Micro and macro views of discrete state Markov models and their application to efficient simulation with Phase-type distributions

Philipp Reinecke[1], Miklós Telek[2], Katinka Wolter[1]

[1] Institut für Informatik, Freie Universität Berlin,

[2] Department of Telecommunications, Technical University of Budapest,

e-mail: {philipp.reinecke,katinka.wolter}@fu-berlin.de, telek@hit.bme.hu

Part 1: Outline

- Starting point: CTMC

- Processes with matrix exponential functions
  - Phase type distributions
  - Matrix exponential distributions
  - Markov arrival process
  - Rational arrival process

- Compositional models
  - Markovian/non-Markovian components
  - Equivalence relations
  - Congruence results

Starting point: CTMC

$X(t) \in S$ is a CTMC.

$S = \{1, 2, \dots, n\}$: discrete finite state space.

$\boldsymbol{Q} = \{q_{ij}\}$ infinitesimal generator matrix.

q_{ij}: transition rate from state i to state j $(i \neq j)$.

$-q_{ii}$: departure rate from state i.

For a regular CTMC $q_{ii} = -\sum_{j \in S} q_{ij} \quad \Rightarrow \quad \boldsymbol{Q}\mathbb{1} = \boldsymbol{0}$,

where $\mathbb{1}$ is a column vector of ones.

$Pr(X(t) = j | X(0) = i) = \left[e^{\boldsymbol{Q}t}\right]_{ij}$

$e^{\boldsymbol{Q}t}$ is a stochastic matrix: $e^{\boldsymbol{Q}t}\mathbb{1} = \boldsymbol{I}\mathbb{1} + \underbrace{\sum_{i=1}^{\infty} \boldsymbol{Q}^i\mathbb{1}\ t^i/i!}_{\boldsymbol{0}} = \mathbb{1}$

Starting point: transient CTMC

$X(t) \in S$ is a transient CTMC.

$S = \{1, 2, \ldots, n\}$: discrete finite state space.

$\boldsymbol{A} = \{a_{ij}\}$ transient infinitesimal generator matrix.

a_{ij}: transition rate from state i to state j $(i \neq j)$.

$-a_{ii}$: departure rate from state i.

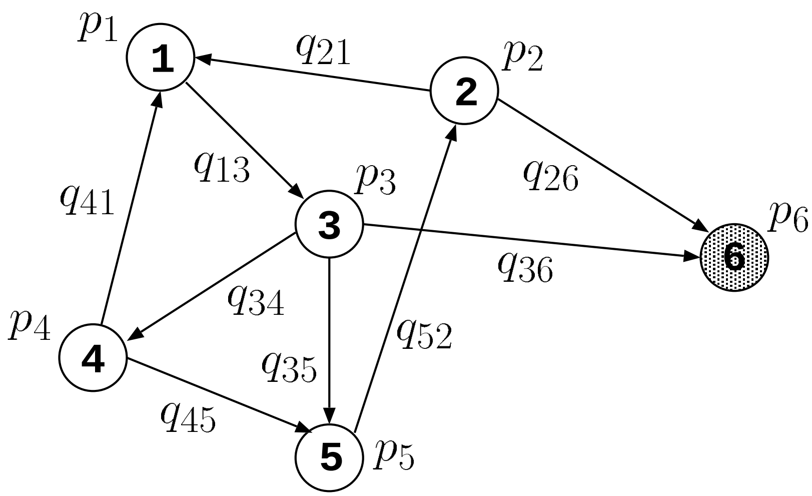
For a transient CTMC $a_{ii} \leq -\sum_{j \in S} a_{ij} \quad \Rightarrow \quad \boldsymbol{A}\mathbb{1} \leq \mathbf{0}$.

$Pr(X(t) = j | X(0) = i) = \left[e^{\boldsymbol{A}t}\right]_{ij}$

$e^{\boldsymbol{A}t}$ is a sub-stochastic matrix: $e^{\boldsymbol{A}t}\mathbb{1} \leq \mathbb{1}$

Phase type distributions

T: time to absorption in a Markov chain with n transient, 1 absorbing state, initial probability vector α and transient generator $\boldsymbol{A}$.



Generator matrix: $\mathbf{Q} = \begin{bmatrix} \mathbf{A} & \mathbf{a} \\ \mathbf{0} & 0 \end{bmatrix}$ $\quad (\mathbf{a} = -\mathbf{A}\mathbb{1})$

Properties of the generator matrix

Generator matrix: $\mathbf{Q} = \begin{bmatrix} \mathbf{A} & \mathbf{a} \\ \mathbf{0} & 0 \end{bmatrix}$ $\qquad (\mathbf{a} = -\mathbf{A}\mathbb{1})$

Transition probability matrix: $e^{\mathbf{Q}t} = \begin{bmatrix} e^{\mathbf{A}t} & \star \\ \mathbf{0} & 1 \end{bmatrix}$

For $i, j \leq n$:

$Pr(X(t) = j | X(0) = i) = [e^{\boldsymbol{Q}t}]_{ij} = [e^{\boldsymbol{A}t}]_{ij}$

Properties of the generator matrix

States $1, 2, \ldots, n$ are transient

$\Rightarrow \lim_{t \to \infty} Pr(X(t) < n + 1) = 0$

$\Rightarrow$ the eigenvalues of $\mathbf{A}$ have negative real part

$\Rightarrow$ $\mathbf{A}$ is non-singular

$\Rightarrow$ $(-\mathbf{A})^{-1}$ has an important stochastic interpretation

Assumption: the CTMC starts from a transient state ($\alpha \mathbb{1} = 1$).

Properties of phase type distributions

$$
\begin{aligned}
Pr(T < t) &= Pr(X(t) = n+1) = 1 - \sum_{i=1}^{n} Pr(X(t) = i) = \\
&= 1 - \sum_{k=1}^{n} \sum_{i=1}^{n} \underbrace{Pr(X(0) = k)}_{\alpha_k} \underbrace{Pr(X(t) = i | X(0) = k)}_{[e^{\boldsymbol{A}t}]_{ki}} \\
&= 1 - \alpha e^{\boldsymbol{A}t} \mathbb{1}
\end{aligned}
$$

Representation: PH($\alpha, \boldsymbol{A}$)
initial probability distribution ($\boldsymbol{\alpha}$) /$n-1$ parameters/ +
transient infinitesimal generator matrix ($\boldsymbol{A}$) /n^2/

