A Coinductive Topology for Reasoning about Markov Processes

Giorgio Bacci

Aalborg University

based on joint work with Giovanni Bacci, Kim G. Larsen, Radu Mardare, Prakash Panangaden and Daniele Toller

OPCT 2023 - June 27th 2023

Motivations

- Many models of computations deal with numerical values
- Reasoning about equivalence of systems is not enough
- We would like to quantify the differences and/or tell which behaviour is closest to a given one

Approximate behavioural reasoning

- **A necessity:** inherent errors in measurements, partial knowledge of the models, imprecise specifications, etc…
- **An opportunity:** faster approximate solutions, enhanced model reductions, data extrapolation, working simpler approximations, etc…

Good Behavioural distances

- They should differentiate processes only on their **behaviour**

$$d(p,q) = 0 \text{ iff } p \sim q$$

- They should differentiate on **logical properties**

$$d(p,q) = \sup_{\phi \in \mathscr{L}} \phi(p) - \phi(q)$$

typically, fuzzy-logics
$\phi : X \to [0,1]$

- It should come with **algorithms** to compute $d(p,q)$ (ideally, with low time-complexity)

- small differences in the processes = small variation in the distances

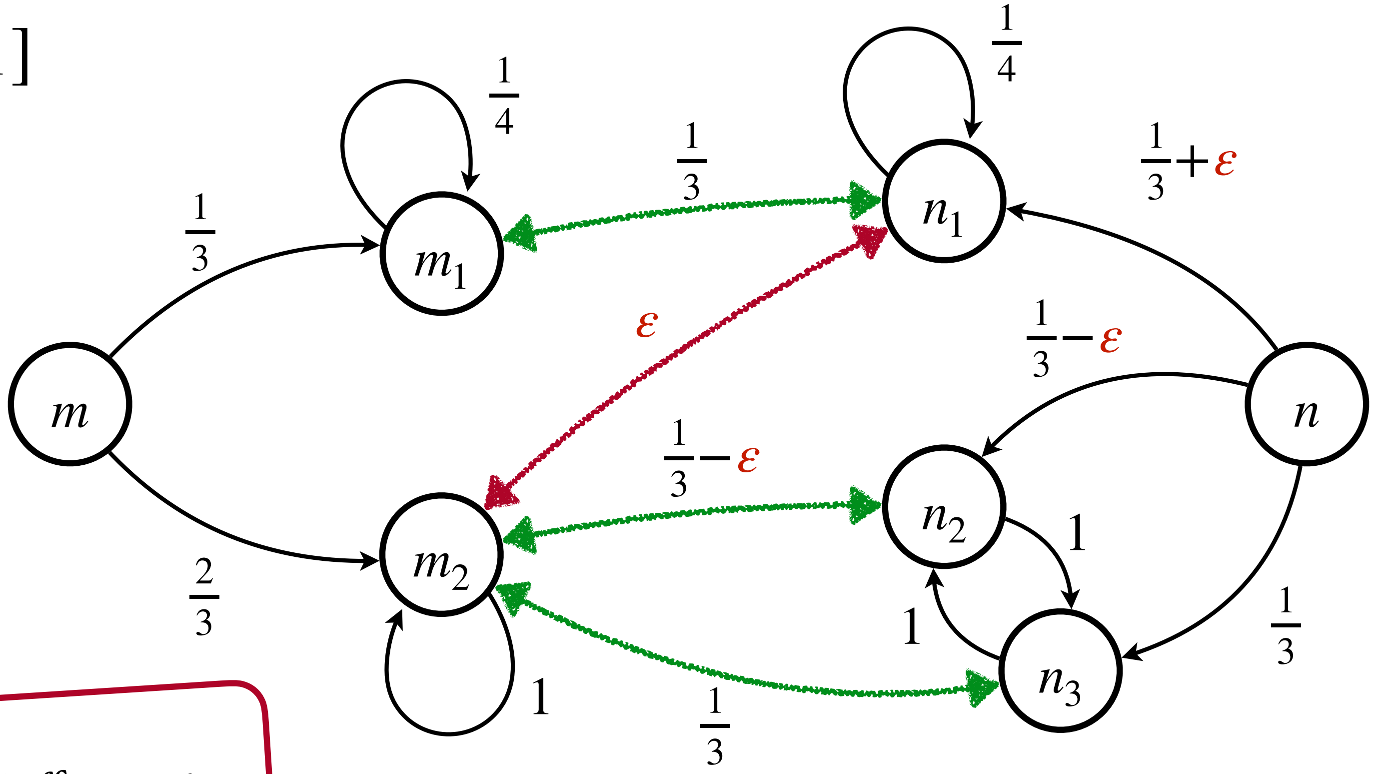
An extensive literature on the topic, especially on probabilistic systems

