Building Accurate Workload Models Using Markovian Arrival Processes

Giuliano Casale
Department of Computing
Imperial College London
Email: g.casale@imperial.ac.uk

ACM SIGMETRICS Tutorial – June 7th, 2011

Outline

Main topics covered in this tutorial:

- Phase-type (PH) distributions
- Moment matching
- Markovian arrival processes (MAP)
- Inter-arrival process fitting

Important topics not covered in this tutorial:

- Queueing applications, matrix-geometric method, ...
- Non-Markovian workload models (e.g., Pareto, matrix exponential process, ARMA processes, fBm, wavelets, ...)
- Maximum-Likelihood (ML) methods, EM algorithm, ...
- ...

1. PH DISTRIBUTIONS

Continuous-Time Markov Chain (CTMC) Notation

- m states
- $\lambda_{i,j} \geq 0$: (exponential) transition rate from state i to j
- $\lambda_i = \sum_{j=1}^{m} \lambda_{i,j}$: total outgoing rate from state i
- Infinitesimal generator matrix:

$$\mathbf{Q} = \begin{bmatrix} -\lambda_1 & \lambda_{1,2} & \dots & \lambda_{1,m} \\ \lambda_{2,1} & -\lambda_2 & \dots & \lambda_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{n,1} & \lambda_{n,2} & \dots & -\lambda_m \end{bmatrix}, \ \mathbf{Q}\mathbb{1} = \mathbf{0}, \ \mathbb{1} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

- $\boldsymbol{\pi}(t) = \boldsymbol{\pi}(0)e^{\mathbf{Q}t}$: state probability vector at time t
  - $\pi_i(t)$: probability of the CTMC being in state i at time t
  - $\boldsymbol{\pi}(0)$: initial state probability vector

Example 1.1: CTMC Transient Analysis

$$\mathbf{Q} = \begin{bmatrix} -4 & 4 & 0 & 0 \\ 4 & -7 & 2 & 1 \\ 2 & 3 & -5 & 0 \\ 2 & 0 & 0 & -2 \end{bmatrix}$$

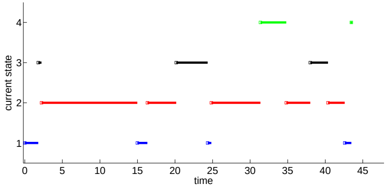
Initial state:

$$\pi(0) = [0.9, 0.0, 0.1, 0.0]$$

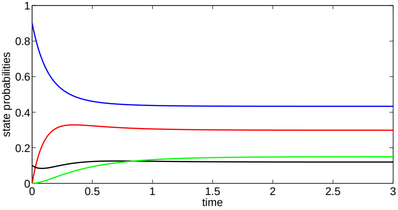
Transient analysis:

$$\pi(t) = \pi(0)e^{\mathbf{Q}t} = \pi(0)\sum_{k=0}^{\infty}\frac{(\mathbf{Q}t)^k}{k!}$$

A Sample Path



Transient Probabilities



Phase-Type Distribution (PH)

- $\mathbf{Q}$: CTMC with $m = n + 1$ states, last is absorbing ($\lambda_{n+1} = 0$)

$$\mathbf{Q} = \left[\begin{array}{c:c} \mathbf{T} & \mathbf{t} \\ \hdashline \mathbf{0} & 0 \end{array}\right] = \left[\begin{array}{cccc:c} -\lambda_1 & \lambda_{1,2} & \dots & \lambda_{1,n} & \lambda_{1,n+1} \\ \lambda_{2,1} & -\lambda_2 & \dots & \lambda_{2,n} & \lambda_{2,n+1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \lambda_{n,1} & \lambda_{n,2} & \dots & -\lambda_n & \lambda_{n,n+1} \\ \hdashline 0 & 0 & \dots & 0 & 0 \end{array}\right]$$

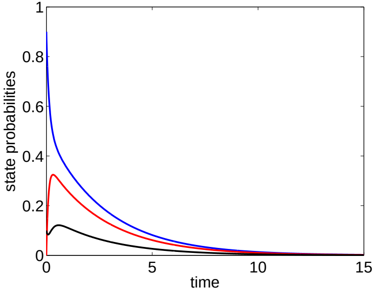
- $\mathbf{T}$: PH subgenerator matrix, $\mathbf{T}\mathbb{1} < 0$
- $\mathbf{t} = -\mathbf{T}\mathbb{1}$: exit vector
- "No mass at zero" assumption: $\boldsymbol{\pi}(0) = [\boldsymbol{\alpha}, 0]$, $\boldsymbol{\alpha}\mathbb{1} = 1$

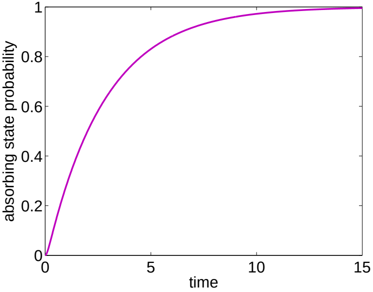
References: [Neu89]

Example 1.2: Absorbing State Probability

Initial state: $\boldsymbol{\pi}(0) = [\boldsymbol{\alpha}, 0.0] = [0.9, 0.0, 0.1, 0.0]$

$$\mathbf{Q} = \left[\begin{array}{ccc|c} -4 & 4 & 0 & 0 \\ 4 & -7 & 2 & 1 \\ 2 & 3 & -5 & 0 \\ \hline 0 & 0 & 0 & 0 \end{array}\right], \mathbf{T} = \begin{bmatrix} -4 & 4 & 0 \\ 4 & -7 & 2 \\ 2 & 3 & -5 \end{bmatrix}, \mathbf{t} = -\mathbf{T}\mathbb{1} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$





PH: Fundamental Concepts

Basic idea: model $F(t) = \Pr(\text{event occurs in } X \leq t \text{ time units})$ as the probability mass absorbed in t time units in $\mathbf{Q}$.

- Semantics: entering the absorbing state models the occurrence of an event, *e.g.*, the arrival of a TCP packet, the completion of a job, a non-exponential state transition in a complex system
- Understanding the absorption dynamics:

$$\boldsymbol{\pi}(t) = \boldsymbol{\pi}(0)e^{\mathbf{Q}t} = \boldsymbol{\pi}(0)\left[\begin{array}{c:c} \sum_{k=0}^{\infty} \frac{(\mathbf{T}t)^k}{k!} & \sum_{k=1}^{\infty} \frac{\mathbf{T}^{k-1}t^k}{k!}\mathbf{t} \\ \hdashline \mathbf{0} & 1 \end{array}\right]$$

- Using the definition $\mathbf{t} = -\mathbf{T}\mathbb{1}$, we get

$$\boldsymbol{\pi}(t) = [\boldsymbol{\alpha} e^{\mathbf{T}t}, \pi_{n+1}(t)], \quad \pi_{n+1}(t) = 1 - \boldsymbol{\alpha} e^{\mathbf{T}t}\mathbb{1}$$

where $\pi_{n+1}(t)$ is the probability mass absorbed in t time units.

PH: Fundamental Formulas

PH distribution: $F(t) = \Pr(\text{event occurs in } X \leq t) \stackrel{def}{=} 1 - \boldsymbol{\alpha} e^{\mathbf{T}t} \mathbb{1}$

- PH representation: $(\boldsymbol{\alpha}, \mathbf{T})$
- Probability Density function: $f(t) = \boldsymbol{\alpha} e^{\mathbf{T}t} (-\mathbf{T}) \mathbb{1}$
- (Power) Moments: $E[X^k] = \int_0^{+\infty} t^k f(t) dt = k! \boldsymbol{\alpha} (-\mathbf{T})^{-k} \mathbb{1}$
- $(-\mathbf{T})^{-1} = [\tau_{i,j}] \geq 0$
  - $\tau_{i,j}$: mean time spent in state j if the PH starts in state i
- Median/Percentiles: no simple form, determined numerically.

Example 1.3: a 3-state PH distribution

$$\boldsymbol{\alpha} = [1,0,0],\ \mathbf{T} = \begin{bmatrix} -(\lambda+\mu) & \lambda & \mu \\ 0 & -\lambda & \lambda \\ 0 & 0 & -\lambda \end{bmatrix}, \mathbf{t} = \begin{bmatrix} 0 \\ 0 \\ \lambda \end{bmatrix},\ \lambda \geq 0.$$

- Case $\mu \to +\infty$: $E[X^k] = k!\lambda^{-k}$ (Exponential, $c^2 = 1$).
- Case $\mu = \lambda$: $E[X^k] = (k+1)!\lambda^{-k}$ (Erlang-2, $c^2 = 1/2$)
- Case $\mu = 0$: $E[X^k] = \frac{(k+2)!}{2}\lambda^{-k}$ (Erlang-3, $c^2 = 1/3$).
- No choice of μ delivers $c^2 > 1$

