Credal Networks

Specification, Algorithms, Complexity

Fabio G. Cozman
Denis D. Mauá

Table of Contents

Table of Contents

A credal set

- United States National Intelligence Estimate report number 29-51:
  an attack on Yugoslavia in 1951 should be considered a serious possibility.
- Authors provided odds that ranged from 20/80 to 80/20 in favor of an attack. That is,
$$0.2 \leq \mathbb{P}(\text{attack}) \leq 0.8\,.$$

Easy warm-up

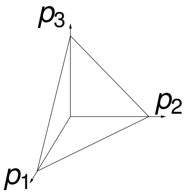
- Possibility space Ω with states ω: everything is FINITE in this tutorial!
- Events are subsets of Ω.
- Random variables are functions from Ω to real numbers.
- Expectations $\mathbb{E}_{\mathbb{P}}[X]$, probabilities $\mathbb{P}(A)$.
- Conditional expectations $\mathbb{E}_{\mathbb{P}}[X|B]$, probabilities $\mathbb{P}(A|B)$.

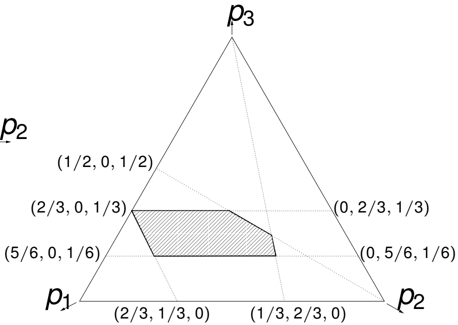
Credal sets

- A credal set is a set of probability measures on a common algebra.
- A credal set is usually defined by a set of *assessments*.

Example:

- $\Omega = \{\omega_1, \omega_2, \omega_3\}$.
- $\mathbb{P}(\omega_i) = p_i$.
- $p_1 \geq p_3$, $2p_1 \geq p_2$, $p_1 \leq 2/3$ and $p_3 \in [1/6, 1/3]$.
- Take points $\mathbb{P} = (p_1, p_2, p_3)$.





Properties of credal sets

- Credal set with distributions for X is denoted $\mathbb{K}(X)$.

Properties of credal sets

- Credal set with distributions for X is denoted $\mathbb{K}(X)$.
- Given credal set $\mathbb{K}(X)$:
  - $\underline{\mathbb{E}}[X] = \inf_{\mathbb{P} \in \mathbb{K}(X)} \mathbb{E}_{\mathbb{P}}[X]$.
  - $\overline{\mathbb{E}}[X] = \sup_{\mathbb{P} \in \mathbb{K}(X)} \mathbb{E}_{\mathbb{P}}[X]$.
  - $\underline{\mathbb{P}}(A) = \inf_{\mathbb{P} \in \mathbb{K}(X)} \mathbb{P}(A)$.
  - $\overline{\mathbb{P}}(A) = \sup_{\mathbb{P} \in \mathbb{K}(X)} \mathbb{P}(A)$.

Properties of credal sets

- Credal set with distributions for X is denoted $\mathbb{K}(X)$.
- Given credal set $\mathbb{K}(X)$:
  - $\underline{\mathbb{E}}[X] = \inf_{\mathbb{P} \in \mathbb{K}(X)} \mathbb{E}_{\mathbb{P}}[X]$.
  - $\overline{\mathbb{E}}[X] = \sup_{\mathbb{P} \in \mathbb{K}(X)} \mathbb{E}_{\mathbb{P}}[X]$.
  - $\underline{\mathbb{P}}(A) = \inf_{\mathbb{P} \in \mathbb{K}(X)} \mathbb{P}(A)$.
  - $\overline{\mathbb{P}}(A) = \sup_{\mathbb{P} \in \mathbb{K}(X)} \mathbb{P}(A)$.
- For closed convex credal sets, lower and upper expectations are attained at vertices.

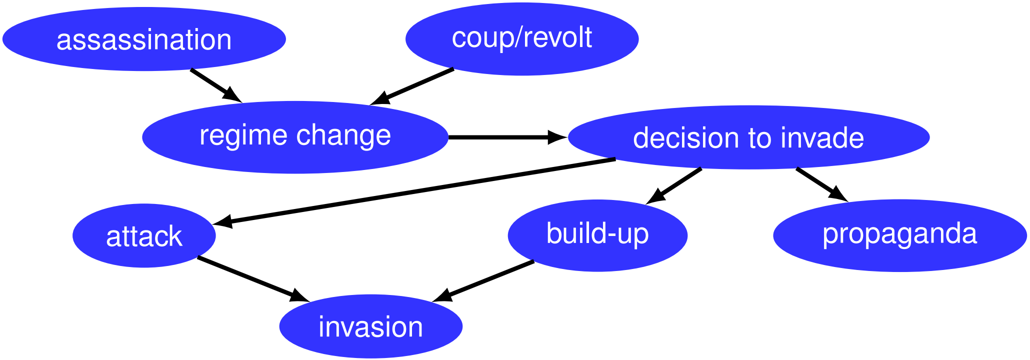
Properties of credal sets

- Credal set with distributions for X is denoted $\mathbb{K}(X)$.
- Given credal set $\mathbb{K}(X)$:
  - $\underline{\mathbb{E}}[X] = \inf_{\mathbb{P} \in \mathbb{K}(X)} \mathbb{E}_{\mathbb{P}}[X]$.
  - $\overline{\mathbb{E}}[X] = \sup_{\mathbb{P} \in \mathbb{K}(X)} \mathbb{E}_{\mathbb{P}}[X]$.
  - $\underline{\mathbb{P}}(A) = \inf_{\mathbb{P} \in \mathbb{K}(X)} \mathbb{P}(A)$.
  - $\overline{\mathbb{P}}(A) = \sup_{\mathbb{P} \in \mathbb{K}(X)} \mathbb{P}(A)$.
- For closed convex credal sets, lower and upper expectations are attained at vertices.
- A closed convex credal set is completely characterized by the associated lower expectations.

Independence

- Complete independence of X and Y: elementwise stochastic independence.
- X and Y are *strongly independent* with respect to a credal set $\mathbb{K}(X, Y)$ if the latter can be written as the convex hull of a credal set for which X and Y are completely independent.
- Y is *epistemically irrelevant* to X if $\underline{\mathbb{E}}[f(X)|Y = y] = \underline{\mathbb{E}}[f(X)]$ for every function f, every value y.

Report 29-51, continued

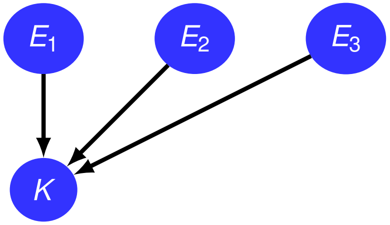


Another example: The three prisoners

- Three prisoners await execution, in separate cells.
- They may be pardoned by the king, independently, with $\mathbb{P}(E_i) = 1/2$.

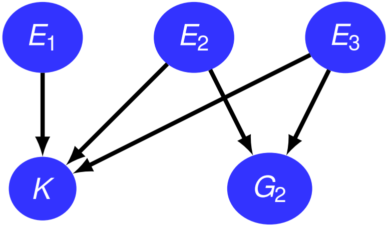
The three prisoners and the king

- Three prisoners await execution, in separate cells.
- They may be pardoned by the king, independently, with $\mathbb{P}(E_i) = 1/2$.
- Possible event: only one prisoner is executed.



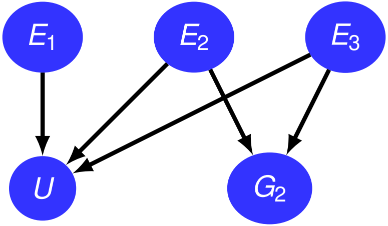
The three prisoners and the guard

- Three prisoners await execution, in separate cells.
- They may be pardoned, independently, with $\mathbb{P}(E_i) = 1/2$.
- Guard knows that only one prisoner is to be executed.
- Guard may say to Prisoner 1: “Prisoner 2 will be released”.
- What is $\mathbb{P}(E_1|G_2 \text{ and } K)$?



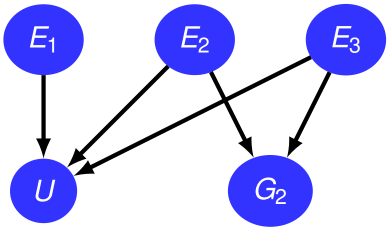
The three prisoners, the king, the guard

- Note: G_2 is: $(\neg E_2 \wedge E_3) \vee (\neg E_2 \wedge \neg E_3 \wedge D)$.
- If $\mathbb{P}(D) = 1/2$, then $\mathbb{P}(E_i | G_2 \text{ and } K) = \mathbb{P}(E_i | K) = 1/3$.



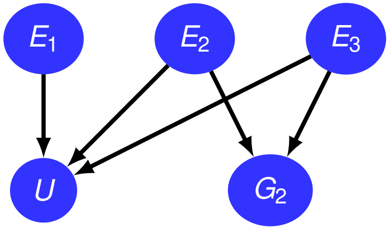
The three prisoners, the king, the guard

- If $\mathbb{P}(D) \in [0, 1]$, then $\mathbb{P}(E_i | G_2 \text{ and } K) \in [0, 1/2]$.



The three prisoners, the king, the guard

- If $\mathbb{P}(D) \in [0, 1]$, then $\mathbb{P}(E_i | G_2 \text{ and } K) \in [0, 1/2]$.
- And if, in addition, $\mathbb{P}(E_i) \in [19/40, 21/40]$, then $\mathbb{P}(E_i | G_2 \text{ and } K) \in [0, 441/802]$.



Credal networks until 2005: a quick summary

- Started around 1990: desire to increase representation power of Bayesian networks to handle uncertainty in Artificial Intelligence (AI).

