An Imprecise Probabilistic Estimator for the Transition Rate Matrix of a Continuous-Time Markov Chain

Thomas Krak, Alexander Erreygers, and Jasper De Bock

Ghent University, ELIS, SYSTeMS
{thomas.krak,alexander.erreygers,jasper.debock}@ugent.be

Abstract We consider the problem of estimating the transition rate matrix of a continuous-time Markov chain from a finite-duration realisation of this process. We approach this problem in an imprecise probabilistic framework, using a set of prior distributions on the unknown transition rate matrix. The resulting estimator is a set of transition rate matrices that, for reasons of conjugacy, is easy to find. To determine the hyperparameters for our set of priors, we reconsider the problem in discrete time, where we can use the well-known Imprecise Dirichlet Model. In particular, we show how the limit of the resulting discrete-time estimators is a continuous-time estimator. It corresponds to a specific choice of hyperparameters and has an exceptionally simple closed-form expression.

1 Introduction

Continuous-time Markov chains (CTMCs) are mathematical models that describe the evolution of dynamical systems under (stochastic) uncertainty [9]. They are pervasive throughout science and engineering, finding applications in areas as disparate as medicine, mathematical finance, epidemiology, queueing theory, and others. We here consider time-homogeneous CTMCs that can only be in a finite number of states.

The dynamics of these models are uniquely characterised by a single *transition rate matrix* Q. This Q describes the (locally) linearised dynamics of the model, and is the generator of the semi-group of transition matrices $T_t = \exp(Qt)$ that determines the conditional probabilities $P(X_t = y \mid X_0 = x) = T_t(x, y)$. In this expression, X_t denotes the uncertain state of the system at time t, and so T_t contains the probabilities for the system to move from any state x at time zero to any state y at time t.

In this work, we consider the problem of estimating the matrix Q from a single realisation of the system up to some finite point in time. This problem is easily solved in both the classical frequentist and Bayesian frameworks, due to the likelihood of the corresponding CTMC belonging to an exponential family; see e.g. the introductions of [3, 7]. The novelty of the present paper is that we instead consider the estimation of Q in an *imprecise probabilistic* [1, 14] context.

Specifically, we approach this problem by considering an entire *set* of Bayesian priors on the likelihood of Q, leading to a *set-valued* estimator for Q. In order to obtain well-founded hyperparameter settings for this set of priors, we recast the problem by interpreting a continuous-time Markov chain as a limit of *discrete*-time Markov chains. This allows us to consider the imprecise-probabilistic estimators of these discrete-time Markov chains, which are described by the popular Imprecise Dirichlet Model (IDM) [10]. The upshot of this approach is that the IDM has well-known prior hyperparameter settings which can be motivated from first principles [4,15].

This leads us to the two main results of this work. First of all, we show that the limit of these IDM estimators is a set of transition rate matrices that can be described in closed-form using a very simple formula. Secondly, we identify the hyperparameters of our imprecise CTMC prior such that the resulting estimator is equivalent to the estimator obtained from this discrete-time limit. For reasons of brevity, the proofs of our results are omitted. They are available in the appendix of an online extended version.[1]

The immediate usefulness of our results is two-fold. From a domain-analysis point of view, where we are interested in the parameter values of the process dynamics, our imprecise estimator provides prior-insensitive information about these values based on the data. If we are instead interested in robust inference about the future behaviour of the system, our imprecise estimator can be used as the main parameter of an *imprecise continuous-time Markov chain* [5,6,8,13].

2   A Brief Refresher on Stochastic Processes

Intuitively, a stochastic process describes the uncertainty in a stochastic system's behaviour as it moves through some state space $\mathcal{X}$ as time t progresses over some time dimension $\mathbb{T}$. A fundamental choice is whether we are considering processes in discrete time, in which case typically $\mathbb{T} = \mathbb{N}_0$, or in continuous time, in which case $\mathbb{T} = \mathbb{R}_{\geq 0}$. Here we write $\mathbb{N}$ for the natural numbers, and let $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$. The real numbers are denoted by $\mathbb{R}$, the positive reals by $\mathbb{R}_{>0}$, and the non-negative reals by $\mathbb{R}_{\geq 0}$. We briefly recall the basic definitions of stochastic processes below; for an introductory work we refer to e.g. [9].