Only for transient states. /$n^2 + n - 1$/

Properties of phase type distributions

CDF: $F(t) = 1 - \boldsymbol{\alpha} e^{\mathbf{A}t} \mathbb{1}$

PDF: $f(t) = \boldsymbol{\alpha} e^{\mathbf{A}t} \mathbf{a}$

moments: $\mu_k = E(T^k) = k!\ \boldsymbol{\alpha}(-\mathbf{A})^{-k} \mathbb{1}$

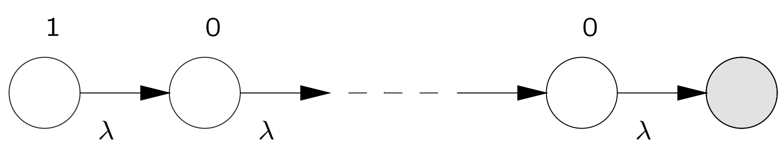
LST:

$$
\begin{aligned}
f^*(s) &= \boldsymbol{\alpha}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{a} = \boldsymbol{\alpha} \left[\frac{det(s\mathbf{I} - \mathbf{A})_{ji}}{det(s\mathbf{I} - \mathbf{A})} \right] \mathbf{a} = \\
&= \frac{s^{n-1} + a_{n-2}s^{n-2} + \ldots + a_1 s + a_0}{s^n + b_{n-1}s^{n-1} + \ldots + b_1 s + b_0}
\end{aligned}
$$

$$
f^*(s)|_{s \to 0} = \int_0^\infty f(t)dt = 1 \qquad \Rightarrow \qquad a_0 = b_0 \qquad /2n-1/
$$

Properties of phase type distributions

- rational Laplace tr.
- closed for min/max, mixture, summation, ...
- $f(t) > 0$
- support on $(0, \infty)$
- exponential tail decay
- $CV_{min} = \frac{1}{N}$ only for Erlang distribution



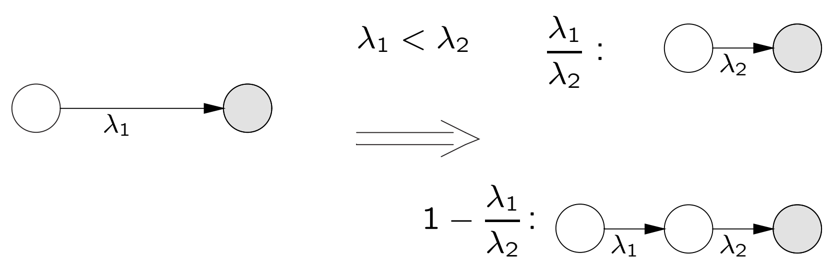
Similar PH distributions

If $\boldsymbol{B}$ is nonsingular, $\boldsymbol{B}\mathbb{1} = \mathbb{1}$, $\gamma = \alpha\boldsymbol{B}$ and $\boldsymbol{G} = \boldsymbol{B}^{-1}\boldsymbol{A}\boldsymbol{B}$

then PH($\alpha, \boldsymbol{A}$) =PH($\gamma, \boldsymbol{G}$)

$$F(t) = 1 - \boldsymbol{\gamma} e^{\mathbf{G}t}\mathbb{1} = 1 - \alpha\boldsymbol{B}\ e^{\boldsymbol{B}^{-1}\boldsymbol{A}\boldsymbol{B}t}\ \boldsymbol{B}^{-1}\mathbb{1} = 1 - \boldsymbol{\alpha} e^{\mathbf{A}t}\mathbb{1}$$

Identity of PH distributions of different sizes:



$$\left(\frac{\lambda_1}{\lambda_2}\right)\ \frac{\lambda_2}{s+\lambda_2} + \left(1 - \frac{\lambda_1}{\lambda_2}\right)\ \frac{\lambda_1}{s+\lambda_1}\ \frac{\lambda_2}{s+\lambda_2} = \frac{\lambda_1}{s+\lambda_1}$$

Special PH classes

A unique and minimal representation (canonical form) of the PH class is not available

$\longrightarrow$ use of simple PH subclasses:

- Acyclic PH distributions
- Hypo-exponential distr. ("series", "$cv < 1$")
- Hyper-exponential distr. ("parallel", "$cv > 1$")
- ...

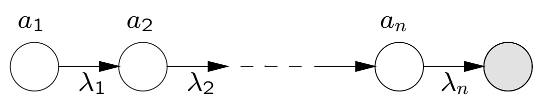
Acyclic PH distributions

Each transient state is visited at most ones

$\Rightarrow$ triangular generator

$\Rightarrow$ real eigenvalues

The acyclic PH class allows a unique and minimal (canonical) representation with only $2N-1$ parameters.



where $\lambda_i < \lambda_{i+1}$ and $\sum_i a_i = 1$ /$2n-1$/.

Matching with PH distributions

Moments matching:
Find a PH distribution with the same first K moments.

- Solution exists for $K = 2n - 1$,

  but the result is not necessarily a distribution.

- Open problem for $3 < K < 2n - 1$.

Fitting with PH distributions

Fitting:
given a non-negative distribution find a "similar" PH distribution.

Formally:

$$\min_{PH parameters} \left\{ \text{Distance}(PH, Original) \right\}$$

Distance:

- squared CDF difference: $\int_0^\infty (F(t) - \widehat{F}(t))^2 dt$
- density difference: $\int_0^\infty |f(t) - \widehat{f}(t)| dt$
- relative entropy: $\int_0^\infty f(t)\ log \left(\frac{f(t)}{\widehat{f}(t)} \right) dt$

Fitting with PH distributions

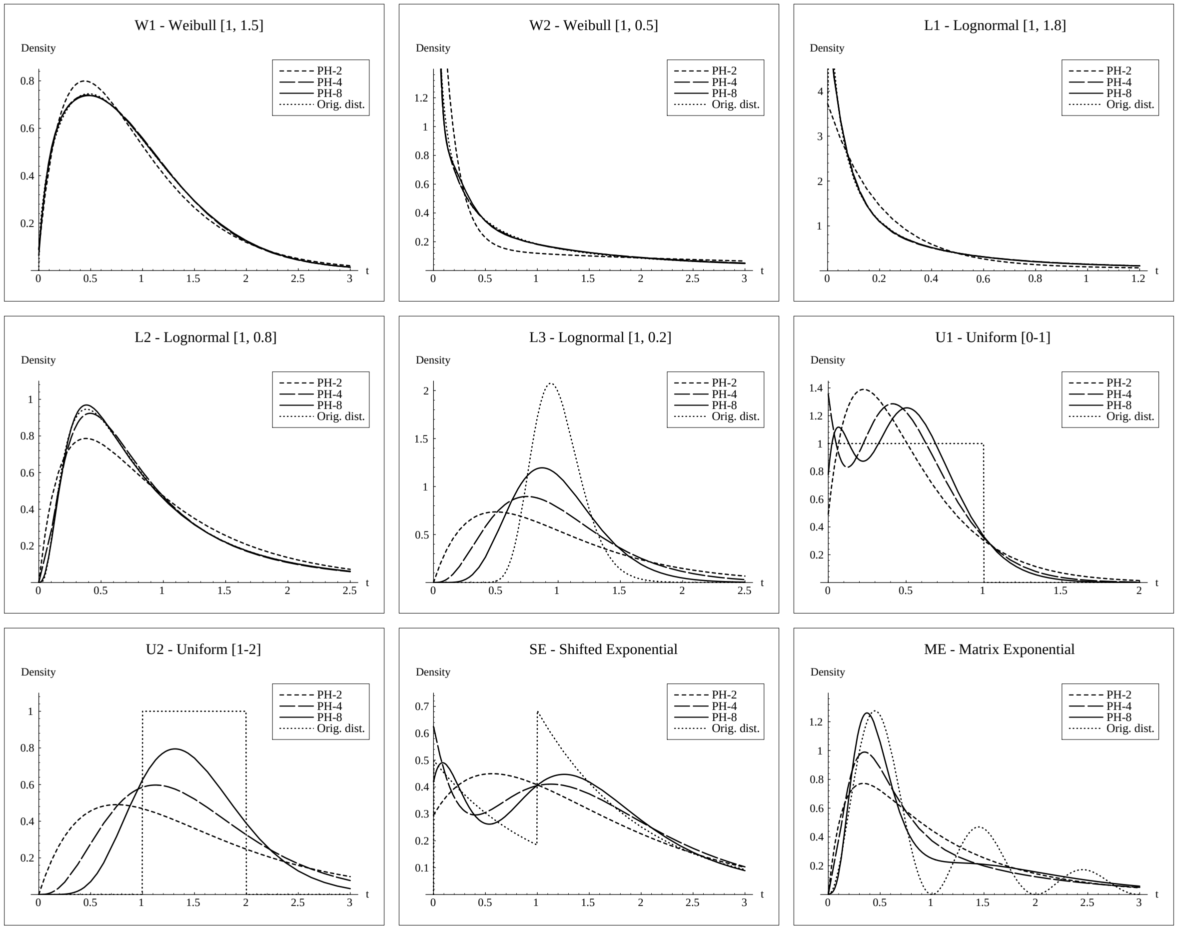
Problems:

- vector-matrix representation:
  - $\sim n^2$ parameters $\rightarrow$ over-parameterized,
  - easy to check the PH conditions,
- moments or Laplace representation:
  - $2n-1$ parameters $\rightarrow$ minimal number of parameters,
  - hard to check the PH conditions.

One possible solution:

- Acyclic PH with canonical representation:
  - $2n-1$ parameters,
  - easy to check the PH conditions,
  - .... but only for a subclass of PH distributions.

Fitting with PH distributions



Applications of Phase type distributions

Non-Markovian (non-exponential) models $\rightarrow$ Markovian analysis (transient $p_0 e^{\boldsymbol{Q}t}$, stationary $p\boldsymbol{Q} = 0, p\,\mathbb{1} = 1$)

- queueing models (matrix geometric methods)
- performance, performability models
- stochastic model description languages (Petri net, process algebra)

Matrix exponential distribution

T has a matrix exponential distribution is its CDF has the form

$$F(t) = 1 - \alpha e^{\boldsymbol{A}t}\mathbb{1}$$

where α is a row vector and $\boldsymbol{A}$ is a square matrix (without any structural restriction).

The vector matrix pair $(\alpha, \boldsymbol{A})$ define a distribution if $F(t) = 1 - \alpha e^{\boldsymbol{A}t}\mathbb{1}$ is monotone increasing.

- Easy to check necessary and sufficient conditions are not available.

- Closed form necessary and sufficient conditions are available for $n = 3$.

Properties of matrix exponential distributions

- rational Laplace tr.
- closed for min/max, mixture, summation, ...
- $f(t) \leq 0$
- support on $(0, \infty)$
- exponential tail decay
- $CV_{min} << \frac{1}{n}$

  ($n = 3$: $CV_{min} \sim 1/5$, $n = 15$: $CV_{min} \sim 1/100$)
- $CV_{min} \leftrightarrow$ only conjectures exit

Applications of matrix exponential distributions

Non-Markovian models $\rightarrow$ easy to compute non-Markovian analysis (transient $p_0 e^{\boldsymbol{Q}t}$, stationary $p\boldsymbol{Q} = 0, p\mathbb{1} = 1$)

- queueing models (matrix geometric methods)
- performance, performability models
- stochastic model description languages (Petri net, process algebra)

Markov arrival process

A point process characterized by a modulating CTMC.

- $\boldsymbol{D_0}$: state (phase) transition rate without arrival
- $\boldsymbol{D_1}$: state (phase) transition rate with arrival
- $\boldsymbol{D}_{1ii}$: arrival rate when the CTMC is in state i.

$\boldsymbol{D} = \boldsymbol{D_0} + \boldsymbol{D_1}$ generator of the modulating CTMC.

$\boldsymbol{D}\mathbb{1} = \boldsymbol{0}$.

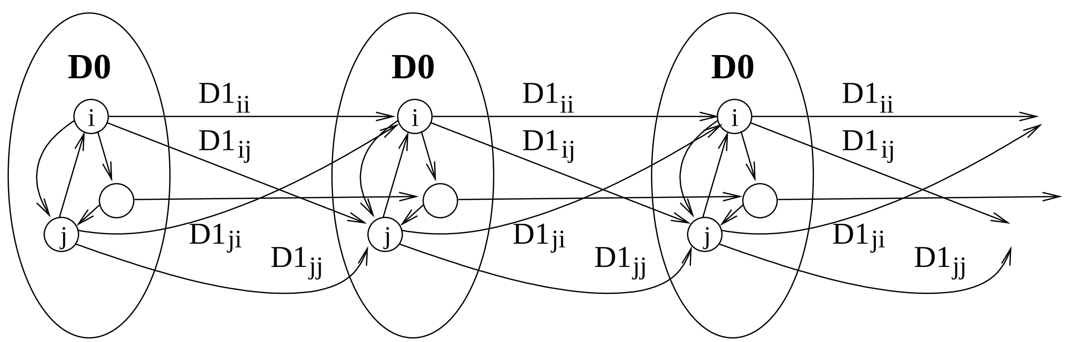
Properties of Markov arrival process

MAP: correlated arrivals

the phase distribution after an arrival depends on the previous inter-arrival time

$\{N(t), J(t)\}$ is a Markov chain, where

- $N(t)$: number of arrivals
- $J(t)$: phase of the CTMC



Markov arrival process

Structure of the generator matrix:

$$
\mathbf{Q} = \begin{pmatrix} \mathbf{D_0} & \mathbf{D_1} & & & \\ & \mathbf{D_0} & \mathbf{D_1} & & \\ & & \mathbf{D_0} & \mathbf{D_1} & \\ & & & \mathbf{D_0} & \mathbf{D_1} \\ & & & & \ddots \end{pmatrix}
$$

On the block level it is similar to the structure of a Poisson process.

$\longrightarrow$ "quasi" birth process.

Properties of Markov arrival process

- the phase distribution at arrival instances form a DTMC with $\mathbf{P} = (-\mathbf{D_0})^{-1}\mathbf{D_1}$
  $\longrightarrow$ correlated initial phase distributions,

- inter-arrival time is PH distributed with representation $(\boldsymbol{\alpha}_0, \mathbf{D_0})$, $(\boldsymbol{\alpha}_1, \mathbf{D_0})$, $(\boldsymbol{\alpha}_2, \mathbf{D_0})$, ...
  $\longrightarrow$ correlated inter-arrival times,

- phase process $(J(t))$ is a CTMC with generator $\mathbf{D} = \mathbf{D_0} + \mathbf{D_1}$

Properties of Markov arrival process

- (embedded) stationary phase distribution after an arrival π is the solution of $\pi \mathbf{P} = \pi, \pi \mathbb{1} = 1$.
- stationary inter arrival time is PH($\pi, \mathbf{D_0}$).
- the stationary arrival intensity is $\lambda = \dfrac{1}{\pi(-\mathbf{D_0})^{-1}\mathbb{1}}$.

