Modelling and Inference for Networks with Repairable Redundant Subsystems

Louis JM Aslett and Simon P Wilson

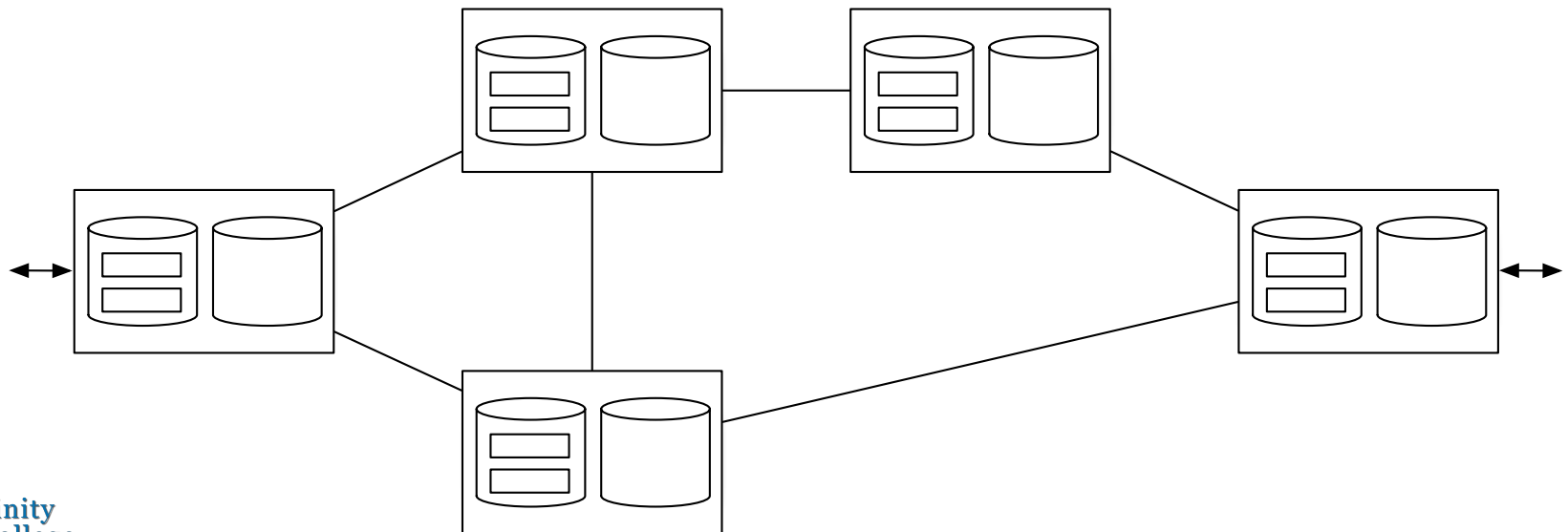
Department of Statistics, Trinity College, University of Dublin

11 February 2010

Problem Specification

The Setting

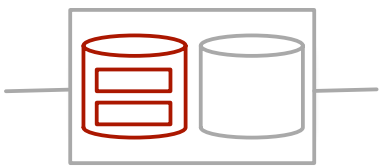
- Network of systems (eg telecoms)
- Each system has critical (repairable/redundant) subsystems
- Varying degrees of data may be available:
  - Only network failure time
  - Individual node failure time
  - Node component failure and repair times



Problem Specification

The Setting

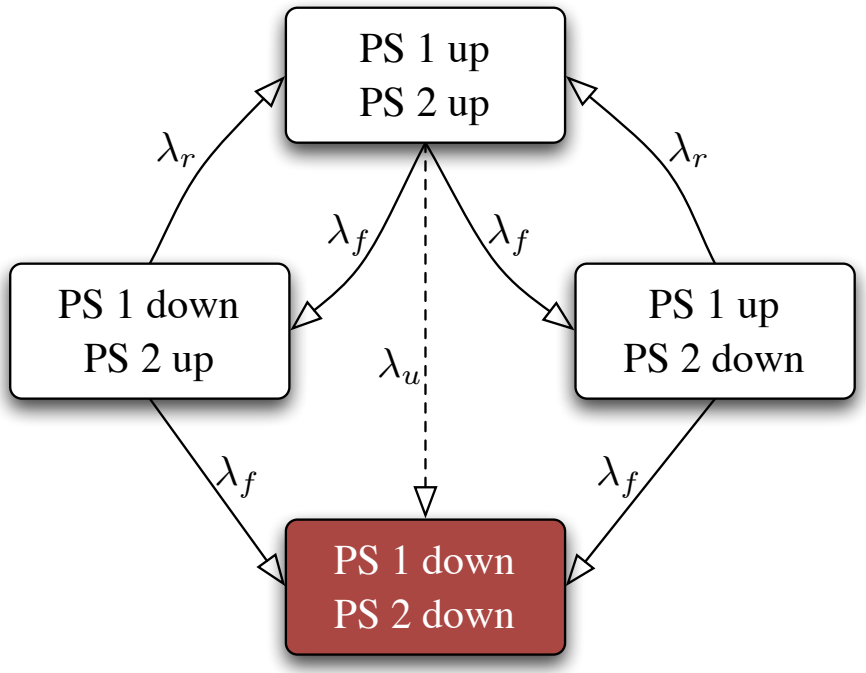
- Network of systems (eg telecoms)
- Each system has critical (repairable/redundant) subsystems
- Varying degrees of data may be available:
  - Only network failure time
  - Individual node failure time
  - Node component failure and repair times



