A minimal representation of Markov arrival processes and a moments matching method [1]

M. Telek, G. Horváth [2]

Department of Telecommunications, Technical University of Budapest, H-1521 Budapest, Hungary

Abstract

The paper investigates the problem of minimal representation of Markov arrival processes of order n (MAP(n)). The minimal representation of MAPs is crucial for developing effective fitting methods. It seems that all existing MAP fitting methods are based on the $\mathbf{D_0}$, $\mathbf{D_1}$ representation which is known to be redundant. We present the minimal number of parameters to define a MAP(n) and provide a numerical moments matching method based on a minimal representation.

The discussion starts with a characterization of phase type (PH) distributions and then the analysis of MAPs follows a similar pattern. This characterization contains essential results on the identity of stationary behaviour of MAPs and on the number of parameters required to describe the stationary behaviour.

The proposed moments matching method is also applicable for PH distributions. In this case it is a unique method that fits a general PH distribution of order n based on $2n-1$ parameters.

Key words: Markov arrival process, parameter fitting, minimal representation, moments of inter-arrival time distribution.

1 Introduction

A set of computationally efficient numerical methods, referred to as matrix geometric methods, were developed during the last 3 decades for the analysis of stochastic systems with some regular Markovian structure [2]. The applicability of these analysis methods depends on the availability of effective matching or fitting methods which are able to describe the empirical behaviour of

Email address: {telek,ghorvath}@hit.bme.hu (M. Telek, G. Horváth).

[1] This work is supported by the OTKA K61709 grant.

[2] The authors thank the help of Tamás Éltető in the proof of Theorem 5.

systems with Markovian stochastic models like phase type (PH) distributions or Markov arrival processes (MAPs).

The research on PH fitting methods has a very long history. The method of phases proposed by A. K. Erlang at the beginning of the 20th century can be considered as the starting point. Since then, several approaches were proposed, but very few of them search over the whole set of PH distributions of order n (PH(n)). The majority of the methods is restricted to some special subsets like the set of acyclic phase type (APH) distributions, or hyper-Erlang distributions, etc.

One of the few exceptions is the EMPHT method proposed by S. Asmussen and O. Nerman [1]. This method optimizes the elements of the initial probability vector and the transition matrix. This method allows to apply the same optimization procedure (expectation-maximization (EM) method) for optimization over the whole PH(n) set and also over some of its particular subsets. The surprising conclusion drawn from EMPHT fitting over the whole PH(n) set and over the APH(n) set using its canonical representation [4] is that the results of APH(n) fitting are commonly better.

Our explanation for this counterintuitive conclusion is that the EMPHT method fails to find a "good" optimum over the PH(n) set because it is based on a redundant representation (the number of parameters to optimize is n^2+n-1), and, even if it is not obvious for the first sight, the same method is based on a minimal representation when it searches over the APH(n) set using its canonical representation (the number of parameters to optimize is $2n-1$, since the EMPHT method leaves the 0 elements of the transition matrix unchanged). We believe that this is the main reason behind the relative success of APH(n) fitting methods [7,10,13].

The research on MAP fitting methods has a much shorter history. The methods proposed so far are all based on redundant representations, e.g., [17,3]. According to our best knowledge the question of MAP fitting with minimal number of parameters was not considered because the minimal number of parameters to describe order n MAPs (MAP(n)) was not known.

In this paper we discuss the minimal number of parameters and the minimal representations of PH(n) distributions and MAP(n) processes and propose a moment matching method that is based on a minimal representation, and performs moments matching over the whole PH(n) or MAP(n) set up to the accuracy of a numerical transformation procedure.

The research papers about MAP fitting methods (including the ones of the authors) seem to suggest a misleading idea. In several cases the authors evaluated the goodness of the fitting methods by comparing the long range behaviour

of the original model (data set) with the one of the fitting MAP, e.g. in [12]. The most common parameter of long range behaviour of MAPs is the lag-k correlation for large k values. This evaluation suggests that the parameters of high lag values contain some information that is not present in the parameters of the low lag values, which is not the case. The minimal representation provided in this paper also demonstrates this fact.

The rest of the paper is organized as follows. Section 2 and 3 investigate the properties of PH distributions and MAPs. We refer to [14] for a detailed introduction of PH distributions and MAPs. Here we summarize only those properties which play role in the subsequent discussion. Section 4 and 5 present various representations of PH distributions and MAPs and discuss the transformation between these representations. Section 6 presents a procedure to obtain a Markovian representation based on non-Markovian ones. Finally, Section 7 demonstrates the properties of the procedure through numerical examples.

2 Phase type distributions

Let X be a phase type (PH) distributed random variable with cumulated distribution function

$$F(t) = Pr(X < t) = 1 - \pi e^{At}\mathbb{1},$$

where π is the initial probability vector, $\mathbf{A}$ is the generator of the PH distribution and $\mathbb{1}$ is the column vector of ones. We say that X is PH$(\pi, \mathbf{A})$ distributed.

π and $\mathbf{A}$ have the following properties:

- $\pi_i \geq 0$, $\pi\mathbb{1} = 1$ (there is no probability mass at $t = 0$),
- $A_{ii} < 0$, $A_{ij} \geq 0$ for $i \neq j$, $\mathbf{A}\mathbb{1} \leq 0$,
- $\mathbf{A}$ is non-singular.

The density, the Laplace transform and the moments of X are

$$f(t) = \pi e^{\mathbf{A}t}(-\mathbf{A})\mathbb{1}, \tag{1}$$

$$f^*(s) = E(e^{-sX}) = \pi(s\mathbf{I} - \mathbf{A})^{-1}(-\mathbf{A})\mathbb{1} = \sum_{i=0}^{\infty} s^i(-1)^i\pi(-\mathbf{A})^{-i}\mathbb{1}, \tag{2}$$

$$\mu_n = E(X^n) = n!\pi(-\mathbf{A})^{-n}\mathbb{1}. \tag{3}$$

Definition 1 *PH($\pi, \mathbf{A}$) is* non-redundant *if its rank equals to its order*[3]*, where the* rank *of PH($\pi, \mathbf{A}$) is the size of vector π and square matrix $\mathbf{A}$ and the* order *of PH($\pi, \mathbf{A}$) is the degree of the denominator of $f^*(s)$ (which is a rational function of s).*

Let X' be PH($\pi', \mathbf{A}'$) distributed with cdf $F'(t) = Pr(X' < t) = 1 - \pi' e^{\mathbf{A}'t}\mathbb{1}$.

Theorem 1 *Let PH($\pi, \mathbf{A}$) and PH($\pi', \mathbf{A}'$) be two non-redundant PH distributions with cdf $F(t)$ and $F'(t)$, respectively. $F(t) \equiv F'(t)$ if and only if there exist a non-singular matrix $\mathbf{B}$ such that $\pi' = \pi\mathbf{B}$, $\mathbf{A}' = \mathbf{B}^{-1}\mathbf{A}\mathbf{B}$ and $\mathbf{B}^{-1}\mathbb{1} = \mathbb{1}$.*

Proof: If $\pi' = \pi\mathbf{B}$, $\mathbf{A}' = \mathbf{B}^{-1}\mathbf{A}\mathbf{B}$ and $\mathbf{B}^{-1}\mathbb{1} = \mathbb{1}$ then

$$F'(t) = 1 - \pi' e^{\mathbf{A}'t}\mathbb{1} = 1 - \pi\mathbf{B}e^{\mathbf{B}^{-1}\mathbf{A}\mathbf{B}t}\mathbf{B}^{-1}\mathbb{1} =$$
$$1 - \pi\mathbf{B}\mathbf{B}^{-1}e^{\mathbf{A}t}\mathbf{B}\mathbf{B}^{-1}\mathbb{1} = 1 - \pi e^{\mathbf{A}t}\mathbb{1} = F(t).$$

If $F(t) \equiv F'(t)$ then from (3) we have

$$\pi(-\mathbf{A})^{-i}\mathbb{1} = \pi'(-\mathbf{A}')^{-i}\mathbb{1} \quad \text{for } i \geq 0, \tag{4}$$

since the moments of PH($\pi, \mathbf{A}$) and PH($\pi', \mathbf{A}'$) are identical.

Let $(-\mathbf{A})^{-1} = \mathbf{\Gamma}^{-1}\mathbf{E}\mathbf{\Gamma}$ and $(-\mathbf{A}')^{-1} = \mathbf{\Gamma}'^{-1}\mathbf{E}'\mathbf{\Gamma}'$ be the ordered (according to the real and than the imaginary parts of the eigenvalues) Jordan decomposition[4] of $(-\mathbf{A})^{-1}$ and $(-\mathbf{A}')^{-1}$, respectively, normalized such that $\mathbf{\Gamma}\mathbb{1} = \mathbb{1}$ and $\mathbf{\Gamma}'\mathbb{1} = \mathbb{1}$. This normalization is always possible according to Theorem 5 in Appendix A. Matrix $\mathbf{E}$ has the Jordan-block structure $\mathbf{E} = \text{diag}\{\mathbf{E}_j\}$ with

$$\mathbf{E}_j = \begin{pmatrix} \lambda_j & 1 & 0 & \dots & 0 \\ 0 & \lambda_j & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & \ddots & \lambda_j & 1 \\ 0 & 0 & \dots & 0 & \lambda_j \end{pmatrix}, \tag{5}$$

[3] This definition of non-redundant PH distributions has crucial consequences. It excludes PH distributions which have a lower order matrix exponential representation and ensures that the Jordan decomposition of $\mathbf{A}$ is such that all identical eigenvalues belong to the same Jordan block.

