Markov-chain Monte Carlo for Phase-type Models

Louis JM Aslett and Simon P Wilson

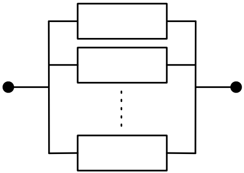
Trinity College Dublin

17$^{\text{th}}$ May 2012

Trinity College Dublin

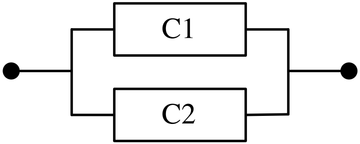
Motivation in Reliability Theory

- Simplest situation: single component modelled with lifetime distribution.
- Redundant collection of components: e.g. components in parallel.



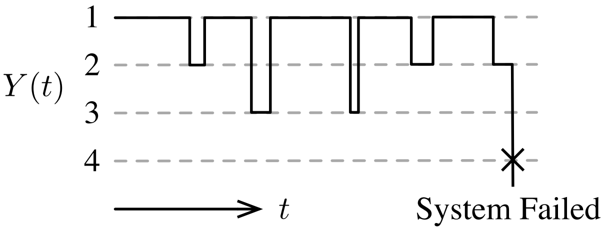
  Often assume no repair. Once component goes down, it stays down.
- Repairable redundant collection of components $\implies$ need to consider a general stochastic process.

Toy Example : Redundant Repairable Components



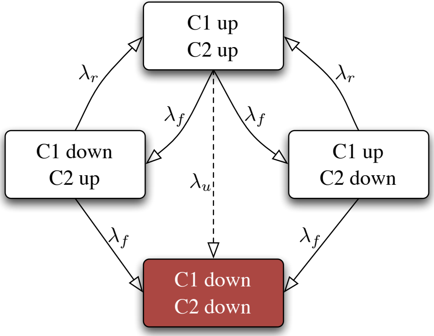
State	Meaning
1	both C1 and C2 work
2	C1 failed, C2 working
3	C1 working, C2 failed
4	system failed

∴ a general stochastic process, e.g.



Continuous-time Markov Chain Model

State	Meaning
1	both C1 and C2 work
2	C1 failed, C2 working
3	C1 working, C2 failed
4	system failed



$$\implies \boldsymbol{\pi} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \mathbf{T} = \begin{pmatrix} -2\lambda_\mathrm{f} & \lambda_\mathrm{f} & \lambda_\mathrm{f} & 0 \\ \lambda_\mathrm{r} & -\lambda_\mathrm{r} - \lambda_\mathrm{f} & 0 & \lambda_\mathrm{f} \\ \lambda_\mathrm{r} & 0 & -\lambda_\mathrm{r} - \lambda_\mathrm{f} & \lambda_\mathrm{f} \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Trinity College Dublin

Inferential Setting

Cano & Rios (2006) provide conjugate posterior calculations in the context of analysing repairable systems when the stochastic process leading to absorption is observed.

Data

For each system failure time, one has:

- Starting state
- Length of time in each state
- Number of transitions between each state
- Ultimate system failure time

Inferential Setting

Cano & Rios (2006) provide conjugate posterior calculations in the context of analysing repairable systems when the stochastic process leading to absorption is observed.

Data
For each system failure time, one has:

- Starting state
- ~~Length of time in each state~~
- ~~Number of transitions between each state~~
- Ultimate system failure time

Reduced information scenario $\implies$ Bladt et al. (2003) provide a Bayesian MCMC algorithm, or Asmussen et al. (1996) provide a frequentist EM algorithm.

Definition of Phase-type Distributions

An absorbing continuous time Markov chain is one in which there is a state that, once entered, is never left. That is, the $n+1$ state intensity matrix can be written:

$$\mathbf{T} = \begin{pmatrix} \mathbf{S} & \mathbf{s} \\ \mathbf{0} & 0 \end{pmatrix}$$

where $\mathbf{S}$ is $n \times n$, $\mathbf{s}$ is $n \times 1$ and $\mathbf{0}$ is $1 \times n$, with

