Health Care Manag Sci (2009) 12:11–26
DOI 10.1007/s10729-008-9070-y

Modelling healthcare systems with phase-type distributions

Mark Fackrell

Received: 17 October 2007 / Accepted: 17 April 2008 / Published online: 7 June 2008
© Springer Science + Business Media, LLC 2008

Abstract Phase-type distributions constitute a very versatile class of distributions. They have been used in a wide range of stochastic modelling applications in areas as diverse as telecommunications, finance, biostatistics, queueing theory, drug kinetics, and survival analysis. Their use in modelling systems in the healthcare industry, however, has so far been limited. In this paper we introduce phase-type distributions, give a survey of where they have been used in the healthcare industry, and propose some ideas on how they could be further utilized.

Keywords Phase-type distribution · Coxian distribution · Markov chain · Patient flow · Healthcare modelling

1 Introduction

Since their introduction by Neuts [38] in 1975, *phase-type* (*PH*) distributions have been used in a wide range of stochastic modelling applications in areas as diverse as telecommunications, finance, teletraffic modelling, biostatistics, queueing theory, drug kinetics, reliability theory, and survival analysis (see Fackrell [12, Chapter 1] for a survey). *PH* distributions have enjoyed such popularity because they constitute a very versatile class of distributions defined on the nonnegative real numbers that lead to models which are algorithmically tractable. Their formulation also allows the Markov structure of stochastic models to be retained when they replace the familiar exponential distribution.

Erlang [11], in 1917, was the first person to extend the exponential distribution with his "method of stages". He defined a nonnegative random variable as the time taken to move through a fixed number of stages (or states), spending an exponential amount of time with a fixed rate in each one. Nowadays we refer to distributions defined in this manner as *Erlang* distributions. Neuts [38] generalized Erlang's method of stages by defining a *PH* random variable as the time spent in the transient states of a finite-state continuous-time Markov chain with one absorbing state, until absorption.

Prior to Neuts's work much of the research in stochastic modelling and queueing theory relied on random variables of interest and service times being modelled by the exponential or Erlang distributions, and point and interarrival processes by the Poisson or Erlang renewal processes. *PH* distributions constitute a much more useful class of distributions for a number of reasons. First, they form a versatile class of distributions that are dense in the class of all distributions defined on the nonnegative real numbers. That is, they can approximate any nonnegative distribution arbitrarily closely (see Asmussen [2, Theorem 4.2]), although the number of states needed may be large. Second, since they have a simple probabilistic interpretation in terms of continuous-time Markov chains, they exhibit

M. Fackrell (✉)
Department of Mathematics and Statistics,
University of Melbourne, Melbourne,
Victoria 3010, Australia
e-mail: mfackrel@ms.unimelb.eu.au

a Markov structure which enables an easier analysis of models that use them instead of general distributions. Lastly, the use of *PH* distributions in stochastic models often enables algorithmically tractable solutions to be found. If *PH* distributions are used, many quantities of interest that are used in algorithms to compute performance measures can be expressed simply in terms of the inverse and exponential of matrices that contain only real entries. These calculations can nowadays be done relatively easily using a suitable software package such as MATLAB®.

For a comprehensive theoretical treatment of *PH* distributions see Neuts [39, Chapter 2]. Latouche and Ramaswami [29, Chapter 2] is a very readable introduction to the topic. The literature on the theory and application of *PH* distributions is vast and both of the abovementioned books provide extensive bibliographies. The two entries in the *Encyclopedia of Statistical Science* on *PH* distributions, Shaked and Shanthikumar [50], and Asmussen and Olsson [4], also provide excellent introductions to the subject.

Over the last two decades *PH* distributions have been used to some extent in healthcare modelling, and their usage has increased over the last two or three years. However, the extent to which they have been implemented is somewhat limited. First Coxian (see Cox [8]) distributions (a subclass of *PH* distributions) have been used almost exclusively. Most researchers have avoided using general *PH* type distributions because they present some problems, but at the expense of developing more versatile models. One particular problem with general *PH* representations is that they are considerably overparameterized, whereas Coxian representations are not. Fitting Coxian distributions of the same order to data requires a lot less parameter estimations (see Section 4 for a discussion of this aspect). Second, the healthcare area in which *PH* (Coxian) distributions have been used most widely has been in modelling the length of stay of patients in geriatric facilities. While this work has been quite good it is hoped that with this paper *PH* distributions can be better understood by modellers and used in a broader variety of ways in the healthcare sector.

The paper is organized as follows. In Section 2 we introduce the exponential distribution and continuous-time Markov chains before formally defining *PH* distributions. We include a subsection on fitting *PH* distributions. Section 3 contains a detailed discussion on where *PH* distributions have been used in the healthcare industry. In Section 4 we propose some ways in which *PH* distributions could be further utilized in healthcare modelling. The paper concludes with Section 5.

2 Phase-type distributions

2.1 The exponential distribution

The exponential distribution is ubiquitous in stochastic modelling, mainly because of its simplicity and ability to model random lengths of time reasonably well. For example, it has been used to model the length of stay in a hospital bed, or the time between presentations to an emergency department. In this short subsection we introduce the exponential distribution and list some of its properties. Refer to Norris [43, Section 2.3] or Ross [47, Section 5.2] for further information about the exponential distribution.

A continuous nonnegative random variable T is distributed according to an *exponential distribution* with parameter $\lambda > 0$, if its *distribution* (or *cumulative distribution*) *function*, defined for $t \geq 0$, is given by

$$F(t) = P(T \leq t) = 1 - e^{-\lambda t}. \tag{1}$$

The *density* (or *probability density*) *function* of T, defined for $t \geq 0$, is given by

$$f(t) = F'(t) = \lambda e^{-\lambda t}. \tag{2}$$

The expected value of T, or its mean, is $\mathbb{E}(T) = \frac{1}{\lambda}$, and its variance is $\mathbb{V}(T) = \frac{1}{\lambda^2}$.

The simplicity in using the exponential distribution in stochastic modelling is not only due to its formulation in terms of a single parameter λ, but also because of the so called *memoryless property*. That is, for $s, t \geq 0$, $P(T > s + t | T > t) = P(T > s)$, see Norris [43, pages 70–71] or Ross [47, pages 201–204]. The memoryless property enables simple expressions for many performance measures of stochastic models that use the exponential distribution to be given. We also remark here that the exponential distribution is the only continuous distribution that exhibits the memoryless property.

While the exponential distribution has been used extensively in stochastic modelling, its main drawback is its lack of versatility, being characterized by only one parameter. We need to seek another, more versatile class of distributions which exhibit some of the

favourable properties of the exponential distribution. *PH* distributions are one such class.

2.2 Markov chains

Before we formally define *PH* distributions in the next subsection, we introduce, by way of an example, the finite-state continuous-time Markov chain, one of the most powerful tools used in stochastic modelling, see Norris [43, Chapter 2] or Ross [47, Chapter 6] for further properties of Markov chains.

Figure 1 shows the *state transition diagram* for a *finite-state continuous-time Markov chain*. The Markov chain consists of four states labelled 0, 1, 2, and 3. States 1, 2, and 3 are called *transient* states, and state 0 an *absorbing* state. A state is *transient* if once it has been reached, the probability of returning to it is less than one, and a state is *absorbing* if once it has been reached the process stops. We choose any of states 0, 1, 2, and 3, according to the probabilities $\frac{1}{10}$, $\frac{1}{2}$, $\frac{3}{10}$, and $\frac{1}{5}$, respectively. The probability of being instantaneously absorbed, that is $\frac{1}{10}$, is known as the *point mass at zero*. Suppose that state 1 has been chosen. We spend an exponentially distributed length of time with parameter $\lambda = 12$ there. This parameter can be interpreted as the (average) rate of movement out of state 1. Once we have completed this time we move to either state 0 or state 2 with (average) rates 8 and 4, respectively. Alternatively, we move from state 1 to state 0 with probability $\frac{8}{12} = \frac{2}{3}$, or to state 2 with probability $\frac{4}{12} = \frac{1}{3}$. If we chose state 0 we stop, but if we chose state 2 we spend an exponentially distributed length of time with

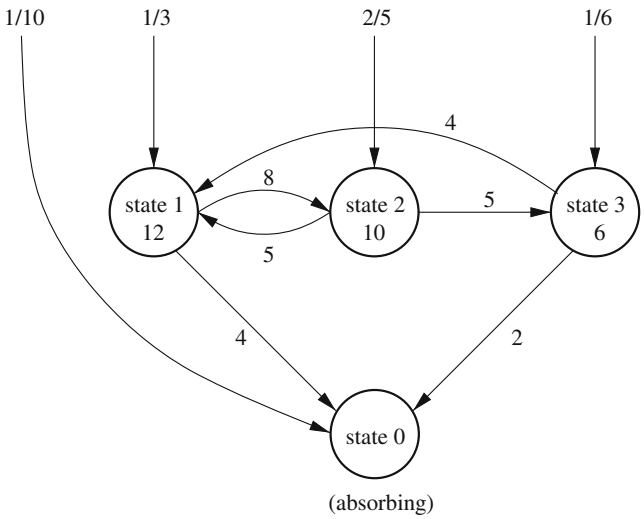

Fig. 1 State transition diagram of a 4-state continuous-time Markov chain with one absorbing state

$\lambda = 10$ there, and so on until absorption. The various rates have been chosen so that absorption occurs *with probability one*.

In order to describe the Markov chain we need three descriptors.

1. A *state space* $S = \{0, 1, 2, 3\}$.
2. An *initial state probability distribution*

$$(\alpha_0, \boldsymbol{\alpha}) = \left(\frac{1}{10}\ \frac{1}{3}\ \frac{2}{5}\ \frac{1}{6}\right) \tag{3}$$

   which governs the selection of the initial state, α_0 being the point mass at zero.
3. An *infinitesimal generator*

$$\boldsymbol{Q} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 4 & -12 & 8 & 0 \\ 0 & 5 & -10 & 5 \\ 2 & 4 & 0 & -6 \end{pmatrix} \tag{4}$$

   which governs the transitions between states.