[4] The Jordan normal form is an efficient tool to describe simple (PH distribution with distinct eigenvalues) and complicated (PH distribution with degenerate eigenvectors) cases in the same framework. The scalar expansion of the Jordan normal form is discussed in Appendix A and the Jordan decomposition is commonly available in mathematical program packages.

where λ_j is an eigenvalue of $(-\mathbf{A})^{-1}$.

Based on the Jordan decomposition of matrix $(-\mathbf{A})^{-1}$ the $\pi(-\mathbf{A})^{-i}\mathbb{1}$ term has the following form

$$\pi(-\mathbf{A})^{-i}\mathbb{1} = \pi\mathbf{\Gamma}^{-1}\mathbf{E}^i\mathbf{\Gamma}\mathbb{1} = \pi\mathbf{\Gamma}^{-1}\mathbf{E}^i\mathbb{1} \tag{6}$$

where the columns of $\mathbf{\Gamma}^{-1}$ are the normalized generalized right eigenvectors of matrix $(-\mathbf{A})^{-1}$. $\mathbf{\Gamma}^{-1}$ is normalized such that $\mathbf{\Gamma}^{-1}\mathbb{1} = \mathbb{1}$.

From (4) and (6), we have

$$\pi\mathbf{\Gamma}^{-1}\mathbf{E}^i\mathbb{1} = \pi'\mathbf{\Gamma}'^{-1}\mathbf{E}'^i\mathbb{1}, \tag{7}$$

for $i \geq 0$, which implies $\mathbf{E} = \mathbf{E}'$ and $\pi\mathbf{\Gamma}^{-1} = \pi'\mathbf{\Gamma}'^{-1}$, since $\mathrm{PH}(\pi, \mathbf{A})$ and $\mathrm{PH}(\pi', \mathbf{A}')$ are non-redundant and $\mathbf{E}$ and $\mathbf{E}'$ are ordered Jordan matrices.

Introducing $\mathbf{B} = \mathbf{\Gamma}^{-1}\mathbf{\Gamma}'$ we have

$$\mathbf{B}^{-1}\mathbf{A}\mathbf{B} = \mathbf{\Gamma}'^{-1}\mathbf{\Gamma}\mathbf{A}\mathbf{\Gamma}^{-1}\mathbf{\Gamma}' = \mathbf{\Gamma}'^{-1}\mathbf{\Gamma}\mathbf{\Gamma}^{-1}\mathbf{E}\mathbf{\Gamma}\mathbf{\Gamma}^{-1}\mathbf{\Gamma}' = \mathbf{\Gamma}'^{-1}\mathbf{E}\mathbf{\Gamma}' = \mathbf{A}',$$

$$\mathbf{B}\mathbb{1} = \mathbf{\Gamma}^{-1}\mathbf{\Gamma}'\mathbb{1} = \mathbf{\Gamma}^{-1}\mathbb{1} = \mathbb{1},$$

where in the last step we used that $\mathbf{\Gamma}^{-1}\mathbb{1} = \mathbb{1}$ since $\mathbf{\Gamma}$ is non-singular and $\mathbf{\Gamma}\mathbb{1} = \mathbb{1}$. Finally multiplying $\pi\mathbf{\Gamma}^{-1} = \pi'\mathbf{\Gamma}'^{-1}$ with $\mathbf{\Gamma}'$ from the right we obtain $\pi\mathbf{B} = \pi'$. ■

Theorem 2 *The distribution of the non-redundant PH($\pi, \mathbf{A}$) of order n is determined by $2n$ independent parameters.*

Proof: It is visible from (2) that the $\pi(-\mathbf{A})^{-i}\mathbb{1}$ series determines the distribution. The number of parameters of

$$\pi(-\mathbf{A})^{-i}\mathbb{1} = \pi\mathbf{\Gamma}^{-1}\mathbf{E}^i\mathbf{\Gamma}\mathbb{1} = v\mathbf{E}^i\mathbb{1} \tag{8}$$

is $2n$ since vector $v = \pi\mathbf{\Gamma}^{-1}$ is composed by n potentially complex elements and $\mathbf{E}$ is determined by the n potentially complex and potentially coincident eigenvalues of $(-\mathbf{A})^{-1}$. ■

Note that, Theorem 1 and 2 are known, see e.g., [15,16]. We present them, together with the given proofs, which differ from the existing ones, for easier understanding of the new results presented in the next section.

In case of phase type distributions without probability mass at zero these $2n$ parameters are redundant since

$$\pi(-\mathbf{A})^{0}\mathbb{1} = \pi\mathbf{\Gamma}^{-1}\mathbf{E}^0\mathbf{\Gamma}\mathbb{1} = \pi\mathbf{\Gamma}^{-1}\mathbb{1} = v\mathbb{1} = 1.$$

In this case the number of independent parameters are $2n-1$.

If $\mathbf{E}$ is a diagonal matrix, i.e., the eigenvalues of $\mathbf{A}$ are distinct, (8) simplifies to

$$\pi(-\mathbf{A})^{-i}\mathbb{1} = \sum_{j=1}^{n} v_j \lambda_j^i, \ i \geq 1, \tag{9}$$

where v_j is the jth element of vector v.

There are alternative proofs of Theorem 2 based on the number of non-trivial coefficients of the order n rational Laplace transform of the distribution, $f^*(s)$, and based on the number of independent moments for a long time. The importance of the parameters applied in this proof is motivated in the next section.

A redundant description of the distribution requires the definition of more than $2n$ parameters. E.g., the $\pi, \mathbf{A}$ representation is defined by n^2+n elements.

3 Markov arrival processes

Let $X(t)$ be a stationary Markov arrival process [14]. The double transform of the number of arrivals in the $(0,t)$ interval starting from an arrival at 0 is

$$f(s,z) = \int_{t=0}^{\infty} e^{-st} E(z^{X(t)}) dt = \pi(s\mathbf{I} - \mathbf{D_0} - z\mathbf{D_1})^{-1}\mathbb{1},$$

where $\mathbf{D_0}$ describes the changes of the phase process without arrival and $\mathbf{D_1}$ describes the changes of the phase process with arrival. We say that $X(t)$ is a MAP$(\mathbf{D_0}, \mathbf{D_1})$ process.

$\mathbf{D_0}$ and $\mathbf{D_1}$ have the following properties:

- $\mathbf{D_0}_{ii} < 0$, $\mathbf{D_0}_{ij} \geq 0$ for $i \neq j$, $\mathbf{D_0}\mathbb{1} \leq 0$,
- $\mathbf{D_1}_{ij} \geq 0$,
- $\mathbf{D_0}\mathbb{1} = -\mathbf{D_1}\mathbb{1}$,
- $\mathbf{D_0}$ is non-singular.

The MAP$(\mathbf{D_0}, \mathbf{D_1})$ process has the following properties:

- The phase process of MAP$(\mathbf{D_0}, \mathbf{D_1})$ is a CTMC with generator $\mathbf{D_0} + \mathbf{D_1}$.
- The phase process embedded into the arrival instances form a DTMC with transition probability matrix $\mathbf{P} = (-\mathbf{D_0})^{-1}\mathbf{D_1}$.
- π is the stationary distribution of the embedded DTMC, i.e., $\pi = \pi\mathbf{P}$, $\pi\mathbb{1} = 1$.
- The stationary inter-arrival time distribution is PH$(\pi, \mathbf{D_0})$.

- The joint density of the $X_0, X_1, \ldots, X_k$ inter-arrival times is

$$f(x_0, x_1, \ldots, x_k) = \pi e^{\mathbf{D_0}x_0}\mathbf{D_1}e^{\mathbf{D_0}x_1}\mathbf{D_1}\ldots e^{\mathbf{D_0}x_k}\mathbf{D_1}\mathbb{1}.$$

- The joint moments of the $a_0 = 0 < a_1 < a_2 < \ldots < a_k$-th inter-arrival times are

$$\begin{aligned} &E(X_0^{i_0}X_{a_1}^{i_1}\ldots X_{a_k}^{i_k}) = \\ &\quad \pi\ i_0!(-\mathbf{D_0})^{-i_0}\mathbf{P}^{a_1-a_0}i_1!(-\mathbf{D_0})^{-i_1}\ldots\mathbf{P}^{a_k-a_{k-1}}i_k!(-\mathbf{D_0})^{-i_k}\mathbb{1}. \end{aligned} \tag{10}$$

Definition 2 *MAP($\mathbf{D_0}, \mathbf{D_1}$) is* irreducible *if its phase process (characterized by $\mathbf{D_0} + \mathbf{D_1}$) is irreducible.*

If MAP($\mathbf{D_0}, \mathbf{D_1}$) is irreducible then the embedded DTMC (characterized by $\mathbf{P}$) has got exactly one recurrent block and at most $n-1$ transient states.

Definition 3 *MAP($\mathbf{D_0}, \mathbf{D_1}$) is* non-redundant *if its rank equals to its order, where the* rank *of MAP($\mathbf{D_0}, \mathbf{D_1}$) is the size the square matrices $\mathbf{D_0}$ and $\mathbf{D_1}$, and the* order *of MAP($\mathbf{D_0}, \mathbf{D_1}$) is the degree of the denominator of $f(s,z)$ as a polynomial of s.*

Let $X'(t)$ be MAP($\mathbf{D'_0}, \mathbf{D'_1}$) with distribution $f'(s,z)$.