Remarks: Squared coefficient of variation: $c^2 \stackrel{def}{=} Var[X]/E[X]^2$

PH: "Family Picture" - $n \leq 2$ states

	c^2	$\boldsymbol{\alpha}$	$\mathbf{T}$	*Subset of*
Exponential	1	$[1]$	$\begin{bmatrix} -\lambda \end{bmatrix}$	Hyper-Exp.
Erlang	$\frac{1}{2}$	$[1, 0]$	$\begin{bmatrix} -\lambda & \lambda \\ 0 & -\lambda \end{bmatrix}$	Hypo-Exp.
Hypo-Exp.	$[\frac{1}{2}, 1)$	$[1, 0]$	$\begin{bmatrix} -\lambda_1 & \lambda_1 \\ 0 & -\lambda_2 \end{bmatrix}$	Coxian/APH
Hyper-Exp.	$[1, +\infty)$	$[\alpha_1, \alpha_2]$	$\begin{bmatrix} -\lambda_1 & 0 \\ 0 & -\lambda_2 \end{bmatrix}$	Coxian/APH
Coxian/APH	$[\frac{1}{2}, +\infty)$	$[1, 0]$	$\begin{bmatrix} -\lambda_1 & p_1\lambda_1 \\ 0 & -\lambda_2 \end{bmatrix}$	General
General	$[\frac{1}{2}, +\infty)$	$[\alpha_1, \alpha_2]$	$\begin{bmatrix} -\lambda_1 & p_1\lambda_1 \\ p_2\lambda_2 & -\lambda_2 \end{bmatrix}$	

Remarks: $\alpha_1 + \alpha_2 = 1$, $c^2 \stackrel{def}{=} Var[X]/E[X]^2$

PH: "Family Picture" - Examples — $n = 3$ states

	c^2	$\boldsymbol{\alpha}$	$\mathbf{T}$
Hyper-Erlang	$[\frac{1}{2}, +\infty)$	$[\alpha_1, \alpha_2, 0]$	$\begin{bmatrix} -\lambda_1 & 0 & 0 \\ 0 & -\lambda_2 & \lambda_2 \\ 0 & 0 & -\lambda_2 \end{bmatrix}$
Coxian/APH	$[\frac{1}{3}, +\infty)$	$[1, 0, 0]$	$\begin{bmatrix} -\lambda_1 & p_1\lambda_1 & 0 \\ 0 & -\lambda_2 & p_2\lambda_2 \\ 0 & 0 & -\lambda_3 \end{bmatrix}$
Circulant	$[\frac{1}{3}, +\infty)$	$[\alpha_1, \alpha_2, \alpha_3]$	$\begin{bmatrix} -\lambda_1 & p_{12}\lambda_1 & 0 \\ 0 & -\lambda_2 & p_{23}\lambda_2 \\ p_{31}\lambda_3 & 0 & -\lambda_3 \end{bmatrix}$
General	$[\frac{1}{3}, +\infty)$	$[\alpha_1, \alpha_2, \alpha_3]$	$\begin{bmatrix} -\lambda_1 & p_{12}\lambda_1 & p_{13}\lambda_1 \\ p_{21}\lambda_2 & -\lambda_2 & p_{23}\lambda_2 \\ p_{31}\lambda_3 & p_{32}\lambda_3 & -\lambda_3 \end{bmatrix}$

Remarks: $\alpha_1 + \alpha_2 + \alpha_3 = 1$, $c^2 \stackrel{def}{=} Var[X]/E[X]^2$

Example 1.4: Reducing to Coxian Form

Algorithms exist to reduce a PH to Coxian form.

- With $n = 2$ states (PH(2) models) this can be done analytically
- Hyper-Exponential: $\boldsymbol{\alpha}' = [0.99, 0.01]$, $\mathbf{T}' = \begin{bmatrix} -25 & 0 \\ 0 & -5 \end{bmatrix}$, $F'(t) = 1 - 0.99e^{-25t} - 0.01e^{-5t}$
- Coxian: $\boldsymbol{\alpha} = [1, 0]$, $\mathbf{T} = \begin{bmatrix} -\lambda_1 & p_1\lambda_1 \\ 0 & -\lambda_2 \end{bmatrix}$
- Symbolic analysis gives that in the 2-state Coxian

$$F(t) = 1 - M_1 e^{-\lambda_1 t} - (1 - M_1)e^{-\lambda_2 t}, \quad M_1 \stackrel{def}{=} 1 - \frac{\lambda_1 p_1}{\lambda_1 - \lambda_2}$$

- Thus the two models are equivalent if $\lambda_1 = 25$, $\lambda_2 = 5$, and $p_1 = 0.008$ such that $M_1 = 0.99$.

Example 1.4: Reducing to Coxian Form

Compare moments of Hyper-Exponential and Coxian:

	Hyper-Exp	Coxian
$E[X]$	$41.600 \cdot 10^{-3}$	$41.600 \cdot 10^{-3}$
$E[X^2]$	$3.968 \cdot 10^{-3}$	$3.968 \cdot 10^{-3}$
$E[X^3]$	$860.160 \cdot 10^{-6}$	$860.160 \cdot 10^{-6}$
$E[X^4]$	$444.826 \cdot 10^{-6}$	$444.826 \cdot 10^{-6}$
$\vdots$	$\vdots$	$\vdots$

Key message: differences between $(\boldsymbol{\alpha}', \mathbf{T}')$ and $(\boldsymbol{\alpha}, \mathbf{T})$ are deceptive! In general, $(\boldsymbol{\alpha}, \mathbf{T})$ is a redundant representation.

Redundancy problem: how many degrees of freedom in PH distributions? How to cope with redundant parameters?

Example 1.5: Fallacies About Degrees of Freedom

- Coxian, 3 parameters: $\alpha' = [1, 0]$, $\mathbf{T'} = \begin{bmatrix} -1.0407 & 0.3264 \\ 0 & -8.0181 \end{bmatrix}$
- Fit PH(2) with 4 parameters: $\alpha = [1, 0]$, $\mathbf{T} = \begin{bmatrix} -\lambda_1 & p_1\lambda_1 \\ p_2\lambda_2 & -\lambda_2 \end{bmatrix}$
- Numerically search $(\lambda_1, \lambda_2, p_1, p_2)$ that minimize the distance from Coxian's $E[X]$,$E[X^2]$,$E[X^3]$ and from $E[X^4] = 50$

$$\mathbf{T} = \begin{bmatrix} -13.4252 & 13.3869 \\ 0.0018 & -1.0431 \end{bmatrix}$$

... returned PH has $E[X^4] = 21.34$, why is it approximately the same as the Coxian's $E[X^4] = 21.41$?

- Key message: 4 parameters $\neq$ freedom to assign $E[X^4]$
- For fixed $E[X]$,$E[X^2]$,$E[X^3]$, the feasible region of the PH(2) parameters yields the same $E[X^4]$ (up to numerical tolerance).

PH: Degrees of Freedom

- PH Moments: $E[X^k] = k!\boldsymbol{\alpha}(-\mathbf{T})^{-k}\mathbb{1}$
- A of order n, characteristic polynomial:
  - $\phi(\theta) = \det(\theta I - A)$
  - $\phi(A) = A^n + m_1 A^{n-1} + \ldots + m_{n-1}A + m_n I = 0$
- $A = (-\mathbf{T})^{-1}$ implies that PH moments are linearly-dependent

$$\frac{E[X^n]}{n!} + m_1 \frac{E[X^{n-1}]}{(n-1)!} + \ldots + m_{n-1}E[X] + m_n = 0$$

- Thus, a PH offers up to $2n-1$ degrees of freedom (df) for fitting a workload distribution ($n-1$ moments + n terms m_j).
- $n = 2$ states $\Rightarrow$ 3 df, $n = 3 \Rightarrow$ 5 df, $n = 4 \Rightarrow$ 7 df, ...

References: [TelH07],[CasZS07]

PH: Algebra of Random Variables and Closure Properties

- PH: the smallest family of distributions on $\Re^+$ that is closed under a finite number of mixtures and convolutions.
- $X_1 \sim (\boldsymbol{\alpha}, \mathbf{T})$ of order n, $\mathbf{t} = -\mathbf{T}\mathbb{1}$
- $X_2 \sim (\boldsymbol{\beta}, \mathbf{S})$ of order m, $\mathbf{s} = -\mathbf{S}\mathbb{1}$
- $Z = g(X_1, X_2) \sim (\boldsymbol{\gamma}, \mathbf{R})$

Z	Convolution $\sum_i X_i$	Mixture X_i w.p. p_i	Minimum $\min(X_i)$	Maximum $\max(X_i)$
$\boldsymbol{\gamma}$	$[\boldsymbol{\alpha}, 0]$	$[p_1\boldsymbol{\alpha}, p_2\boldsymbol{\beta}]$	$[\boldsymbol{\alpha} \otimes \boldsymbol{\beta}]$	$[\boldsymbol{\alpha} \otimes \boldsymbol{\beta}, 0, 0]$
$\mathbf{R}$	$\begin{bmatrix} \mathbf{T} & \mathbf{t} \cdot \boldsymbol{\beta} \\ \mathbf{0} & \mathbf{S} \end{bmatrix}$	$\begin{bmatrix} \mathbf{T} & \mathbf{0} \\ \mathbf{0} & \mathbf{S} \end{bmatrix}$	$\mathbf{T} \oplus \mathbf{S}$	$\begin{bmatrix} \mathbf{T} \oplus \mathbf{S} & \mathbf{t} \otimes \mathbf{I}_m & \mathbf{I}_n \otimes \mathbf{s} \\ \mathbf{0} & \mathbf{S} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{T} \end{bmatrix}$

- $\otimes \stackrel{def}{=}$ Kronecker product, $\oplus \stackrel{def}{=}$ Kronecker sum

References: [MaiO92],[Neu89]