Current focus of attention is inference for repairable redundant subsystems *in this context* $\Longrightarrow$ censoring & highly reliable.

The Simple Example : Dual Redundant Power Supplies

State	Meaning
1	both PS working
2	1 failed, 2 working
3	1 working, 2 failed
4	subsystem failed



Inferential Setting

Cano & Rios (2006) provide conjugate posterior calculations in the context of analysing repairable systems when failure/repair schedule is observed.

Data

For each observed failure, one has:

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

Inferential Setting

Cano & Rios (2006) provide conjugate posterior calculations in the context of analysing repairable systems when failure/repair schedule is observed.

Data

For each observed failure, one has:

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

In absence of schedule of repairs, we require an inferential procedure on the stochastic process in the absence of actually observing that process, only the ultimate failure time.
Bladt et al. (2003) provides a Bayesian MCMC 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

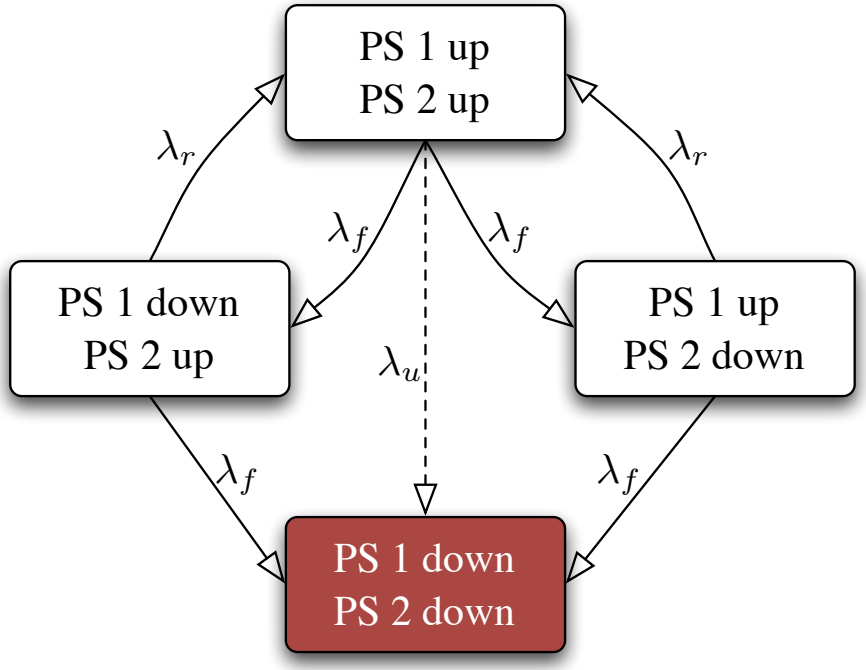
$$\mathbf{s} = -\mathbf{S}\mathbf{e}$$

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

$$X \sim \text{PHT}(\boldsymbol{\pi}, \mathbf{S}) \implies \begin{cases} F_X(x) &= 1 - \boldsymbol{\pi} \exp\{x\mathbf{S}\}\mathbf{e} \\ \\ f_X(x) &= \boldsymbol{\pi} \exp\{x\mathbf{S}\}\mathbf{s} \end{cases}$$

Example

State	Meaning
1	both PS working
2	1 failed, 2 working
3	1 working, 2 failed
4	subsystem failed



$$\Longrightarrow \mathbf{T} = \begin{pmatrix} -2\lambda_\mathrm{f} & \lambda_\mathrm{f} & \lambda_\mathrm{f} & 0 \\ \lambda_\mathrm{r} & -\lambda_\mathrm{r} - \lambda_\mathrm{f} & 0 & \lambda_\mathrm{f} \\ \lambda_\mathrm{r} & 0 & -\lambda_\mathrm{r} - \lambda_\mathrm{f} & \lambda_\mathrm{f} \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

The Method of Bladt et al. (2003)

A Metropolis-Hastings within Gibbs MCMC sampler for inference on the parameters of a phase-type distribution with only absorption times as data.

