Sensitivity Analysis of Decision-Theoretic Networks

Financial support for the publication of this thesis by the Department of Public Health of the Erasmus University Rotterdam and of the Patent Agency LIOC is gratefully acknowledged.

The research reported in this thesis was supported by the Netherlands Computer Science Research Foundation with financial support from the Netherlands Organization for Scientific Research (NWO).

ISBN 90-9013846-3

Printed by: Print Partners Ipskamp. Enschede

Sensitivity Analysis

of

Decision-Theoretic Networks

Sensitiviteitsanalyse van Besliskundige Netwerken

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de Rector Magnificus
Prof.dr.ir. J.H. van Bemmel
en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op
woensdag 27 september 2000 om 13.45 uur

door

Veerle Marleen Herman Coupé

geboren te Gent (België)

Promotiecommissie

Promotores	:	Prof.dr. J.D.F. Habbema
		Prof.dr.ir. L.C. van der Gaag
Overige leden	:	Prof.dr. B.C.J.M. Fauser
		Prof.dr.ir. R. Dekker
		Prof.dr. F.V. Jensen

Contents

Publications and manuscripts reprinted in this thesis

Chapter 2

V.M.H. Coupé, L.C. van der Gaag, and J.D.F. Habbema. Sensitivity analysis: an aid for belief-network quantification, *Knowledge Engineering Review*, to appear.

Chapter 3

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*, Vol. 17, pp. 223-247.

Chapter 4

V.M.H. Coupé and L.C. van der Gaag. Properties of sensitivity analysis of Bayesian belief networks, *Annals of Mathematics and Artificial Intelligence*, to appear.

Chapter 5

V.M.H. Coupé, F.V. Jensen, U. Kjærulff, and L.C. van der Gaag. A computational architecture for n-way sensitivity analysis of Bayesian networks, submitted for publication.

Chapter 6

L.C. van der Gaag and V.M.H. Coupé (1999). Sensitivity analysis for threshold decision making with Bayesian belief networks, E. Lamma and P. Mello (eds.), *AI*IA 99: Advances in Artificial Intelligence*, Lecture Notes in Artificial Intelligence, Springer-Verlag, Berlin, pp. 37 – 48.

Chapter 7

V.M.H. Coupé, L.C. van der Gaag, F.V. Jensen, and U. Kjærulff. Properties of sensitivity analysis of influence diagrams, submitted for publication.

Part I

Introduction

Chapter 1

Introduction

1.1 Background

Clinical Decision Sciences

Clinical decision sciences is a field of research that aims to analyse and support clinical decision making for individual patients. Three important aspects of clinical decision making are diagnosis, therapy planning and prognosis. In brief, diagnosis is concerned with finding an explanation for an observed set of symptoms and clinical findings in terms of diagnostic categories. In therapy planning, the focus is on comparing the benefits and risks of various treatment modalities to arrive at a therapy choice that is optimal in view of the uncertainties involved. Prognosis is concerned with assessing the likelihood of possible outcomes that may occur during a disease process, such as mortality and morbidity. Central to the field of clinical decision making is reasoning under uncertainty and reasoning about preferences. For this purpose, in clinical decision sciences various techniques are available. An overview of general decision analytic techniques can be found in [Von Winterfeldt & Edwards, 1986, Raiffa & Schlaifer, 1961, Keeney & Raiffa, 1976, Chernoff & Moses, 1959]; techniques tailored to clinical decision problems are discussed in [Weinstein & Fineberg, 1980, Pauker & Kassirer, 1980, Sox *et al.*, 1988, Habbema & Hilden, 1979, Habbema *et al.*, 1981, Habbema *et al.*, 1990].

In prognosis and diagnosis, reasoning with uncertainty plays an important role. To support the prognostic and diagnostic task, in classical decision analysis, regression techniques are used. They provide, for an individual patient, a quantitative assessment of the probability of the outcome under consideration. Regression techniques establish from data the relation between a clinically relevant outcome measure, such as one year mortality as a result of a specific disease, and a set of determinants. The resulting regression model gives, for an individual patient, the probability of the outcome by filling

in the appropriate values of the determinants. Examples of regression techniques are *probit analysis* and *logistic regression analysis* [Hosmer & Lemeshow, 1989]. In a probit analysis, the linear predictor, which is a multilinear combination of the determinants under consideration, is assumed to obey a normal probability distribution. Under this assumption, the unknown parameters in the functional relation between this linear predictor and the probability of the outcome are established from data. In a logistic regression analysis, the unknown parameters relating the determinants to the outcome are established in a similar way, this time under the assumption that, conditional on the values of the determinants included, the outcome satisfies a binomial distribution. Both probit and logistic regression analysis focus on the occurrence of a specific event. However also the time until the occurrence of an event may be of interest. The statistical technique establishing the relation between the time until occurrence of an event and a set of determinants is called a *survival analysis* [Lawless, 1982]. Two examples of regression methods in survival analysis are *Cox regression* and *Poisson regression*. For details on the underlying assumptions of both types of analysis and their appropriate use, we refer to [Kalbfleisch & Prentice, 1980] and [Breslow & Day, 1987].

In therapy planning, the two 'classical' techniques for supporting decision making under uncertainty are the *threshold model* and the *decision tree*. In a threshold analysis, various probabilities of disease are established that divide the probability range into intervals where a specific action, be it treatment, further testing, or neither of the two, is considered to be optimal [Pauker & Kassirer, 1980]. For example, the treatment threshold probability of disease is the probability at which a physician is indifferent between giving treatment and withholding treatment. If, for a specific patient, the probability of disease exceeds the treatment threshold probability, then this physician will decide to treat the patient as if the disease were known to be present with certainty. Alternatively, the physician will withhold treatment from the patient. In addition, various threshold probabilities for testing may be established. The various threshold probabilities of disease may be obtained from a physician. In establishing the thresholds, the physician will carefully weigh the (expected) benefits of treatment or testing and the risks involved. In the assessments for the threshold probabilities, the *utilities* or *preference values* of all possible outcomes are implicitly incorporated. The resulting threshold model may support an attending physician to decide upon treatment or testing based upon the probability of disease for a specific patient under consideration.

The decision tree is the best known, and probably most used, analytic tool for supporting decision making under uncertainty. A decision tree provides for displaying the temporal sequence of decision moments and events in a clinical decision problem. Its form reveals the decision alternatives that are available to the decision maker or physician, the events that may follow from each alternative and that hence may affect the decision, and the outcomes that are associated with each possible scenario of

decision alternatives and consequences. A decision tree is built from left to right. Usually, its first branching point, at the extreme left, is a *decision node*. A decision node, indicated by a square, represents a decision moment at which one of various decision alternatives can be selected. Each alternative is associated with a branch emanating from the decision node. Depending on the alternative chosen at the first decision node, a sequence of future decisions and uncertain events may follow. Each uncertain event is represented by a *stochastic variable*, whose corresponding node is depicted as a circle in the decision tree. The possible values a stochastic variable can take are again explicitly depicted by the branches emanating from its associated node. The leaves, at the extreme right, of a decision tree represent the ultimate outcome to the patient of following a specific path consisting of decision alternatives and values for chance events from left to right in the tree.

The structure of the decision tree, as described above, is supplemented with conditional probabilities and utilities. These are called the *parameters* of the decision tree. For every stochastic variable, the probabilities of its values conditional on the path in the decision tree to the left of the associated node are required. For all possible outcomes, furthermore, a preference value, called *utility*, should be established. The conditional probabilities in a decision tree can be provided by one or more physicians – these are called subjective probability estimates – or they can be estimated from data. To obtain a patient-specific probability from data often logistic regression models or other regression models are used. In that case, prediction models developed for a specific patient population provide the input for a decision tree describing the clinical decision problem for an individual patient. The utilities in a decision tree are usually provided by physicians, but they can also be assessed by patients to whom the decision problem is relevant [Jansen *et al.*, 1998]. For probability and utility elicitation, various techniques are available [Keeney & Raiffa, 1976, Stiggelbout, 1995].

A quantified decision tree is evaluated by averaging out and folding back the tree. Averaging out is the process of removing stochastic nodes from right to left in the decision tree by multiplying the probabilities of each branch emanating from a stochastic node by the utility attached to it and summing the resulting values of all branches at the node. Folding back refers to pruning all but the most preferred branch at a decision node and assigning the value that is attached to this branch to the decision node. Averaging out and folding back is described in detail in [Weinstein & Fineberg, 1980]. As a result of this sequential process of averaging out and folding back along the branches of the decision tree, the preferred decision alternative at the first decision node is obtained. This preferred decision alternative, for the patient under consideration, maximises expected utility.

Constructing a decision tree for a decision problem gives insight into the structure of the problem and the preferences of the decision maker, that is, the person supplying the utility estimates. To obtain more detailed insight into the various aspects of the prob-

lem and its representation, additional techniques for analysing a decision tree are available. A *sensitivity analysis*, for example, provides insight into the effect of varying one or more parameters, that is probability or utility estimates, on the decision alternative with maximum expected utility [Weinstein & Fineberg, 1980, Habbema *et al.*, 1990]. It reveals the sensitivity of the preferred decision alternative to parameter variation, and, as such, gives an indication of the robustness of the preferred decision. Moreover, it reveals *which* parameters upon variation have a large effect on the preferred decision alternative and should therefore be assessed with care. An *uncertainty analysis*, also called a probabilistic or stochastic sensitivity analysis, shows the impact of uncertainty in the parameters in the decision tree on the preferred decision alternative [Morgan & Henrion, 1990, Habbema *et al.*, 1990]. Instead of taking a fixed value for each parameter in the tree, for each parameter a value is drawn from a distribution, for example the lognormal distribution. For a set of drawn values for all parameters in the tree the preferred decision alternative is computed. This is repeated a large number of times to obtain an empirical probability distribution over the decision alternatives available. The results of an uncertainty analysis show the joint effect of uncertainty in all parameter estimates on the preferred decision alternative.

Two other insightful analyses of a decision tree are the computation of the *expected value of perfect information*, or EVPI, and the *expected value of perfect control*, or EVPC. The EVPI is a measure of the importance of knowing the value of an, either observable or unobservable, stochastic variable. It is computed as the *expected* difference in maximum expected utility with and without knowing the value of this variable. The EVPI provides an upper bound for the value of additional testing, before this test is actually carried out. The EVPC is closely related to the EVPI. It is a measure of the importance of controlling the value of a stochastic variable. The EVPC thus provides an upper bound for the gain in expected utility that can be achieved from improving tests or treatments. For example, by computing the expected value of having an operation mortality of 0%, we obtain insight in the maximum gain that can be expected from trying to decrease operation mortality.

