Sensitivity Analysis of Probabilistic Networks

Linda C. van der Gaag[1], Silja Renooij[1], and Veerle M.H. Coupé[2]

[1] Department of Information and Computing Sciences, Utrecht University, Utrecht, The Netherlands — {silja,linda}@cs.uu.nl
[2] Department of Clinical Epidemiology and Biostatistics, VU University Medical Centre, Amsterdam, The Netherlands — v.coupe@vumc.nl

Abstract. Sensitivity analysis is a general technique for investigating the robustness of the output of a mathematical model and is performed for various different purposes. The practicability of conducting such an analysis of a probabilistic network has recently been studied extensively, resulting in a variety of new insights and effective methods, ranging from properties of the mathematical relation between a parameter and an output probability of interest, to methods for establishing the effects of parameter variation on decisions based on the output distribution computed from a network. In this paper, we present a survey of some of these research results and explain their significance.

1 Introduction

Sensitivity analysis is a general technique for investigating the effects of inaccuracies in the parameters of a mathematical model on the model's output. As a mathematical model, a probabilistic network can also be subjected to such an analysis. The basic idea of the analysis then is to systematically vary the assessments for the network's parameter probabilities over a plausible interval and study the effects on the output computed from the network.

Sensitivity analysis of a probabilistic network can be performed for various different purposes. The parameter probabilities for a network are generally estimated from statistical data or assessed by human experts in the domain of application. As a consequence of incompleteness of data and partial knowledge of the domain, the assessments obtained inevitably are inaccurate [11]. Since the output probabilities of a network are built from these assessments, they may be sensitive to the inaccuracies involved and may even be unreliable. In general, however, not every parameter will require the same level of accuracy to arrive at satisfactory behaviour of the network; some probabilities will typically have more impact on the network's output than others. During model construction, therefore, sensitivity analysis can be used to gain detailed insight into the level of accuracy that is required for the various parameters and to guide further knowledge elicitation efforts [7]. Moreover, if during an initial evaluation of the network with real-life data, the output of the network is different from what is expected based upon knowledge of the domain, then sensitivity analysis can be used to identify which parameters can be changed to arrive at the expected output [2–4]. Furthermore, upon

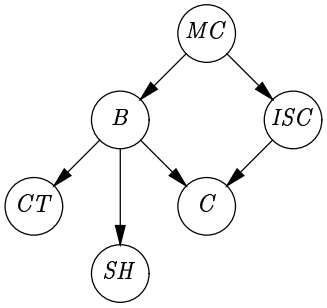
using the network in a real-life setting, sensitivity analysis can be used to gain insight into the robustness of an output probability of interest as well as of a decision based upon this probability. The analysis thereby reveals the range of parameter values for which the network's output is valid.

For a probabilistic network, the simplest type of analysis is to systematically vary one of the network's parameter probabilities while keeping all other parameters fixed. Such an analysis serves to reveal the effect of just the parameter whose assessment is being varied, on the output probability of interest. A sensitivity analysis in which a single parameter is varied, is termed a *one-way sensitivity analysis*. In a *two-way sensitivity analysis* of a probabilistic network, two parameters are varied simultaneously. In addition to the separate effects of variation of the two parameters, a two-way sensitivity analysis reveals the joint effect of their variation on a probability of interest. In essence, it is also possible to systematically vary more than two parameters at the same time. The results of such an *n-way sensitivity analysis*, however, are often hard to interpret.

In recent years, one-way sensitivity analysis of probabilistic networks has been studied extensively, which has resulted in practicable methods for performing such an analysis. In this paper, we present a survey of some of the recent research results and explain their significance. The paper is organised as follows. In Sect. 2, we introduce some preliminaries on probabilistic networks and introduce our running example. Sect. 3 describes the functional relationship between a parameter being varied and the output probability computed from a network; this section further discusses the computation of these sensitivity functions and presents some bounds on the effects of parameter variation. In Sect. 4, we discuss the application of sensitivity analysis to study the robustness of output probabilities computed from a probabilistic network on the one hand, and the robustness of decisions based upon these probabilities on the other hand. The paper ends with our concluding remarks and some directions for further research in Sect. 5.

2 Preliminaries

A *probabilistic network* basically is a representation of a joint probability distribution Pr over a set of stochastic variables. It consists of a qualitative part and an associated quantitative part. The network's qualitative part takes the form of an acyclic directed graph. Each node in this digraph represents a variable that takes its value from a finite set of discrete values. The arcs in the digraph model the influential relationships among the represented variables; more specifically, they capture independence by means of the d-separation criterion [20]. The strengths of the influential relationships are described by conditional probability distributions: for each variable V, conditional distributions $p(V \mid \pi(V))$ over its values are specified conditional on the various possible combinations of values for its set of parents $\pi(V)$ in the digraph.



$p(mc) = 0.20$

$p(b \mid mc) = 0.20$
$p(b \mid \overline{mc}) = 0.05$

$p(c \mid b, isc) = 0.80$
$p(c \mid \overline{b}, isc) = 0.80$
$p(c \mid b, \overline{isc}) = 0.80$
$p(c \mid \overline{b}, \overline{isc}) = 0.05$

$p(isc \mid mc) = 0.80$
$p(isc \mid \overline{mc}) = 0.20$

$p(ct \mid b) = 0.95$
$p(ct \mid \overline{b}) = 0.10$

$p(sh \mid b) = 0.80$
$p(sh \mid \overline{b}) = 0.60$

Fig. 1. The *Brain tumour* network.

The specified probabilities are often referred to as the network's *parameters*. A probabilistic network in essence allows for the computation of any probability of interest over its variables. For this purpose, various efficient algorithms are available [15, 18, 20, 23], which we will refer to as the (*standard*) *propagation algorithms*. In the sequel, we will explicitly distinguish computed probabilities, written as Pr, from the parameter probabilities specified in the network, which are denoted by p.

For our running example we consider the *Brain tumour* network shown in Fig. 1. The network captures some (fictitious and incomplete) medical knowledge, adapted from [5]. It describes the problems associated with metastatic cancer for an arbitrary patient in oncology. Metastatic cancer (modelled by the variable MC) may lead to the development of a brain tumour (B) and typically gives rise to an increased level of serum calcium (ISC). The presence of a brain tumour can be established from a CT scan (CT). Severe headaches (SH) may also be indicative of the presence of a brain tumour. Either a brain tumour or an increased level of serum calcium are likely to ultimately cause a patient to fall into a coma (C). The digraph modelling the relationships among the six variables is shown on the left of the figure. In the network, all variables V are binary, taking one of the values *true* and *false*; we write v if V has the value *true* and $\overline{v}$ if it has the value *false*. The various parameter probabilities associated with the digraph are shown on the right of the figure. The probabilities specified for the variable ISC, for example, express that knowing whether or not metastatic cancer is present has a considerable influence on the probability of finding an increased level of serum calcium in an arbitrary patient. On the other hand, severe headaches are expressed as being quite common in both patients with and without a brain tumour.

