Imprecise stochastic processes

Alexander Erreygers
11th November 2019

Ghent University, ELIS, Foundations Lab for imprecise probabilities
FLip // Foundations Lab for imprecise probabilities
stochastic processes
discrete-time

stochastic processes
We consider an infinite sequence

\[ X_1, X_2, X_3, \ldots, X_n, \ldots \]

of uncertain variables that take values in the finite state space \( \mathcal{X} \).
We consider an infinite sequence

\[ X_1, X_2, X_3, \ldots, X_n, \ldots \]

of uncertain variables that take values in the finite state space \( \mathcal{X} \).

**Example**

\( X_n \) is the weather in Oviedo \( n \) days from now, and

\[ \mathcal{X} = \{\text{☀️, ☃️, ☂️} \}. \]

We want to make inferences, for example answer the following questions:

- What is the probability of ☃️ in 4 days?
- What is the expected number of days until the next ☀️ day?
- Should I bring an ☂️ tomorrow?
Modelling our uncertainty

First, we construct a tree with nodes (or situations)

\[ s = (x_1, \ldots, x_n), \quad x_i \in \mathcal{X}. \]

For example,

\[ (x_1, x_2, x_3) = (\text{clouds}, \text{rain}, \text{thunder}). \]
\[ X_1 \quad X_1, \ X_2 \]
Modelling our uncertainty

First, we construct a tree with nodes (or situations)

\[ s = (x_1, \ldots, x_n), \quad x_i \in \mathcal{X}. \]

Second, we turn this into a probability tree by specifying a local probability mass function \( p_s: \mathcal{X} \rightarrow [0, 1] \) for every situation \( s = (x_1, \ldots, x_n) \):

\[
P(X_{n+1} = x_{n+1} \mid X_1 = x_1, \ldots, X_n = x_n) = p_s(x_{n+1}), \quad x_{n+1} \in \mathcal{X}.
\]
This way, we construct a probability measure $P$ and

- we can make inferences
  - that is, compute $E_P(f \mid s)$ for sufficiently nice functions $f$ on $\Omega$—by using backwards recursion due to the law of total probability (aka the law of iterated expectation);

- we need to specify a countable number of local probability mass functions: one $p_s$ for every situation $s$. 
To make this tractable, one may assume that the local models

\[ p(x_1, \ldots, x_n) = p_{n,x_n} \]

1. only depend on the present, \([\text{Markovianity}]\)
To make this tractable, one may assume that the local models

\[ p(x_1, \ldots, x_n) = p_{n,x_n} = p_{x_n} \]

1. only depend on the present, \([\text{Markovianity}]\)
2. do not change over time. \([\text{time homogeneity}]\)
To make this tractable, one may assume that the local models

\[ p(x_1, \ldots, x_n) = p_{n,x_n} = p_{x_n} \]

1. only depend on the present, \([\text{Markovianity}]\)
2. do not change over time. \([\text{time homogeneity}]\)

This way, we end up with a homogeneous Markov chain and

we only need to specify \(|X^*| + 1\) local probability mass functions:
the initial one \(p_{\square}\) and one \(p_x\) for every state \(x\).
What if we cannot specify the local models $p_s$ precisely?
What if we cannot specify the local models $p_s$ precisely?

¡imprecise probabilities!
Imprecise probabilities is a collection of theories that aim to generalise classical probability theory to allow for partial specification.
Let $\mathcal{L}$ denote the real vector space of all real-valued functions on $\mathcal{X}$, and let $\Sigma_{\mathcal{X}}$ denote the subset of all probability mass functions on $\mathcal{X}$:

$$\Sigma_{\mathcal{X}} = \left\{ p \in \mathcal{L} : p \geq 0, \sum_{x \in \mathcal{X}} p(x) = 1 \right\}.$$
Let $\mathcal{L}$ denote the real vector space of all real-valued functions on $\mathcal{X}$, and let $\Sigma_{\mathcal{X}}$ denote the subset of all probability mass functions on $\mathcal{X}$:

$$\Sigma_{\mathcal{X}} = \left\{ p \in \mathcal{L} : p \geq 0, \sum_{x \in \mathcal{X}} p(x) = 1 \right\}.$$

