Generalised Co-variation for Sensitivity Analysis in Bayesian Networks

Silja Renooij
Utrecht University, The Netherlands
S.Renooij@uu.nl

Abstract

Upon varying parameters in a sensitivity analysis of a Bayesian network, the standard approach is to co-vary the parameters from the same conditional distribution such that their proportions remain the same. Alternative co-variation schemes are, however, possible. We theoretically investigate the effects of using alternative co-variation schemes on the so-called sensitivity function, and conclude that its general form remains the same under any linear co-variation scheme. In addition, we generalise the CD-distance for bounding global belief change, and prove a tight lower bound on this distance for parameter changes in single conditional probability tables.

1 Introduction

Sensitivity analysis is a general technique for studying the effects of parameter changes on the output of a mathematical model. In the context of Bayesian networks the effect of changes, applied to one or more probabilities from the network's conditional probability tables, on computed probabilities is determined. The results can be captured in detail by means of a *sensitivity function*, describing an output probability of interest in terms of one or more parameter probabilities. More global effects can be described by the *CD-distance*, which is a measure for bounding probabilistic belief change and complements the sensitivity function by giving insight in the effect of parameter changes on the global joint distribution, rather than on a specific (posterior) output probability of interest.

Upon varying a probability from a conditional distribution, the remaining probabilities from the same distribution need to be co-varied. The *proportional scheme* has been adopted as the standard scheme for co-variation in Bayesian networks, and various sensitivity analysis algorithms build upon this scheme. The proportional co-variation scheme is one of numerous alternatives for co-varying parameters from the same distribution. The mere fact that it is the standard co-variation scheme used, does not imply that there are no situations in which alternative schemes are suitable. However, the known standard form of the sensitivity function is based on proportional co-variation, and the proportional scheme is known to be optimal when varying a single parameter, in the sense that it minimises the CD-distance (Chan & Darwiche, 2005).

It is as of yet unknown if the proportional scheme is optimal when multiple, independent parameters from are varied. Moreover, we may not be interested in minimising the CD-distance: for example, we may be interested in minimising KL-divergence, which is not equivalent to minimising CD-distance (Chan & Darwiche, 2005); or we may want to perform our analyses in the context of large disturbances, rather than minimal ones. In this paper we will therefore investigate exactly how both the sensitivity function and the CD-distance depend on the co-variation scheme used. We show that the general form of the sensitivity function is maintained as long as the co-variation scheme is linear in the parameter(s) varied. In addition, we generalise the CD-distance to arbitrary co-variation schemes, and prove that a previously suggested approximation of this distance is in fact a lower bound.

This paper is organised as follows. Section 2 provides preliminaries on sensitivity analysis and co-variation. Section 3 generalises the sensitivity function to arbitrary co-variation schemes; the consequences for computing the functions are discussed in Section 4. Section 5 likewise generalises the CD-distance. The paper ends with conclusions and directions for future research in Section 6.

2 Preliminaries

A Bayesian network compactly represents a joint probability distribution $\Pr$ over a set of stochastic variables $\mathbf{V}$ (Jensen & Nielsen, 2007). It combines an acyclic directed graph G, that captures the variables and their dependencies as nodes and arcs respectively, with conditional probability distributions $\Theta_{V_i|\pi(V_i)}$ for each variable V_i and its parents $\pi(V_i)$ in the graph, such that $\Pr(\mathbf{V}) = \prod_i \Theta_{V_i|\pi(V_i)}$. We will refer to $\Theta_{V_i|\pi(V_i)}$ as the conditional probability table (CPT) of V_i; entries θ of Θ are called parameter probabilities, or parameters for short. Variables are denoted by capital letters and their values or instantiations by lower case; bold face is used for sets.

Probabilities computed from a Bayesian network are affected by the inaccuracies in the network's parameters. To investigate the extent of these effects, a sensitivity analysis can be performed in which $n \geq 1$ network parameters are varied simultaneously and the effect on output probabilities of interest are studied. The effects of such n*-way* parameter variation can be described by *sensitivity functions*. Such a function is *multilinear* in the varied parameters in case of a prior probability of interest, and *rational* (quotient of two multilinear functions) in the posterior case (Coupé & Van der Gaag, 2002). For example, the 1-way sensitivity function $f_a^{\mathbf{e}}(x)$ describing the posterior probability $\Pr(a \mid \mathbf{e})$ as a function of parameter x is given by

$$f_a^{\mathbf{e}}(x) = \frac{f_{a,\mathbf{e}}(x)}{f_{\mathbf{e}}(x)} = \frac{c_1 x + c_0}{d_1 x + d_0} \quad (1)$$

with constants c_i, d_i, $i = 0, 1$, built from non-varied network parameters. The general form of the sensitivity function was established under the assumption of *proportional* co-variation (Castillo, Gutiérrez & Hadi, 1997; Coupé & Van der Gaag, 2002).

Co-variation Consider a binary-valued variable V with values v and $\bar{v}$, and parent configuration $\mathbf{u}$. Since $\theta_{v|\mathbf{u}} + \theta_{\bar{v}|\mathbf{u}} = 1$, f, we have that if $\theta_{v|\mathbf{u}}$ varies, $\theta_{\bar{v}|\mathbf{u}}$ should be co-varied to ensure that their sum remains 1. Therefore, if $\theta_{v|\mathbf{u}}$ varies as x in a sensitivity analysis, $\theta_{\bar{v}|\mathbf{u}}$ should co-vary as $1 - x$.

