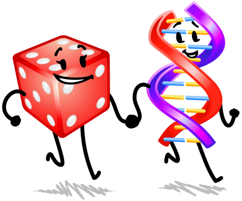
Matrix Exponential Distributions and their Generalisations

Nigel Bean

School of Mathematical Sciences
University of Adelaide

SMMP, UTas, 11/2015



Stochastic Modelling

- Taking a stochastic process and using it to answer questions in the real-world.
- What do we need for a distribution to be useful in stochastic modelling?
  - Efficient techniques for fitting the distribution to collected data.
  - Efficient analysis (numerical, simulation,..)
- Mark Fackrell, Peter Taylor and Bo Friis Nielsen.

Exponential Distributions

- The Exponential Distribution of rate λ is defined by

$$P(X \leq x) = 1 - e^{-\lambda x}.$$

- Too restrictive for many practical purposes.
- In a Markov process the time between events is given by the Exponential Distribution.
- Need greater flexibility, but don't want to lose tractability!

Absorbing Finite-state Markov Processes

- $m + 1$ states: state 0 is absorbing
- $m \times m$ transition matrix T governs transitions between non-absorbing states
- $Q = \begin{pmatrix} 0 & 0 \\ -T\boldsymbol{e} & T \end{pmatrix}$
- $[-T\boldsymbol{e}]_i$ is the rate of moving from state i to the absorbing state 0

Phase-Type Distributions

- The distribution of time to absorption for an absorbing finite-state Markov process.
- Described by a matrix T and a vector α detailing the probability of starting in each of the states.
- The phase-type distribution with representation (α, T) is given by

$$P(X \leq x) = 1 - \alpha e^{Tx} \boldsymbol{e}$$

  with density $f(x) = \alpha e^{Tx}(-T\boldsymbol{e})$
- Obvious multi-state analogue of the exponential distribution.

Phase-Type Distributions: Properties

- Dense in all distributions on the nonnegative real numbers.
- Highly flexible.
- Highly tractable: keep the Markov structure.
- Easy to take a physical description and determine if it represents a Phase-Type distribution.
- Hard to determine a representation from a distribution function or data.

Phase-Type Distributions: Fitting

- Many representations for the one distribution.
- Different orders for the one distribution.
- What is the minimal order? Can be really high.
- Objective function is not uni-modal!
- Problem of determining a representation from a distribution.
- Summary: fitting phase-type distributions to data is difficult to do well.

Phase-Type Distributions: Analysis

- Remains a Markov chain
- Increased size of state space
- Sometimes efficient algorithms can exploit the structure, e.g. QBD, SFM

Matrix Exponential Distributions: What are they

- The matrix-exponential (ME) distribution with representation $(\alpha, T, \boldsymbol{s})$ is given by

$$P(X \leq x) = 1 + \alpha e^{Tx} T^{-1} \boldsymbol{s},$$

  with density $f(x) = \alpha e^{Tx} \boldsymbol{s}$.
- If $\boldsymbol{s} = -T\boldsymbol{e}$ then reduces to the algebraic form of the Phase-type expressions.
- Phase-type distributions are a proper subset of matrix exponential distributions.
- Equivalent to Cox's class of distributions with rational Laplace transform.

ME Distributions - Restrictions

- What restrictions do we impose on α, T and $\boldsymbol{s}$?

Answer: $f(x) \geq 0, \quad \text{for all } \; x \geq 0.$

- Don't need α to represent a probability distribution.
- Don't need T to represent an absorbing Markov process.
- Don't need $\boldsymbol{s} = -T\boldsymbol{e}$.
- Generalization of phase-type distributions away from the natural physical description.

ME Distributions – Example

- order 3
- $\alpha = (3, -1, -1)$.
- $T = \begin{pmatrix} -1 & 0 & 0 \\ -2/3 & -1 & 1 \\ 2/3 & -1 & -1 \end{pmatrix}$.
- $\boldsymbol{s} = \begin{pmatrix} 1 \\ 2/3 \\ 4/3 \end{pmatrix}$.

ME Distributions: Properties

- Many representations of same/different orders exist for the one distribution.
- Laplace transform of any matrix exponential distribution is a rational function.
- Class of distributions with rational Laplace transforms is identical to the class of matrix exponential distributions.
- Canonical minimal-order representation is trivially deduced from the transform.
- The minimal order is often much lower than for the phase-type representation.