The Method of Bladt et al. (2003)

A Metropolis-Hastings within Gibbs MCMC sampler for inference on the parameters of a phase-type distribution with only absorption times as data.

Summary of the approach:

- Given
  1. the full stochastic process (all states and sojourns);
  2. a Dirichlet prior for $\boldsymbol{\pi}$;
  3. a gamma prior for each element of $\mathbf{s}$
  4. a gamma prior for each off-diagonal element of $\mathbf{S}$

  the full conditional posteriors are conjugate.
  $\Longrightarrow$ Gibbs sampler.

The Method of Bladt et al. (2003)

A Metropolis-Hastings within Gibbs MCMC sampler for inference on the parameters of a phase-type distribution with only absorption times as data.

Summary of the approach:

- Given
  1. the full stochastic process (all states and sojourns);
  2. a Dirichlet prior for $\boldsymbol{\pi}$;
  3. a gamma prior for each element of $\mathbf{s}$
  4. a gamma prior for each off-diagonal element of $\mathbf{S}$

  the full conditional posteriors are conjugate.
  $\Longrightarrow$ Gibbs sampler.
- A Metropolis-Hastings algorithm can be constructed to simulate the unobserved stochastic process.

Algorithm 1: Metropolis-Hastings Simulation of Process

In summary:

- Can simulate chain from

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

  trivially by rejection sampling.
- A Metropolis-Hastings acceptance ratio (ratio of exit rates) exists such that for sufficient draws, the final chain when truncated to time y_i (at which point it absorbs) will be a draw from

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

Algorithm 2: Gibbs Sampling from Posterior

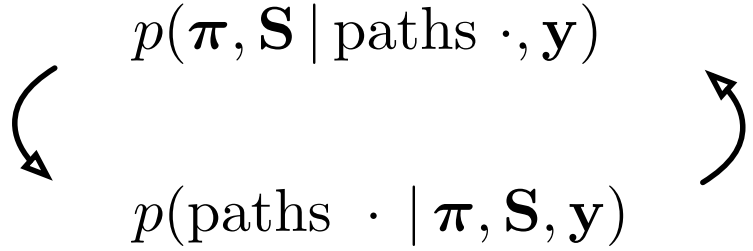
The Gibbs step achieves the goal of simulating from

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

by sampling from

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

through the iterative process



Motivation for Modifications

- Certain state transitions make no physical sense. (eg $2 \rightarrow 3$ in earlier example)

State	Meaning
1	both PS working
2	1 failed, 2 working
3	1 working, 2 failed
4	subsystem failed

$$\implies \mathbf{T} = \begin{pmatrix} -2\lambda_f & \lambda_f & \lambda_f & 0 \\ \lambda_r & -\lambda_r - \lambda_f & 0 & \lambda_f \\ \lambda_r & 0 & -\lambda_r - \lambda_f & \lambda_f \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Motivation for Modifications

- Certain state transitions make no physical sense. (eg $2 \rightarrow 3$ in earlier example)
- When part of a larger system, it is highly likely there will be censored observations.

Motivation for Modifications

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

Motivation for Modifications

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

I. Prohibited Transitions

Straight-forward solution.

To prohibit a transition $i \to j$,

- Do not draw S_{ij} from the prior in algorithm 2.
- Fix S_{ij} at zero when running algorithm 1.
- Do not draw values for S_{ij} from the full conditional posterior in algorithm 2.

Effectively ceasing to consider this entry of $\mathbf{T}$ as a parameter and exclude it from all inferential procedures.

II. Right-censored Observations

The data now consist of actual and censored observations:

$$\mathbf{y} = \{y_1, \ldots, y_m, y^{\mathrm{c}}_{m+1}, \ldots, y^{\mathrm{c}}_n\}$$

To account for censoring:

- In algorithm 1, if y_i is a censored observation, return immediately once rejection sampling from

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

  is complete, ignoring the MH steps. Else, use algorithm 1 as normal.

This effectively incorporates the fact that the unit survived to this time in the likelihood.

