Likelihood-based Robust Classification with Bayesian Networks

Alessandro Antonucci[*], Marco Cattaneo[†], Giorgio Corani[*]

[*] Istituto "Dalle Molle" di Studi sull'Intelligenza Artificiale - Lugano (Switzerland)

[†] Ludwig-Maximilians-Universität - München (Germany)

IPMU'12 - Catania, July 19, 2012

Classification

- Complete data $\mathcal{D} = \{(x_0^{(i)}, \mathbf{x}^{(i)})\}_{i=1}^{n}$ about class variable X_0 and (discrete) features $\mathbf{X} := (X_1, \ldots, X_n)$
- Class label $x_0^* \in \Omega_{X_0}$ of a new instance $\tilde{\mathbf{x}}$ of the features?
- Probabilistic approaches learn $P(X_0, \mathbf{X})$ from $\mathcal{D}$

  optimal class has the highest posterior $x_0^* := \arg\max_{x_0} P(x_0|\tilde{\mathbf{x}})$
- Equivalently, dominance test: $\forall x_0', x_0'' \in \Omega_{X_0}$ check

$$\boxed{\frac{P(x_0'|\tilde{\mathbf{x}})}{P(x_0''|\tilde{\mathbf{x}})} = \frac{P(x_0', \tilde{\mathbf{x}})}{P(x_0'', \tilde{\mathbf{x}})} > 1}$$

x_0^* is the only undominated class

- Posterior probabilities $\propto$ joint probabilities

Classification

- Complete data $\mathcal{D} = \{(x_0^{(i)}, \mathbf{x}^{(i)})\}_{i=1}^n$ about class variable X_0 and (discrete) features $\mathbf{X} := (X_1, \ldots, X_n)$
- Class label $x_0^* \in \Omega_{X_0}$ of a new instance $\tilde{\mathbf{x}}$ of the features?
- Probabilistic approaches learn $P(X_0, \mathbf{X})$ from $\mathcal{D}$

  optimal class has the highest posterior $x_0^* := \arg\max_{x_0} P(x_0|\tilde{\mathbf{x}})$
- Equivalently, dominance test: $\forall x_0', x_0'' \in \Omega_{X_0}$ check

$$\boxed{\frac{P(x_0'|\tilde{\mathbf{x}})}{P(x_0''|\tilde{\mathbf{x}})} = \frac{P(x_0', \tilde{\mathbf{x}})}{P(x_0'', \tilde{\mathbf{x}})} > 1}$$

  x_0^* is the only undominated class
- Posterior probabilities $\propto$ joint probabilities

Classification

- Complete data $\mathcal{D} = \{(x_0^{(i)}, \mathbf{x}^{(i)})\}_{i=1}^{n}$ about class variable X_0 and (discrete) features $\mathbf{X} := (X_1, \ldots, X_n)$
- Class label $x_0^* \in \Omega_{X_0}$ of a new instance $\tilde{\mathbf{x}}$ of the features?
- Probabilistic approaches learn $P(X_0, \mathbf{X})$ from $\mathcal{D}$

  optimal class has the highest posterior $x_0^* := \arg\max_{x_0} P(x_0|\tilde{\mathbf{x}})$
- Equivalently, dominance test: $\forall x_0', x_0'' \in \Omega_{X_0}$ check

$$\boxed{\frac{P(x_0'|\tilde{\mathbf{x}})}{P(x_0''|\tilde{\mathbf{x}})} = \frac{P(x_0', \tilde{\mathbf{x}})}{P(x_0'', \tilde{\mathbf{x}})} > 1}$$

x_0^* is the only undominated class

- Posterior probabilities $\propto$ joint probabilities

Classification

- Complete data $\mathcal{D} = \{(x_0^{(i)}, \mathbf{x}^{(i)})\}_{i=1}^{n}$ about class variable X_0 and (discrete) features $\mathbf{X} := (X_1, \ldots, X_n)$
- Class label $x_0^* \in \Omega_{X_0}$ of a new instance $\tilde{\mathbf{x}}$ of the features?
- Probabilistic approaches learn $P(X_0, \mathbf{X})$ from $\mathcal{D}$

  optimal class has the highest posterior $x_0^* := \arg\max_{x_0} P(x_0|\tilde{\mathbf{x}})$
- Equivalently, dominance test: $\forall x_0', x_0'' \in \Omega_{X_0}$ check

$$\boxed{\frac{P(x_0'|\tilde{\mathbf{x}})}{P(x_0''|\tilde{\mathbf{x}})} = \frac{P(x_0', \tilde{\mathbf{x}})}{P(x_0'', \tilde{\mathbf{x}})} > 1}$$

