Imprecise stochastic processes

Alexander Erreygers

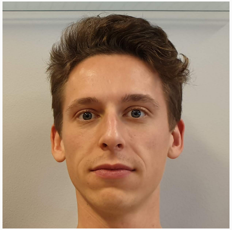
11th November 2019

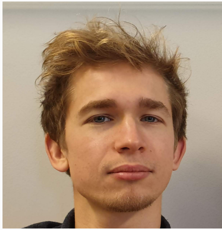
Ghent University, ELIS, Foundations Lab for imprecise probabilities

FLip // Foundations Lab for imprecise probabilities

GHENT UNIVERSITY

FACULTY OF ENGINEERING AND ARCHITECTURE





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 $\mathscr{X}$.

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 $\mathscr{X}$.

Example
X_n is the weather in Oviedo n days from now, and

$$\mathscr{X} = \{☀️, ☁️, ⛈️\}.$$

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?

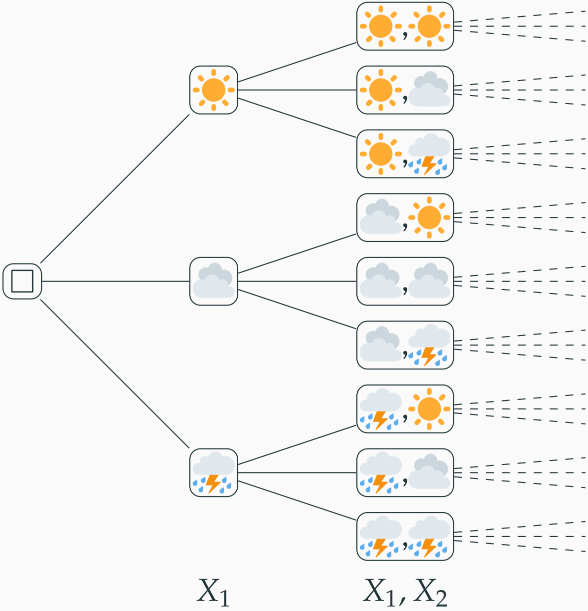
Modelling our uncertainty

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

$$s = (x_1, \ldots, x_n), \quad x_i \in \mathscr{X}.$$

For example,

$$(x_1, x_2, x_3) = (\text{⛈}, \text{⛈}, \text{⛈}).$$


X_1
X_1, X_2

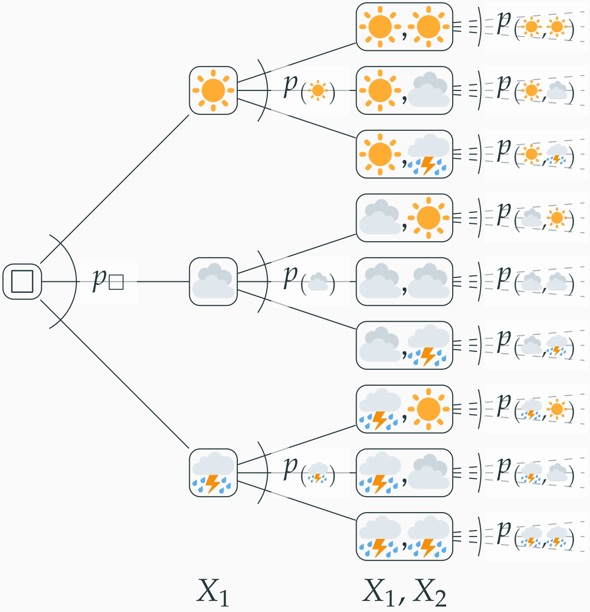
Modelling our uncertainty

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

$$s = (x_1, \ldots, x_n), \quad x_i \in \mathscr{X}.$$

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

$$P(X_{n+1} = x_{n+1} \mid X_1 = x_1, \ldots, X_n = x_n) = p_s(x_{n+1}), \quad x_{n+1} \in \mathscr{X}.$$



X_1
X_1, X_2

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 Ω—
  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, [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, [Markovianity]
2. do not change over time. [time homogeneity]