A probability mass function $p$ induces an expectation operator $E_p : \mathcal{L} \to \mathbb{R}$, defined by

$$E_p(f) = \sum_{x \in \mathcal{X}} p(x)f(x) \quad \text{for all } f \in \mathcal{L}.$$
Let $\mathcal{L}$ denote the real vector space of all real-valued functions on $\mathcal{X}$, and let $\Sigma_\mathcal{X}$ denote the subset of all probability mass functions on $\mathcal{X}$:

$$\Sigma_\mathcal{X} = \left\{ p \in \mathcal{L} : p \geq 0, \sum_{x \in \mathcal{X}} p(x) = 1 \right\}.$$ 

A probability mass function $p$ induces an expectation operator $E_p : \mathcal{L} \to \mathbb{R}$, defined by

$$E_p(f) = \sum_{x \in \mathcal{X}} p(x)f(x) \quad \text{for all } f \in \mathcal{L}.\]

Recall that

E1. $E_p(f) \geq \min f$ for all $f \in \mathcal{L}$; \hspace{1cm} [boundedness]

E2. $E_p(f + g) = E_p(f) + E_p(g)$ for all $f, g \in \mathcal{L}$; \hspace{1cm} [additivity]

E3. $E_p(\lambda f) = \lambda E_p(f)$ for all $f \in \mathcal{L}$ and $\lambda \in \mathbb{R}$; \hspace{1cm} [homogeneity]
Instead of a single probability mass function $p$, we now consider a credal set $\mathcal{M} \subseteq \Sigma_X$, a non-empty, closed and convex set of probability mass functions.
Instead of a single probability mass function $p$, we now consider a credal set $\mathcal{M} \subseteq \Sigma_X$, a non-empty, closed and convex set of probability mass functions.

A credal set is defined by constraints of the form

$$c_f \leq \sum_{x \in X} p(x)f(x) = E_p(f).$$
Credal set

Instead of a single probability mass function $p$, we now consider a credal set $\mathcal{M} \subseteq \Sigma_X$, a non-empty, closed and convex set of probability mass functions.

The credal set $\mathcal{M}$ induces a set of expectations:

$$\{E_p(f) : p \in \mathcal{M}\}.$$  

Specifically of interest are the bounds

$$\underline{E}_\mathcal{M}(f) := \min\{E_p(f) : p \in \mathcal{M}\} \quad \text{and} \quad \overline{E}_\mathcal{M}(f) := \max\{E_p(f) : p \in \mathcal{M}\}.$$  

Note that these are conjugate: $\overline{E}_\mathcal{M}(f) = -\underline{E}_\mathcal{M}(-f)$.  

Lower expectation

An operator $E : \mathcal{L} \to \mathbb{R}$ is called a lower expectation (coherent lower prevision) if

LE1. $E(f) \geq \min f$ for all $f \in \mathcal{L}$; \hfill \text{[boundedness]}

LE2. $E(f + g) \geq E(f) + E(g)$ for all $f, g \in \mathcal{L}$; \hfill \text{[super-additivity]}

LE3. $E(\lambda f) = \lambda E(f)$ for all $f \in \mathcal{L}$ and $\lambda \in \mathbb{R}_{>0}$. \hfill \text{[positive homogeneity]}

Theorem

An operator $E : \mathcal{L} \to \mathbb{R}$ is a lower expectation if and only if it is the lower envelope of some credal set $\mathcal{M} \subseteq \Sigma_X$, meaning that $E(f) = E_M(f) = \min \{E_p(f) : p \in \mathcal{M}\}$ for all $f \in \mathcal{L}$. 


An operator $\mathbb{E}: \mathcal{L} \to \mathbb{R}$ is called a lower expectation (coherent lower prevision) if

1. $\mathbb{E}(f) \geq \min f$ for all $f \in \mathcal{L}$; \hspace{3cm} \text{[boundedness]}
2. $\mathbb{E}(f + g) \geq \mathbb{E}(f) + \mathbb{E}(g)$ for all $f, g \in \mathcal{L}$; \hspace{3cm} \text{[super-additivity]}
3. $\mathbb{E}(\lambda f) = \lambda \mathbb{E}(f)$ for all $f \in \mathcal{L}$ and $\lambda \in \mathbb{R}_{>0}$. \hspace{3cm} \text{[positive homogeneity]}

**Theorem**

An operator $\mathbb{E}: \mathcal{L} \to \mathbb{R}$ is a lower expectation if and only if it is the lower envelope of some credal set $\mathcal{M} \subseteq \Sigma_{\mathcal{X}}$, meaning that

$$\mathbb{E}(f) = \mathbb{E}_{\mathcal{M}}(f) = \min\{E_p(f) : p \in \mathcal{M}\} \quad \text{for all } f \in \mathcal{L}.$$
An operator $E: \mathcal{L} \to \mathbb{R}$ is called a lower expectation (coherent lower prevision) if

**LE1.** $E(f) \geq \min f$ for all $f \in \mathcal{L}$;  
**LE2.** $E(f + g) \geq E(f) + E(g)$ for all $f, g \in \mathcal{L}$;  
**LE3.** $E(\lambda f) = \lambda E(f)$ for all $f \in \mathcal{L}$ and $\lambda \in \mathbb{R} > 0$. 