3 Sensitivity Functions and their Computation

Sensitivity analysis in general is a technique for investigating the effects of inaccuracies in the parameters of a mathematical model on its output. For a

probabilistic network, more specifically, sensitivity analysis amounts to studying the effects of variation of the network's parameter probabilities on the output probabilities computed from the network. In this section, we will argue that any output probability of interest can be expressed as a simple mathematical function in the parameter under study. We further review the computational issues involved in establishing such a function.

3.1 Sensitivity functions

Performing a sensitivity analysis of a probabilistic network in essence amounts to establishing, for each parameter and each output probability of interest, the *sensitivity function* that expresses this output probability in terms of the parameter being varied [6]. These sensitivity functions have a highly constrained functional form as a result of the graphical structure of a probabilistic network. Before elaborating on this functional form, we introduce some further notational conventions. We will denote a probability of interest by $\Pr(A = a \mid e)$, or $\Pr(a \mid e)$ for short, where a is a specific value of a variable A of interest and e denotes the available (possibly compound) evidence. A parameter under study will be denoted by $x = p(b_i \mid \pi)$, where b_i is a value of a variable B and π is a combination of values for the parents of B. We use $f_{\Pr(a|e)}(x)$ to denote the function that expresses the probability $\Pr(a \mid e)$ in terms of the parameter x; we often omit the subscript for the function symbol f, as long as ambiguity cannot occur.

In a one-way sensitivity analysis, a single parameter $x = p(b_i \mid \pi)$ for some variable B is varied. Upon varying this parameter, the other parameters $p(b_j \mid \pi)$, $j \neq i$, specified for the same variable need be *co-varied* to ensure that the parameters from the same distribution keep summing to one. Each such parameter $p(b_j \mid \pi)$ can thus be seen as a function $p(b_j \mid \pi)(x)$ of the parameter x under study. We assume that the parameters $p(b_j \mid \pi)$ are covaried with $p(b_i \mid \pi)$ in such a way that their mutual proportional relationship is kept constant, that is,

$$p(b_j \mid \pi)(x) = p(b_j \mid \pi) \cdot \frac{1 - x}{1 - p(b_i \mid \pi)}$$

for $p(b_i \mid \pi) < 1$. This scheme of *proportional co-variation* has been shown to result in the smallest change, given the variation of the parameter x under study, in the output distribution [3].

Now, under the assumption of proportional co-variation, any sensitivity function $f_{\Pr(a|e)}(x)$ is a quotient of two functions that are linear in the parameter x under study [1,9]. More formally, the function takes the form

$$f_{\Pr(a|e)}(x) = \frac{c_1 \cdot x + c_2}{c_3 \cdot x + c_4}$$

where the constants $c_1, \ldots, c_4$ are built from the original assessments for the parameters that are not being varied. The numerator of the function in

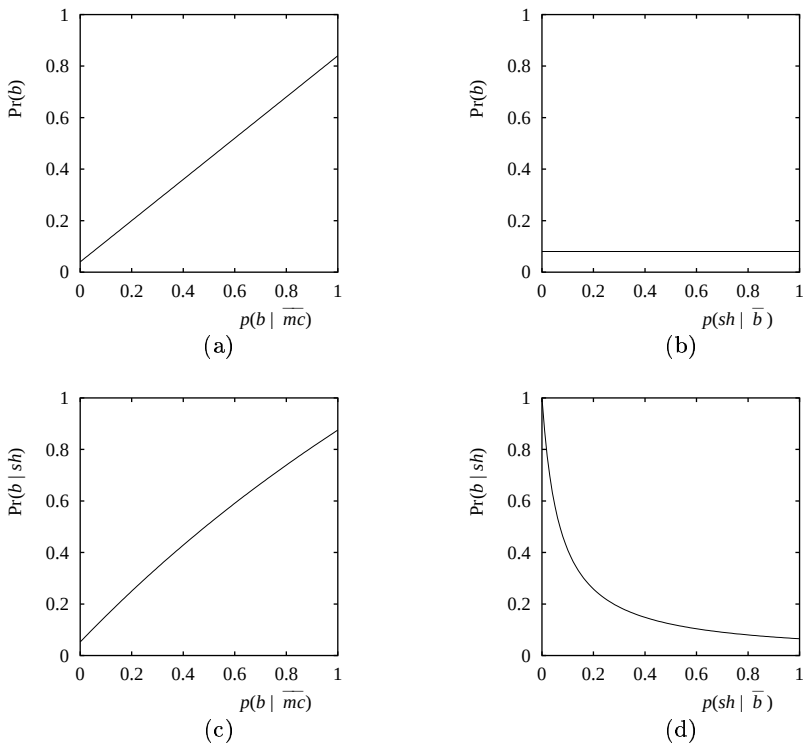


Fig. 2. Some sensitivity functions for the *Brain tumour* network; the effects of varying the parameters $p(b \mid \overline{mc})$ and $p(sh \mid \overline{b})$, respectively, on the prior probability $\Pr(b)$ (**a**,**b**) and on the posterior probability $\Pr(b \mid sh)$ (**c**,**d**) are shown.

essence describes the probability $\Pr(a, e)$ as a function of the parameter x; its denominator describes $\Pr(e)$ as a function of x. We observe that the probability distribution Pr defined by a probabilistic network can be written as a product of its parameter distributions. From the property of marginalisation, it then follows that both $\Pr(a, e)$ and $\Pr(e)$ can be written as a sum of products of parameters, one of which is the parameter under study.

We illustrate the form of a sensitivity function in general by studying, for the *Brain tumour* network, the sensitivity functions that describe the effects of varying the parameter probabilities $p(b \mid \overline{mc})$ and $p(sh \mid \overline{b})$ on the output probabilities $\Pr(b)$ and $\Pr(b \mid sh)$, respectively; the four functions are shown in Fig. 2. Fig. 2(c) and (d) show the sensitivity functions that express the output probability $\Pr(b \mid sh)$ in terms of the two separate parameters. Both functions are non-linear and monotone, and can in fact be written as a quotient of two linear functions. The sensitivity function expressing $\Pr(b \mid sh)$

in terms of $x = p(b \mid \overline{mc})$, for example, equals

$$f_{\Pr(b|sh)}(x) = \frac{0.640 \cdot x + 0.032}{0.160 \cdot x + 0.608} = \frac{4.0 \cdot x + 0.2}{x + 3.8}$$

Both sensitivity functions reveal that varying the parameter under study can have a considerable effect on the output probability of interest. We consider, for example, smaller values of the parameter $p(sh \mid \overline{b})$, meaning that it is becoming less likely to find severe headaches in patients without a brain tumour. If severe headaches then are found in a particular patient, this finding becomes more indicative of the presence of a brain tumour. This type of sensitivity is commonly found in real-life diagnostic networks [14].