Now if V has $t > 2$ values $v_1, \ldots, v_t$, and $x = \theta_{v_1|\mathbf{u}}$ is the parameter varied in a sensitivity analysis, then there are endless ways in which the parameters $\theta_{v_k|\mathbf{u}}$, $1 < k \leq t$ can co-vary with x. Using the above mentioned proportional co-variation scheme, the parameters $\theta_{v_k|\mathbf{u}}$, $k \neq 1$, get the same proportion of the remaining mass of $1 - x$, as they had originally:

$$\theta^*_{v_k|\mathbf{u}} = \frac{\theta_{v_k|\mathbf{u}}}{1 - \theta_{v_1|\mathbf{u}}} \cdot (1 - x)$$

where $\theta^*_{v_k|\mathbf{u}}$ is the new value of the parameter, and $\theta_{v_k|\mathbf{u}}$, $\theta_{v_1|\mathbf{u}}$ indicate the original values that were specified in the network. We typically assume that parameters with an original value of 0 or 1 are not varied, so the above denominator is in $\langle 0, 1 \rangle$.

The proportional scheme has been adopted as the standard scheme for co-variation in Bayesian networks, and various sensitivity analysis algorithms build upon this scheme (Chan & Darwiche (2002; 2004; 2005), Coupé & Van der Gaag (2002), Kjærulff & Van der Gaag (2000)). In fact, the proportional scheme minimises the CD-distance between the new distribution $\Pr^*$ and the original distribution $\Pr$ (Chan & Darwiche, 2005).

3 Co-variation in the Sensitivity Function

The proportional co-variation scheme, although standard, is merely one of many alternatives for co-varying parameters from the same distribution. In this section, we will take a fresh look at sensitivity functions without restricting ourselves to a particular co-variation scheme.

3.1 Co-variation schemes

Consider a t-valued variable V and suppose we vary a parameter from the distribution $\Theta_{V|\mathbf{u}}$ as x. The $t-1$ remaining parameters from this distribution must co-vary; more specifically, each of these parameters should get a portion, or cut, of the remaining mass $1 - x$. We define a valid co-variation scheme based on these cuts.

Definition 1. Consider $k \geq 1$ parameters θ_k from the same distribution and let $m \leq 1$ be the total probability mass available for these parameters. A *co-variation cut* $\gamma : \{\theta_k\} \rightarrow [0, 1]$ defines the share of each parameter θ_k in m. A *co-variation scheme* $s(k) = \gamma(k) \cdot m$ now maps each θ_k to a new value θ^*_k. A co-variation scheme is called *valid* if $\sum_k \gamma(k) = 1$.

From here on we will write γ_k rather than $\gamma(k)$. Note that a valid co-variation scheme ensures that the entire distribution under consideration sums to $(1-m) + \sum_k \gamma_k \cdot m = 1$. The following example gives two valid co-variation schemes.

Example 1. The standard proportional co-variation scheme,

$$\gamma_{v_k|\mathbf{u}} \cdot (1-x) = \frac{\theta_{v_k|\mathbf{u}}}{1-\theta_{v_1|\mathbf{u}}} \cdot (1-x) \tag{2}$$

is indeed valid:

$$\sum_{k \neq 1} \gamma_{v_k|\mathbf{u}} = \sum_{k=2}^{t} \frac{\theta_{v_k|\mathbf{u}}}{1-\theta_{v_1|\mathbf{u}}} = \frac{1-\theta_{v_1|\mathbf{u}}}{1-\theta_{v_1|\mathbf{u}}} = 1$$

We can also think of other valid co-variation schemes. Consider for example a *uniform* co-variation scheme:

$$\gamma_{v_k|\mathbf{u}} \cdot (1-x) = \frac{1}{t-1} \cdot (1-x) \tag{3}$$

which uniformly distributes the remaining mass of $1-x$ over the $t-1$ co-varying parameters. This scheme is also valid:

$$\sum_{k \neq 1} \gamma_{v_k|\mathbf{u}} = \sum_{k=2}^{t} \frac{1}{t-1} = (t-1) \cdot \frac{1}{1-t} = 1$$

■

3.2 The generalised 1-way form

In the remainder of this paper we focus on a probability of interest $\Pr(\mathbf{w})$[1] as a function of a parameter $\theta_{v_1|\mathbf{u}}$ of a t-valued variable V. In addition, without loss of generality, we often assume V to have a single, binary-valued parent U with values u and $\bar{u}$.

The following proposition explicitly captures how the general form of the 1-way sensitivity function depends on the co-variation scheme.

Proposition 1. *Probability* $\Pr(\mathbf{w})$ *as a function of* $x = \theta_{v_1|u}$ *of* t*-valued variable* V *is given by*

$$f_{\mathbf{w}}(x) = (\alpha - \beta^{\gamma}) \cdot x + (\beta^{\gamma} + \delta)$$

where

$$\alpha = \Pr(\mathbf{w}|v_1, u) \cdot \Pr(u), \; \delta = \Pr(\mathbf{w}, \bar{u}),$$
$$\beta^{\gamma} = \sum_{k=2}^{t} \Pr(\mathbf{w}|v_k, u) \cdot \Pr(u) \cdot \gamma_{v_k|u},$$

and $\gamma_{v_k|u} \cdot (1-x)$*, for all* $\theta_{v_k|u}$*,* $1 < k \leq t$*, is a co-variation scheme.*