The probabilistic bisimilarity distances

(a.k.a. the *fixed point Kantorovich distance*)

Desharnais et al. (CONCUR'99)
van Breugel-Worrell (ICALP'01)

$\lambda \in (0,1]$



- $\mathbf{d}_\lambda(m,n) = 0$ iff $m \sim n$
- logical characterization
- polytime computable

Kantorovich lifting

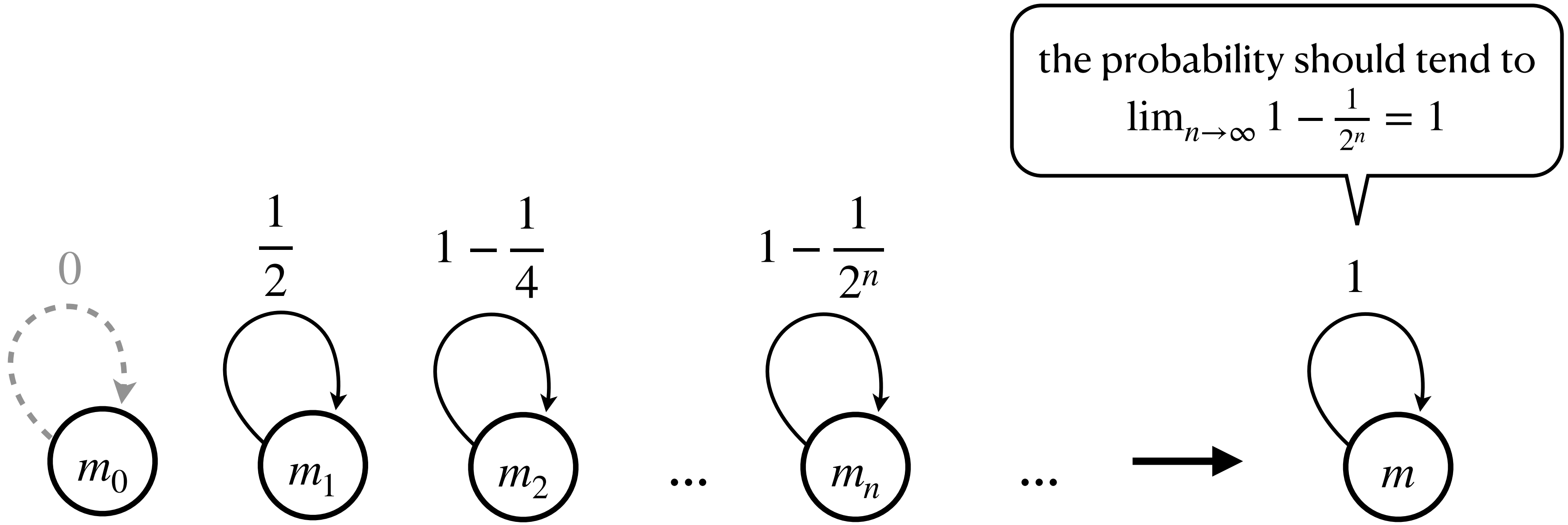
transition sub-probabilities

$$\mathbf{d}_\lambda(m,n) = \lambda \cdot \mathscr{K}(\mathbf{d}_\lambda)(\theta(m), \theta(n))$$

Converging Behaviours

"processes that are close should have probability that are close"