Credal networks until 2005: a quick summary

- Started around 1990: desire to increase representation power of Bayesian networks to handle uncertainty in Artificial Intelligence (AI).
- Around 1993: message passing algorithms.
- Around 1997: 2U algorithm, epistemic extensions.

Credal networks until 2005: a quick summary

- Started around 1990: desire to increase representation power of Bayesian networks to handle uncertainty in Artificial Intelligence (AI).
- Around 1993: message passing algorithms.
- Around 1997: 2U algorithm, epistemic extensions.
- From 1999 to 2005: focus on credal sets rather than intervals and other formalisms; discussion of independence concepts.

A few tutorials

- A. Cano, S. Moral. Algorithms for imprecise probabilities. In: J. Kohlas, S. Moral, *Handbook of Defeasible and Uncertainty Management Systems*, pp. 369–420, 2000.
- F. G. Cozman. Graphical models for imprecise probabilities. *International Journal of Approximate Reasoning*, 39:167–184, 2005.

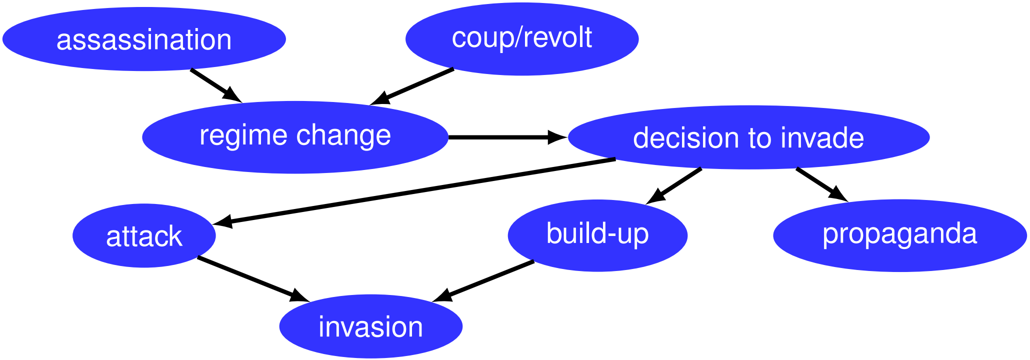
A few tutorials

- A. Cano, S. Moral. Algorithms for imprecise probabilities. In: J. Kohlas, S. Moral, *Handbook of Defeasible and Uncertainty Management Systems*, pp. 369–420, 2000.
- F. G. Cozman. Graphical models for imprecise probabilities. *International Journal of Approximate Reasoning*, 39:167–184, 2005.
- A. Piatti, A. Antonucci, M. Zaffalon. Building knowledge-based systems by credal networks: A tutorial. *Nova Science*, pp. 227–279, 2010.
- A. Antonucci, C. de Campos, M. Zaffalon. Probabilistic graphical models. In: T. Augustin, F. P. A. Coolen, G. de Cooman, M. C. M. Troffaes, *Introduction to Imprecise Probabilities*, pp. 207–229, 2014.

A few tutorials

- A. Cano, S. Moral. Algorithms for imprecise probabilities. In: J. Kohlas, S. Moral, *Handbook of Defeasible and Uncertainty Management Systems*, pp. 369–420, 2000.
- F. G. Cozman. Graphical models for imprecise probabilities. *International Journal of Approximate Reasoning*, 39:167–184, 2005.
- A. Piatti, A. Antonucci, M. Zaffalon. Building knowledge-based systems by credal networks: A tutorial. *Nova Science*, pp. 227–279, 2010.
- A. Antonucci, C. de Campos, M. Zaffalon. Probabilistic graphical models. In: T. Augustin, F. P. A. Coolen, G. de Cooman, M. C. M. Troffaes, *Introduction to Imprecise Probabilities*, pp. 207–229, 2014.
- J. de Bock. Credal networks under epistemic independence. *International Journal of Approximate Reasoning*, 85:107–138, 2017.
- Denis D. Mauá, Fabio G. Cozman. Thirty years of credal networks: Specification, algorithms and complexity. *International Journal of Approximate Reasoning*, 126:133–157, 2020.

Graphs: nodes, edges, cycles



Bayesian networks and the Markov condition

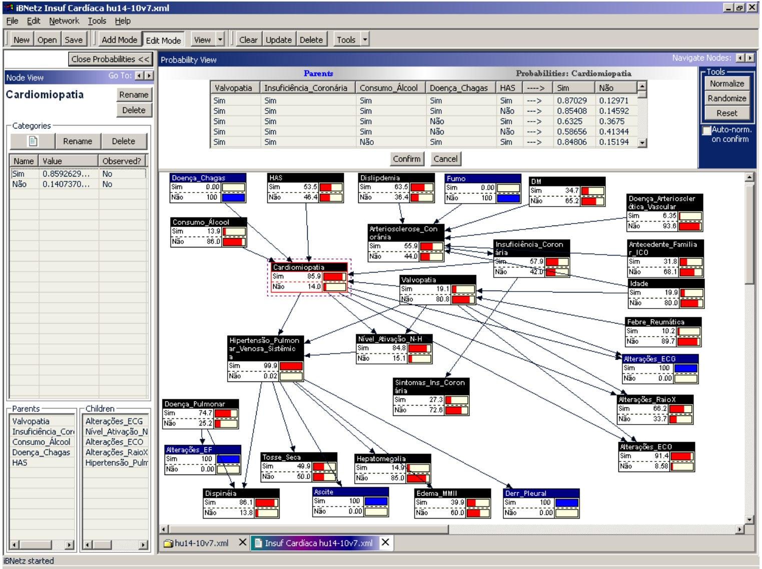
- A Bayesian network consists of a pair:
  - a directed acyclic graph where each node is a random variable,
  - a joint distribution over those random variables,

  such that they satisfy the Markov condition:
  - Every variable X is independent from its nondescendants nonparents given its parents.

- Consequence:

$$\mathbb{P}(X_1 = x_1, \dots, X_n = x_n) = \prod_{i=1}^{n} \mathbb{P}\left(X_i = x_i | \mathrm{Pa}_{X_i} = \pi_i\right).$$

Example:



Credal networks and the Markov condition(S)

- A credal network consists of a pair:
  - a directed acyclic graph where each node is a random variable,
  - a credal set collecting joint distributions over those random variables,

  such that they satisfy the Markov condition:
  - Every variable X is independent from its nondescendants nonparents given its parents.

- Note: there are several definitions of “independence” for credal sets.

Credal Networks:

Structure: the acyclic directed graph whose nodes are random variables.

Quantification: the credal set of joint probability distributions over the random variables.

Example: separately specified credal network

$$\mathbb{P}(X=1) \in [1/10, 1/3], \qquad \mathbb{P}(Y=1) = 4/5,$$
$$\mathbb{P}(Z=1|Y=0) \in [2/5, 3/5], \quad \mathbb{P}(Z=1|Y=1) \in [7/10, 9/10].$$

Digression...

- One might have non-separately specified credal networks.
- Not very common.
- Example: qualitative networks with positive additive synergy.

$$\begin{aligned}\mathbb{P}(X = 1 | Y = 1, Z = 1) + \mathbb{P}(X = 1 | Y = 0, Z = 0) \geq \\ \mathbb{P}(X = 1 | Y = 0, Z = 1) + \mathbb{P}(X = 1 | Y = 1, Z = 0)\,,\end{aligned}$$

Complete Extension

- Suppose we have a directed acyclic graph with associated random variables,
- and separately specified local credal sets,
- and we take the credal set $\mathbb{K}_C(X_1, \ldots, X_n)$:

$$\left\{ \mathbb{P} : \quad \mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) = \prod_{i=1}^{n} \mathbb{P}\big(X_i = x_i | \mathrm{Pa}_{X_i} = \pi_i\big) \right. ,$$

$$\left. \text{with } \mathbb{P}\big(X_i | \mathrm{Pa}_{X_i} = \pi_i\big) \in \mathbb{K}(X_i | \mathrm{Pa}_{X_i} = \pi_i) \right\} .$$

Largest joint credal set for Markov condition with respect to elementwise independence!

Convexifying

- Suppose we have a directed acyclic graph with associated random variables,
- and separately specified local credal sets,
- and we take the convex hull of $\mathbb{K}_C(X_1, \ldots, X_n)$.

Strong Extension

Theorem

The convex hull of the complete extension of a credal network is the largest credal set that satisfies the Markov condition with respect to strong independence. This is also true when we replace each local credal set by the set of its extreme points:

$$\textit{convex-hull-of}\ \Bigg\{\mathbb{P} : \quad \mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) = \prod_{i=1}^{n} \mathbb{P}(X_i = x_i | \mathrm{Pa}_{X_i} = \pi_i)\ ,$$

$$\textit{with}\ \mathbb{P}(X_i | \mathrm{Pa}_{X_i} = \pi_i) \in\ \textit{extreme-points-of}\ \mathbb{K}(X_i | \mathrm{Pa}_{X_i} = \pi_i)\Bigg\}\ .$$

Convexifying: the Strong Extension

- Suppose we have a directed acyclic graph with associated random variables,
- and separately specified local credal sets,
- and we take the convex hull of $\mathbb{K}_C(X_1, \ldots, X_n)$.

Largest joint credal set for Markov condition with respect to strong independence!

Nice results

Theorem

Any combination of extreme points from the local credal sets is an extreme point of the strong extension.

Nice results

Theorem

Any combination of extreme points from the local credal sets is an extreme point of the strong extension.

Theorem

If variables X and Y are d-separated by variables Z, then X and Y are strongly independent given Z in the strong extension of the credal network.

The number of extreme points of a strong extension

$$\mathbb{P}(X=1) \in [1/10, 1/3], \qquad \mathbb{P}(Y=1) = 4/5,$$
$$\mathbb{P}(Z=1|Y=0) \in [2/5, 3/5], \quad \mathbb{P}(Z=1|Y=1) \in [7/10, 9/10].$$

Epistemic Extension

- Suppose we have a directed acyclic graph with associated random variables,
- and separately specified local credal sets,
- and we take the largest set of joint distributions such that, for each variable X, the nondescendants nonparents of X are epistemically irrelevant to X given the parents of X.

