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

Matthias C. M. Troffaes MATTHIAS.TROFFAES@DURHAM.AC.UK
Department of Mathematical Sciences, Durham University, UK

Thomas Krak THOMAS.KRAK@UGENT.BE
ELIS – FLip, Gent University, Belgium

Henna Bains HENNA.BAINS@DURHAM.AC.UK
Department of Engineering, Durham University, UK

Abstract

This paper proposes a method for fitting a two-state imprecise Markov chain to time series data from a two-state non-Markovian process. Such non-Markovian processes are common in practical applications. We focus on how to fit modelling parameters based on data from a process where time to transition is not exponentially distributed, thereby violating the Markov assumption. We do so by first fitting a many-state (i.e. having more than two states) Markov chain to the data, through its associated phase-type distribution. Then, we lump the process to a two-state imprecise Markov chain. In practical applications, a two-state imprecise Markov chain might be more convenient than a many-state Markov chain, as we have closed analytic expressions for typical quantities of interest (including the lower and upper expectation of any function of the state at any point in time). A numerical example demonstrates how the entire inference process (fitting and prediction) can be done using Markov chain Monte Carlo, for a given set of prior distributions on the parameters. In particular, we numerically identify the set of posterior densities and posterior lower and upper expectations on all model parameters and predictive quantities. We compare our inferences under a range of sample sizes and model assumptions.

Keywords: imprecise Markov chain, estimation, reliability, Markov assumption, MCMC

1. Introduction

Imprecise Markov chains provide a popular model for stochastic processes under severe uncertainty [13, 28, 7, 5, 21, 7, 6, 29, 27, 26, 18, 20, 24, 8, 10, 14, 15, 11, 16]. From a sensitivity analysis point of view, imprecise Markov chains model a set of stochastic processes, where only the bounds of this set are required to satisfy the Markov and stationarity conditions. Whilst one might think that computing with such a wide range of models is difficult, it turns out that computing expectation bounds is often extremely efficient, and nearly as easy as working with standard Markov chains [7]. This property makes imprecise Markov chains especially attractive for practical applications. Indeed, Markov and stationarity assumptions are frequently made for computational and conceptual ease, even in the face of clear violations of such assumptions [31].

Two-state imprecise Markov chains are particularly attractive as they allow closed analytic expressions for many quantities of interest. For instance, they admit a closed analytic expression for the expectation of any function of the state at any fixed point in time. Moreover, in practical applications, very often we are interested in stochastic processes that can be in one of two states, but where the Markov and stationarity conditions are violated to some extent.

In this paper, we demonstrate how imprecise two-state continuous time Markov chains can be fitted to time series data to model stochastic processes that violate the Markov assumption. We do so by combining two well-known techniques from the literature: the idea that arbitrary distributions of transition times can be modelled using phase-type distributions, which always have a representation as a Markov chain with potentially many states [17], and the idea that we can lump states together to reduce these complicated models back to a simpler two-state model [11]. In doing so, we also gain insight into what shape of processes imprecise Markov chains can model, potentially further opening up new applications for imprecise Markov chains.

We also manage to perform a full uncertainty quantification on the inferences from the fitted imprecise Markov chain, in the sense that the imprecise Markov chain is directly part of the likelihood of the model, and sets of distributions over the parameters of the imprecise Markov chain are fully propagated. This stands in contrast with how imprecise Markov chains have been fitted elsewhere in the literature, where uncertainty in the bounds on the transition rates is normally not propagated [25, 18, 24, 16].

We find that directly fitting a standard two-state Markov chain (precise or imprecise) does not predict system behaviour as correctly as the lumped phase-type model.

© M.C.M. Troffaes, T. Krak & H. Bains.

Moreover, under a standard model, the influence of the prior will vanish as we get more data, and thereby, so will any imprecision in the posterior. The lumped phase-type model on the other hand will always retain some level of imprecision, as this imprecision is a result of structural uncertainty, and not just due to a lack of data.

The paper is organised as follows. Section 2 gives a brief introduction to Markov chains. Section 3 details how to reduce a many-state Markov chain to a two-state imprecise Markov chain, using lumping. Section 4 presents a simple model that serves as an example of this reduction method. This simple model is also used throughout the remainder of the paper. Section 5 explains how to fit the modelling parameters based on time series data. Section 6 presents a fully worked numerical example. Section 7 concludes the paper.

2. Markov Chains

In this section, we present a minimal introduction to Markov chains.

Let $S := \{1, \dots, n\}$ denote a state space, which we assume to be finite. A stochastic process $(X_t)_{t \geq 0}$ is said to be a (homogeneous, continuous time) Markov chain on S if

$$P(X_{t+\delta t} = j \mid X_{t_1} = i_1, \dots, X_{t_k} = i_k, X_t = i) = P(X_{\delta t} = j \mid X_0 = i) \quad (1)$$

for every $0 < \delta t \in \mathbb{R}$, every $0 \leq t_1 < \cdots < t_k < t \in \mathbb{R}$, and every $i_1, \dots, i_k, i, j \in S$. It can be shown that a Markov chain is uniquely determined by its rate matrix $Q \in \mathbb{R}^{n \times n}$:

$$Q_{ij} := \lim_{\delta t \to 0} \frac{P(X_{\delta t} = j \mid X_0 = i) - P(X_0 = j \mid X_0 = i)}{\delta t} \quad (2)$$

The laws of probability imply that the rows of Q need to sum to zero, the diagonal elements need to be non-positive, and all other entries need to be non-negative. The expectation of any function f of X_t, given X_0, is given by:

$$E(f(X_t) \mid X_0 = i) = [T_t f]_i \quad (3)$$

where

$$T_t := \exp(Qt) = \lim_{n \to \infty} \left(I + \frac{t}{n} Q \right)^n \quad (4)$$

For a two-state Markov chain, the rate matrix Q is determined by just two parameters $\lambda \geq 0$ and $\mu \geq 0$:

$$Q = \begin{bmatrix} -\lambda & \lambda \\ \mu & -\mu \end{bmatrix} \quad (5)$$

and we have a closed expression for T_t (see Theorem 2 in the appendix):

$$T_t = I + \tfrac{1 - e^{-(\lambda+\mu)t}}{\lambda+\mu} Q. \quad (6)$$

3. Induced Imprecise Markov Chain By Lumping

Following [11, Sec. 3], we now explain how to induce imprecise Markov chains through lumping.

