ j \neq i \end{array} \right\}. \qquad (3.25)
\end{aligned}
$$

Maximization Step: Replace S by $E_{(\boldsymbol{\alpha},\boldsymbol{T})^{(m)}}(S|\boldsymbol{x})$ in equations (3.24), that is:

$$
\begin{aligned}
\alpha_i^{(m+1)} &= \frac{E_{(\boldsymbol{\alpha},\boldsymbol{T})^{(m)}}(B_i|\boldsymbol{x})}{n} = \frac{B_i^{(m+1)}}{n}, && i = 1\ldots, k, \\
T_{ij}^{(m+1)} &= \frac{E_{(\boldsymbol{\alpha},\boldsymbol{T})^{(m)}}(N_{ij}|\boldsymbol{x})}{E_{(\boldsymbol{\alpha},\boldsymbol{T})^{(m)}}(Z_i|\boldsymbol{x})} = \frac{N_{ij}^{(m+1)}}{Z_i^{(m+1)}}, && j = 1, \ldots, k+1,\ j \neq i, \qquad (3.26) \\
T_{ii}^{(m+1)} &= -\sum_{\substack{j=1 \\ j\neq i}}^{k+1} T_{ij}^{(m+1)}, && i = 1, \ldots, k.
\end{aligned}
$$

In order to evaluate the conditional expectations in equations (3.26), note that S can be decomposed as $S = S^{(1)} + \cdots + S^{(n)}$, where $S^{(r)}$ is the contribution from the r^{th} Markov process $\{J(t)^{[r]},\ 0 \leq t < x_r\}$. Then

$$
E_{(\boldsymbol{\alpha},\boldsymbol{T})^{(m)}}(S|\boldsymbol{x}) = \sum_{r=1}^{n} E_{(\boldsymbol{\alpha},\boldsymbol{T})^{(m)}}(S^{(r)}|x_r). \qquad (3.27)
$$

It is also shown in Asmussen and Nerman [5] and Olsson [64], that

$$\begin{aligned}
E_{(\boldsymbol{\alpha},\boldsymbol{T})^{(m)}}(B_i|x_r) &= \frac{\alpha_i \boldsymbol{b}_i(x_r|\boldsymbol{T})}{\boldsymbol{\alpha}\boldsymbol{b}(x_r|\boldsymbol{T})}, && \text{for} \quad i = 1,\ldots,k,\\
E_{(\boldsymbol{\alpha},\boldsymbol{T})^{(m)}}(Z_i|x_r) &= \frac{\boldsymbol{c}_i(x_r;i|\boldsymbol{\alpha},\boldsymbol{T})}{\boldsymbol{\alpha}\boldsymbol{b}(x_r|\boldsymbol{T})}, && \text{for} \quad i = 1,\ldots,k,\\
E_{(\boldsymbol{\alpha},\boldsymbol{T})^{(m)}}(N_{ij}|x_r) &= \frac{T_{ij}\boldsymbol{c}_j(x_r;i|\boldsymbol{\alpha},\boldsymbol{T})}{\boldsymbol{\alpha}\boldsymbol{b}(x_r|\boldsymbol{T})}, && \text{for} \quad i,j = 1,\ldots,k,\ j \neq i,\\
E_{(\boldsymbol{\alpha},\boldsymbol{T})^{(m)}}(N_{i,k+1}|x_r) &= \frac{\boldsymbol{a}_i(x_r|\boldsymbol{\alpha},\boldsymbol{T})t_i}{\boldsymbol{\alpha}\boldsymbol{b}(x_r|\boldsymbol{T})}, && \text{for} \quad i = 1,\ldots,k.
\end{aligned} \tag{3.28}$$

where $t_i = T_{i,k+1}$ is the i^{th} element of the exit rate vector $\boldsymbol{t}_0'$, $\boldsymbol{e}_i'$ is the i^{th} unit vector and $\boldsymbol{a}_i(x|\boldsymbol{\alpha},\boldsymbol{T})$, $\boldsymbol{b}_i(x|\boldsymbol{T})$, $\boldsymbol{c}_i(x;i|\boldsymbol{\alpha},\boldsymbol{T})$ are the i^{th} elements of the vector functions defined by

$$\begin{aligned}
\boldsymbol{a}(x|\boldsymbol{\alpha},\boldsymbol{T}) &= \boldsymbol{\alpha}\mathrm{e}^{\boldsymbol{T}x},\\
\boldsymbol{b}(x|\boldsymbol{T}) &= \mathrm{e}^{\boldsymbol{T}x}\boldsymbol{t}_0',\\
\boldsymbol{c}(x;i|\boldsymbol{\alpha},\boldsymbol{T}) &= \int_0^x \boldsymbol{\alpha}\mathrm{e}^{\boldsymbol{T}u}\boldsymbol{e}_i'\mathrm{e}^{\boldsymbol{T}(x-u)}\boldsymbol{t}_0'\,du,\\
&= \int_0^x \boldsymbol{a}(u|\boldsymbol{\alpha},\boldsymbol{T})\boldsymbol{e}_i'\boldsymbol{b}(x-u|\boldsymbol{T})\,du, \qquad \text{for} \quad i = 1,\ldots,k.
\end{aligned} \tag{3.29}$$

Using this in equations (3.24), we obtain:

Theorem 3.2 (Asmussen and Nerman [5]) *The EM algorithm for phase-type distributions is given by*

$$\begin{aligned}
\alpha_i^{(m+1)} &= \alpha_i^{(m)} \frac{\sum_{r=1}^{n} \frac{\boldsymbol{b}_i(x_r|\boldsymbol{T}^{(m)})}{\boldsymbol{\alpha}^{(m)}\boldsymbol{b}(x_r|\boldsymbol{T}^{(m)})}}{n}, && \text{for} \quad i = 1, \ldots, k, \\
T_{ij}^{(m+1)} &= T_{ij}^{(m)} \frac{\sum_{r=1}^{n} \frac{\boldsymbol{c}_j(x_r; i|(\boldsymbol{\alpha}, \boldsymbol{T})^{(m)})}{\boldsymbol{\alpha}^{(m)}\boldsymbol{b}(x_r|\boldsymbol{T}^{(m)})}}{\sum_{r=1}^{n} \frac{\boldsymbol{c}_i(x_r; i|(\boldsymbol{\alpha}, \boldsymbol{T})^{(m)})}{\boldsymbol{\alpha}^{(m)}\boldsymbol{b}(x_r|\boldsymbol{T}^{(m)})}}, && \text{for} \quad i, j = 1, \ldots, k, \; j \neq i, \\
T_{i,k+1}^{(m+1)} &= \frac{\sum_{r=1}^{n} \frac{\boldsymbol{a}_i(x_r|(\boldsymbol{\alpha}, \boldsymbol{T})^{(m)})t_i}{\boldsymbol{\alpha}^{(m)}\boldsymbol{b}(x_r|\boldsymbol{T}^{(m)})}}{\sum_{r=1}^{n} \frac{\boldsymbol{c}_i(x_r; i|(\boldsymbol{\alpha}, \boldsymbol{T})^{(m)})}{\boldsymbol{\alpha}^{(m)}\boldsymbol{b}(x_r|\boldsymbol{T}^{(m)})}}, && \text{for} \quad i = 1, \ldots, k, \\
T_{ii}^{(m+1)} &= -\sum_{\substack{j=1 \\ j \neq i}}^{k+1} T_{ij}^{(m+1)}, && \text{for} \quad i = 1, \ldots, k.
\end{aligned} \tag{3.30}$$

For the computations of the vector functions $\boldsymbol{a}(x|\boldsymbol{\alpha}, \boldsymbol{T}) = \boldsymbol{a}(x)$, $\boldsymbol{b}(x|\boldsymbol{T}) = \boldsymbol{b}(x)$, $\boldsymbol{c}(x; i|\boldsymbol{\alpha}, \boldsymbol{T}) = \boldsymbol{c}(x; i)$, note that, for fixed $(\boldsymbol{\alpha}, \boldsymbol{T})$, they satisfy the $k(k+2)$ dimensional system of homogeneous linear differential equations

$$\begin{aligned}
\dot{\boldsymbol{a}}(x) &= \boldsymbol{a}(x)\boldsymbol{T}, \\
\dot{\boldsymbol{b}}(x) &= \boldsymbol{T}\boldsymbol{b}(x), \\
\dot{\boldsymbol{c}}(x; i) &= \boldsymbol{a}_i(x)\boldsymbol{t}_0' + \boldsymbol{T}\boldsymbol{c}(x; i), \qquad \text{for} \quad i = 1, \ldots, k
\end{aligned} \tag{3.31}$$

with initial conditions

$$\begin{aligned}
\boldsymbol{a}(0) &= \boldsymbol{\alpha}, \\
\boldsymbol{b}(0) &= \boldsymbol{t}_0', \\
\boldsymbol{c}(0; i) &= \boldsymbol{0}, \qquad \text{for} \quad i = 1, \ldots k.
\end{aligned} \tag{3.32}$$

In the m^{th} step of the EM algorithm, the equations (3.31) are solved numerically using the Runge-Kutta method of 4^{th} order with $\boldsymbol{\alpha} = \boldsymbol{\alpha}^{(m)}$ and $\boldsymbol{T} = \boldsymbol{T}^{(m)}$.

For fitting a continuous distribution $G(\cdot)$ by a phase-type distribution, note that the sums in equations (3.30) can be interpreted as an integral of the empirical distribution. Thus, by inserting the corresponding integrals with respect to $G(\cdot)$, the EM algorithm takes on the form

$$
\begin{aligned}
\alpha_i^{(m+1)} &= \alpha_i^{(m)} \int_0^\infty \frac{\boldsymbol{b}_i(u|\boldsymbol{T}^{(m)})}{\boldsymbol{\alpha}^{(m)}\boldsymbol{b}(u|\boldsymbol{T}^{(m)})}\, dG(u), && \text{for} \quad i = 1, \ldots, k, \\
T_{ij}^{(m+1)} &= T_{ij}^{(m)} \frac{\displaystyle\int_0^\infty \frac{\boldsymbol{c}_j(u; i|(\boldsymbol{\alpha}, \boldsymbol{T})^{(m)})}{\boldsymbol{\alpha}^{(m)}\boldsymbol{b}(u|\boldsymbol{T}^{(m)})}\, dG(u)}{\displaystyle\int_0^\infty \frac{\boldsymbol{c}_i(u; i|(\boldsymbol{\alpha}, \boldsymbol{T})^{(m)})}{\boldsymbol{\alpha}^{(m)}\boldsymbol{b}(u|\boldsymbol{T}^{(m)})}\, dG(u)}, && \text{for} \quad i, j = 1, \ldots, k, \; j \neq i, \\
T_{i,k+1}^{(m+1)} &= \frac{\displaystyle\int_0^\infty \frac{\boldsymbol{a}_i(u|(\boldsymbol{\alpha}, \boldsymbol{T})^{(m)})t_i}{\boldsymbol{\alpha}^{(m)}\boldsymbol{b}(u|\boldsymbol{T}^{(m)})}\, dG(u)}{\displaystyle\int_0^\infty \frac{\boldsymbol{c}_i(u; i|(\boldsymbol{\alpha}, \boldsymbol{T})^{(m)})}{\boldsymbol{\alpha}^{(m)}\boldsymbol{b}(u|\boldsymbol{T}^{(m)})}\, dG(u)}, && \text{for} \quad i = 1, \ldots, k, \\
T_{ii}^{(m+1)} &= -\sum_{\substack{j=1 \\ j \neq i}}^{k+1} T_{ij}^{(m+1)} && \text{for} \quad i = 1, \ldots, k.
\end{aligned}
\tag{3.33}
$$

We note that the development above is the only one known to date dealing with general phase-type distributions. No special structure is assumed, e.g. no restriction (selection subset) on the class of phase-type distributions is necessary. Because of the structure of the EM algorithm, there is no danger in converging to a minimum of the likelihood function. Converging to local maxima, however, is possible. Each step of the EM algorithm increases the likelihood function. The first moment of the approximating phase-type distribution is unbiased, as shown in Asmussen and Nerman [5]. However, one cannot match more than the first moment using the EM

algorithm, and the convergence rate of the algorithm is quite slow. Furthermore, the EM algorithm does not allow for the computation of the standard errors of the approximated phase-type parameters.

The algorithm described above was implemented under the name EMPHT by Häggström, Asmussen, and Nerman [26] and further developed by Olsson [64].

Chapter 4

Distributions, Data and Test Criteria

This chapter presents the test data used in the computational evaluation of the parameter approximation methods described in chapter 3. Section 4.1 describes a set of known theoretical probability distributions to which phase-type distributions are to be fit. Section 4.2 presents four sets of empirical data which are suitable for phase-type approximations; these data were supplied by various researchers in the area of stochastic modelling. Section 4.3 introduces a set of test criteria for the goodness-of-fit of the phase-type approximations presented in chapter 5.

This set of test cases and performance measures is based largely on benchmark recommendations developed at the international workshop for fitting phase-type distributions held at Aalborg University, Denmark, in 1991. The motivation for including the Weibull (section 4.1.1) and lognormal distributions (section 4.1.2) is that their support are the nonnegative real numbers, as is the case for phase-type distributions. In addition, they are often encountered in the interpretation of experimental data from various areas in engineering, such as queueing applications, data transmission or reliability (see [12]). These distributions should therefore be fairly easy to approximate by phase-type distributions. The uniform (section 4.1.3) and the mixed-shifted exponential distributions (section 4.1.5) are typical examples of non-phase-type distributions and therefore present a greater challenge for a phase-type approximation. The matrix-exponential distribution (section 4.1.4) has a rational Laplace-Stieltjes transform, but is not a phase-type distribution.

4.1 Theoretical Distributions

This section describes the theoretical distributions used for the computational evaluation of the phase-type parameter approximation methods. For each distribution described, the formulæ for the cumulative distribution function $F(x)$ (cdf), the probability density function $f(x)$ (pdf), and the noncentral (or central) moments are presented (when applicable), together with plots of these functions. The r^{th} *noncentral moment* of a continuous random variable X with distribution $F(x)$ on the range $[0, \infty)$ is

$$\mu'_r = E[X^r] = \int_0^\infty x^r dF(x). \tag{4.1}$$

The r^{th} *central moment* about the mean μ'_1 is given by

$$\mu_r = E[(X - \mu'_1)^r] = \int_0^\infty (x - \mu'_1)^r dF(x). \tag{4.2}$$

4.1.1 Weibull Distributions

The Weibull distribution commonly occurs as the failure time distribution of components in microelectronics. Distributions modelling wear-out behavior in the area of reliability are often well represented by the Weibull distribution.

If a random variable X is distributed according to a Weibull distribution with scale parameter λ and shape parameter β, we write $X \sim W(\lambda, \beta)$, where $\lambda > 0$ and $\beta > 0$.

$$\begin{aligned} F(x) &= 1 - e^{-(\lambda x)^\beta} && \text{for} \quad x \geq 0 \\ f(x) &= \beta \lambda^\beta x^{\beta-1} e^{-(\lambda x)^\beta} && \text{for} \quad x \geq 0 \\ \mu'_r &= \lambda^{-r} \Gamma(1 + \tfrac{r}{\beta}) \end{aligned} \tag{4.3}$$

The following sets of parameters are used for the computational analysis in chapter 5.

$$\begin{aligned} W1: &\quad \lambda = 1 \quad \beta = 1.5 \\ W2: &\quad \lambda = 1 \quad \beta = 0.5 \end{aligned} \tag{4.4}$$

The first three noncentral moments are

$$\begin{aligned} W1: \quad \mu_1' &= \Gamma(1+\tfrac{1}{1.5}) = 0.9027453 \\ \mu_2' &= \Gamma(1+\tfrac{2}{1.5}) = 1.190639 \\ \mu_3' &= \Gamma(1+\tfrac{3}{1.5}) = 2 \\ W2: \quad \mu_1' &= \Gamma(3) = 2 \\ \mu_2' &= \Gamma(5) = 24 \\ \mu_3' &= \Gamma(7) = 720 \end{aligned} \tag{4.5}$$

Figures 4.1 and 4.2 show pdf and cdf plots of the Weibull test distributions with these parameters.

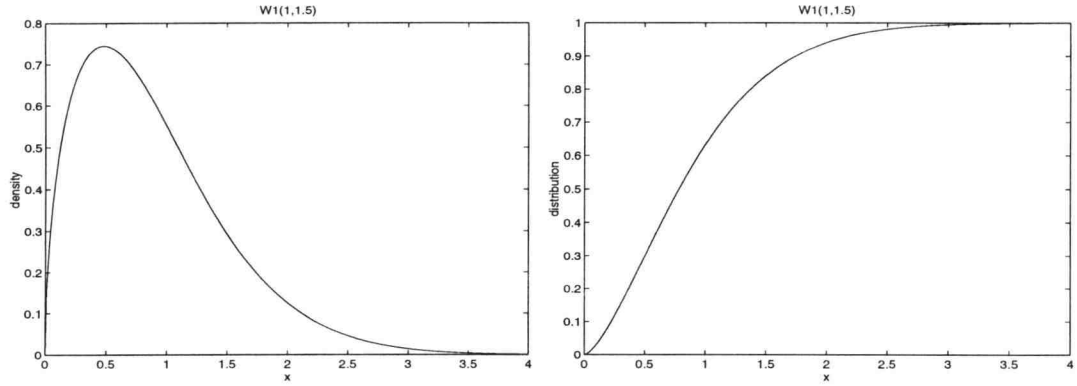

Figure 4.1. Density and distribution of $W1$: Weibull(1,1.5).

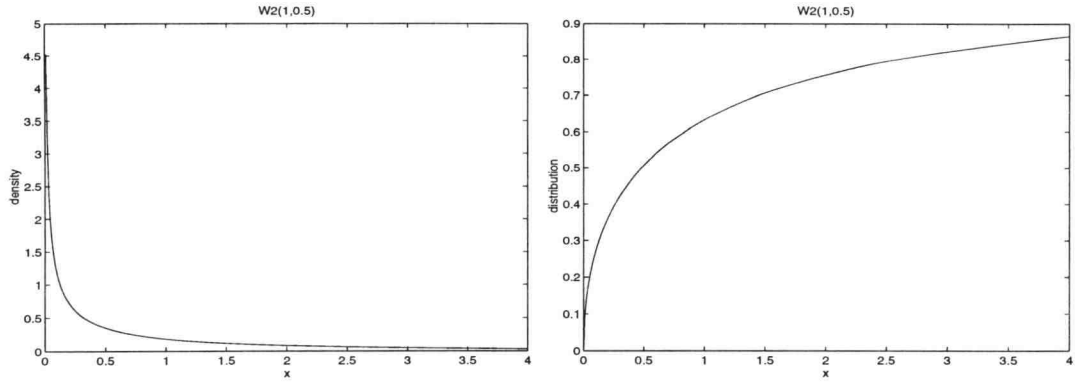

Figure 4.2. Density and distribution of $W2$: Weibull(1,0.5).

4.1.2 Lognormal Distributions

Lognormal distributions are also frequently used to model failure time distributions of electronic devices and are often encountered in the interpretation of experimental data in engineering applications, particularly in reliability and queueing.

If a random variable X is distributed according to a lognormal distribution with scale parameter α and shape parameter β, we write $X \sim L(\alpha, \beta)$, where $-\infty < \alpha < \infty$ and $\beta > 0$.

$$\begin{aligned} f(x) &= \frac{1}{\beta x\sqrt{2\pi}} e^{-\frac{(\ln(x)-\alpha)^2}{2\beta^2}} \quad \text{for} \quad x > 0 \\ \mu'_r &= e^{r\alpha + \frac{1}{2}r^2\beta^2} \end{aligned} \tag{4.6}$$

The following sets of parameters are used for the computational analysis in chapter 5.

$$\begin{aligned} L1: &\quad \alpha = -1.62 \quad \beta = 1.8 \\ L2: &\quad \alpha = -0.32 \quad \beta = 0.8 \\ L3: &\quad \alpha = -0.02 \quad \beta = 0.2 \end{aligned} \tag{4.7}$$

These parameter settings were chosen in accordance with the Aalborg benchmark recommendations. We note that the parameter values in all three cases above are such that $\alpha = -\frac{\beta^2}{2}$, so that the mean $\mu'_1 = 1$.

The first three noncentral moments are

$$\begin{aligned} L1: &\quad \mu'_1 = e^0 = 1 \\ &\quad \mu'_2 = e^{\beta^2} = 25.53372175 \\ &\quad \mu'_3 = e^{3\beta^2} = 16647.24473 \\ L2: &\quad \mu'_1 = 1 \\ &\quad \mu'_2 = 1.896480879 \\ &\quad \mu'_3 = 6.820958469 \\ L3: &\quad \mu'_1 = 1 \\ &\quad \mu'_2 = 1.040810774 \\ &\quad \mu'_3 = 1.127496852 \end{aligned} \tag{4.8}$$

Figures 4.3, 4.4 and 4.5 show pdf and cdf plots of the lognormal test distributions with these parameters.

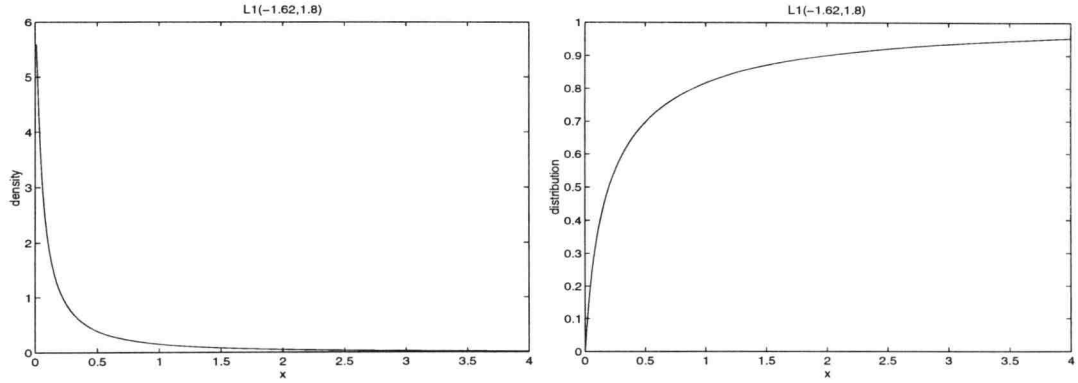

Figure 4.3. Density and distribution of $L1$: lognormal(-1.62,1.8).

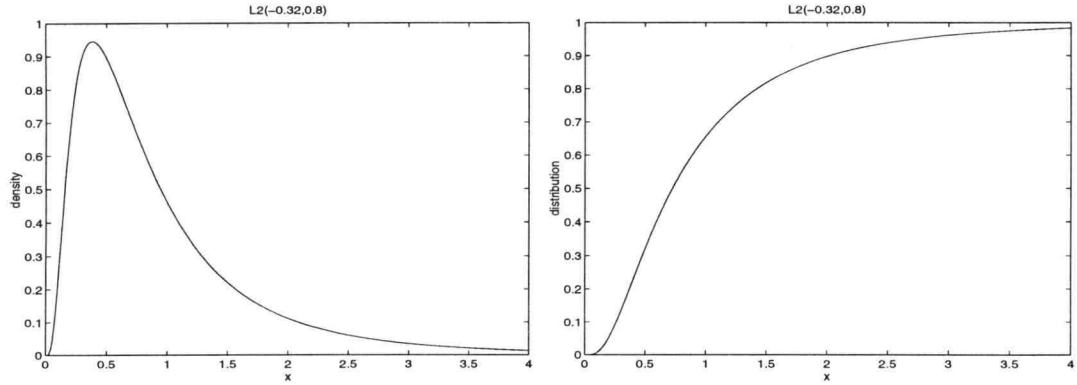

Figure 4.4. Density and distribution of $L2$: lognormal(-0.32,0.8).

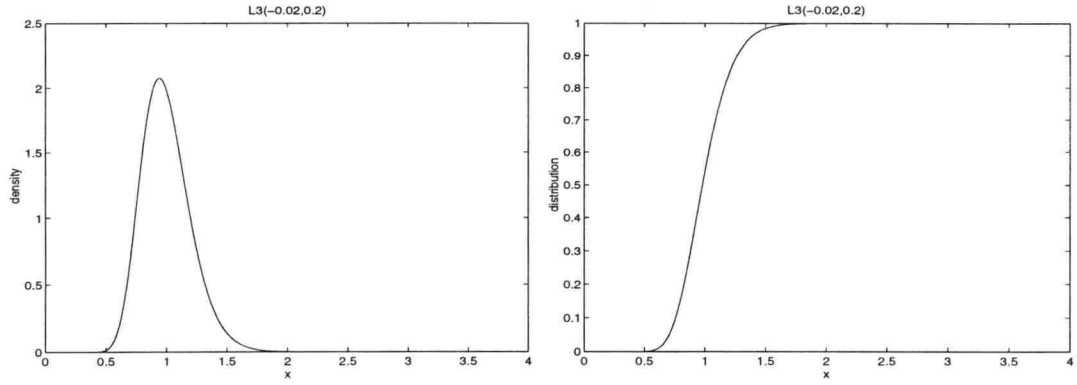

Figure 4.5. Density and distribution of $L3$: lognormal(-0.02,0.2).

4.1.3 Uniform Distributions

If a random variable X is distributed according to a Uniform (or rectangular) distribution with location parameter a (the lower limit of the range) and scale parameter b (the upper limit of the range), we write $X \sim U(a, b)$, where $-\infty < a < b < \infty$.

$$\begin{aligned} F(x) &= \tfrac{x-a}{b-a} && \text{for} \quad a \le x \le b \\ f(x) &= \tfrac{1}{b-a} && \text{for} \quad a \le x \le b \\ \mu_r &= \begin{cases} 0 & \text{for } r \text{ odd} \\ \frac{(b-a)^r}{2^r(r+1)} & \text{for } r \text{ even} \end{cases} \end{aligned} \tag{4.9}$$

The following sets of parameters are used for the computational analysis in chapter 5.

$$\begin{aligned} U1: &\quad a = 0 \quad b = 1 \\ U2: &\quad a = 1 \quad b = 2 \end{aligned} \tag{4.10}$$

The first three noncentral moments are

$$\begin{aligned} U1: &\quad \mu_1' = 0.5 \\ &\quad \mu_2' = 0.33333 \\ &\quad \mu_3' = 0.25 \\ U2: &\quad \mu_1' = 1.5 \\ &\quad \mu_2' = 2.33333 \\ &\quad \mu_3' = 3.75 \end{aligned} \tag{4.11}$$

Figures 4.6 and 4.7 show pdf and cdf plots of the uniform test distributions with these parameters.

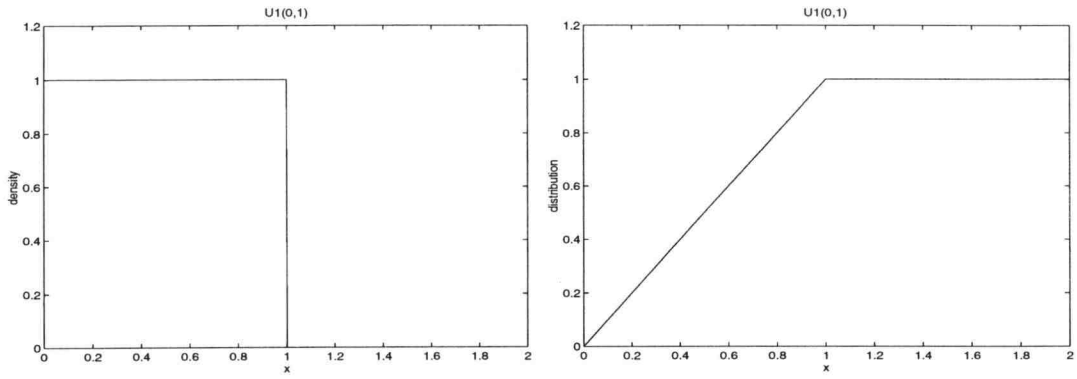

Figure 4.6. Density and distribution of $U1$: uniform(0,1).

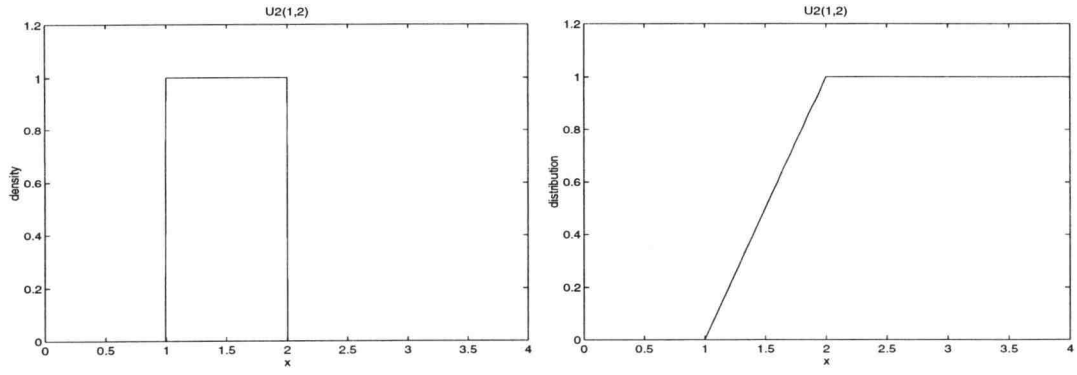

Figure 4.7. Density and distribution of $U2$: uniform(1,2).

4.1.4 Matrix-Exponential Distribution

The class of matrix-exponential distributions has been proposed by Lipsky [42] and consists of all distributions having a rational Laplace-Stieltjes transform, but are not necessarily represented by a scheme of phases. The class of phase-type distributions is contained in the class of matrix-exponential distributions. The matrix-exponential distribution chosen for this analysis was selected for exhibiting multimodal behavior, but is not a phase-type distribution.

The density function is

$$f(x) = \left(1 + \frac{1}{(2\pi)^2}\right)(1 - \cos(2\pi x))e^{-x} \qquad \text{for} \quad x \geq 0. \tag{4.12}$$

The noncentral moments are

$$\begin{aligned} \mu_1' &= \tfrac{4\pi^2+3}{4\pi^2+1} &&= 1.049409046 \\ \mu_2' &= 4\tfrac{8\pi^4+6\pi^2+3}{(4\pi^2+1)^2} &&= 2.054291554 \\ \mu_3' &= 12\tfrac{32\pi^6+32\pi^4+10\pi^2+5}{(4\pi^2+1)^3} &&= 6.148950858 \end{aligned} \tag{4.13}$$

Figure 4.8 shows pdf and cdf plots of the matrix-exponential distribution.

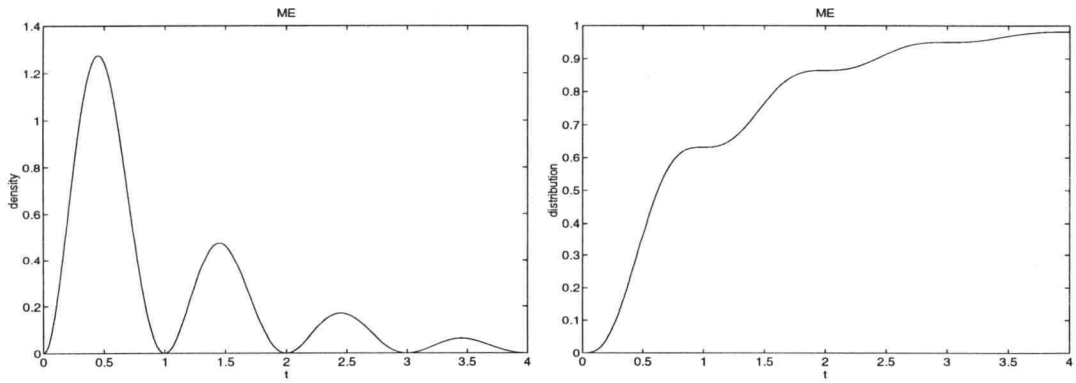

Figure 4.8. Density and distribution of ME: matrix-exponential.

4.1.5 Mixed Shifted Exponential Distribution

The distribution described in this section is an equal-probability mixture of an exponential and a shifted exponential distribution. This distribution was selected for

exhibiting a jump in the density function, which is presumably difficult to represent by a phase-type distribution.

The density is

$$f(x) = \frac{1}{2}e^{-x}I_{\{x\geq 0\}} + \frac{1}{2}e^{-(x-1)}I_{\{x\geq 1\}} \tag{4.14}$$

The noncentral moments are

$$\begin{aligned} \mu_1' &= 1.5 \\ \mu_2' &= 3.5 \\ \mu_3' &= 11.0 \end{aligned} \tag{4.15}$$

Figure 4.9 shows pdf and cdf plots of the mixed shifted exponential distribution.

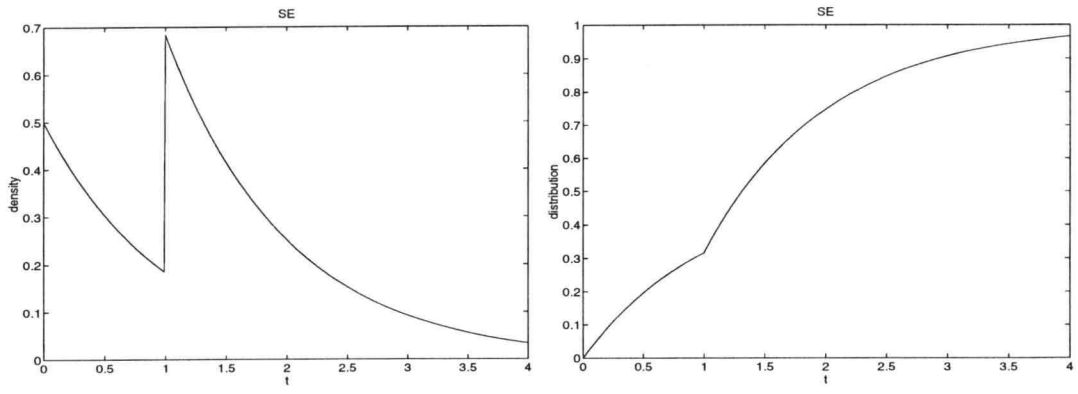

Figure 4.9. Density and distribution of SE: mixed shifted exponential.

4.1.6 Data Generation

A set of 200 observations was generated from each of the theoretical distributions described in the previous sections by dividing the range of the respective distribution (0,1) into 200 equal intervals and calculating the inverse cdf function at the midpoints of these intervals. The following table 4.1 summarizes these test cases:

Table 4.1. Summary of the test cases based on the theoretical distributions.