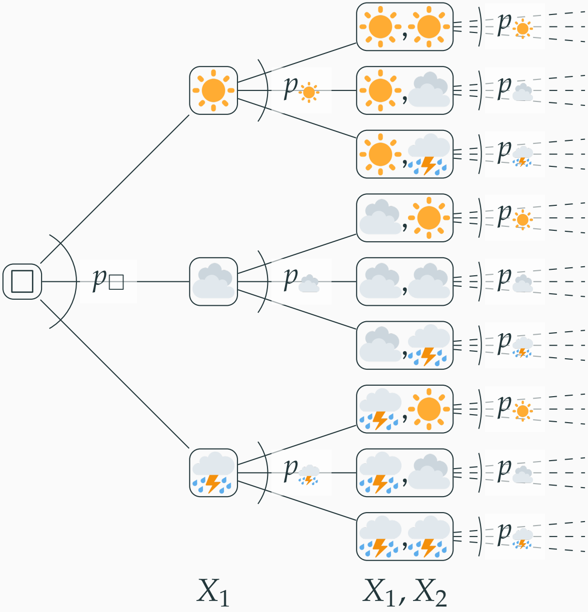
To make this tractable, one may assume that the local models

$$p_{(x_1,\ldots,x_n)} = p_{n,x_n} = p_{x_n}$$

1. only depend on the present, [Markovianity]
2. do not change over time. [time homogeneity]

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

😃 we only need to specify $|\mathscr{X}| + 1$ local probability mass functions: the initial one $p_\Box$ and one p_x for every state x.

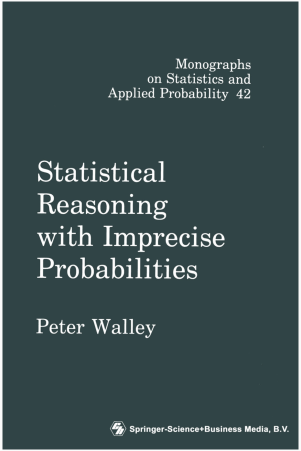

X_1
X_1, X_2

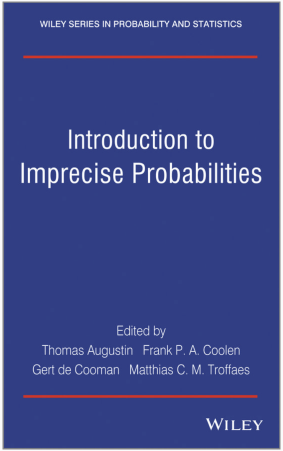
What if we cannot specify the local models p_s precisely?

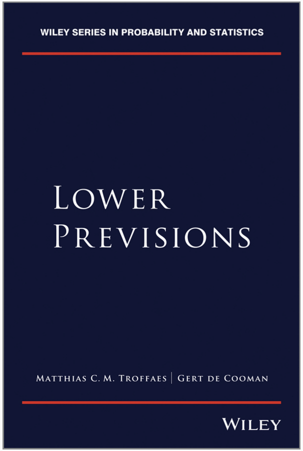
What if we cannot specify the local models p_s precisely?

¡imprecise probabilities!

Imprecise probabilities is a *collection* of theories that aim to generalise classical probability theory to allow for *partial specification*.







Let $\mathscr{L}$ denote the real vector space of all real-valued functions on $\mathscr{X}$, and let $\Sigma_{\mathscr{X}}$ denote the subset of all probability mass functions on $\mathscr{X}$:

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

Let $\mathscr{L}$ denote the real vector space of all real-valued functions on $\mathscr{X}$, and let $\Sigma_{\mathscr{X}}$ denote the subset of all probability mass functions on $\mathscr{X}$:

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

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

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

Let $\mathscr{L}$ denote the real vector space of all real-valued functions on $\mathscr{X}$, and let $\Sigma_{\mathscr{X}}$ denote the subset of all probability mass functions on $\mathscr{X}$:

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

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

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

Recall that

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

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

E3. $E_p(\lambda f) = \lambda E_p(f)$ for all $f \in \mathscr{L}$ and $\lambda \in \mathbb{R}$. [homogeneity]

Credal set

Instead of a single probability mass function p, we now consider

a credal set $\mathscr{M} \subseteq \Sigma_{\mathscr{X}}$,

a non-empty, closed and convex set of probability mass functions.

Credal set

Instead of a single probability mass function p, we now consider

a credal set $\mathscr{M} \subseteq \Sigma_{\mathscr{X}}$,

a non-empty, closed and convex set of probability mass functions.