(Giacalone, Jou, Smolka '90)



… the (undiscounted) probabilistic bisimilarity distance does not make this sequence of behaviours converge

$$\forall n \,.\, \mathbf{d}_1(m_n, m) = 1$$

Toward a notion of approximation

- It's **a topological concept** (not necessary a metric one)
  Indeed, many natural notions of convergence are non metrizable (e.g., point-wise convergence)

- To define a notion of approximation is to give a neighbourhood system (the neighbourhood filters of each point)

- Should be driven by a notion of observation

a set $\mathscr{F}$ of observations $f\colon X \to O$

domain of observed properties
(may be a metric space)

"processes that are close should have probability that are close"

we require that all $f \in \mathscr{F}$ **are continuous** (i.e., preserve similarity)

Our case study: Markov Processes

generic measurable space with Σ a σ-algebra on X

Definition

A **Markov process** on (X, Σ) is $\theta \colon X \times \Sigma \to [0,1]$ such that

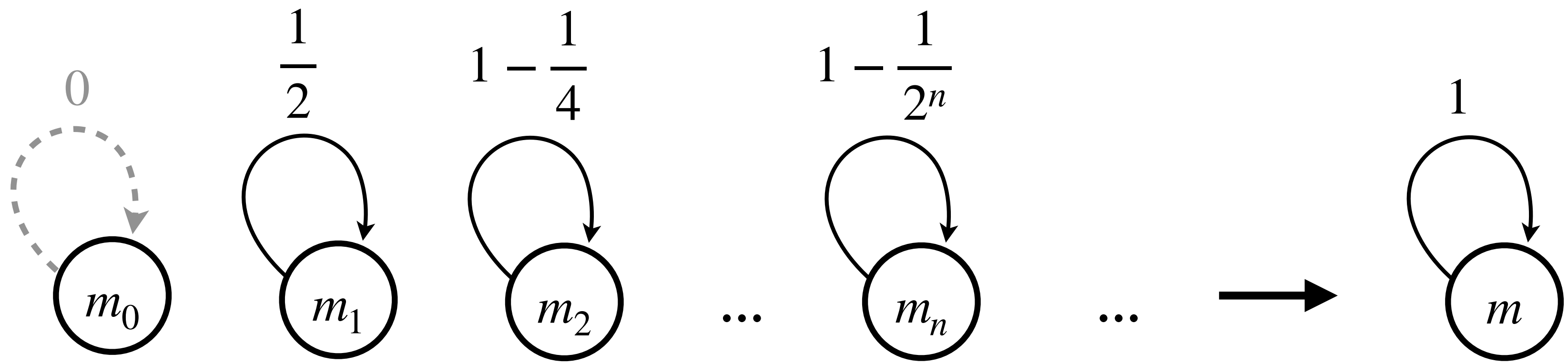
- for all $x \in X$, $\theta(x, \_)$ is a sub-probability distribution on (X, Σ)
- for all $E \in \Sigma$, $\theta(\_, E)$ is a measurable function

sub-probabilistic Giry functor
$\Delta(X) = \{\mu \mid \mu \text{ sub-probability on } X\}$

- Equivalent to the coalgebras of ΔX

$$X \to \Delta X \quad \text{in } \mathbf{Meas}$$

- Markov chains are a special case (with discrete-state)
- *we don't assume (X, Σ) comes from a topological space*



The formal Markov process

Let $M = \{m_n \mid n \in \mathbb{N}\} \cup \{m\}$ and $\theta\colon M \times \mathscr{P}(M) \to [0,1]$ the Markov process where

discrete σ-algebra

$$\theta(m_n, E) = \begin{cases} 1 - \frac{1}{2^n} & m_n \in E \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad \theta(m, E) = \begin{cases} 1 & m \in E \\ 0 & \text{otherwise} \end{cases}$$

