Sensitivity analysis: an aid for belief-network quantification

V. H. M. Coupe,
L. C. van der Gaag,
J. D. F. Habbema

UU-CS-1999-13

Sensitivity Analysis: an Aid for Belief-network Quantification*

Veerle M.H. Coupé[1], *Linda C. van der Gaag*[2], and *J. Dik F. Habbema*[1]

[1] Center for Clinical Decision Sciences,
Department of Public Health, Erasmus University Rotterdam,
P.O. Box 1738, 3000 DR Rotterdam, The Netherlands
e-mail: `coupe@mgz.fgg.eur.nl`

[2] Department of Computer Science, Utrecht University,
P.O. Box 80.089, 3508 TB Utrecht, The Netherlands
e-mail: `linda@cs.uu.nl`

Abstract

When building a Bayesian belief network, usually a large number of probabilities have to be assessed by experts in the domain of application. Experience shows that experts often are reluctant to assess all probabilities required, feeling that they are unable to give assessments with a high level of accuracy. We argue that the elicitation of probabilities from experts can be supported to a large extent by iteratively performing *sensitivity analyses* of the belief network in the making, starting with rough, initial assessments. Since it gives insight into which probabilities require a high level of accuracy and which do not, performing a sensitivity analysis allows for focusing further elicitation efforts. We propose an elicitation procedure in which, alternatingly, sensitivity analyses are performed and probability assessments are refined, until satisfactory behaviour of the belief network is obtained, until the costs of further elicitation outweigh the benefits of higher accuracy, or until higher accuracy can no longer be attained due to lack of knowledge.

*The investigations were (partly) supported by the Netherlands Computer Science Research Foundation with financial support from the Netherlands Organization for Scientific Research (NWO).

1 Introduction

Bayesian belief networks are widely accepted in artificial-intelligence research as intuitively appealing, valuable representations of knowledge, tailored to domains in which uncertainty is predominant [Pearl, 1988]. A Bayesian belief network basically is a concise representation of a joint probability distribution, consisting of a qualitative part and an associated quantitative part. The qualitative part of the network encodes the variables of importance in the domain that is being represented, along with their influential interrelationships. The strengths of the relationships between the variables are quantified by conditional probabilities. These probabilities constitute the quantitative part of the network. An increasing number of knowledge-based systems nowadays builds on the framework of Bayesian belief networks for knowledge representation and inference, thereby demonstrating the usefulness of belief networks for addressing real-life problem domains; applications range from medical diagnosis, prognosis, and treatment planning, to probabilistic information retrieval [Andreassen *et al.*, 1987, Heckerman *et al.*, 1992, Bruza & Van der Gaag, 1994].

Bayesian belief networks are usually constructed with the help of *domain experts*. Experience shows that, although it may require considerable effort, building the qualitative part of a belief network is quite practicable. In fact, building the qualitative part has parallels to designing a domain model for a more traditional knowledge-based system. To a large extent, therefore, well-known knowledge-engineering techniques can be employed for this purpose [McGraw & Harbison-Briggs, 1989, Studer *et al.*, 1998]. Unfortunately, similar observations do not hold with regard to the quantitative part of a Bayesian belief network. Constructing the quantitative part is generally considered a far harder task, not in the least because it tends to consume much more time. It amounts to assessing various conditional probabilities for the variables represented in the belief network's qualitative part. Although for most domains of application abundant probabilistic information is available from literature or from statistical data, it often turns out that this information is not directly amenable to encoding in a belief network [Druzdzel & Van der Gaag, 1995, Jensen, 1995]. Medical literature, for example, often reports conditional probabilities of the presence of symptoms given a disease, but not always the probabilities of these symptoms occurring in the absence of disease; moreover, conditional probabilities for unobservable intermediate disease states are usually lacking in the literature. The majority of the probabilities required will therefore have to be assessed by domain experts. The problems encountered when eliciting probabilities from experts are widely known [Tversky *et al.*, 1982]; an expert's assessments may for example reflect various biases and may not be properly calibrated. Acknowledging these problems, in the field of decision analysis various different techniques have been developed for the elicitation of well-calibrated probabilities from experts [Von Winterfeldt & Edwards, 1986]. These techniques, however, tend to be quite time-consuming. As for a belief network generally a large number of probabilities is required, employing these techniques may not be feasible. For probability elicitation for Bayesian belief networks, therefore, supplementary techniques are being sought [Druzdzel & Van der Gaag, 1995, Van der Gaag *et al.*, 1999].

Experience with eliciting probabilities from domain experts for a Bayesian belief network shows that experts often are reluctant to assess all conditional probabilities required [Van der Gaag *et al.*, 1999]. At least part of their uneasiness stems from their feeling that they are compelled to give exact numbers having a high level of accuracy while they know they are unable to. In general, however, not every probability assessment will require the same level of accuracy to arrive at satisfactory behaviour of the knowledge-based system that is being developed; some probabilities have more impact on the system's behaviour than others. For gaining detailed insight into the level of accuracy that is required for the various conditional probabilities of a Bayesian belief network, a *sensitivity analysis* can be performed [Morgan & Henrion, 1990].

The basic idea of performing a sensitivity analysis of a Bayesian belief network is to systematically vary initial assessments for the network's conditional probabilities over a plausible interval and study the effects on the behaviour of the system being developed. Some probabilities are likely to show a considerable effect, while others will hardly reveal any influence. For the less influential probabilities, the initial assessments may suffice. For the more influential probabilities, however, refinement of the initial assessments may be worthwhile. For these probabilities, for example, elaborate elicitation techniques may be applied to obtain a more accurate assessment. Given the limited and costly time of experts, attention can thus be focused on the probabilities to which the system's behaviour shows the highest sensitivity. As assessments are refined, the belief network under construction may again show various different sensitivities. To gain insight in these, possibly new, sensitivities, the network can once again be subjected to a sensitivity analysis. Based upon these observations, we propose a procedure for eliciting probabilities that builds upon the idea of stepwise refining probability assessments. The procedure sets out with the elicitation of initial, probably highly uncertain, assessments for all conditional probabilities required for the belief network under construction. Starting with these initial assessments, a sensitivity analysis of the network is performed, upon which the assessments for the most influential probabilities are refined. Iteratively performing sensitivity analyses and refining probabilities is pursued until satisfactory behaviour of the belief network is obtained, until the costs of further elicitation outweigh the benefits of higher accuracy, or until higher accuracy can no longer be attained due to lack of knowledge.

We would like to note that, when performed straightforwardly, a sensitivity analysis of a Bayesian belief network may require considerable computational effort. The computational burden involved, however, can be reduced to a large extent by exploiting the probabilistic relationships among the variables that are represented in the belief network under study. These relationships allow for distinguishing between conditional probabilities that may influence the system's behaviour and those that cannot. In addition, these relationships induce simple mathematical functions describing the network's behaviour in terms of its conditional probabilities [Coupé & Van der Gaag, 1998]. Nevertheless, iteratively performing sensitivity analyses of a Bayesian belief network with respect to all potentially influential probabilities will remain computationally expensive. Considering that these analyses are performed only when constructing and validating a belief network, their computational burden is well outweighed by their benefits in probability elicitation.

In this paper, we outline sensitivity analysis of Bayesian belief networks and discuss its use in probability elicitation. In Section 2, we introduce the formalism of Bayesian belief networks. We address the construction of a belief network in Section 3. In Section 4, we illustrate the technique of sensitivity analysis. In Section 5, we detail our elicitation procedure exploiting sensitivity analysis as an aid for belief-network quantification. In Section 6, we address the computational issues involved in performing a sensitivity analysis of a Bayesian belief network. The paper ends with some conclusions and directions for further research in Section 7.

2 Bayesian belief networks

Informally speaking, a *Bayesian belief network* is a representation of domain knowledge. It consists of a qualitative part and an associated quantitative part. The network's qualitative part takes the form of an acyclic directed graph, or digraph, for short. Each node in this digraph represents a domain variable that takes its value from a finite set of discrete values. In this paper we will restrict the discussion to binary variables, taking one of the values *true* and *false*; our results, however, are generalised straightforwardly to account for variables taking their value from larger discrete sets. If a variable V has the value *true*, we will write v; the notation $\neg v$ is used to indicate that $V = \mathit{false}$. The arcs in the network's digraph represent influential relationships among the represented variables. The tail of an arc indicates the cause of the effect at the head of the arc. Absence of an arc between two variables means that these variables do not influence each other directly.

