Local Sensitivity of Bayesian Networks to Multiple Simultaneous Parameter Shifts

Janneke H. Bolt and Silja Renooij

Department of Information and Computing Sciences
Utrecht University, Utrecht, The Netherlands

Abstract. The robustness of the performance of a Bayesian network to shifts in its parameters can be studied with a sensitivity analysis. For reasons of computational efficiency such an analysis is often limited to studying shifts in only one or two parameters at a time. The concept of sensitivity value, an important notion in sensitivity analysis, captures the effect of local changes in a single parameter. In this paper we generalise this concept to an *n-way sensitivity value* in order to capture the local effect of multiple simultaneous parameters changes. Moreover, we demonstrate that an n-way sensitivity value can be computed efficiently, even for large n. An n-way sensitivity value is direction dependent and its maximum, minimum, and direction of maximal change can be easily determined. The direction of maximal change can, for example, be exploited in network tuning. To this end, we introduce the concept of *sliced sensitivity function* for an n-way sensitivity function restricted to parameter shifts in a fixed direction. We moreover argue that such a function can be computed efficiently.

1 Introduction

The robustness of Bayesian networks to changes in their parameter probabilities can be studied with a sensitivity analysis. To this end, a function which describes the effect of varying one or more parameters on an output probability of interest can be established. From such a sensitivity function, various sensitivity properties can be derived that give insight into the effects of the parameter changes [6].

Most research has focused on one-way sensitivity analyses in which only a single parameter is varied at a time. These one-way analyses, however, do not provide full insight into the effects of multiple simultaneous parameter shifts; to study such effects, an n-way sensitivity analysis is required. To this end, we can establish an n-way sensitivity function. Unfortunately, the computation of multi-dimensional functions is generally expensive. Existing algorithms for n-way sensitivity analysis are only computationally feasible for larger n under certain conditions. For example, the efficient method for computing n-way sensitivity functions by Kjærulff and van der Gaag [8] assumes that the n parameters all belong to the same clique in the network's junction tree representation. Another example is the method introduced by Chan and Darwiche [2] for assessing which parameter shifts will enforce a given constraint with respect to some outcome probability: this method is feasible if all parameters concern the same CPT, and quickly becomes infeasible otherwise.

In this paper we are interested in studying the local effects of multiple simultaneous parameters changes in a Bayesian network. To this end we generalise the concept of *sensitivity value* [9] — well-known in the context of one-way sensitivity analysis — to an *n-way sensitivity value*. The sensitivity value captures the effect of local parameter changes by means of the derivative of the sensitivity function in the point corresponding with the original parameter assessment specified in the Bayesian network. We generalise this concept to multiple dimensions by using a directional derivative of the n-way sensitivity function. Moreover, we prove that computing this directional derivative can be done efficiently, due to the fact that we do not need the n-way sensitivity function: availability of the one-way sensitivity values of the parameters under consideration suffices. The n-way sensitivity value is direction dependent, but its maximum and minimum can be easily determined, together with the corresponding direction of maximal change. The latter information is not only useful for studying the robustness of a Bayesian network, but is also useful in the context of parameter tuning. In parameter tuning, network parameters are changed in order to fulfill constraints with respect to outcome probabilities. Assuming that small perturbations are preferred, we argue that tuning parameters by shifting them in the direction of maximal change will yield a good approximation of the optimal parameter change necessary to meet a given constraint. Moreover, since a fixed vector direction ties together all parameters linearly, we can efficiently establish the effect of such a combined parameter shift. To this end, we introduce the concept of *sliced sensitivity function*.

The remainder of the paper is organised as follows. In Section 2 we introduce our notational conventions, briefly review sensitivity analysis in Bayesian networks and review the mathematical notion of directional derivatives. In Section 3 we define the n-way sensitivity value and its bounds, and in Section 4 we address the question of how to compute an n-way sensitivity value efficiently. In Section 5 we discuss the use of our concepts in the context of parameter tuning and we conclude our paper with a discussion in Section 6.

2 Preliminaries

2.1 Bayesian Networks and Sensitivity Analysis

A Bayesian network compactly represents a joint probability distribution Pr over a set of stochastic variables $\mathbf{A}$ [7]. It combines an acyclic directed graph G, that captures the variables and their dependencies as nodes and arcs respectively, with conditional probability distributions for each variable A_i and its parents $\pi(A_i)$ in the graph, such that

$$\Pr(\mathbf{A}) = \prod_i \Pr(A_i \mid \pi(A_i))$$

Variables are denoted by capital letters, which are boldfaced in case of sets; specific values or instantiations are written in lower case. In examples we restrict ourselves to binary variables, writing a and $\bar{a}$ to denote the two possible instantiations of a variable A. We assume the conditional distributions are specified as tables (CPTs) and use the term *parameter* to refer to a CPT entry. The superscript 'o' is used to indicate that a probability is an original parameter value, or is computed from the network with parameter values as originally specified.

To investigate the effects of inaccuracies in its parameters, a Bayesian network can be subjected to a sensitivity analysis. In a sensitivity analysis, parameters of a network are varied and a probability of interest as a function of the varied parameters is computed.

General n-way analysis In an n-way sensitivity analysis, simultaneous perturbations of multiple parameters are considered. The effect of varying the parameters $x_1, \ldots, x_n$ on a probability of interest $\Pr(y \mid \mathbf{e})$ is captured by a function of the form

$$f_{\Pr(y|\mathbf{e})}(x_1, \ldots, x_n) = \frac{f_{\Pr(y\,\mathbf{e})}(x_1, \ldots, x_n)}{f_{\Pr(\mathbf{e})}(x_1, \ldots, x_n)} = \frac{\sum_{\mathbf{X}_k \in \mathcal{P}(\{x_1,\ldots,x_n\})} c_k \cdot \prod_{x_i \in \mathbf{X}_k} x_i}{\sum_{\mathbf{X}_k \in \mathcal{P}(\{x_1,\ldots,x_n\})} d_k \cdot \prod_{x_i \in \mathbf{X}_k} x_i}$$

where $\mathcal{P}$ denotes the powerset, and c_k and d_k, $k = 0, \ldots, 2^n - 1$, are constants constructed from the non-varied network parameter [1]. A two-way function, for example, takes the following form:

$$f_{\Pr(y|\mathbf{e})}(x_1, \ldots, x_2) = \frac{c_0 + c_1 \cdot x_1 + c_2 \cdot x_2 + c_3 \cdot x_1 \cdot x_2}{d_0 + d_1 \cdot x_1 + d_2 \cdot x_2 + d_3 \cdot x_1 \cdot x_2}$$

The n parameters of an n-way sensitivity function are typically assumed to be independent, that is, parameters from the same CPT must come from different conditional distributions. Upon varying a parameter $x = \Pr(a_i \mid \boldsymbol{\pi})$, all probabilities $\Pr(a_j \mid \boldsymbol{\pi})$, $j \neq i$, pertaining to the same conditional distribution are assumed to co-vary proportionally.

