Explicit identification of the class of order 3 matrix exponential distributions

YEQT V Workshop

István Kolossváry and Miklós Telek

Technical University of Budapest, Hungary

24 October, 2011, at Eurandom, Eindhoven, Netherlands.

Motivation

- Analysis of queuing systems with generally distributed inter-arrival and/or service time.
- Problem: absence of efficient methods for analysis, only empirical data, computational complexity too large.
- $\Rightarrow$ use of simpler models, different distributions can be used for approximation.

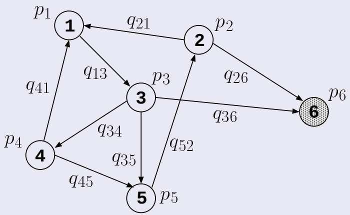
Outline

- Define phase type (PH) and matrix exponential (ME) distributions, quasi-birth-death processes.
- Introduce new approach to identify ME distributions.
- Explicit bounds for ME(3) membership, explore higher orders.
- Greater fitting accuracy showed on a particular example.

Phase type (PH) distributions I.

Stochastic interpretation

Time to absorbtion in a Markov chain with *n* transient and an absorbing state.



Cumulative distribution function

$$F(t) = \mathbf{P}\left(X < t\right) = 1 - \alpha e^{\mathbf{A}t}\mathbb{1}, \quad t \geq 0,$$

where α is the *initial stochastic vector* $\mathbf{A}$ is the *transient* *$(n \times n)$ Markovian generator* and $\mathbb{1}$ is the *closing vector*.

PH distributions II.

Probability density function (PDF)

$$f(t) = \alpha e^{\mathbf{A}t}(-\mathbf{A})\mathbb{1} \quad (1)$$

This is always a valid PDF thanks to the stochastic interpretation.

Different representations

- Vector-matrix representation Defined by the pair $(\alpha, \mathbf{A})$. Problems: similarity transformation, too many parameters.
- PDF representation Defined by (1). Minimal & unique.
- Moments representation The first 2n moments define a non-redundant PH(n) distribution. Minimal & unique.
- Laplace representation The Laplace transform $f^*(s) = \mathbf{E}\left(e^{-sX}\right)$ is an order n rational function. Properly normalized Laplace representation minimal & unique.

Matrix exponential (ME) distributions

The vector matrix pair $(\alpha, \mathbf{A})$ of size $(1 \times n)$ and $(n \times n)$ defines a matrix exponential (ME) distribution iff

$$F(t) = \mathbf{P}(X < t) = 1 - \alpha e^{\mathbf{A}t}\mathbb{1} \quad , t \geq 0$$

is a valid cumulative distribution function, i.e., $F(0) \geq 0$, $\lim_{t\to\infty} F(t) = 1$ and $F(t)$ is monotone increasing.

Advantages

- $ME(n) \supset PH(n)$,
- similar efficient numerical approaches are applicable (matrix geometric methods),

Disadvantages

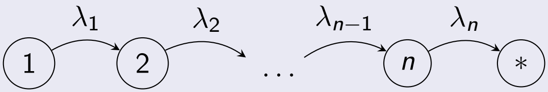
- no simple stochastic interpretation,
- negative numbers in numerical procedures.

Special class: acyclic PH distributions

The absorption time of a CTMC, where any given transient state is visited at most once (hence the name acyclic).

- Generator **A** of the transient states is a triangular matrix $\Rightarrow$ all the eigenvalues of **A** are real $\Rightarrow$ APH $\subset$ PH
- Efficient methods for analysis, which differ from the ones used to analyze ME or PH distributions.

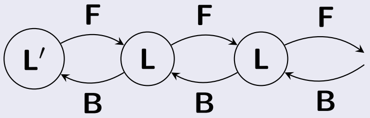
The Markov chain in canonical form



$\lambda_n \geq \ldots \geq \lambda_1 > 0$ are the intensities

Quasi-birth-death (QBD) processes I.

Only one demand can arrive or leave the system at a time. Consider the CTMC:



F: phase transition with arrival regulated by background CTMC
B: phase transition with service regulated by another CTMC
L: local behavior without arrival or service

Special case: birth-death processes

Inter-arrival time $\sim \text{Exp}(\lambda)$, service time $\sim \text{Exp}(\mu)$ $\Rightarrow$ $\mathbf{F} = \lambda$, $\mathbf{B} = \mu$, $\mathbf{L} = -\lambda - \mu$ and there is only a 1 state background MC involved

QBD processes II.

