A tutorial on

Imprecise Markov chains

by Jasper De Bock & Thomas Krak

SMPS/BELIEF 2018

September 17

now :-(
Jasper De Bock & Thomas Krak
A tutorial on

Imprecise Markov chains

by Jasper De Bock & Thomas Krak
September 17, 2018
now :-)

?
Imprecise Markov chains

by Jasper De Bock & Thomas Krak

September 17

now :-)

(Walley 1991)
(Augustin et al. 2014)
Imprecise Markov chains

by Jasper De Bock & Thomas Krak

A tutorial on SMPS/BELIEF 2018

September 17 2018

now :-)
A tutorial on

Imprecise Markov chains

by Jasper De Bock & Thomas Krak

SMPS/BELIEF 2018

September 17

now :-)

GHENT UNIVERSITY
A tutorial on Imprecise Markov chains by Jasper De Bock & Thomas Krak

SMPS/BELIEF 2018

September 17

now :-}
stochastic process
\[ R_{\geq 0} \]

\{\text{continuous, discrete}\} \text{-time stochastic process}

\[ X_0 \rightarrow X_t \]

\[ 0 \rightarrow t \]
\( \mathbb{R}_{\geq 0} \)

\[
\left\{ \begin{array}{c}
\text{continuous} \\
\text{discrete}
\end{array} \right\} \text{-time stochastic process}
\]

\[
P(X_0 = x_0)
\]

\[
P(X_{t_{n+1}} = y \mid X_{t_1} = x_{t_1}, \ldots, X_{t_{n-1}} = x_{t_{n-1}}, X_{t_n} = x)
\]

\[
X_0 \quad X_{t_1} \quad X_{t_{n-1}} \quad X_{t_n} \quad X_{t_{n+1}}
\]

0 \quad t_1 \quad t_{n-1} \quad t_n \quad t_{n+1}
\[ P(X_0 = x_0) \]

\[ P(X_{t_{n+1}} = y \mid X_{t_1} = x_{t_1}, \ldots, X_{t_{n-1}} = x_{t_{n-1}}, X_{t_n} = x) = P(X_{t_{n+1}} = y \mid X_{t_1} = X_{t_n} = x) \]
We consider continuous and discrete $\mathbb{N}$-time Markov chains, with the homogeneous property:

- $\mathbb{R}_{\geq 0} \ni X \in \{\text{continuous, discrete}\}$
- $X_{t_{n+1}} = y \mid X_{t_1} = x_{t_1}, \ldots, X_{t_{n-1}} = x_{t_{n-1}}, X_{t_n} = x$

only the time difference $\Delta = t_{n+1} - t_n$ matters!

$P(X_0 = x_0)$

$P(X_{t_{n+1}} = y \mid X_{t_1} = x_{t_1}, \ldots, X_{t_{n-1}} = x_{t_{n-1}}, X_{t_n} = x)$

$= P(X_{t_{n+1}} = y \mid X_{t_1} = X_{t_n} = x)$

$= T_\Delta(x, y)$
\( \mathbb{R}_{\geq 0} \)

\( \{ \text{continuous} \} \cup \{ \text{discrete} \} \)

\( \text{homogeneous} \)

\( \gamma \)-time Markov chain

\[
P(X_0 = x_0)
\]

\[
P(X_{t_{n+1}} = y \mid X_{t_1} = x_{t_1}, \ldots, X_{t_{n-1}} = x_{t_{n-1}}, X_{t_n} = x)
\]

\[
= P(X_{t_{n+1}} = y \mid X_{t_1} = X_{t_n} = x)
\]

\[
= T_\Delta(x, y)
\]

only the time difference \( \Delta = t_{n+1} - t_n \) matters!
$P(X_0 = x_0)$ that's just a probability mass function $\pi_0(x_0)$
\[ P(X_0 = x_0) \text{ is just a probability mass function } \pi_0(x_0) \]

that's just a probability mass function

\[ \sum_y T(x, y) = 1 \]

\[ (\forall y) \; T(x, y) \geq 0 \]

transition matrix

\[ T_{\Delta} = T^\Delta \quad \text{with} \quad T := T_1 \]

\[ \mathbb{R} \geq 0 \]

\{ \text{continuous} \}

\{ \text{discrete} \} \quad \gamma

- time Markov chain

\[ X_0 \quad X_{t_1} \quad X_{t_{n-1}} \quad X_{t_n} \quad X_{t_{n+1}} \quad 0 \quad t_1 \quad t_{n-1} \quad t_n \quad t_{n+1} \]
\[ T = \begin{bmatrix} 0.6 & 0.3 & 0.1 \\ 0.2 & 0.3 & 0.5 \\ 0.4 & 0.1 & 0.5 \end{bmatrix} \]

\[ \sum_y T(x, y) = 1 \]

\[ (\forall y) \ T(x, y) \geq 0 \]

transition matrix

\[ T_{\Delta} = T^{\Delta} \quad \text{with} \quad T := T_1 \]

\[
X_0 \quad X_{t_1} \quad X_{t_{n-1}} \quad X_{t_n} \quad X_{t_{n+1}}
\]

0 \quad t_1 \quad t_{n-1} \quad t_n \quad t_{n+1}
$$P(X_0 = x_0) \quad \text{that's just a probability mass function } \pi_0(x_0)$$

initial distribution
A homogeneous \( n \)-time Markov chain is a model for systems with transitions between states. The probability of starting in state \( x_0 \) is given by \( P(X_0 = x_0) \), which is just a probability mass function \( \pi_0(x_0) \).

