ITC 2/47
Journal of Information Technology and Control
Vol. 47 / No. 2 / 2018
pp. 197-208
DOI 10.5755/j01.itc.47.2.18169
© Kaunas University of Technology

On Structured Initial Solution Generation for Phase-Type Fitting with EM Method

Received 2017/09/08 | Accepted after revision 2018/01/25

http://dx.doi.org/10.5755/j01.itc.47.2.18169

On Structured Initial Solution Generation for Phase-Type Fitting with EM Method

M. Bražėnas
Department of Mathematical Modelling; Kaunas University of Technology; Studentų Str. 50-143, LT-51368, Kaunas, Lithuania; phone: +370 608 90779; e-mail: mindaugas.brazenas@yahoo.com

E. Valakevičius
Department of Mathematical Modelling; Kaunas University of Technology; Studentų Str. 50-140c, LT-51368, Kaunas, Lithuania; phone: +370 610 31290; e-mail: eimval@ktu.lt

Corresponding author: mindaugas.brazenas@yahoo.com

The choice of initial solutions when fitting with a phase-type distribution (PH), using the expectation maximization method (EM) is investigated. It is known that the EM method can converge to a local solution, especially when fitting with a general structure. The problem of how to choose an initial solution for which the EM method would converge to a global solution is still open. We contribute to the research of this problem by studying the use of structures for initial solution generation. The proposed approach is tested by fitting with four state phase-type distributions (PH(4)). Numerical results show that the EM method converges faster from initial solutions of various structures.

KEYWORDS: Phase-type fitting, expectation maximization, structure generation, sparse representation.

1. Introduction

Phase-type distributions are analytically tractable and can be used to simplify the analysis of various models, to mention a few: queuing [2, 9, 18], calculation methods in risk theory [3], telecommunication [13, 20] and others. Any positive distribution can be arbitrary precisely approximated by PH distribution when a number of states n tends to infinity [17]. However, when n increases, the complexity of fitting rapidly grows to the extent it becomes impractical. The number of states necessary to achieve a certain approximation precision depends on the properties of the target distribution, for example, on the coefficient

of variation [1]. The question of how to approximate with PH efficiently is still an open problem.

The general structure of PH distribution representation is over-parameterized. It can be shown that $2n-1$ parameters are sufficient to describe any PH distribution. However, canonical forms of phase-type distribution representations are only known for $n = 2,3$ [7, 8, 10]. Analytical analysis of the representation structures, even for $n = 4$, is quite complex [5], partly due to the fact that explicit expression for eigenvalues is unknown. Another option, which is not much explored in the scientific literature, is to investigate the representation structures numerically.

The expectation maximization method (EM) is widely used for phase-type fitting and its procedure for general representation structure is presented in [2]. The method aims to maximize the likelihood that an observed sample is generated by a phase-type distribution, in which, parameters are to be found. Fitting with a general representation structure, using the EM method is not effective [16], because, depending on the choice of the initial solution, a local solution can be reached. Therefore, an undefined number of initial solutions have to be tested and the best resulting PH distribution parameters would have to be picked.

It was observed that EM method can be more robust when fitting with a restricted class PH distributions. For example, an EM method procedure for fitting by hyper-exponential distributions is given in [11], which was extended for fitting with hyper-Erlang in [22]. However, one specific structure may not cover the whole class of PH distributions [5], even though its usage is more practical.

Our contribution is an attempt to generate a set of PH(4) representation structures, and then, use it for the initial solutions generation. The generation of representation structures is based on the canonical forms of PH distribution with two [7, 8] and three [16] states, non-stationary Markov arrival process (NMAP) with two states [14] and other observations. The randomly generated initial PH representations of those structures will be used as the initial solutions for the EM procedure. It is expected that the EM method would be more effective when fitting by a number of different PH sub-classes, rather than by the general representation structure. This assumption will be checked numerically by performing phase-type fitting the benchmark distributions [4].

The rest of the paper is organized as follows. The basic theory is given in Section 2; an algorithm to generate the PH representation structures is discussed in Section 3; the results of phase-type fitting are presented in Section 4 and the conclusions are given in Section 5.

2. Basic Theory

2.1. Phase-Type Distributions

Definition 1. [19] The time to reach an absorbing state in a continuous time Markov chain, with n transient states and one absorbing state, is PH distributed.

The continuous time Markov chain (CTMC) used in the Definition 1 has n transient states and one absorbing state described with parameters

$$\underline{\alpha} = [\alpha_1 \quad \alpha_2 \quad \dots \quad \alpha_n], \underline{a} = [a_1 \quad a_2 \quad \dots \quad a_n]^T, \quad \boldsymbol{A} = \begin{bmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ a_{2,1} & a_{2,2} & \dots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \dots & a_{n,n} \end{bmatrix}, \tag{1}$$

where α_i is a probability that the process will start in state i, a_i denotes the rate at which the absorbing state is reached from state i and $a_{i,j}$ for $i \neq j$ is the transition rate from state i to state j.

PH distribution is defined via a continuous time Markov chain, therefore the sum of rates for state i is zero

$$\sum_{j=1}^{n} a_{i,j} + a_i = 0. \tag{2}$$

From (2) the definition of $a_{i,i}$ follows, that is

$$a_{i,i} = -\sum_{j=1, j\neq i}^{n} a_{i,j} - a_i. \tag{3}$$

The vector $\underline{\alpha}$ can be expressed as $\underline{\alpha} = -\boldsymbol{A}\mathbb{1}$, because, from (2) we have

$$a_i = -\sum_{j=1}^{n} a_{i,j}, \tag{4}$$

where $\mathbb{1}$ is a column vector of ones of appropriate size. Thus, the CTMC can be uniquely specified by a pair $(\underline{\alpha}, \boldsymbol{A})$, which represents a specific PH distribution.

Definition 2. A representation $(\underline{\alpha}, \boldsymbol{A})$ is Markovian if $\underline{\alpha}\mathbb{1} = 1$, $\alpha_i \geq 0$ for $i = \overline{1,n}$; $a_{i,j} \geq 0$ for $i \neq j$ and $i, j = \overline{1,n}$; and $a_{i,i} < 0$ for $i = \overline{1,n}$. Otherwise, the representation $(\underline{\alpha}, \boldsymbol{A})$ is non-Markovian.