Decision-theoretic Networks

From the discussion above, it will be clear that there is a wealth of decision-analytic techniques available for analysing and solving decision problems in medical practise. They form the theoretical basis of the art of clinical decision analysis as practised today. However, the techniques described also have various shortcomings. A regression model for prognosis, for example, has a relatively simple structure in which complicated interactions among determinants or between determinants and the outcome under consideration cannot be captured. Although the values of the regression coefficients often have an intuitive interpretation in terms of the mechanism that relates the determi-

nants to the outcome, there is no detailed model of the underlying disease. A decision tree, although insightful for small problems, may rapidly become 'bushy' for complicated problems in which many decision moments and chance events are taken into account. For complicated problems, the decision tree may give rise to confusion rather than giving insight in the structure of the problem. Therefore, during the past two decades, there has been an increasing interest to supplement classical decision-analytic techniques with methods of the field of computer science and from artificial intelligence in particular. Especially, the area of knowledge-based systems has received a lot of attention. The term *knowledge-based system* refers to an information system in which human knowledge is represented symbolically and is applied to give advice concerning a specific problem at a level comparable to that of an expert in the same field. Knowledge-based systems that are tailored to reasoning and decision making nowadays often build upon the framework of *decision-theoretic networks.* Decision-theoretic networks combine a graphical representation of relations between stochastic variables, decision variables and a utility variable with the use of probability theory and utility theory. Two types of decision-theoretic network can be discerned; the *belief network* and the *influence diagram.*

A belief network, basically, is a representation of a joint probability distribution on a set of stochastic variables. It consists of a qualitative part and an associated quantitative part. The qualitative part of a belief network takes the form of a directed graph consisting of nodes and arcs. The nodes represents the stochastic variables that are important in the problem under study and the arcs represent the influential relationships between the variables. The direction of the arcs is often interpreted as having a causal meaning; the variable at the head of the arc is designated as the effect of the cause at the tail of the arc. This notion of causality should, however, be interpreted broadly. Absence of an arc between two variables indicates that there is no direct relation between the two variables; the variables are conditionally independent. The quantitative part of a belief network consists of a set of conditional probabilities. These conditional probabilities quantify the strenghts of the influential relationships among the variables. For each node in the network, the probabilities of its values conditional on every possible combination of values for the direct predecessors in the graph is specified. The conditional probabilities of a belief network are termed the parameters of the network. The qualitative and quantitative part of a belief network together allow for computing any prior or posterior probability from the network. For this purpose, several algorithms have been developed that exploit the independence relations represented in the qualitative part of the network [Pearl, 1988, Lauritzen & Spiegelhalter, 1988].

An influence diagram is a concise representation of a decision problem involving uncertainty. Its qualitative part consists of various types of node and arc. The nodes not only represent stochastic variables but decision variables and a utility variable as well [Howard & Matheson, 1981, Shachter, 1986]. As in a decision tree, the decision

variables represent the various moments of choice for the decision maker and the utility node represents the possible outcomes and their associated preference values. The arcs between the stochastic variables in an influence diagram have the same meaning as in a belief network. Arcs pointing from a stochastic variable into a decision node indicate that the values of the variables at the tail of the arcs should be observed before a decision can be made. The nodes at the tail of the arcs pointing into the utility node represent the variables that directly affect utility. The quantitative part of an influence diagram consists of a set of conditional probabilities for the stochastic variables in the diagram and a set of utilities encoding the desirability of the different outcome scenarios represented by the utility node. Note that influence diagrams and belief networks are closely related; in fact, an influence diagram that contains only stochastic variables is a belief network. To compute from an influence diagram the decision alternative with maximum expected utility, various algorithms are available [Shachter, 1986, Cooper, 1988].

Since a belief network provides for the computation of the posterior probability of an outcome of interest given a set of observations for a patient under consideration, its envisaged use is comparable to that of regression models in prognosis and diagnosis. Likewise, the objective of constructing an influence diagram is the same as of building a decision tree: to obtain insight in and support for a decision problem. Although different graphical representations of a decision problem, an influence diagram and a decision tree are based on the same underlying mathematical model. In fact, any influence diagram can, in principle, be translated into a decision tree and vice versa [Owens *et al.*, 1997, Nease & Owens, 1997]. However, there are considerable differences between influence diagrams and decision trees. Influence diagrams represent the structure of a decision problem more compactly than decision trees do. In an influence diagram the relationships between the stochastic variables of importance are depicted without explicitly enumerating their values; in a decision tree, on the other hand, all possible scenarios, consisting of combinations of values for stochastic variables and decision nodes, are represented explicitly. For large and complex problems, decision trees may rapidly become messy and incomprehensible. An influence diagram may then give a more transparent image of the decision problem and may, moreover, facilitate the communication between the decision analyst and physician. This is actually what we experienced when building both a decision tree and an influence diagram for therapeutic planning in children with a ventricular septal defect. During construction, the decision tree grew so complex that it was hard to get insight into the decision problem. Building the influence diagram was more successful, as will be shown in Chapter 3 of this thesis. Influence diagrams, however, have also limitations compared to decision trees. In a decision tree asymmetries in the decision problem under study – that is, different chance events or decision moments following two different decision alternatives – are easily indentified from the graphical structure.

In an influence diagram, asymmetries are hidden in the probability estimates for the chance nodes.

Between belief networks and regression models many differences exist. A belief network's graphical structure provides a picture of the represented disease process. It allows for easily understanding how observations for a specific patient influence intermediate, unobservable variables and may lead to a specific outcome. Regression models in this sense are black boxes. Furthermore, in a belief network all possible synergistic effects between the stochastic variables that have an arc into the same node are taken into account. In a regression model, usually, only synergistic effects between two or maximally three variables are considered. Another difference is that, in computing the probability of the outcome from a belief network any available patient information can be taken into account. When patient information is lacking, the belief network is still capable of providing an estimate of the outcome. The uncertainty about the lacking information is automatically incorporated in the probability of the outcome. Regression models are not flexible in that sense. In principle, for all determinants in the model observations should be available. Furthermore, regression models are based on patient data, whereas belief networks can be quantified with expert estimates. In problem domains where there is too little data to build a regression model but where expert knowledge is available, it is thus still possible to model the problem using a belief network. On the other hand, it is easier to have an impression of the correctness of the results from a regression model than of a belief network. As regression models are fitted on data, automatically a first validation is carried out in the form of the goodness of fit of the model.

As we see from the similarities and differences between belief networks and influence diagrams on the one hand and regression models and decision trees on the other hand, decision-theoretic networks appear to have the ability to become a useful complement to existing decision-analytic techniques. They solve some of the shortcomings of decision trees and regression models that were discussed earlier. By now, various medical applications based on decision-theoretic networks are reported in the literature. Applications that build on the framework of belief networks for knowledge representation and inference are described in for example [Andreassen *et al.*, 1987, Heckerman & Nathwani, 1992, Heckerman *et al.*, 1992, Korver & Lucas, 1993, Shwe *et al.*, 1991, Peek & Ottenkamp, 1997, Van der Gaag *et al.*, 1999]. Applications using influence diagrams are found in for example [Andreassen *et al.*, 1991, Quaglini *et al.*, 1994].

Although various applications of the framework of decision-theoretic networks have been developed, until now, no applications are known to have been implemented in clinical practice. Probably, the main reason for this is that building a decision-theoretic network for a complex problem is a difficult and time-consuming task. Belief networks and influence diagrams are usually constructed with the help of experts in the field of application. Two closely related tasks can be discerned: the con-

struction of the qualitative part of the network and its subsequent quantification. Building the qualitative part has parallels to designing a model for a more traditional knowledge-based system [Henrion, 1989, Goldman & Charniak, 1993]. Therefore, well-known knowledge-engineering techniques can be employed to a large extent [Schreiber *et al.*, 1994, McGraw & Harbison-Briggs, 1989, Studer *et al.*, 1998]. Detailed methodologies for handling specific problems encountered when building decision-theoretic networks are being developed [Peek & Ottenkamp, 1997]. Although building the qualitative part of a decision-theoretic network may require considerable effort, generally this task is considered to be doable. Unfortunately, this does not hold with regard to the quantitative part of a decision-theoretic network. Quantifying a network is, in general, considered a hard task.

To quantify a belief network or an influence diagram, a local conditional probability distribution for each stochastic variable in the network is to be assessed. For an influence diagram also a set of utilities is to be established. To assess the utilities in an influence diagram, often outcome scenarios have to be evaluated with which the assessor is not acquainted. This is generally considered to be hard. Moreover, evaluating the calibration of utility assessments is very difficult, if not practically impossible. To quantify the stochastic variables in a decision-theoretic network, often a very large number of conditional probabilities have to be assessed. The probability distribution of a stochastic variable in the network is conditioned on its predecessors in the qualitative part; the number of conditional probabilities to be assessed thus grows exponentially in the number of predecessors. For most applications abundant probabilistic information is available from literature or from statistical data. However, it often turns out that this information is not directly amenable to encoding in a decision-theoretic 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. Furthermore, a substantial number of network variables may represent unobservable intermediate disease states for which empirical probabilistic information is difficult or impossible to obtain. The majority of the probabilities required will therefore have to be assessed by domain experts. Subjective probability assessments, however, are known to suffer from several types of bias and may not be properly calibrated [Kahneman *et al.*, 1982]. Although various elicitation techniques have been designed to avert these problems to at least some extent, the large number of probabilities required for a decision-theoretic network often prohibits the use of these rather time-consuming methods.

In [Van der Gaag *et al.*, 1999], experience with the elicitation of probabilities for a belief network for oesophageal carcinoma is described. The experience of the authors confirms the above-mentioned observations. Initially, the authors started with the elicitation of the three thousand probabilities pertaining to fourty network nodes using

the well-known methods of the probability scale for marking estimates, the frequency method and lotteries [Morgan & Henrion, 1990]. Driven by the serious biases observed and the time required for these methods, they developed a new elicitation technique. Their technique is based on the idea of presenting conditional probabilities as fragments of text and providing a scale for marking estimates with both numerical and verbal anchors. In quantifying the network for ventricular septal defects, described in Chapter 3, we encountered similar obstacles as described in [Van der Gaag *et al.*, 1999]. In particular, we experienced the limited value of available data and medical literature. Furthermore, establishing the probability estimates from the physician involved was a time-consuming task.

