

# Using Phase-type Distributions in Health and Social Care Modelling

Mark Fackrell

Department of Mathematics and Statistics  
University of Melbourne  
Victoria 3010  
AUSTRALIA

email: mfackrel@ms.unimelb.eu.au

## 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 health and social care 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 health and social care industry, and propose some ideas on how they could be further utilized.

Keywords: phase-type distribution, Coxian distribution, patient flow

## 1 Introduction

Since their introduction by Neuts [35] 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.  $PH$  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 [10], 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 [35] 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 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<sup>®</sup>.

For a comprehensive theoretical treatment of  $PH$  distributions see Neuts [36, Chapter 2]. Latouche and Ramaswami [26, 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 [47], 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 health and social care 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 [7]) 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 (for example, in fitting), but at the expense of developing more versatile models. Second, the health and social care 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 practitioners and used in a broader variety of ways in the health and social care 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 health and social care industry. In Section 4 we propose some ways in which  $PH$  distributions could be further utilized in health and social care 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 [40, Section 2.3] or Ross [44, 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

$$P(T \leq t) = F(t) = 1 - e^{-\lambda t}. \quad (2.1)$$

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

$$f(t) = F'(t) = \lambda e^{-\lambda t}. \quad (2.2)$$

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  $\lambda$ , 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 [40, pages 70–71] or Ross [44, 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 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 [40, Chapter 2] or Ross [44, Chapter 6] for further properties of Markov chains.

Figure 2.1 shows the *state transition diagram* for a *finite-state continuous-time Markov chain*. The Markov chain consists of four states, 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  $\frac{4}{12} = \frac{1}{3}$ , respectively. If we chose state 0 we stop, but if we chose state 2 we spend an exponentially distributed length of time with  $\lambda = 10$  there,

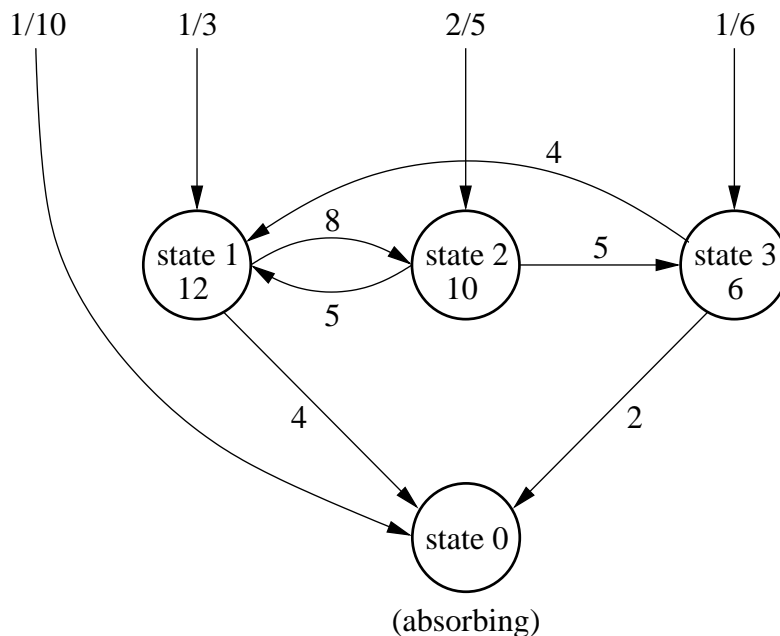


Figure 2.1: State transition diagram of a 4-state continuous-time Markov chain with one absorbing state.

and so on until absorption. The various rates have been chosen so that *with probability one* absorption occurs eventually.

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*

$$(\alpha_0, \boldsymbol{\alpha}) = \left( \frac{1}{10} \quad \frac{1}{3} \quad \frac{2}{5} \quad \frac{1}{6} \right) \tag{2.3}$$

which governs the selection of the initial state,  $\alpha_0$  being the point mass at zero.

3. An *infinitesimal generator*

$$\mathbf{Q} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 4 & -12 & 8 & 0 \\ 0 & 5 & -10 & 5 \\ 2 & 4 & 0 & -6 \end{pmatrix} \tag{2.4}$$

which governs the transitions between states.

The rows (labelled 0, 1, 2, and 3) of  $\mathbf{Q}$  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  $\alpha_0$  is rarely necessary.

## 2.3 Phase-type Distributions

Consider a continuous-time Markov chain on a finite state space  $S = \{0, 1, 2, \dots, p\}$  where state 0 is absorbing. Let the initial state probability distribution be  $(\alpha_0, \boldsymbol{\alpha}) = (\alpha_0, \alpha_1, \dots, \alpha_p)$  (with  $\sum_{i=0}^p \alpha_i = 1$ ) and the infinitesimal generator be  $\mathbf{Q}$ . The random variable that is defined as the time to absorption, is said to have a (*continuous*) *PH* distribution, see Neuts [36].

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

$$\mathbf{Q} = \begin{pmatrix} 0 & \mathbf{0} \\ \mathbf{t} & \mathbf{T} \end{pmatrix}. \quad (2.5)$$

Here,  $\mathbf{0}$  is a  $1 \times p$  vector of zeros. The vector  $\mathbf{t} = (t_{10}, t_{20}, \dots, t_{p0})'$  (the prime denoting transpose) where, for  $i = 1, 2, \dots, p, t_{i0} \geq 0$ , with at least one of the  $t_{i0}$ s positive, is the absorption rate from state  $i$ . The  $p \times p$  matrix  $\mathbf{T} = [t_{ij}]$  is such that, for  $i, j = 1, 2, \dots, p$ , with  $i \neq j$ ,

$$t_{ij} \geq 0, \quad (2.6)$$

and

$$t_{ii} = - \sum_{\substack{j=0 \\ j \neq i}}^p t_{ij}, \quad (2.7)$$

that is,  $\mathbf{t} = -\mathbf{T}\mathbf{e}$  where  $\mathbf{e}$  is a  $p \times 1$  vector of ones. The *PH* distribution is said to have a *representation*  $(\boldsymbol{\alpha}, \mathbf{T})$  of *order*  $p$ . The matrix  $\mathbf{T}$  is referred to as a *PH generator*. The point mass at zero  $\alpha_0$  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 non-absorbing state is transient. This statement is equivalent to  $\mathbf{T}$  being invertible, see Neuts [36, Lemma 2.2.1], or Latouche and Ramaswami [26, Theorem 2.4.3]. An additional requirement on the *PH* representation  $(\boldsymbol{\alpha}, \mathbf{T})$  is that there are no superfluous phases. That is, each phase in the Markov chain defined by  $\boldsymbol{\alpha}$  and  $\mathbf{T}$  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 [36, page 48]. This condition is equivalent to the matrix

$$\mathbf{T}^* = \mathbf{T} - (1 - \alpha_0)^{-1} \mathbf{T} \mathbf{e} \boldsymbol{\alpha}, \quad (2.8)$$

being irreducible. For the definition of an *irreducible matrix* see Seneta [46, pages 18 and 46].