For a LOT MORE:

- J. de Bock. Credal networks under epistemic independence. *International Journal of Approximate Reasoning*, 85:107–138, 2017.

Marginal inference:

- compute bounds for the marginal probability of a value of a *query* variable conditional on some *evidence* on some other variables.

Marginal inference:

- compute bounds for the marginal probability of a value of a *query* variable conditional on some *evidence* on some other variables.

- Given a credal network over variables $\mathbf{X} = \{X_1, \ldots, X_n\}$, an event of interest $Z = z$ and some evidence $\mathbf{Y} = \mathbf{y}$, we are interested in obtaining:

$$\underline{\mathbb{P}}(Z = z | \mathbf{Y} = \mathbf{y}) \quad \text{and} \quad \overline{\mathbb{P}}(Z = z | \mathbf{Y} = \mathbf{y}) \, .$$

Basic result

Theorem

The latter lower/upper probabilities are obtained, respectively, by

$$\text{inf} \,/\, \text{sup} \quad \frac{\sum_{\mathbf{x}'} \prod_{i=1}^{n} \mathbb{P}\left(X_i = x_i | \mathrm{Pa}_{X_i} = \pi\right)}{\sum_{z} \sum_{\mathbf{x}'} \prod_{i=1}^{n} \mathbb{P}\left(X_i = x_i | \mathrm{Pa}_{X_i} = \pi\right)} \,,$$

Basic result

Theorem

The latter lower/upper probabilities are obtained, respectively, by

$$\inf / \sup \quad \frac{\sum_{\mathbf{x}'} \prod_{i=1}^{n} \mathbb{P}(X_i = x_i | \mathrm{Pa}_{X_i} = \pi)}{\sum_{z} \sum_{\mathbf{x}'} \prod_{i=1}^{n} \mathbb{P}(X_i = x_i | \mathrm{Pa}_{X_i} = \pi)},$$

where the sum in the numerator and the inner sum in the denominator are over the values of the set of marginalized variables $\mathbf{X}' = \mathbf{X} \setminus (\{Z\} \cup \mathbf{Y})$, the outer sum in the denominator is over the values of the query variable Z, and the optimization is over the distributions from the complete extension such that $\mathbb{P}(\mathbf{Y} = \mathbf{y}) > 0$.

Key result

Theorem

Suppose a credal network is separately specified with closed convex local credal sets. Then the multilinear optimization for lower/upper probability is attained at some extreme point $\mathbb{P}$ *of the strong extension such that* $\mathbb{P}(\mathbf{Y} = \mathbf{y}) > 0$*; hence it is attained at some combination of extreme points of the local credal sets* $\mathbb{K}(X|\mathrm{Pa}_X = \pi)$*.*

Computational complexity

- Marginal inference with separately specified network is NP^{PP}-complete.
  - ...and is NP-complete if network has a tree-like undirected structure.

- There are many other results, some of them indicating tractable cases.

Table of Contents

Marginal inference

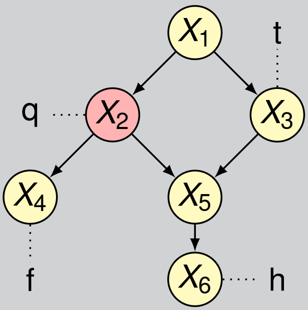
Example

Compute

$$\inf / \sup \ \mathbb{P}(X_2 = q | X_3 = t, X_4 = f, X_6 = h) \ ,$$

where the optimization is over probability measures $\mathbb{P}$ satisfying:

- Regular condition.: $\mathbb{P}(X_3 = t, X_4 = f, X_6 = h) > 0$
- Markov property: $\mathbb{P}(X_i | \mathrm{Nd}_i) = \mathbb{P}(X_i | \mathrm{Pa}_i)$
- Assessments: $\mathbb{P}(X_i | \mathrm{Pa}_i = \pi) \in \mathbb{K}(X_i | \pi)$



Marginal inference

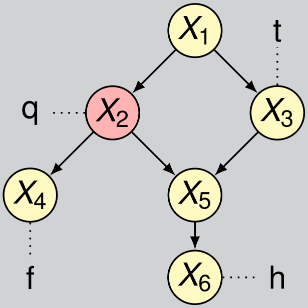
Example

Compute

$$\inf / \sup \mathbb{P}(X_2 = q | X_3 = t, X_4 = f, X_6 = h) \ ,$$

where the optimization is over probability measures $\mathbb{P}$ satisfying:

- Regular condition.: $\mathbb{P}(X_3 = t, X_4 = f, X_6 = h) > 0$
- Markov property: $\mathbb{P}(X_i | \mathrm{Nd}_i) = \mathbb{P}(X_i | \mathrm{Pa}_i)$
- Assessments: $\mathbb{P}(X_i | \mathrm{Pa}_i = \pi) \in \mathbb{K}(X_i | \pi)$



Marginal inference

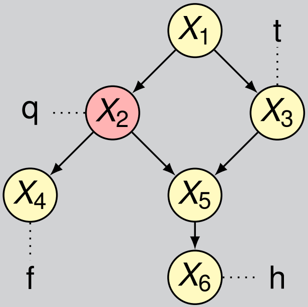
Example

Compute

$$\inf / \sup \, \mathbb{P}(X_2 = q | X_3 = t, X_4 = f, X_6 = h) \, ,$$

where the optimization is over probability measures $\mathbb{P}$ satisfying:

- Regular condition.: $\mathbb{P}(X_3 = t, X_4 = f, X_6 = h) > 0$
- Factorization: $\mathbb{P}(\{X_i\}) = \prod_i \mathbb{P}(X_i | \mathrm{Pa}_i)$
- Assessments: $\mathbb{P}(X_i | \mathrm{Pa}_i = \pi) \in \mathbb{K}(X_i | \pi)$



Marginal inference

- Mathematical programming
  - Multilinear programming
  - Linear programming relaxation
  - Integer programming
- Combinatorial optimization
  - Variable elimination
- Message passing
  - 2U, L2U, GL2U
  - iHMMs
  - TAN

Multilinear programming

[de Campos & Cozman, 2004]

Given: Credal net $(G, \mathbb{K})$, query $X_q = x_q$, evidence $\{X_e = x_e\}$

Optimize: $p(x_q|\{x_e\})$

Subject to: $p(x_q|\{x_e\})p(\{x_e\}) = \sum_{\{x_i\} \sim \{x_q, x_e\}} p(x_1, \ldots, x_n)$

[Markov Property] $p(x_1, \ldots, x_n) = \prod_i p(x_i|\pi_i)$

[Regular cond.] $p(\{x_e\}) > 0$

[Assessments] $A(X_i|\pi_i)p(X_i|\pi_i) \leq b(X_i|\pi_i) \quad [\forall i, \pi_i \sim \mathrm{Pa}_i]$

Multilinear programming

[de Campos & Cozman, 2004]

Given: Credal net $(G, \mathbb{K})$, query $X_q = x_q$, evidence $\{X_e = x_e\}$

Optimize: $p(x_q | \{x_e\})$

Subject to: $p(x_q | \{x_e\}) p(\{x_e\}) = \sum_{\{x_i\} \sim \{x_q, x_e\}} p(x_1, \ldots, x_n)$

[Markov Property] $p(x_1, \ldots, x_n) = \prod_i p(x_i | \pi_i)$

[Regular cond.] $p(\{x_e\}) > 0$

[Assessments] $A(X_i | \pi_i) p(X_i | \pi_i) \leq b(X_i | \pi_i) \quad [\forall i, \pi_i \sim \mathrm{Pa}_i]$

Charnes-Cooper Transformation

$$\begin{aligned} \text{optimize} \quad & \frac{\mathbf{c}'\mathbf{x}}{\mathbf{d}'\mathbf{x}} \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} \leq \mathbf{b} \\ & \mathbf{d}'\mathbf{x} > 0 \end{aligned}$$

$$\xrightarrow{\hspace{2em}} \quad \mathbf{y} = \frac{\mathbf{x}}{\mathbf{d}'\mathbf{x}}$$

$$\begin{aligned} \text{optimize} \quad & \mathbf{c}'\mathbf{y} \\ \text{subject to} \quad & \mathbf{A}\mathbf{y} \leq \mathbf{b}z \\ & \mathbf{d}'\mathbf{y} = 1 \\ & z \geq 0 \\ & \mathbf{y} = z\mathbf{x} \end{aligned}$$

Multilinear programming

[de Campos & Cozman, 2004]

$$
\begin{aligned}
\text{Optimize:} \quad & q(x_q) \\
\text{Subject to:} \quad & \sum_{x'_q} q(x'_q) = 1 \\
[\forall x'_q] \quad & q(x'_q) = \sum_{\{x_i\} \sim \{x'_q, x_e\}} z \prod_i p(x_i|\pi_i) \\
[\forall i, \pi_i] \quad & A(X_i|\pi_i) p(X_i|\pi_i) \leq b(X_i|\pi_i) \\
& z \geq 0
\end{aligned}
$$

Note: $z = 1/\mathbb{P}(\{x_e\})$

Example

$$\text{opt}\quad \frac{\sum_{x_1,x_4} p(X_2=t|x_1)p(X_3=f|x_1)p(X_5=h|X_2=t,X_3=f)p(x_4|X_2=t)p(X_6=m|X_5=h)}{\sum_{x_1,x_2,x_4} p(X_2=x_2|x_1)p(X_3=f|x_1)p(X_5=h|x_2,X_3=f)p(x_4|x_2)p(X_6=m|X_5=h)}$$