A credal set is defined by constraints of the form

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

Credal set

Instead of a single probability mass function p, we now consider

a **credal set** $\mathscr{M} \subseteq \Sigma_{\mathscr{X}}$,

a non-empty, closed and convex set of probability mass functions.

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

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

Specifically of interest are the bounds

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

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

Lower expectation

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

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

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

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

Lower expectation

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

LE1. $\underline{E}(f) \geq \min f$ for all $f \in \mathscr{L}$; [boundedness]
LE2. $\underline{E}(f+g) \geq \underline{E}(f) + \underline{E}(g)$ for all $f, g \in \mathscr{L}$; [super-additivity]
LE3. $\underline{E}(\lambda f) = \lambda \underline{E}(f)$ for all $f \in \mathscr{L}$ and $\lambda \in \mathbb{R}_{>0}$. [positive homogeneity]

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

$$\underline{E}(f) = \underline{E}_{\mathscr{M}}(f) = \min\{E_p(f)\colon p \in \mathscr{M}\} \quad \textit{for all } f \in \mathscr{L}.$$

Lower expectation

An operator $\underline{E}\colon \mathscr{L} \to \mathbb{R}$ is called a lower expectation (coherent lo prevision) if

LE1. $\underline{E}(f) \geq \min f$ for all $f \in \mathscr{L}$; undedness]

LE2. $\underline{E}(f+g) \geq \underline{E}(f) + \underline{E}(g)$ for all $f, g \in$ additivity]

LE3. $\underline{E}(\lambda f) = \lambda \underline{E}(f)$ for all $f \in \mathscr{L}$ mogeneity]

Theorem

An operat and only if it is the lower envelope of some cred at

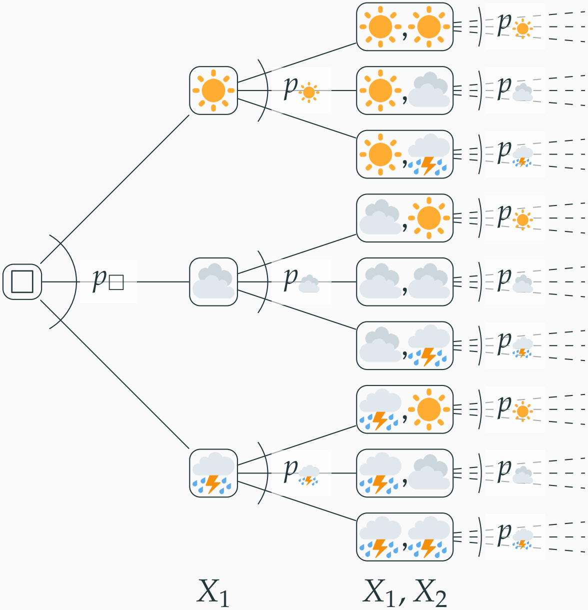
$$\underline{E}_{\mathscr{M}}(f) = \min\{E_p(f)\colon p \in \mathscr{M}\} \quad \textit{for all } f \in \mathscr{L}.$$

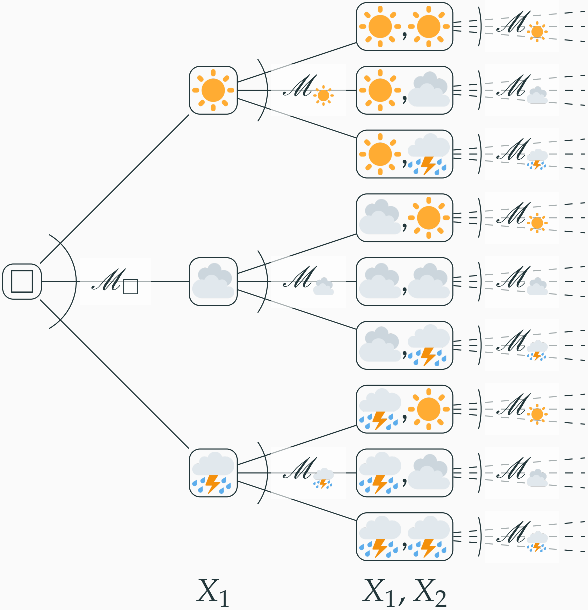
🤔 What does this have to do with stochastic processes? 🤔

Assume that we can only assess that

$$p_{\Box} \in \mathscr{M}_{\Box} \qquad \text{and} \qquad p_{(x_1,\dots,x_n)} \in \mathscr{M}_{x_n}, \tag{1}$$

where $\mathscr{M}_{\Box}$ and $\mathscr{M}_x$ for all $x \in \mathscr{X}$ are credal sets.


