Markov chains

An introduction

Consider a generic continuous-time stochastic process $(X_t)_{t\in\mathbb{R}_{\geq 0}}$, where for all $t\in\mathbb{R}_{\geq 0}$ the state X_t is a random variable that takes values x in the finite state space $\mathscr{X}$. We provide $\mathscr{X}$ with some ordering, such that any real-valued function f on $\mathscr{X}$ can be identified with a row vector. We furthermore let $\mathscr{L}(\mathscr{X})$ denote the set of all real-valued functions on $\mathscr{X}$. Then any linear operator $T\colon \mathscr{L}(\mathscr{X})\to\mathscr{L}(\mathscr{X})$ can be identified with a matrix.

Precise Markov chains

The stochastic process $(X_t)_{t\in\mathbb{R}_{\geq 0}}$ is a *precise (continuous-time) Markov chain* (pMC) if it satisfies the *Markov property*: where $n\geq 0$ is an integer and $\{t_1,\dots,t_n,s,t\}$ is a strictly increasing sequence of non-negative time points. The *transition matrix* T_s^t thus defined satisfies

$$\begin{aligned}[T_s^t f](x_s) &= \mathrm{E}(f(X_t)|X_s=x_s)\\ &= \mathrm{E}(f(X_t)|X_{t_1}=x_1,\dots,X_{t_n}=x_n,X_s=x_s).\end{aligned} \qquad \text{(P1)}$$

A pMC is called *stationary* if it satisfies $T_t^{t+\Delta}=T_0^{\Delta}=:T_\Delta$ for all $t,\Delta\in\mathbb{R}_{\geq 0}$. In this case, there is a unique *transition rate matrix* Q—a matrix with non-negative off-diagonal elements and rows that sum up to zero—such that

$$(\forall t\in\mathbb{R}_{\geq 0})\; T_\Delta = T_t^{t+\Delta}\approx I+\Delta Q \quad \text{for } \Delta \text{ suff. small.}$$

Furthermore, T_t then satisfies the differential equation

$$\frac{\mathrm{d}}{\mathrm{d}t}T_t = QT_t, \qquad \text{with } T_0=I. \qquad \text{(P2)}$$

Similarly, for any non-stationary pMC there is a time-dependent transition rate matrix Q_t such that

$$(\forall t\in\mathbb{R}_{\geq 0})\; T_t^{t+\Delta}\approx I+\Delta Q_t \quad \text{for } \Delta \text{ suff. small.}$$

Imprecise Markov chains

It is often infeasible to precisely specify the transition rate matrix Q of a stationary pMC. Furthermore, assuming stationarity is not always justified. Therefore, we here consider the case where the (time-dependent) transition rate matrix Q_t of a (non-stationary) pMC is only known to be contained in some (non-empty and bounded) set $\mathscr{Q}$. In other words, we consider the set $\mathbb{P}_{\mathscr{Q}}$ of all pMCs that are consistent with $\mathscr{Q}$, in the sense that

$$(\forall t\in\mathbb{R}_{\geq 0})(\exists Q_t\in\mathscr{Q})\; T_t^{t+\Delta}\approx I+\Delta Q_t \quad \text{for } \Delta \text{ suff. small.}$$

This set $\mathbb{P}_{\mathscr{Q}}$ characterises an *imprecise (continuous-time) Markov chain* (iMC) as follows. Analogous to (P1), we define a *lower transition operator* $\underline{T}_s^t$ as

$$\begin{aligned}[\underline{T}_s^t f](x_s) &:= \underline{\mathrm{E}}(f(X_t)|X_s=x_s)\\ &= \underline{\mathrm{E}}(f(X_t)|X_{t_1}=x_1,\dots,X_{t_n}=x_n,X_s=x_s),\end{aligned} \qquad \text{(I1)}$$

where $\underline{\mathrm{E}}(\cdot|\cdot)$ is the minimum of the conditional expectations that are induced by the set of consistent processes.