The rows (labelled 0, 1, 2, and 3) of $\boldsymbol{Q}$ correspond to the state we move from, and the columns (labelled 0, 1, 2, and 3) correspond to the state we move to. The zeroth row consists of all zeros because once we have reached state 0 (absorption) we stay there. The remaining diagonal entries are negative and the off diagonal entries nonnegative, with all row sums equal to zero. The absorption rates from states 1, 2, and 3 are 4, 0, and 2, respectively. The distribution of time from start to finish (absorption), in the Markov chain, is said to have a *PH* distribution which we formally define in Subsection 2.3. We also note that, in practice, the point mass at zero α_0 is rarely necessary and is set to zero.

2.3 Phase-type distributions

Consider a continuous-time Markov chain on a finite state space $S = \{0, 1, 2, \ldots, p\}$ where state 0 is absorbing. Let the initial state probability distribution be $(\alpha_0, \boldsymbol{\alpha}) = (\alpha_0, \alpha_1, \ldots, \alpha_p)$ $\left(\text{with } \sum_{i=0}^{p} \alpha_i = 1\right)$ and the infinitesimal generator be $\boldsymbol{Q}$. The random variable that is defined as the time to absorption, is said to have a (*continuous*) *PH* distribution, see Neuts [39].

The infinitesimal generator for the Markov chain can be written in block-matrix form as

$$\boldsymbol{Q} = \begin{pmatrix} 0 & \mathbf{0} \\ \boldsymbol{t} & \boldsymbol{T} \end{pmatrix}. \tag{5}$$

Here, $\mathbf{0}$ is a $1 \times p$ vector of zeros. The vector $\boldsymbol{t} = (t_{10}, t_{20}, \ldots, t_{p0})'$ (the prime denoting transpose) where,

for $i = 1, 2 \ldots p, t_{i0} \geq 0$, with at least one of the t_{i0}s positive, is the absorption rate from state i. The $p \times p$ matrix $\boldsymbol{T} = [t_{ij}]$ is such that, for $i, j = 1, 2, \ldots, p$, with $i \neq j$,

$$t_{ij} \geq 0, \tag{6}$$

and

$$t_{ii} = -\sum_{\substack{j=0 \\ j \neq i}}^{p} t_{ij}, \tag{7}$$

that is, $\boldsymbol{t} = -\boldsymbol{T}\boldsymbol{e}$ where $\boldsymbol{e}$ is a $p \times 1$ vector of ones. The *PH* distribution is said to have a *representation* $(\boldsymbol{\alpha}, \boldsymbol{T})$ of *order* p. The matrix $\boldsymbol{T}$ is referred to as a *PH generator*. The point mass at zero α_0 is completely determined by $\boldsymbol{\alpha}$ and therefore does not need to appear in the expression for the representation. Typically representations are not unique and there must exist at least one representation of minimal order. Such a representation is known as a *minimal representation*, and the *order of the PH distribution* itself is defined to be the order of any of its minimal representations.

To ensure absorption in a finite time with probability one, we require that every nonabsorbing state is transient. This statement is equivalent to $\boldsymbol{T}$ being invertible, see Neuts [39, Lemma 2.2.1], or Latouche and Ramaswami [29, Theorem 2.4.3]. An additional requirement on the *PH* representation $(\boldsymbol{\alpha}, \boldsymbol{T})$ is that there are no superfluous phases. That is, each phase in the Markov chain defined by $\boldsymbol{\alpha}$ and $\boldsymbol{T}$ has a positive probability of being visited before absorption. If this is the case, then we say that the *PH* representation is *irreducible*, see Neuts [39, page 48]. This condition is equivalent to the matrix

$$\boldsymbol{T}^* = \boldsymbol{T} - (1 - \alpha_0)^{-1}\boldsymbol{T}\boldsymbol{e}\boldsymbol{\alpha}, \tag{8}$$

being irreducible. For the definition of an *irreducible matrix* see Seneta [49, pages 18 and 46]. If the representation is reducible, we can form an irreducible representation by simply deleting those states that are superfluous.

A *PH* distribution with representation $(\boldsymbol{\alpha}, \boldsymbol{T})$ has distribution function, defined for $t \geq 0$, given by

$$F(t) = \begin{cases} \alpha_0, & t = 0 \\ 1 - \boldsymbol{\alpha}\exp(\boldsymbol{T}t)\boldsymbol{e}, & t > 0, \end{cases} \tag{9}$$

see Neuts [39, Lemma 2.2.2], or Latouche and Ramaswami [29, Theorem 2.4.1]. Differentiating Eq. 9 with respect to t gives the corresponding density function, defined for $t > 0$,

$$f(t) = -\boldsymbol{\alpha}\exp(\boldsymbol{T}t)\boldsymbol{T}\boldsymbol{e}. \tag{10}$$

The Laplace-Stieltjes transform of Eq. 9, which is defined for $s \in \mathbb{C}$ such that $\Re(s) > \delta$ where δ is the real and negative eigenvalue of maximal real part of $\boldsymbol{T}$ (see Neuts [40]), is given by

$$\begin{aligned} \phi(s) &= \int_0^\infty e^{-st} dF(t) \\ &= -\boldsymbol{\alpha}(s\boldsymbol{I} - \boldsymbol{T})^{-1}\boldsymbol{T}\boldsymbol{e} + \alpha_0. \end{aligned} \tag{11}$$

For $k = 1, 2, \ldots$, differentiating Eq. 11 k times with respect to s and letting $s \to 0$ gives the kth noncentral moment

$$m_k = (-1)^k k!\boldsymbol{\alpha}\boldsymbol{T}^{-k}\boldsymbol{e}. \tag{12}$$

In particular the mean of a *PH* distribution with representation $(\boldsymbol{\alpha}, \boldsymbol{T})$ is

$$m_1 = -\boldsymbol{\alpha}\boldsymbol{T}^{-1}\boldsymbol{e},$$

and its variance is

$$m_2 - m_1^2 = 2\boldsymbol{\alpha}\boldsymbol{T}^{-2}\boldsymbol{e} - (\boldsymbol{\alpha}\boldsymbol{T}^{-1}\boldsymbol{e})^2.$$

We now give some examples of *PH* distributions.

1. The exponential distribution. The minimal representation is

$$\boldsymbol{\alpha} = (1) \tag{13}$$

$$\boldsymbol{T} = (-\lambda). \tag{14}$$

2. The order p *generalized Erlang* distribution. This distribution can be described using a state transition diagram that has p states in series, see Fig. 2. It is easy to see, without loss of generality, that the states can be ordered so that the rates $0 < \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_p$.

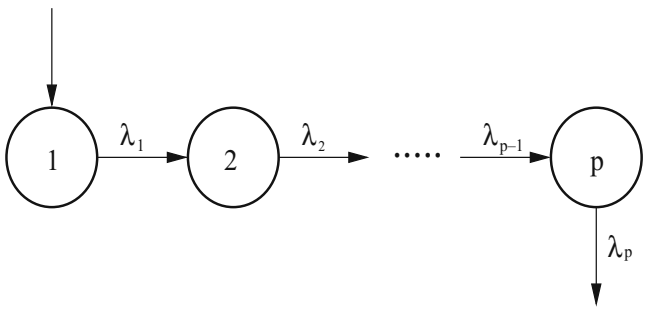

Fig. 2 State transition diagram for an order p generalized Erlang distribution

The representation for the generalized Erlang distribution corresponding to the state transition diagram is

$$\boldsymbol{\alpha} = \begin{pmatrix} 1 & 0 & \dots & 0 \end{pmatrix} \tag{15}$$

$$\boldsymbol{T} = \begin{pmatrix} -\lambda_1 & \lambda_1 & 0 & \cdots & 0 \\ 0 & -\lambda_2 & \lambda_2 & \cdots & 0 \\ 0 & 0 & -\lambda_3 & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & -\lambda_p \end{pmatrix}. \tag{16}$$

Figure 3 shows the density function for an order 5 *Erlang* distribution (that is, all rates are equal). The density function for an Erlang distribution of order p, defined for $t > 0$, is given by

$$f(t) = \frac{\lambda^p t^{p-1} e^{-\lambda t}}{p!}. \tag{17}$$

The expression for the density function of a generalized Erlang distribution is complicated by the fact that some of the rates may be unequal.

All generalized Erlang distributions have *coefficient of variation* (that is, the ratio of the variance to the mean squared) less than or equal to one. In fact, the exponential distribution, which is a degenerate case of the generalized Erlang distribution, is the only one that has coefficient of variation equal to one.

3. The order p hyperexponential distribution. This distribution can be described using a state transition diagram with p states in parallel, see Fig. 4. Clearly, without loss of generality, the states can be ordered so that the rates $0 < \lambda_1 < \lambda_2 < \ldots < \lambda_p$.

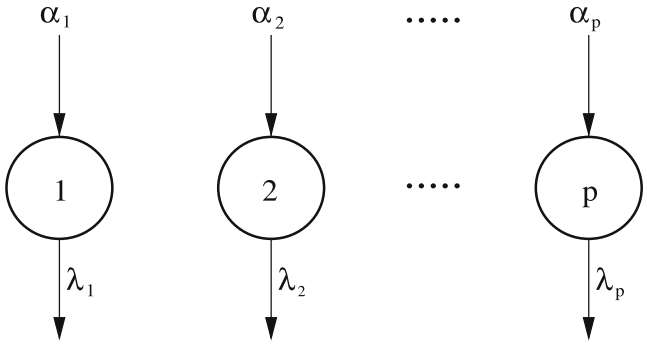

Fig. 4 State transition diagram for an order p hyperexponential distribution

The corresponding representation is

$$\boldsymbol{\alpha} = \begin{pmatrix} \alpha_1 & \alpha_2 & \dots & \alpha_p \end{pmatrix} \tag{18}$$

$$\boldsymbol{T} = \begin{pmatrix} -\lambda_1 & 0 & \dots & 0 \\ 0 & -\lambda_2 & \ddots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & -\lambda_p \end{pmatrix}, \tag{19}$$

with density function, defined for $t > 0$, given by

$$f(t) = \sum_{i=1}^{p} \alpha_i \lambda_i e^{-\lambda_i t}, \tag{20}$$

where, for $i = 1, 2, \ldots, p$, $\alpha_i > 0$ and $\sum_{i=1}^{p} \alpha_i = 1$.