Corollary 1 *If the joint moments of the $a_0 = 0 < a_1 < a_2 < \ldots < a_k$-th inter-arrival times of MAP($\mathbf{D_0}, \mathbf{D_1}$) and MAP($\mathbf{D'_0}, \mathbf{D'_1}$) are identical for all $k \geq 0$; $i_0, \ldots, i_k$ and $a_1, \ldots, a_k$ then $f(s,z) \equiv f'(s,z)$.*

Proof: In the convergence region of $f(s,z)$ we have

$$\begin{aligned} &f(s,z) = \pi(s\mathbf{I} - \mathbf{D_0} - z\mathbf{D_1})^{-1}\mathbb{1} = \pi\Big(s(-\mathbf{D_0})^{-1} + \mathbf{I} - z\mathbf{P}\Big)^{-1}(-\mathbf{D_0})^{-1}\mathbb{1} = \\ &\quad \sum_{i=0}^{\infty}\pi\Big(-s(-\mathbf{D_0})^{-1} + z\mathbf{P}\Big)^{i}(-\mathbf{D_0})^{-1}\mathbb{1}\ . \end{aligned} \tag{11}$$

The ith term of the above sum, $\pi(-s(-\mathbf{D_0})^{-1} + z\mathbf{P})^i(-\mathbf{D_0})^{-1}\mathbb{1}$, is composed by the permutations of the $(-\mathbf{D_0})^{-1}$ and the $\mathbf{P}$ matrices. The permutations that starts with $\pi\mathbf{P}^j$ can be simplified to the

$$\pi(-\mathbf{D_0})^{-i_0}\mathbf{P}^{j_0}(-\mathbf{D_0})^{-i_1}\ldots\mathbf{P}^{j_{k-1}}(-\mathbf{D_0})^{-i_k}\mathbb{1} \tag{12}$$

form, since $\pi = \pi\mathbf{P}$. Indeed, (12) is $E(X_0^{i_0}X_{a_1}^{i_1}\ldots X_{a_k}^{i_k})/(i_0!i_1!\ldots i_k!)$, where $a_k = \sum_{\ell=0}^{k-1} j_\ell$. Due to the equality of the joint moments of MAP($\mathbf{D_0}, \mathbf{D_1}$) and MAP($\mathbf{D'_0}, \mathbf{D'_1}$) all terms of the (11) composition of $f(s,z)$ and $f'(s,z)$ are identical, which implies the theorem. ■

Theorem 3 *MAP($\mathbf{D_0}, \mathbf{D_1}$) and MAP($\mathbf{D'_0}, \mathbf{D'_1}$) are two non-redundant irreducible MAPs with distributions $f(s,z)$ and $f'(s,z)$, respectively. $f(s,z) \equiv f'(s,z)$ if and only if there exist a non-singular matrix $\mathbf{B}$ such that $\mathbf{D'_0} = \mathbf{B}^{-1}\mathbf{D_0}\mathbf{B}$, $\mathbf{D'_1} = \mathbf{B}^{-1}\mathbf{D_1}\mathbf{B}$ and $\mathbf{B}^{-1}\mathbb{1} = \mathbb{1}$.*

Proof: If $\mathbf{D'_0} = \mathbf{B}^{-1}\mathbf{D_0}\mathbf{B}$, $\mathbf{D'_1} = \mathbf{B}^{-1}\mathbf{D_1}\mathbf{B}$ and $\mathbf{B}^{-1}\mathbb{1} = \mathbb{1}$ then

$$\mathbf{P'} = (-\mathbf{D'_0})^{-1}\mathbf{D'_1} = (-\mathbf{B}^{-1}\mathbf{D_0}\mathbf{B})^{-1}\mathbf{B}^{-1}\mathbf{D_1}\mathbf{B} = \mathbf{B}^{-1}(-\mathbf{D_0})^{-1}\mathbf{B}\mathbf{B}^{-1}\mathbf{D_1}\mathbf{B} = \mathbf{B}^{-1}\mathbf{P}\mathbf{B}.$$

Multiplying $\pi = \pi\mathbf{P}$ with $\mathbf{B}$ from the right gives

$$\pi\mathbf{B} = \pi\mathbf{P}\mathbf{B} = \pi\mathbf{B}\mathbf{B}^{-1}\mathbf{P}\mathbf{B} = \pi\mathbf{B}\mathbf{P'},$$

which means that the stationary solution associated with $\mathbf{P'}$ is $\pi' = \pi\mathbf{B}$. Using these we have

$$f'(s,z) = \pi'(s\mathbf{I} - \mathbf{D'_0} - z\mathbf{D'_1})^{-1}\mathbb{1} = \pi\mathbf{B}(s\mathbf{I} - \mathbf{B}^{-1}\mathbf{D_0}\mathbf{B} - z\mathbf{B}^{-1}\mathbf{D_1}\mathbf{B})^{-1}\mathbf{B}^{-1}\mathbb{1} = \pi\mathbf{B}\mathbf{B}^{-1}(s\mathbf{I} - \mathbf{D_0} - z\mathbf{D_1})^{-1}\mathbf{B}\mathbf{B}^{-1}\mathbb{1} = f(s,z).$$

If $f(s,z) \equiv f'(s,z)$ then the marginal moments $E(X_0^i)$ and the joint moments $E(X_0^i X_1^j)$ are identical with $E(X_0'^i)$ and $E(X_0'^i X_1'^j)$, respectively, for $i, j = 0, 1, 2, \ldots$.

Composing matrix $\mathbf{B}$ based on the identity of the marginal moments, $\pi(-\mathbf{D_0})^{-i}\mathbb{1} = \pi'(-\mathbf{D'_0})^{-i}\mathbb{1}$, as in Theorem 1 results that $\mathbf{D'_0} = \mathbf{B}^{-1}\mathbf{D_0}\mathbf{B}$, $\pi' = \pi\mathbf{B}$ and $\mathbf{B}^{-1}\mathbb{1} = \mathbb{1}$. Substituting these into the joint moment expression we have

$$\begin{aligned} E(X_0'^i X_1'^j)/i!j! &= \pi'(-\mathbf{D'_0})^{-i}\mathbf{P'}(-\mathbf{D'_0})^{-j}\mathbb{1} \\ &= \pi\mathbf{B}(-\mathbf{B}^{-1}\mathbf{D_0}\mathbf{B})^{-i}\mathbf{P'}(-\mathbf{B}^{-1}\mathbf{D_0}\mathbf{B})^{-j}\mathbb{1} \\ &= \pi(-\mathbf{D_0})^{-i}\mathbf{B}\mathbf{P'}\mathbf{B}^{-1}(-\mathbf{D_0})^{-j}\mathbb{1}\ , \end{aligned} \tag{13}$$

and the identity of $E(X_0^i X_1^j)$ and $E(X_0'^i X_1'^j)$ results

$$\pi(-\mathbf{D_0})^{-i}\mathbf{P}(-\mathbf{D_0})^{-j}\mathbb{1} = \pi(-\mathbf{D_0})^{-i}\mathbf{B}\mathbf{P'}\mathbf{B}^{-1}(-\mathbf{D_0})^{-j}\mathbb{1}, \qquad \forall i, j \geq 0. \tag{14}$$

(14) implies $\mathbf{P} = \mathbf{B}\mathbf{P'}\mathbf{B}^{-1}$ since $\mathbf{D_0}$ is non-singular, π and $\mathbb{1}$ are non-zero and are not orthogonal with the relevant eigenvectors of $\mathbf{D_0}$.

Finally, using $\mathbf{P'} = \mathbf{B}^{-1}\mathbf{P}\mathbf{B}$ we have

$$\mathbf{D'_1} = \mathbf{D'_0}\mathbf{P'} = \mathbf{B}^{-1}\mathbf{D_0}\mathbf{B}\mathbf{B}^{-1}\mathbf{P}\mathbf{B} = \mathbf{B}^{-1}\mathbf{D_0}\mathbf{P}\mathbf{B} = \mathbf{B}^{-1}\mathbf{D_1}\mathbf{B}.$$

■

Theorem 4 *The distribution of an order n non-redundant irreducible MAP is determined by at most n^2 independent parameters.*

Proof: To prove the theorem we provide a description of all joint moments based on n^2 parameters and Corollary 1 ensures that this description also defines the distribution.

Let $-\mathbf{D_0}^{-1} = \mathbf{\Gamma}^{-1}\mathbf{E}\mathbf{\Gamma}$ be the Jordan decomposition of $-\mathbf{D_0}^{-1}$ normalized such that $\mathbf{\Gamma}\mathbb{1} = \mathbb{1}$ and $\mathbf{R} = \mathbf{\Gamma}\mathbf{P}\mathbf{\Gamma}^{-1}$. The $\mathbf{E}$ matrix has the Jordan-block structure $\mathbf{E} = \mathrm{diag}\{\mathbf{E}_j\}$ and $\mathbf{R}$ satisfies $\mathbf{R}\mathbb{1} = \mathbb{1}$ since $\mathbf{\Gamma}\mathbf{P}\mathbf{\Gamma}^{-1}\mathbb{1} = \mathbf{\Gamma}\mathbf{P}\mathbb{1} = \mathbf{\Gamma}\mathbb{1} = \mathbb{1}$.

