Imprecise Markov chains
From basic theory to applications I
prof. Gert de Cooman
Imprecise discrete-time Markov chains
Precise probability models
Mass functions and expectations

Assume we are uncertain about:

- the value or a variable $X$
- in a set of possible values $\mathcal{X}$.

This is usually modelled by a probability mass function $p$ on $\mathcal{X}$:

$$p(x) \geq 0 \text{ and } \sum_{x \in \mathcal{X}} p(x) = 1;$$

With $p$ we can associate an expectation operator $E_p$:

$$E_p(f) := \sum_{x \in \mathcal{X}} p(x)f(x) \text{ where } f : \mathcal{X} \to \mathbb{R}.$$

If $A \subseteq \mathcal{X}$ is an event, then its probability is given by

$$P_p(A) = \sum_{x \in A} p(x) = E_p(I_A).$$
Consider the simplex $\Sigma_{\mathcal{X}}$ of all mass functions on $\mathcal{X}$:

$$\Sigma_{\mathcal{X}} := \left\{ p \in \mathbb{R}_{+}^{\mathcal{X}} : \sum_{x \in \mathcal{X}} p(x) = 1 \right\}.$$
Geometrical interpretation of expectation

Assessments lead to constraints

Specifying an expectation $E(f)$ for a map $f : \mathcal{X} \to \mathbb{R}$

\[
\sum_{x \in \mathcal{X}} p(x)f(x) = E(f)
\]

imposes a linear constraint on the possible values for $p$ in $\Sigma_\mathcal{X}$.

It corresponds to intersecting the simplex $\Sigma_\mathcal{X}$ with a hyperplane, whose direction depends on $f$:

\[
E(I_{\{b\}} - I_{\{c\}}) = 0
\]

\[
E(I_{\{a\}}) = \frac{1}{2}
\]
Imprecise probability models
Linear inequality constraints

More flexible assessments
Impose linear inequality constraints on $p$ in $\Sigma \mathcal{X}$:

$$E(f) \leq \sum_{x \in \mathcal{X}} p(x)f(x) \quad \text{or} \quad \sum_{x \in \mathcal{X}} p(x)f(x) \leq \bar{E}(f).$$

Corresponds to intersecting $\Sigma \mathcal{X}$ with affine semi-spaces:
More flexible assessments

Impose linear inequality constraints on $p$ in $\Sigma \mathcal{X}$:

$$E(f) \leq \sum_{x \in \mathcal{X}} p(x)f(x) \quad \text{or} \quad \sum_{x \in \mathcal{X}} p(x)f(x) \leq \bar{E}(f).$$

Corresponds to intersecting $\Sigma \mathcal{X}$ with affine semi-spaces:
Linear inequality constraints

More flexible assessments
Impose linear inequality constraints on \( p \) in \( \Sigma \mathcal{X} \):

\[
E(f) \leq \sum_{x \in \mathcal{X}} p(x)f(x) \quad \text{or} \quad \sum_{x \in \mathcal{X}} p(x)f(x) \leq \bar{E}(f).
\]

Corresponds to intersecting \( \Sigma \mathcal{X} \) with affine semi-spaces:
Linear inequality constraints

More flexible assessments
Impose linear inequality constraints on $p$ in $\Sigma \mathcal{X}$:

$$
\mathbb{E}(f) \leq \sum_{x \in \mathcal{X}} p(x)f(x) \quad \text{or} \quad \sum_{x \in \mathcal{X}} p(x)f(x) \leq \mathbb{E}(f).
$$

Corresponds to intersecting $\Sigma \mathcal{X}$ with affine semi-spaces:
Linear inequality constraints

More flexible assessments

Impose linear inequality constraints on \( p \) in \( \Sigma \mathcal{X} \):

\[
E(f) \leq \sum_{x \in \mathcal{X}} p(x)f(x) \quad \text{or} \quad \sum_{x \in \mathcal{X}} p(x)f(x) \leq \bar{E}(f).
\]

Corresponds to intersecting \( \Sigma \mathcal{X} \) with affine semi-spaces:
Credal sets

Any such number of assessments leads to a credal set $\mathcal{M}$.