Generator matrix

$$
\mathbf{Q} = \begin{matrix} & 0 & 1 & 2 & 3 & \dots \\ 0 & \mathbf{L}' & \mathbf{F} & & & \\ 1 & \mathbf{B} & \mathbf{L} & \mathbf{F} & & \\ 2 & & \mathbf{B} & \mathbf{L} & \mathbf{F} & \\ \vdots & & & \ddots & \ddots & \ddots \end{matrix}
$$

Inter-arrival time $\sim ME(\alpha, \mathbf{A})$ and service times $\sim ME(\tilde{\alpha}, \tilde{\mathbf{A}})$

$$
\mathbf{L} = \mathbf{A} \oplus \tilde{\mathbf{A}}, \qquad \mathbf{L}' = \mathbf{A} \otimes \mathbf{I},
$$

$$
\mathbf{F} = (-\mathbf{A}\mathbb{1})\alpha \otimes \mathbf{I}, \qquad \mathbf{B} = \mathbf{I} \otimes (-\tilde{\mathbf{A}}\mathbb{1})\tilde{\alpha},
$$

Available: Matrix geometric methods for analysis!

Previous results I.

- $ME(2) \equiv PH(2)$
- $PH(n)$ is a real subset of $ME(n)$ for $n \geq 3$
- minimal coefficient of variation
  $\min CV_{PH(3)} = 0.3333 = \frac{1}{3}, \min CV_{ME(3)} = 0.2009 \approx \frac{1}{5}$
- characterization of ME(3) in Laplace transform domain - Mark Fackrell
  the transform of the density function is

$$f^*(s) = \frac{x_2 s^2 + x_1 s + b_1}{s^3 + b_3 s^2 + b_2 s + b_1}$$

for a given $\{b_1, b_2, b_3\}$ a parametric equations of x_1 and x_2 define the border of the ME(3) class.

The parametric equations involve transcendental equations $\Rightarrow$ there is no closed form solution.

Previous results II. / Time domain analysis

Eigenvalue structure of **A**

- case A: distinct real eigenvalues: $\lambda_3 < \lambda_2 < \lambda_1 < 0$

$$f(t) = a_1 e^{\lambda_1 t} + a_2 e^{\lambda_2 t} + a_3 e^{\lambda_3 t},$$

- case B: 2 identical real eigenvalues:
  $\lambda_3 = \lambda_2 < \lambda_1 < 0$ or $\lambda_3 < \lambda_2 = \lambda_1$
- case C: 3 identical real eigenvalues: $\lambda_3 = \lambda_2 = \lambda_1 \in \mathbb{R}^-$,

$$f(t) = (a_1 + a_2 t + a_3 t^2) e^{\lambda_1 t},$$

- case D: complex eigenvalues: $\lambda_1 \in \mathbb{R}^-$, $\lambda_2 = \overline{\lambda}_3 \in \mathbb{C}^-$.

Horvath at al. Closed form expressions: Case A, Case C;
Numerical method: Case B, Case D.

General principle of the method

Goal: efficient method to decide non-negativity of ME functions

Assume all eigenvalues are distinct $\rightsquigarrow$ PDF: $f(t) = \sum_{i=1}^{n} a_i e^{\lambda_i t}$

Order reduction step

$$f(t) \geq 0 \quad \rightsquigarrow \quad \tilde{f}(t) = \sum_{i=1}^{n-1} a_i e^{(\lambda_i - \lambda_n)t} \geq -a_n.$$

where the transformed problem has a lower order.

This way we need to study the following problem:

$$\tilde{f}(t) \geq b, \quad \forall t \geq 0, \tag{2}$$

which is based on the solution of

$$\tilde{f}(t) = 0,$$

$$\tilde{f}(t) = b.$$

Case B: 2 different real eigenvalues I.

$$f(t) = a_1 e^{\lambda_1 t} + (a_2 + a_{21} t) e^{\lambda_2 t},$$

where $a_1, a_{21} \neq 0$.
Dividing the density function by $e^{\lambda_1 t}$ gives the following problem of type (2):

$$\hat{f}(t) = (g_1 + g_2 t) e^{\gamma t} \geq b \;\; \forall t \geq 0,$$

where $b, g_2 \neq 0$.
For which we have following closed form expressions

$$t^* = \frac{-g_1}{g_2}, \;\text{ where } \hat{f}(t^*) = 0,$$

$$t_{opt} = -\frac{g_2 + g_1\gamma}{g_2\gamma} = \frac{-1}{\gamma} + t^*, \;\text{ where } \frac{d\hat{f}(t)}{dt}\Big|_{t=t_{opt}} = 0,$$

$$f_{opt} = \hat{f}(t_{opt}) = -\frac{g_2}{\gamma} \cdot \exp\left(-\frac{1}{\gamma} - \frac{g_1}{g_2}\right).$$

