Two-State Imprecise Markov Chains for Statistical Modelling of Two-State Non-Markovian Processes

Matthias C. M. Troffaes — Durham University, UK — matthias.troffaes@durham.ac.uk

Thomas Krak — Gent University, Belgium — thomas.krak@ugent.be

Henna Bains — Durham University, UK — henna.bains@durham.ac.uk

1. Aims

We aim to present a new framework for

- fitting an imprecise two-state Markov chain
- to data generated by a two-state non-Markovian process
- via an imprecise version of MCMC

2. Old approach

1. transition times $T_i \sim \mathrm{Exp}(\lambda_i)$ (i.e. precise Markov chain assumed)
2. set of priors for λ_i
3. combine with data to get set of posteriors for λ_i
4. posterior predictive bounds on λ_i
5. use these bounds to fix an imprecise Markov chain

3. What is wrong with the old approach?

- Model = set of distributions on parameters of a precise Markov chain. Imprecise Markov chains are not equivalent to this model.
- Sampling uncertainty ignored: only posterior predictive bounds are used. No full uncertainty quantification.
- Imprecision does not reflect violations of stationarity & Markovianity
- Non-Bayesian analysis: inferences are not coherent with the model

4. New approach

1. transition times $T_i \sim f_i(\lambda_i)$ where f_i is a non-Markovian and/or non-stationary process and λ_i are unknown parameters of this process
2. set of prior distributions on λ_i
3. for each prior, use standard MCMC to sample posterior realizations of λ_i i.e. fit the non-Markovian process to the data
4. for each posterior sample of λ_i bound process by an imprecise Markov chain
5. produce sample of posterior bounds for any predictive quantity

5. Advantages to the new approach

- a form of imprecise MCMC: imprecision built into the predictive part this works because of Walley's marginal extension theorem
- fully coherent robust Bayesian approach: predictions are directly derived from posterior distribution
- imprecision reflects lack of data & non-Markovianity & non-stationarity (even if there is lot's of data, inferences can remain imprecise)

6. How did we do it?

Step 1

Model non-Markovian two-state process by a many-state Markov chain.
For example (non-exponential transition from state 2 to state 1):

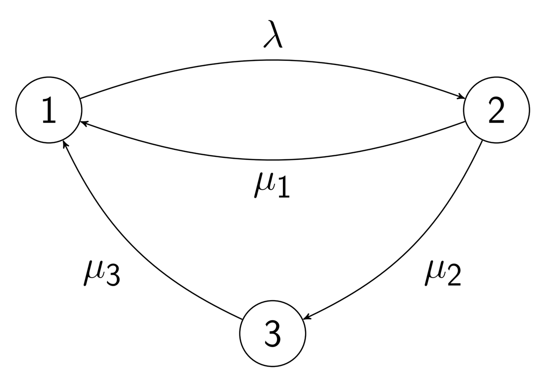


Figure 1: Example of a Markov chain.

with the following transition rate matrix:

$$Q = \begin{bmatrix} -\lambda & \lambda & 0 \\ \mu_1 & -\mu_1 - \mu_2 & \mu_2 \\ \mu_3 & 0 & -\mu_3 \end{bmatrix} \quad (1)$$

Step 2

Fit this many-state Markov chain to the data through its associated phase-type distribution.
How? Use time series data to estimate the parameters λ, μ_1, μ_2 and μ_3.