In case $\mathscr{Q}$ has separately specified rows, Krak et al. (2017) show that $\underline{T}_t^{t+\Delta}=\underline{T}_0^{\Delta}=:\underline{T}_\Delta$ for all $t,\Delta\in\mathbb{R}_{\geq 0}$. Moreover, they show that $\underline{T}_\Delta$ is the unique operator that satisfies

$$\frac{\mathrm{d}}{\mathrm{d}t}\underline{T}_t = \underline{Q}\,\underline{T}_t, \qquad \text{with } \underline{T}_0=I. \qquad \text{(I2)}$$

In (I2), $\underline{Q}$ is the so-called *lower transition rate operator* of $\mathscr{Q}$, which, for any $f\in\mathscr{L}(\mathscr{X})$ and $x\in\mathscr{X}$, is defined as

$$[\underline{Q}f](x) := \min\{[Qf](x)\colon Q\in\mathscr{Q}\}. \qquad \text{(I3)}$$

Ergodicity

We are often interested in the long-term limit behaviour of stationary pMCs and iMCs. For iMCs, a special case is when

$$\lim_{t\to+\infty}[\underline{T}_t f](x) = \underline{\mathrm{E}}_\infty(f) \quad \text{for all } f\in\mathscr{L}(\mathscr{X}) \text{ and } x\in\mathscr{X}.$$

If this is the case, then the iMC is said to be *ergodic* and $\underline{\mathrm{E}}_\infty(f)$ is called the *limit lower expectation*. Similarly, a stationary pMC is ergodic if

$$\lim_{t\to+\infty}[T_t f](x) = \mathrm{E}_\infty(f) \quad \text{for all } f\in\mathscr{L}(\mathscr{X}) \text{ and } x\in\mathscr{X},$$

where E_∞ is now called the limit expectation.

Imprecise continuous-time Markov chains

Efficient computational methods with guaranteed error bounds

Alexander Erreygers
SMACS Research Group, Ghent University

Jasper De Bock
IDLab, Ghent University-imec

Guaranteed approximation methods

From an application point of view on imprecise (continuous-time) Markov chains (or iMCs, as introduced in **Markov chains: An introduction**), it is essential to have an efficient computational method to numerically approximate $\underline{T}_t f$ for some $f\in\mathscr{L}(\mathscr{X})$ and some $t\in\mathbb{R}_{>0}$. We are specifically interested in methods that yield an approximation $\Phi_t f$ of $\underline{T}_t f$ such that the error $\|\underline{T}_t f-\Phi_t f\|$ is lower than some desired maximal error ε. For ergodic iMCs, it is often also essential to approximate $\underline{\mathrm{E}}_\infty(f)$, see for instance **Modelling spectrum assignment in a two-service flexi-grid optical link**.

Some theoretical results

Throughout this poster, we let $\mathscr{X}$ be a finite and ordered state space, and $\underline{Q}\colon\mathscr{L}(\mathscr{X})\to\mathscr{L}(\mathscr{X})$ a generic lower transition rate operator. For any $f\in\mathscr{L}(\mathscr{X})$, we define

$$\|f\| := \max\{|f(x)|\colon x\in\mathscr{X}\} \quad \text{and} \quad \|f\|_c := (\max f-\min f)/2.$$

A first—although minor—result we prove is that

$$\|\underline{Q}\| := \sup\{\|\underline{Q}f\|\colon f\in\mathscr{L}(\mathscr{X}),\|f\|=1\} = 2\max\{|[\underline{Q}\mathbb{I}_x](x)|\colon x\in\mathscr{X}\}.$$

The two computational methods with guaranteed error bounds we consider are based on the following theorem.

Theorem 1. *Fix some $f\in\mathscr{L}(\mathscr{X})$ and $t\in\mathbb{R}_{\geq 0}$. Let $\Phi_t f$ be an approximation of $\underline{T}_t f$. Then for any $\delta\in\mathbb{R}_{\geq 0}$ such that $\delta\|Q\|\leq 2$ and any $m\in\mathbb{N}$,*

$$\|\underline{T}_{t+m\delta}f-(I+\delta\underline{Q})^m\Phi_t f\| \leq \|\underline{T}_t f-\Phi_t f\|+m\delta^2\|Q\|^2\|\Phi_t f\|_c.$$

For any *lower transition operator* $\underline{T}$ (a super-additive, positively homogeneous operator that dominates the minimum), Škulj and Hable (2013) define the *coefficient of ergodicity*

$$\rho(\underline{T}) := \max\{2\|\underline{T}f\|_c\colon f\in\mathscr{L}(\mathscr{X}), 0\leq f\leq 1\}. \qquad (1)$$