The initial distribution is defined by:

\[
P(X_0 = x_0) = \pi_0(x_0)
\]

The transition rate matrix \( Q \) is defined as:

\[
Q := \lim_{\Delta \to 0} \frac{T_\Delta - I}{\Delta}
\]

where \( T_\Delta := e^{Q\Delta} := \lim_{n \to \infty} \left( I + \frac{\Delta}{n} Q \right)^n \) is the transition matrix. The condition that ensures the chain is homogeneous is:

\[
\sum_y Q(x, y) = 0 \quad (\forall y \neq x)
\]

and the transition rate matrix must be non-negative:

\[
Q(x, y) \geq 0
\]
Transition rate matrix

\[ Q = \begin{bmatrix}
-4 & 3 & 1 & 0 \\
4 & -6 & 2 & 0 \\
2 & 3 & -6 & 1 \\
0 & 0 & 2 & -2 \\
\end{bmatrix} \]

\[ Q(x, y) = 0 \quad (\forall y \neq x) \quad Q(x, y) \geq 0 \]

\[ Q := \lim_{\Delta \to 0} \frac{T_{\Delta} - I}{\Delta} \]
The set \( \mathbb{R}_{\geq 0} \) consists of continuous \{ continuous, discrete \} -time Markov chains.

The initial distribution is given by \( \pi_0 \).

The transition rate matrix is denoted by \( Q \), or the transition matrix by \( T \).

The random variables are \( X_0 \) and \( X_t \) at times 0 and \( t \), respectively.
\[ E(f(X_t)|X_0 = x) = [T_t f](x) \]
\[ = \sum_y T_t(x, y) f(y) = \begin{cases} 
\sum_y e^{Qt(x, y)} f(y) \\
\sum_y T^t(x, y) f(y)
\end{cases} \]

\[ x = 0 \]
\[ f(X_t) = X_t \]
\[ E(f(X_t)|X_0 = x) = [T_t f](x) \]
\[ = \sum_y T_t(x, y) f(y) = \begin{cases} \sum_y e^{Qt(x, y)} f(y) \\ \sum_y T^t(x, y) f(y) \end{cases} \]

\[ P(X_t = y|X_0 = x) = E(\mathbb{I}_y(X_t)|X_0 = x) = [T_t \mathbb{I}_y](x) \]

\[
\begin{array}{l}
y = x = 0 \\
f(X_t) = \mathbb{I}_y(X_t) = \begin{cases} 1 & \text{if } X_t = y \\ 0 & \text{otherwise} \end{cases}
\end{array}
\]
\[ E(f(X_t)|X_0 = x) = [T_t f](x) \]

\[
E_\infty(f) := \lim_{t \to +\infty} E(f(X_t)|X_0 = x)
\]

\[ P(X_t = y|X_0 = x) = E(\mathbb{I}_y(X_t)|X_0 = x) = [T_t \mathbb{I}_y](x) \]

\[
\pi_\infty(y) := \lim_{t \to +\infty} P(X_t = y|X_0 = x)
\]
Reliability engineering (failure probabilities, …)

Queuing theory (waiting in line …)
- optimising supermarket waiting times
- dimensioning of call centers
- airport security lines
- router queues on the internet

Chemical reactions (time-evolution …)

Pagerank

...
So how about imprecision?
So how about imprecision?

What if we don’t know $T$ or $Q$ exactly?
Sets of transition (rate) matrices

Don’t know $T$ (or $Q$) exactly

But confident that $T \in \mathcal{T}$ for some set $\mathcal{T}$ of transition matrices
  - (or that $Q \in \mathcal{D}$ for some set $\mathcal{D}$ of rate matrices)

Induces imprecise Markov chain; set of processes compatible with $\mathcal{T}$.  

Jasper De Bock, Thomas Krak
Imprecise Markov Chains
Sets of transition (rate) matrices

Don’t know \( T \) (or \( Q \)) exactly

But confident that \( T \in \mathcal{T} \) for some set \( \mathcal{T} \) of transition matrices

\( \) (or that \( Q \in \mathcal{D} \) for some set \( \mathcal{D} \) of rate matrices)

Induces *imprecise Markov chain*; set of processes *compatible* with \( \mathcal{T} \).

Different versions:

\( \mathcal{P}^{\text{HM}}_{\mathcal{T}} \): all homogeneous Markov chains with \( T \in \mathcal{T} \)
Sets of transition (rate) matrices

Don’t know $T$ (or $Q$) exactly

But confident that $T \in \mathcal{T}$ for some set $\mathcal{T}$ of transition matrices

- (or that $Q \in \mathcal{Q}$ for some set $\mathcal{Q}$ of rate matrices)

Induces *imprecise Markov chain; set of processes compatible* with $\mathcal{T}$.

Different versions:

- $\mathcal{P}^{\text{HM}}_{\mathcal{T}}$: all homogeneous Markov chains with $T \in \mathcal{T}$
- $\mathcal{P}^{\text{M}}_{\mathcal{T}}$: all (non-homogeneous) Markov chains with $T^{(t)} \in \mathcal{T}$
Sets of transition (rate) matrices

Don’t know $T$ (or $Q$) exactly

But confident that $T \in \mathcal{T}$ for some set $\mathcal{T}$ of transition matrices

- (or that $Q \in \mathcal{Q}$ for some set $\mathcal{Q}$ of rate matrices)

Induces *imprecise Markov chain; set of processes compatible* with $\mathcal{T}$.

Different versions:

- $\mathbb{P}^{\text{HM}}_{\mathcal{T}}$: all homogeneous Markov chains with $T \in \mathcal{T}$
- $\mathbb{P}^{\text{M}}_{\mathcal{T}}$: all (non-homogeneous) Markov chains with $T(t) \in \mathcal{T}$
- $\mathbb{P}_{\mathcal{T}}$: all (non-Markov) processes with $T(t,x_u) \in \mathcal{T}$
Sets of transition (rate) matrices

Don’t know $T$ (or $Q$) exactly

But confident that $T \in \mathcal{T}$ for some set $\mathcal{T}$ of transition matrices

- (or that $Q \in \mathcal{D}$ for some set $\mathcal{D}$ of rate matrices)

Induces *imprecise Markov chain*; set of processes *compatible* with $\mathcal{T}$.