The problems encountered in the construction and quantification of a belief network or influence diagram partly accord with those come across in building a decision tree. Building the structure of either an influence diagram or a decision tree requires the identification of the variables and decision moments relevant to the decision problem and their influential relations. Likwise, both for an influence diagram and a decision tree, utilities have to be assessed by experts. The main difference lies in the scale, or level of detail, at which a decision problem is modeled. In general, the probabilistic part of an influence diagram contains far more stochastic variables and, hence, far more probabilities than a decision tree. Also, usually, far more different outcome scenarios are being represented in an influence diagram; a larger amount of utilities thus have to be assessed. Moreover, although over the years a large number of techniques have been developed for supporting the analysis of decision problems using decision trees, for building influence diagrams or belief networks such methods are as yet lacking. The research presented in this thesis is motivated by this observation.

1.2 Rationale for this thesis

To enhance the applicability of decision-theoretic networks for clinical decision making, techniques to assist in their construction and evaluation should be designed. Acknowledging this, the objectives of this thesis are:

- The development of efficient methods for sensitivity analysis of decision-theoretic networks;
- Exploring the use of sensitivity analysis in building decision-theoretic networks.

As indicated before, in building decision-theoretic networks generally a large number of parameters, that is, conditional probabilities and utilities, have to be assessed. These assessments inevitably are inaccurate. To evaluate the possible effect of inaccuracies in a network's assessments, a sensitivity analysis can be performed. For a decision-theoretic network, sensitivity analysis amounts to varying the assessments for

one or more parameters of the network's quantitative part simultaneously and investigating the effects on a probability of interest, on a diagnosis or on a decision computed from the network [Laskey, 1995, Castillo *et al.*, 1997b]. Upon such an analysis, some parameters will show a considerable effect, while others will hardly reveal any influence.

Performing sensitivity analysis of a decision-theoretic network in a straightforward way, unfortunately, is highly time-consuming. The parameters under study are varied stepwise and in a systematic way. The effect of these variations on the output of the network are evaluated by using any standard evaluation algorithm. Even for rather small decision-theoretic networks, the computational burden of this method of performing sensitivity analysis is considerable. In fact, it is prohibitive when sensitivity analysis is to be used for verifying the effect of inaccuracies in, for example, daily medical practice. To be of practical use, therefore, more efficient methods for sensitivity analysis of decision-theoretic networks are indispensable. In this thesis, more efficient methods for sensitivity analysis of both belief networks and influence diagrams are presented.

Apart from studying the effect of inaccuracies on the network's output, sensitivity analysis can also be a useful tool in quantifying decision-theoretic networks. Sensitivity analysis reveals the conditional probabilities and utilities in a decision-theoretic network that independently or jointly have a large effect on the outcome of the network. As such, it can be used to facilitate the quantification of a probabilistic network by focusing on the most influential assessments. In this thesis, the role of sensitivity analysis in quantifying belief networks is also investigated.

1.3 Outline of the thesis

This thesis consists of four different parts. Part I is a general introduction, Part II focusses on sensitivity analysis of belief networks, Part III extends some of the methods introduced in part II to influence diagrams, and in Part IV a general discussion and summary of the results presented in this thesis is given.

In Chapter 2 of Part II, it is investigated how sensitivity analysis can be used in making the process of quantifying a belief network more effective. We propose a procedure of iteratively performing sensitivity analyses of an initially roughly quantified network. The results from the sensitivity analyses reveal the influential conditional probabilities in the network upon which subsequent refinement efforts can be focussed. In Chapter 3, subsequently, this iterative procedure is tested. For the study, a belief network describing the pathophysiology of the congenital heart disease *ventricular septal defect* [Peek & Ottenkamp, 1997] is used. Chapter 4 describes an efficient method for performing one-way sensitivity analysis of a belief network. The method exploits the qualitative part of the network to identify conditional probabilities that cannot

influence the network's output given a specific set of observed variables. Furthermore, the method builds upon the property of a belief network that the network's output can be expressed as a quotient of two functions that are linear in a conditional probability under study in a sensitivity analysis. The observed characteristics of belief networks allow to reduce the computational burden of a sensitivity analysis considerably. In Chapter 5, the last chapter of part II, the method for one-way sensitivity analysis presented in Chapter 4 is extended to sensitivity analysis with respect to an arbitrary number of probability assessments in the network. We propose in this chapter to adapt an existing evaluation algorithm for belief networks such that the functional relation expressing the network's output in terms of a subset of conditional probabilities is easily and efficiently obtained.

In Part III, the method for sensitivity of belief networks presented in Chapter 4 is extended to include decision making under uncertainty. First, in Chapter 6 decision making with belief networks by using the threshold model [Pauker & Kassirer, 1980] is considered. The focus is on the computation of bounds between which a network's probability assessment can be varied without inducing a change in recommended decision. The *minimum deviation* in a probability assessment under study that does induce a change in the currently recommended decision is taken as a measure of the network's robustness. In Chapter 7, sensitivity analysis of influence diagrams is considered. Similar functional relations as presented in Chapter 4 are presented. In general, expected utility in an influence diagram can be expressed as a quotient of two functions that are linear in either a conditional probability or a utility in the diagram. These relations allow for easily computing the minimum deviation in a single probability assessment (or set of assessments) that would induce a change in recommended decision.

The thesis ends with Part IV. In Part IV, consisting of Chapter 8, the theoretical results presented in the thesis are summarized and compared with related research. Furthermore, relevant experience with respect to using sensitivity analysis in building belief networks is discussed as well as some ideas for further research.

Part II

Sensitivity analysis of belief networks

Chapter 2

Sensitivity analysis: an aid for belief-network quantification

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.

2.1 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 quanti-

tative 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 & Nathwani, 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 [Kahneman *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 (see Chapter 4). 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 chapter, we outline sensitivity analysis of Bayesian belief networks and dis-

cuss its use in probability elicitation. In Section 2.2, we introduce the formalism of Bayesian belief networks. We address the construction of a belief network in Section 2.3. In Section 2.4, we illustrate the technique of sensitivity analysis. In Section 2.5, we detail our elicitation procedure exploiting sensitivity analysis as an aid for belief-network quantification. In Section 2.6, we address the computational issues involved in performing a sensitivity analysis of a Bayesian belief network. The chapter ends with some conclusions and directions for further research in Section 2.7.

2.2 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 chapter 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 2.1. 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.

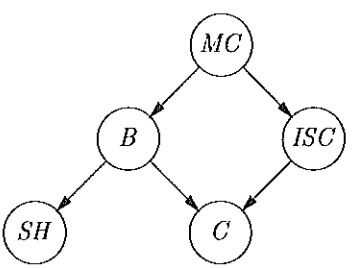


Figure 2.1: The digraph of an example Bayesian belief network, encoding information concerning the presence of a brain tumour and its associated problems in an arbitrary patient (the meanings of the variables are given in the text).

The absence of the arc $MC \rightarrow C$ in the digraph shown in Figure 2.1 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 chapter 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.

2.3 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 2.2 illustrates the basic idea. For each probability required, a 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

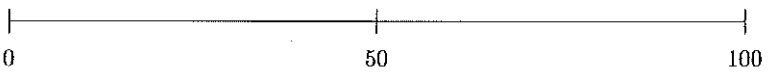


Figure 2.2: A numerical scale for probability elicitation.

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 2.3 depicts this choice between lotteries. The domain expert is asked to adjust the value of p in the reference lottery until 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

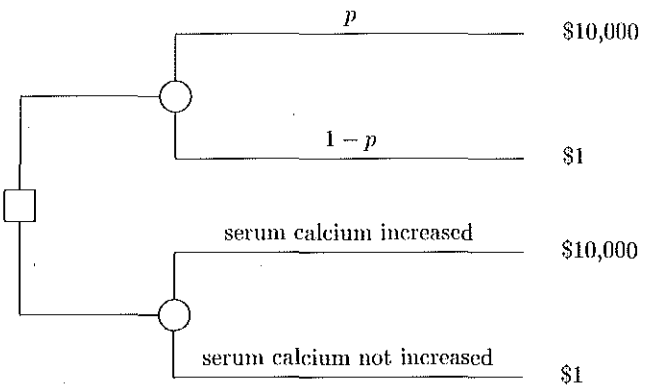
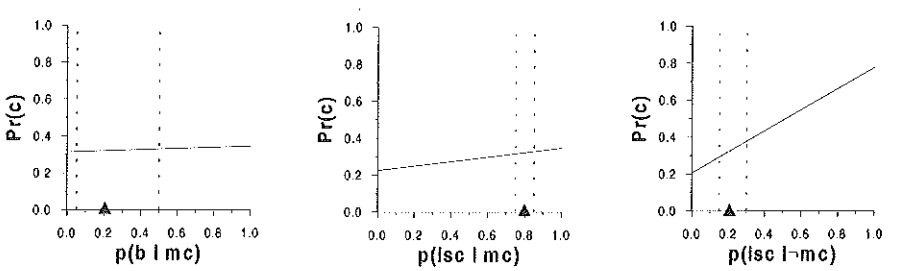


Figure 2.3: An example reference lottery, tailored to assessment of the conditional probability of the presence of an increased level of serum calcium in a patient with metastatic cancer.

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.

2.4 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 re-



(a) (b) (c)

Figure 2.4: A one-way sensitivity analysis of the example belief network from Figure 2.1. 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.

veal 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 chapter, 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 2.4. Figure 2.4(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 2.4(b) shows that varying the initial assessment 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 2.4(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 2.4 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 2.4 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 2.5.

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 2.5. The contour lines in the figure connect the combinations of values for the two probability assessments that result in the same

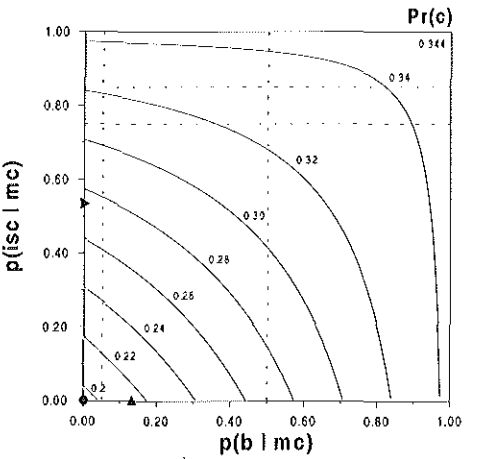


Figure 2.5: A two-way sensitivity analysis of the example belief network. The joint effect of varying the assessments for the conditional probabilities $p(b \mid mc)$ and $p(isc \mid mc)$ simultaneously on the prior probability $\Pr(c)$ is shown by contour lines.