Properties of Markov arrival process

The joint pdf of X_0 and X_k is

$$f_{X_0,X_k}(x,y) = \pi e^{\mathbf{D}_0 x}\mathbf{D_1}\mathbf{P}^{k-1}e^{\mathbf{D}_0 y}\mathbf{D_1}\mathbb{1}.$$

Due to the Markovian behaviour of MAPs X_0 and X_k depend only via their initial states !!

Lag k joint moment ($\rightarrow$ correlation):

$$\begin{aligned} E(X_0X_k) &= \int_{t=0}^{\infty}\int_{\tau=0}^{\infty} t\ \tau\ \pi e^{\mathbf{D}_0 t}\mathbf{D_1}\mathbf{P}^{k-1}e^{\mathbf{D}_0\tau}\mathbf{D_1}\mathbb{1}\ d\tau\ dt \\ &= \pi \underbrace{\int_{t=0}^{\infty} t\ e^{\mathbf{D}_0 t}\ dt}_{(-\mathbf{D}_0)^{-2}}\mathbf{D_1}\mathbf{P}^{k-1}\underbrace{\int_{\tau=0}^{\infty}\tau\ e^{\mathbf{D}_0\tau}\ d\tau}_{(-\mathbf{D}_0)^{-2}}\mathbf{D_1}\mathbb{1} \\ &= \pi(-\mathbf{D_0})^{-1}\mathbf{P}^k(-\mathbf{D_0})^{-1}\mathbb{1} \end{aligned}$$

Properties of Markov arrival process

Generally for $a_0 = 0 < a_1 < a_2 < \ldots < a_k$
the joint density is:

$$f_{X_{a_0},X_{a_1},\ldots,X_{a_k}}(x_0, x_1, \ldots, x_k) =$$

$$= \pi e^{\mathbf{D_0} x_0} \mathbf{D_1} \mathbf{P}^{a_1 - a_0 - 1} e^{\mathbf{D_0} x_1} \mathbf{D_1} \mathbf{P}^{a_2 - a_1 - 1} \ldots e^{\mathbf{D_0} x_k} \mathbf{D_1} \mathbb{1} \ ,$$

and the joint moment is:

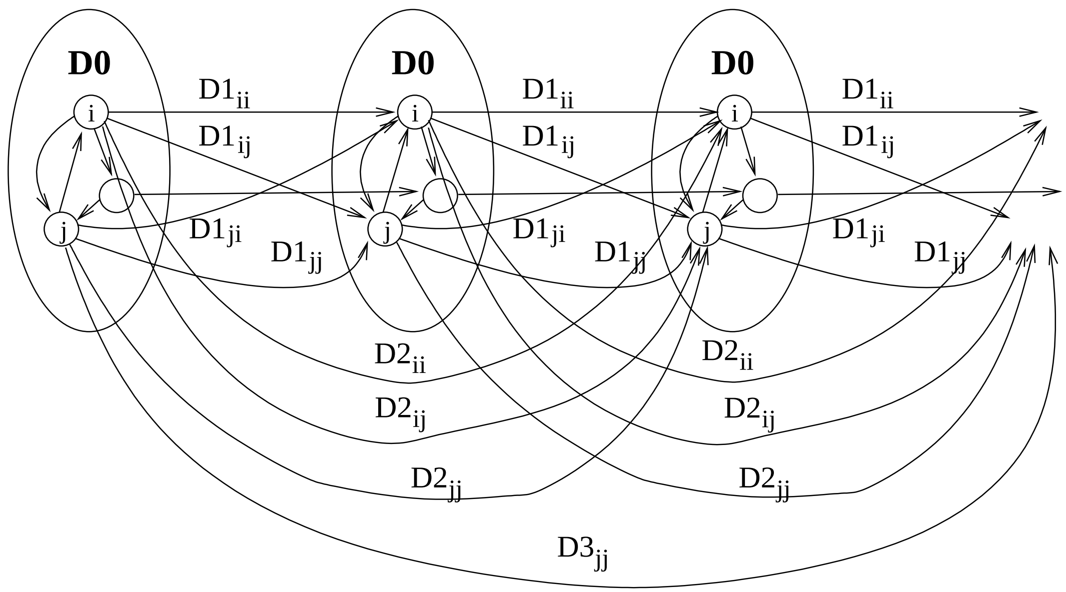
$$E(X_{a_0}^{i_0}, X_{a_1}^{i_0}, \ldots, X_{a_k}^{i_0}) =$$

$$= \pi i_0! (-\mathbf{D_0})^{-i_0} \mathbf{P}^{a_1 - a_0} i_1! (-\mathbf{D_0})^{-i_1} \mathbf{P}^{a_2 - a_1} \ldots i_k! (-\mathbf{D_0})^{-i_k} \mathbb{1} \ .$$

Batch Markov arrival process

MAP with batch arrivals

- $\mathbf{D_0}$ – phase transitions without arrival
- $\mathbf{D_k}$ – phase transitions with k arrivals



$\longrightarrow \{N(t), J(t)\}$ is still a Markov chain.

Batch Markov arrival process

Structure of the generator matrix:

$$\mathbf{Q} = \begin{pmatrix} \mathbf{D_0} & \mathbf{D_1} & \mathbf{D_2} & \mathbf{D_3} & \mathbf{D_4} \\ & \mathbf{D_0} & \mathbf{D_1} & \mathbf{D_2} & \mathbf{D_3} \\ & & \mathbf{D_0} & \mathbf{D_1} & \mathbf{D_2} \\ & & & \mathbf{D_0} & \mathbf{D_1} \\ & & & & \ddots \end{pmatrix}$$

Properties of matrices $\mathbf{D_k}$:

- $\mathbf{D_0}$: $\mathbf{D_0}_{ij} \geq 0$ for $i \neq j$, and $\mathbf{D_0}_{ii} \leq 0$
- for $k \geq 1$: $\mathbf{D_k}_{ij} \geq 0$

Examples of (batch) Markov arrival processes

- bath PH renewal process:
  $\mathbf{D_0} = \mathbf{A}$, $\mathbf{D_k} = p_k \mathbf{a}\boldsymbol{\alpha}$.

- MMPP (Markov modulated Poisson process):
  $\mathbf{D_0} = \mathbf{Q} - \text{diag}{<}\boldsymbol{\lambda}{>}$, $\mathbf{D_1} = \text{diag}{<}\boldsymbol{\lambda}{>}$.