An n-way sensitivity function in general requires the computation of 2^n constants and is thus computationally expensive; an algorithm to this end can be found in [8].

Single CPT analysis In the special case where all n parameters are independent parameters from the *same* CPT, the interaction terms in the n-way sensitivity function become zero and the function reduces to the following form [2]:

$$f_{\Pr(y|\mathbf{e})}(x_1, \ldots, x_n) = \frac{c_0 + \sum_i c_i \cdot x_i}{d_0 + \sum_i d_i \cdot x_i}$$

One-way analysis Most research has focused on one-way sensitivity analysis, in which just a single parameter x is varied. In this case the sensitivity function becomes [4]:

$$f_{\Pr(y|\mathbf{e})}(x) = \frac{c_0 + c_1 \cdot x}{d_0 + d_1 \cdot x}$$

The constants of the one-way functions $f_{\Pr(y|\mathbf{e})}(x_i)$ for output probability $\Pr(y \mid \mathbf{e})$ can be established efficiently for *all* network parameters x_i simultaneously from just one inward and two outward propagations in the junction tree representation of the Bayesian network [8].

From the one-way sensitivity function, several sensitivity properties can be established [6]. The most well-known sensitivity property is the *sensitivity value* [9]. This value captures the sensitivity of an outcome probability of interest to small perturbations of the parameter under consideration. The sensitivity value of the one-way sensitivity function $f(x)$ for parameter x with original assessment x^o is defined as the absolute value of the first derivative of the function at $x = x^o$:

$$\left| \frac{df}{dx}(x^o) \right|$$

High sensitivity values indicate that the output probability of interest may change considerably as a result of small parameter changes. The one-way sensitivity function takes the form of a rectangular hyperbola, with its vertex (in which the derivative is $+1$ or -1) marking the transition from low to possibly high sensitivity.

2.2 Directional Derivatives

The sensitivity value is defined in terms of the first derivative of the one-way sensitivity function. For a one-dimensional function $f(x)$ we can refer to *the* derivative at $x = a$, since $\frac{df}{dx}(a)$ is a single value. The multi-dimensional analogue of the derivative is the *directional derivative*. The directional derivative of an n-dimensional function depends on the direction $\mathbf{v}$ and the specific point $\mathbf{x}$ of the function that is considered. To compute the directional derivative of a function $f(\mathbf{x})$ for $\mathbf{x} = (x_1, \ldots, x_n)$, we can use its *gradient* ∇f, that is, the vector of partial derivatives $(\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n})$ of f. The directional derivative at $\mathbf{x} = \mathbf{a}$ in the direction $\mathbf{v}$ now equals the following dot product

$$D_{\mathbf{u}} f(\mathbf{a}) = \nabla f(\mathbf{a}) \bullet \mathbf{u}$$

where unit vector $\mathbf{u}$ is the normalised vector of $\mathbf{v}$, that is, $\mathbf{u}$ is the vector in the direction of $\mathbf{v}$ that has length 1.

Although the directional derivative varies depending on the chosen direction, we can establish bounds on its value. The maximum directional derivative of f at $\mathbf{x} = \mathbf{a}$ is found in the direction of the gradient vector at $\mathbf{a}$, $\nabla f(\mathbf{a})$, and equals the length of the gradient vector at $\mathbf{a}$, $|\nabla f(\mathbf{a})|$. Similarly, the minimum directional derivative occurs in the opposite direction.

Example 1. Suppose we are interested in the directional derivative of $f(x, y) = x^2 + 4 \cdot x \cdot y$ at $(1, 2)$, in the direction $(-2, 1)$. We have that $\nabla f = (2 \cdot x + 4 \cdot y, 4 \cdot x)$ which yields $\nabla f(1, 2) = (10, 4)$. Vector $(-2, 1)$ has length $\sqrt{5}$ and is normalised to $\mathbf{u} = (\frac{-2}{\sqrt{5}}, \frac{1}{\sqrt{5}})$. The requested directional derivative thus equals $D_{\mathbf{u}} f(1, 2) = (10, 4) \bullet (\frac{-2}{\sqrt{5}}, \frac{1}{\sqrt{5}}) = \frac{-16}{\sqrt{5}}$. The maximum directional derivative at $(1, 2)$ occurs in the direction $(10, 4)$ and equals $\sqrt{10^2 + 4^2} \approx 10.77$; the minimum directional derivative at this point equals -10.77 and occurs in the direction $(-10, -4)$.

3 Defining an n-way Sensitivity Value

The sensitivity value as defined in [9] reflects the local sensitivity of some outcome of interest to a single parameter shift. In this section we generalise the definition of sensitivity value in order to capture the local sensitivity given multiple simultaneous parameter shifts.

A (one-way) sensitivity value is defined in terms of the first derivative of a one-way sensitivity function. In mathematics, the notion of *first derivative* of a function with a single variable generalises to the notion of *directional derivative* for a function with multiple variables. We therefore define an n-way sensitivity value in terms of a directional derivative. In contrast to the definition of the one-way sensitivity value, we will not consider the n-way sensitivity value to be an absolute value. In our opinion, using the absolute value results in loss of useful information concerning the direction of change in the output of interest upon local perturbation of the parameters. For this reason, we also introduce a signed version of the one-way sensitivity value, which equals the sensitivity value prior to taking the absolute value.

Definition 1 (signed sensitivity value). *Let $f(x)$ be a one-way sensitivity function and x^o the original value for parameter x. The* signed sensitivity value *for $f(x)$, denoted sv^x, equals the first derivative of f at x^o:*

$$sv^x = \frac{df}{dx}(x^o)$$

We now generalise the concept of (signed) sensitivity value to multiple dimensions.