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 2.5 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 2.2. 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 2.5, 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

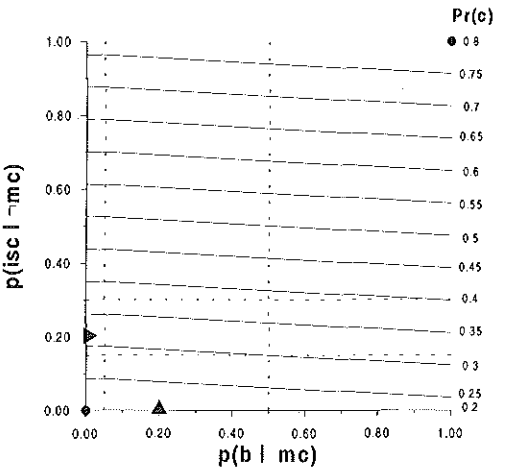
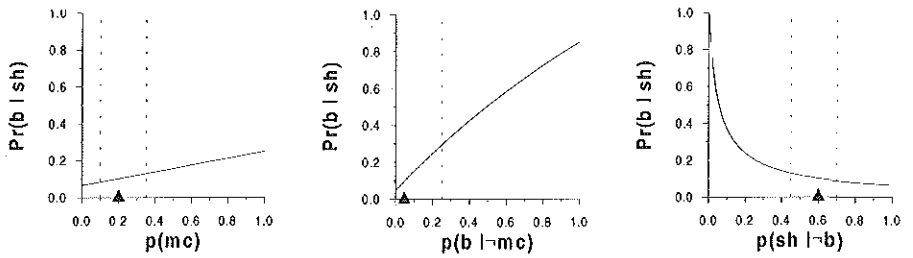


Figure 2.6: A two-way sensitivity analysis of the example belief network. The joint effect of varying the assessments for the conditional probabilities $p(b \mid mc)$ and $p(isc \mid \neg mc)$ simultaneously on the prior probability $\Pr(c)$ is shown by contour lines.

Figure 2.6. We observe from the figure that the contour lines, once again indicating values for the 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



(a) (b) (c)

Figure 2.7: 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.

$p(mc)$, $p(b \mid \neg mc)$, and $p(sh \mid \neg b)$, respectively. The results of these analyses are shown in Figure 2.7. Note that, in contrast with the one-way analyses discussed before, the analyses for the posterior 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 2.8. 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 2.8, however, the probability of interest is relatively stable.

2.5 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 2.9. We elaborate on the various different steps of the procedure separately.

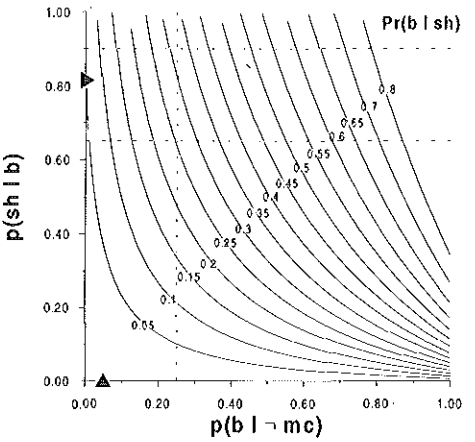


Figure 2.8: A two-say sensitivity analysis of the example belief network. The effect of varying the assessments for the conditional probabilities $p(b \mid \neg mc)$ and $p(sh \mid b)$ simultaneously, on the posterior probability $\Pr(b \mid sh)$ is shown by contour lines.

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 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.

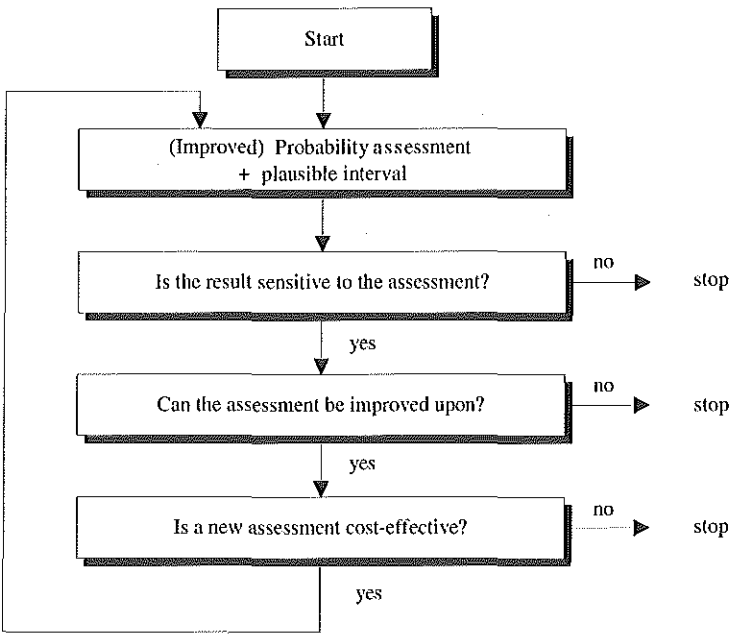


Figure 2.9: A procedure for probability elicitation for Bayesian belief networks.

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 (see Chapter 3). 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].

2.6 Computational issues

Straightforwardly performing all sensitivity analyses, as described in Section 2.5, 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. These network calculations are computationally expensive. To give some impression of the number of network computations that may 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] (see Figure 4.2). 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] (see Figure 3.1). 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 con-

siderably. 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 2.7 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

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

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 2.4 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 2.1 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*).

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 2.1: The reduction in computational effort obtained for three different realistic belief networks by using the properties of a one-way sensitivity analysis. Note the decrease in the number of computations in going from column 4 to column 6 (the meanings of the numbers are given in the text).

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 2.8 showing for our example belief 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 2.5 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 2.6 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 2.1 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.

2.7 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 chapter, 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 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 chapter, we have briefly presented several properties that allow for reducing the computational burden involved to at least some extent. In Chapter 4, we have detailed these and additional properties that will allow for practical use of sensitivity analysis. Building on 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 (Chapter 3) 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.

Chapter 3

Using sensitivity analysis for efficient quantification of a belief network

Abstract

Sensitivity analysis is a method to investigate the effects of varying a model's parameters on its predictions. It was previously suggested as a suitable means to facilitate quantifying the joint probability distribution of a Bayesian belief network. This chapter presents practical experience with performing sensitivity analyses on a belief network in the field of medical prognosis and treatment planning. Three network quantifications with different levels of informedness were constructed. Two poorly-informed quantifications were improved by replacing the most influential parameters with the corresponding parameter estimates from the well-informed network quantification; these influential parameters were found by performing one-way sensitivity analyses. Subsequently, the results of the replacements were investigated by comparing network predictions. It was found that it may be sufficient to gather a limited number of highly-informed network parameters to obtain a satisfying network quantification. It is therefore concluded that sensitivity analysis can be used to improve the efficiency of quantifying a belief network.

3.1 Introduction

The framework of *Bayesian belief networks* was introduced in the late 1980s [Pearl, 1988] for reasoning with uncertainty in a mathematically correct manner. It owes much of its popularity to the use of probability theory combined with an appealing graphical representation of conditional independence relations. As such, belief networks allow for explicit and declarative modelling of a problem domain, capturing domain knowl-

edge that is relevant for solving knowledge-intensive problems [Andreassen *et al.*, 1987, Heckerman & Nathwani, 1992].

In building a belief network, two closely related tasks can be discerned; the construction of the graphical part of the network, and its subsequent quantification. Building the graphical part of a network consists of identifying variables in the domain under study, and assessing the conditional independence relations that exist between these variables; these relations are represented by an acyclic, directed graph. The quantification of a belief network amounts to assessing a local conditional probability distribution for each variable in the network. These local distributions uniquely define a joint probability distribution on the variables discerned, that respects the independence relations portrayed by the graph. A variable's distribution is conditioned on its parents in the graph; the number of parameters (conditional probabilities) that need to be assessed for the distribution grows exponentially in the number of parents. The total number of parameters that is needed to quantify a belief network may therefore be considerable. Furthermore, a substantial number of network variables may be hidden from direct observation; it is then very difficult, if not impossible, to collect quantitative data on these variables. For these reasons, quantification of a belief network is often considered a hard task.

To facilitate the quantification of belief networks several methods have been proposed in the literature [Jensen, 1995, Monti & Carenini, 1995, Druzdzel & Van der Gaag, 1995]. In the previous chapter, we have described how *sensitivity analysis* can be used to reduce the quantification effort. Sensitivity analysis is a method to investigate the effects of varying a model's parameters on its predictions. For a belief network, it can reveal which parameters have a large effect on posterior probabilities, and, therefore, on which parameters the quantification effort should be focused. In Chapter 2, we propose a procedure of iteratively performing sensitivity analyses of an initially roughly quantified network, in order to stepwise refine the quantification.

This chapter presents an empirical investigation regarding the viability of this procedure. As a case study, we selected a belief network that describes the pathophysiology of ventricular septal defect (VSD), a frequently occurring congenital cardiac anomaly. It was developed as part of a larger decision-theoretic application for treatment planning and prognosis in the field of paediatric cardiology [Peek & Ottenkamp, 1997]. For the quantification, we have obtained subjective probability estimates. The use of subjective probabilities is indispensable in domains where there is a shortage of clinical data and many variables cannot be measured. Unfortunately, this may require a massive amount of probabilities to be estimated by field experts, a difficult and time-consuming task. The main objective of our research is to establish whether it is possible to reduce the number of parameters that have to be estimated by field experts. That is, we want to obtain a network quantification that gives predictions comparable to an expert-quantified network, without having to elicit all the network's parameters from

the expert.

The following procedure was used in the investigation. Three network quantifications were obtained, differing in the level of *informedness* of the estimates. The term informedness refers to the knowledge about the problem domain of the person supplying the estimates; we assume that accuracy of network predictions increases with the level of informedness. Extensive sensitivity analyses were performed on all three network quantifications, yielding a set of most influential parameters. In the two least informed quantifications, the estimates for these influential parameters were replaced, stepwise, with the estimates in the quantification of the field expert. The predictions of the resulting improved quantifications are compared with the predictions of a network that was completely quantified by the field expert.

Our results show that the procedure contributes to efficient quantification of a belief network: if, in a poorly-informed quantification, a limited number of highly influential parameters are replaced by more precise estimates, then the network gives predictions that are comparable to the network that is completely quantified with precise estimates. This means that we can avoid lengthy elicitation procedures, and focus the quantification effort on parameters to which network predictions are found to be most sensitive when varied.

The chapter is organised as follows. In Section 3.2, we briefly discuss the problem of treatment planning for patients with a ventricular septal defect, and present the qualitative part of the VSD network. Section 3.3 gives formal backgrounds of sensitivity analysis in belief networks, and describes the method of investigation to test the refinement procedure. Then, in Section 3.4, the results of the sensitivity analyses and subsequent refinements of the network quantification are presented. Discussion and conclusions are given in Section 3.5.