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

$$F(u) = \begin{cases} \alpha_0, & t = 0 \\ 1 - \boldsymbol{\alpha} \exp(\mathbf{T}t) \mathbf{e}, & t > 0, \end{cases} \quad (2.9)$$

see Neuts [36, Lemma 2.2.2], or Latouche and Ramaswami [26, Theorem 2.4.1]. Differentiating (2.9) with respect to  $t$  gives the corresponding density function, defined for  $t > 0$ ,

$$f(t) = -\boldsymbol{\alpha} \exp(\mathbf{T}t) \mathbf{T} \mathbf{e}. \quad (2.10)$$

The Laplace-Stieltjes transform (*LST*) of (2.9), which is defined for  $s \in \mathbb{C}$  such that  $\Re(s) > \delta$  where  $\delta$  is the real and negative eigenvalue of maximal real part of  $\mathbf{T}$  (see Neuts [37]), is given by

$$\begin{aligned} \phi(s) &= \int_0^\infty e^{-st} dF(t) \\ &= -\boldsymbol{\alpha}(s\mathbf{I} - \mathbf{T})^{-1} \mathbf{T} \mathbf{e} + \alpha_0. \end{aligned} \quad (2.11)$$

For  $k = 1, 2, \dots$ , differentiating (2.11)  $k$  times with respect to  $s$  and letting  $s \rightarrow 0$  gives the  $k$ th noncentral moment

$$m_k = (-1)^k k! \boldsymbol{\alpha} \mathbf{T}^{-k} \mathbf{e}. \quad (2.12)$$

We now give some examples of *PH* distributions.

1. The exponential distribution. The minimal representation is

$$\boldsymbol{\alpha} = (1) \quad (2.13)$$

$$\mathbf{T} = (-\lambda). \quad (2.14)$$

2. The order  $p$  *generalized Erlang* distribution. This distribution can be described using a state transition diagram that has  $p$  states in series, see Figure 2.2. It is easy to see, without loss of generality, that the states can be ordered so that the rates  $0 < \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_p$ . The representation for the generalized Erlang distribution

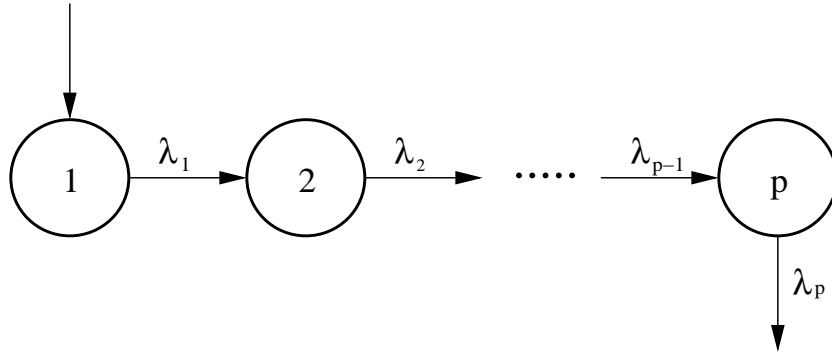


Figure 2.2: State transition diagram for an order  $p$  generalized Erlang distribution.

corresponding to the state transition diagram is

$$\boldsymbol{\alpha} = (1 \ 0 \ \dots \ 0) \quad (2.15)$$

$$\mathbf{T} = \begin{pmatrix} -\lambda_1 & \lambda_1 & 0 & \dots & 0 \\ 0 & -\lambda_2 & \lambda_2 & \dots & 0 \\ 0 & 0 & -\lambda_3 & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & -\lambda_p \end{pmatrix}. \quad (2.16)$$

Figure 2.3 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

$$f(t) = \frac{\lambda^p t^{p-1} e^{-\lambda t}}{p!}. \quad (2.17)$$

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

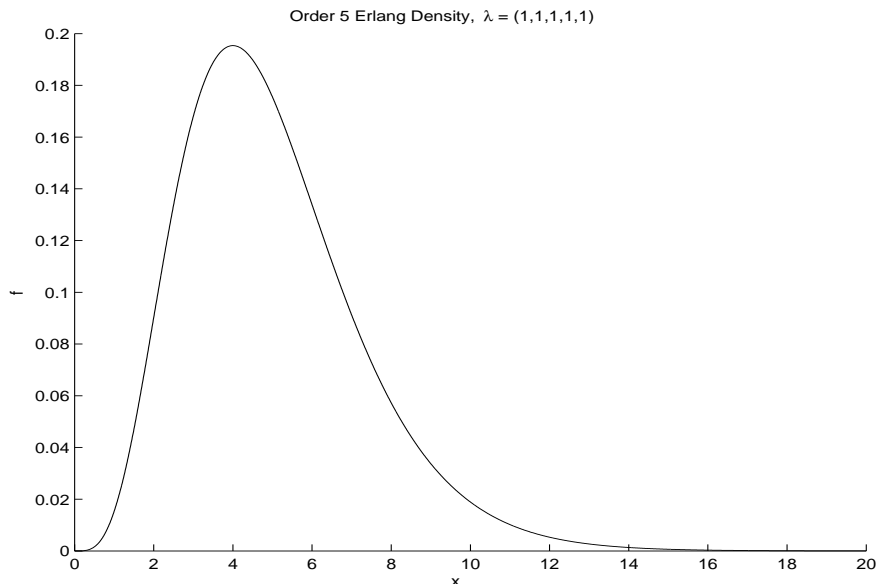


Figure 2.3: Density function of a order 5 Erlang distribution with  $\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = \lambda_5 = 1$ .

