Markov chain Monte Carlo for Inference on Phase-type Models

Louis JM Aslett (louis@maths.tcd.ie) and Simon P Wilson (swilson@tcd.ie)
Trinity College, University of Dublin

1. The Problem

Phase-type (PHT) distributions are a natural choice for modelling stochastic processes where interest lies in first passage time to an absorbing state.

Bladt et al. (2003) present a scheme for Bayesian inference on general PHT distributions. There are some key areas where there is scope to extend this work when using PHTs for modelling as opposed to distribution fitting:

i) the need to account for censoring;

ii) the need to impose special structures on the Continuous-Time Markov Chain (CTMC) generators parameters;

iii) the sampling scheme is intractably slow for many applications the author's encountered (reliability theory).

2. PHT Distributions

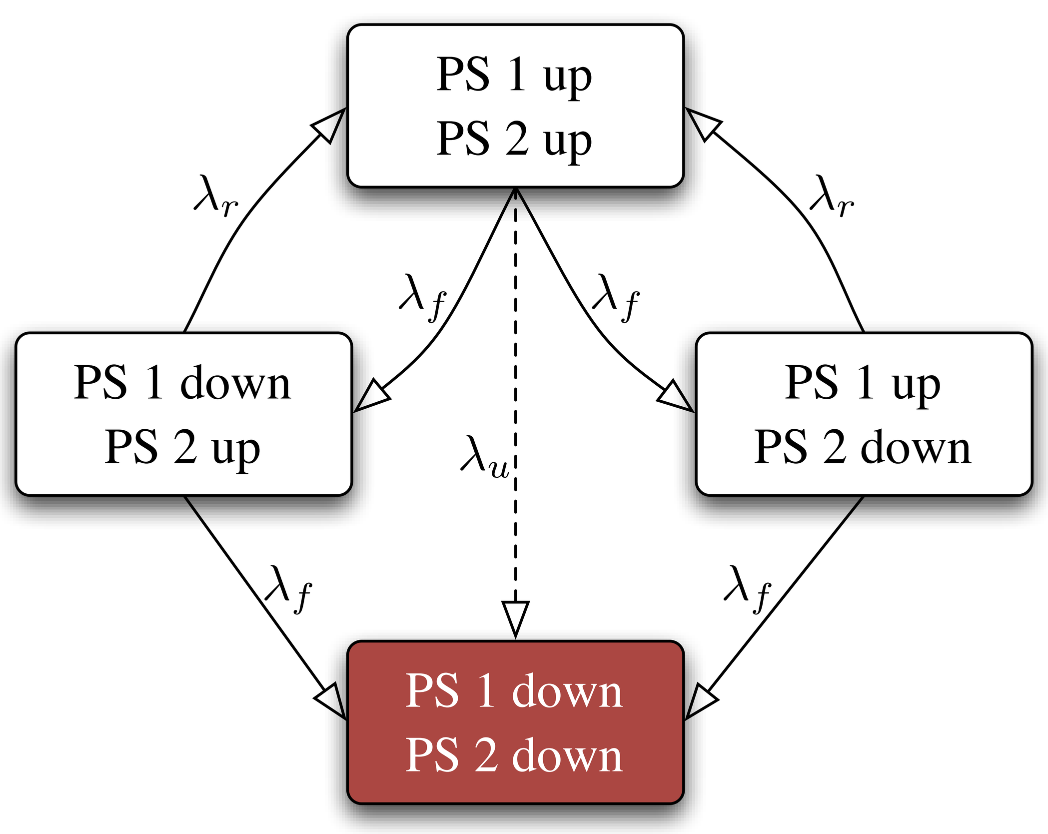
Consider a CTMC with an absorbing state. Without loss of generality, write generator as:

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

Then, if X is the random variable denoting time to entering the absorbing state, $X \sim \mathrm{PHT}(\boldsymbol{\pi}, \mathbf{T})$ and

$$\begin{aligned} F_X(x) &= 1 - \boldsymbol{\pi}^{\mathrm{T}} \exp\{x\mathbf{S}\}\mathbf{e}, \quad \mathbf{e} = (1, \dots, 1)^{\mathrm{T}} \\ f_X(x) &= \boldsymbol{\pi}^{\mathrm{T}} \exp\{x\mathbf{S}\}\mathbf{s} \end{aligned}$$

Simplest Example: (reliability theory)
Consider a dual redundant hot-swappable power supply (PS) subsystem.



λ_r : Repair rate; λ_f : Failure Rate; λ_u : Uncovered Failure Rate (ignore for simple case).