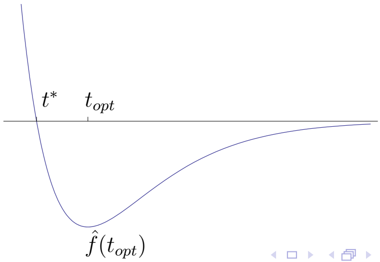
Case B II.

Depending on the sign of γ and g_2 there are four cases to consider.

- $\gamma < 0,\ g_2 < 0.$
  The possible values of b depend on the sign of t_{opt}.

$$\gamma < 0,\ g_2 < 0,\ b \le g_1 \le \frac{-g_2}{\gamma} \quad \text{if } t_{opt} \le 0,$$

$$\gamma, g_2 < 0,\ g_1 > \frac{-g_2}{\gamma},\ b \le -\frac{g_2}{\gamma} \cdot e^{-\frac{1}{\gamma} - \frac{g_1}{g_2}} \quad \text{if } t_{opt} > 0.$$



Case B III.

There are explicit limits also in the other cases.

- $\gamma < 0,\ g_2 > 0.$

$$\gamma < 0,\ b \le g_1 < 0 < g_2 \quad \text{if } t^* > 0,$$
$$\gamma < 0,\ g_2 > 0,\ g_1 \ge 0,\ b \le 0 \quad \text{if } t^* \le 0.$$

- $\gamma > 0,\ g_2 < 0.$
  $\hat{f}(t) \ge b$ can't hold for any b since $\lim_{t\to\infty} \hat{f}(t) = -\infty.$
- $\gamma > 0,\ g_2 > 0.$

$$\gamma > 0,\ g_2 > 0,\ g_1 \ge \frac{-g_2}{\gamma},\ b \le g_1 \quad \text{if } t_{opt} \le 0,$$
$$\gamma, g_2 > 0,\ g_1 < \frac{-g_2}{\gamma},\ b \le -\frac{g_2}{\gamma} \cdot e^{-\frac{1}{\gamma} - \frac{g_1}{g_2}} \quad \text{if } t_{opt} > 0.$$

Results for Case B

Theorem $\left(f_1(t) = a_1 e^{\lambda_1 t} + (a_2 + a_{21} t) e^{\lambda_2 t},\ \lambda_2 < \lambda_1 < 0\right)$

$f_1(t) \geq 0\ (\forall t \geq 0) \quad \Leftrightarrow \quad$ *one of the following holds:*

i) $a_{21} < 0,\ -a_1 \leq a_2 \leq \frac{-a_{21}}{\lambda_2 - \lambda_1}$

ii) $a_{21} < 0,\ a_2 > \frac{-a_{21}}{\lambda_2 - \lambda_1},\ -a_1 \leq -\frac{-a_{21}}{\lambda_2 - \lambda_1} \exp\left(-\frac{1}{\lambda_2 - \lambda_1} - \frac{a_2}{a_{21}}\right)$

iii) $-a_1 \leq a_2 < 0 < a_{21}$

iv) $a_{21} > 0,\ -a_1 < 0,\ a_2 \geq 0$

Theorem $\left(f_2(t) = (a_1 + a_{11} t) e^{\lambda_1 t} + a_2 e^{\lambda_2 t},\ \lambda_2 < \lambda_1 < 0\right)$

$f_2(t) \geq 0\ (\forall t \geq 0) \quad \Leftrightarrow \quad$ *on of the following holds:*

i) $a_{11} > 0,\ a_1 \geq \frac{-a_{11}}{\lambda_1 - \lambda_2},\ -a_2 \leq a_1$

ii) $a_{11} > 0,\ a_1 < \frac{-a_{11}}{\lambda_1 - \lambda_2},\ a_2 \geq \frac{-a_{11}}{\lambda_1 - \lambda_2} \exp\left(-\frac{1}{\lambda_1 - \lambda_2} - \frac{a_1}{a_{11}}\right)$

Case D: complex eigenvalues I.