PH distribution [19] probability density function (PDF) is

$$f(t) = -\underline{\alpha}e^{\boldsymbol{A}t}\boldsymbol{A}\mathbb{1} = \underline{\alpha}e^{\boldsymbol{A}t}\underline{a} \tag{5}$$

and cumulative distribution function (CDF) is

$$F(t) = 1 - \underline{\alpha}e^{\boldsymbol{A}t}\underline{a}. \tag{6}$$

In general, the PH distribution has many representations [18, 19]. Let T be a non-singular matrix with unit row sums, i.e., $\boldsymbol{T}\mathbb{1}$=1. Then, any representation $(\underline{\alpha}, \boldsymbol{A})$ can be transformed into another representation $(\underline{\beta}, \boldsymbol{B})$ by

$$\underline{\beta} = \underline{\alpha}\boldsymbol{T}, \boldsymbol{B} = \boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T}. \tag{7}$$

Both $(\underline{\alpha}, \boldsymbol{A})$ and $(\underline{\beta}, \boldsymbol{B})$ represent the same PH distribution, because

$$\begin{aligned} f(t) &= -\underline{\beta}e^{\boldsymbol{B}t}\boldsymbol{B}\mathbb{1} = -\underline{\alpha}\boldsymbol{T}e^{\boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T}t}\boldsymbol{A}\boldsymbol{T}\mathbb{1} \\ &= -\underline{\alpha}\boldsymbol{T}\boldsymbol{T}^{-1}e^{\boldsymbol{A}t}\boldsymbol{T}\boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T}\mathbb{1} = -\underline{\alpha}e^{\boldsymbol{A}t}\boldsymbol{A}\mathbb{1}. \end{aligned}$$

By applying transformation (7) infinitely, many equivalent representations $(\underline{\beta}, \boldsymbol{B})$ can be found, also including the non-Markovian ones [15].

Proposition 1. *Phase-type distribution PH(n) can be uniquely specified with 2_{n-1} parameters.*

Proof. The proof of Proposition 1, using Laplace transform, is of particular interest, because it will be used later. The Laplace transform of (5) is

$$\begin{aligned} f^*(t) &= \int_0^\infty e^{-st}f(t)dt \\ &= \underline{\alpha}(s\boldsymbol{I} - \boldsymbol{A})^{-1}\underline{a} = \underline{\alpha}\frac{[\gamma_{j,i}(s)]}{det(s\boldsymbol{I} - \boldsymbol{A})}\underline{a} \\ &= \frac{c_{n-1}s^{n-1} + \cdots + c_1 s + c_0}{d_n s^n + \cdots + d_1 s + d_0}, \end{aligned} \tag{8}$$

where $\gamma_{i,j}(s)$ is an adjunct of the matrix $s\boldsymbol{I}$–$\boldsymbol{A}$ element at i, j. The adjunct $\gamma_{i,j}(s)$ is found by calculating a determinant of a sub-matrix, which is obtained by removing the row i and column j from the matrix $s\boldsymbol{I}$–$\boldsymbol{A}$. Notice that the matrix $s\boldsymbol{I}$–$\boldsymbol{A}$ has terms with s only in the diagonal elements. Therefore, adjunct $\gamma_{i,i}(s)$ is a polynomial in s of order n–1 at most, because, in the calculation of sub-matrix determinant, only one diagonal element of the matrix $s\boldsymbol{I}$–$\boldsymbol{A}$ is removed. Similarly, the adjunct $\gamma_{i,j}(s)$ for $i \neq j$ is a polynomial in s of order n–2 at most, because two diagonal elements with term s are removed. Thus, the degree of polynomial in the numerator of (9) can be n–1 at most, and the degree of denominator is n at most, when the rank of matrix $\boldsymbol{A}$ is n. Therefore, there are $2n$+1 coefficients $c_{n-1}, \ldots, c_0$, $d_n, \ldots, d_0$ in Laplace transform expression (8). Without the loss of generality, the coefficient d_n can be set to one by dividing the numerator and the denominator by $d_n \neq 0$.

From the observation that $\lim_{t\to+\infty} F(t) = 1$, we have that $f^*(0) = \int_0^\infty f(t)dt = c_0/d_0 = 1$, from which $c_0 = d_0$. Thus, the Laplace transform of the PH distribution is uniquely specified by $2n - 1$ free coefficients.

2.2. Markov Arrival Processes

Markov arrival process (MAP) is a generalization of the PH distribution, and is able to model the dependent inter-arrival times. MAP is a process of $\{N(t), J(t)\}$, where $N(t) \in \{0, 1, \ldots\}$ is a level and $J(t) \in \{0, 1, \ldots n\}$ is an index of an active state within level. MAP is characterized by the initial probability vector

$\underline{\alpha} = [\alpha_1\ \alpha_2\ \alpha_n]$ and two matrices $\boldsymbol{D}_0, \boldsymbol{D}_1$ of size $n \times n$. The entry α_i denotes a probability that process will start in state i at zero level; the matrix $\boldsymbol{D}_0$ is a transient generator for the states within the level and $\{\boldsymbol{D}_1\}_{i,j}$ for $i, j = \overline{1, n}$ denotes the transition rate from the current level state i to the next level state j.

The stationary distribution $\underline{\pi}$ of a phase process $J(t)$ can be found by solving the following system of equations

$$\begin{cases} \underline{\pi}(\boldsymbol{D}_0 + \boldsymbol{D}_1) = 0, \\ \underline{\pi}\mathbb{1} = 1. \end{cases} \tag{9}$$

Definition 3. MAP process is stationary if $\underline{\alpha} \equiv \underline{\pi}$, otherwise, such process is non-stationary (NMAP).

The probability density function of NMAP is

$$F(t) = \underline{\alpha} e^{\boldsymbol{D}_0 t} \boldsymbol{D}_1 \mathbb{1}. \tag{10}$$

The canonical form of NMAP(2) has two structures [16], the first one is

$$\begin{aligned} \boldsymbol{D}_0 &= \begin{bmatrix} -\lambda_1 & (1-a)\lambda_1 \\ 0 & -\lambda_2 \end{bmatrix}, \\ \boldsymbol{D}_1 &= \begin{bmatrix} a\lambda_1 & 0 \\ (1-b)\lambda_2 & b\lambda_2 \end{bmatrix}, \end{aligned} \tag{11}$$

where $0 < \lambda_1 \leq \lambda_2, 0 < a < 1,\ 0 < b < 1, b \geq a\frac{\lambda_1}{\lambda_2}$. And the second form is given by

$$\begin{aligned} \boldsymbol{D}_0 &= \begin{bmatrix} -\lambda_1 & (1-a)\lambda_1 \\ 0 & -\lambda_2 \end{bmatrix}, \\ \boldsymbol{D}_1 &= \begin{bmatrix} 0 & a\lambda_1 \\ b\lambda_2 & (1-b)\lambda_2 \end{bmatrix}, \end{aligned} \tag{12}$$

where $0 < \lambda_1 \leq \lambda_2, 0 < a \leq 1,\ 0 < b \leq 1, b \geq a\frac{\lambda_1}{\lambda_2}$.