Imagine that we would like to reduce the state space to $\{1, \dots, m\}$ for some $m < n$. For this purpose, we define a lumping map $\Lambda \colon \{1, \dots, n\} \to \{1, \dots, m\}$, as well as the following non-linear operator $\underline{Q} \colon \mathbb{R}^m \to \mathbb{R}^m$ [11, Sec. 3.3]:

$$[\underline{Q} f]_k := \min_{i \in \Lambda^{-1}(k)} [Q(f \circ \Lambda)]_i \quad (7)$$

for all $f \in \mathbb{R}^m$ and $k \in \{1, \dots, m\}$.

The lumped process $(Y_t)_{t \geq 0}$, defined by $Y_t := \Lambda(X_t)$ then satisfies

$$[\underline{T}_t f]_i \leq E(f(Y_t) \mid Y_0 = i) \leq [\overline{T}_t f]_i \quad (8)$$

where

$$\underline{T}_t := \lim_{r \to \infty} \left(I + \frac{t}{r} \underline{Q} \right)^r \quad (9)$$

and $\overline{T}_t(f) := -\underline{T}_t(-f)$ for every $f \in \mathbb{R}^m$. Whilst the process Y_t is under normal circumstances not Markovian, the expectation bounds in Eq. (8) do correspond to the lower and upper expectation induced by the imprecise Markov chain with lower transition rate operator $\underline{Q}$.

For example, consider a situation where we leave state 1 as it is, and lump all states $\{2, \dots, n\}$ into a single state. This corresponds to the lumping map $\Lambda(1) := 1$ and $\Lambda(2) = \cdots = \Lambda(n) := 2$. As this is the lumping we will use throughout the remainder of the paper, for clarity of exposition, we will denote state 1 of the lumped process by α and state 2 of the lumped process by β. Equation (7) becomes:

$$[\underline{Q} f]_\alpha := Q_{11} f_\alpha + \sum_{j=2}^{n} Q_{1j} f_\beta = -Q_{11}(f_\beta - f_\alpha) \quad (10)$$

$$[\underline{Q} f]_\beta := \min_{i=2}^{n} \left\{ Q_{i1} f_\alpha + \sum_{j=2}^{n} Q_{ij} f_\beta \right\} = \min_{i=2}^{n} Q_{i1}(f_\alpha - f_\beta) \quad (11)$$

Further, in this case, because the reduced state space has only two states, we know that we have a closed analytic expression for the lower transition operator $\underline{T}_t$.

More precisely, any two-state imprecise Markov chain can be described by just 4 parameters, λ_* and λ^* (with $0 \leq \lambda_* \leq \lambda^*$) which bound the transition rate from state α to state β, and μ_* and μ^* (with $0 \leq \mu_* \leq \mu^*$), which bound the transition rate from state β to state α:

$$[\underline{Q} f]_\alpha = \min\{\lambda_*(f_\beta - f_\alpha), \lambda^*(f_\beta - f_\alpha)\}, \quad (12)$$

$$[\underline{Q} f]_\beta = \min\{\mu_*(f_\alpha - f_\beta), \mu^*(f_\alpha - f_\beta)\}. \quad (13)$$

Then, with the λ and μ that achieve the minimum in Eqs. (12) and (13):

$$(\lambda_f, \mu_f) := \begin{cases} (\lambda_*, \mu^*) & \text{if } f_\alpha \le f_\beta \\ (\lambda^*, \mu_*) & \text{if } f_\alpha > f_\beta \end{cases} \tag{14}$$

and

$$Q_f := \begin{bmatrix} -\lambda_f & \lambda_f \\ \mu_f & -\mu_f \end{bmatrix} \tag{15}$$

it can be shown that (see [10] or Theorem 5 in the appendix for a shorter proof)

$$\underline{T}_t f = f + \tfrac{1-e^{-(\lambda_f+\mu_f)t}}{\lambda_f+\mu_f} Q_f f. \tag{16}$$

Comparing Eqs. (10) and (11) with Eqs. (12) and (13), we see that we can interpret our lumped process as one with a precise transition rate

$$\lambda_* = \lambda^* = -Q_{11} \tag{17}$$

for going from state α to state β, and lower and upper transition rates

$$\mu_* = \min_{i=2}^{n} Q_{i1} \qquad\qquad \mu^* = \max_{i=2}^{n} Q_{i1} \tag{18}$$

for going from state β to state α.

In conclusion, provided that the bounds in Eq. (8) are sufficiently tight for our needs, we can use an induced two-state imprecise Markov chain to analytically bound arbitrarily complex many-state Markov chains. In particular, this allows us to analytically bound any two-state process where transitions between states follow far more complex distributions, and specifically, phase-type distributions, which are distributions that arise from transition times between any two states in an arbitrary Markov chain.

4. A Simple Model

In this section, we demonstrate, on a very simple model, how we can use the ideas explained so far. We will study how a three-state Markov chain can be used to model a two-state non-Markovian process, and how the lumping of this three-state Markov chain can bound this two-state process. The analysis presented here carries over easily to models that incorporate many more states to capture non-Markovian behaviour.

Before carrying on with the model, we emphasize that with this approach we often do not have a semantics for the additional states in the many-state model. Their existence is only motivated by obtaining a better fit to the observed transition times. For example, we could compute a stationary distribution that also includes those auxiliary states, however this would not necessarily have any meaning to an end user. For ease of presentation and interpretation, in the model that is presented next, the additional state is physically motivated (and is thereby provided with a semantics), although such interpretation is not strictly necessary.

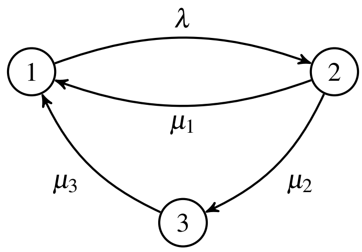

Figure 1: Example of a Markov chain.