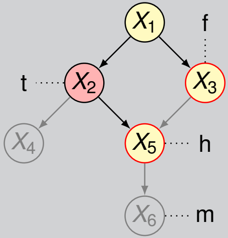
s.t.

$$\sum_{x_1,x_2} p(x_2|x_1)p(X_3=f|x_1)p(X_5=h|X_3=f)p(x_4|x_2)p(X_6=m|X_5=h) > 0$$

$$p(X_1) \in \mathbb{K}(X_1), \qquad [\forall x_1]\ p(X_2|x_1) \in \mathbb{K}(X_2|x_1)$$

$$[\forall x_1]\ p(X_3|x_1) \in \mathbb{K}(X_3|x_1), \qquad [\forall x_2]\ p(X_4|x_2) \in \mathbb{K}(X_4|x_2)$$

$$[\forall x_2, x_3]\ p(X_5|x_2,x_3) \in \mathbb{K}(X_5|x_2,x_3), \qquad [\forall x_5]\ p(X_6|x_5) \in \mathbb{K}(X_6|x_5)$$



Example

$$\text{opt} \quad q(t)$$

$$\text{s.t.} \quad \sum_{x_2} q(x_2) = 1$$

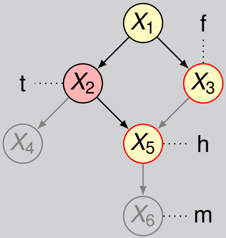
$$[\forall x_2] \quad q(x_2) = \sum_{x_1, x_2, x_4} z \cdot p(X_2 = t | x_1) p(X_3 = f | x_1) p(X_5 = h | X_2 = t, X_3 = f) p(x_4 | X_2 = t) p(X_6 = m | X_5 = h)$$

$$t \geq 0$$

$$p(X_1) \in \mathbb{K}(X_1), \qquad [\forall x_1]\ p(X_2 | x_1) \in \mathbb{K}(X_2 | x_1)$$

$$[\forall x_1]\ p(X_3 | x_1) \in \mathbb{K}(X_3 | x_1), \qquad [\forall x_2]\ p(X_4 | x_2) \in \mathbb{K}(X_4 | x_2)$$

$$[\forall x_2, x_3]\ p(X_5 | x_2, x_3) \in \mathbb{K}(X_5 | x_2, x_3), \qquad [\forall x_5]\ p(X_6 | x_5) \in \mathbb{K}(X_6 | x_5)$$



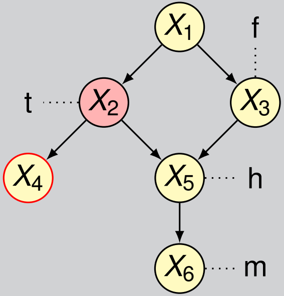
Multilinear programming: Improvements

- Requisite graph
  - Prune parts that are irrelevant to inference
- Symbolic variable elimination
  - Reduce size and degree of constraints

Requisite Graph

Preprocessing

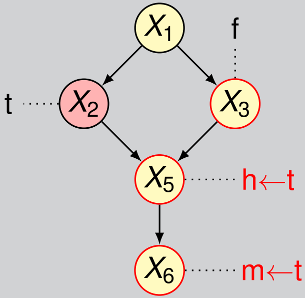
- Remove barren nodes
- Binarize evidence variables
- Drop arcs leaving observed variables
- Remove disconnected nodes from query



Requisite Graph

Preprocessing

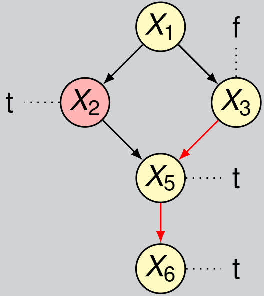
- Remove barren nodes
- Binarize evidence variables
- Drop arcs leaving observed variables
- Remove disconnected nodes from query



Requisite Graph

Preprocessing

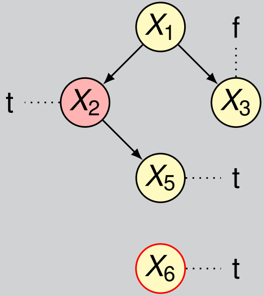
- Remove barren nodes
- Binarize evidence variables
- Drop arcs leaving observed variables
- Remove disconnected nodes from query



Requisite Graph

Preprocessing

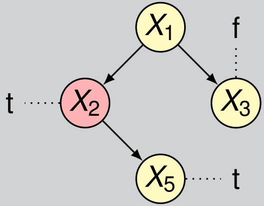
- Remove barren nodes
- Binarize evidence variables
- Drop arcs leaving observed variables
- Remove disconnected nodes from query



Requisite Graph

Preprocessing

- Remove barren nodes
- Binarize evidence variables
- Drop arcs leaving observed variables
- Remove disconnected nodes from query



Example (Requisite Graph)

Minimize: $q(t)$

s.t.: $\sum_{x_2} q(x_2) = 1$

$[\forall x_2]\ q(x_2) = \sum_{x_1} z \cdot p(x_1)p(x_2|x_1)p(e_3|x_1)p(e_5|x_2, e_3)$

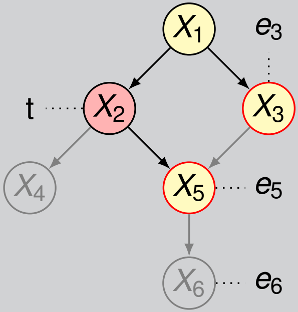
$p(X_1) \in \mathbb{K}(X_1)$

$[\forall x_1]\ p(X_2|x_1) \in \mathbb{K}(X_2|x_1)$

$[\forall x_1]\ p(e_3|x_1) \in [\underline{\mathbb{P}}(e_3|x_1), \overline{\mathbb{P}}(e_3|x_1)]$

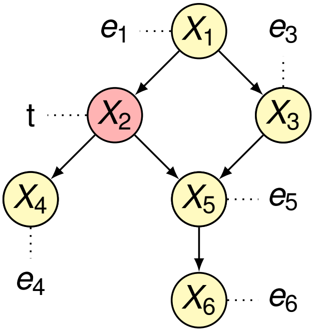
$[\forall x_2]\ p(e_5|x_2, e_3) \in [\underline{\mathbb{P}}(e_5|x_2, e_3), \overline{\mathbb{P}}(e_5|x_2, e_3)]$

$z \geq 0$



Exercise

Obtain requisite graph and write the multilinear program of the inference



Requisite graph

- Remove barren nodes
- Binarize evidence variables
- Drop arcs leaving observed variables
- Remove disconnected nodes from query

Multilinear program

Optimize: $q(x_q)$

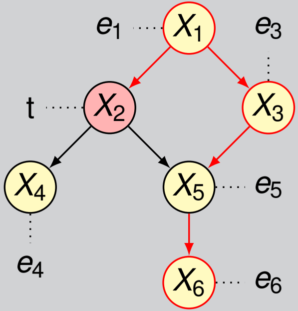
Subject to: $\sum_{x'_q} q(x'_q) = 1$

$[\forall x'_q] \quad q(x'_q) = \sum_{x_m} z \prod_i p(x_i|\pi_i)$

$[\forall i, \pi_i] \quad p(X_i|\pi_i) \in \mathbb{K}(X_i|\pi_i)$

$z \geq 0$

Exercise: Fully observed evidence



Exercise: Fully observed evidence

Minimize: $q(t)$

subject to: $\sum_{x_2} q(x_2) = 1$

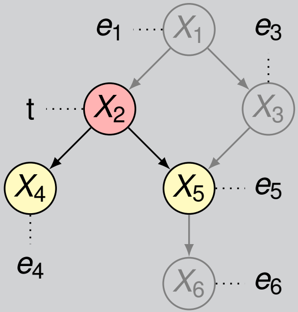
$$[\forall x_2]\ q(x_2) = z \cdot p(x_2|e_1)p(e_4|x_2)p(e_5|x_2, e_3)$$

$$p(X_2|e_1) \in \mathbb{K}(X_2|e_1)$$

$$[\forall x_2]\ p(e_4|x_2) \in [\underline{\mathbb{P}}(e_4|x_2)\,, \overline{\mathbb{P}}(e_4|x_2)]$$

$$[\forall x_2]\ p(e_5|x_2, e_3) \in [\underline{\mathbb{P}}(e_5|x_2, e_3)\,, \overline{\mathbb{P}}(e_5|x_2, e_3)]$$

$$z \geq 0$$



Exercise: Fully observed evidence

Minimize: $q(t)$

subject to: $q(t) = 1 - \sum_{x_2 \neq t} q(x_2)$

$$q(t) = z \cdot p(t|e_1)p(e_4|t)p(e_5|t, e_3)$$

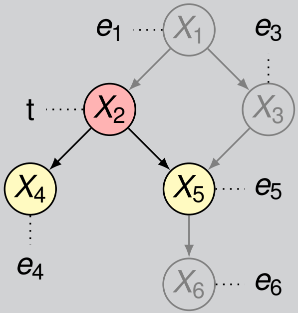
$$[\forall x_2 \neq t]\ q(x_2) = z \cdot p(x_2|e_1)p(e_4|x_2)p(e_5|x_2, e_3)$$

$$p(X_2|e_1) \in \mathbb{K}(X_2|e_1)$$

$$[\forall x_2]\ p(e_4|x_2) \in [\underline{\mathbb{P}}(e_4|x_2), \overline{\mathbb{P}}(e_4|x_2)]$$

$$[\forall x_2]\ p(e_5|x_2, e_3) \in [\underline{\mathbb{P}}(e_5|x_2, e_3), \overline{\mathbb{P}}(e_5|x_2, e_3)]$$

$$z \geq 0$$



Exercise: Fully observed evidence

Minimize: $q(t)$

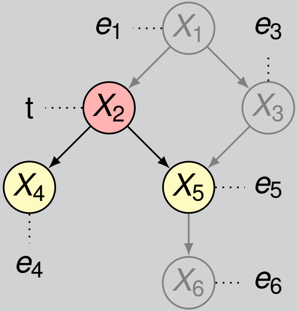
subject to: $q(t) = 1 - \sum_{x_2 \neq t} q(x_2)$