For our running example we consider the following fragment of (fictitious and incomplete) medical information, adapted from [Cooper, 1984]; the example is meant for illustrative purposes only.

> "Consider a primary tumour with an uncertain prognosis in an arbitrary patient. It is known that the cancer can spread to the brain and to other sites. We are interested in the course of the cancer within the next three years, especially with regard to the development of a brain tumour and its associated problems. The probability of a metastatic cancer developing from the primary tumour is estimated to be 0.2. The probability that the metastatic cancer will be in the brain is also estimated at 0.2. Also in the absence of metastatic cancer, there is a small probability of development of a (primary) brain tumour; this probability is assessed to be 0.05.
>
> Metastatic cancer may be detected by an increased level of serum calcium. In fact, serum calcium will be increased with probability 0.8 in the presence of metastatic cancer and only with probability 0.2 in the absence of metastatic cancer. It is estimated that a patient will fall into a coma within the next three years with probability 0.8 in the case that a brain tumour is present and/or the level of serum calcium is increased. Otherwise, there is only a probability of 0.05 of falling into a coma. Severe headaches are likely to

develop both with a brain tumour, with probability 0.8, and without a brain tumour, with probability 0.6."

In this fragment of information, five variables are identified: the presence or absence of metastatic cancer in a patient (indicated by MC), the presence or absence of a brain tumour (B), an increased level of serum calcium (ISC), a patient falling into a coma within the next three years (C), and the presence or absence of severe headaches (SH). The relationships among these variables are encoded in the digraph depicted in Figure 1. The graph for example reflects, by means of the arc $B \rightarrow SH$, that the presence of a brain tumour is a possible cause of severe headaches.

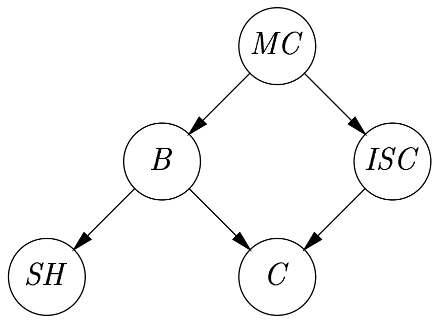

Figure 1: The digraph of an example Bayesian belief network, encoding information concerning the presence of a brain tumour and its associated problems in an arbitrary patient (the meanings of the variables are given in the text).

The absence of the arc $MC \rightarrow C$ in the digraph shown in Figure 1 indicates that the presence or absence of metastatic cancer does not directly influence whether or not a patient will fall into a coma within the next three years. Metastatic cancer influences coma only indirectly, through brain tumour and increased total serum calcium. Once the presence or absence of a brain tumour and the level of serum calcium have been established in a patient, the presence or absence of metastatic cancer no longer influences the probability of this patient falling into a coma. Metastatic cancer and coma therefore are conditionally independent given brain tumour and increased total serum calcium. For further details on the independences that are read from a belief network's qualitative part, we refer the reader to [Pearl, 1988].

The influential relationships that are represented in the qualitative part of a Bayesian belief network generally are probabilistic in nature. To describe the 'strengths' of these relationships, several conditional probabilities are provided. For each node in the network's digraph, the probabilities of its values are specified, conditional on the various possible combinations of values for its immediate predecessors in the graph. For our running example, we have the following probabilities:

$p(mc) =$	0.20	$p(c \mid b, isc) =$	0.80
		$p(c \mid \neg b, isc) =$	0.80
$p(b \mid mc) =$	0.20	$p(c \mid b, \neg isc) =$	0.80
$p(b \mid \neg mc) =$	0.05	$p(c \mid \neg b, \neg isc) =$	0.05
$p(isc \mid mc) =$	0.80	$p(sh \mid b) =$	0.80
$p(isc \mid \neg mc) =$	0.20	$p(sh \mid \neg b) =$	0.60

From the conditional probabilities specified for node ISC, it is readily seen that knowing whether or not metastatic cancer is present has a considerable influence on the probability of an increased level of serum calcium in an arbitrary patient: the relationship between metastatic cancer and increased total serum calcium is a fairly strong dependence. On the other hand, severe headaches are expressed as being quite common in both patients with and without a brain tumour; severe headaches therefore have a low predictive value for the presence or absence of a brain tumour. From the conditional probabilities specified for node C, we see that in the absence of both a brain tumour and an increased level of serum calcium, there is only a very small probability of a patient falling into a coma. The presence of either one of these causes in an arbitrary patient, however, suffices to render the probability of this patient falling into a coma in the near future quite high. Note that the two causes do not contribute to this probability independently: if one of the causes is present, then the presence of the other cause has no further influence on the probability of a patient falling into a coma. The two causes are said to exhibit a (negative) *synergistic influence* on their common effect.

The probabilities associated with the qualitative part of a Bayesian belief network constitute the network's quantitative part. An important property of the belief-network formalism is that a network's quantitative part defines a unique joint probability distribution that respects the independences that are portrayed by the network's qualitative part [Pearl, 1988]. In our examples, we will write Pr to denote the probability distribution that is defined by a belief network. To explicitly distinguish between the probabilities that are derived from the distribution and the conditional probabilities that are specified in the network's quantitative part, we will write p to denote the latter probabilities.

So far we have treated the probability assessments of a Bayesian belief network as exact point probabilities. For most applications, however, the initially obtained assessments will be quite uncertain. To capture this uncertainty, each probability assessment is supplemented with a *plausible interval* that defines a range of values in which the 'true' probability lies with reasonable certainty. For our example belief network, for instance, we assume the plausible interval (0.75 – 0.85) for the assessment $p(isc \mid mc) = 0.80$ for the probability of an increased level of serum calcium in the presence of metastatic cancer, indicating that $0.75 \leq p(isc \mid mc) \leq 0.85$ with reasonable certainty. In the sequel, we will use the following plausible intervals for the various assessments of our example network:

$p(mc)$:	$(0.1 - 0.35)$	$p(c \mid b, isc)$:	$(0.6 - 0.9)$
		$p(c \mid \neg b, isc)$:	$(0.75 - 0.85)$
$p(b \mid mc)$:	$(0.05 - 0.5)$	$p(c \mid b, \neg isc)$:	$(0.65 - 0.85)$
$p(b \mid \neg mc)$:	$(0 - 0.25)$	$p(c \mid \neg b, \neg isc)$:	$(0 - 0.1)$
$p(isc \mid mc)$:	$(0.75 - 0.85)$	$p(sh \mid b)$:	$(0.65 - 0.9)$
$p(isc \mid \neg mc)$:	$(0.15 - 0.3)$	$p(sh \mid \neg b)$:	$(0.45 - 0.7)$

The plausible intervals for the assessments that are initially obtained for a belief network's conditional probabilities can be quite large. Upon refining the various assessments, as proposed in the sequel, the plausible intervals involved typically become smaller.

To conclude, associated with the belief-network formalism are procedures for computing probabilities of interest from a belief network and for processing evidence, that is, for entering evidence into the network and subsequently computing the revised probability distribution given the evidence [Pearl, 1988, Lauritzen & Spiegelhalter, 1988]. These procedures are the basic building blocks for *reasoning* with a Bayesian belief network. The details involved are not relevant for the present paper and therefore are not reviewed here. We would like to note, however, that the belief-network formalism accommodates for various types of reasoning, among which are diagnostic and prognostic reasoning.

3 Building a Bayesian belief network

Building a Bayesian belief network for a domain of application involves various tasks. First, the variables that are of importance in the domain are identified, along with their values. Since a belief network, as any model, necessarily is a simplification of reality, well-founded decisions have to be taken as to which variables and values should be included in the network and which may be omitted. The important variables and values are typically identified with the help of experts in the domain under study. We would like to emphasise that this task is not reserved for building Bayesian belief networks, but instead is quite common in engineering knowledge-based systems in general, for which purpose several methodologies and techniques are available [McGraw & Harbison-Briggs, 1989, Studer *et al.*, 1998]. Once the variables of importance have been identified, the construction of the qualitative part of the belief network commences. For acquiring the topology of the network's digraph, domain experts are interviewed. In the interviews, the concept of causality is generally used as a heuristic guiding principle. Elicited causalities are expressed in graphical terms by taking the direction of causation for directing arcs. Building on the concept of causality has the advantage that domain experts may express their knowledge in either causal or diagnostic direction. Since they are allowed to express their knowledge in a form they feel comfortable with, the experts' statements and, hence, the qualitative part of a belief network, tend to be quite robust [Druzdzel & Van der Gaag, 1995]. The task of eliciting relationships among variables from domain experts once again has parallels with engineering knowledge-based systems in general.