Using these notations the joint moments can be written as

$$\begin{aligned} &E(X_{a_0}^{i_0}X_{a_1}^{i_1}\dots X_{a_k}^{i_k})/(i_0!i_1!\dots i_k!) = \\ &\pi(-\mathbf{D_0})^{-i_0}\mathbf{P}^{a_1-a_0}(-\mathbf{D_0})^{-i_1}\dots\mathbf{P}^{a_k-a_{k-1}}(-\mathbf{D_0})^{-i_k}\mathbb{1} = \\ &\pi\ \mathbf{\Gamma}^{-1}\mathbf{E}^{i_0}\mathbf{\Gamma}\ \mathbf{P}^{a_1-a_0}\ \mathbf{\Gamma}^{-1}\mathbf{E}^{i_1}\mathbf{\Gamma}\dots\mathbf{P}^{a_k-a_{k-1}}\ \mathbf{\Gamma}^{-1}\mathbf{E}^{i_k}\mathbf{\Gamma}\ \mathbb{1} = \\ &v\mathbf{E}^{i_0}\ \mathbf{R}^{a_1-a_0}\ \mathbf{E}^{i_1}\dots\mathbf{R}^{a_k-a_{k-1}}\ \mathbf{E}^{i_k}\mathbb{1}\ . \end{aligned} \tag{15}$$

where $v = \pi\mathbf{\Gamma}^{-1}$. v is determined by $\mathbf{R}$ because $v\mathbf{R} = v$ and $v\mathbb{1} = 1$, since

$$v = \pi\mathbf{\Gamma}^{-1} = \pi\mathbf{P}\mathbf{\Gamma}^{-1} = \pi\mathbf{\Gamma}^{-1}\mathbf{\Gamma}\mathbf{P}\mathbf{\Gamma}^{-1} = v\mathbf{R},$$

and

$$v\mathbb{1} = \pi\mathbf{\Gamma}^{-1}\mathbb{1} = \pi\mathbb{1} = \mathbb{1}.$$

Based on (15) any joint moment can be determined by $\mathbf{E}$ and $\mathbf{R}$. Matrix $\mathbf{E}$ is determined by the n (potentially partially coinciding, potentially complex) eigenvalues of $(-\mathbf{D_0})^{-1}$. Matrix $\mathbf{R}$ is determined by its $n(n-1)$ (potentially complex) elements, since $\mathbf{R}\mathbb{1} = \mathbb{1}$. All together these give n^2 parameters. ■

4 Transformation between PH representations

4.1 Representations of Phase type distributions

Markovian representation

One of the main problem of PH fitting methods is the redundancy of the representation $\{\pi, \mathbf{A}\}$. This vector-matrix representation is composed by $n^2 + n$ elements and satisfies some simple constraints ($\pi_i \geq 0$, $A_{ii} < 0$, $A_{ij} \geq 0$ for $i \neq j$, $\mathbf{A}\mathbb{1} \leq 0$). We refer to this representation as Markovian representation. The Markovian representation is real valued, but not unique and not

minimal. A representation is called minimal if it defines the distribution of a non-redundant PH distribution of order n based on $2n$ parameters.

Jordan representation

In Section 2 we obtained a further vector-matrix representation, the representation $\{v, \mathbf{E}\}$. This vector-matrix representation is composed by $2n$ non-trivial elements. We refer to this representation as Jordan representation. Assuming the eigenvalues are ordered (according to the real and than the imaginary parts) the Jordan representation is unique and minimal, but it is complex valued if the distribution has a complex eigenvalue.

Laplace representation

The Laplace transform of a non-redundant PH distribution of order n is an order n rational function. The $2n$ coefficients of the properly normalized Laplace transform is referred to as the Laplace representation. The Laplace representation is minimal, unique and real valued.

Moments representation

The first $2n$ moments $(\mu_0, \ldots, \mu_{2n-1})$ define a non-redundant PH distribution of order n [18]. We refer to vector $\mu = \{\mu_0, \ldots, \mu_{2n-1}\}$ as the moments representation a PH distribution. The moments representation is minimal, unique and positive real valued.

MRP representation

In [18], Appie van de Liefvoort recognized the "minimal realization problem" and the algorithm provided by Gragg and Lindquist [8] can be applied to compose a real valued matrix $\mathbf{K}$ such that $\mu_i/i! = e_1\mathbf{K}^i e_1^T$, where e_1 is the row vector whose only non-zero element is the first element which equals to one (i.e., $e_1 = \{1, 0, \ldots, 0\}$). The $\{e_1, \mathbf{K}'\}$ vector-matrix representation, defined below based on matrix $\mathbf{K}$, is referred to as MRP representation. The MRP representation is unique, real, but not minimal [5].

4.2 Transformation methods

From Markovian representation to Jordan representation

The Jordan representation of the non-redundant phase type distribution with Markovian representation $\{\pi, \mathbf{A}\}$ is obtained using the Jordan decomposition

[5] The number of parameters of the MRP representation depends on the eigenvalue structure of the transition matrix. If the eigenvalues are distinct matrix $\mathbf{K}$ is tridiagonal, otherwise it has non-zero elements elsewhere as well.

$(-\mathbf{A})^{-1} = \mathbf{\Gamma}^{-1}\mathbf{E}\mathbf{\Gamma}$, where the similarity matrix $\mathbf{\Gamma}$ is normalized such that $\mathbf{\Gamma}\mathbb{1} = \mathbb{1}$ and $\mathbf{E}$ is the Jordan canonical form of $(-\mathbf{A})^{-1}$ such that the diagonal elements of $\mathbf{E}$ (eigenvalues of $(-\mathbf{A})^{-1}$) are ordered. From this decomposition $v = \pi\mathbf{\Gamma}^{-1}$ and $\mathbf{E}$ provides the Jordan representation.

From Markovian, Jordan and MRP representations to Laplace and moments representations

The properties of the PH distribution can be evaluated from the Markovian, the Jordan and the MRP representations in a similar way.

$$f(t) = \pi\ e^{At}(-\mathbf{A})\mathbb{1} = v\ e^{(-\mathbf{E})^{-1}t}\mathbf{E}^{-1}\mathbb{1} = e_1\ e^{(-\mathbf{K}')^{-1}t}\mathbf{K}'^{-1}\mathbb{1}, \tag{16}$$

$$f^*(s) = \pi\left(s(-\mathbf{A})^{-1} + \mathbf{I}\right)^{-1}\mathbb{1} = v(s\mathbf{E}+\mathbf{I})^{-1}\mathbb{1} = e_1(s\mathbf{K}'+\mathbf{I})^{-1}\mathbb{1}, \tag{17}$$

$$\mu_n = n!\ \pi(-\mathbf{A})^{-n}\mathbb{1} = n!\ v\mathbf{E}^n\mathbb{1} = n!\ e_1\mathbf{K}'^n\mathbb{1}. \tag{18}$$

From Laplace transform to moments representation

$$\mu_i = (-1)^i\ \frac{d^i}{ds^i}f^*(s)\bigg|_{s=0}.$$

From moments representation to MRP representation

The first step of this transformation is the application of the procedure provided in [18] and the second step is a simple transformation to obtain a closing vector equals to $\mathbb{1}$. The procedure in [18] generates the real valued matrix $\mathbf{K}$ such that $\mu_i/i! = e_1\mathbf{K}^i e_1^T$ for $i = 1, 2, \ldots, 2n-1$ and $\mathbf{K}$ is non-redundant. As a consequence any representation obtained from matrix $\mathbf{K}$ via similarity transforms is non-redundant as well. Let $\mathbf{T}$ be the square matrix, whose elements are $T_{ij} = \begin{cases} 1\ i \geq j \\ 0\ otherwise \end{cases}$. Introducing $\mathbf{K}' = \mathbf{T}\mathbf{K}\mathbf{T}^{-1}$, we have $\mu_n/n! = e_1\mathbf{K}^n e_1^T = e_1\mathbf{K}'^n\mathbb{1}$. We refer to vector e_1 and matrix $\mathbf{K}'$ as the MRP representation a PH distribution.

From MRP representation to Jordan representation

The Jordan decomposition of $\mathbf{K}'$, $\mathbf{K}' = \mathbf{\Gamma}^{-1}\mathbf{E}\mathbf{\Gamma}$, results the Jordan representation as $v = e_1\mathbf{\Gamma}^{-1}$ and $\mathbf{E}$.

From Laplace transform to Jordan representation

The previous 3 transformation methods already allow to transform the Laplace representation to the Jordan representation. The *companion form* provides a

simpler way of this transformation.

Based on the coefficients of the rational Laplace transform

$$f^*(s) = \frac{\varphi_n s^{n-1} + \varphi_{n-1} s^{n-2} + \ldots + \varphi_2 s + \varphi_1}{s^n + \varrho_n s^{n-1} + \varrho_{n-1} s^{n-2} + \ldots + \varrho_2 s + \varrho_1} + \varphi_0,$$

we can generate the *companion form* [6]

$$f^*(s) = \nu(s\mathbf{I} - \mathbf{C})^{-1}(-\mathbf{C})e_1^T,$$

where $\nu = \{\frac{\varphi_1}{\varrho_1}, \frac{\varphi_2}{\varrho_1}, \ldots, \frac{\varphi_n}{\varrho_1}\}$ and

$$\mathbf{C} = \begin{pmatrix} 0 & 1 & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \ddots & 1 & 0 \\ 0 & 0 & \ldots & 0 & 1 \\ -\varrho_1 & -\varrho_2 & \ldots & -\varrho_{n-1} & -\varrho_n \end{pmatrix}.$$

Applying the $\mathbf{C}' = \mathbf{T}\mathbf{C}\mathbf{T}^{-1}$ transformation and the Jordan decomposition of $(-\mathbf{C}')^{-1}$, $(-\mathbf{C}')^{-1} = \mathbf{\Gamma}^{-1}\mathbf{E}\mathbf{\Gamma}$, we have the elements of the Jordan representation, $v = \nu\mathbf{T}^{-1}\mathbf{\Gamma}^{-1}$ and $\mathbf{E}$.