Theoretical Test Distributions			
data set	density	theoretical moments	empirical moments
W1	Weibull($\lambda = 1, \beta = 1.5$)	$\mu'_1 = 0.90275$	$\hat{\mu}'_1 = 0.90216$
		$\mu'_2 = 1.19064$	$\hat{\mu}'_2 = 1.18613$
		$\mu'_3 = 2$	$\hat{\mu}'_3 = 1.97406$
W2	Weibull($\lambda = 1, \beta = 0.5$)	$\mu'_1 = 2$	$\hat{\mu}'_1 = 1.97407$
		$\mu'_2 = 24$	$\hat{\mu}'_2 = 20.974924$
		$\mu'_3 = 720$	$\hat{\mu}'_3 = 430.16871$
L1	Lognormal($\alpha = -1.62, \beta = 1.8$)	$\mu'_1 = 1$	$\hat{\mu}'_1 = 0.93382$
		$\mu'_2 = 25.53372$	$\hat{\mu}'_2 = 8.61377$
		$\mu'_3 = 16647.24473$	$\hat{\mu}'_3 = 185.52266$
L2	Lognormal($\alpha = -0.32, \beta = 0.8$)	$\mu'_1 = 1$	$\hat{\mu}'_1 = 0.99618$
		$\mu'_2 = 1.89648$	$\hat{\mu}'_2 = 1.81672$
		$\mu'_3 = 6.82096$	$\hat{\mu}'_3 = 5.4754$
L3	Lognormal($\alpha = -0.02, \beta = 0.2$)	$\mu'_1 = 1$	$\hat{\mu}'_1 = 0.99987$
		$\mu'_2 = 1.04081$	$\hat{\mu}'_2 = 1.04018$
		$\mu'_3 = 1.1275$	$\hat{\mu}'_3 = 1.1256$
U1	Uniform($a = 0, b = 1$)	$\mu'_1 = 0.5$	$\hat{\mu}'_1 = 0.5$
		$\mu'_2 = 0.33333$	$\hat{\mu}'_2 = 0.33333$
		$\mu'_3 = 0.25$	$\hat{\mu}'_3 = 0.24999$
U2	Uniform($a = 1, b = 2$)	$\mu'_1 = 1.5$	$\hat{\mu}'_1 = 1.5$
		$\mu'_2 = 2.33333$	$\hat{\mu}'_2 = 2.33333$
		$\mu'_3 = 3.75$	$\hat{\mu}'_3 = 3.74999$
ME	Matrix-Exponential	$\mu'_1 = 1.04941$	$\hat{\mu}'_1 = 1.04721$
		$\mu'_2 = 2.05429$	$\hat{\mu}'_2 = 2.01987$
		$\mu'_3 = 6.14895$	$\hat{\mu}'_3 = 5.77110$
SE	Mixed-Shifted Exponential	$\mu'_1 = 1.5$	$\hat{\mu}'_1 = 1.49827$
		$\mu'_2 = 3.5$	$\hat{\mu}'_2 = 3.47191$
		$\mu'_3 = 11.0$	$\hat{\mu}'_3 = 10.65516$

4.2 Empirical Data

The empirical data decribed in this section and used in the computational evaluation were supplied by Prof. Søren Asmussen, Institute of Electronic Systems, Aalborg University, Denmark, who originally received them from Prof. O. Kella, Hebrew University, Jerusalem, and Prof. A. Mandelbaum, Technion, Haifa, Israel. They represent measured lengths of incoming telephone calls to the service center of one of Israel's major television cable companies. The following types of telephone calls are recorded:

- type 1 (home services): notices from subscribers on problems and transferring the information to technicians.
- type 2 (sales): notices on sales actions, including seeking help on prices, times, clarifications with sales representatives, etc.
- type 3 (billings): providing information to customers on payment procedures.
- type 4 (general information): including change of address, private calls.

Figures 4.10-4.13 present empirical pdf and cdf plots of the data. Table 4.2 reports the number of observations and the first three noncentral empirical moments for each data set.

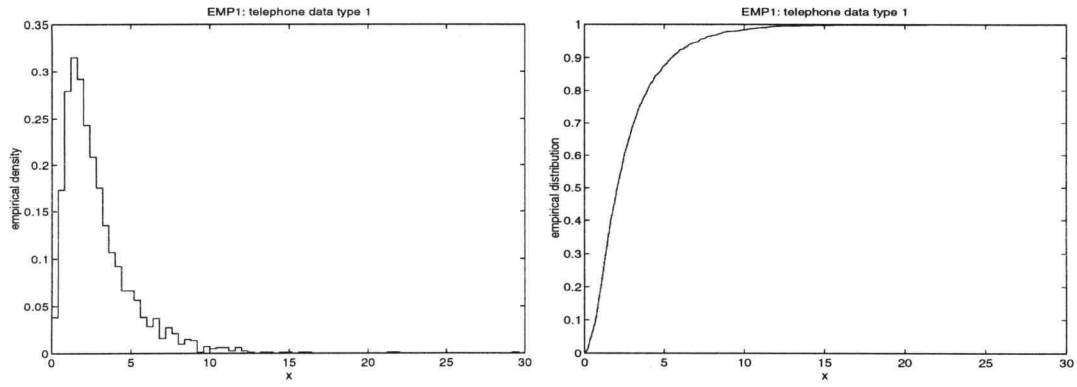

Figure 4.10. Density and distribution of $EMP1$: telephone data of type 1.

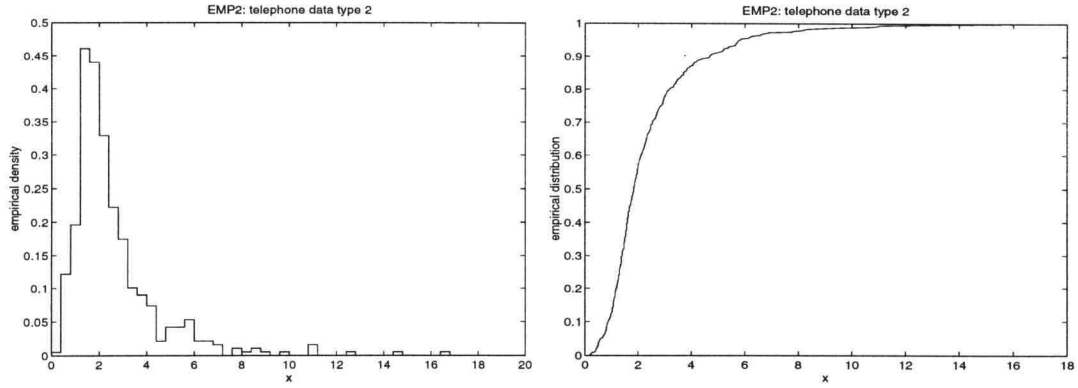

Figure 4.11. Density and distribution of $EMP2$: telephone data of type 2.

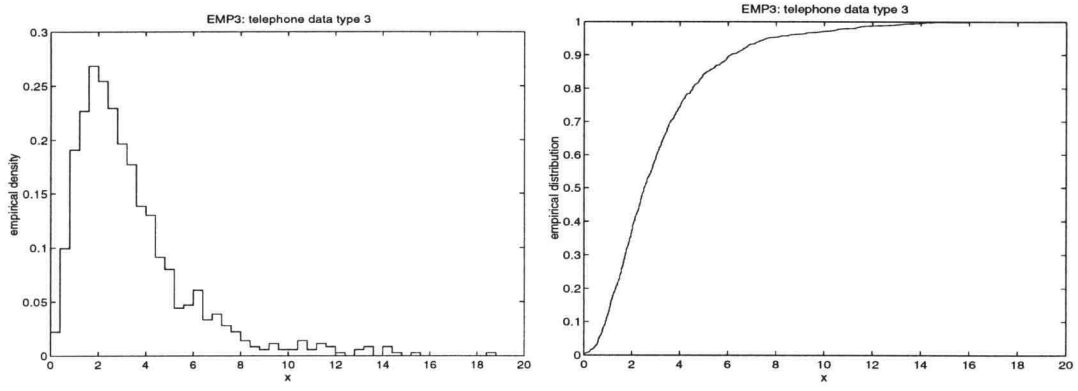

Figure 4.12. Density and distribution of $EMP3$: telephone data of type 3.

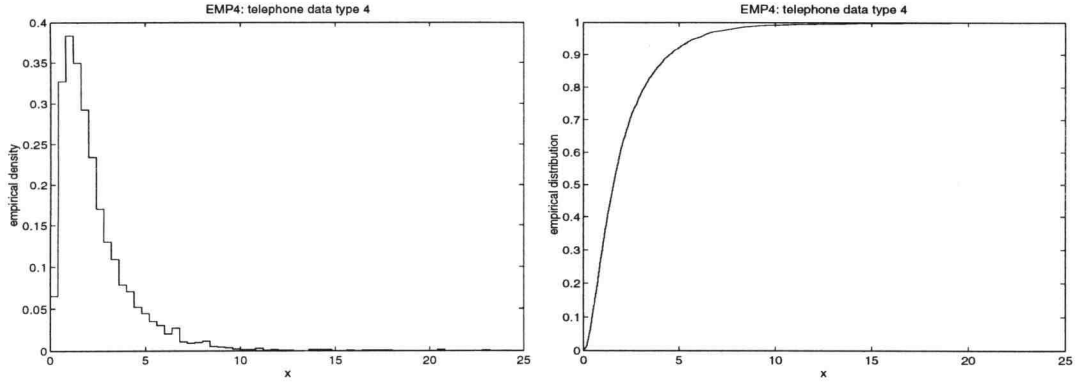

Figure 4.13. Density and distribution of $EMP4$: telephone data of type 4.

Table 4.2. Summary of the empirical test data.

Empirical Test Data		
data set	# of observations	empirical moments
EMP1	2039	$\hat{\mu}'_1 = 2.69321$
		$\hat{\mu}'_2 = 12.60459$
		$\hat{\mu}'_3 = 96.7668$
EMP2	472	$\hat{\mu}'_1 = 2.40502$
		$\hat{\mu}'_2 = 9.51204$
		$\hat{\mu}'_3 = 61.43339$
EMP3	904	$\hat{\mu}'_1 = 3.17723$
		$\hat{\mu}'_2 = 16.13827$
		$\hat{\mu}'_3 = 118.96632$
EMP4	3189	$\hat{\mu}'_1 = 2.14653$
		$\hat{\mu}'_2 = 8.65458$
		$\hat{\mu}'_3 = 59.84374$

4.3 Test Criteria

Malhotra and Reibman [46] mention the following concerns that influence the choice of phase-type approximations:

- *The achievable goodness-of-fit.* Measures of *closeness* include matching several moments and incorporating shape constraints of the cdf or pdf to be approximated.
- *The number of transient states of the phase-type approximation.* If phase-type distributions are applied - for instance in matrix-analytic methods - it is important that the order is reasonably small.

- *The ease of generation of the resulting Markovian model.* It is convenient to obtain *sparse* phase-type distributions, i.e. the structure of possible transitions in the Markov chain should be kept reasonably simple.

- *The ease of approximating the parameters.* Time and memory requirements of the approximation procedure, as well as robustness and accuracy, are important.

Let $F(x)$ and $f(x)$ be the cdf and pdf of the (empirical) distribution to be approximated and $\hat{F}(x)$ and $\hat{f}(x)$ be the cdf and pdf of the approximating phase-type distribution. The r^{th} noncentral moments of $F(x)$ and $\hat{F}(x)$ are denoted by μ'_r and $\hat{\mu}'_r$, respectively. The centered moments of $F(x)$ and $\hat{F}(x)$ are denoted by μ_r and $\hat{\mu}_r$, respectively. The following set of performance measures, based largely on the 1991 conference on phase-type fitting held at Aalborg University, Denmark, is used for the evaluation of the phase-type approximations to the theoretical distributions and the empirical data sets described in the previous sections:

- Area difference between the distributions:

$$\Delta cdf = \frac{1}{\mu'_1} \int_0^\infty |\hat{F}(t) - F(t)| \, dt \tag{4.16}$$

- Area difference between the densities:

$$\Delta pdf = \int_0^\infty |\hat{f}(t) - f(t)| \, dt \tag{4.17}$$

- Cross entropy:

$$\hat{H} = -\int_0^\infty \log(\hat{f}(t)) \, dF(t) \tag{4.18}$$

- Relative entropy:

$$H_r = \int_0^\infty \log \frac{f(t)}{\hat{f}(t)} \, dF(t) \tag{4.19}$$

  The relative entropy (information divergence or Kullback-Leibler information [40]) is defined as the difference between the cross entropy (as defined in 4.18 above) between two probability densities, and the intrinsic entropy

$H = -\int_0^\infty \log(f(t))\, dF(t)$ of the probability density to be approximated. The cross entropy $\hat{H}$ converges from above to the intrinsic entropy H. As a measure of deviance between the two densities $f(x)$ and $\hat{f}(x)$, the relative entropy provides a measure of the goodness of the phase-type approximation.

- Relative error on the mean:

$$e_1 = \frac{|\hat{\mu}_1' - \mu_1'|}{\mu_1'} \tag{4.20}$$

- Relative error on the variance:

$$e_2 = \frac{|\hat{\mu}_2 - \mu_2|}{\mu_2} \tag{4.21}$$

- Relative error on the coefficient of skewness:

$$e_3 = \frac{|\hat{\mu}_3 - \mu_3|}{\mu_3} \tag{4.22}$$

The performance measures (4.16-4.19) were computed for the phase-type approximation results obtained from the two maximum-likelihood based methods (MLAPH and EMPHT) and the two moment-matching methods (MEFIT and MEDA) to the theoretical test distributions ($W1$-SE). Moreover, the relative moment errors (4.20-4.22) were obtained for the maximum-likelihood based approximations. The relative moment errors were not investigated for the moment-matching approximations, however, since the first three moments of the empirical distributions are matched exactly in MEDA and up to a user-specified tolerance in MEFIT.

The performance measure for phase-type approximations based on the empirical test cases $EMP1$-$EMP4$ is the area difference between the distributions (4.16). In addition, the relative moment errors (4.20-4.22) were computed for the maximum-likelihood based approximations.

The computation of the performance measures and the generation of the plots presented in chapter 5 was carried out in MATLAB [49]. The integrals involved in the area difference and entropy measures were computed numerically using the function

quad8, a quadrature algorithm involving an adaptive recursive Newton-Cotes 8-panel rule. For comparisons between phase-type approximations to the theoretical test distributions, the area integrals and entropy measures were generally computed between the origin and $50 \times \mu_1$. An exception was test case $L1$, whose pdf function has an extremely long tail; area measures for $L1$ were computed between the origin and $150 \times \mu_1$. For comparisons of phase-type aproximations to the empirical data sets, the area difference between the distributions was computed between the origin and the maximum observation.

In addition to the performance measures, the following results of the phase-type approximations are reported:

- Order of the approximating phase-type distribution.
- General attributes:
  - *Generality*: applicability of the parameter approximation methods to a wide variety of fitting problems.
  - *Reliability*: proper run behavior of the algorithms in the test cases.
  - *Stability*: small changes in the data (or input parameters) should not cause large changes in the results.
  - *Accuracy*: refers to errors introduced by truncating infinite series or terminating iterations.
  - *Efficiency*: amount of CPU time necessary for the parameter approximation runs.
- Plots of the approximated and approximating distributions and densities.

Chapter 5

Parameter Approximation

Thou com'st in such a questionable shape.
Shakespeare, Hamlet, Act I, Scene IV.

This chapter presents the results of the computational evaluation of fitting phase-type approximations to the test distributions described in chapter 4. Sections 5.1 and 5.2 contain results from the moment-matching approaches MEFIT and MEDA, sections 5.3 and 5.4 contain results from the maximum-likelihood based methods MLAPH and EMPHT, respectively. Section 5.5 concludes this chapter by presenting a comparison between the analyzed phase-type fitting methods; the moment-matching approaches are discussed in subsection 5.5.1 and the maximum-likelihood based methods in subsection 5.5.2.

All computational experiments, except for MEDA parameter approximations, were conducted on SPARC-IPX workstations (40 MHz clockspeed, 32 MB RAM) running under SunOS 4.1.3. MEDA experiments were run on a PC with an 80486 CPU (33MHz clockspeed, 8MB RAM) running under MSDOS 5.0. It should be pointed out that the run times presented for the approximation runs performed on the SPARC-IPX workstations are estimates based on user time computed by the SunOS operating system. Since these computers allow for multiple users and processes, some variation on the true run times should be expected.

5.1 MEFIT

This section contains the experimental results obtained from approximating the parameters of phase-type distributions based on the test distributions $W1$-SE and empirical data sets $EMP1$-$EMP4$ (see chapter 4) using the moment-matching method MEFIT (Johnson [33, 35]). The following subsections 5.1.1 and 5.1.2 present the general attributes of MEFIT and the numerical results obtained from our experiments.

MEFIT is written in FORTRAN and requires the NLP solver NPSOL [24] (written in FORTRAN as well) for the solution of the nonlinear equations involved in the moment-matching fitting algorithm. However, the MEFIT source code is adjustable for using other NLP solvers.

MEFIT consists of the main executable program (compiled with the *f77* SPARC FORTRAN compiler from SunPro, release 2.0.1, on a SPARC-IPX computer) and several input and output files. Input files consist of the empirical data (observations) in file *xdf.mat* with corresponding cdf or pdf values in the files *cdf.mat* or *pdf.mat*, respectively. Output files (if requested by the user) are *xdfme.mat, cdfme.mat* and *pdfme.mat*, which contain a vector of x-values and corresponding cdf and pdf values, respectively. The file extension *.mat* refers to the naming convention used in MATLAB [49], which can be used for plotting purposes.

At the program start, the user is asked to supply several parameters (from the standard input) which determine the process of the fitting run:

- *parameters of the Erlang mixture:* the number and orders of the Erlang distributions in the mixture, initial values of and bounds for the means and mixing probabilities of each Erlang distribution, and bounds on differences of the means of consecutively specified Erlang distributions, therefore preserving the ordering of the individual means. Up to six Erlang distributions may be mixed. For our analysis, however, we fit only Erlang mixtures with two, three and four branches.

- *fitting specifications:* up to six moments of order six or lower may be specified for inclusion in the fitting algorithm. Furthermore, the user may also specify values of the approximated cdf or pdf at up to 200 points. This limit was changed to 3200 points in order to accommodate data files containing a larger number of observations. Matching the moments (up to a user-specified tolerance level) may be carried out by the algorithm either by including the corresponding equations in the objective function or in the form of constraints. For our analysis, we fit all examples with the first three noncentral empirical moments of the test data specified as constraints, using tolerance level=0.0001. The empirical cdf values (in file *cdf.mat*) are incorporated in the objective function by minimizing the sum of the squared differences between the empirical and approximating cdf at the observations (in file *xdf.mat*).

- *request for optional output:* output consists of x-values (in file *xdfme.mat*), cdf values (in file *cdfme.mat*) and pdf values (in file *pdfme.mat*) of the approximating Erlang mixture at up to 200 equally spaced points within a user-specified interval $[0, b]$.

- *NPSOL optional parameters:* The NPSOL code allows the user to specify several parameters guiding its search algorithm. These parameters are read from the file *options*. MEFIT uses this feature to specify that all gradients are to be approximated by finite differences.

In addition to the aforementioned output files, MEFIT echoes the user-specified input parameters and reports the final distributional parameters (mixing probabilities and means) of the Erlang mixture on the standard output.

5.1.1 MEFIT General Results

- *Generality:* MEFIT approximations incorporate only *mixtures of Erlang* distributions as the selection subset of the class of phase-type distributions. Other structures, such as Coxian or general phase-type distributions, can not be fit.

The hyperexponential distribution (a mixture of exponential distributions) is contained in the class of Erlang mixtures by mixing Erlang distributions of order one.

- *Reliability:* The moment-matching algorithm contained in MEFIT was very reliable in all our experiments. In no cases did the program terminate abnormally. For particular initial parameter value settings, however, the NPSOL routine did terminate reporting a suboptimal solution. Reasons were: (i) exceeding the maximum allowable number of iterations (may be changed by the user in the file *options*); (ii) finding a solution but without reaching the prescribed accuracy; or (iii) inability to improve on a current solution. Changing the initial parameter value settings yielded improvements (optimal solutions reported by the NPSOL routine) in all such cases.

- *Stability:* MEFIT requires a large amount of user interaction (mostly based on heuristics and experience) for finding phase-type approximations. The number r of Erlang branches as well as the order k_i of each branch must be determined by the user. A separate program *minord* assists the user in finding the minimum order k^* of at least one of the Erlang branches, based on the first three empirical moments of the data (see section 3.2.1). Moreover, initial values for the mixing probabilities p_i and the means $\beta_i = \frac{k_i}{\lambda_i}$ (λ_i is the parameter of the i^{th} Erlang distribution in the mixture) must be supplied.

  It was found during our experimentation that even small changes in these initial parameter settings yielded significantly different final parameter values. Potential users of MEFIT should have a certain amount of experience in the properties of phase-type distributions in order to determine appropriate initial parameter settings. All our results presented in section 5.1.2 are based on approximation runs which might be improved by a more experienced user. With our limited experience, we tried to find the best possible phase-type approximations.

- *Accuracy:* The NPSOL routine reports in some cases that the maximum number of iterations was exceeded or that a presented optimal solution was found without achieving the required accuracy. It was possible in all such cases to find better solutions by changing the initial parameter values. If an optimal solution is found by MEFIT, it is ensured that the user-specified empirical moments are matched by the corresponding moments of the approximating Erlang mixture up to a certain tolerance level (in our experiments the tolerance was set at 0.0001).

- *Efficiency:* Parameter approximations using MEFIT generally required a small amount of CPU time, in the range of 1-10 seconds. One exception was a fit of a four branch Erlang mixture with orders 5, 40, 50, 50 to the ME test data set, which required 64.5 seconds.

Table 5.1 presents the orders and run times of the mixture of Erlang approximations obtained from MEFIT.

Table 5.1. MEFIT PH approximations: run times and orders of the approximating mixtures of Erlang distributions.

		Mixture of Erlang distributions		
Test distribution		2 branches	3 branches	4 branches
W1	run time (sec.)	0.7	1.9	4.2
	orders	2, 4	2, 3, 4	2, 3, 4, 2
W2	run time (sec.)	0.8	1.8	8.8
	orders	1, 1	1, 1, 1	1, 1, 2, 1
L1	run time (sec.)	0.7	5.2	10.2
	orders	1, 1	1, 1, 1	1, 1, 1, 1
L2	run time (sec.)	2.2	4.5	9.9
	orders	2, 2	3, 3, 5	3, 3, 5, 1
L3	run time (sec.)	1.2	2.3	7.6
	orders	2, 25	2, 2, 25	2, 2, 26, 10
U1	run time (sec.)	1.7	3.5	8.6
	orders	2, 23	2, 2, 25	2, 2, 25, 5
U2	run time (sec.)	3.1	6.3	9.7
	orders	1, 30	1, 1, 30	1, 1, 30, 22
ME	run time (sec.)	0.9	5.9	64.5
	orders	8, 2	5, 50, 10	5, 40, 50, 50
SE	run time (sec.)	2.5	6.9	11.9
	orders	2, 6	1, 3, 2	1, 3, 3, 3
EMP1	run time (sec.)	2.1	2.7	4.0
	orders	2, 2	1, 2, 2	1, 2, 2, 1
EMP2	run time (sec.)	0.9	2.3	3.1
	orders	5, 2	1, 5, 2	1, 5, 2, 1
EMP3	run time (sec.)	0.9	1.9	5.8
	orders	3, 2	1, 3, 3	1, 3, 3, 2
EMP4	run time (sec.)	2.4	7.6	4.4
	orders	2, 1	1, 2, 1	1, 2, 1, 1

5.1.2 MEFIT Experimental Results

Density and distribution plots: theoretical test cases

Figure 5.1 shows plots of the original theoretical densities ($W1$-SE) together with the approximating mixture of Erlang densities involving two, three and four Erlang branches. Figure 5.2 presents plots of the original theoretical distributions along with the approximating mixture of Erlang distributions involving two, three and four Erlang branches.

A visual inspection shows that the MEFIT approximations involving two Erlang branches are in most cases not satisfactory, whereas the fits involving three or four Erlang branches tend to approximate the original densities and distributions more closely. This is particularly noticeable in the test cases $W1$, $W2$, $L1$, $L2$, $L3$, ME and SE. Obvious disagreements between the original densities and approximating phase-type densities are evident in the test cases $U1$, $U2$, ME and SE.

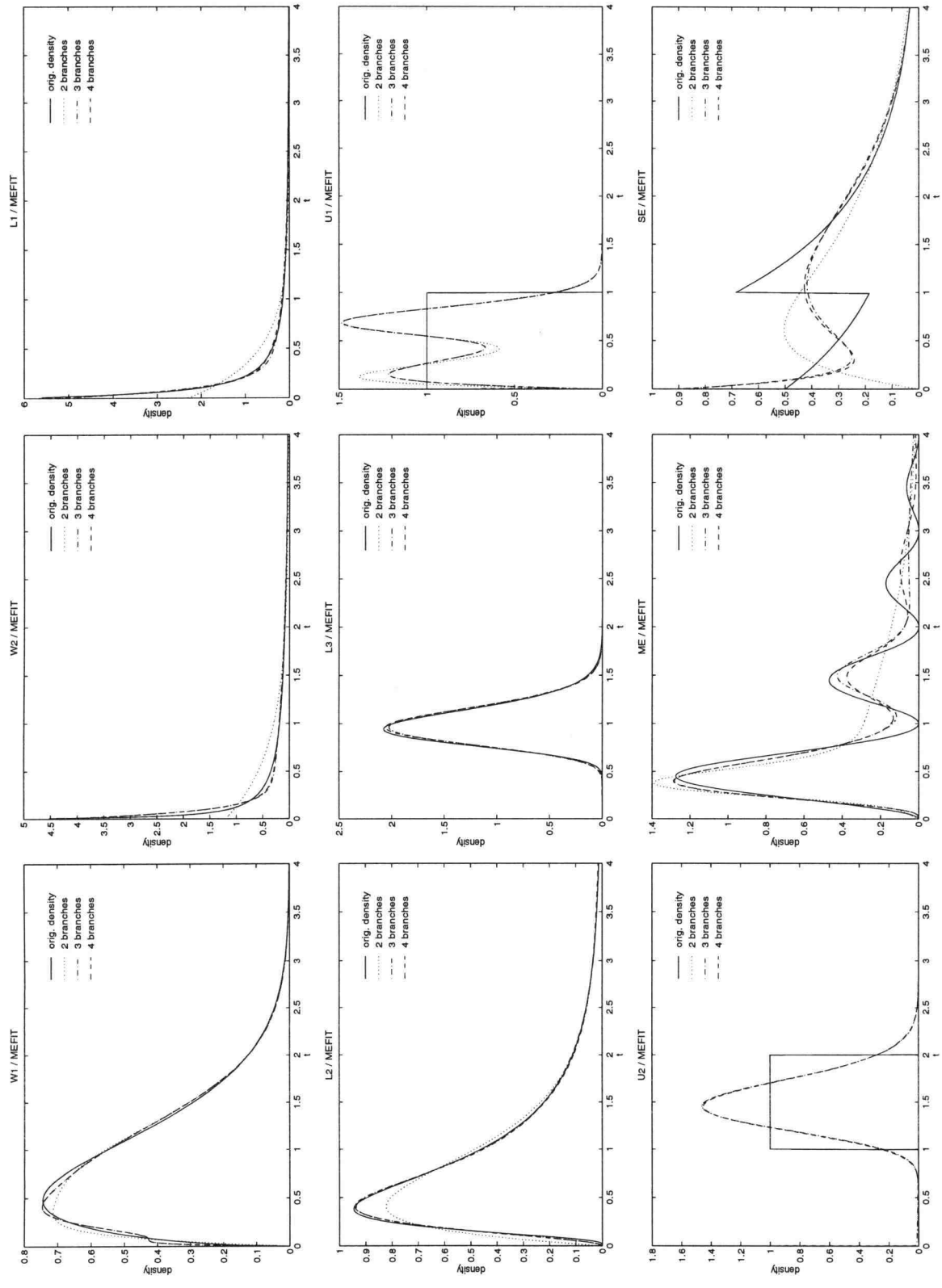

Figure 5.1. MEFIT approximations to the theoretical densities.

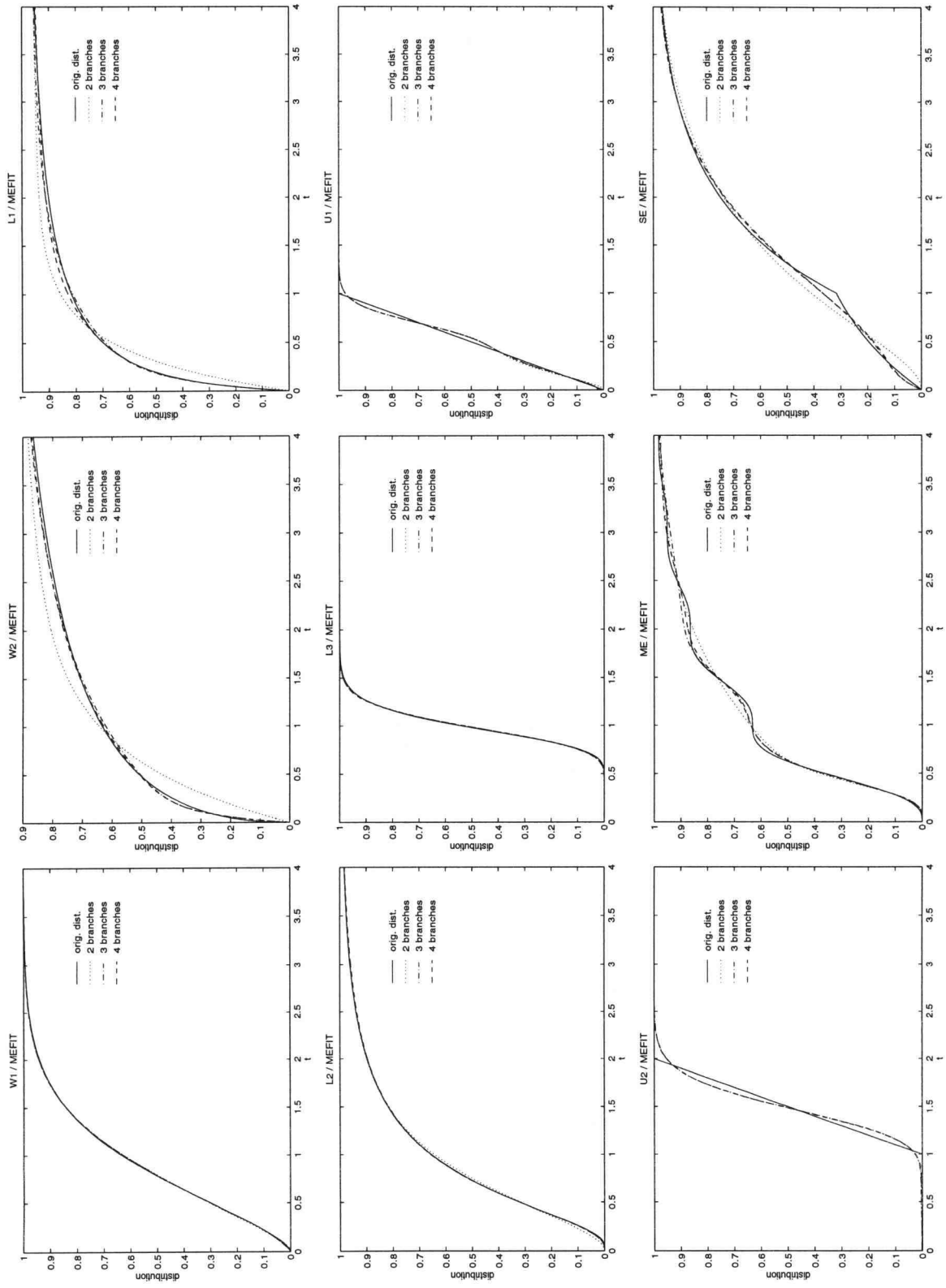

Figure 5.2. MEFIT approximations to the theoretical distributions.

Performance measures: theoretical test cases

Table 5.2 presents the area differences as well as the entropy measures between the theoretical test densities and distributions ($W1$-SE) and the approximating mixture of Erlang densities and distributions involving two, three and four branches. Also presented is the order of the phase-type distribution corresponding to the mixture of Erlang distributions (see appendix B for phase-type representations of Erlang mixtures).

The pdf and cdf area differences as well as the entropy measures are generally decreasing with increasing number of branches and orders of the approximating Erlang mixture. Slight inconsistencies are present in test cases $W1$, $L2$, $L3$, $U1$ and SE, but improvements of the fits based on user experience are possible, as mentioned in section 5.1.1.

It was necessary to choose high orders of the approximating Erlang mixture in order to attempt to recover the density shapes of the more challenging distributions $U1$, $U2$, ME and SE. Of particular interest are the fits to test case ME, where high orders were purposely selected to approximate the multiple modes and poles of the original density. The area differences and entropy measures show, however, that this approach was only partially successful.

The large order fits to $L3$ were necessary because this distribution has low variability (squared coefficient of variation = 0.041). Aldous and Shepp [2] show that the minimal squared coefficient of variation of an order-k phase-type distribution is equal to $1/k$ and is attained by the Erlang distribution of order k. For the $L3$ test distribution it was therefore necessary to use at least an order-25 phase-type distribution for an adequate approximation.

Large pdf area differences resulted from MEFIT approximations to test cases $U1$, $U2$, ME and SE, because of their challenging density function shapes. Moreover, surprisingly large pdf area differences resulted from the MEFIT approximations to test cases $W2$ and $L1$, although the visual fit does seem to be satisfactory for three and four Erlang branches in both cases. A reason for this phenomenon is that

the $L1$ and the $W2$ distributions have an extremely long tail ($\mu_3' = 16647.24$ for $L1$, $\mu_3' = 720$ for $W2$). Phase-type distributions, however, have an exponentially decaying tail, which makes the approximation of long-tailed distributions difficult.

Table 5.2. MEFIT PH approximations to the theoretical distributions: area differences and entropy measures.