Figure 5 shows an example of the density function of a order 3 hyperexponential distribution.

Hyperexponential distributions have coefficient of variation greater than or equal to one (the

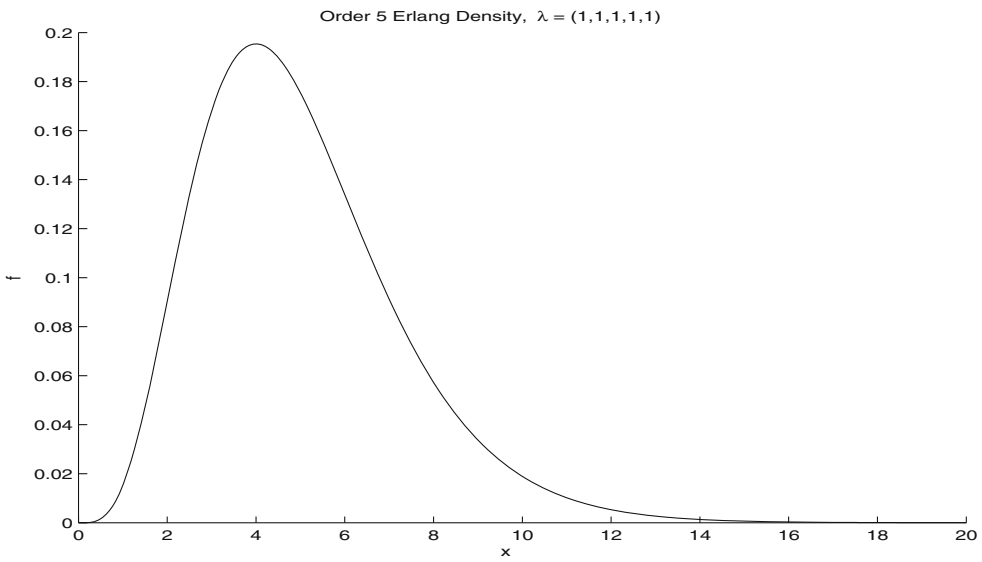

Fig. 3 Density function of a order 5 Erlang distribution with $\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = \lambda_1 = 1$

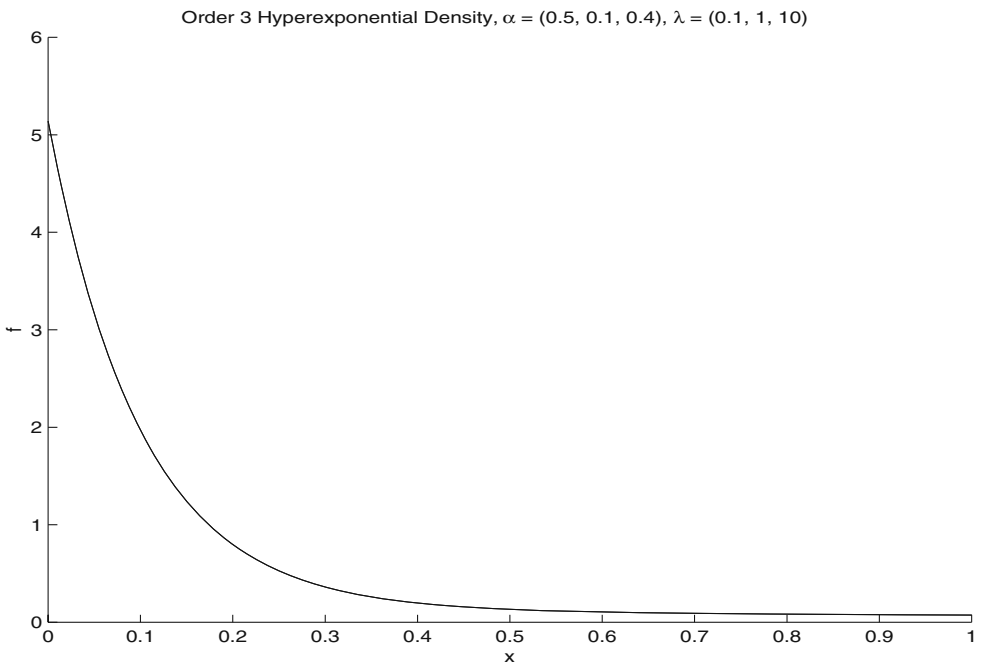

Fig. 5 Density function of an order 3 hyperexponential distribution with $\boldsymbol{\alpha} = (0.5, 0.1, 0.4)$, and $\lambda_1 = 0.1$, $\lambda_2 = 1$, and $\lambda_3 = 10$

exponential distribution is the only one that attains equality here).

4. The order p *Coxian* distribution. The state transition diagram for this distribution is shown in Fig. 6.

   These distributions have representations of the form

$$\boldsymbol{\alpha} = (\alpha_1\ \alpha_2\ \ldots\ \alpha_p) \tag{21}$$

$$\boldsymbol{T} = \begin{pmatrix} -\lambda_1 & \lambda_1 & 0 \ldots & 0 \\ 0 & -\lambda_2 & \lambda_2 \ldots & 0 \\ 0 & 0 & -\lambda_3 \ddots & 0 \\ \vdots & \vdots & \ddots\ \ddots & \vdots \\ 0 & 0 & 0 \ldots & -\lambda_p \end{pmatrix}. \tag{22}$$

   Although it is not obvious, in this case, without loss of generality, the states can be ordered so that the rates $0 < \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_p$, see Cumani [9] or O'Cinneide [44]. Figure 7 shows the density function for an order 4 Coxian distribution. Its shape exemplifies the extra flexibility Coxian distributions exhibit over generalized Erlang and hyperexponential distributions.

5. The *acyclic*, or *triangular PH* (*TPH*) distribution. This type of *PH* distribution have generators that are upper triangular matrices. Cumani [9] (see also O'Cinneide [44]) showed that any *TPH* representation has a Coxian representation of the same or lower order.

6. The order p *unicyclic* distribution. These distributions have state transition diagrams as shown in Fig. 8.
   They have representations of the form

$$\boldsymbol{\alpha} = (\alpha_1\ \alpha_2\ \ldots\ \alpha_p) \tag{23}$$

$$\boldsymbol{T} = \begin{pmatrix} -\lambda_1 & \lambda_1 & 0 \ldots & 0 & 0 \\ 0 & -\lambda_2 & \lambda_2 \ldots & 0 & 0 \\ 0 & 0 & -\lambda_3 \ddots & 0 & 0 \\ \vdots & \vdots & \ddots\ \ddots & \vdots & \vdots \\ 0 & 0 & 0 \ldots & -\lambda_{p-1} & \lambda_{p-1} \\ \mu_1 & \mu_2 & \mu_3 \ldots & \mu_{p-1} & -\lambda_p \end{pmatrix}, \tag{24}$$

   where, for $i = 1, 2, \ldots, p-1$, $\mu_i \geq 0$, $0 < \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_p$, and $\lambda_p > \sum_{i=1}^{p-1} \mu_i$, see O'Cinneide

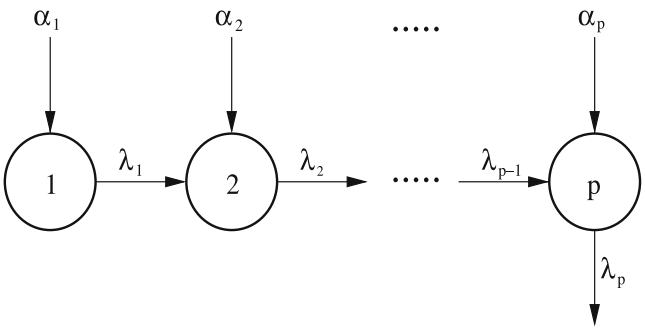

Fig. 6 State transition diagram for an order p Coxian distribution

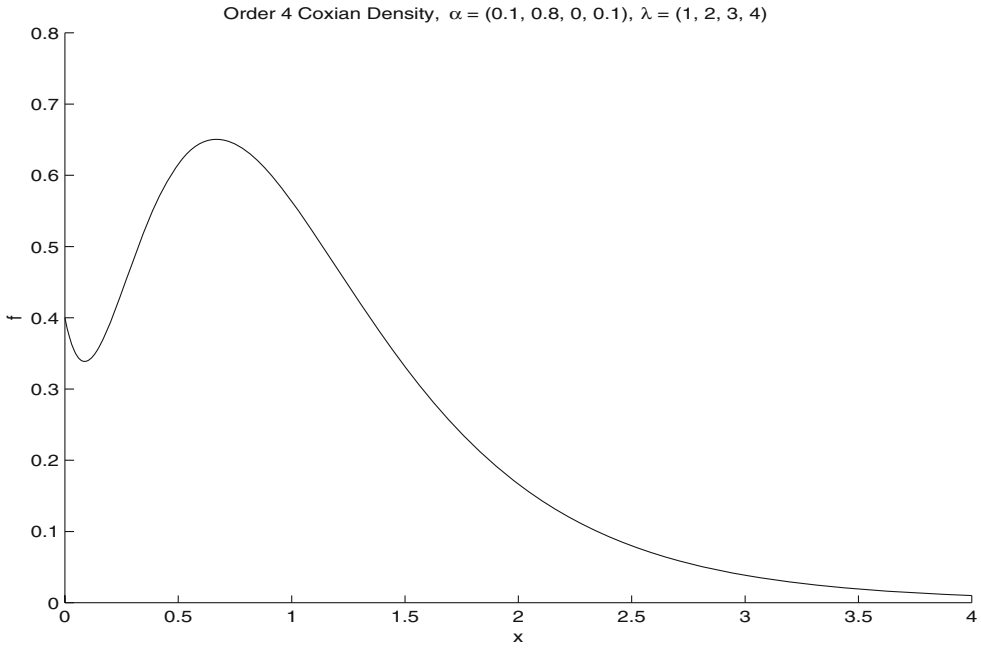

Fig. 7 Density function of an order 4 Coxian distribution with $\boldsymbol{\alpha} = (0.1, 0.8, 0, 0.1)$, and $\lambda_1 = 1, \lambda_2 = 2, \lambda_3 = 3, \lambda_4 = 4$

[45, Section 7]. Figure 9 shows the density function for an order 5 unicyclic distribution.