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 Figure 2.4. Clearly, without loss of generality, the states can be ordered so that the rates  $0 < \lambda_1 < \lambda_2 < \dots < \lambda_p$ . The corresponding representation is

$$\boldsymbol{\alpha} = ( \alpha_1 \quad \alpha_2 \quad \dots \quad \alpha_p ) \quad (2.18)$$

$$\mathbf{T} = \begin{pmatrix} -\lambda_1 & 0 & \dots & 0 \\ 0 & -\lambda_2 & \ddots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & -\lambda_p \end{pmatrix}, \quad (2.19)$$

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

$$f(t) = \sum_{i=1}^p \alpha_i \lambda_i e^{-\lambda_i t}, \quad (2.20)$$

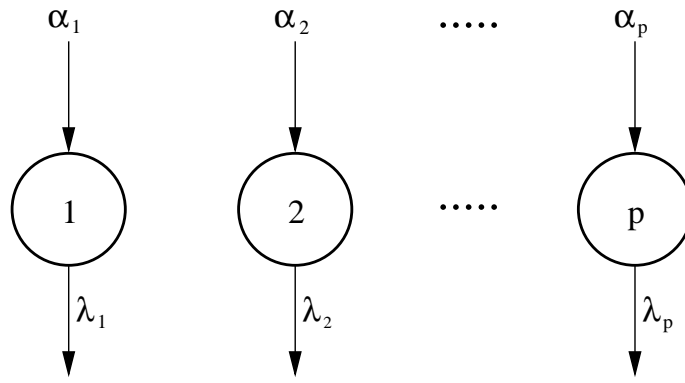


Figure 2.4: State transition diagram for an order  $p$  hyperexponential distribution.

where, for  $i = 1, 2, \dots, p$ ,  $\alpha_i > 0$  and  $\sum_{i=1}^p \alpha_i = 1$ . Figure 2.5 shows an example of the density function of a order 3 hyperexponential distribution. Hyperexponential

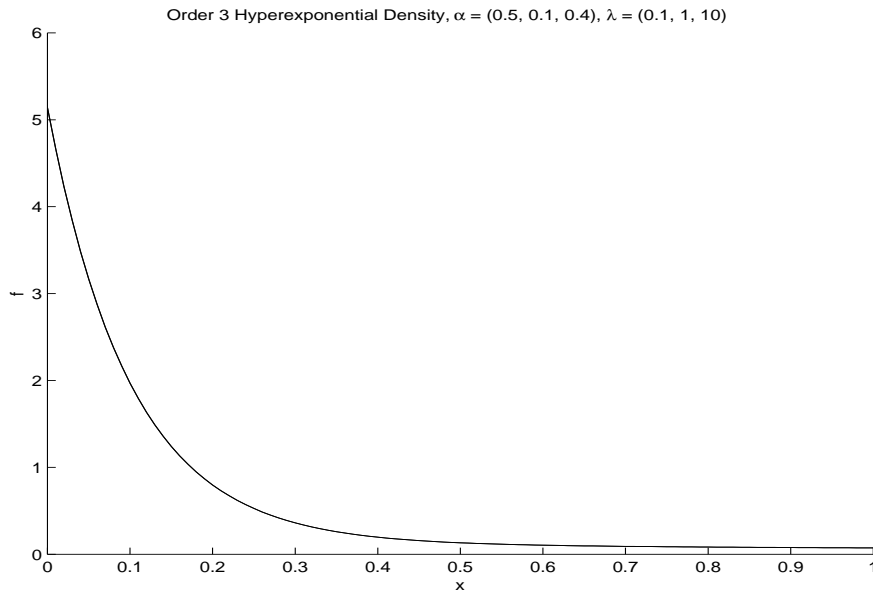


Figure 2.5: Density function of an order 3 hyperexponential distribution with  $\alpha = (0.5, 0.1, 0.4)$ , and  $\lambda_1 = 0.1$ ,  $\lambda_2 = 1$ , and  $\lambda_3 = 10$ .

distributions have coefficient of variation greater than or equal to one (the 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 Figure 2.6. They have representations of the form

$$\boldsymbol{\alpha} = ( \alpha_1 \quad \alpha_2 \quad \dots \quad \alpha_p ) \tag{2.21}$$

$$\mathbf{T} = \begin{pmatrix} -\lambda_1 & \lambda_1 & 0 & \dots & 0 \\ 0 & -\lambda_2 & \lambda_2 & \dots & 0 \\ 0 & 0 & -\lambda_3 & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & -\lambda_p \end{pmatrix}. \tag{2.22}$$



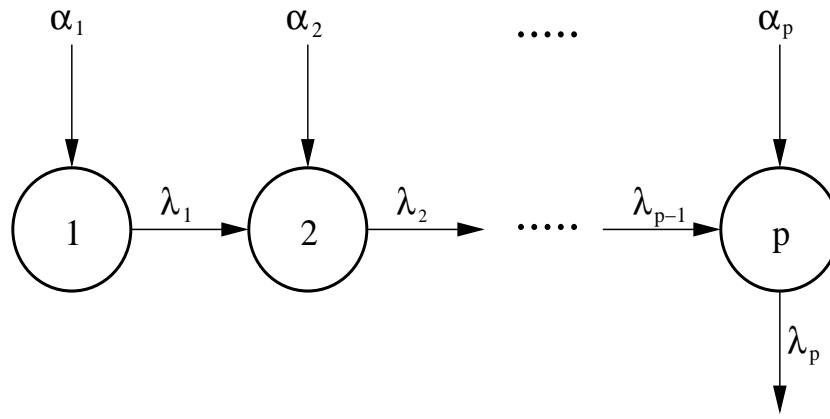


Figure 2.6: State transition diagram for an order  $p$  Coxian distribution.

Although it is not obvious, in this case, without loss of generality, the states can be ordered so that the rates  $0 < \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_p$ , see Cumani [8] or O’Cinneide [41]. Figure 2.7 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.