Test distribution		Mixture of Erlang distributions		
		2 branches	3 branches	4 branches
W1	order of PH	6	9	11
	Δpdf	0.040617	0.031482	0.032369
	Δcdf	0.009917	0.007460	0.007637
	cross entropy	0.788568	0.788146	0.788248
	relative entropy	0.001693	0.001271	0.001374
W2	order of PH	2	3	5
	Δpdf	0.433801	0.206894	0.188909
	Δcdf	0.148747	0.058638	0.050888
	cross entropy	1.351233	1.186433	1.184311
	relative entropy	0.236146	0.071346	0.069224
L1	order of PH	2	3	4
	Δpdf	0.419492	0.114236	0.087368
	Δcdf	0.253672	0.134939	0.133350
	cross entropy	0.514507	0.408139	0.405828
	relative entropy	0.124785	0.018417	0.016106
L2	order of PH	4	11	12
	Δpdf	0.088871	0.022855	0.024819
	Δcdf	0.023073	0.009491	0.014912
	cross entropy	0.890987	0.878713	0.878826
	relative entropy	0.015193	0.002919	0.003032
L3	order of PH	27	29	40
	Δpdf	0.050417	0.050745	0.046719
	Δcdf	0.006428	0.006495	0.005863
	cross entropy	-0.206902	-0.206780	-0.207329
	relative entropy	0.003597	0.003720	0.003171
U1	order of PH	25	29	34
	Δpdf	0.329649	0.307500	0.307176
	Δcdf	0.050932	0.044412	0.044625
	cross entropy	0.110721	0.111215	0.105112
	relative entropy	0.110721	0.111215	0.105112
U2	order of PH	31	32	54
	Δpdf	0.422486	0.422486	0.417376
	Δcdf	0.036150	0.036150	0.035062
	cross entropy	0.188491	0.188491	0.184671
	relative entropy	0.188491	0.188491	0.184671
ME	order of PH	10	65	145
	Δpdf	0.381293	0.236773	0.239328
	Δcdf	0.054537	0.042760	0.038591
	cross entropy	0.882325	0.825052	0.823965
	relative entropy	0.154594	0.097321	0.096234
SE	order of PH	8	6	10
	Δpdf	0.336804	0.215890	0.213871
	Δcdf	0.063206	0.032675	0.029273
	cross entropy	1.459252	1.333473	1.334810
	relative entropy	0.164290	0.038511	0.039848

Density and distribution plots: empirical test cases

Figure 5.3 shows plots of the empirical test densities and distributions ($EMP1$-$EMP4$) together with the approximating mixture of Erlang densities and distributions involving two, three and four Erlang branches.

The visual inspection shows that the number of Erlang branches does not influence the goodness-of-fit of phase-type approximations to the empirical data significantly. The approximating densities and distributions are virtually indistinguishable in terms of their number of Erlang branches. The empirical densities as well as distribution functions are quite well approximated in all four test cases.

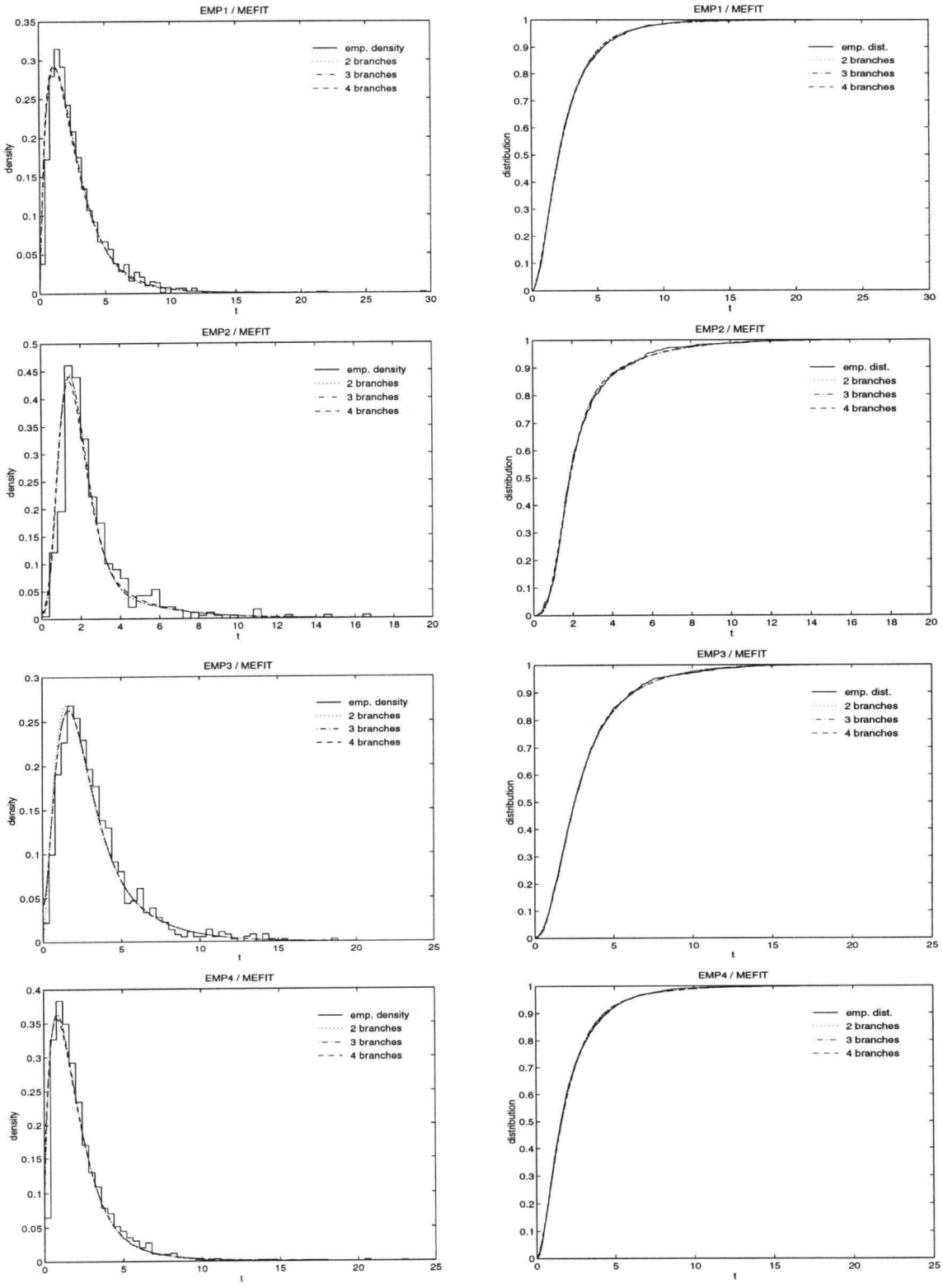

Figure 5.3. MEFIT approximations to the empirical densities and distributions.

Table 5.3. MEFIT PH approximations to the empirical data: area differences between the approximated and the approximating distributions and orders of the phase-type distributions corresponding to a mixture of Erlang distributions.

		Mixture of Erlang distributions		
Test distribution		2 branches	3 branches	4 branches
EMP1	order of PH	4	5	6
	Δcdf	0.027957	0.027904	0.015755
EMP2	order of PH	7	8	9
	Δcdf	0.032903	0.032736	0.025915
EMP3	order of PH	5	7	9
	Δcdf	0.022752	0.017951	0.018360
EMP4	order of PH	3	4	5
	Δcdf	0.029693	0.029626	0.026982

Performance measures: empirical test cases

Table 5.3 presents the cdf area differences between the empirical test distributions ($EMP1$-$EMP4$) and the approximating mixture of Erlang distributions. Also presented is the order of the phase-type distribution corresponding to the Erlang mixtures (see appendix B).

The cdf area differences are slightly decreasing with increasing orders and number of branches of the approximating Erlang mixture. An exception is test case $EMP3$, where a minor increase in cdf area difference is reported between 3 and 4 Erlang branches.

5.2 MEDA

This section presents the experimental results obtained from approximating the parameters of phase-type distributions based on the test distributions $W1$-SE and empirical data sets $EMP1$-$EMP4$ (see chapter 4) using the moment-matching method MEDA (Schmickler [72, 73]). The following subsections 5.2.1 and 5.2.2 discuss the general attributes of MEDA and the numerical results obtained from our approximation runs.

The version of MEDA used in this analysis was written by L. Schmickler [72] in Turbo-Pascal and was compiled and run on a MSDOS PC with an 80486 CPU. There also exists a C version of MEDA (translated from Turbo-Pascal to C by the program *p2c*), but during the initial experiments with this program, it was found that the results were not always consistent with those obtained from the Turbo-Pascal version.

MEDA consists of the main executable program and several input and output files. Input files are *meda.in*, containing parameters which guide the program run and *medacdf.in*, containing the data to be approximated by a mixture of Erlang distributions in two columns: (i) the observations and (ii) the corresponding empirical cdf values. Moreover, *medacdf.in* contains the first three noncentral empirical moments of the data, which must be computed and supplied by the user. Output files are *meda.out*, containing the final results of a fitting run and *medalog.out*, a log file containing intermediate parameter values of the approximating Erlang mixture at all iterations of a particular fitting run.

Several parameters for a particular program run have to be specified by the user in the file *meda.in*: (i) the approximation method, either GURU (used in former versions of MEDA) or GAMMA-FPS (the *Flexible Polyhedron Search* (FPS) nonlinear programming algorithm by Nelder and Mead [50]), which was used for our experiments; (ii) the number r of Erlang branches in the mixture; (iii) the objective function criterion (choices are: the area difference between the empirical and the ap-

proximating phase-type cdf (used for our approximations), the sum of the squared area differences between observations, the sum of the differences in abscissas, the sum of the squared differences in abscissas and a chi-squared test criterion); (iv) a parameter determining whether the fourth branch (in parameter approximations involving four Erlang branches) is computed and optimized using Erlang or Gamma distributions; (v) the number of Erlang branches to be read from the file *meda.out* in the case where a previous solution is to be improved by adding another Erlang branch (necessary for approximations involving more than three Erlang branches); and (vi) the minimal height of a jump in the empirical cdf to be significant (set to 1.00 in our experiments, i.e. no jump is considered to be significant in our test cases).

After starting the program, MEDA echoes the user-specified input parameters and prints intermediate parameter values at every iteration on the screen and into the file *medalog.out.* After an approximation run has been successfully completed, MEDA prints the input parameters, the final distributional parameters (mixing probabilities p_i, inverse values λ_i^{-1} of the individual Erlang parameters, orders of the individual Erlang distributions k_i) and some auxiliary parameters (number of iterations, elapsed CPU time, final objective function value) in the file *meda.out.*

It should be pointed out that MEDA does not require the user to specify the orders of the individual Erlang branches or any other initial distributional parameter values. Regarding this fact, the MEDA algorithm is unique among all the parameter approximation methods considered in this study. As opposed to either having the user specifying initial parameter values or resorting to a random number generator, MEDA makes use of information contained in the lower range of the empirical cdf to find initial parameter values. A detailed discussion on how this is done can be found in Schmickler [72, 73].

5.2.1 MEDA General Results

- *Generality:* As with MEFIT (see section 5.1), MEDA approximations incorporate only *mixtures of Erlang* distributions as the selection subset of the class of phase-type distributions. No other phase-type structures can be used.

- *Reliability:* The algorithm contained in MEDA is generally reliable in that Erlang mixture approximations are obtained in most cases. Problems arose during initial experimentations with MEDA, where occasionally the program would terminate with a run time error. Some research on this matter revealed that the Turbo-Pascal compiler puts restrictions on some of MEDA's internal variables, which in turn lead to numerical instabilities (e.g. over- or underflows). A few changes in the code implemented by L. Schmickler significantly improved the reliability of MEDA. Another restriction imposed by the compiler is the maximum size of 64 KB of the data segment (arrays and vectors) within the program, which limits the maximum possible number of observations in the empirical data set to be approximated to 2300. This fact did not present a problem in our approximation runs. Although data set $EMP4$, for instance, consists of 3189 observations, many observations in this data set are of equal value. Deleting several repeated observations (while adjusting the corresponding cdf values) resulted in a smaller data set with 2038 observations.

- *Stability:* Some problems with MEDA arose during the part of the approximation run where the order of a particular Erlang branch is estimated. During the initial experimentation with MEDA, the program approximated very high orders, in some cases even getting into an infinite loop. The maximum order for any Erlang branch in our current version is thus limited to 100. It should be noted, however, that even in cases where an Erlang branch does achieve the maximum order, the improvement over intermediate fits (parameter values at intermediate iterations of the program run) involving lower order Erlang branches is usually minor.

- *Accuracy:* MEDA offers two different possibilities to approximate mixtures of Erlang distributions with three branches: (i) fitting three branches directly; and (ii) fitting two branches first and using this solution to obtain a three branch solution by adding the third branch. These two methods generally do not lead to the same Erlang mixture. It should also be noted that adding another Erlang branch to an existing solution does not always lead to an improvement of the fit; this is further discussed in the following section.

  During several fitting runs, it was noted that MEDA sometimes achieves intermediate parameter values which constitute a better fit than the final parameter values. This fact together with the above mentioned inconsistencies supports the conjecture that inaccuracies internal to the program sometimes lead away from a good approximation already obtained. In general, however, it must be noted that these inconsistencies and inaccuracies are often of minor relevance in terms of the resulting goodness-of fit, except in the cases where very high order approximations are obtained.

- *Efficiency:* The CPU requirements of MEDA are generally low, considering that our experiments were carried out on a PC with a 33 MHz clockspeed 80486 processor. Faster machines would speed up the run times considerably. As shown in table 5.4, the run times as well as the number of iterations are generally increasing with the number of Erlang branches. The run times are also increasing with increasing number of observations in the data sets, as shown in the empirical test cases $EMP1$-$EMP4$.

Table 5.4 presents the number of iterations, run times and orders of the individual Erlang distributions in the resulting mixtures obtained from approximation runs using MEDA.

Table 5.4. MEDA PH approximations: number of iterations, run times and orders of the approximating mixtures of Erlang distributions.

Test distribution		Mixture of Erlang distributions		
		2 branches	3 branches	4 branches
W1	# of iterations	14	219	477
	run time (sec.)	4.3	94.1	256.0
	orders	2, 5	1, 6, 2	1, 6, 2, 3
W2	# of iterations	8	205	597
	run time (sec.)	3.8	106.4	458.5
	orders	1, 1	1, 1, 1	1, 2, 1, 1
L1	# of iterations	8	154	609
	run time(sec.)	3.9	74.7	381.5
	orders	1, 1	1, 1, 1	1, 2, 1, 1
L2	# of iterations	12	402	685
	run time (sec.)	4.2	188.8	455.9
	orders	2, 3	3, 100, 2	3, 100, 2, 3
L3	# of iterations	398	362	868
	run time (sec.)	37.6	127.3	462.2
	orders	22, 30	1, 9, 40	1, 9, 40, 41
U1	# of iterations	57	343	549
	run time (sec.)	13.9	158.6	354.6
	orders	2, 25	3, 31, 2	3, 31, 1, 2
U2	# of iterations	116	361	687
	run time (sec.)	5.9	96.2	297.9
	orders	3, 27	1, 28, 100	1, 28, 100, 101
ME	# of iterations	17	175	368
	run time (sec.)	5.8	84.0	215.7
	orders	2, 8	1, 5, 6	1, 5, 5, 6
SE	# of iterations	10	198	457
	run time (sec.)	3.5	102.9	292.0
	orders	1, 4	2, 12, 2	2, 12, 1, 2
EMP1	# of iterations	11	215	499
	run time (sec.)	41.6	443.9	1233.7
	orders	2, 2	3, 7, 2	3, 7, 2, 3
EMP2	# of iterations	15	207	454
	run time (sec.)	12.3	161.0	444.0
	orders	2, 5	1, 7, 4	1, 7, 4, 5
EMP3	# of iterations	11	326	555
	run time (sec.)	18.4	391.5	793.5
	orders	2, 3	5, 21, 2	5, 21, 2, 3
EMP4	# of iterations	8	235	786
	run time (sec.)	38.8	460.5	1910.3
	orders	1, 2	1, 2, 1	2, 6, 2, 1

5.2.2 MEDA Experimental Results

Density and distribution plots: theoretical test cases

Figure 5.4 shows plots of the original theoretical densities ($W1$-SE) together with the approximating mixture of Erlang densities involving two, three and four Erlang branches. Figure 5.5 presents plots of the original theoretical distributions along with the approximating mixture of Erlang distributions involving two, three and four Erlang branches.

The visual inspection shows that an improvement of MEDA approximations involving three or four Erlang branches over two-branch approximations is not always given. This behavior is clearly shown in test cases $L2$ and $L3$. Furthermore, approximations to test case $L2$ show inconsistent density function behavior.

The multimodality of test distribution ME could not be recovered by MEDA approximations. The most obvious disagreements between the original densities and the Erlang mixture approximations are present in test cases $L2$, $L3$, $U1$, $U2$, ME and SE.

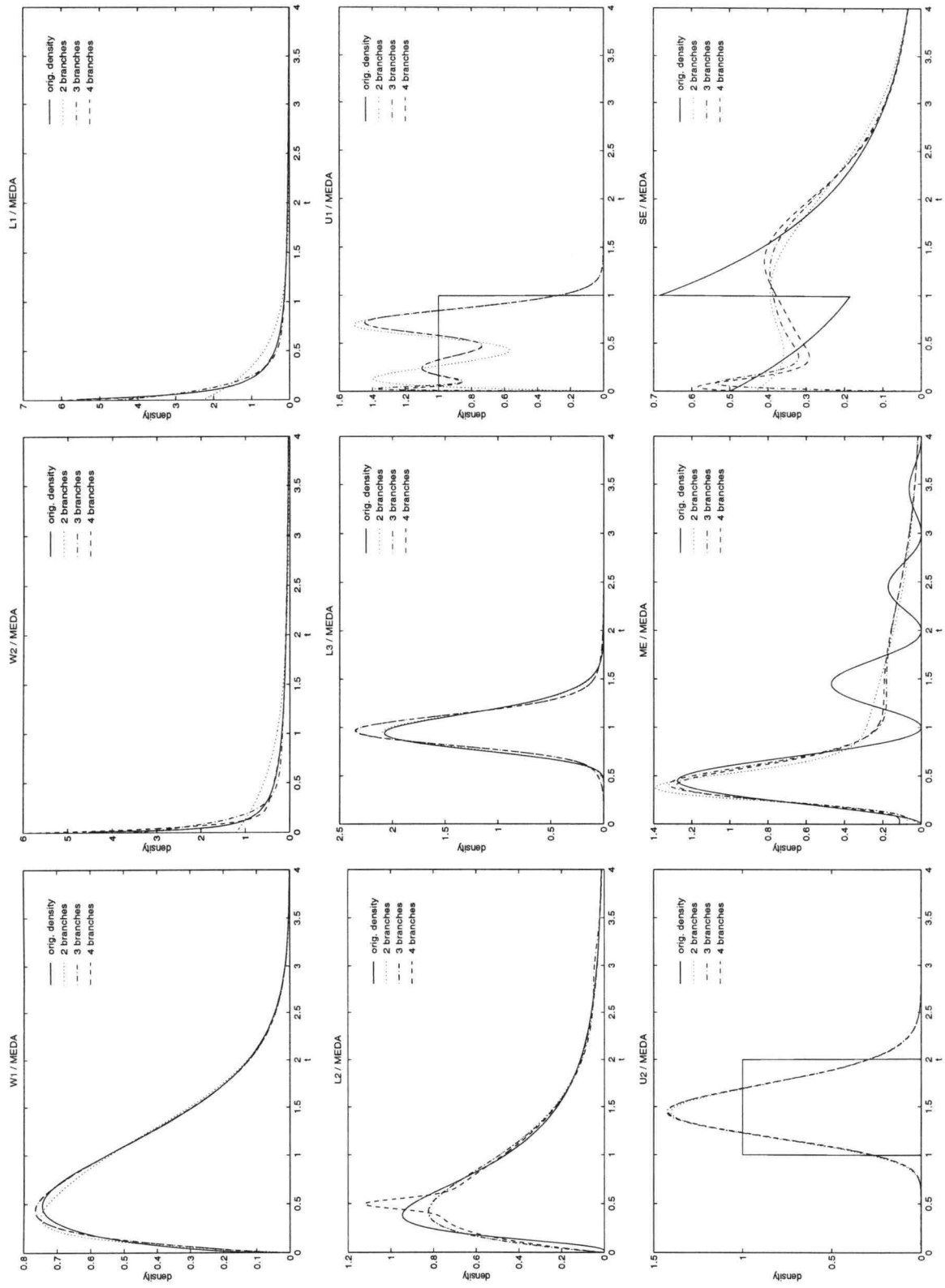

Figure 5.4. MEDA approximations to the theoretical densities.

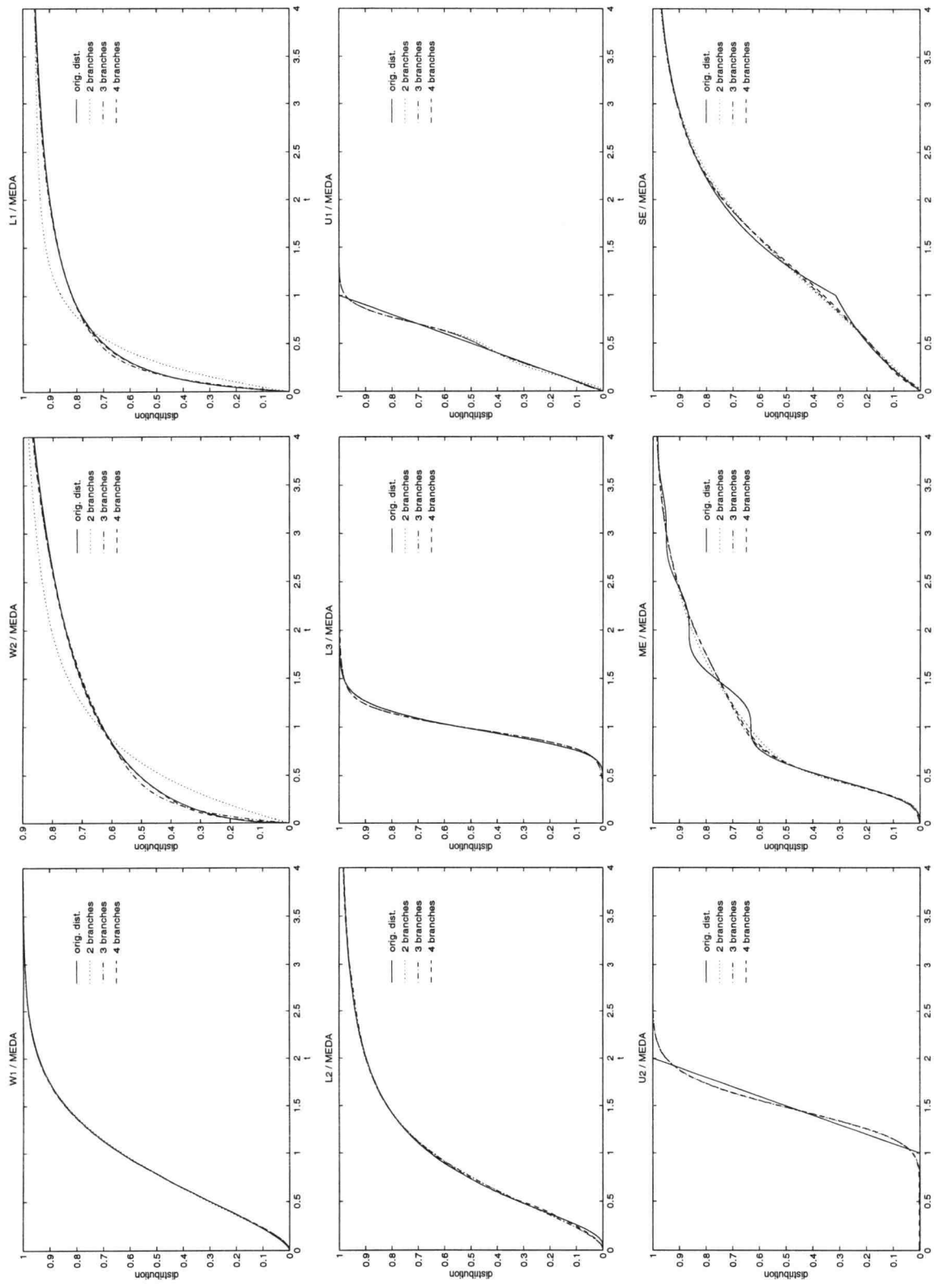

Figure 5.5. MEDA approximations to the theoretical distributions.

Performance measures: theoretical test cases

Table 5.5 presents the area differences as well as the entropy measures between the theoretical test densities and distributions ($W1$-SE) and the approximating mixture of Erlang densities and distributions involving two, three and four Erlang branches. Also presented is the order of the phase-type distribution corresponding to the mixture of Erlang distributions (see appendix B).

The pdf and cdf area differences as well as the entropy measures are generally decreasing with increasing order and number of branches of the approximating Erlang mixture. Inconsistencies of this observation are given in test cases $W1$, $L2$, $L3$, $U2$ and ME. Noteworthy are test cases $L2$ and $U2$, where additional Erlang branches of order 100 resulted in no improvement of the approximation. Another inconsistency is test case $L3$, where the approximations become worse with increasing number of Erlang branches.

MEDA determined high order approximations for test cases $L3$, $U1$ and $U2$, which is explained by their low variability and density function shapes. The high order approximations in test case $L2$ cannot be explained by properties of the original distribution, but are due to the MEDA algorithm, which increases the orders of the approximating Erlang mixture as long as the objective function criterion (cdf area difference) can be improved. Noteworthy in test case $L2$ is the fact that the third branch adds an extra 100 phases over the two-branch approximation, but in terms of the performance measures the fit becomes worse. Another example is the fit to data set $L3$, where the two-branch approximation represents a much better fit than the three-branch approximation.

Large pdf area differences resulted from the MEDA approximations to $W2$, $L1$, $U1$, $U2$, ME and SE. As mentioned previously in the discussion on MEFIT, these differences are explained by the challenging shapes of these distributions, and in cases $W2$ and $L1$ are explained by their long tail, which cannot be closely approximated by a phase-type distribution.

Table 5.5. MEDA PH approximations to the theoretical distributions: area differences and entropy measures.

Test distribution		Mixture of Erlang distributions		
		2 branches	3 branches	4 branches
W1	order of PH	7	9	12
	Δpdf	0.039652	0.022811	0.022750
	Δcdf	0.009218	0.006609	0.006732
	cross entropy	0.788988	0.788317	0.788297
	relative entropy	0.002113	0.001442	0.001422
W2	order of PH	2	3	5
	Δpdf	0.433804	0.215892	0.139228
	Δcdf	0.148736	0.049617	0.042059
	cross entropy	1.351265	1.201522	1.168134
	relative entropy	0.236178	0.086434	0.053046
L1	order of PH	2	3	5
	Δpdf	0.419565	0.137204	0.050589
	Δcdf	0.253667	0.119979	0.104857
	cross entropy	0.514546	0.408583	0.404330
	relative entropy	0.124824	0.018862	0.014608
L2	order of PH	5	105	108
	Δpdf	0.092715	0.096080	0.098034
	Δcdf	0.024391	0.024880	0.017140
	cross entropy	0.891104	0.891457	0.891098
	relative entropy	0.015310	0.015663	0.015304
L3	order of PH	52	50	91
	Δpdf	0.019711	0.143756	0.144622
	Δcdf	0.002146	0.017008	0.017213
	cross entropy	-0.210004	-0.189669	-0.189597
	relative entropy	0.000496	0.020830	0.020902
U1	order of PH	27	36	37
	Δpdf	0.336802	0.240572	0.230672
	Δcdf	0.052292	0.033993	0.033738
	cross entropy	0.112617	0.074914	0.069494
	relative entropy	0.112617	0.074914	0.069494
U2	order of PH	30	129	230
	Δpdf	0.401158	0.407668	0.410454
	Δcdf	0.031631	0.032900	0.034066
	cross entropy	0.178022	0.179994	0.182172
	relative entropy	0.178022	0.179994	0.182172
ME	order of PH	10	12	17
	Δpdf	0.382380	0.331878	0.335461
	Δcdf	0.054655	0.059120	0.057902
	cross entropy	0.882539	0.881090	0.879278
	relative entropy	0.154809	0.153360	0.151547
SE	order of PH	5	16	17
	Δpdf	0.204821	0.187586	0.178352
	Δcdf	0.037276	0.026806	0.024616
	cross entropy	1.331007	1.379746	1.324498
	relative entropy	0.036045	0.084784	0.029536

Density and distribution plots: empirical test cases

Figure 5.6 shows plots of the empirical densities and distributions ($EMP1$-$EMP4$) together with the approximating mixture of Erlang densities and distributions involving two, three and four Erlang branches.

The visual inspection of MEDA approximations to the empirical data sets reveals that all empirical densities and distributions are adequately fit. While the approximating densities show slight differences in their shape depending on the number of Erlang branches in the mixture, the corresponding distributions are virtually indistinguishable.

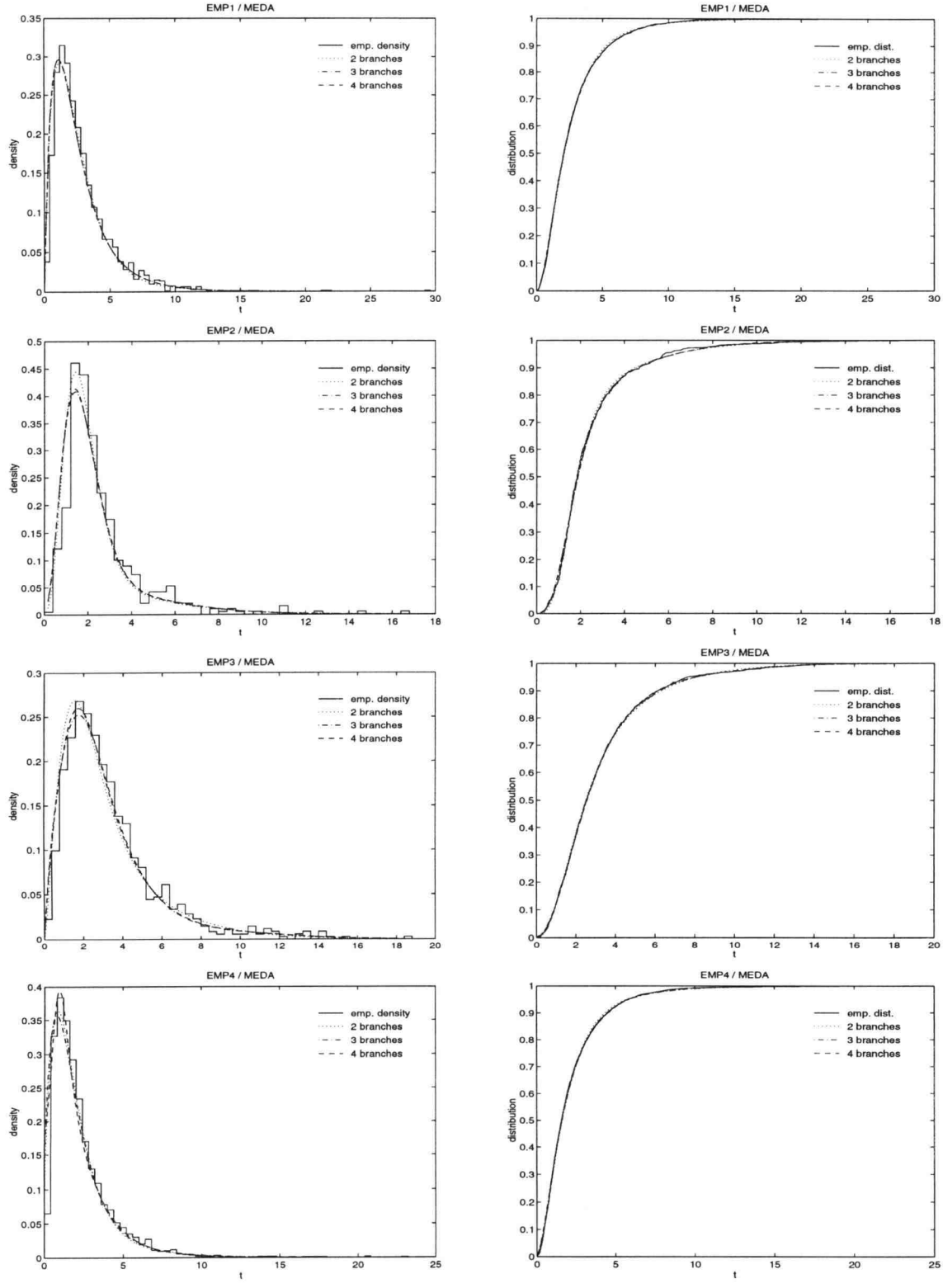

Figure 5.6. MEDA approximations to the empirical densities and distributions.

Table 5.6. MEDA PH approximations to the empirical data: area differences between the approximated and the approximating distributions and orders of the phase-type distributions corresponding to a mixture of Erlang distributions.

		Mixture of Erlang distributions		
Test distribution		2 branches	3 branches	4 branches
EMP1	order of PH	4	12	15
	Δcdf	0.028085	0.013471	0.013111
EMP2	order of PH	7	12	17
	Δcdf	0.031534	0.029557	0.028627
EMP3	order of PH	5	28	31
	Δcdf	0.023363	0.014002	0.012765
EMP4	order of PH	3	4	11
	Δcdf	0.029821	0.023471	0.013831

Performance measures: empirical test cases

Table 5.6 presents the cdf area differences between the empirical test distributions ($EMP1$-$EMP4$) and the approximating mixture of Erlang distributions involving two, three and four Erlang branches. Also presented is the order of the phase-type distribution corresponding to the mixture of Erlang distributions (see appendix B).

The cdf area differences are slightly decreasing with increasing order and number of branches of the approximating Erlang mixture. It should be pointed out that even approximations involving only two branches and low orders yield satisfactory results.

5.3 MLAPH

This section presents the experimental results obtained from approximating the parameters of acyclic phase-type distributions (APH) based on the test distributions $W1$-SE and empirical data sets $EMP1$-$EMP4$ (see chapter 4) using the maximum-likelihood based approach MLAPH (see Bobbio and Cumani [10, 11]). The following subsections 5.3.1 and 5.3.2 discuss the general attributes of MLAPH and the numerical results obtained from our approximation runs.

MLAPH is written in FORTRAN and was compiled with the *f77* SPARC FORTRAN compiler from SunPro, release 2.0.1. The program consists of the main executable, which handled in its original form all input and output via the standard input/output and internally generated all data to be approximated by the maximum-likelihood based algorithm, incorporating the known theoretical distributions mentioned in chapter 4. The MLAPH code was modified so that parameters guiding the approximation run and empirical data are now supplied by external files.

The following input parameters have to be specified by the user: (i) a known distribution together with its parameters (choices are: Weibull, lognormal, uniform, matrix-exponential, mixed-shifted exponential) or specification that empirical data are to be read in; (ii) the number of iterations to be performed; (iii) the order of the approximating APH distribution; (iv) the number of generated observations from the aforementioned known probability distributions (if selected); (v) the number of sets of initial APH parameter values to be randomly generated; and (vi) a seed for the random number generator.

For our experiments, we have generated 200 observations from each of the nine theoretical test distributions ($W1$-SE) and used the four empirical data sets ($EMP1$-$EMP4$), and therefore always specified external data to be used for the parameter approximation runs. The maximum number of iterations of the MLAPH algorithm was limited to 200. APH distributions have been approximated of order 2, order 4 and order 8. In its present form, MLAPH limits the maximum order

of the approximating APH distribution to 10 because of numerical instabilities and the low convergence rate of the algorithm. In order to find a good starting point for the approximation algorithm, a user-specified number of sets of initial APH parameter values ($\boldsymbol{\alpha}^{(0)}$ and $\boldsymbol{\lambda}^{(0)}$) is randomly generated based on a user-specified seed. The set of initial APH parameter values with the largest value of the log-likelihood function is selected as the starting point for the approximation algorithm. In our experiments, the number of initial APH parameter sets was set to 200.