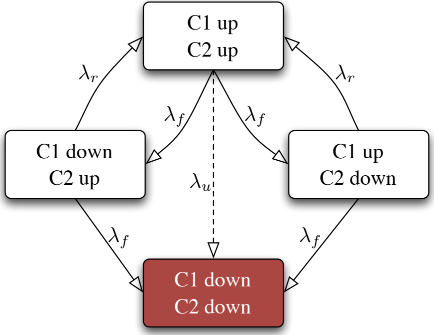
$$\mathbf{s} = -\mathbf{S}\mathbf{e}$$

Then, a *Phase-type distribution* (PHT) is defined to be the distribution of the time to entering the absorbing state.

$$Y \sim \mathrm{PHT}(\boldsymbol{\pi}, \mathbf{S}) \implies \begin{cases} F_Y(y) = 1 - \boldsymbol{\pi}^{\mathrm{T}} \exp\{y\mathbf{S}\}\mathbf{e} \\ f_Y(y) = \boldsymbol{\pi}^{\mathrm{T}} \exp\{y\mathbf{S}\}\mathbf{s} \end{cases}$$

Relating to the Toy Example

State	Meaning
1	both C1 and C2 work
2	C1 failed, C2 working
3	C1 working, C2 failed
4	system failed



$$\implies \mathbf{T} = \left(\begin{array}{ccc|c} -2\lambda_f & \lambda_f & \lambda_f & 0 \\ \lambda_r & -\lambda_r - \lambda_f & 0 & \lambda_f \\ \lambda_r & 0 & -\lambda_r - \lambda_f & \lambda_f \\ \hline 0 & 0 & 0 & 0 \end{array}\right)$$

$$f_Y(y) = \boldsymbol{\pi}^{\mathrm{T}} \exp\{y\mathbf{S}\}\mathbf{s} \qquad F_Y(y) = 1 - \boldsymbol{\pi}^{\mathrm{T}} \exp\{y\mathbf{S}\}\mathbf{e}$$

Bladt et al: Gibbs Sampling from Posterior

Strategy is a Gibbs MCMC algorithm which achieves the goal of simulating from

$$p(\boldsymbol{\pi}, \mathbf{S} \,|\, \mathbf{y})$$

by sampling from

$$p(\boldsymbol{\pi}, \mathbf{S}, \text{paths } \cdot \,|\, \mathbf{y})$$

through the iterative process

$$p(\boldsymbol{\pi}, \mathbf{S} \,|\, \text{paths } \cdot, \mathbf{y})$$

$$p(\text{paths } \cdot \,|\, \boldsymbol{\pi}, \mathbf{S}, \mathbf{y})$$

Bladt et al: Metropolis-Hastings Simulation of Process

In summary:

- Can simulate chain from

$$p(\text{path } \cdot \mid Y_i \geq y_i)$$

  trivially by rejection sampling.
- A Metropolis-Hastings acceptance ratio (ratio of exit rates) exists st truncating chain to time y_i (at which point it absorbs) will be a draw from

$$p(\text{path } \cdot \mid Y_i = y_i)$$

Metropolis-Hastings
$p(\text{path } \cdot \mid \boldsymbol{\pi}, \mathbf{S}, Y = y)$

↑

Rejection Sampling
$p(\text{path } \cdot \mid \boldsymbol{\pi}, \mathbf{S}, Y \geq y)$

↑

CTMC Sampling
$p(\text{path } \cdot \mid \boldsymbol{\pi}, \mathbf{S})$

Motivation for Modifications

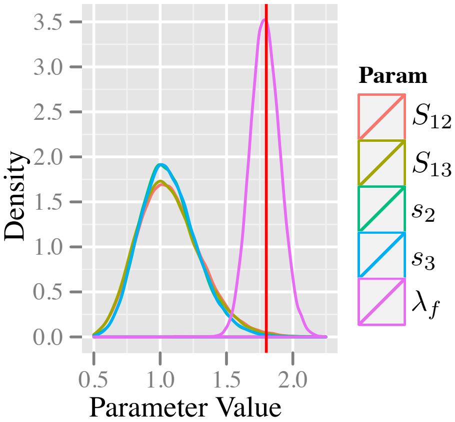
1. Certain state transitions make no physical sense. (eg $2 \to 3$ in earlier example)
2. When part of a larger system, it is highly likely there will be censored observations.
3. Where there is no reason to believe distributional differences between parameters, they should (in idealised modelling sense) be constrained to be equal. This is as much to assist with reducing parameter dimensionality.
4. Computation time!