From MRP representation to Markovian representation

The above presented transformations allow to transform from any mentioned representation to any other except to a Markovian representation. The Markovian representation is not unique, hence we need a procedure that finds one of the Markovian representations of the given distribution.

According to Theorem 1, all Markovian representations of a PH distribution defined by $\{e_1, \mathbf{K}'\}$ are similar to e_1 and $\mathbf{K}'$ in the sense that there exists a matrix, $\mathbf{B}$, such that $\pi' = e_1\mathbf{B}$ and $\mathbf{A}' = \mathbf{B}^{-1}(-\mathbf{K}')^{-1}\mathbf{B}$ are proper probability vector and transient generator matrix, respectively. Section 6 presents a method to find such matrix $\mathbf{B}$.

4.3 Unique, minimal and real valued representations of PH distributions

From the above list of representations the moments and the Laplace representations are unique, minimal and real valued. For the use of these representation in PH fitting we are in favour of using the moments representation because it

carries more practical information. The main drawback of these two representations is that they do not indicate if a given set of $2n$ moments or the $f^*(s)$ function defines a PH distribution or not. The numerical procedure presented in Section 6 can be applied to answer this question (up to the precision of the numerical procedure).

5 Transformation between MAP representations

5.1 Representations of Markov arrival processes

Markovian representation of MAPs

We refer to the matrix pair $\{\mathbf{D_0}, \mathbf{D_1}\}$ as the Markovian representation of an order n MAP. The number of parameters of this representation is $2n^2 - n$. The Markovian representation is real valued but redundant and not unique.

Jordan representation of MAPs

We refer to the matrix pair $\{\mathbf{E}, \mathbf{R}\}$ as the Jordan representation of the order n MAP. The number of parameters of this representation is n^2, since $\mathbf{E}$ is a Jordan matrix of size n (n parameters), and $\mathbf{R}$ is an $n \times n$ matrix which satisfies $\mathbf{R}\mathbb{1} = \mathbb{1}$ ($n(n-1)$ parameters). The Jordan representation is minimal and unique but potentially complex valued.

Laplace representation of MAPs

We refer to the vector of coefficients of the rational function $f^*(s, z)$ as the Laplace representation of the order n MAP. The number of parameters of this representation is not larger than $2n^2$, since both the numerator and the denominator are at most order n polynomials of s and z. The Laplace representation is unique (with proper normalization), real valued but not minimal.

Moments representation of MAPs

We refer to the $\mu_i = E(X_0^i), i = 1, \ldots, 2n-1$ moments of the inter-arrival time distribution and the $\eta_{ij} = E(X_0^i X_1^j), i, j = 1, \ldots, n-1$ joint moments ($(2n-1) + (n-1)^2 = n^2$ parameters) as the moments representation of the order n MAP. The moments representation is unique, minimal and positive real valued.

MRP representation of MAPs

We refer to the real valued matrix pair $\{\mathbf{K}', \mathbf{R}'\}$ as the MRP representation of the order n MAP, where matrix $\mathbf{K}'$ has the same role as in the case of PH

distributions and matrix $\mathbf{R}'$ represent the dependency of consecutive arrivals. The definitions of these matrices are given in the next subsection. The number of parameters of this representation is usually more than n^2. The MRP representation is unique and real valued but redundant.

5.2 Transformation methods

From Markovian representation to Jordan representation

The Jordan representation of the non-redundant MAP with Markovian representation $\mathbf{D_0}, \mathbf{D_1}$ is obtained using the Jordan decomposition $(-\mathbf{D_0})^{-1} = \mathbf{\Gamma}^{-1}\mathbf{E}\mathbf{\Gamma}$, where the eigenvalues are ordered in $\mathbf{E}$, $\mathbf{\Gamma}$ is normalized such that $\mathbf{\Gamma}\mathbb{1} = \mathbb{1}$ and $\mathbf{R} = \mathbf{\Gamma}\mathbf{P}\mathbf{\Gamma}^{-1}$.

From Markovian, Jordan and MRP representations to Laplace and moments representations

$$\begin{aligned} f^*(s,z) &= \pi(s\mathbf{I}-\mathbf{D_0}-z\mathbf{D_1})^{-1}\mathbb{1} = \pi\left(s(-\mathbf{D_0})^{-1}+\mathbf{I}-z\mathbf{P}\right)^{-1}(-\mathbf{D_0})^{-1}\mathbb{1} \\ &= v(s\mathbf{E}+\mathbf{I}-z\mathbf{R})^{-1}\mathbf{E}\mathbb{1} = v(s\mathbf{K}'+\mathbf{I}-z\mathbf{R}')^{-1}\mathbf{E}'\mathbb{1}, \end{aligned} \tag{19}$$

$$\begin{aligned} \frac{\mathbf{E}(X_0^{i_0}X_1^{i_1}\dots X_k^{i_k})}{i_0!i_1!\dots i_k!} &= \pi(-\mathbf{D_0})^{-i_0}\mathbf{P}(-\mathbf{D_0})^{-i_1}\dots\mathbf{P}(-\mathbf{D_0})^{-i_k}\mathbb{1} \\ &= v\ \mathbf{E}^{i_0}\mathbf{R}\ \mathbf{E}^{i_1}\dots\mathbf{R}\mathbf{E}^{i_k}\mathbb{1} = v\ \mathbf{K}'^{i_0}\mathbf{R}'\ \mathbf{K}'^{i_1}\dots\mathbf{R}'\mathbf{K}'^{i_k}\mathbb{1}. \end{aligned} \tag{20}$$

Further more we have

$$\begin{aligned} f(x_0,x_1,\dots,x_k) &= \pi e^{\mathbf{D_0}x_0}\mathbf{D_1}e^{\mathbf{D_0}x_1}\mathbf{D_1}\dots e^{\mathbf{D_0}x_k}\mathbf{D_1}\mathbb{1} \\ &= \pi e^{\mathbf{D_0}x_0}(-\mathbf{D_0})\mathbf{P}e^{\mathbf{D_0}x_1}(-\mathbf{D_0})\mathbf{P}\dots e^{\mathbf{D_0}x_k}(-\mathbf{D_0})\mathbf{P}\mathbb{1} \\ &= ve^{(-\mathbf{E})^{-1}x_0}\mathbf{E}^{-1}\mathbf{R}e^{(-\mathbf{E})^{-1}x_1}\mathbf{E}^{-1}\mathbf{R}\dots e^{(-\mathbf{E})^{-1}x_k}\mathbf{E}^{-1}\mathbb{1} \\ &= e_1e^{(-\mathbf{K}')^{-1}x_0}\mathbf{K}'^{-1}\mathbf{R}'e^{(-\mathbf{K}')^{-1}x_1}\mathbf{K}'^{-1}\mathbf{R}'\dots e^{(-\mathbf{K}')^{-1}x_k}\mathbf{K}'^{-1}\mathbb{1}. \end{aligned} \tag{21}$$

From moments to MRP representation

The first steps of the transformation is to generate the $\{e_1, \mathbf{K}'\}$ MRP representation of the inter-arrival time distribution based on the moments $\mu_i = E(X_0^i), i = 0, 1, \dots, 2n-1$ as it is discussed in the previous section, and the second step is to obtain matrix $\mathbf{R}'$ based on the $\eta_{ij} = E(X_0^iX_1^j), i, j = 1, \dots, n-1$ joint moments. Note that $\eta_{i0} = E(X_0^iX_1^0) = \eta_{0i} = E(X_0^0X_1^i) = \mu_i$.

The second step is as follows. Based on the $\{e_1, \mathbf{K}'\}$ representation, the $\mu_i = E(X_0^i), i = 1, \ldots, n-1$ moments and the $\eta_{ij} = E(X_0^i X_1^j), i, j = 1, \ldots, n-1$ joint moment, we compose 3 matrices of size $n \times n$. Matrix $\mathbf{N}$ contains the moments such that $N_{ij} = \eta_{i-1,j-1}$, matrix $\mathbf{\Lambda}_{e_1}$ and $\mathbf{\Lambda}_{\mathbf{1}}$ are such that the ith row of $\mathbf{\Lambda}_{e_1}$ is $e_1(i-1)!\mathbf{K}'^{i-1}$ and the jth column of $\mathbf{\Lambda}_{\mathbf{1}}$ is $(j-1)!\mathbf{K}'^{j-1}\mathbb{1}$. That is

$$\mathbf{N} = \begin{pmatrix} 1 & \mu_1 & \mu_2 & \ldots \\ \mu_1 & \eta_{1,1} & \eta_{1,2} & \ldots \\ \mu_2 & \eta_{2,1} & \eta_{2,2} & \ldots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \mathbf{\Lambda}_{e_1} = \left(\begin{array}{c} e_1 \\ \hline e_1\mathbf{K}' \\ \hline e_1 2!\mathbf{K}'^2 \\ \hline \vdots \end{array}\right), \mathbf{\Lambda}_{\mathbf{1}} = \left(\begin{array}{c|c|c|c} \mathbb{1} & \mathbf{K}'\mathbb{1} & 2!\mathbf{K}'^2\mathbb{1} & \ldots \end{array}\right).$$

Based on these matrices, $\mathbf{R}'$ is obtained as $\mathbf{R}' = \mathbf{\Lambda}_{e_1}^{-1}\mathbf{N}\mathbf{\Lambda}_{\mathbf{1}}^{-1}$, since from

$$\mathbf{N}_{ij} = \eta_{i-1,j-1} = E(X_0^{i-1}X_1^{j-1}) = e_1(i-1)!\mathbf{K}'^{i-1}\ \mathbf{R}'\ (j-1)!\mathbf{K}'^{j-1}\mathbb{1},$$

we have $\mathbf{\Lambda}_{e_1}\mathbf{R}'\mathbf{\Lambda}_{\mathbf{1}} = \mathbf{N}$.