After the qualitative part of a Bayesian belief network has been constructed, its quantitative part is specified. Quantifying a belief network's qualitative part amounts to assessing various conditional probabilities for the represented variables. The assessment of all probabilities required tends to be, if not the hardest, then certainly the most time-consuming task in building a belief network. Although for most domains of application abundant probabilistic information is available from literature, it rarely turns out to be directly amenable to encoding in a belief network [Druzdzel & Van der Gaag, 1995]. Medical literature, for example, often reports conditional probabilities of the presence of symptoms given a disease, but not always the probabilities of these symptoms occurring in the absence of disease; moreover, conditional probabilities for unobservable intermediate disease states are usually lacking. If literature does not provide sufficient and reliable probability assessments, estimates may be obtained from statistical data. Experience shows, however, that even if comprehensive data collections are available, they very seldom contribute to the entire quantification task [Jensen, 1995, Korver & Lucas, 1993]. In a medical data collection, for example, unobservable intermediate pathophysiological states are typically not recorded. As a consequence, a large number of probabilities remain to be assessed by domain experts.

The field of *decision analysis* offers various techniques for the elicitation of judgemental probabilities from experts [Von Winterfeldt & Edwards, 1986, Morgan & Henrion, 1990]. We briefly review the two techniques that are most often used for eliciting probabilities for Bayesian belief networks. The simplest technique is the use of a numerical *probability scale*. A probability scale is a horizontal or vertical line with the endpoints denoting a 0% and a 100% chance, respectively, and a few numerical anchors in between, for example to denote a 50% chance; Figure 2 illustrates the basic idea. For each probability required, a

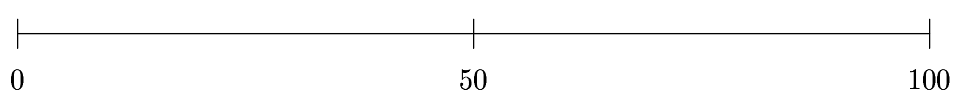

Figure 2: A numerical scale for probability elicitation.

domain expert is asked to indicate his or her assessment on a separate scale. In communicating a probability to be assessed to a domain expert, the probability is often transcribed verbally in terms of frequencies. The expert is asked to envisage one hundred cases with a specific context and assess the number of cases that exhibit a certain characteristic. For our example belief network, for instance, the domain expert may be asked to envisage a population of one hundred arbitrary patients with metastatic cancer and to assess the number of patients from among this population who will show an increased level of serum calcium upon examination. Experience shows that the use of a probability scale along with the frequency method provides experts little to go by and may result in highly inaccurate probability assessments [Van der Gaag *et al.*, 1999].

A more elaborate technique for the elicitation of judgemental probabilities is the use of *reference lotteries*. A domain expert is presented with a choice between two lotteries, one of which pertains to a probability to be assessed and the other one serves as a reference.

The reference lottery yields a desired reward with probability p and a less desired outcome with probability $1 - p$. The second lottery yields the same desired reward if a specific case exhibits a certain characteristic and the less desired outcome otherwise. For our example belief network, the domain expert is presented, for instance, with a choice between a reference lottery and the lottery that yields $10,000 if a specific patient with metastatic cancer upon examination shows an increased level of serum calcium and $1 if the level of serum calcium is not increased in this patient; Figure 3 depicts this choice between lotteries. The domain expert is asked to adjust the value of p in the reference lottery until

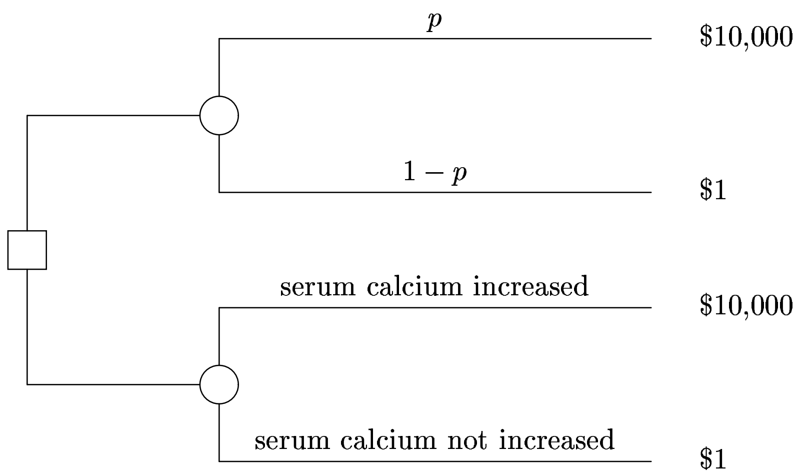

Figure 3: An example reference lottery, tailored to assessment of the conditional probability of the presence of an increased level of serum calcium in a patient with metastatic cancer.

he or she is indifferent between the two lotteries. The resulting value of p then is taken to be the conditional probability that had to be assessed. Experience shows that the use of reference lotteries for eliciting probabilities from domain experts may avert to at least some extent the problems of bias and poor calibration that are typically found in human probability assessment. The use of lotteries, however, tends to be quite time-consuming and, in fact, often turns out to be infeasible for probability elicitation for belief networks as a result of the large size and complexity of a network in the making.

While the use of lotteries for probability elicitation on the one hand tends to be infeasible for quantifying a Bayesian belief network, the use of a probability scale on the other hand tends to yield assessments that are too much inaccurate. The use of a probability scale, nonetheless, serves to yield an assessment, be it an inaccurate one, for every conditional probability required. In the sequel, we will argue that these assessments may be used as a starting point for further refinement. We will propose an elicitation procedure in which, starting with uncertain assessments, sensitivity analyses of a belief network in the making are performed and assessments are refined alternatingly. Our elicitation procedure builds upon plausible intervals capturing the uncertainties in the various assessments as described in the previous section. Various techniques are available for acquiring plausible intervals with probability assessments [Morgan & Henrion, 1990]. The simplest technique is to elicit the plausible intervals directly with the probability assessments from domain experts. An expert is asked to envisage one hundred cases with a specific context and provide a lower bound and an upper bound on the number of cases that exhibit a certain characteristic; these bounds are selected so that the expert is relatively certain that the true number of cases exhibiting the characteristic lies between the bounds. Plausible intervals thus obtained may not be very robust. It is not clear, for example, whether to interpret these intervals as 90%, 95%, or 100% confidence intervals. To obtain more robust plausible intervals, more elaborate techniques are available. These techniques once again tend to be quite time-consuming and therefore are not reviewed here.

4 Sensitivity analysis

Sensitivity analysis is a general technique for studying the effects of the uncertainties in the parameters of a mathematical model on this model's outcome [Morgan & Henrion, 1990, Habbema *et al.*, 1990]. For a Bayesian belief network, sensitivity analysis provides for example for studying the effects of the uncertainties in the assessments for the network's conditional probabilities on a probability of interest. There are various different types of sensitivity analyses. For a belief network, the simplest type of sensitivity analysis amounts to systematically varying the assessment for one of the network's probabilities while keeping all other assessments fixed. Such an analysis serves to reveal the effect of just the conditional probability whose assessment is being varied, on a probability of interest. A sensitivity analysis in which a single assessment is varied, is termed a *one-way sensitivity analysis*. In a *two-way sensitivity analysis* of a Bayesian belief network, two probability assessments are varied simultaneously. In addition to the separate effects of variation of the two assessments, 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 probability assessments at the same time. The results of such an analysis, however, are often hard to interpret. In this paper, we will therefore restrict the discussion to one-way and two-way sensitivity analyses.