PH: Kronecker operators

- $\mathbf{A}$ of order n, $\mathbf{B}$ of order m
- Kronecker sum: $\mathbf{A} \oplus \mathbf{B} = \mathbf{I}_n \otimes \mathbf{B} + \mathbf{A} \otimes \mathbf{I}_m$

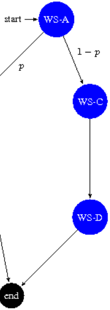
$$\mathbf{A} \oplus \mathbf{B} = \begin{bmatrix} a_{1,1} + b_{1,1} & b_{1,2} & a_{1,2} & 0 \\ b_{2,1} & a_{1,1} + b_{2,2} & 0 & a_{1,2} \\ a_{2,1} & 0 & a_{2,2} + b_{1,1} & b_{1,2} \\ 0 & a_{2,1} & b_{2,1} & a_{2,2} + b_{2,2} \end{bmatrix}$$

- Kronecker product:

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{1,1} b_{1,1} & a_{1,1} b_{1,2} & a_{1,2} b_{1,1} & a_{1,2} b_{1,2} \\ a_{1,1} b_{2,1} & a_{1,1} b_{2,2} & a_{1,2} b_{2,1} & a_{1,2} b_{2,2} \\ a_{2,1} b_{1,1} & a_{2,1} b_{1,2} & a_{2,2} b_{1,1} & a_{2,2} b_{1,2} \\ a_{2,1} b_{2,1} & a_{2,1} b_{2,2} & a_{2,2} b_{2,1} & a_{2,2} b_{2,2} \end{bmatrix}$$

References: [Bre78]

Example 1.6: BPEL Workflow



End-to-end resp. times:

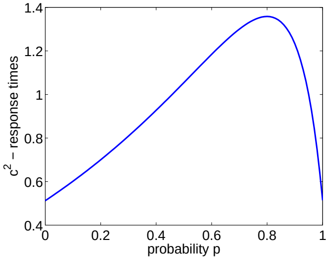
$$\boldsymbol{\alpha} \equiv (1, 0, 0, 0, 0), q \stackrel{def}{=} 1 - p$$

$$\mathbf{T} = \left[\begin{array}{c|c|cc|c} -5 & 5p & \frac{5}{3}q & \frac{10}{3}q & 0 \\ \hline 0 & -7 & 0 & 0 & 0 \\ \hline 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & -2 & 2 \\ \hline 0 & 0 & 0 & 0 & -3 \end{array}\right]$$

Web service response time distributions:

- A: $\boldsymbol{\alpha} \equiv [1], \mathbf{T} \equiv [-5]$
- B: $\boldsymbol{\alpha} \equiv [1], \mathbf{T} \equiv [-7]$
- C: $\boldsymbol{\alpha} \equiv [\frac{1}{3}, \frac{2}{3}], \mathbf{T} \equiv \begin{bmatrix} -1 & 1 \\ 0 & -2 \end{bmatrix}$
- D: $\boldsymbol{\alpha} \equiv [1], \mathbf{T} \equiv [-3]$

Prediction:



2. MOMENT MATCHING

Includes joint work with Evgenia Smirni (IBM Research, William & Mary)

PH: Moment Bounds and Approximate Fitting

- Feasibility constraints for PH fitting:
  - $\lambda_{i,j} \in \Re_0^+$, $\alpha_i \in \Re_0^+$, $\boldsymbol{\alpha}\mathbb{1} = 1$, $\lambda_i = \sum_{j=1}^{m} \lambda_{i,j}$
  - $\mathbf{t} = -\mathbf{T}\mathbb{1} \geq 0$ and $\mathbb{1}^T\mathbf{t} > 0$
  - $\mathbf{T} + \mathbf{t} \cdot \boldsymbol{\alpha}$ is irreducible
- Moment bounds are available for some PH models to determine if a set of empirical moments can be fitted exactly, e.g.:
  - PH(n): $c^2 \geq \frac{1}{n}$ (and Erlang has the smallest c^2)
  - PH(2): $c^2 > 1 \Rightarrow E[\widetilde{X}^3] > \frac{3}{2}\frac{E[\widetilde{X}^2]^2}{E[\widetilde{X}]}$
  - ...
- What can we fit with a PH? What is the best approximating PH for an infeasible set of moments?

References: [AldS87],[TelH03],[OsoH06]

PH: Spectral Characterization

- Spectral analysis based on Jordan canonical forms
- θ_i: eigenvalue of $(-\mathbf{T})^{-1}$ with algebraic multiplicity q_i
- For diagonalizable $(-\mathbf{T})^{-1}$

$$E[X^k] = k! \sum_{i=1}^{n} M_{i,1}\theta_i^k, \qquad F(t) = 1 - \sum_{i=1}^{n} M_{i,1} e^{-t/\theta_i}$$

where $\sum_{i=1}^{m} M_{i,1} = 1$

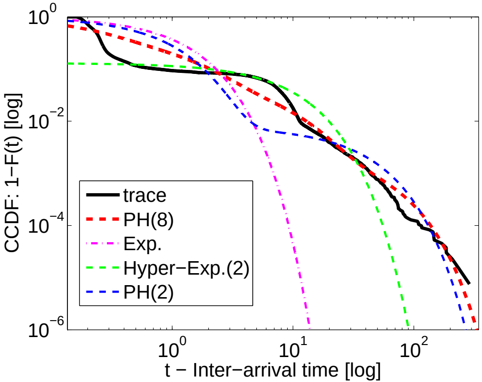
- Special case: $M_{i,1} = \alpha_i$ for hyper-exponential distributions.
- Tail Behavior: $F(t) \approx 1 - M_{i_{max},1} e^{-t/\theta_{i_{max}}}$, $\theta_{i_{max}} \geq \theta_i \ \forall i$

References: [CasZS07]

Example 2.1: Approximating Heavy-Tail Distributions

- Heavy-Tail distribution: $\lim_{t\to\infty} e^{\mu t} F(t) = \infty,\ \forall \mu > 0$
- Multiple decay rates in PH enable approximating non-exponential tails
- Moment matching usually fits the tail better than the body

Example: Radius-Auth trace 08-30-07.12-59-AM, http://iotta.snia.org



PH: Exact Moment Matching Method

- For $n \leq 3$ states, $\boldsymbol{\alpha}$ and $\mathbf{T}$ can be expressed directly as a function of $2n-1$ empirical moments $E[\widetilde{X}^k]$ by means of canonical forms.
- Canonical form: non-redundant form of $\boldsymbol{\alpha}$ and $\mathbf{T}$, same expressive power but $2n-1$ parameters.

	$n = 2$ states	$n = 3$ states	$n \geq 4$ states
$\boldsymbol{\alpha}$	$(1, 0)$	$(\alpha_1, \alpha_2, 1-\alpha_1-\alpha_2)$	unknown
$\mathbf{T}$	$\begin{bmatrix} -\lambda_1 & p\lambda_1 \\ 0 & -\lambda_2 \end{bmatrix}$	$\begin{bmatrix} -\lambda_1 & 0 & q\lambda_1 \\ \lambda_2 & -\lambda_2 & 0 \\ 0 & \lambda_3 & -\lambda_3 \end{bmatrix}$	unknown

- For $n = 3$, it is possible to reduce the number of parameters to $2n-1$ by setting either $q = 0$, $\lambda_1 = \lambda_2$, or $\alpha_2 = 0$ depending on the numerical values of the moments.

References: [HorT09]

Example 2.2: Exact Moment Matching – PH(2)

- Symbolic analysis of $\phi(A)$, $A = (-\mathbf{T})^{-1}$, shows that

$$\begin{cases} m_1 = -(\lambda_1^{-1} + \lambda_2^{-1}) = \frac{E[\widetilde{X}^3] - 3E[\widetilde{X}]E[\widetilde{X}^2]}{3(2E[\widetilde{X}]^2 - E[\widetilde{X}^2])} \\ m_2 = (\lambda_1 \lambda_2)^{-1} = \frac{\frac{3}{2}E[\widetilde{X}^2]^2 - E[\widetilde{X}]E[\widetilde{X}^3]}{3(2E[\widetilde{X}]^2 - E[\widetilde{X}^2])} \\ p = \lambda_2(E[\widetilde{X}] - \lambda_1^{-1}) \end{cases}$$

- Solving for $(\lambda_1, \lambda_2, p)$ we obtain the canonical form
- Symbol solution is feasible, but yields very complex expressions

MATLAB Code: same model of Example 1.4

```
E=[41.600e-3,3.968e-3,860.160e-6]; % 2n-1=3 independent moments
D=3*(2*E(1)^2-E(2)); m1=(E(3)-3*E(1)*E(2))/D, m2=(1.5*E(2)^2-E(1)*E(3))/D,
[lam,fval] = fsolve(@(lam) [-sum(1./lam) - m1;1./prod(lam) - m2],rand(1,2)),
p = lam(2)*(E(1)-1/lam(1))
Output: lam = [ 24.9948; 4.9999], p = 0.0080
```

PH: Prony's Method

Exact moment matching method for $c^2 > 1$

- Obtain m_1,...,m_n by solving the linear system

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

- Obtain θ_i as roots of $\phi(\theta) = \theta^n + m_1\theta^{n-1} + \ldots + m_{n-1}\theta + m_n$
- Obtain $M_{i,1}$ from the spectral charization given θ_i and $E[\widetilde{X}^k]$.
- Output: $\boldsymbol{\alpha} = (M_{1,1}, \ldots, M_{K,1})$, $\mathbf{T} = -\text{diag}(\theta_1^{-1}, \ldots, \theta_K^{-1})$