Fig. 2(a) and (b) show the sensitivity functions that express the prior probability $\Pr(b)$ in terms of the two separate parameters. We observe that both functions are linear in the parameter under study. The sensitivity function expressing $\Pr(b)$ in terms of $x = p(b \mid \overline{mc})$, for example, equals

$$f_{\Pr(b)}(x) = 0.8 \cdot x + 0.04$$

In the absence of evidence, any probability of interest relates linearly to any network parameter. Linear functions are also found if the parameter under study pertains to an ancestor of the variable of interest and the parameter's variable has no observed descendants in the network's qualitative part. From Fig. 2(b) we further observe that variation of the parameter $p(sh \mid \overline{b})$ has no effect at all on our probability of interest; the associated sensitivity function is a constant function. If no information is available about the presence or absence of severe headaches, varying the diagnostic weight of headaches cannot have any impact on the output. In general, the sensitivity function is a constant function for any parameter associated with a variable that is not included in the sensitivity set of the variable of interest. This set of variables whose parameters may affect the probability of interest upon variation, is readily identified by simple inspection of the network's qualitative part [9].

From the four sensitivity functions in Fig. 2 we note that by entering evidence into a network, the exhibited sensitivities may change. While the observation of the presence of severe headaches hardly influences the effect of varying the parameter associated with the variable B, for example, we find that this same evidence changes the effect of varying the parameter associated with SH to a considerable extent. In general, quite different patterns of sensitivity may arise for different profiles of evidence.

For our discussion in the subsequent sections, it is convenient to observe that a sensitivity function is either a *linear* function or a fragment of a *rectangular hyperbola*. A rectangular hyperbola takes the general form

$$f(x) = \frac{r}{x - s} + t$$

where, for a hyperbolic sensitivity function, we have that

$$s = -\frac{c_4}{c_3}, \quad t = \frac{c_1}{c_3}, \quad \text{and} \quad r = \frac{c_2}{c_3} + s \cdot t$$

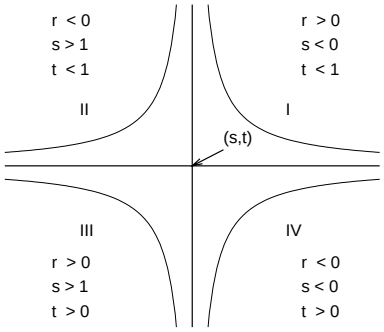


Fig. 3. Hyperbolas and their constants (specific for sensitivity functions).

with $c_1, \ldots, c_4$ as before. A rectangular hyperbola has two branches and two asymptotes. Fig. 3 illustrates the locations of the possible hyperbola branches relative to the two asymptotes. For $r < 0$, the branches lie in the second (II) and fourth (IV) quadrants relative to the asymptotes $x = s$ and $f(x) = t$; for $r > 0$, the branches are found in the first (I) and third (III) quadrants. Since any sensitivity function is well-defined for $x \in [0, 1]$, a hyperbolic sensitivity function is a fragment of one of the four possible hyperbola branches; the area with $0 \leq x \leq 1$ and $0 \leq f(x) \leq 1$ that defines the fragment, is called the *unit window* for the function. The vertical asymptote $x = s$ of the hyperbola lies either to the left of the unit window or to the right, that is, we have that $s < 0$ or $s > 1$. For first- and fourth-quadrant sensitivity functions more specifically we find $s < 0$, while for second- and third-quadrant functions we have that $s > 1$. In addition, the horizontal asymptote $f(x) = t$ of the hyperbola lies below $f(1)$ for first-quadrant functions and below $f(0)$ for second-quadrant functions; we then have that $t < 1$. The horizontal asymptote lies above $f(0)$ for third-quadrant and above $f(1)$ for fourth-quadrant functions, which implies that $t > 0$. Note that the horizontal asymptote of the hyperbola not necessarily lies within the unit window for the sensitivity function: for a first-quadrant function, for example, negative values of t are possible and for a fourth-quadrant function values of t larger than 1 can be found.

We illustrate the hyperbolic form by writing the sensitivity function for the posterior probability $\Pr(b \mid sh)$ and the parameter $x = p(b \mid \overline{mc})$ as a hyperbolic function:

$$f_{\Pr(b|sh)}(x) = -\frac{15.0}{x + 3.8} + 4.0$$

The three constants $s = -3.8, t = 4.0$, and $r = -15.0$ indicate that the sensitivity function is a fourth-quadrant function and, hence, is monotonically increasing. The vertical asymptote $x = -3.8$ is located at some distance to the left of the unit window; the horizontal asymptote $f(x) = 4.0$, moreover, is located at some distance above the unit window. These properties indicate that no major changes in the derivative of the function are expected within

the unit window, as is confirmed by Fig. 2(c). For the first-quadrant function from Fig. 2(d) in contrast, the constants $s = -0.07$ and $t = 0$ are found.

3.2 Computing sensitivity functions

In the previous subsection, we argued that the effect of varying a single parameter upon an output probability computed from a probabilistic network can be described by a highly constrained mathematical function. The major advantage of this constrained form is that any sensitivity function can be established by computing just its constants.

The simplest method of determining the constants of a sensitivity function amounts to computing, from the network, the probability of interest for up to three values for the parameter under study; using the functional form of the function to be established, a system of linear equations is obtained, which can subsequently be solved [9]. For the network computations, any of the standard propagation algorithms can be used. With this method, the sensitivity function for a single parameter and a single output probability is established. A more efficient method determines the required constants by propagating information through a junction tree, similar to the standard junction-tree propagation algorithm [16]. This method requires a very small number of inward and outward propagations in the tree to determine either the constants of *all* sensitivity functions that relate the probability of interest to any one of the network parameters, or to determine the sensitivity functions for any output probability in terms of a single parameter.

In a full sensitivity analysis, the effects of varying all parameter probabilities of a network on all output probabilities of interest are investigated. From the example sensitivity functions for the *Brain tumour* network, we have seen that the effect of varying a particular parameter can change considerably for different output probabilities. For a full analysis, therefore, all sensitivity functions of interest need be established explicitly, for which purpose the two methods outlined above need be applied multiple times. Although the concept of sensitivity set can be used to forestall some of the computations involved, performing a full sensitivity analysis of a probabilistic network of realistic size is highly time consuming.

3.3 Bounding sensitivity functions

For real-life probabilistic networks, performing a full sensitivity analysis is infeasible in practice, mostly as a consequence of the large range of evidence profiles to be studied. Recent research therefore focused on the derivation of general bounds for sensitivity functions. These bounds provide for selecting, without actually performing the analysis for the full range of profiles, the sensitivity functions that are the most likely to reveal high sensitivities; we will return to this observation in Sect. 4.