Different versions:

- $\mathcal{P}_{\mathcal{T}}^{HM}$: all homogeneous Markov chains with $T \in \mathcal{T}$
- $\mathcal{P}_{\mathcal{T}}^{M}$: all (non-homogeneous) Markov chains with $T(t) \in \mathcal{T}$
- $\mathcal{P}_{\mathcal{T}}$: all (non-Markov) processes with $T^{(t,x_u)} \in \mathcal{T}$

Clearly

\[
\mathcal{P}_{\mathcal{T}}^{HM} \subseteq \mathcal{P}_{\mathcal{T}}^{M} \subseteq \mathcal{P}_{\mathcal{T}}
\]
Lower expectations and lower probabilities

Given an imprecise Markov chain $\mathbb{P}^*$, we are interested in

$$\mathbb{E}^*_{\mathcal{F}}[f(X_t) \mid X_0 = x] = \inf_{P \in \mathbb{P}^*} \mathbb{E}_P[f(X_t) \mid X_0 = x]$$

(And $\mathbb{E}^*_{\mathcal{F}}[f(X_t) \mid X_0 = x]$ by conjugacy)

Lower- (and upper) probabilities a special case:

$$P^*_{\mathcal{F}}(X_t = y \mid X_0 = x) = \inf_{P \in \mathbb{P}^*} P(X_t = y \mid X_0 = x) = \mathbb{E}^*_{\mathcal{F}}[\mathbb{I}_y(X_t) \mid X_0 = x]$$

Because different types are nested,

$$\mathbb{E}_{\mathcal{F}}[f(X_t) \mid X_0 = x] \leq \mathbb{E}^M_{\mathcal{F}}[f(X_t) \mid X_0 = x] \leq \mathbb{E}^{HM}_{\mathcal{F}}[f(X_t) \mid X_0 = x]$$
Computing lower expectations, first try

Recall that for a homogeneous Markov chain $P$ with transition matrix $T$,

$$\mathbb{E}_P[f(X_1) | X_0 = x] = [Tf](x).$$

Now consider $P_{\mathcal{F}}^H$. Then,

$$\mathbb{E}_{P_{\mathcal{F}}}^H[f(X_1) | X_0 = x] := \inf_{P \in P_{\mathcal{F}}^H} \mathbb{E}_P[f(X_1) | X_0 = x]$$

$$= \inf_{T \in \mathcal{F}} [Tf](x)$$

- Linear optimisation problem with constraints given by $\mathcal{F}$
- Relatively straightforward if $\mathcal{F}$ is “nice”
- Essentially solving a linear programming problem
Computing lower expectations, first try

Recall that for a homogeneous Markov chain $P$ with transition matrix $T$,

$$
\mathbb{E}_P[f(X_t) \mid X_0 = x] = [T^t f](x).
$$

Now consider $\mathcal{P}^{HM}$. Then,

$$
\mathbb{E}_{\mathcal{T}}^{HM}[f(X_t n) \mid X_0 = x] := \inf_{P \in \mathcal{P}^{HM}} \mathbb{E}_P[f(X_t) \mid X_0 = x]
$$

$$
= \inf_{T \in \mathcal{T}} [T^t f](x)
$$

- **Non**-linear optimisation problem with constraints given by $\mathcal{T}$
- **Not** straightforward even if $\mathcal{T}$ is “nice”
Computing lower expectations, first try

Recall that for a homogeneous Markov chain $P$ with transition matrix $T$,

$$E_P[f(X_t) | X_0 = x] = [T^t f](x).$$

Now consider $\mathbb{P}^{HM}$ . Then,

$$\mathbb{E}^{HM}_{\mathcal{I}}[f(X_t^n) | X_0 = x] := \inf_{P \in \mathbb{P}^{HM}} \mathbb{E}_P[f(X_t) | X_0 = x]$$

$$= \inf_{T \in \mathcal{T}} [T^t f](x)$$

- **Non**-linear optimisation problem with constraints given by $\mathcal{T}$
- **Not** straightforward even if $\mathcal{T}$ is “nice”

See e.g. (Kozine and Utkin, 2002) and (Campos et al., 2003) for analyses of this approach.
What about the non-Markov case?

\[ \mathbb{P}_T: \text{all (non-Markov) processes with } T(t, x_u) \in T \]

How to interpret this?
⇒ Helps to draw a picture
What about the non-Markov case?

\[ \mathbb{P}_T : \text{all (non-Markov) processes with } T(t, x_u) \in T \]

How to interpret this?
⇒ Helps to draw a picture

Example with binary state space \( \mathcal{X} = \{a, b\} \)

Use event tree / probability tree
Illustration of behaviour over time

Need notation
\[ \mathcal{T}_x := \{ T(x, \cdot) | T \in T \} \quad \forall x \in \mathcal{X} \]
What about the non-Markov case?

\[ \mathbb{P}_T: \text{all (non-Markov) processes with } T(t,x_u) \in T \]

How to interpret this?
\[ \Rightarrow \text{Helps to draw a picture} \]

Example with binary state space \( \mathcal{X} = \{a, b\} \)

Use event tree / probability tree
Illustration of behaviour over time

This setting explored by (De Cooman et al., 2009).