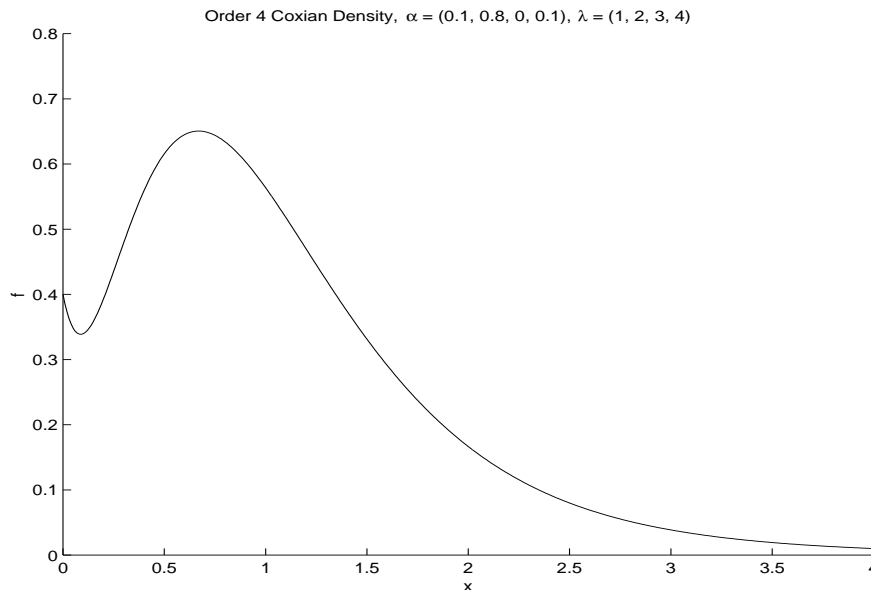


Figure 2.7: Density function of an order 4 Coxian distribution with  $\alpha = (0.1, 0.8, 0, 0.1)$ , and  $\lambda_1 = 1$ ,  $\lambda_2 = 2$ ,  $\lambda_3 = 3$ ,  $\lambda_4 = 4$ .

5. The *acyclic*, or *triangular PH* (*TPH*) distribution. This type of *PH* distribution have generators that are upper triangular matrices. Cumani [8] (see also O’Cinneide [41]) 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 Figure 2.8. They have representations of the form

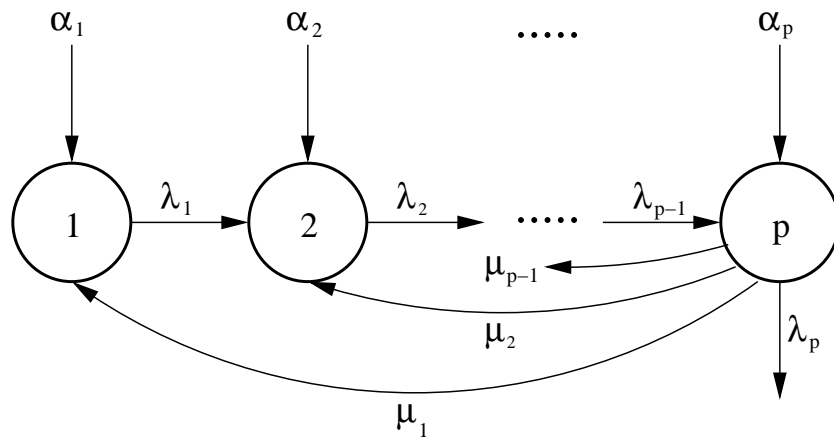


Figure 2.8: State transition diagram for an order  $p$  unicyclic distribution.

$$\boldsymbol{\alpha} = (\alpha_1 \ \alpha_2 \ \dots \ \alpha_p) \quad (2.23)$$

$$\mathbf{T} = \begin{pmatrix} -\lambda_1 & \lambda_1 & 0 & \dots & 0 & 0 \\ 0 & -\lambda_2 & \lambda_2 & \dots & 0 & 0 \\ 0 & 0 & -\lambda_3 & \ddots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -\lambda_{p-1} & \lambda_{p-1} \\ \mu_1 & \mu_2 & \mu_3 & \dots & \mu_{p-1} & -\lambda_p \end{pmatrix}, \quad (2.24)$$

where for  $i = 1, 2, \dots, p - 1$ ,  $\mu_i \geq 0$ ,  $0 < \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_p$ , and  $\lambda_p > \sum_{i=1}^{p-1} \mu_i$ , see O’Cinneide [42, Section 7]. Figure 2.9 shows the density function for an order 5 unicyclic distribution. It was conjectured in O’Cinneide [42, Conjecture 4] that

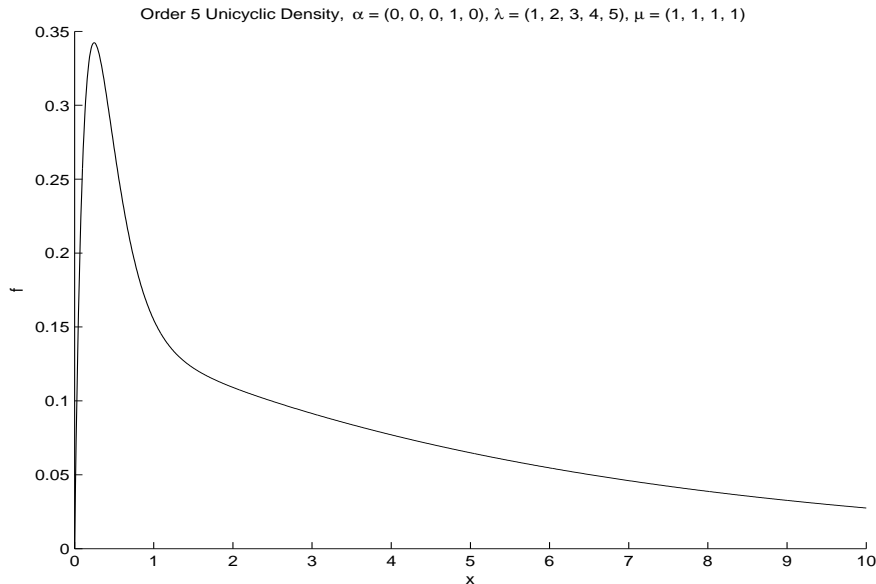


Figure 2.9: Density function of an order 5 unicyclic distribution with  $\boldsymbol{\alpha} = (0, 0, 0, 1, 0)$ ,  $\lambda_1 = 1$ ,  $\lambda_2 = 2$ ,  $\lambda_3 = 3$ ,  $\lambda_4 = 4$  and  $\lambda_5 = 5$ , and  $\mu_1 = \mu_2 = \mu_3 = \mu_4 = 1$ .

every  $PH$  distribution of order  $p$  has a unicyclic representation of the same order,

however, He and Zhang [23] showed that this is not, in general, the case.

## 2.4 Fitting 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  $\mathbf{T}$  so that they best fit the data in some sense. In *approximating* a probability distribution with a *PH* distribution the parameters  $\boldsymbol{\alpha}$  and  $\mathbf{T}$  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, Nerman, and Olsson [3], Bobbio and Cumani [6], and Johnson [25] for more information.