$q(t) = z \cdot p(t|e_1)\mathbb{P}(e_4|t)\,\mathbb{P}(e_5|t, e_3)$

$[\forall x_2 \neq t]\ q(x_2) = z \cdot p(x_2|e_1)\overline{\mathbb{P}}(e_4|x_2)\,\overline{\mathbb{P}}(e_5|x_2, e_3)$

$p(X_2|e_1) \in \mathbb{K}(X_2|e_1)$

$z \geq 0$



Exercise: Fully observed evidence

Minimize: $q(t)$

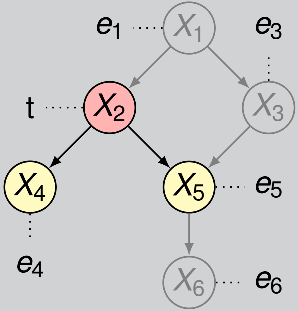
subject to: $q(t) = 1 - \sum_{x_2 \neq t} q(x_2)$

$$q(t) = z \cdot p(t|e_1)\underline{\mathbb{P}}(e_4|t)\,\underline{\mathbb{P}}(e_5|t, e_3)$$

$$[\forall x_2 \neq t]\; q(x_2) = z \cdot p(x_2|e_1)\overline{\mathbb{P}}(e_4|x_2)\,\overline{\mathbb{P}}(e_5|x_2, e_3)$$

$$p(X_2|e_1) \in \mathbb{K}(X_2|e_1)$$

$$z \geq 0$$



Exercise: Fully observed evidence

Minimize: $q(t)$

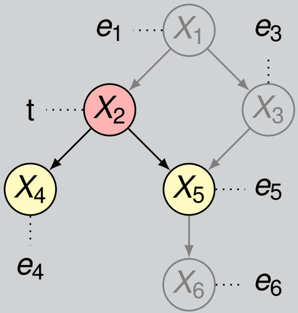
subject to: $q(t) = 1 - \sum_{x_2 \neq t} q(x_2)$

$q(t) = p'(t|e_1)\underline{\mathbb{P}}(e_4|t)\,\underline{\mathbb{P}}(e_5|t, e_3)$

$[\forall x_2 \neq t]\; q(x_2) = p'(x_2|e_1)\overline{\mathbb{P}}(e_4|x_2)\,\overline{\mathbb{P}}(e_5|x_2, e_3)$

$p'(X_2|e_1) \in z\mathbb{K}(X_2|e_1)$

$z \geq 0$



Exercise: Fully observed evidence

Minimize: $q(t)$

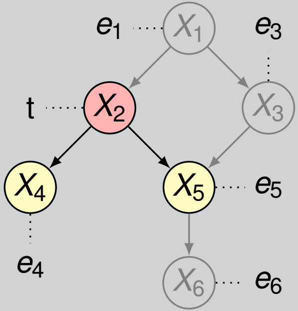
subject to: $q(t) = 1 - \sum_{x_2 \neq t} q(x_2)$

$$q(t) = p'(t|e_1)\mathbb{P}(e_4|t)\,\mathbb{P}(e_5|t, e_3)$$

$$[\forall x_2 \neq t]\; q(x_2) = p'(x_2|e_1)\overline{\mathbb{P}}(e_4|x_2)\,\overline{\mathbb{P}}(e_5|x_2, e_3)$$

$$p'(X_2|e_1) \in z\mathbb{K}(X_2|e_1)$$

$$z \geq 0$$



This is a linear program!

Symbolic Variable Elimination

$$\text{Minimize: } q(t)$$

$$\text{subject to: } \sum_y q(y) = 1$$

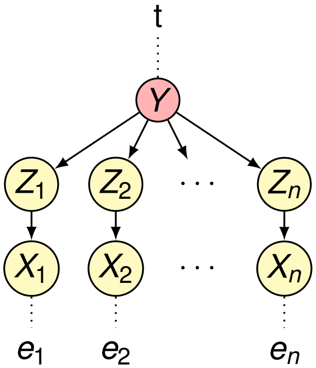
$$[\forall y]\ q(y) = \underbrace{\sum_{z_1,\ldots,z_n} p'(y) \prod_i p(e_i|z_i)p(z_i|y)}_{O(2^n)\text{ terms}}$$

$$p'(Y) \in z\mathbb{K}(Y)$$

$$[\forall i, y]\ p(Z_i|y) \in \mathbb{K}(Z_i|y)$$

$$[\forall i, z_i]\ p(e_i|z_i) \in [\underline{\mathbb{P}}(e_i|z_i), \overline{\mathbb{P}}(e_i|z_i)]$$

$$z \geq 0$$



Symbolic Variable Elimination

- Use dynamic programming to rewrite joint distribution expression as set of constraints of smaller size

- Main idea: exploit distributivity of multiplication over addition

$$\sum_y [f(x)g(x,y) + f(x)h(x,y)] = f(x) \sum_y [g(x,y) + h(x,y)]$$
$$= f(x)s(x),$$

$$s(x) = \sum_y [g(x,y) + h(x,y)]$$

Example

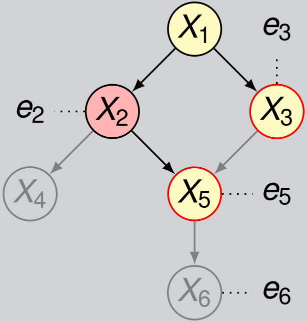
Optimize: $q(t)$

subject to: $\sum_{x_2} q(x_2) = 1$

$$[\forall x_2]\ q(x_2) = \sum_{x_1} z \cdot p(x_2|x_1)p(e_3|x_1)p(e_5|x_2, e_3)$$

$\vdots$

$z \geq 0$



Example

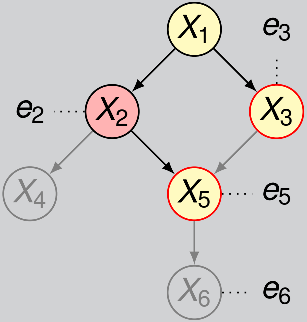
Optimize: $q(t)$

subject to: $\sum_{x_2} q(x_2) = 1$

$[\forall x_2]\ q(x_2) = \sum_{x_1} z \cdot p(x_2|x_1)p(e_3|x_1)p(e_5|x_2, e_3)$

$\vdots$

$z \geq 0$



Example

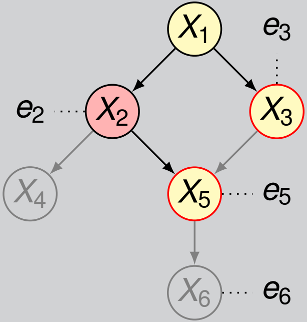
Optimize: $q(t)$

subject to: $\sum_{x_2} q(x_2) = 1$

$$[\forall x_2]\ q(x_2) = z \cdot p(e_5 | x_2, e_3) \sum_{x_1} p(x_2 | x_1) p(e_3 | x_1)$$

$$\vdots$$

$$z \geq 0$$



Example

Optimize: $q(t)$

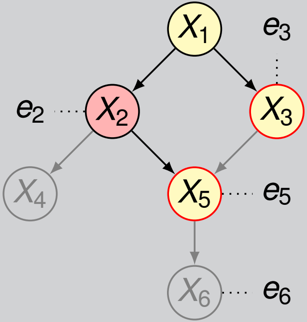
subject to: $\sum_{x_2} q(x_2) = 1$

$[\forall x_2]\ q(x_2) = z \cdot p(e_5 | x_2, e_3) \cdot f_1(x_2)$

$f_1(x_2) = \sum_{x_1} p(x_2|x_1) p(e_3|x_1)$

$\vdots$

$z \geq 0$



Example

Optimize: $q(t)$

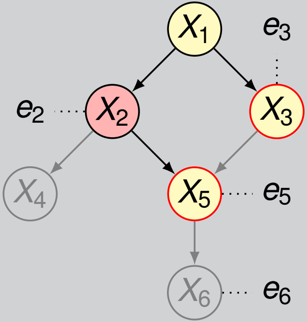
subject to: $\sum_{x_2} q(x_2) = 1$

$$[\forall x_2]\ q(x_2) = z \cdot p(e_5|x_2, e_3) \cdot f_1(x_2)$$

$$f_1(x_2) = \sum_{x_1} p(x_2|x_1)p(e_3|x_1)$$

$$\vdots$$

$$z \geq 0$$



Example

Optimize: $q(t)$

subject to: $\sum_{x_2} q(x_2) = 1$

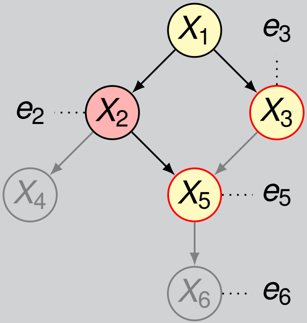
$[\forall x_2]\ q(x_2) = z \cdot f_2(x_2)$

$f_1(x_2) = \sum_{x_1} p(x_2|x_1)p(e_3|x_1)$

$f_2(x_2) = p(e_5|x_2, e_3) \cdot f_1(x_2)$

$\vdots$

$z \geq 0$



Example (Tree-structured Credal Network)

Optimize: $q(t)$

subject to: $\sum_{y} q(y) = 1$

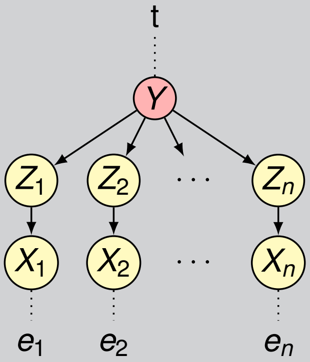
$[\forall y]\ q(y) = p'(y) \sum_{z_1,\ldots,z_n} \prod_{i} p(e_i|z_i) p(z_i|y)$