Within a given network, we consider a parameter x for which the value x_0 is specified. We are interested in some prior or posterior probability p, for which the value p_0 is computed from the network, using the value x_0 for the parameter x. We will refer to the values x_0 and p_0 as the *original* values of the parameter and of the output probability, respectively; we assume that neither x_0 nor p_0 is equal to zero or one. Without any further knowledge of the network, we only know that the sensitivity function for x and p passes through the point (x_0, p_0). Now, any sensitivity function through this point is bounded by the two rectangular hyperbola's $i(x)$ and $d(x)$ [21], with

$$i(x) = \frac{p_0 \cdot (1 - x_0) \cdot x}{(p_0 - x_0) \cdot x + (1 - p_0) \cdot x_0} \text{ and } d(x) = \frac{p_0 \cdot x_0 \cdot (1 - x)}{(1 - p_0 - x_0) \cdot x + p_0 \cdot x_0}$$

These bounds follow from the observation that an increasing rectangular hyperbola can, in the most extreme case, pass through the points $(0, 0)$ and $(1, 1)$; a decreasing hyperbola may pass through $(0, 1)$ and $(1, 0)$.

For any sensitivity function $f(x)$ with $f(x_0) = p_0$, we now have that

$$\min\{i(x_j), d(x_j)\} \leq f(x_j) \leq \max\{i(x_j), d(x_j)\}$$

for all $x_j \in [0, 1]$. From the bounding hyperbolas, we can readily establish numerical bounds on the new value p_1 of the probability of interest that results from varying the parameter x from its original value x_0 to x_1. More specifically, these bounds are given by [3]:

$$\frac{p_0 \cdot e^{-\delta}}{p_0 \cdot (e^{-\delta} - 1) + 1} \leq p_1 \leq \frac{p_0 \cdot e^{\delta}}{p_0 \cdot (e^{\delta} - 1) + 1}$$

where

$$\delta = \left| \ln \frac{x_1}{1 - x_1} - \ln \frac{x_0}{1 - x_0} \right|$$

We would like to note that, if we know that a sensitivity function is linear in the parameter under study, we can also establish linear bounding functions that may lead to tighter numerical bounds. These linear bounding functions also pass through the point (x_0, p_0). Since both functions moreover should be well-defined within the unit window, the increasing bounding function further passes through either $(0, 0)$ or $(1, 1)$, and the decreasing bounding function goes through $(0, 1)$ or $(1, 0)$ [21].

To illustrate the various functions, Fig. 4 depicts two sets of bounding functions for the *Brain tumour* network. Given the original value $x_0 = 0.05$ for the parameter $x = p(b \mid \overline{mc})$, the posterior probability $p = \Pr(b \mid sh)$ equals $p_0 = 0.10$. Fig. 4(a) now depicts the increasing and decreasing bounding hyperbolas through the point $(x_0, p_0) = (0.05, 0.10)$; the true sensitivity function for x and p is shown in the figure by a dashed curve. For the prior probability $\Pr(b)$, we know that the sensitivity function that expresses this

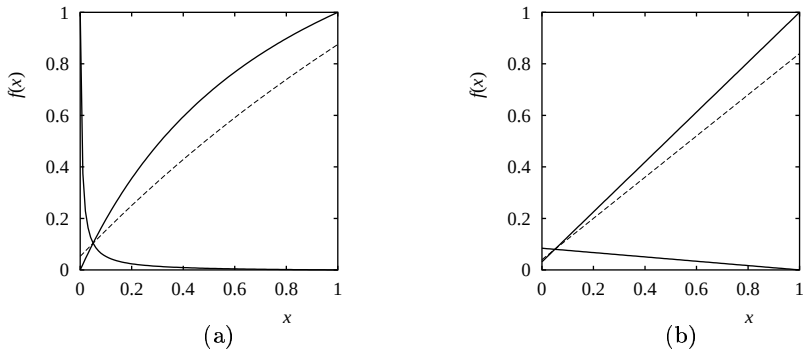


Fig. 4. Bounds on hyperbolic sensitivity functions through $(x_0, p_0) = (0.05, 0.10)$ (**a**) and on linear sensitivity functions through $(x_0, p_0) = (0.05, 0.08)$ (**b**); example sensitivity functions for the *Brain tumour* network are also shown (dashed).

probability in terms of the same parameter x is a linear function. Fig. 4(b) now depicts the linear bounding functions established for this function; the true sensitivity function again is shown by a dashed line. Note that knowledge of the sensitivity function being linear leads to tighter bounds than the bounding hyperbolas would give.

The bounding functions introduced above all depend on the original values of the parameter and of the probability of interest, but are independent of any other aspect of the network under study. The bounds therefore apply to any pair (x_0, p_0) for any network. Their computation, moreover, does not require any network propagations, except for establishing the value p_0. For a given profile of evidence, the bounds on the sensitivity function can be further tightened. To this end, the probability of the evidence $\Pr(e)$ is expressed as a linear function $f_{\Pr(e)}(x) = c_3 \cdot x + c_4$ of the parameter under study; this function is readily computed using the junction-tree algorithm outlined in the previous subsection. The two constants c_3 and c_4 now determine the constant $s = -\frac{c_4}{c_3}$ in the hyperbolic form of any sensitivity function passing through the point (x_0, p_0) and thereby constrain the space of feasible functions [22].

4 Applications of Sensitivity Analysis

One of the purposes to which a sensitivity analysis is performed of a probabilistic network, is to study the robustness of the network's output. We thereby distinguish between the output probabilities that are computed from the network and the decisions based upon these probabilities.

4.1 Robustness of output probabilities

In the previous section, we argued that a sensitivity function $f(x)$ provides for establishing the change in the output probability of interest that is occasioned

by a shift in the value of the parameter x under study. For the *Brain tumour* network, for example, we have from Fig. 2(c) that a shift in the value of the parameter $p(b \mid \overline{mc})$ occasions a more or less proportional shift in the output probability $\Pr(b \mid sh)$. A shift in the value of the parameter $p(sh \mid \overline{b})$ may have quite a considerable effect on this posterior probability; small deviations from the original value 0.60 for the parameter, however, will have little effect. The change in the output probability that is occasioned by a shift in the value of a parameter under study is indicative of the robustness of the output. Different concepts have been designed to measure this robustness. We discuss for this purpose the concepts of *sensitivity value*, which describes the change in an output probability for infinitesimally small shifts in a parameter under study, and *vertex proximity*, which indicates potential effects of larger shifts.

Sensitivity value We consider a parameter x with an original value of x_0. We further consider an output probability p and the sensitivity function $f(x)$ that expresses p in terms of x. The *sensitivity value* of x and p now describes the effect of infinitesimally small shifts in the parameter's original value on the probability of interest. It is defined as $|f'(x_0)|$, that is, as the absolute value of the first derivative of the sensitivity function at the original value x_0 of the parameter. The sensitivity value of the parameter x and the output probability of interest can be established analytically by computing the sensitivity function from the network, taking its first derivative

$$f'(x) = \frac{c_1 \cdot c_4 - c_2 \cdot c_3}{(c_3 \cdot x + c_4)^2}$$

and filling in the original value x_0 for the parameter under study. Alternatively, the sensitivity value can be established directly from the network by performing two network propagations [10, 17].