Consider the three-state Markov chain depicted in Fig. 1. Here, we could imagine state 1 to correspond to a working state, and states 2 and 3 to correspond to non-working states. The transition from 1 to 2 represents failure. Time to failure in this model is exponential, as is commonly the case in applications. For repair however, we consider two distinct situations. Either, the system can be quickly repaired remotely with mean time $1/\mu_1$, or, the system needs to be serviced locally to fix the problem, at a much slower repair time $1/\mu_2 + 1/\mu_3 \gg 1/\mu_1$. The ratio $\frac{\mu_1}{\mu_1+\mu_2}$ determines how often quick remote repair is successful.

Although here, we clearly have reliability in mind, these many-state Markov models are commonly used for modelling non-exponentially distributed transition times, for instance in human resource planning [31] and medical sciences [2, 3].

Our process has the following transition rate matrix:

$$Q = \begin{bmatrix} -\lambda & \lambda & 0 \\ \mu_1 & -\mu_1-\mu_2 & \mu_2 \\ \mu_3 & 0 & -\mu_3 \end{bmatrix} \tag{19}$$

Time to transition from state 2 to state 1 is no longer exponential, but follows a so-called phase-type distribution, because the process potentially has to pass through state 3 before reaching state 1. The specific phase-type distribution for our model here has a cumulative distribution function equal to [17, Lemma 2.2.2]

$$\Phi_{21}(t) = 1 - \begin{bmatrix} 1 & 0 \end{bmatrix} \exp\left(\begin{bmatrix} -\mu_1-\mu_2 & \mu_2 \\ 0 & -\mu_3 \end{bmatrix} t \right) \begin{bmatrix} 1 \\ 1 \end{bmatrix} \tag{20}$$

and so, by Theorem 1 (see the appendix),

$$= 1 - \left(e^{-(\mu_1+\mu_2)t} + \mu_2 \tfrac{e^{-(\mu_1+\mu_2)t} - e^{-\mu_3 t}}{-\mu_1-\mu_2+\mu_3} \right). \tag{21}$$

This distribution is depicted in Fig. 2. The corresponding probability density function is given by

$$\phi_{21}(t) = \begin{bmatrix} 1 & 0 \end{bmatrix} \exp\left(\begin{bmatrix} -\mu_1-\mu_2 & \mu_2 \\ 0 & -\mu_3 \end{bmatrix} t \right) \begin{bmatrix} \mu_1 \\ \mu_3 \end{bmatrix} \tag{22}$$

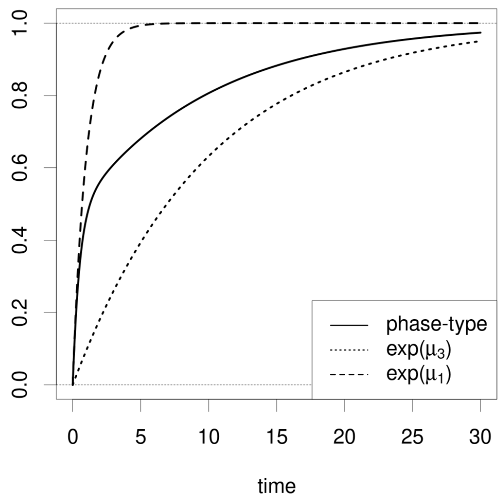

Figure 2: Phase-type distribution for moving from state 2 to state 1, for $\mu_1 = \mu_2 = 1$ and $\mu_3 = 0.1$. Exponential distribution bounds, identified through lumping, are plotted as well.

$$= \mu_1 e^{-(\mu_1+\mu_2)t} + \mu_2\mu_3 \frac{e^{-(\mu_1+\mu_2)t} - e^{-\mu_3 t}}{-\mu_1-\mu_2+\mu_3} \tag{23}$$

For the lumped process, note that

$$Q(f \circ \Lambda) = \begin{bmatrix} \lambda(f_\beta - f_\alpha) \\ \mu_1(f_\alpha - f_\beta) \\ \mu_3(f_\alpha - f_\beta) \end{bmatrix} \tag{24}$$

so,

$$[\underline{Q}f]_\alpha = \lambda(f_\beta - f_\alpha) \tag{25}$$

$$[\underline{Q}f]_\beta = \min\{\mu_1(f_\alpha - f_\beta), \mu_3(f_\alpha - f_\beta)\} \tag{26}$$

We can interpret this as follows: by lumping states 2 and 3, we no longer describe the transition from 2 to 3 (at rate μ_2). If μ_2 is very low, then we will jump to state 1 at rate μ_1. If μ_2 is very high, then we will nearly instantly jump to state 3, to arrive back at state 1 at rate μ_3. Figure 2 confirms that μ_1 and μ_3 bound the possible rates at which we can go from the lumped state back to state 1. Finally, because there is only a single way to jump from state 1 to the lumped state, that rate remains exactly λ.

5. Fitting Parameters

In reality, we do not know the rates λ, μ_1, μ_2, and μ_3. However, we may have some time to failure and time to repair data, and these data can be used to estimate these parameters.

For λ, if we have an i.i.d. sequence $x_1, \ldots, x_N$ of observed times to failure, the maximum likelihood estimate is:

$$\hat{\lambda} = \frac{N}{\sum_{i=1}^N x_i} \tag{27}$$

If we have a Gamma$(s, s\tau)$ prior on λ, where the hyperparameter s determines the strength of the prior, and $1/\tau$ is our prior expectation of λ before having seen the data, then we can use the posterior expectation of λ as an estimate:

$$\hat{\lambda} = \frac{s+N}{s\tau + \sum_{i=1}^N x_i} \tag{28}$$

For full uncertainty quantification, we can propagate the entire posterior distribution:

$$\lambda \mid x_1 \ldots x_N \sim \text{Gamma}\left(s+N, s\tau + \sum_{i=1}^N x_i\right) \tag{29}$$

In either case, we can use intervals for the hyperparameters (for example, as in [24]), in case we prefer to represent our prior knowledge more robustly using a set of prior distributions.

