Structure-based Categorisation of Bayesian Network Parameters

Janneke H. Bolt & Silja Renooij

Department of Information and Computing Sciences
Utrecht University, The Netherlands

Universiteit Utrecht

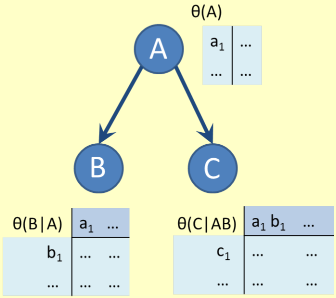
ECSQARU 2017, Lugano

Context I

Multi-parameter tuning in Bayesian networks

[Bolt & van der Gaag, ECSQARU 2015]

- change $\Pr(\mathbf{h} \mid \mathbf{e})$ by changing $\theta_1, \ldots, \theta_n$ simultaneously
- most effective if parameters work in same direction



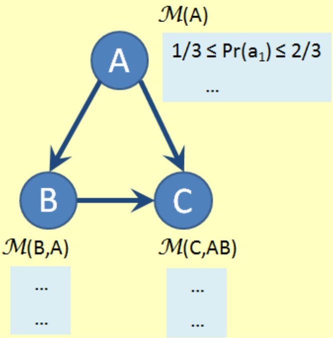
Exploitable knowledge: qualitative effect ('$+$', '$-$') on $\Pr(\mathbf{h} \mid \mathbf{e})$ of change in θ_i for any value $\theta_j \neq \theta_i$

Context II

Pre-processing for inference in credal networks

[Bolt, De Bock & Renooij, ECAI 2016]

- ▷ inference: compute $\overline{\Pr}(\mathbf{h}\,|\,\mathbf{e})$ and $\underline{\Pr}(\mathbf{h}\,|\,\mathbf{e})$
- ▷ simplify computation: replace credal set $\mathcal{M}(V, pa(V))$ with single $\theta(V\,|\,pa(V))$



Exploitable knowledge: qualitative effect ('$+$', '$-$') on $\Pr(\mathbf{h}\,|\,\mathbf{e})$ for choice of $\theta(v\,|\,pa(V))$

Relating parameters to output probabilities

This is exactly what sensitivity functions for Bayesian networks $\mathcal{B}$ do:

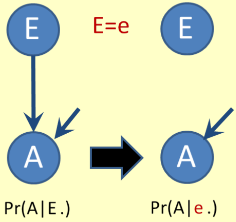
$$\Pr(\mathbf{h} \,|\, \mathbf{e}) = \frac{\Pr(\mathbf{h}, \mathbf{e})}{\Pr(\mathbf{e})} = \frac{\sum_{\mathbf{R}} \prod_V \theta(v \,|\, \boldsymbol{\pi})}{\sum_{\mathbf{H},\mathbf{R}} \prod_V \theta(v \,|\, \boldsymbol{\pi})}$$

- fraction of multi-linear expressions in the parameters
- exact function $f_{\Pr(\mathbf{h} \,|\, \mathbf{e})}(\theta)$ depends on (fixed) values of other parameters
- whether $f(\theta)$ is increasing or decreasing may be independent of those other parameters

Which parameters are not used?

We consider parameters in the query-dependent backbone $\mathcal{B}_q$ of $\mathcal{B}$:

- no d-separated nodes (V_i s.t. $\langle V_i, \mathbf{H} | \mathbf{E} \rangle_d$)
- no barren nodes
  ($V_i \in \mathbf{R}$ s.t. V_i is leaf or $de(V_i) \subseteq$ barren)
- evidence absorption
  ($\Longrightarrow \forall\ V_i : pa(V_i) \cap \mathbf{E} = \emptyset$)



Which parameters *are* used?

$$\Pr(\mathbf{h}, \mathbf{e}) = \sum_{\mathbf{R}} \prod_{V} \theta(v|\boldsymbol{\pi}) \qquad \Big| \qquad \Pr(\mathbf{e}) = \sum_{\mathbf{H},\mathbf{R}} \prod_{V} \theta(v|\boldsymbol{\pi})$$

Computations include parameters $\theta(v \mid \boldsymbol{\pi})$ obeying

- in general: $v \sim \mathbf{her}$ and $\boldsymbol{\pi} \sim \mathbf{h}\cancel{\mathbf{e}}\mathbf{r}$

Which parameters *are* used?

$$\Pr(\mathbf{h}, \mathbf{e}) = \sum_{\mathbf{R}} \prod_{V} \theta(v|\boldsymbol{\pi}) \qquad \Big| \qquad \Pr(\mathbf{e}) = \sum_{\mathbf{H},\mathbf{R}} \prod_{V} \theta(v|\boldsymbol{\pi})$$