3.2 The VSD network

Ventricular septum defect (VSD) is the most frequently occurring congenital heart disease; approximately 2 to 3 out of each 1000 infants is born with this cardiac anomaly. It is a relatively well-understood disorder with many clinical features that are characteristic for congenital heart disease in general. We are currently developing an application to support the management of VSD patients in clinical practice, based on recently developed techniques from uncertainty reasoning and decision theory. Our aim is to deliver a 'white-box' system, in which the user can perceive what is going on, and can interact by proposing alternatives or adjust admissible treatment plans [Peek, 1999]. The core of the system is formed by a sequence of Bayesian belief networks that model VSD pathophysiology and its clinical findings at different stages of infant development. Heart, lungs and vessels are subject to a number of changes during the first years of

life, which make it impractical to use the same belief network for each development stage. Instead, we chose to employ different belief networks for different stages; there is, of course, substantial overlap between these networks.

In this chapter, we focus on the belief network that models VSD pathophysiology for patients aged 3 to 6 months. After a VSD has been diagnosed (usually in the first weeks of life), the patient is monitored during the subsequent months. The age of 3 to 6 months is crucial from a clinical perspective, as it provides the first opportunity to establish the severity of disease. It is therefore the point in time where the clinician will want to assess a preliminary prognosis of the patient's further development, and may already want to decide upon the treatment plan that is to be followed.

This section describes the *qualitative part* of the belief network for VSD patients aged 3–6 months. We take this part to comprise the directed graph that models conditional independence relations between domain variables, and various types of constraint on the probability distribution modelled by the network. Before we elaborate on the network itself, we first briefly review the domain under consideration.

3.2.1 VSD

A VSD is an abnormal opening in the ventricular septum, the fibromuscular wall that separates the heart's two ventricles [Graham & Gutgesell, 1995]. The main pathophysiological consequence of the presence of a VSD is blood flow ("shunt") from the left to the right ventricle due to ventricular pressure differences. Left-to-right ventricular shunting renders typical murmurs that can be heard by auscultation of the heart, and abnormal vibrations of the heart (called a "thrill") that can be felt at the chest. The *shunt size*, i.e., the amount of blood flowing through the defect, depends primarily on the size of the defect and the pulmonary vascular resistance. The consequence of shunting is that oxygenous blood is recirculated through the lungs. As a result, pulmonary vascular pressure will rise, and systemic cardiac output will decrease. With large defects, the large shunt size and high pulmonary arterial pressure may lead to *heart failure*: the heart is unable to adequately fulfil its primary function, the circulation of blood through the body. Heart failure accounts for most of the typical symptoms associated with VSDs, such as shortness of breath, feeding problems, oedema, and growth arrearage. Severe heart failure may result in cardiomegaly (enlarged heart), hepatomegaly (enlarged liver), and pulmonary infections.

About 70% of all VSDs close spontaneously by normal tissue growth, [Krovetz, 1998], where small defects are more likely to close spontaneously than large ones. This development may take several months or even years, but it precludes the need for surgical intervention. Unfortunately, continual pulmonary overflow and hypertension may cause severe, irreversible damage to the pulmonary arterioles; this is termed *Eisenmenger's reaction*, and represents the primary risk to VSD patients. Eisenmenger's reaction is

detected at early stages by considering the ratio of pulmonary and systemic vascular resistances; increasing pulmonary vascular resistance is indicative for the reaction. However, there exist no means to measure this resistance in clinical practice; it can only be estimated from related signs. For this reason, early surgical intervention is recommended for patients with large VSDs that are unlikely to close spontaneously. The majority of patients with timely repair of uncomplicated VSDs in infancy or early childhood have an excellent result with no clinical signs or symptoms and apparently normal life-expectancy [Moller *et al.*, 1991].

For the clinician, the main problem is to decide if and when to submit a patient to surgery. Usually, the patient's condition is monitored without surgical intervention during the first year of life. During this period, non-invasive diagnostic tests such as auscultation and echocardiography are conducted repeatedly to gain insight into the shunt size and the pulmonary vascular resistance. When necessary, medical treatment is given to reduce heart failure. After the first year of life, the risks associated with surgical intervention have dropped, and a decision whether surgery is necessary is made. In cases of doubt concerning the state of the pulmonary arterioles, cardiac catheterisation or pulmonary biopsy may be performed prior to that decision to obtain more information on the severity of disease. Therapy is considered completed after closure of the defect, either spontaneously or by surgical intervention.

3.2.2 Qualitative part of the VSD network

A Bayesian belief network represents a joint probability distribution over a set of discrete, stochastic variables. Each node in the graph represents one of the domain variables, and the absence of an arc between two nodes means that there exists no direct dependency between the variables represented by these nodes; the variables are only dependent via intermediate nodes in the graph. In practice, the most commonly used heuristic to assess a network's structure is the concept of *causality*: an arc is drawn between two nodes if there is known to exist a direct, causal relation between the corresponding variables. This often provides a sound overall representation of probabilistic dependencies in the problem domain.

For the VSD application, a network structure was hand-crafted with the aid of a field expert, a senior paediatric cardiologist. First, a set of variables jointly describing the VSD domain was selected, and then the graphical part of the belief network was assessed, using causality as the principle modelling heuristic. For a more elaborate description of this development process, we refer to [Peek & Ottenkamp, 1997]. The resulting network is depicted in Figure 3.1.

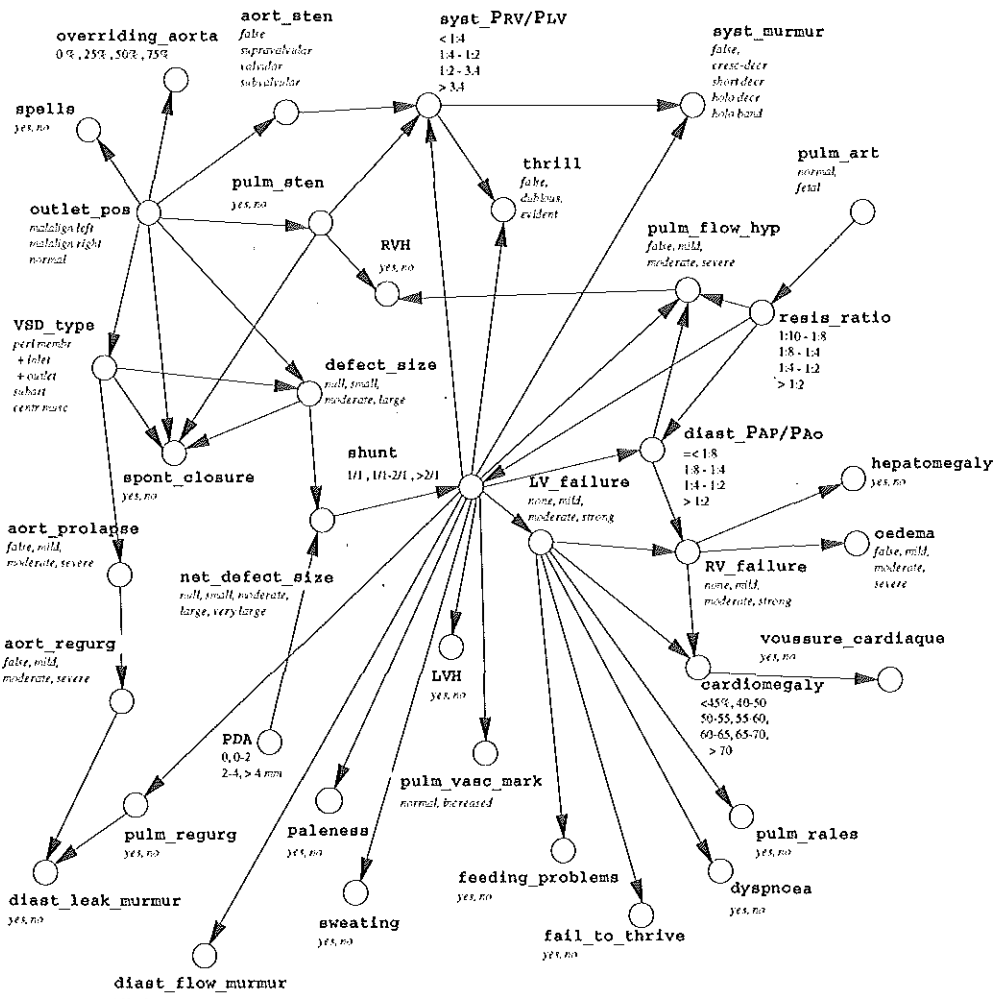


Figure 3.1: The VSD network.

Once the structure of a belief network has been established, the next task is to estimate the parameters for the quantitative part of the network; this part encodes a joint probability distribution over the network's variables. For the VSD network, which consists of 38 variables, 1298 parameters are needed to fully quantify the probability distribution. A significant number of variables cannot be observed in clinical practice, making it very difficult to obtain estimates of the required parameters. In order to alleviate the quantification burden for the VSD network, we decided to collect three types of qualitative information on the probability distributions prior to the precise estimation of parameters:

1. *functional dependencies*, modelling deterministic relations between variables,
2. *consistency constraints*, describing which combinations of values cannot occur, and
3. *qualitative influences*, expressing the sign of probabilistic interactions.

Each of these information types can be interpreted as expressing a constraint on one of the local probability distributions in the network [Druzdzel & Van der Gaag, 1995]. Below, we elaborate on each of them.

The first type of qualitative information consists of *functional dependencies*. These dependencies effectively model deterministic relations, as they express that a variable will necessarily take a certain value when its conditioning variables (i.e., its parent nodes in the graph) have a particular configuration of values. For example, in the VSD domain, we know that if there is severe aortic regurgitation (i.e., the aortic valve is very leaky), then cardiac auscultation will definitely reveal a typical kind of murmur (called a *leak murmur*) during the diastole, the phase of relaxation in the cardiac cycle. Formally, we have that

$$\mathrm{p}(\texttt{diast\_leak\_murmur} = \textit{yes} \mid \texttt{aort\_regurg} = \textit{severe}, \texttt{pulm\_regurg} = x) \;=\; 1,$$

for any legal value of x for `pulm_regurg`, the second conditioning variable of `diast_leak_murmur`. Furthermore, the probabilities of other values for `diast_leak_murmur` (in this case only the value *no*) are 0 under these conditions. For the VSD network, 8 variables were completely described by functional dependencies; 10 additional variables were partially described by them. These dependencies, provided by the field expert, were all assumed to be reliable.