x_0^* is the only undominated class

- Posterior probabilities $\propto$ joint probabilities

Classification

- Complete data $\mathcal{D} = \{(x_0^{(i)}, \mathbf{x}^{(i)})\}_{i=1}^{n}$ about class variable X_0 and (discrete) features $\mathbf{X} := (X_1, \ldots, X_n)$
- Class label $x_0^* \in \Omega_{X_0}$ of a new instance $\tilde{\mathbf{x}}$ of the features?
- Probabilistic approaches learn $P(X_0, \mathbf{X})$ from $\mathcal{D}$

  optimal class has the highest posterior $x_0^* := \arg\max_{x_0} P(x_0|\tilde{\mathbf{x}})$
- Equivalently, dominance test: $\forall x_0', x_0'' \in \Omega_{X_0}$ check

$$\boxed{\frac{P(x_0'|\tilde{\mathbf{x}})}{P(x_0''|\tilde{\mathbf{x}})} = \frac{P(x_0', \tilde{\mathbf{x}})}{P(x_0'', \tilde{\mathbf{x}})} > 1}$$

x_0^* is the only undominated class

- Posterior probabilities $\propto$ joint probabilities

Classification

- Complete data $\mathcal{D} = \{(x_0^{(i)}, \mathbf{x}^{(i)})\}_{i=1}^{n}$ about class variable X_0 and (discrete) features $\mathbf{X} := (X_1, \ldots, X_n)$
- Class label $x_0^* \in \Omega_{X_0}$ of a new instance $\tilde{\mathbf{x}}$ of the features?
- Probabilistic approaches learn $P(X_0, \mathbf{X})$ from $\mathcal{D}$

  optimal class has the highest posterior $x_0^* := \arg\max_{x_0} P(x_0|\tilde{\mathbf{x}})$
- Equivalently, dominance test: $\forall x_0', x_0'' \in \Omega_{X_0}$ check

$$\boxed{\frac{P(x_0'|\tilde{\mathbf{x}})}{P(x_0''|\tilde{\mathbf{x}})} = \frac{P(x_0', \tilde{\mathbf{x}})}{P(x_0'', \tilde{\mathbf{x}})} > 1}$$

  x_0^* is the only undominated class
- Posterior probabilities $\propto$ joint probabilities

Classification with Bayesian networks

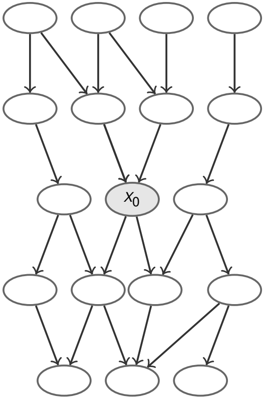
- Bayesian networks: a graph $\mathcal{G}$ to depict conditional independencies in $(X_0, \mathbf{X})$
- $\mathcal{G}$ induces a factorization in the joint

$$P(x_0, \mathbf{x}) = \prod_{i=0}^{n} P(x_i | \pi_i)$$

- Dominance test rewrites as

$$\boxed{\frac{P(x_0', \tilde{\mathbf{x}})}{P(x_0', \tilde{\mathbf{x}})} = \prod_{i=0}^{n} \frac{P(x_i' | \pi_i')}{P(x_i'' | \pi_i'')} > 1}$$

- Factors not including X_0 are equal to one focusing on the Markov blanket of X_0



Classification with Bayesian networks

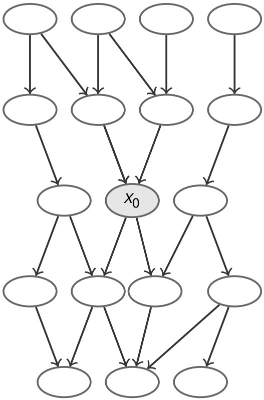
- Bayesian networks: a graph $\mathcal{G}$ to depict conditional independencies in $(X_0, \mathbf{X})$
- $\mathcal{G}$ induces a factorization in the joint

$$P(x_0, \mathbf{x}) = \prod_{i=0}^{n} P(x_i | \pi_i)$$

- Dominance test rewrites as

$$\boxed{\frac{P(x_0', \tilde{\mathbf{x}})}{P(x_0', \tilde{\mathbf{x}})} = \prod_{i=0}^{n} \frac{P(x_i' | \pi_i')}{P(x_i'' | \pi_i'')} > 1}$$

- Factors not including X_0 are equal to one focusing on the Markov blanket of X_0



Classification with Bayesian networks

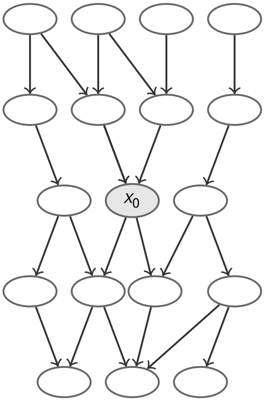
- Bayesian networks: a graph $\mathcal{G}$ to depict conditional independencies in $(X_0, \mathbf{X})$
- $\mathcal{G}$ induces a factorization in the joint

$$P(x_0, \mathbf{x}) = \prod_{i=0}^{n} P(x_i|\pi_i)$$

- Dominance test rewrites as

$$\boxed{\frac{P(x_0', \tilde{\mathbf{x}})}{P(x_0', \tilde{\mathbf{x}})} = \prod_{i=0}^{n} \frac{P(x_i'|\pi_i')}{P(x_i''|\pi_i'')} > 1}$$