ME Distributions – Physical Interpretation

- A matrix exponential distribution can be thought of as being generated by a piecewise deterministic Markov process $\{\boldsymbol{A}(t)\}_{t\geq 0}$ on a compact subset of $\{\boldsymbol{a} \in \mathbb{R}^m : \boldsymbol{ae} = 1\}$, such that the lifetime is identical to the time of the first jump in $\boldsymbol{A}(\cdot)$.
- $\boldsymbol{A}(\cdot)$ evolves before a jump according to:

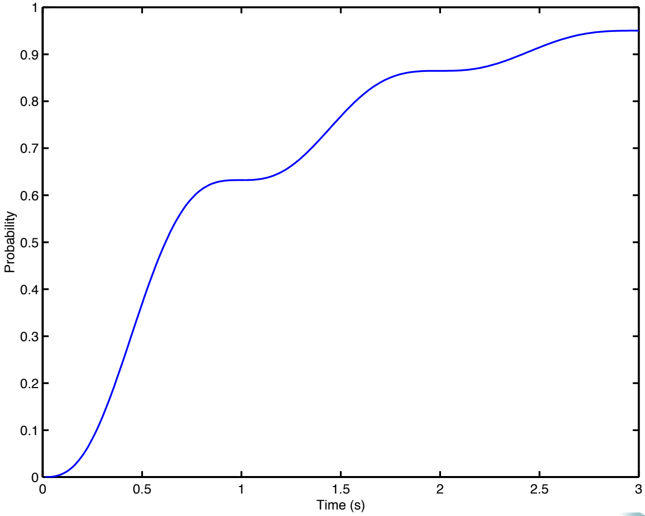
$$\frac{\mathrm{d}\boldsymbol{a}(t)}{\mathrm{d}t} = \boldsymbol{a}(t)T - \boldsymbol{a}(t)T\boldsymbol{e} \cdot \boldsymbol{a}(t)$$

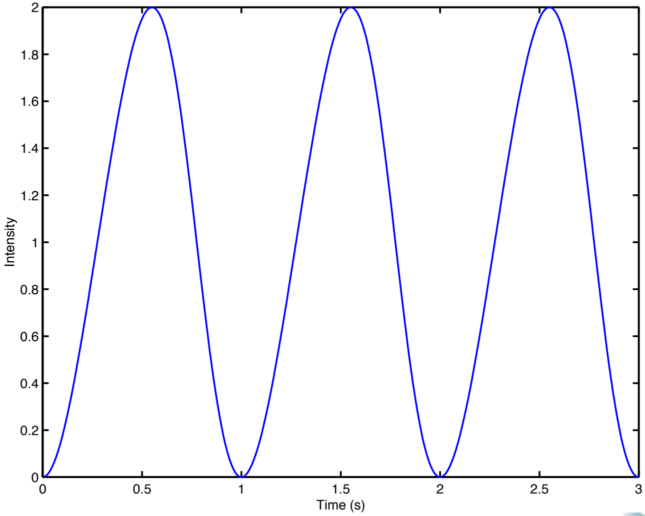
- In state $\boldsymbol{a}$, a jump occurs with stochastic intensity $\boldsymbol{as}$.
- After a jump the deterministic Markov process $\{\boldsymbol{A}(t)\}_{t\geq 0}$ expires.

ME Distributions – Example

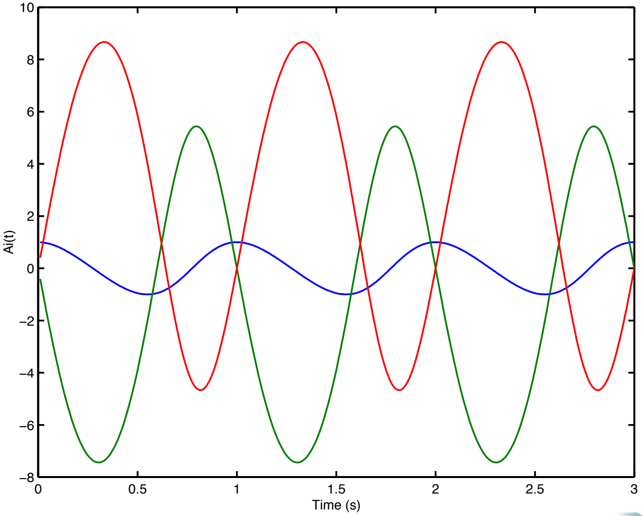
- $\alpha = (1, 0, 0)$.
- $T = \begin{pmatrix} 0 & -1-4\pi^2 & 1+4\pi^2 \\ 3 & 2 & -6 \\ 2 & 2 & -5 \end{pmatrix}$ and $\boldsymbol{s} = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$.
- Canonical minimal-order representation: $\alpha = (1+4\pi^2, 0, 0)$,

  $T = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1-4\pi^2 & -3-4\pi^2 & -3 \end{pmatrix}$ and $\boldsymbol{s} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$.