References: [CasZS08b]

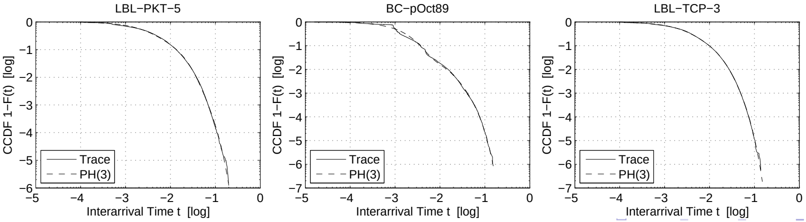
Example 2.3: Prony's Method

Trace: LBL-TCP-3, http://ita.ee.lbl.gov
MATLAB Code, PH(3):

```
E=[1,5.3109e-003,89.7845e-006,3.0096e-006,163.1872e-009,12.9061e-009];
f=factorial(0:6);
A=[E(4:-1:1)./f(4:-1:1);E(5:-1:2)./f(5:-1:2); E(6:-1:3)./f(6:-1:3);1,0,0,0]; b=[0;0;0;1];
m=(A\b)', theta=roots(m)'
M=([theta*f(2);theta.^2*f(3);theta.^3*f(4)]\E(2:4)')'
```

$$\boldsymbol{\alpha} = [M_{1,1}, M_{2,1}, M_{3,1}] = [26.2074, 384.7137, 589.0789] \cdot 10^{-3};$$

$$\mathbf{T} = -\mathrm{diag}(\theta_1^{-1}, \theta_2^{-1}, \theta_3^{-1}) = -\mathrm{diag}(49.9628, 110.5089, 451.3770)$$



PH: Approximate Moment Matching

- Relatively few techniques for approximate moment matching
- Limited understanding of moment bounds for $n \geq 3$
- Kronecker product composition for PH (KPC-PH):
  - $X_1 \sim (\boldsymbol{\alpha}, \mathbf{T})$ of order n, $X_2 \sim (\boldsymbol{\beta}, \mathbf{S})$ of order m
  - $X = X_1 \otimes X_2 \sim (\gamma, \mathbf{R})$, $\gamma \stackrel{def}{=} \boldsymbol{\alpha} \otimes \boldsymbol{\beta}$, $\mathbf{R} \stackrel{def}{=} (-\mathbf{T}) \otimes \mathbf{S}$
  - $(\gamma, \mathbf{R})$ is PH only if $\mathbf{S} = -\text{diag}(\lambda_1, \ldots, \lambda_n)$
- Divide-and-conquer approximate moment matching:

$$\begin{aligned} E[X^k] =& k!(\boldsymbol{\alpha} \otimes \boldsymbol{\beta})(-((-\mathbf{T}) \otimes \mathbf{S}))^{-k}(\mathbb{1}_n \otimes \mathbb{1}_m) \\ =& k!(\boldsymbol{\alpha}(-\mathbf{T})^{-k}\mathbb{1}_n)(\boldsymbol{\beta}(-\mathbf{S})^{-k}\mathbb{1}_m) \\ =& E[X_1^k]E[X_2^k]/k! \end{aligned}$$

References: [CasZS08]

Example 2.4: KPC-PH – Increased Degrees of Freedom

- X_1: PH(2), $E[X_1] = 1$, $E[X_1^2] = 10$, $E[X_1^3] = 200$
- X_2: PH(2), $E[X_2] = 1$, $E[X_2^2] = 10$, $E[X_2^3] = 3200$
- Y_1: PH(2), $E[Y_1] = 1$, $E[Y_1^2] = \mathbf{3.2691}$, $E[Y_1^3] = 200$
- Y_2: PH(2), $E[Y_2] = 1$, $E[Y_2^2] = \mathbf{30.589}$, $E[Y_2^3] = 3200$
- Z_1: PH(2), $E[Z_1] = 1$, $E[Z_1^2] = \mathbf{20}$, $E[Z_1^3] = 200$
- Z_2: PH(2), $E[Z_2] = 1$, $E[Z_2^2] = \mathbf{5}$, $E[Z_2^3] = 3200$

	$X_1 \otimes X_2$	$Y_1 \otimes Y_2$	$Z_1 \otimes Z_2$
$E[X]$	1.0000	1.0000	1.0000
$E[X^2]$	$5.0000 \cdot 10^1$	$5.0000 \cdot 10^1$	$5.0000 \cdot 10^1$
$E[X^3]$	$1.0667 \cdot 10^5$	$1.0667 \cdot 10^5$	$1.0667 \cdot 10^5$
$E[X^4]$	$7.2362 \cdot 10^8$	$7.2362 \cdot 10^8$	$\mathbf{3.7813 \cdot 10^8}$
$E[X^5]$	$5.2745 \cdot 10^{12}$	$\mathbf{6.3130 \cdot 10^{12}}$	$\mathbf{1.6796 \cdot 10^{12}}$
$E[X^6]$	$4.8979 \cdot 10^{16}$	$\mathbf{6.6129 \cdot 10^{16}}$	$\mathbf{8.9531 \cdot 10^{15}}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$

PH: Generalized KPC-PH Technique

Generalization:

- $X = \bigotimes_{j=1}^{J} X_j \sim (\bigotimes_{j=1}^{J} \boldsymbol{\alpha}_j, (-1)^{J-1} \bigotimes_{j=1}^{J} \mathbf{T}_j)$
- X is PH if $J-1$ subgenerators are diagonal
- $E[X^k] = k! \prod_{j=1}^{J} \frac{E[X_j^k]}{k!}$

KPC-Toolbox: http://www.cs.wm.edu/MAPQN/kpctoolbox.html

- Support for exact and approximate moment matching
- Determine $(\boldsymbol{\alpha}_j, \mathbf{T}_j)$ by numerical optimization
- Search on moments directly instead of $(\boldsymbol{\alpha}, \mathbf{T})$ parameters

References: [CasZS08]

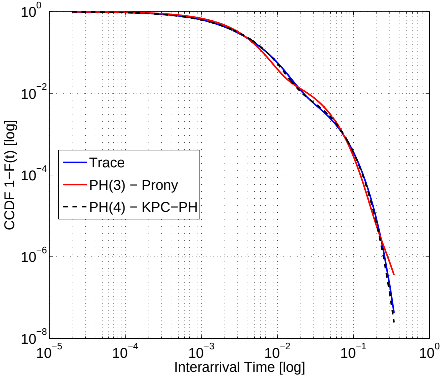
Example 2.5: KPC-Toolbox Fitting

- Internet Traffic Archive Trace: BC-Aug-p89, http://ita.ee.lbl.gov
- Prony's method fails if $n > 3$: $\theta_2 = 0.0026 + i0.0179$
- How to find a perturbation of the moment sets that delivers more accurate result?

KPC-PH, PH(4):

$$\mathbf{T} = -\,\mathrm{diag}(283.677, 1114.3, 39.630, 155.67)$$

$$\boldsymbol{\alpha} = [0.66082, 0.31138, 0.01889, 0.00890]$$



PH: Other Publicly Available Tools for PH Fitting

EMpht (1996) — http://home.imf.au.dk/asmus/pspapers.html:

- EM algorithm for ML fitting, based on Runge-Kutta methods
- Local optimization technique

jPhase (2006) — http://copa.uniandes.edu.co/software/jmarkov/index.html:

- Java library — ML and canonical form fitting algorithms

PhFit (2002) — http://webspn.hit.bme.hu/~telek/tools.htm:

- Separate fit of distribution body and tail
- Both continuous and discrete ML distributions

G-FIT (2007) — http://ls4-www.cs.uni-dortmund.de/home/thummler/gfit.tgz:

- Hyper-Erlang PHs used as building block
- Automatic aggregation of large traces, dramatic speed-up of computational times compared to EMpht

3. MARKOVIAN ARRIVAL PROCESS

Time Series Analysis

Notation:

- $\{t_0 \stackrel{def}{=} 0, t_1, t_2, t_3, \ldots\}$: sequence of arrival times of events in the real system
- $X_k \stackrel{def}{=} t_k - t_{k-1}$: inter-arrival time between arrival of the $(k-1)$-th and the k-th events.
- t_k and X_k may not be directly observable, e.g., aggregate data

Stochastic Process Descriptions

- Inter-arrival process: models sequence of values X_k
  - Natural description for unaggregated traces
  - Enables reasoning on individual events (e.g., response time distributions, covariance of successive arrivals, ...)
- Counting process: models number of arrivals in interval $[0, t]$
  - Preferred for aggregate data (e.g., packet counts)
  - Enables reasoning on the volumes of events at timescale t
- The two descriptions are equivalent in theory; in practice they carry independent information when fitted on a dataset

Sequence of PH Inter-Arrival Times

- PH-Renewal process:
  - for $k = 1, 2, 3, \ldots$
    - initialize a PH with subgenerator $\mathbf{T}$ according to $\boldsymbol{\alpha}$
    - generate X_k as the time to absorption in $(\boldsymbol{\alpha}, \mathbf{T})$
    - $t_k = t_{k-1} + X_k$ is the arrival time of the k-th event
- Limitation: every time the PH is restarted independently of the past. No way to define time-varying patterns, e.g., periodicities, burstiness, ...