- Factors not including X_0 are equal to one focusing on the Markov blanket of X_0



Classification with Bayesian networks

- Bayesian networks: a graph $\mathcal{G}$ to depict conditional independencies in $(X_0, \mathbf{X})$
- $\mathcal{G}$ induces a factorization in the joint

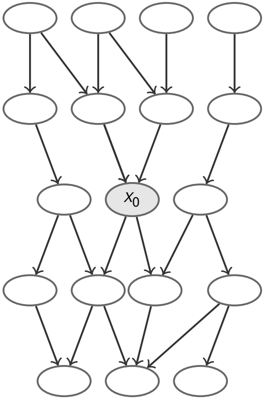
$$P(x_0, \mathbf{x}) = \prod_{i=0}^{n} P(x_i|\pi_i)$$

- Dominance test rewrites as

$$\boxed{\frac{P(x_0', \tilde{\mathbf{x}})}{P(x_0', \tilde{\mathbf{x}})} = \prod_{i=0}^{n} \frac{P(x_i'|\pi_i')}{P(x_i''|\pi_i'')} > 1}$$

- Factors not including X_0 are equal to one

focusing on the Markov blanket of X_0



Classification with Bayesian networks

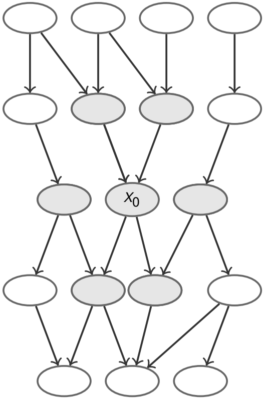
- Bayesian networks: a graph $\mathcal{G}$ to depict conditional independencies in $(X_0, \mathbf{X})$
- $\mathcal{G}$ induces a factorization in the joint

$$P(x_0, \mathbf{x}) = \prod_{i=0}^{n} P(x_i|\pi_i)$$

- Dominance test rewrites as

$$\boxed{\prod_{X_i \in \mathrm{Blank}(X_0)} \frac{P(x_i'|\pi_i')}{P(x_i''|\pi_i'')} > 1}$$

- Factors not including X_0 are equal to one focusing on the Markov blanket of X_0

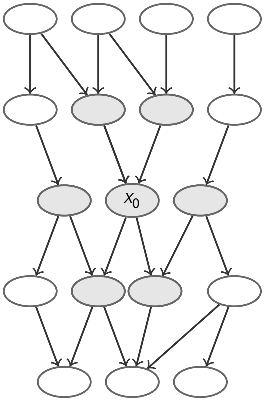


Classification with credal networks

- Few data $\Rightarrow$ unreliable learning of $P(X_i|\pi_i)$
- More reliable with sets of prob functions
- A credal set of joint $\mathbf{P}(X_0, \mathbf{X})$, instead of a single $P(X_0, \mathbf{X})$ (all Bayesian nets over $\mathcal{G}$)
- How to classify instances?
- x_0' dominates x_0'' iff this happens for each Bayesian net (maximality), i.e.,

$$\min_{P(X_0,\mathbf{X})\in\mathbf{P}(X_0,\mathbf{X})} \prod_{X_i\in\mathrm{Blank}(X_0)} \frac{P(x_i'|\pi_i')}{P(x_i''|\pi_i'')} > 1$$

- Not always a single optimal class, can be also a set of undominated classes
- This is a credal classifier possibly assigning multiple classes to test instance

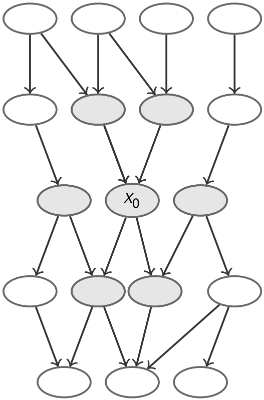


Classification with credal networks

- Few data $\Rightarrow$ unreliable learning of $P(X_i|\pi_i)$
- More reliable with sets of prob functions
- A credal set of joint $\mathbf{P}(X_0, \mathbf{X})$, instead of a single $P(X_0, \mathbf{X})$ (all Bayesian nets over $\mathcal{G}$)
- How to classify instances?
- x_0' dominates x_0'' iff this happens for each Bayesian net (maximality), i.e.,

$$\min_{P(X_0,\mathbf{X})\in\mathbf{P}(X_0,\mathbf{X})} \prod_{X_i\in\mathrm{Blank}(X_0)} \frac{P(x_i'|\pi_i')}{P(x_i''|\pi_i'')} > 1$$