It was conjectured in O'Cinneide [45, Conjecture 4] that every *PH* distribution of order p has a unicyclic representation of the same order, however, He and Zhang [25] showed that this is not, in general, the case.

2.4 Fitting and approximating with phase-type distributions

In order to use *PH* distributions to model real world phenomena we need reliable methods to fit empirical data and approximate probability distributions with them. The aim of any *fitting* procedure is to estimate the parameters $\boldsymbol{\alpha}$ and $\boldsymbol{T}$ so that they best fit the data in some sense. In *approximating* a probability distribution with a *PH* distribution the parameters $\boldsymbol{\alpha}$ and $\boldsymbol{T}$ need to be selected so that a predetermined function of the *approximated* distribution and the *approximating PH* distribution is minimized. In this subsection we give a brief overview of some of the *PH* fitting and distribution approximation algorithms found in the literature. The survey is by no means complete and we refer the reader to the comprehensive reference lists given in Asmussen, et al. [3], Bobbio and Cumani [7], and Johnson [28] for more information.

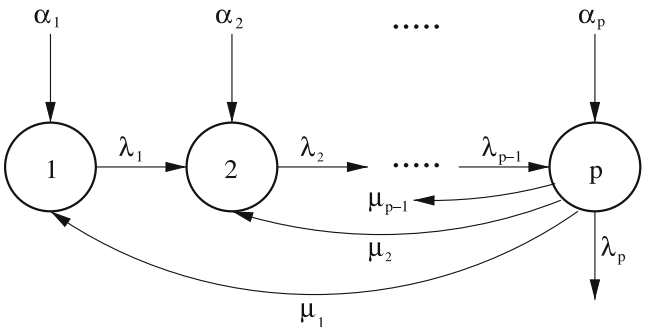

Fig. 8 State transition diagram for an order p unicyclic distribution

Maximum likelihood estimation has been the most popular method used to fit data and approximate distributions with *PH* distributions. Asmussen et al. [3] (see also Asmussen [1]) developed an expectation-maximization (*EM*) algorithm (see Dempster et al. [10], or McLachlan and Krishnan [36]) to calculate maximum likelihood parameter estimates for general *PH* distributions when fitted to empirical data. In a companion paper Olsson [46] extended the algorithm so that it could be used with right-censored and interval-censored data. The original and extended algorithms are available as the downloadable package EMpht.[1] Bobbio and Cumani [7] developed an algorithm to fit Coxian distributions to empirical data, with the option of including right-censored data, using maximum likelihood estimation. In order to find the parameters that maximized the loglikelihood function the resulting nonlinear program was solved by combining a linear program with a line search at each iteration. Faddy [15, 16], and [17], Faddy and McClean [19], and

[1] http://home.imf.au.dk/asmus/pspapers.html.

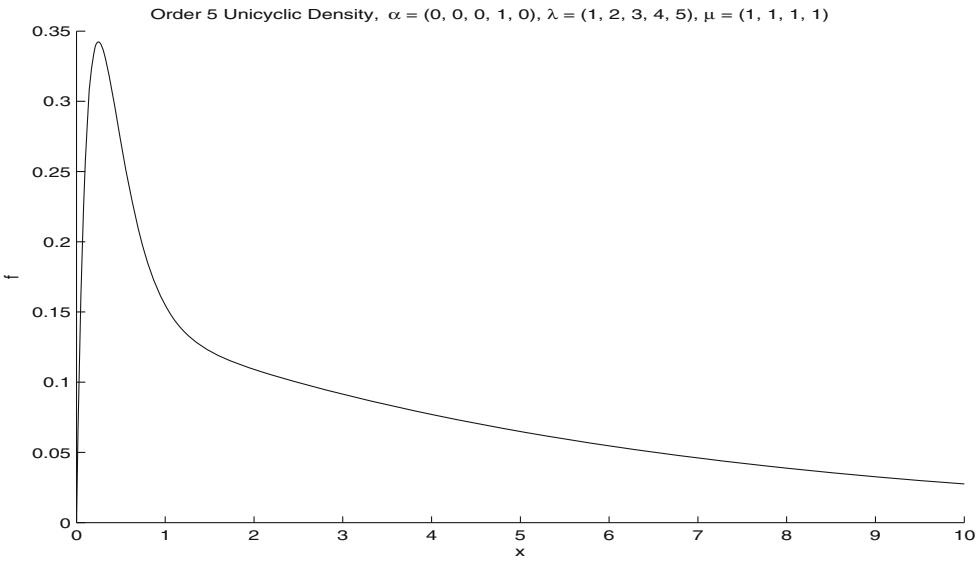

Fig. 9 Density function of an order 5 unicyclic distribution with $\boldsymbol{\alpha} = (0, 0, 0, 1, 0)$, $\lambda_1 = 1$, $\lambda_2 = 2$, $\lambda_3 = 3$, $\lambda_4 = 4$ and $\lambda_5 = 5$, and $\mu_1 = \mu_2 = \mu_3 = \mu_4 = 1$

Hampel [24] used maximum likelihood estimation to fit Coxian distributions to real data. They used existing MATLAB® or S-PLUS® routines (for example the Nelder-Mead algorithm in MATLAB®, see Nelder and Mead [37]) to perform the required parameter estimation. In Faddy [18] a penalized maximum likelihood method was developed to fit Coxian distributions to data. Coxian representations where $\boldsymbol{T}$ had disparate eigenvalues (that is, diagonal entries for upper triangular matrices) were penalized in the fitting process. This restriction resulted in smoother fitted density functions.

The method of *moment matching* has also been used to fit *PH* distributions to data and approximate probability density functions. Johnson [28] developed an algorithm that matched the first three moments of a mixture of Erlang distributions to the respective moments of empirical data or a distribution. The nonlinear optimization program, which resulted from the parameter estimation or distribution approximation technique, was solved using sequential quadratic programming. Schmickler [48] also developed a moment matching algorithm where the first three moments of a mixture of two or more Erlang distributions were matched exactly to the respective moments of an empirical distribution function. The Nelder-Mead algorithm was used to solve the resulting nonlinear program.

Other methods for *PH* fitting and approximation have also been used. Hórvath and Telek [26] developed a method which separately approximated the main part and the tail of an arbitrary distribution defined on the nonnegative real numbers with a *PH* distribution. The main part of the distribution was approximated with a Coxian distribution by minimizing any distance (goal) function of the approximated and approximating densities. A nonlinear programming procedure similar to that of Bobbio and Cumani [7] was used to perform the minimization. The tail was approximated with a hyperexponential distribution using a heuristic method. Faddy [13] and [14] used least squares to fit Coxian distributions to real sample data in order to estimate the parameters for a compartmental model used in drug kinetics. A quasi-Newton minimization algorithm was used to perform the parameter estimation.

3 Phase-type distributions in the healthcare industry

There have been a number of papers written on the application of *PH* distributions in the healthcare literature, but as we shall see, the number of areas where they have been used is rather limited. Most papers concern the modelling of the length of stay in geriatric facilities, and these papers have been written by a relatively small pool of researchers. In this section we present a literature review on the use of *PH* distributions in the healthcare sector.

Faddy [13] used a two-compartment model, such as the one depicted in Fig. 10, to model the outflow of labelled red blood cells injected into a rat liver. Each compartment represents a body organ, and the residency time the labelled cells spend in the body before being excreted was modelled with a generalized Erlang distribution. In the resultant fit $m = 35, n = 1, \lambda = 7.60$, and $\mu = 0.22$. The value of m was large because of a delay of about 3 seconds before any outflow was recorded. The method of least squares was used to fit the data.

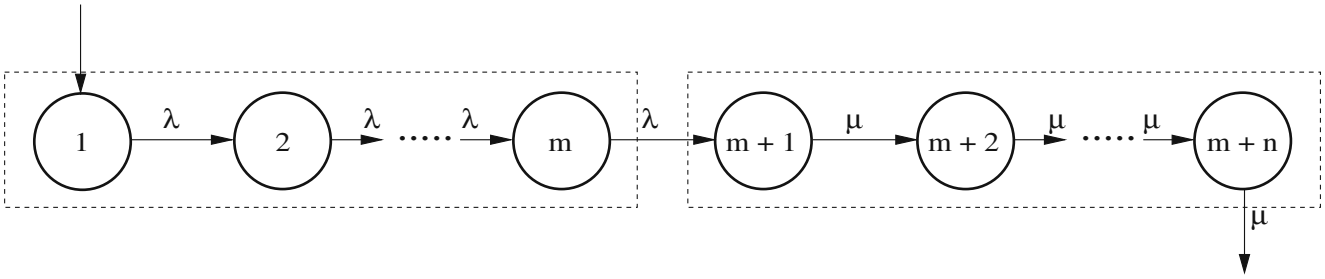

Fig. 10 Two-compartment model to model the outflow of labelled red blood cells injected into a rat liver

In Faddy [14] a more complex two-compartment model (see Fig. 11) was used to model the retention time of a drug injected into an organ. The cycling in the first compartment models diffusion, and the second compartment models the drug's clearance from the body. The model was fitted to the renal concentrations of an antibiotic drug in four sheep that were given differing doses at $t = 0$. For $n = 1$ and $m \geq 4$ it was reported that $\lambda = 0.028$, $\mu = 0.018$, and $\nu = 0.235$. Again, the method of least squares was used to fit the data.