Proof. First we rewrite $\Pr(\mathbf{w})$ to include the parameters under consideration:

$$\begin{aligned} \Pr(\mathbf{w}) &= \Pr(\mathbf{w}, \bar{u}) + \Pr(\mathbf{w}, u) \\ &= \Pr(\mathbf{w}, \bar{u}) + \Pr(\mathbf{w}|v_1, u) \cdot \Pr(u) \cdot \theta_{v_1|u} \\ &\quad + \sum_{k=2}^{t} \Pr(\mathbf{w}|v_k, u) \cdot \Pr(u) \cdot \theta_{v_k|u} \end{aligned}$$

The proposition now follows by replacing $\theta_{v_1|u}$ by x and each $\theta_{v_k|u}$ by its value according to the co-variation scheme. □

3.3 The generalised n-way form

In this section we consider the explicit simultaneous variation of $n > 1$ parameters from either a single distribution, or from an entire CPT. Taking parameters from multiple CPTs is also possible (Chan & Darwiche, 2004; Kjærulff & Van der Gaag, 2000). Although our results extend to this latter case, it will not be considered here: the approach is not often used in practice since it quickly becomes computationally infeasible.

The following proposition explicitly captures how the general form of the n-way sensitivity function for n parameters from a single distribution $\Theta_{V|u}$, depends on the co-variation scheme.

Proposition 2. *Probability* $\Pr(\mathbf{w})$ *as a function of* n *parameters* $x_i = \theta_{v_i|u}$*,* $1 \leq i \leq n$*, of* t*-valued variable* V*,* $t > n$*, is given by*

$$f_{\mathbf{w}}(x_1, \ldots, x_n) = \sum_{i=1}^{n} (\alpha_i - \beta^{\gamma}) \cdot x_i + (\beta^{\gamma} + \delta)$$

where

$$\alpha_i = \Pr(\mathbf{w}|v_i, u) \cdot \Pr(u), \; \delta = \Pr(\mathbf{w}, \bar{u}),$$
$$\beta^{\gamma} = \sum_{k=n+1}^{t} \Pr(\mathbf{w}|v_k, u) \cdot \Pr(u) \cdot \gamma_{v_k|u},$$

and $\gamma_{v_k|u} \cdot (1 - \sum_{i=1}^{n} x_i)$*, for all* $\theta_{v_k|u}$*,* $n < k \leq t$*, is a co-variation scheme.*

[1] Note from Equation 1 that $\Pr(\mathbf{w})$ is general enough to represent any probability of interest. For example, $\Pr(a \mid \mathbf{e}) = \Pr(a, \mathbf{e}) / \Pr(\mathbf{e})$, where both numerator and denominator are of the form $\Pr(\mathbf{w})$.

Proof. The proof is analogous to that of Proposition 1, with each $\theta_{v_i|u}$, $1 \leq i \leq n$, replaced by x_i and each $\theta_{v_k|u}$, $n < k \leq t$, replaced by its value according to the co-variation scheme. □

Note that in the above case we need to ensure that $\sum_{i=1}^{n} x_i \leq 1$, which means we have to impose an additional constraint on the sensitivity analysis. For this reason, an n-way analysis typically considers parameters from n different distributions, often constituting an entire CPT. That is, from a single CPT $\Theta_{V|\mathbf{U}}$, exactly one parameter for each conditional distribution $\Theta_{V|\mathbf{u}_j}$ is varied, and all other parameters are co-varied (Chan & Darwiche, 2004). For such single-CPT n-way analyses, the following proposition describes how the general form of the sensitivity function depends on the co-variation scheme.

Proposition 3. *Probability* $\Pr(\mathbf{w})$ *as a function of* n *parameters* $x_j = \theta_{v_1|\mathbf{u}_j}$, $1 \leq j \leq n$, *of* t*-valued variable* V*, is given by*

$$f_{\mathbf{w}}(x_1, \ldots, x_n) = \sum_{j=1}^{n} (\alpha_j - \beta_j^{\gamma}) \cdot x_j + \beta_j^{\gamma}$$

where

$$\begin{aligned} \alpha_j &= \Pr(\mathbf{w}|v_1,\, \mathbf{u}_j) \cdot \Pr(\mathbf{u}_j), \\ \beta_j^{\gamma} &= \sum_{k=2}^{t} \Pr(\mathbf{w}|v_k, \mathbf{u}_j) \cdot \Pr(\mathbf{u}_j) \cdot \gamma_{v_k|\mathbf{u}_j}, \end{aligned}$$

and $\gamma_{v_k|\mathbf{u}_j} \cdot (1 - x_j)$*, for all* $\theta_{v_k|\mathbf{u}_j}$, $1 < k \leq t$*, are co-variation schemes.*

Proof. The proof is analogous to that of Proposition 1, except that for each parent configuration $\mathbf{u}_j$, the term $\Pr(\mathbf{w}, \mathbf{u}_j)$ now depends on an x_j and its co-varying parameters. □

3.4 Fixing parameters

In the above, we assumed that we varied one or more parameters from a t-valued variable and let the remaining parameters co-vary in some way. We may, however, want one or more parameters to stick to their original value. For example, parameters with an original value of zero indicate an impossibility that should remain impossible.

We now consider the effect of fixing parameters on the sensitivity function. Without loss of generality we assume that upon varying parameter $\theta_{v_1|\mathbf{u}}$ of a t-valued variable V, $t \geq 3$, a single parameter $\theta_{v_t|\mathbf{u}}$ should remain unchanged.