Bisimulation Topology

The type of observations that we are interested in are of the form

$$\mathbb{E}_\theta[f] \colon X \to [0,1]$$

with euclidean metric

A random variable
$f \colon (X, \Sigma) \to [0,1]$

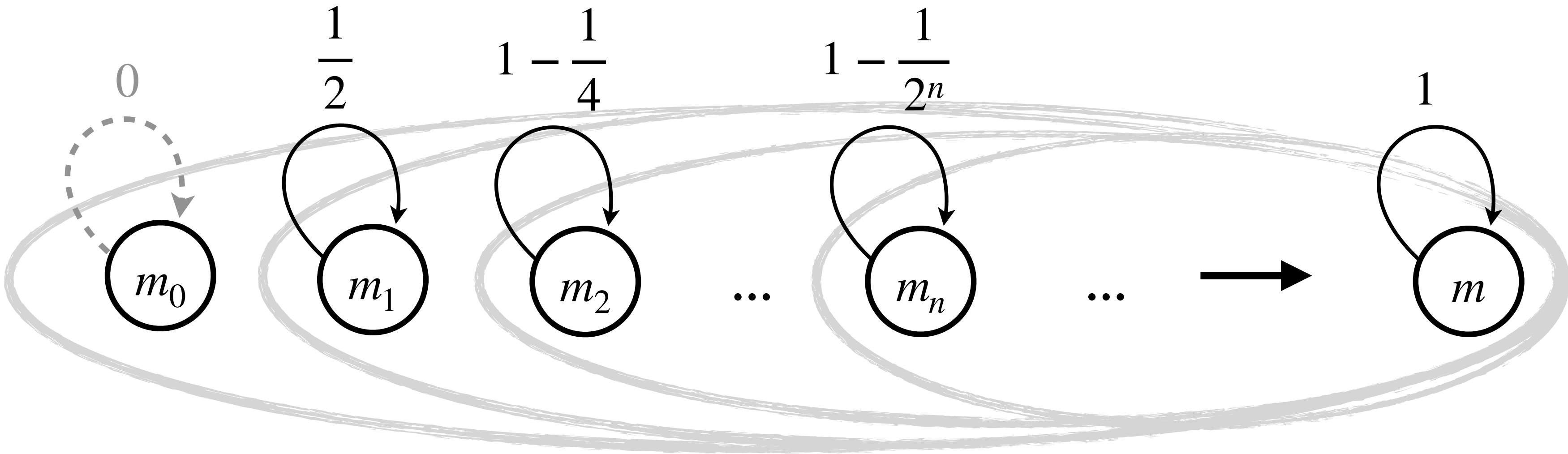
$$\mathbb{E}_\theta[f](x) \overset{\text{def}}{=} \int f \, d\theta(x, \_)$$

Definition (bisimulation topology)

Let $\theta \colon X \times \Sigma \to [0,1]$ be a Markov process. A topology τ on X is a **bisimulation topology** if the following implication holds

$$f \in \mathscr{C}_\Sigma(X) \implies \mathbb{E}_\theta[f] \in \mathscr{C}_\Sigma(X)$$

τ-continuous random variables



Let $f\colon (M, \mathscr{P}(M)) \to [0,1]$ a random variable (any function!), then

$$\forall n \in M\,.\; \mathbb{E}_\theta[f](n) = f(n) \cdot \theta(n, M)$$

If we force $\theta(\_, X)$ to be continuous then, f continuous iff $\mathbb{E}_\theta[f]$ continuous

A bisimulation topology

$$\tau = \Big\{ \{n \mid \theta(n, M) \in O \subseteq [0,1] \text{ open}\} \Big\}$$

the smallest topology on X that makes $\theta(\_, X)\colon X \to [0,1]$ continuous

A coinductive proof principle

Smallest bisimulation topology

Let $\theta\colon X \times \Sigma \to [0,1]$ be a Markov process, and $\mathscr{T}$ a family of bisimulation topologies on X, then $\bigcap \mathscr{T}$ is bisimulation topology.