Faddy [15] fitted Coxian distributions of increasing order to the length of treatment for patients at risk of suicide using maximum likelihood estimation. In this case it was deemed that an order three Coxian distribution was sufficient to model the data. The representation for the Coxian distribution was given as

$$\boldsymbol{\alpha} = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix} \tag{25}$$

$$\boldsymbol{T} = \begin{pmatrix} -(\lambda_1 + \mu_1) & \lambda_1 & 0 \\ 0 & -(\lambda_2 + \mu_2) & \lambda_2 \\ 0 & 0 & -\mu_3 \end{pmatrix}, \tag{26}$$

which corresponds to the state transition diagram shown in Fig. 12.

This "drop out" representation is equivalent to the "drop in" representation given by Eqs. 21 and 22. Here, it was noticed that in the fitted $\boldsymbol{T}$, $\lambda_1 + \mu_1 \approx \lambda_2 + \mu_2$. A further order three fit with the constraint $\lambda_1 + \mu_1 = \lambda_2 + \mu_2$ was made and compared with the original order three fit. It was this phenomenon that lead to the penalized maximum likelihood estimation described in Faddy [18]. A similar approach was taken by Faddy and Taylor [21] to model the time to onset of *bronchiolitis obliterans syndrome* (*BOS*) for lung transplant patients. In this case, three covariates were also included in the model. They were the number of rejections (x_1), the number of infections (x_2), and cytomegalovirus episodes (x_3) in the post operative period before the onset of *BOS*. They were incorporated via the parameters of $\boldsymbol{T}$, that is,

$$\lambda_i = \exp(c_i - b_1 x_1 - b_2 x_2 - b_3 x_3), \quad i = 1, 2, \ldots, p-1, \tag{27}$$

$$\mu_i = \exp(d_i - b_1 x_1 - b_2 x_2 - b_3 x_3), \quad i = 1, 2, \ldots, p, \tag{28}$$

where the parameters b_1, b_2, b_3, $c_1, c_2, \ldots, c_{p-1}$, and $d_1, d_2, \ldots, d_p$ are real numbers. In the example given, the order of the most suitable *PH* fit was $p = 2$.

McClean and Millard [35], while not specifically referring to *PH* distributions, fitted an order two hyperexponential distribution to the length of stay of patients in a geriatric medicine department. That is, the density function for the length of stay distribution, for $t \geq 0$, $\lambda_1, \lambda_2 > 0$, and $0 < \rho < 1$, was given by

$$f(t) = \rho e^{-\lambda_1 t} + (1 - \rho) e^{-\lambda_2 t}. \tag{29}$$

They fitted the data for male and female patients separately. The two states in the model represented acute/rehabilitative (short stay) patients, and long stay

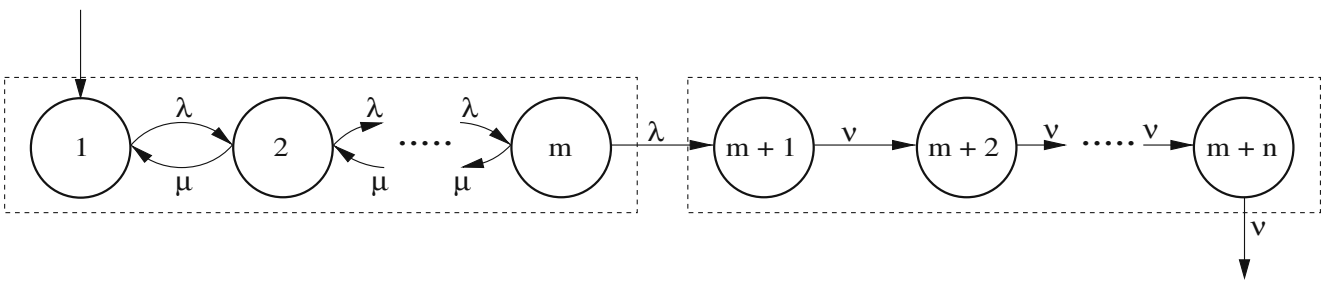

Fig. 11 Two-compartment model to model the diffusion and clearance of a drug injected into a body organ

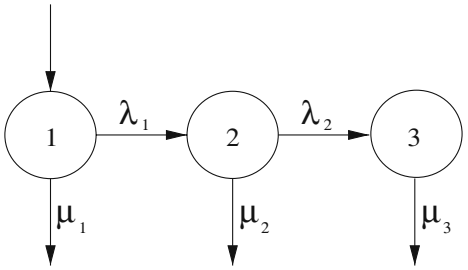

Fig. 12 State transition diagram for a "drop out" order three Coxian distribution

patients. Patients who left the system by either being discharged or dying were categorized as short stay, and those who left by being transferred elsewhere, as long stay. The parameter ρ was estimated by the proportion of (either male or female) short stay patients, and λ_1 and λ_2 by the reciprocal of the mean length of stay for short and long stay patients, respectively. The model was improved by fitting a mixture of a lognormal distribution (for short stay) and an exponential distribution (for long stay).

In Faddy and McClean [19] Coxian distributions of increasing order were fitted to the male patient data used by McClean and Millard [35] using maximum likelihood estimation. Two covariates, the age of patient at admission, and the year of admission were incorporated into the model in the same way as in Faddy [15]. Unlike McClean and Millard [35], here the Coxian distribution was fitted first, and then an interpretation sought. If, for example, a three state model (see Eqs. 25 and 26) was fitted, the states could be interpreted as representing severity of illness, leading to a characterization for "short stay", "medium stay", and "long stay" patients. If absorption takes place from the first state then the patient could be classified as short stay, from the second state, medium stay, and so on. If more states are used a similar interpretation can be given. Hence, the Coxian distribution can be interpreted as a mixture of generalized Erlang distributions with, for the order three case, mixing coefficients

$$p_1 = \frac{\mu_1}{\mu_1 + \lambda_1} \tag{30}$$

$$p_2 = \frac{\lambda_1 \mu_2}{(\mu_1 + \lambda_1)(\mu_2 + \lambda_2)} \tag{31}$$

$$p_2 = \frac{\lambda_1 \lambda_2}{(\mu_1 + \lambda_1)(\mu_2 + \lambda_2)}. \tag{32}$$

These mixing coefficients from the fitted distribution would model the proportion of short, medium, and long stay patients, respectively. Faddy and McClean [19] fitted the abovementioned dataset with an order four Coxian distribution (loglikelihood $= -9332.5$) without including the two covariates, but noted that when they were included the loglikelihood increased to -9310.2. They also remarked observing the same kind of parameter redundancy mentioned in Faddy [14].

Faddy and McClean [20] extended their earlier work by not only modelling the length of stay in geriatric care, but also the length of stay for geriatric patients in community care. Penalized maximum likelihood estimation was used to fit the data. In McClean et al. [33] a similar approach was taken to assign patients to clusters (for example, short stay, medium stay, and long stay if an order three Coxian distribution is used) based on the time already spent in care and the two covariates, age at admission and year of admission. The aim of this study was to develop a model to predict how long a patient remains in care given information about these three predictors.

Gorunescu et al. [22] modelled the patient flow through a hospital department using a steady state $M/PH/c/c$ queue (the authors refer to this queue as a $M/PH/c$ queue). That is, the interarrival time is exponentially distributed, the service time (time spent occupying a bed) is distributed according to a PH distribution, there are c beds in the department, and the capacity of the system (queue plus beds) is c. We can see that there is no queuing, that is, in theory, if a patient arrives to find the ward fully occupied they are lost to the system. In reality, depending on their condition, such a patient would be found a bed elsewhere. We note that, despite modelling the service time with a PH distribution, the performance measures that were considered, that is, the probability of all beds being occupied (the blocking probability), and the mean number of occupied beds, only depend on the mean service time and not on the service time distribution. The authors did mention this fact. Using some real data from a geriatric department, the authors presented an example where the minimum number of beds was calculated given that a specified blocking probability cannot be exceeded. The number of beds that minimized the average cost per unit time was also calculated.

Marshall and McClean [30] fitted conditional Coxian distributions to the length of stay data for geriatric patients. The term *conditional* was used because the data was first categorized according to a Bayesian belief network, and then fitted using maximum likelihood estimation. A *Bayesian belief network* is a model that links various causal characteristics of the data in some meaningful way. For example, in the paper, each

patient's age, gender, and admission method contributed to their Barthel grade (heavily dependent, very dependent, slightly dependent, or independent), and anticipated final destination (death, home, or transfer). This information, all determined beforehand, enabled the patient length of stay data to be categorized into 12 groups. A Coxian distribution was fitted to each categorized dataset in turn. The aim of this approach was to be better able to predict a patient's length of stay by utilizing prior information, in this case, Barthel grade and anticipated final destination.

Xie et al. [52] modelled the length of stay of geriatric patients in residential and nursing home care with a more complex *PH* representation than had previously been used. The times spent in residential and nursing home care were both modelled with a Coxian distribution. Figure 13 shows the state transition diagram for the model. Patients enter the system via the residential home care block consisting of states 1 and 2, and spend either a short time (state 1), or a long time (states 1 and 2) there before either leaving the system (state 0), or progressing to nursing home care, where, again, they can spend a short time (state 3), or a long time (states 3 and 4) before leaving. The *PH* representation for the model depicted in Fig. 13 is

$$\boldsymbol{\alpha} = \begin{pmatrix} 1 & 0 & 0 & 0 \end{pmatrix} \tag{33}$$

$$\boldsymbol{T} = \begin{pmatrix} -(\lambda_1+\mu_1+\nu_1) & \lambda_1 & \nu_1 & 0 \\ 0 & -(\lambda_2+\mu_2) & \lambda_2 & 0 \\ 0 & 0 & -(\lambda_3+\mu_3) & \lambda_3 \\ 0 & 0 & 0 & -\mu_4 \end{pmatrix}. \tag{34}$$

A Coxian distribution of more than order two may be used to model the length of stay in residential or nursing home care, resulting in a model of higher order. We remark here that the distribution with representation given by Eqs. 33 and 34 is Coxian since $\boldsymbol{T}$ is an upper triangular matrix.