The second type of qualitative information on probability distributions consists of *consistency constraints*. These constraints exclude certain combinations of values that cannot possibly occur in practice. For instance, once the pulmonary arterioles have reached their normal state (6 to 12 weeks after birth), then the pulmonary vascular resistance cannot exceed 1/4 of the systemic vascular resistance. Formally:

$$\mathrm{p}(\texttt{resis\_ratio} > 1:4 \mid \texttt{pulm\_art} = \textit{normal}) \;=\; 0. \tag{3.1}$$

For the VSD network, 102 consistency constraints were found, pertaining to 10 different variables. Overall, 560 parameters were determined by a functional dependency or a consistency constraint, reducing the number of parameters that need estimation to 738.

The third and last type of qualitative information on probability distributions consists of *qualitative influences* [Wellman, 1990] A qualitative influence is a symmetric property describing the sign of probabilistic interaction between two variables, building on orderings of these variables' value domains. A positive (negative) qualitative influence, indicated by attaching the label '+' ('−') to an arc in the graph, expresses that higher values of the one variable makes higher (lower) values of the other more likely, and vice versa. For instance, if the amount of blood that flows through the VSD (the shunt size) grows, then increasing failure of the left ventricle becomes more likely. This is expressed by a positive qualitative influence between the variables `shunt` and `LV_failure`:

$$\texttt{shunt} \xrightarrow{+} \texttt{LV\_failure} \tag{3.2}$$

which induces the following inequalities:

$$\begin{aligned}
&\mathrm{p}(\texttt{LV_failure} = \textit{none} \mid \texttt{shunt} \geq 2:1) \quad \leq \\
&\qquad \mathrm{p}(\texttt{LV_failure} = \textit{none} \mid 1:1 < \texttt{shunt} < 2:1) \\
&\mathrm{p}(\texttt{LV_failure} = \textit{moderate} \mid \texttt{shunt} \geq 2:1) \\
&+\mathrm{p}(\texttt{LV_failure} = \textit{severe} \mid \texttt{shunt} \geq 2:1) \quad \geq \\
&\qquad \mathrm{p}(\texttt{LV_failure} = \textit{moderate} \mid 1:1 < \texttt{shunt} < 2:1) \\
&\qquad +\mathrm{p}(\texttt{LV_failure} = \textit{severe} \mid 1:1 < \texttt{shunt} < 2:1) \\
&\mathrm{p}(\texttt{LV_failure} = \textit{severe} \mid \texttt{shunt} \geq 2:1) \quad \geq \\
&\qquad \mathrm{p}(\texttt{LV_failure} = \textit{severe} \mid 1:1 < \texttt{shunt} < 2:1) \ .
\end{aligned} \tag{3.3}$$

These inequalities constrain the local probability distribution for the variable `LV_failure`. In co-operation with the field expert, a total of 24 positive and 5 negative qualitative influences was found for the VSD network.

3.3 Quantification and sensitivity analysis

When assessing the quantitative part of a belief network, numerous network parameters have to be estimated, either from frequencies found in statistical data, or subjectively by experts in the field of application. But often such statistical data are difficult, if not impossible to obtain, and gathering estimates from field experts is very time-consuming. In this chapter, we experimentally assess the viability of a procedure to facilitate the quantification task, proposed in Chapter 2. This procedure, which is based on performing sensitivity analyses, is reviewed in Section 3.3.1. For the experimental investigation, where the VSD network is used as a case study, we have collected three

quantifications that differ with respect to informedness of the estimates; these are described in Section 3.3.2. Furthermore, two variables in the network that are indicative for its performance, and five case profiles, describing typical patterns of observations, have been identified; these are described in Section 3.3.3.

3.3.1 One-way sensitivity analysis of a belief network

Sensitivity analysis is a technique to systematically study the effects of variations in the parameters of a mathematical model on this model's predictions. It is widely used in the fields of decision theory and mathematical modelling, [Habbema *et al.*, 1990, Morgan & Henrion, 1990, Von Winterfeldt & Edwards, 1986]. For a Bayesian belief network, sensitivity analysis provides for studying the effects of variations in the estimates of the network's parameters on one or more posterior probabilities of interest. As such, sensitivity analysis allows for identifying network parameters that are highly influential, and should therefore be estimated with the highest accuracy. For less influential parameters, rough estimates may suffice. Sensitivity analysis can thus be used to increase the efficiency of quantifying a belief network, as it directs the quantification effort towards crucial parameters.

The simplest type of sensitivity analysis is a *one-way sensitivity analysis.* In a one-way sensitivity analysis of a belief network, the estimates of the network's parameters are varied one at a time, keeping all others fixed. The analysis then reveals the separate effect of variation of a parameter estimate on posterior probabilities. In this investigation, we used the method for one-way sensitivity analysis proposed in [Coupé & Van der Gaag, 1998] (see also Chapter 4).

Coupé et al. show that in a sensitivity analysis of a belief network, it is not necessary to vary all network parameters, given a particular posterior probability of interest. Only a subset of parameters will influence the posterior probability; this subset can be derived solely from the graphical structure of the network. We will refer to the *sensitivity set* as the set of variables whose parameters may be influential; the constitution of this set depends on the evidence entered into the network and the posterior probability one is interested in. For details concerning the identification of the sensitivity set, we refer Chapter 4.

Furthermore, Coupé et al. show that there exist functional relationships between individual parameters and posterior probabilities in a belief network. Any posterior probability is a rational polynomial over the parameter under study:

$$\Pr^{\theta=x}(V = v \mid \xi) \;=\; \frac{a \cdot x + b}{x + c} \tag{3.4}$$

where θ is the parameter under study, and a, b, and c are real-valued constants. $\Pr^{\theta=x}(V = v \mid \xi)$ is the posterior probability of the value v for the variable V given evidence ξ. We refer to the right hand side of Eq. 3.4 as a *sensitivity function.* It is easily

seen that systematic variation of the parameter under study is not necessary to determine its associated sensitivity function: if we compute the posterior $\Pr^{\theta=x}(V = v \mid \xi)$ with three different values for θ, the constants in the functional relationship can be determined. These constants are now used to calculate the first order derivative of the sensitivity function:

$$\frac{d}{dx}\Pr^{\theta=x}(V = v \mid \xi) = \frac{a \cdot c - b}{(x + c)^2} \tag{3.5}$$

If we apply this derivative function to the original parameter estimate, we obtain the gradient of the sensitivity function at that point. This quantity gives an impression of the influence of (small) variations in the estimate on the posterior $\Pr(V = v \mid \xi)$; see Figure 3.2 for illustrations.

As the influences of parameter estimates on posterior probabilities may vary with the evidence ξ, sensitivity analyses should be performed for several evidence sets. We will refer to these sets as *case profiles*; they preferably consist of realistic patterns of observations. Furthermore, it should be established which posterior probabilities are indicative for the performance of the belief network. The variables to which these probabilities pertain will be called the *variables of interest*. Both realistic case profiles and variables of interest depend on the envisioned application of the belief network under consideration.

To facilitate the quantification of a belief network, Coupé et al. now propose the following two-stage procedure. After the graphical part of the network has been assessed, a rough quantification is established. Such a rough quantification can be based, for instance, on a small collection of statistical data, or order-of-magnitude estimates derived from qualitative descriptions of the relations involved. The second step consists of performing one-way sensitivity analyses on the network; this requires the identification of several realistic case profiles, and one or more variables of interest. Finally, those parameter estimates that turn out to be highly influential are refined. Improved estimates are obtained, for instance, by gathering more statistical data on the variables involved, or by eliciting them from experts in the field. The effort of obtaining highly accurate parameter estimates is thus limited to a subset of network parameters.

We conclude this section by noting that the refinement of influential parameters in a given quantification will generally not reduce the sensitivity of posterior probabilities to parameter variation. A quantification solely comprising highly accurate parameter estimates will often contain just as much influential parameters as a completely random quantification: there exists no relation between sensitivity and 'quality' of a quantification.

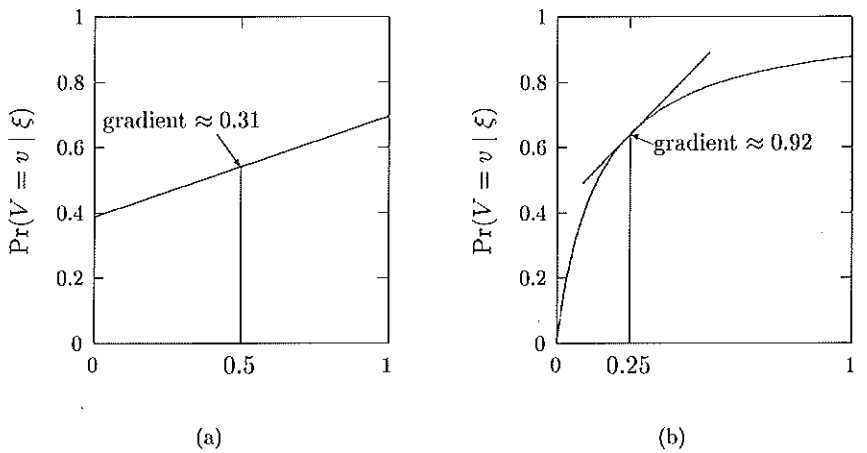


Figure 3.2: Two sample sensitivity functions. (a) A linear sensitivity function (i.e., $c = 0$). The gradient is independent of the actual point estimate. (b) A nonlinear sensitivity function. The gradient of the function is computed for the point estimate (0.25), providing an indication of the influence of small variations in the parameter estimate on the posterior probability.

3.3.2 Three quantifications of the VSD network

The main question now is: how compares a network quantification obtained by following the procedure described above to a network quantification that consists completely of accurate parameter estimates? Furthermore, what is the *efficiency gain*, in terms of quantification effort, yielded by the procedure? This chapter presents an experimental investigation into these issues, using the VSD network as a case study.

To be able to answer the questions above, we have acquired three different quantifications of the VSD network. A total of 738 parameters had to be established. The following network quantifications were obtained, listed in order of increasing informedness:

Q1, consisting of completely uninformative parameter estimates,

Q2, consisting of parameter estimates supplied by a non-expert, on the basis of qualitative characterisations of the uncertain relations, and

Q3, consisting of parameter estimates supplied by a field expert.

In network quantification Q1, a uniform probability distribution was used for each variable in the network. The assessed deterministic relations, as described in Section 3.2.2, were however preserved in this otherwise uninformative network quantification.