Theorem

An operator $E: \mathcal{L} \to \mathbb{R}$ is a lower expectation if and only if it is the lower envelope of some credal set $\mathcal{M} \subseteq \Sigma_X$, meaning that 

$$E(f) = E_M(f) = \min \{E_p(f): p \in \mathcal{M}\}$$

for all $f \in \mathcal{L}$.

What does this have to do with stochastic processes?
Assume that we can only assess that
\[ p \in \mathcal{M} \quad \text{and} \quad p(x_1, \ldots, x_n) \in \mathcal{M}_{x_1, \ldots, x_n}, \quad (1) \]
where \( \mathcal{M} \) and \( \mathcal{M}_x \) for all \( x \in \mathcal{X} \) are credal sets.
Assume that we can only assess that
\[ p \Box \in \mathcal{M} \Box \quad \text{and} \quad p(x_1, \ldots, x_n) \in \mathcal{M}_{x_n}. \] (1)

We consider three nested sets of probability trees that satisfy (1):

- \( \mathcal{P}^{\text{CHM}} \): all compatible homogeneous Markov chains;
- \( \mathcal{P}^{\text{CM}} \): all compatible Markov chains;
- \( \mathcal{P}^{\text{C}} \): all compatible probability trees.

Can we compute
\[ E_{\mathcal{P}}(f \mid s) := \inf \{ E_P(f \mid s) : P \in \mathcal{P} \} \]
and
\[ \bar{E}_{\mathcal{P}}(f \mid s) := \sup \{ E_P(f \mid s) : P \in \mathcal{P} \}? \]
Computing these tight lower and upper bounds turns out to be:

- Intractable for $\mathcal{P}^{\text{CHM}}$,
- Intractable for $\mathcal{P}^{\text{CM}}$—at least in general,
- Tractable for $\mathcal{P}^{\text{C}}$, because we can use 	extit{backwards recursion} due to the 	extit{imprecise law of iterated expectation}. 


Damjan Škulj. “Discrete time Markov chains with interval probabilities”. In: IJAR 50.8 (2009), pp. 1314–1329


continuous-time
stochastic processes
We consider the collection

\[ \{ X_\tau : \tau \in \mathbb{R}_{\geq 0} \} \]

of uncertain variables that take values in the finite state space \( \mathcal{X} \).
We consider the collection

\[ \{ X_\tau : \tau \in \mathbb{R}_{\geq 0} \} \]

of uncertain variables that take values in the finite state space \( \mathcal{X} \).

**Example**

\( X_\tau \) is the weather in Oviedo \( \tau \) time units from now, and

\[ \mathcal{X} = \{ \text{\sunny}, \text{\cloudy}, \text{\rainy} \}. \]

We want to make inferences, for example answer questions like:

- What is the probability of \( \text{\cloudy} \) after 4 days?
- How long do I have to wait until it is \( \text{\sunny} \) again?
- Do I have to bring an \( \text{\umbrella} \) tomorrow?
Imprecise continuous-time Markov chains

DTMC
\[ p_0, \{ p_x \}_{x \in X} \]

imprecise DTMC
\[ M_0, \{ M_x \}_{x \in X} \]

CTMC
\[ p_0, \{ q_x \}_{x \in X} \]

imprecise CTMC
\[ M_0, \{ Q_x \}_{x \in X} \]
Imprecise continuous-time Markov chains

Similar results as for imprecise discrete-time Markov chains, but—for now—limited to inferences that depend on a finite number of time points.


A **counting process** is a model for a stream of events $X_\tau$: the number of events that have occurred up to time $\tau$, so $X = \mathbb{Z}_{\geq 0}$.

Example $X_\tau$: the number of lightning strikes that have hit the cathedral of Oviedo.

We want to answer questions like:

- What is the probability of at least one?
- What is the expected number in the following year?
- What is the expected time until the next?
A **counting process** is a model for a stream of events $X_\tau$: the number of events that have occurred up to time $\tau$, so $X = \mathbb{Z}_{\geq 0}$.