From MRP representation to Markovian representation

All Markovian representations of the MAP defined by the $\{\mathbf{K}', \mathbf{R}'\}$ MRP representation are similar to $-\mathbf{K}'^{-1}$ and $\mathbf{R}'$ in the sense that there exists a real valued matrix, $\mathbf{B}$, such that $\mathbf{D_0}' = \mathbf{B}^{-1}(-\mathbf{K}')^{-1}\mathbf{B}$ and $\mathbf{P}' = \mathbf{B}^{-1}\mathbf{R}'\mathbf{B}$ are proper transient generator matrix and transition probability matrix (or equivalently, the $\mathbf{D_0}' = \mathbf{B}^{-1}(-\mathbf{K}')^{-1}\mathbf{B}$ and $\mathbf{D_1}' = \mathbf{B}^{-1}\mathbf{K}'^{-1}\mathbf{R}'\mathbf{B}$ matrices form a proper MAP). Section 6 provides a numerical procedure to obtain matrix $\mathbf{B}$.

5.3 A note on MAP representations

Similar to the case of PH distributions the moments representation is a unique, minimal and positive real valued representation. It has the same drawback that the representation does not indicate if the process is a valid MAP.

According to our knowledge, the Laplace representation of MAPs is far less studied than the Laplace representation of PH distributions. The authors believe that there should be a way to obtain a minimal description of MAPs also in the Laplace domain and to transform the $f^*(s, z)$ function to another representation, but the solutions of these problems are not known for us and are out of the scope of the present paper.

6 Finding a Markovian representation

In this section we focus on order n non-redundant MAPs and matrices of size $n \times n$.

Starting from the real valued matrices $\mathbf{H_0}$ and $\mathbf{H_1}$ which are such that the joint moments

$$E(X_0^i X_1^j) = \rho i!(-\mathbf{H_0})^{-i}\,(-\mathbf{H_0})^{-1}\mathbf{H_1}\; j!(-\mathbf{H_0})^{-j}\mathbb{1}, \qquad i,j \in \{0,1,\ldots\} \quad (22)$$

are the joint moments of a MAP, we look for a non-singular transformation matrix $\mathbf{B}$, for which $\mathbf{B}\mathbb{1} = \mathbb{1}$, and $\mathbf{D_0} = \mathbf{B}^{-1}\mathbf{H_0}\mathbf{B}$ and $\mathbf{D_1} = \mathbf{B}^{-1}\mathbf{H_1}\mathbf{B}$ are valid generator matrices of a MAP.

Indeed, the transformation matrix has to be such that the elements of $\mathbf{B}^{-1}\mathbf{H_1}\mathbf{B}$ and the off-diagonal elements of $\mathbf{B}^{-1}\mathbf{D_0}\mathbf{B}$ are non-negative. The non-positive constraint on the diagonal elements of $\mathbf{B}^{-1}\mathbf{D_0}\mathbf{B}$ is automatically fulfilled in this case, since for a valid MAP $(\mathbf{H_0} + \mathbf{H_1})\mathbb{1} = 0$ and from $\mathbf{B}\mathbb{1} = \mathbb{1}$ we have $(\mathbf{D_0} + \mathbf{D_1})\mathbb{1} = \mathbf{B}^{-1}(\mathbf{H_0} + \mathbf{H_1})\mathbf{B}\mathbb{1} = 0$.

We apply an iterative numerical optimization method to find such matrix $\mathbf{B}$, starting from $\mathbf{H_0}$ and $\mathbf{H_1}$. The goal function and the elementary step of the procedure are as follows.

The goal function

To obtain a numerically stable procedure we apply continuous, many times differentiable error functions that penalize the negative elements, e.g.:

$$\begin{aligned} e_1(a,x) &= e^{-a\cdot x} && a > 0, \\ e_2(a,x) &= e^{-(x-1)^{2a+3}} && a \in \{1,2,\ldots\}, \end{aligned} \quad (23)$$

where a is referred to as the penalty parameter. Based on these error functions the overall goal function is

$$\mathcal{E}_k(\mathbf{H_0},\mathbf{H_1}) = \sum_{i,j,i\neq j} e_k(a,\mathbf{H_0}_{i,j}) + \sum_{i,j} e_k(a,\mathbf{H_1}_{i,j}), \qquad k \in \{1,2\}. \quad (24)$$

The elementary transformation

The elementary transformation selects the best infinitesimal matrix of step size b according to the goal function (if it is better than the current value). The step size indicates how different is the transformed representation from

the original. If $b = 0$ the transformed and the original representations are identical. The infinitesimal transformation matrix of step size b is:

$$\mathbf{B}_{i,j}(b) = \begin{matrix} \\ i \\ \\ \\ \\ j \end{matrix} \begin{matrix} \begin{bmatrix} 1 & 0 & 0 & 0 & \dots \\ 0 & 1 & 0 & 0 & \dots \\ \vdots & \vdots & \ddots & \ddots & \dots \\ 0 & b & 0 & 1-b & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{matrix}, \tag{25}$$

and the elementary transformation selects the i, j pair that minimizes

$$\min_{i,j,i \neq j} \left(\mathcal{E}_k(\mathbf{H_0}_{i,j}(b), \mathbf{H_1}_{i,j}(b)), \mathcal{E}_k(\mathbf{H_0}_{i,j}(-b), \mathbf{H_1}_{i,j}(-b)) \right),$$

if it is less than $\mathcal{E}_k(\mathbf{H_0}, \mathbf{H_1})$, where $\mathbf{H_0}_{ij}(b) = \mathbf{B}^{-1}_{i,j}(b)\mathbf{H_0}\mathbf{B}_{i,j}(b)$ and $\mathbf{H_1}_{ij}(b) = \mathbf{B}^{-1}_{i,j}(b)\mathbf{H_1}\mathbf{B}_{i,j}(b)$ are the results of the infinitesimal transformation of step size b.

6.1 Behaviour of the algorithm

Theoretically, any matrix $\mathbf{B}$ that transforms $\mathbf{H_0}$ and $\mathbf{H_1}$ into a Markovian representation can be composed by a series of elementary transformations with adequately chosen step sizes, but unfortunately, such matrix $\mathbf{B}$ is not known beforehand. By this reason we use a numerical optimization to find a Markovian representation. In practice, we need to minimize the number of iterations and the numerical errors. The penalty parameter dominates the level of numerical error, since at large a values the procedure adds numbers of different orders of magnitude, and the step size parameter dominates the required number of iterations. A large step size ($b > 0.2$) results in a faster convergence, but the algorithm may stop quickly without finding a Markovian solution. The algorithm stops when an elementary step does not improve the goal function.

Based on these properties we use a heuristic approach that starts with a large b ($b = 0.25$) and geometrically increases a when the convergence slows down or stops.

According to our experiences there are cases when the convergence is rather slow (with a given a, b setting) using error function e_1, in these cases the use of error function e_2 might help to move out of this region. The adjustment of

the parameters of our implementation is outlined in Figure 1.

```
PROCEDURE Minimize(H0,H1,errfunc,a,b,iter);
   FOR i=1 TO iter DO
     (H0new,H1new)= ElementaryStep(H0,H1,errfunc,a,b);
     IF Markovian(H0new,H1new) THEN Exit(H0new,H1new);
   IF GoalFunc(H0,H1,errfunc,a,b)=GoalFunc(H0new,H1new,errfunc,a,b) // if no improvement
      THEN Return(H0,H1);                                          // then return
     ELSE (H0,H1)=(H0new,H1new);                                   // else move to the new point
   END;
END;

PROCEDURE Main(H0,H1);
   IF Markovian(H0,H1) THEN Exit(H0,H1);
   b=0.25;
  FOR i=1 TO 10 DO                          // optimization with decreasing step size b
     (H0,H1)= Minimize(H0,H1,e1,2,b,200);        // goal function 1
     (H0,H1)= Minimize(H0,H1,e1,1000,b,200);     // goal function 2
     (H0,H1)= Minimize(H0,H1,e2,1,b,200);        // goal function 3
     b=b/2;                                      // decreasing the step size
   END;
END;
```

Fig. 1. Adjustment of the a and b parameters and the error function

A detailed investigation of this algorithm is provided in [5]. [5] recommends to cut the procedure into 2 passes. During the first pass the goal function is insensitive to the $\mathbf{H_1}$ matrix. This way it might find a transformation that results in a Markovian $\mathbf{H_0'} = \mathbf{B'}^{-1}\mathbf{H_0}\mathbf{B'}$ matrix faster. If $\mathbf{H_0}$ cannot be transformed into a Markovian representation it already indicates that the $\mathbf{H_0}$ and $\mathbf{H_1}$ matrices do not represent a MAP. If the first pass is successful, the second pass starts from the obtained Markovian $\mathbf{H_0'}$ matrix and the associated $\mathbf{H_1'} = \mathbf{B'}^{-1}\mathbf{H_1}\mathbf{B'}$ matrix. In the second pass the goal function is sensitive to both matrices.