Three different criteria are used to terminate the iterative approximation algorithm: (i) the maximum number of iterations is reached (200 in our experiments), (ii) the relative Euclidean norm (REN) between successive APH parameter estimates becomes smaller than $\epsilon = 10^{-6}$, or (iii) the REN$\times\alpha$ (denoting the angle between the search directions of the NLP problem at successive iterations of the algorithm) becomes smaller than $\epsilon = 10^{-6}$.

After starting the program, MLAPH echoes the user-specified input parameters and reports the initial set of APH parameter values determined by the program on the screen. After an approximation run has been successfully completed, MLAPH prints the final APH parameters ($\boldsymbol{\alpha}$ and $\boldsymbol{\lambda}$) as well as the first three central moments of the resulting APH distribution on the screen.

5.3.1 MLAPH General Results

- *Generality:* MLAPH uses only acyclic phase-type distributions (APH) as the selection subset of the class of phase-type distributions. Under the constraint that the parameters of the individual exponential phases are in increasing order, the APH class in general contains phase-type subclasses such as the (generalized) Erlang distribution, the hyperexponential distribution and the Coxian distribution (see formulæ (2.15, 2.16) and figure 2.6). The parameter fitting algorithm, however, uses a canonical form for APH distributions (see formulæ (2.13, 2.14) and figure 2.5) that does not include hyperexponential

or Coxian representations. Due to numerical instabilities in the fitting algorithm and due to the low convergence rate of finding the maximum of the log-likelihood surface, the algorithm does not allow for approximation of APH distributions of higher order than 10.

- *Reliability:* MLAPH performed generally reliable in most of our experiments. In some cases, however, the program terminated abnormally and generated NaN's instead of numerical values for the APH parameters. Using a different seed for the random number generator and therefore finding another starting point for the algorithm lead in all cases to solutions.

- *Stability:* The situation mentioned above supports the conjecture that the algorithm is not stable for all possible starting points of a given approximation problem. It is not clear how to find initial APH parameter values that lead to a solution except for trying different seeds for the random number generator.

  It should also be noted that using different seeds (i.e. different initial APH parameters) always lead to different final parameter values, which supports the hypothesis that local maxima are present in the log-likelihood surface. Different sets of phase-type parameters do not necessarily lead to different phase-type distributions (non-uniqueness property of phase-type distributions), but in the cases where we observed different final parameter values based on different seeds, we always found the shape of the corresponding APH distributions to vary significantly.

- *Accuracy:* A general observation based on our experiments is that, as the order of the approximating APH distribution increases, the algorithm usually terminates because the maximum number of iterations (200 in all our cases) is reached. In order to be able to compare the results from different fitting runs, we needed to define common termination criteria. Fitting runs with a larger maximum number of iterations, however, might improve the results we obtained in our experiments.

- *Efficiency:* MLAPH usage of CPU time is generally moderate and increasing as the number of phases of the approximating APH distribution increases. As shown in table 5.7, the run times were in the range of 3-11 seconds for order-2 APH, in the range of 7-35 seconds for order-4 APH and in the range of 65-165 seconds for order-8 APH distributions. Furthermore, the number of observations has a significant influence on the CPU time requirements, as shown in table 5.8. The run times for fitting empirical data set $EMP1$ with 2039 observations ranged from 14 seconds (order-2 APH) to 674 seconds (order-8 APH), the run times for fitting empirical data set $EMP4$ with 3189 observations were between 21 seconds (order-2 APH) to 1003 seconds (order-8 APH).

Tables 5.7 and 5.8 present the termination criteria (number of iterations, relative Euclidean norm or relative Euclidean norm $\times$ the angle between successive search directions α, indicated by an asterisk), the final log-likelihood after completion of the approximation run and run times of the APH parameter approximations obtained from MLAPH.

Table 5.7. MLAPH APH approximations to the theoretical distributions: general results.

Test distribution		Acyclic Phase-Type distribution order 2	order 4	order 8
W1	# of iterations	101	200*	200*
	REN	5.64585e-06	0.00101637	0.00529277
	REN×α	0*	0.00101637	0.00529277
	log-likelihood	-158.262	-157.447	-157.427
	run time (sec.)	10.2	29	106.2
W2	# of iterations	135	200*	200*
	REN	8.80129e-07*	0.0142989	0.00364286
	REN×α	8.80129e-07	6.43503e-03	0.00036429
	log-likelihood	-252.969	-235.058	-236.173
	run time (sec.)	10.5	26.5	70.9
L1	# of iterations	94	200*	200*
	REN	8.4538e-07*	0.00228365	0.00223394
	REN×α	8.4538e-08	0.00228365	0.00223394
	log-likelihood	-85.3005	-77.3606	-76.4461
	run time (sec.)	8.3	26.4	155.2
L2	# of iterations	50	200*	200*
	REN	1.14509e-06	0.00579694	0.00166704
	REN×α	8.04856e-07*	0.00025886	0.00166704
	log-likelihood	-180.922	-175.712	-175.884
	run time (sec.)	6.8	26.5	74.3
L3	# of iterations	7	26	98
	REN	0.00476907	0.0040067	0.00214247
	REN×α	8.98994e-08*	8.82047e-07*	2.76481e-07*
	log-likelihood	-126.741	-61.3164	-27.387
	run time (sec.)	4.9	8.7	65.2
U1	# of iterations	34	127	154
	REN	2.12587e-06	5.66857e-06	0.000580294
	REN×α	8.81751e-07*	4.94923*	0*
	log-likelihood	-40.3394	-27.7809	-20.3886
	run time (sec.)	6.8	34.9	154.8
U2	# of iterations	7	30	100
	REN	0.00465074	0.00308789	0.000181472
	REN×α	9.06039e-08*	5.28786e-07*	2.52547e-09*
	log-likelihood	-207.124	-141.493	-107.279
	run time (sec.)	2.8	6.6	65.9
ME	# of iterations	60	200*	200*
	REN	1.95084e-06	0.0139045	0.000211028
	REN×α	9.72096e-07*	0.0123811	1.34491e-04
	log-likelihood	-193.114	-190.575	-177.038
	run time (sec.)	6.7	28.5	164.4
SE	# of iterations	51	200*	200*
	REN	2.12301e-06	0.00407422	0.00242575
	REN×α	8.65948e-09*	0.00271884	0.00024258
	log-likelihood	-270.449	-265.165	-263.54
	run time (sec.)	6.4	26.3	159.6

Table 5.8. MLAPH APH approximations to the empirical distributions: general results.

Test distribution		Acyclic Phase-Type distribution		
		order 2	order 4	order 8
EMP1	# of iterations	2	200*	200*
	REN	0.173641	0.00184473	0.000437371
	REN$\times\alpha$	0*	0.00184473	0.000437371
	run time (sec.)	13.6	225.8	673.4
EMP2	# of iterations	32	200*	200*
	REN	1.69618e-06	2.85536e-05	0.0061579
	REN$\times\alpha$	7.91558e-07*	2.85536e-05	0.00519881
	run time (sec.)	6.4	46.6	187.8
EMP3	# of iterations	200*	200*	200*
	REN	4.05842e-05	0.149802	0.00343751
	REN$\times\alpha$	3.07139e-05	0.1320729	0.00343751
	run time (sec.)	40.9	94.5	298.8
EMP4	# of iterations	2	200*	200*
	REN	0.17582	0.00350216	0.00320486
	REN$\times\alpha$	0*	0.00350216	0.00320486
	run time (sec.)	21.0	350.9	1003.4

5.3.2 MLAPH Experimental Results

Density and distribution plots: theoretical test cases

Figure 5.7 shows plots of the original theoretical densities ($W1$-SE) together with the approximating APH densities of order 2, 4 and 8. Figure 5.8 presents plots of the original theoretical distributions along with the approximating APH distributions of order 2, 4 and 8.

The plots suggest that, while order-2 APH approximations are generally not satisfactory, adequate APH approximations of order 4 and order 8 are found in test cases $W1$, $W2$, $L1$ and $L2$. Visual disagreements between the original distributions and the approximating APH distributions are evident in test cases $L3$, $U1$, $U2$, ME and SE. The APH approximations are generally improving with increasing order.

The inability of APH approximations to match the shape of test distribution $L3$ is based on its low variability (squared coefficient of variation = 0.041). The minimal order needed to achieve the variability of test distribution $L3$ is therefore 25. Our APH approximations of order 2, 4 and 8 are inadequate in this case.

From the APH approximation plots to test case ME, we note that the multimodality of the original density is not matched by the approximating APH densities. In test case SE, the bimodality appears for order-4 and order-8 APH approximations.

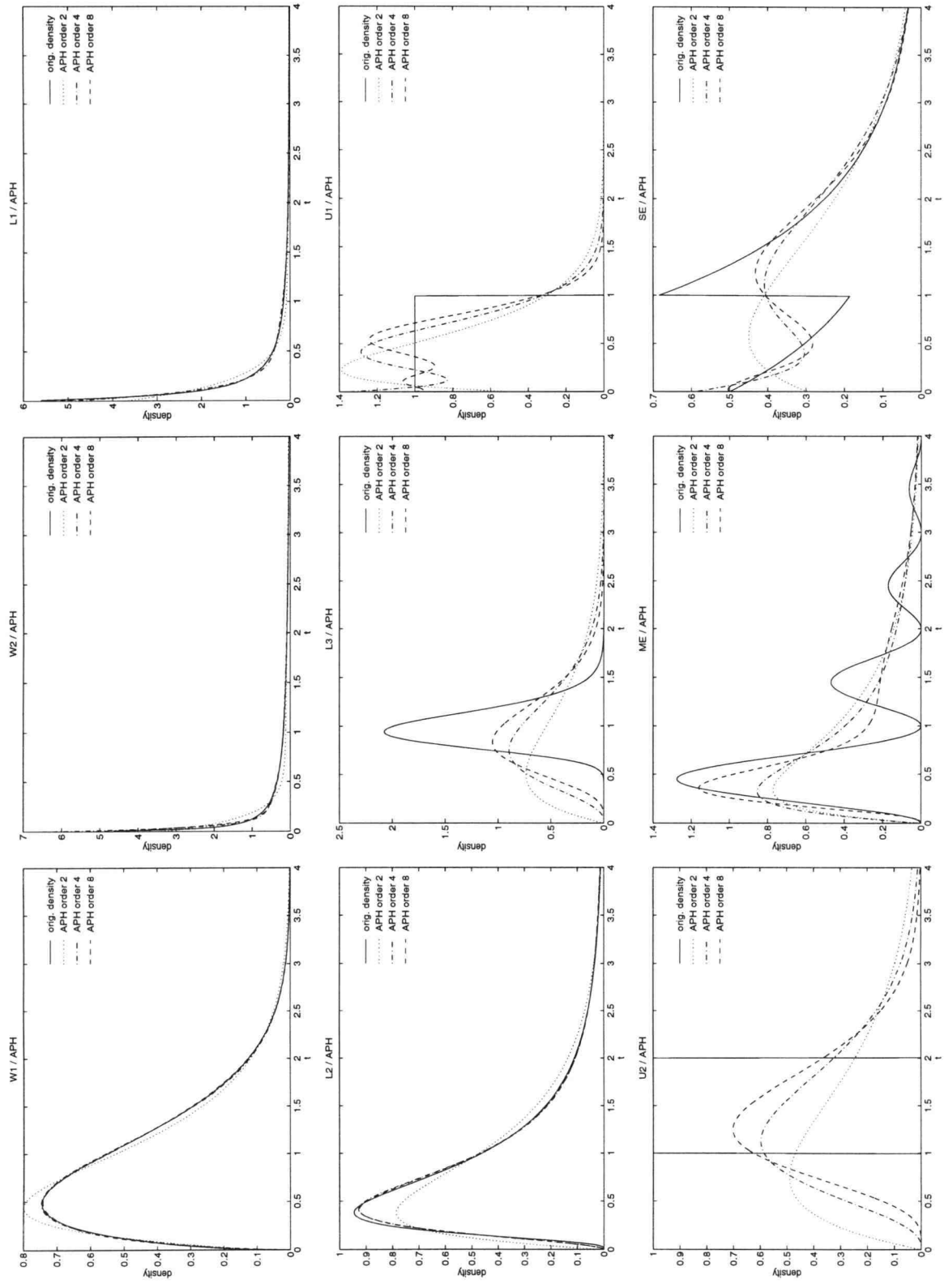

Figure 5.7. MLAPH approximations to the theoretical densities.

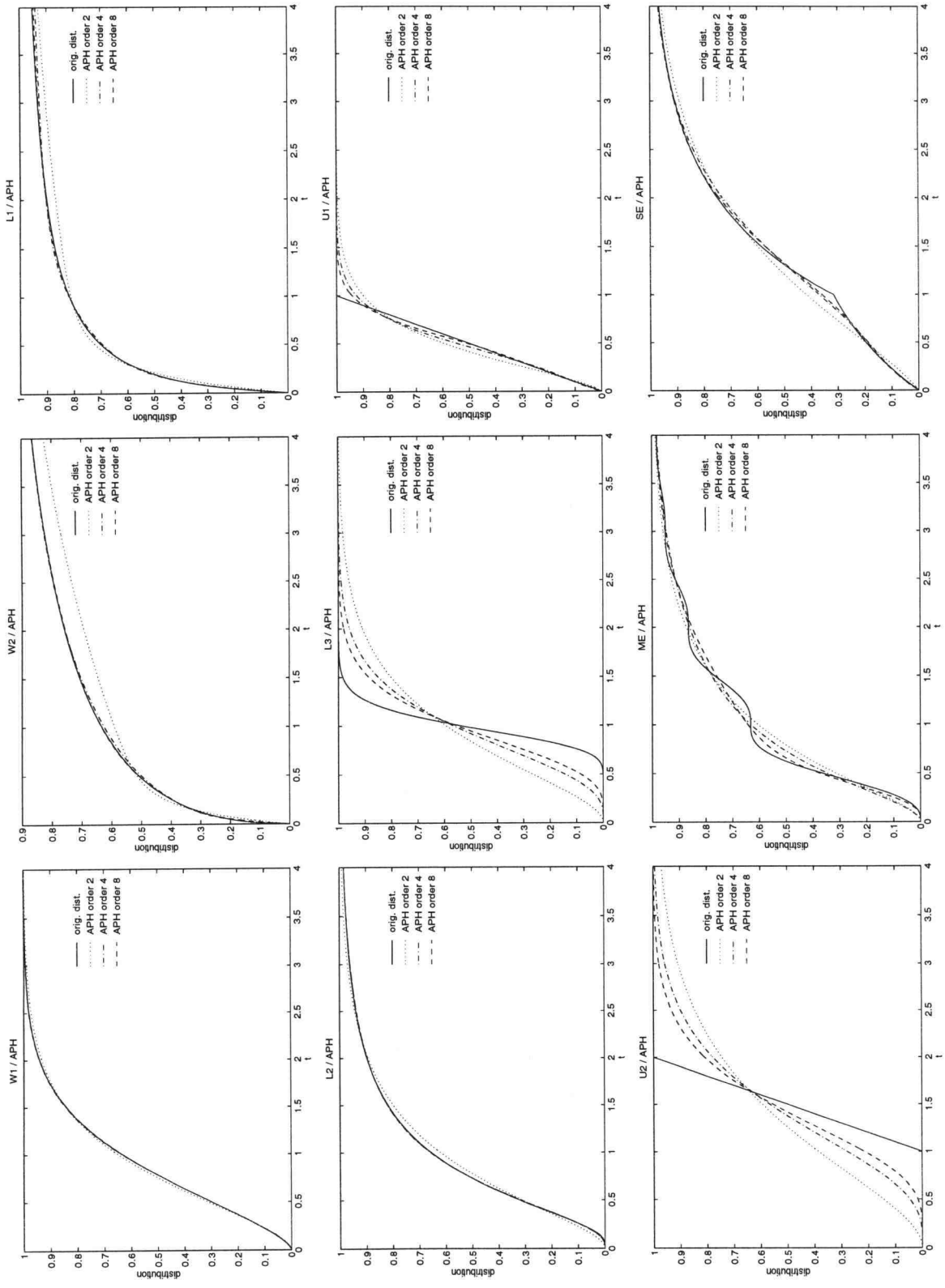

Figure 5.8. MLAPH approximations to the theoretical distributions.

Performance measures: theoretical test cases

Table 5.9 presents the area differences as well as the entropy measures between the theoretical test densities and distributions ($W1$-SE) and the approximating APH densities and distributions of order 2, 4 and 8, respectively.

The APH approximations to the theoretical densities and distributions are generally improving with increasing order, based on the pdf and cdf area differences and the entropy measures. Slight inconsistencies are present in test cases $W2$ and $L2$.

The difficulty of APH distributions to match distributions with long tails ($W2$ and $L1$) is evident in the performance measures in these test cases. Although the visual approximation seems adequate for order-4 and order-8 APH distributions, the performance measures present a more accurate understanding of the actual goodness-of-fit.

Table 5.10 presents the relative errors between the first three central moments of the original theoretical distributions and the approximating APH distributions of order 2, 4 and 8, respectively. The relative error on the third moment is undefined for test distributions $U1$ and $U2$ since their central third moment μ_3 equals zero (see equation 4.22). The performance measure reported (and marked with an asterisk) in these two cases is the central third moment $\hat{\mu}_3$ of the approximating APH distribution.

We note that the APH fits generally match the first moments of the test distributions well. The second and third moments, however, are not adequately matched in most test cases.

The expectation that the quality of APH approximations improves with increasing orders is generally not met in terms of the relative moment errors. The relative error on the first moment is inconsistent or increasing with increasing APH orders in test cases $W2$, $L1$, $L2$, $L3$, $U1$ and ME. The relative error on the second moment is inconsistent with increasing order in test cases $L1$, $L2$ and ME. Inconsistencies in the relative error on the third moment are present in test cases $L1$ and $L2$.

The large relative errors on the second and third moments in test cases $W2$ and $L1$ are again based on the inability of the APH distribution to match their long tail. The large second and third moment errors in test cases $L3$ and $U2$ are explained by the inability of the approximating APH distributions to match their low squared coefficient of variation.

In some test cases the relative moment errors are higher for visually good approximations than for fits with large visual discrepancies. This is particularly noticeable when comparing the APH fits of test cases $W2$ or $L1$ to the fits of test cases ME or SE.

Table 5.9. MLAPH APH approximations to the theoretical distributions: area differences and entropy measures.

Test distribution		Acyclic Phase-Type distribution		
		order 2	order 4	order 8
W1	Δpdf	0.065167	0.010957	0.008084
	Δcdf	0.032067	0.004244	0.001875
	cross entropy	0.791275	0.787202	0.787085
	relative entropy	0.004400	0.000327	0.000210
W2	Δpdf	0.342083	0.130881	0.125713
	Δcdf	0.251956	0.057097	0.033007
	cross entropy	1.256353	1.162110	1.169144
	relative entropy	0.141266	0.047022	0.054056
L1	Δpdf	0.258706	0.118305	0.056054
	Δcdf	0.303565	0.182292	0.158850
	cross entropy	0.460052	0.414839	0.412513
	relative entropy	0.07033	0.025118	0.022791
L2	Δpdf	0.146074	0.033419	0.031227
	Δcdf	0.073554	0.013513	0.014766
	cross entropy	0.904610	0.878561	0.879422
	relative entropy	0.028816	0.002767	0.003628
L3	Δpdf	1.112908	0.848279	0.676640
	Δcdf	0.384072	0.231995	0.162118
	cross entropy	0.633706	0.306582	0.136935
	relative entropy	0.844205	0.517081	0.347434
U1	Δpdf	0.422574	0.314775	0.234067
	Δcdf	0.167583	0.089365	0.051738
	cross entropy	0.201676	0.138886	0.101952
	relative entropy	0.201676	0.138886	0.101952
U2	Δpdf	1.278881	0.999510	0.812080
	Δcdf	0.383727	0.230028	0.159766
	cross entropy	1.038318	0.709540	0.538237
	relative entropy	1.038318	0.709540	0.538237
ME	Δpdf	0.544412	0.525272	0.390993
	Δcdf	0.099559	0.080372	0.058505
	cross entropy	0.965572	0.952876	0.885189
	relative entropy	0.237842	0.225146	0.157458
SE	Δpdf	0.282667	0.185023	0.140752
	Δcdf	0.065891	0.032966	0.016235
	cross entropy	1.352755	1.326338	1.318209
	relative entropy	0.057793	0.031376	0.023247

Table 5.10. MLAPH APH approximations to the theoretical distributions: relative moment errors.

		Acyclic Phase-Type distribution		
Test distribution		order 2	order 4	order 8
W1	e_1	0.0006464	0.0005765	0.0001740
	e_2	0.1285957	0.0060030	0.0052559
	e_3	0.5837212	0.0707516	0.0180982
W2	e_1	0.0129671	0.0020182	0.0119791
	e_2	0.5254806	0.1886697	0.1606635
	e_3	0.8708110	0.5344133	0.4665294
L1	e_1	0.0661893	0.0709493	0.1374785
	e_2	0.8371037	0.7708373	0.8199352
	e_3	0.9980452	0.9955274	0.9969084
L2	e_1	0.0038253	0.0048197	0.0074579
	e_2	0.2615398	0.0936450	0.1227760
	e_3	0.6838972	0.3690537	0.4344384
L3	e_1	0.0001350	0.0001325	0.0001333
	e_2	11.248359	5.1242103	3.0828001
	e_3	97.685891	23.671658	9.9651538
U1	e_1	5.49724e-7	0.0001106	0.0024947
	e_2	0.7034906	0.2643689	0.1161154
	e_3*	0.0765430	0.0275195	0.0120981
U2	e_1	0.0000025	1.25e-6	0.0000000
	e_2	12.500068	5.7499831	3.5000000
	e_3*	1.6875127	0.4218734	0.1875000
ME	e_1	0.0020990	0.0026553	0.0026347
	e_2	0.1519981	0.0042582	0.0083248
	e_3	0.3022478	0.0789958	0.0673206
SE	e_1	0.0011538	0.0001872	0.0008039
	e_2	0.1399388	0.0316543	0.0011091
	e_3	0.2529391	0.2108032	0.0327994

Density and distribution plots: empirical test cases

Figure 5.9 shows plots of the empirical test densities and distributions ($EMP1$-$EMP4$) together with the approximating APH densities and distributions of order 2, 4 and 8, respectively.

We note that while order-2 APH approximations generally do not satisfactorily match the empirical test densities and distributions, order-4 and order-8 APH approximations present visually adequate fits. Noticeable disagreements between the empirical density and distribution and the APH approximations are evident in test case $EMP2$.

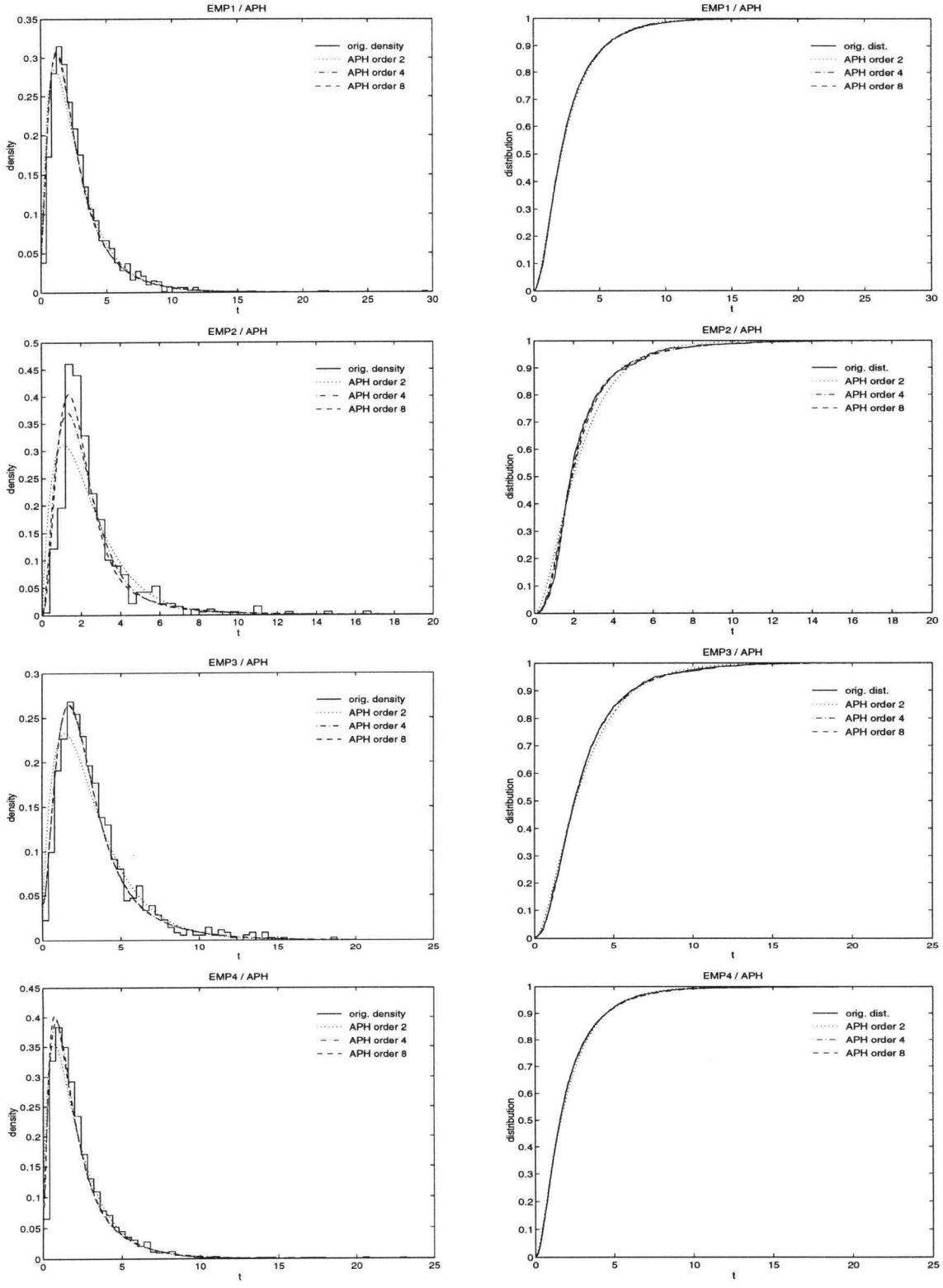

Figure 5.9. MLAPH approximations to the empirical densities and distributions.

Performance measures: empirical test cases

Table 5.11 presents the cdf area differences between the empirical test distributions ($EMP1$-$EMP4$) and the approximating APH distributions of order 2, 4 and 8. Also presented are the relative moment errors between the distributions.

The visual observation that the APH fits to test case $EMP2$ are not as good as in the other three test cases is verified by the cdf area difference measure.

A surprising observation is that, while the first empirical central moments are matched quite well, the relative errors on the first moment are increasing with increasing order of the approximation APH distribution in test cases $EMP1$, $EMP2$ and $EMP4$. Inconsistencies in the relative error on the second moment are reported for test cases $EMP3$ and $EMP4$, and only test case $EMP4$ shows inconsistencies in the relative error on the third moment.

Table 5.11. MLAPH APH approximations to the empirical distributions: cdf area differences and relative moment errors.

		Acyclic Phase-Type distribution		
Test distribution		order 2	order 4	order 8
EMP1	Δcdf	0.0341720	0.0130350	0.0116340
	e_1	0.0000002	0.0002550	0.0005070
	e_2	0.1159429	0.0475135	0.0240227
	e_3	0.4564210	0.2780476	0.1442145
EMP2	Δcdf	0.1097040	0.0433670	0.0316260
	e_1	0.0000005	0.0001580	0.0070688
	e_2	0.1425094	0.0400540	0.0218277
	e_3	0.5562170	0.1190947	0.0518619
EMP3	Δcdf	0.0502450	0.0155480	0.0152830
	e_1	0.0000016	0.0018729	0.0001485
	e_2	0.0489114	0.0175157	0.0192200
	e_3	0.2438591	0.1638906	0.1343859
EMP4	Δcdf	0.0450390	0.0204420	0.0176860
	e_1	0.0000005	0.0005963	0.0106432
	e_2	0.1565780	0.0532796	0.1309954
	e_3	0.5073439	0.2564726	0.4168254

5.4 EMPHT

This section contains the experimental results of approximating the parameters of phase-type distributions based on the test distributions $W1$-SE and empirical data sets $EMP1$-$EMP4$ (see chapter 4) using the maximum-likelihood based method EMPHT (Asmussen et al. [5, 6], see also [26, 64]). The following subsections 5.4.1 and 5.4.2 describe the general attributes of EMPHT and the numerical results of our experiments.

EMPHT consists of four programs: EMPHTenter, EMPHTdensity, EMPHT-main and EMPHTgraphics. The first three programs are written in C and were compiled with the *cc* C compiler included within the SunOS 4.1.3 installation on a SPARC-IPX workstation. EMPHTgraphics is a MATLAB [49] program used for plotting purposes.

EMPHTenter specifies the data (*sample* or *density*) to be fit by a phase-type distribution. If *density* is selected, there is a choice of 6 predefined densities: uniform, normal, lognormal, Weibull, inverse Gaussian and phase-type. The exponential and (generalized) Erlang distributions are contained within the class of phase-type distributions. The user must write the specification for all other density functions in C code and implement it within the program EMPHTdensity, which works similar to EMPHTenter, described below.

EMPHTenter converts the specified density into a weighted sample of n observations $\boldsymbol{x} = x_1, \ldots, x_n$ and n weights $\boldsymbol{w} = w_1, \ldots, w_n$. The weights represent the probability mass of each observation. The number of observations n is determined by a user-specified point where the density can be truncated, the maximum acceptable probability (weight) in one point and the maximum time interval corresponding to one point. The interval between zero and the truncation point is then discretized into intervals $0, x'_1, \ldots, x'_n$ and the weights w_i are calculated according to Simpson's approximation formula of $\int_{x'_{i-1}}^{x'_i} g(u)du$, where $g(u)$ is the density selected, i.e.

$$w_i = \left(g(x'_{i-1}) + 4\, g(x'_{i-1} + \frac{x'_i - x'_{i-1}}{2}) + g(x'_i)\right)(x'_i - x'_{i-1}) \,/\, 6. \qquad (5.1)$$

If the weight becomes larger than the maximum specified by the user, the interval $[x'_{i-1}, x'_i]$ is shortened by choosing a smaller value for x'_i and a new weight is calculated. The corresponding observation x_i is approximated as being the center of mass of the interval $[x'_{i-1}, x'_i]$ by

$$x_i = \frac{x'_{i-1}\, g(x'_{i-1}) + 4\,(x'_{i-1} + \frac{x'_i - x'_{i-1}}{2})\, g(x'_{i-1} + \frac{x'_i - x'_{i-1}}{2}) + x'_i\, g(x'_i)}{g(x'_{i-1}) + 4\, g(x'_{i-1} + \frac{x'_i - x'_{i-1}}{2}) + g(x'_i)}. \tag{5.2}$$

The total number n of observations is successively adjusted to obtain x'_n larger than or equal to the truncation point.

In the case where *sample* is selected, there is a choice between an unweighted and a weighted sample, which are either read as the vector of observations $\boldsymbol{x}$ by EMPHTenter from the file *unweighted* or two vectors (observations $\boldsymbol{x}$ and weights $\boldsymbol{w}$) from file *observations*. If the sample is unweighted, EMPHTenter converts it into a weighted sample by adding on the weight vector $\boldsymbol{w}$ with all weights equal to one, and stores it in the file *observations*.

An output file of EMPHTenter (or EMPHTdensity) is the file *exactdistr*, which contains the cdf, pdf and intensity values along with corresponding x-values of the distribution to be fit in 4 columns, each of length 400. These values are used by EMPHTgraphics for plotting purposes.

The main program which contains the EM algorithm is EMPHTmain. Input is the file *observations* (containing vectors $\boldsymbol{x}$ and $\boldsymbol{w}$), which is created by EMPHTenter or EMPHTdensity. The user specifies the structure of the phase-type distribution to be fit, using either the predefined choices (see section 5.4.1 for all possibilities), or supplying a file called *distrtype*, which contains a user-defined phase-type structure. In addition, the number of EM iterations, the step-length for the Runge-Kutta procedure (either default or user-specified) and a seed for the random number generator needed to determine the initial phase-type parameter values $\boldsymbol{\alpha}^{(0)}$ and $\boldsymbol{T}^{(0)}$ must be specified. Output is a file called *phases*, which contains the final parameter values for the initial probability vector $\boldsymbol{\alpha}$ and the subintensity $\boldsymbol{T}$. This output file *phases* can be used as input (instead of using one of the predefined phase-type structures or

the user-specified phase-type structure in file *distrtype*) determining the phase-type structure and initial parameter values for additional iterations of the algorithm.

The second output file of EMPHTmain is the file *approxdistr*. Its structure is equivalent to the file *exactdistr*, containing x-values and corresponding pdf, cdf and intensity values for the approximating phase-type distribution.

The last part of EMPHT is a MATLAB [49] program called EMPHTgraphics. Running it under MATLAB, it reads as input the files *exactdistr* and *approxdistr* and generates plots of the approximated and approximating cdf, pdf and intensity.

For our experiments, we have externally generated 200 observations from each of the nine test distributions ($W1 - SE$) and used the four empirical data sets ($EMP1 - EMP4$) for phase-type approximations. These data were stored in the file *unweighted* before each program run and converted to a weighted sample with all weights equal to one by EMPHTenter. The option of using the internal data generator (in EMPHTenter) based on the test distributions was therefore not used. This procedure provided the possibility to use the same data sets for all parameter approximation methods discussed in this analysis.

The following phase-type structures were used for the parameter approximation runs (see section 5.4.1 for all possible phase-type structures in EMPHT): (i) sum of exponentials (generalized Erlang); (ii) acyclic phase-type; and (iii) general phase-type. The sum of exponentials structure was chosen because it was intended to provide the closest possible comparisons to the mixture of Erlang distributions fit by the moment-matching methods MEFIT and MEDA (see results in sections 5.1 and 5.2). It also contains the smallest number of parameters among the selected subsets of phase-type distributions in this analysis; k parameters for an order-k sum of exponentials distribution. The acyclic phase-type structure was selected in order to provide the closest possible comparisons to the acyclic phase-type (APH) distributions fit by method MLAPH (see results in section 5.3). The difference, however, between the APH structure fit by MLAPH and the acyclic phase-type structure used in EMPHT is that the parameters of the individual exponential phases are not

restricted to be in increasing order. EMPHT does not allow for restrictions on the parameters to be approximated. The general phase-type structure was intended to provide the highest versatility among all phase-type approximations in this analysis. While subclasses of the family of phase-type distributions impose restrictions upon the possible transitions in their underlying Markov process, the general phase-type distribution does not imply any such restrictions and provides therefore the highest flexibility. The number of parameters to be approximated for a general phase-type distribution, however, is the largest among all the phase-type structures considered in this study, namely $k^2 + k - 1$ for an order-k general phase-type distribution.