Proposition 4. *Probability* $\Pr(\mathbf{w})$ *as a function of* $x = \theta_{v_1|u}$ *of* t*-valued variable* V, $t \geq 3$*, with parameter* $\theta_{v_t|u}$ *fixed, is given by*

$$f_{\mathbf{w}}(x) = (\alpha - \beta^{\gamma}) \cdot x + (\mu \cdot \beta^{\gamma} + \delta)$$

where

$$\begin{aligned} \alpha &= \Pr(\mathbf{w}|v_1,\, u) \cdot \Pr(u), \\ \delta &= \Pr(\mathbf{w}|v_t, u) \cdot \Pr(u) \cdot \theta_{v_t|u} + \Pr(\mathbf{w}, \bar{u}), \\ \beta^{\gamma} &= \sum_{k=2}^{t-1} \Pr(\mathbf{w}|v_k, u) \cdot \Pr(u) \cdot \gamma_{v_k|u}, \\ \mu &= 1 - \theta_{v_t|u} \ \textit{(mass for co-variation)}, \end{aligned}$$

and $\gamma_{v_k|\mathbf{u}} \cdot (\mu - x)$*, for all* $\theta_{v_k|\mathbf{u}}$, $1 < k < t$*, is a co-variation scheme.*

Proof. Reconsider the proof of Proposition 1. The proposition follows directly by taking into account that we no longer co-vary all $t-1$ remaining parameters, and that the co-varied parameters together only have a remaining mass of $(1 - \theta_{v_t|u}) - x = \mu - x$ to divide. □

Proposition 4 generalises to any fixed mass $1-\mu$, including the special case $\mu = 1$ where no parameters are fixed, as is the standard case. Note that fixing parameters with an original value of zero also gives $\mu = 1$, so we may have to exclude them explicitly from the co-variation. This is the case when using, for example, the uniform co-variation scheme. Using proportional co-variation, however, parameters with an original value of zero keep that value upon co-variation: $\frac{0}{1-\theta_{v_1|\mathbf{u}}} \cdot (1-x) = 0$. Proposition 4 also generalises to the n-way functions considered in Propositions 2 and 3.

3.5 Generalised versus standard functions

We note that in the propositions above, the term β^{γ} (or β_j^{γ}), sums over all the co-varying parameters. It is exactly this term that depends on the co-variation scheme used. The propositions therefore explicitly indicate which part of the sensitivity function depends on the co-variation scheme used. We now note that β^{γ} can be considered constant with respect to the varied parameter(s) x, only if $\gamma_{v_k|\mathbf{u}}$ is independent of x, i.e. if the co-variation scheme $\gamma_{v_k|\mathbf{u}} \cdot (1-x)$ is linear in x.

Corollary 1. *A sensitivity function $f_a^{\mathbf{e}}(x)$ is of the standard form (Equation* (1)*), iff the co-variation scheme used is valid and linear in x.*

Proof. The result follows directly from Definition 1 and Proposition 1. □

The above corollary generalises to n-way sensitivity functions, as well as to sensitivity functions with fixed parameters. From the corollary we have that, contrary to the usual assumption in literature, the general form of the sensitivity function does not necessarily require a proportional co-variation scheme: any valid scheme, linear in x will do.

Example 2. The uniform co-variation scheme (Equation (3)) is clearly linear in x. The standard proportional co-variation scheme (Equation (2)) is also linear in x, since $\gamma_{v_k|\mathbf{u}}$ only depends on the *original* value of $\theta_{v_1|\mathbf{u}}$. ■

4 Using Another Co-variation Scheme

In the previous section we concluded that sensitivity functions keep their standard form as long as a valid and linear co-variation scheme is used. We now investigate the implications of this for existing algorithms that establish sensitivity functions, and provide a preliminary comparison of such functions for alternative co-variation schemes.

4.1 Computing the constants

Algorithms that build upon the analytic expression of the constants of the sensitivity function for their computation, such as e.g. Kjærulff & Van der Gaag (2000), cannot be applied if we use a co-variation scheme that is different from proportional co-variation. This is simply because the co-variation scheme affects the analytic form of the constants, as we have demonstrated in the previous section.

Another class of algorithms constructs and solves a system of r equations, one for each required constant, by computing the output probability of interest for r different values of x (see e.g. Coupé & Van der Gaag (2002)). These algorithms do not depend on the analytic form of the constants, and can therefore be applied regardless of the co-variation scheme used.

A note on uniform co-variation As illustrated by the following example, uniform co-variation can be implemented by applying proportional co-variation to a pre-processed CPT. Hence, algorithms that build on the analytic form of the constants *can* be applied in the context of uniform co-variation.