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

3. Bladt et al. (2003) Algorithm

Full stochastic process to absorption observed
$\implies \exists$ conjugate priors $\boldsymbol{\pi} \sim \mathrm{Dir}$; $S_{ij}, s_i \sim \mathrm{Gam}$

Bladt et al. (2003) proposed a Metropolis-Hastings (MH) within Gibbs sampler for the unobserved process case.

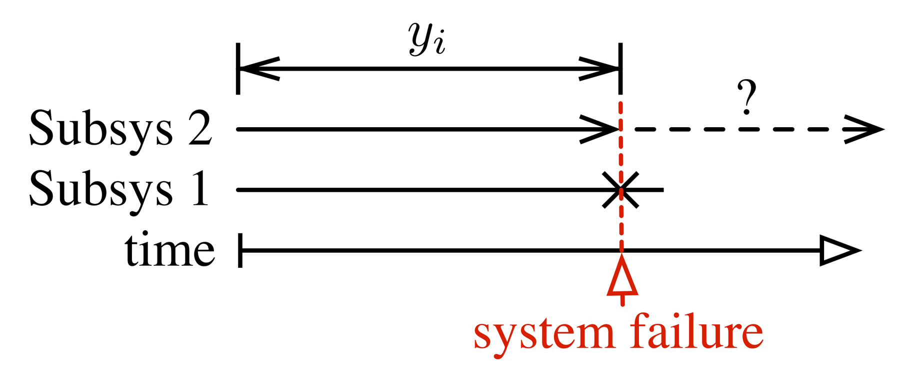
- MH proposal is draw from:
  $$p(\text{path} \cdot \,|\, \boldsymbol{\pi}, \mathbf{S}, Y \geq y_i)$$
  by rejection sampling.

  Acceptance ratio $\implies$ last sample is from
  $$p(\text{path} \cdot \,|\, \boldsymbol{\pi}, \mathbf{S}, Y = y_i)$$
  after truncating to y_i.
- sample from unobserved process in MH step gives conjugacy for Gibbs step.

$$\begin{gathered} p(\boldsymbol{\pi}, \mathbf{S} \,|\, \text{paths} \cdot, \mathbf{y}) \\ \circlearrowleft \quad \circlearrowright \\ p(\text{paths} \cdot \,|\, \boldsymbol{\pi}, \mathbf{S}, \mathbf{y}) \end{gathered}$$

4. Censoring/Constraints

Censoring: arises through competing risks:



Elegantly dealt with by performing just rejection sampling part of MH step. Also, a faster exact sampling method has been developed.

Parameter Constraints: We have shown that Gibbs step conjugacy can be maintained — and indeed, prior specification can be relaxed slightly — when imposing constraints such as:

$$\begin{aligned} \mathcal{C}_1 &: \quad S_{12} = S_{13} = s_2 = s_3 = \lambda_{\mathrm{f}} \\ \mathcal{C}_2 &: \quad S_{23} = 0 \end{aligned}$$

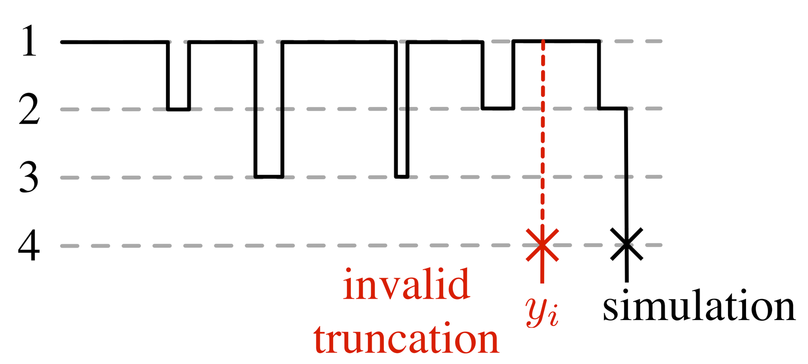
This is desirable for applications and also reduces the dimension of the parameter space.

5. Exact Conditional Sampling of Latent Process

The most significant advance is computational. There are 3 key computational issues with sampling from $p(\text{path} \cdot \,|\, \boldsymbol{\pi}, \mathbf{S}, Y = y_i)$

(i) Exploring parameter space: The Gibbs sampler can explore the space of $(\boldsymbol{\pi}, \mathbf{S})$, selecting parameters of low posterior density for which the Metropolis-Hastings step must then simulate a latent process. This 'tail' exploration can stall the rejection sampler.

