Fitting with matrix exponential mixtures generated by discrete probabilistic scaling

Julianna Bor, Giuliano Casale, William Knottenbelt
Department of Computing
Imperial College London
{jb1316,gcasale,wjk}@ic.ac.uk

Evgenia Smirni, Andreas Stathopoulos
Department of Computer Science
William and Mary
{esmirni,andreas}@cs.wm.edu

ABSTRACT
Matrix exponential (ME) distributions generalize phase-type distributions; however, their use in queueing theory is hampered by the difficulty of checking their feasibility. We propose a novel ME fitting algorithm that produces a valid distribution by construction. The ME distribution used during the fitting is a product of independent random variables that are easy to control in isolation. Consequently, the calculation of the CDF and the Mellin transform factorizes, making it possible to use these measures for the fitting without significant restriction on the distribution order. Trace-driven queueing simulations indicate that the resulting distributions yield highly accurate results.

1. INTRODUCTION
Markovian models are ubiquitous tools for the performance evaluation of network traffic as their queueing behavior is well understood [14, 20]. However, these models usually assume correct parameters that are hard to obtain in practice. Therefore, a notable branch of stochastic modeling is focused on fitting distributions to real-world data traces.

Existing fitting methods can be placed in two categories based on whether they use the whole trace or only certain parameters of the trace as their input. Most methods in the former category are of the expectation-maximization (EM) type [17, 4]. The main limitation of EM-based methods is their efficiency. These methods are not suitable to fit longer traces that are required to capture heavy-tailed distributions. Most fitting methods in the latter category use moment matching techniques [16, 18]. Within moment matching techniques, we can distinguish between exact moment matching [19] and approximate moment matching [7]. These methods are more scalable in theory; however, scaling them requires the estimation of higher moments from traces to be provided as inputs for the fitting algorithm. In practice, empirical higher moments are often inaccurate [6].

Another classification of fitting methods is based on the type of distribution that is fitted. One of the most commonly used family of distributions is phase-type (PH) distributions [17, 18]. Their main advantage is that the set of PH distributions is dense in the set of non-negative distributions; therefore, any non-negative distribution can be arbitrarily well-approximated by a PH distribution. Their limitation comes from the fact that PH distributions impose non-negativity constraints on the off-diagonal elements of their corresponding state transition matrix. These constraints add significant complexity to the optimization problem of finding a PH distribution close to a given empirical distribution. PH distributions underpin Markovian arrival processes (MAPs) which are used to model the burstiness of network traffic [8]. The matrix-exponential (ME) distribution is another possible generalization of PH distributions that relaxes the above mentioned non-negativity constraints on the parameters of PH. In terms of the optimization problem, this is very beneficial. Furthermore, queueing results stated for PH and MAP distributions stay intact when using ME distributions [2]. In the case of PH distributions the non-negativity constraints guarantee that any vector-matrix parameter pair produces a valid (non-negative) density function. However, this is no longer the case for ME distributions. It is difficult to check if a given set of parameters produces a valid ME density function. Therefore, whilst PH fitting algorithms are ubiquitous, there are very few fitting algorithm for fitting ME distributions. One example of an ME fitting algorithm using an exact moment matching technique is [12]. Laplace-Stieltjes transforms (LST) are used for fitting in [10]; however, it is not guaranteed that the resulting LST belongs to a valid distribution. This issue is resolved in [9] but the resulting algorithm converges even slower than EM algorithms.

We propose a novel ME fitting algorithm that has the ability to produce high order ME distributions and produces a valid density function at every step of the fitting process. The rest of the paper is structured as follows. Theoretical results on how the proposed family of ME mixtures is constructed are presented in Section 2. Section 3 details how these distributions are used in the fitting algorithm. Numerical results are presented in Section 4.

2. MIXTURE OF ME DISTRIBUTIONS
The family of distributions used for the fitting is constructed as the product of a valid ME distribution of low order and a finite number of independent random discrete scaling factors. Even though the original ME distribution has few degrees of freedom, the scaling factors ensure that the resulting ME mixture is significantly more flexible. Due to this increased flexibility and the fact that it is a product of independent random variables, each of which is easy to control in isolation, the constructed ME mixture is an excellent candidate for fitting. Formally, we have:

Definition 1. Let C be a random variable that takes values $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n) \in \mathbb{R}$ with respective probabilities $\boldsymbol{p} = (p_1, \ldots, p_n)$, where $\sum_{i=1}^{n} p_i = 1$ and $p_i \in [0, 1]$. Then $C \sim CAT(\boldsymbol{\lambda}, \boldsymbol{p})$ is called a categorical distribution, .