Example 3. Consider parameters $\theta_{v_1|\mathbf{u}}$, $\theta_{v_2|\mathbf{u}}$ and $\theta_{v_3|\mathbf{u}}$ with values 0.3, 0.5 and 0.2, respectively. Suppose we vary $\theta_{v_1|\mathbf{u}}$ to 0.5 and co-vary the other two: under uniform co-variation this results in a value of 0.25 for both. The same result is achieved by using proportional co-variation, *after* first uniformly distributing $(1-\theta_{v_1|\mathbf{u}})$ over $\theta_{v_2|\mathbf{u}}$ and $\theta_{v_3|\mathbf{u}}$. This approach is valid, since equal parameter values remain equal under proportional co-variation. ■

4.2 Comparing co-variation schemes

In this section we provide a preliminary analysis of the impact of using different co-variation schemes. Example 3 points to a potential problem with alternative co-variation schemes, especially those independent of the original value of $\theta_{v_1|\mathbf{u}}$: if a parameter x, upon variation, takes on a value that corresponds to the original value x^0 of the parameter, then the co-varying parameters may take on different values from their original ones. As a result, for a probability of interest $\Pr(a|e)$ with original value p^0 we will find that $f_a^{\mathbf{e}}(x^0) \neq p^0$. This may seem counter-intuitive, but the sensitivity function describes the relation between an output probability and the value of parameter x, *in the context of the values for the co-varied parameters*. Upon choosing a uniform co-variation scheme, for example, $f(x^0)$ is computed assuming that the co-varying parameters are distributed uniformly over $1-x^0$, even if they originally weren't. Note that for the proportional co-variation scheme, we indeed have that $\gamma_{v_k|\mathbf{u}}\cdot(1-x^0) = \theta_{v_k|\mathbf{u}}\ \forall k$. From the above observations, we have that properties that assume the sensitivity function to include (x^0, p^0) may give corrupt results upon applying alternative co-variation schemes; this includes properties such as the sensitivity value and various bounds on the sensitivity function (Chan & Darwiche, 2002; Van der Gaag, Renooij & Coupé, 2007).

With respect to sensitivity functions which are defined for $x \in [0,1]$, i.e. none of the co-varying parameters are fixed to a non-zero value, we observe the following:

- if a parameter is varied to $x = 1$, there is no more mass left for co-variation, so $f_{\mathbf{w}}(1)$ is independent of the scheme used; this also holds for rational functions, and for n-way functions.
- for 1-way linear sensitivity functions we have that for any co-variation scheme which preserves p^0 for $x = x^0$, the function is independent of the co-variation scheme. This observation, however, does not necessarily hold for 1-way rational functions, which are uniquely defined by three points.
- the maximum difference in 1-way linear sensitivity functions, caused by using different co-variation schemes, is thus found for $x = 0$; whether the same can be said for 1-way rational functions requires further investigation.

From these observations we conjecture that considering alternative co-variation schemes may be most interesting for parameters with smaller values. In the context of parameter tuning, it could turn out that a 'single' parameter change under an alternative co-variation scheme can accomplish an effect that is not possible with a 'single' parameter change under proportional co-variation.

5 Co-variation and the CD-distance

The CD-distance measures the distance between two probability distributions $\Pr$ and $\Pr^*$. If $\Pr^*$ is the result of making changes in a single CPT $\Theta_{V|\mathbf{U}}$ in a Bayesian network, the CD-distance is given by (Chan & Darwiche, 2004):

$$D(\Theta_{V|\mathbf{U}}, \Theta^*_{V|\mathbf{U}}) = \ln \max_{v_i, \mathbf{u}_j} \frac{\theta^*_{v_i|\mathbf{u}_j}}{\theta_{v_i|\mathbf{u}_j}} - \ln \min_{v_i, \mathbf{u}_j} \frac{\theta^*_{v_i|\mathbf{u}_j}}{\theta_{v_i|\mathbf{u}_j}}$$

In this section we will analyse how the CD-distance depends on the co-variation scheme used.

5.1 Single parameter co-variation

Chan & Darwiche (2005) demonstrated that upon changing a single parameter $\theta_{v_1|\mathbf{u}}$ from a distribution $\Theta_{V|\mathbf{u}}$, the proportional co-variation scheme is optimal in the sense that it minimises the CD-distance D between the original distribution $\Theta_{V|\mathbf{u}}$ and the new distribution $\Theta^*_{V|\mathbf{u}}$. In addition, they showed that this distance has the following closed form:[2]

$$D_p(\Theta_{V|\mathbf{u}}, \Theta^*_{V|\mathbf{u}}) = \left| \ln \frac{\theta^*_{v_1|\mathbf{u}}}{\theta_{v_1|\mathbf{u}}} - \ln \frac{1-\theta^*_{v_1|\mathbf{u}}}{1-\theta_{v_1|\mathbf{u}}} \right| \quad (4)$$

Opting for a different co-variation scheme will therefore necessarily increase the distance between original and new distribution. It is unknown, however, how the CD-distance is exactly affected by the co-variation scheme used. The following proposition provides us with that information.

Proposition 5. *Consider changing a parameter $\theta_{v_1|\mathbf{u}}$ for a t-valued variable, $t \geq 2$, to $\theta^*_{v_1|\mathbf{u}}$. Let $\gamma_{v_k|\mathbf{u}} \cdot (1-\theta^*_{v_1|\mathbf{u}})$, $1 < k \leq t$, be the new values of the co-varied parameters. Then, $D_\gamma(\Theta_{V|\mathbf{u}}, \Theta^*_{V|\mathbf{u}}) =$*

$$= \ln \max \left\{ \frac{\theta^*_{v_1|\mathbf{u}}}{\theta_{v_1|\mathbf{u}}}, \frac{1-\theta^*_{v_1|\mathbf{u}}}{\min\limits_{1<k\leq t} \gamma^{-1}_{v_k|\mathbf{u}} \cdot \theta_{v_k|\mathbf{u}}} \right\} - \ln \min \left\{ \frac{\theta^*_{v_1|\mathbf{u}}}{\theta_{v_1|\mathbf{u}}}, \frac{1-\theta^*_{v_1|\mathbf{u}}}{\max\limits_{1<k\leq t} \gamma^{-1}_{v_k|\mathbf{u}} \cdot \theta_{v_k|\mathbf{u}}} \right\}$$