(ii) Zero constraints for absorbing moves: With $s_1 = 0$ as in the simple example, there are significant issues with 'invalid' MH proposals: when truncating to time y_i, if the CTMC is in state 1 (highly common) an invalid absorbing move $1 \to 4$ is inserted.

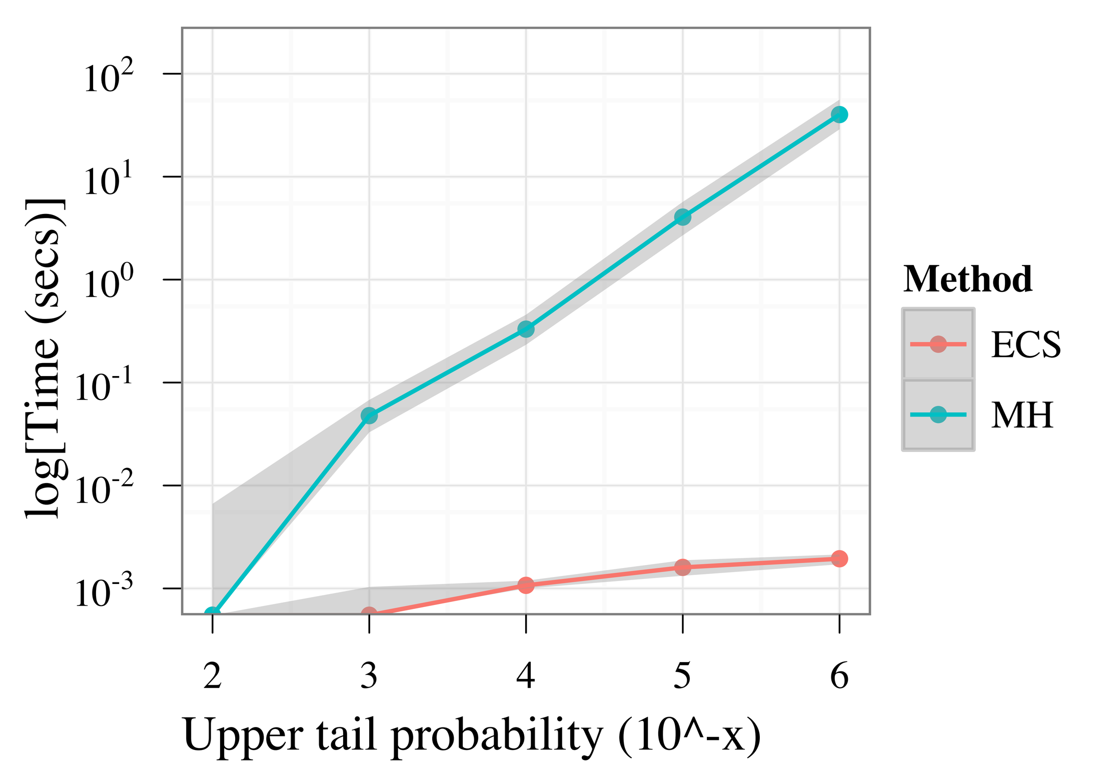


(iii) Reaching stationarity: We have shown for this MH sampler, the total variation distance from stationarity can be large initially, requiring extensive burnin on every outer MCMC iteration. This extensive resampling can compound issues (i) and (ii).

We propose an algorithm to sample the latent process which remedies these issues.

Exact Conditional Sampling (ECS): Our approach is to adapt the standard algorithm for unconditional simulation of an absorbing CTMC by explicitly conditioning on an absorption time $Y = y$. To do so, we derive new expressions from which to sample the starting state, sojourn times and state moves (sadly the poster margin is too small to contain them!)

Tail Depth Speed Comparison: The speed difference due to issue (i) can be graphed. Here $\lambda_f = 478^{-1}, \lambda_r = 39^{-1}$ for the simple example. Shading is bootstrapped 95% CI. ***N.B:*** *log scale*



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

$$\mathbf{T} = \begin{pmatrix} -3 & 1 & 1 & 1 \\ 1 & -3 & 1 & 1 \\ 1 & 1 & -3 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \qquad \mathbf{T} = \begin{pmatrix} -2 & 0.01 & 1.99 & 0 \\ 1 & -300 & 0 & 299 \\ 299 & 0 & -300 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

	No problems i-iii		All problems i-iii	
	MH	ECS	MH	ECS
$\bar{t}$	1.6 μs	7.2 μs	10.2 hours	0.016 secs
s_t	104 μs	19 μs	9.4 hours	0.015 secs

References

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.

Funding

Louis Aslett is an IRCSET postgraduate scholar.