Tree representation related to (Shafer and Vovk, 2001) game-theoretic probabilities. Connection to (Walley’s) imprecise probabilities in (De Cooman and Hermans, 2008).
Visualising a stochastic process

\[ \begin{align*}
\emptyset & \quad \pi_0 \\
a & \quad P(X_1 | X_0 = a) \\
\quad & \quad P(X_2 | X_0 = a, X_1 = a) \\
ab & \quad P(X_2 | X_0 = a, X_1 = b) \\
ba & \quad P(X_2 | X_0 = b, X_1 = a) \\
bb & \quad P(X_2 | X_0 = b, X_1 = b) \\
b & \quad P(X_1 | X_0 = b) \\
\end{align*} \]
Visualising a stochastic process in $\mathbb{P}_\mathcal{I}$

\begin{align*}
\emptyset & \quad \pi_0 \\
 a & \quad \begin{array}{c}
\text{time} \\
0 \quad 1 \quad 2
\end{array} \\
\quad & \quad \begin{array}{c}
P(X_1 \mid X_0 = a) \quad \in \mathcal{I}_a \\
P(X_2 \mid X_0 = a, X_1 = a) \quad \in \mathcal{I}_a
\end{array} \\
\quad & \quad \begin{array}{c}
P(X_1 \mid X_0 = b) \quad \in \mathcal{I}_b \\
P(X_2 \mid X_0 = a, X_1 = b) \quad \in \mathcal{I}_b
\end{array} \\
\quad & \quad \begin{array}{c}
P(X_2 \mid X_0 = b, X_1 = a) \quad \in \mathcal{I}_a \\
P(X_2 \mid X_0 = b, X_1 = b) \quad \in \mathcal{I}_b
\end{array}
\end{align*}

Jasper De Bock, Thomas Krak

Imprecise Markov Chains
Computations by iterated lower expectation

For the set $\mathbb{P}_{\mathbb{T}}$, it can be shown that

$$
\mathbb{E}_{\mathbb{T}}[f(X_t) \mid X_0 = x] = \mathbb{E}_{\mathbb{T}}\left[\mathbb{E}_{\mathbb{T}}[f(X_t) \mid X_0 = x, X_s] \mid X_0 = x\right] \quad \forall s \leq t
$$

This is the *law of iterated lower expectation*.

Provides *backwards recursive* scheme for computations.

$\Rightarrow$ Intuitive in the tree representation

Example: compute $\mathbb{E}_{\mathbb{T}}[\mathbb{I}_b(X_2) \mid X_0 = a] = \mathbb{P}(X_2 = b \mid X_0 = a)$

$$
\mathcal{T}_a := \left\{ T(a, \cdot) \mid T(a, a) \in [0.4, 0.6] \right\}
$$

$$
\mathcal{T}_b := \left\{ T(b, \cdot) \mid T(b, a) \in [0.1, 0.3] \right\}
$$
Example $\mathbb{E}_\mathcal{F} \left[ I_b(X_2) \mid X_0 = a \right]$ base case
Example $\mathbb{E}_{\mathcal{F}}[\mathbb{I}_b(X_2) \mid X_0 = a]$ base case

\begin{align*}
\mathbb{I}_b(a) &= 0 \\
\mathbb{I}_b(b) &= 1
\end{align*}

\text{time}
Example $\mathbb{E}_T \left[ \mathbb{I}_b(X_2) \mid X_0 = a \right]$ base case

$\mathbb{I}_b(a) = 0$

$\mathbb{I}_b(b) = 1$

Jasper De Bock, Thomas Krak

Imprecise Markov Chains
Example $\mathbb{E}_{\mathcal{G}} [\mathbb{I}_b(X_2) \mid X_0 = a]$ base case

\[
\mathbb{E}_{\mathcal{G}} [\mathbb{I}_b(X_2) \mid X_0 = a, X_1 = a] = \inf_{T(a, \cdot) \in \mathcal{T}_a} T(a, a) \mathbb{I}_b(a) + T(a, b) \mathbb{I}_b(b)
\]

$\mathbb{I}_b(a) = 0$

$\mathbb{I}_b(b) = 1$
Example $\mathbb{E}_{\mathcal{F}} [\mathbb{I}_b(X_2) \mid X_0 = a]$ base case

$\mathcal{F}_a = \{ T(a, \cdot) \mid T(a, a) \in [0.4, 0.6] \}$

$\mathbb{E}_{\mathcal{F}} [\mathbb{I}_b(X_2) \mid X_0 = a, X_1 = a] = \inf_{T(a, \cdot) \in \mathcal{F}_a} T(a, a) \mathbb{I}_b(a) + T(a, b) \mathbb{I}_b(b) = 0.4$

$\mathbb{I}_b(a) = 0$

$\mathbb{I}_b(b) = 1$

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>aa</td>
<td>$\mathbb{I}_b(a) = 0$</td>
</tr>
<tr>
<td>2</td>
<td>aaa</td>
<td>$\mathbb{I}_b(a) = 0$</td>
</tr>
<tr>
<td>2</td>
<td>aab</td>
<td>$\mathbb{I}_b(b) = 1$</td>
</tr>
</tbody>
</table>
Recursive, local computations

\[ \mathbb{E}_{\mathcal{G}} [ I_b(X_2) | X_0 = a, X_1 = a ] = 0.4 \]

\[ \mathbb{E}_{\mathcal{G}} [ I_b(X_2) | X_0 = a, X_1 = b ] = 0.7 \]

\[ \mathbb{E}_{\mathcal{G}} [ I_b(X_2) | X_0 = a, X_1 = a ] = 0.4 \]

\[ \mathbb{E}_{\mathcal{G}} [ I_b(X_2) | X_0 = a, X_1 = b ] = 0.7 \]
Recursive, local computations

\[ \inf_{T(a, \cdot) \in \mathcal{T}_a} \sum_{T(a, \cdot) \in \mathcal{T}_a} T(a, a) \times 0.4 + T(a, b) \times 0.7 \]

\[ \mathbb{E}_{\mathcal{J}} [\mathbb{I}_b(X_2) | X_0 = a, X_1 = a] = 0.4 \]

\[ \mathbb{E}_{\mathcal{J}} [\mathbb{I}_b(X_2) | X_0 = a, X_1 = b] = 0.7 \]

\[ \mathbb{E}_{\mathcal{J}} [\mathbb{I}_b(X_2) | X_0 = a, X_1 = a] = 0.4 \]

\[ \mathbb{E}_{\mathcal{J}} [\mathbb{I}_b(X_2) | X_0 = a, X_1 = b] = 0.7 \]