Definition 2. Let Y be a random variable with density function

$$f(t) = -\boldsymbol{\alpha} \boldsymbol{A} e^{\boldsymbol{A}t} \mathbf{1} \quad t \geq 0 \tag{1}$$

where $\boldsymbol{\alpha} \in \mathbb{R}^{1\times n}$, $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ and $\mathbf{1} \in \mathbb{R}^{n\times 1}$ is a vector of ones. This is called a matrix exponential distribution of order n, $Y \sim ME(\boldsymbol{\alpha}, \boldsymbol{A})$.

When n is large calculating the density is inefficient due to the matrix exponential component.

PROPOSITION 2.1. *Let $X = YC_1 \cdots C_n$, where Y and C_i are independent random variables, $i = 1, \ldots, n$. Furthermore, $Y \sim ME(\boldsymbol{\alpha}, \mathbf{A})$ and $C_i \sim CAT(\boldsymbol{\lambda}^{(i)}, \boldsymbol{p}^{(i)})$, where $\boldsymbol{\lambda}^{(i)} > 0 \quad i = 1, \ldots, n$. Then X is a finite mixture of matrix exponential distributions.*

PROOF. The product of independent categorical distributions is categorical; therefore, it is enough to prove for $X = YC$ where $C \sim CAT(\boldsymbol{\lambda}, \mathbf{p})$ and $\boldsymbol{\lambda}$ is a strictly positive vector. Conditioning on the values of C, the conditional distribution is $X|\{C = \lambda_j\} = Y\lambda_j \sim ME(\boldsymbol{\alpha}, \frac{1}{\lambda_j}\mathbf{A})$. Therefore, X is the mixture of ME distributions, where each ME in the mixture is a scaled version of Y. □

An immediate consequence of Proposition 2.1 is that X is matrix exponentially distributed since the class of ME distributions is closed under finite mixtures [3]. The proposition above assumes that Y has a valid ME distribution. As stated before, there is not an efficient way to check if a vector $\boldsymbol{\alpha}$ and a matrix $\boldsymbol{A}$ produces a valid non-negative density function. In order to overcome this, we restrict Y to a subclass of matrix exponential distributions which produces a valid density by definition [5, 13]. Here we only include the definition for odd n:

Definition 3. $Y \sim ME(\omega, \boldsymbol{\phi})$ with $\omega \in \mathbb{R}$ and $\boldsymbol{\phi} \in R^n$ has a matrix exponential distribution of order $n = 2k - 1$ with unnormalized density $f(t) = e^{-t} \prod_{i=1}^{(n-1)/2} \cos^2(\omega t - \phi_i)$.

PROPOSITION 2.2. *In the special case where Y has a PH distribution, $Y \sim PH(\boldsymbol{\alpha}, \mathbf{A})$, X defined as in Proposition 2.1 has a KPC-PH distribution,*

$$X \sim Y \otimes C_1^* \otimes \ldots \otimes C_n^*$$

where $C_i^* \sim HyperExp(\boldsymbol{p}^{(i)}, \boldsymbol{\Lambda}^{(i)})$ and $\boldsymbol{\Lambda}^{(i)}$ is a diagonal matrix with $\operatorname{diag}(\boldsymbol{\Lambda}^{(i)}) = -1/\boldsymbol{\lambda}^{(i)}$.

It is easy to see that X has a phase-type distribution, since the class of PH distributions is also closed with respect to finite mixtures [15]. Showing that it has a KPC-PH distribution with the above parameters is through the equality of Mellin transforms. The detailed proof is omitted to conserve space.

3. FITTING ALGORITHM

The next step is to choose the measures based on which the fitting is performed. Defining X as in Proposition 2.1 has several advantages. First, the CDF of the chosen subclass of ME distributions have a closed-form. Furthermore, the scale factors can only take a finite number of different values. Therefore F_X, the CDF of X, can be calculated efficiently even for a large number of states. Next, since X is a product of independent random variables, $E(X)$ also factorizes, or more generally, the Mellin transform, $M_X(s) = E(X^{s-1})$ and $s \in \mathbb{C}$, factorizes. This is important as the Mellin transform of an ME distribution has a matrix exponential component. We use the Mellin transform as it allows the calculation of half moments (i.e. $E(X^{1.5})$) on top of ordinary moments; therefore, higher moments can be avoided by instead using the first $\lfloor n/2 \rfloor$ moments and half moments for the fitting. Both the empirical CDF (F_{tr}) and Mellin transform (M_{tr}) are easily obtainable from a trace. Therefore the objective function C to be minimized during the fitting is

$$C(\overline{x}, \overline{s}) = \|F_{tr}(\overline{x}) - F_X(\overline{x})\| + \|M_{tr}(\overline{s}) - M_X(\overline{s})\| \tag{2}$$

