UNIVERSITEIT GENT

FACULTY OF ENGINEERING

Epistemic irrelevance in credal networks

Gert de Cooman

Ghent University, SYSTeMS
gert.decooman@UGent.be

Imprecise Probability Workshop
Université Paris Dauphine
28 October 2008

Before we start

Precise probability models

Mass functions and expectations

Assume we are uncertain about:

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

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

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

With p we can associate an expectation operator E_p:

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

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

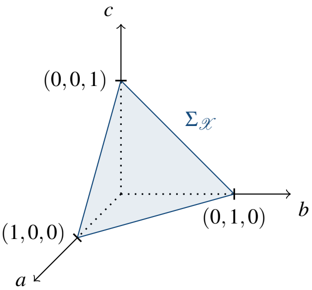
$$P_p(A) = \sum_{x\in A} p(x) = E_p(I_A).$$

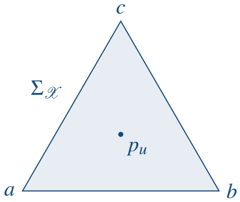
Precise probability models

The simplex of all probability mass functions

Consider the simplex $\Sigma_{\mathscr{X}}$ of all mass functions on $\mathscr{X}$:

$$\Sigma_{\mathscr{X}} := \left\{ p \in \mathbb{R}_{+}^{\mathscr{X}} : \sum_{x \in \mathscr{X}} p(x) = 1 \right\}.$$



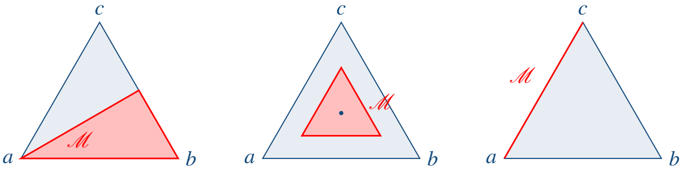


Imprecise probability models

Credal sets

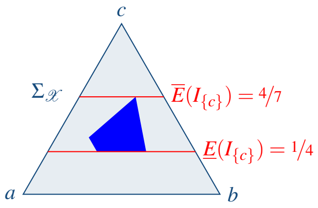
Definition

A credal set $\mathscr{M}$ is a convex closed subset of $\Sigma_{\mathscr{X}}$.



Imprecise probability models

Lower and upper expectations



Equivalent model

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

$$\underline{E}_{\mathscr{M}}(f) = \min\left\{E_p(f) : p \in \mathscr{M}\right\} \text{ lower expectation}$$
$$\overline{E}_{\mathscr{M}}(f) = \max\left\{E_p(f) : p \in \mathscr{M}\right\} \text{ upper expectation.}$$

Observe that [conjugacy]: $\overline{E}_{\mathscr{M}}(f) = -\underline{E}_{\mathscr{M}}(-f).$

Imprecise probability models

Basic properties of upper expectations

Definition

We call a real functional $\overline{E}$ on $\mathscr{L}(\mathscr{X})$ an upper expectation if it satisfies the following properties:
For all f and g in $\mathscr{L}(\mathscr{X})$ 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 (Other properties)

Let $\overline{E}$ be an upper expectation, with conjugate lower expectation $\underline{E}$. Then for all real numbers μ and all f and g in $\mathscr{L}(\mathscr{X})$:

1. $\underline{E}(f) \leq \overline{E}(f)$;
2. $\underline{E}(f) + \underline{E}(g) \leq \underline{E}(f+g) \leq \underline{E}(f) + \overline{E}(g) \leq \overline{E}(f+g) \leq \overline{E}(f) + \overline{E}(g)$;
3. $\overline{E}(f+\mu) = \overline{E}(f) + \mu$;
4. $\overline{E}(|f|) \geq |\underline{E}(f)|$ *and* $\overline{E}(|f|) \geq |\overline{E}(f)|$.

Theorem (Lower Envelope Theorem)

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

Proof.