*Maximum likelihood estimation* has been the most popular method used to fit data and approximate distributions with *PH* distributions. Asmussen, Nerman, and Olsson [3] (see also Asmussen [1]) developed an expectation-maximization (*EM*) algorithm (see Dempster, Laird, and Rubin [9], or McLachlan and Krishnan [33]) to calculate maximum likelihood parameter estimates for general *PH* distributions when fitted to empirical data. In a companion paper Olsson [43] 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<sup>1</sup>. Bobbio and Cumani [6] 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 [13], [14], and [15], Faddy and McClean [17], and Hampel [22] used maximum likelihood estimation to fit Coxian distributions to real data. They used existing MATLAB<sup>®</sup> or S-PLUS<sup>®</sup> routines (for example the Nelder-Mead algorithm in MATLAB<sup>®</sup>, see Nelder and Mead [34]) to perform the required parameter estimation. In Faddy [16] a penalized maximum likelihood method was developed to fit Coxian distributions to data. Coxian representations where  $\mathbf{T}$  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 [25] 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 [45] 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. Horvath and

---

<sup>1</sup><http://home.imf.au.dk/asmus/pspapers.html>

Telek [24] 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 [6] was used to perform the minimization. The tail was approximated with a hyperexponential distribution using a heuristic method. Faddy [11] and [12] 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 Health and Social Care Industry

There have been a number of papers written on the application of *PH* distributions in the health and social care 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 health and social care sector.

Faddy [11] used a two-compartment model, such as the one depicted in Figure 3.1, 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. In Faddy [12] a more complex two-compartment model (see Figure 3.2) was

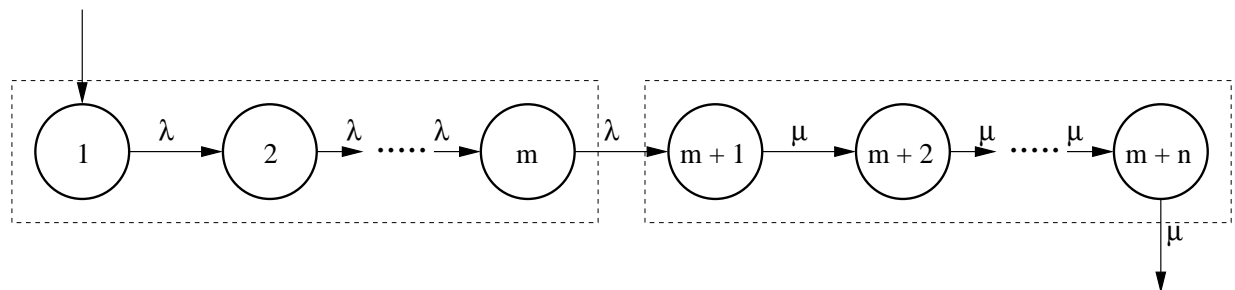


Figure 3.1: Two-compartment model to model the outflow of labelled red blood cells injected into a rat liver.

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.

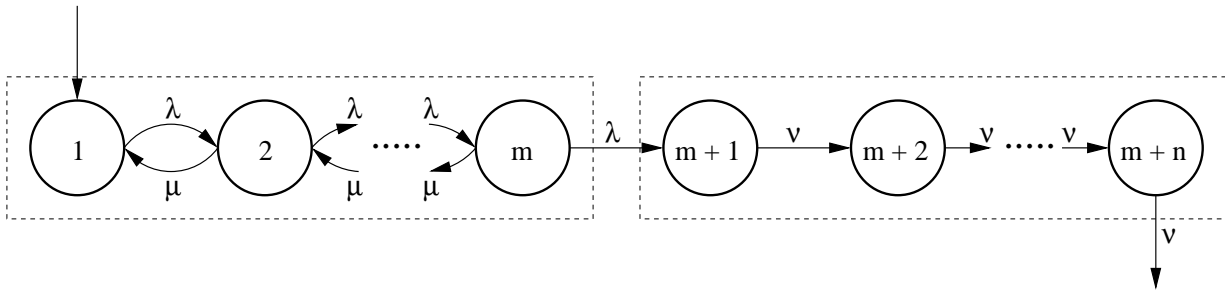


Figure 3.2: Two-compartment model to model the diffusion and clearance of a drug injected into a body organ.

Faddy [13] 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

$$\boldsymbol{\alpha} = ( 1 \ 0 \ 0 ) \quad (3.1)$$

$$\mathbf{T} = \begin{pmatrix} -(\lambda_1 + \mu_1) & \lambda_1 & 0 \\ 0 & -(\lambda_2 + \mu_2) & \lambda_2 \\ 0 & 0 & -\mu_3 \end{pmatrix}, \quad (3.2)$$

which corresponds to the state transition diagram shown in Figure 3.3. This “drop out”

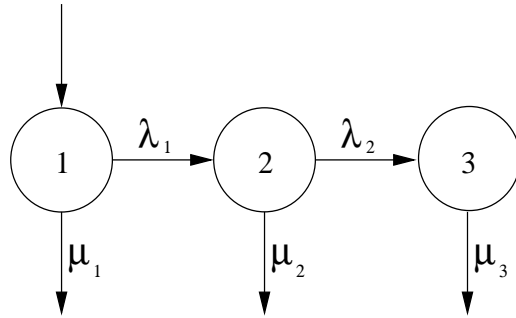


Figure 3.3: State transition diagram for a “drop out” order three Coxian distribution.

representation is equivalent to the “drop in” representation given by equations (2.21) and (2.22). Here, it was noticed that in the fitted  $\mathbf{T}$ ,  $\lambda_1 + \mu_1 \approx \lambda_2 + \mu_2$ . 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 [16]. A similar approach was taken by Faddy and Taylor [19] 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  $\mathbf{T}$ , that is,

$$\lambda_i = \exp(c_i - b_1x_1 - b_2x_2 - b_3x_3), \quad i = 1, 2, \dots, p-1, \quad (3.3)$$

$$\mu_i = \exp(d_i - b_1x_1 - b_2x_2 - b_3x_3), \quad i = 1, 2, \dots, p, \quad (3.4)$$

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

McClellan and Millard [32], 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

$$f(t) = \rho e^{-\lambda_1 t} + (1 - \rho)e^{-\lambda_2 t}. \quad (3.5)$$

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 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  $\rho$  was estimated by the proportion of (either male or female) short stay patients, and  $\lambda_1$  and  $\lambda_2$  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 McClellan [17] Coxian distributions of increasing order were fitted to the male patient data used by McClellan and Millard [32] 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 [13]. Unlike McClellan and Millard [32], here the Coxian distribution was fitted first, and then an interpretation sought. If, for example, a three state model (see equations (3.1) and (3.2)) 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

$$p_1 = \frac{\mu_1}{\mu_1 + \lambda_1} \quad (3.6)$$

$$p_2 = \frac{\lambda_1 \mu_2}{(\mu_1 + \lambda_1)(\mu_2 + \lambda_2)} \quad (3.7)$$

$$p_3 = \frac{\lambda_1 \lambda_2}{(\mu_1 + \lambda_1)(\mu_2 + \lambda_2)}. \quad (3.8)$$

These mixing coefficients from the fitted distribution would model the proportion of short, medium, and long stay patients, respectively. Faddy and McClellan [17] 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 [12].