For the three phase-type structures used in this analysis, we fit orders 2, 4, 8 and 16. Higher orders were not fit, partially because the number of parameters to be approximated becomes very large and the run times prohibitive. EMPHT restricts the maximum number of phases to 50 in its current version.

Three different criteria were adopted to terminate the iterative EM algorithm: (i) the maximum number of iterations is reached (in our experiments the limit was set to 200); (ii) the relative Euclidean norm (REN) between successive parameter estimates becomes smaller than a user-specified level (in our experiments $\epsilon = 10^{-4}$); or (iii) the relative change in log-likelihood between successive iterations becomes smaller than a user-specified level (in our experiments $\epsilon = 10^{-6}$). The last two termination criteria were not included in the original program.

After starting the program, EMPHTmain echoes the user-specified input parameters and prints the approximated phase-type parameters at every iteration of the EM algorithm on the screen. A successful approximation run is completed with printing the final phase-type parameters in the file *phases*.

5.4.1 EMPHT General Results

- *Generality:* EMPHT allows a large variety of subsets of phase-type distributions to be fit to empirical data. Five structures of the initial probability vector $\boldsymbol{\alpha}$ and the subintensity $\boldsymbol{T}$ are predefined:

  1. *General phase-type:* all elements of $\boldsymbol{\alpha}$ and $\boldsymbol{T}$ are allowed to be non-zero. The underlying Markov process may start at any transient state, and movements from any transient state to any other transient state are allowed in one transition, until the absorbing state is reached.

  2. *Hyperexponential:* only the main diagonal elements of $\boldsymbol{T}$ and all elements of $\boldsymbol{\alpha}$ are allowed to be non-zero. The underlying Markov process may start at any transient state. After an exponential sojourn time, the Markov process jumps immediately to the absorbing state without visiting any other transient state (see figure 2.4).

  3. *Sum of exponentials:* always starts at the first transient state. Transitions are restricted to jumps from state i to state $i + 1$ only (see figure 2.3). Note that this form of a phase-type distribution is also called a *generalized Erlang distribution.*

  4. *Coxian distribution:* as with a generalized Erlang distribution, starts always at the first transient state, but termination from any state is allowed (see figure 2.7).

  5. *Generalized Coxian:* as the Coxian distribution described above, but start from any transient state is allowed.

  Other structures of phase-type distributions may be specified by the user in the file *distrtype.* It should be noted, however, that only the location of zero and non-zero elements in $\boldsymbol{\alpha}$ and $\boldsymbol{T}$ may be specified. There is no possibility of specifying any restricting structure on the non-zero elements. Thus it is not possible to fit, for instance, a regular Erlang distribution, for which the

parameters λ of each transient state of the underlying Markov process must be equal. We used the possibility of defining a special phase-type structure in our experiments by incorporating the acyclic phase-type distribution mentioned in the previous section.

- *Reliability:* EMPHT obtains in most cases phase-type fits for the test distributions. Problems were encountered while fitting sum of exponential (generalized Erlang) distributions. In several cases it happened that the Runge-Kutta procedure, implemented to solve the vector-differential equations at each E-step of the EM algorithm, generated results that were not numbers (NaN) at the first iteration, which indicated a division by zero. An investigation showed that the step-size for the Runge-Kutta procedure – which is by default set to $0.1/|t_{max}|$, t_{max} being the largest diagonal element of $\boldsymbol{T}$ – sometimes becomes too large, which in turn sets some of the vector-differential equations equal to zero. Another version of EMPHT (received from M. Olsson [64]), which allows the user to specify a fixed step-size, resolves most of the problems in these circumstances. However, by choosing a step-size small enough for the Runge-Kutta algorithm to work, the run times might become prohibitively large. We had to terminate some program runs due to this situation after obtaining no EM iterations with run times of up to several weeks.

  A minor inconvenience with EMPHT is that observations with a value of zero are not allowed. The program terminates abnormally in such a situation. A small amount (10^{-7} in our experiments) must be added to zero-valued observations in order for the program to run.

- *Stability:* In order to obtain initial phase-type parameter values $\boldsymbol{\alpha}^{(0)}$ and $\boldsymbol{T}^{(0)}$, the program generates random values based on a user-specified seed for the random number generator. These initial parameter values are important, since the EM algorithm may be at a saddlepoint of the likelihood surface, or it may converge to a local instead of a global optimum. Trial runs with different seeds for the random number generator resulted in different solutions for the final

parameters as well as the final phase-type distribution, but there is limited experience yet as to how many different program runs are required to obtain a reasonably good solution. Because of prohibitively long run times of the EM algorithm, we chose to run only one parameter approximation per test case.

- *Accuracy:* It is shown in the following tables 5.12-5.17 that the termination criterion in general is the maximum number of iterations (200 in our experiments), particularly for parameter approximations of orders larger than two. It is also shown in section 5.4.2 that results involving higher order phase-type distributions often do not improve the fit. This supports the conjecture that, in most cases, more than 200 iterations are needed to obtain reasonable phase-type approximations with EMPHT. Asmussen, Nerman and Olsson [6] report 2000 - 10000 iterations necessary for their results.

- *Efficiency:* The amount of CPU time required to obtain reasonable fits of phase-type distributions of moderate order (for instance order 4 or 8) is generally very large. As reported in the following tables, run times for fits using sum of exponentials range from 2 to 132000 seconds for the test cases where results could be obtained. Several fitting runs for the sum of exponentials structure did not lead to solutions in any reasonable run time. For the acyclic phase-type structure, run times range from 3 to 240000 seconds. Parameter approximations involving general phase-type structures required between 4 and almost 300000 seconds CPU time. Run times are generally increasing with increasing order and/or number of parameters of the approximating phase-type distribution. Exceptions are, for instance, the acyclic structure order-2 fit to the $L1$ data (see table 5.13), which took much longer than the order-4 fit of the same category. One possible explanation of this phenomenon might be that the starting point (initial phase-type parameter values) for the EM algorithm was poorly chosen by the random number generator in the order-2 case.

  The number of observations does seem to have some influence on the run times. However, other factors such as the initial starting point, the order and

the structure (i.e. the number of parameters) of the approximating phase-type distribution, and the step-size of the Runge-Kutta method seem to be more significant regarding the performance of EMPHT.

Tables 5.12, 5.13, 5.14, 5.15, 5.16 and 5.17 present the termination criteria (maximum number of iterations, relative Euclidean norm or relative change in log-likelihood, indicated by an asterisk), the final log-likelihood after completion of the parameter approximation run and the run times of the phase-type parameter approximations obtained from EMPHT. Tables 5.12, 5.13 and 5.14 contain results from sum of exponentials, acyclic and general phase-type approximations to the test data generated from the theoretical distributions ($W1$-SE), respectively. Tables 5.15, 5.16 and 5.17 present results based on the empirical data sets ($EMP1$-$EMP4$), respectively.

Tables 5.12 and 5.15 contain empty cells. The reason for the absence of results in these cases is that these parameter approximation runs did not lead to final solutions in any reasonable run time.

Table 5.12. EMPHT PH approximations (SUMEX structure) to the theoretical distributions: general results.

Test distribution		Sum of Exponentials distribution order 2	order 4	order 8	order 16
W1	# of iterations	112	200*	200*	200*
	REN	9.75054e-05*	0.00885854	0.00121975	0.00262119
	Δlog-likelihood	3.93935e-06	0.0688436	0.0036327	0.0200654
	log-likelihood	-158.319	-165.297	-174.56	-177.986
	run time (sec.)	11.9	315.3	776.5	41878
W2	# of iterations	200*			
	REN	0.0101023			
	Δlog-likelihood	0.0200558			
	log-likelihood	-339.326			
	run time (sec.)	189782			
L1	# of iterations	200*	200*	200*	
	REN	0.00347872	0.00345372	0.000125183	
	Δlog-likelihood	0.00240029	0.00709896	2.18358e-05	
	log-likelihood	-186.543	-187.074	-195.844	
	run time (sec.)	20705.6	255902	64481.7	
L2	# of iterations	76	94	200*	200*
	REN	9.16182e-05*	9.52409e-05*	0.000344812	0.00027488
	Δlog-likelihood	1.87232e-06	5.65514e-06	0.00016582	0.000203736
	log-likelihood	-179.879	-176.106	-176.54	-178.437
	run time (sec.)	13.7	173.6	6234.6	9515.7
L3	# of iterations	16	30	52	95
	REN	9.5617e-05*	9.27109e-05*	9.91396e-05*	9.81608e-05*
	Δlog-likelihood	3.29035e-06	6.35417e-06	1.51423e-05	3.05641e-05
	log-likelihood	-126.688	-61.2102	-5.71355	32.8536
	run time (sec.)	2.2	13.6	140.7	2220.6
U1	# of iterations	39	200*	200*	200*
	REN	9.61731e-05*	0.0082806	0.00527849	0.00502599
	Δlog-likelihood	3.54874e-06	0.0280985	0.040492	0.0790557
	log-likelihood	-45.1359	-48.7139	-66.6019	-72.0423
	run time (sec.)	4.2	693.2	2234.2	132058
U2	# of iterations	16	30	51	87
	REN	8.67897e-05*	8.35886e-05*	9.2808e-05*	9.90694e-05*
	Δlog-likelihood	2.90632e-06	5.20706e-06	1.32658e-05	3.10786e-05
	log-likelihood	-207.669	-141.912	-85.8572	-46.1747
	run time (sec.)	2.3	13.3	127	1653.8
ME	# of iterations	74	119	200*	200*
	REN	9.80767e-05*	9.96408e-05*	0.00017971	0.000362747
	Δlog-likelihood	1.98329e-06	5.92268e-06	4.36546e-05	0.000378176
	log-likelihood	-192.511	-186.91	-186.495	-188.793
	run time (sec.)	13.9	206.2	5201.5	8585.5
SE	# of iterations	200*	200*	200*	200*
	REN	0.000104091	0.00553702	0.00516948	0.00518389
	Δlog-likelihood	3.20844e-06	0.0182948	0.038635	0.0811265
	log-likelihood	-281.841	-283.253	-286.046	-291.689
	run time (sec.)	23.7	1563.5	16695.8	101336

Table 5.13. EMPHT PH approximations (APH structure) to the theoretical distributions: general results.

Test distribution		Acyclic Phase-Type distribution order 2	order 4	order 8	order 16
W1	# of iterations	131	200*	200*	200*
	REN	9.66292e-05*	0.000627423	0.000497799	0.00111084
	Δlog-likelihood	4.67731e-06	0.000554762	0.000555898	0.000495575
	log-likelihood	-157.978	-157.269	-157.096	-157.127
	run time (sec.)	13.9	96.9	663.6	11274.2
W2	# of iterations	130	200*	200*	200*
	REN	0.015905	0.00172963	0.0102058	0.0056545
	Δlog-likelihood	9.66121e-07*	0.000342883	0.0305327	0.0107041
	log-likelihood	-336.019	-250.002	-228.17	-227.237
	run time (sec.)	12042.3	1338.5	20179.3	241076.2
L1	# of iterations	200*	200*	200*	200*
	REN	0.00418133	0.000362862	0.000458669	0.000309582
	Δlog-likelihood	4.22602e-05	9.68702e-05	0.000311791	0.000321437
	log-likelihood	-186.309	-87.9987	-77.3226	-77.3253
	run time (sec.)	25994.4	1248.9	6919.3	55617.8
L2	# of iterations	80	200*	200*	200*
	REN	9.91329e-05*	0.00134594	0.000435572	0.00130754
	Δlog-likelihood	0.00115067	0.0106496	0.000790951	0.00733606
	log-likelihood	-179.891	-177.865	-174.877	-175.391
	run time (sec.)	14.8	381.8	1168.2	6434.9
L3	# of iterations	27	68	200*	200*
	REN	7.77078e-05*	8.83315e-05*	0.00315864	0.0014936
	Δlog-likelihood	2.74154e-05	0.000261199	0.096413	0.0632474
	log-likelihood	-126.688	-61.2111	-11.6336	-13.3391
	run time (sec.)	3.4	30.9	834.1	21816.3
U1	# of iterations	61	200*	200*	200*
	REN	9.33381e-05*	0.000105017	0.00152838	0.00109911
	Δlog-likelihood	7.72492e-06	0.000844176	0.00871282	0.0104846
	log-likelihood	-40.3345	-27.8268	-22.3244	-22.7459
	run time (sec.)	6.3	74.8	748.8	14217.6
U2	# of iterations	27	67	200*	200*
	REN	7.13694e-05*	9.82179e-05*	0.0032624	0.00146285
	Δlog-likelihood	2.69158e-05	0.000332237	0.102412	0.0672871
	log-likelihood	-207.669	-141.913	-91.7147	-93.5093
	run time (sec.)	3.3	29.3	642	17502.7
ME	# of iterations	77	200*	200*	200*
	REN	9.99659e-05*	0.000366306	0.0021233	0.00249889
	Δlog-likelihood	0.00254155	0.000137471	0.03501	0.0259235
	log-likelihood	-192.541	-189.942	-177.438	-180.875
	run time (sec.)	14.8	123.1	1185.3	6368.6
SE	# of iterations	123	200*	200*	200*
	REN	9.93443e-05*	0.00177208	0.00108578	0.000600533
	Δlog-likelihood	5.11497e-06	0.00416903	0.00184798	0.00313063
	log-likelihood	-270.199	-265.07	-264.422	-264.234
	run time (sec.)	13.6	124.7	656.3	8692.8

Table 5.14. EMPHT PH approximations (GPH structure) to the theoretical distributions: general results.

Test distribution		General Phase-Type distribution order 2	order 4	order 8	order 16
W1	# of iterations	134	200*	200*	200*
	REN	9.70096e-05*	0.000672326	0.00107421	0.00118417
	Δlog-likelihood	0.000421197	0.00175892	0.0131097	0.0663471
	log-likelihood	-157.988	-157.801	-158.298	-161.066
	run time (sec.)	14.3	101.4	907	10660.6
W2	# of iterations	174	200*	200*	200*
	REN	9.94863e-05*	0.00342279	0.000366215	0.000212669
	Δlog-likelihood	4.24666e-06	0.00725737	0.00243754	0.00607134
	log-likelihood	-250.125	-232.437	-250.126	-250.175
	run time (sec.)	182.2	2425.5	14514.5	203000
L1	# of iterations	86	200*	127	152
	REN	9.65244e-05*	0.00166392	9.90391e-05*	9.67236e-05*
	Δlog-likelihood	4.27639e-06	0.00378723	0.00142494	0.00966223
	log-likelihood	-88.0763	-87.7712	-88.0815	-88.1169
	run time (sec.)	81.5	1980.7	18609.6	293662.2
L2	# of iterations	120	200*	200*	200*
	REN	9.96391e-05*	0.000687156	0.000819834	0.000493552
	Δlog-likelihood	1.75503e-05	0.00133034	0.0105677	0.012423
	log-likelihood	-179.879	-179.761	-177.918	-177.133
	run time (sec.)	20.1	177.3	1941	22551.9
L3	# of iterations	30	96	200*	200*
	REN	7.80921e-05*	8.55741e-05*	0.00497911	0.00677898
	Δlog-likelihood	0.00164439	0.000376106	0.14386	0.348719
	log-likelihood	-126.692	-61.2117	-11.5862	-5.27324
	run time (sec.)	3.6	42.6	554.2	4735.4
U1	# of iterations	61	200*	200*	200*
	REN	9.63428e-05*	0.00411071	0.00266779	0.00284403
	Δlog-likelihood	0.00117951	0.026984	0.057672	0.295628
	log-likelihood	-40.3419	-31.6851	-29.4094	-50.3886
	run time (sec.)	6.3	75.2	500.7	5643.5
U2	# of iterations	29	94	200*	200*
	REN	9.99646e-05*	8.49245e-05*	0.00309656	0.00682164
	Δlog-likelihood	0.00215683	0.000395133	0.13813	0.352977
	log-likelihood	-207.673	-141.913	-88.1061	-84.3575
	run time (sec.)	3.6	40.5	509	3811.4
ME	# of iterations	134	200*	200*	200*
	REN	9.80935e-05*	0.000445533	0.00093485	0.00121819
	Δlog-likelihood	1.60686e-05	0.000795603	0.0195473	0.0609494
	log-likelihood	-192.511	-192.443	-189.796	-189.734
	run time (sec.)	23.2	180.9	1944.7	22941.7
SE	# of iterations	130	200*	200*	200*
	REN	9.60392e-05*	0.00403005	0.00317891	0.000474048
	Δlog-likelihood	0.000681611	0.0429781	0.096801	0.00185604
	log-likelihood	-270.213	-267.966	-272.659	-279.928
	run time (sec.)	14.6	113	1153.5	13490.3

Table 5.15. EMPHT PH approximations (SUMEX structure) to the empirical distributions: general results.

		Sum of Exponentials distribution			
Test distribution		order 2	order 4	order 8	order 16
EMP1	# of iterations	98	200*		
	REN	9.75332e-05*	0.00119341		
	Δlog-likelihood	2.56128e-05	0.00827848		
	log-likelihood	-3927.48	-4013.59		
	run time (sec.)	28.2	714.6		
EMP2	# of iterations	72	110	200*	200*
	REN	9.5782e-05*	9.9935e-05*	0.00194325	0.00156821
	Δlog-likelihood	7.76269e-06	1.43075e-05	0.0106687	0.02653
	log-likelihood	-810.904	-790.381	-795.535	-814.682
	run time (sec.)	9.3	128.9	3899.6	20482.5
EMP3	# of iterations	101	200*		
	REN	9.99486e-05*	0.00119082		
	Δlog-likelihood	1.75632e-05	0.00363218		
	log-likelihood	-1887.87	-1980.21		
	run time (sec.)	16.4	311.2		
EMP4	# of iterations	110	200*		
	REN	9.5025e-05*	0.00036463		
	Δlog-likelihood	3.16482e-05	0.00117988		
	log-likelihood	-5560.92	-5825.61		
	run time (sec.)	39.6	935.5		

Table 5.16. EMPHT PH approximations (APH structure) to the empirical distributions: general results.

Test distribution		Acyclic Phase-Type distribution order 2	order 4	order 8	order 16
EMP1	# of iterations	140	200*	200*	200*
	REN	9.62509e-05*	0.00124957	0.000370313	0.0011946
	Δlog-likelihood	0.000151024	0.00867982	0.00457382	0.00592014
	log-likelihood	-3904.27	-3901.49	-3891.3	-3894.22
	run time (sec.)	38.2	419	1454.6	6223.6
EMP2	# of iterations	63	200*	200*	200*
	REN	9.658e-05*	0.0011322	0.00134341	0.00130645
	Δlog-likelihood	0.000331288	0.0144834	0.0128354	0.0240403
	log-likelihood	-810.905	-792.638	-780.583	-781.756
	run time (sec.)	8.8	287.3	1237.9	8279.4
EMP3	# of iterations	65	200*	200*	200*
	REN	9.39546e-05*	0.00100344	0.000320488	0.000437576
	Δlog-likelihood	0.000187674	0.0258219	0.00241836	0.0174643
	log-likelihood	-1851.54	-1842.55	-1836.89	-1837.82
	run time (sec.)	11.1	261.6	860.2	13676.9
EMP4	# of iterations	156	200*	200*	200*
	REN	9.69242e-05*	0.00185763	0.000971627	0.000652069
	Δlog-likelihood	0.000130798	0.00468772	0.021572	0.00681012
	log-likelihood	-5455.4	-5436.24	-5430.3	-5433.28
	run time (sec.)	49.3	281.5	1757.6	10062.7

Table 5.17. EMPHT PH approximations (GPH structure) to the empirical distributions: general results.

Test distribution		General Phase-Type distribution order 2	order 4	order 8	order 16
EMP1	# of iterations	194	200*	200*	200*
	REN	9.75038e-05*	0.00151504	0.000379073	0.000309255
	Δlog-likelihood	5.01458e-05	0.0312489	0.00699739	0.012484
	log-likelihood	-3904.27	-3901.5	-3902.95	-3899.68
	run time (sec.)	46.6	194.3	1518	17538.7
EMP2	# of iterations	76	200*	200*	200*
	REN	9.72195e-05*	0.00267294	0.001409	0.000572453
	Δlog-likelihood	3.33624e-05	0.0686931	0.0397706	0.0186008
	log-likelihood	-810.904	-792.298	-791.321	-791.215
	run time (sec.)	9.8	135.7	1340.4	14393.9
EMP3	# of iterations	92	200*	200*	200*
	REN	9.95736e-05*	0.00124276	0.00104457	0.000449464
	Δlog-likelihood	4.25107e-05	0.0160355	0.0382445	0.0167992
	log-likelihood	-1851.54	-1850.26	-1843.18	-1840.16
	run time (sec.)	14.8	134.2	1009.2	11786.8
EMP4	# of iterations	200*	200*	200*	200*
	REN	0.000442049	0.000663317	0.000178792	0.000334678
	Δlog-likelihood	0.00120905	0.00961837	0.00253298	0.0136401
	log-likelihood	-5455.42	-5438.98	-5455.33	-5452.78
	run time (sec.)	53.4	186.9	1271.8	14684.3

5.4.2 EMPHT Experimental Results

Density and distribution plots: theoretical test cases

A visual inspection of the order-2 EMPHT phase-type approximations to the theoretical test distributions $W1$-SE (figures 5.10 and 5.11) shows that the different structures (sum of exponentials, acyclic phase-type, general phase-type) cannot be distinguished in several cases, e.g. $L2$, $L3$, $U2$ and ME. It is shown in tables 5.12, 5.13 and 5.14 that the EM algorithm converged in most of these cases. A possible explanation is therefore the non-uniqueness of phase-type distributions. The EM algorithm, although fitting different phase-type structures and therefore resulting in different phase-type representations, ended up finding the same phase-type distribution. More obvious differences between the different phase-type structures are seen in test cases $W2$ and $L1$. The acyclic and sum of exponentials structure seem to represent the same phase-type distribution, but do not fit the original distributions well. The general phase-type approximation performs better in both these cases.

Visual differences become more pronounced in the plots of order-4 EMPHT approximations (figures 5.12 and 5.13), except in test cases $L3$ and $U2$. A slight trend towards the acyclic phase-type structure providing better fits than the other two structures can be seen in test cases $W1$, $L1$ (together with the general phase-type structure), $U1$, ME and SE. We note that the plots for test case $W2$ in figures 5.12 and 5.13 do not contain approximations based on the sum of exponential structure, because no fit could be obtained.

From the plots based on order-8 EMPHT approximations (figures 5.14 and 5.15), we see that the acyclic phase-type structure provides better fits over the other two structures in test cases $W1$, $W2$, $L1$, $L2$, $U1$, ME (second mode slightly present) and SE (bimodality present). For test cases $L3$ and $U2$, however, the sum of exponentials structure tends to approximate the original distributions slightly better than the other two structures. The plots for test case $W2$ in figures 5.14 and 5.15 do not contain approximations based on the sum of exponential structure.

These observations become more pronounced in the plots based on order-16 EMPHT approximations (figures 5.16 and 5.17). The acyclic phase-type structure approximates the original distributions better than the sum of exponentials and general phase-type structure in all test cases except $L3$ and $U2$, in which the sum of exponentials structure clearly outperforms the other two structures. The plots for test cases $W2$ and $L1$ in figures 5.16 and 5.17 do not contain approximations based on the sum of exponential structure.

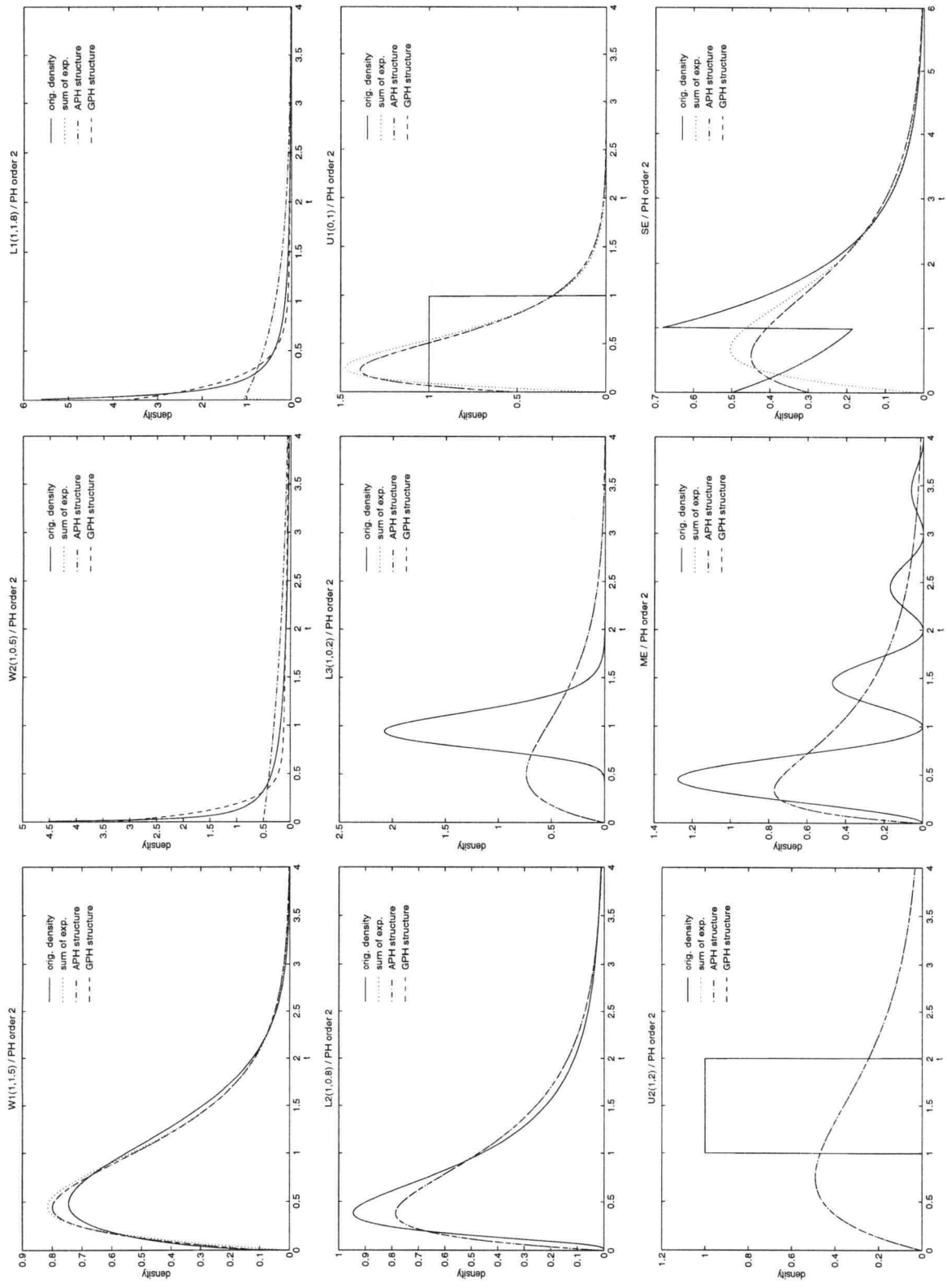

Figure 5.10. EMPHT order-2 approximations to the theoretical densities.

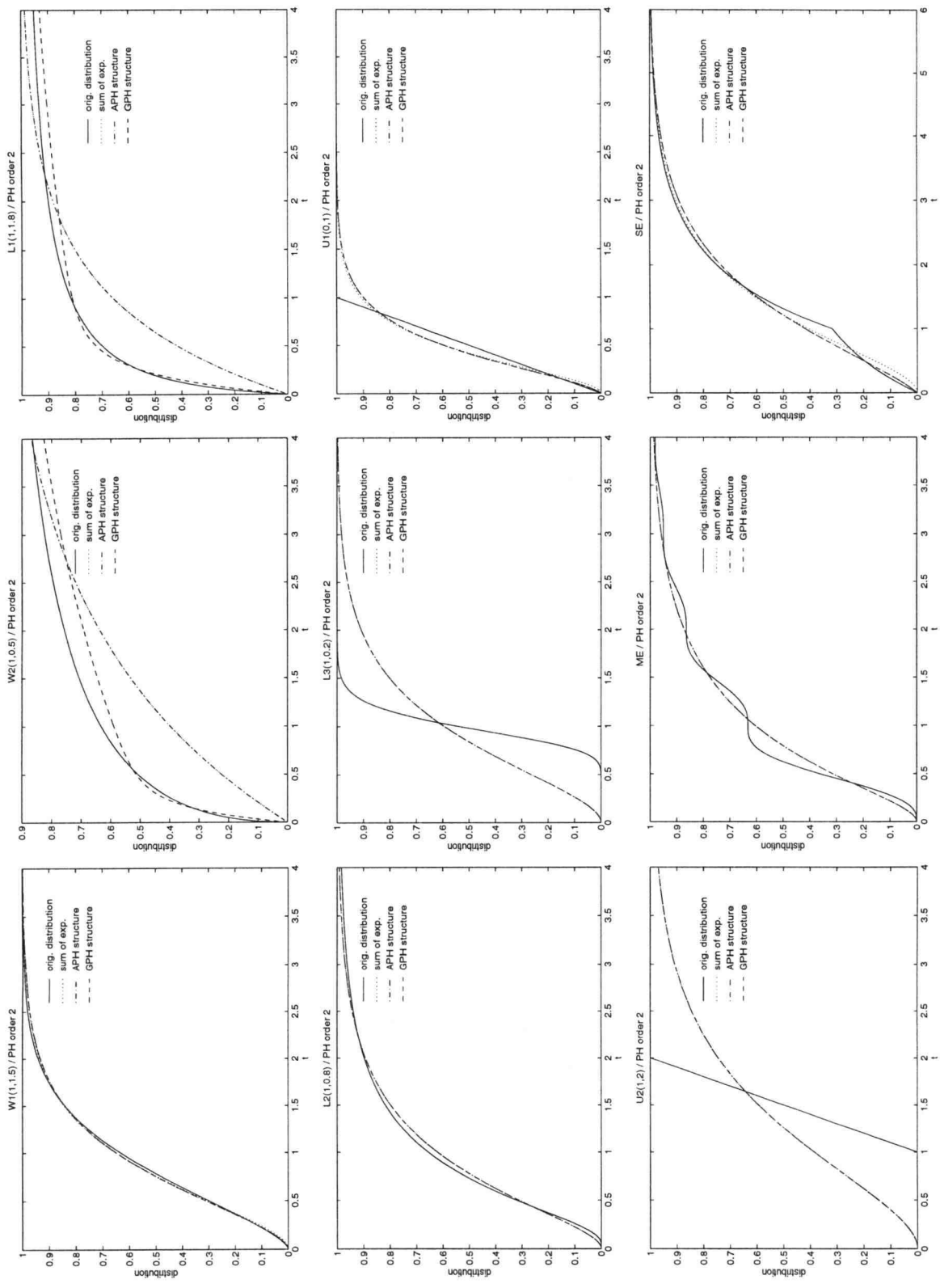

Figure 5.11. EMPHT order-2 approximations to the theoretical distributions.

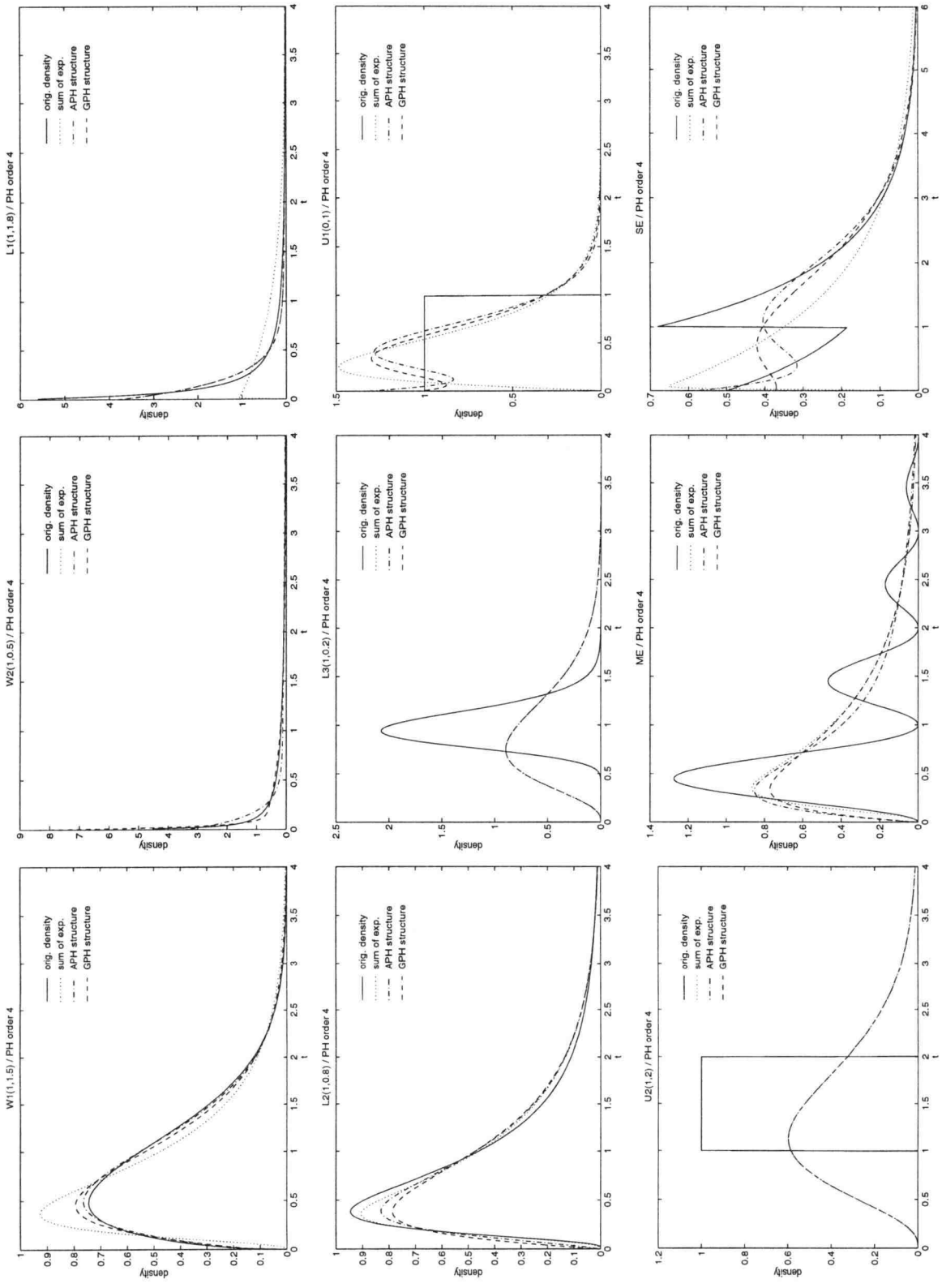

Figure 5.12. EMPHT order-4 approximations to the theoretical densities.