Computations include parameters $\theta(v \mid \boldsymbol{\pi})$ obeying

- in general: $v \sim \mathbf{her}$ and $\boldsymbol{\pi} \sim \mathbf{h}\cancel{\mathbf{e}}\mathbf{r}$
- !note: always $\exists \mathbf{r}(\mathbf{h}^*) : v \sim \mathbf{r}(\mathbf{h}^*)$ and $\boldsymbol{\pi} \sim \mathbf{r}(\mathbf{h}^*)$

Which parameters *are* used?

$$\Pr(\mathbf{h}, \mathbf{e}) = \sum_{\mathbf{R}} \prod_{V} \theta(v|\boldsymbol{\pi}) \qquad \Big| \qquad \Pr(\mathbf{e}) = \sum_{\mathbf{H},\mathbf{R}} \prod_{V} \theta(v|\boldsymbol{\pi})$$

Computations include parameters $\theta(v \mid \boldsymbol{\pi})$ obeying

- in general: $v \sim \mathbf{her}$ and $\boldsymbol{\pi} \sim \mathbf{h}\cancel{\mathbf{e}}\mathbf{r}$
- !note: always $\exists \mathbf{r}(\mathbf{h}^*) : v \sim \mathbf{r}(\mathbf{h}^*)$ and $\boldsymbol{\pi} \sim \mathbf{r}(\mathbf{h}^*)$
- more specifically:

$\Pr(\mathbf{h}, \mathbf{e})$	$\Pr(\mathbf{e})$
$V \in \mathbf{E}$: $v \sim \mathbf{e}$ and $\boldsymbol{\pi} \sim \mathbf{h}$	$V \in \mathbf{E}$: $v \sim \mathbf{e}$
$V \in \mathbf{H}$: $v \sim \mathbf{h}$ and $\boldsymbol{\pi} \sim \mathbf{h}$	$V \in \mathbf{H}$:
$V \in \mathbf{R}$: $\boldsymbol{\pi} \sim \mathbf{h}$	$V \in \mathbf{R}$:

Which parameters *are* used?

$$\Pr(\mathbf{h}, \mathbf{e}) = \sum_{\mathbf{R}} \prod_{V} \theta(v|\boldsymbol{\pi}) \qquad \Big| \qquad \Pr(\mathbf{e}) = \sum_{\mathbf{H},\mathbf{R}} \prod_{V} \theta(v|\boldsymbol{\pi})$$

Computations include parameters $\theta(v \mid \boldsymbol{\pi})$ obeying

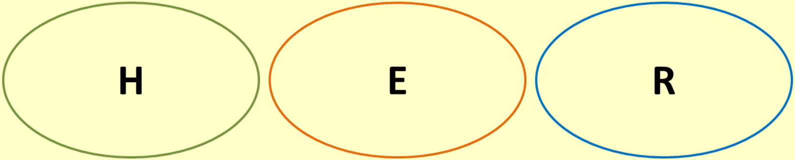
- in general: $v \sim \mathbf{her}$ and $\boldsymbol{\pi} \sim \mathbf{h}\cancel{\mathbf{e}}\mathbf{r}$
- !note: always $\exists \mathbf{r}(\mathbf{h}^*) : v \sim \mathbf{r}(\mathbf{h}^*)$ and $\boldsymbol{\pi} \sim \mathbf{r}(\mathbf{h}^*)$
- more specifically:

$V \in \mathbf{E}$: $v \sim \mathbf{e}$ and $\boldsymbol{\pi} \sim \mathbf{h}$	$V \in \mathbf{E}$: $v \sim \mathbf{e}$
$V \in \mathbf{H}$: $v \sim \mathbf{h}$ and $\boldsymbol{\pi} \sim \mathbf{h}$	$V \in \mathbf{H}$:
$V \in \mathbf{R}$: $\boldsymbol{\pi} \sim \mathbf{h}$	$V \in \mathbf{R}$:

Roughly speaking:
more constraints $\rightarrow$ more knowledge of $f_{\Pr(\mathbf{h}|\mathbf{e})}(\theta)$

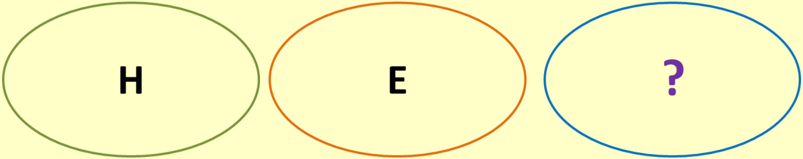
Qualitative effects of parameters

Consider $\Pr(\mathbf{h}|\mathbf{e})$ and all variables V in $\mathcal{B}_q$:



Qualitative effects of parameters

Consider $\Pr(\mathbf{h} | \mathbf{e})$ and all variables V in $\mathcal{B}_q$:

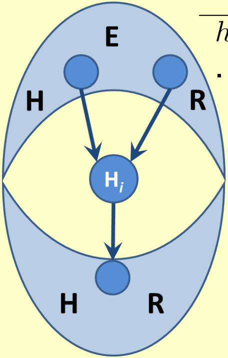


For parameters of $V \in \mathbf{R}$ the qualitative effect is *unknown*

($\triangleright$ i.e. cannot be established from merely structural properties)

Parameters of variables in $\mathbf{H}$ (I)

Consider $\Pr(\mathbf{h}|\mathbf{e})$ and $\theta(v|\boldsymbol{\pi})$ for $V = H_i \in \mathbf{H}$



	π_1	$\ldots$	π_m
h_1	$+$		$*$
$\ldots$		$*$	

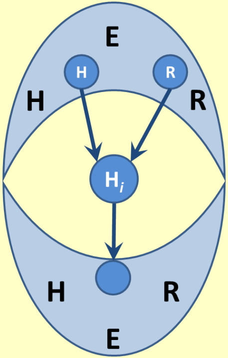
Assume $de(H_i) \cap \mathbf{E} = \emptyset$

- H_i's parameters are not used for computing $\Pr(\mathbf{e})$
- if $v \sim \mathbf{h}$ and $\boldsymbol{\pi} \sim \mathbf{h}$ $\rightarrow$ category $+$[1]
- otherwise, H_i's parameters are not used for computing $\Pr(\mathbf{h}, \mathbf{e})$ either $\rightarrow$ category $*$

[1] Proof: sign of first derivative of sensitivity function under given constraints

Parameters of variables in $\mathbb{H}$ (II)

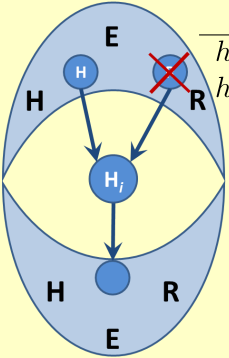
Consider $\Pr(\mathbf{h}|\mathbf{e})$ and $\theta(v|\boldsymbol{\pi})$ for $V = H_i \in \mathbf{H}$



Assume H_i is binary-valued

Parameters of variables in $\mathbb{H}$ (II)

Consider $\Pr(\mathbf{h}|\mathbf{e})$ and $\theta(v|\boldsymbol{\pi})$ for $V = H_i \in \mathbf{H}$



	$\pi_1 \ldots \pi_m$	
h_1	$+$	$?$
h_2	$-$	$?$

Assume H_i is binary-valued

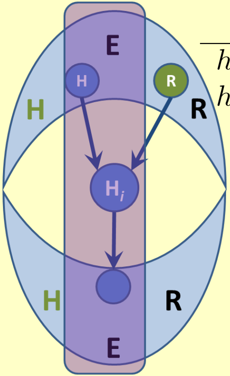
If $\boldsymbol{\pi} \sim \mathbf{h}$ then

- if $pa(H_i) \cap \mathbf{R} = \emptyset$ [ECAI 2016]
  - $\triangleright$ $v \sim \mathbf{h} \rightarrow$ category $+$
  - $\triangleright$ $v \nsim \mathbf{h} \rightarrow$ category $-$

Otherwise: category $?$

Parameters of variables in $\mathbb{H}$ (II)

Consider $\Pr(\mathbf{h}|\mathbf{e})$ and $\theta(v|\boldsymbol{\pi})$ for $V = H_i \in \mathbf{H}$



	$\pi_1 \ldots \pi_m$	
h_1	$+$	$?$
h_2	$-$	$?$

Assume H_i is binary-valued

If $\boldsymbol{\pi} \sim \mathbf{h}$ then

- if $pa(H_i) \cap \mathbf{R}$ d-separated[2]
  - $\triangleright$ $v \sim \mathbf{h} \rightarrow$ category $+$
  - $\triangleright$ $v \nsim \mathbf{h} \rightarrow$ category $-$

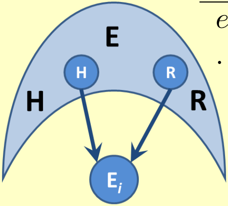
Otherwise: category $?$

[2] $pa(H_i) \cap \mathbf{R}$ and $\mathbf{H} \setminus (\{H_i\} \cup pa(H_i))$ d-sep given $\mathbf{E} \cup ((\{H_i\} \cup pa(H_i)) \cap \mathbf{H})$