Obtaining the solution of the optimisation problem in (1) is, in general, infeasible. However, we prove that a computable upper bound is

$$\rho(\underline{T})\leq\overline{\rho}(\underline{T}) := \max\left\{\max_{x,y\in\mathscr{X}}\left([\overline{T}\mathbb{I}_A](x)-[\underline{T}\mathbb{I}_A](y)\right)\colon \emptyset\neq A\subset\mathscr{X}\right\}, \qquad (2)$$

where $\overline{T}\mathbb{I}_A := -\underline{T}(-\mathbb{I}_A)$. The following novel theorem is useful because, for all $f\in\mathscr{L}(\mathscr{X})$, $m\in\mathbb{N}$ and $\delta\in\mathbb{R}_{\geq 0}$ such that $\delta\|\underline{Q}\|\leq 2$,

$$\|(I+\delta\underline{Q})^m f\|_c \leq \rho((I+\delta\underline{Q})^m)\|f\|_c \leq \overline{\rho}((I+\delta\underline{Q})^m)\|f\|_c. \qquad (3)$$

Theorem 2. *If $\underline{Q}$ is ergodic (De Bock, 2017), then there is some $n<|\mathscr{X}|$ such that, for any $m\geq n$ and any $\delta\in\mathbb{R}_{>0}$ that satisfies $\delta\|\underline{Q}\|<2$,*

$$\rho((I+\delta\underline{Q})^m)\leq\overline{\rho}((I+\delta\underline{Q})^m)<1.$$

Uniform approximation method

The uniform approximation method was introduced by (Krak et al., 2017). They suggest to approximate $\underline{T}_t f$ with $\Psi(\delta,n)f$, where

$$\Psi(\delta,n) := (I+\delta\underline{Q})^n$$

and $t=n\delta$. Given some desired maximal error $\varepsilon\in\mathbb{R}_{>0}$, they propose a way to select the required number of grid steps n—or equivalently, the step size $\delta=t/n$—which a priori guarantees that $\|\underline{T}_t f-\Psi(\delta,n)f\|\leq\varepsilon$. We modify their method in two ways:

(i) we use a less conservative lower bound for n; and
(ii) we a posteriori compute a tighter guaranteed error bound ε'.

Algorithm 1: Uniform approximation

$g_0\leftarrow f$, $\varepsilon'\leftarrow 0$
$n\leftarrow\lceil\max\{t\|\underline{Q}\|/2, t^2\|\underline{Q}\|^2\|f\|_c/\varepsilon\}\rceil$
$\delta\leftarrow t/n$
for $i=0,\dots,n-1$ **do**
  $\varepsilon'\leftarrow\varepsilon'+\delta^2\|\underline{Q}\|^2\|g_i\|_c$ ▷ *If interested in a tighter error bound*
  $g_{i+1}\leftarrow g_i+\delta\underline{Q}g_i$
return $\underline{T}_t f=g_n\pm\varepsilon$ (or $\underline{T}_t f=g_n\pm\varepsilon'$)

As a consequence of Theorems 1 and 2 and Eqn. (3), in case $\underline{Q}$ is ergodic, an alternative a priori guaranteed upper bound for the error is

$$\|\underline{T}_t f-\Psi(\delta,n)f\| \leq \delta^2\|\underline{Q}\|^2\|f\|_c\frac{1-\alpha^k}{1-\alpha} \leq \delta^2\|\underline{Q}\|^2\|f\|_c\frac{1-\beta^k}{1-\beta}, \qquad (4)$$

where $k:=\lceil n/m\rceil$, $\alpha:=\rho((I+\delta\underline{Q})^m)$ and $\beta:=\overline{\rho}((I+\delta\underline{Q})^m)$.

Adaptive approximation method

We observe that in practice, the a posteriori determined error bound ε' is often much smaller than the desired maximal error ε. By combining Theorems 1 and 2, we find that one way to get the posterior error bound closer to ε is to increase the step size δ over time.

In the adaptive approximation method we propose, we achieve this by re-evaluating the step size after every m iterations.