Jasper De Bock, Thomas Krak

Imprecise Markov Chains
Recursive, local computations

$$\mathbb{E}_\mathcal{T} [\mathbb{I}_b(X_2) | X_0 = a]$$

$$= \inf_{T(a', a) \in \mathcal{T}_a} T(a, a) \times 0.4 + T(a, b) \times 0.7$$

$$= 0.52$$
Local computations in operator form

Consider $\mathcal{T}_X$, and define for any $f : X \to \mathbb{R}$,

$$[Tf](x) := \inf_{T(x, \cdot) \in \mathcal{T}_X} \sum_y T(x, y)f(y)$$

Linear optimisation problem, and

$$[Tf](x) = \inf_{T \in \mathcal{T}} [Tf](x)$$

We call $\mathcal{T}$ the lower transition operator for $\mathcal{T}$.

We can write

$$\mathbb{E}_{\mathcal{T}} \left[ f(X_{t+1}) \mid X_{0:t} = x_{0:t} \right] = [Tf](x_t)$$
Local computations in operator form

Consider $\mathcal{T}_X$, and define for any $f : \mathcal{X} \to \mathbb{R}$,

$$[\underline{T}f](x) := \inf_{T(x, \cdot) \in \mathcal{T}_X} \sum_y T(x, y)f(y)$$

Linear optimisation problem, and

$$[\underline{T}f](x) = \inf_{\underline{T} \in \mathcal{T}} [\underline{T}f](x)$$

We call $\underline{T}$ the lower transition operator for $\mathcal{T}$.

We can write

$$\mathbb{E}_{\mathcal{T}}[f(X_{t+1}) | X_{0:t} = x_{0:t}] = [\underline{T}f](x_t)$$

We find

$$\mathbb{E}_{\mathcal{T}}[f(X_{t+1}) | X_{0:t} = x_{0:t}] = [\underline{T}f](x_t) = \mathbb{E}_{\mathcal{T}}[f(X_{t+1}) | X_t = x_t]$$

Lower envelope for imprecise Markov chain $\mathbb{P}_\mathcal{T}$ has “Markov” property

- But contains non-Markov models!

Similarly the lower envelope is also homogeneous!
Multiple time steps

By repeating the local computations,

\[ \mathbb{E}_\mathcal{T} [f(X_2) | X_0 = x] = [T T f](x), \]

if the set \( \mathcal{T} \) has separately specified rows:

\[
\begin{bmatrix}
1 & 2 \\
3 & 4 \\
5 & 6
\end{bmatrix},
\begin{bmatrix}
7 & 8 \\
9 & 10 \\
11 & 12
\end{bmatrix} \in \mathcal{T} \quad \Rightarrow \quad
\begin{bmatrix}
1 & 2 \\
3 & 4 \\
5 & 6
\end{bmatrix} \in \mathcal{T}
\]

(\( \mathcal{T} \) is closed under recombination of rows)
Multiple time steps

By repeating the local computations,

\[ \mathbb{E}_\mathcal{F} [f(X_2) | X_0 = x] = [TTf](x), \]

if the set \( \mathcal{F} \) has separately specified rows:

\[
\begin{bmatrix}
\text{Orange} \\
\text{Red} \\
\text{Yellow}
\end{bmatrix}
, \quad
\begin{bmatrix}
\text{Green} \\
\text{Teal} \\
\text{Blue}
\end{bmatrix}
\in \mathcal{F}
\implies
\begin{bmatrix}
\text{Blue} \\
\text{Red} \\
\text{Orange}
\end{bmatrix}
\in \mathcal{F}
\]

(\( \mathcal{F} \) is closed under recombination of rows)

By induction we get

\[ \mathbb{E}_\mathcal{F} [f(X_t) | X_0 = x] = [T^tf](x) \]

- Local, linear optimisations only
- Can be efficiently computed
Multiple time steps

By repeating the local computations,

\[
\mathbb{E}_\mathcal{T} [f(X_2) \mid X_0 = x] = [T^T f](x),
\]

if the set \( \mathcal{T} \) has separately specified rows:

\[
\begin{bmatrix}
\text{orange} \\
\text{red} \\
\text{yellow}
\end{bmatrix}, \quad \begin{bmatrix}
\text{green} \\
\text{blue} \\
\text{red}
\end{bmatrix} \in \mathcal{T} \quad \Rightarrow \quad \begin{bmatrix}
\text{green} \\
\text{blue} \\
\text{red}
\end{bmatrix} \in \mathcal{T}
\]

(\( \mathcal{T} \) is closed under recombination of rows)

By induction we get

\[
\mathbb{E}_\mathcal{T} [f(X_t) \mid X_0 = x] = [T^t f](x)
\]

Imprecise Markov chain \( \mathbb{P}_\mathcal{T} \) can be seen as credal network under epistemic irrelevance. Gives a graphical model representation.

“Separately specified rows” is a well-known condition in that context.
That’s two extremes. What about the intermediate one?

So far ignored $\mathbb{P}_{F}^{M}$
That’s two extremes. What about the intermediate one?

So far ignored $\mathbb{P}^M_{\mathcal{F}}$

Turns out that if $\mathcal{F}$ has separately specified rows, then

$$\mathbb{E}^M_{\mathcal{F}}[f(X_t) \mid X_0 = x] = [T^t f](x)$$

If follows that

$$\mathbb{E}^M_{\mathcal{F}}[f(X_t) \mid X_0 = x] = \mathbb{E}_{\mathcal{F}}[f(X_t) \mid X_0 = x]$$

- Does not hold for functions on multiple time points
- Then only $\mathbb{P}_{\mathcal{F}}$ remains tractable
That’s two extremes. What about the intermediate one?