Proof. For the CD-distance we need to compute

$$\ln \max_{1\leq i\leq t} \frac{\theta^*_{v_i|\mathbf{u}}}{\theta_{v_i|\mathbf{u}}} - \ln \min_{1\leq i\leq t} \frac{\theta^*_{v_i|\mathbf{u}}}{\theta_{v_i|\mathbf{u}}}$$

We therefore consider all ratios $\theta^*_{v_i|\mathbf{u}}/\theta_{v_i|\mathbf{u}}$, $1 \leq i \leq t$ and determine their maximum and minimum, respectively. For $i \neq 1$, we have that

$$\frac{\theta^*_{v_i|\mathbf{u}}}{\theta_{v_i|\mathbf{u}}} = \frac{\gamma_{v_i|\mathbf{u}} \cdot (1-\theta^*_{v_1|\mathbf{u}})}{\theta_{v_i|\mathbf{u}}} = \frac{1-\theta^*_{v_1|\mathbf{u}}}{\gamma^{-1}_{v_i|\mathbf{u}} \cdot \theta_{v_i|\mathbf{u}}}$$

It is obvious that

$$\max_{1<k\leq t} \frac{1-\theta^*_{v_1|\mathbf{u}}}{\gamma^{-1}_{v_k|\mathbf{u}} \cdot \theta_{v_k|\mathbf{u}}} = \frac{1-\theta^*_{v_1|\mathbf{u}}}{\min\limits_{1<k\leq t} \gamma^{-1}_{v_k|\mathbf{u}} \cdot \theta_{v_k|\mathbf{u}}}$$

and

$$\min_{1<k\leq t} \frac{1-\theta^*_{v_1|\mathbf{u}}}{\gamma^{-1}_{v_k|\mathbf{u}} \cdot \theta_{v_k|\mathbf{u}}} = \frac{1-\theta^*_{v_1|\mathbf{u}}}{\max\limits_{1<k\leq t} \gamma^{-1}_{v_k|\mathbf{u}} \cdot \theta_{v_k|\mathbf{u}}}$$

[2] The subscript for D indicates the assumed co-variation scheme: γ is general, p is proportional, u is uniform.

Comparing the above minimum and maximum to the ratio $\theta^*_{v_1|\mathbf{u}}/\theta_{v_1|\mathbf{u}}$ for $i = 1$, we straightforwardly find that there exist values for $\theta_{v_k|\mathbf{u}}$ such that $\theta^*_{v_1|\mathbf{u}}/\theta_{v_1|\mathbf{u}}$ is the largest term, but there are also values for $\theta_{v_k|\mathbf{u}}$ such that $\theta^*_{v_1|\mathbf{u}}/\theta_{v_1|\mathbf{u}}$ is the smallest term (see example below). The result follows from these observations. □

The following example illustrates the various possibilities.

Example 4. Consider a distribution over a variable V with 3 values. Suppose we vary parameter $\theta_{v_1|\mathbf{u}}$ and co-vary the remaining two using the *uniform* co-variation scheme. Let the original assessments for v_1, v_2 and v_3 be 0.25, 0.7 and 0.05, respectively. Then $\gamma^{-1}_{v_2|\mathbf{u}} \cdot \theta_{v_2|\mathbf{u}} = (3-1)\cdot 0.7 = 1.4$ and $\gamma^{-1}_{v_3|\mathbf{u}} \cdot \theta_{v_3|\mathbf{u}} = (3-1)\cdot 0.05 = 0.1$.

Case 1 ($\max = \theta_{v_1|\mathbf{u}}/\theta^*_{v_1|\mathbf{u}}$)**:** we vary $\theta_{v_1|\mathbf{u}} = 0.25$ to $\theta^*_{v_1|\mathbf{u}} = 0.8$; then $\theta^*_{v_1|\mathbf{u}}/\theta_{v_1|\mathbf{u}} = 0.8/0.25 = 3.2$. Therefore $D_u(\Theta_{V|\mathbf{u}}, \Theta^*_{V|\mathbf{u}}) = \ln\max\{3.2, 0.2/0.1\} - \ln\min\{3.2, 0.2/1.4\} = 3.109$.

Case 2 ($\max \geq \theta_{v_1|\mathbf{u}}/\theta^*_{v_1|\mathbf{u}} \geq \min$)**:** we vary $\theta_{v_1|\mathbf{u}}$ to $\theta^*_{v_1|\mathbf{u}} = 0.5$; then $\theta^*_{v_1|\mathbf{u}}/\theta_{v_1|\mathbf{u}} = 0.5/0.25 = 2$. Therefore $D_u(\Theta_{V|\mathbf{u}}, \Theta^*_{V|\mathbf{u}}) = \ln\max\{2, 0.5/0.1\} - \ln\min\{2, 0.5/1.4\} = 2.639$.

Case 3 ($\min = \theta_{v_1|\mathbf{u}}/\theta^*_{v_1|\mathbf{u}}$)**:** we vary $\theta_{v_1|\mathbf{u}}$ to 0.10; then $\theta^*_{v_1|\mathbf{u}}/\theta_{v_1|\mathbf{u}} = 0.1/0.25 = 0.4$. Therefore $D_u(\Theta_{V|\mathbf{u}}, \Theta^*_{V|\mathbf{u}}) = \ln\max\{0.4, 0.9/0.1\} - \ln\min\{0.4, 0.9/1.4\} = 3.114$.