- IPP (Interrupted Poisson process):

$$\mathbf{D_0} = \begin{array}{|c|c|}\hline -\alpha-\lambda & \alpha \\ \hline 0 & -\beta \\ \hline \end{array}, \qquad \mathbf{D_1} = \begin{array}{|c|c|}\hline \lambda & 0 \\ \hline 0 & 0 \\ \hline \end{array}.$$

- batch MMPP :
  $\mathbf{D_0} = \mathbf{Q} - \text{diag}{<}\boldsymbol{\lambda}{>}$, $\mathbf{D_k} = p_k \ \text{diag}{<}\boldsymbol{\lambda}{>}$.

Examples of (batch) Markov arrival processes

- filtered MAP (arrivals discarded with probability p):
  $\mathbf{D_0} = \hat{\mathbf{D}}_0 + p\hat{\mathbf{D}}_1$, $\mathbf{D_1} = (1-p)\hat{\mathbf{D}}_1$.
- cyclicly filtered MAP (every second arrivals are discarded with probability p):

$$\mathbf{D_0} = \begin{array}{|c|c|}\hline \hat{\mathbf{D}}_0 & 0 \\ \hline p\hat{\mathbf{D}}_1 & \hat{\mathbf{D}}_0 \\ \hline \end{array}, \qquad \mathbf{D_1} = \begin{array}{|c|c|}\hline 0 & \hat{\mathbf{D}}_1 \\ \hline (1-p)\hat{\mathbf{D}}_1 & 0 \\ \hline \end{array}.$$

- superposition of BMAPs:
  $\mathbf{D_k} = \hat{\mathbf{D}}_\mathbf{k} \bigoplus \tilde{\mathbf{D}}_\mathbf{k}$,

Kronecker product: $\mathbf{A} \bigotimes \mathbf{B} = \begin{array}{|ccc|}\hline A_{11}\mathbf{B} & \ldots & A_{1n}\mathbf{B} \\ \vdots & & \vdots \\ A_{n1}\mathbf{B} & \ldots & A_{nn}\mathbf{B} \\ \hline \end{array}$

Kronecker sum: $\mathbf{A} \bigoplus \mathbf{B} = \mathbf{A} \bigotimes \mathbf{I_B} + \mathbf{I_A} \bigotimes \mathbf{B}$

Examples of (batch) Markov arrival processes

- Departure process of an M/M/1/2 queue:

$\mathbf{D_0} =$

$-\lambda$	λ	
	$-\lambda-\mu$	λ
		$-\mu$

$\mathbf{D_1} =$

μ		
	μ	

- Overflow process of an M/M/1/2 queue:

$\mathbf{D_0} =$

$-\lambda$	λ	
μ	$-\lambda-\mu$	λ
	μ	$-\lambda-\mu$

$\mathbf{D_1} =$

		λ

- Correlated inter-arrivals ($\lambda_1 \neq \lambda_2$):

$\mathbf{D_0} =$

$-\lambda_1$	0
0	$-\lambda 2$

$\mathbf{D_1} =$

$p\lambda_1$	$(1-p)\lambda_1$
$(1-p)\lambda_2$	$p\lambda_2$

$p \sim 1 \rightarrow$ positive correlated consecutive inter-arrivals

$p \sim 0 \rightarrow$ negative correlated consecutive inter-arrivals

Rational arrival process

A point process with inter-arrival time $X_0, X_1, \ldots$ is a Rational arrival process if its joint density for $a_0 = 0 < a_1 < a_2 < \ldots < a_k$ has the form:

$$f_{X_{a_0}, X_{a_1}, \ldots, X_{a_k}}(x_0, x_1, \ldots, x_k) =$$

$$= \pi e^{\mathbf{D_0} x_0} \mathbf{D_1} \mathbf{P}^{a_1 - a_0 - 1} e^{\mathbf{D_0} x_1} \mathbf{D_1} \mathbf{P}^{a_2 - a_1 - 1} \ldots e^{\mathbf{D_0} x_k} \mathbf{D_1} \mathbb{1} \ ,$$

The matrix pair $\mathbf{D_0}, \mathbf{D_1}$ (without any structural description) define a Rational arrival process if

$$f_{X_{a_0}, X_{a_1}, \ldots, X_{a_k}}(x_0, x_1, \ldots, x_k)$$

is non-negative for $\forall k, a_0 < a_1 < a_2 < \ldots < a_k, x_0, x_1, \ldots, x_k$.

Queues with PH, MAP arrival/departure

Example: PH/M/1 queue

- arrival process: PH($\tau, \mathbf{T}$) renewal process ($t = -\mathbf{T}\mathbb{1}$)
- service time: exponentially distributed with parameter μ.

$$\mathbf{Q} = \begin{pmatrix} \mathbf{T} & t\tau & & & \\ \mu\mathbf{I} & \mathbf{T}-\mu\mathbf{I} & t\tau & & \\ & \mu\mathbf{I} & \mathbf{T}-\mu\mathbf{I} & t\tau & \\ & & \mu\mathbf{I} & \mathbf{T}-\mu\mathbf{I} & t\tau \\ & & & \ddots & \ddots \end{pmatrix}$$

$\longrightarrow \{N(t), J(t)\}$ is a Markov chain with generator

Queues with PH, MAP arrival/departure

Example: MAP/PH/1 queue

- arrival process: MAP($\mathbf{D_0}, \mathbf{D_1}$),
- service time: PH($\tau, \mathbf{T}$), ($t = -\mathbf{T}\mathbb{1}$).

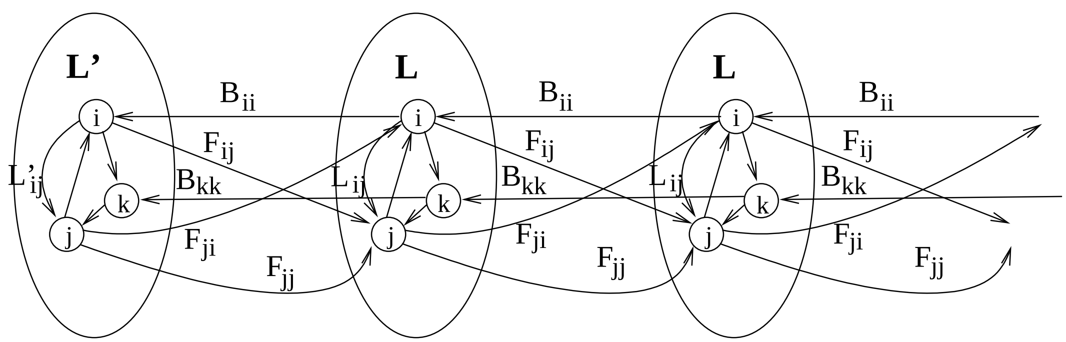
$$\mathbf{Q} = \begin{array}{|c|c|c|c|}\hline \mathbf{L}' & \mathbf{F}' & & \\ \hline \mathbf{B}' & \mathbf{L} & \mathbf{F} & \\ \hline & \mathbf{B} & \mathbf{L} & \ddots \\ \hline & & \ddots & \ddots \\ \hline \end{array}$$

where
$\mathbf{F} = \mathbf{D_1} \bigotimes \mathbf{I}$, $\mathbf{L} = \mathbf{D_0} \bigoplus \mathbf{T}$, $\mathbf{B} = \mathbf{I} \bigotimes t\tau$,
$\mathbf{F}' = \mathbf{D_1} \bigotimes \tau$, $\mathbf{L}' = \mathbf{D_0}$, $\mathbf{B}' = \mathbf{I} \bigotimes \mathbf{T}$.