Toy Example Results

100 uncensored observations simulated from PHT with

$$\mathbf{S} = \begin{pmatrix} -3.6 & 1.8 & 1.8 \\ 9.5 & -11.3 & 0 \\ 9.5 & 0 & -11.3 \end{pmatrix}$$

$$\implies \lambda_f = 1.8, \ \lambda_r = 9.5$$

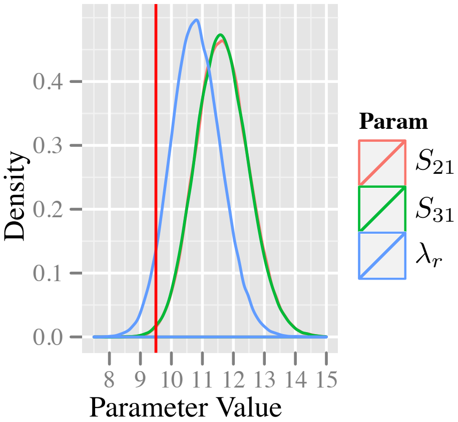


Toy Example Results

100 uncensored observations simulated from PHT with

$$\mathbf{S} = \begin{pmatrix} -3.6 & 1.8 & 1.8 \\ 9.5 & -11.3 & 0 \\ 9.5 & 0 & -11.3 \end{pmatrix}$$

$$\implies \lambda_f = 1.8, \ \lambda_r = 9.5$$



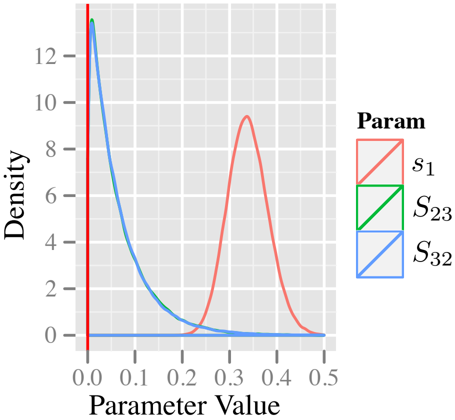
Reliability less sensitive to λ_r (Daneshkhah & Bedford 2008)

Toy Example Results

100 uncensored observations simulated from PHT with

$$\mathbf{S} = \begin{pmatrix} -3.6 & 1.8 & 1.8 \\ 9.5 & -11.3 & 0 \\ 9.5 & 0 & -11.3 \end{pmatrix}$$

$$\implies \lambda_f = 1.8, \ \lambda_r = 9.5$$



The Big Issue

Intractable computation time for many applications!

The Big Issue

Intractable computation time for many applications!

1. longer chains and MCMC jumps to states for which observations are far in the tails can stall rejection sampling step of MH algorithm.

$$\boxed{\mathbb{P}(Y_i \geq y_i \,|\, \boldsymbol{\pi}, \mathbf{S}) = 10^{-6}}$$

$$p(\boldsymbol{\pi}, \mathbf{S} \,|\, \text{paths } \cdot, \mathbf{y})$$

$$p(\text{paths } \cdot \,|\, \boldsymbol{\pi}, \mathbf{S}, \mathbf{y})$$

The Big Issue

Intractable computation time for many applications!

1. longer chains and MCMC jumps to states for which observations are far in the tails can stall rejection sampling step of MH algorithm.

$$\mathbb{P}(Y_i \geq y_i \,|\, \boldsymbol{\pi}, \mathbf{S}) = 10^{-6}$$

$p(\boldsymbol{\pi}, \mathbf{S} \,|\, \text{paths } \cdot, \mathbf{y})$

$p(\text{paths } \cdot \,|\, \boldsymbol{\pi}, \mathbf{S}, \mathbf{y})$

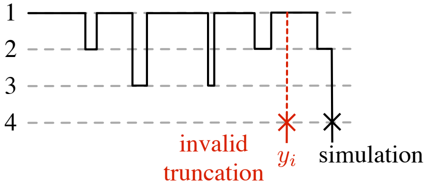
$\mathbb{E}(\text{RS iter.}) = 10^6$

$95\% \text{ CI} = [25317, 3688877]$

The Big Issue

Intractable computation time for many applications!

1. longer chains and MCMC jumps to states for which observations are far in the tails can stall rejection sampling step of MH algorithm.
2. states from which absorption impossible – wasteful to resample whole chain because state at time y_i unsuitable for truncation.



The Big Issue

Intractable computation time for many applications!