We illustrate performing a *one-way sensitivity analysis* of our example belief network. We begin by taking the probability of falling into a coma, $\Pr(c)$, for our probability of interest. By doing so, we assess the robustness of the *prognosis* of falling into a coma for an arbitrary patient with a primary tumour. Such an analysis may be useful in predicting, for example, the expected number of patients that will fall into a coma within the next three years in a population of patients with primary tumours. We address the one-way analyses with respect to the assessments for the conditional probabilities $p(b \mid mc)$, $p(isc \mid mc)$, and $p(isc \mid \neg mc)$, respectively. The results of these three analyses are shown in Figure 4. Figure 4(a) shows that systematically varying, from 0 to 1, the assessment for the probability $p(b \mid mc)$ of the presence of a brain tumour given, with certainty, that a patient has a metastatic cancer, has a negligible effect on the probability of interest $\Pr(c)$: the prior probability of a patient falling into a coma within the next three years increases from 0.31 to 0.34, approximately. Figure 4(b) shows that varying the initial assessment

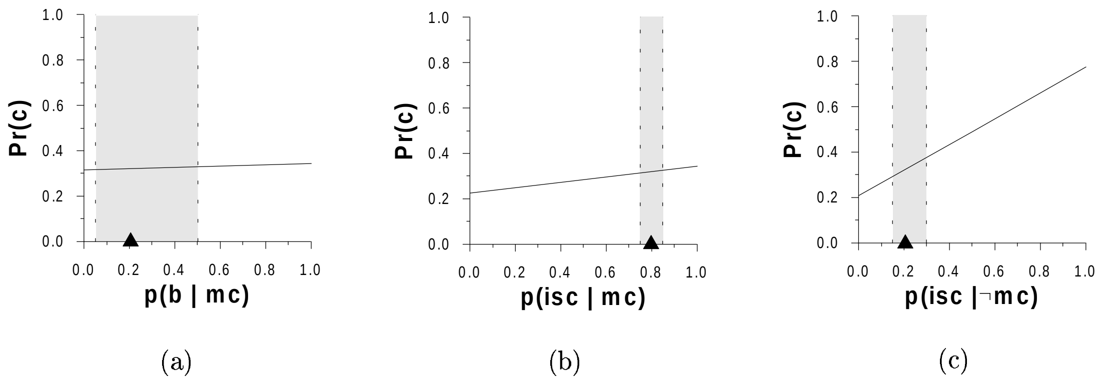

Figure 4: A one-way sensitivity analysis of the example belief network from Figure 1. The effects of varying the assessments for the probabilities $p(b \mid mc)$, $p(isc \mid mc)$, and $p(isc \mid \neg mc)$, respectively, on the prior probability $\Pr(c)$ are shown; the plausible intervals for the assessments are indicated by shading.

for the probability $p(isc \mid mc)$ of an increased level of total serum calcium conditional on the presence of a metastatic cancer has a somewhat stronger effect on the probability of interest: $\Pr(c)$ now ranges from 0.22 to 0.34. From Figure 4(c), to conclude, it is seen that varying the assessment for the probability $p(isc \mid \neg mc)$ of an increased level of total serum calcium in the absence of a metastatic cancer has an even stronger effect on $\Pr(c)$: the prior probability of a coma ranges from 0.21 to 0.78. Note that the three analyses reveal a linear relationship between the probability assessment that is being varied and the probability of interest.

The three example analyses and their results have so far been discussed without taking into consideration the plausible interval of the probability assessment that is being varied. To decide whether or not further elicitation efforts are worthwhile for the three assessments, their plausible range of values needs to be considered, however. By including the plausible interval for a probability assessment into a one-way sensitivity analysis, a *plausible effect* of this assessment on the probability of interest is yielded. The sensitivity analysis may now show, for example, that a probability assessment that is not very influential yet is rather uncertain can have a stronger effect on the probability of interest than a probability assessment that is very accurate and quite influential. For the three probability assessments under study for our example belief network, the plausible intervals are indicated in Figure 4 by shading. The figure shows that plausible variation of the assessment for the conditional probability $p(isc \mid \neg mc)$ has the strongest effect on the probability of interest $\Pr(c)$. It may therefore be worthwhile to try and obtain a more accurate assessment for this conditional probability. Varying the assessments for the probabilities $p(b \mid mc)$ and $p(isc \mid mc)$, respectively, within their plausible intervals results in a rather small effect on the probability of interest. We recall from Figure 4 that the effect on $\Pr(c)$ of varying the assessment for $p(b \mid mc)$ from 0 to 1 is smaller than the effect of varying $p(isc \mid mc)$

from 0 to 1. By taking the plausible intervals into consideration, however, variation of the assessment for $p(b \mid mc)$ has the stronger plausible effect. Especially since the plausible interval for this assessment is quite large, further elicitation efforts may better be directed at the probability $p(b \mid mc)$ than at the probability $p(isc \mid mc)$. We will return to this observation in Section 5.

We now address a *two-way sensitivity analysis* of a Bayesian belief network. In a two-way sensitivity analysis, two probability assessments are varied simultaneously to reveal their joint effect on a probability of interest. We illustrate performing a two-way sensitivity analysis for our example belief network. For our probability of interest, we once again take the prior probability $\Pr(c)$ of an arbitrary patient with a primary tumour falling into a coma within the next three years. We address the analysis with respect to the assessments for the conditional probabilities $p(b \mid mc)$ and $p(isc \mid mc)$. The result of the analysis, varying the assessments under study from 0 to 1 simultaneously, is shown in Figure 5. The contour

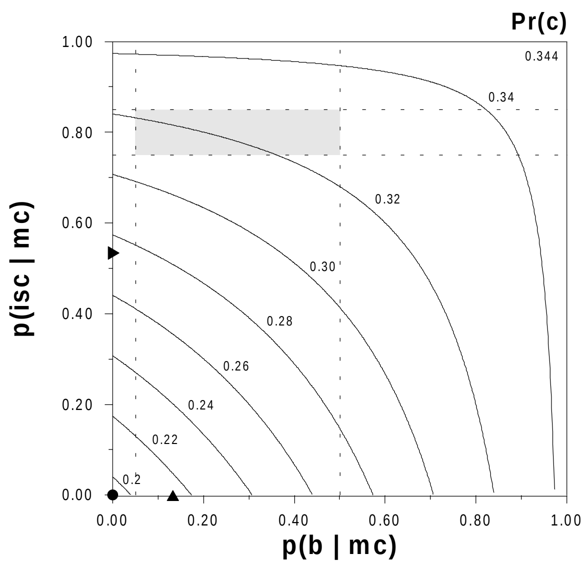

Figure 5: A two-way sensitivity analysis of the example belief network. The joint effect of varying the assessments for the conditional probabilities $p(b \mid mc)$ and $p(isc \mid mc)$ simultaneously on the prior probability $\Pr(c)$ is shown by contour lines.

lines in the figure connect the combinations of values for the two probability assessments that result in the same value for the probability of interest $\Pr(c)$. The distance between two contour lines indicates the variation necessary in the two assessments to shift the probability of interest from one contour line to another. If the contour lines are very close to one another, then a small variation in the probability assessments under study suffices to have a strong effect on the probability of interest; if, in contrast, the contour lines are further apart, then the probability of interest is not very sensitive to variation of the two assessments. We observe from Figure 5 that the distances between the contour lines differ, indicating that varying the assessments for $p(b \mid mc)$ and $p(isc \mid mc)$ simultaneously has a joint effect on the probability of interest beyond the effects of their separate variation; this joint effect is due to the synergistic influence of the variables B and ISC on the variable C outlined before in Section 2. We further observe that the contour lines are closer to one another in the lower left part of the figure than in the upper right part. If the assessments for the conditional probabilities $p(b \mid mc)$ and $p(isc \mid mc)$ are both quite small, therefore, their variation will have a stronger effect on the probability of interest than if the initial assessments have a higher value. To variation within the plausible intervals of the assessments $p(b \mid mc) = 0.2$ and $p(isc \mid mc) = 0.8$, as indicated by shading in Figure 5, the probability of interest shows a relatively low sensitivity. We further observe that the absolute effect of their joint variation on $\Pr(c)$ is not too strong.

