Generalised Co-variation for Sensitivity Analysis in Bayesian Networks

Silja Renooij

Universiteit Utrecht

Presentation for PGM 2012

Contents

- Sensitivity analysis
  - ▷ in general
  - ▷ in Bayesian networks
- Co-variation
  - ▷ what, why and how?
- Examples
- Summary of contributions

Sensitivity analysis in Bayesian networks

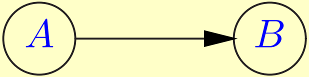
- Sensitivity analysis: a standard technique for studying effect of changes in model parameters on model output
- in Bayesian Networks: output probabilities are simple, multi-linear functions of network parameters (CPT entries)

Example:
A probability $\Pr(v)$ as a function of 2 network parameters x_1, x_2:

$$f_{\Pr(v)}(x_1, x_2) \; = \; c^{11} \cdot x_1 \cdot x_2 + c^{10} \cdot x_1 + c^{01} \cdot x_2 + c^{00}$$

- posterior $\rightarrow$ quotient
- assumption: (proportional) co-variation of other entries from same distribution

Co-variation in 1-way analysis



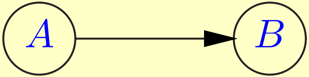
Varying a single parameter for a binary-valued variable:

	a_1	a_2
b_1	0.8	0.4
b_2	0.2	0.6

$\Longrightarrow$

	a_1	a_2
b_1	x	0.4
b_2	$1 - x$	0.6

Co-variation in 1-way analysis



Varying a single parameter for a binary-valued variable:

	a_1	a_2
b_1	0.8	0.4
b_2	0.2	0.6

$\Longrightarrow$

	a_1	a_2
b_1	x	0.4
b_2	$1-x$	0.6

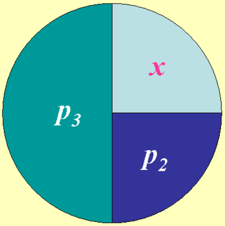
Varying a single parameter for a multi-valued variable:

	a_1	a_2
b_1	0.5	0.1
b_2	0.2	0.5
b_3	0.3	0.4

$\Longrightarrow$

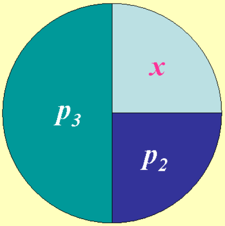
<table>
<tr><th></th><th>a_1</th><th>a_2</th></tr>
<tr><td>b_1</td><td>x</td><td>0.1</td></tr>
<tr><td>b_2</td><td rowspan="2">$\}\ 1-x$</td><td>0.5</td></tr>
<tr><td>b_3</td><td>0.4</td></tr>
</table>

Proportional co-variation



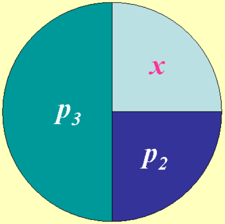
		$1 - x$
p_2	$= 0.25$	0.75
p_3	$= 0.50$	0.75

Proportional co-variation

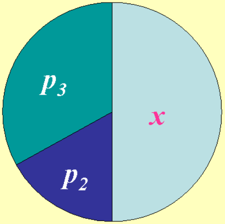


$$
\begin{aligned}
& \qquad\qquad\qquad\qquad 1-x \\
p_2 &= 0.25 = 0.33 \cdot 0.75 \\
p_3 &= 0.50 = 0.67 \cdot 0.75
\end{aligned}
$$

Proportional co-variation

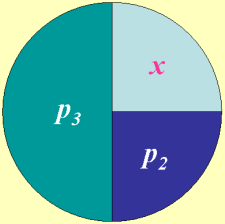


$$\begin{aligned} & \qquad\qquad\qquad\quad 1-x \\ p_2 &= 0.25 = 0.33 \cdot 0.75 \\ p_3 &= 0.50 = 0.67 \cdot 0.75 \end{aligned}$$

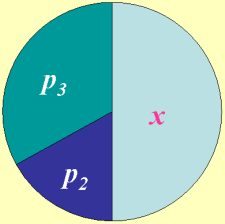


$$\begin{aligned} & \qquad\qquad 1-x \\ p_2 &= 0.33 \cdot 0.50 \\ p_3 &= 0.67 \cdot 0.50 \end{aligned}$$

Proportional co-variation



$$\begin{aligned} p_2 &= 0.25 = \underbrace{0.33}_{} \cdot \overbrace{0.75}^{1-x} \\ p_3 &= 0.50 = 0.67 \cdot 0.75 \end{aligned}$$



$$\begin{aligned} p_2 &= 0.17 = 0.33 \cdot \overbrace{0.50}^{1-x} \\ p_3 &= 0.33 = 0.67 \cdot 0.50 \end{aligned}$$

Motivation I

Why use proportional co-variation?

- standard approach
- assumed by sensitivity functions, algorithms & properties
- seems sensible
- works with any parameter
- optimal

Motivation I

Why use proportional co-variation?

- standard approach
- assumed by sensitivity functions, algorithms & properties
- seems sensible
- works with any parameter
- optimal:

> The CD-distance between the original distribution Pr and the new distribution Pr*
>
> $$D(\text{Pr}, \text{Pr}^*) = \ln \max_w \frac{\text{Pr}^*(w)}{\text{Pr}(w)} - \ln \min_w \frac{\text{Pr}^*(w)}{\text{Pr}(w)}$$
>
> is smallest under a proportional co-variation scheme [Chan & Darwiche, 2002]

- ...

Motivation II

Why use an alternative co-variation scheme?

- is standard most appropriate?
- do functions, algorithms & properties depend on scheme?
- is CD-distance really optimal?