$p'(Y) \in z\mathbb{K}(Y)$

$[\forall i, y]\ p(Z_i|y) \in \mathbb{K}(Z_i|y)$

$[\forall i, z_i]\ p(e_i|z_i) \in [\underline{\mathbb{P}}(e_i|z_i), \overline{\mathbb{P}}(e_i|z_i)]$

$z \geq 0$



Example (Tree-structured Credal Network)

Optimize: $q(t)$

subject to: $\sum_y q(y) = 1$

$[\forall y]\ q(y) = p'(y) \sum_{z_2,\ldots,z_n} \prod_{i>1} p(e_i|z_i)p(z_i|y)f_1(y)$

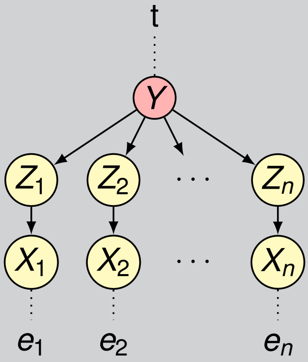
$f_1(y) = \sum_{z_1} p(e_n|z_n)p(z_n|y)$

$p'(Y) \in z\mathbb{K}(Y)$

$[\forall i, y]\ p(Z_i|y) \in \mathbb{K}(Z_i|y)$

$[\forall i, z_i]\ p(e_i|z_i) \in [\underline{\mathbb{P}}(e_i|z_i), \overline{\mathbb{P}}(e_i|z_i)]$

$z \geq 0$



Example (Tree-structured Credal Network)

Optimize: $q(t)$

subject to: $\sum_y q(y) = 1$

$[\forall y]\ q(y) = p'(y) f_n(y)$

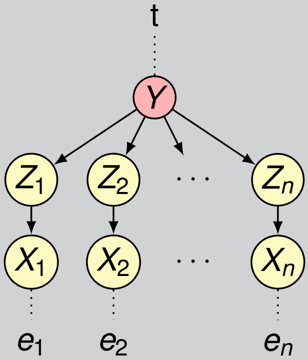
$[\forall i, y]\ f_i(y) = \sum_{z_i} f_{i-1}(y) p(e_i|z_i) p(z_i|y)$

$p'(Y) \in z\mathbb{K}(Y)$

$[\forall i, y]\ p(Z_i|y) \in \mathbb{K}(Z_i|y)$

$[\forall i, z_i]\ p(e_i|z_i) \in [\underline{\mathbb{P}}(e_i|z_i), \overline{\mathbb{P}}(e_i|z_i)]$

$z \geq 0$



Multilinear programming

Symbolic Variable Elimination

$$
\begin{aligned}
\text{Optimize:}\quad & q(t) \\
\text{Subject to:}\quad & \sum_{x_q} q(x_q) = 1 \\
[\forall x_q]\quad & q(x_q) = z \cdot f_j(x_q) \\
[\forall j]\quad & f_j(\mathbf{x}_j) = \sum_{\mathbf{x}_l} \prod_k f_k(\mathbf{x}_k) \\
[\forall i, \pi_i]\quad & f_i(x_1, \pi_i) = p(X_i|\pi_i) \in \mathbb{K}(X_i|\pi_i) \\
& z \geq 0
\end{aligned}
$$

Multilinear Programming

Caveats:

- Number of optimization variables is exponential in the network treewidth (measure treelikeness of the network)
  - Usually much smaller than number of requisite variables
- Not many solvers available; numerical problems often arise
- Scales to at most a few dozens of variables

Linear Programming Relaxation

[Antonucci et al., 2013; 2015]

1. Find feasible solution $\{p(x_i|\pi_i) \in \mathbb{K}(X_i|\mathrm{Pa}_{X_i}=\pi_i)\}$
2. Repeat until convergence:
   1. For $k = 1, \ldots, n$:

$$\text{Optimize: } q(t)$$

$$\text{Subject to: } \sum_{x_q} q(x_q) = 1$$

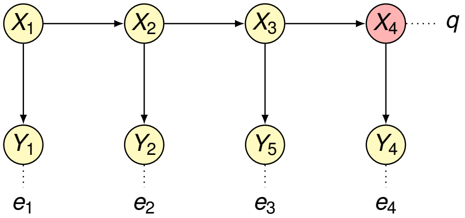
$$[\forall x_q] \quad Q(x_q) = z \cdot f_1(x_q), \quad z \geq 0 \qquad [\forall j] \quad f_j(\mathbf{x}_j) = \sum_{\mathbf{x}_l} \prod_{\ell} f_k(\mathbf{x}_\ell)$$

$$[\forall \pi_k] \quad p(X_k|\pi_k) \in \mathbb{K}(X_k|\pi_k) \qquad [\forall i \neq k] \quad p(X_i|\mathrm{Pa}_i) \text{ is fixed}$$

Linear Programming Relaxation

- Produces sequence of monotonically improving solutions, leading to inner approximation
- Can exploit efficient algorithms for Bayesian networks to pre-solve constraints
- Can approximate such constraints if treewidth is too large (loosing inner approximation guarantees)
- Best performing approximate algorithm
  - Greedy Lazy selection of variable to optimize
  - Stop criterion
  - Random restarts

Message Passing



Imprecise Hidden Markov Models

- Hidden variables: $X_1, \ldots, X_n$
- Manifest variables: $Y_1, \ldots, Y_n$

- Set of non-stationary (precise) hidden Markov models
- Time series (robust) prediction
- Time series (robust) classification

Message Passing in iHMMs

[Mauá, de Campos & Antonucci, 2015]

$$
\begin{aligned}
\text{Optimize: } & p(q, e_1, \ldots, e_n)/p(e_1, \ldots, e_n) \\
\text{Subject to: } & p(q, e_1, \ldots, e_n) = \sum_{\{x_i\}} \prod_i p(x_i|x_{i-1})p(e_i|x_i) \\
& p(e_1, \ldots, e_n) > 0 \\
[\forall i] \quad & p(X_i|x_{i-1}) \in \mathbb{K}(X_i|x_{i-1}), \quad p(e_i|x_i) \in \mathbb{K}(Y_i|x_i)
\end{aligned}
$$

Note: $p(x_1|x_0) = p(x_1)$

Message Passing in iHMMs

Assume that $\underline{\mathbb{P}}(e_1, \ldots, e_n) > 0 \implies \forall \mathbb{P}\, \exists! \gamma : \mathbb{P}(q, e_1, \ldots, e_n) = \gamma \mathbb{P}(e_1, \ldots, e_n)$

Message Passing in iHMMs

Assume that $\underline{\mathbb{P}}(e_1, \ldots, e_n) > 0 \implies \forall \mathbb{P}\, \exists! \gamma : \mathbb{P}(q, e_1, \ldots, e_n) = \gamma \mathbb{P}(e_1, \ldots, e_n)$

Generalized Bayes Rule

Repeat until convergence: Guess $\gamma \in (0, 1]$, and solve

Minimize: $p(q, e_1, \ldots, e_n) - \gamma p(e_1, \ldots, e_n)$

Subject to:

$$[\forall x_n] \quad p(x_n, e_1, \ldots, e_n) = \sum_{\{x_i\}_{i<n}} \prod_i p(x_i|x_{i-1}) p(e_i|x_i)$$

$$[\forall i, x_{i-1}] \quad p(X_i|x_{i-1}) \in \mathbb{K}(X_i|x_{i-1}) \qquad [\forall i, x_i] \quad p(e_i|x_i) \in [\underline{\mathbb{P}}(Y_i|x_i), \overline{\mathbb{P}}(Y_i, x_i)]$$

Message Passing in iHMMs

Assume that $\underline{\mathbb{P}}(e_1, \ldots, e_n) > 0 \implies \forall \mathbb{P}\, \exists! \gamma : \mathbb{P}(q, e_1, \ldots, e_n) = \gamma \mathbb{P}(e_1, \ldots, e_n)$

Generalized Bayes Rule

Repeat until convergence: Guess $\gamma \in (0, 1]$, and solve

$$\text{Minimize: } \underbrace{[1-\gamma]}_{f_{n+1}(q)} p(q, e_1, \ldots, e_n) + \sum_{x_n \neq q} \underbrace{[-\gamma]}_{f_{n+1}(x_n)} p(x_n, e_1, \ldots, e_n)$$

Subject to:

$$[\forall x_n] \quad p(x_n, e_1, \ldots, e_n) = \sum_{\{x_i\}_{i<n}} \prod_i p(x_i | x_{i-1}) p(e_i | x_i)$$

$$[\forall i, x_{i-1}] \quad p(X_i | x_{i-1}) \in \mathbb{K}(X_i | x_{i-1}) \qquad [\forall i, x_i] \quad p(e_i | x_i) \in [\underline{\mathbb{P}}(Y_i | x_i), \overline{\mathbb{P}}(Y_i, x_i)]$$

Message Passing in iHMMs

Assume that $\underline{\mathbb{P}}(e_1, \ldots, e_n) > 0 \implies \forall \mathbb{P}\, \exists! \gamma : \mathbb{P}(q, e_1, \ldots, e_n) = \gamma \mathbb{P}(e_1, \ldots, e_n)$

Generalized Bayes Rule

Repeat until convergence: Guess $\gamma \in (0, 1]$, and solve

Minimize: $\sum_{x_n} f_{n+1}(x_n) p(x_n, e_1, \ldots, e_n)$

Subject to:

$$[\forall x_n] \quad p(x_n, e_1, \ldots, e_n) = \sum_{\{x_i\}_{i<n}} \prod_i p(x_i | x_{i-1}) p(e_i | x_i)$$

$$[\forall i, x_{i-1}] \quad p(X_i | x_{i-1}) \in \mathbb{K}(X_i | x_{i-1}) \qquad [\forall i, x_i] \quad p(e_i | x_i) \in [\underline{\mathbb{P}}(Y_i | x_i), \overline{\mathbb{P}}(Y_i, x_i)]$$

