Universiteit Utrecht

Silja Renooij

Decision Support Systems
Department of Information and Computing Sciences
Utrecht University, The Netherlands
Generalised Co-variation for Sensitivity Analysis in Bayesian Networks
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:

\[
\begin{array}{c|cc}
 b_1 & a_1 & a_2 \\
 0.8 & 0.4 \\
 b_2 & 0.2 & 0.6 \\
\end{array}
\quad \Rightarrow \quad
\begin{array}{c|cc}
 b_1 & a_1 & a_2 \\
 x & 0.4 \\
 b_2 & 1 - x & 0.6 \\
\end{array}
\]

Varying a single parameter for a multi-valued variable:

\[
\begin{array}{c|cc}
 b_1 & a_1 & a_2 \\
 0.5 & 0.1 \\
 b_2 & 0.2 & 0.5 \\
 b_3 & 0.3 & 0.4 \\
\end{array}
\quad \Rightarrow \quad
\begin{array}{c|cc}
 b_1 & a_1 & a_2 \\
 x & 0.1 \\
 b_2 & 1 - x & 0.5 \\
 b_3 & \}
\end{array}
\]
Proportional co-variation

\[ p_2 = 0.25 = 0.33 \cdot 0.75 \]
\[ p_3 = 0.50 = 0.67 \cdot 0.75 \]

\[ p_2 = 0.17 = 0.33 \cdot 0.50 \]
\[ p_3 = 0.33 = 0.67 \cdot 0.50 \]
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 \( \Pr^* \)

$$D(\Pr, \Pr^*) = \ln \max_w \frac{\Pr^*(w)}{\Pr(w)} - \ln \min_w \frac{\Pr^*(w)}{\Pr(w)}$$

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?

*this was only proven for single parameter changes!*

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

- 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 $x = p(b_1 | a)$ 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:
$Pr(a | c)$ with $p_0 = 0.79$
Conclusions from examples

Consider a 1-way sensitivity function $f(x)$:

- alternatives to proportional co-variation don’t necessarily ensure that $f(x_0) = p_0$
- $f(1)$, if defined, is independent of the co-variation scheme
- for linear $f(x)$: largest effect of alternative schemes is found for small values of $x$;
- for hyperbolic $f(x)$: this is not necessarily the case
- fixing parameters preserves the standard form of $f(x)$, but constrains its domain

\(^1\) (i.e. preventing a co-varying parameter from varying)
Conclusions and contributions

Consider a 1-way sensitivity function $f(x)$:

- the standard form of the sensitivity function is preserved for co-variation schemes linear in $x$:

| $\Pr(w)$ as a function of $x = \theta_{v_1|u}$ of $t$-valued variable $V$ : |
| --- |
| $f_w(x) = (\alpha - \beta \gamma) \cdot x + (\beta \gamma + \delta)$ |
| $\alpha = \Pr(w|v_1, u) \cdot \Pr(u)$, $\beta \gamma = \sum_{k=2}^{t} \Pr(w|v_k, u) \cdot \Pr(u) \cdot \gamma_{v_k|u}$, $\delta = \Pr(w, \bar{u})$, and $\gamma_{v_k|u} \cdot (1 - x)$, for all $\theta_{v_k|u}$, $1 < k \leq t$, is a co-variation scheme. |

- this result extends to $n$-way functions

- constants can be computed with existing algorithms that construct and solve systems of equations.
Further contributions

Consider the CD-distance $D$:

- the **generalised distance** between old CPT $\Theta_V|U$ and new CPT $\Theta_V^*|U$

(for each $u_j$ we change parameter $\theta_{v_1|u_j}$ and co-vary all other $\theta_{v_k|u_j}$):

\[
D_\gamma(\Theta_V|U, \Theta_V^*|U) = \ln \max_{u_j} \left\{ \frac{\theta_{v_1|u_j}^*}{\theta_{v_1|u_j}}, \frac{1 - \theta_{v_1|u_j}^* \cdot \gamma_{v_k|u_j}^{-1} \cdot \theta_{v_k|u_j}}{\min_{k \neq 1} \gamma_{v_k|u_j}^{-1} \cdot \theta_{v_k|u_j}} \right\}
- \ln \min_{u_j} \left\{ \frac{\theta_{v_1|u_j}^*}{\theta_{v_1|u_j}}, \frac{1 - \theta_{v_1|u_j}^* \cdot \gamma_{v_k|u_j}^{-1} \cdot \theta_{v_k|u_j}}{\max_{k \neq 1} \gamma_{v_k|u_j}^{-1} \cdot \theta_{v_k|u_j}} \right\}
\]

where $\gamma_{v_k|u_j} \cdot (1 - x)$ is a co-variation scheme.

- similar result for single parameter variation in $\Theta_V|u$
- is this optimal for $\gamma = \text{‘proportional’}$?
Further contributions and future

Consider the CD-distance $D$

- the lowerbound for $n$-way, single CPT co-variation under the proportional scheme:

$$D_p(\Theta_V|U, \Theta^*_V|U) \geq \max_{u_j} \left| \ln \frac{\theta^*_v|u_j}{\theta_v|u_j} - \ln \frac{1 - \theta^*_v|u_j}{1 - \theta_v|u_j} \right|$$

- NB Chan (2005) introduced this expression as an approximation of the true distance

- What if we find an alternative co-variation scheme with

$$D_\gamma(\Theta_V|U, \Theta^*_V|U) \leq \max_{u_j} \left| \ln \frac{\theta^*_v|u_j}{\theta_v|u_j} - \ln \frac{1 - \theta^*_v|u_j}{1 - \theta_v|u_j} \right|$$