III. Constrained Parameter Equality

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

III. Constrained Parameter Equality

$$p(\boldsymbol{\pi}, \mathbf{S} \,|\, \text{paths } \cdot) \propto \left(\prod_{i=1}^{p} \pi_i^{\beta_i - 1} \prod_{i=1}^{p} s_i^{\nu_{i0} - 1} e^{-s_i \zeta_i} \prod_{i=1}^{p} \prod_{\substack{j=1 \\ j \neq i}}^{p} S_{ij}^{\nu_{ij} - 1} e^{-S_{ij} \zeta_i} \right) \times \left(\prod_{i=1}^{p} \pi_i^{B_i} \prod_{i=1}^{p} s_i^{N_{i0}} e^{-s_i Z_i} \prod_{i=1}^{p} \prod_{\substack{j=1 \\ j \neq i}}^{p} S_{ij}^{N_{ij}} e^{-S_{ij} Z_i} \right)$$

III. Constrained Parameter Equality

$$p(\boldsymbol{\pi}, \mathbf{S} \,|\, \text{paths } \cdot) \propto \left(\prod_{i=1}^{p} \pi_i^{\beta_i - 1} \prod_{i=1}^{p} s_i^{\nu_{i0} - 1} e^{-s_i \zeta_i} \prod_{i=1}^{p} \prod_{\substack{j=1 \\ j \neq i}}^{p} S_{ij}^{\nu_{ij} - 1} e^{-S_{ij} \zeta_i} \right) \times \left(\prod_{i=1}^{p} \pi_i^{B_i} \prod_{i=1}^{p} s_i^{N_{i0}} e^{-s_i Z_i} \prod_{i=1}^{p} \prod_{\substack{j=1 \\ j \neq i}}^{p} S_{ij}^{N_{ij}} e^{-S_{ij} Z_i} \right)$$

Example Constraint $\mathcal{C} : S_{12} = S_{13} = s_2 = s_3 = \lambda_{\text{f}}$ results in full conditional posterior for λ_{f} becoming

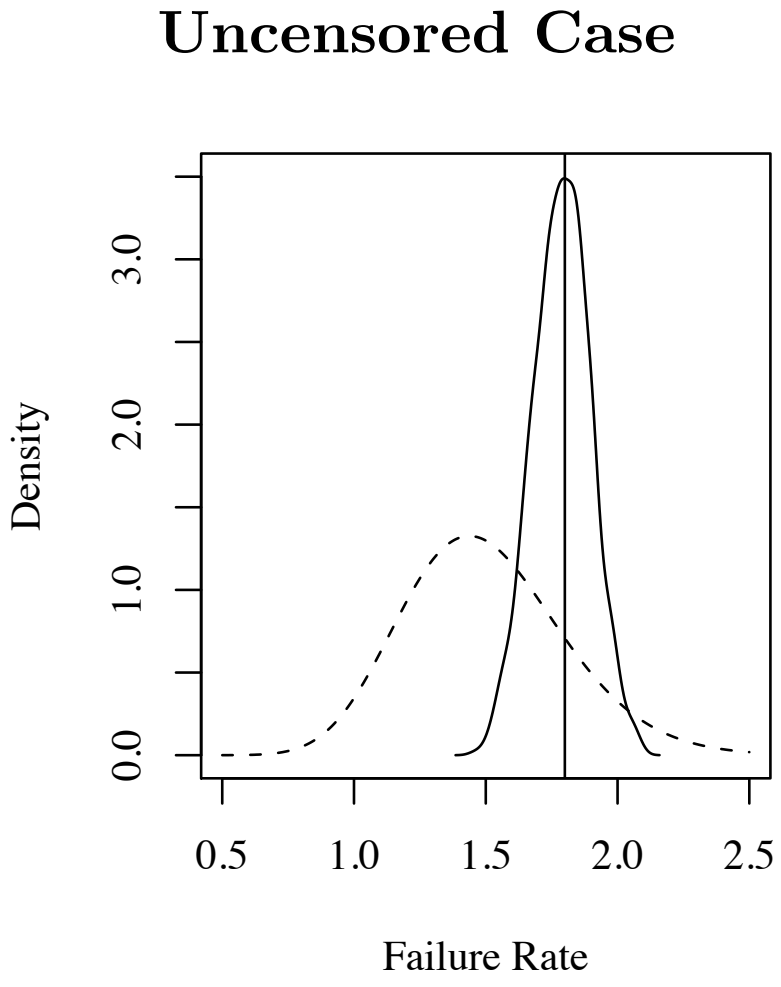
$$p(\lambda_{\text{f}} \,|\, \text{paths } \cdot, \mathcal{C}) \quad \propto \quad \lambda_{\text{f}}^{N_{12} + N_{13} + N_{20} + N_{30} + \nu - 1} e^{\lambda_{\text{f}} (2Z_1 + Z_2 + Z_3 + \zeta)}$$

where ν and ζ were the prior parameters for λ_{f}.