Formally, a realisation of a stochastic process is a *sample path*, which is a map $\omega : \mathbb{T} \to \mathcal{X}$. Here $\omega(t) \in \mathcal{X}$ represents the state of the process at time $t \in \mathbb{T}$. We collect all sample paths in the set Ω and, when $\mathbb{T} = \mathbb{R}_{\geq 0}$, these paths are assumed to be càdlàg under the discrete topology on $\mathcal{X}$. With this domain in place, we then consider some abstract underlying probability space $(\Omega, \mathcal{F}, P)$, where $\mathcal{F}$ is some appropriate (σ-)algebra on Ω, and where P is a (countably-)additive probability measure.

The stochastic process can now finally be defined as a family of random variables $\{X_t\}_{t \in \mathbb{T}}$ associated with this probability space. In particular, for fixed $t \in \mathbb{T}$, the quantity X_t is a random variable $\Omega \to \mathcal{X} : \omega \mapsto \omega(t)$. Conversely, for a fixed realisation $\omega \in \Omega$, $X_t(\omega)$ is a deterministic map $\mathbb{T} \to \mathcal{X} : t \mapsto \omega(t)$.

---

[1] URL: `https://arxiv.org/abs/1804.01330`

Well-known and popular kinds of stochastic processes are *Markov chains*:

Definition 1 (Markov Chain). *Fix $\mathbb{T} \in \{\mathbb{N}_0, \mathbb{R}_{\geq 0}\}$, and let $\{X_t\}_{t \in \mathbb{T}}$ be a stochastic process. We call this process a* Markov chain *if, for all $s_0, \ldots, s_n, s, t \in \mathbb{T}$ for which $s_0 < \cdots < s_n < s < t$, it holds that $P(X_t = x_t \mid X_{s_0} = x_{s_0}, \ldots, X_{s_n} = x_{s_n}, X_s = x_s) = P(X_t = x_t \mid X_s = x_s)$ for all $x_{s_0}, \ldots, x_{s_n}, x_s, x_t \in \mathcal{X}$. If then $\mathbb{T} = \mathbb{N}_0$, we call $\{X_t\}_{t \in \mathbb{T}}$ a* discrete-time Markov chain *(DTMC). If instead $\mathbb{T} = \mathbb{R}_{\geq 0}$, we call it a* continuous-time Markov chain *(CTMC).*

Furthermore, attention is often restricted to *homogenous* Markov chains:

Definition 2 (Homogeneous Markov Chain). *Let $\{X_t\}_{t \in \mathbb{T}}$ be a Markov chain. We call this Markov chain* (time-)homogeneous *if, for all $s, t \in \mathbb{T}$, $s \leq t$, and all $x, y \in \mathcal{X}$, it holds that $P(X_t = y \mid X_s = x) = P(X_{(t-s)} = y \mid X_0 = x)$.*

This homogeneity property makes such processes particularly easy to describe.

In what follows, we will say that a $|\mathcal{X}| \times |\mathcal{X}|$ matrix T is a *transition matrix*, if it is a real-valued and row stochastic matrix, i.e. if $T(x,y) \geq 0$ and $\sum_{z \in \mathcal{X}} T(x,z) = 1$ for all $x, y \in \mathcal{X}$. We write $\mathfrak{T}$ for the space of all transition matrices. The elements T of $\mathfrak{T}$ can be used to describe the single-step conditional probabilities of a (homogeneous) DTMC:

Proposition 3 ([9]). *Let $\{X_t\}_{t \in \mathbb{N}_0}$ be a homogeneous DTMC. Then this process is completely and uniquely characterised by a probability mass function p on $\mathcal{X}$ and some $T \in \mathfrak{T}$. In particular, $P(X_0) = p$ and, for all $t \in \mathbb{N}_0$ and all $x, y \in \mathcal{X}$, $P(X_t = y \mid X_0 = x) = T^t(x,y)$, where T^t is the tth matrix power of T.*

On the other hand, to describe CTMCs we need the concept of a *(transition) rate matrix*: a $|\mathcal{X}| \times |\mathcal{X}|$ real-valued matrix Q with non-negative off-diagonal elements and zero row-sums, i.e. $Q(x,y) \geq 0$ and $\sum_{z \in \mathcal{X}} Q(x,z) = 0$ for all $x, y \in \mathcal{X}$ such that $x \neq y$. We write $\mathfrak{Q}$ for their entire space. A rate matrix describes the "speed" with which a CTMC moves between its states:

Proposition 4 ([9]). *Let $\{X_t\}_{t \in \mathbb{R}_{\geq 0}}$ be a homogeneous CTMC. Then this process is completely and uniquely characterised by a probability mass function p on $\mathcal{X}$ and some $Q \in \mathfrak{Q}$. In particular, $P(X_0) = p$ and, for all $t \in \mathbb{R}_{\geq 0}$ and all $x, y \in \mathcal{X}$, $P(X_t = y \mid X_0 = x) = \exp(Qt)(x,y)$, where $\exp(Qt)$ is the matrix exponential of Qt. Furthermore, for small enough $\Delta \in \mathbb{R}_{\geq 0}$ and all $x, y \in \mathcal{X}$, it holds that $P(X_\Delta = y \mid X_0 = x) \approx (I + \Delta Q)(x,y)$, where I is the identity matrix.*

3 Estimation of a CTMC's Rate Matrix

In what follows, we will derive methods to estimate the rate matrix Q of a homogeneous CTMC from a realisation $\omega \in \Omega$ that was observed up to some finite point in time $t_{\max} \in \mathbb{R}_{\geq 0}$. We denote with $\widetilde{\omega}$ the restriction of ω to this interval $[0, t_{\max}] \subset \mathbb{R}_{\geq 0}$, and we consider this (finite-duration) observation to be fixed throughout the remainder of this paper.

For any $x, y \in \mathcal{X}$ such that $x \neq y$, we let n_{xy} denote the number of transitions from state x to state y in $\widetilde{\omega}$. Furthermore, we let d_x denote the total duration

spent in state x, that is, we let $d_x := \int_0^{t_{\max}} \mathbb{I}_x(\widetilde{\omega}(t))\,\mathrm{d}t$, where $\mathbb{I}_x$ is the indicator of $\{x\}$, defined by $\mathbb{I}_x(\widetilde{\omega}(t)) := 1$ if $\widetilde{\omega}(t) = x$ and $\mathbb{I}_x(\widetilde{\omega}(t)) := 0$ otherwise. We assume in the remainder that $d_x > 0$ for all $x \in \mathcal{X}$. Finally, for notational convenience, we define $q_{xy} := Q(x, y)$ for all $x, y \in \mathcal{X}$.

3.1 Precise Estimators

Under the assumption that the realisation ω was generated by a homogeneous continuous-time Markov chain with rate matrix Q, it is well known that the process dynamics can be modelled using exponentially distributed random variables whose parameters are given by the elements of Q. For various of such interpretations, we refer to e.g. [9]. What matters to us here is that, regardless of the interpretation, we can use this to obtain the following likelihood result (see e.g. [7]): for a given $\widetilde{\omega}$, the likelihood for a rate matrix Q is

$$L(\widetilde{\omega} \mid Q) = \prod_{\substack{x,y\in\mathcal{X} \\ x\neq y}} (q_{xy})^{n_{xy}} e^{-q_{xy}d_x}. \tag{1}$$

The corresponding maximum-likelihood estimator Q^{ML} is easily found [7]: $q_{xy}^{\mathrm{ML}} = n_{xy}/d_x$ if $x \neq y$ and $q_{xx}^{\mathrm{ML}} = -\sum_{y\in\mathcal{X}\setminus\{x\}} q_{xy}^{\mathrm{ML}}$, where the final expression follows from the (implicit) constraint that the rows of a rate matrix should sum to zero.

Inspection of the likelihood in (1) reveals that it belongs to an exponential family. This implies that there exists a conjugate prior for the rate matrix Q, such that its posterior distribution, given $\widetilde{\omega}$, belongs to the same family as this prior. This prior is given by a product of Gamma distributions, specifically on the off-diagonal elements q_{xy}, $x \neq y$, of the corresponding rate matrix [3]. We here use a slightly more general joint prior on Q whose "density" f is given by

$$f(Q \mid \boldsymbol{\alpha}, \boldsymbol{\beta}) := \prod_{\substack{x,y\in\mathcal{X} \\ x\neq y}} (q_{xy})^{\alpha_{xy}-1} e^{-q_{xy}\beta_x} \propto \prod_{\substack{x,y\in\mathcal{X} \\ x\neq y}} \mathrm{Gamma}(q_{xy} \mid \alpha_{xy}, \beta_x), \tag{2}$$

with shapes α_{xy} and rates β_x in $\mathbb{R}_{\geq 0}$; we write $\boldsymbol{\alpha}, \boldsymbol{\beta}$ for the joint parameters.