Probability
Time (s)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0.5
1.5
2
2.5
3



Intensity
Time (s)
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2
0
0.5
1
1.5
2
2.5
3


Ai(t)
Time (s)
10
8
6
4
2
0
−2
−4
−6
−8
0
0.5
1
1.5
2
2.5
3

ME Distributions: Analysis

- No longer a discrete-space Markov chain
- Piecewise Deterministic Markov Process
- Sometimes we can exploit the structure, e.g. QBD - see later

Poisson Process of rate λ

- Renewal process with exponentially distributed inter-arrival time
- Counting process has finite-dimensional density:

$$f(x_1, x_2, \ldots, x_n) = \lambda e^{-\lambda x_1} \lambda e^{-\lambda x_2} \ldots \lambda e^{-\lambda x_n}$$

- Easily incorporated into models of queueing systems through $M/M/\cdot$ queues, or Birth-and-Death processes

Phase-type Renewal process — (α, T)

- Renewal process with phase-type distributed inter-arrival time, with representation (α, T)
- Counting process has finite-dimensional density:

$$f(x_1, x_2, \ldots, x_n) = \alpha e^{Tx_1}(-T\boldsymbol{e})\alpha e^{Tx_2}(-T\boldsymbol{e})\ldots\alpha e^{Tx_n}(-T\boldsymbol{e})$$

- Easily incorporated into models of queueing systems in the form of Quasi-Birth-and-Death processes (QBDs)

Markovian Arrival Process (MAP) — (C, D)

- Non-renewal process, but with (conditional) phase-type distributed inter-arrival times
- Counting process has finite-dimensional density:

$$f(x_1, x_2, \ldots, x_n) = \alpha e^{Cx_1} D e^{Tx_2} D \ldots e^{Tx_n} D\boldsymbol{e}$$

- If $D = (-T\boldsymbol{e})\alpha$, then a phase-type renewal process
- Easily incorporated into models of queueing systems in the form of Quasi-Birth-and-Death processes (QBDs)

Quasi-Birth-and-Death Process (QBD)

- 2-dimemsional Markov chain
  - Phase lies on the state space $\{1, 2, \ldots, m\}$ – records the state of the MAP/PH distribution/environment
  - Level lies on the state space $\mathbb{Z}_+$ but restricted to changes in $\{-1, 0, 1\}$ – records the number in the queue, for example.
- has Markovian representation:

$$Q = \begin{bmatrix} B_0 & A_1 & 0 & 0 & 0 & \cdots \\ A_{-1} & A_0 & A_1 & 0 & 0 & \ddots \\ 0 & A_{-1} & A_0 & A_1 & 0 & \ddots \\ 0 & 0 & A_{-1} & A_0 & A_1 & \ddots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots \end{bmatrix}$$

Quasi-Birth-and-Death Process (QBD) – Analysis

- Equilibrium distribution given by

$$\pi = (\pi_0, \pi_1, \pi_2, \ldots)$$

where

$$\pi_n = \pi_{n-1} R$$

$$\pi_0 (B_0 + RA_{-1}) = \mathbf{0}$$

and R is the minimal nonnegative solution to the matrix-quadratic equation

$$A_1 + RA_0 + R^2 A_{-1} = 0$$

- Efficient (and stable) algorithms to solve this matrix quadratic equation

Rational Arrival Process (RAP) - fundamentals

- Class of all point processes on a finite state space is the class of MAPs
- Class of all point processes on a finite dimensional space is the class of RAPs

Asmussen and Bladt (1999):

Definition 1.1. We call a point process N a rational arrival process (RAP) if $\mathbb{P}(N(0,\infty)=\infty)=1$ and there exists a finite-dimensional subspace V of $\mathscr{M}(\mathscr{N})$ such that for any t, $\mathbb{P}(\theta_t N \in \cdot \,|\, \mathscr{F}_t)$ has a version $\mu(t,\cdot)$ with $\mu(t,\omega) \in V$ for all $\omega \in \Omega$.