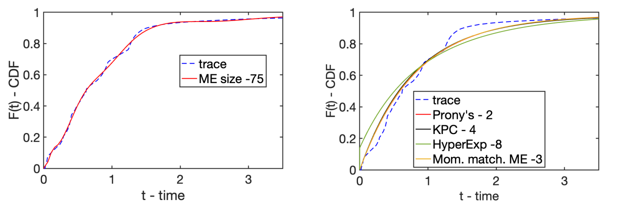
where $\overline{x}, \overline{s}$ are the inputs at which the CDF and Mellin transform are calculated, respectively.

4. NUMERICAL TESTS

Since the optimization problem is highly non-convex the fitting is run several times and the best distribution is chosen in the case of all fitting methods. Two traces have been chosen to demonstrate the adequacy of the proposed method. The first one is the Bellcore pAug89 (`BCAug89`) trace which is often used in literature for the evaluation of the accuracy of fitting methods [1, 11], the second one is the `BuildServer00` trace[1] recorded on 11-28-2007 at 7:55PM. Both traces were randomly shuffled in order to remove temporal dependence.

Figure 1 shows the empirical CDF of the `BCAug89` trace and the fitted ME distribution on the left 1(a), where size-75 refers to the number of states in the distribution. It is compared against the KPC-PH method, moments based ME fitting [12], Prony's method and hyperexponential parameter fitting on the right 1(b). The mixture of ME distributions produces a much closer fit than any of the other methods considered. Figure 2 shows the same for the `BuildServer00` trace. Once more, the mixture of ME distributions considerably outperforms the other methods.

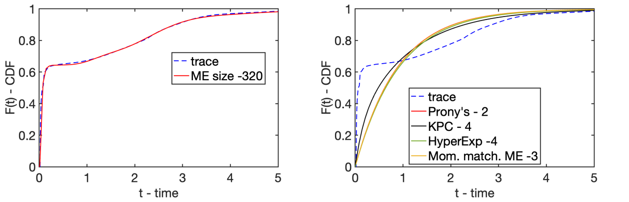
Next, the ME mixture fitted based on the empirical Mellin transform and the empirical CDF (ECDF) is compared to the ME mixture that is fitted based on the ECDF alone. Table 1 shows how including the Mellin transform in the fitting makes sure the moments are matched correctly. This in turn improves how well the tail of the distribution is fitted (Figure 3).



(a) ME mixture (b) KPC and other methods

Figure 1: ECDF of the `BCAug89` trace and CDF of the fitted distributions.

[1] http://iotta.snia.org/traces/block-io/158



(a) ME mixture (b) KPC and other methods

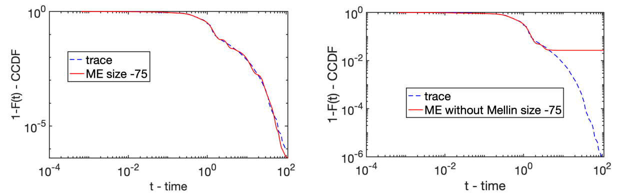
Figure 2: ECDF of the `BuildServer00` trace and CDF of the fitted distributions.

Table 1: Log-moments of `BCAug89` trace and ME dists (fitted with and without the Mellin transform).

$\log(E(X^n))$	Trace	ME w/ Mellin	ME w/o Mellin
$n=1$	0	0	0
$n=2$	1.44	1.44	24.53
$n=3$	4.17	4.17	49.58
$n=4$	7.53	7.56	75.04
$n=5$	11.32	11.65	100.81

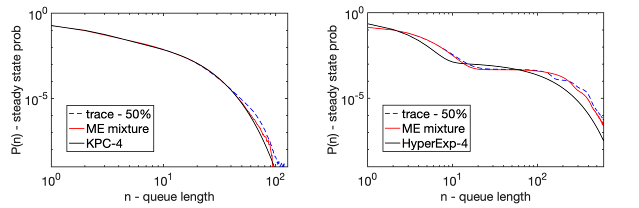
Finally, the fitted ME distributions are sampled and a queueing process is simulated with a fixed hyperexponential arrival process and the fitted distribution as the service process at 50% utilization. Figure 4 shows the steady state queueing probabilities obtained during the simulation first using the shuffled `BCAug89` and `BuildServer00` traces, followed by the corresponding ME mixture samples. Queueing results solved analytically for a fitted KPC-PH and hyperexponential distribution are also displayed for comparison. As expected, the increased accuracy of the fitted distributions translates to the increased accuracy of the queueing probabilities.

