KPC-Toolbox: Fitting Markovian Arrival Processes and Phase-Type Distributions with MATLAB

Giuliano Casale*
Imperial College London
Dept. of Computing
London, SW7 2AZ, UK
g.casale@imperial.ac.uk

Evgenia Smirni
College of William & Mary
Computer Science Dept.
Williamsburg, 23187, VA, US
esmirni@cs.wm.edu

1. INTRODUCTION

The KPC-Toolbox[1] is a library of MATLAB scripts for exact and approximate fitting of empirical datasets using Markovian Arrival Processes (MAPs) and Phase-Type distributions (PHs). The approximate fitting methods integrated in the toolbox are based on the Kronecker Product Composition (KPC) operator proposed in [1]. Let us consider a set of MAP random variables $X_j \sim (\boldsymbol{D}_0^j, \boldsymbol{D}_1^j)$, $1 \leq j \leq J$, KPC defines a new random variable

$$X = X_1 \otimes X_2 \otimes \cdots \otimes X_J \sim ((-1)^{J-1} \bigotimes_{j=1}^{J} \boldsymbol{D}_0^j, \bigotimes_{j=1}^{J} \boldsymbol{D}_1^j)$$

where $\otimes$ denotes the Kronecker product. It can be verified that the representation of X specifies a valid MAP if at least $J-1$ matrices $\boldsymbol{D}_0^j$ are diagonal. This constraint holds also for PHs, where $X \sim (\bigotimes_{j=1}^{J} \boldsymbol{\alpha}_j, (-1)^{J-1} \bigotimes_{j=1}^{J} \boldsymbol{T}_j)$, being $(\boldsymbol{\alpha}_j, \boldsymbol{T}_j)$ the representation of PH j and $\boldsymbol{T}_j$ the subgenerator corresponding to $\boldsymbol{D}_0^j$ [2].

The immediate consequence of the KPC definition is that moments and joint moments of X can be recursively computed from the corresponding moments and joint moments of the X_j random variables. For example, for $J = 3$, $E[X^k] = (k!)^{1-J} E[X_1^k]E[X_2^k]E[X_3^k]$. Similar expressions can be derived for joint moments, autocorrelation coefficients, squared coefficient of variation, and decay rates of the autocorrelation function [1]. This decomposition property of KPC offers the opportunity to define divide-and-conquer algorithms for MAP/PH fitting, where numerical optimization is used to set the moments of the variables X_j instead of those of X directly. In the KPC toolbox, the X_j random variables are MAP(2)s, which allows us to search in the space of MAP(2) moments to fit X to a dataset, and this space is well characterized in terms of moment bounds. Thus, analytical constraints can be integrated in the optimization programs.

2. KPC-TOOLBOX

The latest version 0.3.0 of the KPC-Toolbox includes several improvements, especially to the KPC-PH submodule that focuses on PH fitting. First, we have integrated optimization based on the barrier method, which enables to search in the moment space of PH(2)s without exiting the feasible region. Constraints such as $E[X_j^3] > \frac{3}{2} E[X_j^2]^2 / E[X_j]$ for hyper-exponential PH(2) distributions are now expressed in logarithmic form in such a way to define a convex set of linear equations for the $J-1$ PHs with diagonal subgenerator. Furthermore, for hyper-exponential random variables, we have integrated in the KPC-Toolbox an efficient technique to compute the moments of order $k > 3$ based on the linear dependence of moments in MAPs/PHs [1].

[1] http://www.cs.wm.edu/MAPQN/kpctoolbox.html

Another improvement is that KPC-PH is now able to perform exact moment matching (non-KPC-based) for moment sets that belong to the PH(n), Erlang-n, and Hyper-Exponential-n classes. In particular, hyper-exponential distributions are fitted automatically using an adaptation of Prony's method to PH models. Let $\phi(\theta) = \theta^n + m_1\theta^{n-1} + \ldots + m_{n-1}\theta + m_n$ be the characteristic polynomial of $(-\boldsymbol{T})^{-1}$. Then we estimate the m_k coefficients from the empirical moments $E[\widetilde{X}]$ by the following system of linear equations:

$$\begin{cases} \frac{E[\widetilde{X}^n]}{k!} + m_1 \frac{E[\widetilde{X}^{n-1}]}{(n-1)!} + \ldots + m_n = 0 \\ \frac{E[\widetilde{X}^{n+1}]}{(k+1)!} + m_1 \frac{E[\widetilde{X}^{n}]}{n!} + \ldots + m_n E[\widetilde{X}] = 0 \\ \quad\vdots \qquad\qquad \vdots \\ \frac{E[\widetilde{X}^{2n-1}]}{(2k-1)!} + m_1 \frac{E[\widetilde{X}^{2n-2}]}{(2K-2)!} + \ldots + m_n \frac{E[\widetilde{X}^{n-1}]}{(n-1)!} = 0 \end{cases}$$

This enables the parameterization of $\phi(\theta)$ which provides the eigenvalues θ_k of $(-\boldsymbol{T})^{-1}$ and thus the diagonal elements of $\boldsymbol{T}$. Finally, the elements of the $\boldsymbol{\alpha}$ vector are obtained by direct application of the spectral formulas for $E[X^k]$ in MAPs/PHs [1]. Figure 1 shows an example of PH(3) fitted by Prony' method on a network trace from ita.ee.lbl.gov.

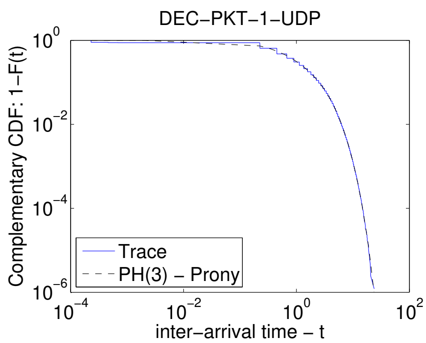

Figure 1: Moment matching by Prony's method

3. REFERENCES

[1] G. Casale, E. Zhang, and E. Smirni. Characterization and synthesis of markovian workload models. In *Proc. GLOBECOM EFSOI workshop*, 2007.

[2] G. Casale, E. Zhang, and E. Smirni. KPC-toolbox: Simple yet effective trace fitting using markovian arrival processes. In *Proc. of QEST*, 83–92, 2008.