In order to fit the model to the data the authors first established the number of states needed to model the residential and nursing home length of stays separately. They did this by fitting mixtures (not necessarily convex) of exponential distributions of increasing order to the data, and then used both the Aikaike and Bayesian information criteria to determine the optimal number of states required. Then, a Coxian distribution with the appropriate structure was fitted to the whole dataset. The authors fitted the model to some real data, and found that the residential home care length of stay was modelled sufficiently well with an exponential distribution, and the nursing home length of stay with an order two Coxian distribution.

A summary of the current state of affairs with modelling the length of stay in hospital departments was given in Marshall et al. [31], and Vasilakis and Marshall [51]. Both papers explained the various techniques for modelling length of stay including descriptive statistics, survival analysis, compartmental models, simulation modelling, mixed exponential distributions, *PH* distributions, and conditional *PH* distributions. In Vasilakis and Marshall [51] some of the methods were illustrated by modelling the length of hospital stay of stroke patients over the age of 65 in the UK.

Gribbin and McClean [23] (see also, McClean [32], and McClean and Gribbin [34]) modelled the length of time nurses took to return to service after a temporary interlude with compartment models. The models were fitted to data from the Northern Ireland nursing service and then analysed. Here, it is interesting to note that the focus is on human resource management in the

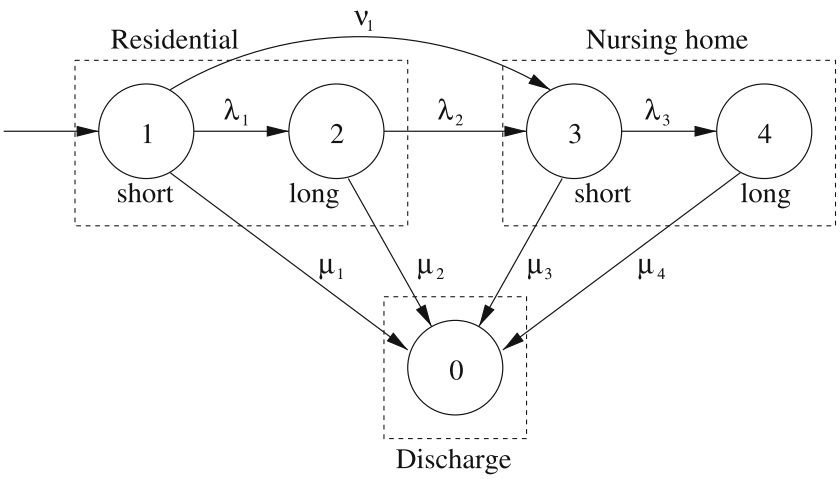

Fig. 13 State transition diagram to model the length of stay in residential and nursing home care

healthcare industry, rather than on the patient care process as have been most of the other applications of *PH* distributions in healthcare modelling.

This discourse on the use of *PH* distributions in the health and social care sector is by no means complete, and the reference lists given in the abovementioned papers should be referred to for further information.

4 Phase-type distributions and modelling healthcare processes

As mentioned in the previous section, the use of *PH* distributions in the healthcare sector has been limited, not only because of the relatively few areas in which they have been applied, but also because of the almost exclusive use of Coxian distributions to model lengths of stay. In this section we propose some ways in which the use of *PH* distributions could be extended in healthcare modelling.

Phase-type distributions can be used to fit any length of stay or interarrival data, not just the ones that have been discussed in the previous section. It appears that Coxian distributions are used because they are easy to fit, and also offer a simple interpretation for the length of stay. However, if general, or even unicyclic, phase-type distributions are used, fits with smaller order may be achieved. For example, Fig. 14 shows an order 6 general *PH* distribution fitted (using EMpht) to some data (truncated at 30 days) consisting of 4696 lengths of stay of patients at the Royal Melbourne Hospital that were transferred from other hospitals. The resultant representation is

$$\boldsymbol{\alpha} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \tag{35}$$

$$\boldsymbol{T} = \begin{pmatrix} -3.21 & 3.21 & 0 & 0 & 0 & 0 \\ 0 & -3.21 & 0 & 3.21 & 0 & 0 \\ 0.61 & 0 & -0.63 & 0 & 0.02 & 0 \\ 0 & 0 & 0 & -3.21 & 0 & 3.21 \\ 0 & 0 & 0.81 & 0 & -0.81 & 0 \\ 0 & 0 & 0 & 0 & 1.65 & -3.21 \end{pmatrix}. \tag{36}$$

The algorithm took approximately 3.5 min to perform 30,000 iterations, and the loglikelihood was −11706.92. The fit is shown in Fig. 14. As we can see it looks very good. It is not Coxian as the eigenvalues of $\boldsymbol{T}$ are not all real. In fact, a Coxian distribution of order 25 is required to achieve a fit with a greater loglikelihood! In this situation a non-Coxian distribution gives a *PH* representation of much lower order which will be easier to use in the calculation of any performance measures such as the mean and standard deviation of the length of stay. Here the mean of the fitted *PH* distribution is 5.6382 and its standard deviation is 6.5654. These quantities compare favourably with the sample mean of 5.6382, and the sample standard deviation of 6.2959.

To compare the performance of general *PH* distributions with other *PH* distributions, the same dataset was fitted (again using EMpht) with an exponential distribution, and order 6 hyperexponential, generalized Erlang, and Coxian distributions. Table 1 shows the

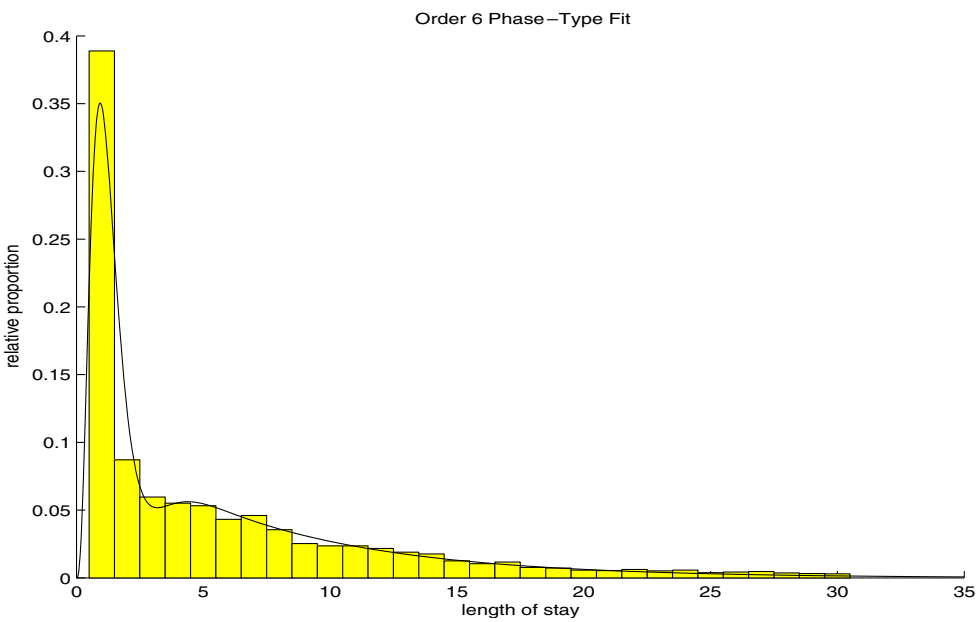

Fig. 14 Order 6 *PH* fit to the length of stay histogram

Table 1 Various *PH* fits to the length of stay histogram

Fit	Loglikelihood	order
Exponential	−12818.04	1
Hyperexponential	−12712.44	2
Generalized Erlang	−12437.75	6
Coxian	−12174.63	6
General *PH*	−11706.92	6

loglikelihoods for the five fits. We note here that the resultant hyperexponential fit produced by EMpht is of order 2. Figure 15 shows the five fits. The graph is truncated at 12 days to give a clear picture of how well the fits perform on the main body of the histogram.

As expected, the general *PH* fit is the best followed by the Coxian fit. These two distributions have the most flexibility. The generalized Erlang fit is better than the hyperexponential fit because, first, it is zero at the origin, and second, it has a higher order. Curiously, EMpht, even when fitting an order 25 hyperexponential distribution, still produces the same order 2 fit - this may be something to do with the expectation-maximization algorithm itself but further investigation is beyond the scope of this paper. The exponential distribution, of course, performs the worst.

One of the major problems with using general *PH* distributions to fit data is that they are considerably overparameterized. A general order p *PH* representation (with $\alpha_0 = 0$) requires $p^2 + p - 1$ parameters, but an argument using their Laplace transforms, or their moments (see Asmussen et al. [3]), shows that only $2p - 1$ parameters are required. Apart from needing to estimate more parameters than is necessary when fitting general *PH* distributions to data, the overparameterization problem means that it is virtually impossible to develop any asymptotic properties of the estimators and give confidence intervals for the parameter estimates. Bootstrapping is one alternative method for producing confidence intervals but requires many iterations of the expectation-maximization algorithm, see Asmussen [1]. On the other hand, Coxian distributions are not overparameterized, but the development of any asymptotic properties has not yet been undertaken.