Faddy and McClellan [18] 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 McClellan, Faddy, and Millard [30] 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, McClean, and Millard [20] 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 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 [27] 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, Chausaulet, and Millard [49] 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 3.4 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 Figure 3.4 is

$$\boldsymbol{\alpha} = ( 1 \ 0 \ 0 \ 0 ) \tag{3.9}$$

$$\mathbf{T} = \begin{pmatrix} -(\lambda_1 + \mu_1 + \nu_1) & \lambda_1 & \nu_1 & 0 \\ 0 & -(\lambda_2 + \mu_2) & \lambda_2 & 0 \\ 0 & 0 & -(\lambda_3 + \mu_3) & \lambda_3 \\ 0 & 0 & 0 & -\mu_4 \end{pmatrix}. \tag{3.10}$$

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 equations (3.9) and (3.10) is Coxian since  $\mathbf{T}$  is an upper triangular matrix.

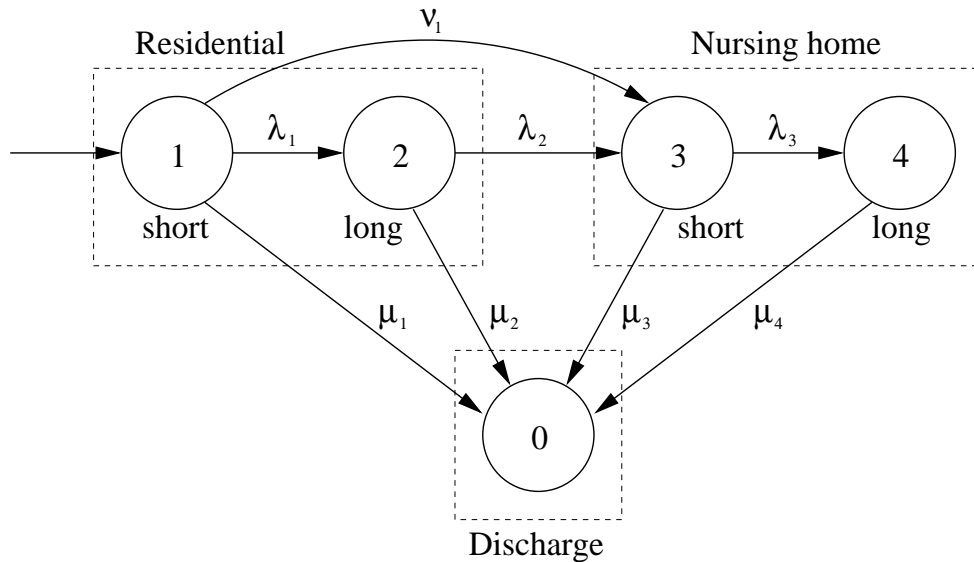


Figure 3.4: State transition diagram to model the length of stay in residential and nursing home care.

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, Vasilakis, and El-Darzi [28], and Vasilakis and Marshall [48]. 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 [48] 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 [21] (see also, McClean [29], and McClean and Gribbin [31]) 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 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 the Patient Care Process

As mentioned in the previous section, the use of *PH* distributions in the health and social care 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 health and social care modelling, in particular in modelling the patient care process.

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, Figure 4.1 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

$$\boldsymbol{\alpha} = ( 1 \ 0 \ 0 \ 0 \ 0 \ 0 ) \quad (4.1)$$

$$\mathbf{T} = \begin{pmatrix} -3.21 & 3.21 & 0 & 0 & 0 & 0 \\ 0 & -3.21 & 0 & 3.21 & 0 & 0 \\ 0.61 & 0 & -0.63 & 0 & 0.02 & 0 \\ 0 & 0 & 0 & -3.21 & 0 & 3.21 \\ 0 & 0 & 0.81 & 0 & -0.81 & 0 \\ 0 & 0 & 0 & 0 & 1.65 & -3.21 \end{pmatrix}. \quad (4.2)$$

The algorithm took approximately 23 minutes to perform 30,000 iterations, and the log-likelihood was  $-11706.92$ . The fit is shown in Figure 4.1. As we can see it looks very good. It is not Coxian as the eigenvalues of  $\mathbf{T}$  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.

More sophisticated models could be used to model lengths of stay and interarrival times in the health and social care industry. Consider the (simplified) schematic diagram for patient flow in a hospital shown in Figure 4.2. 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

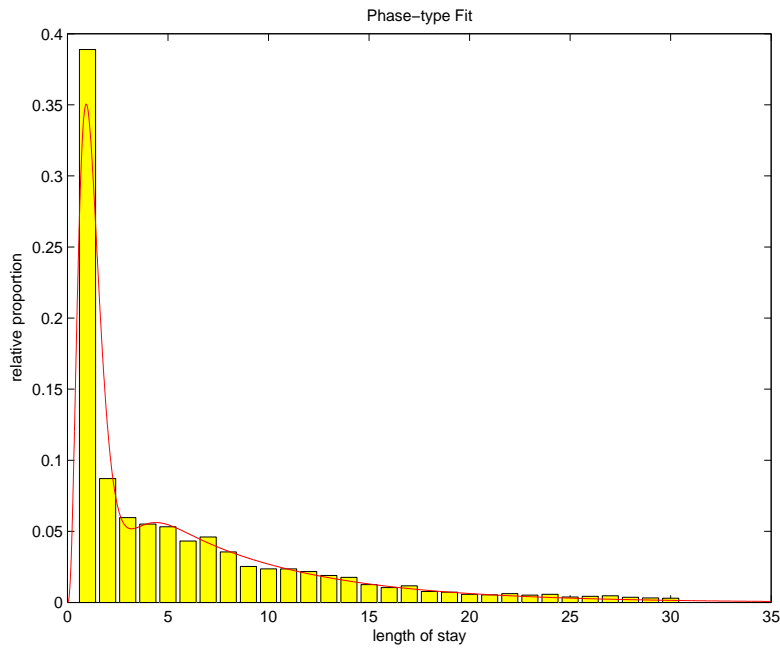


Figure 4.1: Order 6 *PH* fit to the length of stay histogram.