Toy Example Results

100 uncensored observations simulated from PHT with

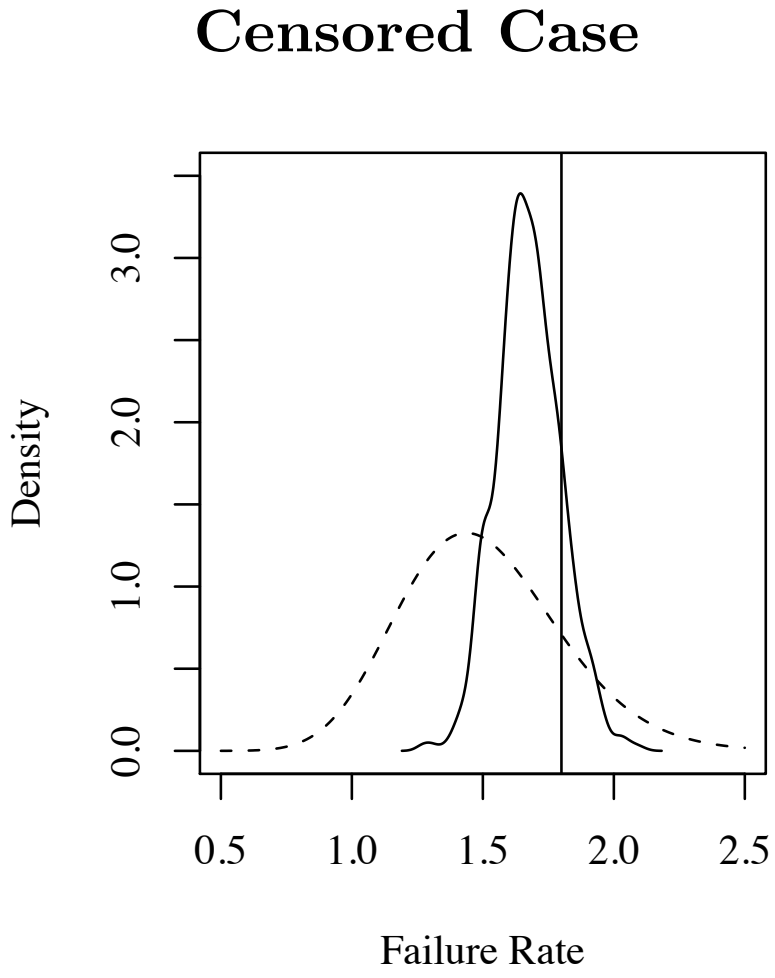
$$\mathbf{S} = \begin{pmatrix} -3.6 & 1.8 & 1.8 \\ 9.5 & -11.3 & 0 \\ 9.5 & 0 & -11.3 \end{pmatrix}$$



Toy Example Results

100 (25%) censored observations simulated from PHT with

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



The Big Issue

Intractable computation time for realistic parameter values!
Obviously any slowdown is from Metropolis-Hastings algorithm.
Realistic values $\Longrightarrow$

The Big Issue

Intractable computation time for realistic parameter values!
Obviously any slowdown is from Metropolis-Hastings algorithm.
Realistic values $\Longrightarrow$

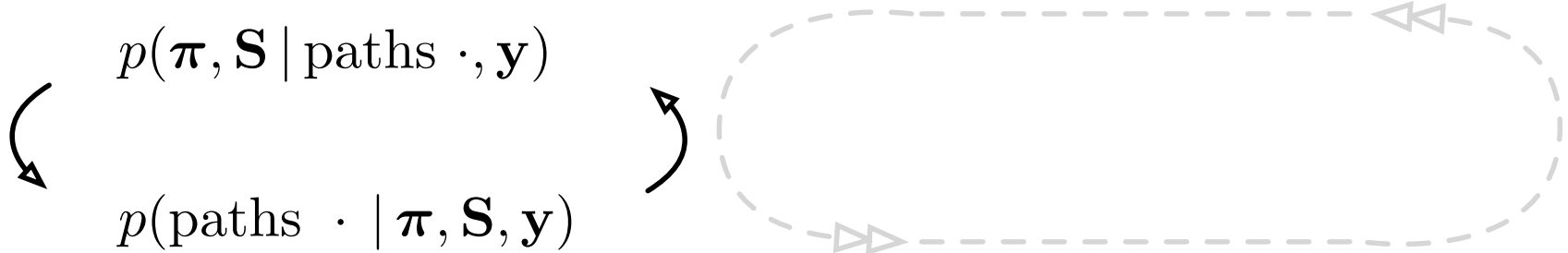
1. longer chains and MCMC jumps to states for which observations are far in the tails stalls rejection sampling step of MH algorithm