Note that we have only defined the prior to equal a product of Gamma distributions up to normalisation, so that the prior $f(Q \mid \boldsymbol{\alpha}, \boldsymbol{\beta})$ may be improper. This has the advantage that it allows us to close the parameter domains and allow prior hyperparameters $\alpha_{xy} = 0$ and $\beta_x = 0$, for which the Gamma distribution is not properly defined. We acknowledge that the use of such improper priors is not entirely uncontroversial, and that their interpretation as a prior probability (which it indeed is not) leaves something to be desired. We will nevertheless, in this specific setting, be able to motivate their use here as a consequence of Theorem 5 further on.

Also, despite being improper, we can of course combine the prior (2) with the likelihood (1) and fix the normalisation in the posterior. As is well known, the means of the marginals of this posterior are then of the form[2]

$$\mathbb{E}\big[q_{xy} \mid \boldsymbol{\alpha}, \boldsymbol{\beta}, \widetilde{\omega}\big] = \frac{\alpha_{xy} + n_{xy}}{\beta_x + d_x} \qquad \forall x, y \in \mathcal{X}, x \neq y. \tag{3}$$

---

[2] The assumption $d_x > 0$ prevents division by zero in (3). However, n_{xy} might be zero and, if then also $\alpha_{xy} = 0$, the posterior cannot be normalised and will still be

Furthermore, the (joint) posterior mean is well-known to be a Bayes estimator for Q under quadratic loss and given the prior $f(\,\cdot\,|\,\boldsymbol{\alpha},\boldsymbol{\beta})$ [2].

The question now remains of how to a priori settle on a "good" choice for these hyperparameters $\boldsymbol{\alpha},\boldsymbol{\beta}$, in the sense that they should adequately represent our prior beliefs. This is a non-trivial problem, and no general solution can be given. A popular (but not uncontroversial) attempt to characterise a non-informative prior consists in choosing the improper prior with $\boldsymbol{\alpha} = \boldsymbol{\beta} = 0$; the posterior mean (Bayes) estimator then equals Q^{ML}.

3.2 An Imprecise Probabilistic Estimator

Generalising the above Bayesian approach, we here suggest an imprecise probabilistic treatment. Following for example [11, 14], this approach consists in using an entire *set* of prior distributions. Specifically, we consider a set of the form

$$\left\{ f(\,\cdot\,|\,\boldsymbol{\alpha},\boldsymbol{\beta}) \,\Big|\, (\boldsymbol{\alpha},\boldsymbol{\beta}) \in C \right\}, \tag{4}$$

with $f(\,\cdot\,|\,\boldsymbol{\alpha},\boldsymbol{\beta})$ as in (2), and where C is a set of possible prior parameters. In this way, we do not have to restrict our attention to one specific choice of the parameters $\boldsymbol{\alpha},\boldsymbol{\beta}$; rather, we can include all the parameter settings that we deem reasonable, by collecting them in C. Inference from $\widetilde{\omega}$ is then performed by point-wise updating each of these priors; we thereby obtain a set of posterior distributions on the space of all rate matrices. Each of these posteriors has a mean of the form (3), which is a Bayes estimator for Q under a specific prior in the set (4). This leads us to consider the imprecise, i.e., *set-valued*, estimator

$$\mathcal{Q}_C := \left\{ Q \in \mathfrak{Q} \,\middle|\, \left(\forall x, y \in \mathcal{X}, x \neq y : q_{xy} = \frac{\alpha_{xy} + n_{xy}}{\beta_x + d_x} \right), (\boldsymbol{\alpha},\boldsymbol{\beta}) \in C \right\}.$$