- Not always a single optimal class, can be also a set of undominated classes
- This is a credal classifier possibly assigning multiple classes to test instance

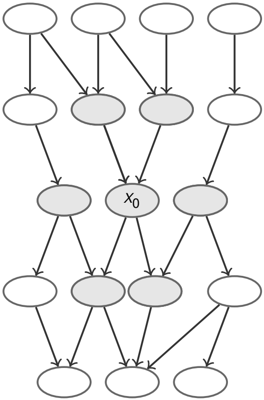


Classification with credal networks

- Few data $\Rightarrow$ unreliable learning of $P(X_i|\pi_i)$
- More reliable with sets of prob functions
- A credal set of joint $\mathbf{P}(X_0, \mathbf{X})$, instead of a single $P(X_0, \mathbf{X})$ (all Bayesian nets over $\mathcal{G}$)
- How to classify instances?
- x_0' dominates x_0'' iff this happens for each Bayesian net (maximality), i.e.,

$$\min_{P(X_0,\mathbf{X})\in\mathbf{P}(X_0,\mathbf{X})} \prod_{X_i\in\mathrm{Blank}(X_0)} \frac{P(x_i'|\pi_i')}{P(x_i''|\pi_i'')} > 1$$

- Not always a single optimal class, can be also a set of undominated classes
- This is a credal classifier possibly assigning multiple classes to test instance

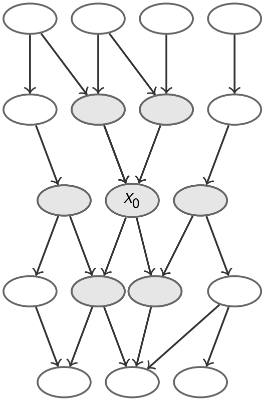


Classification with credal networks

- Few data $\Rightarrow$ unreliable learning of $P(X_i|\pi_i)$
- More reliable with sets of prob functions
- A credal set of joint $\mathbf{P}(X_0, \mathbf{X})$, instead of a single $P(X_0, \mathbf{X})$ (all Bayesian nets over $\mathcal{G}$)
- How to classify instances?
- x_0' dominates x_0'' iff this happens for each Bayesian net (maximality), i.e.,

$$\min_{P(X_0,\mathbf{X}) \in \mathbf{P}(X_0,\mathbf{X})} \prod_{X_i \in \mathrm{Blank}(X_0)} \frac{P(x_i'|\pi_i')}{P(x_i''|\pi_i'')} > 1$$

- Not always a single optimal class, can be also a set of undominated classes
- This is a credal classifier possibly assigning multiple classes to test instance

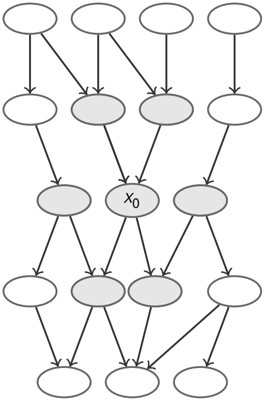


Classification with credal networks

- Few data $\Rightarrow$ unreliable learning of $P(X_i|\pi_i)$
- More reliable with sets of prob functions
- A credal set of joint $\mathbf{P}(X_0, \mathbf{X})$, instead of a single $P(X_0, \mathbf{X})$ (all Bayesian nets over $\mathcal{G}$)
- How to classify instances?
- x_0' dominates x_0'' iff this happens for each Bayesian net (maximality), i.e.,

$$\min_{P(X_0,\mathbf{X})\in\mathbf{P}(X_0,\mathbf{X})} \prod_{X_i\in\mathrm{Blank}(X_0)} \frac{P(x_i'|\pi_i')}{P(x_i''|\pi_i'')} > 1$$

- Not always a single optimal class, can be also a set of undominated classes
- This is a credal classifier possibly assigning multiple classes to test instance

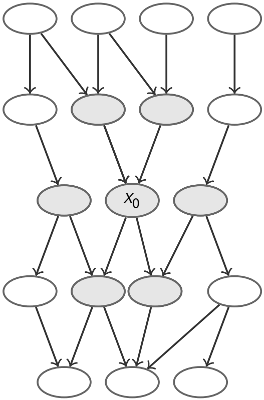


Classification with credal networks

- Few data $\Rightarrow$ unreliable learning of $P(X_i|\pi_i)$
- More reliable with sets of prob functions
- A credal set of joint $\mathbf{P}(X_0, \mathbf{X})$, instead of a single $P(X_0, \mathbf{X})$ (all Bayesian nets over $\mathcal{G}$)
- How to classify instances?
- x_0' dominates x_0'' iff this happens for each Bayesian net (maximality), i.e.,

$$\boxed{\min_{P(X_0,\mathbf{X})\in\mathbf{P}(X_0,\mathbf{X})} \prod_{X_i\in\mathrm{Blank}(X_0)} \frac{P(x_i'|\pi_i')}{P(x_i''|\pi_i'')} > 1}$$