Definition 2 (n-way sensitivity value). *Let $f(\mathbf{x})$ be an n-way sensitivity function and let $\mathbf{x^o}$ be the vector of original parameter settings. Consider a shift of the parameters in the direction $\mathbf{v}$. The* n-way sensitivity value *for $f(\mathbf{x})$, denoted $sv_{\mathbf{v}}^{\mathbf{x}}$, equals the directional derivative of f at the original parameter assessments $\mathbf{x^o}$ in the direction $\mathbf{v}$:*

$$sv_{\mathbf{v}}^{\mathbf{x}} = D_{\mathbf{u}}\, f(\mathbf{x^o})$$

where unit vector $\mathbf{u}$ is the normalised vector of $\mathbf{v}$.

Note that sv^x is a special case of $sv_{\mathbf{v}}^{\mathbf{x}}$ for $n = 1$ and $\mathbf{u} = (1)$.

Whereas a single parameter can only be changed to lower or higher values, multiple simultaneous parameters shifts can occur in an infinite number of directions. Hence the dependence on $\mathbf{v}$ in our definition of n-way sensitivity value. Fortunately, the n-way sensitivity values have an upper- and lowerbound.

Definition 3 ($sv_{\max}^{\mathbf{x}}$). *Let $f(\mathbf{x})$ be an n-way sensitivity function and $\mathbf{x^o}$ the vector of original parameter settings. The* maximum n-way sensitivity value, *denoted $sv_{\max}^{\mathbf{x}}$, equals*

$$sv_{\max}^{\mathbf{x}} = \max_{\mathbf{v}}\ sv_{\mathbf{v}}^{\mathbf{x}} = \max_{\mathbf{u}}\ D_{\mathbf{u}}\, f(\mathbf{x^o})$$

where unit vector $\mathbf{u}$ is the normalised vector of $\mathbf{v}$.

Since the n-way sensitivity value is defined as a directional derivative, its maximum value in fact equals the length of the gradient vector of f at $\mathbf{x^o}$, that is, $sv^{\mathbf{x}}_{\max} = |\nabla f(\mathbf{x^o})|$; moreover, $sv^{\mathbf{x}}_{\max}$ is obtained in the direction $\nabla f(\mathbf{x^o})$. We can similarly define $sv^{\mathbf{x}}_{\min} = -sv^{\mathbf{x}}_{\max}$ as the minimum n-way sensitivity value, which occurs in the opposite direction $-\mathbf{1} \bullet \nabla f(\mathbf{x^o})$.

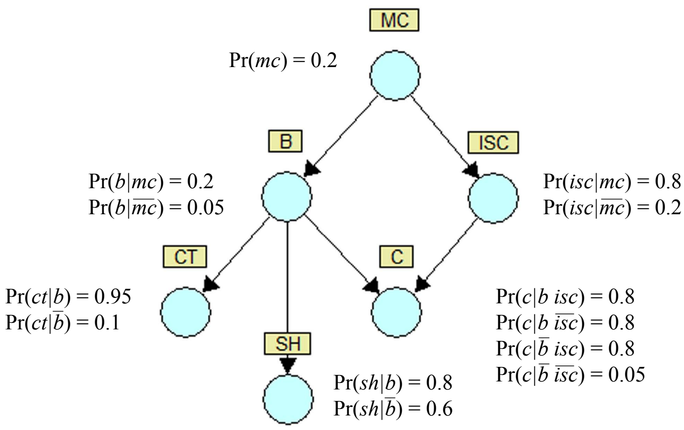

Fig. 1. An example Bayesian network.

Example 2. Consider the example network from Fig. 1, representing some (fictitious) medical information. For a patient, the variables MC, B and SH represent the presence or absence of metastatic cancer, a brain tumour, and severe headaches, respectively. Variable ISC captures the presence or absence of an increased serum calcium level, variable C represents whether or not a patient is comatose, and CT whether or not the outcome of a CT scan is positive. Suppose that we are interested in the output probability of a brain tumour in a patient with a positive CT-scan, severe headaches, but who is not in a coma, that is, $\Pr(b \mid ct\, sh\, \bar{c})$. In addition, suppose that the assessments of the parameters $x = \Pr(mc)$ and $y = \Pr(sh \mid \bar{b})$ might be inaccurate. We now find the following sensitivity function, depicted in Fig. 2:

$$f_{\Pr(b|ct\,sh\,\bar{c})}(x, y) = \frac{0.76 + 2.28 \cdot x}{0.76 + 2.28 \cdot x + 7.6 \cdot y - 4.8 \cdot x \cdot y}$$

with gradient $\nabla f = (\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y})$, where

$$\frac{\partial f}{\partial x} = \frac{20.976 \cdot y}{(0.76 + x \cdot (2.28 - 4.8 \cdot y) + 7.6 \cdot y)^2}$$

and

$$\frac{\partial f}{\partial y} = \frac{-5.776 - 13.68 \cdot x + 10.944 \cdot x^2}{(0.76 + x \cdot (2.28 - 4.8 \cdot y) + 7.6 \cdot y)^2}$$

The gradient at $(x^o, y^o) = (0.2, 0.6)$ then equals $\nabla f(0.2, 0.6) \approx (0.465, -0.299)$.

Now consider a parameter shift from $(x^o, y^o) = (0.2, 0.6)$ to $(0.1, 0.7)$, that is, a shift in the direction $(-0.1, 0.1)$. The directional derivative at $(0.2, 0.6)$ in this direction is $(0.465, -0.299) \bullet (\frac{-0.1}{\sqrt{0.02}}, \frac{0.1}{\sqrt{0.02}}) \approx -0.540$ and equals the sensitivity value $sv^{x,y}_{\mathbf{v}}$ for this direction. The maximum sensitivity value $sv^{x,y}_{\max}$ $= |(0.465, -0.299)| \approx 0.553$ and is found in the direction $(0.465, -0.299)$.

We can also compute the directional derivative at some other point (x, y) than the original parameter assessments. For example, the gradient at $(x, y) = (0.1, 0.1)$ equals $\nabla f(0.1, 0.1) \approx (0.581, -0.414)$. For a shift from this point, in the direction $(0.4, 0.2)$, we have a directional derivative of $(0.581, -0.414) \bullet (\frac{0.4}{\sqrt{0.2}}, \frac{0.2}{\sqrt{0.2}}) \approx -0.414$.