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

$$\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1 \quad \boldsymbol{\alpha}_2 \quad \mathbf{0} \quad \mathbf{0} \quad \mathbf{0} \quad \mathbf{0}) \quad (4.3)$$

$$\mathbf{T} = \begin{pmatrix} \mathbf{T}_{11} & \mathbf{T}_{12} & \mathbf{0} & \mathbf{0} & \mathbf{T}_{15} & \mathbf{T}_{16} \\ \mathbf{0} & \mathbf{T}_{22} & \mathbf{T}_{23} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{T}_{33} & \mathbf{T}_{34} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{T}_{43} & \mathbf{T}_{44} & \mathbf{T}_{45} & \mathbf{T}_{46} \\ \mathbf{0} & \mathbf{0} & \mathbf{T}_{53} & \mathbf{0} & \mathbf{T}_{55} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{T}_{63} & \mathbf{0} & \mathbf{0} & \mathbf{T}_{66} \end{pmatrix}. \quad (4.4)$$

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

$$\mathbf{t} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{t}_5 \\ \mathbf{t}_6 \end{pmatrix} \quad (4.5)$$

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

The simplest *PH* distribution with representation  $(\boldsymbol{\alpha}, \mathbf{T})$  given by (4.3) and (4.4) is where the length of stay in each unit is exponentially distributed. In this case the *PH*

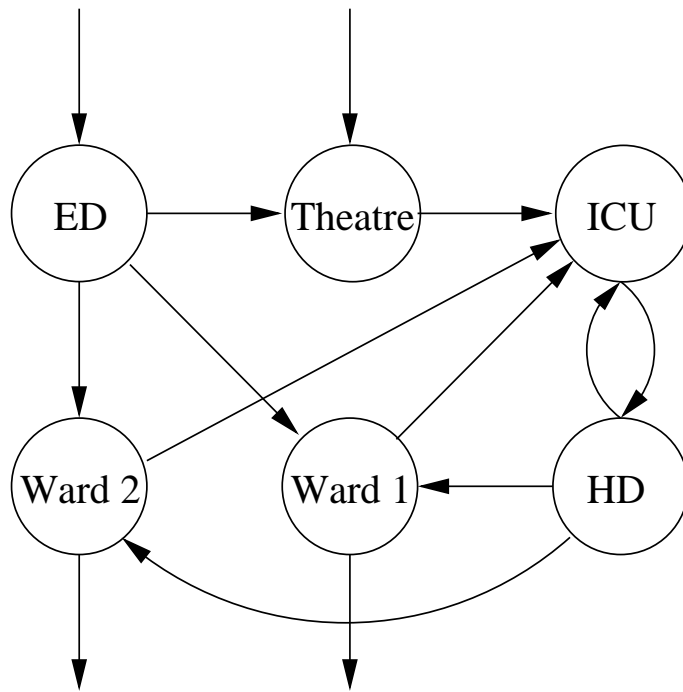


Figure 4.2: Schematic diagram for patient flow in a hospital.

distribution that models the overall length of stay would be an order 6 non-Coxian distribution 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  $\mathbf{T}$ . 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 Gorenescu, McClean, and Millard [20] 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 [36], or Latouche and Ramaswami [26]) contains a vast literature on stochastic models that use  $PH$  distributions. Matrix-analytic methods deals 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 ( $MMPP$ ), the Markovian arrival process ( $MAP$ ), and the quasi-birth-and-death ( $QBD$ ) process (see Neuts [36], [38], [39], Latouche and Ramaswami [26], Asmussen

[2], and Bini, Latouche, and Meini [5], 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.

## 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 health and social care sector. In the last section we suggested some ways in which  $PH$  distributions could be further utilized in health and social care modelling.

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 health and social care sector. It is hoped that, with this paper, practitioners 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

The author would like to thank the two anonymous referees for their invaluable comments and suggestions for this paper.

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

## References

- [1] Asmussen, S. Phase-type distributions and related point processes: Fitting and recent advances. In *Matrix-analytic Methods in Stochastic Models*, Chakravarty, S. R., Alfa, A. S., editors; Lecture Notes in Pure and Applied Mathematics; Marcel Dekkar: New York, 1997; Volume 183, 137–149.
- [2] Asmussen, S. *Applied Probability and Queues*, 2nd edition; Springer-Verlag: New York, 2003.
- [3] Asmussen, S.; Nerman, O.; Olsson, M. Fitting phase-type distributions via the  $EM$  algorithm. *Scandinavian Journal of Statistics* **1996**, *23*, 419–441.

- [4] Asmussen, S.; Olsson, M. Phase-type distributions (update). In *Encyclopedia of Statistical Science Update*, Kotz, S., Read, C. B., Banks, D. L., editors; John Wiley and Sons: New York, 1998; Volume 2, 525–530.
- [5] Bini, D. A.; Latouche, G.; Meini, B. *Numerical Methods for Structured Markov Chains*; Oxford University Press: Oxford, 2005.
- [6] Bobbio, A.; Cumani, A. *ML* estimation of the parameters of a *PH* distribution in triangular canonical form. In *Computer Performance Evaluation*, Balbo, G., Serazzi, G., editors; Elsevier Science Publishers, 1992; 33–46.
- [7] Cox, D. R. A use of complex probabilities in the theory of stochastic processes. *Proceedings of The Cambridge Philosophical Society* **1955**, *51*, 313–319.
- [8] Cumani, A. On the canonical representation of homogeneous Markov processes modelling failure-time distributions. *Microelectronics and Reliability* **1982**, *22*, 583–602.
- [9] Dempster, A. P.; Laird, N. L.; Rubin, D. B. Maximum likelihood from incomplete data via the *EM* algorithm. With discussion. *Journal of the Royal Statistical Society. Series B* **1977**, *39*, 1–38.
- [10] Erlang, A. K. Solution of some problems in the theory of probabilities of significance in automatic telephone exchanges. *The Post Office Electrical Engineer’s Journal* **1917–18**, *10*, 189–197.
- [11] Faddy, M. J. Compartmental models with phase-type residence-time distributions. *Applied Stochastic Models and Data Analysis* **1990**, *6*, 121–127.
- [12] Faddy, M. J. A structured compartmental model for drug kinetics. *Biometrics* **1993**, *49*, 243–248.
- [13] Faddy, M. J. Examples of fitting structured phase-type distributions. *Applied Stochastic Models and Data Analysis* **1994**, *10*, 247–255.
- [14] Faddy, M. J. Phase-type distributions for failure times. *Mathematical and Computational Modelling* **1995**, *22*, 63–70.
- [15] Faddy, M. J. On inferring the number of phases in a Coxian phase-type distribution. *Communications in Statistics – Stochastic Models* **1998**, *14*, 407–417.
- [16] Faddy, M. J. Penalized maximum likelihood estimation of the parameters in a Coxian phase-type distribution. In *Matrix-analytic Methods: Theory and Applications*, Latouche, G., Taylor, P., editors; World Scientific: Singapore, 2002; 107–114.
- [17] Faddy, M. J.; McClean, S. I. Analysing data on lengths of stay of hospital patients using phase-type distributions. *Applied Stochastic Models in Business and Industry* **1999**, *15*, 311–317.
- [18] Faddy, M. J.; McClean, S. I. Markov chain modelling for geriatric patient care. *Methods of Information in Medicine* **2005**, *44*, 369–373.
- [19] Faddy, M. J.; Taylor, G. T. Stochastic modelling of the onset of bronchiolitis obliterans syndrome following lung transplantations: An analysis of risk factors. *Mathematical and Computer Modelling* **2003**, *38*, 1185–1189.