For quantification Q2, the parameter estimates were provided by a non-medical researcher who was involved in the construction of the graphical part of the network. For this quantification, the information on qualitative probabilistic influences between the variables in the network was used. Where possible, a linear model was assumed for the dependency between the parameter estimates for a variable and the values of the parents of that variable. That is, the difference in probability estimates for successive values of the conditioning parents is taken to be equal. In estimating the required prior distributions for variables without ascendants in the graph, both medical literature and qualitative statements of the field expert were used. For the estimated occurrence of different types of VSD, incidence figures found in the literature were used; for the various complications of disease, statements such as "common", "rare", "very rare", given by the field expert during the elicitation of the qualitative part of the network, were translated to probability estimates. Furthermore, as in quantification Q1, the deterministic relations were also ensured in quantification Q2. In total, seven hours were spent on establishing this quantification of the VSD network.

For the network quantification Q3, parameter estimates were provided by a senior paediatric cardiologist. For each distribution, the cardiologist was asked to provide the expected number of patients out of a hundred with a specific value for the variable under consideration, given a configuration of its parents in the graph. Initially, the clinician felt reluctant to give such precise numbers; he was therefore asked to provide 95% and 50% confidence intervals in addition to the point estimates. As the confidence intervals allowed him to express his own uncertainty regarding the estimates, he felt more comfortable with this procedure. For a more detailed description of the elicitation procedure followed, we refer the reader to [Coupé *et al.*, 2000]. The total amount of time the cardiologist spent on the quantification of the network was approximately twenty-five hours.

Our objective now was to assess whether it is possible to improve quantifications Q1 and Q2 up to the level of Q3, where the improvements consist of selective revisions of influential parameters. These influential parameters are found by performing one-way sensitivity analyses of the network.

3.3.3 The variables of interest and case profiles under consideration

Two variables of interest were indicated by the field expert to be used in the sensitivity analyses: the variable `shunt` and the variable `resis_ratio`. These are the two most important variables in the network, in the sense that the patient's prognosis largely depends on their values; a clinician usually bases his management decisions on estimates of these variables. We have therefore assumed that the performance of the VSD network can be measured by testing the accuracy of predictions for these variables. For the

Evidence	profile 1	profile 2	profile 3	profile 4	profile 5
syst_murmur	*no*	*short_decr*	*holo_band*	*holo_band*	*holo_band*
thrill	*none*	*none*	*none*	—	*evident*
diast_flow_murmur	*no*	*no*	*no*	*yes*	*yes*
paleness	*yes*	*no*	—	*no*	*yes*
sweating	*yes*	*no*	—	*no*	*yes*
hepatomegaly	*no*	*no*	—	—	*yes*
dyspnoea	*yes*	*no*	*no*	*no*	*yes*
feeding_problems	*yes*	*no*	*no*	*no*	*yes*
fail_to_thrive	*yes*	*no*	—	*no*	*yes*

Table 3.1: The case profiles for the VSD network. A maximum of nine observations is available for every case profile; '—' means that the value for that specific variable is unknown. The profiles are ordered according to the severity of disease. From left to right the VSD size increases and, therefore, also the likelihood of an unfavourable outcome.

variable shunt, we have focused on the value shunt $\geq 2:1$, corresponding to a strongly increased pulmonary blood flow. Continual pulmonary overflow increases the risk of damage to the pulmonary circulation. For the variable resis_ratio, taking one of the four values 1:10–1:8, 1:8–1:4, 1:4–1:2, and >1:2, the value 1:10–1:8 was taken as the value of interest. This value reflects normal vascular resistance in the pulmonary circulation and therefore corresponds to a favourable situation; the remaining values for resis_ratio correspond to increasing pulmonary damage.

In analysing the network quantifications, we thus focused on the posterior probabilities

- $\Pr(\text{shunt} \geq 2:1 \mid \textit{profile}_i)$ and
- $\Pr(1:10 \leq \text{resis\_ratio} \leq 1:8 \mid \textit{profile}_i)$,

and their respective sensitivity to parameter variations, where $\textit{profile}_i$ stands for a particular case profile (i.e., a set of observations). Five case profiles were selected in cooperation with the field expert; they are shown in Table 3.1. Except for profile 1, each of these profiles reflects a realistic pattern of observations found in a particular type of VSD patient. Some profiles provide strong, unequivocal evidence towards certain predictions and pose no interpretation problems to clinicians in the field. For instance, profile 2 represents a patient with few, yet compelling findings, providing strong evidence for a small VSD. In profile 5, all symptoms related to a VSD are present, providing strong evidence for a large defect. For other profiles, the prognosis is more difficult to assess, and therefore more uncertain. Profiles 3 and 4 are less evident than profiles 2 and 5, and even somewhat contradictory. Typical signs of a

VSD are absent, but still a holosystolic murmur is audible, which is symptomatic for the disease. In profile 4, moreover, a diastolic flow murmur is present, increasing the evidence for a large VSD. The profile that is listed first, finally, shows many symptoms common for a VSD patient, but necessary findings such as systolic murmur and thrill are absent; this profile corresponds to a patient not having a VSD, but some other, unknown disease.

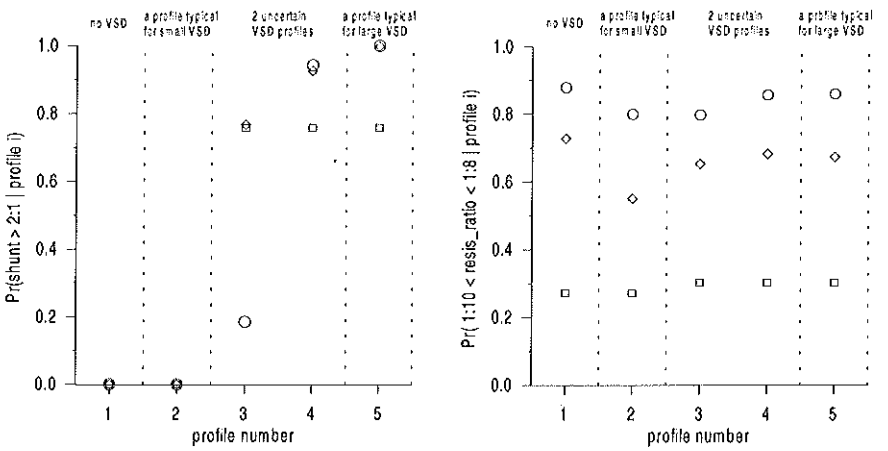
3.4 Results

For each of the network quantifications Q1, Q2, and Q3, the sensitivity of the selected posterior probabilities to variations in the network parameters was analysed. Subsequently, quantifications Q1 and Q2 were refined by replacing the probability distributions of the most influential variables with distributions from quantification Q3. This section discusses the results of the sensitivity analyses, and the effects of refining network quantifications. First, in Section 3.4.1, we compare the predictions of the three network quantifications for the five case profiles. The results of the sensitivity analyses are presented in Section 3.4.2. In Section 3.4.3, we detail the refinement procedure, and compare the predictions of refined network quantifications to the predictions of quantification Q3.

3.4.1 Predictions of the three network quantifications

For each of the network quantifications and each of the case profiles, the predictions for the variables `shunt` and `resis_ratio` were computed; they are shown in Figures 3.3a and 3.3b, respectively.

First, consider Figure 3.3a, showing the results for the variable `shunt`. We see that all quantifications assign zero probability to a large shunt, given profiles 1 and 2. This is due to consistency constraints encoded in each of the quantifications: the lack of loud heart murmurs and thrill precludes existence of a large shunt. These profiles therefore provide no basis for comparison here. For the other profiles, we find that the quantifications Q1 through Q3 have increasingly more discriminative power; this is in line with the increasing level of informedness of the quantifications. Quantification Q1 gives the same prediction for each of these profiles, due to the uniform distributions used in this quantification. The predictions of quantifications Q2 and Q3 are more pronounced and they do provide the same ranking of profiles. For profiles 4 and 5, the predictions from Q2 and Q3 agree well, but for profile 3, however, a large difference is seen in the predictions. This corresponds well to the fact that profile 3 provides contradicting observations and is therefore hard to interpret. The field expert indicated, however, that quantification Q3's prediction was best in line with his intuition for this profile.



(a) (b)

Figure 3.3: Predictions for the posterior probabilities (a) Pr(**shunt** $\geq 2:1 \mid \mathit{profile}_i$) and (b) Pr($1:10 \leq$ **resis_ratio** $\leq 1:8 \mid \mathit{profile}_i$) for each of the case profiles, computed from the three network quantifications (□ for Q1, ◊ for Q2 and ○ for Q3). The vertical dotted lines separate VSD profiles with typical characteristics.

Turning to the predictions for the variable **resis_ratio**, shown in Figure 3.3b, we see that for each of the quantifications the predictions remain more or less constant over the case profiles. However, the average level of the posteriors differs considerably per quantification: whereas quantification Q1 assigns a low probability of normal pulmonary vascular resistance, quantification Q3 is fairly confident about this circumstance; quantification Q2 is located in between. These differences can be traced back to prior (unconditional) distributions for variables without ascendants in the graph, which are quite different for the three quantifications. In discussing these results, the field expert again confirmed the predictions of his own quantification (Q3): at the age of three to six months, VSD patients usually have a normal pulmonary vascular resistance, regardless of the severity of disease.

3.4.2 Results of the sensitivity analyses

A one-way sensitivity analysis of both posterior probabilities Pr(**shunt** $\geq 2:1 \mid \mathit{profile}_i$) and Pr($1:10 \leq$ **resis_ratio** $\leq 1:8 \mid \mathit{profile}_i$) was performed for every case profile, for each of the three quantifications. From the total of 1298 parameters in the network, 560 parameters that are determined by a functional relationship or a consistency constraint were excluded from the analysis. Table 3.2 shows the sensitivity of the posteriors given

	Profile	Q1	Q2	Q3
	1		insensitive	
shunt	2		insensitive	
=	3	0.01172 (87%)	0.01685 (84%)	0.01446 (83%)
$\geq 2:1$	4	0.01222 (83%)	0.00716 (83%)	0.00882 (85%)
	5	0.01256 (82%)	0.00071 (78%)	0.00000 (77%)
	1	0.00793 (70%)	0.01230 (80%)	0.02173 (80%)
resis_ratio	2	0.00638 (76%)	0.01058 (77%)	0.01292 (82%)
=	3	0.00483 (88%)	0.00738 (85%)	0.00857 (85%)
$1:10-1:8$	4	0.00485 (85%)	0.00726 (83%)	0.00704 (85%)
	5	0.00506 (79%)	0.00715 (79%)	0.00732 (78%)

Table 3.2: The sensitivities of the posterior probabilities $\Pr(\texttt{shunt} \geq 2:1 \mid \mathit{profile}_i)$ and $\Pr(1:10 \leq \texttt{resis\_ratio} \leq 1:8 \mid \mathit{profile}_i)$ to parameter variations, averaged out over the network parameters considered in the analysis. The percentage of uninfluential parameters is given between parentheses.

each case profile, averaged over the 738 parameters considered in the analysis. Between parentheses, the percentage of uninfluential network parameters is written. The high percentages illustrate that there is often a considerable number of network parameters inside the sensitivity set, that yet turn out to be uninfluential when varied. In Table 3.3 the maximum sensitivity found for each case profile and network quantification is listed.