X_1
X_1, X_2


X_1
X_1, X_2

Assume that we can only assess that

$$p_\square \in \mathscr{M}_\square \qquad \text{and} \qquad p_{(x_1,\ldots,x_n)} \in \mathscr{M}_{x_n}. \tag{1}$$

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

- $\mathscr{P}^{\mathrm{CHM}}$: all compatible homogeneous Markov chains;
- $\mathscr{P}^{\mathrm{CM}}$: all compatible Markov chains;
- $\mathscr{P}^{\mathrm{C}}$: all compatible probability trees.

Can we compute

$$\underline{E}_{\mathscr{P}}(f \mid s) := \inf\{E_P(f \mid s) \colon P \in \mathscr{P}\}$$

and

$$\overline{E}_{\mathscr{P}}(f \mid s) := \sup\{E_P(f \mid s) \colon P \in \mathscr{P}\}?$$

Imprecise Markov chains

Computing these tight lower and upper bounds turns out to be

- 🙁 intractable for $\mathscr{P}^{\mathrm{CHM}}$,
- 😟 intractable for $\mathscr{P}^{\mathrm{CM}}$—at least in general,
- 😯 tractable for $\mathscr{P}^{\mathrm{C}}$, because we can use *backwards recursion* due to the *imprecise law of iterated expectation*.

- Gert de Cooman and Filip Hermans. "Imprecise probability trees: Bridging two theories of imprecise probability". In: *AI* 172.11 (2008), pp. 1400–1427
- Gert de Cooman, Filip Hermans, and Erik Quaeghebeur. "Imprecise Markov chains and their limit behavior". In: *PEIS* 23.4 (2009), pp. 597–635
- Damjan Škulj. "Discrete time Markov chains with interval probabilities". In: *IJAR* 50.8 (2009), pp. 1314–1329
- Stavros Lopatatzidis. "Robust modelling and optimisation in stochastic processes using imprecise probalities, with an application to queueing theory". PhD thesis. Ghent University, 2017
- Natan T'Joens, Jasper De Bock, and Gert de Cooman. "In Search of a Global Belief Model for Discrete-Time Uncertain Processes". In: *Proc. of ISIPTA 2019*. Vol. 103. PMLR. 2019, pp. 377–385

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 $\mathscr{X}$.

We consider the collection

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

of uncertain variables that take values in the finite state space $\mathscr{X}$.

Example

X_τ is the weather in Oviedo τ time units from now, and

$$\mathscr{X} = \{☀️, ☁️, ⛈️\}.$$

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

- What is the probability of ⛈️ after 4 days?
- How long do I have to wait until it is ☀️ again?
- Do I have to bring an ☂️ tomorrow?

Imprecise continuous-time Markov chains

DTMC
$p_\square, \{p_x\}_{x \in \mathscr{X}}$

→

imprecise DTMC
$\mathscr{M}_\square, \{\mathscr{M}_x\}_{x \in \mathscr{X}}$

CTMC
$p_0, \{q_x\}_{x \in \mathscr{X}}$

→

imprecise CTMC
$\mathscr{M}_0, \{\mathscr{Q}_x\}_{x \in \mathscr{X}}$

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.

📄 Damjan Škulj. "Efficient computation of the bounds of continuous time imprecise Markov chains". In: *AMC* 250 (2015), pp. 165–180

📄 Thomas Krak, Jasper De Bock, and Arno Siebes. "Imprecise continuous-time Markov chains". In: *IJAR* 88 (2017), pp. 452–528

A **counting process** is a model for a stream of events

X_τ: the number of events that have occurred up to time τ, so $\mathscr{X} = \mathbb{Z}_{\geq 0}$.

A **counting process** is a model for a stream of events

X_τ: the number of events that have occurred up to time τ, so $\mathscr{X} = \mathbb{Z}_{\geq 0}$.