Definition
A credal set $\mathcal{M}$ is a convex closed subset of $\Sigma X$. 
Lower and upper expectations

Equivalent model
Consider the set $\mathcal{L}(\mathcal{X}) = \mathbb{R}^\mathcal{X}$ of all real-valued maps on $\mathcal{X}$. We define two real functionals on $\mathcal{L}(\mathcal{X})$: for all $f : \mathcal{X} \to \mathbb{R}$

\[
\begin{align*}
\underline{E}_\mathcal{M}(f) &= \min \{ E_p(f) : p \in \mathcal{M} \} \quad \text{lower expectation} \\
\overline{E}_\mathcal{M}(f) &= \max \{ E_p(f) : p \in \mathcal{M} \} \quad \text{upper expectation}.
\end{align*}
\]

Observe that

\[
\overline{E}_\mathcal{M}(f) = -\underline{E}_\mathcal{M}(-f).
\]
Basic properties of upper expectations

Definition
We call a real functional \( \overline{E} \) on \( \mathcal{L}(\mathcal{X}) \) an upper expectation if it satisfies the following properties:
for all \( f \) and \( g \) in \( \mathcal{L}(\mathcal{X}) \) and all real \( \lambda \geq 0 \):

1. \( \overline{E}(f) \leq \max f \) [boundedness];
2. \( \overline{E}(f + g) \leq \overline{E}(f) + \overline{E}(g) \) [sub-additivity];
3. \( \overline{E}(\lambda f) = \lambda \overline{E}(f) \) [non-negative homogeneity].

Theorem
A real functional \( \overline{E} \) is an upper expectation if and only if it is the upper envelope of some credal set \( \mathcal{M} \).

Proof.
Use \( \mathcal{M} = \{ p \in \Sigma_{\mathcal{X}} : (\forall f \in \mathcal{L}(\mathcal{X}))(E_{p}(f) \leq \overline{E}(f)) \} \). \( \square \)
Discrete-time uncertain processes
Precise probability trees

We consider an uncertain process with variables $X_1, X_2, \ldots, X_n, \ldots$ assuming values in a finite set of states $\mathcal{X}$.

This leads to a standard event tree with nodes

$$s = (x_1, x_2, \ldots, x_n), \quad x_k \in \mathcal{X}, \quad n \geq 0.$$
Precise probability trees

The standard event tree becomes a probability tree by attaching to each node a local probability mass function $p_s$ on $X$ with associated expectation operator $E_s$. 
Precise probability trees

We consider an uncertain process with variables $X_1, X_2, \ldots, X_n, \ldots$ assuming values in a finite set of states $\mathcal{X}$.

This leads to a standard event tree with nodes

$$s = (x_1, x_2, \ldots, x_n), \quad x_k \in \mathcal{X}, \quad n \geq 0.$$ 

The standard event tree becomes a probability tree by attaching to each node $s$ a local probability mass function $p_s$ on $\mathcal{X}$ with associated expectation operator $E_s$. 
Calculating global expectations from local ones

Consider a function $g: \mathcal{X}^n \rightarrow \mathbb{R}$ of the first $n$ variables:

$$g = g(X_1, X_2, \ldots, X_n)$$

We want to calculate its expectation $E(g|s)$ in $s = (x_1, \ldots, x_k)$.

**Theorem (Law of Iterated Expectation)**

*Suppose we know $E(g|s, x)$ for all $x \in \mathcal{X}$, then we can calculate $E(g|s)$ by backwards recursion using the local model $p_s$:*

$$E(g|s) = E_s (E(g|s, \cdot)) = \sum_{x \in \mathcal{X}} p_s(x) E(g|s, x).$$

$$E(g|s) = p_s(a)E(g|s,a) + p_s(b)E(g|s,b)$$

---

$(s,a) \rightarrow E(g|s,a)$

$(s,b) \rightarrow E(g|s,b)$
Calculating global expectations from local ones

All expectations $E(g|x_1, \ldots, x_k)$ in the tree can be calculated from the local models as follows:

1. start in the final cut $\mathcal{X}^n$ and let:

   $$E(g|x_1, x_2, \ldots, x_n) = g(x_1, x_2, \ldots, x_n);$$

2. do backwards recursion using the Law of Iterated Expectation:

   $$E(g|x_1, \ldots, x_k) = E(x_1, \ldots, x_k)(E(g|x_1, \ldots, x_k, \cdot))$$

3. go on until you get to the root node $\square$, where:

   $$E(g|\square) = E(g).$$
Christiaan Huygens (1656–1657)
Sets of mass functions

Major restrictive assumption
Until now, we have assumed that we have sufficient information in order to specify, in each node \( s \), a probability mass function \( p_s \) on the set \( \mathcal{X} \) of possible values for the next state.