- $\mu(t,\omega) = \boldsymbol{A}(t)W$, where the rows of W are the basis vectors of the finite dimensional space V.
- Hence we can focus on $\boldsymbol{A}(\cdot)$ in all our analysis

Rational Arrival Process (RAP) — (C, D)

- Finite-dimensional density given by

$$f(x_1, x_2, \ldots, x_n) = \alpha e^{Cx_1} D e^{Cx_2} D \ldots e^{Cx_n} D \boldsymbol{e} \, .$$

- Non-renewal process, but with (conditional) ME distributed inter-arrival times
- Essentially same physical interpretation as for the ME distribution except
  - $\boldsymbol{A}(\cdot)$ evolves before a jump according to:

$$\frac{\mathrm{d}\boldsymbol{a}(t)}{\mathrm{d}t} = \boldsymbol{a}(t)C - \boldsymbol{a}(t)C\boldsymbol{e} \cdot \boldsymbol{a}(t)$$

  - In state $\boldsymbol{a}$, a jump occurs with stochastic intensity $\boldsymbol{a}D\boldsymbol{e}$.
  - After a jump $\boldsymbol{A}(\cdot)$ jumps to $\dfrac{\boldsymbol{a}D}{\boldsymbol{a}D\boldsymbol{e}}$

$A(\cdot)$ - fundamentals

- $A(\cdot)$ is a piece-wise deterministic Markov process on a subset of $\mathbb{R}^m$.
- If we embed a RAP in a queueing model, we need to use general state-space Markov process theory to analyse such a process
- or do you??
- Let's consider a special case

QBD with RAP components

- Use the RAP process and ME distributions to define a QBD with RAP components
- Clearly such a process exists since a MAP is a RAP and phase-type distribution is an ME distribution – MAP/PH/1 queue
- Representation could be (C, D_{-1}, D_1) where
  - D_{-1} corresponds to jumps that take the process down one level
  - D_1 corresponds to jumps that take the process up one level

QBD with RAP components

- Bean and Nielsen (2010) developed an analysis by exploiting the physical interpretation (above)
- Extended the Matrix-Analytic Methods argument to this domain
- Relied on the fact that $\mathbb{E}[\boldsymbol{A}(t+s)|\mathcal{F}_t] = \boldsymbol{A}(t)e^{(C+D_1+D_{-1})s}$
- Essentially, the expectation of $\boldsymbol{A}(\cdot)$ contains all the essential information because of the inherent linearity of the process

Richard Tweedie (1982) - framework

- Considered the QBD concept where the underlying space was a general state space
- Two dimensional state space $\mathbb{Z}_+ \times \mathbb{J}$: Phase – $\mathbb{J} \subset \mathbb{R}^m$
- Kernel (in discrete-time) of the form

$$Q = \begin{bmatrix} B_0(x,J) & A_1(x,J) & 0 & 0 & 0 & \cdots \\ A_{-1}(x,J) & A_0(x,J) & A_1(x,J) & 0 & 0 & \ddots \\ 0 & A_{-1}(x,J) & A_0(x,J) & A_1(x,J) & 0 & \ddots \\ 0 & 0 & A_{-1}(x,J) & A_0(x,J) & A_1(x,J) & \ddots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots \end{bmatrix}.$$

Richard Tweedie (1982) - results

- Invariant measure of the form $\boldsymbol{\mu} = (\mu_0, \mu_1, \mu_2, \ldots)$ where

$$\mu_{n+1}(J) = \int_{\mathbb{J}} \mu_n(\mathrm{d}x)\tilde{S}(x, J)$$

$$\tilde{S}(x, J) = A_1(x, J) + \int_{\mathbb{J}} \tilde{S}(x, \mathrm{d}y)A_0(y, J) + \int_{\mathbb{J}} \tilde{S}^2(x, \mathrm{d}y)A_{-1}(y, J)$$

$$\tilde{S}^2(x, J) = \int_{\mathbb{J}} \tilde{S}(x, \mathrm{d}y)\tilde{S}(y, J)$$

and

$$\mu_0(J) = \int_{\mathbb{J}} \mu_0(\mathrm{d}x)B_0(x, J) + \int_{\mathbb{J}}\int_{\mathbb{J}} \mu_0(\mathrm{d}x)S(x, \mathrm{d}y)A_{-1}(y, J)$$