Use $\mathscr{M} = \{p \in \Sigma_{\mathscr{X}} : (\forall f \in \mathscr{L}(\mathscr{X}))(E_p(f) \leq \overline{E}(f))\}$. □

Types of independence

Three possible definitions

Epistemic irrelevance

X_2 is epistemically irrelevant to X_1, conditional on X_3:

$$\overline{E}(f(X_1)|X_2,X_3) = \overline{E}(f(X_1)|X_3)$$

Epistemic independence

X_1 and X_2 are epistemically independent, conditional on X_3:

$$\overline{E}(f(X_1)|X_2,X_3) = \overline{E}(f(X_1)|X_3) \text{ and } \overline{E}(g(X_2)|X_1,X_3) = \overline{E}(g(X_2)|X_3)$$

Strong independence

Model $\overline{E}(h(X_1,X_2)|X_3)$ is an upper envelope of precise independent models

Discrete-time uncertain processes

Precise probability trees

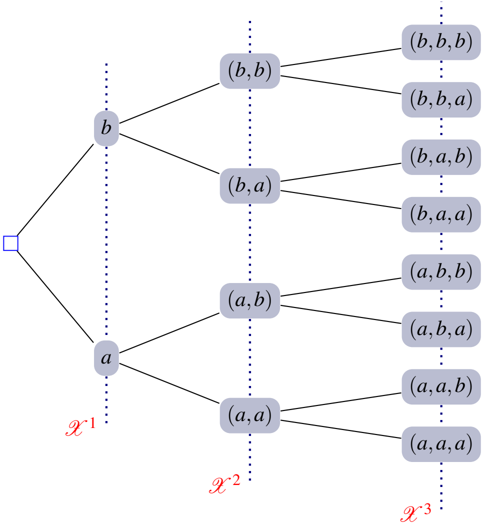
We consider an uncertain process with variables $X_1, X_2, \ldots, X_n, \ldots$ assuming values in a finite set of states $\mathscr{X}$.

This leads to a standard event tree with nodes

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

Discrete-time uncertain processes

Precise probability trees



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 $\mathscr{X}$.

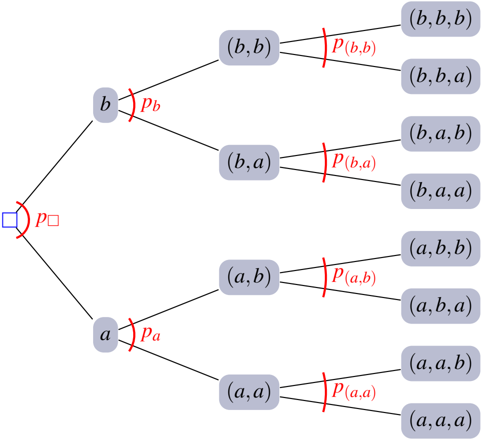
This leads to a standard event tree with nodes

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

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

Discrete-time uncertain processes

Precise probability trees



Precise probability trees

Calculating global expectations from local ones

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

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

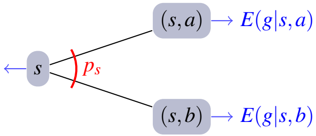
We want to calculate its expectation $E(g|s)$ in $s = (x_1, \dots, x_k)$.

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:

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

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



Precise probability trees

Calculating global expectations from local ones

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

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

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

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

$$E(g|x_1,\dots,x_k) = \underbrace{E_{(x_1,\dots,x_k)}}_{\text{local}}(E(g|x_1,\dots,x_k,\cdot))$$

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

$$E(g|\square) = E(g).$$

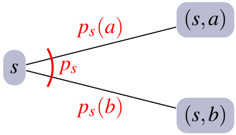
UNIVERSITEIT GENT

Imprecise probability models

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 $\mathscr{X}$ of possible values for the next state.



More general uncertainty models

We consider credal sets as more general uncertainty models: closed convex subsets of $\Sigma_{\mathscr{X}}$.