Note that if we use the *proportional* co-variation scheme, then $\gamma^{-1}_{v_2|\mathbf{u}} \cdot \theta_{v_2|\mathbf{u}} = (0.7/0.75)^{-1} \cdot 0.7 = 0.75$ and $\gamma^{-1}_{v_3|\mathbf{u}} \cdot \theta_{v_3|\mathbf{u}} = (0.05/0.75)^{-1} \cdot 0.05 = 0.75$, which indeed both equal $1 - \theta_{v_1|\mathbf{u}}$, as in Equation (4). For **case 1**, for example, we then get $D_p(\Theta_{V|\mathbf{u}}, \Theta^*_{V|\mathbf{u}}) = \ln\max\{3.2, 0.2/0.75\} - \ln\min\{3.2, 0.2/0.75\} = 2.485$ which is indeed less than the distance found with uniform co-variation. ■

Note from the above example that the minimisation and maximisation terms over the co-varied parameters depend on $\gamma_{v_k|\mathbf{u}}$, but not on $\theta^*_{v_1|\mathbf{u}}$; hence they can be pre-computed and used for computing the CD-distance for any new distribution $\Theta^*_{V|\mathbf{u}}$ obtained by changing parameter $\theta_{v_1|\mathbf{u}}$ to $\theta^*_{v_1|\mathbf{u}}$.

5.2 Single CPT co-variation

Chan & Darwiche (2004) erroneously introduced the following closed form for the CD-distance in case the parameters of an entire CPT $\Theta_{V|\mathbf{U}}$ are varied as described in Section 3.3 (just above Proposition 3), i.e. a single parameter from *each* $\Theta_{V|\mathbf{u}_j}$ is varied, while co-varying all others:

$$\max_{\mathbf{u}_j} \left| \ln \frac{\theta^*_{v_1|\mathbf{u}_j}}{\theta_{v_1|\mathbf{u}_j}} - \ln \frac{1-\theta^*_{v_1|\mathbf{u}_j}}{1-\theta_{v_1|\mathbf{u}_j}} \right|$$

In Chan (2005) the error was corrected by stating that the above expression is an approximation $\widetilde{D}$ of the true distance $D_p(\Theta_{V|\mathbf{U}}, \Theta^*_{V|\mathbf{U}})$.

If we extend Proposition 5 by maximising and minimising over all parent configurations $\mathbf{u}_j$ as well, we get the exact expression for the distance between two CPTs: $D_\gamma(\Theta_{V|\mathbf{U}}, \Theta^*_{V|\mathbf{U}}) =$

$$\ln \max_{\mathbf{u}_j} \left\{ \frac{\theta^*_{v_1|\mathbf{u}_j}}{\theta_{v_1|\mathbf{u}_j}}, \frac{1-\theta^*_{v_1|\mathbf{u}_j}}{\min\limits_{1<k\leq t} \gamma^{-1}_{v_k|\mathbf{u}_j} \cdot \theta_{v_k|\mathbf{u}_j}} \right\}$$
$$- \quad \ln \min_{\mathbf{u}_j} \left\{ \frac{\theta^*_{v_1|\mathbf{u}_j}}{\theta_{v_1|\mathbf{u}_j}}, \frac{1-\theta^*_{v_1|\mathbf{u}_j}}{\max\limits_{1<k\leq t} \gamma^{-1}_{v_k|\mathbf{u}_j} \cdot \theta_{v_k|\mathbf{u}_j}} \right\}$$

which in the case of proportional co-variation reduces to: $D_p(\Theta_{V|\mathbf{U}}, \Theta^*_{V|\mathbf{U}}) =$

$$\ln \max_{\mathbf{u}_j} \left\{ \frac{\theta^*_{v_1|\mathbf{u}_j}}{\theta_{v_1|\mathbf{u}_j}}, \frac{1-\theta^*_{v_1|\mathbf{u}_j}}{1-\theta_{v_1|\mathbf{u}_j}} \right\}$$
$$- \quad \ln \min_{\mathbf{u}_j} \left\{ \frac{\theta^*_{v_1|\mathbf{u}_j}}{\theta_{v_1|\mathbf{u}_j}}, \frac{1-\theta^*_{v_1|\mathbf{u}_j}}{1-\theta_{v_1|\mathbf{u}_j}} \right\}$$

The following example illustrates the difference between this expression and the approximation $\widetilde{D}$.

Example 5. Consider a table $\Theta_{V|\mathbf{U}}$ for binary-valued V and U. Let $\theta_{v_1|u_1} = 0.8$ and $\theta_{v_1|u_2} = 0.6$, and suppose these parameters are decreased to $\theta^*_{v_1|u_1} = 0.6$ and $\theta^*_{v_1|u_2} = 0.3$, respectively. Let R denote the set of all ratios under consideration, i.e. $R = \{\frac{0.6}{0.8}, \frac{1-0.6}{1-0.8}, \frac{0.3}{0.6}, \frac{1-0.3}{1-0.6}\}$. Then $D_p(\Theta_{V|\mathbf{U}}, \Theta^*_{V|\mathbf{U}}) = \ln\max R - \ln\min R = \ln 2 - \ln 0.5 = 1.386$. $\widetilde{D}(\Theta_{V|\mathbf{U}}, \Theta^*_{V|\mathbf{U}})$, however, equals the maximum of $|\ln 0.75 - \ln 2|$ and $|\ln 0.5 - \ln 1.75|$, which is 1.252. ■

In the above example the approximated CD-distance is smaller than the true distance. In fact,

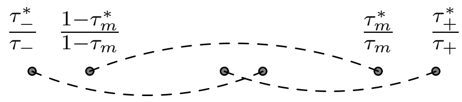

Figure 1: Illustration of distances between (the $\ln$ of) various parameter ratios. Dashed lines link $\ln \frac{\tau^*}{\tau}$-terms to their corresponding $\ln \frac{1-\tau^*}{1-\tau}$-terms.

the approximate distance is a lower bound on the true distance.