**Example**

$X_\tau$: the number of ⚡ lightning strikes that have hit the cathedral of Oviedo.

We want to answer questions like:

- What is the probability of at least one ⚡ in some time period?
- What is the expected number of ⚡ in the following year?
- What is the expected time until the next ⚡?
Counting processes in general

\[ X_{t_1} = x_1 \quad X_{t_n} = x_n \quad X_t = x \quad X_{t+\Delta} = y \]

We model our beliefs by means of the transition probabilities

\[ P(X_{t+\Delta} = y \mid X_t = x, X_{t_n} = x_n, \ldots, X_{t_1} = x_1, X_u = x_u). \]
For the Poisson process, we additionally assume that the transition probabilities

\[ P(X_{t+\Delta} = y \mid X_t = x, X_u = x_u) \]
The Poisson process in particular

\[
X_{t_1} = x_1 \quad \ldots \quad X_{t_n} = x_n \quad X_t = x \quad X_{t+\Delta} = y
\]

For the Poisson process, we additionally assume that the transition probabilities

\[
P(X_{t+\Delta} = y \mid X_t = x, X_u = x_u) = P(X_{t+\Delta} = y \mid X_t = x)
\]

1. only depend on the present, \[\text{[Markovianity]}\]
The Poisson process in particular

\[ X_{t_1} = x_1 \quad X_{t_n} = x_n \quad X_0 = x \quad X_\Delta = y \]

For the Poisson process, we additionally assume that the transition probabilities

\[ P(X_{t+\Delta} = y \mid X_t = x, X_u = x_u) = P(X_\Delta = y \mid X_0 = x) \]

1. only depend on the present, \[ \text{[Markovianity]} \]
2. only depend on the length of the time period, \[ \text{[time homogeneity]} \]
The Poisson process in particular

For the Poisson process, we additionally assume that the transition probabilities

\[ P(X_{t+\Delta} = y \mid X_t = x, X_u = x_u) = P(X_{\Delta} = y - x \mid X_0 = 0) \]

1. only depend on the present, \[ \text{[Markovianity]} \]
2. only depend on the length of the time period, \[ \text{[time homogeneity]} \]
3. only depend on the number of new events. \[ \text{[state homogeneity]} \]
The rate parameter

A Poisson process is uniquely characterised by a single parameter: the rate $\lambda$!
The rate parameter

A Poisson process is uniquely characterised by a single parameter: the rate $\lambda$.

It has multiple interpretations, for instance:

- the expected number of new events in any time period is proportional to $\lambda$:

  $$E_P(X_{t+\Delta} \mid X_t = x, X_u = x_u) = x + \lambda \Delta;$$
A Poisson process is uniquely characterised by a single parameter: the rate $\lambda$!

It has multiple interpretations, for instance:

- The expected number of new events in any time period is proportional to $\lambda$:
  \[ E_P(X_{t+\Delta} \mid X_t = x, X_u = x_u) = x + \lambda\Delta; \]

- $\lambda$ is the (initial) rate at which the probability of a single event increases:
  \[ P(X_{t+\Delta} = x + 1 \mid X_t = x, X_u = x_u) = \lambda\Delta + o(\Delta). \]
What if we do not know the rate $\lambda$ precisely, but only know that it belongs to the rate interval $[\lambda, \bar{\lambda}]$?
The general approach

Let $\mathcal{P}$ be a set of counting processes characterised by the rate interval $[\lambda, \bar{\lambda}]$, and define the lower expectation

$$E_{\mathcal{P}}(f \mid X_t = x, X_u = x_u) := \inf \{ E_P(f \mid X_t = x, X_u = x_u) : P \in \mathcal{P} \}.$$ 

Choose $\mathcal{P}$ such that

(i) computing $E_{\mathcal{P}}(f \mid X_t = x, X_u = x_u)$ is tractable, 

(ii) $E_{\mathcal{P}}(\cdot \mid \cdot)$ is Poisson-like, in the sense that

(a) $E_{\mathcal{P}}(g(X_{t+\Delta}) \mid X_t = x, X_u = x_u)$ is Markov and homogeneous, 

(b) $E_{\mathcal{P}}(X_{t+\Delta} \mid X_t = x, X_u = x_u) = x + \lambda \Delta$. 