To illustrate the concept of sensitivity value, we establish, for our *Brain tumour* network, the derivative of the sensitivity function that expresses the output probability $\Pr(b \mid sh)$ in terms of the parameter $x = p(b \mid \overline{mc})$. We find that

$$f'(x) = \frac{0.384}{(0.16 \cdot x + 0.608)^2}$$

Since the parameter x has an original value of 0.05, we find a sensitivity value of $|f'(0.05)| = 1.01$. This sensitivity value indicates that small shifts in the parameter under study will induce a more or less similar change in the output probability. For the parameter $x = p(sh \mid \overline{b})$ having an original value of 0.60, in contrast, we find a sensitivity value equal to

$$\left| \frac{-0.059}{(0.92 \cdot 0.60 + 0.064)^2} \right| = 0.16$$

which indicates that small shifts in the parameter's original value will hardly have any effect on the output probability of interest. We note that these observations are confirmed by Fig. 2(c) and (d).

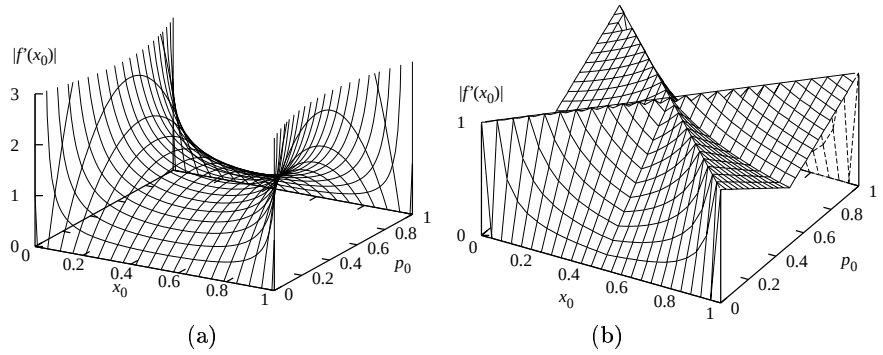


Fig. 5. The upper bound on the sensitivity value as a function of x_0 and p_0, for a hyperbolic sensitivity function (**a**) and for a linear sensitivity function (**b**).

In the previous section we argued that for a real-life probabilistic network it is generally infeasible to compute sensitivity functions given all profiles of evidence. By building upon the bounding functions for a sensitivity function, however, it is relatively straightforward to establish an upper bound on the sensitivity value [2, 21]. By taking the first derivatives of the hyperbolic bounding functions $i(x)$ and $d(x)$ for a parameter x with an original value of x_0 and an output probability with an original value of p_0, we find the following upper bound on the sensitivity value of x and p:

$$|f'(x_0)| \leq \frac{p_0 \cdot (1 - p_0)}{x_0 \cdot (1 - x_0)}$$

Fig 5(a) depicts this upper bound as a function of x_0 and p_0. The figure reveals that large sensitivity values are expected only for the more extreme values of x_0. For the sensitivity functions from Fig. 2(c) and (d), for example, with $p_0 = 0.10$, and $x_0 = 0.05$ for the parameter $p(b \mid \overline{mc})$ and $x_0 = 0.60$ for the parameter $p(sh \mid \overline{b})$, we find that the parameter with the smaller original value will have a sensitivity value of at most 1.96, whereas the less extreme value of the other parameter results in a bound of 0.39 on the sensitivity value. Note that the actual sensitivity values 1.01 and 0.16 for these parameters and the probability of interest are indeed below the established bounds.

We recall from the previous section that, if a sensitivity function is known to be linear in a parameter under study, then linear bounding functions can be established. From these linear bounding functions, also an upper bound on the sensitivity value can be computed. This bound will in general be tighter than the upper bound that is found from hyperbolic bounding functions [22]. Fig. 5(b) depicts the upper bound as a function of x_0 and p_0. As an example, we consider for our *Brain tumour* network the sensitivity function that describes the prior probability $\Pr(b)$ as a function of the parameter $x = p(b \mid \overline{mc})$; recall that this function is depicted in Fig. 2(a). With

$x_0 = 0.05$ and $p_0 = 0.08$, we find that the sensitivity value of this parameter and output probability is at most 0.97; the actual sensitivity value equals 0.80. Note that the sensitivity value for a linear function never exceeds 1.0, that is, a shift in the parameter under study can never result in a larger shift in the output probability of interest. Further note that the surface from Fig. 5(b) can be placed underneath that of Fig. 5(a), which confirms that knowledge of a sensitivity function being linear allows for tighter bounds.

In a full sensitivity analysis of a probabilistic network, typically a large range of different profiles of evidence need be studied, which renders the analysis infeasible in practice. The bounds on the sensitivity value established above can now be exploited for selecting the parameters for which a detailed analysis indeed is useful. We consider, as an example, a parameter with an original value of 0.5. From the bounds established above, we have that any output probability will be quite insensitive to small shifts in this parameter, regardless of the evidence profile under study. For such a parameter, therefore, a more detailed analysis is not required.

Vertex proximity The concept of sensitivity value reviewed above has been designed to give insight in the effect of very small parameter variations. As the initial assessments for the parameters of a probabilistic network may be highly inaccurate, however, we are interested in the effects of larger parameter shifts as well. We observe that for linear sensitivity functions, the sensitivity value in essence is a constant function in the value of the parameter under study. The computed sensitivity value therefore remains valid also for larger parameter shifts. For hyperbolic sensitivity functions, this property does not hold. In fact, the sensitivity value can strongly differ for two values of the parameter under study. To illustrate this observation, we consider again, for our *Brain tumour* network, the sensitivity function from Fig. 2(d) which expresses the output probability $p = \Pr(b \mid sh)$ in terms of the parameter $x = p(sh \mid \overline{b})$. The sensitivity value of x and p, that is, the absolute value of the function's first derivative, is depicted in Fig. 6 as a function of the value of x. The figure reveals that the sensitivity value of x and p equals at most 1.0 if the parameter adopts a value within the interval $[0.2, 1]$. For smaller values of the parameter, the sensitivity value rapidly grows to infinity. Now, if the original value of the parameter would be slightly larger than $x_0 = 0.2$, we would find a relatively small sensitivity value which would be interpreted as indicating that the output probability is not very sensitive to variation of the parameter. Yet, if a more accurate value for the parameter would be slightly smaller than 0.2, we would conclude that the network's output is not very robust as a result of the larger sensitivity value found.