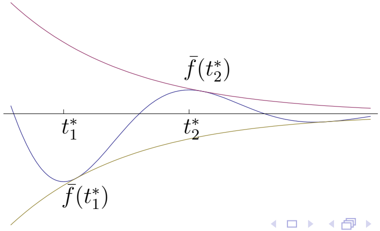
Order reduction step

$$f(t) \;=\; a_1\, e^{\lambda_1 t} + a_2 \cos(\omega t + \phi)\, e^{\lambda_c t} \geq 0$$

$$\Leftrightarrow$$

$$\bar{f}(t) = \cos(\omega t + \phi) e^{\lambda t} \geq b$$

where $\lambda = \lambda_2 - \lambda_1$, $b = \frac{-a_1}{a_2}$ and $a_1 > 0$.



Case D II., solving the inequality

The extreme points of $\bar{f}(t)$ are obtained at $\bar{f}(t)' = 0$ which are

$$t_k^* = \frac{\tan^{-1}\left(\frac{\lambda}{\omega}\right) - \phi + k\pi}{\omega},$$

where $k \in \mathbb{Z}$.

It is enough to consider the t_k^* which fall into $[0, 2\pi/\omega]$

which are explicitly given by

$$t_i^* = \left(\tan^{-1}\left(\frac{\lambda}{\omega}\right) - \phi + (k^* + i - 1)\pi\right)/\omega \text{ for } i = 1, 2.$$

It only remans to check if $\bar{f}(t_i^*) \geq b$, $i = 1, 2$.

Applying the method to higher order ME functions

- Can the order reduction approach give results in higher dimensions?
  Yes, but mainly No
- Show an explicit case: four real eigenvalues, one of which has multiplicity of three.
- Another case where we stumble into a transcendental equation that can't be explicitly solved.
- What are the other cases where we can expect to obtain explicit conditions?

An explicit case ME(4) distribution

Eigenvalue structure of **A**

Assume **A** has two different real eigenvalues $\lambda_2 < \lambda_1 < 0$, one of which has a multiplicity of three.

ME functions

There are two different cases to consider:

- The multiplicity of λ_1 is one

$$f_1(t) = a_1 e^{\lambda_1 t} + (a_{20} + a_{21} t + a_{22} t^2) e^{\lambda_2 t}, \text{ where } a_1, a_{22} \neq 0.$$

- The multiplicity of the dominant eigenvalue is three

$$f_2(t) = (a_{10} + a_{11} t + a_{12} t^2) e^{\lambda_1 t} + a_2 e^{\lambda_2 t}, \text{ where } a_{12}, a_2 \neq 0.$$

We deal with $f_1(t)$. $f_2(t)$ can be dealt with the same way.

Result of order reduction

Divide with the exponential term of the eigenvalue with single multiplicity $\Rightarrow$

$$\hat{f}(t) := (at^2 + bt + c)e^{\lambda t} \geq d \ \forall t \geq 0, \text{ where } a, d \neq 0.$$

In our case (a, b, c, d, λ) equal $(a_{22}, a_{21}, a_{20}, -a_1, \lambda_2 - \lambda_1)$.

Roots $t^*_{1,2}$ and extreme points $t'_{1,2}$ of $\hat{f}(t)$

$$t^*_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}, \text{ and} \quad (3)$$

$$t'_{1,2} = \frac{-\tilde{b} \pm \sqrt{\tilde{b}^2 - 4\tilde{a}\tilde{c}}}{2\tilde{a}}, \text{ where } \begin{cases} \tilde{a} &= \lambda a \\ \tilde{b} &= \lambda b + 2a \\ \tilde{c} &= \lambda c + b \end{cases} \quad (4)$$

Solving the inequality $\hat{f}(t) \geq d$

Lemma

For $\lambda < 0$ and $a, d \neq 0$, the inequality

$$\hat{f}(t) = (at^2 + bt + c)e^{\lambda t} \geq d,$$

holds for every $t \geq 0$ if and only if one of the following holds

i) $a < 0,\ t_2' < 0,\ d \leq c$

ii) $a < 0,\ t_2' \geq 0,\ d \leq \min\left(c, \hat{f}(t_2')\right)$

iii) $a > 0,\ b^2 \leq 4ac,\ d < 0$

iv) $a > 0,\ b^2 > 4ac,\ t_2^* \leq 0,\ d < 0$

v) $a > 0,\ b^2 > 4ac,\ t_1' < 0 < t_2^*,\ d \leq c$

vi) $a > 0,\ b^2 > 4ac,\ t_1' \geq 0,\ d \leq \hat{f}(t_1')$

where t_2^*, t_1' and t_2' are defined in (3) and (4).