- [20] Gorunescu, F.; McClean, S. I.; Millard, P. H. A queueing model for bed-occupancy management and planning of hospitals. *Journal of the Operational Research Society* **2002**, *53*, 19–24.
- [21] Gribbin, J. O.; McClean, S. I. Modelling the duration of spells of withdrawal from a female dominated labour force. *Journal of Applied Statistics* **1990**, *17*, 369–381.
- [22] Hampel, K. *Modelling Phase-Type Distributions*. Honours Thesis, Department of Statistics, University of Adelaide, South Australia, 1997.
- [23] He, Q-M.; Zhang, H. A note on unicyclic representations of phase type distributions. *Stochastic Models* **2005**, *21*, 465–483.
- [24] Horváth, A.; Telek, M. Approximating heavy tailed behaviour with phase type distributions. In *Advances in Algorithmic Methods for Stochastic Models*, Latouche, G., Taylor, P., editors; Notable Publications: New Jersey, 2000; 191–213.
- [25] Johnson, M. A. Selecting parameters of phase distributions: Combining nonlinear programming, heuristics, and Erlang distributions. *ORSA Journal on Computing* **1993**, *5*, 69–83.
- [26] Latouche, G.; Ramaswami, V. *Introduction to Matrix Analytic Methods in Stochastic Models*; SIAM: Philadelphia, 1999.
- [27] Marshall, A. H.; McClean, S. I. Conditional phase-type distributions for modelling length of stay in hospital. *International Transactions in Operational Research* **2003**, *10*, 565–576.
- [28] Marshall, A. H.; Vasilakis, C.; El-Darzi, E. Length of stay-based patient flow models: Recent developments and future directions. *Health Care Management Science* **2005**, *8*, 213–220.
- [29] McClean, S. I. The two-stage model of personnel behaviour. *Journal of the Royal Statistical Society, Series A* **1976**, *139*, 205–217.
- [30] McClean, S. I.; Faddy, M. J.; Millard, P. H. Markov model-based clustering for efficient patient care. *Proceedings of the 18th IEEE Symposium on Computer-Based Medical Systems* **2005**, 467–472.
- [31] McClean, S. I.; Gribbin, J. O. Estimation for incomplete manpower data. *Applied Stochastic Models and Data Analysis* **1987**, *3*, 13–25.
- [32] McClean, S. I.; Millard, P. H. Patterns of length of stay after admission in geriatric medicine: An event history approach. *The Statistician* **1993**, *42*, 263–274.
- [33] McLachlan, G. J.; Krishnan, T. *The EM Algorithm and Extensions*; John Wiley and Sons: New York, 1997.
- [34] Nelder, J. A.; Mead, R. A simplex method for function minimization. *The Computer Journal* **1965**, *7*, 308–313.
- [35] 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.

- [36] Neuts, M. F. *Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach*; The John Hopkins University Press: Baltimore, 1981.
- [37] Neuts, M. F. The abscissa of convergence of the Laplace-Stieltjes transform of a *PH* distribution. *Communications in Statistics – Simulation and Computation* **1984**, *13*, 367–373.
- [38] Neuts, M. F. *Structured Stochastic Matrices of M/G/1 Type and their Applications*; Marcel Dekkar: New York, 1989.
- [39] Neuts, M. F. *Algorithmic Probability*; Chapman and Hall: London, 1995.
- [40] Norris, J. R. *Markov Chains*; Cambridge University Press: Cambridge, 1997.
- [41] O’Cinneide, C. A. On non-uniqueness of representations of phase-type distributions. *Communications in Statistics – Stochastic Models* **1989**, *5*, 247–259.
- [42] O’Cinneide, C. A. Phase-type distributions: Open problems and a few properties. *Communications in Statistics – Stochastic Models* **1999**, *15*, 731–757.
- [43] Olsson, M. Estimation of phase-type distributions from censored data. *Scandinavian Journal of Statistics* **1996**, *23*, 443–460.
- [44] Ross, S. M. *Introduction to Probability Models*, 5th edition; Academic Press: San Diego, 1993.
- [45] Schmickler, L. MEDA: Mixed Erlang distributions as phase-type representations of empirical distribution functions. *Communications in Statistics – Stochastic Models* **1992**, *8*, 131–156.
- [46] Seneta, E. *Non-negative Matrices and Markov Chains*, revised printing; Springer: New York, 2006.
- [47] Shaked, M.; Shantikumar, J. G. Phase-type distributions. In *Encyclopedia of Statistical Science*, Kotz, S., Johnson, N. L., Read, C. B., editors; John Wiley and Sons: New York, 1985; Volume 6, 709–715.
- [48] Vasilakis, C.; Marshall, A. H. Modelling nationwide hospital length of stay: Opening the black box. *Journal of the Operational Research Society* **2005**, *56*, 862–869.
- [49] Xie, H.; Chausalet, T. J.; Millard, P. H. A continuous time Markov model for the length of stay of elderly people in institutional long-term care. *Journal of the Royal Statistical Society. Series A* **2005**, *168*, 51–61.