More general uncertainty models
We consider credal sets as more general uncertainty models: closed convex subsets of \( \Sigma_\mathcal{X} \).
Imprecise probability trees
Definition and interpretation

Definition
An imprecise probability tree is a probability tree where in each node $s$ the local uncertainty model is an imprecise probability model $\mathcal{M}_s$, or equivalently, its associated upper expectation $\overline{E}_s$:

$$\overline{E}_s(f) = \max \{E_p(f) : p \in \mathcal{M}_s\} \text{ for all real maps } f \text{ on } \mathcal{X}. $$
Definition and interpretation

An imprecise probability tree can be seen as an infinity of compatible precise probability trees: choose in each node a probability mass function \( p_s \) from the set \( M_s \).
Definition and interpretation

Definition
An imprecise probability tree is a probability tree where in each node $s$ the local uncertainty model is an imprecise probability model $\mathcal{M}_s$, or equivalently, its associated upper expectation $\bar{E}_s$:

$$\bar{E}_s(f) = \max \{E_p(f) : p \in \mathcal{M}_s\} \text{ for all real maps } f \text{ on } \mathcal{X}.$$

An imprecise probability tree can be seen as an infinity of compatible precise probability trees: choose in each node $s$ a probability mass function $p_s$ from the set $\mathcal{M}_s$. 
An imprecise probability tree can be seen as an infinity of compatible precise probability trees: choose in each node a probability mass function $p_s$ from the set $\mathcal{M}_s$. 

### Definition and interpretation
Associated lower and upper expectations

For each real map $g = g(X_1, \ldots, X_n)$, each node $s = (x_1, \ldots, x_k)$, and each such compatible precise probability tree, we can calculate the expectation

$$E(g|x_1, \ldots, x_k)$$

using the backwards recursion method described before.

By varying over each compatible probability tree, we get a closed real interval:

$$[E(g|x_1, \ldots, x_k), \bar{E}(g|x_1, \ldots, x_k)]$$

We want a better, more efficient method to calculate these lower and upper expectations $\underline{E}(g|x_1, \ldots x_k)$ and $\bar{E}(g|x_1, \ldots, x_k)$. 
The Law of Iterated Expectation

Theorem (Law of Iterated Expectation)

Suppose we know $E(g|s,x)$ for all $x \in \mathcal{X}$, then we can calculate $E(g|s)$ by backwards recursion using the local model $E_s$:

$$E(g|s) = E_s(E(g|s,\cdot)) = \max_{p_s \in \mathcal{M}_s} \sum_{x \in \mathcal{X}} p_s(x) E(g|s,x).$$

The complexity of calculating the $E(g|s)$, as a function of $n$, is therefore essentially the same as in the precise case!
Imprecise Markov chains
Precise Markov chains: definition

Definition
The uncertain process is a stationary precise Markov chain when all $\mathcal{M}_s$ are singletons (precise), and

1. $\mathcal{M}_\square = \{m_1\}$,
2. the Markov Condition is satisfied:

$$\mathcal{M}(x_1,...,x_n) = \{q(\cdot|x_n)\}.$$
Precise Markov chains: definition

For each \( x \in X \), the transition mass function \( q(\cdot|x) \) corresponds to an expectation operator:

\[
E(f|x) = \sum_{z \in X} q(z|x) f(z).
\]
Precise Markov chains: definition

Definition
The uncertain process is a stationary precise Markov chain when all $M_s$ are singletons (precise), and