Quasi birth-death process

- $N(t)$ is the "level" process (e.g., number of customers in a queue),
- $J(t)$ is the "phase" process (e.g., state of the environment).

The CTMC $\{N(t), J(t)\}$ is a Quasi birth-death process if transitions are restricted to one level up or down or inside the same level.



Level 0 is irregular (e.g., no departure).

Quasi birth-death process

Structure of the transition probability matrix:

$$\mathbf{Q} = \begin{array}{|c|c|c|c|c|}\hline \mathbf{L}' & \mathbf{F} & & & \\ \hline \mathbf{B} & \mathbf{L} & \mathbf{F} & & \\ \hline & \mathbf{B} & \mathbf{L} & \mathbf{F} & \\ \hline & & \mathbf{B} & \mathbf{L} & \mathbf{F} \\ \hline & & & \ddots & \ddots \\ \hline \end{array}$$

On the block level it has a birth-death structure

$\longrightarrow$ “quasi” birth-death process.

Matrix geometric distribution

Stationary solution: $\boldsymbol{\pi}\mathbf{Q} = \mathbf{0}$, $\boldsymbol{\pi}\mathbb{1} = 1$.

Partitioning $\boldsymbol{\pi}$: $\boldsymbol{\pi} = \{\boldsymbol{\pi}_0, \boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \ldots\}$

Decomposed stationary equations:

$$\boldsymbol{\pi}_0\mathbf{L}' + \boldsymbol{\pi}_1\mathbf{B} = \mathbf{0}$$

$$\boldsymbol{\pi}_{n-1}\mathbf{F} + \boldsymbol{\pi}_n\mathbf{L} + \boldsymbol{\pi}_{n+1}\mathbf{B} = \mathbf{0} \qquad \forall n \geq 1$$

$$\sum_{n=0}^{\infty} \boldsymbol{\pi}_n\mathbb{1} = 1$$

Conjecture: $\boldsymbol{\pi}_n = \boldsymbol{\pi}_{n-1}\mathbf{R} \quad \rightarrow \quad \boldsymbol{\pi}_n = \boldsymbol{\pi}_0\mathbf{R}^n$ and

$$\boldsymbol{\pi}_0\mathbf{L}' + \boldsymbol{\pi}_0\mathbf{R}\mathbf{B} = \mathbf{0}$$

$$\boldsymbol{\pi}_0\mathbf{R}^{n-1}\mathbf{F} + \boldsymbol{\pi}_0\mathbf{R}^n\mathbf{L} + \boldsymbol{\pi}_0\mathbf{R}^{n+1}\mathbf{B} = \mathbf{0} \qquad \forall n \geq 1$$

$$\sum_{n=0}^{\infty} \boldsymbol{\pi}_0\mathbf{R}^n\mathbb{1} = \boldsymbol{\pi}_0(\mathbf{I} - \mathbf{R})^{-1}\mathbb{1} = 1$$

Matrix geometric distribution

The solution is defined by vector π_0 and matrix $\mathbf{R}$:

Matrix $\mathbf{R}$ is the solution of the matrix equation:

$$\mathbf{F} + \mathbf{R}\mathbf{L} + \mathbf{R}^2\mathbf{B} = \mathbf{0}$$

Vector π_0 is the solution of linear system:

$$\boldsymbol{\pi}_0(\mathbf{L}' + \mathbf{R}\mathbf{B}) = \mathbf{0}$$

$$\boldsymbol{\pi}_0(\mathbf{I} - \mathbf{R})^{-1}\mathbb{1} = 1$$

Minimal solution of the quadratic equation

From

$$\mathbf{F} + \mathbf{R}\mathbf{L} + \mathbf{R}^2\mathbf{B} = \mathbf{0}$$

we have

$$\mathbf{R} = \mathbf{F}\left(-\mathbf{L} - \mathbf{R}\mathbf{B}\right)^{-1}$$

A simple numerical algorithm to calculate $\mathbf{R}$:

$$\begin{array}{l} \mathbf{R} := \mathbf{0}; \\ \textbf{REPEAT} \\ \quad \mathbf{R}_{old} := \mathbf{R}; \\ \quad \mathbf{R} := \mathbf{F}\left(-\mathbf{L} - \mathbf{R}\mathbf{B}\right)^{-1}; \\ \textbf{UNTIL} ||\mathbf{R} - \mathbf{R}_{old}|| \leq \epsilon \end{array}$$

Performance measures

The typical performance measures can be computed in an efficient way based on the stationary distribution.

For example, the mean number of customers in the queue is

$$\sum_{i=0}^{\infty} i\pi_i \mathbb{1} = \pi_0 \sum_{i=0}^{\infty} i\boldsymbol{R}^i \mathbb{1} = \pi_0 \boldsymbol{R}(\boldsymbol{I} - \boldsymbol{R})^{-2} \mathbb{1}$$

Queues with ME, RAP arrival/departure

Example: RAP/ME/1 queue

- arrival process: RAP$(\mathbf{D_0}, \mathbf{D_1})$,
- service time: ME$(\tau, \mathbf{T})$, $(t = -\mathbf{T}\mathbb{1})$.

$$\mathbf{Q} = \begin{array}{|c|c|c|c|} \hline \mathbf{L}' & \mathbf{F}' & & \\ \hline \mathbf{B}' & \mathbf{L} & \mathbf{F} & \\ \hline & \mathbf{B} & \mathbf{L} & \ddots \\ \hline & & \ddots & \ddots \\ \hline \end{array}$$

where
$\mathbf{F} = \mathbf{D_1} \bigotimes \mathbf{I}$, $\mathbf{L} = \mathbf{D_0} \bigoplus \mathbf{T}$, $\mathbf{B} = \mathbf{I} \bigotimes t\tau$,
$\mathbf{F}' = \mathbf{D_1} \bigotimes \tau$, $\mathbf{L}' = \mathbf{D_0}$, $\mathbf{B}' = \mathbf{I} \bigotimes \mathbf{T}$.

The same analysis applies as for the Markovian models!!!

Open problems

- Markovian models
  - canonical representation of the PH class
  - structural restrictions of MAPs
  - efficient PH fitting (whole PH class)
  - efficient MAP fitting
- non-Markovian models
  - efficient check if $(\alpha, \mathbf{A})$ defines an ME distribution.
  - efficient check if $(\mathbf{D_0}, \mathbf{D_1})$ defines a RAP.
  - structural restrictions of RAPs
  - ME fitting
  - RAP fitting

Compositional models

A wide range of complex stochastic models are composed by components which form a common stochastic model through simple interactions.