The Big Issue

Intractable computation time for realistic parameter values!
Obviously any slowdown is from Metropolis-Hastings algorithm.
Realistic values $\Longrightarrow$

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

$$\mathbb{P}(Y_i \geq y_i \,|\, \text{true params}) = 0.01$$

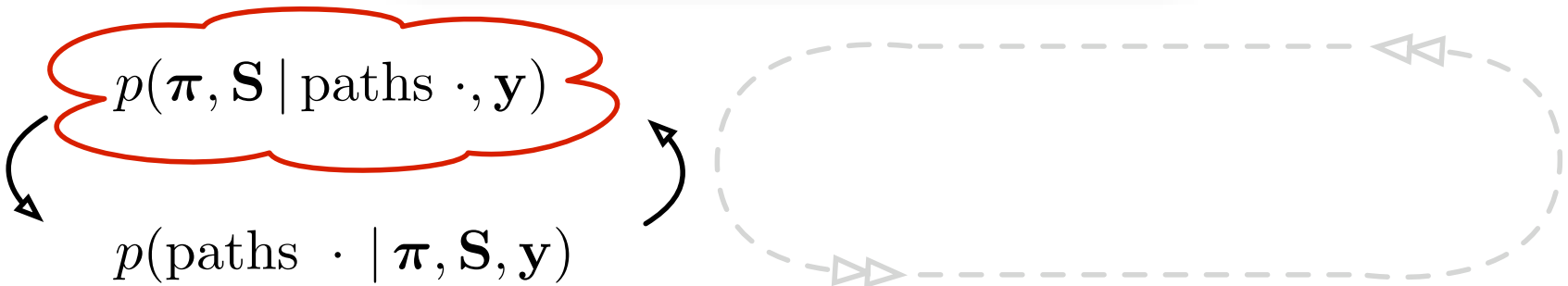


The Big Issue

Intractable computation time for realistic parameter values!
Obviously any slowdown is from Metropolis-Hastings algorithm.
Realistic values $\Longrightarrow$

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

$$\mathbb{P}(Y_i \geq y_i \,|\, \boldsymbol{\pi}, \mathbf{S}) = 0.09$$

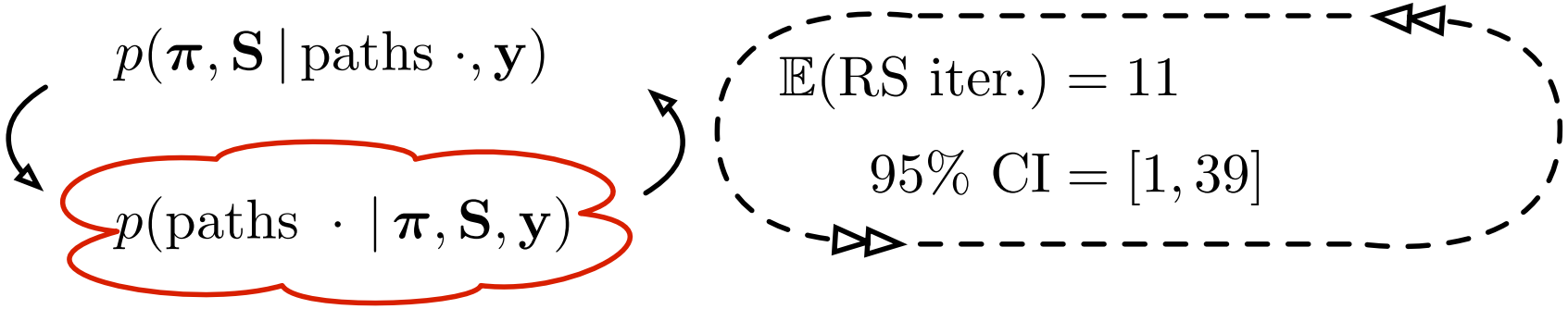


The Big Issue

Intractable computation time for realistic parameter values!
Obviously any slowdown is from Metropolis-Hastings algorithm.
Realistic values $\Longrightarrow$