Imprecise probability trees

Definition and interpretation

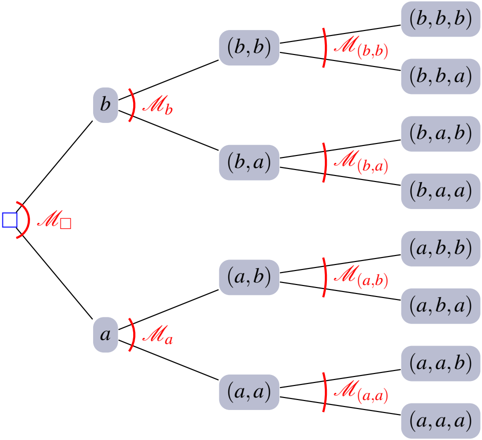
Definition

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

$$\overline{E}_s(f) = \max\left\{E_p(f) \colon p \in \mathscr{M}_s\right\} \text{ for all real maps } f \text{ on } \mathscr{X}.$$

Imprecise probability trees

Definition and interpretation



UNIVERSITEIT GENT

Imprecise probability trees

Definition and interpretation

Definition

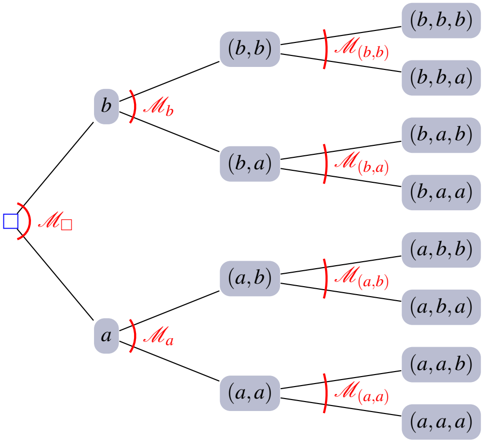
An imprecise probability tree is a probability tree where in each node s the local uncertainty model is an imprecise probability model $\mathscr{M}_s$, or equivalently, its associated upper expectation $\overline{E}_s$:

$$\overline{E}_s(f) = \max\left\{E_p(f)\colon p \in \mathscr{M}_s\right\} \text{ for all real maps } f \text{ on } \mathscr{X}.$$

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 $\mathscr{M}_s$.

Imprecise probability trees

Definition and interpretation



Imprecise probability trees

Associated lower and upper expectations

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

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

using the backwards recursion method described before.

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

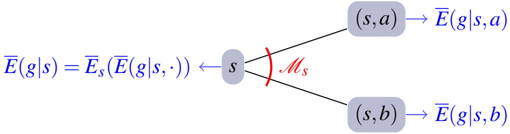
$$[\underline{E}(g|x_1, \dots, x_k), \overline{E}(g|x_1, \dots, x_k)]$$

We want a better, more efficient method to calculate these lower and upper expectations $\underline{E}(g|x_1, \dots x_k)$ and $\overline{E}(g|x_1, \dots, x_k)$.

Theorem (Law of Iterated Expectation)

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

$$\overline{E}(g|s) = \underbrace{\overline{E}_s}_{\text{local}}(\overline{E}(g|s,\cdot)) = \max_{p_s \in \mathscr{M}_s} \sum_{x \in \mathscr{X}} p_s(x)\overline{E}(g|s,x).$$



The complexity of calculating the $\overline{E}(g|s)$, as a function of n, is therefore essentially the same as in the precise case!

Precise Markov chains

Definition

Definition

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

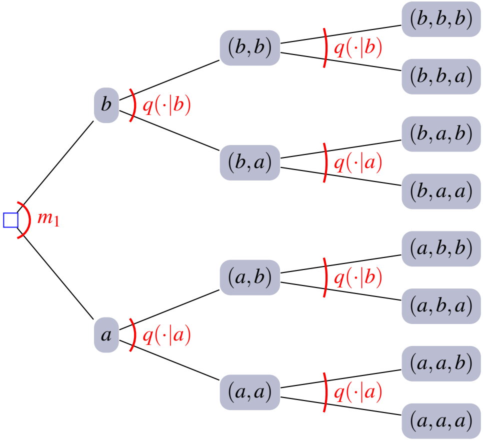
1. $\mathscr{M}_\square = \{m_1\}$,
2. the Markov Condition is satisfied:

$$\boxed{\mathscr{M}_{(x_1,\dots,x_n)} = \{q(\cdot|x_n)\}.}$$

UNIVERSITEIT GENT

Precise Markov chains

Definition



Precise Markov chains

Definition

Definition

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

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

$$\boxed{\mathscr{M}_{(x_1,\dots,x_n)} = \{q(\cdot|x_n)\}.}$$

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

$$E(f|x) = \sum_{z\in\mathscr{X}} q(z|x)f(z).$$

UNIVERSITEIT GENT

Precise Markov chains

Transition operators

Definition

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

$$\mathrm{T}\colon \mathscr{L}(\mathscr{X}) \to \mathscr{L}(\mathscr{X}) : f \mapsto \mathrm{T}f$$

where $\mathrm{T}f$ is the real map given by, for any $x \in \mathscr{X}$:

$$\boxed{\mathrm{T}f(x) := E(f|x) = \sum_{z\in\mathscr{X}} q(z|x)f(z)}$$

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

UNIVERSITEIT GENT

Precise Markov chains

Transition operators

Definition

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

$$\mathrm{T}\colon \mathscr{L}(\mathscr{X}) \to \mathscr{L}(\mathscr{X}) : f \mapsto \mathrm{T}f$$

where $\mathrm{T}f$ is the real map given by, for any $x \in \mathscr{X}$:

$$\boxed{\mathrm{T}f(x) := E(f|x) = \sum_{z \in \mathscr{X}} q(z|x) f(z)}$$

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:

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

Complexity is linear in the number of time steps n.

Imprecise Markov chains

Definition

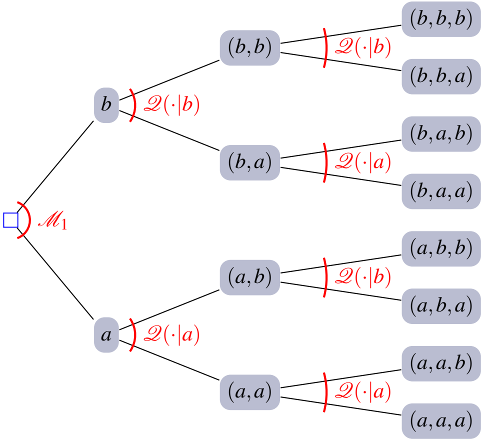
Definition

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

$$\boxed{\mathscr{M}_{(x_1,\dots,x_n)} = \mathscr{Q}(\cdot|x_n).}$$

Imprecise Markov chains

Definition



Imprecise Markov chains

Definition

Definition

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

$$\boxed{\mathscr{M}_{(x_1,\dots,x_n)} = \mathscr{Q}(\cdot|x_n).}$$

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

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

$$\underline{E}(f|x) = \min\left\{E_p(f) \colon p \in \mathscr{Q}(\cdot|x)\right\}$$
$$\overline{E}(f|x) = \max\left\{E_p(f) \colon p \in \mathscr{Q}(\cdot|x)\right\}.$$

Imprecise Markov chains

Lower and upper transition operators

Definition

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

$$\underline{\mathrm{T}}\colon \mathscr{L}(\mathscr{X}) \to \mathscr{L}(\mathscr{X}) : f \mapsto \underline{\mathrm{T}}f$$
$$\overline{\mathrm{T}}\colon \mathscr{L}(\mathscr{X}) \to \mathscr{L}(\mathscr{X}) : f \mapsto \overline{\mathrm{T}}f$$

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

$$\boxed{\begin{aligned}\underline{\mathrm{T}}f(x) &:= \underline{E}(f|x) = \min\left\{E_p(f) : p \in \mathscr{Q}(\cdot|x)\right\}\\ \overline{\mathrm{T}}f(x) &:= \overline{E}(f|x) = \max\left\{E_p(f) : p \in \mathscr{Q}(\cdot|x)\right\}\end{aligned}}$$

Imprecise Markov chains

Lower and upper transition operators

Definition

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

$$\underline{\mathrm{T}}\colon \mathscr{L}(\mathscr{X}) \to \mathscr{L}(\mathscr{X}) : f \mapsto \underline{\mathrm{T}}f$$
$$\overline{\mathrm{T}}\colon \mathscr{L}(\mathscr{X}) \to \mathscr{L}(\mathscr{X}) : f \mapsto \overline{\mathrm{T}}f$$