For μ_1, μ_2, and μ_3, given an i.i.d. sequence $y_1, \ldots, y_M$ of observed times to repair, a variety of advanced methods have been developed in the literature in order to fit phase-type distributions [1, 22, 12, 23, 30, 19]. However, in general, there is no closed expression for the maximum likelihood estimate, or even for the posterior distribution. In our case, we know that the likelihood does have a closed form:

$$f(y_1 \ldots y_M \mid \mu_1, \mu_2, \mu_3) = \prod_{i=1}^M \phi_{21}(y_i) \tag{30}$$

where we derived ϕ_{21} in Eq. (23). Using this expression, we can easily obtain (correlated) samples from the posterior joint distribution of (μ_1, μ_2, μ_3) using Markov chain Monte Carlo, as long as the prior also has a simple analytic form (e.g. we could also use Gamma distributions for these). Here too, if we prefer to represent our prior knowledge more robustly, we could use a set of Gamma distributions, and then simply take lower and upper envelopes over different Markov chain Monte Carlo runs to obtain lower and upper expectations. This allows us to do a full uncertainty quantification.

6. Numerical Example

For reasons of confidentiality, here, we use simulated data based on an approximate fit to data from an actual reliability system. This ensures that our example is still somewhat representative of a real reliability system. At the same time, the use of simulated data allows us to explore a wider range of sample sizes, and allows us to investigate how our fits evolve as more data is available.

Specifically, we studied data generated from the Markov chain depicted in Fig. 1, with $\lambda = 0.01$, $\mu_1 = 4$, $\mu_2 = 0.6$, and $\mu_3 = 0.2$. The generated data for y_i is depicted in Fig. 3.

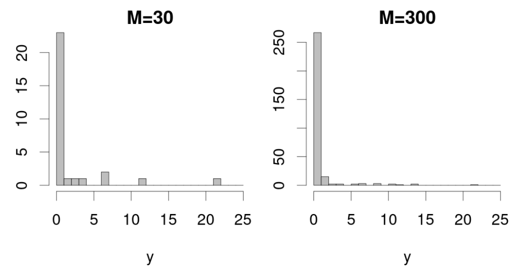

Figure 3: Histogram of times to transition data from state 2 to state 1. The left figure depicts just the first 30 observations, whilst the right figure depicts all 300 observations.

For convenience, all parameters (including λ) were fitted using Markov chain Monte Carlo. The model was programmed in Stan [4] (see Appendix B). The sampler drew 10000 samples from the posterior, of which the first 1000 samples were discarded to allow warmup. For diagnostics, we checked the trace plots, effective sample sizes, and partial auto-correlation plots. The chain was found to mix really well, so for brevity, we have omitted these diagnostics here. Stan conveniently allows direct programming of the phase-type distribution, as well as calculating the limiting lower and upper probabilities for being in state α i.e. the working state, which are typical inferential quantities of interest:

$$\underline{\pi} := \lim_{t\to\infty} \underline{T}_t \begin{bmatrix}1\\0\end{bmatrix} = \frac{\mu_*}{\lambda+\mu_*}\begin{bmatrix}1\\1\end{bmatrix} \tag{31}$$

$$\overline{\pi} := \lim_{t\to\infty} \overline{T}_t \begin{bmatrix}1\\0\end{bmatrix} = \frac{\mu^*}{\lambda+\mu^*}\begin{bmatrix}1\\1\end{bmatrix} \tag{32}$$

where

$$\mu_* := \min\{\mu_1,\mu_3\} \qquad \mu^* := \max\{\mu_1,\mu_3\}. \tag{33}$$

At this point, we emphasize that because 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:

$$\lambda \sim \text{Gamma}(s, s\tau_0) \tag{34}$$

$$\mu_i \sim \text{Gamma}(s, s\tau_i) \qquad i\in\{1,2,3\} \tag{35}$$

our aim is to derive a posterior distribution for $\underline{\pi}$ and $\overline{\pi}$, or, more precisely, a set of posterior distributions, as we will use a set of priors to deal with the lack of prior information.

For our prior, we fix $s=1$. This means that our parameters have, a priori, equal mean and standard deviation, which seems reasonable. The τ_i parameter then represents the mean of the distribution. For our specific reliability application, we deem, a priori, that

$$\tau_0\in[50,250] \quad \tau_1\in[0.1,1] \quad \tau_2\in[0.2,4] \quad \tau_3\in[2,40] \tag{36}$$

Due to the structure of Eqs. (31) and (32), intuitively, we expect that the relevant extreme cases are the ones obtained for

$$(\tau_0,\tau_1,\tau_2,\tau_3)\in\{(50,1,4,40),(250,0.1,0.2,2)\} \tag{37}$$

so, for ease of presentation, we simply plot the results under these two priors, as these cover the two relevant extreme cases within our analysis.

Marginal posterior densities are depicted in Fig. 4, where we had a relatively small sample size ($N=M=30$). We note in particular that μ_2 appears to be the hardest parameter to estimate, compared to the other parameters, both in terms of relative posterior variance and in terms of relative posterior imprecision. This is interesting because, through the lumping, our inferences from the two-state imprecise Markov chain do not depend on μ_2. The posterior distributions are generally skewed. The limiting posterior lower expectation of being in state α is around 0.87 (solid line on bottom left), whereas the posterior upper expectation is around 0.9975 (dashed line on bottom right).

Marginal posterior density plots for a larger sample size ($N=M=300$) are depicted in Fig. 6. Here, the posterior distributions are less skewed, and the influence of the prior has been reduced. The limiting posterior lower expectation of being in state α is around 0.94, whereas the posterior upper expectation is around 0.9976.

For comparison, we also fitted a model where, instead of a phase-type distribution for repair, we simply assumed exponential repair, with $\mu\sim\text{Gamma}(s,s\tau)$ where $\tau\in\{0.1,40\}$ (this roughly corresponds to the range for the parameters of the phase-type model). The marginal posterior distributions for μ and $\pi := \frac{\mu}{\lambda+\mu}$ are depicted in Figs. 5 and 7. We can see that the fit is contained within the fit of the lumping model, as we would expect.