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

$$\mathbb{P}(Y_i \geq y_i \,|\, \boldsymbol{\pi}, \mathbf{S}) = 0.09$$

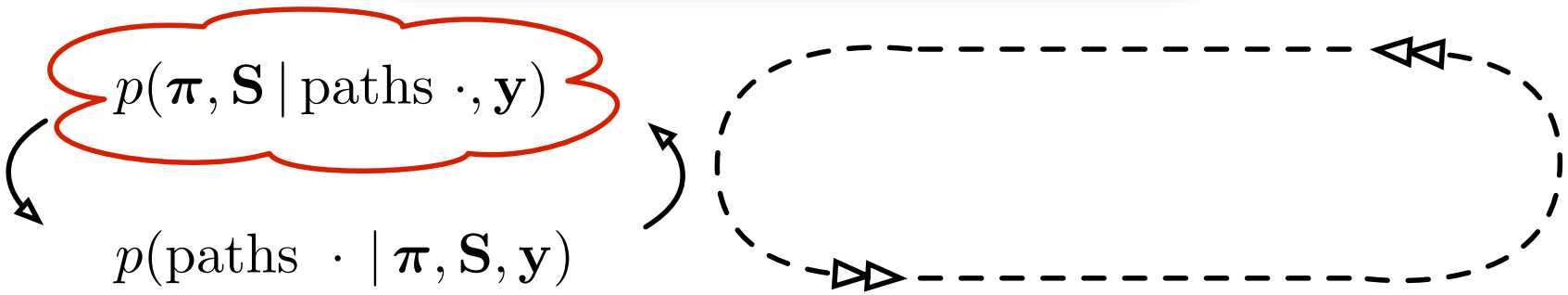


The Big Issue

Intractable computation time for realistic parameter values!
Obviously any slowdown is from Metropolis-Hastings algorithm.
Realistic values $\Longrightarrow$

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

$$\mathbb{P}(Y_i \geq y_i \,|\, \boldsymbol{\pi}, \mathbf{S}) = 0.001$$

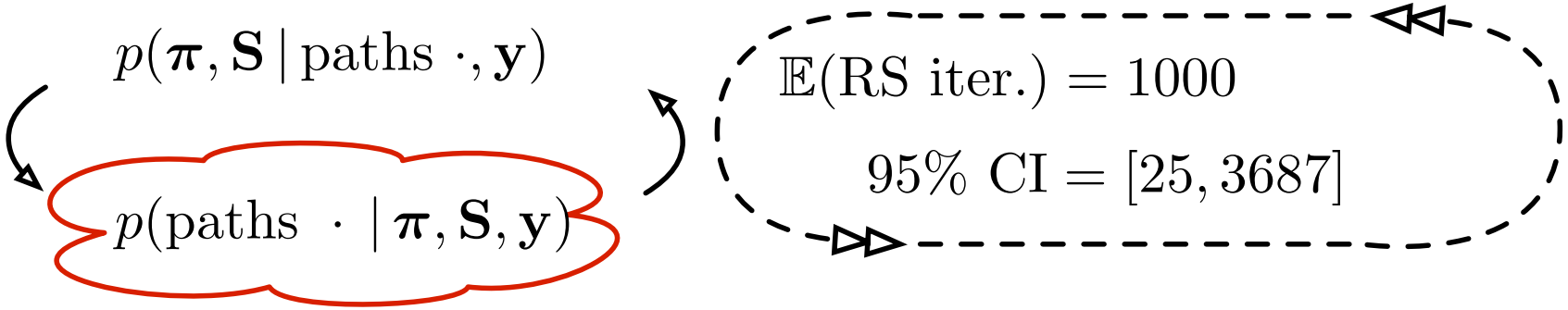


The Big Issue

Intractable computation time for realistic parameter values!
Obviously any slowdown is from Metropolis-Hastings algorithm.
Realistic values $\Longrightarrow$