So far ignored $\mathbb{P}^M_{\mathcal{F}}$

Turns out that if $\mathcal{F}$ has separately specified rows, then

$$\mathbb{E}^M_{\mathcal{F}}[f(X_t) \mid X_0 = x] = [T^t f](x)$$

If follows that

$$\mathbb{E}^M_{\mathcal{F}}[f(X_t) \mid X_0 = x] = \mathbb{E}_{\mathcal{F}}[f(X_t) \mid X_0 = x]$$

- Does not hold for functions on multiple time points
- Then only $\mathbb{P}_{\mathcal{F}}$ remains tractable


⇒ No explicit connection to imprecise probabilities

Exploration with imprecise probabilities by (Škulj, 2009)
Limit behaviour?

Limit inference often of interest:

$$\mathbb{E}[f(X_\infty) | X_0 = x] = \lim_{t \to +\infty} \mathbb{E}[f(X_t) | X_0 = x]$$

In imprecise setting, often exists:

$$\mathbb{E}_\mathcal{F}[f(X_\infty) | X_0 = x] := \lim_{t \to +\infty} \mathcal{T}^t f(x),$$

and often independent of $x$.

See e.g. (De Cooman et al., 2009) and (Škulj, 2009)
Summary for imprecise Markov chains in discrete time

Parameterisation through set $\mathcal{I}$ of transition matrices.

Can induce three different \textit{imprecise Markov chains}:

- $\mathbb{P}_{\mathcal{I}}^{\text{HM}}$: all homogeneous Markov chains compatible with $\mathcal{I}$
- $\mathbb{P}_{\mathcal{I}}^{\text{M}}$: all (non-)homogeneous Markov chains compatible with $\mathcal{I}$
- $\mathbb{P}_{\mathcal{I}}$: all (non-)Markov processes compatible with $\mathcal{I}$

For $\mathbb{P}_{\mathcal{I}}^{\text{HM}}$, computations are difficult.

For $\mathbb{P}_{\mathcal{I}}^{\text{M}}$ and $\mathbb{P}_{\mathcal{I}}$, computations using \textit{lower transition operator}

$$E_{\mathcal{I}}^M[f(X_t) | X_0 = x] = E_{\mathcal{I}}[f(X_t) | X_0 = x] = [T^t f](x)$$

The imprecise Markov chain $\mathbb{P}_{\mathcal{I}}$ satisfies an \textit{imprecise Markov property}

The limit $\lim_{t \to +\infty} [T^t f](x)$ often exists, and often independent of $x$. 
Imprecise Continuous-Time Markov Chains

Going to go a bit faster with more intuition

We use the same basic approach:

- Uncertain about $Q$, but consider a set $\mathcal{Q}$
- Three imprecise (continuous-time) Markov chains, compatible with $\mathcal{Q}$:
  - $\mathbb{P}^{HM}_{\mathcal{Q}}$: all homogeneous Markov chains with $Q \in \mathcal{Q}$
  - $\mathbb{P}^{M}_{\mathcal{Q}}$: all (non-)homogeneous Markov chains with $Q_t \in \mathcal{Q}$
  - $\mathbb{P}_{\mathcal{Q}}$: all (non-)Markov processes with $Q_{t,x_u} \in \mathcal{Q}$

Similar to discrete-time case,

$$\mathbb{E}^{HM}_{\mathcal{Q}} \left[ f(X_t) \mid X_0 = x \right] = \inf_{Q \in \mathcal{Q}} [e^{Qt}f](x)$$

which is difficult due to nonlinearities in the optimisation.
Imprecise Continuous-Time Markov Chains

Going to go a bit faster with more intuition

We use the same basic approach:

- Uncertain about $Q$, but consider a set $\mathcal{Q}$
- Three imprecise (continuous-time) Markov chains, compatible with $\mathcal{Q}$:
  - $\mathbb{P}^{\text{HM}}_{\mathcal{Q}}$: all homogeneous Markov chains with $Q \in \mathcal{Q}$
  - $\mathbb{P}^{\text{M}}_{\mathcal{Q}}$: all (non-)homogeneous Markov chains with $Q_t \in \mathcal{Q}$
  - $\mathbb{P}_{\mathcal{Q}}$: all (non-)Markov processes with $Q_{t,x_u} \in \mathcal{Q}$

Similar to discrete-time case,

$$\mathbb{E}^{\text{HM}}_{\mathcal{Q}} [f(X_t) \mid X_0 = x] = \inf_{Q \in \mathcal{Q}} [e^{Q_t} f](x)$$

which is difficult due to nonlinearities in the optimisation.

See e.g. (Goldsztejn and Neumaier, 2014) and (Oppenheimer and Michel, 1988) for details on this homogeneous setting.
Non-homogeneous case in continuous-time

$P_M$: all (non-)homogeneous Markov chains with $Q_t \in \mathcal{D}$

How to interpret this?

Homogeneous case, rate matrix is just a derivative,

$$Q := \lim_{\Delta \to 0} \frac{T_\Delta - I}{\Delta} \quad \text{where} \quad T_\Delta(x, y) := P(X_\Delta = y | X_0 = x)$$
Non-homogeneous case in continuous-time

$P_M^Q$: all (non-homogeneous) Markov chains with $Q_t \in \mathcal{Q}$

How to interpret this?

Homogeneous case, rate matrix is just a derivative,

$$Q := \lim_{\Delta \to 0} \frac{T_\Delta - I}{\Delta} \quad \text{where} \quad T_\Delta(x, y) := P(X_\Delta = y \mid X_0 = x)$$

For non-homogeneous case we write

$$T_{t+\Delta}(x, y) := P(X_{t+\Delta} = y \mid X_t = x),$$

which has a time-dependent derivative,

$$Q_t := \lim_{\Delta \to 0} \frac{T_{t+\Delta} - T_t}{\Delta} = \lim_{\Delta \to 0} \frac{T_{t+\Delta} - I}{\Delta}$$
Non-homogeneous case in continuous-time

$\mathbb{P}_Q^M$: all (non-homogeneous) Markov chains with $Q_t \in \mathcal{Q}$

How to interpret this?