References: [Neu89]

PH-R: Counting Process

- PH-Renewal Process State $(N(t), J(t))$
- $N(t)$: event counter increased upon arrival/restart events
- $J(t)$: PH state within current level

$$\boldsymbol{\pi}^c(0) = \begin{bmatrix} \boldsymbol{\alpha} \\ \mathbf{0} \\ \mathbf{0} \\ \vdots \end{bmatrix} \quad \mathbf{Q}^c = \begin{bmatrix} \mathbf{T} & \mathbf{t}\cdot\boldsymbol{\alpha} & \mathbf{0} & \mathbf{0} & \dots \\ \mathbf{0} & \mathbf{T} & \mathbf{t}\cdot\boldsymbol{\alpha} & \mathbf{0} & \dots \\ \mathbf{0} & \mathbf{0} & \mathbf{T} & \mathbf{t}\cdot\boldsymbol{\alpha} & \dots \\ \vdots & \vdots & \vdots & \ddots & \ddots \end{bmatrix} \begin{matrix} \rightarrow N(t) = 0 \\ \rightarrow N(t) = 1 \\ \rightarrow N(t) = 2 \\ \vdots \end{matrix}$$

Markovian Arrival Process (MAP)

- MAP: generalizes the PH-Renewal construction by considering restarts that depend on the exit state of the previous sample.
  - A technical device to introduce "memory" in the time series.
- MAP Counting Process:

$$\boldsymbol{\pi}^c(0) = \begin{bmatrix} \boldsymbol{\alpha} \\ \mathbf{0} \\ \mathbf{0} \\ \vdots \end{bmatrix} \quad \mathbf{Q}^c = \begin{bmatrix} \mathbf{D}_0 & \mathbf{D}_1 & \mathbf{0} & \mathbf{0} & \dots \\ \mathbf{0} & \mathbf{D}_0 & \mathbf{D}_1 & \mathbf{0} & \dots \\ \mathbf{0} & \mathbf{0} & \mathbf{D}_0 & \mathbf{D}_1 & \dots \\ \vdots & \vdots & \vdots & \ddots & \ddots \end{bmatrix} \begin{matrix} \rightarrow N(t) = 0 \\ \rightarrow N(t) = 1 \\ \rightarrow N(t) = 2 \\ \vdots \end{matrix}$$

- $\mathbf{D}_0$: same as $\mathbf{T}$, transitions do not increase $N(t)$
- $\mathbf{D}_1$: generalizes $\mathbf{t} \cdot \boldsymbol{\alpha}$, transitions increase $N(t)$ by 1
- Batch MAP (BMAP): $\mathbf{D}_b$, $b > 1$, increasing $N(t)$ by b units

References: [Neu89]

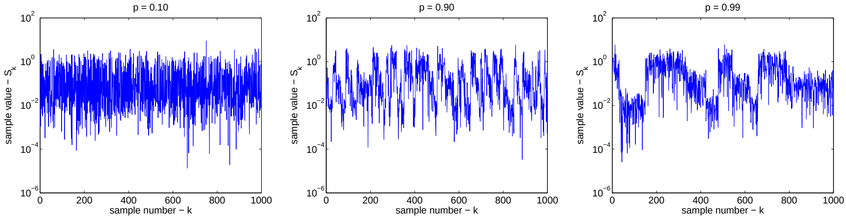
Example 3.1: Sequence of MAP(2) Inter-Arrival Times

$$\mathbf{D}_0 = \begin{bmatrix} -10 & 3 \\ 5 & -5 \end{bmatrix}, \mathbf{D}_1 = \begin{bmatrix} 5 & 2 \\ 0 & 0 \end{bmatrix}$$

- $k = 1$; start in a random state according to $\boldsymbol{\alpha}$, e.g., state 2.
- $X_k = 0$
- $X_k = X_k + r$, $r \sim \exp(5)$. Jump to state 1.
- $X_k = X_k + r$, $r \sim \exp(10)$.
- $u \sim \text{uniform}(0, 1)$.
  - if $u \in [0, \frac{3}{10})$ jump to state 2
  - if $u \in [\frac{3}{10}, \frac{3+5}{10})$: save X_k; $X_{k+1} = 0$; restart from state 1.
  - if $u \in [\frac{3+5}{10}, \frac{3+5+2}{10}]$: save X_k; $X_{k+1} = 0$; restart from state 2
- ...

Example 3.2: Temporal Dependence in MAPs

$$\mathbf{D}_0 = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -10 & 0 \\ 0 & 0 & -100 \end{bmatrix}, \mathbf{D}_1 = \begin{bmatrix} p & 1-p & 0 \\ 0 & 10p & 10(1-p) \\ 100(1-p) & 0 & 100p \end{bmatrix}$$



MAP: "Family Picture" - $n \leq 2$ states

Name	$\mathbf{D}_0$	$\mathbf{D}_1$
Poisson	$\begin{bmatrix} -\lambda \end{bmatrix}$	$\begin{bmatrix} \lambda \end{bmatrix}$
Erlang Renewal Process	$\begin{bmatrix} -\lambda & \lambda \\ 0 & -\lambda \end{bmatrix}$	$\begin{bmatrix} 0 & 0 \\ \lambda & 0 \end{bmatrix}$
Hyper-exp. Renewal Process	$\begin{bmatrix} -\lambda_1 & 0 \\ 0 & -\lambda_2 \end{bmatrix}$	$\begin{bmatrix} p\lambda_1 & q\lambda_1 \\ p\lambda_2 & q\lambda_2 \end{bmatrix}$
Interrupted Poisson Process	$\begin{bmatrix} -\lambda_1 & \lambda_{1,2} \\ \lambda_{2,1} & -\lambda_2 \end{bmatrix}$	$\begin{bmatrix} \lambda^*_{1,1} & \lambda^*_{1,2} \\ 0 & 0 \end{bmatrix}$
MMPP	$\begin{bmatrix} -\lambda_1 & \lambda_{1,2} \\ \lambda_{2,1} & -\lambda_2 \end{bmatrix}$	$\begin{bmatrix} \lambda^*_{1,1} & 0 \\ 0 & \lambda^*_{2,2} \end{bmatrix}$
Acyclic MAP(2)	$\begin{bmatrix} -\lambda_1 & \lambda_{1,2} \\ 0 & -\lambda_2 \end{bmatrix}$	$\begin{bmatrix} \lambda^*_{1,1} & 0 \\ \lambda^*_{2,1} & \lambda^*_{2,2} \end{bmatrix}$
MAP(2)	$\begin{bmatrix} -\lambda_1 & \lambda_{1,2} \\ \lambda_{2,1} & -\lambda_2 \end{bmatrix}$	$\begin{bmatrix} \lambda^*_{1,1} & \lambda^*_{1,2} \\ \lambda^*_{2,1} & \lambda^*_{2,2} \end{bmatrix}$

Remarks: $p + q = 1$

MAP: "Family Picture" - Examples – $n = 3$ states

Name	$\mathbf{D}_0$	$\mathbf{D}_1$
Hyper-Exp. Renewal	$\begin{bmatrix} -\lambda_1 & 0 & 0 \\ 0 & -\lambda_2 & 0 \\ 0 & 0 & -\lambda_3 \end{bmatrix}$	$\begin{bmatrix} \lambda_1 p & \lambda_1 q & \lambda_1 r \\ \lambda_1 p & \lambda_2 q & \lambda_2 r \\ \lambda_3 p & \lambda_3 q & \lambda_3 r \end{bmatrix}$
Circulant MMPP(3)	$\begin{bmatrix} -\lambda_1 & 0 & \lambda_{1,1} \\ \lambda_{2,1} & -\lambda_2 & 0 \\ 0 & \lambda_{3,2} & -\lambda_3 \end{bmatrix}$	$\begin{bmatrix} \lambda^*_{1,1} & 0 & 0 \\ 0 & \lambda^*_{2,2} & 0 \\ 0 & 0 & \lambda^*_{3,3} \end{bmatrix}$
MMPP(3)	$\begin{bmatrix} -\lambda_1 & \lambda_{1,2} & \lambda_{1,3} \\ \lambda_{2,1} & -\lambda_2 & \lambda_{2,3} \\ \lambda_{3,1} & \lambda_{3,2} & -\lambda_3 \end{bmatrix}$	$\begin{bmatrix} \lambda^*_{1,1} & 0 & 0 \\ 0 & \lambda^*_{2,2} & 0 \\ 0 & 0 & \lambda^*_{3,3} \end{bmatrix}$
Hyper-Exp. MAP	$\begin{bmatrix} -\lambda_1 & 0 & 0 \\ 0 & -\lambda_2 & 0 \\ 0 & 0 & -\lambda_3 \end{bmatrix}$	$\begin{bmatrix} \lambda^*_{1,1} & \lambda^*_{1,2} & \lambda^*_{1,3} \\ \lambda^*_{2,1} & \lambda^*_{2,2} & \lambda^*_{2,3} \\ \lambda^*_{3,1} & \lambda^*_{3,2} & \lambda^*_{3,3} \end{bmatrix}$
MAP(3)	$\begin{bmatrix} -\lambda_1 & \lambda_{1,2} & \lambda_{1,3} \\ \lambda_{2,1} & -\lambda_2 & \lambda_{2,3} \\ \lambda_{3,1} & \lambda_{3,2} & -\lambda_3 \end{bmatrix}$	$\begin{bmatrix} \lambda^*_{1,1} & \lambda^*_{1,2} & \lambda^*_{1,3} \\ \lambda^*_{2,1} & \lambda^*_{2,2} & \lambda^*_{2,3} \\ \lambda^*_{3,1} & \lambda^*_{3,2} & \lambda^*_{3,3} \end{bmatrix}$

Remarks: $p + q + r = 1$

MAP: Stationarity

- What is the (marginal) distribution of each sample?

$$X_1 \sim (\boldsymbol{\alpha}_1, \mathbf{D}_0),\ \boldsymbol{\alpha}_1 = \boldsymbol{\alpha}$$

$$X_2 \sim (\boldsymbol{\alpha}_2, \mathbf{D}_0),\ \boldsymbol{\alpha}_2 = \boldsymbol{\alpha}_1 e^{\mathbf{D}_0 x_1} \mathbf{D}_1 \text{ if } X_1 = x_1$$

$$\ldots$$

- Since $(\boldsymbol{\alpha}_1, \mathbf{D}_0) \neq (\boldsymbol{\alpha}_2, \mathbf{D}_0)$ in general, how to choose $\boldsymbol{\alpha}$ to generate stationary and identically distributed samples?