2.3. Phase-Type Structures

The structure of a PH representation is denoted by a triple $(\dot{\underline{\alpha}}, \dot{\boldsymbol{A}}, \dot{\underline{a}})$, which describes the distribution of non-zero elements in $(\underline{\alpha}, \boldsymbol{A})$, where $\dot{\alpha}_i \in \{0,1\}$, $\dot{A}_{i,j} \in \{0,1\}$, $\dot{a}_i \in \{0,1\}$ for $i, j = \overline{1,n}$ are defined as

$$\begin{aligned} \dot{\alpha}_i &= \begin{cases} 1, if\ \alpha_i > 0, \\ 0, otherwise, \end{cases} \\ \dot{a}_{i,j} &= \begin{cases} 1, if\ a_{i,j} > 0\ for\ i \neq j, \\ 0, otherwise, \end{cases} \\ \dot{a}_i &= \begin{cases} 1, if\ \sum_{j=1}^{n} a_{i,j} < 0, \\ 0, otherwise, \end{cases} \end{aligned} \tag{13}$$

The values of $\dot{\underline{\alpha}}$ indicate non-zero elements of the initial probability vector $\underline{\alpha}$. The matrix $\dot{\boldsymbol{A}}$ denotes the positive transition rates between the transient states. Note that $\dot{a}_{i,j}$ is always equal to zero because $\dot{\boldsymbol{A}}$ accounts only for the transitions between the transient states.

Definition 4. An independent structural element is a variable in $(\underline{\alpha}, \boldsymbol{A})$ that does not depend on any other variable in $(\underline{\alpha}, \boldsymbol{A})$.

The number of independent structural elements in $(\underline{\alpha}, \boldsymbol{A})$, with structure $(\dot{\underline{\alpha}}, \dot{\boldsymbol{A}}, \dot{\underline{a}})$, is equal to $\sum_{i=1}^{n} \dot{\alpha}_i - 1 + \sum_{i=1}^{n} \sum_{j=1}^{n} \dot{a}_{i,j} + \sum_{i=1}^{n} \dot{a}_i$, where the term -1 is necessary, because the initial probability vector has one dependent element, which can be found from the condition $\underline{\alpha}\mathbb{1} = 1$.

Example 1. The structure $(\dot{\underline{\alpha}}, \dot{\boldsymbol{A}}, \dot{\underline{a}})$ with

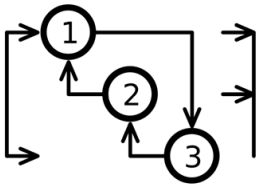
$$\dot{\underline{\alpha}} = [1 \quad 0 \quad 1], \quad \dot{\boldsymbol{A}} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad \dot{\underline{a}} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$$

stands for all the PH representations $(\underline{\alpha}, \boldsymbol{A})$ which can be expressed in the following form

$$\begin{aligned} \underline{\alpha} &= [\alpha_1 \quad 0 \quad (1-\alpha_1)], \quad \underline{a} = \begin{bmatrix} a_1 \\ a_2 \\ 0 \end{bmatrix}, \\ \boldsymbol{A} &= \begin{bmatrix} -(a_{1,3} + a_1) & 0 & a_{1,3} \\ a_{2,1} & -(a_{2,1} + a_2) & 0 \\ 0 & a_{3,2} & -a_{3,2} \end{bmatrix}, \end{aligned}$$

where α_1, $\alpha_{1,3}$, $\alpha_{2,1}$, $\alpha_{3,2}$, a_1, a_2 are positive and real numbers. The initial probability vector $\underline{\alpha}$ has two non-zero elements, the one at position (1, 1) is independent. The transition matrix $\boldsymbol{A}$ has three independent non-zero elements at positions (2, 1), (3, 1) and (1, 3). The exit rate vector $\underline{a}$ has two independent non-zero elements at positions (1, 1) and (2, 1). In total, there are 6 independent structural elements. Note that (by Proposition 1) any PH distribution can be specified with $2n$ –1 parameters, therefore, this structure is over-parameterized.

Figure 1
Graphical representation of the PH(3) structure given in Example 1



The structure given in Example 1 is graphically depicted in Figure 1. The diagonally placed circles represent the transient states. The arrows between the circles show the possible transitions. The horizontal arrows on the left indicate the states for which the initial probability is non-zero. Similarly, the horizontal arrows on the right indicate the states from which there is a transition to the absorbing state.

3. Phase-Type Structure Generation

In this section, we present an algorithm to generate a set of PH representation structures with the specified number of independent structural elements.

Input: the number of states n, the number of independent structural elements m.

Output: the set of PH structures.

Step 1. Generate a set of all possible valid structures with m independent structural elements (Sub-section 3.1).

Step 2. Remove structures until the resulting set does not contain at least two trivially equivalent structures (Sub-section 3.2).

Step 3. Remove all the structures which can be trivially transformed into a structure with acyclic generator (Sub-section 3.3).

Step 4. If $m = 2n - 1$, then add a Coxian structure (Sub-section 3.3).

Step 5. Remove the constrained structures, i.e., which stand for the PH distributions with $2n - 2$ parameters at most (Sub-section 3.4).

Step 6. Remove the structures which have not obviously constrained and non-canonical lower order PH sub-structures (Sub-section 3.6 and 3.7).

Step 7. Remove structures which have not obviously constrained and non-canonical NMAP(2) sub-structures (Sub-section 3.8).

Definition 5. The structure of PH distribution (or NMAP process) representation is obviously constrained if it has less structurally independent elements than the number of parameters necessary to fully specify the distribution/process.

A structure containing a lower order PH or NMAP sub-structure, that is not obviously constrained and of non-canonical form, is not preferred and therefore is not considered, as such a structure duplicates the other similar structure which has a canonical form sub-structure.