26
A naive imprecise Poisson process

If $\mathcal{P}$ is the set of all Poisson processes with rate $\lambda$ in the rate interval $[\lambda, \bar{\lambda}]$, then

- computing $E_{\mathcal{P}}(f \mid X_t = x, X_u = x_u)$ is a one-parameter optimisation problem;
- $E_{\mathcal{P}}(\cdot \mid \cdot)$ is Poisson-like;
- every $P$ in $\mathcal{P}$ is Markov and homogeneous.
An alternative condition

\[(\forall P \in \mathcal{P})(\exists \lambda \in [\underline{\lambda}, \overline{\lambda}]) (\forall t, \Delta, x, x_u \ldots)\]

\[P(X_{t+\Delta} = x + 1 \mid X_t = x, X_u = x_u) = \lambda \Delta + o(\Delta)\]
An alternative condition

\((\forall P \in \mathcal{P})(\exists \lambda \in [\lambda, \bar{\lambda}])(\forall t, \Delta, x, x_u \ldots)\)

\[ P(X_{t+\Delta} = x + 1 \mid X_t = x, X_u = x_u) = \lambda \Delta + o(\Delta) \]

\[ \Downarrow \]

\((\forall P \in \mathcal{P})(\exists \lambda \in [\lambda, \bar{\lambda}])(\forall t, \Delta, x, x_u \ldots)\)

\[ \lambda \Delta + o(\Delta) \leq P(X_{t+\Delta} = x + 1 \mid X_t = x, X_u = x_u) \leq \bar{\lambda} \Delta + o(\Delta) \]
A more involved imprecise Poisson process

If $\mathcal{P}$ is the set of processes that are consistent with the rate interval $[\lambda, \bar{\lambda}]$, in the sense that

$$\lambda \Delta + o(\Delta) \leq P(X_{t+\Delta} = x + 1 \mid X_t = x, X_u = x_u) \leq \bar{\lambda} \Delta + o(\Delta),$$

then

😊 a $P$ in $\mathcal{P}$ is not necessarily Markov nor homogeneous;
😊 computing $E_{\mathcal{P}}(f \mid X_t = x, X_u = x_u)$ is non-trivial (if not infeasible).
A more involved imprecise Poisson process

If $\mathcal{P}$ is the set of processes that are consistent with the rate interval $[\underline{\lambda}, \overline{\lambda}]$, in the sense that

$$\underline{\lambda}\Delta + o(\Delta) \leq P(X_{t+\Delta} = x + 1 \mid X_t = x, X_u = x_u) \leq \overline{\lambda}\Delta + o(\Delta),$$

then

😊 a $P$ in $\mathcal{P}$ is not necessarily Markov nor homogeneous;
🎉 computing $\mathbb{E}_\mathcal{P}(f \mid X_t = x, X_u = x_u)$ is non-trivial (if not infeasible).

However, it turns out that

Laughing face with sunglasses 🙈 computing $\mathbb{E}_\mathcal{P}(g(X_{t+\Delta}) \mid X_t = x, X_u = x_u)$ is tractable;
Smiling face with sunglasses 😎 $\mathbb{E}_\mathcal{P}(\cdot \mid \cdot)$ is Poisson-like.

As a consequence of our assumptions on $F$, we show that this requires the solution of a one-parameter transformation explained by Krak et al. [8, Section 7.3], the two results above allow us to define the corresponding lower transition without any difficulty. Moreover, the third property follows immediately from Theorem 15, Equation (20) and Lemma 53.

Finally, we verify that the unique extension $\tilde{\sigma}(u) = \Lambda(u)$ for all $u \in X$.

We need to show that $\tilde{\sigma}(u)$ is the unique solution of the equation $\tilde{\sigma}(u) = \Lambda(u)$ for all $u \in X$. This follows immediately from Proposition 98.

Finally, we consider the case $\tilde{\sigma}(u) = \Lambda(u)$. Then for any $f$ in $L(X)$, we have $\tilde{\sigma}(u) = \Lambda(u)$ for all $u \in X$.

2. Fix any $\tilde{\sigma}(u) = \Lambda(u)$. Then for any $f$ in $L(X)$, we have $\tilde{\sigma}(u) = \Lambda(u)$ for all $u \in X$.

We end this section with our strongest argument for using the Poisson process. This follows immediately from Proposition 98.
I have not mentioned that

- the parameters of imprecise Markov chains can be learned;
- hidden imprecise Markov chains have been studied as well;
- if state space explosion occurs in a precise Markov chain, we can use a coarser imprecise Markov chain to tractably bound inferences.

I should also mention that

- more work is needed to allow for a larger class of inferences;
- infinite state spaces are largely unexplored.