MAP: Interval-Stationary Initialization

- MAP samples $X_1, X_2, \ldots$ are stationary and identically distributed as $(\boldsymbol{\alpha}, \mathbf{T})$ if and only if

$$\boldsymbol{\alpha} = \int_0^{+\infty} \boldsymbol{\alpha} e^{\mathbf{D}_0 t} \mathbf{D}_1 dt = \boldsymbol{\alpha}(-\mathbf{D}_0)^{-1}\mathbf{D}_1 \stackrel{def}{=} \boldsymbol{\alpha}\mathbf{P}$$

- $\boldsymbol{\alpha}$: eigenvector corresponding to the unit eigenvalue of $\mathbf{P}$
- $\mathbf{P} = [p_{i,j}]$: discrete-time Markov chain (DTMC) embedded at restart instants, i.e.,
  $p_{i,j} = \Pr[X_{k+1} \text{ starts in state } j | X_k \text{ starts in state } i]$
- $\boldsymbol{\alpha}$: equilibrium probability vector of the DTMC $\mathbf{P}$
- $\mathbf{P}^h = [q_{i,j}]$: estimate initial state for non-successive samples
  $q_{i,j} = \Pr[X_{k+h} \text{ starts in state } j | X_k \text{ starts in state } i]$

MAP: Key Formulas for Inter-Arrival Times

- MAP representation: $(\mathbf{D}_0, \mathbf{D}_1)$
- Embedded chain: $\mathbf{P} = (-\mathbf{D}_0)^{-1}\mathbf{D}_1$
- Interval-stationary initial vector: $\boldsymbol{\alpha} = \boldsymbol{\alpha}\mathbf{P}$

Distribution of samples:

- (Marginal) Distribution: $F(t) = 1 - \boldsymbol{\alpha} e^{\mathbf{D}_0 t}\mathbb{1}$
- (Marginal) Density: $f(t) = \boldsymbol{\alpha} e^{\mathbf{D}_0 t}(-\mathbf{D}_0)\mathbb{1} = \boldsymbol{\alpha} e^{\mathbf{D}_0 t}\mathbf{D}_1\mathbb{1}$
- (Marginal) Moments: $E[X^k] = k!\boldsymbol{\alpha}(-\mathbf{D}_0)^{-k}\mathbb{1}$
- Degrees of freedom of distribution: $2n - 1$

MAP: Key Formulas for Inter-Arrival Times

Sequence of samples:

- Joint Density Function:

  $$\Pr(X_1 = x_1, X_2 = x_2, \ldots, X_q = x_q) = \boldsymbol{\alpha} e^{\mathbf{D}_0 x_1} \mathbf{D}_1 e^{\mathbf{D}_0 x_2} \mathbf{D}_1 \cdots e^{\mathbf{D}_0 x_q} \mathbf{D}_1 \mathbb{1}$$

  - $\mathbf{D}_1 \to \mathbf{D}_1 \mathbf{P}^{h-1}$ for samples spaced by h lags
- Joint Moments:

  $$E[X_1^{k_1} X_2^{k_2} \cdots X_q^{k_q}] = k_1! \cdots k_q! \boldsymbol{\alpha} (-\mathbf{D}_0)^{-k_1} \mathbf{P} (-\mathbf{D}_0)^{-k_2} \cdots (-\mathbf{D}_0)^{-k_q} \mathbb{1}$$

  - $\mathbf{P} \to \mathbf{P}^h$ for samples spaced by h lags
- Analysis often limited to second-order properties.
- Autocorrelation function:

  $$\rho_h = \frac{E[X_1 X_{1+h}] - E[X]^2}{E[X^2] - E[X]^2} = \frac{\boldsymbol{\alpha}(-\mathbf{D}_0)^{-1} \mathbf{P}^h (-\mathbf{D}_0)^{-1} \mathbb{1} - \boldsymbol{\alpha}(-\mathbf{D}_0)^{-1} \mathbb{1}}{2\boldsymbol{\alpha}(-\mathbf{D}_0)^{-2} \mathbb{1} - \boldsymbol{\alpha}(-\mathbf{D}_0)^{-1} \mathbb{1}}$$

References: [TelH07]

MAP: Spectral Analysis

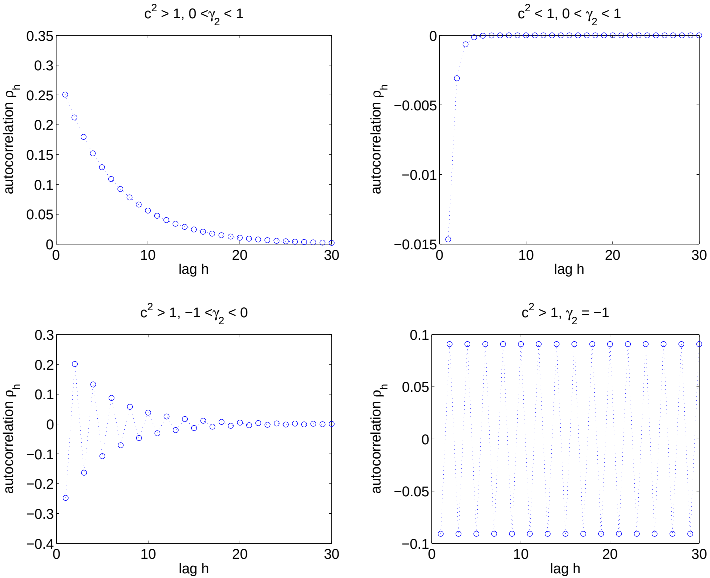
- Characteristic polynomial method applies also to powers $\mathbf{P}^h$
- γ_i: i-th largest eigenvalue of $\mathbf{P}$, algebraic multiplicity r_i
- for $k = 0$ it can be shown that $\rho_0 = \frac{1}{2}\left(1 - \frac{1}{c^2}\right) \neq 1$
- Assume $\mathbf{P}$ diagonalizable, then

$$\rho_k = \sum_{i=2}^{m} A_{t,1}\gamma_i^k, \qquad \sum_{j=2\ldots r_i} A_{t,1} = \rho_0$$

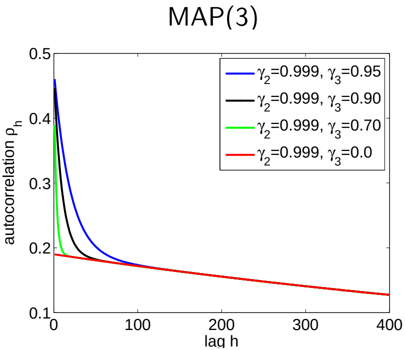
- Assume $n = 2$ states, then $\rho_k = \rho_0 \gamma_2^k$ (geometric decay)
- Degrees of freedom for autocorrelation coefficients: $2n - 3$

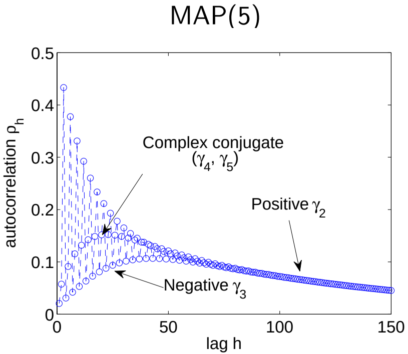
References: [CasMS07]

Example 3.3: MAP(2) Autocorrelation Patterns



Example 3.4: MAP(n) Autocorrelation Patterns





Examples exist also for cases such as:

- $c^2 > 1,\ \rho_1 > 0.50,\ c^2 < 1,\ \rho_1 > 0.30$
- Exponential distribution — $c^2 = 1$, but not Poisson $\rho_k \neq 0$

References: [Nie98]

4. INTER-ARRIVAL PROCESS FITTING

Includes joint work with Evgenia Smirni (IBM Research, William & Mary)

Example 4.1: Redundancy of MAP Representation

- MAP $(\mathbf{D}_0, \mathbf{D}_1)$ defined by $2n^2 - n$ parameters
- Degrees of freedom: difficult problem, typically at most n^2
- Example: redundancy in MAP(2)s

$$\mathbf{D}_0 = \begin{bmatrix} -2 & 1 \\ 5 & -10 \end{bmatrix}, \mathbf{D}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 5 \end{bmatrix}$$

$$\mathbf{D}'_0 = \begin{bmatrix} -1.4174 & 0 \\ 0 & -10.5826 \end{bmatrix}, \mathbf{D}'_1 = \begin{bmatrix} 1.2543 & 0.1632 \\ 5.8368 & 4.7457 \end{bmatrix}$$