3.1. Generation of All Valid Structures

Basically, from a set of all possible structures, the ones with m independent structural elements are taken for further consideration. Let $\tilde{v}$ be a set of all possible binary vectors of size $n \times n$

$$\tilde{v} = \left\{ [\dot{v}_1 \quad \dots \quad \dot{v}_n] \middle| \dot{v}_i \in \{0,1\},\ i = \overline{1,n} \right\}.$$

Similarly, let $\tilde{A}$ be a set of all possible binary matrices of size $n \times n$

$$\tilde{A} = \left\{ \begin{bmatrix} \dot{a}_{1,1} & \cdots & \dot{a}_{1,n} \\ \vdots & \ddots & \vdots \\ \dot{a}_{n,1} & \cdots & \dot{a}_{n,n} \end{bmatrix} \middle| \begin{matrix} \dot{a}_{i,j} \in \{0,1\},\ \dot{a}_{i,i} = 0 \\ for\ i \neq j,\ i,j = \overline{1,n} \end{matrix} \right\}.$$

Then the initial set of all structures, with m independent structural elements, is

$$\left\{ (\underline{\dot{\alpha}}, \dot{\boldsymbol{A}}, \underline{\dot{a}}) \middle| \begin{matrix} \underline{\dot{\alpha}} \in \tilde{v}, \dot{\boldsymbol{A}} \in \tilde{A}, \underline{\dot{a}} \in \tilde{v}\,, \\ \sum_{i=1}^{n} \dot{\alpha}_i - 1 \sum_{i=1}^{n} \sum_{j=1}^{n} \dot{a}_{i,j} + \sum_{i=1}^{n} \dot{a}_i = m \end{matrix} \right\}$$

Definition 6. A structure is considered to be valid if the following conditions are met: a) there exists a path to visit every transient state, b) there exists a path to reach an absorbing state from every transient state.

This set may contain many structures which do not define a valid CTMC. All such invalid structures are removed from the initial set. The resulting set still contains many equivalent structures, which are investigated in the following sections.

3.2. Trivial Structure Equivalence

We are interested in a set of PH distributions that contain all the distributions which have at least one representation $(\underline{\alpha}, \boldsymbol{A})$ with the structure $(\underline{\dot{\alpha}}, \dot{\boldsymbol{A}}, \underline{\dot{a}})$. Such a set is denoted as $PH_{(\underline{\dot{\alpha}}, \dot{\boldsymbol{A}}, \underline{\dot{a}})}$.

Two structures $(\underline{\dot{\alpha}}, \dot{\boldsymbol{A}}, \underline{\dot{a}})$ and $\left(\underline{\dot{\beta}}, \dot{\boldsymbol{B}}, \underline{\dot{b}}\right)$ are equivalent if

$$PH_{(\underline{\dot{\alpha}}, \dot{\boldsymbol{A}}, \underline{\dot{a}})} \subseteq PH_{\left(\underline{\dot{\beta}}, \dot{\boldsymbol{B}}, \underline{\dot{b}}\right)}$$

and

$$PH_{\left(\underline{\dot{\beta}}, \dot{\boldsymbol{B}}, \underline{\dot{b}}\right)} \subseteq PH_{(\underline{\dot{\alpha}}, \dot{\boldsymbol{A}}, \underline{\dot{a}})}.$$

The structure with reordered states stands for the same set of PH distributions. This comes from the fact that renumbering states of a certain representation results in another representation, which represents the same PH distribution. Formally, given a list of

unique new states indices$(l_1, l_2, \dots, l_n)$ the equivalent structure $\left(\underline{\dot{\beta}}, \dot{\boldsymbol{B}}, \underline{\dot{b}}\right)$ obtained from $\left(\underline{\dot{\alpha}}, \dot{\boldsymbol{A}}, \underline{\dot{a}}\right)$ by renumbering is defined as

$$\dot{\beta}_i = \dot{\alpha}_{l_i}, \dot{b}_{i,j} = \dot{a}_{l_i,l_j}, \dot{b}_i = \dot{a}_{l_i} \quad for\ i,j = \overline{1,n}.$$

Similarly, the time-reversed structure has the same set of PH distributions, as time-reversed representation represents the same PH distribution. Formally, given a structure $\left(\underline{\dot{\alpha}}, \dot{\boldsymbol{A}}, \underline{\dot{a}}\right)$, the equivalent time-reversed structure $\left(\underline{\dot{\beta}}, \dot{\boldsymbol{B}}, \underline{\dot{b}}\right)$ is obtained by

$$\dot{\beta}_i = \dot{\alpha}_i, \dot{b}_{i,j} = \dot{a}_{j,i}, \dot{b}_i = \dot{a}_i \quad for\ i,j = \overline{1,n}.$$

Definition 7. A trivial transformation of a structure or a representation is such transformation which involves the state reordering and/or time reversal.

In addition, if a structure or a representation was obtained by the trivial transformation, those are said to be trivially equivalent.

3.3. Structures with Acyclic Generator and Coxian Structure

Definition 8. The structure $\left(\underline{\dot{\alpha}}, \dot{\boldsymbol{A}}, \underline{\dot{a}}\right)$ represents the acyclic PH distribution if it can be trivially transformed into the structure $\left(\underline{\dot{\beta}}, \dot{\boldsymbol{B}}, \underline{\dot{b}}\right)$ with an upper triangular matrix $\dot{\boldsymbol{B}}$.

Any PH distribution representation with a triangular generator matrix can be transformed into an ordered Coxian representation structure [8]. The ordered Coxian representation structure has the following form [7]

$$\underline{\alpha} = [1 \quad 0 \quad \dots \quad 0],$$
$$\underline{a} = [a_1 \quad a_2 \quad \dots \quad a_n]^T,$$
$$\boldsymbol{A} = \begin{bmatrix} a_{1,1} & a_{1,2} & 0 & \dots & 0 \\ 0 & a_{2,2} & a_{2,3} & \dots & 0 \\ \vdots & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & 0 & a_{n,n} \end{bmatrix},$$

where $\alpha_{i,i} = -(\alpha_{i,i+1} + \alpha_i)$ and $\alpha_{1,1} \le \dots \le \alpha_{n,n}$ for $i = \overline{1, n}$.

3.4. Structurally Induced Zero Density of PH Distribution at Zero

Proposition 2. *If the representation $\left(\underline{\alpha}, \boldsymbol{A}\right)$ of the structure $\left(\underline{\dot{\alpha}}, \dot{\boldsymbol{A}}, \underline{\dot{a}}\right)$ is such that there is no immediate exit from the states for which the initial probability is non-zero, i.e., for all $i = \overline{1,n}$: $\alpha_i a_i = 0$, the represented PH distribution is defined by $2n - 2$ parameters at most.*

Proof. From (8) we have that

$$det(s\boldsymbol{I} - \boldsymbol{A}) f^*(t) = \underline{\alpha}\left[\gamma_{j,i}(s)\right]\underline{a}$$
$$= [\alpha_i \quad \dots \quad \alpha_n] \begin{bmatrix} \gamma_{1,1}(s) & \dots & \gamma_{n,1}(s) \\ \vdots & \ddots & \vdots \\ \gamma_{1,n}(s) & \dots & \gamma_{n,n}(s) \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$$
$$= \alpha_1\left(\gamma_{1,1}(s)a_1 + \dots + \gamma_{n,1}(s)a_n\right) + \dots$$
$$+\alpha_n\left(\gamma_{1,n}(s)a_1 + \dots + \gamma_{n,n}(s)a_n\right).$$

From the condition of Proposition 1 it follows that adjuncts $\gamma_{i,i}(s)$ are multiplied by zero. Since these are the only adjuncts that are polynomials in s of $n-1$ degree, the numerator in (8) does not contain term s^{n-1}, i.e., the coefficient c_{n-1} is equal to zero. Consequently, the Laplace transform (and the corresponding PH) is defined by $2n-2$ parameters.