For a hyperbolic sensitivity function, we now take the point $(x_v, f(x_v))$ where the first derivative $|f'(x_v)|$ equals 1.0 as the point that marks the transition from large sensitivity values to small ones, and vice versa. This point is called the *vertex* of the hyperbola branch under study and can easily

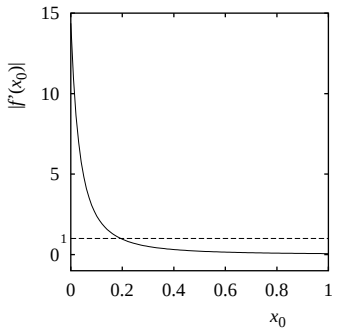


Fig. 6. The sensitivity value for the parameter $x = p(sh \mid \overline{b})$ and the output probability $\Pr(b \mid sh)$, as a function of x.

be computed from the constants of the sensitivity function using

$$x_v = \begin{cases} s + \sqrt{|r|}, \text{ if } s < 0 \\ s - \sqrt{|r|}, \text{ if } s > 1 \end{cases}$$

Now, if the original value x_0 of a parameter lies close to the x-value of the hyperbola's vertex, then the output probability of interest may be quite sensitive to variation of the parameter even if just a small sensitivity value is found. We say that the parameter's original value exhibits *vertex proximity*.

To illustrate the concept of vertex proximity, we consider again, for our *Brain tumour* network, the two sensitivity functions from Fig. 2(c) and (d). For the hyperbola branch that describes the output probability of interest as a function of the parameter $p(b \mid \overline{mc})$, with an original value of 0.05, we find that $x_v = -3.8 + \sqrt{15} = 0.07$. The vertex of the hyperbola branch thus is quite close to the parameter's original value. Since the sensitivity function is a fourth-quadrant hyperbolic function and $x_0 < x_v$, we conclude that the sensitivity value for values of x larger than x_v will be smaller than the sensitivity value of 1.01 computed for x_0; note that this observation is confirmed by Fig. 2(c). For the parameter $p(sh \mid \overline{b})$, the x-value of the vertex of the hyperbola branch is found around 0.20, which can be considered quite distant from the parameter's original value of 0.60. Both the sensitivity value of 0.16 and the lack of vertex proximity thus indicate that the output probability is not highly sensitivity to shifts in the parameter; note that this observation is confirmed by Fig. 2(d). If its original value had been 0.25, for example, then we would have found a sensitivity value of 0.68. The property of vertex proximity would then have been indicative of possibly significant effects of variation of the parameter to smaller values.

4.2 Robustness of decisions

More and more, probabilistic networks are used within decision-support systems, where decisions are recommended based upon some probability dis-

tribution established from the network. In such a system, robustness of the recommended decision is often more important than robustness of the output probabilities themselves. Now, for some network parameters variation will have a considerable effect on an output probability and yet not induce a change in decision; for other parameters, variation will show little effect on the output probability and nonetheless result in a different decision. For studying the robustness of a recommended decision, therefore, studying the effects of parameter variation on an output probability no longer suffices: the effects on the decision itself need be taken into consideration explicitly. We discuss to this end the issue of output robustness in view of the *threshold model for decision making*, where a decision is based upon an output probability computed from the network, and in view of a model for decision making that is based upon the *most likely value* for a variable of interest.

Threshold decision making The *threshold model* has been designed to support decision making for diagnostic problems in which the decision maker has to choose between gathering further evidence and acting without acquiring additional information. Although generally applicable, the model is used most notably for patient management in medical applications [19]. In such an application, the probability of disease is generally taken by an attending physician to decide upon management of a patient under consideration. The physician may decide to start treatment rightaway, or to withhold treatment altogether. Alternatively, if a diagnostic test can provide additional information which may affect the patient's probability of disease, then the physician may defer the treatment decision until this information has become available.

To support choosing among the various decision alternatives, the threshold model builds upon three threshold probabilities of disease. The *treatment threshold probability* $P^*(d)$ of a disease d being present, is the probability at which the physician is indifferent between giving treatment and withholding treatment; the *no treatment-test threshold probability* $P^-(d)$ is the probability at which the physician is indifferent between the decision to withhold treatment and the decision to obtain additional diagnostic information; and the *test-treatment threshold probability* $P^+(d)$ is the probability at which the physician is indifferent between obtaining further information and starting treatment rightaway. Now, as long as not all possible diagnostic tests have been performed, a physician has three decision alternatives at his disposal. Given the probability of disease $\Pr(d)$ for a patient, the model recommends the physician to withhold treatment if $\Pr(d) < P^-(d)$, to start treatment if $\Pr(d) > P^+(d)$, and to perform a diagnostic test if $P^-(d) \leq \Pr(d) \leq P^+(d)$. If no further tests are available, the physician has to choose between only two alternative decisions. The model recommends to start treatment if $\Pr(d) > P^*(d)$ and to withhold treatment if $\Pr(d) \leq P^*(d)$. Fig. 7 summarises the basic idea of the threshold model.

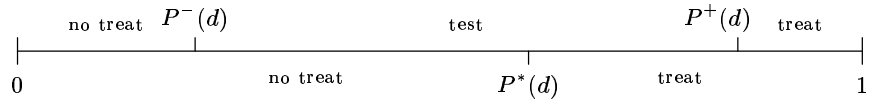


Fig. 7. The threshold model for patient management, indicating three threshold probabilities and the various decision alternatives at a physician's disposal.

For studying output robustness of a probabilistic network in view of the threshold model for decision making, the various threshold probabilities employed by the model should be taken into consideration [12]. More specifically, the effects of parameter variation on the recommended decision should be analysed. We consider an output probability of disease $\Pr(d \mid e)$ computed from the probabilistic network and suppose that $P^-(d) \leq \Pr(d \mid e) \leq P^+(d)$. Assuming that an appropriate diagnostic test is still available, the threshold model thus recommends to gather additional information. We further consider a parameter x from the network and the sensitivity function $f(x)$ that expresses the posterior probability of disease in terms of the parameter. We compute the values x^- and x^+ for the parameter x such that

$$f(x^-) = P^-(d) \quad \text{and} \quad f(x^+) = P^+(d)$$

Note that these values can be readily established from the sensitivity function. We now have that the decision to gather additional information remains unaltered as long as the original value of the parameter is not varied beyond the interval $[x^-, x^+]$. Similar bounds on variation can be computed for an output probability of disease smaller than $P^-(d)$ or larger than $P^+(d)$.