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

$$\mathbb{P}(Y_i \geq y_i \,|\, \boldsymbol{\pi}, \mathbf{S}) = 0.001$$

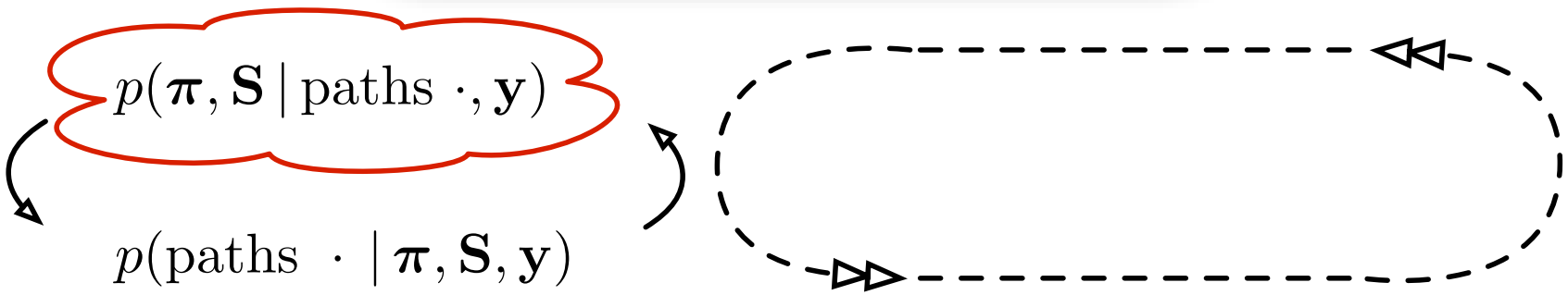


The Big Issue

Intractable computation time for realistic parameter values!
Obviously any slowdown is from Metropolis-Hastings algorithm.
Realistic values $\Longrightarrow$

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

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

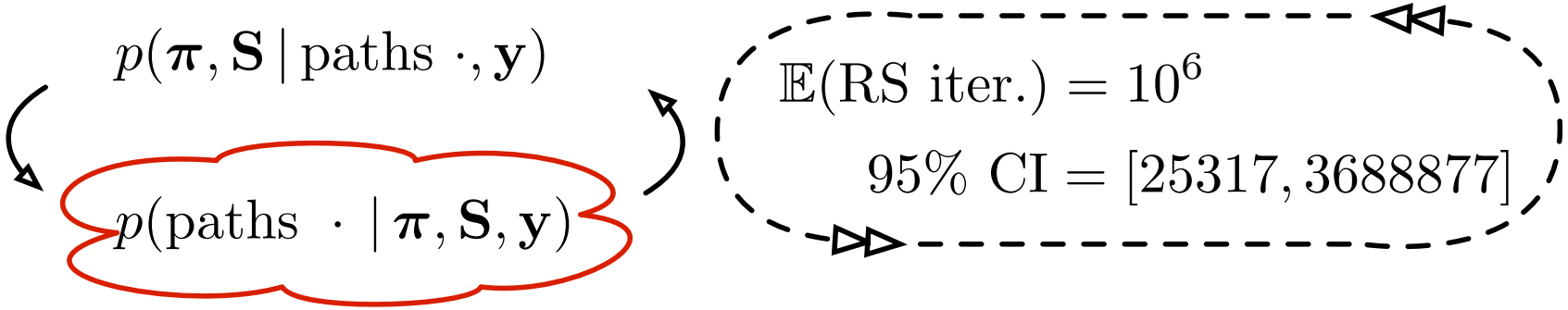


The Big Issue

Intractable computation time for realistic parameter values!
Obviously any slowdown is from Metropolis-Hastings algorithm.
Realistic values $\Longrightarrow$

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

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



The Big Issue

Intractable computation time for realistic parameter values!
Obviously any slowdown is from Metropolis-Hastings algorithm.
Realistic values $\Longrightarrow$

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

The Big Issue

Intractable computation time for realistic parameter values!
Obviously any slowdown is from Metropolis-Hastings algorithm.
Realistic values $\Longrightarrow$

1. longer chains and MCMC jumps to states for which observations are far in the tails stalls rejection sampling step of MH algorithm
2. little time in degraded states – 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% CI $= [36, 5267]$

Proposed Solutions

1. Instead of rejection sampling for $\mathbb{P}(\text{path } \cdot \mid Y_i \geq y_i)$, sample directly conditional upon absorption after time y_i.

   Thus, jump times no longer necessarily be Exponential.

   Guarantees first sample is a sample from $\mathbb{P}(\text{path } \cdot \mid Y_i \geq y_i)$

   Done

Proposed Solutions

1. Instead of rejection sampling for $\mathbb{P}(\text{path } \cdot \,|\, Y_i \geq y_i)$, sample directly conditional upon absorption after time y_i.

   Thus, jump times no longer necessarily be Exponential.

   Guarantees first sample is a sample from $\mathbb{P}(\text{path } \cdot \,|\, Y_i \geq y_i)$

   Done

2. **Nascent Solution:** Exploit reversibility to sample from absorption back to starting time. Since start in operational state and that is the most common state, unlikely to require many outer MH-iterations.

   Exploring quasi-stationary Markov chain theory, since traditionally reversibility requires a chain in stationarity and no guarantee of that here due to absorbing state.

Bladt, M., Gonzalez, A. & Lauritzen, S. L. (2003), 'The estimation of phase-type related functionals using Markov chain Monte Carlo methods', *Scandinavian Journal of Statistics* **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)* .