However, the exponential model will obviously not predict system behaviour as accurately as the lumped phase-type model. Further, asymptotically, as we get more data, eventually, the influence of the prior will vanish, and thereby, under the exponential model, the imprecision in the posterior, will vanish. The lumped phase-type model on the other hand will never rid of all imprecision, as there is imprecision as a result of structural uncertainty (which will never vanish), and not just due to a lack of data.

At this point, the reader may wonder why we would not use the three-state model directly to perform our inferences, since we estimate μ_2 anyway. Whilst this could be workable in the simple three-state model, in a much more complicated setting, working with a simpler (fewer state) model can be advantageous, as mentioned earlier. First, the interpretation of the two-state model may be more convenient, as the auxiliary states may not have a direct semantics. Secondly, a two-state model has a computational advantage. For example, in reliability, many thousands of components may need to be considered at once, and in this case, having a simple two-state model can be particularly attractive.

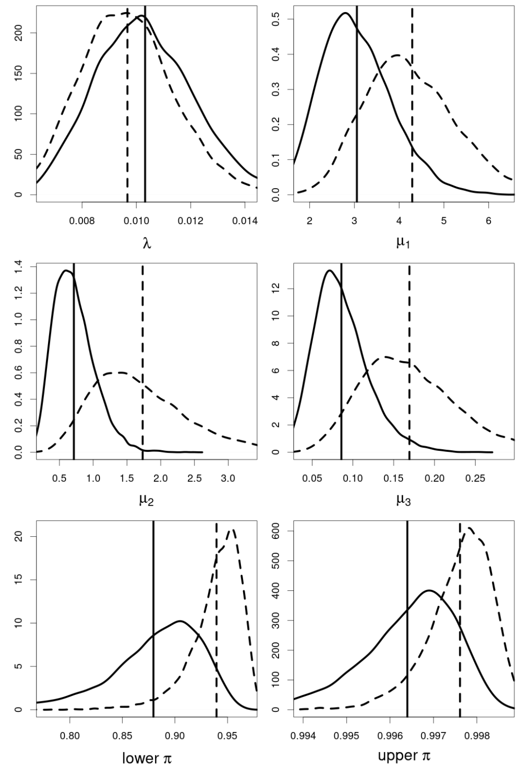

Figure 4: Posterior densities for the lumped model under both priors (solid and dashed line), for small sample size ($N = M = 30$). The vertical lines indicate posterior expectations.

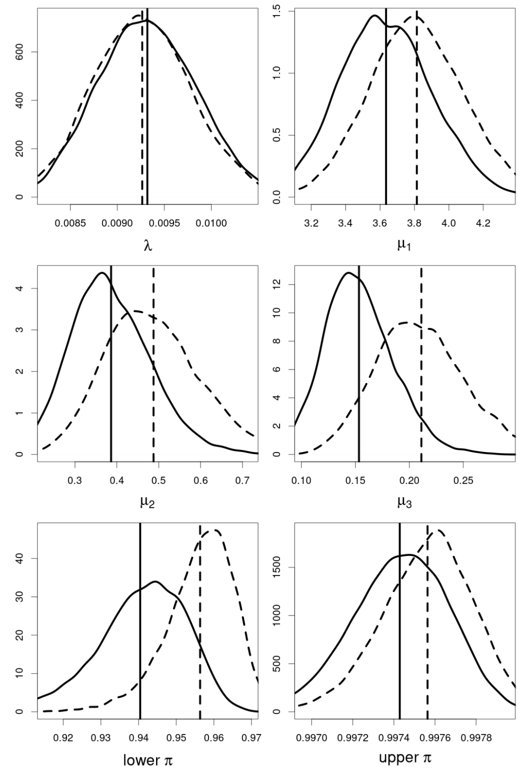

Figure 6: Posterior densities for the lumped model under both priors (solid and dashed line), for large sample size ($N = M = 300$). The vertical lines indicate posterior expectations.

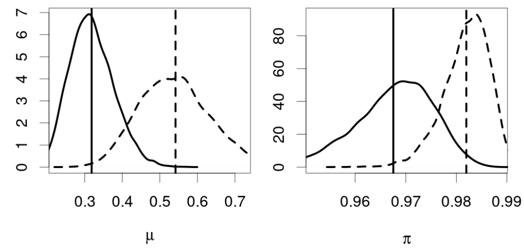

Figure 5: Posterior densities for the exponential model under both priors (solid and dashed line), for small sample size ($N = M = 30$). The vertical lines indicate posterior expectations.

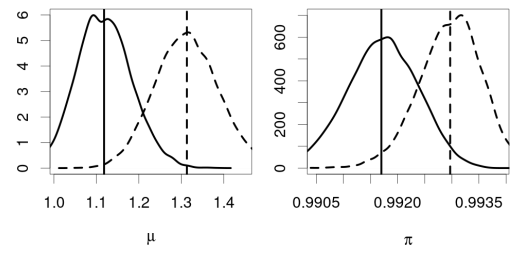

Figure 7: Posterior densities for the exponential model under both priors (solid and dashed line), for large sample size ($N = M = 300$). The vertical lines indicate posterior expectations.

7. Conclusion

In this paper, we explored how to combine fitting of phase-type distributions and lumping of Markov chains into imprecise Markov chains, to fit two-state imprecise Markov chains to data from a non-Markovian two-state processes. Through Markov chain Monte Carlo simulation, we demonstrated how to perform a full uncertainty quantification, deriving sets of posterior distributions on relevant quantities of interest, from a set of prior distributions on all parameters. In doing so, we have shown how non-Markovian processes can be statistically fitted, in one fell swoop, through robust Bayesian analysis.