Example

X_τ: the number of ⚡ lightning strikes that have hit the cathedral of Oviedo.

We want to answer questions like:

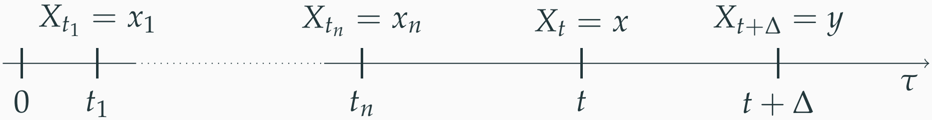
- What is the probability of at least one ⚡ in some time period?
- What is the expected number of ⚡ in the following year?
- What is the expected time until the next ⚡?

Counting processes in general

We model our beliefs by means of the transition probabilities

$$P(X_{t+\Delta} = y \mid X_t = x, \underbrace{X_{t_n} = x_n, \ldots, X_{t_1} = x_1}_{X_u = x_u}).$$

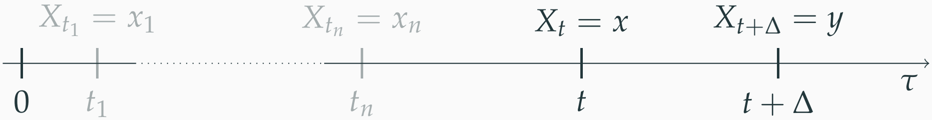
The Poisson process in particular



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

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

The Poisson process in particular

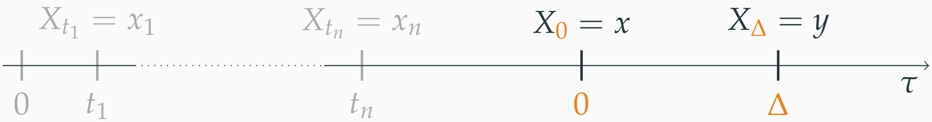


For the Poisson process, we additionally assume that the transition probabilities

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

1. only depend on the present, [Markovianity]

The Poisson process in particular

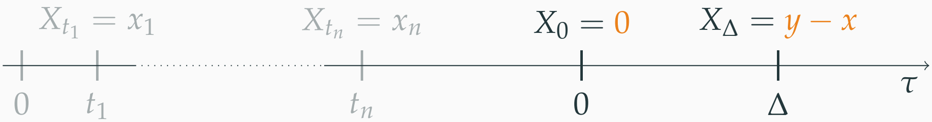


For the Poisson process, we additionally assume that the transition probabilities

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

1. only depend on the present, [Markovianity]
2. only depend on the length of the time period, [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, [Markovianity]
2. only depend on the length of the time period, [time homogeneity]
3. only depend on the number of new events. [state homogeneity]

The rate parameter

A Poisson process is uniquely characterised by a single parameter:

the **rate** λ!

The rate parameter

A Poisson process is uniquely characterised by a single parameter:

the **rate** λ!

It has multiple interpretations, for instance:

- the expected number of new events in any time period is proportional to λ:

$$E_P(X_{t+\Delta} \mid X_t = x, X_u = x_u) = x + \lambda\Delta;$$

The rate parameter

A Poisson process is uniquely characterised by a single parameter:

the **rate** λ!

It has multiple interpretations, for instance:

- the expected number of new events in any time period is proportional to λ:

$$E_P(X_{t+\Delta} \mid X_t = x, X_u = x_u) = x + \lambda\Delta;$$

- λ 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 λ precisely, but only know that it belongs to the rate interval $[\underline{\lambda}, \overline{\lambda}]$?

The general approach

Let $\mathscr{P}$ be a set of counting processes characterised by the rate interval $[\underline{\lambda}, \overline{\lambda}]$,

and define the lower expectation

$$\underline{E}_{\mathscr{P}}(f \mid X_t = x, X_u = x_u) := \inf\{E_P(f \mid X_t = x, X_u = x_u) \colon P \in \mathscr{P}\}.$$

Choose $\mathscr{P}$ such that

(i) computing $\underline{E}_{\mathscr{P}}(f \mid X_t = x, X_u = x_u)$ is tractable,

(ii) $\underline{E}_{\mathscr{P}}(\cdot \mid \cdot)$ is Poisson-like, in the sense that