1. longer chains and MCMC jumps to states for which observations are far in the tails can stall rejection sampling step of MH algorithm.
2. states from which absorption impossible – wasteful to resample whole chain because state at time y_i unsuitable for truncation.

State	**Meaning**	$\mathbb{P}$(state)
1	both PS working	0.9986
2	1 failed, 2 working	0.0007
3	1 working, 2 failed	0.0007

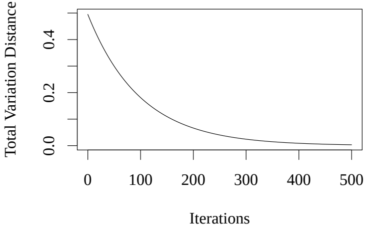
$\Longrightarrow$ $\mathbb{E}(\text{MH iter}) = 1429$

$95\%\ \text{CI} = [36, 5267]$

The Big Issue

Intractable computation time for many applications!

1. longer chains and MCMC jumps to states for which observations are far in the tails can stall rejection sampling step of MH algorithm.
2. states from which absorption impossible – wasteful to resample whole chain because state at time y_i unsuitable for truncation.
3. time for MH algorithm to reach stationarity can grow rapidly.



Solution: "Exact Conditional Sampling"

Metropolis-Hastings

$$p(\text{path} \ \cdot \mid \boldsymbol{\pi}, \mathbf{S}, Y = y)$$

$\uparrow$

Rejection Sampling

$$p(\text{path} \ \cdot \mid \boldsymbol{\pi}, \mathbf{S}, Y \geq y)$$

$\uparrow$

CTMC Sampling

$$p(\text{path} \ \cdot \mid \boldsymbol{\pi}, \mathbf{S})$$

Solution: "Exact Conditional Sampling"

Metropoli[...]
$p(\text{path} \;\cdot\; | \;\boldsymbol{\pi}$ [...]

Rejection [...]
$p(\text{path} \;\cdot\; | \;\boldsymbol{\pi}$ [...]

CTMC S[...]
$p(\text{path} \;\cdot\; | \;\boldsymbol{\pi}, \mathrm{S})$

i) Starting state ~ discrete

$$Y\{0\} \sim \boldsymbol{\pi}$$

ii) Advance time ~ Exponential

$$\delta \sim \mathrm{Exp}(-S_{Y\{t\},Y\{t\}})$$

iii) Select next state ~ discrete

$$Y\{t+\delta\} \sim \{-S_{Y\{t\},-Y\{t\}} / S_{Y\{t\},Y\{t\}}\}$$

iv) If not absorbed, $\quad t = t + \delta$, loop to ii

Solution: "Exact Conditional Sampling"

~~Metropoli~~

~~$p(\text{path} \cdot \mid$~~

~~Rejection~~

~~$p(\text{path} \cdot \mid$~~

CTMC S

$p(\text{path} \cdot \mid \boldsymbol{\pi}, \mathbf{S}, Y = y)$

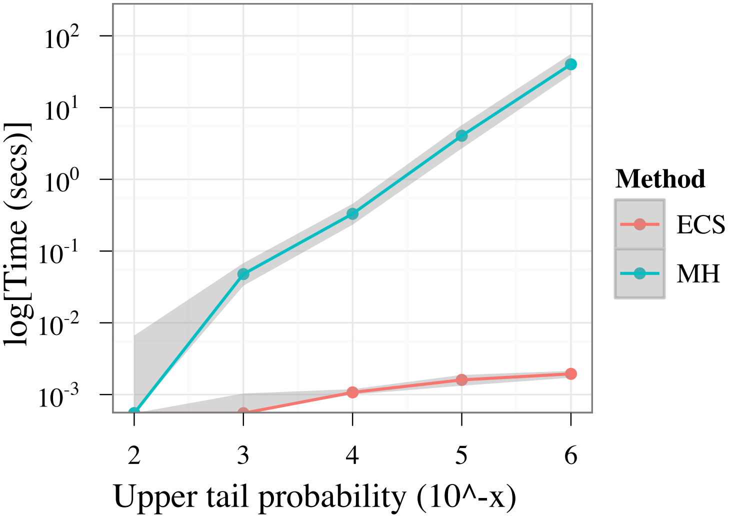
e.g. Starting state mass function changes with conditioning:

$$\mathbb{P}(Y\{0\} = i \mid \boldsymbol{\pi}, \mathbf{S}) = \pi_i$$

$$\downarrow$$

$$\mathbb{P}(Y\{0\} = i \mid \boldsymbol{\pi}, \mathbf{S}, Y = y) = \frac{\mathbf{e}_i^{\mathrm{T}} \exp\{\mathbf{S}y\}\mathbf{s} \; \pi_i}{\boldsymbol{\pi}^{\mathrm{T}} \exp\{\mathbf{S}y\}\mathbf{s}}$$