3.5. The Sub-chains

Recall that PH distribution is defined via a CTMC. Choose a few transient states for analysis. Partition the whole original CTMC into two chains. The first one contains the chosen transient states and is called a sub-chain. The second one contains all the remaining states and is called a complimentary chain.

A particular sub-chain with p states is given by a list of unique state indices $(l_1, \dots l_p)$. There are the remaining $q = n + 2 - p$ states in the complimentary chain, where $+2$ is for the starting and for the absorbing states. For the given sub-chain state index list $(l_1, \dots l_p)$, the generator matrix of CTMC can be reordered to the form

$$\begin{bmatrix} \widetilde{\boldsymbol{C}} & \boldsymbol{E} \\ \boldsymbol{X} & \boldsymbol{C} \end{bmatrix}, \tag{14}$$

where the lower right block $\boldsymbol{C}$ is a transient generator matrix of the sub-chain. The input $\boldsymbol{E}$ and output $\boldsymbol{X}$ matrices are the upper right and lower left blocks, respectively. The complimentary chain transient generator $\widetilde{\boldsymbol{C}}$ matrix is not investigated further. The structure of the $\boldsymbol{C}$, $\boldsymbol{E}$ and $\boldsymbol{X}$ matrices is encoded by the binary element matrices $\dot{\boldsymbol{C}}$, $\dot{\boldsymbol{E}}$ and $\dot{\boldsymbol{X}}$

$$\dot{\boldsymbol{C}} = \begin{bmatrix} \dot{c}_{1,1} & \cdots & \dot{c}_{1,p} \\ \vdots & \ddots & \vdots \\ \dot{c}_{p,1} & \cdots & \dot{c}_{p,p} \end{bmatrix},$$
$$\dot{\boldsymbol{E}} = \begin{bmatrix} \dot{e}_{1,1} & \cdots & \dot{e}_{1,p} \\ \vdots & \ddots & \vdots \\ \dot{e}_{q,1} & \cdots & \dot{e}_{q,p} \end{bmatrix}, \quad (15)$$
$$\dot{\boldsymbol{X}} = \begin{bmatrix} \dot{x}_{1,1} & \cdots & \dot{x}_{1,q} \\ \vdots & \ddots & \vdots \\ \dot{x}_{q,1} & \cdots & \dot{x}_{q,q} \end{bmatrix},$$

where $\dot{c}_{i,j} \in \{0,1\}$ for $i \neq j,\ i,j = \overline{1,p}$ indicate the transitions between sub-chain states; $\dot{e}_{i,j} \in \{0,1\}$ for $i = \overline{1,q}$, $j = \overline{1,p}$ indicate the transitions from complimentary chain to sub-chain; $\dot{x}_{i,j} \in \{0,1\}$ for $i \neq j,\ i,j = \overline{1,q}$ indicate the transitions from sub-chain to complimentary chain.

The following characteristics based on the matrices $\dot{\boldsymbol{E}}$ and $\dot{\boldsymbol{X}}$ are defined:

- E_c / E_r is the number of states to/from which sub-chain can be entered,
- X_c / X_r is the number of states to/from which sub-chain can be left,
- E_{rank} / X_{rank} is the maximal possible rank of input/output matrices, found by (16)

$$E_{rank} = \min\{E_c, E_r\},$$
$$X_{rank} = \min\{X_c, X_r\}. \quad (16)$$

Example 2. Let us study a sub-chain, specified by the state indices list (2,3), of the structure $(\underline{\alpha}, \dot{\boldsymbol{A}}, \underline{a})$ given in Example 1. The index of the source state is 0 and the index of the absorbing state is 4. To have a partition of the CTMC generator matrix as shown in (15), the states are reordered to be in the sequence of (0,4,1,2,3). After state reordering, the structure of the resulting generator matrix is

$$\begin{bmatrix} 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

Consequently, the structure of the sub-chain is

$$\dot{\boldsymbol{C}} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix},$$

and the structures of the input and output matrices are

$$\dot{\boldsymbol{E}} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}, \dot{\boldsymbol{X}} = \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

Based on the matrices $\dot{\boldsymbol{E}}$, $\dot{\boldsymbol{X}}$, the following characteristics are found

$$E_r = 2, \quad E_c = 1, \quad E_{rank} = 1,$$
$$X_r = 1, \quad X_c = 2, \quad X_{rank} = 1.$$

The given structure $(\underline{\alpha}, \dot{\boldsymbol{A}}, \underline{a})$ is investigated by analyzing the structure of its sub-chains obtained from the corresponding CTMC (or time-reversed CTMC) structure. Note that the order of states in the sub-chain is not important, because the state renumbering in CTMC does not affect the PH distribution that it represents.

3.6. The PH(2) Case

A two state sub-chain that is characterized by $E_r = 1$ and $X_r = 1$ can be considered as a PH(2) distribution. To determine whether such a sub-chain is a constrained PH distribution, the statistic T, which denotes the number of independent structural elements in the sub-chain, is defined as

$$T = E_c - 1 + \sum_{i=1}^{p} \sum_{j=1}^{p} \dot{c}_{i,j} + X_c. \quad (17)$$

If $T = 2$, then the sub-chain is a constrained PH(2) and no reasonable conclusion can be drawn about the original PH structure.

It is known that any PH(2) distribution can be transformed into the Coxian representation structure [7]. Therefore, if $T = 3$ and the structure of sub-chain is not trivially equivalent to the Coxian structure, the original PH structure is not preferred.

If > 3, then the sub-chain represents a redundant PH(2) distribution and the original PH structure is not preferred as well.

3.7. The PH(3) Case

If a three state sub-chain is such that $E_r = 1$ and $X_r = 1$, it can be considered as a PH(3) distribution. If the sub-chain structure is not constrained, it is necessary

to check if it is of canonical form, as given in [10].

If $T < 5$, then the sub-chain is a constrained PH(3) and no reasonable conclusion can be drawn about the original PH structure.