Compositional models

- describe the components $\mathcal{A}^{(i)}$ and
- composition roles the way as they form the system model $(\mathcal{A}^{(1)} \|_{\mathcal{C}} \mathcal{A}^{(2)}) \|_{\mathcal{C}} \mathcal{A}^{(3)} \ldots$

Compositional models

To avoid state space explosion the components are represented in a compact way using an equivalence relation

$\mathcal{A}^{(1)}$	$\sim$	$\mathcal{A}^{(1')}$
of size m_1		of size $n_1 < m_1$

such that this relation is preserved during the composition components

$$\mathcal{A}^{(1)} \sim \mathcal{A}^{(1')} \quad \Rightarrow \quad \mathcal{A}^{(1)} \|_{\mathcal{C}} \mathcal{A}^{(2)} \sim \mathcal{A}^{(1')} \|_{\mathcal{C}} \mathcal{A}^{(2)}.$$

Compositional models

The currently applied compositional models uses

- Markovian components,
- stochastic bisimulation (different forms of lumpability) as equivalence relation ($\dot{\sim}$),
- Kronecker operators for composition of components.

The nice properties of setting are that

- the composed model is Markovian and
- the equivalence relation is preserved by composition of components.

Compositional models

An extension of Markovian compositional models

- non-Markovian components,
- a more general equivalence relation (similarity transformation) ($\simeq$) and
- the same Kronecker operators for composition of components.

The resulted compositional model

- is a non-Markovian system model, which can be computed by similar ODEs (transient) or linear system of equations (stationary) and
- the equivalence relation is preserved by composition of components.

Compositional models

When to use the proposed compositional model?

When $\mathcal{A}^{(1)} \dot{\sim} \mathcal{A}^{(1')}$ of size $m_1 \to n_1$,

but $\mathcal{A}^{(1)} \simeq \mathcal{A}^{(1'')}$ of size $m_1 \to g_1 < n_1$.

Markovian components

A Markovian component is $\mathcal{A} = (\mathcal{S}, \pi, \mathbf{E}_e(e \in \mathcal{E}), \Lambda)$, where

- $\mathcal{S} = \{0, \ldots, m-1\}$ is the finite state space,
- $\pi \in \mathbb{R}^{1,m}$ is the initial probability distribution,
- $\mathcal{E}$ is a finite set of events,
- $\mathbf{E}_e \in \mathbb{R}^{m,m}$ is the non-negative transition weight matrix according to event e
- $\Lambda = (\lambda_e(e \in \mathcal{E}))$ is a positive rate vector.

$\mathcal{E}$ contains a specific event ϵ (local event of the component) that is not observable and

$\mathcal{E}_s = \mathcal{E} \setminus \{\epsilon\}$.

Markovian components

Based on this description we define diagonal matrix

$$\mathbf{D}_e = diag(\mathbf{E}_e \mathbb{1}).$$

The generator matrix of a Markov component is

$$\mathbf{Q} = \underbrace{\lambda_\epsilon (\mathbf{E}_\epsilon - \mathbf{D}_\epsilon)}_{\mathbf{Q}_\epsilon} + \sum_{e \in \mathcal{E}_s} \lambda_e (\mathbf{E}_e - \mathbf{D}_e)$$

with a unique stationary vector ψ

$$\psi \mathbf{Q} = \mathbf{0} \text{ and } \psi \mathbb{1} = 1.$$

The transient and the stationary throughput of event e are

$$\pi e^{\mathbf{Q}t} \mathbf{D}_e \mathbb{1} \quad \text{and} \quad \psi \mathbf{D}_e \mathbb{1}.$$

Markovian components

The joint density for a sequence of k observations $(e_1, t_1, e_2, t_2, \ldots, e_k, t_k)$ is given by

$$f_{\mathcal{A}}(e_1, t_1, \ldots, e_k, t_k) = \pi \left(\prod_{i=1}^{k} e^{\mathbf{R}t_i} \lambda_{e_i} \mathbf{E}_{e_i} \right) 1\!\!1,$$

where

$$\mathbf{R} = \mathbf{Q}_{\epsilon} - \sum_{e \in \mathcal{E}_s} \lambda_e \mathbf{D}_e.$$

In case of Markov components

$$f_{\mathcal{A}}(e_1, t_1, \ldots, e_k, t_k) \geq 0$$

due to the non-negativity of π, $\mathbf{E}_e (e \in \mathcal{E})$ and Λ.

Non-Markovian components

A non-Markovian component is $\mathcal{A} = (\mathcal{S}, \pi, \mathbf{E}_e(e \in \mathcal{E}), \Lambda)$, where

- $\mathcal{S} = \{0, \ldots, m-1\}$ is the finite state space,
- $\pi \in \mathbb{R}^{1,m}$ is a vector with possibly negative elements,
- $\mathcal{E}$ is a finite set of events,
- $\mathbf{E}_e \in \mathbb{R}^{m,m}$ is the transition weight matrix according to event e with possibly negative elements,
- $\Lambda = (\lambda_e(e \in \mathcal{E}))$ is a positive rate vector.

AND

$$f_{\mathcal{A}}(e_1, t_1, \ldots, e_k, t_k) \geq 0$$

for every sequence of $k > 0$ observations $(e_1, t_1, e_2, t_2, \ldots, e_k, t_k)$.

Composition of components

To compose $\mathcal{A}^{(1)}$ and $\mathcal{A}^{(2)}$, without loss of generality, we assume that the event sets $\mathcal{E}$ and the rate vectors Λ of length $|\mathcal{E}|$ are identical for all events.

Composition is performed over the set of signals $\mathcal{E}$:

- signals from $\mathcal{C} \subseteq \mathcal{E}_s$ occur as synchronized signals in both components,

- signals from $\mathcal{N} = \mathcal{E}_s \setminus \mathcal{C}$ and signal ϵ occur independently.

Composition of components

The composed model $\mathcal{A}^{(0)} = \mathcal{A}^{(1)} \|_{\mathcal{C}} \mathcal{A}^{(2)}$ is defined by

- state space $\mathcal{S} = \{0, \ldots, m_1 m_2 - 1\}$,
- vector $\pi^{(0)} = \pi^{(1)} \otimes \pi^{(2)}$.
- weight matrices

$$\mathbf{E}_e^{(0)} = \begin{cases} \mathbf{E}_e^{(1)} \oplus \mathbf{E}_e^{(2)} & \text{if } e \in \mathcal{N} \cup \{\epsilon\}, \\ \mathbf{E}_e^{(1)} \otimes \mathbf{E}_e^{(2)} & \text{if } e \in \mathcal{C}, \end{cases}$$

- rate vector $\Lambda = (\lambda_e(e \in \mathcal{E}))$.

Observed equivalence

Definition 1 *Two components $\mathcal{A}^{(1)}$ and $\mathcal{A}^{(2)}$ observed to be equivalent, if and only if*

$$f_{\mathcal{A}^{(1)}}(e_1, t_1, \ldots, e_k, t_k) = f_{\mathcal{A}^{(2)}}(e_1, t_1, \ldots, e_k, t_k)$$

for all $k > 0$, $e_i \in \mathcal{E}_s$ and $t_i > 0$.