This analysis stands in contrast with how imprecise Markov chains have been fitted elsewhere in the literature, which typically comprised of identifying lower and upper bounds on the transition rates of a precise Markov chain model, and then using those lower and upper bounds as bounds on the rates of an imprecise Markov chain model (see for instance [25, 18, 24, 16]). Those approaches cannot model non-Markovian behaviour as correctly as the lumped phase-type model. Also, in those approaches, the influence of the prior will vanish as we get more data, and thereby, so will any imprecision in the posterior. The lumped phase-type model on the other hand will always retain some level of imprecision, as this imprecision is a result of structural uncertainty, and not just due to a lack of data.

In our numerical example, the many-state model that we used was still quite simple. In general, to model more complex phase-type distributions for transition times, we may not have a fast closed form expression for the phase-type density. Either, one would have to resort to direct evaluation of a matrix exponential as part of the Markov chain Monte Carlo run, or, one could simply sample directly from the likelihood (which can be done quickly for any phase-type distribution), and then use approximate Bayesian computation [9] to fit the model.

Acknowledgments

We thank all reviewers for their constructive comments on the paper. 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.

Appendix A. Technical Results

Throughout, we use the fact that for any matrix A:

$$\exp(A) = \sum_{n=0}^{\infty} \frac{1}{n!} A^n \tag{38}$$

Theorem 1 *For every a, b, $c \in \mathbb{R}$ such that $a \neq c$, we have that*

$$\exp\left(\begin{bmatrix} a & b \\ 0 & c \end{bmatrix}\right) = \begin{bmatrix} e^a & b\frac{e^a - e^c}{a-c} \\ 0 & e^c \end{bmatrix} \tag{39}$$

Proof We first show by induction that, for every $n \in \mathbb{N}$,

$$\begin{bmatrix} a & b \\ 0 & c \end{bmatrix}^n = \begin{bmatrix} a^n & b\frac{a^n - c^n}{a-c} \\ 0 & c^n \end{bmatrix} \tag{40}$$

Indeed, clearly, Eq. (40) holds for $n = 0$. Assume Eq. (40) holds for a particular fixed value of n. Then,

$$\begin{bmatrix} a & b \\ 0 & c \end{bmatrix}^{n+1} = \begin{bmatrix} a & b \\ 0 & c \end{bmatrix}^n \begin{bmatrix} a & b \\ 0 & c \end{bmatrix} \tag{41}$$

$$= \begin{bmatrix} a^n & b\frac{a^n - c^n}{a-c} \\ 0 & c^n \end{bmatrix} \begin{bmatrix} a & b \\ 0 & c \end{bmatrix} \tag{42}$$

$$= \begin{bmatrix} a^{n+1} & ba^n + cb\frac{a^n - c^n}{a-c} \\ 0 & c^{n+1} \end{bmatrix} \tag{43}$$

Now note that

$$ba^n + cb\frac{a^n - c^n}{a-c} = b\frac{a^n(a-c) + c(a^n - c^n)}{a-c} \tag{44}$$

$$= b\frac{a^{n+1} - c^{n+1}}{a-c} \tag{45}$$

So, by induction, Eq. (40) must hold for all $n \in \mathbb{N}$.

We can now prove Eq. (39). Indeed,

$$\exp\left(\begin{bmatrix} a & b \\ 0 & c \end{bmatrix}\right) = \sum_{n=0}^{\infty} \frac{1}{n!} \begin{bmatrix} a & b \\ 0 & c \end{bmatrix}^n \tag{46}$$

$$= \sum_{n=0}^{\infty} \frac{1}{n!} \begin{bmatrix} a^n & b\frac{a^n - c^n}{a-c} \\ 0 & c^n \end{bmatrix} \tag{47}$$

$$= \begin{bmatrix} e^a & b\frac{e^a - e^c}{a-c} \\ 0 & e^c \end{bmatrix} \tag{48}$$

■

Theorem 2 *For every a and $b \in \mathbb{R}$ such that $a + b \neq 0$,*

$$\exp\left(\begin{bmatrix} -a & a \\ b & -b \end{bmatrix}\right) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + c \begin{bmatrix} -a & a \\ b & -b \end{bmatrix} \tag{49}$$

where $c := \frac{1 - e^{-a-b}}{a+b}$.

Proof We first show by induction that, for every $n \in \mathbb{N}$, $n \geq 1$,

$$\begin{bmatrix} -a & a \\ b & -b \end{bmatrix}^n = c_n \begin{bmatrix} -a & a \\ b & -b \end{bmatrix} \tag{50}$$

where $c_n := \frac{-(-a-b)^n}{a+b}$. Indeed, clearly Eq. (50) holds for $n = 1$. Assume Eq. (50) holds for a particular fixed value of $n \geq 1$. Then,

$$\begin{bmatrix} -a & a \\ b & -b \end{bmatrix}^{n+1} = \begin{bmatrix} -a & a \\ b & -b \end{bmatrix}^n \begin{bmatrix} -a & a \\ b & -b \end{bmatrix} \tag{51}$$

$$= c_n \begin{bmatrix} -a & a \\ b & -b \end{bmatrix} \begin{bmatrix} -a & a \\ b & -b \end{bmatrix} \tag{52}$$

$$= c_n \begin{bmatrix} a^2 + ab & -a^2 - ab \\ -ab - b^2 & ab + b^2 \end{bmatrix} \tag{53}$$

$$= c_n \begin{bmatrix} -a(-a-b) & a(-a-b) \\ b(-a-b) & -b(-a-b) \end{bmatrix} \tag{54}$$

$$= c_{n+1} \begin{bmatrix} -a & a \\ b & -b \end{bmatrix} \tag{55}$$

So, by induction, Eq. (50) must hold for all $n \in \mathbb{N}$, $n \geq 1$.