In the case of $T = 5$, the original PH structure is not preferred if it cannot be trivially transformed into one of the structures (18):

$$\begin{aligned}(\underline{\dot{\alpha}}_1, \dot{\boldsymbol{A}}_1, \underline{\dot{\alpha}}_1) &= \left([1\ 1\ 1], \begin{bmatrix}0 & 1 & 0\\ 0 & 0 & 1\\ 0 & 0 & 0\end{bmatrix}, \begin{bmatrix}0\\0\\1\end{bmatrix}\right),\\ (\underline{\dot{\alpha}}_2, \dot{\boldsymbol{A}}_2, \underline{\dot{\alpha}}_2) &= \left([1\ 0\ 1], \begin{bmatrix}0 & 1 & 0\\ 0 & 0 & 1\\ 1 & 0 & 0\end{bmatrix}, \begin{bmatrix}0\\0\\1\end{bmatrix}\right).\end{aligned} \quad (18)$$

Similarly, if $T = 6$ and the structure of the sub-chain is not trivially equivalent to structure (19), the original PH structure is not preferred:

$$(\underline{\dot{\alpha}}_3, \dot{\boldsymbol{A}}_3, \underline{\dot{\alpha}}_3) = \left([1\ 1\ 1], \begin{bmatrix}0 & 1 & 0\\ 0 & 0 & 1\\ 1 & 0 & 0\end{bmatrix}, \begin{bmatrix}0\\0\\1\end{bmatrix}\right). \quad (19)$$

Finally, in the case of $T > 0$ the whole original structure is not preferred, since such PH(3) distribution is redundant.

3.8. The NMAP(2) Case

A two state sub-chain can be considered as a NMAP(2) process if it can be entered from one complimentary chain state ($E_r = 1$) and exited to two complimentary chain states ($X_r = 2$).

It is necessary to check if NMAP(2) structure is not constrained, to do that the statistic Z is defined as

$$Z = z_E + z_X + z_C, \quad (20)$$

where z_E / z_X is the number of zeros in the matrix $\dot{\boldsymbol{E}}$ / $\dot{\boldsymbol{X}}$ excluding rows/columns with only zero elements and z_C is the number of zeros in the matrix $\dot{\boldsymbol{C}}$ excluding diagonal elements.

If , such a sub-chain is a constrained NMAP(2) process and the original PH structure has to be left as possibly contributing.

Otherwise, if and sub-chain cannot be trivially transformed into one of the two NMAP(2) canonical form structures (11), (12), then the original PH structure is not preferred.

4. Numerical Experiments

4.1. The Generated PH(4) Structures

The algorithm presented in Section 3 is used to generate PH(4) representation structures. We are interested only in the structures which have $m = 7, 8, ..., 19$ independent structural elements. The structures with $m < 7$ are not significant, because these stand for PH distributions parameterized with less than 7 parameters (i.e., less than $2n - 1$ parameters). The number of generated structures for $m = 7, 8, ..., 19$ is given in Table 1.

Table 1
Number of generated structures (M) for very number of independent structural elements (m)

m	7	8	9	10	11	12
M	101	441	1136	1752	1885	1433

m	13	14	15	16	17	18	19
M	829	362	132	37	10	2	1

Table 2
Target distributions and their density functions

Distribution	Density function
Weibull	$\begin{cases}\frac{k}{\lambda}\left(\frac{t}{\lambda}\right)^{k-1} e^{-\left(\frac{t}{\lambda}\right)^k}, if\ t \geq 0,\\ 0, otherwise.\end{cases}$
Log-normal	$\begin{cases}\frac{1}{t\sigma\sqrt{2\pi}} e^{\frac{-(ln(t)-\mu)^2}{2\sigma^2}}, if\ t \geq 0\\ 0, otherwise.\end{cases}$
Uniform	$\begin{cases}\frac{1}{b-a}, \text{ if } a \leq t \leq b\\ 0, \quad \text{otherwise.}\end{cases}$
Shifted exponential	$\begin{cases}\frac{1}{2}e^{-t} + \frac{1}{2}e^{-(t-1)}, if\ t \geq 1,\\ \frac{1}{2}e^{-t}, if\ 0 \leq t < 1,\\ 0, \quad \text{otherwise.}\end{cases}$
Matrix exponential	$\begin{cases}\left(1 + \frac{1}{4\pi^2}\right)\left(1 - cos(2\pi t)\right)e^{-t}, if\ t \geq 0,\\ 0, \quad \text{otherwise.}\end{cases}$

Table 3
Parameters of target distributions

Distribution	Code	Parameters
Weibull	**W1**	λ = 1, k = 1.5
	W2	λ = 1, k = 0.5
Log-normal	**L1**	μ = –1.8, σ = 1.8
	L2	μ = –0.32, σ = 0.8
	L3	μ = –0.02, σ = 0.2
Uniform	**U1**	a = 0, b = 1
	U2	a = 1, b = 2
Shifted exponential	**SE**	
Matrix exponential	**ME**	

Table 4
Target distribution discretization

Code	Distribution truncation time, t_{end}	Number of observations, N
W1	2.768	277
W2	8.974	897
L1	10.885	1089
L2	4.670	467
L3	1.561	156
U1	1	100
U2	2	200
SE	5.225	523
ME	4.578	458

The generated PH(4) structures, in total 8340, are used to randomly generate the initial solutions for the EM method.

4.2. The Experiment Setup

The EM algorithm presented in [2] is used to approximate the target distributions (Tables 2 and 3) by PH(4) distribution. These benchmark distributions are taken from [4]. The EM algorithm has a property to maintain the zeros in the refined solutions. Therefore, the structure of initial solution is preserved. EM method maximizes the log-likelihood function

$$L = \sum_{i=1}^{N} w_i \ln\left(f(t_i;\theta)\right), \tag{21}$$

where $f(t_i; \theta)$is the density function of target distribution, t_i are the density function discretization time instants with the weights w_i for $i = \overline{1, N}$.

The observations are obtained by the evenly spaced time discretization of the target distributions in Table 3.

A certain target distribution is discretized in the time range $[0, t_{end}]$ by dividing it into N equal intervals. The middle points of these intervals are the time instants t_i and the probability of the interval is a weight W_i. The distributions are truncated at 0.99 quantile, except for the **U1** and **U2** cases.

The two sets of PH(4) representations are generated and used as the initial solutions for the EM method. The first set (further referred to by **A-set**) consists of one randomly generated representation for every structure generated in the Sub-section 4.1. The second one (further referred to by **B-set**) consists of 8340 randomly generated PH(4) representations of a general structure. These two initial solution generation strategies are compared by comparing the highest log-likelihood function values.