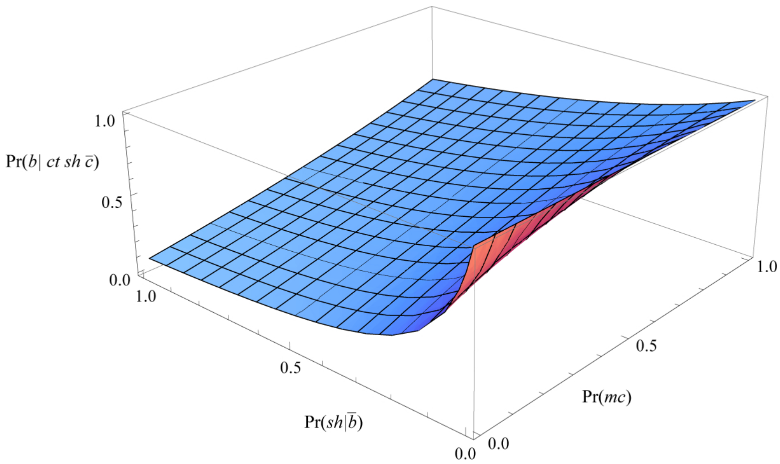

Fig. 2. $\Pr(b \mid ct\, sh\, \bar{c})$ as function of $\Pr(mc)$ and $\Pr(sh \mid \bar{b})$ given the network from Figure 1.

4 Computing an n-way Sensitivity Value

For the computation of an n-way sensitivity value, the partial derivatives $\frac{\partial f}{\partial x_i}$ of the n-way sensitivity function at $\mathbf{x^o}$ are required. These partial derivatives can be established in various ways. In Section 4.1 we will identify the possibilities and drawbacks of using various existing algorithms. In Section 4.2 we will subsequently point out a relation between the n-way sensitivity value and one-way sensitivity values that allows for efficiently computing the former.

4.1 Computing Partial Derivatives for Sensitivity Functions

There are basically two approaches that we can employ for computing our partial derivatives $\frac{\partial f}{\partial x_i}$ for sensitivity function $f(\mathbf{x})$: a direct and an indirect approach.

Indirect Approach One approach is to establish the complete sensitivity function $f(\mathbf{x})$, using one of the available algorithms for computing its constants from the Bayesian network (see Section 2). Subsequently, the partial derivative with respect to x_i can be computed from the resulting function. This approach allows for computing partial derivatives at any value of x_i and not only at x_i^o. The major drawback of this approach, however, is that the currently most efficient algorithm to compute an n-way sensitivity function requires in the order of $2^n/n$ full junction tree propagations to establish 2^n equations from which the required constants can be solved; this can only be done more efficiently if all n parameters are in the same clique [8]. We note that in the special case where all n parameters are independent parameters from the same CPT, the n-way sensitivity function requires a linear rather than exponential number of constants. In that case, it is doable to compute the n-way sensitivity function.

For the special case where all parameters are from the same CPT, we can express $sv_{\max}^{\mathbf{x}}$ in terms of the constants of the n-way sensitivity function, the original probability of interest and the original probability of the evidence, as stated in the following proposition.

Proposition 1 **($sv_{\max}^{\mathbf{x}}$; x in single CPT).** *Consider an n-way sensitivity function $f_{\Pr(y|\mathbf{e})}(\mathbf{x}) = f_{\Pr(y\,\mathbf{e})}(\mathbf{x})/f_{\Pr(\mathbf{e})}(\mathbf{x})$ for output probability $\Pr(y \mid \mathbf{e})$ and n independent parameters $\mathbf{x} = (x_1, \ldots, x_n)$ from a single CPT with original values $\mathbf{x^o} = (x_1^o, \ldots, x_n^o)$. Let $f_{\Pr(y\,\mathbf{e})}(\mathbf{x}) = c_0 + \sum_i c_i \cdot x_i$ and $f_{\Pr(\mathbf{e})}(\mathbf{x}) = d_0 + \sum_i d_i \cdot x_i$ for constants c_i, d_i, $i = 1, \ldots, n$. Then the maximum n-way sensitivity value equals*

$$sv_{\max}^{\mathbf{x}} = \frac{1}{\Pr^o(\mathbf{e})} \cdot \sqrt{\sum_{i=1}^{n} \Big(c_i - d_i \cdot \Pr^o(y \mid \mathbf{e})\Big)^2}$$

Proof. The value $sv_{\max}^{\mathbf{x}}$ is the length of the gradient vector in $\mathbf{x^o}$. To compute the gradient vector we compute the partial derivatives of f for each x_k, $k = 1, \ldots, n$:

$$\frac{\partial f}{\partial x_k} = \frac{c_k \cdot (d_0 + \sum_{i,i\neq k} d_i \cdot x_i) - d_k \cdot (c_0 + \sum_{i,i\neq k} c_i \cdot x_i)}{(d_0 + \sum_{i=1}^{n} d_i \cdot x_i)^2}$$

At $\mathbf{x^o}$ this partial derivative equals:

$$\begin{aligned}\frac{\partial f}{\partial x_k}(\mathbf{x^o}) &= \frac{c_k \cdot (d_0 + \sum_{i,i\neq k} d_i \cdot x_i^o) - d_k \cdot (c_0 + \sum_{i,i\neq k} c_i \cdot x_i^o)}{(d_0 + \sum_{i=1}^{n} d_i \cdot x_i^o)^2} \\ &= \frac{c_k \cdot (\Pr^o(\mathbf{e}) - d_k \cdot x_k^o) - d_k \cdot (\Pr^o(y\,\mathbf{e}) - c_k \cdot x_k^o)}{\Pr^o(\mathbf{e})^2} \\ &= \frac{c_k \cdot \Pr^o(\mathbf{e}) - d_k \cdot \Pr^o(y\,\mathbf{e})}{\Pr^o(\mathbf{e})^2} = \frac{c_k - d_k \cdot \Pr^o(y \mid \mathbf{e})}{\Pr^o(\mathbf{e})}\end{aligned}$$

The result now follows directly. □

Direct Approach The second approach is far more elegant for our purposes. A *differential* approach can be used to compute partial derivatives from the so-called *canonical network polynomial* $\mathcal{F}$. For $\frac{\partial f}{\partial x_i}$ a closed form in terms of first and second order partial derivatives of polynomial $\mathcal{F}$ exists; details are beyond the scope of this paper and can be found in [5]. As an alternative, for any parameter $x_i = \Pr(a \mid \boldsymbol{\pi})$ with $x_i^o \neq 0$, we can use the following probabilistic closed form, which is equivalent to the above-mentioned one based on partial derivatives [5]:

$$\frac{\partial f_{\Pr(y|\mathbf{e})}}{\partial x_i}(\mathbf{x^o}) = \frac{\Pr^o(y\, a\, \boldsymbol{\pi} \mid \mathbf{e}) - \Pr^o(y \mid \mathbf{e}) \cdot \Pr^o(a\, \boldsymbol{\pi} \mid \mathbf{e})}{x_i^o}$$