	$(\mathbf{D}_0, \mathbf{D}_1)$	$(\mathbf{D}'_0, \mathbf{D}'_1)$
$E[X]$	0.6000	0.6000
$E[X^2]$	0.8267	0.8267
$E[X^3]$	1.7440	1.7440
$\vdots$	$\vdots$	$\vdots$

	$(\mathbf{D}_0, \mathbf{D}_1)$	$(\mathbf{D}'_0, \mathbf{D}'_1)$
ρ_1	0.0381	0.0381
ρ_2	0.0127	0.0127
ρ_3	0.0042	0.0042
$\vdots$	$\vdots$	$\vdots$

References: [TelH07]

MAP(2): 3 Canonical Forms

- PH-Renewal–($\gamma_2 = 0$): $\mathbf{D}_0 = \mathbf{T}$, $\mathbf{D}_1 = -\mathbf{T}\mathbb{1}\boldsymbol{\alpha}$, $\mathbf{T}$ is Coxian
- Positive autocorrelation decay– $\gamma_2 \stackrel{def}{=} pq > 0$:

$$\mathbf{D}_0 = \begin{bmatrix} -\lambda_1 & (1-p)\lambda_1 \\ 0 & -\lambda_2 \end{bmatrix}, \mathbf{D}_1 = \begin{bmatrix} p\lambda_1 & 0 \\ (1-q)\lambda_2 & q\lambda_2 \end{bmatrix}$$

$$\boldsymbol{\alpha} = [(1-q),(q-pq)]/(1-pq),\ p,q \neq 1$$

- Negative autocorrelation decay – $\gamma_2 \stackrel{def}{=} pq < 0$:

$$\mathbf{D}_0 = \begin{bmatrix} -\lambda_1 & (1-p)\lambda_1 \\ 0 & -\lambda_2 \end{bmatrix}, \mathbf{D}_1 = \begin{bmatrix} 0 & p\lambda_1 \\ q\lambda_2 & (1-q)\lambda_2 \end{bmatrix}$$

$$\boldsymbol{\alpha} = [q,(1-q+pq)]/(1+pq),\ q \neq 0$$

References: [BodHGT08],[HeiML06],[HeiHG06]

MAP(2): Exact Fitting

- Symbolic analysis of $\phi(A)$, $A = (-\mathbf{D}_0)^{-1}$, shows that

$$\begin{cases} -(\lambda_1^{-1} + \lambda_2^{-1}) = \frac{E[\widetilde{X}^3] - 3E[\widetilde{X}]E[\widetilde{X}^2]}{3(2E[\widetilde{X}]^2 - E[\widetilde{X}^2])} \\ (\lambda_1 \lambda_2)^{-1} = \frac{\frac{3}{2}E[\widetilde{X}^2]^2 - E[\widetilde{X}]E[\widetilde{X}^3]}{3(2E[\widetilde{X}]^2 - E[\widetilde{X}^2])} \\ (1-p)\lambda_2^{-1} + (1-q)\lambda_1^{-1} = E[\widetilde{X}](1-\gamma_2) \\ pq = \gamma_2 \end{cases}$$

- Generalization of PH(2) moment matching formulas
- Solving for $(\lambda_1, \lambda_2, p, q)$ gives the canonical form
- Symbol solution is feasible and useful for fast evaluation, but yields very complex expressions

Example 4.2: Reducing to MAP(2) Canonical Form

- Same MAPs of Example 4.1:

```
E=[0.6000, 0.8267, 1.7440]; % 2n-1=3 independent moments
g2=1/3; % autocorrelation decay rate
D=3*(2*E(1)^2-E(2)); m1=(E(3)-3*E(1)*E(2))/D, m2=(1.5*E(2)^2-E(1)*E(3))/D,
[x,fval] = fsolve(@(x) [-sum(1./x(1:2)) - m1;1./prod(x(1:2)) - m2; prod(x(3:4))-g2;
((1-x(3)./x(2)) + (1-x(4))./x(1))/(1-g2)-E(1)],rand(1,4));
lam = x(1:2); p=x(3); q=x(4);
Output: lam = [ 10.5826 1.4174], p = 0.4725, q = 0.7055
```

- Canonical form $\gamma_2 > 0$

$$\mathbf{D}_0 = \begin{bmatrix} -10.5826 & 5.5826 \\ 0 & -1.4174 \end{bmatrix}, \ \mathbf{D}_1 = \begin{bmatrix} 5 & 0 \\ 0.4174 & 1 \end{bmatrix}$$

- Moments: $E[X] = 0.6000$, $E[X^2] = 0.8267$, $E[X^3] = 1.7440$
- Autocorrelation decay rate: $\gamma_2 = 0.333$

MAP: Approximate Inter-Arrival Process Fitting

KPC readily generalizes to MAP random variables

- $X = \bigotimes_j^J X_j \sim ((-1)^{J-1} \bigotimes_j^J \mathbf{D}_0^{(j)}, \bigotimes_j^J \mathbf{D}_1^{(j)})$
- X is a MAP if $J-1$ $\mathbf{D}_0^{(j)}$ are diagonal (hyper-exp. moments)

Moments and Joint Moments:

- $E[X^k] = k! \prod_{j=1}^J \frac{E[X_j^k]}{k!}$
- $E[X_i^k X_{i+h}^u] = k!u! \prod_{j=1}^J \frac{E[X_{j,i}^k X_{j,i+h}^u]}{k!u!}$, $\forall$ lags i, h
- $E[X_i^k X_{i+h}^u X_{i+h+l}^v] = k!u!v! \prod_{j=1}^J \frac{E[X_{j,i}^k X_{j,i+h}^u X_{j,i+h+l}^v]}{k!u!v!}$, $\forall$ lags i, h, l
- ...

References: [CasZS07]

MAP: KPC-Toolbox

Autocorrelation decay rates and embedded chain:

- $\gamma_i = \prod_{j=1}^{J} \gamma_{j,i}, \;\; \mathbf{P} = \bigotimes_{j=1}^{J} \mathbf{P}_j$

Second-order properties follow recursively from those of $X \otimes Y$:

$$
\begin{aligned}
1 + c^2 &= \tfrac{1}{2}(1 + c_X^2)(1 + c_Y^2) \\
\rho_k &= \left(\frac{c_X^2}{c^2}\right)\rho_k^X + \left(\frac{c_Y^2}{c^2}\right)\rho_k^Y + \left(\frac{c_X^2 c_Y^2}{c^2}\right)\rho_k^X \rho_k^Y
\end{aligned}
$$

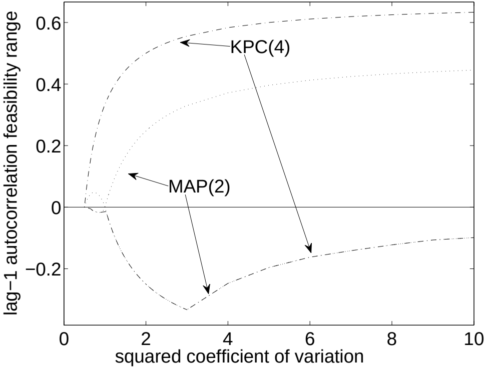
KPC-Toolbox – MAP Fitting:

- Optimization-based second-order and third-order fitting
- $X_j \sim$ MAP(2), known feasibility region for c^2 and ρ_k
- Fitting of c^2 and ρ_k disjoint from first-order and third-order
- Residual df spent to fit third-order moments $E[X_i X_{i+h} X_{i+h+l}]$

References: [CasZS08]

Example 4.3: Feasible ρ_1 Values

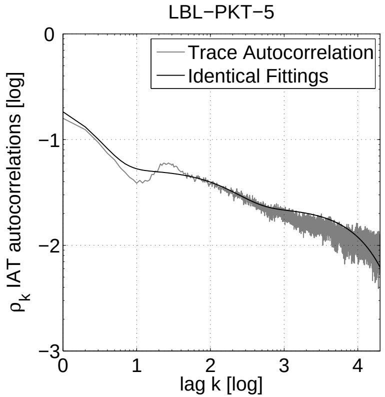
- KPC trades number of states for greater fitting flexibility, e.g.,
- X – MAP(2): $\rho_1 \leq \frac{1}{2}$
- X, Y – MAP(2)s, $X \otimes Y$ – KPC(4): $\rho_1 \leq \frac{2}{3}$

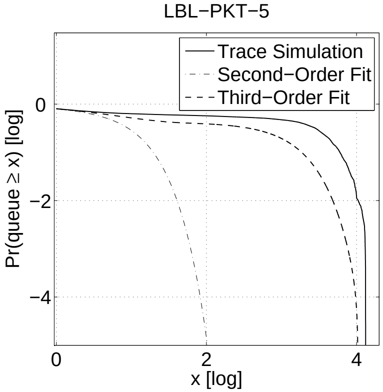


References: [CasZS08]

Example 4.4: Second-Order vs Third-Order Fitting

- Second-Order: approximately match $E[X_i X_{i+h}]$ (equiv. ρ_h)
- Third-Order: approx. match $E[X_i X_{i+h}]$ and $E[X_i X_{i+h} X_{i+h+l}]$