4.3. The Results

The maximum log-likelihood function values obtained after running the EM method for 600 iterations on the initial solutions sets **A-set**, **B-set** are given in Table 5.

Table 5
The best log-likelihood function values for two initial solution generation strategies after running the EM method for 600 iterations

Code	Initial solution of various structures (A-set)	Initial solutions of general structure (B-set)
W1	-0.739508894546	-0.739855890106
W2	-0.799975887679	-0.814530004973
L1	-0.121300277809	-0.121300291222
L2	-0.809498393517	-0.810271592230
L3	-0.292479164600	-0.292479164600
U1	-0.138916558371	-0.138917012205
U2	-0.709553122213	-0.709553122213
SE	-1.263328058400	-1.263473379700
ME	-0.857747924918	-0.857810154292

5. Conclusions

The strategy of randomly generating initial solutions of various structures helps to improve results of the EM method. The 8340 PH(4) distribution representation structures were generated by the developed algorithm. Randomly generated representations of these structures were used as the initial solutions for the EM method. Phase-type fitting was performed and the results compared with ones obtained when the initial solutions were of a general structure. A faster convergence to the solution was observed in Figures 2 and 6-10.

Figure 2
Approximation of W1 distribution

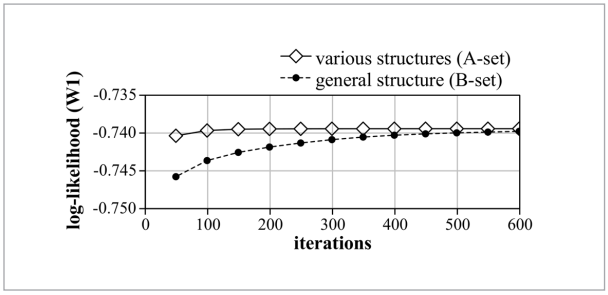


Figure 3
Approximation of W2 distribution

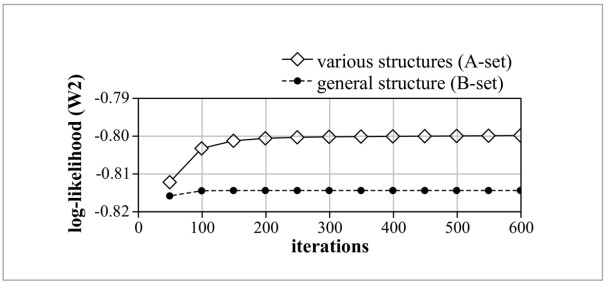


Figure 4
Approximation of L1 distribution

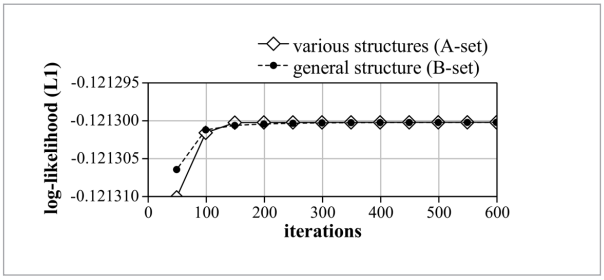


Figure 5
Approximation of L2 distribution

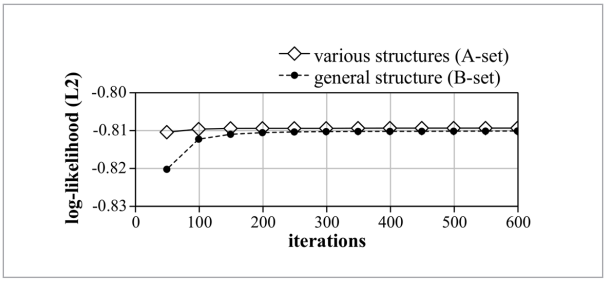


Figure 6
Approximation of L3 distribution

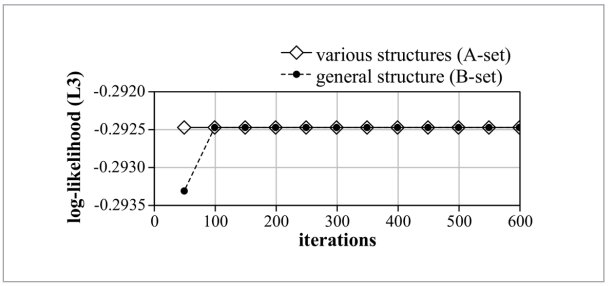


Figure 7
Approximation of U1 distribution

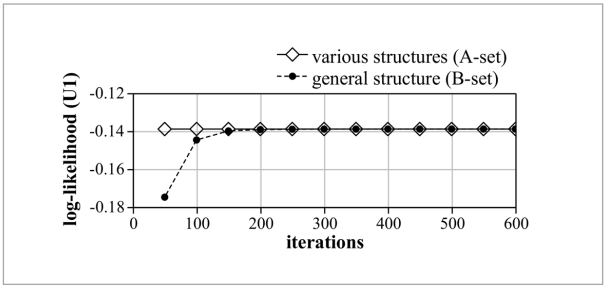


Figure 8
Approximation of U2 distribution

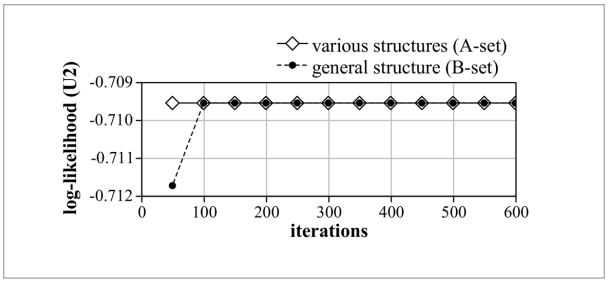


Figure 9
Approximation of SE distribution

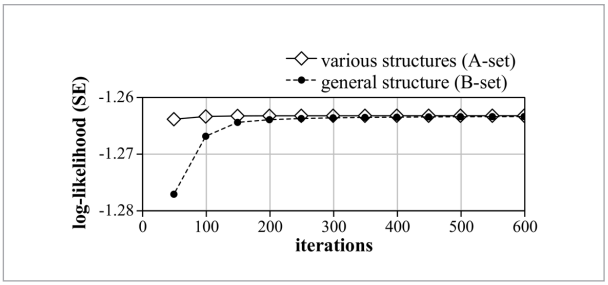
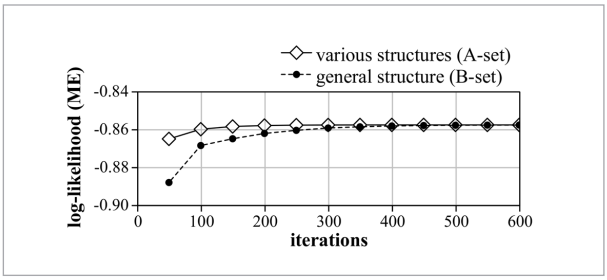


Figure 10
Approximation of ME distribution



In a few cases (Figures 3 and 5), the proposed strategy gave distinctly higher log-likelihood values. These results suggest that the initial solutions of various structures are able to explore the parameter space better.