Since the closed forms allow for direct computation of partial derivatives, albeit at $x_i = x_i^o$ only, they are more efficient to compute than the approach using the n-way sensitivity function: rather than computing a number of constants that is exponential in n, we compute n partial derivatives. A drawback of using the partial-derivative-based closed form is that it cannot be computed using classical inference algorithms and requires the computation of both first and second order partial derivatives. A minor drawback of the probabilistic closed form is that the expression requires the computation of several probabilities per parameter for which it is not immediately clear what their relation to sensitivity analysis or sensitivity properties is.

4.2 n-way Partial Derivatives From One-way Functions

In the previous section we argued that the direct computation of partial derivatives is much more efficient than establishing them from an n-way sensitivity function. In this section we demonstrate that there is a simple correspondence between partial derivatives of n-way sensitivity functions and derivatives for one-way functions[1]. This provides us with an alternative way of efficiently establishing n-way sensitivity values during a sensitivity analysis.

Proposition 2. *Let $x_1, \ldots, x_n$ be $n > 1$ network parameters with original assessments x_i^o, $i = 1, \ldots, n$, and let P be an output probability of interest. Consider the n-way sensitivity function $f_P(x_1, \ldots, x_n)$ and the one-way sensitivity function $f_P^*(x_k)$, $k \in \{1, \ldots, n\}$. Then*

$$\frac{\partial f_P}{\partial x_k}(x_1^o, \ldots, x_n^o) = \frac{d\, f_P^*}{d x_k}(x_k^o)$$

Proof. Consider an output probability $P = \Pr(y \mid \mathbf{e}) = \frac{\Pr(y\,\mathbf{e})}{\Pr(\mathbf{e})}$. As a result of the factorisation defined by a Bayesian network, both numerator and denominator can be written as an expression of all network parameters consistent with y and/or $\mathbf{e}$ [1]. Suppose these expressions contain m independent parameters (the remaining ones will co-vary). A sensitivity function for $n < m$ of these independent parameters then basically is the m-dimensional sensitivity function with $m - n$ independent parameters fixed at their original value. This also holds for

[1] We note that this correspondence, formally stated in Proposition 2, has been implicitly exploited in, for example, [5]; to the best of our knowledge, however, it has not been formalised explicitly before.

$n = 1$. The partial derivative w.r.t x_k of an n-way function $f_P(x_1, \ldots, x_k, \ldots, x_n)$ with parameters $x_i \neq x_k$ kept at x_i^o, is therefore the same as the derivative of the one-way sensitivity function $f_P^*(x_k)$. This proves the proposition. $\Box$

To assess the partial derivatives of an n-way sensitivity function given the original parameter assessments, we thus just need the appropriate one-way sensitivity functions. The above proposition thus gives a computationally feasible way of computing the n-way sensitivity value, since the constants of the one-way functions can be established efficiently. Note that if we are not interested in an n-way sensitivity value, but in the directional derivative at some other point than the original parameters assessments, then the one-way sensitivity functions will not suffice.

Example 3. Consider again the outcome of interest $\Pr(b \mid ct\, sh\, \bar{c})$ and the parameters $x = \Pr(mc)$ and $y = \Pr(sh \mid \bar{b})$ from *Example 2* and Fig. 1. The one-way sensitivity functions are given by

$$f^*_{\Pr(b|ct\,sh\,\bar{c})}(x) = \frac{0.76 + 2.28 \cdot x}{5.32 - 0.6 \cdot x} \quad \text{and} \quad f^{\diamondsuit}_{\Pr(b|ct\,sh\,\bar{c})}(y) = \frac{1.216}{1.216 + 6.64 \cdot y}$$

Their derivatives equal

$$\frac{df^*}{dx}(x^o) = \frac{12.586}{(5.32 - 0.6 \cdot x^o)^2} = 0.465, \; \frac{df^{\diamondsuit}}{dy}(y^o) = \frac{-8.074}{(1.216 + 6.64 \cdot y^o)^2} = -0.299$$

at x^o and y^o, respectively. We observe that indeed

$$\Big(\frac{\partial f}{\partial x}(x^o, y^o), \frac{\partial f}{\partial y}(x^o, y^o)\Big) = \Big(\frac{df^*}{dx}(x^o), \frac{df^{\diamondsuit}}{dy}(y^o)\Big)$$

Using Proposition 2, we can express $sv^{\mathbf{x}}_{\max}$ in terms of the constants of the one-way sensitivity functions and the original probability of the evidence.

Proposition 3 ($sv^{\mathbf{x}}_{\max}$ in general). *Consider $n > 1$ network parameters $\mathbf{x} = (x_1, \ldots, x_n)$ with original values $\mathbf{x^o} = (x_1^o, \ldots, x_n^o)$, and let $\Pr(y \mid \mathbf{e})$ be an output probability of interest. In addition, consider the n one-way sensitivity functions $f^{(i)}_{\Pr(y|\mathbf{e})}(x_i) = f^{(i)}_{\Pr(y\,\mathbf{e})}(x_i) \,/\, f^{(i)}_{\Pr(\mathbf{e})}(x_i)$, $i = 1, \ldots, n$, where $f^{(i)}_{\Pr(y\,\mathbf{e})}(x_i) = c_0^i + c_1^i \cdot x_i$, with constants c_0^i, c_1^i, and $f^{(i)}_{\Pr(\mathbf{e})}(x_i) = d_0^i + d_1^i \cdot x_i$, with constants d_0^i, d_1^i. Then the maximum n-way sensitivity value for the n-way function $f_{\Pr(y|\mathbf{e})}(\mathbf{x})$ equals*

$$sv^{\mathbf{x}}_{\max} = \frac{1}{\Pr^o(\mathbf{e})^2} \cdot \sqrt{\sum_{i=1}^{n} (c_1^i \cdot d_0^i - c_0^i \cdot d_1^i)^2}$$