Note that even in this imprecise probabilistic approach, we still need to somehow specify the (now set-valued) prior model. That is, we need to be specific about the set C. Inspired by the well-known imprecise Dirichlet model [15], we may choose an "imprecision parameter" $s \in \mathbb{R}_{\geq 0}$, which can be interpreted as a number of "pseudo-counts", to constrain $0 \leq \sum_{y \in \mathcal{X} \setminus \{x\}} \alpha_{xy} \leq s$ for all $x \in \mathcal{X}$, and to then vary all β_x over their domain $\mathbb{R}_{\geq 0}$. Unfortunately, similar to what is noted in [11], this leads to undesirable behaviour. For example, as is readily seen from e.g. (3), including unbounded β_x allows the off-diagonal elements q_{xy} to get arbitrarily close to zero, causing the model to a posteriori believe that transitions leaving x may be impossible, no matter the number of such transitions that we actually observed in $\widetilde{\omega}$! Hence, we prefer a different choice of C.

One way to circumvent this undesired behaviour is to constrain the range within which each β_x may be varied, to some interval $[0, \overline{\beta}_x]$, say. The downside

---

improper. Nevertheless, using an intuitive (but formally cumbersome) argument we can still identify this posterior for q_{xy} with the (discrete) distribution putting all mass at zero. Alternatively, we can motivate (3) by continuous extension from the cases where $\alpha_{xy} > 0$, similarly yielding the estimate $\hat{q}_{xy} = 0$ at $\alpha_{xy} = n_{xy} = 0$.

is that this introduces a large number of additional hyperparameters; we then need to ("reasonably") choose a value $\overline{\beta}_x \in \mathbb{R}_{\geq 0}$ for each $x \in \mathcal{X}$. Fortunately, our main result – Theorem 5 further on – suggests that setting $\overline{\beta}_x = 0$ (and therefore $\beta_x = 0$) is in fact a very reasonable choice. This identification is obtained in the next section, using a limit result of discrete-time estimators, for which the hyperparameter settings follow entirely from first principles.

In summary, we keep the "imprecision parameter" $s \in \mathbb{R}_{\geq 0}$ and the constraint $0 \leq \sum_{y \in \mathcal{X} \setminus \{x\}} \alpha_{xy} \leq s$ for all $x \in \mathcal{X}$, and simply set $\beta_x = 0$ for all $x \in \mathcal{X}$. We then define C_s to be the largest set of parameters that satisfies these properties. Every $\boldsymbol{\alpha}$ in this set can be conveniently identified with the off-diagonal elements of a matrix sA, with $A \in \mathfrak{T}$ a transition matrix. Our set-valued estimator $\mathcal{Q}_s$ can thus be written as

$$\mathcal{Q}_s := \left\{ Q \in \mathfrak{Q} \,\middle|\, \left(\forall x, y \in \mathcal{X}, x \neq y \,:\, q_{xy} = \frac{sA(x,y) + n_{xy}}{d_x} \right), A \in \mathfrak{T} \right\}. \quad (5)$$

4 Discrete-Time Estimators and Limit Relations

A useful intuition is that we can consider a CTMC as a limit of DTMCs, where we assign increasingly shorter durations to the time steps at which the latter operate. In this section, we will use this connection to relate estimators for DTMCs to estimators for CTMCs. We start by discretising the observed path.

Because the realisation ω was only observed up to some time $t_{\max} \in \mathbb{R}_{\geq 0}$, we can discretise the (finite-duration) realisation $\widetilde{\omega}$ into a finite number of steps. For any $m \in \mathbb{N}$, we write $\delta^{(m)} := t_{\max}/m$, and we define the discretised path $w^{(m)} : \{0, \ldots, m\} \to \mathcal{X}$ as $w^{(m)}(i) := \widetilde{\omega}(i\delta^{(m)})$ for all $i \in \{0, \ldots, m\}$.

For any $m \in \mathbb{N}$ and $x, y \in \mathcal{X}$, we let $n_{xy}^{(m)} := \sum_{i=1}^{m} \mathbb{I}_x(w^{(m)}(i-1)) \mathbb{I}_y(w^{(m)}(i))$ denote the number of transitions from state x to y in $w^{(m)}$, and we let $n_x^{(m)} := \sum_{y \in \mathcal{X}} n_{xy}^{(m)}$ denote the total number of time steps that started in state x.