6.2 Existence of the solution and application of the algorithm

Theorem 3 and its proof remains valid when $\mathbf{D_0'}$ and $\mathbf{D_1'}$ do not meet the non-negativity (non-positivity) requirements of the Markovian generators. According to this extended version of Theorem 3, for a given $\mathbf{H_0}$ and $\mathbf{H_1}$ there is a solution matrix $\mathbf{B}$ if the moments defined by (22) are the joint moments of a non-redundant MAP of the same order and there is no such matrix $\mathbf{B}$ otherwise. The procedure usually finds a solution in the former case and never finds a solution in the later case as it is indicated by the numerical examples in Section 7.

In case of PH distribution fitting the initial $\mathbf{H_0}$ and $\mathbf{H_1}$ matrices can be obtained, e.g., from the MRP representation as $\mathbf{H_0} = (-\mathbf{K'})^{-1}$ and $\mathbf{H_1} = \mathbb{1}e_1$; and in case of MAP fitting as $\mathbf{H_0} = (-\mathbf{K'})^{-1}$ and $\mathbf{H_1} = \mathbf{R'}$.

6.3 A moments matching method

Based on the pieces presented in the previous sections and subsections we can compose a moments matching method for PH distributions and MAPs:

- transformation from the moments representation to the MRP representation,
- transformation from the MRP representation to the Markovian representation.

To the best of our knowledge this is the first moments matching approach for fitting general PH distributions and MAPs of order n based on the minimal number of parameters.

7 Numerical examples

We have preformed a set of various tests to investigate the properties of the presented moments matching method. It seems to provide a Markovian representation in a wide range of cases. Some of these tests are summarized below.

7.1 Random MAP generation

We applied the transformation methods of Section 5 and the numerical procedure of Section 6 to perform the following series of computations:

step 1) random MAP generation,
step 2) computation of its moments representation,
step 3) computation of the MRP representation from the moments representation,
step 4) finding a MAP representation of the MRP representation.

First we applied the following simple random MAP generation method. The off-diagonal elements of $\mathbf{D_0}$ and elements of $\mathbf{D_1}$ are uniformly distributed samples between 0 and 1 and the diagonal elements of $\mathbf{D_0}$ are set such that $(\mathbf{D_0} + \mathbf{D_1})\mathbb{1} = 0$.

Theoretically we should be able to find a MAP representation from the MRP representation in all cases if our numerical method was perfect. Using the random MAPs generated in the above described way, our method always found a MAP representation. We performed this test up to order 8, where the computation time of our Mathematica implementation increased to the order of a

minute due to the increased complexity of the elementary iteration step.

We also test this procedure with a modified random MAP generator. Instead of the uniformly distributed random samples between 0 and 1, the modified procedure sets the matrix elements to 0 with probability p and to a uniformly distributed random sample between 0 and 1 with probability $1-p$. We dropped the MAPs whose generator matrix, $\mathbf{D_0} + \mathbf{D_1}$, was reducible.

For $p = 0$ the modified random MAP generator is the same as the original one. When p is increasing the number of zero elements increases in $\mathbf{D_0}$ and $\mathbf{D_1}$. In case of $p < 0.5$ the procedure still obtained MAP representation almost always. When p increased above 0.5 there were less and less cases where the procedure managed to find a MAP representation.

Based on this experience we conclude that the random MAPs with non-zero elements allow a wider set of MAP representations than the ones that have several zero elements, and our numerical procedure is more effective to find a MAP representation when there is a wide range of possible MAP representations.

7.2 Random similarity transformation

We check the effect of initial representation on the performance of our numerical procedure according to the following steps:

step 1) initial MAP representation, $(\mathbf{D_0}, \mathbf{D_1})$,
step 2) random matrix generation, $\mathbf{B}$, such that $\mathbf{B}\mathbb{1} = \mathbb{1}$,
step 3) similarity transform, $\mathbf{H_0} = \mathbf{B}^{-1}\mathbf{D_0}\mathbf{B}$, $\mathbf{H_1} = \mathbf{B}^{-1}\mathbf{D_1}\mathbf{B}$,
step 4) finding a MAP representation of $(\mathbf{H_0}, \mathbf{H_1})$.

The applied random matrix generator sets the elements of the first $n-1$ columns of $\mathbf{B}$ to be uniformly distributed samples between $\check{b}$ and $\hat{b}$ and the elements of the last column to satisfy $\mathbf{B}\mathbb{1} = \mathbb{1}$. (Singular matrices were dropped.) The similarity transform with the obtained random matrices almost always resulted non-markovian representations, even with $\check{b} = 0, \hat{b} = 1$.

Tuning the $\check{b}$ and $\hat{b}$ parameters of the random matrix generation we found that the procedure is more or less insensitive to the similarity transformation. When the original $(\mathbf{D_0}, \mathbf{D_1})$ matrices had mainly non-zero elements the procedure found a MAP representation, and when the original $(\mathbf{D_0}, \mathbf{D_1})$ matrices had several zero elements non of the random similarity transformation helped to find a MAP representation.

7.3 Approaching known limits of the Markovian set

We start from the following matrix exponential representation:

$$v = [0.2, 0.3, 0.5], \qquad \mathbf{H} = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -3 & h \\ 0 & -h & -3 \end{pmatrix}.$$

The eigenvalues of $\mathbf{H}$ are $-1, -3 \pm hi$. $F(x) = 1 - ve^{Ht}\mathbb{1}$ is a (matrix exponential, see e.g. [6]) distribution function when $h \leq 2.30033340858$ and $F(x)$ is a Phase type distribution (i.e., has a Markovian representation [11]) when $h \leq 0.552748375$. Our transformation method finds a Markovian representation in 57 iterations when $h = 0.5$, in 3155 iterations when $h = 0.54$ and it fails to find a Markovian representation when $h = 0.545$. Modifying the error functions to e^{-50x}, e^{-200x}, e^{-1000x}, (compared to the one presented in Figure 1) we obtain a Markovian representation:

$$\pi = (0.000476066, 0.0144194, 0.985105)\ ,$$
$$\mathbf{A} = \begin{pmatrix} -1.161615445405 & 1.152943843029 & 0.001001275793779 \\ 0.00547506636475 & -2.91106856299 & 2.833291906724 \\ 0.1783951586341 & 0.0000208821133978 & -2.92731599159 \end{pmatrix}.$$

It seems that increasing h towards 0.552748375 has the same effect as increasing the number of zeros, i.e., it reduces the range of potential Markovian representations of the $F(x)$ distribution. At $h = 0.552748375$ there is one representation (apart of the order of the states), whose structure is the same as the canonical from presented in [9]:

$$\pi = (0, 0.01153060576692, 0.98846939298)\ ,$$
$$\mathbf{A} = \begin{pmatrix} -1.162691021171 & 1.162691021171 & 0 \\ 0 & -2.91865448941 & 2.91865448941 \\ 0.1764859465857 & 0 & -2.91865448941 \end{pmatrix}.$$

7.4 Computational complexity

The computational complexity of the procedure is determined by the complexity of an iteration step and by the number of iterations. The first one depends

only on the size of the problem ($\mathcal{O}(n^2)$, where n is size of the matrices), while the second one depends on the input of the method. When the input is such that it has a wide range of Markovian representations the procedure finds a solution in some 100 iterations. If the input is close to the border of the set of Markovian processes then the number of required iterations is increasing and occasionally the process stops before finding a Markovian representation. When the procedure is called with an input without a Markovian representation it performs the predefined number of iterations and never finds a false Markovian representation.

Considering the robust and run time limited version of our procedure as it is in Figure 1 we conclude that the procedure is able to find a Markovian representation for the major part of the set of Markovian models and it fails only around the border of this set. Whenever it fails it is due to the presence of local optima of the considered goal functions.

8 Conclusions

This paper recommends applying unique and minimal representations of PH and MAP processes instead of the redundant and non-unique Markovian representations and presents a numerical procedure for deciding if a minimal set of moments (composed of $2n-1$ moments in case of continuous PH distributions and n^2 moments and joint moments in case of MAPs) represents a Markovian process or not.

Assuming that the moments set is transformed to MRP representation using the method presented in [18] we have a non-redundant representation which allows using the presented methodology. Currently we are working on the enhancement of the presented numerical method for improving its performance around the border of the Markovian sets. Future research plans include the application of this transformation method to moments based MAP fitting.

Appendix

A Normalized Jordan decomposition

Theorem 5 *If PH($\pi, \mathbf{A}$) is a non-redundant PH distribution then there is an* $(-\mathbf{A})^{-1} = \hat{\mathbf{\Gamma}}^{-1}\mathbf{E}\hat{\mathbf{\Gamma}}$ *Jordan decomposition of* $(-\mathbf{A})^{-1}$ *such that* $\hat{\mathbf{\Gamma}}\mathbb{1} = \mathbb{1}$.

Proof: Let $(-\mathbf{A})^{-1} = \mathbf{\Gamma}^{-1}\mathbf{E}\mathbf{\Gamma}$ be an arbitrary Jordan decomposition of

$(-\mathbf{A})^{-1}$, x_i the ith column of $\mathbf{\Gamma}^{-1}$ and y_i the ith row of $\mathbf{\Gamma}$. We need to prove that $\{\mathbf{\Gamma}\mathbb{1}\}_i = y_i\mathbb{1} \neq 0$ or if $\{\mathbf{\Gamma}\mathbb{1}\}_i = y_i\mathbb{1} = 0$ then there is a $\mathbf{\Gamma}^{-1}\mathbf{E}\mathbf{\Gamma} = \mathbf{\Gamma}^{-1}\mathbf{G}^{-1}\mathbf{E}\mathbf{G}\mathbf{\Gamma}$ transformation for which $\{\mathbf{G}\mathbf{\Gamma}\mathbb{1}\}_i \neq 0$.