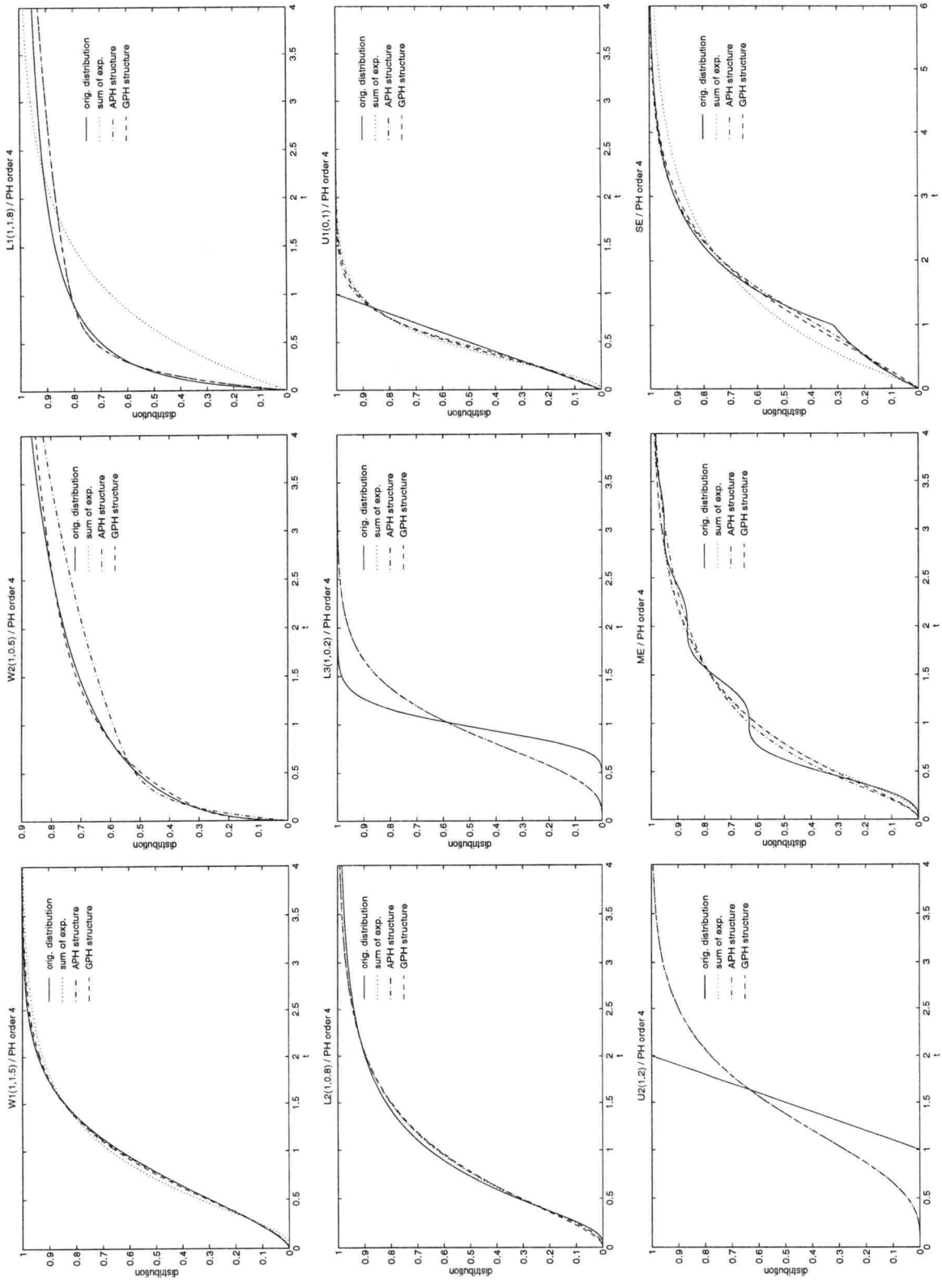

Figure 5.13. EMPHT order-4 approximations to the theoretical distributions.

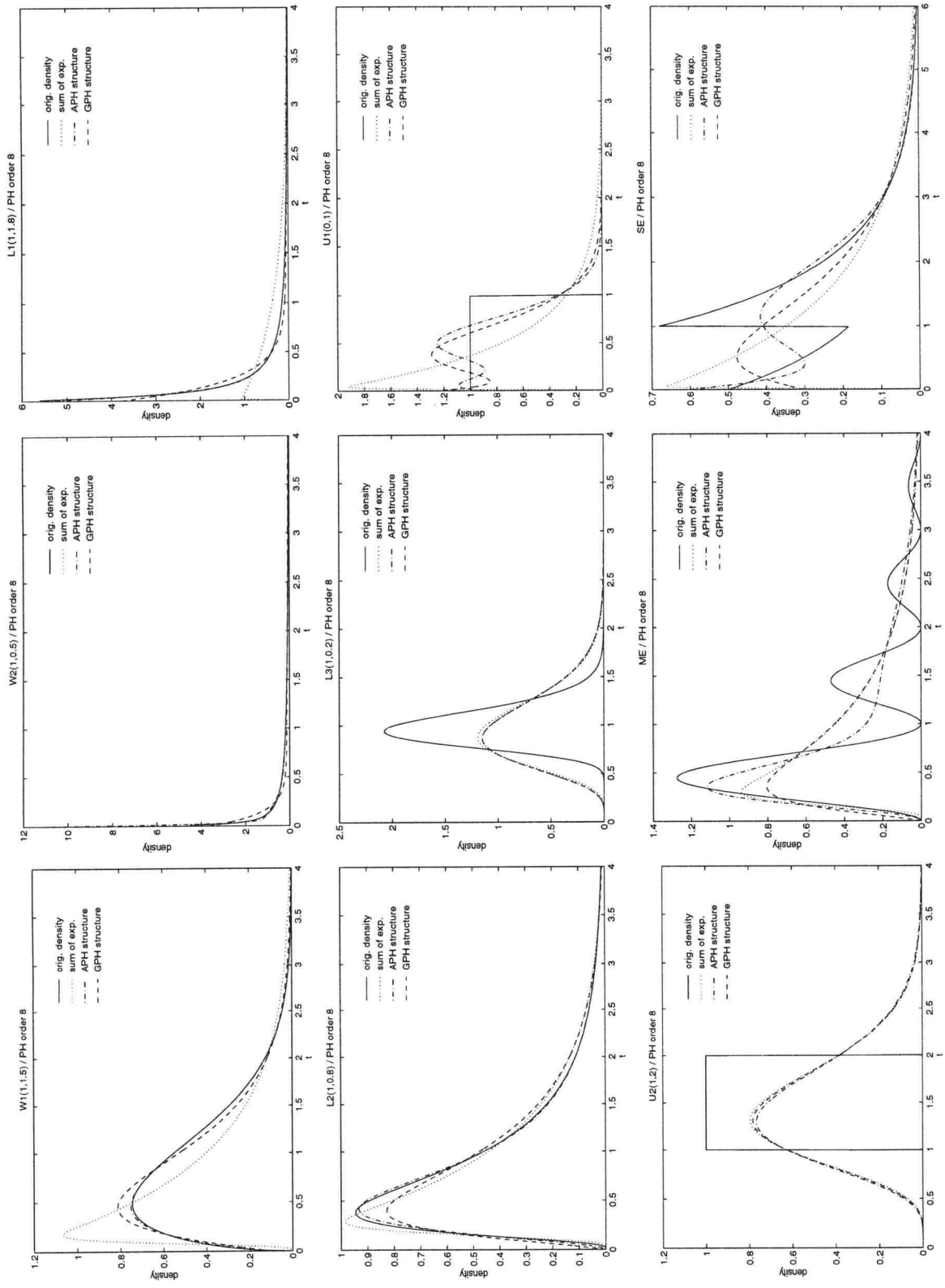

Figure 5.14. EMPHT order-8 approximations to the theoretical densities.

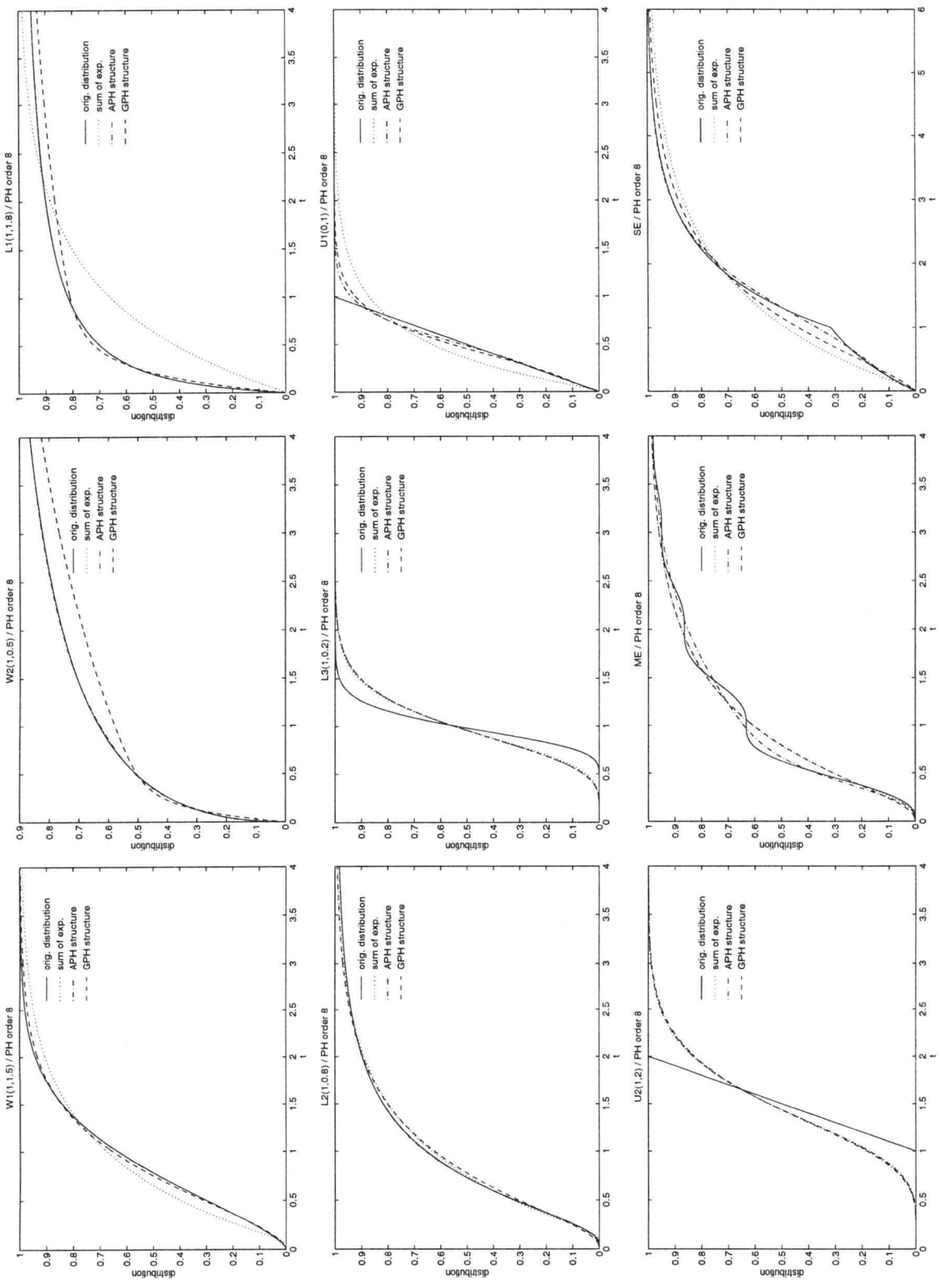

Figure 5.15. EMPHT order-8 approximations to the theoretical distributions.

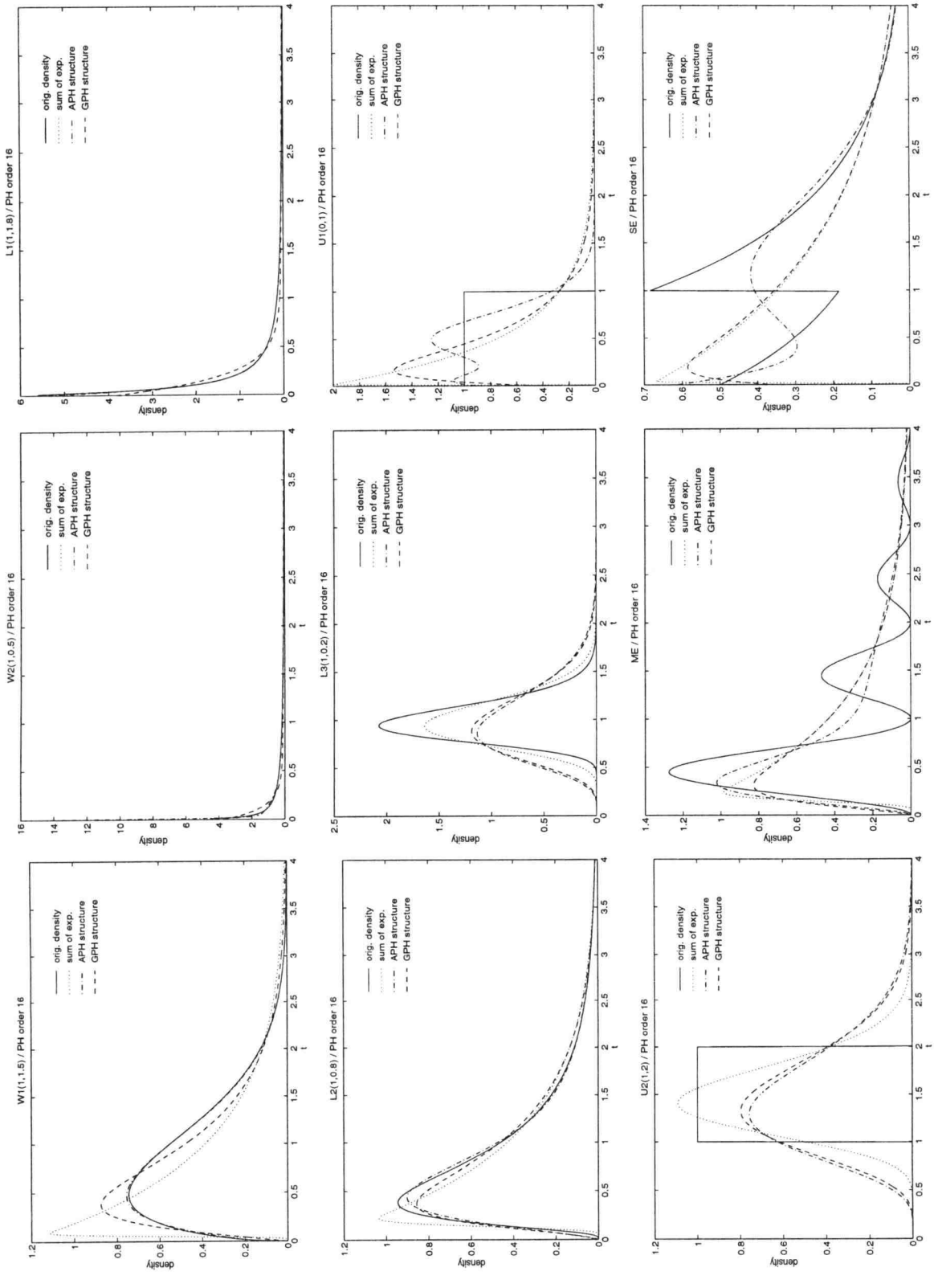

Figure 5.16. EMPHT order-16 approximations to the theoretical densities.

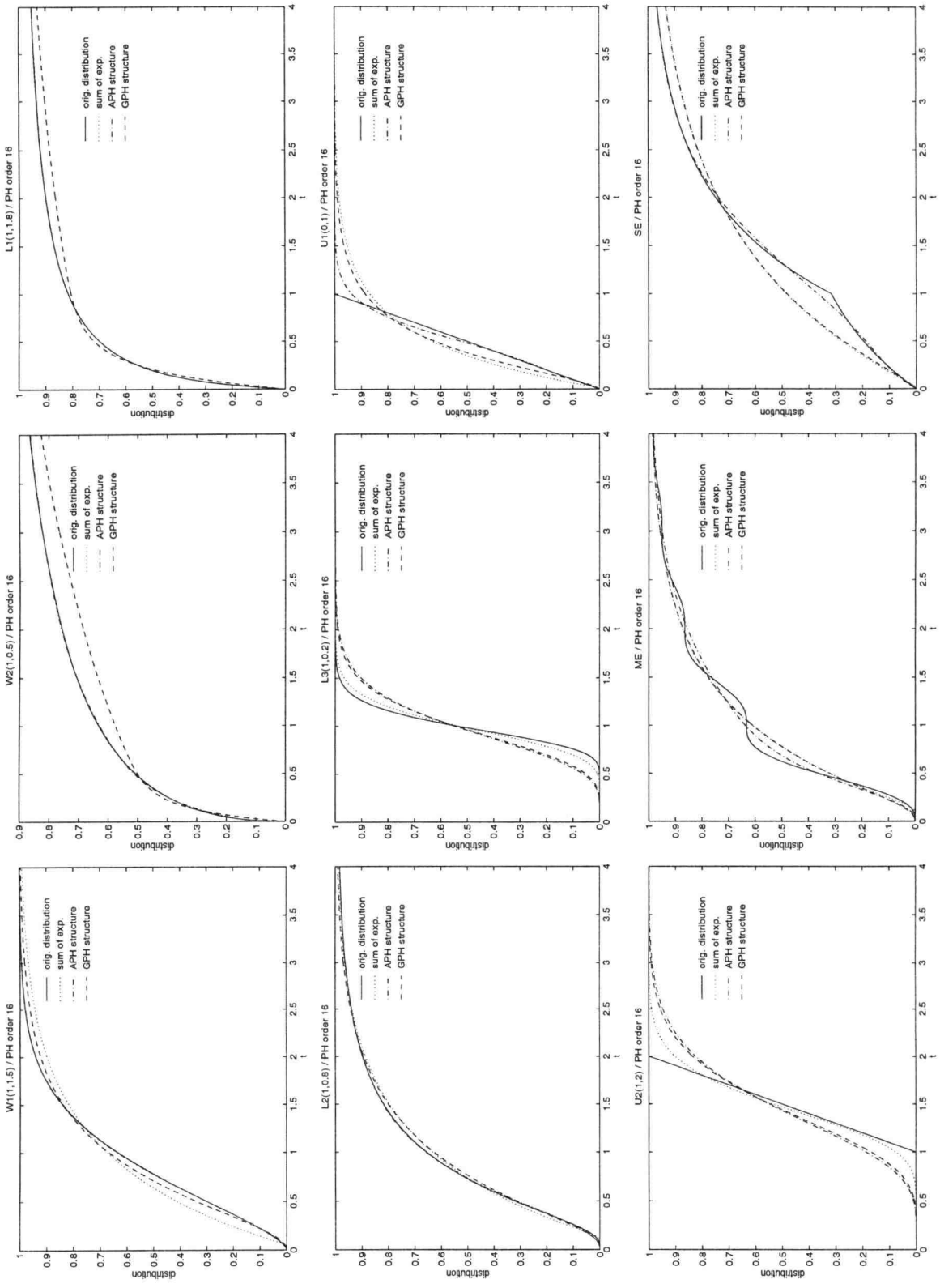

Figure 5.17. EMPHT order-16 approximations to the theoretical distributions.

Performance measures: theoretical test cases

Table 5.18 presents the area differences and entropy measures between the original theoretical test distributions ($W1$-SE) and the EMPHT sum of exponentials approximations. Table 5.19 presents the relative errors of the first three central moments between the original theoretical test distributions and the EMPHT sum of exponentials approximations. These tables do not contain the performance measures in test cases $W2$ (orders 4, 8 and 16) and $L1$ (order 16) because of our inability to obtain sum of exponential fits in these cases. The relative error on the third moment is undefined for test distributions $U1$ and $U2$ since their central third moment μ_3 equals zero (see equation 4.22). The performance measure reported (and marked with an asterisk) in these two cases is the central third moment $\hat{\mu}_3$ of the approximating phase-type distribution.

The area differences and entropy measures for the sum of exponentials structure (see table 5.18) generally do not follow the expected behavior of decreasing values for increasing orders. While the results in test cases $L3$ and $U2$ do follow this rule, test case $W1$ shows increasing area differences and entropy measures for increasing orders, i.e. the phase-type fit gets worse with increasing orders of the approximating sum of exponential distribution.

The relative moment errors (see table 5.19) show very inconsistent behavior, but in most cases do not improve with increasing orders of the approximating sum of exponentials distribution (see test cases $L1$, $L2$, ME). Some improvements with increasing orders are evident in test cases $L3$ and $U2$, in particular with respect to the second and third central moment. Unusual behavior is exhibited by test case SE, where the order-2 sum of exponential fit shows much better moment approximation than higher order fits in this case.

Table 5.18. EMPHT PH approximations (SUMEX structure) to the theoretical distributions: area differences and entropy measures.

Test distribution		Sum of Exponentials distribution			
		order 2	order 4	order 8	order 16
W1	Δpdf	0.074916	0.180690	0.291007	0.323317
	Δcdf	0.029881	0.079488	0.153521	0.179364
	cross entropy	0.793768	0.830337	0.876873	0.904406
	relative entropy	0.006893	0.043462	0.089998	0.117531
W2	Δpdf	0.655295			
	Δcdf	0.538502			
	cross entropy	1.710735			
	relative entropy	0.595647			
L1	Δpdf	0.702173	0.699201	0.699658	
	Δcdf	0.702885	0.703158	0.704245	
	cross entropy	0.969929	0.969267	0.967176	
	relative entropy	0.580207	0.579546	0.577455	
L2	Δpdf	0.146091	0.070631	0.102072	0.143289
	Δcdf	0.073592	0.056754	0.054396	0.060565
	cross entropy	0.904611	0.886218	0.889258	0.900457
	relative entropy	0.028817	0.010424	0.013464	0.024663
L3	Δpdf	1.112907	0.848279	0.548338	0.224156
	Δcdf	0.384074	0.231995	0.119900	0.039208
	cross entropy	0.633706	0.306582	0.029630	-0.16215
	relative entropy	0.844205	0.517082	0.240129	0.048355
U1	Δpdf	0.452214	0.457487	0.578961	0.582297
	Δcdf	0.157477	0.156968	0.302080	0.309961
	cross entropy	0.243759	0.296220	0.454894	0.393363
	relative entropy	0.243759	0.296220	0.454894	0.393363
U2	Δpdf	1.278881	0.999510	0.672859	0.386033
	Δcdf	0.383727	0.230028	0.117668	0.042066
	cross entropy	1.038318	0.709540	0.429274	0.230865
	relative entropy	1.038318	0.709540	0.429274	0.230865
ME	Δpdf	0.544396	0.487496	0.480222	0.507312
	Δcdf	0.099600	0.100283	0.095944	0.089267
	cross entropy	0.965573	0.938556	0.938198	0.952795
	relative entropy	0.237842	0.210826	0.210467	0.225064
SE	Δpdf	0.311310	0.433832	0.435122	0.435987
	Δcdf	0.056959	0.181962	0.183090	0.183489
	cross entropy	1.457939	1.558526	1.554611	1.722817
	relative entropy	0.162977	0.263564	0.259649	0.427855

Table 5.19. EMPHT PH approximations (SUMEX structure) to the theoretical distributions: relative moment errors.

Test distribution		Sum of Exponentials distribution			
		order 2	order 4	order 8	order 16
W1	e_1	0.0006462	0.0006462	0.0004842	0.0006463
	e_2	0.0832136	0.3139494	0.7589373	0.9163598
	e_3	0.4873199	1.6061395	3.3373523	3.9473611
W2	e_1	0.0129666			
	e_2	0.8051679			
	e_3	0.9740137			
L1	e_1	0.0661845	0.0661845	0.0661372	
	e_2	0.9645307	0.9645796	0.9647656	
	e_3	0.9999020	0.9999022	0.9999030	
L2	e_1	0.0038236	0.0038236	0.0038236	0.0038238
	e_2	0.2618068	0.2663248	0.2279865	0.1828077
	e_3	0.6841261	0.6684063	0.6354891	0.6006270
L3	e_1	0.0001304	0.0001326	0.0001327	0.0001327
	e_2	11.248472	5.1242103	2.0621081	0.5310847
	e_3	97.687266	23.671666	5.1679371	0.5420769
U1	e_1	6.6684e-09	9.72951e-11	1.40615e-06	1.534528e-8
	e_2	0.5000009	0.4739451	1.7931919	1.8685305
	e_3*	0.0625001	0.0608714	0.2244690	0.2337347
U2	e_1	2.87443e-06	1.28347e-07	5.08805e-08	4.88868e-09
	e_2	12.500078	5.7500029	2.3750037	0.6875275
	e_3*	1.6875115	0.4218754	0.1054691	0.0263685
ME	e_1	0.0020953	0.0020954	0.0020954	0.0020973
	e_2	0.1522279	0.2034203	0.1975544	0.1573123
	e_3	0.3025947	0.3499875	0.3346733	0.2798463
SE	e_1	0.0011551	0.0011551	0.0011552	0.0011567
	e_2	0.0314707	0.7403451	0.7462973	0.7483915
	e_3	0.2160013	2.2082131	2.2249506	2.2308369

Table 5.20 presents the area differences and entropy measures between the original theoretical test distributions ($W1$-SE) and the EMPHT acyclic phase-type approximations. Table 5.21 presents the relative errors of the first three central moments between the original theoretical test distributions and the EMPHT acyclic phase-type approximations.

The EMPHT acyclic phase-type approximations exhibit more consistent behavior in term of decreasing area differences and entropy measures with increasing orders. However, the order-16 approximations are often worse than the order-8 approximations. A possible explanation is that a larger number of parameters to be approximated requires more iterations of the EM algorithm and termination after 200 iterations results in solutions far away from the optimum.

The relative moment errors (see table 5.21) show this observation too. Order 16 acyclic phase-type approximations are generally worse than order-8 approximations. The first moments, however, are virtually not improving at all with increasing orders, except in test case $U2$, where consistent improvement with increasing order is obtained.

Table 5.22 presents the area differences and entropy measures between the original theoretical test distributions ($W1$-SE) and the EMPHT general phase-type approximations. Table 5.23 presents the relative errors of the first three central moments between the original theoretical test distributions and the EMPHT general phase-type approximations.

The EMPHT general phase-type approximations show the trend of increasing area differences and entropy measures with orders 8 and 16 even more pronounced than the previously discussed acyclic phase-type approximations, except in test cases $L2$ and $L3$, where consistently decreasing values are observed.

As in the results based on the acyclic structure, the first central moments are generally well approximated. The relative errors on the second and third central moment, however, are getting worse with order-8 and order-16 general phase-type approximations. A possible explanation is the insufficient number of EM iterations.

Table 5.20. EMPHT PH approximations (APH structure) to the theoretical distributions: area differences and entropy measures.

Test distribution		Acyclic Phase-Type distribution			
		order 2	order 4	order 8	order 16
W1	Δpdf	0.065136	0.025720	0.008282	0.012273
	Δcdf	0.032092	0.009718	0.002611	0.004415
	cross entropy	0.791274	0.787834	0.787170	0.787271
	relative entropy	0.004400	0.000959	0.000295	0.000396
W2	Δpdf	0.655241	0.340090	0.094570	0.100402
	Δcdf	0.538482	0.250298	0.028371	0.027903
	cross entropy	1.690403	1.255721	1.143268	1.138608
	relative entropy	0.575315	0.140634	0.028180	0.023520
L1	Δpdf	0.701682	0.257388	0.044179	0.040023
	Δcdf	0.702476	0.303387	0.065487	0.065023
	cross entropy	0.970900	0.459684	0.400151	0.400314
	relative entropy	0.581178	0.069963	0.010429	0.010592
L2	Δpdf	0.146224	0.111108	0.026193	0.049243
	Δcdf	0.073604	0.066949	0.007774	0.012130
	cross entropy	0.904670	0.894642	0.877807	0.882528
	relative entropy	0.028876	0.018849	0.002013	0.006735
L3	Δpdf	1.112908	0.848283	0.585591	0.596055
	Δcdf	0.384074	0.231997	0.131474	0.134801
	cross entropy	0.633706	0.306587	0.059146	0.067651
	relative entropy	0.844205	0.517086	0.269645	0.278150
U1	Δpdf	0.422528	0.314000	0.255496	0.258553
	Δcdf	0.167626	0.089862	0.061663	0.063757
	cross entropy	0.201677	0.139112	0.111618	0.113725
	relative entropy	0.201677	0.139112	0.111618	0.113725
U2	Δpdf	1.278881	0.999516	0.712297	0.723972
	Δcdf	0.383727	0.230030	0.128825	0.132275
	cross entropy	1.038318	0.709546	0.458561	0.467533
	relative entropy	1.038318	0.709546	0.458561	0.467533
ME	Δpdf	0.544621	0.525132	0.403744	0.447769
	Δcdf	0.099551	0.080448	0.060780	0.067500
	cross entropy	0.965716	0.952838	0.890392	0.908554
	relative entropy	0.237985	0.225108	0.162661	0.180823
SE	Δpdf	0.282692	0.187886	0.170921	0.167840
	Δcdf	0.065940	0.033834	0.022770	0.021010
	cross entropy	1.352755	1.327307	1.323870	1.323324
	relative entropy	0.057793	0.032345	0.028908	0.028362

Table 5.21. EMPHT PH approximations (APH structure) to the theoretical distributions: relative moment errors.

Test distribution		Acyclic Phase-Type distribution order 2	order 4	order 8	order 16
W1	e_1	0.0006462	0.0006462	0.0006462	0.0006462
	e_2	0.1288516	0.0215842	0.003186	0.0068040
	e_3	0.5842943	0.1546853	0.0373382	0.0656494
W2	e_1	0.0129666	0.0129666	0.0129666	0.0129666
	e_2	0.8051530	0.5236228	0.1526079	0.1491846
	e_3	0.9740107	0.8696883	0.44688633	0.4397835
L1	e_1	0.0661845	0.0661845	0.0661845	0.0661845
	e_2	0.9644576	0.8373351	0.6726668	0.6727947
	e_3	0.9999017	0.9980503	0.9878719	0.9878983
L2	e_1	0.0038235	0.0038236	0.0038237	0.0038236
	e_2	0.2616123	0.2758888	0.0678652	0.0781492
	e_3	0.6840064	0.6851092	0.2829589	0.3369267
L3	e_1	0.0001304	0.0001326	0.0001327	0.0001327
	e_2	11.248476	5.1242737	2.3262098	2.4044680
	e_3	97.687303	23.672048	6.2343031	6.5934678
U1	e_1	4.42761e-9	2.888e-10	1.05414e-9	6.79e-12
	e_2	0.7040267	0.2665294	0.1472548	0.1556543
	e_3*	0.0765837	0.0278425	0.0155061	0.0164044
U2	e_1	0.0000029	0.0000001	6.82414e-9	2.9e-13
	e_2	12.500082	5.7500918	2.6588291	2.7481826
	e_3*	1.6875154	0.4218838	0.1231677	0.1295141
ME	e_1	0.0020953	0.0020951	0.0020954	0.0020954
	e_2	0.1517101	0.0053725	0.0069423	0.0073460
	e_3	0.3019492	0.0814578	0.0658463	0.1084807
SE	e_1	0.0011552	0.0011551	0.0011551	0.0011551
	e_2	0.1402241	0.0301253	0.0014090	0.0010130
	e_3	0.2535244	0.2066769	0.0440679	0.0723611

In general, the visual observations from the plots presented in figures 5.10-5.17 are matched by the performance measures presented in the previous tables 5.18-5.23. The non-uniqueness of approximating phase-type distributions, in particular the order-2 approximations, is clearly shown by the virtually equal performance measures in the corresponding test cases. The acyclic phase-type approximations are performing better than the sum of exponentials and general phase-type structures, except in test cases $L3$ and $U2$, where the sum of exponentials structure is superior. These two test cases are distributions with small squared coefficient of variation ($c^2 = 0.041$ for $L3$, $c^2 = 0.037$ for $U2$). A possible explanation is that the sum of exponential distribution is the closest relative to an Erlang distribution, which achieves the smallest squared coefficient of variation among all phase-type distributions (see Aldous and Shepp [2]). Since the orders of the approximating phase-type distributions in our experiments are not large enough to achieve the low variability of the original distributions (at least order 25 necessary for $L3$, at least order 27 necessary for $U2$), these phase-type approximations are driven by the EM algorithm to achieve their lowest possible variability, i.e. become Erlang distributions. This goal is easier to achieve for sum of exponential distributions than for any other phase-type structure.

Table 5.22. EMPHT PH approximations (GPH structure) to the theoretical distributions: area differences and entropy measures.

Test distribution		General Phase-Type distribution			
		order 2	order 4	order 8	order 16
W1	Δpdf	0.065561	0.057066	0.076065	0.149521
	Δcdf	0.032261	0.026830	0.036020	0.076737
	cross entropy	0.791325	0.790362	0.792753	0.807101
	relative entropy	0.004451	0.003488	0.005878	0.020226
W2	Δpdf	0.342189	0.181521	0.344900	0.345712
	Δcdf	0.252359	0.107848	0.263970	0.267922
	cross entropy	1.256360	1.165922	1.256532	1.256840
	relative entropy	0.141272	0.050834	0.141445	0.141752
L1	Δpdf	0.258603	0.252652	0.256603	0.253402
	Δcdf	0.303722	0.304273	0.306501	0.311434
	cross entropy	0.460064	0.458748	0.460312	0.460870
	relative entropy	0.070343	0.069027	0.070591	0.071148
L2	Δpdf	0.146111	0.144128	0.112104	0.098451
	Δcdf	0.073636	0.073465	0.068118	0.061594
	cross entropy	0.904613	0.904034	0.894932	0.890999
	relative entropy	0.028820	0.028241	0.019138	0.015205
L3	Δpdf	1.112921	0.848286	0.585290	0.545227
	Δcdf	0.384085	0.231998	0.131332	0.118914
	cross entropy	0.633724	0.306590	0.058911	0.027443
	relative entropy	0.844223	0.517089	0.269410	0.237942
U1	Δpdf	0.422605	0.339885	0.324185	0.499564
	Δcdf	0.167660	0.115197	0.099828	0.233900
	cross entropy	0.201714	0.158403	0.147022	0.251964
	relative entropy	0.201714	0.158403	0.147022	0.251964
U2	Δpdf	1.278897	0.999518	0.688366	0.663913
	Δcdf	0.383742	0.230031	0.121897	0.114419
	cross entropy	1.038342	0.709548	0.440518	0.421775
	relative entropy	1.038342	0.709548	0.440518	0.421775
ME	Δpdf	0.544378	0.543697	0.519345	0.524633
	Δcdf	0.099648	0.099868	0.101374	0.092276
	cross entropy	0.965574	0.965234	0.952207	0.951885
	relative entropy	0.237844	0.237504	0.224476	0.224155
SE	Δpdf	0.282874	0.244040	0.321990	0.415497
	Δcdf	0.066049	0.048908	0.095821	0.174655
	cross entropy	1.352822	1.341779	1.364666	1.400787
	relative entropy	0.057860	0.046817	0.069704	0.105825

Table 5.23. EMPHT PH approximations (GPH structure) to the theoretical distributions: relative moment errors.