Parameters of variables in $\mathbf{E}$

Consider $\Pr(\mathbf{h}|\mathbf{e})$ and $\theta(v|\boldsymbol{\pi})$ for $V = E_i \in \mathbf{E}$

- If $v \nsim \mathbf{e}$ then E_i's parameters are not used for computing $\Pr(\mathbf{h}, \mathbf{e})$ or $\Pr(\mathbf{e}) \rightarrow$ category $*$ [ECAI 2016]
- If $v \sim \mathbf{e}$ and $\boldsymbol{\pi} \nsim \mathbf{h}$ then E_i's parameters are only used for computing $\Pr(e) \rightarrow$ category $-$ [ECAI 2016]

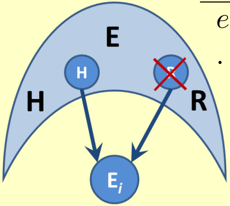


	π_1	$\dots$	π_m
e_1	$+$		$-$
$\dots$		$*$	

Parameters of variables in $\mathbf{E}$

Consider $\Pr(\mathbf{h}|\mathbf{e})$ and $\theta(v|\boldsymbol{\pi})$ for $V = E_i \in \mathbf{E}$

- If $v \nsim \mathbf{e}$ then E_i's parameters are not used for computing $\Pr(\mathbf{h}, \mathbf{e})$ or $\Pr(\mathbf{e}) \rightarrow$ category $*$
- If $v \sim \mathbf{e}$ and $\boldsymbol{\pi} \nsim \mathbf{h}$ then E_i's parameters are only used for computing $\Pr(e) \rightarrow$ category $-$



	$\pi_1 \ldots \pi_m$	
e_1	$+$	$-$
$\cdots$	$*$	

If $v \sim \mathbf{e}$ and $\boldsymbol{\pi} \sim \mathbf{h}$ then

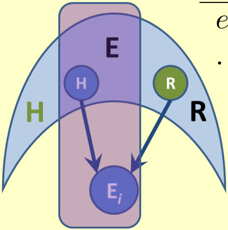
- if $pa(E_i) \cap \mathbf{R} = \emptyset$ [ECAI 2016] $\rightarrow$ category $+$

Otherwise: category $?$

Parameters of variables in $\mathbb{E}$

Consider $\Pr(\mathbf{h}|\mathbf{e})$ and $\theta(v|\boldsymbol{\pi})$ for $V = E_i \in \mathbf{E}$

- If $v \nsim \mathbf{e}$ then E_i's parameters are not used for computing $\Pr(\mathbf{h}, \mathbf{e})$ or $\Pr(\mathbf{e})$ → category $*$
- If $v \sim \mathbf{e}$ and $\boldsymbol{\pi} \nsim \mathbf{h}$ then E_i's parameters are only used for computing $\Pr(e)$ → category $-$



	$\pi_1 \ldots \pi_m$	
e_1	$+$	$-$
$\ldots$	$*$	

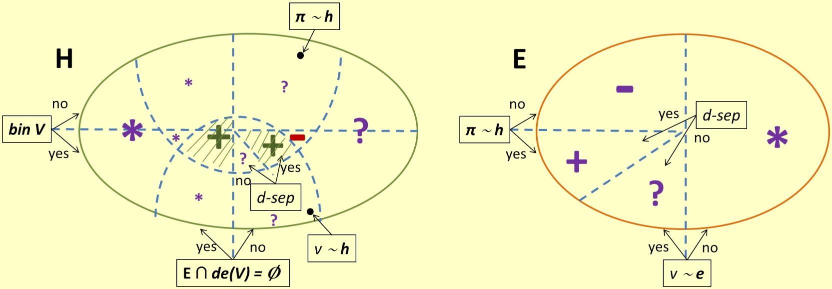
If $v \sim \mathbf{e}$ and $\boldsymbol{\pi} \sim \mathbf{h}$ then

- if $pa(E_i) \cap \mathbf{R}$ d-separated[3] → category $+$

Otherwise: category $?$

[3] $pa(E_i) \cap \mathbf{R}$ and $\mathbf{H} \setminus pa(H_i)$ d-sep given $\mathbf{E} \cup (pa(H_i) \cap \mathbf{H})$

Summary and conclusions



Recall: qualitative effect on $\Pr(\mathbf{h}|\mathbf{e})$ for choice of $\theta(v|pa(V))$ is exploitable knowledge

- all network parameters are now categorised
- more parameters have a meaningful category ($\neq ?$)
- category $?$ leaves some room for further study