1. $M_{\square} = \{m_1\}$,
2. the Markov Condition is satisfied:

$$M(x_1, \ldots, x_n) = \{q(\cdot|x_n)\}.$$

For each $x \in X$, the transition mass function $q(\cdot|x)$ corresponds to an expectation operator:

$$E(f|x) = \sum_{z \in X} q(z|x) f(z).$$
Precise Markov chains: transition operators

Definition
Consider the linear transformation $T$ of $\mathcal{L}(\mathcal{X})$, called transition operator:

$$T: \mathcal{L}(\mathcal{X}) \to \mathcal{L}(\mathcal{X}): f \mapsto Tf$$

where $Tf$ is the real map given by, for any $x \in \mathcal{X}$:

$$Tf(x) := E(f|x) = \sum_{z \in \mathcal{X}} q(z|x)f(z)$$

$T$ is the dual of the linear transformation with Markov matrix $M$, with elements $M_{xy} := q(y|x)$. 
Precise Markov chains: transition operators

Definition
Consider the linear transformation $T$ of $\mathcal{L}(\mathcal{X})$, called transition operator:

$$T: \mathcal{L}(\mathcal{X}) \rightarrow \mathcal{L}(\mathcal{X}): f \mapsto Tf$$

where $Tf$ is the real map given by, for any $x \in \mathcal{X}$:

$$Tf(x) := E(f|x) = \sum_{z \in \mathcal{X}} q(z|x)f(z)$$

$T$ is the dual of the linear transformation with Markov matrix $M$, with elements $M_{xy} := q(y|x)$.

Then the Law of Iterated Expectation yields:

$$E_n(f) = E_1(T^{n-1}f), \text{ and dually, } m_n = M^{n-1}m_1.$$ 

Complexity is linear in the number of time steps $n$. 
Imprecise Markov chains: definition

Definition
The uncertain process is a stationary imprecise Markov chain when the Markov Condition is satisfied:

\[ M(x_1, \ldots, x_n) = \mathcal{D}(\cdot | x_n). \]
Imprecise Markov chains: definition

An imprecise Markov chain can be seen as an infinity of probability trees. For each $x \in X$, the local transition model $Q(\cdot|x)$ corresponds to lower and upper expectation operators:

$$E(f|x) = \min\{E_p(f) : p \in Q(\cdot|x)\}$$

$$E(f|x) = \max\{E_p(f) : p \in Q(\cdot|x)\}.$$
Imprecise Markov chains: definition

Definition

The uncertain process is a stationary imprecise Markov chain when the Markov Condition is satisfied:

$$M(x_1,...,x_n) = \mathcal{D}(\cdot|x_n).$$

An imprecise Markov chain can be seen as an infinity of probability trees.

For each $$x \in \mathcal{X}$$, the local transition model $$\mathcal{D}(\cdot|x)$$ corresponds to lower and upper expectation operators:

$$E(f|x) = \min \{ E_p(f) : p \in \mathcal{D}(\cdot|x) \}$$

$$\bar{E}(f|x) = \max \{ E_p(f) : p \in \mathcal{D}(\cdot|x) \}.$$
Imprecise Markov chains: transition operators

Definition
Consider the non-linear transformations $T$ and $\bar{T}$ of $\mathcal{L}(\mathcal{X})$, called lower and upper transition operators:

$$T: \mathcal{L}(\mathcal{X}) \rightarrow \mathcal{L}(\mathcal{X}): f \mapsto Tf$$
$$\bar{T}: \mathcal{L}(\mathcal{X}) \rightarrow \mathcal{L}(\mathcal{X}): f \mapsto \bar{T}f$$

where the real maps $Tf$ and $\bar{T}f$ are given by:

$$Tf(x) := \mathbb{E}(f|x) = \min \{ E_p(f) : p \in \mathcal{Q}(\cdot|x) \}$$
$$\bar{T}f(x) := \overline{\mathbb{E}}(f|x) = \max \{ E_p(f) : p \in \mathcal{Q}(\cdot|x) \}$$
Imprecise Markov chains: transition operators

Definition
Consider the non-linear transformations $\mathbf{T}$ and $\overline{T}$ of $\mathcal{L}(\mathcal{X})$, called lower and upper transition operators:

$\mathbf{T}: \mathcal{L}(\mathcal{X}) \rightarrow \mathcal{L}(\mathcal{X}): f \mapsto T f$

$\overline{T}: \mathcal{L}(\mathcal{X}) \rightarrow \mathcal{L}(\mathcal{X}): f \mapsto \overline{T} f$

where the real maps $T f$ and $\overline{T} f$ are given by:

$T f(x) := E(f|x) = \min \{ E_p(f) : p \in \mathcal{Q}(\cdot|x) \}$

$\overline{T} f(x) := \overline{E}(f|x) = \max \{ E_p(f) : p \in \mathcal{Q}(\cdot|x) \}$

Then the Law of Iterated Expectation yields:

$E_n(f) = E_1(T^{n-1}f)$ and $\overline{E}_n(f) = \overline{E}_1(\overline{T}^{n-1}f)$.

Complexity is still linear in the number of time steps $n$. 
Lower and upper mass functions
Another example with $\mathcal{X} = \{a, b, c\}$

$$\begin{bmatrix} \mathcal{T}I \{a\} & \mathcal{T}I \{b\} & \mathcal{T}I \{c\} \end{bmatrix} = \begin{bmatrix} q(a | a) & q(b | a) & q(c | a) \\ q(a | b) & q(b | b) & q(c | b) \\ q(a | c) & q(b | c) & q(c | c) \end{bmatrix} = \frac{1}{200} \begin{bmatrix} 9 & 9 & 162 \\ 144 & 18 & 18 \\ 9 & 162 & 9 \end{bmatrix}$$

$$\begin{bmatrix} \overline{\mathcal{T}}I \{a\} & \overline{\mathcal{T}}I \{b\} & \overline{\mathcal{T}}I \{c\} \end{bmatrix} = \begin{bmatrix} \overline{q}(a | a) & \overline{q}(b | a) & \overline{q}(c | a) \\ \overline{q}(a | b) & \overline{q}(b | b) & \overline{q}(c | b) \\ \overline{q}(a | c) & \overline{q}(b | c) & \overline{q}(c | c) \end{bmatrix} = \frac{1}{200} \begin{bmatrix} 19 & 19 & 172 \\ 154 & 28 & 28 \\ 19 & 172 & 19 \end{bmatrix}$$
Another example with $\mathcal{X} = \{a, b, c\}$
A non-linear Perron–Frobenius Theorem
Generalising the linear case

Theorem (De Cooman, Hermans and Quaeghebeur, 2008)
Consider a stationary imprecise Markov chain with finite state set $\mathcal{X}$ and an upper transition operator $\overline{T}$. Suppose that $\overline{T}$ is regular, meaning that there is some $n > 0$ such that $\min \overline{T}^n I_{\{x\}} > 0$ for all $x \in \mathcal{X}$. Then for every initial upper expectation $\overline{E}_1$, the upper expectation $\overline{E}_n = \overline{E}_1 \circ \overline{T}^{n-1}$ for the state at time $n$ converges point-wise to the same upper expectation $\overline{E}_\infty$:

$$
\lim_{n \to \infty} \overline{E}_n(h) = \lim_{n \to \infty} \overline{E}_1(\overline{T}^{n-1}h) := \overline{E}_\infty(h)
$$

for all $h$ in $\mathcal{L}(\mathcal{X})$. Moreover, the corresponding limit upper expectation $\overline{E}_\infty$ is the only $\overline{T}$-invariant upper expectation on $\mathcal{L}(\mathcal{X})$, meaning that $\overline{E}_\infty = \overline{E}_\infty \circ \overline{T}$. 
Ergodicity

And in that case also

**Theorem (De Cooman, De Bock and Lopatatzidis, 2016)**

Consider a stationary imprecise Markov chain with finite state set $\mathcal{X}$ and an upper transition operator $\overline{T}$. Suppose that $\overline{T}$ is Perron–Frobenius-like with stationary upper expectation $\overline{E}_\infty$. Then

$$\overline{E}_\infty(h) \leq \liminf_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} h(X_k) \leq \limsup_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} h(X_k) \leq \overline{E}_\infty(h)$$

almost surely for all $h$ in $\mathcal{L}(\mathcal{X})$. 
Literature
Literature