The results clearly show the potential of the proposed fitting method. The CDF is followed much more closely by the ME mixture than any of the other fitted distributions. The practical advantage of the improved fitting is demonstrated by the strong improvement in the accuracy of queueing prediction. Given these promising results, the next step is to extend the method to Random Arrival Processes (RAP) in order to be able to fit the auto correlation of traces.



(a) ME with Mellin (b) ME without Mellin

Figure 3: Empirical CCDF of the `BCAug89` trace and CCDF of the fitted ME distribution.



(a) BCAug89 (b) Buildserver00

Figure 4: Steady state queueing probabilities for `BCAug89` and `BuildServer00` trace at 50% utilization level.

5. REFERENCES

[1] A. Andersen et al. A Markovian approach for modeling packet traffic with long-range dependence. *IEEE JSAC*, 16(5):719–732, 1998.

[2] N. G. Bean and B. F. Nielsen. Analysis of queues with rational arrival process components: A general approach. *Perform. Eval. Rev.*, 39(4):31, 2012.

[3] M. Bladt and B. F. Nielsen. *Matrix-Exponential Distributions in Applied Probability*. Springer, 2017.

[4] P. Buchholz. An EM-algorithm for MAP fitting from real traffic data. In *Comp. Perf. Eval. Modelling Tech. and Tools*, pages 218–236. Springer, 2003.

[5] P. Buchholz et al. Stochastic Petri nets with low variation matrix exponentially distributed firing time. *Int. Journal of Perf. Eng.*, 7:441–454, 2011.

[6] P. Buchholz and J. Kriege. A heuristic approach for fitting MAPs to moments and joint moments. In *Proc. of QEST*, pages 53–62, 2009.

[7] G. Casale et al. Characterization of moments and autocorrelation in MAPs. *Perform. Eval. Rev.*, 35(2):27–29, sep 2007.

[8] G. Casale, N. Mi, and E. Smirni. Model-driven system capacity planning under workload burstiness. *IEEE Trans. Comp.*, 59(1):66–80, 2010.

[9] M. Fackrell. Fitting with matrix-exponential distributions. *Stoch. Models*, 21(2-3):377–400, 2005.

[10] C. M. Harris and W. G. Marchal. Distribution estimation using Laplace transforms. *INFORMS J. Comput.*, 10:448–458, 1998.

[11] A. Horváth et al. Markovian modeling of real data traffic: Heuristic phase type and MAP fitting of heavy tailed and fractal like samples. In *Perf. Eval. of Comp. Sys.: Tech. and Tools*, pages 405–434. Springer, 2002.

[12] G. Horváth and M. Telek. Butools 2: a rich toolbox for Markovian perf. eval. In *Proc. of VALUETOOLS*. ACM, 2017.

[13] I. Horváth et al. Concentrated matrix exponential distributions. In *Comp. Perf. Eng.*, pages 18–31. Springer, 2016.

[14] G. Latouche and V. Ramaswami. *Introduction to Matrix Analytic Methods in Stochastic Modeling*. SIAM, 1999.

[15] R. S. Maier and C. A. O'Cinneide. A closure characterisation of phase-type distributions. *J. App. Prob.*, 29(1):92–103, 1992.

[16] T. Osogami and M. Harchol-Balter. Closed form solutions for mapping general distributions to quasi-minimal PH distributions. *Perf. Eval.*, 63(6):524–552, 2006.

[17] A. Riska, V. Diev, and E. Smirni. An EM-based technique for approximating long-tailed data sets with PH distributions. *Perf. Eval.*, 55(1):147–164, 2004.

[18] M. Telek et al. Matching moments for acyclic discrete and continuous phase-type distributions of second order. *Int. Jour. of Sim. Systems, Sci. and Tech.*, 3, 04 2003.

[19] M. Telek and G. Horváth. A minimal representation of Markov arrival processes and a moments matching method. *Perf. Eval.*, 64(9):1153–1168, 2007.

[20] B. Van Houdt and J. S. van Leeuwaarden. Triangular M/G/1-type and tree-like quasi-birth-death Markov chains. *INFORMS J. Computing*, 23(1):165–171, 2011.