Then, the smallest bisimilarity topology is

bisimilarity topology

$$\hat{\tau} = \bigcap \{\tau \mid \tau \text{ bisimulation topology}\}$$

$$\frac{(x_t)_{t\in T} \xrightarrow{\tau} x \qquad \tau \text{ bisimulation topology}}{(x_t)_{t\in T} \xrightarrow{\hat{\tau}} x}$$

net convergence

It's a behavioural topology!

$$x \equiv_{\hat{\tau}} y \quad \text{iff} \quad x \equiv_{\mathscr{E}} y$$

topologically indistinguishable

event bisimilar

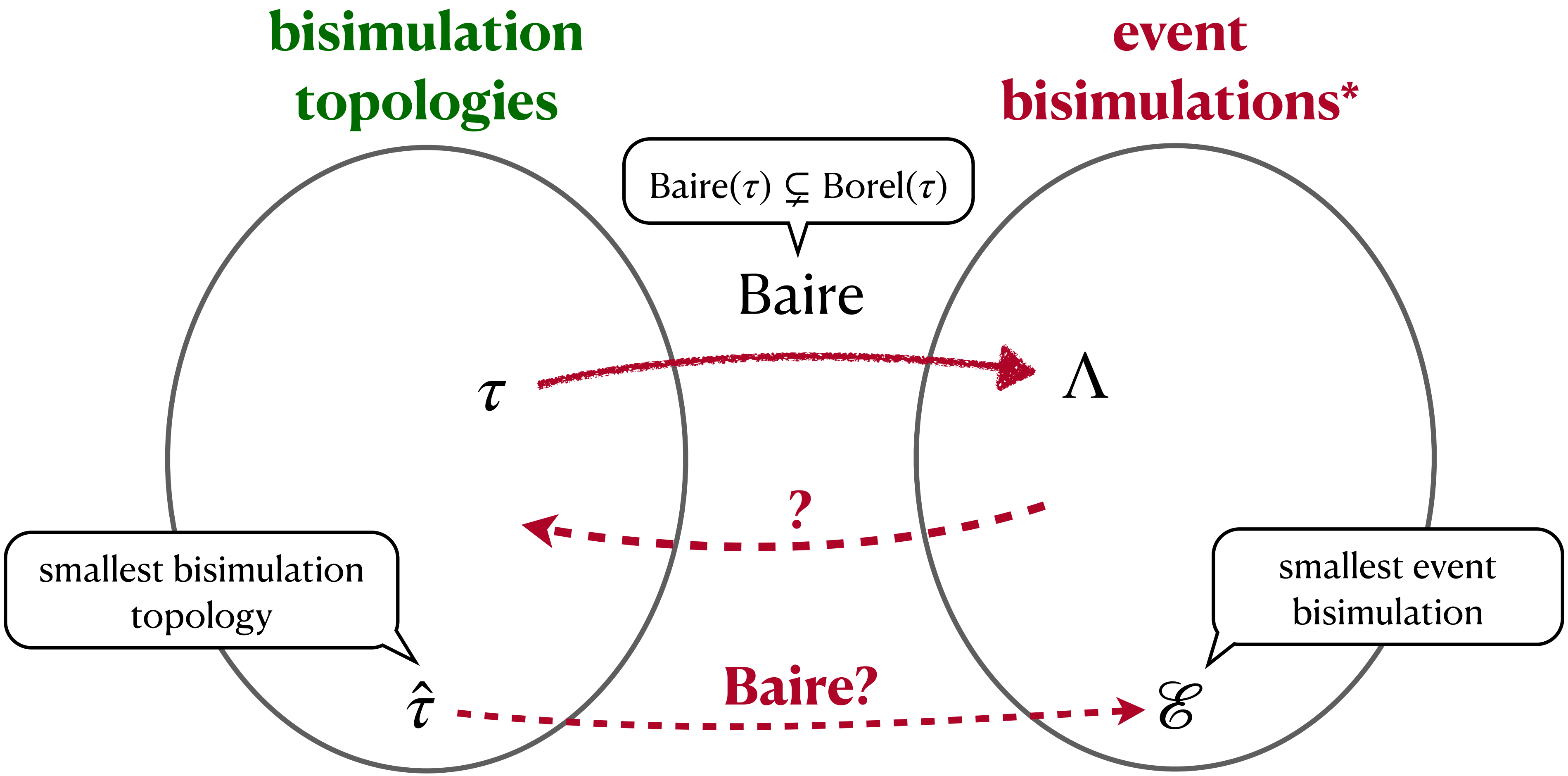
(Danos et al. '06)