Tail Depth Performance Improvement



Overall Performance Improvement

This shows the new method keeping pace in 'nice' problems and significantly outperforming otherwise.

$$\mathbf{T} = \begin{pmatrix} -3 & 1 & 1 & 1 \\ 1 & -3 & 1 & 1 \\ 1 & 1 & -3 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

No problems i-iii

	MH	ECS
$\bar{t}$	1.6 µs	7.2 µs
s_t	104 µs	19 µs

$$\mathbf{T} = \begin{pmatrix} -2 & 0.01 & 1.99 & 0 \\ 1 & -300 & 0 & 299 \\ 299 & 0 & -300 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

All problems i-iii

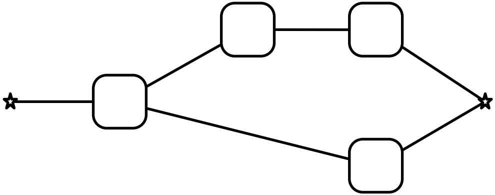
	MH	ECS
$\bar{t}$	10.2 hours	0.016 secs
s_t	9.4 hours	0.015 secs

2,300,000 $\times$ faster on average in hard problem

Trinity College Dublin

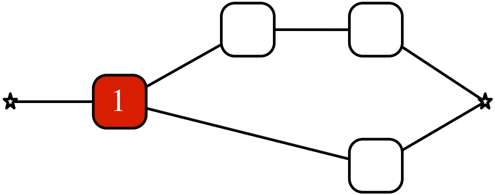
NEW! Networks/Systems of Components

Research focus is now on inference for networks/systems comprising Phase-type nodes/components.



NEW! Networks/Systems of Components

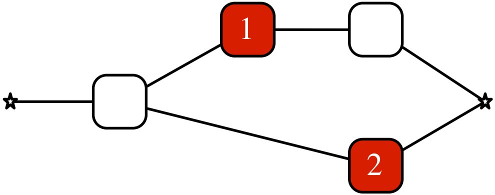
Research focus is now on inference for networks/systems comprising Phase-type nodes/components.



Trinity College Dublin

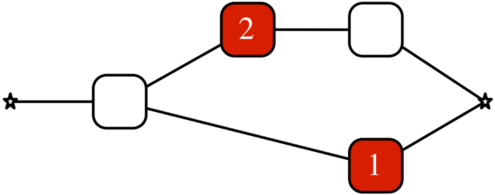
NEW! Networks/Systems of Components

Research focus is now on inference for networks/systems comprising Phase-type nodes/components.



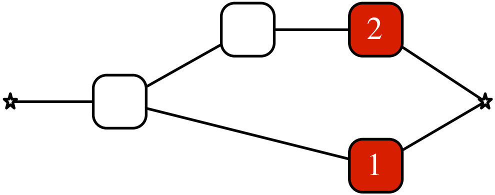
NEW! Networks/Systems of Components

Research focus is now on inference for networks/systems comprising Phase-type nodes/components.



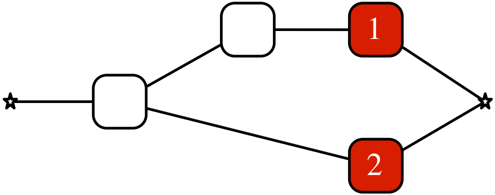
NEW! Networks/Systems of Components

Research focus is now on inference for networks/systems comprising Phase-type nodes/components.



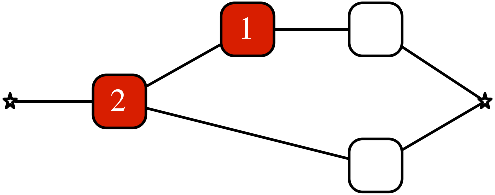
NEW! Networks/Systems of Components

Research focus is now on inference for networks/systems comprising Phase-type nodes/components.