Proof. The derivative of the one-way sensitivity function $f^{(i)}_{\Pr(y|\mathbf{e})}(x_i)$ at the original parameter assessment x_i^o equals

$$\frac{df^{(i)}_{\Pr(y|\mathbf{e})}}{dx_i}(x_i^o) = \frac{c_1^i \cdot d_0^i - c_0^i \cdot d_1^i}{(d_0^i + d_1^i \cdot x_i^o)^2} = \frac{c_1^i \cdot d_0^i - c_0^i \cdot d_1^i}{\Pr^o(\mathbf{e})^2}$$

Since this derivative is equal to the partial derivative $\frac{\partial f_{\Pr(y|\mathbf{e})}}{\partial x_i}$ at $\mathbf{x^o}$ (Proposition 2), and $sv^{\mathbf{x}}_{\max}$ is the length of the gradient vector in $\mathbf{x^o}$, the proposition follows. □

The following corollary states two convenient properties for $sv^{\mathbf{x}}_{\max}$. In addition, it states properties that can be used in case we are interested in the n-way sensitivity value in an arbitrary direction, rather than just in the maximum value.

Corollary 1. *Consider $n > 1$ network parameters $\mathbf{x} = (x_1, \ldots, x_n)$ with original values $\mathbf{x^o} = (x_1^o, \ldots, x_n^o)$, and let P be an output probability of interest. Consider the n-way sensitivity function $f_P(\mathbf{x})$ with n-way sensitivity value $sv^{\mathbf{x}}_{\mathbf{v}}$ in direction $\mathbf{v}$ of at most $sv^{\mathbf{x}}_{\max}$. In addition, consider the n one-way sensitivity functions $f_P^{(i)}(x_i)$, $i = 1, \ldots, n$, and let $\mathbf{s} = (sv^{x_1}, \ldots, sv^{x_n})$ be a vector of their one-way signed sensitivity values $sv^{x_i} = \frac{d\,f_P^{(i)}}{dx_i}(x_i^o)$. Then*

1. $sv^{\mathbf{x}}_{\max} = |\mathbf{s}| = \sqrt{\sum_i (sv^{x_i})^2}$
2. $\mathbf{s} = \nabla f_P(\mathbf{x^o})$
3. $sv^{\mathbf{x}}_{\mathbf{v}} = \mathbf{s} \bullet \mathbf{u}$, *where unit vector* $\mathbf{u}$ *is the normalised vector of* $\mathbf{v}$
4. *if* $|sv^{x_i}| < \frac{1}{\sqrt{n}}$ $\forall i$, *then* $sv^{\mathbf{x}}_{\max} < 1$ *and* $sv^{\mathbf{x}}_{\min} > -1$

Proof. Equalities *1.* and *2.* follow directly from the definition of the signed one-way sensitivity value and Proposition 2. Equality *3.* then follows directly from the definition of $sv^{\mathbf{x}}_{\mathbf{v}}$. Inequality *4.* follows directly from Equality *1.* □

Note that the above corollary can be exploited both in the context of an indirect and a direct approach to computing partial derivatives. Moreover, inequality *4.* enables us to analyse what combinations of parameters may *not* be interesting enough to investigate further during a sensitivity analysis, allowing us to focus on more important parameters.

Example 4. We illustrate, using Fig. 3, the fact that some combinations of parameters may not be interesting enough for further investigation. This figure gives $sv^{x_1,x_2}_{\max}$ as a function of sv^{x_1} and sv^{x_2} for sensitivity values $|sv^{x_i}| < 1$. The figure in addition shows the plane $sv^{x_1,x_2}_{\max} = 1$. The fraction of combinations of sv^{x_1} and sv^{x_2} that result in $sv^{x_1,x_2}_{\max} < 1$ is found below the plane $sv^{x_1,x_2}_{\max} = 1$ and equals $\frac{\pi}{4} \approx 0.785$. From inequality *4.* of Corollary 1 it follows that, in order to result in a two-way sensitivity value ≥ 1, the absolute value of at least one of the individual values has to be $\geq \frac{1}{\sqrt{2}}$. Thus whenever both $|sv^{x_1}|$ and $|sv^{x_2}|$ are < 0.71, we can be sure that $sv^{x_1,x_2}_{\max} < 1$ and that $sv^{x_1,x_2}_{\min} > -1$, that is,

$|sv_{\mathbf{v}}^{x_1,x_2}| < 1$ for any $\mathbf{v}$. If one of the two parameters, however, has a one-way sensitivity value ≥ 0.71, it depends on the sensitivity value of the other parameter whether $sv_{\max}^{x_1,x_2} \geq 1$ or not.

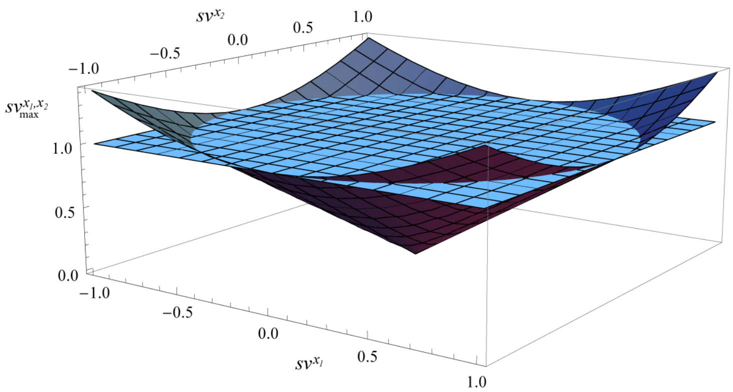

Fig. 3. $sv_{\max}^{x_1,x_2}$, as a function of $sv^{x_1}, sv^{x_2} \in \langle -1, 1 \rangle$ and the plane $sv_{\max}^{x_1,x_2} = 1$.

4.3 Joint vs Synergistic Effect in n-way Analyses

We observe that the absolute value $|sv_{\mathbf{v}}^{\mathbf{x}}|$ may be higher (or lower) than each of the individual absolute values $|sv^{x_i}|$ of which it is composed. For example, in the network from Fig. 1 (see *Examples 2* and *3*), we had that $|sv^x| = 0.465$, $|sv^y| = 0.299$ and that a shift in the direction of $\mathbf{v} = (-0.1, 0.1)$ resulted in $|sv_{\mathbf{v}}^{x,y}| = 0.540$. A simultaneous shift thus may have a larger (or smaller) local effect on the outcome probability than each parameter shift separately.