- Not always a single optimal class, can be also a set of undominated classes
- This is a credal classifier possibly assigning multiple classes to test instance

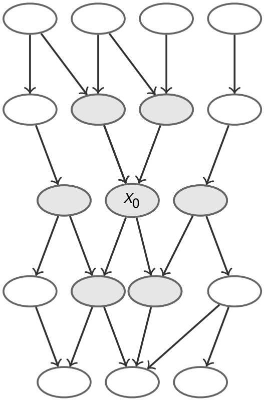


Classification with credal networks

- Few data $\Rightarrow$ unreliable learning of $P(X_i|\pi_i)$
- More reliable with sets of prob functions
- A credal set of joint $\mathbf{P}(X_0, \mathbf{X})$, instead of a single $P(X_0, \mathbf{X})$ (all Bayesian nets over $\mathcal{G}$)
- How to classify instances?
- x_0' dominates x_0'' iff this happens for each Bayesian net (maximality), i.e.,

$$\min_{P(X_0,\mathbf{X}) \in \mathbf{P}(X_0,\mathbf{X})} \prod_{X_i \in \mathrm{Blank}(X_0)} \frac{P(x_i'|\pi_i')}{P(x_i''|\pi_i'')} > 1$$

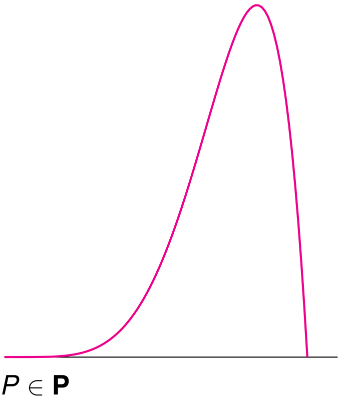
- Not always a single optimal class, can be also a set of undominated classes
- This is a credal classifier possibly assigning multiple classes to test instance



Likelihood-based learning sets of distributions

Likelihood-based learning sets of distributions

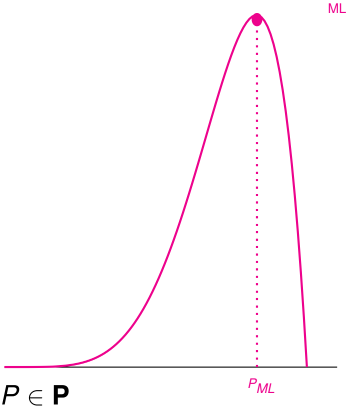
- Frequentist: no priors, learning from the likelihood $P(\mathcal{D})$



Likelihood-based learning sets of distributions

- Frequentist: no priors, learning from the likelihood $P(\mathcal{D})$
- Given a set of models $\mathbf{P}$,

$$P_{ML} := \arg\max_{P \in \mathbf{P}} P(\mathcal{D})$$



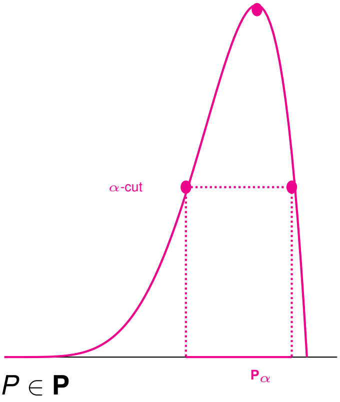
Likelihood-based learning sets of distributions

- Frequentist: no priors, learning from the likelihood $P(\mathcal{D})$
- Given a set of models $\mathbf{P}$,

$$P_{ML} := \arg\max_{P \in \mathbf{P}} P(\mathcal{D})$$

- More robust: all the models s.t. likelihood $\geq$ threshold

$$\mathbf{P}_\alpha := \{ P \in \mathbf{P} \,|\, P(\mathcal{D}) \geq \alpha P_{ML}(\mathcal{D}) \}$$



Likelihood-based learning sets of distributions

- Frequentist: no priors, learning from the likelihood $P(\mathcal{D})$
- Given a set of models $\mathbf{P}$,

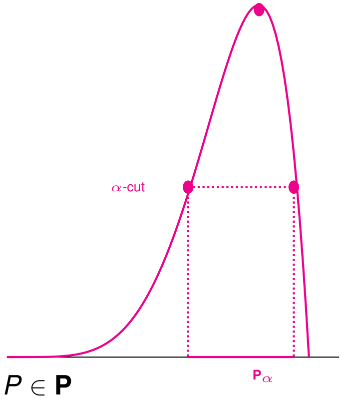
$$P_{ML} := \arg\max_{P \in \mathbf{P}} P(\mathcal{D})$$