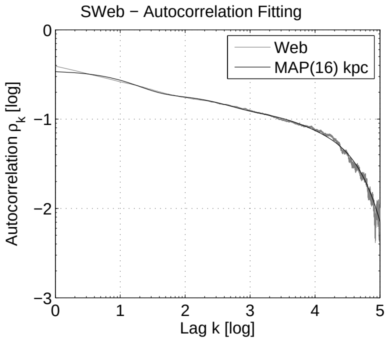


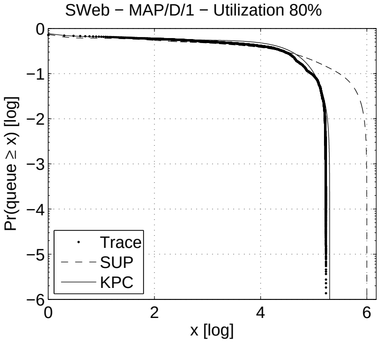
References: [AndN02],[CasZS07]

Example 4.5: KPC MAP Fitting

Comparison of inter-Arrival and counting process methods:

- KPC: second-order and third-order fit of inter-arrival process
- Superposition (cf. Appendix): fit Hurst coefficient for counts





References: [AndN02],[CasZS07]

CONCLUSION

Summary

- PHs and MAPs are tractable models for characterizing empirical data using Markov chains
- PHs and MAPs closed under several operations (mixture, convolution, KPC, ...)
- Models with $n \leq 3$ states are analytically tractable by means of canonical forms
- Larger models can be dealt with using divide-and-conquer approximate moment matching

APPENDIX
COUNTING PROCESS FITTING

MAP: Counting Process Statistics

- $\mathbf{Q} = \mathbf{D}_0 + \mathbf{D}_1$: CTMC for state $J(t)$
- Time-stationary initialization: $\boldsymbol{\pi}$, equilibrium solution of $\mathbf{Q}$
- $\boldsymbol{\pi}_j^c(t)$: state probabilities for level $N(t) = j$
- Kolmogorov forward equations for $\mathbf{Q}^c$:

$$\begin{cases} \dot{\boldsymbol{\pi}}_0^c(t) = \boldsymbol{\pi}_0^c(t)\mathbf{D}_0 \\ \dot{\boldsymbol{\pi}}_j^c(t) = \boldsymbol{\pi}_j^c(t)\mathbf{D}_0 + \boldsymbol{\pi}_{j-1}^c(t)\mathbf{D}_1, \quad j \geq 1 \end{cases}$$

- Solving with $\dot{\boldsymbol{\pi}}_0^c(0) = \boldsymbol{\pi}$, $\dot{\boldsymbol{\pi}}_j^c(0) = \mathbf{0}$, $j \geq 1$ yields

$$E[N(t)] = \lambda t = t/E[X], \qquad \lambda \stackrel{def}{=} \boldsymbol{\pi}\mathbf{D}_1\mathbb{1}, \qquad \mathbf{d}_i \stackrel{def}{=} (\mathbb{1}\boldsymbol{\pi} - \mathbf{Q})^{-i}\mathbf{D}_1\mathbb{1}$$

$$Var[N(t)] = (\lambda - 2\lambda^2 + 2\boldsymbol{\pi}\mathbf{D}_1\mathbf{d}_1)t - 2\boldsymbol{\pi}\mathbf{D}_1(\mathbf{I} - e^{\mathbf{Q}t})\mathbf{d}_2$$

- Covariance of counts in slots of length u spaced by $k-1$ slots:

$$Cov[N(u), N((k+1)u) - N(ku)] = \boldsymbol{\pi}\mathbf{D}_1(\mathbf{I} - e^{\mathbf{Q}u})e^{\mathbf{Q}(k-1)u}(\mathbf{I} - e^{\mathbf{Q}u})\mathbf{d}_2$$

References: [Neu89]

MAP: Process Superposition

MAP is closed under superposition, not true for PH-Renewal

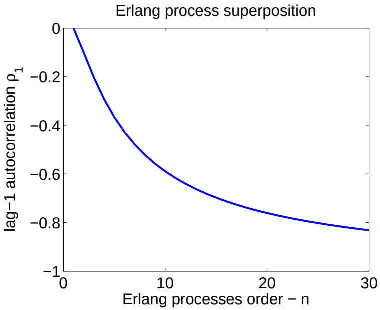
- $N(t)$: counting process for MAP $(\mathbf{D}_0, \mathbf{D}_1)$
- $N'(t)$: counting process for MAP $(\mathbf{D}'_0, \mathbf{D}'_1)$
- $(\mathbf{D}_0 \oplus \mathbf{D}'_0, \mathbf{D}_1 \oplus \mathbf{D}'_1)$ has counting process $N(t) + N'(t)$
- Widely applicable, e.g., superposition of network traffic flows

Example: superposition of Erlang-n flows (PH-Renewal) $\neq$ i.i.d.

- e.g., for $n = 3$

$$\mathbf{D}_0 = \mathbf{D}'_0 = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \end{bmatrix}$$

$$\mathbf{D}_1 = \mathbf{D}'_1 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$



References: [Neu89]

MAP: Approximate Counting Process Fitting

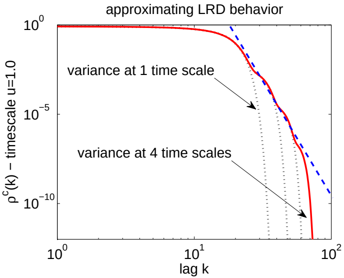
- Autocorrelation of counts in arrivals in slots of length u:

$$\rho^c(k;u) \stackrel{def}{=} Cov[N(u), N((k+1)u) - N(ku)]/Var[N(u)]$$

- Long-Range Dependence (LRD) / variance at multiple time-scales: $\rho^c(k;u) \sim k^{-2(1-H)}$ as $k \to +\infty$ (Hurst coeff.)
- Andersen-Nielsen: Match H by superposing 2-states processes

Example:

- Time scale: $\tau_j = 10^{-j}$ units
- Superposed process: $(\bigoplus_{j=0}^{3} \tau_j \mathbf{D}_0, \bigoplus_{j=0}^{3} \tau_j \mathbf{D}_1)$



References: [AndN98]

REFERENCES

References and Additional Readings 1/2

[AldS87] D. Aldous and L. Shepp. The least variable phase type distribution is Erlang. *Comm. Stat.-Stochastic Models*, 3(3):467–473, 1987.

[AndN98] A. T. Andersen and B. F. Nielsen. A Markovian approach for modeling packet traffic with long-range dependence. *IEEE JSAC*, 16(5):719–732, 1998.

[AndN02] A. T. Andersen and B. F. Nielsen. On the use of second-order descriptors to predict queueing behavior of MAPs. *Naval Res. Log.*, 49(4):391–409, 2002.

[BauBK10] Falko Bause, Peter Buchholz, Jan Kriege. ProFiDo - The Processes Fitting Toolkit Dortmund. QEST 2010: 87-96.

[BuchK09] P. Buchholz, Jan Kriege. A Heuristic Approach for Fitting MAPs to Moments and Joint Moments. In *Proc of QEST*, 2009: 53-62, 2009.

[BodHGT08] L. Bodrog and A. Heindl and G. Horváth and M. Telek. A Markovian Canonical Form of Second-Order Matrix-Exponential Processes. *European Journal of Operations Research*, 190(457-477), 2008.

[Bre78] J. Brewer. Kronecker products and matrix calculus in system theory. *IEEE Trans. on Circuits and Systems*, 25(9):772–781, 1978.

[CasZS07] G. Casale, E. Zhang, and E. Smirni. Characterization of moments and autocorrelation in MAPs. *ACM Perf. Eval. Rev.*, 35(1):27–29, 2007.

[CasZS08] 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.

[CasZS08b] G. Casale, E. Z. Zhang, and E. Smirni. Interarrival times characterization and fitting for Markovian traffic analysis. Technical Report WM-CS-2008-02, College of William and Mary, 2008.

References and Additional Readings 2/2

[HeiHG06] A. Heindl, G. Horvath, and K. Gross. Explicit Inverse Characterizations of Acyclic MAPs of Second Order. EPEW workshop, 108-122, 2006.

[HeiML06] A. Heindl, K. Mitchell, and A. van de Liefvoort. Correlation bounds for second-order MAPs with application to queueing network decomposition. *Performance Evaluation*, 63(6):553–577, 2006.

[HorT09] G. Horvath, M. Telek. On the canonical representation of phase type distributions. *Performance Evaluation*, 66(8):396–409, (2009).

[MaiO92] R. S. Maier and C. A O'Cinneide. A closure characterisation of phase-type distributions. *J. App. Probab.*, 29:92–103, 1992.

[Neu89] M. F. Neuts. *Structured Stochastic Matrices of M/G/1 Type and Their Applications*. Marcel Dekker, New York, 1989.

[Nie98] B. F. Nielsen. Note on the Markovian Arrival Process, Nov 1998. `http://www2.imm.dtu.dk/courses/04441/map.pdf`

[OsoH06] T. Osogami, M. Harchol-Balter. Closed form solutions for mapping general distributions to quasi-minimal PH distributions. *Performance Evaluation*, 63(6): 524-552, 2006.

[TelH03] M. Telek and A. Heindl. Matching moments for acyclic discrete and continuous phase-type distributions of second order. International Journal of Simulation, 3:47–57, 2003.

[TelH07] M. Telek and G. Horváth. A minimal representation of Markov arrival processes and a moments matching method. *Performance Evaluation*, 64(9-12):1153-1168, 2007.