Corollary

$$\exists (z_t)_{t \in T} .\ \left((z_t)_{t \in T} \xrightarrow{\hat{\tau}} x \text{ and } (z_t)_{t \in T} \xrightarrow{\hat{\tau}} y \right) \quad \text{iff} \quad x \sim_e y$$

a net of approximants that witnesses the similarity in the behaviours

...some more on behaviours



(*) σ-algebras $\Lambda \subseteq \Sigma$ such that $E \in \Lambda \Rightarrow \forall r.\ \{x \mid \theta(x, E) \geq r\} \in \Lambda$

Attacking $\mathscr{E} = \text{Baire}(\hat{\tau})$

Let $\mathscr{G}$ be the family of functions from X to $[0,1]$ generated by grammar

$$g, f ::= \mathbf{1} \mid r \cdot g \mid g \oplus f \mid 1 - g \mid \min(g, f) \mid \max(g, f) \mid \mathbb{E}_\theta[g]$$

Theorem

- $\mathscr{E}$ is the smallest σ-algebra making all $g \in \mathscr{G}$ measurable
- $\mathscr{G}$ is dense in $\mathscr{M}(\mathscr{E})$ wrt point-wise convergence

Open problem

Stone-Weierstrass like result

$\mathscr{G}$ is dense in $\mathscr{C}(\hat{\tau})$ wrt uniform convergence

it implies $\mathscr{E} = \text{Baire}(\hat{\tau})$

On pseudometrizability

Not all topologies come from (pseudo)metric, i.e., are the open ball topologies of some (pseudo)metric

Proposition

a transfinite construction

Base case: $d_0 = \sqcap \{d \in \mathbf{Pmet}(X) \mid \mathscr{G} \subseteq \mathscr{C}(X, d)\}$

Inductive step:

$$d_{n+1} = \sqcap \{d \in \mathbf{Pmet}(X) \mid \mathscr{C}(X, d_n) \cup \mathbb{E}_\theta(\mathscr{C}(X, d_n)) \subseteq \mathscr{C}(X, d)\}$$

Limit step (α limit ordinal): $d_\alpha = \sqcap \{d_\beta \mid \beta < \alpha\}$

If $d_\kappa = d_{\kappa+1}$, then τ_{d_κ} is a bisimulation topology

Open problem

$\hat{\tau}$ is pseudo-metrizable

Let κ the smallest ordinal such that $d_\kappa = d_{\kappa+1}$. Then, $\hat{\tau} = \tau_{d_\kappa}$.

Approximations & Logical properties

Theorem (Mardare et al. '12)

If d is a *dynamically continuous* bisimulation pseudometric,

$$(m_n)_{n\in\mathbb{N}} \xrightarrow{d} m \wedge (\phi_n)_{n\in\mathbb{N}} \xrightarrow{H(d)} \phi \wedge (\forall n \,.\, m_n \vDash \phi_n) \Longrightarrow m \vDash \phi$$

positive logical formulas in $\mathscr{L}^+$

A more abstract equivalent statement

The satisfiability map $\vDash \,:\, X \times \mathscr{L}^+ \to \{0,1\}$ is continuous

Sierpiński space with topology $\{\varnothing, \{0\}, \{0,1\}\}$

$$\frac{\vDash \,:\, X \times \mathscr{L} \to \{0,1\}}{\vDash^\dagger \,:\, \mathscr{L} \to \{0,1\}^X}$$

continuous maps to Sierpiński space (in bijection with closed sets of X)

Conclusions

- We proposed a **coinductive topology** for reasoning about approximations of behaviours of Markov processes
- Still an **ongoing work** with lots of unresolved problems

- Our way of investigating the limits of behavioural distances
- The same approach can be **relevant for other types of models** (we played already bit with stream systems)