Message Passing in iHMMs

Assume that $\underline{\mathbb{P}}(e_1, \ldots, e_n) > 0 \implies \forall \mathbb{P}\, \exists! \gamma : \mathbb{P}(q, e_1, \ldots, e_n) = \gamma \mathbb{P}(e_1, \ldots, e_n)$

Generalized Bayes Rule

Repeat until convergence: Guess $\gamma \in (0, 1]$, and solve

Minimize: $$\sum_{x_{n-1}} p(x_{n-1}, e_1, \ldots, e_{n-1}) \sum_{x_n} f_{n+1}(x_n) p(x_n|x_{n-1}) p(e_n|x_n)$$

Subject to:

$$[\forall x_{n-1}] \quad p(x_{n-1}, e_1, \ldots, e_{n-1}) = \sum_{\{x_i\}_{i<n-1}} \prod_{i<n} p(x_i|x_{i-1}) p(e_i|x_i)$$

$$[\forall i, x_{i-1}] \quad p(X_i|x_{i-1}) \in \mathbb{K}(X_i|x_{i-1}) \qquad [\forall i, x_i] \quad p(e_i|x_i) \in [\underline{\mathbb{P}}(Y_i|x_i), \overline{\mathbb{P}}(Y_i, x_i)]$$

Message Passing in iHMMs

Assume that $\underline{\mathbb{P}}(e_1, \ldots, e_n) > 0 \implies \forall \mathbb{P}\, \exists! \gamma : \mathbb{P}(q, e_1, \ldots, e_n) = \gamma \mathbb{P}(e_1, \ldots, e_n)$

Generalized Bayes Rule

Repeat until convergence: Guess $\gamma \in (0, 1]$, and solve

$$\text{Minimize: } \sum_{x_{n-1}} p(x_{n-1}, e_1, \ldots, e_{n-1}) f_n(x_{n-1})$$

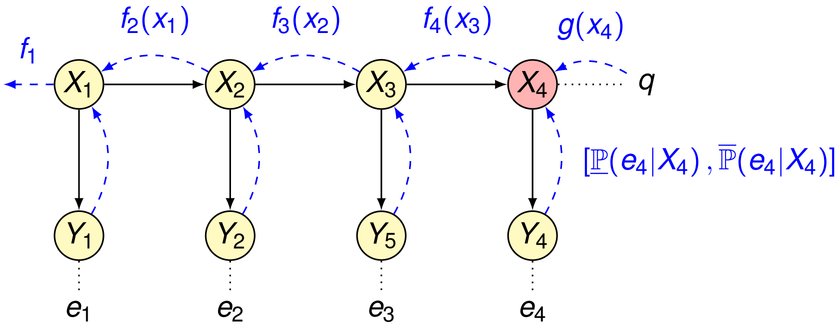
$$\text{Subject to: } f_n(x_{n-1}) = \min \sum_{x_n} f_{n+1}(x_n) p(x_n | x_{n-1}) p(e_n | x_n)$$

$$[\forall x_{n-1}] \quad p(x_{n-1}, e_1, \ldots, e_{n-1}) = \sum_{\{x_i\}_{i<n-1}} \prod_{i<n} p(x_i | x_{i-1}) p(e_i | x_i)$$

$$[\forall i, x_{i-1}] \quad p(X_i | x_{i-1}) \in \mathbb{K}(X_i | x_{i-1}) \qquad [\forall i, x_i] \quad p(e_i | x_i) \in [\underline{\mathbb{P}}(Y_i | x_i), \overline{\mathbb{P}}(Y_i, x_i)]$$

Message Passing in iHMMs

Assuming $\mathbb{P}(e_1, \ldots, e_n) > 0$



$$f_i(x_{i-1}) = \min \sum_{x_i} f_{i+1}(x_i) p(x_i|x_{i-1}) p(e_i|x_i)$$

Message Passing

- Very efficient
- Exact for:
  - Evidence likelihood $\underline{\mathbb{P}}(e_1, \ldots, e_n)$ in tree-shaped nets
  - Prediction in iHMMs (requisite graph)
  - Marginal inference in polytree-shaped nets with binary variables (2U)

Loopy approximate 2U for arbitrary topology; GL2U for non-binary variables

Software

- CREMA: https://github.com/IDSIA/crema
  - Exact inference by vertex propagation
  - Approximate inference by linear relaxation
- OpenCossan: https://github.com/cossan-working-group/OpenCossan
  - Matlab plugin for risk analysis
  - Approximate inference
- iHMM: https://github.com/denismaua/ihmm
  - Exact predictive filtering and evidence loglikelihood
  - Approximate robustness analysis of sequence prediction

Decision making

Precise probability

- Utility function $U(s, a)$
- Probability measure $\mathbb{P}(s)$
- MEU principle:

$$a^* = \arg\max_a \mathbb{E}_{\mathbb{P}}[U(S, a)]$$

Imprecise probability

- Inderteminate
  - Admissibility
  - Maximality
  - Interval dominance
- Determinate: Maximin

Admissibility

0/1 Utility $\implies$ classification accuracy
Target variable Z, evidence $\mathbf{Y} = \mathbf{y}$

Definition

z^* is admissible if

$$\exists \mathbb{P} \in \mathbb{K}(Z|\mathbf{y}) : \left[\mathbb{P}(Z = z^* | \mathbf{Y} = \mathbf{y}) - \max_z \mathbb{P}(Z = z | \mathbf{Y} = \mathbf{y}) \right] \geq 0 \,.$$

- Intuition: z^* is admissible if it is expected utility maximizer for some measure in the set
- Certificate $\mathbb{P} \in \mathbb{K}(Z|\mathbf{y})$ might not be an extreme point (hence doesn't factorize)

Maximality

0/1 Utility $\implies$ classification accuracy
Target variable Z, evidence $\mathbf{Y}=\mathbf{y}$

Definition

z^* is maximal if

$$\max_{z} \min_{\mathbb{P}\in\mathbb{K}(Z|\mathbf{y})} : \left[\mathbb{P}(Z=z^*|\mathbf{Y}=\mathbf{y}) - \max_{z} \mathbb{P}(Z=z|\mathbf{Y}=\mathbf{y})\right] \leq 0\,.$$

- Intuition: z^* is maximal if it is not strictly less probable than some other configuration under *all* measures
- Certificate $\mathbb{P} \in \mathbb{K}(Z|\mathbf{y})$ might not is an extreme point
  - Multilinear program with different objective

Interval dominance

0/1 Utility $\implies$ classification accuracy
Target variable Z, evidence $\mathbf{Y} = \mathbf{y}$

Definition

z^* is dominant if

$$\overline{\mathbb{P}}(Z = z^* | \mathbf{Y} = \mathbf{y}) \geq \max_{z} \underline{\mathbb{P}}(Z = z | \mathbf{Y} = \mathbf{y}) \leq 0 \, .$$

- Intuition: z^* is dominant if we cannot rule out the possibility that it might have greater probability than some other value under some measure
- Usually taken as heuristic
  - Reduces to solving marginal inference

Maximin

0/1 Utility $\implies$ classification accuracy
Target variable Z, evidence $\mathbf{Y}=\mathbf{y}$

Definition

z^* is maximin if

$$\underline{\mathbb{P}}(Z=z^*|\mathbf{Y}=\mathbf{y}) = \max_z \underline{\mathbb{P}}(Z=z|\mathbf{Y}=\mathbf{y}) \ .$$

- Intuition: z^* maximizes the worst-case scenario
- Very pessimistic/cautious
  - Reduces to solving marginal inference

Decision Making

We have that:

$$\text{admissible} \implies \text{maximal} \implies \text{dominant}$$

and

$$\text{maximin} \implies \text{maximal}$$

Elicitation

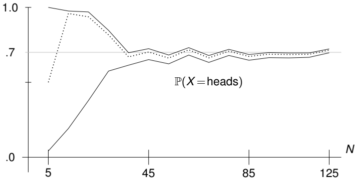
- Expert knowledge
- Imprecise Dirichlet Model

$$\mathbb{P}(X=x) \in \left[\frac{N[X=x]}{N+s}, \frac{N[X=x]+s}{N+s}\right]$$

- Nonparametric predictive inference [Augustin & Coolen 2004], ϵ-contamination, etc.
- Unreliable observations

$$X \to e\,, \quad \mathbb{P}(e|X=x) \in [l, u]$$

Elicitation



- Expert knowledge
- Imprecise Dirichlet Model

$$\mathbb{P}(X=x) \in \left[\frac{N[X=x]}{N+s}, \frac{N[X=x]+s}{N+s}\right]$$

- Nonparametric predictive inference [Augustin & Coolen 2004], ϵ-contamination, etc.
- Unreliable observations

$$X \to e\,, \quad \mathbb{P}(e|X=x) \in [l, u]$$

Elicitation

- Expert knowledge
- Imprecise Dirichlet Model

$$\mathbb{P}(X=x) \in \left[\frac{N[X=x]}{N+s}, \frac{N[X=x]+s}{N+s}\right]$$

- Nonparametric predictive inference [Augustin & Coolen 2004], ϵ-contamination, etc.
- Unreliable observations

$$X \to e\,, \quad \mathbb{P}(e|X=x) \in [l, u]$$

Elicitation

- Expert knowledge
- Imprecise Dirichlet Model

$$\mathbb{P}(X = x) \in \left[\frac{N[X = x]}{N + s}, \frac{N[X = x] + s}{N + s} \right]$$

- Nonparametric predictive inference [Augustin & Coolen 2004], ϵ-contamination, etc.
- Unreliable observations

$$X \to e\,, \quad \mathbb{P}(e|X = x) \in [l, u]$$

Applications

- Debris Flow hazard assessment [Antonucci et al., 2004]
- Analysis of place of death in cancer patients [Kern et al., 2020]
- Action Recognition [Mauá et al., 2015]

Debris flows hazard assessment



- Debris flows are very destructive natural hazards
- Still partially understood
- Human expertise is still fundamental in assessment
- Decision support system: aggregates physical theories, historical data, expert knowledge

Debris flows harzard assessment

Variables

- Movable debris thickness: proxy of debris flow hazard
- Geology: characteristics of bedrock
- Available debris thickness: propensity to generate sendiment
- Permeability: rate of liquid flow
- Hydrologic soil type
- Soil water capacity
- Rainfall
- . . .