In addition to the analysis discussed before, we address another two-way analysis of our example belief network, this time pertaining to the assessments for the conditional probabilities $p(b \mid mc)$ and $p(isc \mid \neg mc)$. The result of this analysis is shown in Figure 6. We observe from the figure that the contour lines, once again indicating values for the

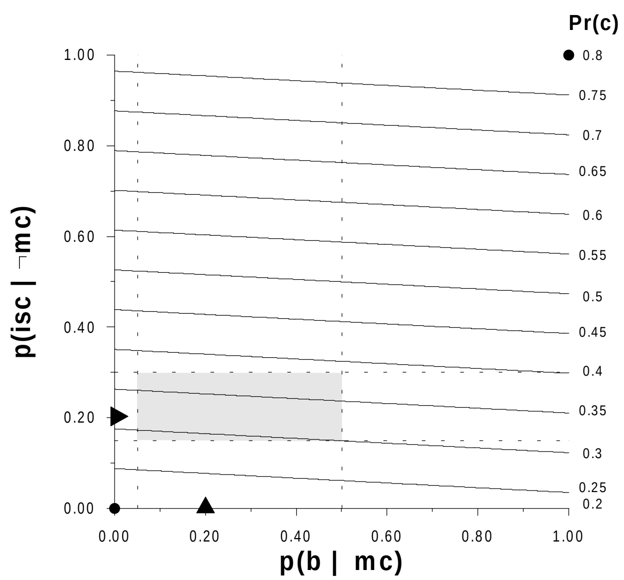

Figure 6: A two-way sensitivity analysis of the example belief network. The joint effect of varying the assessments for the conditional probabilities $p(b \mid mc)$ and $p(isc \mid \neg mc)$ simultaneously on the prior probability $\Pr(c)$ is shown by contour lines.

probability of interest $\Pr(c)$, are equidistant. Equidistance of contour lines indicates that simultaneously varying the probability assessments under study has no joint effect on the probability of interest beyond the effects of their separate variation. The two-way analysis therefore does not provide any information in addition to the information yielded by one-way analyses for the separate assessments.

In the sensitivity analyses for our example belief network described so far, we have taken for the probability of interest a *prior* probability. A sensitivity analysis with respect to a prior probability of interest allows for assessing the quality and robustness of a

Bayesian belief network in its reflecting a prior probability distribution for the domain of application. In the presence of case-specific observations, however, a belief network may show very different sensitivities. To reveal these sensitivities, a sensitivity analysis may be performed with respect to a *posterior*, or conditional, probability. Such an analysis allows for validating the network's behaviour for specific cases or profiles, for example, profiles of typical patient populations in a medical application.

For our example belief network, we have once again performed a one-way sensitivity analysis, this time taking for the probability of interest the *posterior* probability $\Pr(b \mid sh)$ of the presence of a brain tumour in a patient who is *known* to suffer from severe headaches. By doing so, we assess the robustness of the *diagnosis* of a brain tumour for an arbitrary patient with a primary tumour who is suffering from severe headaches. We address the one-way analyses with respect to the assessments for the probabilities $p(mc)$, $p(b \mid \neg mc)$, and $p(sh \mid \neg b)$, respectively. The results of these analyses are shown in Figure 7. Note that, in contrast with the one-way analyses discussed before, the analyses for the posterior

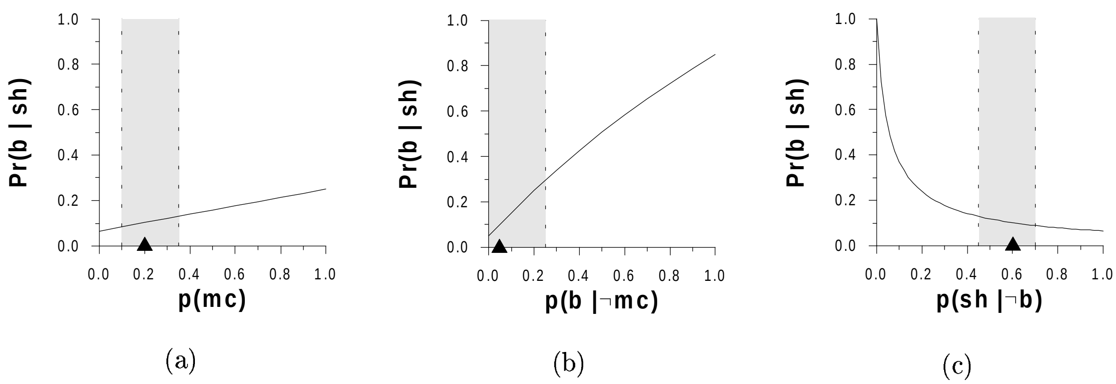

Figure 7: A one-way sensitivity analysis of the example belief network. The effects of varying the assessments for the probabilities $p(mc)$, $p(b \mid \neg mc)$, and $p(sh \mid \neg b)$, respectively, on the posterior probability $\Pr(b \mid sh)$ are shown.

probability of interest reveal a non-linear relationship between the probability assessment that is being varied and the probability of interest.

To conclude, we have performed a two-way sensitivity analysis of our example belief network with respect to the posterior probability of interest $\Pr(b \mid sh)$. We address the analysis of varying the assessments for the conditional probabilities $p(b \mid \neg mc)$ and $p(sh \mid b)$ simultaneously. The result of this analysis is shown in Figure 8. Note that the contour lines are closest to one another in the lower right part of the figure, indicating a high sensitivity of the posterior probability of interest to high values for the probability $p(b \mid \neg mc)$ and low values for the probability $p(sh \mid b)$. For variation, within the plausible intervals of the initial assessments $p(b \mid \neg mc) = 0.05$ and $p(sh \mid b) = 0.8$, as indicated by shading in Figure 8, however, the probability of interest is relatively stable.

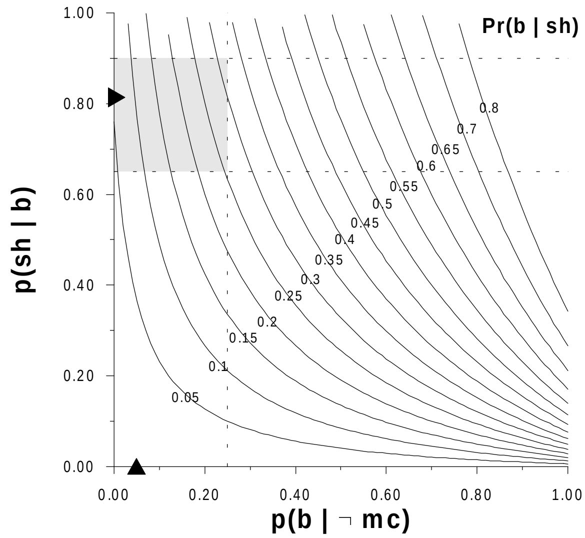

Figure 8: A two-say sensitivity analysis of the example belief network. The effect of varying the assessments for the conditional probabilities $p(b \mid \neg mc)$ and $p(sh \mid b)$ simultaneously, on the posterior probability $\Pr(b \mid sh)$ is shown by contour lines.

5 Sensitivity analysis and probability refinement

Sensitivity analysis of a Bayesian belief network provides for studying the effects of the uncertainties in the various probability assessments of the network on a probability of interest, as demonstrated in the previous section. Studying these effects can to a large extent serve to support the elicitation of probabilities, as it gives detailed insight into the level of accuracy that is required for the various probabilities of the network and as a result provides for focusing further elicitation efforts. We envisage an *elicitation procedure* in which, alternatingly, sensitivity analyses are performed of a belief network in the making and probability assessments are refined; our procedure is summarised in Figure 9. We elaborate on the various different steps of the procedure separately.

In the first step of the elicitation procedure, *initial assessments* are acquired for all conditional probabilities for a belief network in the making. As we have argued before, for most domains of application, experts will have to provide the majority of these initial assessments. Applying an elaborate elicitation technique, such as the use of reference lotteries, for this purpose may in this phase of the construction of the belief network be too much time-consuming to be practicable; also, it would take much more effort than probably necessary, as for several probabilities less accurately obtained assessments will suffice. The first step of the elicitation procedure therefore aims at acquiring assessments *within a short period of time*. To this end, in a limited number of interview sessions, experts are asked to assess all probabilities required off the top of their heads, for example using a numerical probability scale. In addition, they are asked to indicate plausible intervals along with their assessments. These intervals should be pessimistic rather than optimistic to ensure that