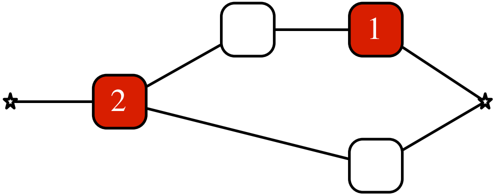
NEW! Networks/Systems of Components

Research focus is now on inference for networks/systems comprising Phase-type nodes/components.



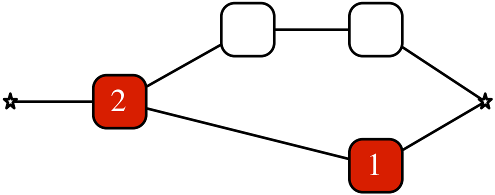
NEW! Networks/Systems of Components

Research focus is now on inference for networks/systems comprising Phase-type nodes/components.



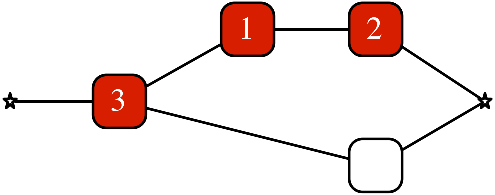
NEW! Networks/Systems of Components

Research focus is now on inference for networks/systems comprising Phase-type nodes/components.



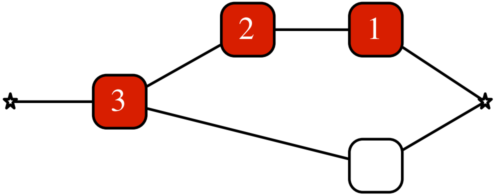
NEW! Networks/Systems of Components

Research focus is now on inference for networks/systems comprising Phase-type nodes/components.



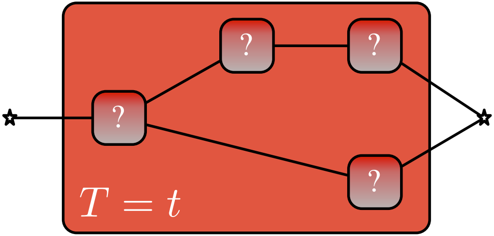
NEW! Networks/Systems of Components

Research focus is now on inference for networks/systems comprising Phase-type nodes/components.



NEW! Networks/Systems of Components

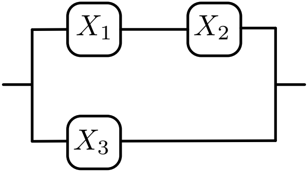
Research focus is now on inference for networks/systems comprising Phase-type nodes/components.



Again, reduced information setting: overall network failure time. ‘Masked system lifetime data’.

Direct Masked System Lifetime Inference

Even ‘simple’ setting quite hard to tackle directly.



$X_i \overset{\text{iid}}{\sim} \text{Weibull}(\text{scale} = \alpha, \text{shape} = \beta)$

System lifetimes $\mathbf{t} = \{t_1, \dots, t_n\}$

$\bar{F}_T(t) = 1 - (1 - \bar{F}_{X_1}(t)\bar{F}_{X_2}(t))(1 - \bar{F}_{X_3}(t))$

$$\implies L(\alpha, \beta; \mathbf{t}) = \prod_{i=1}^{m} t_i^{-1} \beta (t_i/\alpha)^{\beta} \exp\left\{-3(t_i/\alpha)^{\beta}\right\} \left[2 \exp\left\{(t_i/\alpha)^{\beta}\right\} + \exp\left\{2(t_i/\alpha)^{\beta}\right\} - 3\right]$$

$\therefore \quad p(\alpha, \beta \mid \mathbf{t}) \propto L(\alpha, \beta; \mathbf{t}) p(\alpha, \beta)$ awkward.

MCMC Solution (Independent Case)

Proposed solution in the tradition of Tanner & Wong (1987), since inference easy in the presence of (augmented) component lifetimes.
Thus, for $X_i \stackrel{\text{iid}}{\sim} F_X(\,\cdot\,;\psi)$ sample from the natural completion of the posterior distribution:

$$p(\psi, \mathbf{x}_{1\cdot}, \dots, \mathbf{x}_{n\cdot} \,|\, \mathbf{t})$$

by blocked Gibbs sampling using the conditional distributions:

$$p(\mathbf{x}_{1\cdot}, \dots, \mathbf{x}_{n\cdot} \,|\, \psi, \mathbf{t})$$
$$p(\psi \,|\, \mathbf{x}_{1\cdot}, \dots, \mathbf{x}_{n\cdot}, \mathbf{t})$$

where $\mathbf{x}_{i\cdot} = \{x_{i1}, \dots, x_{im}\}$ are the m component failure times for the i^{th} system ($x_{ij} = t_j$ some j)

$p(\psi \,|\, \mathbf{x}_{1\cdot}, \dots, \mathbf{x}_{n\cdot}, \cancel{\mathbf{t}})$ is now simple Bayesian inference for system lifetime distribution — well understood.