Debris flows harzard assessment

Quantification

- Physical mechanisms, e.g.

$$\text{Theoretical thickness} = \text{water depth}\left(k\frac{\tan(f(\text{granulometry}))}{\tan(\text{slope})} - 1\right)^{-1}$$

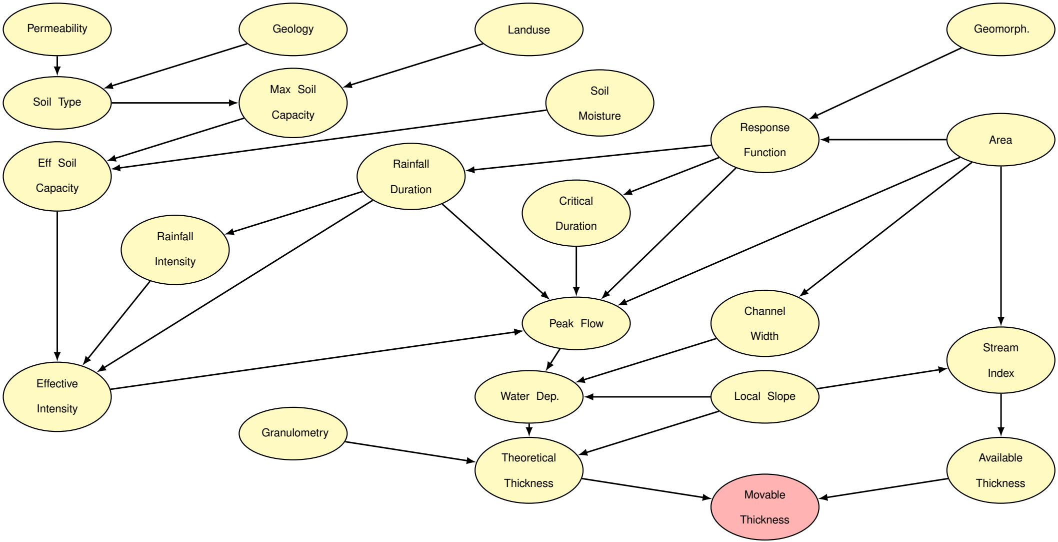
- IDM estimates from data, e.g.

$$\mathbb{P}(\text{Hydrologic soil type} = \text{low infiltration} \mid g_2, t_4) \in [0.0988, 0.1235]$$

- Expert knowledge, e.g.

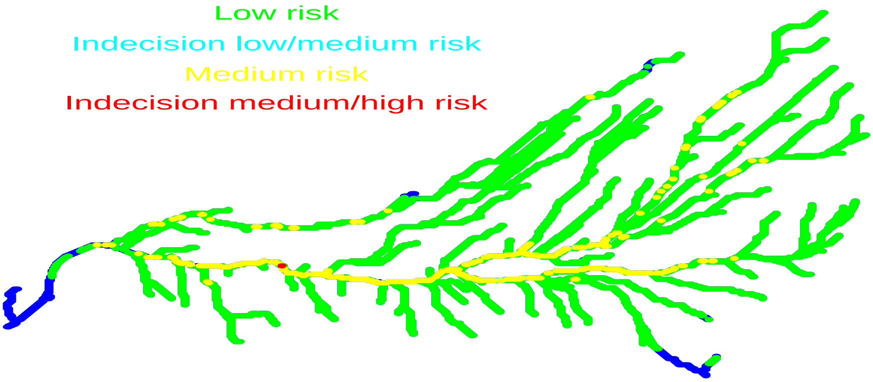
$$\mathbb{P}(\text{Granulometry} \in [10, 100]) \in [0.1, 0.2]$$

Debris flows hazard assessment



Debris flows hazard assessment

- Extensive simulations in a debris flow prone watershed
  Acquarossa Creek Basin (area 1.6 Km^2, length 3.1 Km)

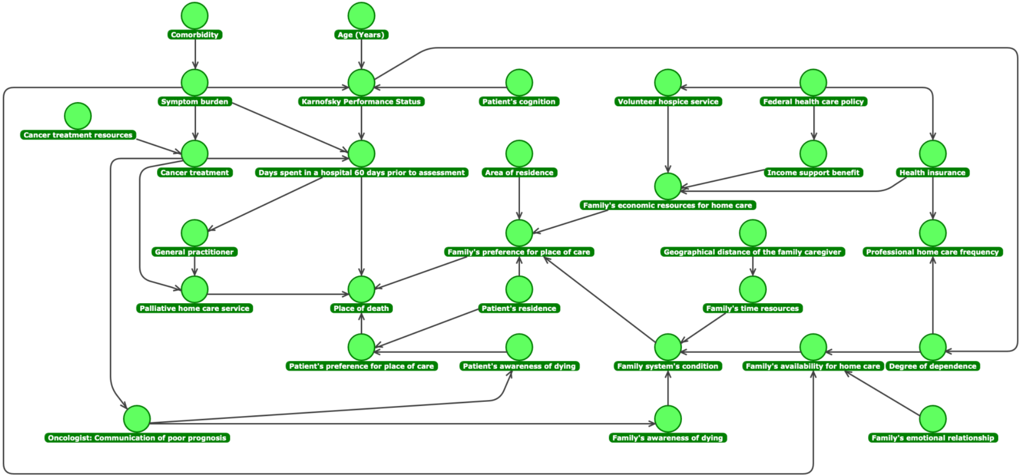


Place of Death Analysis

- Demographic and clinical data about 116 adult patients who died from cancer
  - Place of Death: hospital (78%), home (18%), nursing home (4%)
  - Days spent in a hospital 60 days prior to assessment: 0-20, 21-40, 41-60
  - Age: 20-40; 41-65; 66-80; >80
  - Comorbidity: mild or non-existent, severe
  - Cancer treatment: ongoing; discontinued
  - Area of residence: rural, urban
  - . . .

- Assessments:
  - “Ongoing cancer treatment decreases the probability of open communication by 40% compared with that of treatment discontinuation”
  - “Of the patients, 60 to 80% were probably only partially informed about the proximity of death when undergoing anticancer treatments”
  - “patients and their relatives are more likely to remain in a closed rather than open state of awareness of dying”

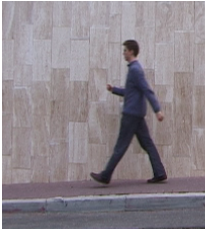
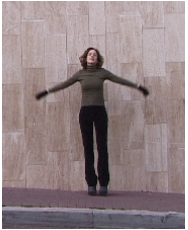
Place of Death Analysis



Place of Death Analysis

“if the family’s preference for the POD is the hospital, despite full access to an interdisciplinary home care network, the probability of dying at home drops from 76–83% to 19–40%”

Action Recognition

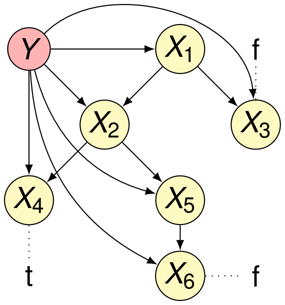




- Learn one iHMM for each action
- Classify video by interval dominance
  - E.g. $\underline{\mathbb{P}}(e_1, \ldots, e_n|\text{walk}) > \overline{\mathbb{P}}(e_1, \ldots, e_n|\text{jump})$
  - Indeterminate classifications

Credal Classifiers

Specially designed credal networks for cautious classification

- Makes fewer assumptions than Bayesian network classifiers (e.g. Naive Bayes Classifier)
- Allow/produce indeterminacy in classifications



Evaluating set-valued classifications

True class	Credal classifier	Precise classifier
red	red, yellow	red
red	yellow, green	yellow
yellow	green	green
green	red, green, yellow	yellow
green	green	green

- Determinacy: 2/5
- Average output size: 7/3
- Precise accuracy: 1/2
- Set accuracy: 2/3
- Discounted acc.: $(1/2 + 1/3 + 1)/5$

Action Recognition

Determinacy	77.5%	(4.6%)	(62/80)
Average output size	2.4		(out of 10)
Single accuracy	35.5%	(6.0%)	(22/62)
Set accuracy	44.4%	(11.4%)	(8/18)
Discounted accuracy	32.1%		
Utility-based accuracy $u_{.65}$	33.5%		
Utility-based accuracy $u_{.80}$	35.0%		
Accuracy precise counterpart	31.3%	(5.2%)	(25/80)
Precise single accuracy	35.5%	(6.0%)	(22/62)
Precise set accuracy	16.7%	(8.5%)	(3/18)

Table of Contents

Final Remarks

- Already thirty years of development around credal networks,
  - mostly centered on strong extensions,
  - but quite a bit on epistemic extensions as well.

Final Remarks

- Already thirty years of development around credal networks,
  - mostly centered on strong extensions,
  - but quite a bit on epistemic extensions as well.
- Two questions:
  - What is the best extension?
  - Is convexity so important?

Final Remarks

- Already thirty years of development around credal networks,
  - mostly centered on strong extensions,
  - but quite a bit on epistemic extensions as well.
- Two questions:
  - What is the best extension?
  - Is convexity so important?
- We now have a solid set of algorithms, a comprehensive set of results on computational complexity.
- However, still space to produce faster algorithms with guarantees, and in particular algorithms for decision-making.