- More robust: all the models s.t. likelihood $\geq$ threshold

$$\mathbf{P}_\alpha := \{ P \in \mathbf{P} \,|\, P(\mathcal{D}) \geq \alpha P_{ML}(\mathcal{D}) \}$$

threshold $\alpha \in [0,1]$, $\mathbf{P}_\alpha \subseteq \mathbf{P}$

$(\mathbf{P}_{\alpha=0} \equiv \mathbf{P}$, $\mathbf{P}_{\alpha=1} = P_{ML})$



Likelihood-based classification

- $\mathbf{P}$ is any BN quantification
- $\mathcal{D}$ and α to shrink $\mathbf{P}$ to $\mathbf{P}_\alpha$
- Dominance test

$$\boxed{\min_{P(X_0,\mathbf{X})\in\mathbf{P}_\alpha} \log \frac{P(x_0'|\tilde{\mathbf{x}})}{P(x_0''|\tilde{\mathbf{x}})} > 0}$$

- Monte Carlo approach
  Sampling P from $\mathbf{P}$
  if P: $\frac{P(\mathcal{D})}{P_{ML}(\mathcal{D})} > \alpha$ AND $\frac{P(x_0'|\tilde{\mathbf{x}})}{P(x_0''|\tilde{\mathbf{x}})} < 1$
  then no dominance
- Analytical methods
  Profile lik (upper envelope)
  detect (numerically) α-cut

Likelihood-based classification

- **P** is any BN quantification
- $\mathcal{D}$ and α to shrink **P** to $\mathbf{P}_\alpha$
- Dominance test

$$\boxed{\min_{P(X_0,\mathbf{X})\in\mathbf{P}_\alpha} \log \frac{P(x_0'|\tilde{\mathbf{x}})}{P(x_0''|\tilde{\mathbf{x}})} > 0}$$

- Monte Carlo approach
  Sampling P from **P**
  if P: $\frac{P(\mathcal{D})}{P_{ML}(\mathcal{D})} > \alpha$ AND $\frac{P(x_0'|\tilde{\mathbf{x}})}{P(x_0''|\tilde{\mathbf{x}})} < 1$
  then no dominance
- Analytical methods
  Profile lik (upper envelope)
  detect (numerically) α-cut

Likelihood-based classification

- **P** is any BN quantification
- $\mathcal{D}$ and α to shrink **P** to $\mathbf{P}_\alpha$
- Dominance test

  $$\boxed{\min_{P(X_0,\mathbf{X})\in\mathbf{P}_\alpha} \log \frac{P(x_0'|\tilde{\mathbf{x}})}{P(x_0''|\tilde{\mathbf{x}})} > 0}$$

- Monte Carlo approach
  Sampling P from **P**
  if P: $\frac{P(\mathcal{D})}{P_{ML}(\mathcal{D})} > \alpha$ AND $\frac{P(x_0'|\tilde{\mathbf{x}})}{P(x_0''|\tilde{\mathbf{x}})} < 1$
  then no dominance
- Analytical methods
  Profile lik (upper envelope)
  detect (numerically) α-cut

Likelihood-based classification

- **P** is any BN quantification
- $\mathcal{D}$ and α to shrink **P** to $\mathbf{P}_\alpha$
- Dominance test

$$\boxed{\min_{P(X_0,\mathbf{X})\in\mathbf{P}_\alpha} \log \frac{P(x_0'|\tilde{\mathbf{x}})}{P(x_0''|\tilde{\mathbf{x}})} > 0}$$

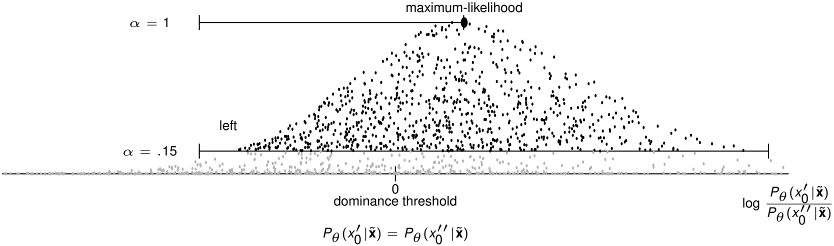
- Monte Carlo approach
  Sampling P from **P**
  if P: $\frac{P(\mathcal{D})}{P_{ML}(\mathcal{D})} > \alpha$ AND $\frac{P(x_0'|\tilde{\mathbf{x}})}{P(x_0''|\tilde{\mathbf{x}})} < 1$
  then no dominance
- Analytical methods
  Profile lik (upper envelope)
  detect (numerically) α-cut