In order to make the introduced technique more applicable for the practical use, a few problems should be solved. First of all, an upper bound for the number of independent structural elements necessary to represent the whole class of PH(n) distributions has to be determined. This might significantly reduce the number of structures. Secondly, a more efficient structure generation algorithm could be developed.

6. Discussion

For example, the phase-type distributions can be used in modelling particular control systems. In practice, time intervals between events or duration of operations have a distribution which is not Markovian or unknown. In general, the analysis of such processes is complex and specific. However, it is possible to approximate such distributions by a Markovian (i.e. phase-type distribution) and build a model which is analytically tractable. This approach is applied to modelling queueing systems [2, 9, 18], inventory control systems [21] and others.

In addition, there is more direct relation between the phase-type representation and a positive realization in control theory. For example, the Perron-Frobenius theorem can be used in order to derive a transformation from positive realization to a phase-type representation [6, 12]. Currently, the phase-type representation canonical forms for are unknown. The investigated PH(4) could be contributed to the field of positive realization systems.

References

1. Aldous, D., Shepp, L. The Least Variable Phase-Type Distribution is Erlang. Communications in Statistics. Stochastic Models, 1987, 3(3), 467-473. https://doi.org/10.1080/15326348708807067
2. Asmussen, S., Nerman, O., Olsson, M. Fitting Phase-Type Distributions via the EM Algorithm. Scandinavian Journal of Statistics, 1996, 23(4), 419-441. http://dx.doi.org/10.2307/4616418
3. Asmussen, S., Rolski, T. Computational Methods in Risk Theory: A Matrix-algorithmic Approach. Insurance: Mathematics and Economics, 1992, 10(4), 259-274. https://doi.org/10.1016/0167-6687(92)90058-J
4. Bobbio, A., Telek, M. A Benchmark for PH Estimation Algorithms: Results for Acyclic-PH. Stochastic Models, 1994, 10, 661-677. https://doi.org/10.1080/15326349408807315
5. Bodrog, L., Heindl, A., Horváth, G., Telek, M., Horváth, A. Current Results and Open Questions on PH and MAP Characterization. In: Numerical Methods for Structured Markov Chains, 2008, 07461. http://drops.dagstuhl.de/opus/volltexte/2008/1401
6. Commault, C., Mocanu, S. Phase-Type Distributions and Representations: Some Results and Open Problems for System Theory. International Journal of Control, 2003, 76(6), 566-580. https://doi.org/10.1080/0020717031000114986
7. Cox, D. R. A Use of Complex Probabilities in the Theory of Stochastic Processes. Mathematical Proceedings of

the Cambridge Philosophical Society, 1955, 51, 313-319. https://doi.org/10.1017/S0305004100030231

8. Cumani, A. On the Canonical Representation of Homogeneous Markov Processes Modelling Failure-time Distributions. Microelectronics Reliability, 1982, 22(3), 583-602. https://doi.org/10.1016/j.peva.2008.11.002
9. Elteto, T., Telek, M. Numerical Analysis of M/G/1 Type Queueing Systems with Phase-Type Transition Structure. Journal of Computational and Applied Mathematics, 2008, 212(2), 331-340. https://doi.org/10.1016/j.cam.2006.12.019
10. Horváth, G., Telek, M. A Canonical Representation of Order 3 Phase Type Distributions. In: Formal Methods and Stochastic Models for Performance Evaluation: EPEW 2007, 2007, 48-62. https://doi.org/10.1007/978-3-540-75211-0
11. Khayari el Abdouni, R., Sadre, R., Haverkort, B. R. Fitting World-Wide Web Request Traces with the EM-algorithm. Performance Evaluation, 2003, 52, 175-191. https://doi.org/10.1016/S0166-5316(02)00179-7
12. Kim, K. On the Relation Between Phase-Type Distributions and Positive Systems. Abstract and Applied Analysis, 2015, 731261. https://doi.org/10.1155/2015/731261
13. Lakatos, L., Szeidl, L., Telek, M. Introduction to Queueing Systems with Telecommunication Applications. Springer, New York, 2013. https://doi.org/10.1007/978-1-4614-5317-8
14. Lang, A., Arthur, J. Parameter Approximation for Phase-Type Distributions. Lecture Notes in Pure and Applied Mathematics, 1996, 151-206. https://doi.org/10.1201/b17050-11
15. Lipsky, L. Queueing Theory: A Linear Algebraic Approach. Springer, New York, 2008. https://doi.org/10.1007/978-0-387-49706-8
16. Mészáros, A., Telek, M. Canonical Form of Order-2 Non-stationary Markov Arrival Processes. In: Computer Performance Engineering: EPEW2015, 2015, 163-176. https://doi.org/10.1007/978-3-319-23267-6
17. Nelson, R. Probability, Stochastic Processes, and Queueing Theory: The Mathematics of Computer Performance Modeling. Springer, New York, 1995. https://doi.org/10.1007/978-1-4757-2426-4
18. Neuts, M. F. Matrix-geometric Solutions in Stochastic Models: an Algorithmic Approach. Johns Hopkins University Press, Baltimore, 1981. https://doi.org/10.1002/net.3230130219
19. Neuts, M. F. Phase-type Probability Distributions. Encyclopedia of Operations Research and Management Science, 2001, 614-616. https://doi.org/10.1007/978-1-4419-1153-7
20. Reinecke, P., Wolter, K. Phase-Type Approximations for Message Transmission Times in Web Services Reliable Messaging. In: Performance Evaluation: Metrics, Models and Benchmarks: SIPEW2008, 2008, 191-207. https://doi.org/10.1007/978-3-540-69814-2
21. Ramaswami, V. Algorithms for a Continuous-Review (s, S) Inventory System. Journal of Applied Probability, 1981, 18(2), 461-472. https://doi.org/10.2307/3213292
22. Thummler, A., Buchholz, P., Telek, M. A Novel Approach for Phase-Type Fitting with the EM Algorithm. IEEE Transactions on Dependable and Secure Computing, 2006, 3(3), 245-258. https://doi.org/10.1109/TDSC.2006.27