Result for $f_1(t)$

Theorem

To obtain the desired explicit result for the non-negativity of

$$f_1(t) = a_1 e^{\lambda_1 t} + (a_{20} + a_{21} t + a_{22} t^2) e^{\lambda_2 t}, \text{ where } a_1, a_{22} \neq 0,$$

combine results of previous Lemma with (3) *and* (4) *for parameters* (a, b, c, d, λ) *equal to* $(a_{22}, a_{21}, a_{20}, -a_1, \lambda_2 - \lambda_1)$.

When to expect explicit conditions?

- The multiplicity of the dominant eigenvalue is three. ✓
- ME(4) case $(a_0 + a_1 t + a_2 t^2 + a_3 t^3) e^{\lambda t}$, determine the roots of cubic functions ✓
- ME(5) two different real eigenvalues, one with multiplicity of four ✓

Two real eigenvalues and a complex conjugate pair

Most likely eigenvalue structure for a random matrix (we can rule out two complex conjugate pairs).

The general form of the ME function is

$$f(t) = a_1 e^{\lambda_1 t} + a_2 e^{\lambda_2 t} + a_c \cos(\omega t + \phi) e^{\lambda_c t},$$

where $t \geq 0$, $a_c > 0$, $-\pi < \phi < \pi$ and $\lambda_2 < \lambda_1 < 0$ denote the real eigenvalues and $\lambda_c < 0$ denotes the real part of the complex eigenvalue.

Identifying the transcendental equation

Dividing by $a_c e^{\lambda_c t}$ we get

$$\hat{f}(t) := \frac{a_1}{a_c} e^{(\lambda_1 - \lambda_c)t} + \frac{a_2}{a_c} e^{(\lambda_2 - \lambda_c)t} \geq -\cos(\omega t + \phi).$$

Non-trivial necessary/sufficient conditions

We can still give non-trivial necessary/sufficient conditions.

Bounding from below and above

$$f_-(t) := a_1 e^{\lambda_1 t} + a_2 e^{\lambda_2 t} - a_c e^{\lambda_c t} \leq f(t)$$
$$f_+(t) := a_1 e^{\lambda_1 t} + a_2 e^{\lambda_2 t} + a_c e^{\lambda_c t} \geq f(t)$$

Sufficient condition: $f_-(t) \geq 0 \;\Rightarrow\; f(t) \geq 0$.
Necessary condition: $\exists t \geq 0 : f_+(t) \leq 0 \;\Rightarrow\; f(t) \ngeq 0$.

Comparing ME and APH

- Metric: relative entropy

$$D(\hat{f}, f) := \int_{t=0}^{\infty} f(t) \log\left(\frac{f(t)}{\hat{f}(t)}\right) \mathrm{d}t \geq 0,$$

where $\hat{f}$: PDF of the original distribution, while f is the PDF of the approximating distribution.

- Optimization in
  - ME: moment matching based optimization + check if result is a valid ME distribution with the new theorems
  - APH: use *PhFit*, where fitting goes by relative entropy among APH(3) distributions
- Comparison
  - Calculate relative entropy
  - Plot PDf-s

Weibull distributions in general

- Importance in extreme value theory, relation to other probability distributions
- Two parameters

  shape parameter $k > 0$ and scale parameter $\lambda > 0$.

- PDF

$$\hat{f}_{k,\lambda}(x) = \begin{cases} \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k}, & \text{if } x \geq 0 \\ 0, & \text{if } x < 0. \end{cases}$$

- Moments

$$\mu_n = \mathbf{E}\left(X^n\right) = \lambda^n \Gamma\left(1 + \frac{n}{k}\right),$$

where $\Gamma(t) = \int_0^\infty x^{t-1} e^{-x} \mathrm{d}x$.