Motivation II

Why use an alternative co-variation scheme?

- is standard most appropriate?
- do functions, algorithms & properties depend on scheme?
- is CD-distance really optimal?

this was only proven for single parameter changes !

$n > 1$ simultaneous parameter changes can result in a smaller CD-distance [Chan & Darwiche, 2004]

- again smallest under proportional co-variation ?
- this is unknown, and not obvious. . .

Motivation II

Why use an alternative co-variation scheme?

- is standard most appropriate?
- do functions, algorithms & properties depend on scheme?
- is CD-distance really optimal?

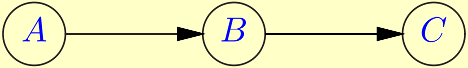
this was only proven for single parameter changes !

$n > 1$ simultaneous parameter changes can result in a smaller CD-distance [Chan & Darwiche, 2004]

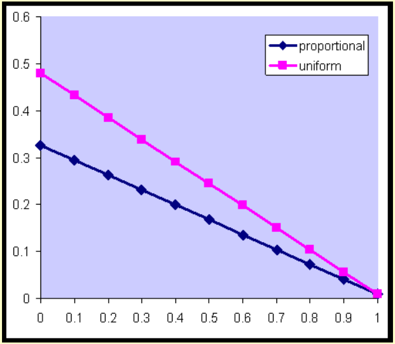
- ▷ again smallest under proportional co-variation ?
- ▷ this is unknown, and not obvious. . .

- who cares about CD-distance? 
- why *minimise 'disturbance'* in a sensitivity analysis?
- . . .

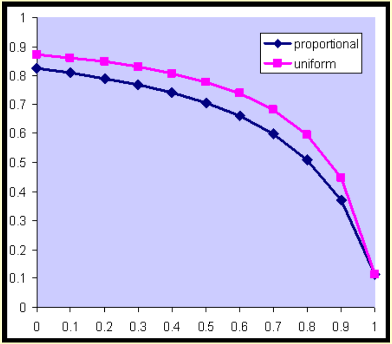
Examples



Parameter $x = p(b_1 \mid a)$ with $x_0 = 0.2$ is varied in steps of 0.1(!)

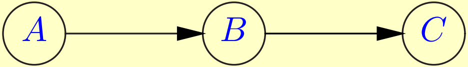


Output of interest:
Pr(a, c) with $p_0 = 0.26$

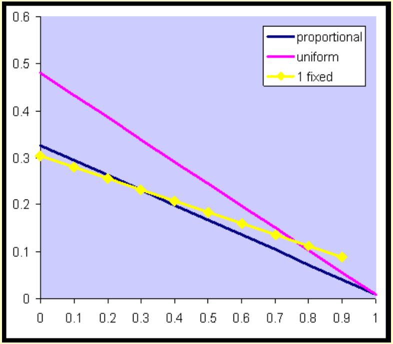


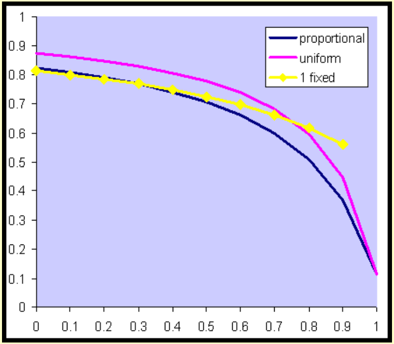
Output of interest:
Pr$(a \mid c)$ with $p_0 = 0.79$

Examples



Parameter $x = p(b_1 \mid a)$ with $x_0 = 0.2$ is varied in steps of 0.1(!)

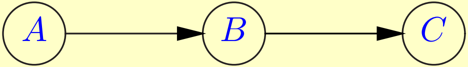




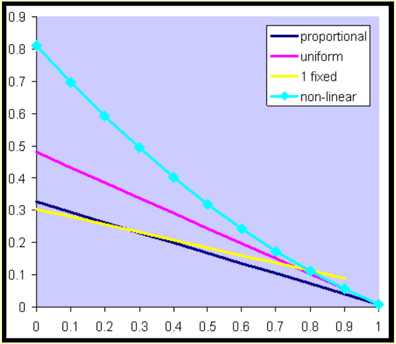
Output of interest:
$\Pr(a, c)$ with $p_0 = 0.26$

Output of interest:
$\Pr(a \mid c)$ with $p_0 = 0.79$

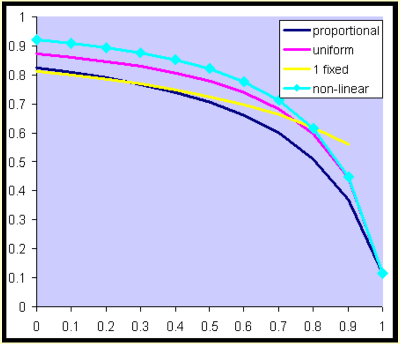
Examples



Parameter $x = p(b_1 \mid a)$ with $x_0 = 0.2$ is varied in steps of 0.1(!)



Output of interest:
$\Pr(a, c)$ with $p_0 = 0.26$



Output of interest:
$\Pr(a \mid c)$ with $p_0 = 0.79$

Contributions

- We provide generalised formulas for the sensitivity function
  - ▷ which explicitly incorporate the co-variation scheme
  - ▷ for both 1-way and n-way functions
  - ▷ also when parameters are fixed
  - ▷ standard form is preserved for co-variation schemes linear in x
- We provide generalised formulas for the CD-distance
  - ▷ which show that optimality of the proportional scheme is not obvious in the n-way case
  - ▷ and prove a lowerbound for single CPT co-variation under the proportional scheme