4.1 Discrete-Time Estimators

For fixed $m \in \mathbb{N}$, we can interpret the discretised path $w^{(m)}$ as a finite-duration ($m{+}1$ steps long) realisation of a homogeneous discrete-time Markov chain with transition matrix $T^{(m)}$, with m keeping track of the discretisation level. Each transition along the path $w^{(m)}$, from state x to y, say, is then a realisation of a categorical distribution with parameters $T^{(m)}(x, \cdot)$. The likelihood for $T^{(m)}$, given $w^{(m)}$, is therefore proportional to a product of independent multinomial likelihoods. Hence, the maximum likelihood estimator follows straightforwardly and as expected: $T^{(m),\mathrm{ML}}(x, y) = n_{xy}^{(m)}/n_x^{(m)}$ for all $x, y \in \mathcal{X}$; see [7] for details.

In a Bayesian analysis, and following e.g. [10], for fixed m we can model our uncertainty about the unknown $T^{(m)}$ by putting independent Dirichlet priors on the rows $T^{(m)}(x, \cdot)$. We write this prior as $g(\cdot \,|\, s, A)$, where $s \in \mathbb{R}_{\geq 0}$ is a "prior strength" parameter, and $A \in \mathrm{int}(\mathfrak{T})$ is a prior location parameter. Note that we take A in the interior of $\mathfrak{T}$ – under the metric topology on $\mathfrak{T}$ – so that each row $A(x, \cdot)$ corresponds to a strictly positive probability mass function.

After updating with $w^{(m)}$, the posterior mean is an estimator for $T^{(m)}$ that is Bayes under quadratic loss and for the specific prior $g(\,\cdot\,|\,s, A)$; due to conjugacy, the posterior is again a product of independent Dirichlet distributions [10], whence the elements of the posterior mean are

$$\mathbb{E}\Big[T(x, y)\,\Big|\,s, A, w^{(m)}\Big] = \frac{sA(x, y) + n_{xy}^{(m)}}{s + n_x^{(m)}} \qquad \forall x, y \in \mathcal{X}\,.$$

What remains is again to determine a good choice for s and A. However, in an imprecise probabilistic context we do not have to commit to any such choice: the popular Imprecise Dirichlet Model generalises the above approach using a *set* of Dirichlet priors. This set is given by $\mathrm{IDM}(\,\cdot\,|\,s) := \big\{g(\,\cdot\,|\,s, A)\,\big|\,A \in \mathrm{int}(\mathfrak{T})\big\}$ and can be motivated from first principles [4,15]. Observe that only a parameter $s \in \mathbb{R}_{\geq 0}$ remains, which controls the "degree of imprecision". In particular, we no longer have to commit to a location parameter A; instead this parameter is freely varied over its entire domain $\mathrm{int}(\mathfrak{T})$.

Element-wise updating with $w^{(m)}$ yields a set of posteriors which, due to conjugacy, are again independent products of Dirichlet distributions. The corresponding set $\mathcal{T}_s^{(m)}$ of posterior means thus contains estimators for $T^{(m)}$ that are Bayes for a specific prior from the IDM, and is easily verified to be

$$\mathcal{T}_s^{(m)} = \left\{T \in \mathfrak{T}\,\middle|\,\left(\forall x, y \in \mathcal{X} : T(x, y) = \frac{sA(x, y) + n_{xy}^{(m)}}{s + n_x^{(m)}}\right), A \in \mathrm{int}(\mathfrak{T})\right\}\,.$$

4.2 Limits of Discrete-Time Estimators

As noted in Proposition 4, a rate matrix Q is connected to the transition probabilities $T_\Delta(x, y) := P(X_\Delta = y\,|\,X_0 = x)$ in the sense that $T_\Delta \approx (I + \Delta Q)$ for small Δ. Hence, for small Δ, we have that $Q \approx (T_\Delta - I)^1/\Delta$. This becomes exact in the limit for Δ going to zero.