However, despite the problem of overparameterization, the above example shows that an order 6 fit performs better (in terms of maximum likelihood estimation) than any Coxian fit of order less than or equal to 24, and takes a lot less time to fit. In fact, $(\boldsymbol{\alpha}, \boldsymbol{T})$ in Eqs. 35–36 has, it appears, only 5 free parameters, less than the required 11 parameters. Faddy [15] and [17], and Hampel [24] observed this phenomenon when fitting Coxian distributions to data using maximum likelihood estimation. That is, many of the parameters in the final fitted representations were equal, leading to representations that depended on very few parameters. The investigation and development of general *PH*

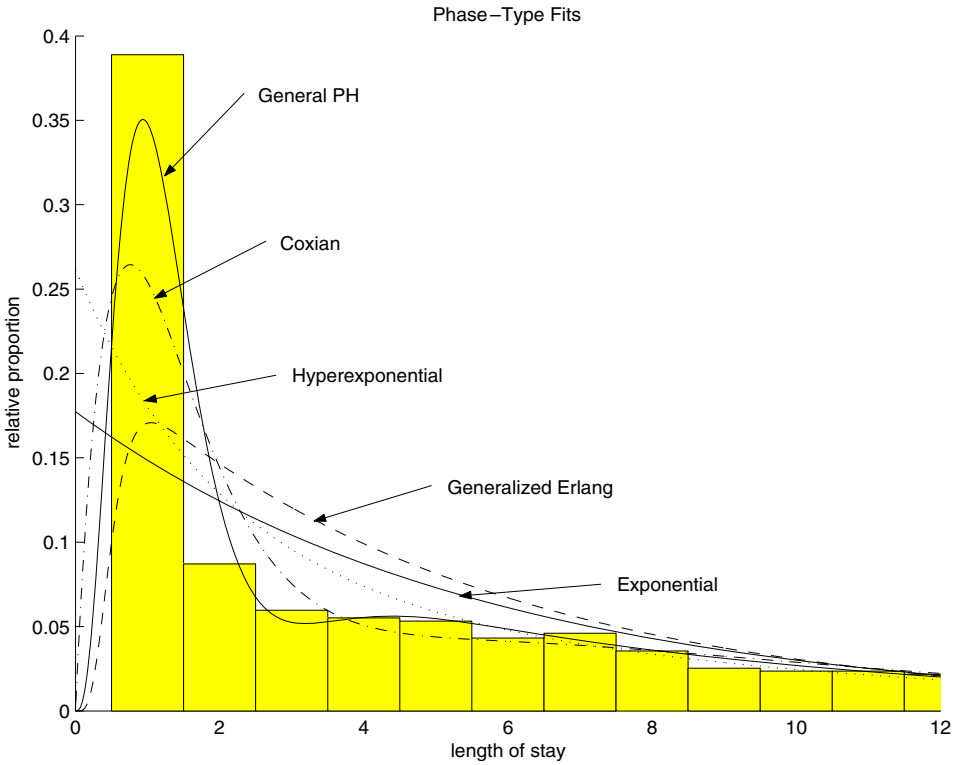

Fig. 15 Order 6 *PH* fit to the length of stay histogram

representations that depend on a minimum number of parameters is the focus of ongoing research, see, for example, He and Zhang [25], and Horváth and Telek [27].

Another way in which *PH* distributions could enhance healthcare modelling is to use them in more sophisticated models for lengths of stay and interarrival times. Consider the (simplified) schematic diagram for patient flow in a hospital shown in Fig. 16.

Patients enter the hospital via the emergency department (unit 1), or as elective patients. Patients who have had surgery (unit 2) enter the intensive care unit (unit 3), and then the high dependency ward (unit 4) before being sent to one of the two wards (units 5 and 6). Emergency patients can either have surgery, or be sent to one of the two wards. Patients may return to the intensive care unit if their condition warrants it. Patients are discharged only from the two wards. To model the length of time a patient stays in hospital we could, for $i = 1, 2, 3, 4, 5, 6$, model the time spent in unit i, with an order p_i *PH* distribution. The overall *PH* distribution would have a representation of the form

$$\boldsymbol{\alpha} = \left(\boldsymbol{\alpha}_1\ \boldsymbol{\alpha}_2\ \mathbf{0}\ \mathbf{0}\ \mathbf{0}\ \mathbf{0}\right) \tag{37}$$

$$\boldsymbol{T} = \begin{pmatrix} \boldsymbol{T}_{11} & \boldsymbol{T}_{12} & \mathbf{0} & \mathbf{0} & \boldsymbol{T}_{15} & \boldsymbol{T}_{16} \\ \mathbf{0} & \boldsymbol{T}_{22} & \boldsymbol{T}_{23} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{T}_{33} & \boldsymbol{T}_{34} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{T}_{43} & \boldsymbol{T}_{44} & \boldsymbol{T}_{45} & \boldsymbol{T}_{46} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{T}_{53} & \mathbf{0} & \boldsymbol{T}_{55} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{T}_{63} & \mathbf{0} & \mathbf{0} & \boldsymbol{T}_{66} \end{pmatrix}. \tag{38}$$

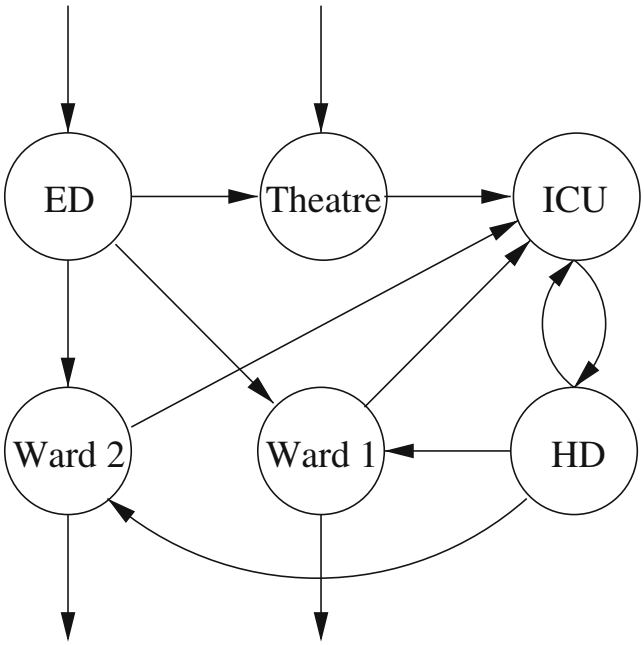

Fig. 16 Schematic diagram for patient flow in a hospital

Here, $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2$ are nonnegative and nonzero vectors, whose dimensions are $1 \times p_1$ and $1 \times p_2$, respectively, with $(\boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2)\boldsymbol{e} = 1$. Also, for $i = 1, 2, 3, 4, 5, 6$, $\boldsymbol{T}_{ii}$ is an order p_i *PH* generator, and for $(i, j) \in \{(1, 2), (1, 5), (1, 6), (3, 4), (4, 3), (4, 5), (4, 6), (5, 3), (6, 3)\}$, $\boldsymbol{T}_{ij}$ is a $p_i \times p_j$ nonnegative and nonzero matrix. For this structure the absorption rate vector has the form

$$\boldsymbol{t} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \boldsymbol{t}_5 \\ \boldsymbol{t}_6 \end{pmatrix} \tag{39}$$

where $\boldsymbol{t}_5 = -(\boldsymbol{T}_{53} + \boldsymbol{T}_{55})\boldsymbol{e}$ and $\boldsymbol{t}_6 = -(\boldsymbol{T}_{63} + \boldsymbol{T}_{66})\boldsymbol{e}$ are nonnegative and nonzero vectors of dimensions $p_5 \times 1$ and $p_6 \times 1$, respectively. All zero matrices and $\boldsymbol{e}$ vectors have the appropriate dimension.

The simplest *PH* distribution with representation $(\boldsymbol{\alpha}, \boldsymbol{T})$ given by Eqs. 37 and 38 is where the length of stay in each unit is exponentially distributed. In this case the *PH* distribution that models the overall length of stay would be an order 6 *PH* distribution that does not have a Coxian representation because of the cycling. In order to estimate the parameters for this *PH* distribution one method is to fit an exponential distribution to the data from each unit, and then determine the proportions of patients that move between the units to estimate the off-diagonal entries of $\boldsymbol{T}$. Another method is to fit an order 6 *PH* distribution with the required structure to the whole length of stay data. The first method is probably the simplest in terms of the fitting procedure, and more accurate, but requires much more patient data. To get a better fit, however, we may decide to use a *PH* or Coxian distribution to model the time spent in each unit, and then combine them to create a larger *PH* distribution. In this case, fitting the data from each unit separately (given that the data is readily available) would be relatively simple using, say, EMpht, but estimating the parameters of the off diagonal matrices could be problematic. Fitting the entire distribution as a whole may be computationally infeasible as the order of the fitting *PH* distribution may be quite high.

More sophisticated methods exist to calculate performance measures such as steady state probabilities, blocking probabilities, expected waiting times, and mean queue lengths, in queues whose arrival and service times are modelled with *PH* distributions. We have already seen in Gorunescu et al. [22] that patient flow through a hospital ward can be modelled with a *M/PH/c/c* queue. The branch of computational probability known as *matrix-analytic methods* (see Neuts [39],

or Latouche and Ramaswami [29]) contains a vast literature on stochastic models that use *PH* distributions. Matrix-analytic methods deal with the analysis of stochastic models, particularly queueing systems, using a matrix formalism to develop algorithmically tractable solutions. The ever-increasing ability of computers to perform numerical calculations has supported the growing interest in this area. More sophisticated models such as the *PH* renewal process, the Markov-modulated Poisson process, the Markovian arrival process, and the quasi-birth-and-death (*QBD*) process (see Neuts [39, 41, 42], Latouche and Ramaswami [29], Asmussen [2], and Bini et al. [6], and the references therein) could be implemented in health and social care modelling. The mathematics for these models is involved but the modelling power is considerable. For example, Au et al. [5] developed and analysed a queueing model that predicts when the emergency department at the Royal Melbourne Hospital needs to go on ambulance bypass.