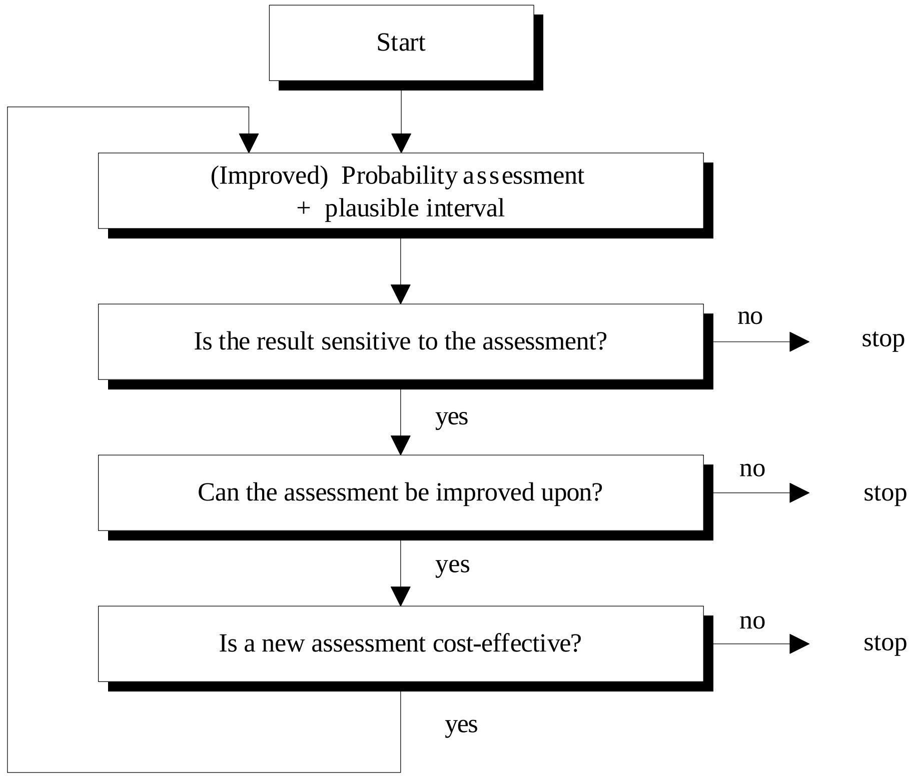

Figure 9: A procedure for probability elicitation for Bayesian belief networks.

the uncertainties in the various different assessments are not underestimated. We would like to note that, since they are allowed to express the uncertainties in their assessments, experts will tend to be less reluctant to provide numerical statements than when they feel compelled to give exact numbers. Instead of eliciting all probabilities required from domain experts, initial assessments, for at least some of these probabilities, may be obtained from data, if available. If the data at hand is known to be imperfect, incomplete, or biased, then the assessments derived should be supplemented with fairly large plausible intervals to capture the uncertainties involved.

The probability assessments obtained in the first step of the elicitation procedure will in general be highly uncertain, that is, these assessments will and should have considerably large plausible intervals. For some of the probabilities, these initial assessments will nonetheless suffice. Other probabilities, however, will require assessments with a higher level of accuracy. The second step of our elicitation procedure is aimed at uncovering the latter probabilities. For this purpose, the belief network in the making is subjected to various *sensitivity analyses*. In these analyses, every single probability assessment for the network is varied, as is every pair of assessments. The effects of varying these assessments within their plausible intervals are studied with respect to one or more prior probabilities of interest as well as several posterior probabilities of interest. From these analyses, the probability assessments that induce the largest plausible effects are identified. We would like to note that, if performing all these sensitivity analyses tends to be too much time-consuming to be practicable, the analyses can be restricted to a moderate number of assessments. To this end, experts may be asked to indicate the probability assessments that are the most likely to affect a probability of interest or, alternatively, to indicate a number of conceptually related probability assessments.

In the second step of the elicitation procedure, various probability assessments have been identified that induce a considerably large plausible effect on a probability of interest. These assessments are potential candidates for further refinement. In the third step of the procedure, the extent to which these assessments can actually be refined is determined. The extent to which a probability assessment lends itself to further refinement depends on the current uncertainty in the assessment, indicated by the width of its plausible interval, and on the elicitation techniques used to arrive at the assessment. A probability assessment with a rather small plausible interval obtained from applying elaborate elicitation techniques may not lend itself to further refinement. An assessment that is not yet very certain, on the other hand, may be more easily improved upon. Refinement of such an uncertain assessment, however, is only actually possible if reliable sources of probabilistic information remain to be explored; examples of such sources of information include further literature search, the use of a panel of experts, the use of more elaborate elicitation techniques, and the study of data, for instance collected in a prospective study. From among the potential candidates for further refinement, therefore, the assessments are identified that have a considerable plausible interval and for which yet unexplored sources of probabilistic information are available.

The probability assessments that have been selected in the third step of the elicitation procedure are assessments that *can* be refined. To actually refine these assessments, an investment of time and money is required. Not for every assessment may this investment be worthwhile, however. In the fourth step of the elicitation procedure, therefore, it is investigated for each of the selected assessments whether refinement is *cost-effective*. The cost-effectiveness of refining a probability assessment is determined by weighing the costs in terms of money and time to be invested with the benefits of higher accuracy. The benefits of having an assessment of higher accuracy in the belief network in the making may be a higher accuracy of a computed probability of interest and an improved performance more in general. For example, for a belief network that is to be used for diagnostic purposes, performance may be measured as the percentage of correctly diagnosed cases. Refining a probability assessment for this network would only be worthwhile if it would increase the number of correct diagnoses. Once a belief network in the making exhibits satisfactory overall behaviour, refining assessments may be found to be no longer cost-effective.

The probability assessments that have been identified in the fourth step of the elicitation procedure are known to induce a considerable plausible effect on a probability of interest; moreover, any such assessment can be cost-effectively refined. For these assessments, the entire elicitation procedure is recursively repeated. The probabilities concerned are assessed anew in the first step of the next cycle of the procedure, using yet unexplored sources of probabilistic information. Since the plausible intervals of the initial assessments for these probabilities do not underestimate the uncertainties involved, refinement is likely to result in smaller plausible intervals for the new assessments. In the second step, once again several sensitivity analyses are performed. These analyses should not only focus on the new assessments, but also on previously investigated assessments as their effect on a probability of interest upon variation may have changed as a result of the refinement of the network. In addition, analyses may be performed with respect to assessments that have not been investigated before. By recursively refining probability assessments, the performance of the belief network in the making is likely to gradually improve. The elicitation procedure is stopped as soon as the costs of further elicitation outweigh the improvement in the network's performance or higher accuracy can no longer be attained due to lack of knowledge.

To conclude, we would like to note that, currently, little practical experience with the elicitation procedure outlined above is available. We have performed a preliminary experiment with the procedure in view of a Bayesian belief network for congenital heart disease [Coupé *et al.*, 1999]. Encouraged by the results from this experiment, we are currently implementing the procedure in the construction of a decision-theoretic network for treatment planning for patients with oesophageal cancer [Van der Gaag *et al.*, 1999].

6 Computational issues

Straightforwardly performing all sensitivity analyses, as described in Section 5, for a Bayesian belief network of realistic size can be quite time-consuming. In principle, every probability assessment in the quantitative part of the network is varied systematically as is every pair of probability assessments. For every probability, a number of deviations from the initial assessment are investigated, twenty values per probability not being an exception. For every value or pair of values under study, the probability of interest is computed from the network. To give some impression of the number of network computations that may thus be required, we address a one-way sensitivity analysis of three small realistic belief networks. The well-known ALARM-network for patient monitoring contains 37 nodes, for which 571 (non-redundant) probability assessments are specified [Beinlich *et al.*, 1989]. The WILSON-network is a network for the diagnosis of Wilson's disease; it specifies 21 nodes and 162 assessments [Korver & Lucas, 1993]. The VSD-network, to conclude, provides for prognostic assessment of children with a ventricular septum defect; this network comprises 38 nodes for which 807 probability assessments are specified [Peek & Ottenkamp, 1997]. If each assessment is varied from 0 to 1 in steps of 0.1, a one-way sensitivity analysis of the ALARM-network requires $571 \cdot 11 = 6281$ network computations; the WILSON- and VSD-networks require 1782 and 8877 computations, respectively. The computational burden of performing all sensitivity analyses for a Bayesian belief network, fortunately, can be reduced considerably. In general, not every assessment will be investigated in every analysis, as we have argued before in the previous section. The computational burden can in addition be further reduced by exploiting the probabilistic relationships among the variables that are represented in the network's digraph. In this section, we elaborate on the latter observation.