We would like to note that this joint effect of multiple parameter changes is not the same as the *synergistic effect* of multiple parameter changes, as first described in [3]. A synergistic effect is caused by the fact that the exact form of the one-way sensitivity function of x_i, depends on the original values of the other parameters of the network, and thus may be different for different values of some other parameter x_j. Such a synergistic effect can only be present if the n-way sensitivity function of those parameters includes product terms of x_i and x_j. For a joint effect of multiple simultaneous changes, the presence of such product terms is not necessary. As mentioned in Section 2, given just parameters from a single CPT, a sensitivity function will not include product terms of its parameters. Given parameters from a single CPT, therefore, no synergistic effect will be observed; a joint effect, however, may be present.

5 Parameter Tuning

The theory we discussed in Sections 3 and 4 can be used to study the robustness of a network to small simultaneous parameter changes. Another area of application lies in parameter tuning. In building a network, we may want to adjust parameters in order to meet certain constraints. An example of such a constraint is $\Pr(y \mid \mathbf{e}) > t$ for some probability $\Pr(y \mid \mathbf{e})$ and a desired value t.

In [2], a method for parameter tuning is described in which the parameter adjustment is guided by the log-odd change of the varied parameters in order to keep the distance between the old and the new distribution as small as possible. The paper moreover describes a method to compute the solution space of all possible parameter changes that would fulfill the constraint. This method, although not mentioned as such by the authors, in essence provides for computing the constants of the sensitivity function of the varied parameters, and is exponential in the number of CPTs from which the parameters are chosen. The method thus is feasible only if the adjusted parameters come from a limited number of different CPTs.

We now propose another tuning approach, based on the theory introduced in Sections 3 and 4. In this approach, our goal is to satisfy some constraint by adjusting a given set of n parameters as little as possible, that is, by keeping the sum of the absolute values of the parameter changes as low as possible. In our method the direction of the maximal change is used to guide the parameter changes. Since the gradient $\mathbf{s} = (sv^{x_1}, \ldots, sv^{x_n})$ of an n-way sensitivity function at x^o gives the direction of local maximal increase of the outcome probability, we will simultaneously adjust the n parameters in or against the direction of $\mathbf{s}$ to achieve the desired value. As long as the changes needed are small, this adjustment will be a good approximation of the adjustments required to satisfy the desired constraint with a minimal change of the parameters.

The adjustments needed can be assessed using a sensitivity function in which the parameters are constrained to variation only in or against the direction of $\mathbf{s}$. Below we first define an n-way sensitivity function given parameter changes in or against a fixed direction $\mathbf{v}$ to be a *sliced sensitivity function* in the direction of $\mathbf{v}$.

Definition 4 (sliced sensitivity function). *Let $f_{\Pr(y|\mathbf{e})}(\mathbf{x})$ be an n-way sensitivity function. A sliced sensitivity function of f in the direction of $\mathbf{v}$, denoted $f^{\mathbf{v}}_{\Pr(y|\mathbf{e})}$, expresses $\Pr(y \mid \mathbf{e})$ as a function of the change of the parameters $\mathbf{x}$ in or against direction $\mathbf{v}$ only.*

The following proposition shows that a sliced sensitivity function can be expressed in a single parameter and takes the form of a fraction of two polynomial functions of degree at most the number of CPTs from which the parameters are chosen.

Proposition 4. *Consider an n-way sensitivity function $f_{\Pr(y|\mathbf{e})}(\mathbf{x})$ for an output probability $\Pr(y \mid \mathbf{e})$, and a change of its parameters $\mathbf{x} = (x_1, \ldots, x_n)$ in a fixed direction $\mathbf{v} = (v_1, \ldots, v_n)$. Then for any x_i, $i = 1, \ldots, n$, with $v_i \neq 0$ there exists a sliced sensitivity function $f^{\mathbf{v}}_{\Pr(y|\mathbf{e})}(x_i)$ of the form:*

$$f^{\mathbf{v}}_{\Pr(y|\mathbf{e})}(x_i) = \frac{c_0 + c_1 \cdot x_i^1 + \ldots + c_m \cdot x_i^m}{d_0 + d_1 \cdot x_i^1 + \ldots + d_m \cdot x_i^m}$$

where each x_i^k, $k = 1, \ldots, m$, is a polynomial term of degree k and m is the number of different CPTs from which $x_1, \ldots, x_n$ are chosen.

Proof. Given a change in or against a fixed direction $(v_1, \ldots, v_n)$, we can express all parameters x_j in parameter x_i, since $(x_j - x_j^o) = \frac{v_j}{v_i} \cdot (x_i - x_i^o) \Leftrightarrow x_j = \frac{v_j}{v_i} \cdot (x_i - x_i^o) + x_j^o$, which is linear in x_i. Product terms of parameters in the n-way function $f_{\Pr(y|\mathbf{e})}(\mathbf{x})$ will now result in polynomial terms in $f^{\mathbf{v}}_{\Pr(y|\mathbf{e})}(x_i)$, of which the degree is determined by the number of interacting parameters in $f_{\Pr(y|\mathbf{e})}(\mathbf{x})$. This number equals at most the number of CPTs from which the parameters $x_1, \ldots, x_n$ are chosen. □

The n-way sensitivity function in any fixed vector direction $\mathbf{v}$ thus is a polynomial with as maximum degree the number of CPTs m from which the parameters are chosen and is determined by just $2 \cdot (m+1)$ constants. This observation also holds for the direction $\mathbf{s}$ of maximal increase. A constraint on $\Pr(y \mid \mathbf{e})$ now can be expressed in terms of a sliced sensitivity function in the direction of $\mathbf{s}$, and a feasible solution with minimal parameter change in or against the direction of $\mathbf{s}$ can be derived, if any. The solutions of a polynomial equation can be established analytically for polynomials up to degree 4; solutions for higher degree polynomials can be approximated.

In the above we assumed a *given* set of parameters $\{x_1, \ldots, x_n\}$. Note that a reasonable heuristic for choosing a set of parameters to adjust can be based on the one-way sensitivity values since $sv^{\mathbf{x}}_{\max} = \sqrt{\sum_i (sv^{x_i})^2}$; i.e. selecting the n parameters with highest one-way sensitivity value will allow for the largest possible local shift in the output upon their simultaneous perturbation.