The observable events of the components are stochastically identical.

Observed equivalence

$\mathcal{A}^{(1)} \sim \mathcal{A}^{(2)}$: Two components $\mathcal{A}^{(1)}$ of size m and $\mathcal{A}^{(2)}$ of size $n < m$ are observed equivalent if a matrix $\mathbf{V}$ of size $m \times n$ exists such that

- $\mathbf{V}\mathbb{1}_n = \mathbb{1}_m$, $\pi^{(1)}\mathbf{V} = \pi^{(2)}$,
- $\mathbf{R}^{(1)}\mathbf{V} = \mathbf{V}\mathbf{R}^{(2)}$ and $\mathbf{E}_e^{(1)}\mathbf{V} = \mathbf{V}\mathbf{E}_e^{(2)}$ for $\forall e \in \mathcal{S}$

Then

$$
\begin{aligned}
& f_{\mathcal{A}^{(1)}}((e_1, t_1, \ldots, e_k, t_k) && = \\
& \pi^{(1)} \left(\prod_{i=1}^{k} \sum_{j=0}^{\infty} \frac{(\mathbf{R}^{(1)} t_i)^j}{j!} \lambda_{e_i} \mathbf{E}_{e_i}^{(1)} \right) \mathbb{1}_m && = \\
& \pi^{(1)} \left(\prod_{i=1}^{k} \sum_{j=0}^{\infty} \frac{(\mathbf{R}^{(1)} t_i)^j}{j!} \lambda_{e_i} \mathbf{E}_{e_i}^{(1)} \right) \mathbf{V}\mathbb{1}_n && = \\
& \pi^{(1)}\mathbf{V} \left(\prod_{i=1}^{k} \sum_{j=0}^{\infty} \frac{(\mathbf{R}^{(2)} t_i)^j}{j!} \lambda_{e_i} \mathbf{E}_{e_i}^{(2)} \right) \mathbb{1}_n && = \\
& f_{\mathcal{A}^{(2)}}((e_1, t_1, \ldots, e_k, t_k) \ .
\end{aligned}
$$

Compositional equivalence

Unfortunately, synchronized composition relates the internal (non-observable) structures of the components.

$\Downarrow$

Observed equivalence is not enough for the equivalence of the composed models in case of synchronized composition.

$\mathcal{A}^{(1)} \simeq \mathcal{A}^{(2)}$: Two components $\mathcal{A}^{(1)}$ of size m and $\mathcal{A}^{(2)}$ of size $n < m$ are compositional equivalent if a matrix $\mathbf{V}$ of size $m \times n$ exists such that

- $\mathbf{V}\mathbb{1}_n = \mathbb{1}_m$, $\pi^{(1)}\mathbf{V} = \pi^{(2)}$,
- $\mathbf{R}^{(1)}\mathbf{V} = \mathbf{V}\mathbf{R}^{(2)}$, $\mathbf{E}_e^{(1)}\mathbf{V} = \mathbf{V}\mathbf{E}_e^{(2)}$ for $\forall e \in \mathcal{S}$ and
- $\mathbf{D}_e^{(1)}\mathbf{V} = \mathbf{V}\mathbf{D}_e^{(2)}$ for $\forall e \in \mathcal{C}$.

$\mathcal{A}^{(1)} \simeq \mathcal{A}^{(2)}$ implicitly depends on $\mathcal{C}$!!!!

Congruence of composition equivalence

Main Theorem:

If $\mathcal{A}^{(1)} \simeq \mathcal{A}^{(2)}$,

then $\mathcal{A}^{(1)} \|_{\mathcal{C}} \mathcal{A}^{(3)} \simeq \mathcal{A}^{(2)} \|_{\mathcal{C}} \mathcal{A}^{(3)}$

(and $\mathcal{A}^{(3)} \|_{\mathcal{C}} \mathcal{A}^{(1)} \simeq \mathcal{A}^{(3)} \|_{\mathcal{C}} \mathcal{A}^{(2)}$)

if the same set $\mathcal{C}$ is used.

Core of the proof:

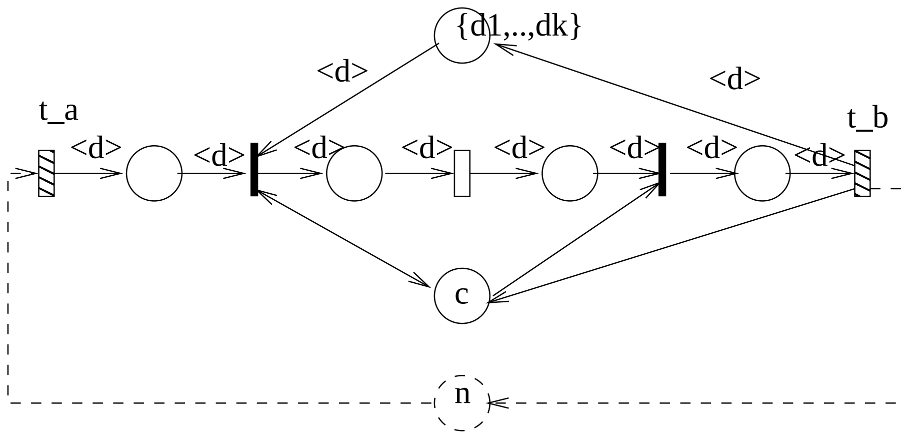
If matrix $\mathbf{V}^{(1,2)}$ relates $\mathcal{A}^{(1)}$ and $\mathcal{A}^{(2)}$

then matrix $\mathbf{V}^{(13,23)} = \mathbf{V}^{(1,2)} \otimes \mathbf{I}_{n^{(3)}}$

relates $\mathcal{A}^{(1)} \|_{\mathcal{C}} \mathcal{A}^{(3)}$ and $\mathcal{A}^{(2)} \|_{\mathcal{C}} \mathcal{A}^{(3)}$.

A disk system

IO system proposed by Balbo, Bruell and Ghanta:



k disks, c channels, n requests

- t_a requests arrival,
- t_1 disk assignment if there is a free channel,
- t_{exp} disk operation,
- t_2 channel allocation,
- t_b data transmission.

A disk system

State space sizes of equivalent representations of the IO system:

Parameters			State space size				
n	k	c	original	ordinary	weak	$\sim$	$\simeq$
4	2	1	59	27	31	27	27
4	2	2	41	23	23	23	23
4	4	1	842	47	61	43	46
4	4	2	444	45	45	43	43
8	2	1	229	101	117	101	101
8	2	2	145	77	77	77	77
8	4	1	15143	541	836	508	524
8	4	2	7779	494	494	433	433
8	6	1	326115	853	1501	738	752
8	6	2	205239	968	971	890	898
8	8	4	444496	530	528	482	482

Conclusions

- Non-exponential $\neq$ non-solvable

  matrix analytic methods

- Parameter estimation, moments matching
  - there are results,
  - but there are also several open problems.

  Non-unique matrix representation.

- Model composition
  - important difference between the internal (micro view) and the external (macro view) transitions

- Efficient simulation ....