We illustrate investigating the robustness of decisions for our *Brain tumour* network. We assume that the threshold probabilities of a brain tumour being present have been set at $P^-(b) = 0.045$ and $P^+(b) = 0.56$. The prior probability of the presence of a brain tumour in an arbitrary patient in oncology is computed from the network to be $\Pr(b) = 0.08$. For this probability, we have that $P^-(b) \leq \Pr(b) \leq P^+(b)$. For a patient about whom no further information is available, therefore, the threshold model recommends to gather additional evidence. We now study the robustness of this decision in terms of variation of the parameter $x = p(sh \mid \overline{b})$. The sensitivity function that expresses the prior probability $\Pr(b)$ in terms of x is a constant function; we recall that the function is depicted in Fig. 2(b). We thus find that the parameter can be varied within the entire $[0, 1]$-interval without inducing a change in decision. The decision to gather further evidence therefore is insensitive to variation of this parameter. Now suppose that the patient complains of severe headaches. From the network, the posterior probability of a brain tumour being present is computed to be $\Pr(b \mid sh) = 0.10$. For this probability, we find that $P^-(b) \leq \Pr(b \mid sh) \leq P^+(b)$ which again results in the recommendation to gather further information, for example by performing a CT scan. We again investigate the robustness of this decision in terms

of variation of the parameter x. The sensitivity function that expresses the output probability $\Pr(b \mid sh)$ in terms of x no longer is a constant function; we recall that the function is shown in Fig. 2(d). The function equals

$$f_{\Pr(b|sh)}(x) = \frac{0.064}{0.92 \cdot x + 0.064}$$

From the intersections of the sensitivity function with the two threshold probabilities, we find that $x^+ = 0.055$ and $x^- = 1.476 > 1.0$. We thus find that the decision to gather additional information is robust to variation of the parameter x within the interval $[0.055, 1]$. Since the original value of the parameter equals $x_0 = 0.60$, we conclude that the decision is quite robust to variation of the parameter under study.

We would like to note that by building upon the bounding functions for a sensitivity function a *cautious* interval for variation of a parameter under study can be established. This interval provides boundaries between which the parameter can at least be safely varied. If the parameter is varied beyond these boundaries, however, robustness is no longer guaranteed and the recommended decision may change.

Most likely value For classification problems, decision making with a probabilistic network often amounts to selecting the value that is the most likely for an output variable of interest, based upon the computed probability distribution. For studying output robustness of the network in view of this model for decision making, the effects of parameter variation on the most likely value of the output variable need be studied. For this purpose, the concept of *admissible deviation* has been designed. An admissible deviation captures the extent of the variation that can be applied to a parameter without changing the most likely value of the output variable [13]. It is a pair of real numbers (α, β) that describe the shifts to smaller values and to larger values, respectively, that are allowed in the parameter under study without inducing a change in the most likely value of the output variable; often the symbols $\leftarrow$ and $\rightarrow$ are used to express that the parameter can be varied as far as the boundaries of the probability interval. For a parameter with an original value of x_0, the admissible deviation (α, β) thus indicates that the parameter can be safely varied within the interval $[x_0 - \alpha, x_0 + \beta]$.

We consider a parameter x with an original value of x_0, and an output variable A; we suppose that, given the available evidence e, the most likely value of A is the value a_k. We observe that, if we compare just the two values a_k and a_i of A, then the intersection of the two sensitivity functions $f_{\Pr(a_k|e)}(x)$ and $f_{\Pr(a_i|e)}(x)$ marks the value of the parameter x at which the value a_i becomes more likely than the value a_k. Based upon this observation, the admissible deviation for the parameter x is now readily established from the intersections of the sensitivity function $f_{\Pr(a_k|e)}(x)$ for the most likely value a_k with the functions $f_{\Pr(a_i|e)}(x)$ pertaining to the other values a_i,

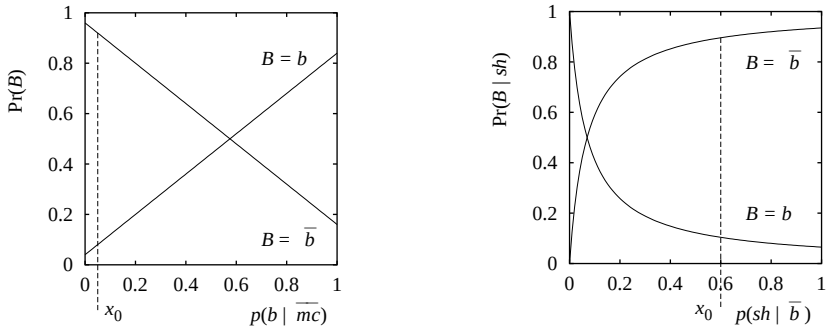


Fig. 8. Sensitivity functions for the possible values of the variable B, in terms of the parameters $p(b \mid \overline{mc})$ (**a**) and $p(sh \mid \overline{b})$ (**b**).

$i \neq k$, of A. If the intersections are found at the values x_i for the parameter x with $x_1 \leq \ldots \leq x_j \leq x_0 \leq x_{j+1} \leq \ldots \leq x_n$, then we can conclude that the parameter can be varied between the values x_j and x_{j+1} without inducing a change in the most likely value of the output variable. The admissible deviation then equals $[x_0 - x_j, x_{j+1} - x_0]$.

To illustrate the concept of admissible deviation, we consider, in the *Brain tumour* network, the output variable B and its two values. For an arbitrary patient, the probability of a brain tumour being present equals $\Pr(b) = 0.08$. The absence of a brain tumour, therefore, is the most likely value of B. We now study the effects of varying the parameter $x = p(b \mid \overline{mc})$, with an original value of $x_0 = 0.05$, on this most likely value. The sensitivity functions that describe the prior probabilities of the two values of B in terms of x are

$$f_{\Pr(b)}(x) = 0.80 \cdot x + 0.04 \quad \text{and} \quad f_{\Pr(\overline{b})}(x) = -0.80 \cdot x + 0.96$$

These functions are depicted in Fig. 8(a). The two sensitivity functions intersect at $x = 0.575$. For the original value of x, we thus find an admissible deviation of $(\leftarrow, 0.525)$. We conclude that the diagnosis of this patient not having a brain tumour is quite robust to variation of the parameter x. We now consider a patient with severe headaches. For this patient, the probability of a brain tumour being present equals $\Pr(b \mid sh) = 0.10$. Again, the absence of such a tumour is the most likely value of the variable B. The effects of varying the parameter $x = p(sh \mid \overline{b})$, with an original value of $x_0 = 0.60$, on the posterior probabilities for B are described by the two sensitivity functions

$$f_{\Pr(b \mid sh)}(x) = \frac{0.064}{0.92 \cdot x + 0.064} \quad \text{and} \quad f_{\Pr(\overline{b} \mid sh)}(x) = \frac{0.92 \cdot x}{0.92 \cdot x + 0.064}$$

These functions are shown in Fig. 8(b). The two sensitivity functions intersect at $x = 0.07$, resulting in an admissible deviation of $(0.53, \rightarrow)$. We conclude

that the most likely value of the output variable again is quite robust to variation of the parameter under study.

We would like to note that by building upon the bounding functions for a sensitivity function a *cautious* interval for variation of a parameter under study can be established. This interval provides boundaries between which the parameter can at least be safely varied. If the parameter is varied beyond these boundaries, however, robustness is no longer guaranteed and the recommended decision may change. From this interval, a *minimal admissible deviation* is established [13].

5 Concluding Observations

Recent research in sensitivity analysis of probabilistic networks has resulted in a variety of new insights and effective methods. The insight that has proved to be most significant, is that the mathematical properties of a probabilistic network strongly constrain the functional form of the relation between a parameter probability and an output probability of interest. In fact, this so-called sensitivity function is a fraction of two linear functions in the parameter under study, taking the form of either a linear function or a fragment of a hyperbola branch. Because of this constrained functional form, any sensitivity function can be established analytically from a probabilistic network by computing just its constants using a limited number of network propagations.