The digraph of a Bayesian belief network captures the influential relationships among the represented variables, or conversely, the independences among these variables. Knowledge of these independences allows for identifying probability assessments that upon variation cannot affect a probability of interest. For example, the non-ancestors of a network's node of interest that are not observed and do not have any observed descendants, cannot exert any diagnostic influence on the probability of interest. Varying the probability assessments for these nodes will therefore not have any effect on the probability of interest. Also, case-specific observations entered into the network may effectively block influences between nodes. In our example belief network, for instance, once the presence or absence of a metastatic cancer has been established in a patient, varying the probability assessments for the node modelling the level of serum calcium no longer has any influence on the probability of a brain tumour. The nodes whose assessments upon variation may influence a network's probability of interest constitute this probability's *sensitivity set*. A sensitivity set depends to a large extent on the case-specific observations that have been entered into the network. Since the probability assessments for the nodes that are not included in a sensitivity set under study cannot influence a network's probability of interest upon variation, these assessments can be excluded from the sensitivity analysis; the analysis can thus be restricted to the sensitivity set. To give some impression of the size of a sensitivity set, we address again the ALARM-, WILSON-, and VSD-networks. If no case-specific observations have been entered, the ALARM-network reveals only three assessments from among its total of 571 assessments to be influential; the WILSON- and VSD-networks reveal 6 and 151 influential probability assessments, respectively. After entering a typical patient profile, the number of influential assessments increases to 54, 32, and 491, for the three networks, respectively. For further details of the sensitivity set and its computation, the reader is referred to [Coupé & Van der Gaag, 1998].

So far, we have exploited the independences among the variables portrayed by the digraph of a Bayesian belief network to identify probability assessments that upon variation cannot influence the network's probability of interest. The independences can even be further exploited as they constrain the relation between the network's probability of interest and an assessment, or pair of assessments, under study to a simple mathematical function. In a one-way sensitivity analysis of a Bayesian belief network, the network's probability of interest relates to a probability assessment under study as a quotient of two functions that are linear in this assessment. We reconsider Figure 7 showing for our example belief network the probability of interest $\Pr(b \mid sh)$ as a function of the assessments for the probabilities $p(mc)$, $p(b \mid \neg mc)$, and $p(sh \mid \neg b)$, respectively; for the assessment $x = p(b \mid \neg mc)$, for example, this function equals

$$\Pr(b \mid sh) \quad = \quad \frac{4 \cdot x + 0.20}{x + 3.80}$$

If a probability assessment under study pertains to a node that is an ancestor of the node of interest and does not have any observed descendants, the function expressing the probability of interest in terms of this assessment reduces to a simple linear function. More specifically, if the probability of interest is a prior probability, it relates linearly to any of the influential assessments. We consider once again Figure 4 showing for our example belief network the probability of interest $\Pr(c)$ as a function of the assessments for the conditional probabilities $p(b \mid mc)$, $p(isc \mid mc)$, and $p(isc \mid \neg mc)$, respectively; for the assessment $x = p(isc \mid \neg mc)$, for example, this function equals

$$\Pr(c) = 0.57 \cdot x + 0.21$$

The properties described above for the functions that are yielded by a one-way sensitivity analysis of a Bayesian belief network, provide for considerably reducing the computational burden of the analysis. Systematic variation of a probability assessment is no longer necessary. The constants in the functional relation between a probability of interest and a probability assessment under study can be determined by computing this probability of interest from the network for a small number of deviations from the assessment and solving the resulting system of linear equations. For an assessment that is related linearly to the probability of interest, two network computations suffice; for all other assessments three network computations are required. To give some impression of the reduction in computational effort thus attained, we address once again the three belief networks we have studied. If no case-specific observations have been entered, the ALARM-network reveals only three influential probability assessments. The probability of interest relates as a linear function to any of these assessments. To compute the constants in these functions, therefore, $2 \cdot 3 = 6$ network computations suffice. After entering a typical patient profile, the network reveals 54 influential probability assessments. To establish the mathematical functions expressing the probability of interest in terms of any of these assessments, $3 \cdot 54 = 162$ network computations are required. Note that this number of computations is just 3% of the number of computations that would have been performed in a straightforward one-way sensitivity analysis of the network. Table 1 summarises these results, along with the results that we have obtained for the WILSON- and VSD-networks; for the number of computations required, the table lists the number of computations for systematic variation of all probability assessments from a network (*variation*), the number of computations for systematic variation of just the assessments pertaining to nodes from the sensitivity set for a probability of interest (*set*), and the number of computations for determining the constants in the mathematical functions relating a probability of interest to an assessment under study (*functions*).

For a two-way sensitivity analysis of a Bayesian belief network similar observations hold as for a one-way sensitivity analysis. As a one-way analysis, a two-way sensitivity analysis can be restricted to the probability assessments pertaining to the nodes from a sensitivity set under study. Also, the independences among the variables portrayed by the network's digraph once again constrain the relation between the probability of interest and two assessments that are being varied to a simple mathematical function. The probability of interest relates to two assessments under study as a quotient of two functions that are bi-linear in these assessments. We reconsider Figure 8 showing for our example belief

network (total # probs)	*# observations*	*# influential assessments*	*# computations*		
			variation	*set*	*functions*
ALARM (571)	0	3	6281	33	6
	6	54	6281	594	162
WILSON (162)	0	6	1782	66	12
	4	32	1782	352	96
VSD (807)	0	151	8877	1661	302
	9	491	8877	5401	1473

Table 1: The reduction in computational effort obtained for three different realistic belief networks by using the properties of a one-way sensitivity analysis (the meanings of the numbers are given in the text).

network the probability of interest $\Pr(b \mid sh)$ as a function of the probability assessments $x = p(b \mid \neg mc)$ and $y = p(sh \mid b)$; the function equals

$$\Pr(b \mid sh) = \frac{1.10005 \cdot x \cdot y - 0.00056 \cdot x + 0.0559 \cdot y - 0.00034}{x \cdot y - 0.6268 \cdot x + 0.0835 \cdot y + 0.7811}$$

In this function, the terms $-0.00056 \cdot x$ and $-0.6268 \cdot x$ pertain to the effect of variation of just the probability assessment $p(b \mid mc)$; the terms $0.0559 \cdot y$ and $0.0835 \cdot y$ pertain to the assessment $p(isc \mid mc)$. The terms $1.10005 \cdot x \cdot y$ and $x \cdot y$ with each other capture the interaction effect of the two assessments on the network's probability of interest. These terms provide information that cannot be revealed by one-way analyses with respect to the two assessments separately. The mathematical function expressing a network's probability of interest in terms of two assessments reduces to a simple bi-linear function for assessments that pertain to nodes that are ancestors of the node of interest and do not have any observed descendants. More specifically, if the probability of interest is a prior probability, it relates bi-linearly to any pair of influential assessments. We consider once again Figure 5 showing for our example belief network the probability of interest $\Pr(c)$ as a function of the assessments $x = p(b \mid mc)$ and $y = p(isc \mid mc)$; the function equals

$$\Pr(c) = -0.15 \cdot x \cdot y + 0.15 \cdot x + 0.15 \cdot y + 0.194$$

From this function it is readily seen that the two probability assessments under study upon variation have a negative interaction effect on the probability of interest. Not every pair of probability assessments that are being varied will have an interaction effect on a probability of interest, however. For example, any two probability assessments that pertain to incompatible probabilities, in the sense of specifying complementary values for the same variable, will not interact. The function expressing the probability of interest in terms of two such assessments will lack a product term. We consider once again Figure 6 showing for our example belief network the probability of interest $\Pr(c)$ as a function of the assessments $x = p(b \mid mc)$ and $y = p(isc \mid \neg mc)$; note that the two assessments under study pertain to incompatible probabilities. The function equals

$$\Pr(c) = 0.03 \cdot x + 0.57 \cdot y + 0.204$$

A two-way sensitivity analysis involving assessments for incompatible probabilities does not reveal any unanticipated effects on a prior probability of interest beyond the effects shown by one-way sensitivity analyses for the two assessments separately. Any such pair of assessments can therefore be excluded from the analysis. The properties described above for the functions that are yielded by a two-way sensitivity analysis of a Bayesian belief network once again allow for considerably reducing the computational burden of the analysis. The constants in the functions can again be determined by computing the network's probability of interest for a small number of deviations from the assessments under study and solving the resulting system of equations. The number of network computations thus required ranges from four to seven per pair of assessments. For an impression of the reduction in computational effort thus attained, we refer once more to Table 1 and observe that the number of network computations required for a two-way sensitivity analysis equals the number of computations for a one-way analysis to the second power, approximately.