5 Conclusion

In this paper we have introduced *PH* distributions, given a brief overview of some *PH* fitting and approximation methods, and presented a comprehensive, although not exhaustive, literature review on where *PH* distributions have been used in the healthcare sector. In the last section we suggested some ways in which *PH* distributions could be further utilized in healthcare modelling. In particular we have made the following recommendations:

1. More general *PH* representations should be considered for fitting length of stay and interarrival time data because of the extra flexibility. The example given, where a general *PH* fit of order 6 was compared to five other *PH* fits, demonstrates clearly what these representations can offer. In fact, the author knows of no other example where a low order (6) non-Coxian *PH* fit is comparable to a high order (25) Coxian fit.
2. More sophisticated models that use general *PH* distributions should be considered in modelling healthcare systems. In Section 3 we saw that all authors used Coxian distributions to fit length of stay data. In most cases this was justified because the process being modelled had patients moving in a left-right manner (that is, with no cycling). However, as suggested in Section 4, healthcare systems have patients moving in a more complicated manner (for example, systems with readmissions to particular units) and non-Coxian *PH* representations can model these situations better than Coxian representations.
3. The wider literature on matrix-analytic methods should be consulted when developing stochastic models used in modelling healthcare systems. Here, there is the potential to develop more powerful models that incorporate *PH* distributions (and other models, see above) to describe systems in healthcare.

It is encouraging to note that, more recently, the use of *PH* distributions has increased. Of the thirteen papers reviewed in Section 3, eight have been published in the last four years. Despite this, the use of *PH* distributions needs to become more widespread, and the level of sophistication of the models where they are used needs to increase. If this is done, as has been the case in other industries (for example, telecommunications and finance), more powerful mathematical models can be developed to address and solve some of the important problems in the healthcare sector. It is hoped that, with this paper, modellers will not only use *PH* distributions more widely, but also look to the established literature on matrix-analytic methods and stochastic modelling for the mathematical tools required to develop suitable models that will help address and solve problems in the sector.

Acknowledgements This research is supported in part by the Australian Research Council linkage grant number LP0349153.

References

1. Asmussen S (1997) Phase-type distributions and related point processes: fitting and recent advances. In: Chakravarthy SR, Alfa AS (eds) Matrix-analytic methods in stochastic models, lecture notes in pure and applied mathematics, vol 183. Marcel Dekkar, New York, pp 137–149
2. Asmussen S (2003) Applied probability and queues, 2nd edn. Springer, Berlin Heidelberg New York
3. Asmussen S, Nerman O, Olsson M (1996) Fitting phase-type distributions via the *EM* algorithm. Scand J Statist 23:419–441
4. Asmussen S, Olsson M (1998) Phase-type distributions (update). In: Kotz S, Read CB, Banks DL (eds) Encyclopedia of statistical science update, vol 2. Wiley, New York, pp 525–530
5. Au L, Byrnes GB, Bain CA, Fackrell M, Brand C, Campbell DA, Taylor PG (2008) Predicting overflow in an emergency department. IMA J Manag Math (in press)
6. Bini DA, Latouche G, Meini B (2005) Numerical methods for structured Markov chains. Oxford University Press, Oxford
7. Bobbio A, Cumani A (1992) *ML* estimation of the parameters of a *PH* distribution in triangular canonical form. In: Balbo G, Serazzi G (eds) Computer performance evaluation. Elsevier, Amsterdam, pp 33–46

8. Cox DR (1955) A use of complex probabilities in the theory of stochastic processes. Proc Camb Philos Soc 5:313–319
9. Cumani A (1982) On the canonical representation of homogeneous Markov processes modelling failure-time distributions. Microelectron Reliab 22:583–602
10. Dempster AP, Laird NL, Rubin DB (1977) Maximum likelihood from incomplete data via the *EM* algorithm. With discussion. J R Stat Soc Ser B 39:1–38.
11. Erlang AK (1917–1918) Solution of some problems in the theory of probabilities of significance in automatic telephone exchanges. Post Off Electr Eng J 10:189–197
12. Fackrell M (2003) Characterization of matrix-exponential distributions. PhD thesis, School of Applied Mathematics, University of Adelaide, South Australia
13. Faddy MJ (1990) Compartmental models with phase-type residence-time distributions. Appl Stoch Models Data Anal 6:121–127
14. Faddy MJ (1993) A structured compartmental model for drug kinetics. Biometrics 49:243–248
15. Faddy MJ (1994) Examples of fitting structured phase-type distributions. Appl Stoch Models Data Anal 10:247–255
16. Faddy MJ (1995) Phase-type distributions for failure times. Math Comput Model 22:63–70
17. Faddy MJ (1998) On inferring the number of phases in a Coxian phase-type distribution. Commun Stat Stoch Models 14:407–417
18. Faddy MJ (2002) Penalized maximum likelihood estimation of the parameters in a Coxian phase-type distribution. In: Latouche G, Taylor P (eds) Matrix-analytic methods: theory and applications. World Scientific, Singapore, pp 107–114
19. Faddy MJ, McClean SI (1999) Analysing data on lengths of stay of hospital patients using phase-type distributions. Appl Stoch Models Bus Ind 15:311–317
20. Faddy MJ, McClean SI (2005) Markov chain modelling for geriatric patient care. Methods Inf Med 44:369–373
21. Faddy MJ, Taylor GT (2003) Stochastic modelling of the onset of bronchiolitis obliterans syndrome following lung transplantations: an analysis of risk factors. Math Comput Model 38:1185–1189
22. Gorunescu F, McClean SI, Millard PH (2002) A queueing model for bed-occupancy management and planning of hospitals. J Oper Res Soc 53:19–24
23. Gribbin JO, McClean SI (1990) Modelling the duration of spells of withdrawal from a female dominated labour force. J Appl Stat 17:369–381
24. Hampel K (1997) Modelling phase-type distributions. Honours Thesis, Department of Statistics, University of Adelaide, South Australia
25. He Q-M, Zhang H (2005) A note on unicyclic representations of phase type distributions. Stoch Models 21:465–483
26. Horváth A, Telek M (2000) Approximating heavy tailed behaviour with phase type distributions. In: Latouche G, Taylor P (eds) Advances in algorithmic methods for stochastic models. Notable Publications, Neshanic Station, pp 191–213
27. Horváth A, Telek M (2007) A canonical representation of order 3 phase type distributions. In: Wolter K (ed) Formal methods and stochastic models for performance evaluation, Lecture notes in computer science, vol 4748. Springer, Berlin Heidelberg New York, pp 48–62
28. Johnson MA (1993) Selecting parameters of phase distributions: combining nonlinear programming, heuristics, and Erlang distributions. ORSA J Comput 5:69–83
29. Latouche G, Ramaswami V (1999) Introduction to matrix analytic methods in stochastic models. SIAM, Philadelphia
30. Marshall AH, McClean SI (2003) Conditional phase-type distributions for modelling length of stay in hospital. Int Trans Oper Res 10:565–576
31. Marshall AH, Vasilakis C, El-Darzi E (2005) Length of stay-based patient flow models: recent developments and future directions. Health Care Manage Sci 8:213–220
32. McClean SI (1976) The two-stage model of personnel behaviour. J R Stat Soc Ser A 139:205–217
33. McClean SI, Faddy MJ, Millard PH (2005) Markov model-based clustering for efficient patient care. In: Proceedings of the 18th IEEE symposium on computer-based medical systems. IEEE, Piscataway, pp 467–472
34. McClean SI, Gribbin JO (1987) Estimation for incomplete manpower data. Appl Stoch Models Data Anal 3:13–25
35. McClean SI, Millard PH (1993) Patterns of length of stay after admission in geriatric medicine: an event history approach. Statistician 42:263–274
36. McLachlan GJ, Krishnan T (1997) The EM algorithm and extensions. Wiley, New York
37. Nelder JA, Mead R (1965) A simplex method for function minimization. Comput J 7:308–313
38. Neuts MF (1975) Probability distributions of phase-type. In: Liber amicorum Prof. Emeritus H. Florin, Department of Mathematics, University of Louvain, Louvain, pp 173–206
39. Neuts MF (1981) Matrix-geometric solutions in stochastic models: an algorithmic approach. The John Hopkins University Press, Baltimore
40. Neuts MF (1984) The abscissa of convergence of the Laplace-Stieltjes transform of a *PH* distribution. Commun Stat Simul Comput 13:367–373
41. Neuts MF (1989) Structured stochastic matrices of $M/G/1$ type and their applications. Marcel Dekkar, New York
42. Neuts MF (1995) Algorithmic probability. Chapman and Hall, London
43. Norris JR (1997) Markov chains. Cambridge University Press, Cambridge
44. O'Cinneide CA (1989) On non-uniqueness of representations of phase-type distributions. Commun Stat Stoch Models 5:247–259
45. O'Cinneide CA (1999) Phase-type distributions: open problems and a few properties. Commun Stat Stoch Models 15:731–757
46. Olsson M (1996) Estimation of phase-type distributions from censored data. Scand J Statist 23:443–460
47. Ross SM (1993) Introduction to probability models, 5th edn. Academic, San Diego
48. Schmickler L (1992) MEDA: mixed Erlang distributions as phase-type representations of empirical distribution functions. Commun Stat Stoch Models 8:131–156
49. Seneta E (2006) Non-negative matrices and markov chains, revised printing. Springer, New York
50. Shaked M, Shanthikumar JG (1985) Phase-type distributions. In: Kotz S, Johnson NL, Read CB (eds) Encyclopedia of statistical science, vol 6. Wiley, New York, pp 709–715
51. Vasilakis C, Marshall AH (2005) Modelling nationwide hospital length of stay: opening the black box. J Oper Res Soc 56:862–869
52. Xie H, Chaussalet TJ, Millard PH (2005) A continuous time Markov model for the length of stay of elderly people in institutional long-term care. J R Stat Soc Ser A 168:51–61