The significance of the sensitivity function lies in that it provides for easy computation of various sensitivity properties pertaining to the robustness of the output of a network. We detailed the concepts of sensitivity value and vertex proximity as giving insight into the effects of parameter variation on an output probability of interest. These concepts can be used during model construction, for example, to gain detailed insight into the level of accuracy that is required for the various parameters and to guide further elicitation efforts. We further discussed the issue of output robustness in view of two different models for decision making that build upon the output distribution computed from a network, and presented methods that give insight into the effects of parameter variation on a recommended decision. These methods reveal the range of parameter values for which a decision based upon the network's output distribution is valid when used in a real-life setting.

Another use of the sensitivity function is for *parameter tuning*. Parameter tuning is typically employed if the output of a probabilistic network differs from what is expected based upon knowledge of the domain and amounts to changing the values of one or more of the network's parameters. Sensitivity functions can then be used to identify which parameters had best be changed to arrive at the expected output. Parameter tuning, however, should be performed with care since changing even a single parameter may easily have unwanted effects over the large range of evidence profiles. We feel that further research into the global effects of local changes is still required.

To conclude, we focused in this paper on *one-way sensitivity analysis*. In essence it is possible to also perform a more general *n-way analysis* of a probabilistic network. Compared to one-way sensitivity analysis, however, higher-order analysis of probabilistic networks has so far received far less attention [4,8]. Since a higher-order analysis is particularly useful for uncovering and studying synergistic effects of variation of the parameter probabilities of a network, we feel that it is worthwhile to direct further research efforts to gaining new insights and developing effective methods for conducting such analyses and interpreting their results.

References

1. E. Castillo, J.M. Gutiérrez and A.S. Hadi (1997). Sensitivity analysis in discrete Bayesian networks. *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 27, pp. 412 – 423.
2. H. Chan and A. Darwiche (2002). When do numbers really matter? *Journal of Artificial Intelligence Research*, vol. 17, pp. 265 – 287.
3. H. Chan and A. Darwiche (2002). A distance measure for bounding probabilistic belief change. *Proceedings of the Eighteenth National Conference on Artificial Intelligence*, AAAI Press, Menlo Park, pp. 539 – 545.
4. H. Chan and A. Darwiche (2004). Sensitivity analysis in Bayesian networks: from single to multiple parameters. In: M. Chickering, J. Halpern (editors), *Proceedings of the Twentieth Conference on Uncertainty in Artificial Intelligence*, AUAI Press, Arlington, VA, pp. 67 – 75.
5. G.F. Cooper (1984). *NESTOR: a Computer-based Medical Diagnostic Aid that Integrates Causal and Probabilistic Knowledge*, Report HPP-84-48, Stanford University.
6. V.M.H. Coupé and L.C. van der Gaag (1998). Practicable sensitivity analysis of Bayesian belief networks. In: M. Hušková, P. Lachout, J.A. Víšek (editors), *Prague Stochastics '98*, Union of Czech Mathematicians and Physicists, Prague, pp. 81 – 86; also available as Report UU-CS-1998-10, Utrecht University.
7. V.M.H. Coupé, L.C. van der Gaag and J.D.F. Habbema (2000). Sensitivity analysis: an aid for probability elicitation. *Knowledge Engineering Review*, vol. 15, pp. 1 – 18.
8. V.M.H. Coupé, F.V. Jensen, U. Kjærulff and L.C. van der Gaag (2000). *A computational architecture for n-way sensitivity analysis of Bayesian networks*. Technical Report, Aalborg University.
9. V.M.H. Coupé and L.C. van der Gaag (2002). Properties of sensitivity analysis of Bayesian belief networks. *Annals of Mathematics and Artificial Intelligence*, vol. 36, pp. 323 – 356.
10. A. Darwiche (2003). A differential approach to inference in Bayesian networks. *Journal of the ACM*, vol. 50, pp. 280 – 305.
11. M.J. Druzdzel and L.C. van der Gaag (1995). Elicitation of probabilities for belief networks: combining qualitative and quantitative information. In: P. Besnard, S. Hanks (editors), *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, Morgan Kaufmann, Palo Alto, pp. 141 – 148.

12. L.C. van der Gaag and V.M.H. Coupé (1999). Sensitivity analysis for threshold decision making with Bayesian belief networks. In: E. Lamma, P. Mello (editors), *AI*IA 99: Advances in Artificial Intelligence*, LNAI, Springer-Verlag, Berlin, pp. 37 – 48.
13. L.C. van der Gaag and S. Renooij (2001). Analysing sensitivity data from probabilistic networks. In: J. Breese, D. Koller (editors), *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*, Morgan Kaufmann, San Francisco, pp. 530 – 537.
14. L.C. van der Gaag and S. Renooij (2006). On the sensitivity of probabilistic networks to reliability characteristics. In: B. Bouchon-Meunier, G. Coletti and R.R. Yager (editors), *Modern Information Processing: From Theory to Applications*, Elsevier B.V., pp. 395-405.
15. F.V. Jensen, S.L. Lauritzen and K.G. Oleson (1990). Bayesian updating in causal probabilistic networks by local computations. *Computational Statistics Quarterly*, vol. 4, pp. 269 – 282.
16. U. Kjærulff and L.C. van der Gaag (2000). Making sensitivity analysis computationally efficient. In: G. Boutilier, M. Goldszmidt (editors), *Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence*, Morgan Kaufmann, San Francisco, pp. 317 – 325.
17. K.B. Laskey (1995). Sensitivity analysis for probability assessments in Bayesian networks. *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 25, pp. 901 – 909.
18. S.L. Lauritzen and D.J. Spiegelhalter (1988). Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society, Series B*, vol. 50, pp. 157 – 224.
19. S.G. Pauker and J.P. Kassirer (1980). The threshold approach to clinical decision making. *New England Journal of Medicine*, vol. 302, pp. 1109 – 1117.
20. J. Pearl (1988). *Probabilistic Reasoning in Intelligent Systems. Networks of Plausible Inference*, Morgan Kaufmann, Palo Alto.
21. S. Renooij and L.C. van der Gaag (2004). Evidence-invariant sensitivity bounds. In: M. Chickering, J. Halpern (editors), *Proceedings of the Twentieth Conference on Uncertainty in Artificial Intelligence*, AUAI Press, Arlington, VA, pp. 479-486.
22. S. Renooij and L.C. van der Gaag (2005). Exploiting evidence-dependent sensitivity bounds. In: F. Bacchus, T. Jaakkola (editors), *Proceedings of the Twenty-First Conference on Uncertainty in Artificial Intelligence*, AUAI Press, Corvallis, OR, pp. 485-492.
23. G. Shafer and P.P. Shenoy (1990) Probability propagation. *Annals of Mathematics in Artificial Intelligence*, vol. 2, pp. 327 – 352.