Let $\#n$ be the number of Jordan blocks of $\mathbf{E}$ and n_j the size of the jth Jordan block ($\sum_{j=1}^{\#n} n_j = n$). We introduce the Jordan block decomposition

$$\mathbf{\Gamma}^{-1}\mathbf{E}\mathbf{\Gamma} = \sum_{j=1}^{\#n} \mathbf{X_j}\mathbf{E_j}\mathbf{Y_j}, \tag{A.1}$$

where $\mathbf{E_j}$ is the jth Jordan block and $\mathbf{X_j}$ and $\mathbf{Y_j}$ are the associated $n \times n_j$ and $n_j \times n$ part of $\mathbf{\Gamma}^{-1}$ and $\mathbf{\Gamma}$, respectively. Without loss of generality we study only the first Jordan block.

Let λ_1 be the eigenvalue of the first Jordan block and x_ℓ and y_ℓ be the ℓth column of $\mathbf{X_1}$ and the ℓth row of $\mathbf{Y_1}$, respectively, $0 \leq \ell \leq n_1$.

We consider two cases, the first is when the ℓth row of $\mathbf{E_1}$ contains a single non-zero element, the $\lambda_1 > 0$ eigenvalue, in the diagonal ($\ell = n_1$) and the second is when the ℓth row of $\mathbf{E_1}$ contains two non-zero elements, the $\lambda_1 > 0$ eigenvalue in the diagonal and 1 in the sub-diagonal ($\ell < n_1$). In the first case the ℓth row is the last row of the Jordan block, in the second case the ℓth row is one of the internal rows of the Jordan block.

According to (A.1) the contribution of the λ_1 eigenvalue to the ith moment, $\mu_i = i!\ \pi(-\mathbf{A})^{-i}\mathbb{1}$, is characterized by $\pi\mathbf{X_1}{\mathbf{E_1}}^i\mathbf{Y_1}\mathbb{1}$, where, for $i \geq n_1$,

$$\pi\mathbf{X_1}{\mathbf{E_1}}^i\mathbf{Y_1}\mathbb{1} =$$

$$\pi \left(\begin{array}{c|c|c|c} x_1 & x_2 & \dots & x_{n_1} \end{array}\right) \left(\begin{array}{cccc} \lambda_1 & 1 & 0 & \dots \\ 0 & \lambda_1 & \ddots & \ddots \\ \vdots & \ddots & \ddots & 1 \\ 0 & 0 & \dots & \lambda_1 \end{array}\right)^i \left(\begin{array}{c} y_1 \\ \hline y_2 \\ \hline \vdots \\ \hline y_{n_1} \end{array}\right) \mathbb{1} =$$

$$\sum_{\ell=1}^{n_1} \left(\sum_{j=1}^{\ell} \pi x_j\ \lambda_1^{i-j+1}\right) y_\ell\mathbb{1} = \sum_{\ell=1}^{n_1} \lambda_1^{i-j+1} \left(\sum_{j=1}^{n_1-\ell+1} \pi x_j\ y_{j+\ell-1}\mathbb{1}\right) = \sum_{\ell=1}^{n_1} \lambda_1^{i-\ell+1}\ c_{1\ell}. \tag{A.2}$$

In the last expression $c_{1\ell}$ denotes the coefficient of $\lambda_1^{i-\ell+1}$, the contribution of the λ_1 eigenvalue according to the given power. Based on (A.2) $y_{n_1}\mathbb{1}$ is non-zero, because if $y_{n_1}\mathbb{1} = 0$ then $c_{1n_1} = 0$ and PH$(\pi, \mathbf{A})$ is redundant since the multiplicity of the contribution of the λ_1 eigenvalue is less than n_1.

If for any $j, 1 \leq j < n_1$ we have $y_j\mathbb{1} = 0$ we apply the transformation $\mathbf{E_1} =$

$\mathbf{G_1}^{-1}\,\mathbf{G_1}\,\mathbf{E_1} = \mathbf{G_1}^{-1}\,\mathbf{E_1}\,\mathbf{G_1}$ with

$$\mathbf{G_1} = \begin{pmatrix} 1 & g_1 & g_2 & \dots \\ 0 & 1 & \ddots & \ddots \\ \vdots & \ddots & \ddots & g_1 \\ 0 & \dots & 0 & 1 \end{pmatrix}, \tag{A.3}$$

which commutes with $\mathbf{E_1}$ due to the structure of the Jordan blocks. It results

$$\pi\mathbf{X}_1\mathbf{E}_1^i\mathbf{Y}_1\mathbb{1} = \pi \left(\begin{array}{c|c|c|c} & & & \\ x_1 & x_2 & \dots & x_{n_1} \\ & & & \end{array}\right) \mathbf{G}_1^{-1} \begin{pmatrix} \lambda_1 & 1 & 0 & \dots \\ 0 & \lambda_1 & \ddots & \ddots \\ \vdots & \ddots & \ddots & 1 \\ 0 & 0 & \dots & \lambda_1 \end{pmatrix}^i \begin{pmatrix} y_1 + \sum_{k=2}^{n_1} y_k g_{k-1} \\ \hline \vdots \\ \hline y_{n_1-1} + g_1 y_{n_1} \\ \hline y_{n_1} \end{pmatrix} \mathbb{1}. \tag{A.4}$$

Based on $y_{n_1}\mathbb{1} \neq 0$ we set g_1 such that $(y_{n_1-1} + g_1 y_{n_1})\mathbb{1} \neq 0$, and successively we set g_k, $1 \leq k \leq n_1 - 1$, such that all row-sums of $\mathbf{G}_1\mathbf{Y}_1$ are non-zero.

Finally, having $\{\mathbf{\Gamma}\mathbb{1}\}_i \neq 0$ or $\{\mathbf{G\Gamma}\mathbb{1}\}_i \neq 0$ the ith row of $\hat{\mathbf{\Gamma}}$ is composed by the ith row of $\mathbf{\Gamma}$ divided by $\{\mathbf{\Gamma}\mathbb{1}\}_i$ or the ith row of $\mathbf{G\Gamma}$ divided by $\{\mathbf{G\Gamma}\mathbb{1}\}_i$, respectively. ■

References

[1] S. Asmussen and O. Nerman. Fitting Phase-type distributions via the EM algorithm. In *Proceedings: "Symposium i Advent Statistik"*, pages 335–346, Copenhagen, 1991.

[2] D. A. Bini, G. Latouche, and B. Meini. *Numerical Methods for Structured Markov Chains*. Oxford University Press, 2005.

[3] P. Buchholz. An EM-algorithm for MAP fitting from real traffic data. In *Computer Performance Evaluation Modelling Techniques and Tools*, pages 218–236, Urbana, IL, USA, 2003. Springer, LNCS 2794.

[4] A. Cumani. On the canonical representation of homogeneous Markov processes modelling failure-time distributions. *Microelectronics and Reliability*, 22:583–602, 1982.

[5] Tamás Demián. Fitting of Markov traffic models. Technical report, Department of Telecommunications, Technical University of Budapest, master thesis, 2006.

[6] Mark William Fackrell. Characterization of matrix-exponential distributions. Technical report, School of Applied Mathematics, The University of Adelaide, Ph. D. thesis, 2003.

[7] A. Feldman and W. Whitt. Fitting mixtures of exponentials to long-tail distributions to analyze network performance models. *Performance Evaluation*, 31:245–279, 1998.

[8] W. B. Gragg and A. Linquist. On the partial realization problem. *Linear Algebra and Applications*, 50:277–319, 1983.

[9] Qi-Ming He and Hanqin Zhang. A note on unicyclic representation of PH-distributions. *Stochastic Models*, 21:465–483, 2005.

[10] A. Horváth and M. Telek. PhFit: A general purpose phase type fitting tool. In *Tools 2002*, pages 82–91, London, England, April 2002. Springer, LNCS 2324.

[11] G. Horváth and M. Telek. A canonical representation order 3 phase type distributions. Technical report, Dept. of Telecommunications, Technical University of Budapest, tech. rep., 2007.

[12] G. Horváth, M. Telek, and P. Buchholz. A MAP fitting approach with independent approximation of the inter-arrival time distribution and the lag correlation. In *2nd Int. conf. on the quantitative evaluation of systems, QEST*, pages 124–133, Torino, Italy, 2005. IEEE CS.

[13] Rachid El Abdouni Khayari, Ramin Sadre, and Boudewijn R. Haverkort. Fitting world-wide web request traces with the EM-algorithm. *Performance Evaluation*, 52(2-3):175–191, 2003.

[14] G. Latouche and V. Ramaswami. *Introduction to matrix analytic methods in stochastic modeling*. SIAM, 1999.

[15] Colm A. O'Cinneide. On non-uniqueness of representations of phase-type distributions. *Communications in Statistics - Stochastic Models*, 5(2):247–259, 1989.

[16] Colm A. O'Cinneide. Phase-type distributions: open problems and a few properties. *Communications in Statistics - Stochastic Models*, 15(4):731–758, 1999.

[17] T. Rydén. An EM algorithm for estimation in Markov-modulated Poisson processes. *Computational Statistics & Data Analysis*, 21:431–447, 1996.

[18] A. van de Liefvoort. The moment problem for continuous distributions. Technical report, University of Missouri, WP-CM-1990-02, Kansas City, 1990.