Picking parameters k and λ

- Coefficient of variation/Relative variance

$$\frac{\mathbf{D}^2(X)}{\mathbf{E}^2 X} = \frac{\left(\Gamma(1+2/k) - \Gamma^2(1+1/k)\right)}{\Gamma^2(1+1/k)}$$

  does not depend on the scale parameter λ.
- Choose parameters so that coefficient of variation is small in one case and large in the other.
-

	k	λ	coefficient of variation
Case 1	1.55	50	0.434
Case 2	1.02	40	0.961

Comparison of relative entropy I.

ME(3) approximation

	$\mathbf{A}$	$D(\hat{f}, f_{\alpha,\mathbf{A}})$
Case 1	$\begin{pmatrix} -0.0539585 & 0.373227 & -0.286393 \\ -0.0082287 & -0.038395 & 0.025353 \\ -0.0065961 & 0.023570 & -0.040178 \end{pmatrix}$	0.0013129
Case 2	$\begin{pmatrix} -0.0842461 & 0.286888 & -0.223354 \\ -0.0043999 & -0.051695 & 0.030653 \\ -0.0124034 & 0.034892 & -0.047457 \end{pmatrix}$	1.307×10^{-5}

- In Case 1 $\mathbf{A}$ has a dominant real and a complex conjugate pair of eigenvalues, while in Case 2 all the eigenvalues are negative real numbers.
- In each case $\alpha = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$.

Comparison of relative entropy II.

APH(3) approximation

	α^T	Intensities	$D(\hat{f}, f_{\alpha,\mathbf{A}})$
Case 1	$\begin{pmatrix} 0.73673 \\ 0.21406 \\ 0.04921 \end{pmatrix}$	$\begin{pmatrix} -0.07766 & 0 & 0 \\ 0.05506 & -0.05506 & 0 \\ 0 & 0.05368 & -0.05368 \end{pmatrix}$	0.005178
Case 2	$\begin{pmatrix} 0.17460 \\ 0.18490 \\ 0.64049 \end{pmatrix}$	$\begin{pmatrix} -0.14452 & 0 & 0 \\ 0.09706 & -0.09706 & 0 \\ 0 & 0.02852 & -0.02852 \end{pmatrix}$	0.002025

- Greater coefficient of variation $\Rightarrow$ better approximation.
- (Significantly) Better results with ME(3) approximation!

Comparison of PDF-s I.

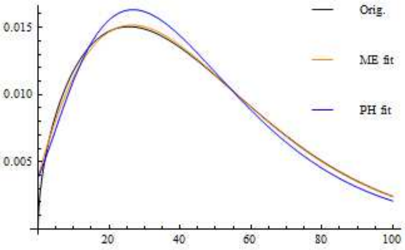


Figure: $k_1 = 1.55$, $\lambda_1 = 50$

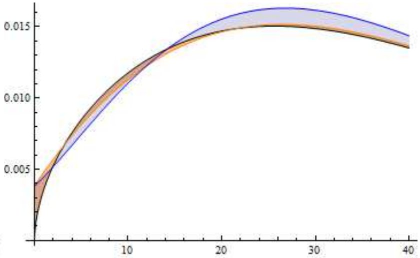


Figure: Inset of Fig. 1

Black: original
Orange: ME(3) approximation
Blue: APH(3) approximation

Comparison of PDF-s II.

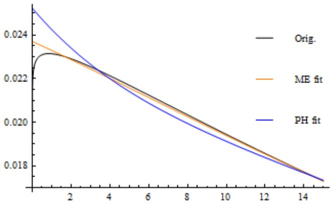


Figure: $k_2 = 1.02$, $\lambda_2 = 40$

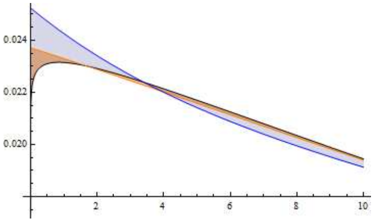


Figure: Inset of Fig. 3

- It is indeed worth using ME distributions for approximating general distributions.
- Unfortunately the moment matching algorithm very rarely gave a valid matrix exponential function as its result.

Conclusions, future plans

- Introduced new order reduction approach to check the non-negativity of ME functions.
- As a result, obtained explicit conditions for ME(3) membership in all different cases.
- Studied order 4 ME functions: showed an explicit case, otherwise stumble into transcendental equations.
- The example showed that greater fitting accuracy can be achieved with ME(3) than with APH(3).

- Finding non-trivial necessary or sufficient conditions for higher order ME membership.
- Implementation using numerical subroutines.
- Develop a moment matching based optimization algorithm that gives valid ME results in all cases.