We can now prove Eq. (49). Indeed, by Eq. (38),

$$\exp\left(\begin{bmatrix} -a & a \\ b & -b \end{bmatrix}\right) - \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{56}$$

$$= \sum_{n=1}^{\infty} \frac{1}{n!} \begin{bmatrix} -a & a \\ b & -b \end{bmatrix}^n \tag{57}$$

$$= \sum_{n=1}^{\infty} \frac{c_n}{n!} \begin{bmatrix} -a & a \\ b & -b \end{bmatrix} \tag{58}$$

$$= \frac{-(e^{-a-b}-1)}{a+b} \begin{bmatrix} -a & a \\ b & -b \end{bmatrix} \tag{59}$$

■

Next, we want to prove that our expression for $\underline{T}_t$ in Eq. (16) indeed corresponds to the lower transition operator. To do so, however, we will not use the definition of Eq. (9). Instead, we will take Eq. (16), that is (repeated here for convenience),

$$\underline{T}_t f := f + \frac{1-e^{-(\lambda_f+\mu_f)t}}{\lambda_f+\mu_f} Q_f f, \tag{60}$$

as a definition of $\underline{T}_t$, and then prove that this operator solves a specific differential equation. It follows then from [20, Sec. 2.6] that the solution to this differential equation is the lower transition operator.

Lemma 3 *For every $f \in \mathbb{R}^2$, we have that*

$$[\underline{T}_t f]_\alpha - [\underline{T}_t f]_\beta = (f_\alpha - f_\beta) e^{-(\lambda_f+\mu_f)t}. \tag{61}$$

Proof

$$[\underline{T}_t f]_\alpha - [\underline{T}_t f]_\beta \tag{62}$$

$$= \left(f_\alpha + \frac{1-e^{-(\lambda_f+\mu_f)t}}{\lambda_f+\mu_f}[Q_f f]_\alpha\right) - \left(f_\beta + \frac{1-e^{-(\lambda_f+\mu_f)t}}{\lambda_f+\mu_f}[Q_f f]_\beta\right) \tag{63}$$

$$\begin{aligned} &= \left(f_\alpha + \frac{1-e^{-(\lambda_f+\mu_f)t}}{\lambda_f+\mu_f}\lambda_f(f_\beta - f_\alpha)\right) \\ &\quad - \left(f_\beta + \frac{1-e^{-(\lambda_f+\mu_f)t}}{\lambda_f+\mu_f}\mu_f(f_\alpha - f_\beta)\right) \end{aligned} \tag{64}$$

$$= f_\alpha - f_\beta + (f_\alpha - f_\beta)(-\lambda_f - \mu_f)\frac{1-e^{-(\lambda_f+\mu_f)t}}{\lambda_f+\mu_f} \tag{65}$$

$$= (f_\alpha - f_\beta) e^{-(\lambda_f+\mu_f)t} \tag{66}$$

■

Lemma 4 *$\lambda_{\underline{T}_t f} = \lambda_f$ and $\mu_{\underline{T}_t f} = \mu_f$.*

Proof λ_f and μ_f are determined solely by the sign of $f_\alpha - f_\beta$. Since, by Lemma 3, the sign of $[\underline{T}_t f]_\alpha - [\underline{T}_t f]_\beta$ is the same as the sign of $f_\alpha - f_\beta$, we have that $(\lambda_f, \mu_f) = (\lambda_{\underline{T}_t f}, \mu_{\underline{T}_t f})$. ■

Theorem 5 *The operator $\underline{T}_t$, as defined in Eq. (60), solves*

$$\frac{\mathrm{d}}{\mathrm{d}t}[\underline{T}_t f] = \underline{Q}[\underline{T}_t f] \tag{67}$$

with initial condition $\underline{T}_0 f = f$.

Proof We see that the initial condition $\underline{T}_0 f = f$ is satisfied. Evaluating the left-hand side of Eq. (67) we have:

$$\frac{\mathrm{d}}{\mathrm{d}t}[\underline{T}_t f] = e^{-(\lambda_f+\mu_f)t} Q_f f \tag{68}$$

Evaluating the right-hand side of Eq. (67) and using Lemma 4 we have:

$$\underline{Q}[\underline{T}_t f] = Q_{\underline{T}_t f}[\underline{T}_t f] \tag{69}$$

$$= Q_f[\underline{T}_f f] \tag{70}$$

$$= \left(Q_f + \frac{1-e^{-(\lambda_f+\mu_f)t}}{\lambda_f+\mu_f} Q_f^2\right) f \tag{71}$$

By Eq. (50), $Q_f^2 = -(\lambda_f + \mu_f)Q_f$. Therefore,

$$\underline{Q}[\underline{T}_t f] = \left(Q_f - \frac{1-e^{-(\lambda_f+\mu_f)t}}{\lambda_f+\mu_f}(\lambda_f+\mu_f)Q_f\right) f \tag{72}$$

$$= e^{-(\lambda_f+\mu_f)t} Q_f f \tag{73}$$

■

Appendix B. Stan model specification

```
functions {
  real phasetype_lpdf(
      real t, real mu1, real mu2, real mu3) {
    real a = -mu1-mu2;
    real b = mu2;
    real c = -mu3;
    real expa = exp(a * t);
    real expc = exp(c * t);
    real pdf =
      mu1 * expa
      + mu3 * b * (expa - expc) / (a - c);
    return log(pdf);
  }
  real min2(real a, real b) {
    vector[2] tmp;
    tmp[1] = a;
    tmp[2] = b;
    return min(tmp);
  }
}
data {
  // data
  int NX;
  vector[NX] X;
  int NY;
  vector[NY] Y;
  // hyperparameters
  real S;
  real T0;
  real T1;
  real T2;
  real T3;
}
parameters {
  real<lower=0> lambda;
  real<lower=0> mu1;
  real<lower=0> mu2;
  real<lower=0> mu3;
}
model {
  lambda ~ gamma(S, S * T0);
  mu1 ~ gamma(S, S * T1);
  mu2 ~ gamma(S, S * T2);
  mu3 ~ gamma(S, S * T3);
  for (n in 1:NX)
    X[n] ~ exponential(lambda);
  for (n in 1:NY)
    Y[n] ~ phasetype(mu1, mu2, mu3);
}
generated quantities {
  real lmu = min2(mu1, mu3);
  real umu = -min2(-mu1, -mu3); // max
  // limiting lower and upper probabilities
  real lpi = lmu / (lambda + lmu);
  real upi = umu / (lambda + umu);
}
```

References

[1] Søren Asmussen, Olle Nerman, and Marita Olsson. Fitting phase-type distributions via the EM algorithm. *Scandinavian Journal of Statistics*, 23(4): 419–441, 1996. URL http://www.jstor.org/stable/4616418.