Problem shifted to sampling $p(\mathbf{x}_{1\cdot}, \dots, \mathbf{x}_{n\cdot} \,|\, \psi, \mathbf{t})$

Trinity College Dublin

$p(\psi \,|\, \mathbf{x}_{1\cdot}, \dots, \mathbf{x}_{n\cdot}, \cancel{\mathbf{t}})$ is now simple Bayesian inference for system lifetime distribution — well understood.

Problem shifted to sampling $p(\mathbf{x}_{1\cdot}, \dots, \mathbf{x}_{n\cdot} \,|\, \psi, \mathbf{t})$

Propose using system signature $s_j = \mathbb{P}(T = X_{j:n})$

$$\text{e.g. } p(x_{i1}, \dots, x_{im} \,|\, \psi, \mathbf{t}) = \sum_{j=1}^{m} \{ p(x_{i1}, \dots, x_{im} \,|\, \psi, X_{j:n} = t_i) \\ \times \mathbb{P}(T = X_{j:n} \,|\, \psi, t_i) \}$$

where

$$\mathbb{P}(T = X_{j:n} \,|\, \psi, t_i) \propto s_j \binom{m-1}{j-1} F_X(t_i)^{j-1} \bar{F}_X(t_i)^{m-j-1}$$

Algorithm to sample $p(\mathbf{x}_{1\cdot}, \dots, \mathbf{x}_{n\cdot} \mid \psi, \mathbf{t})$

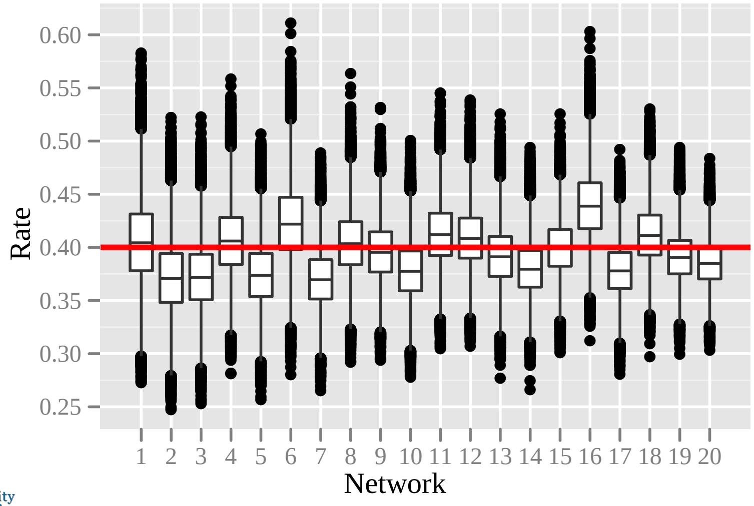
For each system $i = 1, \dots, n$:

1. Sample $j \in \{1, \dots, m\}$ from the discrete probability distribution defined by the conditioned system signature, $\mathbb{P}(T = X_{j:n} \,|\, \psi, t_i)$. This samples the order statistic indicating that the j^{th} failure caused system failure.
2. Sample:
   - $j - 1$ values, $x_{i1}, \dots, x_{i(j-1)}$, from $F_{X \,|\, X<t_i}(\,\cdot\,; \psi)$, the distribution of the component lifetime conditional on failure before t_i
   - $m - j$ values, $x_{i(j+1)}, \dots, x_{in}$, from $F_{X \,|\, X>t_i}(\,\cdot\,; \psi)$, the distribution of the component lifetime conditional on failure after t_i

   and set $x_{ij} = t_i$.