Example 5. Consider again the example network from Fig. 1. Suppose we know that the outcome probability $\Pr(b \mid ct\, sh\, c)$ should be at least 0.80. In the network as it is we find that $\Pr(b \mid ct\, sh\, c) = 0.76$ so some parameter adjustment is required. First the signed one-way sensitivity values sv^x for all independent parameters x of the network are assessed[2]; these are given in Table 1. We observe that the parameters which will affect the outcome probability the most are $x = \Pr(b \mid \overline{mc})$ and $y = \Pr(ct \mid \bar{b})$. Suppose that we want to satisfy our constraint by adjusting those two parameters. The direction of maximal increase is $\mathbf{v} = (1.87, -1.81) \sim (1, -0.97)$. Tying y to x, the sliced sensitivity function in the direction of the maximal change now is

$$f^{\mathbf{v}}_{\Pr(b|ct\, sh\, c)}(x) = \frac{0.02432 + 0.4864 \cdot x}{0.0478424 + 0.318496 \cdot x + 0.09312 \cdot x^2}$$

Solutions to $f^{\mathbf{v}}_{\Pr(b|ct\, sh\, c)}(x) = 0.80$ are $x \approx 0.061$ and $x = 3.047$, where only the former is feasible. Parameter values that will satisfy $\Pr(b \mid ct\, sh\, c) \geq 0.80$ thus are $x = \Pr(b \mid \overline{mc}) = 0.061$ and $y = \Pr(ct \mid \bar{b}) = y^o - 0.97 \cdot (0.061 - x^o) = 0.10 - 0.97 \cdot (0.061 - 0.05) = 0.089$.

[2] Recall that dependent parameters are included in the analysis by covariation.

Table 1. Sensitivity values sv^x for independent parameters of the example network from Fig. 1.

x	x^o	sv^x
$\Pr(mc)$	0.20	0.12
$\Pr(b \mid mc)$	0.20	0.56
$\Pr(b \mid \overline{mc})$	0.05	1.87
$\Pr(isc \mid mc)$	0.80	-0.09
$\Pr(isc \mid \overline{mc})$	0.20	-0.41
$\Pr(ct \mid b)$	0.95	0.19
$\Pr(ct \mid \bar{b})$	0.10	-1.81
$\Pr(c \mid b\ isc)$	0.80	0.11
$\Pr(c \mid b\ \overline{isc})$	0.80	0.11
$\Pr(c \mid \bar{b}\ isc)$	0.80	-0.20
$\Pr(c \mid \bar{b}\ \overline{isc})$	0.05	-0.46
$\Pr(sh \mid b)$	0.80	0.23
$\Pr(sh \mid \bar{b})$	0.60	-0.30

6 Discussion

The robustness of Bayesian networks to changes in their parameter probabilities can be studied with a sensitivity analysis. Since the study of multiple simultaneous parameter shifts is computationally expensive, most research has focused on one-way sensitivity analyses in which only a single parameter is varied at a time. An important notion in sensitivity analysis is the notion of *sensitivity value*, which captures the sensitivity of some outcome of the network to a small change in a single parameter under consideration. In this paper we generalised this concept to multiple dimensions and proved that the computation of such an n-way sensitivity value can be done efficiently from the one-way sensitivity values of the parameters under consideration.

In contrast to a one-way sensitivity value, an n-way sensitivity value varies depending on the direction of shift under consideration. We expressed the direction of maximal change in terms of one-way sensitivity values and provided bounds on the n-way sensitivity value. We argued that the maximal (minimal) sensitivity value and the corresponding direction of maximal change is not only useful for studying the robustness of a Bayesian network, but can also be used in the context of network tuning. For small parameter changes, a shift of the parameters in or against the direction of the maximal increase until some tuning constraint is met will yield a good approximation of the minimal parameter change necessary to meet this constraint. We also proved that, since a fixed vector direction of change ties all parameters linearly, the effect of a parameter shift in the direction of the maximal change on some outcome probability can be efficiently established. To this end we introduced the concept of *sliced sensitivity function* for a sensitivity function that captures such a linearly tied parameter shift.

In a sliced sensitivity function variables are tied linearly. Variables, however, can also be tied by some other meaningful relationship. In [2], for example, parameters are tied by their log-odds ratio changes. In future research, we would like to expand the notion of sliced sensitivity function to more general forms of constrained sensitivity functions and explore the use of these functions both within and outside the field of parameter tuning.

Acknowledgements. This research was supported by the Netherlands Organisation for Scientific Research.

References

1. Castillo, E., Gutiérrez, J.M., Hadi, A.S.: Parametric Structure of Probabilities in Bayesian Networks. In: Froidevaux, C., Kohlas, J. (eds.) ECSQARU 1995. LNAI, vol. 946, pp. 89–98. Springer, Heidelberg (1995)
2. Chan, H., Darwiche, A.: Sensitivity Analysis in Bayesian Networks: From Single to Multiple Parameters. In: Chickering, M., Halpern, J. (eds.) Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence, pp. 67–75. AUAI Press, Arlington, VA (2004)
3. Coupé, V.M.H., Van der Gaag, L.C., Habbema, J.D.F.: Sensitivity Analysis: An Aid for Belief Network Quantification. The knowledge Engineering Review 15, 215–232 (2000)
4. Coupé, V.M.H., Van der Gaag, L.C.: Properties of Sensitivity Analysis of Bayesian Belief Networks. Annals of Mathematics and Artificial Intelligence 36, 323–356 (2002)
5. Darwiche, A.: A Differential Approach to Inference in Bayesian Networks. Journal of the ACM 50, 280–305 (2003)
6. Van der Gaag, L.C., Renooij, S.: Analysing Sensitivity Data from Probabilistic Networks. In: Breese, J., Koller, D. (eds.) Proceedings of the 17th Conference on Uncertainty in Artificial Intelligence, pp. 530–537. Morgan Kaufmann, San Francisco (2001)
7. Jensen, F.V., Nielsen, T.D.: Bayesian Networks and Decision Graphs (2nd ed). Springer Verlag (2007)
8. Kjærulff, U., Van der Gaag, L.C.: Making Sensitivity Analysis Computationally Efficient. In: Boutilier, C., Goldszmidt, M. (eds.) Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence, pp. 317–325. Morgan Kaufmann, San Francisco (2000)
9. Laskey, K.B.: Sensitivity Analysis for Probability Assessments in Bayesian Networks. IEEE Transactions on Systems, Man and Cybernetics 25, 901–909 (1995)