Likelihood-based classification

- **P** is any BN quantification
- $\mathcal{D}$ and α to shrink **P** to $\mathbf{P}_\alpha$
- Dominance test

$$\boxed{\min_{P(X_0,\mathbf{X})\in\mathbf{P}_\alpha} \log \frac{P(x_0'|\tilde{\mathbf{x}})}{P(x_0''|\tilde{\mathbf{x}})} > 0}$$

- Monte Carlo approach
  Sampling P from **P**
  if P: $\frac{P(\mathcal{D})}{P_{ML}(\mathcal{D})} > \alpha$ AND $\frac{P(x_0'|\tilde{\mathbf{x}})}{P(x_0''|\tilde{\mathbf{x}})} < 1$
  then no dominance
- Analytical methods
  Profile lik (upper envelope)
  detect (numerically) α-cut

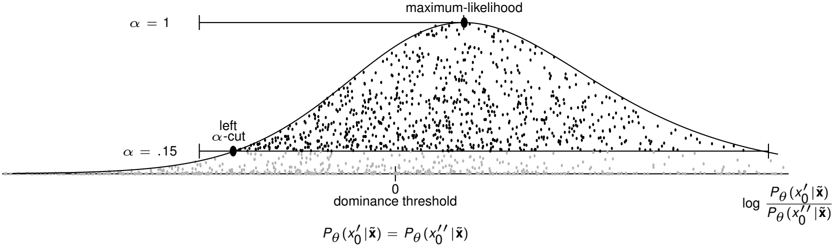


Likelihood-based classification

- **P** is any BN quantification
- $\mathcal{D}$ and α to shrink **P** to $\mathbf{P}_\alpha$
- Dominance test

$$\boxed{\min_{P(X_0,\mathbf{X})\in\mathbf{P}_\alpha} \log \frac{P(x_0'|\tilde{\mathbf{x}})}{P(x_0''|\tilde{\mathbf{x}})} > 0}$$

- Monte Carlo approach
  Sampling P from **P**
  if P: $\frac{P(\mathcal{D})}{P_{ML}(\mathcal{D})} > \alpha$ AND $\frac{P(x_0'|\tilde{\mathbf{x}})}{P(x_0''|\tilde{\mathbf{x}})} < 1$
  then no dominance
- Analytical methods
  Profile lik (upper envelope)
  detect (numerically) α-cut



Our formula

- Parametric formulae for the profile likelihood (complete data)

$$\left\{ \left(\frac{P_t(x_0', \tilde{x}_1, \dots, \tilde{x}_n)}{P_t(x_0'', \tilde{x}_1, \dots, \tilde{x}_n)}, P_t(\mathcal{D}) \right) : t \in [a, b] \right\}$$

- For the naive structure (ISIPTA '11)
- For general topologies (this paper)
- First credal classifier for BNs with general topologies!
- Bayesian-like approaches only for naive Bayes and TAN

Our formula

- Parametric formulae for the profile likelihood (complete data)

$$\left\{ \left(\frac{P_t(x_0', \tilde{x}_1, \ldots, \tilde{x}_n)}{P_t(x_0'', \tilde{x}_1, \ldots, \tilde{x}_n)}, P_t(\mathcal{D}) \right) : t \in [a, b] \right\}$$

- For the naive structure (ISIPTA '11)
- For general topologies (this paper)
- First credal classifier for BNs with general topologies!
- Bayesian-like approaches only for naive Bayes and TAN

Our formula

- Parametric formulae for the profile likelihood (complete data)

$$\left\{ \left(\frac{P_t(x_0', \tilde{x}_1, \ldots, \tilde{x}_n)}{P_t(x_0'', \tilde{x}_1, \ldots, \tilde{x}_n)}, P_t(\mathcal{D}) \right) : t \in [a, b] \right\}$$

- For the naive structure (ISIPTA '11)
- For general topologies (this paper)
- First credal classifier for BNs with general topologies!
- Bayesian-like approaches only for naive Bayes and TAN

Our formula

- Parametric formulae for the profile likelihood (complete data)

$$\left\{ \left(\frac{P_t(x_0', \tilde{x}_1, \ldots, \tilde{x}_n)}{P_t(x_0'', \tilde{x}_1, \ldots, \tilde{x}_n)}, P_t(\mathcal{D}) \right) : t \in [a, b] \right\}$$

- For the naive structure (ISIPTA '11)
- For general topologies (this paper)
- First credal classifier for BNs with general topologies!
- Bayesian-like approaches only for naive Bayes and TAN

Our formula

$$
\begin{aligned}
a &:= -\min\{n(x_0', \tilde{\pi}_0), n(\tilde{x}_1, x_0', \tilde{\pi}_1), \ldots, n(\tilde{x}_k, x_0', \tilde{\pi}_k)\} \\
b &:= \min\{n(x_0'', \tilde{\pi}_0), n(\tilde{x}_1, x_0'', \tilde{\pi}_1), \ldots, n(\tilde{x}_k, x_0'', \tilde{\pi}_k)\}
\end{aligned}
$$