Test distribution		General Phase-Type distribution order 2	order 4	order 8	order 16
W1	e_1	0.0006462	0.0006462	0.0006462	0.0006462
	e_2	0.1294278	0.1011909	0.1389397	0.3419463
	e_3	0.5883139	0.4926943	0.7088512	1.6159294
W2	e_1	0.0129666	0.0129666	0.0129666	0.0129666
	e_2	0.5260276	0.3030702	0.5416983	0.5471147
	e_3	0.8711244	0.6870788	0.8798539	0.88274
L1	e_1	0.0661846	0.0661845	0.0661845	0.0661845
	e_2	0.8372057	0.8381358	0.8390823	0.8422390
	e_3	0.9980481	0.9980731	0.9980998	0.9981847
L2	e_1	0.0038236	0.0038236	0.0038236	0.0038236
	e_2	0.2621265	0.2653929	0.2823159	0.2582906
	e_3	0.6844001	0.6862342	0.6902469	0.6678525
L3	e_1	0.0001304	0.0001326	0.0001327	0.0001327
	e_2	11.249207	5.1243351	2.3209649	2.0333779
	e_3	97.704967	23.672858	6.1580621	4.7905716
U1	e_1	4.43195e-9	6.58497e-9	7.14243e-9	4.871e-10
	e_2	0.7042003	0.3909010	0.3143483	1.2175692
	e_3*	0.0766241	0.0417241	0.0333872	0.1487667
U2	e_1	0.0000029	0.0000001	3.02638e-8	6.67547e-9
	e_2	12.501116	5.7501419	2.4818372	2.2964995
	e_3*	1.6879024	0.4218955	0.1112479	0.0960988
ME	e_1	0.0020953	0.0020953	0.0020954	0.0020954
	e_2	0.1524953	0.1542678	0.1833600	0.1337232
	e_3	0.3029970	0.3051657	0.3349686	0.2628802
SE	e_1	0.0011552	0.0011551	0.0011551	0.0011551
	e_2	0.1410369	0.0578444	0.3132642	0.7065228
	e_3	0.2574707	0.0089833	0.8790306	2.1054250

Density and distribution plots: empirical test cases

Figure 5.18 shows plots of the empirical test densities and distributions ($EMP1$-$EMP4$) together with the approximating EMPHT phase-type densities and distributions of order 2. We note that these order-2 approximations seem to satisfactorily match the shape of the empirical distributions in test cases $EMP1$ and $EMP4$, but not in test cases $EMP2$ and $EMP3$. There is no obvious visual difference between the three different phase-type structures.

Differences between the approximating phase-type structures start to appear in order-4 EMPHT approximations (see figure 5.19). The sum of exponentials structure seems to do worse than the other two structures. The acyclic phase-type structure appears to approximate the shape of test case $EMP3$ better than the other two structures.

The superiority of the acyclic phase-type structure becomes more pronounced in the plots of EMPHT order 8 and order 16 approximations to the empirical data sets (see figures 5.20 and 5.21). We note, however, that order-8 and order-16 sum of exponential fits could not be obtained except for test case $EMP2$.

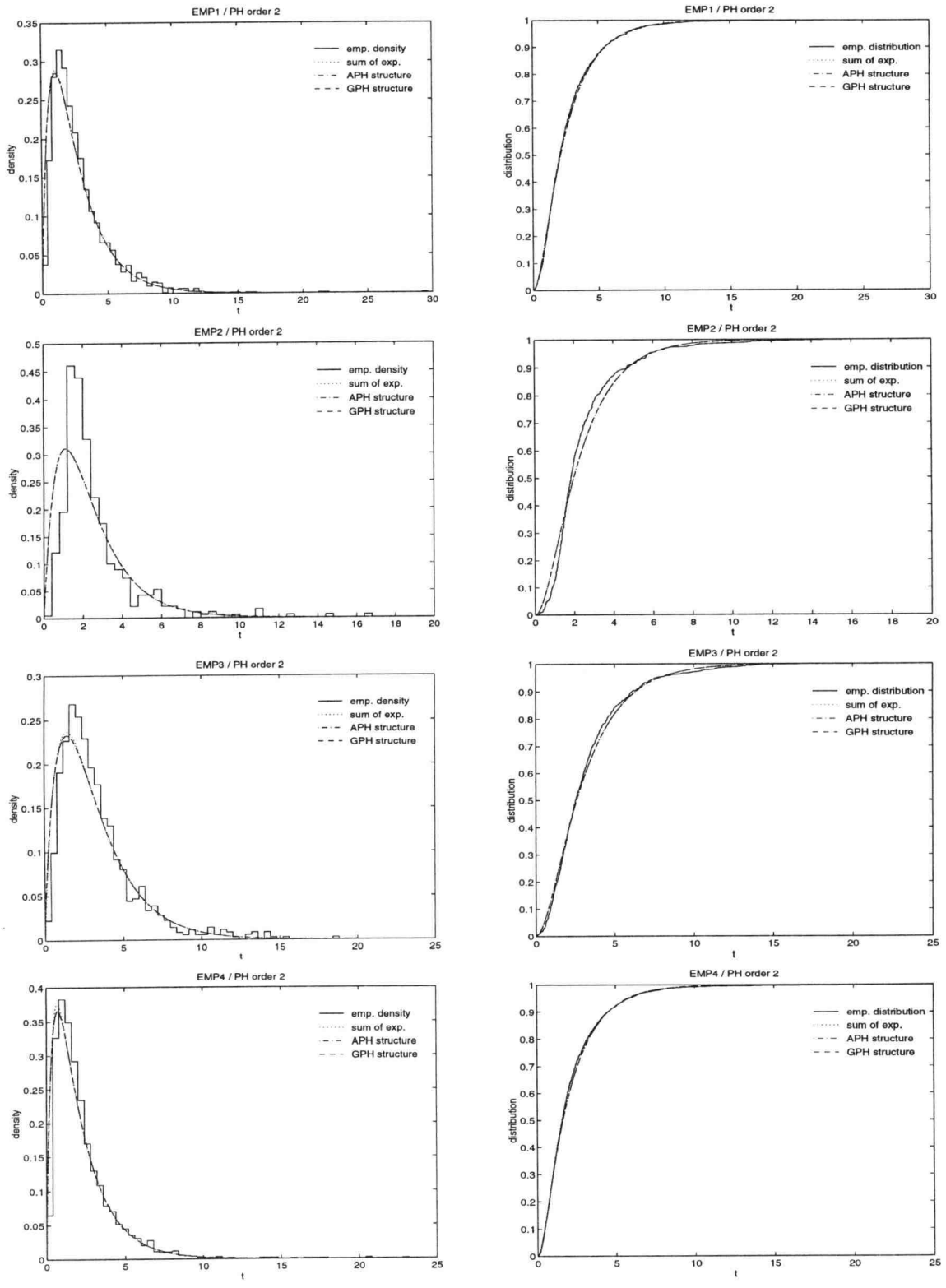

Figure 5.18. EMPHT order-2 approximations to the empirical densities and distributions.

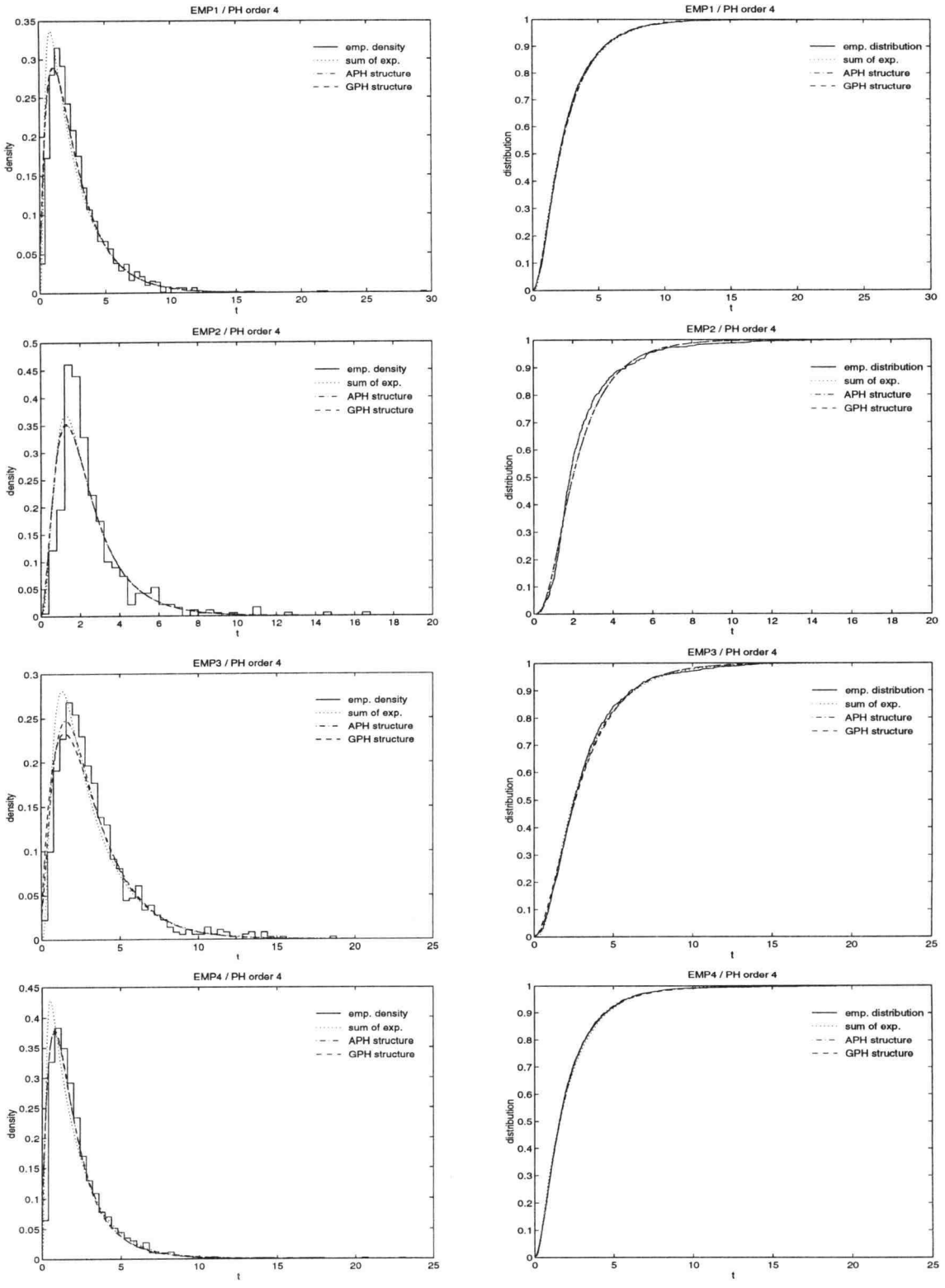

Figure 5.19. EMPHT order-4 approximations to the empirical densities and distributions.

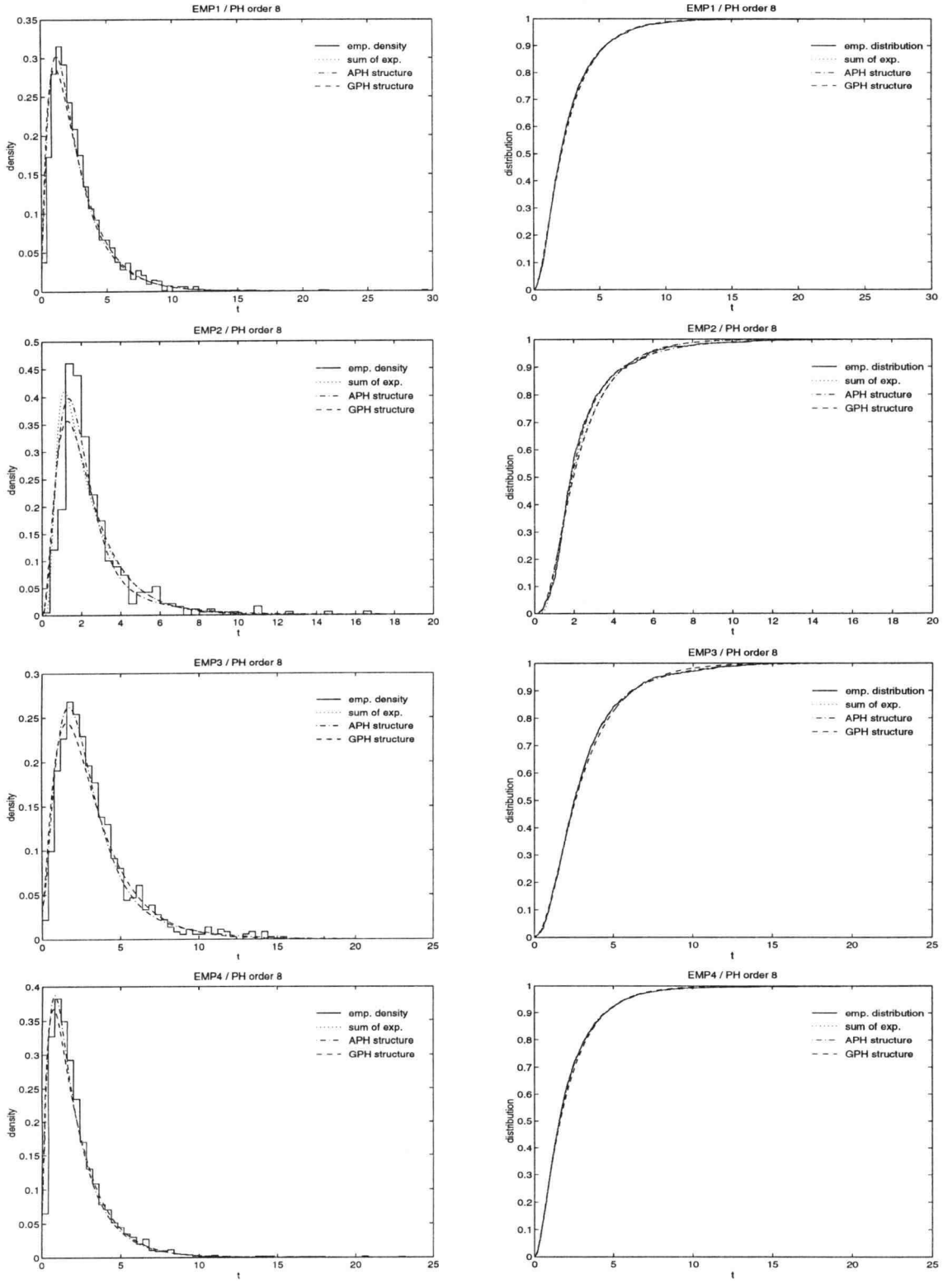

Figure 5.20. EMPHT order-8 approximations to the empirical densities and distributions.

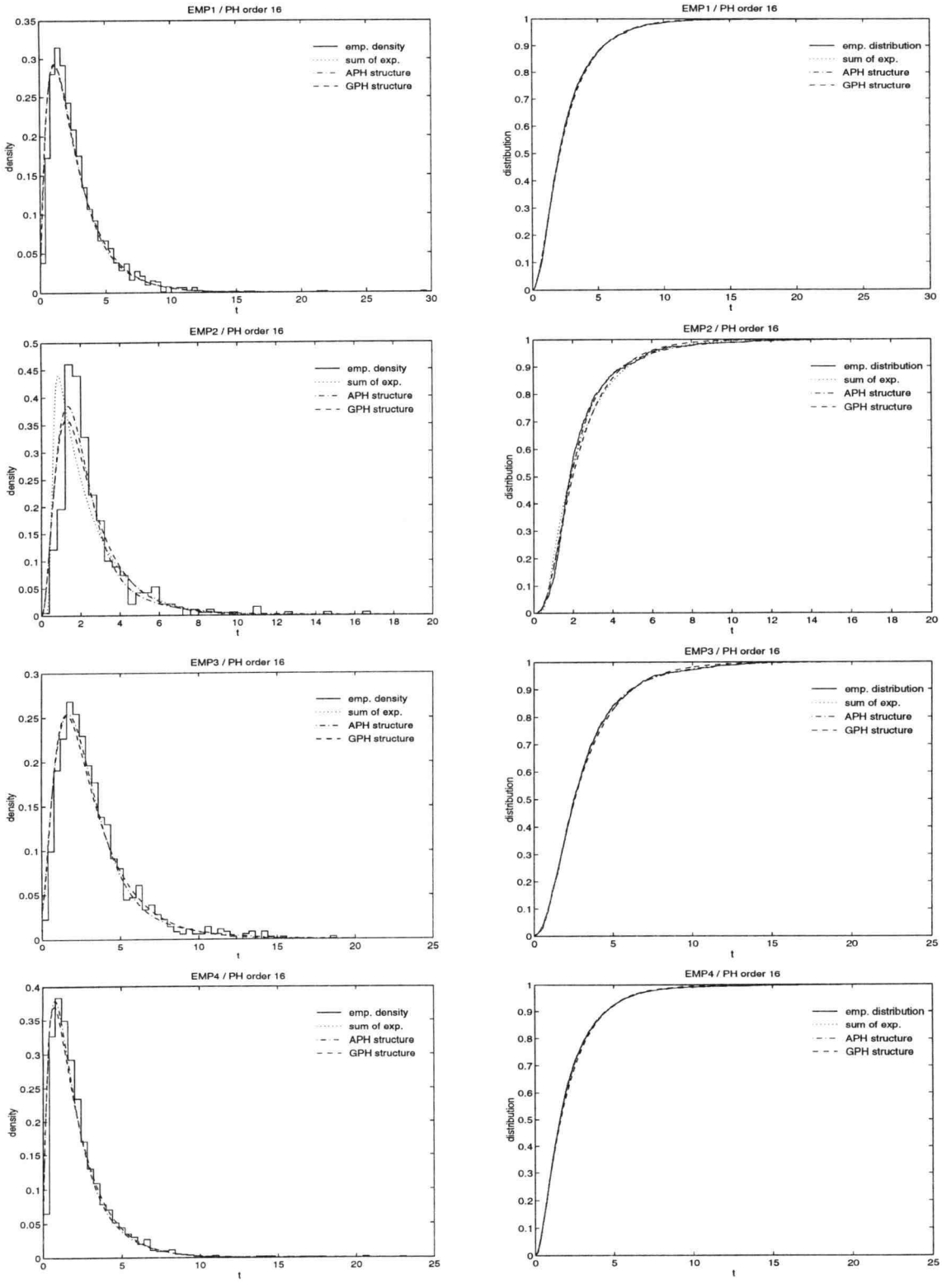

Figure 5.21. EMPHT order-16 approximations to the empirical densities and distributions.

Performance measures: empirical test cases

Tables 5.24, 5.25 and 5.26 present the cdf area differences as well as the errors of the first three central moments between the empirical test distributions ($EMP1$-$EMP4$) and the phase-type approximations involving the sum of exponentials, acyclic and general phase-type structure, respectively.

The results from EMPHT acyclic phase-type approximations to the empirical data sets $EMP1$-$EMP4$ (see table 5.25) show that, in general, order-16 fits are significantly better than lower order approximations, except test case $EMP4$, where the order-4 acyclic fit does not improve with higher orders in terms of the relative moment errors.

This behavior is even more pronounced in EMPHT general phase-type approximations to the empirical data sets $EMP1$-$EMP4$ (see table 5.26). Both, cdf area difference and relative moment errors do not improve with higher orders over order-4 fits in test cases $EMP1$ and $EMP4$.

The visual inspection in figures 5.18-5.21 has shown that the acyclic phase-type structure provides better approximations than the sum of exponentials (where available) and the general phase-type structure. This observation coincides with the performance measures presented in tables 5.24, 5.25 and 5.26.

Table 5.24. EMPHT PH approximations (SUMEX structure) to the empirical distributions: cdf area differences and relative moment errors.

		Sum of Exponentials distribution			
Test distribution		order 2	order 4	order 8	order 16
EMP1	Δcdf	0.031951	0.034259		
	e_1	0	0		
	e_2	0.1182042	0.0335666		
	e_3	0.4531701	0.3202896		
EMP2	Δcdf	0.109733	0.080982	0.071528	0.079987
	e_1	0.0000001	0	0	0.0000003
	e_2	0.1429016	0.2537853	0.2311468	0.0909925
	e_3	0.5567281	0.5956204	0.5502358	0.4015796
EMP3	Δcdf	0.046258	0.034888		
	e_1	0	0.0000001		
	e_2	0.0545642	0.0028242		
	e_3	0.2356102	0.0356121		
EMP4	Δcdf	0.037848	0.040228		
	e_1	0.0000001	0		
	e_2	0.1665944	0.1015455		
	e_3	0.5049057	0.4242830		

Table 5.25. EMPHT PH approximations (APH structure) to the empirical distributions: cdf area differences and relative moment errors.

Test distribution		Acyclic Phase-Type distribution order 2	order 4	order 8	order 16
EMP1	Δcdf	0.034096	0.031142	0.010229	0.014473
	e_1	0	0	0	0
	e_2	0.1191118	0.1221158	0.0115793	0.0097219
	e_3	0.4596452	0.4553544	0.1298289	0.0376401
EMP2	Δcdf	0.109675	0.086731	0.031926	0.035688
	e_1	0.0000001	0	0	0
	e_2	0.1420956	0.2448835	0.0245953	0.0276338
	e_3	0.5556835	0.6004092	0.0954063	0.0471766
EMP3	Δcdf	0.050312	0.034496	0.014562	0.014756
	e_1	0	0	0	0
	e_2	0.0479304	0.0781518	0.0084056	0.0148695
	e_3	0.2422833	0.2291088	0.0738255	0.0737751
EMP4	Δcdf	0.041010	0.018331	0.012338	0.012501
	e_1	0.0000001	0	0	0
	e_2	0.1698722	0.0024688	0.0096299	0.0105603
	e_3	0.5132952	0.0422158	0.0973918	0.0527240

Table 5.26. EMPHT PH approximations (GPH structure) to the empirical distributions: cdf area differences and relative moment errors.

		General Phase-Type distribution			
Test distributions		order 2	order 4	order 8	order 16
EMP1	Δcdf	0.034097	0.026206	0.032897	0.028563
	e_1	0	0	0	0
	e_2	0.1196705	0.0776044	0.1240156	0.1217252
	e_3	0.4602444	0.3827211	0.4614631	0.4484587
EMP2	Δcdf	0.109640	0.085494	0.083674	0.083936
	e_1	0.0000003	0.0000001	0	0
	e_2	0.1416447	0.2350884	0.2409331	0.2518900
	e_3	0.5550894	0.5875040	0.5900672	0.6007408
EMP3	Δcdf	0.050290	0.048002	0.036398	0.030656
	e_1	0	0	0	0
	e_2	0.0504284	0.0643572	0.0876054	0.0800173
	e_3	0.2469997	0.2610495	0.2526335	0.2152230
EMP4	Δcdf	0.041415	0.019806	0.040754	0.039673
	e_1	0.0000001	0.0000001	0.0000001	0
	e_2	0.1727563	0.0444912	0.1685668	0.1721446
	e_3	0.5165841	0.2261944	0.5116540	0.5127965

5.5 Comparison

This section presents a comparison between the moment-matching parameter approximation methods MEFIT and MEDA in subsection 5.5.1 and between the maximum-likelihood based approaches MLAPH and EMPHT in subsection 5.5.2.

5.5.1 Moment Matching Methods

Both moment-matching methods (MEFIT and MEDA) have shown that adequate phase-type approximations to a variety of *phase-type behaved* empirical distributions can be obtained. For *non-phase-type behaved* empirical distribution functions (i.e. sharp jumps in the pdf, low variability, long tails), these approximations are determined at the price of high orders of the resulting phase-type distribution. Both parameter approximation methods use the phase-type subclass of *mixture of Erlang* distributions as the selection subset.

A general observation between the results obtained from MEFIT and MEDA is that MEDA approximations generally estimate higher order mixtures of Erlang distributions than MEFIT. However, the area difference and entropy measures show that MEDA fits are generally better than MEFIT approximations. The higher order fits resulting from MEDA are mainly due to the fact that the orders of the individual Erlang branches in the mixture are increased as long as the objective function criterion (cdf area difference between the empirical and approximating distribution) can be improved. This is the case for empirical distributions with low variability (such as $L3$) or sharp jumps in the pdf or cdf (such as $U1$, $U2$, ME and SE). MEFIT, however, does not estimate the orders of the individual Erlang branches, but requires the user to specify them for a particular fitting run.

In terms of run times, MEFIT approximations require generally less time than MEDA runs. However, this comparison is based on approximation runs carried out on different computing environments (MEFIT runs on SPARC-IPX workstations with 40 MHz clockspeed, MEDA runs on a 80486 PC with 33 MHz clockspeed).

Furthermore, MEDA approximations require only one run to obtain a solution, since it does not depend heavily on user-specifications. MEFIT offers much more flexibility in terms of user-specified input parameters guiding the parameter search. Our experience showed that usually several parameter approximation runs are required to obtain satisfactory solutions, which limits its use in terms of user-friendliness.

5.5.2 Maximum Likelihood Methods

Both maximum-likelihood based parameter approximation methods (MLAPH and EMPHT) have shown that satisfactory low-order approximations to *phase-type behaved* empirical distribution functions can be obtained. Empirical distributions exhibiting long tails, low variability or sharp jumps in their density functions present challenges with which both MLAPH and EMPHT struggle. In particular, the high order phase-type distributions needed to adequately approximate such distributions can not be fit with either of the two maximum-likelihood based methods because of numerical instabilities and/or prohibitively long run times. MLAPH approximates the parameters of the class of *acyclic* phase-type distributions (see section 3.3.1 and experimental results in section 5.3). EMPHT offers a variety of subclasses of the family of phase-type distributions as the selection subset (see section 3.3.2 and experimental results in section 5.4), including the *general* phase-type distribution.

In terms of run time requirements, we found that EMPHT parameter approximation runs generally take a significantly longer time than MLAPH runs. In a few cases, it was impossible to obtain EMPHT solutions at all, even after run times of up to a month. This comparison is based on a maximum of 200 iterations for each run of the algorithms. To provide a starting point for both the algorithms, randomly generated initial phase-type parameter values are generated, based on a user-specified seed. There is evidence that different seeds, i.e. different starting points, lead to different final solutions in both methods, and better starting points might require less iterations. The run times are based on the computational re-

quirements at each iteration and therefore could be reduced by a smaller number of iterations. The experimental results based on MLAPH and EMPHT show, however, that the maximum number of 200 iterations was generally carried out, even for phase-type approximations of moderate order, and it is not expected that the number of iterations can be significantly reduced by finding better initial parameter values. The computational results based on EMPHT also show evidence that the maximum of 200 iterations was not enough to obtain satisfactory phase-type approximations. Asmussen et. al. [6] report up to 10000 iterations necessary for the results reported in their paper.

Regarding the pdf and cdf area differences and entropy measures, we note that MLAPH approximations are generally better than EMPHT results involving the sum of exponentials structure. Comparing the MLAPH results to the EMPHT acyclic structure, the results do not differ to a great degree. A slight superiority of EMPHT order-8 acyclic phase-type approximations can be noted based on the four empirical data sets ($EMP1 - EMP4$). MLAPH approximations generally perform better than EMPHT general phase-type approximations, except in test cases $L3$ and $U2$, where EMPHT order-8 general phase-type approximations gain slight superiority over MLAPH order-8 fits.

In terms of the relative moment errors, similar results are found. The MLAPH approximations perform generally better than EMPHT sum of exponentials and general phase-type approximations. The EMPHT acyclic structure approximations, however, compare well to the MLAPH results. In several cases, EMPHT acyclic order-8 approximations perform even better than MLAPH fits.

These observations lead to the conclusion that both MLAPH and EMPHT acyclic phase-type structure approximations of moderate order (order 4 or order 8) yield satisfactory results in terms of matching the shape of *phase-type behaved* empirical distribution functions. Distributions exhibiting multiple modes and sharp jumps in their pdf or cdf functions are generally not well approximated by any of the phase-type distributions obtained in our experiments. The second and third mo-

ments of the empirical distributions are generally not adequately approximated by their corresponding phase-type moments. Neither the EMPHT sum of exponential structure nor the general phase-type structure has shown to provide better results than the acyclic phase-type structure.

Chapter 6

Conclusions and Recommendations

Phase-type distributions present an important tool in applied stochastic modeling. The development of *matrix-analytic methods* for the solution of very general questions in queueing theory has shown that phase-type distributions are of great significance in this field.

A variety of methods for approximating the parameters of phase-type distributions is currently available and sufficiently well developed to obtain phase-type approximations of moderate dimensions. However, the need to restrict the approximating distribution to subsets of general phase-type distributions, in conjunction with the numerical difficulties inherent to the fitting problem, have deterred the development of a standard approximation method for phase-type distributions.

Based on the experimental results of the phase-type parameter approximation methods MEFIT, MEDA, MLAPH and EMPHT, presented in sections 5.1-5.4, respectively, the following general observations have been obtained.

The run times required to obtain phase-type approximations by the analyzed parameter approximation methods vary, dependent on the difficulty of the fitting problem, in particular: the number of parameters to be approximated, the number of observations in the empirical data set, and the degree of non-phase-type behavior of the empirical distribution (sharp jumps, multimodality, low variability, long tails). In general, it was found that MEFIT tends to be the fastest approximation algorithm, followed by MLAPH. MEFIT, however, usually requires several approximation runs dependent on the user-specified input parameters. MEDA approximation runs seemed to take somewhat longer than MEFIT, but the computing

environment was different (MEDA runs on a MSDOS 80486 PC with 33 MHz clockspeed, all other methods on SPARC-IPX workstations with 40 MHz clockspeed). EMPHT approximates low order phase-type distributions (e.g. order 2 or order 4) generally in a reasonable run time, but higher order approximations take an unreasonable amount of time even when fitting phase-type structures involving a small number of parameters. This is attributable to the low convergence rate of the EM algorithm.

In terms of stability and reliability of the analyzed parameter approximation algorithms, we note that none was free of problems, particularly when a more challenging distribution was to be approximated. MEFIT depends heavily on user specifications guiding the parameter approximation runs and might not be able to obtain a satisfactory solution when these specifications are poorly chosen. MEDA does not rely as greatly on user-specified parameters, but is sometimes troubled by numerical instabilities inherent to the fitting algorithm. The MLAPH and EMPHT parameter approximation algorithms generate random initial phase-type parameter values as the starting point. Dependent on this starting point, we note that problems sometimes occur, either with the inability to obtain solutions or with abnormal program termination. Moreover, solutions obtained by MLAPH and EMPHT might differ significantly, dependent on the initial starting point. In addition, EMPHT showed problems with a particular phase-type structure (the sum of exponentials or generalized Erlang distribution) where no solutions could be obtained in a few of our test cases.

The experimental results based on the four analyzed methods have shown that satisfactory phase-type approximations can be obtained, provided the empirical distributions to be fit do not exhibit properties such as very long tails, low variability and sharp jumps in their density functions. Moreover, the phase-type structures inherent to the fitting methods MEFIT, MEDA and MLAPH, and those chosen for this analysis for EMPHT, were not particularly well-suited to recover the shape of distribution functions with multiple modes.

The moment-matching methods MEFIT and MEDA obtained adequate low-order phase-type approximations for the non-challenging test cases ($W1$ and $L2$), both fitting the shape of the empirical distribution functions and matching the first three empirical moments. Satisfactory phase-type approximations for these test cases were also obtained by the maximum-likelihood based methods MLAPH and EMPHT (acyclic structure) with the area difference and entropy measures in some cases even superior to MEFIT and MEDA, but the second and third moments are not adequately matched.

For the test cases with long tails ($W2$ and $L1$), MEFIT as well as MEDA obtained low order approximations while still matching the first three empirical moments. MLAPH and EMPHT (acyclic structure) shape approximations are comparable to MEFIT and MEDA, but the relative errors on the second and third moment are large.

The test cases exhibiting low variability ($L3$ and $U2$) were adequately approximated by MEFIT and MEDA, but at the price of very high orders of the approximating phase-type distributions. MLAPH and EMPHT lower order approximations neither achieved a satisfactory shape approximation nor a close moment matching in these test cases.

In terms of density shape fitting, the most challenging test cases were $U1$, $U2$, SE (sharp jumps) and ME (multiple modes). None of the four parameter approximation methods was able to achieve adequate fits, which is based on the inability of phase-type distributions to exhibit such pdf shapes. MEFIT approximations in these test cases were generally found with medium order (between order 8 and order 54, except ME fit with order 145) phase-type distributions. MEDA approximated higher order fits for test cases $U1$ and $U2$ (between order 27 and order 230), but was able to approximate medium orders for test cases ME and SE (between orders 5 and 17). Due to these phase-type orders, MEFIT and MEDA were able to match the first three empirical moments and generally obtain similar or better shape approximations than MLAPH and EMPHT.

In order to establish a homogeneous comparison between different parameter approximation methods for phase-type distributions, this analysis has applied a benchmark test partially developed by researchers in the field. In addition to the numerical results obtained, general issues, such as: (i) the order of the approximating phase-type distributions; (ii) the structure and sparsity of the underlying Markov processes (selection subsets); and (iii) the run time requirements for the individual parameter approximation algorithms were investigated.

This analysis leads to the conclusion that *phase-type behaved* distribution functions can adequately be approximated by a variety of subsets of phase-type distributions, such as the *mixture of Erlang* and the *acyclic phase-type distribution (APH)*. The moment-matching parameter approximation methods based on the mixture of Erlang distributions provide phase-type fits with a simple and sparse structure of their underlying Markov processes together with a very effective numerical implementation. Furthermore, these methods match the first three empirical moments up to a user-specified tolerance level. In some cases, however, satisfactory approximations are obtained only at the price of high orders of the approximated Erlang mixture (large number of transient states of the underlying Markov process). The maximum-likelihood based parameter approximation methods seemed to be most effective based on the class of acyclic phase-type distributions. This class also provides a simple and sparse structure of their underlying Markov process and satisfactory approximations of low order can be found. Their numerical implementation, however, is very intensive. Moreover, moments of higher order than the mean cannot be matched, which limits their use in applications where results depend heavily on the moments of the distributions involved.