Apply Tweedie to a QBD with RAP components

- Consider the embedded discrete-time chain at changes in level of the QBD with RAP components
- Discrete-time QBD on a general state-space
- Tweedie says we need to find the operator $\tilde{S}(x, J)$
- Numerically essentially imposible
- Provides detailed information that is essentially meaningless/useless
- Recall that all we really need is the expectation of the phase process $\boldsymbol{A}(\cdot)$
- What would happen if we considered the expectation of the kernels?

Γ-linearity

- Let Γ be some linear continuous operator taking values in a topological finite-dimensional vector space V.

Definition (rough)

An operator Π is said to be Γ-linear if $\Gamma(\Pi(\phi)) = \Gamma(\phi)P$ for all measures ϕ and a given matrix P.

- What happens to Tweedie's results if the kernels are Γ-linear?

Richard Tweedie (1982) - Γ - linearity

- Assume that the kernels are such that

$$\Gamma(A_i(\phi)) = \Gamma(\phi)A_i, \quad i = -1, 0, 1, \quad \textit{and} \quad \Gamma(B_0(\phi)) = \Gamma(\phi)B_0$$

- Invariant measure $\boldsymbol{\mu} = (\mu_0, \mu_1, \mu_2, \ldots)$ satisfies

$$\Gamma(\mu_{n+1}) = \Gamma(\mu_n)S$$

where

$$S = A_1 + SA_0 + S^2A_{-1}$$

and

$$\Gamma(\mu_0) = \Gamma(\mu_0)\left(B_0 + SA_{-1}\right)$$

Expectation-linearity of the (embedded) QBD with RAP components

Theorem

- The operators $A_i(x, J)$ are expectation-linear with matrices $(-C)^{-1}D_i$ for $i = -1, 1$.
- The matrix S is a solution to $S = (-C)^{-1}D_1 + S^2(-C)^{-1}D_{-1}$

- Can undo the embedding and determine the expectation (with respect to the continuous phase variable) of the time-stationary distribution of the level variable

Stationary distribution of QBD with RAP components

- Let $\boldsymbol{\pi}(\cdot) = (\pi_0(\cdot), \pi_1(\cdot), \ldots)$ be the (time) stationary measure of the QBD with RAP components
- Let $\boldsymbol{\theta}_i = \mathbb{E}(\pi_i)$

Theorem

- $\boldsymbol{\theta}_{n+1} = \boldsymbol{\theta}_n R$
- R is a solution to the matrix-quadratic equation

$$0 = A_1 + RA_0 + R^2 A_{-1}$$

- Any algorithm for standard QBDs for solving this matrix-quadratic equation that relies on level-censoring arguments will produce the required solution in the environment of a QBD with RAP components

Conclusions

- Richard Tweedie's operator extension of some Matrix-Analytic Methods results can be directly applied (after embedding at level changes)
- No practical way to compute these results and they provide essentially useless information
- Take the expectation of all results (with respect to the continuous phase-variable) to provide a numerically tractable approach
- Essentially replicates the results of the traditional Matrix-Analytic Methods for a standard QBD
- Equations are identical
- Analysis is very different!

Conclusions

- ME distributions are a generalization of phase-type distributions and are equivalent to Cox's class of distributions with rational Laplace transforms
- ME distributions have some potential practical advantages over phase-type distributions
  - Phase-type distributions have a simpler physical interpretation.
  - Minimal-order representations of matrix exponential distributions are much easier to find.
  - Fitting phase-type distributions to data is difficult: multiple representations, multi-modal objective function,...
  - Fitting matrix exponential distributions is potentially easy via the Laplace Transform: If we can implement Semi-Infinite Programming!
- Rational Arrival Processes are the process extension of ME distributions

Conclusions

Matrix Exponential Distributions versus Phase-type Distributions?

- Can you identify physically the phases to justify the use of phase-type distributions?
- Do you need the extra flexibility of matrix exponential distributions?
- Does the required analysis exist for your model?
  - For some models, the equations are identical so this does not influence the decision.
  - Choose the class that is easier to fit to data in this situation.