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

$$\underline{\mathrm{T}}f(x) := \underline{E}(f|x) = \min\left\{E_p(f) : p \in \mathscr{Q}(\cdot|x)\right\}$$
$$\overline{\mathrm{T}}f(x) := \overline{E}(f|x) = \max\left\{E_p(f) : p \in \mathscr{Q}(\cdot|x)\right\}$$

Then the Law of Iterated Expectation yields:

$$\underline{E}_n(f) = \underline{E}_1(\underline{\mathrm{T}}^{n-1}f) \text{ and } \overline{E}_n(f) = \overline{E}_1(\overline{\mathrm{T}}^{n-1}f).$$

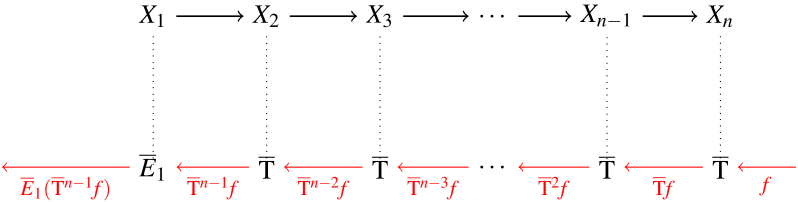
Complexity is still linear in the number of time steps n.

Imprecise Markov chains

Message passing

Important observation

The backpropagation can be seen as message passing.



A special credal network

under epistemic irrelevance

An imprecise Markov chain can also be depicted as follows:

$$X_1 \longrightarrow X_2 \longrightarrow X_3 \longrightarrow \cdots \longrightarrow X_{n-1} \longrightarrow X_n$$

Interpretation of the graph

Conditional on X_k we have that $X_1, \ldots, X_{k-1}$ are epistemically irrelevant to $X_{k+1}, \ldots, X_n$:

$$\overline{E}(f(X_{k+1},\ldots,X_n)|X_1,\ldots,X_{k-1},X_k) = \overline{E}(f(X_{k+1},\ldots,X_n)|X_k)$$

Credal networks under epistemic irrelevance

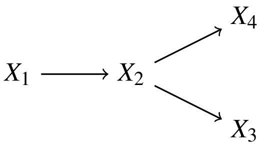
Definition

The graphical structure is interpreted as follows:

Conditional on the parents, the non-parent non-descendants of each node are epistemically irrelevant to it.

Credal networks under epistemic irrelevance

Example



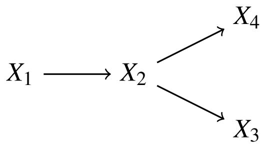
- X_1 is epistemically irrelevant to X_3, conditional on X_2
- X_3 need not be epistemically irrelevant to X_1, conditional on X_2.

Conclusion

X_1 and X_3 need not be epistemically, and certainly not strongly independent, conditional on X_2.

Credal networks under epistemic irrelevance

Example



- X_3 is epistemically irrelevant to X_4, conditional on X_2
- X_4 is epistemically irrelevant to X_3, conditional on X_2.

Conclusion

X_3 and X_4 are epistemically, but not necessarily strongly, independent, conditional on X_2.

UNIVERSITEIT GENT

Credal networks under epistemic irrelevance

Some separation properties

Figure: I_2 separates T from I_1.

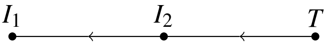


Figure: I_2 doesn't separate T from I_1.

Conclusion

For a variable T to be separated from I_2 by a variable I_1, arrows should point from I_2 to T.