Empirical distributions exhibiting *non-phase-type behavior*, such as long non-exponential tails, low variability or sharp jumps, are very challenging to all the parameter approximation methods considered in this study. Phase-type approximations might be inappropriate in such cases. The question arises, however, regarding the likelihood that such distributions are encountered in practice.

These performance characterictics of the analyzed methods lead to suggestions for future research in the area of parameter approximation for phase-type distributions:

(i) An evaluation of the presented phase-type parameter approximation methods based on random samples from known phase-type distributions. In situations where an underlying Markov process is known, it would be interesting to see how well the parameter approximation methods perform in terms of recovering the original phase-type distribution. Examples of such test cases could include Erlang or hyperexponential distributions with a small number of phases.

(ii) An investigation of alternative selection subsets. Based on discussions with researchers in the area, a possible candidate could be the class of *unicyclic phase-type distributions*, which is similar to the acyclic phase-type distributions, but feedback to previous states is allowed from the last transient state of the underlying Markov process. These distributions exhibit properties such as multimodality, even at low orders.

(iii) A refinement of the present algorithms in terms of robustness and numerical stability. Possible improvements include the use of different NLP algorithms for the solution of the non-linear equations involved and the development of an effective method to generate initial phase-type parameter values. Schmickler's approach [72, 73] might indicate a possibility in this direction.

(iv) The development of a hybrid parameter approximation method, combining the advantages of maximum-likelihood based and moment-matching approaches. Possibilities include using a moment-matching approach and improving the results by maximum-likelihood or using a maximum-likelihood based method with additional moment-matching constraints.

(v) An investigation of the performance of phase-type approximations in applications, such as matrix-analytic methods in stochastic models.

(vi) A statistical perspective is needed for a more complete evaluation of phase-type parameter approximation methods. The quality of the approximated parameters needs to be investigated, either via computation of standard errors or by simulation.

It is hoped that this dissertation has shed some light on the various issues involved in phase-type parameter approximation and will lead to further research in this area. All that remains to be done is the work.

BIBLIOGRAPHY

Be admonished: of making many books there is no end;
but their study wearies the flesh.
Ecclesiastes, Ch. 12, v. 12.

[1] ALBERT, A., Estimating the Infinitesimal Generator of a Continuous Time, Finite State Markov Process, *Annals of Mathematical Statistics*, **38**, (1961), 727-753.

[2] ALDOUS, D. and SHEPP, L., The Least Variable Phase Type Distribution is Erlang, *Stochastic Models*, **3**, (1987), 467-473.

[3] ALTIOK, T., On the Phase-Type Approximations of General Distributions, *IIE Transactions*, **17:2**, (1985), 110-116.

[4] ASMUSSEN, S., *Applied Probability and Queues*, New York: John Wiley & Sons, 1987.

[5] ASMUSSEN, S. and NERMAN, O., Fitting Phase-Type Distributions via the EM Algorithm, K. Vest Nielsen (Ed.), in *Proceedings of Symposium i Anvendt Statistik*, Copenhagen, Denmark (January 21-23, 1991).

[6] ASMUSSEN, S., NERMAN, O. and OLSSON, M., Fitting Phase-Type Distributions via the EM Algorithm, Technical Report No. 1994:23/ISSN 0347-2809, Department of Mathematics, Chalmers University of Göteborg, Sweden, 1994.

[7] ASSAF, D. and LEVIKSON, B., Closure of Phase Type Distributions under Operations Arising in Reliability Theory, *The Annals of Probability*, **10**, (1982), 265-269.

[8] BASAWA, I.V. and PRAKASA RAO, B.L.S., *Statistical Inference for Stochastic Processes*, London: Academic Press, 1980.

[9] BOBBIO, A., CUMANI, A., PREMOLI, A. and SARACCO, O., Modelling and Identification of Non-Exponential Distributions by Homogegeous Markov Processes, in *Proceedings of the 6^{th} Advances in Reliability Technology Symposium*, **1**, Bradford (UK) (April 9-11, 1980), 373-392.

[10] BOBBIO, A. and CUMANI, A., Markov Models: A New Class of Distributions for the Analysis of Lifetime Data Samples, in *Proceedings of the 4^{th} Euredata Conference*, Paper 10.3/1-13, Venice (Italy) (March 23-25, 1983).

[11] BOBBIO, A. and CUMANI, A., ML Estimation of the Parameters of a PH Distribution in Triangular Canonical Form, *Computer Performance Evaluation*, G. Balbo and G. Serazzi (Eds.), 33-46, Elsevier Science Publishers (1992).

[12] BOBBIO, A. and TELEK, M., Parameter Estimation of Phase Type Distributions, Technical Report R.T. 423, Istituto Elettrotecnico Nazionale Galileo Ferraris (Italy) (August 1992).

[13] BOBBIO, A. and TELEK, M., A Benchmark for PH Estimation Algorithms: Results for Acyclic-PH, to appear in *Stochastic Models*, **10:3**, (1994).

[14] BOTTA, R.F., HARRIS, C.M. and MARCHAL, W.G., Characterizations of Generalized Hyperexponential Distribution Functions, *Stochastic Models*, **3**, (1987), 115-148.

[15] BUX, W. and HERZOG, U., Approximation von Verteilungsfunktionen: Ein wichtiger Schritt bei der Modellbildung für Rechensysteme, Workshop über Modelle für Rechensysteme, Institut für Informatik, University of Bonn, Bonn, Germany, March 1977.

[16] BUX, W. and HERZOG, U., The Phase Concept: Approximation of Measured Data and Performance Analysis, in *Computer Performance*, K.M. Chandy and M. Reiser (Eds.), Proceedings of the International Symposium on Computer Performance Modeling, Measurement and Evaluation (August 16-18, 1977), 23-38, North-Holland, 1977.

[17] COX, D.R., A Use of Complex Probabilities in the Theory of Stochastic Processes, *Proceedings of the Cambridge Philosophical Society*, **51**, (1955), 313-319.

[18] CUMANI, A., On the Canonical Representation of Homogeneous Markov Processes Modelling Failure-Time Distributions, *Microelectronics and Reliability*, **22:3**, (1982), 583-602.

[19] DEMPSTER, A.P., LAIRD, N.M. and RUBIN, D.B., Maximum Likelihood from Incomplete Data via the EM Algorithm, *Journal of the Royal Statistical Society*, **39**, (1977), 1-38.

[20] ERLANG, A.K., Solution of some Problems in the Theory of Probabilities of Significance in Automatic Telephone Exchanges, *The Post Office Engineer's Journal*, **10**, (1917-18), 189-197.

[21] FADDY, M.J., Fitting Structured Phase-Type Distributions, submitted to *Applied Stochastic Models and Data Analysis*, 1994.

[22] FADDY, M.J., Compartmental Models with Phase-Type Residence Time Distributions, *Applied Stochastic Models and Data Analysis*, **6**, (1990), 121-127.

[23] FADDY, M.J., A Structured Compartmental Model for Drug Kinetics, *Biometrics*, **49**, (1993), 243-248.

[24] GILL, P.E., MURRAY, M.A., SAUNDERS, M.A. and WRIGHT, M.H., User's Guide for NPSOL (Version 4.0): A Fortran Package for Nonlinear Programming, Technical Report SOL 86-2, Department of Operations Research, Stanford University, 1986.

[25] GRAHAM, A., *Kronecker Products and Matrix Calculus with Applications*, Chichester: Ellis Horwood, 1981.

[26] HÄGGSTRÖM, O., ASMUSSEN, S. and NERMAN, O., EMPHT - A Program for Fitting Phase-Type Distributions, Studies in Statistical Quality, Control and Reliability, Mathematical Statistics, Chalmers University of Technology, The University of Göteborg, 1992:4.

[27] HARRIS, C.M., A Note on Mixed Exponential Approximations for *GI/G/1* Waiting Times, *Computers & Operations Research*, **12:3**, (1985), 285-289.

[28] HARRIS, C.M. and SYKES, E.A., Likelihood Estimation for Generalized Mixed Exponential Distributions, *Naval Research Logistics*, **34**, (1987), 251-179.

[29] HARRIS, C.M., MARCHAL, W.G. and BOTTA, R.F., A Note on Generalized Hyperexponential Distributions, *Stochastic Models*, **8**, (1992), 179-191.

[30] JOHNSON, M.A. and TAAFFE, M.F., Matching Moments to Phase Distributions: Mixtures of Erlang Distributions of Common Order, *Stochastic Models*, **5:4**, (1989), 711-743.

[31] JOHNSON, M.A. and TAAFFE, M.F., Matching Moments to Phase Distributions: Nonlinear Programming Approaches, *Stochastic Models*, **6:2**, (1990), 259-281.

[32] JOHNSON, M.A. and TAAFFE, M.F., Matching Moments to Phase Distributions: Density Function Shapes, *Stochastic Models*, **6:2**, (1990), 283-306.

[33] JOHNSON, M.A., User's Guide for MEFIT (version 1.0): A FORTRAN Package for Fitting Mixtures of Erlang Distributions, Working Paper Nr. 90-004, Department of Systems and Industrial Engineering, University of Arizona, 1990.

[34] JOHNSON, M.A. and TAAFFE, M.F., An Investigation of Phase-Distribution Moment-Matching Algorithms for Use in Queueing Models, *Queueing Systems*, **8:2**, (1991), 129-147.

[35] JOHNSON, M.A., Selecting Parameters of Phase Distributions: Combining Nonlinear Programming, Heuristics, and Erlang Distributions, *ORSA Journal on Computing*, **5:1**, (1993), 69-83.

[36] KAYLAN, A.R. and HARRIS, C.M., Efficient Algorithms to Derive Maximum-Likelihood Estimates for Finite Exponential and Weibull Mixtures, *Computers & Operations Research*, **8**, (1981), 97-104.

[37] KHINCHINE, A.YA., Mathematical Theory of a Stationary Queue, *Math. Sbornik*, **39**, (1932), 73-84.

[38] KHOSHGOFTAAR, T. and PERROS, H.G., A Comparison of Three Methods of Estimation for Approximating General Distributions by a Coxian Distribution, in *Proceedings of the 3rd International Workshop on Modelling Techniques and Performance Evaluation*, Paris (1987), 69-77.

[39] KOTZ, S. and JOHNSON, N.L., *Encyclopedia of Statistical Sciences*, New York: John Wiley, 1982.

[40] KULLBACK, S., *Information Theory and Statistics*, Magnolia, Mass.: Peter Smith, 1978.

[41] VAN DE LIEFVOORT, A., The Moment Problem for Continuous Distributions, Working Paper CM-1990-02, Computer Science Telecommunications Program, University of Missouri, Kansas City, (1990).

[42] LIPSKY, L.R., *Queueing Theory - A Linear Algebraic Approach*, New York: Macmillan, 1992.

[43] LUCANTONI, D.M. and RAMASWAMI, V., Efficient Algorithms for solving the Nonlinear Matrix Equation arising in Phase-Type Queues, *Stochastic Models*, **1**, (1985), 29-51.

[44] MADSEN, H. and NIELSEN, B.F., The Use of Phase Type Distributions for Modelling Packet-Switched Traffic, in *Teletraffic and Datatraffic in a Period of Change (Proceedings of the 13th International Teletraffic Congress ITC-13)*, A. Jensen and V.B. Iversen (Eds.), 593-599, North-Holland: Elsevier Science Publishers, 1991.

[45] MAIER, R.S., The Algebraic Construction of Phase-Type Distributions, *Stochastic Models*, **7**, (1991), 573-602.

[46] MALHOTRA, M. and REIBMANN, A., Selecting and Implementing Phase Approximations for Semi-Markov Models, *Stochastic Models*, **9:4**, (1993), 473-506.

[47] MANDELBAUM, J. and HARRIS, C.M., Parameter Estimation under Progressive Censoring Conditions for a Finite Mixture of Weibull Distributions, *TIMS Studies in the Management Sciences*, **19**, (1982), 239-260.

[48] MARIE, R., Calculating Equilibrium Probabilities for $\lambda(n)/C_k/1/N$ Queues, in *Performance '80*, Proceedings of the International Symposium on Computer Performance Modelling, May 28-30, 1980, 117-125.

[49] THE MATHWORKS, INC., MATLAB User's Guide, 1993.

[50] NELDER, J.A. and MEAD, R., A Simplex Method for Function Minimization, *Computer Journal*, **7**, (1965), 308-313.

[51] NEUTS, M.F., *Matrix-geometric Solutions in Stochastic Models: An Algorithmic Approach*, Baltimore: The Johns Hopkins University Press, 1981.

[52] NEUTS, M.F., *Structured Stochastic Matrices of M/G/1 Type and their Applications*, New York: Marcel Dekker, 1989.

[53] NEUTS, M.F., Probability Distributions of Phase Type, in *Liber Amicorum Prof. Emeritus H. Florin*, Department of Mathematics, University of Louvain (Belgium) (1975), 173-206.

[54] NEUTS, M.F., Phase-Type Distributions: A Bibliography, Department of Systems and Industrial Engineering, University of Arizona, Working Paper Nr. 89-005, February 1989.

[55] NEUTS, M.F., Matrix-Analytic Methods in Queueing Theory, *European Journal of Operational Research*, **15**, (1984), 2-12.

[56] NEUTS, M.F., Computational Uses of the Method of Phases in the Theory of Queues, *Computers Math. Appl.*, **1**, (1975), 151-166.

[57] NEUTS, M.F., Renewal Processes of Phase Type, *Naval Research Logistics Quarterly*, **25**, (1978), 445-454.

[58] NEUTS, M.F., Markov Chains with Applications in Queueing Theory, which have a Matrix-Geometric Invariant Probability Vector, *Advances in Applied Probability*, **10**, (1978), 185-212.

[59] NIELSEN, B.F., Modelling of Multiple Access Systems with Phase Type Distributions, Ph.D. Dissertation No. 49, Institute of Mathematical Statistics and Operations Research (IMSOR), Technical University of Denmark, Lyngby, 1988.

[60] O'CINNEIDE, C.A., On Non-Uniqueness of Representations of Phase-Type Distributions, *Stochastic Models*, **5:2**, (1989), 247-259.

[61] O'CINNEIDE, C.A., Characterization of Phase-Type Distributions, *Stochastic Models*, **6:1**, (1990), 1-57.

[62] O'CINNEIDE, C.A., Phase-Type Distributions and Invariant Polytopes, *Adv. Appl. Prob*, **23**, (1991), 515-535.

[63] O'CINNEIDE, C.A., Triangular Order of Triangular Phase-Type Distributions, *Stochastic Models*, **9:4**, (1993), 507-529.

[64] OLSSON, M., Estimation of Phase Type Distributions from Censored Samples, Technical Report No. 1993:36/ISSN 0347-2809, Department of Mathematics, Chalmers University of Technology, Göteborg, Sweden, 1993.

[65] RAMASWAMI, V. and LUCANTONI, D.M., Stationary Waiting Time Distributions in Queues with Phase-Type Service and in Quasi-Birth-and-Death Processes, *Stochastic Models*, **1**, (1985), 125-136.

[66] REDNER, R.A. and WALKER, H.F., Mixture Densities, Maximum Likelihood and the EM Algorithm, *SIAM Review*, **26:2**, (1984), 195-239.

[67] RYDÉN, T., On Identifiability and Order Of Phase-Type Distributions, Working Paper, Department of Mathematical Statistics, Lund Institute of Technology, Lund, Sweden.

[68] SAUER, C.H. and CHANDY, K.M., Approximate Analysis of Central Server Models, *IBM Journal of Research and Development*, **19**, (1975), 301-312.

[69] SAUER, C.H. and CHANDY, K.M., *Computer Systems Performance Modeling*, Englewood Cliffs, NJ: Prentice-Hall, (1981).

[70] SEVCIK, K.C., LEVY, A.I., TRIPATHI, S.K. and ZAHORJAN, J.L., Improving Approximations of Aggregated Queueing Network Subsystems, in *Computer Performance*, K.M. Chandy and M. Reiser (Eds.), Proceedings of the International Symposium on Computer Performance Modeling, Measurement and Evaluation (August 16-18, 1977), 1-22, North-Holland, 1977.

[71] SCHMICKLER, L., Approximation von empirischen Verteilungsfunktionen mit Erlangmischverteilungen und Coxverteilungen, *Proceedings of 4. GI/ITG-Fachtagung*, Erlangen (September/October 1987), *Informatik Fachberichte* **154**, U. Herzog and M. Paterock (Eds.), Springer-Verlag.

[72] SCHMICKLER, L., Erweiterung des Verfahrens MEDA zur analytischen Beschreibung empirischer Verteilungsfunktionen, *Proceedings of 5. GI/ITG-Fachtagung*, Braunschweig (September 1989), *Informatik Fachberichte* **218**, G. Stiege and J.S. Lie (Eds.), Springer-Verlag.

[73] SCHMICKLER, L., MEDA: Mixed Erlang Distributions as Phase-Type Representations of Empirical Distribution Functions, *Stochastic Models*, **8:1**, (1992), 131-156.

[74] SCHMICKLER, L. and SÜDHOFEN, H.-D., Representation of Data Network Measurement Results by the Mixed Erlang Approximation Method MEDA, in *Teletraffic Science for New Cost-Effective Systems, Networks, and Services: ITC-12*, M. Bonatti (Ed.) Proceedings of the 12^{th} International Teletraffic Conference, June 1-8, 1989, 239-245.

[75] TIJMS, H.C., *Stochastic Modelling and Analysis: A Computational Approach*, New York: John Wiley & Sons, 1986.

[76] WHITT, W., Approximating a Point Process by a Renewal Process, I: Two Basic Methods, *Operations Research*, **30:1**, (1982), 125-147.

[77] WHITT, W., The Queueing Network Analyzer, *The Bell System Technical Journal*, **62:9**, (1983), 2779-2815.

APPENDICES

Appendix A

Related Literature

[1] ASMUSSEN, S., Exponential Families Generated by Phase-Type Distributions and other Markov Lifetimes, *Scandinavian Journal of Statistics*, **16**, (1989), 319-334.

[2] ASMUSSEN, S. and RAMASWAMI, V., Probabilistic Interpretations of Some Duality Results for the Matrix Paradigms in Queueing Theory, *Stochastic Models*, **6**, (1990), 715-733.

[3] ASMUSSEN, S. and ROLSKI, T., Computational Methods in Risk Theory: A Matrix-Algorithmic Approach, *Insurance: Mathematics and Economics*, **10**, (1991), 259-274.

[4] BAUM, L.E., PETRIE, T., SOULES, G. and WEISS, N., A Maximization Technique occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains, *Ann. Math. Statist.*, **41**, (1970), 164-171.

[5] BOBBIO, A. and CUMANI, A., A Markov Approach to Wear-Out Modelling, *Microelectronics and Reliability*, **23:1**, (1983), 113-119.

[6] BOBBIO, A. and TRIVEDI, K.S., Computation of the Distribution of the Completion Time when the Work Requirement is a PH Random Variable, *Stochastic Models*, **6:1**, (1990), 133-150.

[7] BOBBIO, A. and TELEK, M., A Benchmark for PH Estimation Algorithms: Results for Acyclic-PH, in *Proceedings of the 20th European Meeting of Statisticians*, Bath (UK) (September 14-18, 1992).

[8] DEHON, M. and LATOUCHE, G., A Geometric Interpretation of the Relations between the Exponential and the Generalized Erlang Distributions, *Advances in Applied Probability*, **14**, (1982), 885-897.

[9] DAIGLE, J.N. and LUCANTONI, D.M., Queueing Systems having Phase-Dependent Arrival and Service Rates, in *Numerical Solution of Markov Chains*, W.J. Stewart (Ed.), New York: Marcel Dekker Inc., 1991.

[10] DEVROYE, L. and GYÖRFI, L., *Non-Parametric Density Estimation. The L_1 View.* New York: Wiley, 1985.

[11] GÜN, L., Experimental Results of Matrix-Analytical Solution Techniques - Extensions and Comparisons, *Stochastic Models*, **5**, (1989), 669-682.

[12] HASSELBLAD, V., Estimation of Finite Mixtures of Distributions from the Exponential Family, *Journal of the American Statistical Association*, **64:328**, (1966), 1459-1471.

[13] JAMSHIDIAN, M. and JENNRICH, R.I., Conjugate Gradient Acceleration of the EM Algorithm, *Journal of the American Statistical Association*, **88**, (1993), 221-228.

[14] JOHNSON, M.A. and TAAFFE, M.F., A Graphical Investigation of Error Bounds for Moment-Based Queueing Approximations, *Queueing Systems*, **8:3**, (1991), 295-312.

[15] JOHNSON, M.A., An Empirical Study of Queueing Approximations Based on Phase-Type Distributions, *Stochastic Models*, **9:4**, (1993), 531-561.

[16] KAMMLER, D.W., Least Squares Approximation of Completely Monotonic Functions by Sums of Exponentials, *SIAM Journal on Numerical Analysis*, **16**, (1979), 801-818.

[17] KAO, E.P.C., Computing the Phase-Type Renewal and Related Functions, *Technometrics*, **30**, (1988), 87-93.

[18] KLINCEWICZ, J.G. and WHITT, W., On Approximations for Queues, II: Shape Constraints, *AT& T Bell Laboratories Technical Journal*, **63**, (1984), 139-161.

[19] KRIEGER, U. and WAGNER, D., Analysis and Application of an $MMPP/PH/n/m$ Multi-Server Model, Technical Report, Technische Hochschule Darmstadt, Germany.

[20] LATOUCHE, G., A Note on Two Matrices Occuring in the Solution of Quasi-Birth-and-Death Processes, *Stochastic Models*, **3**, (1987), 251-257.

[21] LATOUCHE, G., Comparison of Algorithms for Infinite Markov Chains with Repeating Columns, Technical Report Nr. 91/1, Université Libre de Bruxelles, Belgium, Séminaire de Théorie des Probabilités, (March 1992).

[22] LATOUCHE, G., Algorithms for Evaluating the Matrix G in Markov Chains of $PH/G/1$ Type, Bellcore, Technical Report, (1992).

[23] LATOUCHE, G. and RAMASWAMI, V., A Logarithmic Reduction Algorithm for Quasi Birth and Death Processes, Bellcore, Technical Report, (1993).

[24] LATOUCHE, G., Algorithms for Evaluating the Matrix G in Markov Chains of $PH/G/1$ Type, Bellcore, Technical Report, 1992.

[25] LATOUCHE, G., Newton's Iterations for Nonlinear Equations in Markov Chains, Bellcore, Technical Report, 1992.

[26] LEROUX, B.G., Maximum-Likelihood Estimation for Hidden Markov Chains, *Stochastic Processes and Applications*, **40**, (1992), 127-143.

[27] LIPSKY, L.R. and FANG, Z., Classification of Functions with Rational Laplace Transforms, in *Proceedings of the Summer Computer Simulation Conference*, July 1986, Las Vegas.

[28] LOUIS, T.A., Finding the Observed Information Matrix when using the EM Algorithm, *Journal of the Royal Statistical Society*, **B 44:2**, (1982), 226-233.

[29] LUCANTONI, D.M., MEIER-HELLSTERN, K.S. and NEUTS, M.F., A Single Server Queue with Server Vacations and a Class of Non-Renewal Arrival Processes, *Advances in Applied Probability*, **22**, (1990), 676-705.

[30] LUCANTONI, D.M. New Results on the Single Server Queue with a Batch Markovian Arrival Process, *Stochastic Models*, **7:1**, (1991), 1-46.

[31] LUCANTONI, D.M. The $BMAP/G/1$ Queue: A Tutorial, Working Paper, AT& T Bell Laboratories, (1993).

[32] MEILIJSON, I., A Fast Improvement to the EM Algorithm on its own Terms, *Journal of the Royal Statistical Society*, **B 51:1**, (1989), 127-138.

[33] MOLER, C. and VAN LOAN, C., Nineteen Dubious Ways to Compute the Exponential of a Matrix, *SIAM Review*, **20**, (1978), 801-836.

[34] NEUTS, M.F., A Versatile Markovian Point Process, *Journal of Applied Probability*, **16**, (1979), 764-779.

[35] NEUTS, M.F., The Probabilistic Significance of the Rate Matrix in Matrix-Geometric Invariant Vectors, *Journal of Applied Probability*, **17**, (1980), 291-296.

[36] NEUTS, M.F., The Caudal Characteristic Curve of Queues, *Adv. Appl. Prob*, **18**, (1986), 221-254.

[37] NEUTS, M.F., Models based on the Markovian Arrival Process, *IEICE Trans. Commun.*, **E75-B:12**, (1992), 1255-1265.

[38] RAMASWAMI, V., Non-Linear Matrix Equations in Applied Probability: Solution Techniques and Open Problems, *SIAM Review*, **30**, (1988), 256-263.

[39] RAMASWAMI, V., A Stable Recursion for the Steady-State Vector in Markov Chains of $M/G/1$ Type, *Stochastic Models*, **4**, (1988), 183-188.

[40] RAMASWAMI, V., From the Matrix-Geometric to the Matrix-Exponential, *Queueing Systems Theory Appl.*, **6**, (1990), 229-260.

[41] RAMASWAMI, V., A Duality Theorem for the Matrix Paradigms in Queueing Theory, *Stochastic Models*, **6**, (1990), 151-161.

[42] RAMASWAMI, V. and LATOUCHE, G., A General Class of Markov Processes with Explicit Matrix-Geometric Solutions, *OR Spektrum*, **8**, (1986), 209-218.

[43] RAMASWAMI, V. and LATOUCHE, G., An Experimental Evaluation of the Matrix-Geometric Method for the $GI/PH/1$ Queue, *Stochastic Models*, **5**, (1989), 629-667.

[44] RAMASWAMI, V. and LUCANTONI, D.M., Moments of the Stationary Waiting Time Distribution in the $GI/PH/1$ Queue, *Journal of Applied Probability*, **25**, (1988), 636-641.

[45] RUHE, A., Fitting Empirical Data by Positive Sums of Exponentials, *SIAM Journal on Scientific and Statistical Computing*, **1**, (1980), 481-498.

[46] RYDÉN, T., Estimating the Order of Hidden Markov Models, Working Paper, Department of Mathematical Statistics, Lund Institute of Technology, Lund, Sweden.

[47] RYDÉN, T., Parameter Estimation for Markov Modulated Poisson Processes, Technical Report Nr. 1992:9, Department of Mathematical Statistics, Lund Institute of Technology, Lund, Sweden.

[48] RYDÉN, T., A Note on Parameter Estimation for Multiple State Markov Modulated Poisson Processes, Technical Report Nr. 1992:12, Department of Mathematical Statistics, Lund Institute of Technology, Lund, Sweden.

[49] RYDÉN, T., An EM Algorithm for Multiple State MMPP Parameter Estimation, Technical Report Nr. 1993:2, Department of Mathematical Statistics, Lund Institute of Technology, Lund, Sweden.

[50] RYDÉN, T., Consistent and Asymptotically Normal Parameter Estimates for Markov Modulated Poisson Processes, Technical Report Nr. 1993:5, Department of Mathematical Statistics, Lund Institute of Technology, Lund, Sweden.

[51] RYDÉN, T., Consistent and Asymptotically Normal Parameter Estimates for Hidden Markov Models, Technical Report Nr. 1993:8, Department of Mathematical Statistics, Lund Institute of Technology, Lund, Sweden.

[52] SCHELLHAAS, H., On Ramaswami's Algorithm for the Computation of the Steady State Vector in Markov Chains of $M/G/1$ Type, *Stochastic Models*, **6**, (1990), 541-550.

[53] SCHELLHAAS, H., Single Server Queues with a Batch Markovian Arrival Process and Server Vacations, Technical Report, Preprint Nr. 1566, Technische Hochschule Darmstadt, Germany, (May 1993).

[54] SENGUPTA, B., Markov Processes whose Steady-State Distribution is Matrix-Exponential with an Application to the $GI/PH/1$ Queue, *Advances in Applied Probability*, **21**, (1989), 159-180.

[55] SENGUPTA, B., Phase-Type Representations for Matrix-Geometric Solutions, *Stochastic Models*, **6**, (1990), 163-167.

[56] SENGUPTA, B., The Semi-Markov Queue: Theory and Applications, *Stochastic Models*, **6**, (1990), 383-413.

[57] SILVERMAN, B.W., *Density Estimation for Statistics and Data Analysis*, New York: Chapman and Hall, (1986).

[58] SINGH, C., BILLINGTON, R. and LEE, S., The Method of Stages for Non-Markovian Models, *IEEE Transactions on Reliability*, **R-26**, (1977), 135-137.

[59] VARAH, J.M., On Fitting Exponentials by Nonlinear Least Squares, *SIAM Journal on Scientific and Statistical Computing*, **6**, (1985), 30-44.

[60] WALLACE, V., The Solution of Quasi Birth and Death Processes Arising from Multiple Access Computer Systems, Ph.D. Dissertation, Systems Engineering Laboratory, University of Michigan, Technical Report No. 07742-6-T, 1969.

[61] WHITT, W., Approximating a Point Process by a Renewal Process: The View through a Queue, An Indirect Approach, *Management Science*, **27:6**, (1981), 619-636.

[62] WHITT, W., On Approximations for Queues, I: Extremal Distributions, *AT& T Bell Laboratories Technical Journal*, **63**, (1984), 115-138.

[63] WHITT, W., On Approximations for Queues, III: Mixtures of Exponential Distributions, *AT& T Bell Laboratories Technical Journal*, **63**, (1984), 163-175.

[64] WU, C.F.J., On the Convergence Properties of the EM Algorithm, *Ann. Statist.*, **11**, (1983), 95-103.

Appendix B

Phase-type representation of an Erlang mixture

Both moment-matching methods, MEFIT and MEDA, are approximating the parameters of mixtures of Erlang distributions. This appendix shows how mixtures of Erlang distributions are represented in phase-type notation.

The density $f_{mE}(x)$ of a mixture of r Erlang distributions E_{k_i}, $i = 1, \ldots, r$ with mixing probabilities π_i ($\sum_{i=1}^{r} \pi_i = 1$), orders k_i, and parameters λ_i is

$$f_{mE}(x) = \sum_{i=1}^{r} \pi_i \frac{x^{k_i-1}}{(k_i-1)!} \lambda_i^{k_i} e^{-\lambda_i x} \qquad \text{for} \quad x \geq 0. \tag{B.1}$$

The distribution $F_{mE}(x)$ of such an Erlang mixture is

$$F_{mE}(x) = \sum_{i=1}^{r} \pi_i \left(1 - \sum_{j=0}^{k_i-1} \frac{e^{-\lambda_i x} (\lambda_i x)^j}{j!} \right) \qquad \text{for} \quad x \geq 0. \tag{B.2}$$

Corresponding to each Erlang distribution E_{k_i} in the mixture, let $\boldsymbol{T}_i$ denote the matrix of transition rates (subintensity), $\boldsymbol{t}_i'$ denote the exitrate vector and $\boldsymbol{\alpha}_i$ denote the initial probability vector (for a definition of phase-type representations see section 2.1), i.e. $\boldsymbol{\alpha}_i = (1, 0, \ldots, 0)$ and

$$\boldsymbol{T}_i = \begin{bmatrix} -\lambda_i & \lambda_i & 0 & \cdots & 0 & 0 \\ 0 & -\lambda_i & \lambda_i & \cdots & 0 & 0 \\ \vdots & & & \ddots & & \vdots \\ 0 & 0 & 0 & \cdots & -\lambda_i & \lambda_i \\ 0 & 0 & 0 & \cdots & 0 & -\lambda_i \end{bmatrix}, \quad \boldsymbol{t}_i' = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ \lambda_i \end{bmatrix}, \tag{B.3}$$

where the dimensions of $\boldsymbol{\alpha}_i, \boldsymbol{T}_i$ and $\boldsymbol{t}_i'$ are given by their corresponding orders k_i.

The phase-type representation for such a mixture of Erlang distributions is given by the following:

The initial probability vector $\boldsymbol{\alpha}$ of the phase-type representation is

$$\boldsymbol{\alpha} = ((\pi_1, 0, \ldots, 0), (\pi_2, 0, \ldots, 0), \ldots, (\pi_r, 0, \ldots, 0)) \tag{B.4}$$

where the i^{th} vector-component of vector $\boldsymbol{\alpha}$ is $\pi_i \boldsymbol{\alpha}_i$.

The phase-type subintensity $\boldsymbol{T}$ is

$$\boldsymbol{T} = \begin{bmatrix} \boldsymbol{T}_1 & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{T}_2 & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & & \ddots & & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{T}_{r-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \boldsymbol{T}_r \end{bmatrix} \tag{B.5}$$

The exitrate vector $\boldsymbol{t}'$ is given by

$$\boldsymbol{t}' = ((0, \ldots, 0, \lambda_1), (0, \ldots, 0, \lambda_2), \ldots, (0, \ldots, 0, \lambda_r))' \tag{B.6}$$

where the i^{th} vector-component of $\boldsymbol{t}'$ is $\boldsymbol{t}'_i$.

We note that the order of the resulting phase-type distribution is equal to $\sum_{i=1}^{r} k_i$, the sum of the individual orders of the Erlang distributions in the mixture.

Appendix C

Notation

$F(\cdot)$	A cumulative probability distribution function (cdf).
$f(\cdot)$	A probability density function (pdf).
μ_i'	The i^{th} non-central moment of a probability distribution.
μ_i	The i^{th} central moment of a probability distribution.
PH_k	A probability distribution of phase-type with k phases.
E_k	An Erlang distribution with k phases.
GE_k	A generalized Erlang distribution with k phases.
H_k	A hyperexponential distribution with k phases.
GH_k	A generalized hyperexponential distribution with k phases.
APH_k	An acyclic phase-type distribution with k phases.
C_k	A Coxian distribution with k phases.
$\boldsymbol{A}$	Notation for a matrix.
$\boldsymbol{A}'$	The transpose of the matrix $\boldsymbol{A}$.
$\boldsymbol{a}$	Notation for a row vector.
$\boldsymbol{b}'$	Notation for a column vector.
$\boldsymbol{0}'$	A column vector consisting of zeroes.
$\boldsymbol{e}'$	A column vector consisting of ones.
$\boldsymbol{Q}$	The infinitesimal generator corresponding to a continuous time Markov chain with k transient states and one absorbing state.
$\boldsymbol{T}$	The subintensity of $\boldsymbol{Q}$ that corresponds to the k transient states.
$\boldsymbol{t}_0'$	The exit rate vector.
$(\boldsymbol{\alpha}, \alpha_{k+1})$	The initial probability vector of $\boldsymbol{Q}$.
$(\boldsymbol{\alpha}, \boldsymbol{T})$	The representation of a phase-type distribution with initial probability vector $(\boldsymbol{\alpha}, \alpha_{k+1})$ and subintensity $\boldsymbol{T}$ (or generator $\boldsymbol{Q}$).
λ_j	The rate of the j^{th} exponential phase of a phase-type distribution.