This interpretation can also be used to connect discrete-time estimators for $T^{(m)}$ to estimators for Q. For example, if we let $Q^{(m)} := (T^{(m),\mathrm{ML}} - I)^1/\delta^{(m)}$, then $Q^{\mathrm{ML}} = \lim_{m \to +\infty} Q^{(m)}$. Similarly, we can connect our set-valued estimators for the discretised models to the set-valued continuous-time estimator in (5):

Theorem 5. *For all $m \in \mathbb{N}$, let $\mathcal{Q}_s^{(m)} := \big\{(T - I)^1/\delta^{(m)}\,\big|\,T \in \mathcal{T}_s^{(m)}\big\}$. Then the Painlevé-Kuratowski [12] limit $\lim_{m \to +\infty} \mathcal{Q}_s^{(m)}$ exists, and equals $\mathcal{Q}_s$.*

5 Discussion

We have derived a set-valued estimator $\mathcal{Q}_s$ for the transition rate matrix of a homogeneous CTMC. It can be motivated both as a set of posterior means of a set of Bayesian models in continuous-time, and as a limit of set-valued discrete-time estimators based on the Imprecise Dirichlet Model. The only parameter of the estimator is a scalar $s \in \mathbb{R}_{\geq 0}$ that controls the degree of imprecision. In the special case where $s = 0$ there is no imprecision, and then $\mathcal{Q}_0 = \{Q^{\mathrm{ML}}\}$.

The set-valued representation $\mathcal{Q}_s$ is convenient when one is interested in the numerical values of the transition rates, e.g. for domain-analysis. If one aims to

use the estimator to describe an *imprecise CTMC* [8,13], a representation using the *lower transition rate operator* $\underline{Q}$ is more convenient. This operator is the lower envelope of a set of rate matrices; for $\mathcal{Q}_s$ it is given, for all $h : \mathcal{X} \to \mathbb{R}$, by

$$\left[\underline{Q}h\right](x) := \inf_{Q \in \mathcal{Q}_s} \sum_{y \in \mathcal{X}} Q(x,y)h(y) = \frac{s}{d_x} \min_{y \in \mathcal{X}} \big(h(y) - h(x)\big) + \sum_{y \in \mathcal{X} \setminus \{x\}} \frac{n_{xy}}{d_x} \big(h(y) - h(x)\big),$$

for all $x \in \mathcal{X}$. Hence, $\underline{Q}h$ is straightforward to evaluate. This implies that when our estimator is used to learn an imprecise CTMC from data, the lower expectations of this imprecise CTMC can be computed efficiently [6].

Acknowledgements

References

[1] Augustin, T., Coolen, F.P.A., De Cooman, G., Troffaes, M.C.M. (eds.): Introduction to Imprecise Probabilities. John Wiley & Sons (2014)

[2] Berger, J.O.: Statistical decision theory and Bayesian analysis. Springer (1985)

[3] Bladt, M., Sørensen, M.: Statistical inference for discretely observed Markov jump processes. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 67(3), 395–410 (2005)

[4] De Cooman, G., De Bock, J., Diniz, M.A.: Coherent predictive inference under exchangeability with imprecise probabilities. Journal of Artificial Intelligence Research 52, 1–95 (2015)

[5] De Bock, J.: The limit behaviour of imprecise continuous-time markov chains. Journal of Nonlinear Science 27(1), 159–196 (2017)

[6] Erreygers, A., De Bock, J.: Imprecise continuous-time markov chains: Efficient computational methods with guaranteed error bounds. In: Proceedings of ISIPTA 2017. pp. 145–156 (2017)

[7] Inamura, Y.: Estimating continuous time transition matrices from discretely observed data. Bank of Japan (2006)

[8] Krak, T., De Bock, J., Siebes, A.: Imprecise continuous-time Markov chains. International Journal of Approximate Reasoning 88, 452–528 (2017)

[9] Norris, J.R.: Markov chains. Cambridge university press (1998)

[10] Quaeghebeur, E.: Learning from samples using coherent lower previsions. Ph.D. thesis

[11] Quaehebeur, E., De Cooman, G.: Imprecise probability models for inference in exponential families. In: Proceedings of ISIPTA 2005 (2005)

[12] Rockafellar, T.R., Wets, R.J.B.: Variational Analysis. Springer (1997)

[13] Škulj, D.: Efficient computation of the bounds of continuous time imprecise Markov chains. Applied Mathematics and Computation 250(C), 165–180 (2015)

[14] Walley, P.: Statistical reasoning with imprecise probabilities. Chapman and Hall, London (1991)

[15] Walley, P.: Inferences from multinomial data: learning about a bag of marbles. Journal of the Royal Statistical Society, Series B 58, 3–57 (1996)