Credal networks under epistemic irrelevance

As an expert system

Message passing algorithm

- when the credal network is a (Markov) tree
- treated as an expert system
- linear complexity in the number of nodes

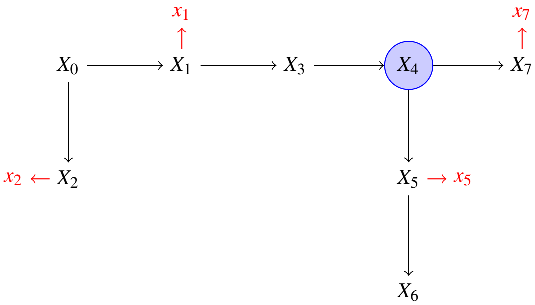
Python code

- written by Filip Hermans
- testing and connection with strong independence by Alessandro Antonucci

UNIVERSITEIT GENT

An example

A particular Markov tree

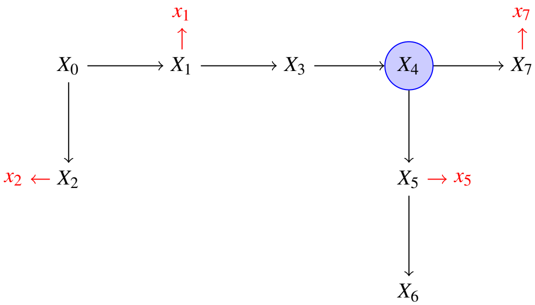


We are looking for:

$$\underline{E}(f(X_4)|x_1, x_2, x_5, x_7)$$

An example

A particular Markov tree

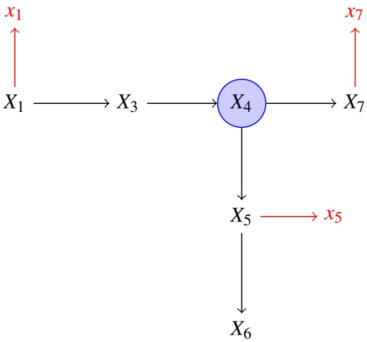


This is the unique μ such that:

$$\underline{E}([f(X_4) - \mu]I_{\{x_1\}}I_{\{x_2\}}I_{\{x_5\}}I_{\{x_7\}}) = 0$$

An example

A particular Markov tree

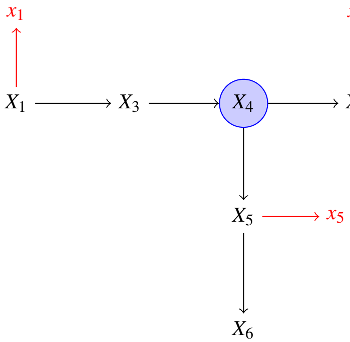


This is the unique μ such that:

$$\underline{E}([f(X_4) - \mu]I_{\{x_5\}}I_{\{x_7\}}|x_1) = 0$$

An example

A particular Markov tree

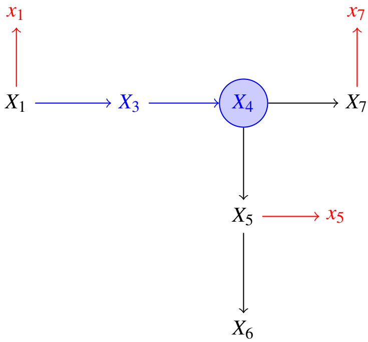


This is the unique μ such that:

$$\underline{E}_3(\underline{E}_4([f(X_4)-\mu]\overline{E}_5(\{x_5\}|X_4)\overline{E}_7(\{x_7\}|X_4)|X_3)|x_1) = 0$$

An example

A particular Markov tree



This is the unique μ such that:

$$\underline{E}_3\left(\underline{E}_4\left(\underbrace{[f(X_4)-\mu]\overbrace{\overline{E}_5(\{x_5\}|X_4)\overline{E}_7(\{x_7\}|X_4)}^{\text{passed to the backbone}}}_{\text{passed along the backbone}} \,\middle|\, X_3\right) \,\middle|\, x_1\right) = 0$$

Literature

Gert de Cooman and Filip Hermans.
Imprecise probability trees: Bridging two theories of imprecise probability.
Artificial Intelligence, 2008, vol. 172, pp. 1400–1427. (arXiv:0801.1196v1).

Gert de Cooman, Filip Hermans, and Erik Quaeghebeur.
Imprecise Markov chains and their limit behaviour.
Submitted for publication (arXiv:0801.0980).

Glenn Shafer and Vladimir Vovk.
Probability and Finance: It's Only a Game!
Wiley, New York, 2001.

Peter Walley.
Statistical Reasoning with Imprecise Probabilities.
Chapman and Hall, London, 1991.