Homogeneous case, rate matrix is just a derivative,

$$Q := \lim_{\Delta \to 0} \frac{T_\Delta - I}{\Delta} \quad \text{where} \quad T_\Delta(x, y) := P(X_\Delta = y | X_0 = x)$$

For non-homogeneous case we write

$$T_{t+\Delta}(x, y) := P(X_{t+\Delta} = y | X_t = x),$$

which has a \textit{time-dependent} derivative,

$$Q_t := \lim_{\Delta \to 0} \frac{T_{t+\Delta} - T_t}{\Delta} = \lim_{\Delta \to 0} \frac{T_{t+\Delta} - I}{\Delta}$$

Setting explored by (Hartfiel, 1985) and (Škulj, 2015)
Continuous-time local models

We have

\[ Q_t = \lim_{\Delta \to 0} \frac{T_{t+\Delta} - I}{\Delta} \]

and so for small \( \Delta \),

\[ T_{t+\Delta} \approx I + \Delta Q_t \]
Continuous-time local models

We have

\[ Q_t = \lim_{\Delta \to 0} \frac{T_{t+\Delta} - I}{\Delta} \]

and so for small \( \Delta \),

\[ T_{t+\Delta} \approx I + \Delta Q_t \]

Then we can write

\[ \mathbb{E}^M_{\mathcal{Q}} [f(X_{t+\Delta}) \mid X_t = x] = \inf_{T_{t+\Delta}} [T_{t+\Delta} f](x) \approx \inf_{Q \in \mathcal{Q}} [(I + \Delta Q) f](x) \]
Continuous-time local models

We have

\[ Q_t = \lim_{\Delta \to 0} \frac{T_t^{t+\Delta} - I}{\Delta} \]

and so for small \( \Delta \),

\[ T_t^{t+\Delta} \approx I + \Delta Q_t \]

Then we can write

\[ \mathbb{E}^{\mathcal{M}}_{\mathcal{Q}}[f(X_{t+\Delta}) \mid X_t = x] = \inf_{T_t^{t+\Delta}} \left[ T_t^{t+\Delta} f \right](x) \approx \inf_{Q \in \mathcal{Q}} [(I + \Delta Q)f](x) \]

We get

\[ \mathbb{E}^{\mathcal{M}}_{\mathcal{Q}}[f(X_{t+\Delta}) \mid X_t = x] \approx [(I + \Delta Q)f](x) \approx \mathbb{E}^{\mathcal{M}}_{\mathcal{Q}}[f(X_{\Delta}) \mid X_0 = x] \]

where we have defined

\[ [Qf](x) := \inf_{Q \in \mathcal{Q}} [Qf](x), \]

Again homogeneous lower expectation!
Arbitrary time points

If $Q$ has separately specified rows,

$$\mathbb{E}^M_{D} [f(X_t) \mid X_0 = x] \approx [(I + t/nQ)^n f](x)$$

and in fact

$$\mathbb{E}^M_{D} [f(X_t) \mid X_0 = x] = \lim_{n \to +\infty} [(I + t/nQ)^n f](x)$$

Allows practical computation

- Solve $\inf_{Q \in Q}[Q \cdot]$ multiple times
- Each is a linear optimisation problem
Arbitrary time points

If $\mathcal{D}$ has separately specified rows,

$$
\mathbb{E}^\mathcal{M}_{\mathcal{D}}[f(X_t) \mid X_0 = x] \approx [(l + t/nQ)^n f](x)
$$

and in fact

$$
\mathbb{E}^\mathcal{M}_{\mathcal{D}}[f(X_t) \mid X_0 = x] = \lim_{n \to +\infty} [(l + t/nQ)^n f](x)
$$

Allows practical computation

- Solve $\inf_{Q \in \mathcal{D}}[Q \cdot]$ multiple times
- Each is a linear optimisation problem

Better computational method in (Erreygers and De Bock, 2017)
The non-Markov case

For the set $\mathbb{P}_Q$, derivative becomes *history* dependent.

Let $x_u = x_{u_1}, \ldots, x_{u_n}$, $0 \leq u_1 < \cdots < u_n < t$. For all $x, y \in \mathcal{X}$,

$$Q_{t,x_u}(x, y) := \lim_{\Delta \to 0} \frac{P(X_{t+\Delta} = y | X_u = x_u, X_t = x) - I(x, y)}{\Delta}$$

This is becoming a bit unwieldy...
The non-Markov case

For the set $\mathbb{P}_Q$, derivative becomes *history* dependent.

Let $x_u = x_{u_1}, \ldots, x_{u_n}$, $0 \leq u_1 < \cdots < u_n < t$. For all $x, y \in \mathcal{X}$,

$$Q_{t,x_u}(x, y) := \lim_{\Delta \to 0} \frac{P(X_{t+\Delta} = y \mid X_u = x_u, X_t = x) - I(x, y)}{\Delta}$$

This is becoming a bit unwieldy...

Turns out that

$$\mathbb{E}_Q[f(X_{s+t}) \mid X_u = x_u, X_s = x] = \lim_{n \to +\infty} [(I + t/nQ)^n f](x)$$
The non-Markov case

For the set $\mathbb{P}_Q$, derivative becomes \textit{history} dependent.
Let $x_u = x_{u_1}, \ldots, x_{u_n}, 0 \leq u_1 < \cdots < u_n < t$. For all $x, y \in \mathcal{X}$,

$$Q_{t,x_u}(x, y) := \lim_{\Delta \to 0} \frac{P(X_{t+\Delta} = y \mid X_u = x_u, X_t = x) - I(x, y)}{\Delta}$$

This is becoming a bit unwieldy...