Proposition 6. *Let $\Theta_{V|\mathbf{U}}$ and $\Theta^*_{V|\mathbf{U}}$ be as before. Then,*

$$D_p(\Theta_{V|\mathbf{U}}, \Theta^*_{V|\mathbf{U}}) \geq \widetilde{D}(\Theta_{V|\mathbf{U}}, \Theta^*_{V|\mathbf{U}})$$

Proof. Let τ^*_+/τ_+ denote the maximum, over all parent configurations $\mathbf{u}_j$, of $\theta^*_{v_1|\mathbf{u}_j}/\theta_{v_1|\mathbf{u}_j}$ and $(1-\theta^*_{v_1|\mathbf{u}_j})/(1-\theta_{v_1|\mathbf{u}_j})$; likewise, let τ^*_-/τ_- denote the minimum (see Figure 1). Note that $\tau^*_+/\tau_+ \geq 1$ and $0 < \tau^*_-/\tau_- \leq 1$. Then, by definition,

$$D_p(\Theta_{V|\mathbf{U}}, \Theta^*_{V|\mathbf{U}}) = \ln \frac{\tau^*_+}{\tau_+} - \ln \frac{\tau^*_-}{\tau_-}$$

Now if τ_- and $1-\tau_+$ actually correspond with the same parameter, then $D_p(\Theta_{V|\mathbf{U}}, \Theta^*_{V|\mathbf{U}}) = \widetilde{D}(\Theta_{V|\mathbf{U}}, \Theta^*_{V|\mathbf{U}})$, since $\ln(\tau^*_+/\tau_+) - \ln((1-\tau^*_+)/(1-\tau_+))$ equals the largest possible distance. Otherwise, suppose that $\widetilde{D} = \ln(\tau^*_m/\tau_m) - \ln((1-\tau^*_m)/(1-\tau_m))$, for some τ_m such that $\tau^*_+/\tau_+ \geq \tau^*_m/\tau_m \geq \tau^*_-/\tau_-$. Then from Figure 1 it is obvious that $\widetilde{D}(\Theta_{V|\mathbf{U}}, \Theta^*_{V|\mathbf{U}}) \leq D_p(\Theta_{V|\mathbf{U}}, \Theta^*_{V|\mathbf{U}})$. □

6 Conclusions

We have generalised both sensitivity functions and CD-distance to cope with arbitrary co-variation schemes. We showed that the sensitivity function remains a rational function as long as a valid and linear co-variation scheme is used. In addition, we discussed the suitability of various algorithms for computing the sensitivity functions under different co-variation schemes. Finally, we proved a lower bound on CD-distance for single CPTs.

In Section 4.2 we provided some preliminary results on comparing sensitivity functions under various co-variation schemes. We plan to study the differences in more detail in the near future. The increased complexity of the formula for the CD-distance for single CPT changes, forestalls straightforward generalisation of the known optimality of the CD-distance. If proportional co-variation turns out not to be necessarily optimal for whole CPTs, then alternative co-variation schemes will become more interesting to consider in e.g. the context of parameter tuning. We suspect that sensible co-variation schemes will be domain-dependent, preserving e.g. known thresholds or relationships between parameters.

References

E. Castillo, J.M. Gutiérrez, A.S. Hadi (1997). Sensitivity analysis in discrete Bayesian networks. *IEEE Transactions on Systems, Man, and Cybernetics*, 27, pp. 412 – 423.

H. Chan (2005). *Sensitivity Analysis of Probabilistic Graphical Models*, Ph.D. thesis, University of California, Los Angeles.

H. Chan, A. Darwiche (2002). When do numbers really matter? *Journal of Artificial Intelligence Research*, 17, pp. 265 – 287.

H. Chan, A. Darwiche (2004). Sensitivity analysis in Bayesian networks: from single to multiple parameters. *Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence*, AUAI Press, pp. 67 – 75.

H. Chan, A. Darwiche (2005). A distance measure for bounding probabilistic belief change. *International Journal of Approximate Reasoning*, 38, pp. 149 – 174.

V.M.H. Coupé, L.C. van der Gaag (2002). Properties of sensitivity analysis of Bayesian belief networks. *Annals of Mathematics and Artificial Intelligence*, 36, pp. 323 – 356.

Th.M. Cover, J.A. Thomas (1991). *Elements of Information Theory*, Wiley-Interscience.

L.C. van der Gaag, S. Renooij, V.M.H. Coupé (2007). Sensitivity analysis of probabilistic networks. *Advances in Probabilistic Graphical Models*, 213, pp. 103 – 124.

F.V. Jensen, T.D. Nielsen (2007). *Bayesian Networks and Decision Graphs*, Springer-Verlag.

U. Kjærulff, L.C. van der Gaag (2000). Making sensitivity analysis computationally efficient. *Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence*, Morgan Kaufmann, pp. 317 – 325.