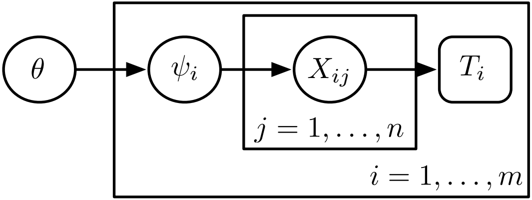
Each iteration provides $\mathbf{x}_{i\cdot}$. ■

All Systems of 4 Components, Exponential ($\lambda = 0.4$)



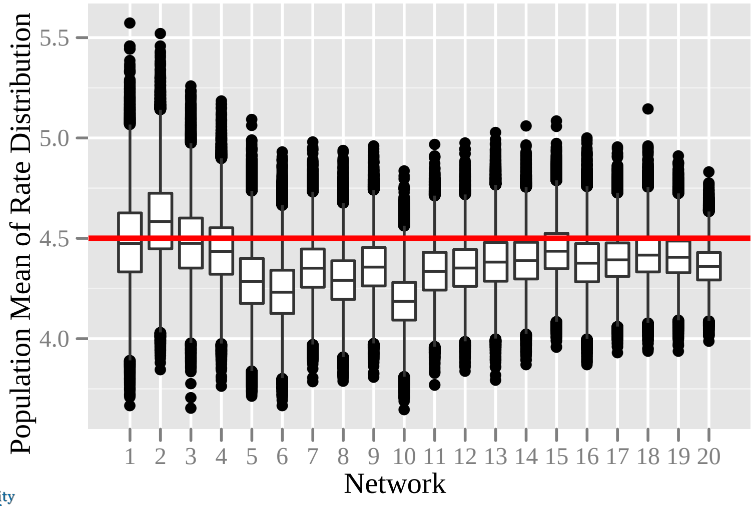
Exchangeable Failure Rate Parameters

It is actually straight-forward to allow the more general setting of exchangeable failure rate parameters between networks:

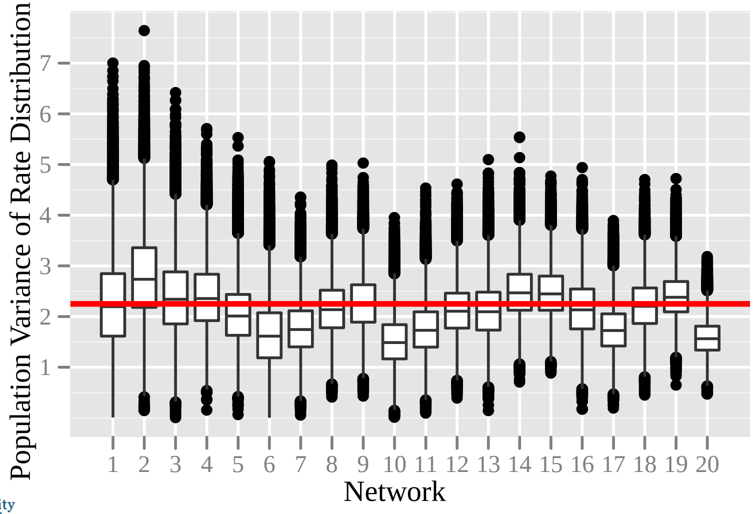


However, exchangeability within network would break the signature-based sampling of node failure times so this is probably as general as this MCMC algorithm can go.

All Systems of 4 Components, $\Psi \sim \text{Gam}(\alpha = 9, \beta = \frac{1}{2})$



All Systems of 4 Components, $\Psi \sim \text{Gam}(\alpha = 9, \beta = \frac{1}{2})$



Asmussen, S., Nerman, O. & Olsson, M. (1996), 'Fitting phase-type distributions via the EM algorithm', *Scand. J. Statist.* **23**(4), 419–441.

Bladt, M., Gonzalez, A. & Lauritzen, S. L. (2003), 'The estimation of phase-type related functionals using Markov chain Monte Carlo methods', *Scand. Actuar. J.* **2003**(4), 280–300.

Cano, J. & Rios, D. (2006), 'Reliability forecasting in complex hardware/software systems', *Proceedings of the First International Conference on Availability, Reliability and Security (ARES'06)* .

Daneshkhah, A. & Bedford, T. (2008), Sensitivity analysis of a reliability system using gaussian processes, *in* T. Bedford, ed., 'Advances in Mathematical Modeling for Reliability', IOS Press, pp. 46–62.

Tanner, M. A. & Wong, W. H. (1987), 'The calculation of posterior distributions by data augmentation', *Journal of the American Statistical Association* **82**(398), 528–540.

Trinity College Dublin