Turns out that

$$\mathbb{E}_Q[f(X_{s+t}) \mid X_u = x_u, X_s = x] = \lim_{n \to +\infty} [(I + t/nQ)^n f](x) = \mathbb{E}_Q^M[f(X_t) \mid X_0 = x]$$

Lower expectation for $\mathbb{P}_Q$ has an \textit{imprecise Markov property}!
- And is time-homogeneous!
- \textbf{Not} the same as $\mathbb{P}_M^Q$ when $f$ depends on multiple time points!
  - Then only $\mathbb{P}_Q$ remains tractable.
The non-Markov case

For the set \( \mathbb{P}_Q \), derivative becomes *history* dependent. Let \( x_u = x_{u_1}, \ldots, x_{u_n}, 0 \leq u_1 < \cdots < u_n < t \). For all \( x, y \in \mathcal{X} \),

\[
Q_{t,x_u}(x,y) := \lim_{\Delta \to 0} \frac{P(X_{t+\Delta} = y \mid X_u = x_u, X_t = x) - I(x,y)}{\Delta}
\]

This is becoming a bit unwieldy...

Turns out that

\[
\mathbb{E}_Q[f(X_{s+t}) \mid X_u = x_u, X_s = x] = \lim_{n \to +\infty} (I + t/nQ)^n f](x) = \mathbb{E}_M[f(X_t) \mid X_0 = x]
\]

Lower expectation for \( \mathbb{P}_Q \) has an *imprecise Markov property*!

- And is time-homogeneous!
- **Not** the same as \( \mathbb{P}_M \) when \( f \) depends on multiple time points!
  - Then only \( \mathbb{P}_Q \) remains tractable.

Explored by (Krak et al., 2017)
Continuous-time limit behaviour?

Limit inference often of interest:

\[
\mathbb{E}[f(X_\infty) | X_0 = x] = \lim_{t \to +\infty} \mathbb{E}[f(X_t) | X_0 = x]
\]

In imprecise setting, limit *always* exists:

\[
\mathbb{E}_\mathcal{Q}[f(X_\infty) | X_0 = x] = \lim_{t \to +\infty} \mathbb{E}_\mathcal{Q}[f(X_t) | X_0 = x]
\]

and often independent of \(x\).

See (De Bock, 2017)
Main take away points

If we do not know $T$ or $Q$, we can consider sets $T$ or $Q$

Gives rise to three different *imprecise* models:
- Set of homogeneous Markov chains
- Set of non-homogeneous Markov chains
- Set of non-Markov processes

For homogeneous Markov chains:
- *Difficult* to work with

For non-homogeneous and non-Markov processes:
- Efficient computations using *local models* $T$ or $Q$
- Have *homogeneous* lower expectations
- Have “*Markov*” lower expectations
That’s all fine and well, but what can you use it for?
Reliability engineering (failure probabilities, …)

Queuing theory (waiting in line …)
- optimising supermarket waiting times
- dimensioning of call centers
- airport security lines
- router queues on the internet

Chemical reactions (time-evolution …)

Pagerank

...

Google
Message passing in optical links

$m_1 \text{ channels} \quad m_2 = \frac{m_1}{n_2} \quad \text{superchannels}$

**type I** messages require 1 channel

**type II** messages require $n_2$ channels

We want to know the **blocking probability** of messages for a given policy, and optimise it
\[ \mathcal{X}_{\text{det}} := \left\{ (i_0, \ldots, i_{n_2}) \in \mathbb{N}^{(n_2+1)} : \sum_{k=0}^{n_2} i_k \leq m_2 \right\} \]

\[ I := \sum_{k=0}^{n_2} i_k \quad R := \sum_{k=0}^{n_2-1} i_k (n_2 - k) \]
Amorous
Bickering
Confusion
Depression

![Amorous Relation Diagram](image)

### Matrix Q

\[
Q = \begin{bmatrix}
-4 & 3 & 1 & 0 \\
4 & -6 & 2 & 0 \\
2 & 3 & -6 & 1 \\
0 & 0 & 2 & -2
\end{bmatrix}
\]

(Erreygers & De Bock 2018)
\( \mathcal{X}_{\text{red}} := \{ (i, j, e) \in \mathbb{N}^3 : m_2 \leq i + j + e, i + (j + e)n_2 \leq m_1 \} \)

\[ R := m_1 - i - jn_2 \]
(Erreygers et al. 2018)
Advantages of imprecise Markov chains over their precise counterpart

- Partially specified $\pi_0$ and $Q/T$ are allowed
- Time homogeneity can be relaxed
- The Markov assumption can be relaxed
- Efficient computations remain possible
- State space explosion can be dealt with
All of this sounds too good to be true! What have you been hiding?
Can we learn these from data?

Initial distribution \( \pi_0 \) +

\( [X_0 \rightarrow X_t] \)

\( [0 \rightarrow t] \)

IDM (Walley 1996)
(Quaeghebeur 2009)
(Krak et al. 2018)

\( \text{transition rate matrix } Q \) or
\( \text{transition matrix } T \)
What if the states can’t be observed directly?

Can we still learn these? 

not yet…
How about more complicated inferences? (Troffaes et al. 2015) (Lopatatzidis 2017) in some cases...
Can we do infinite state spaces?

only in theory...
References (1)


References (2)


D. Hartfiel: On the solutions to $x'(t) = a(t)x(t)$ over all $a(t)$, where $p \leq a(t) \leq q$. Journal of Mathematical Analysis and Applications, 108:230–240, 1985.


References (3)


This work was partially supported by H2020-MSCA-ITN-2016 UTOPIAE, grant agreement 722734.

http://twitter.com/utopiae_network  http://utopiae.eu