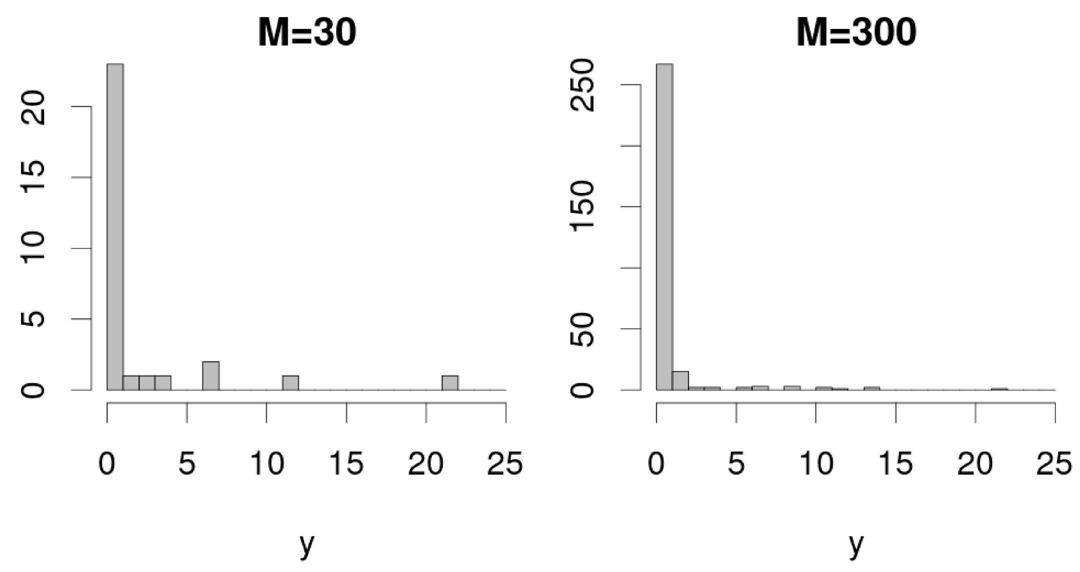


Figure 2: Histogram of times to transition data from state 2 to state 1. The left figure depicts just the first 30 observations, whilst the right figure depicts all 300 observations.

- All parameters fitted using MCMC.
- The model is programmed in Stan.
- Stan conveniently allows direct programming of the phase-type distribution, as well as calculating the limiting lower and upper probabilities for being in state the working state (see step 4).

Step 3

Lump the process to a two-state imprecise Markov chain. Why? Easier to work with.

- We leave state 1 as it is → denoted by α.
- We lump states 2 and 3 together → denoted by β.

$$[\underline{Q}f]_\alpha = \min\{\lambda_*(f_\beta - f_\alpha), \lambda^*(f_\beta - f_\alpha)\}, \quad (2)$$
$$[\underline{Q}f]_\beta = \min\{\mu_*(f_\alpha - f_\beta), \mu^*(f_\alpha - f_\beta)\}. \quad (3)$$

Then, with the λ and μ that achieve the minimum in eqs. (2) and (3):

$$(\lambda_f, \mu_f) \begin{cases} (\lambda_*, \mu^*) & \text{if } f_\alpha \leq f_\beta \\ (\lambda^*, \mu_*) & \text{if } f_\alpha > f_\beta \end{cases} \quad (4)$$

and

$$Q_f := \begin{bmatrix} -\lambda_f & \lambda_f \\ \mu_f & -\mu_f \end{bmatrix} \quad (5)$$

we see that we can interpret our lumped process as one with a precise transition rate

$$\lambda_* = \lambda^* = -Q_{11} \quad (6)$$

for going from state 1 to state 2, and lower and upper transition rates

$$\mu_* = \min_{i=2}^{n} Q_{i1} = \min\{\mu_1, \mu_3\} \qquad \mu^* = \max_{i=2}^{n} Q_{i1} = \max\{\mu_1, \mu_3\} \quad (7)$$

for going from state 2 to state 1.

Step 4

The limiting lower and upper probabilities $(\underline{\pi}, \overline{\pi})$ for being in state 1 i.e. the working state, which are typical inferential quantities of interest:

$$\underline{\pi} := \lim_{t\to\infty} \underline{T}_t \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \frac{\mu_*}{\lambda + \mu_*} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad (8)$$

$$\overline{\pi} := \lim_{t\to\infty} \overline{T}_t \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \frac{\mu^*}{\lambda + \mu^*} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad (9)$$

λ, μ_1, μ_2 and μ_3 are now treated as uncertain quantities, so are $\underline{\pi}$ and $\overline{\pi}$. In particular, given a prior distribution for each of the parameters:

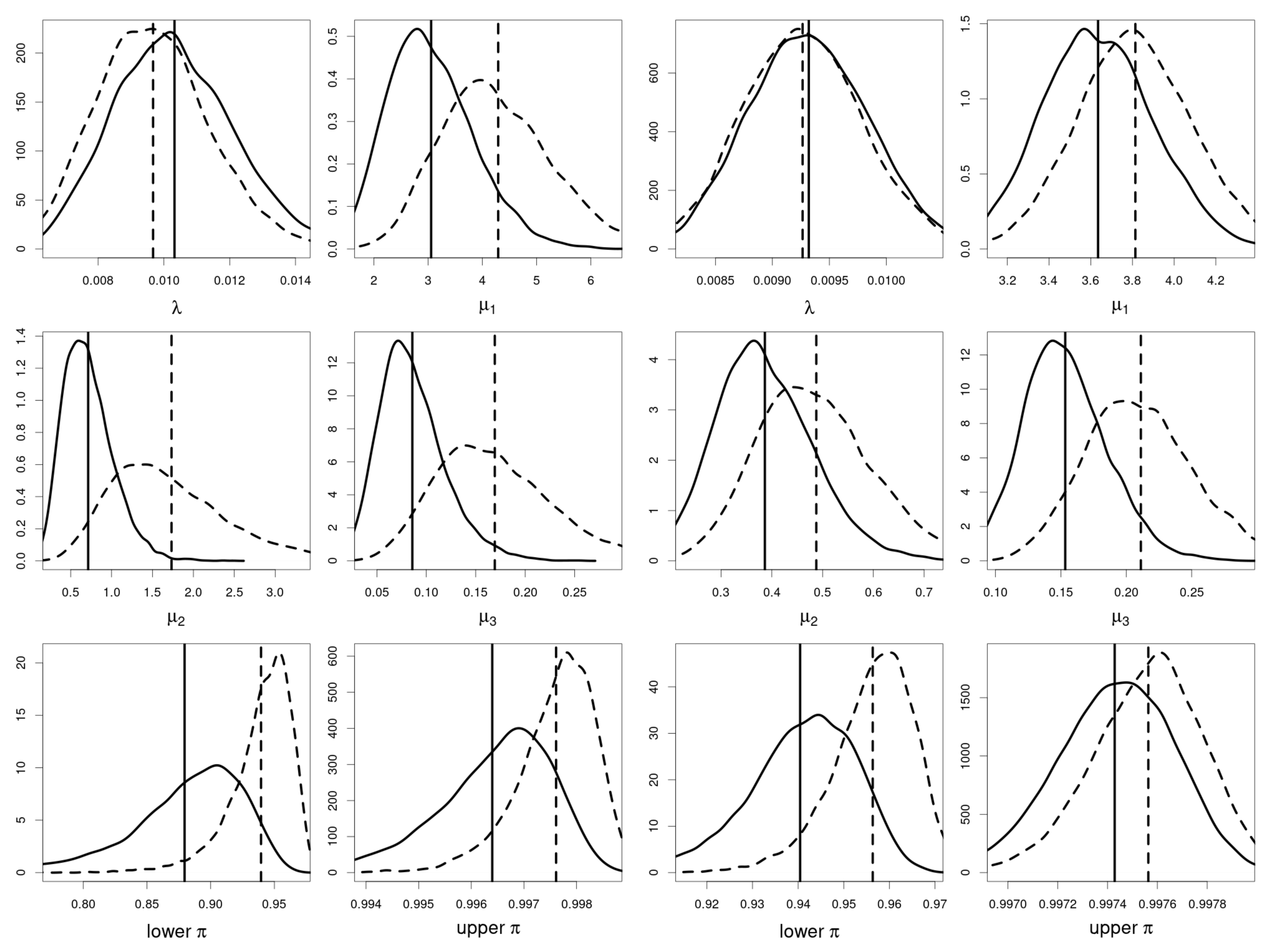
$$\lambda \sim \mathrm{Gamma}(s, s\tau_0) \quad (10)$$
$$\mu_i \sim \mathrm{Gamma}(s, s\tau_i) \qquad i \in \{1, 2, 3\} \quad (11)$$

Aim: Derive a set of posterior distributions for $\underline{\pi}$ and $\overline{\pi}$.
For a specific reliability application, the relevant extreme cases are the ones obtained for

$$(\tau_0, \tau_1, \tau_2, \tau_3) \in \{(50, 1, 4, 40), (250, 0.1, 0.2, 2)\} \quad (12)$$

We simply plot the results under these two priors, as these cover the two relevant extreme cases.

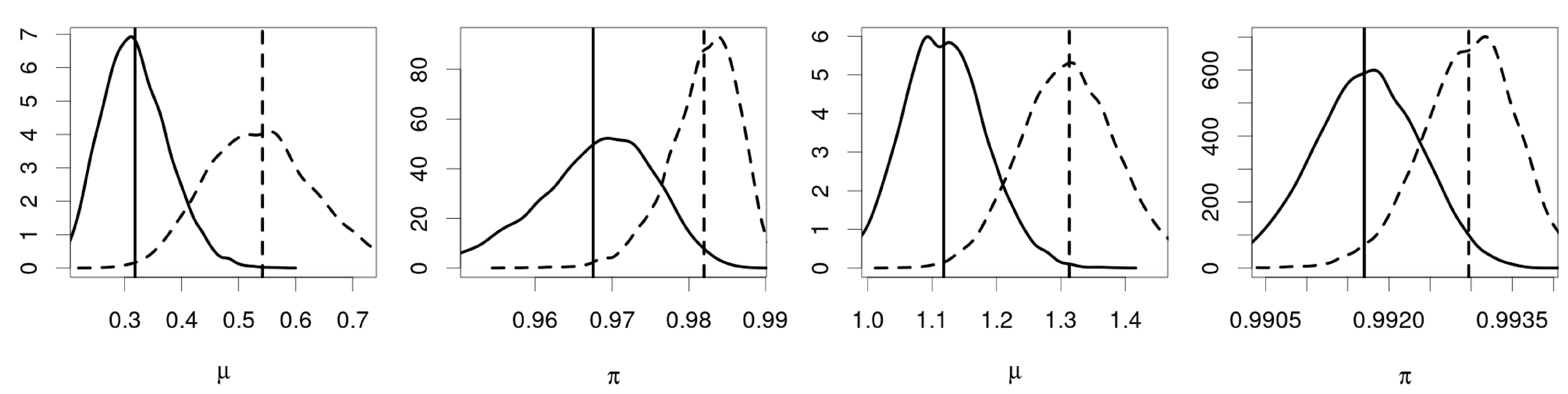
7. Results



(a) For small sample size ($N = M = 30$). (b) For large sample size ($N = M = 300$).