[2] Peter Bacchetti, Ross D. Boylan, Norah A. Terrault, Alexander Monto, and Marina Berenguer. Non-Markov multistate modeling using time-varying covariates, with application to progression of liver fibrosis due to hepatitis C following liver transplant. *The international journal of biostatistics*, 6(1), 2010.

[3] Andrew Briggs and Mark Sculpher. An introduction to Markov modelling for economic evaluation. *Pharmacoeconomics*, 13(4):397–409, 1998.

[4] 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. doi:10.18637/jss.v076.i01.

[5] Richard Crossman, Pauline Coolen-Schrijner, Damjan Škulj, and Frank Coolen. Imprecise Markov chains with an absorbing state. In Thomas Augustin, Frank P. A. Coolen, Serafín Moral, and Matthias C. M. Troffaes, editors, *ISIPTA'09: Proceedings of the Sixth International Symposium on Imprecise Probability: Theories and Applications*, pages 119–128, Durham, UK, July 2009. SIPTA.

[6] Richard J. Crossman and Damjan Škulj. Imprecise Markov chains with absorption. *International Journal of Approximate Reasoning*, 51:1085–1099, 2010. doi:10.1016/j.ijar.2010.08.008.

[7] 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. doi:10.1017/S0269964809990039.

[8] Gert De Cooman, Jasper De Bock, and Stavros Lopatatzidis. Imprecise stochastic processes in discrete time: global models, imprecise Markov chains, and ergodic theorems. *International Journal of Approximate Reasoning*, 76:18–46, 2016. doi:10.1016/j.ijar.2016.04.009.

[9] Peter J. Diggle and Richard J. Gratton. Monte Carlo methods of inference for implicit statistical models. *Journal of the Royal Statistical Society, Series B*, 46 (2):193–227, 1984. URL http://www.jstor.org/stable/2345504.

[10] 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. URL http://proceedings.mlr.press/v62/erreygers17a.html.

[11] 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.

[12] András Horváth and Miklós Telek. PhFit: A general phase-type fitting tool. In *International Conference on Modelling Techniques and Tools for Computer Performance Evaluation*, pages 82–91, 2002.

[13] Igor O. Kozine and Lev V. Utkin. Interval-valued finite Markov chains. *Reliable Computing*, 8:97–113, 2002. doi:10.1023/A:1014745904458.

[14] Thomas Krak, Jasper De Bock, and Arno Siebes. Imprecise continuous-time Markov chains. *International Journal of Approximate Reasoning*, 88:452–528, 2017. ISSN 0888-613X. doi:10.1016/j.ijar.2017.06.012.

[15] Thomas Krak, Jasper De Bock, and Arno Siebes. Efficient computation of updated lower expectations for imprecise continuous-time hidden Markov chains. 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 193–204, July 2017. URL http://proceedings.mlr.press/v62/krak17a.html.

[16] 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.

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

[18] 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. doi:10.1007/978-3-319-08852-5_49.

[19] Philipp Reinecke, Tilman Krauß, and Katinka Wolter. Cluster-based fitting of phase-type distributions to empirical data. *Computers & Mathematics with Applications*, 64(12):3840–3851, 2012. doi:10.1016/j.camwa.2012.03.016. Theory and Practice of Stochastic Modeling.

[20] Damjan Škulj. Efficient computation of the bounds of continuous time imprecise Markov chains. *Applied Mathematics and Computation*, 250:165–180, 2015. doi:10.1016/j.amc.2014.10.092.

[21] Damjan Škulj and Robert Hable. Coefficients of ergodicity for imprecise Markov chains. In Thomas Augustin, Frank P. A. Coolen, Serafín Moral, and Matthias C. M. Troffaes, editors, *ISIPTA'09: Proceedings of the Sixth International Symposium on Imprecise Probability: Theories and Applications*, pages 377–386, Durham, UK, July 2009. SIPTA.

[22] M. Telek and A. Heindl. Matching moments for acyclic discrete and continous phase-type distributions of second order. *International Journal of Simulation: Systems, Science & Technology*, 3:47–57, 2002.

[23] A. Thummler, P. Buchholz, and M. Telek. A novel approach for phase-type fitting with the EM algorithm. *IEEE Transactions on Dependable and Secure Computing*, 3(3):245–258, 2006. doi:10.1109/TDSC.2006.27.

[24] 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. ISBN 978-88-548-8555-4. URL `http://www.sipta.org/isipta15/data/paper/18.pdf`.

[25] 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. URL `http://www.sipta.org/isipta13/index.php?id=paper&paper=031.html`.

[26] Matthias C. M. Troffaes and Lewis Paton. Logistic regression on Markov chains for crop rotation modelling. 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 329–336, Compiègne, France, July 2013. SIPTA. URL `http://www.sipta.org/isipta13/index.php?id=paper&paper=033.html`.

[27] Matthias C. M. Troffaes and Damjan Škulj. Model checking for imprecise Markov chains. 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 337–344, Compiègne, France, July 2013. SIPTA. URL `http://www.sipta.org/isipta13/index.php?id=paper&paper=034.html`.

[28] Damjan Škulj. Discrete time Markov chains with interval probabilities. *International Journal of Approximate Reasoning*, 50(8):1314–1329, 2009. doi:10.1016/j.ijar.2009.06.007.

[29] Damjan Škulj and Robert Hable. Coefficients of ergodicity for Markov chains with uncertain parameters. *Metrika*, 76:107–133, 2013. doi:10.1007/s00184-011-0378-0.

[30] Junfeng Wang, Jin Liu, and Chundong She. Segment-based adaptive hyper-erlang model for long-tailed network traffic approximation. *The Journal of Supercomputing*, 45(3):296–312, 2008. doi:10.1007/s11227-008-0173-5.

[31] Stelios H. Zanakis and Martin W. Maret. A Markov chain application to manpower supply planning. *Journal of the Operational Research Society*, 31(12):1095–1102, 1980.