The sensitivity analyses of the three realistic belief networks we have studied, shows that the computational burden involved can be reduced considerably by exploiting the probabilistic relationships portrayed by a network's digraph. This reduction increases the practicability of the procedure suggested in the previous section for quantifying Bayesian belief networks.

7 Conclusions and Directions for Further Research

When building a Bayesian belief network, a large number of probabilities will have to be assessed by experts in the domain of application. Experience shows that experts often are reluctant to assess all probabilities required, feeling that they are unable to give assessments having a high level of accuracy. However, not every probability in a belief network in the making needs to be assessed with a high level of accuracy to arrive at satisfactory behaviour of the network. In this paper, we have addressed performing a sensitivity analysis of a belief network to gain insight into which probabilities require a higher level of accuracy and which probabilities do not. Insight into which probabilities are the most influential allows for careful focusing of elicitation efforts. We therefore believe sensitivity analysis to be a practical aid in probability elicitation for Bayesian belief networks. We have proposed an elicitation procedure in which sensitivity analysis takes a central role. In the procedure, alternatingly, analyses of a belief network in the making are performed and probability assessments are refined, until satisfactory behaviour of the belief network is obtained, until the costs of further elicitation outweigh the benefits of higher accuracy, or until higher accuracy can no longer be attained due to lack of knowledge.

Sensitivity analysis of a Bayesian belief network allows for assessing the sensitivity of a probability of interest to the various probability assessments specified in the network. When performed with respect to a prior probability of interest, a sensitivity analysis serves to assess the network's quality and robustness in its reflecting a prior probability distribution. The network's robustness in modeling a posterior probability distribution can be assessed by performing a sensitivity analysis with respect to a posterior probability of interest, conditional on case-specific observations; such an analysis may be looked upon as yielding a measure of confidence for the probability of interest computed from the network for a case under consideration. Sensitivity analysis therefore serves for both general and case-specific validation of a belief network under study.

A one-way sensitivity analysis of a Bayesian belief network reveals the effect of varying a single probability assessment on a probability of interest. A two-way sensitivity analysis in addition yields insight into the joint effect of varying two probability assessments simultaneously. Performing a two-way analysis therefore is particularly useful for uncovering and studying synergistic influences among probability assessments. With a two-way sensitivity analysis, however, it is not possible to reveal higher-order synergistic influences involving more than two assessments. For this purpose, a higher-order sensitivity analysis would be required. Interpreting the results of such an analysis is often very hard. To be of practical use, appropriate tools need to be designed for this purpose.

The computational complexity of performing a sensitivity analysis of a Bayesian belief network of realistic size is an issue of major concern. In this paper, we have briefly presented several properties that allow for reducing the computational burden involved to at least some extent. In a forthcoming technical paper, we have detailed these and additional properties that will allow for practical use of sensitivity analysis. Building upon these properties, we have implemented a prototype tool for performing one-way and two-way sensitivity analyses of belief networks. We are currently experimenting with this tool in view of various real-life Bayesian belief networks.

To conclude, we would like to note that, currently, we have little practical experience with our elicitation procedure. We have performed a preliminary experiment with the procedure for a Bayesian belief network for congenital heart disease and are currently using the procedure in the construction of a large decision-theoretic network for treatment planning for patients with oesophageal cancer. The experiences with our elicitation procedure so far encourage us to pursue this line of research in the future.

Acknowledgement

We would like to thank René Eijkemans and anonymous reviewers for their valuable comments on earlier drafts of this paper.

References

[Andreassen *et al.*, 1987] S. Andreassen, M. Woldbye, B. Falck, and S.K. Andersen (1987). MUNIN – A causal probabilistic network for interpretation of electromyographic findings. *Proceedings of the Tenth International Joint Conference on Artificial Intelligence*, pp. 366 – 372.

[Beinlich *et al.*, 1989] I.A. Beinlich, H.J. Suermondt, R.M. Chavez, and G.F. Cooper (1989). The ALARM monitoring system: a case study with two probabilistic inference techniques for belief networks. In: J. Hunter, J. Cookson, and J. Wyatt (Eds).

Proceedings of the Second Conference on Artificial Intelligence in Medicine, Springer-Verlag, Berlin, pp. 247 – 256.

[Bruza & Van der Gaag, 1994] P.D. Bruza and L.C. van der Gaag (1994). Index expression belief networks for information disclosure. *International Journal of Expert Systems: Research and Applications*, vol. 7, pp. 107 – 138.

[Cooper, 1984] G.F. Cooper (1984). *NESTOR: a Computer-based Medical Diagnostic Aid that Integrates Causal and Probabilistic Knowledge.* Report HPP-84-48, Stanford University.

[Coupé *et al.*, 1999] V.M.H. Coupé, N.B. Peek, J. Ottenkamp, and J.D.F. Habbema (1999). Using sensitivity analysis for efficient quantification of a belief network. *Artificial Intelligence in Medicine*, to appear.

[Coupé & Van der Gaag, 1998] 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. *Prague Stochastics '98 – Proceedings of the Joint Session of the 6th Prague Symposium of Asymptotic Statistics and the 13th Prague Conference on Information Theory, Statistical Decision Functions and Random Processes*, Union of Czech Mathematicians and Physicists, pp. 81 – 86.

[Druzdzel & Van der Gaag, 1995] M.J. Druzdzel and L.C. van der Gaag (1995). Elicitation of probabilities for belief networks: combining qualitative and quantitative information. *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, pp. 141 – 148.

[Habbema *et al.*, 1990] J.D.F. Habbema, P.M.M. Bossuyt, and D.W.J. Dippel (1990). Analysing clinical decision analyses. *Statistics in Medicine*, vol. 9, pp. 1229 – 1242.

[Heckerman *et al.*, 1992] D.E. Heckerman, E.J. Horvitz, and B.N. Nathwani (1992). Toward normative expert systems. Part 1: The Pathfinder project. *Methods of Information in Medicine*, vol. 31, pp. 90 – 105.

[Jensen, 1995] A.L. Jensen (1995). Quantification experience of a DSS for mildew management in winter wheat. In: M.J. Druzdzel, L.C. van der Gaag, M. Henrion, and F.V. Jensen. *Working Notes of the Workshop on Building Probabilistic Networks: Where Do the Numbers Come From ?*, pp. 23 –31.

[Korver & Lucas, 1993] M. Korver and P.J.F. Lucas (1993). Converting a rule-based expert system into a belief network. *Medical Informatics*, vol. 18, pp. 219 – 241.

[Lauritzen & Spiegelhalter, 1988] 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.

[McGraw & Harbison-Briggs, 1989] K.L. McGraw and K. Harbison-Briggs (1998). *Knowledge Acquisition: Principles and Guidelines*, Prentice-Hall.

[Morgan & Henrion, 1990] M.G. Morgan and M. Henrion (1990). *Uncertainty, a Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis.* Cambridge University Press, Cambridge.

[Pearl, 1988] J. Pearl (1988). *Probabilistic Reasoning in Intelligent Systems. Networks of Plausible Inference.* Morgan Kaufmann, Palo Alto.

[Peek & Ottenkamp, 1997] N.B. Peek and J. Ottenkamp (1997). Developing a decision-theoretic network for a congenital heart disease. In: E. Keravnou, C. Garbay, R. Baud, and J. Wyatt (Editors). *Proceedings of the Sixth Conference on Artificial Intelligence in Medicine Europe*, Lecture Notes in Artificial Intelligence 1211, Springer Verlag, Berlin, pp. 157 – 168.

[Studer *et al.*, 1998] R. Studer, V.R. Benjamins, and D. Fensel (1998). Knowledge engineering: principles and methods. *Data & Knowledge Engineering*, vol. 25, pp. 161 – 197.

[Tversky *et al.*, 1982] D. Kahneman, P. Slovic, and A. Tversky (1982). *Judgment under Uncertainty: Heuristics and Biases*, Cambridge University Press, Cambridge.

[Van der Gaag *et al.*, 1999] L.C. van der Gaag, S. Renooij, C.L.M. Witteman, B. Aleman, and B.G. Taal (1999). How to elicit may probabilities. *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*, to appear.

[Von Winterfeldt & Edwards, 1986] D. von Winterfeldt and W. Edwards. *Decision Analysis and Behavioral Research.* Cambridge University Press, New York.