For each $t \in [a, b]$, let us consider the following functions:

$$
x(t) := \frac{n(x_0', \tilde{\pi}_0) + t}{\hat{n}(x_0'', \tilde{\pi}_0) - t} \cdot \prod_{i=1}^{k} \frac{\frac{n(\tilde{x}_i, x_0', \tilde{\pi}_i) + t}{n(x_0', \tilde{\pi}_i) + t}}{\frac{n(\tilde{x}_i, x_0'', \tilde{\pi}_i) - t}{n(x_0'', \tilde{\pi}_i) - t}}
$$

$$
\begin{aligned}
y(t) := {} & [n(x_0', \tilde{\pi}_0) + t]^{n(x_0', \tilde{\pi}_0)} \cdot [n(x_0'', \tilde{\pi}_0) - t]^{n(x_0'', \tilde{\pi}_0)} \\
& \cdot \prod_{i=1}^{k} \left[\frac{[n(\tilde{x}_i, x_0', \tilde{\pi}_i) + t]^{n(\tilde{x}_i, x_0', \tilde{\pi}_i)}}{\left[n(x_0', \tilde{\pi}_i) + t\right]^{n(x_0', \tilde{\pi}_i)}} \cdot \frac{[n(\tilde{x}_i, x_0'', \tilde{\pi}_i) - t]^{n(\tilde{x}_i, x_0'', \tilde{\pi}_i)}}{\left[n(x_0'', \tilde{\pi}_i) - t\right]^{n(x_0'', \tilde{\pi}_i)}} \right]
\end{aligned}
$$

Coping with zero counts

- With zero counts, classifier becomes unnecessarily imprecise
- Known issue (also for Bayesian-like approaches)
- Solved by a semi-supervised approach
  - test instance as an incomplete observation ($X_0 = ?, \mathbf{X} = \tilde{\mathbf{x}}$)
  - EM to complete the missing observation with fractionary counts
  - Use formula as for complete data
- This solve the zero count problem

Coping with zero counts

- With zero counts, classifier becomes unnecessarily imprecise
- Known issue (also for Bayesian-like approaches)
- Solved by a semi-supervised approach
  - test instance as an incomplete observation $(X_0 = ?, \mathbf{X} = \tilde{\mathbf{x}})$
  - EM to complete the missing observation with fractionary counts
  - Use formula as for complete data
- This solve the zero count problem

Coping with zero counts

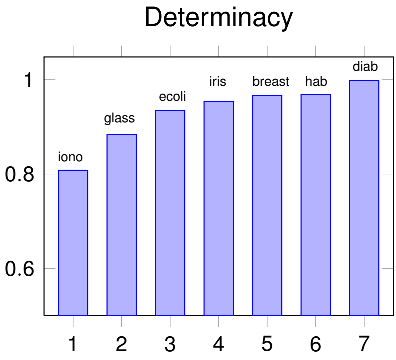
- With zero counts, classifier becomes unnecessarily imprecise
- Known issue (also for Bayesian-like approaches)
- Solved by a semi-supervised approach
  - test instance as an incomplete observation ($X_0 = ?, \mathbf{X} = \tilde{\mathbf{x}}$)
  - EM to complete the missing observation with fractionary counts
  - Use formula as for complete data
- This solve the zero count problem

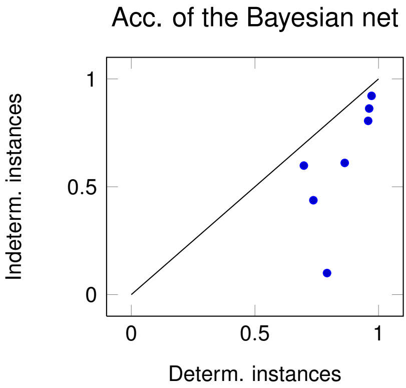
Coping with zero counts

- With zero counts, classifier becomes unnecessarily imprecise
- Known issue (also for Bayesian-like approaches)
- Solved by a semi-supervised approach
  - test instance as an incomplete observation ($X_0 = ?, \mathbf{X} = \tilde{\mathbf{x}}$)
  - EM to complete the missing observation with fractionary counts
  - Use formula as for complete data
- This solve the zero count problem

Experiments

If determinate the class returned by the credal classifier
is the same returned by the Bayesian network





Mostly determinate, successfully detect hard-to-classify instances

Conclusions and Outlooks

- Conclusions
  - A credal classifier for general topologies
  - Solve zero-counts problem with a semi-supervised technique
  - Separate easy-to-classify from hard instances
- Outlooks
  - Comparisons with other credal classifiers on specific topologies
  - Pre-processing for other (precise) classifiers
  - Applications to state-of-the-art approaches (AODE)