Algorithm 2: Adaptive approximation

$g_0\leftarrow f$, $\Delta\leftarrow t$, $i\leftarrow 0$, $\varepsilon'\leftarrow 0$
while $\Delta>0$ **and** $\|g_i\|_c>0$ **do**
  $i\leftarrow i+1$
  $\delta_i\leftarrow\min\{\Delta, 2/\|\underline{Q}\|, \varepsilon/(t\|\underline{Q}\|^2\|g_{i-1}\|_c)\}$
  if $m\delta_i>\Delta$ **then**
    $m_i\leftarrow\lceil\Delta/\delta_i\rceil$
    $\delta_i\leftarrow\Delta/m_i$
  else $m_i\leftarrow m$
  $g_i\leftarrow g_{i-1}$
  repeat m_i **times**
    $\varepsilon'\leftarrow\varepsilon'+\delta_i^2\|\underline{Q}\|^2\|g_i\|_c$ ▷ *If interested in a tighter error bound*
    $g_i\leftarrow g_i+\delta_i\underline{Q}g_i$
  $\Delta\leftarrow\Delta-m_i\delta$
return $\underline{T}_t f=g_i\pm\varepsilon$ (or $\underline{T}_t f=g_i\pm\varepsilon'$)

Computational comparison

We compare the uniform and adaptive approximation methods using the Healthy-Sick model introduced in (Krak et al., 2017). The obtained results are collected in the table below, where n is the number of iterations and D (D') is the duration in seconds of the computations without (with) keeping track of ε'. We chose $\varepsilon=10^{-4}$.

	n	D	D'	ε'
Uniform	80 000	0.414	1.19	4.29×10^{-5}
Adaptive ($m=1$)	34 360	0.593	0.856	1.00×10^{-4}
Adaptive ($m=10$)	34 369	0.224	0.529	1.00×10^{-4}

Approximating $\underline{\mathrm{E}}_\infty(f)$

Let Q be the transition rate matrix of a stationary and ergodic precise Markov chain. Then

$$\lim_{t\to+\infty}[T_t f](x)=\mathrm{E}_\infty(f) \quad \text{for all } f\in\mathscr{L}(\mathscr{X}) \text{ and all } x\in\mathscr{X}.$$

It is well known that E_∞ is the unique expectation operator that satisfies

$$\mathrm{E}_\infty(Qf)=0 \quad \text{for all } f\in\mathscr{L}(\mathscr{X}).$$

Consequently, it is also the unique expectation operator that, for all $\delta\in\mathbb{R}_{>0}$ such that $\delta\|Q\|<2$, satisfies

$$\mathrm{E}_\infty((I+\delta Q)f)=\mathrm{E}_\infty(f) \quad \text{for all } f\in\mathscr{L}(\mathscr{X}).$$

By the theory of discrete-time Markov chains, the above equality actually implies that, for **all** $\delta\in\mathbb{R}_{>0}$ such that $\delta\|Q\|<2$,

$$\mathrm{E}_\infty(f)=\lim_{n\to+\infty}(I+\delta Q)^n f \quad \text{for all } f\in\mathscr{L}(\mathscr{X}).$$

In the imprecise case, all these nice connections do not necessarily hold. Let $\underline{Q}$ be the lower transition rate operator of an ergodic iMC. Then

$$\lim_{t\to+\infty}[\underline{T}_t f](x)=\underline{\mathrm{E}}_\infty(f) \quad \text{for all } f\in\mathscr{L}(\mathscr{X}) \text{ and all } x\in\mathscr{X},$$

where $\underline{\mathrm{E}}_\infty$ is a lower expectation operator. Unfortunately, it does not hold in general that

$$\underline{\mathrm{E}}_\infty(\underline{Q}f)=0 \quad \text{for all } f\in\mathscr{L}(\mathscr{X}),$$

or that, for all $\delta\in\mathbb{R}_{>0}$ such that $\delta\|\underline{Q}\|<2$,

$$\underline{\mathrm{E}}_\infty((I+\delta\underline{Q})f)=\underline{\mathrm{E}}_\infty(f) \quad \text{for all } f\in\mathscr{L}(\mathscr{X}).$$

Therefore, to the best of our knowledge, the only way to approximate $\underline{\mathrm{E}}_\infty(f)=\lim_{t\to+\infty}[\underline{T}_t f](x)$ is to use an approximation $\Phi_t f$ of $\underline{T}_t f$. If $\|\underline{T}_t f-\Phi_t f\|\leq\varepsilon/2$ and $\|\Phi_t f\|_c\leq\varepsilon/2$, then

$$\left|\underline{\mathrm{E}}_\infty(f)-\frac{\max\Phi_t f+\min\Phi_t f}{2}\right|\leq\varepsilon.$$