(a) $\underline{E}_{\mathscr{P}}(g(X_{t+\Delta}) | X_t = x, X_u = x_u)$ is Markov and homogeneous,

(b) $\underline{E}_{\mathscr{P}}(X_{t+\Delta} \mid X_t = x, X_u = x_u) = x + \underline{\lambda}\Delta.$

A naive imprecise Poisson process

If $\mathscr{P}$ is the **set of all Poisson processes** with rate λ in the rate interval $[\underline{\lambda}, \overline{\lambda}]$, then

- 😃 computing $\underline{E}_{\mathscr{P}}(f \mid X_t = x, X_u = x_u)$ is a one-parameter optimisation problem;
- 😃 $\underline{E}_{\mathscr{P}}(\cdot \mid \cdot)$ is Poisson-like;
- 🙁 every P in $\mathscr{P}$ is Markov and homogeneous.

An alternative condition

$(\forall P \in \mathscr{P})(\exists \lambda \in [\underline{\lambda}, \overline{\lambda}])(\forall t, \Delta, x, x_u \dots)$

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

An alternative condition

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

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

$$\Downarrow$$

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

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

A more involved imprecise Poisson process

If $\mathscr{P}$ 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 $\mathscr{P}$ is not necessarily Markov nor homogeneous;
🤯 computing $\underline{E}_{\mathscr{P}}(f \mid X_t = x, X_u = x_u)$ is non-trivial (if not infeasible).

A more involved imprecise Poisson process

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

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

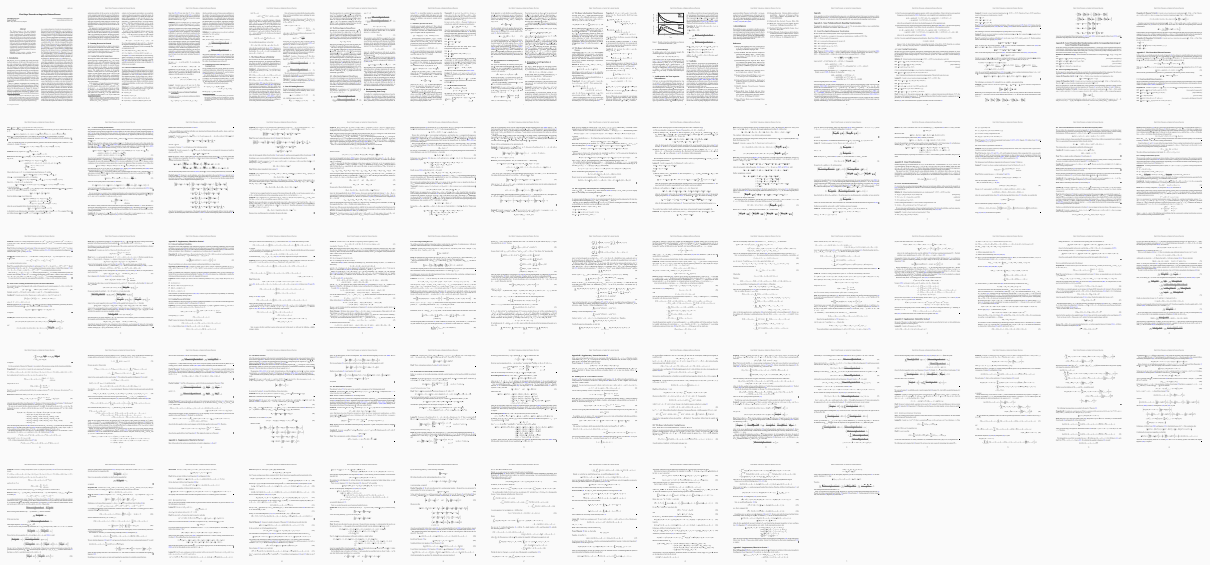
then

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

However, it turns out that

🤓 computing $\underline{E}_{\mathscr{P}}(g(X_{t+\Delta}) \mid X_t = x, X_u = x_u)$ is tractable;
🤓 $\underline{E}_{\mathscr{P}}(\cdot \mid \cdot)$ is Poisson-like.

Alexander Erreygers and Jasper De Bock. “First Steps Towards an Imprecise Poisson Process”. In: *Proc. of ISIPTA 2019*. Vol. 103. PMLR. 2019, pp. 175–184



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.