Figure 3: Posterior densities for the lumped model under both priors (solid and dashed line). The vertical lines indicate posterior expectations.

8. Comparison With Two-State Exponential Model



(a) For small sample size ($N = M = 30$). (b) For large sample size ($N = M = 300$).

Figure 4: Posterior densities for the exponential model under both priors (solid and dashed line). The vertical lines indicate posterior expectations.

9. Conclusions

- We have improved the way imprecise Markov chains are fitted
- Imprecision can result not only from limited data but also from characteristics of the process
- These principles may apply to fitting of general stochastic processes (not only imprecise Markov chains)
- We identified a class of problems where imprecise MCMC is easy
- Larger problems may require imprecise ABC if likelihood of phase-type distribution has no closed form

The first two authors are supported by the H2020 Marie Curie ITN, UTOPIAE, Grant Agreement No. 722734. Henna Bains is funded by an Offshore Renewable Energy (ORE) Catapult Doctoral Studentship.

References

[1] Bob Carpenter, Andrew Gelman, Matthew Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. Stan: A probabilistic programming language. *Journal of Statistical Software*, 76(1):1–32, 2017.

[2] Gert de Cooman, Filip Hermans, and Erik Quaeghebeur. Imprecise Markov chains and their limit behavior. *Probability in the Engineering and Informational Sciences*, 23(4):597–635, October 2009.

[3] Alexander Erreygers and Jasper De Bock. Imprecise continuous-time Markov chains: Efficient computational methods with guaranteed error bounds. In Alessandro Antonucci, Giorgio Corani, Inés Couso, and Sébastien Destercke, editors, *Proceedings of the Tenth International Symposium on Imprecise Probability: Theories and Applications*, volume 62 of *Proceedings of Machine Learning Research*, pages 145–156. PMLR, Jul 2017.

[4] Alexander Erreygers and Jasper De Bock. Computing inferences for large-scale continuous-time Markov chains by combining lumping with imprecision. In Sébastien Destercke, Thierry Denoeux, María Ángeles Gil, Przemyslaw Grzegorzewski, and Olgierd Hryniewicz, editors, *Uncertainty Modelling in Data Science*, pages 78–86. Springer International Publishing, 2019.

[5] Thomas Krak, Jasper De Bock, and Arno Siebes. Imprecise continuous-time Markov chains. *International Journal of Approximate Reasoning*, 88:452–528, 2017.

[6] Thomas Krak, Alexander Erreygers, and Jasper De Bock. An imprecise probabilistic estimator for the transition rate matrix of a continuous-time markov chain. In Sébastien Destercke, Thierry Denoeux, María Ángeles Gil, Przemyslaw Grzegorzewski, and Olgierd Hryniewicz, editors, *Uncertainty Modelling in Data Science*, pages 124–132, 2019.

[7] Marcel F. Neuts. *Matrix-geometric solutions in stochastic models: an algorithmic approach*. Dover, 1981.

[8] Lewis Paton, Matthias C. M. Troffaes, Nigel Boatman, Mohamud Hussein, and Andy Hart. Multinomial logistic regression on Markov chains for crop rotation modelling. In Anne Laurent, Oliver Strauss, Bernadette Bouchon-Meunier, and Ronald R. Yager, editors, *Proceedings of the 15th International Conference IPMU 2014 (Information Processing and Management of Uncertainty in Knowledge-Based Systems, 15–19 July 2014, Montpellier, France)*, volume 444 of *Communications in Computer and Information Science*, pages 476–485. Springer, 2014.

[9] Matthias Troffaes, Jacob Gledhill, Damjan Škulj, and Simon Blake. Using imprecise continuous time Markov chains for assessing the reliability of power networks with common cause failure and non-immediate repair. In Thomas Augustin, Serena Doria, Enrique Miranda, and Erik Quaeghebeur, editors, *ISIPTA'15: Proceedings of the 9th International Symposium on Imprecise Probability: Theories and Applications*, pages 287–294, Pescara, Italy, July 2015. ARACNE.

[10] Matthias C. M. Troffaes and Simon Blake. A robust data driven approach to quantifying common-cause failure in power networks. In F. Cozman, T. Denœux, S. Destercke, and T. Seidenfeld, editors, *ISIPTA'13: Proceedings of the Eighth International Symposium on Imprecise Probability: Theories and Applications*, pages 311–317, Compiègne, France, July 2013. SIPTA.