	Profile	Q1	Q2	Q3
	1	0	0	0
shunt	2	0	0	0
=	3	−0.73372	−1.07992	1.51122
$\geq 2:1$	4	−0.73372	−0.38576	−2.68748
	5	0.36686	−0.38789	−0.00020
	1	0.48170	−0.55913	−5.19101
resis_ratio	2	0.48170	0.53568	0.95637
=	3	0.51299	0.49599	0.95812
$1:10-1:8$	4	0.51299	0.48988	−0.82572
	5	0.51299	0.50188	-0.82463

Table 3.3: The maximum sensitivities of the posterior probabilities $\Pr(\texttt{shunt} \geq 2:1 \mid \mathit{profile}_i)$ and $\Pr(1:10 \leq \texttt{resis\_ratio} \leq 1:8 \mid \mathit{profile}_i)$.

First, consider the results for the variable shunt. We recall from the previous section that for profiles 1 and 2, the posterior $\Pr(\text{shunt} \geq 2 : 1 \mid \textit{profile}_i)$ is determined to be zero by consistency constraints. For this reason, varying parameters estimates will not affect the posterior in any of the network quantifications; it is completely insensitive, given these profiles. We therefore restrict the discussion to profiles 3, 4 and 5. We find that the predictions for shunt in quantifications Q2 and Q3 are significantly more sensitive to parameter variation for profiles 3 and 4 than for profile 5. An explanation for this pattern exists in the fact that profile 5 provides several independent pieces of evidence indicating a large shunt; varying individual parameters therefore hardly influences that prediction. In contrast, profiles 3 and 4 comprise contradicting observations: although heart murmurs indicating a large VSD are observed, none of the symptoms that would then be expected are present. In these cases, varying a single parameter can change the prediction for the shunt variable considerably. The result is not found for the uninformed quantification Q1: as this quantification gives the same prediction for each case profile, these predictions are almost equally sensitive to parameter variation.

For each case profile and each quantification, we identified the thirty parameters showing the largest influence on the posterior $\Pr(\text{shunt} \geq 2 : 1 \mid \textit{profile}_i)$. The variables to which they pertain are listed in Table 3.4, in order of decreasing maximum influence of their parameters. We note that there exists a substantial overlap in the variables to which highly influential parameters pertain, and therefore the number of variables is much smaller than thirty. Furthermore, the selections of variables per profile are roughly the same for all quantifications. This indicates that the structure of the belief network considerably affects the sensitivity of posteriors to parameter variation; the quantification that is used in the analysis is of secondary importance.

For quantification Q1, only the prediction variable itself (shunt), and observed, direct descendants of this variable are selected.[1] This is not surprising, as the uniform distributions used in this quantification eliminate all influences through longer pathways in the graph when only one parameter estimate is varied at a time. Therefore, only higher-order sensitivity analyses can reveal the propagation of influences through the graph for this quantification. The selections for quantifications Q2 and Q3 are supersets of the selection for quantification Q1. Notably, they also comprise variables at a greater distance of the shunt variable, and ascendant variables of shunt in the graph, e.g., `defect_size` and `resis_ratio`. The distributions of these ascendant variables represent prevalences of the disease and its complications, and are therefore influential on the posterior distribution of shunt. We conclude that the selections for quantifications Q2 and Q3 provide more realistic patterns of influential variables.

[1] The variable `LV_failure` is functionally determined by its observed descendants and can therefore itself be regarded as observed.

	Q1	Q2	Q3
profile 3	LV_failure	syst_murmur	LV_failure
	diast_flow_murmur	syst_P_{RV}/P_{LV}	syst_P_{RV}/P_{LV}
	thrill	diast_flow_murmur	diast_flow_murmur
	syst_murmur	LV_failure	syst_murmur
	shunt	thrill	pulm_sten
		shunt	shunt
		resis_ratio	defect_size
			PDA
			outlet_pos
			thrill
profile 4	LV_failure	syst_P_{RV}/P_{LV}	diast_flow_murmur
	paleness	syst_murmur	LV_failure
	diast_flow_murmur	diast_flow_murmur	syst_P_{RV}/P_{LV}
	sweating	paleness	syst_murmur
	syst_murmur	sweating	pulm_sten
	shunt	LV_failure	shunt
		shunt	paleness
		outlet_pos	sweating
		pulm_sten	defect_size
		resis_ratio	PDA
			outlet_pos
profile 5	diast_flow_murmur	diast_P_{AP}/P_{Ao}	diast_flow_murmur
	paleness	thrill	LV_failure
	sweating	syst_P_{RV}/P_{LV}	sweating
	thrill	diast_flow_murmur	RV_failure
	syst_murmur	syst_murmur	thrill
	LV_failure	LV_failure	syst_murmur
	shunt	paleness	paleness
		sweating	VSD_type
		shunt	syst_P_{RV}/P_{LV}
		outlet_pos	pulm_sten
		pulm_sten	shunt
		resis_ratio	

Table 3.4: The variables to which the thirty network parameters pertain that are most influential to the posterior $\Pr(\text{shunt} \geq 2 : 1 \mid \textit{profile}_i)$, for profiles 3, 4, and 5. The variables are ordered with respect to the maximum influence of their parameters.

Consider now the results, in Tables 3.2 and 3.3, for the variable resis_ratio. For this variable, the differences in average and maximum sensitivity between the three quantifications and five case profiles are very small. This illustrates that there is no relation between the level of informedness of a network quantification and the sensitivity for variations in parameter estimates. Furthermore, there exists no distinction between the 'contradicting' profiles 3 and 4 and other profiles; this seems to be correct as these contradictions mainly concern the size of the shunt. At the age of three to six months, only minor differences between patients with regard to the resistance ratio are to be expected. For the posterior $\Pr(1:10 \leq \texttt{resis\_ratio} \leq 1:8 \mid \mathit{profile}_i)$, the variables pertaining to the thirty most influential parameters were also identified. Due to space limitations, these results are not shown here. However, similar observations hold as for the posterior $\Pr(\texttt{shunt} \geq 2:1 \mid \mathit{profile}_i)$. Again, the selections of influential variables per profile are roughly the same for each quantification. Every selection contains the variable resis_ratio itself, as well as its direct ascendant in the graph, the variable pulm_art. The direct descendants diast_P_{AP}/P_{Ao} and shunt also turn out to be very influential. Furthermore, the selected variables at greater distance from resis_ratio partially overlap with those variables selected for shunt.

3.4.3 Predictions of the refined network quantifications

Using the results of the sensitivity analyses, quantifications Q1 and Q2 were stepwise refined with parameter estimates from quantification Q3. Although sensitivity analysis reveals the influence of individual parameters, we chose to substitute, at every refinement step, all parameter estimates pertaining to a network variable. The motivation for this approach is that parameter estimates often have little meaning in isolation: it is their relation with other parameter estimates from the same local distribution that matters. To select variables whose parameters are eligible for substitution, the sets of influential variables per profile (Table 3.4) were compiled to a single, ordered set; the order was determined by averaging the positions of the variables in the original sets.

With respect to the shunt variable, the following sets were thus compiled:

$$V_{Q1}(\texttt{shunt}) = \{\texttt{diast\_flow\_murmur}, \texttt{LV\_failure}, \texttt{syst\_murmur}, \texttt{shunt}, \texttt{paleness}, \texttt{thrill}, \texttt{sweating}\} \quad (3.6)$$

$$V_{Q2}(\texttt{shunt}) = \{\texttt{syst\_P}_{\texttt{RV}}\texttt{/P}_{\texttt{LV}}, \texttt{syst\_murmur}, \texttt{diast\_flow\_murmur}, \texttt{LV\_failure}, \texttt{shunt}, \texttt{resis\_ratio}, \texttt{thrill}, \texttt{paleness}, \texttt{sweating}, \texttt{outlet\_pos}, \texttt{pulm\_sten}, \texttt{diast\_P}_{\texttt{AP}}\texttt{/P}_{\texttt{Ao}}\}. \quad (3.7)$$

With respect to the variable resis_ratio, we have

$$V_{Q1}(\texttt{resis\_ratio}) = \{\texttt{resis\_ratio}, \texttt{shunt}, \texttt{pulm\_art}, \texttt{diast\_P}_{\texttt{AP}}\texttt{/P}_{\texttt{Ao}}, \texttt{LV\_failure}, \texttt{PDA}\} \quad (3.8)$$

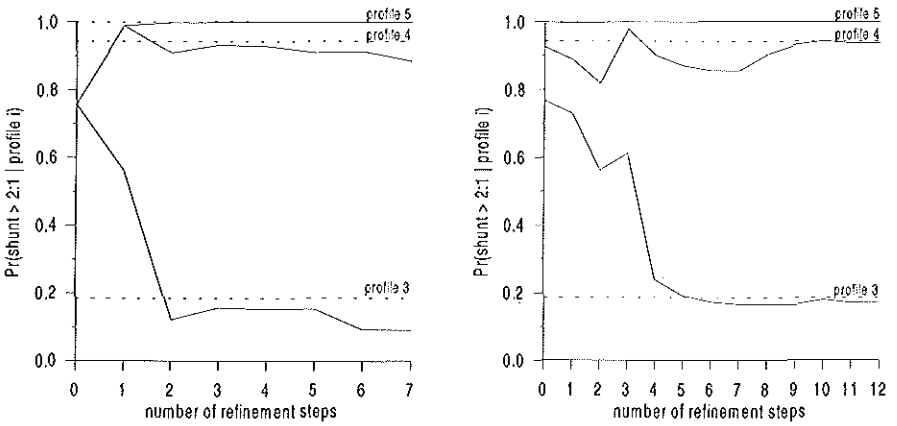


Figure 3.4: The predictions of (a) quantification Q1 and (b) quantification Q2 for the posterior probability $\Pr(\texttt{shunt} \geq 2:1 \mid \mathit{profile}_i)$, after successive refinement steps. The dotted horizontal lines correspond to the predictions of quantification Q3.

$$V_{Q2}(\texttt{resis\_ratio}) = \{\texttt{resis\_ratio}, \texttt{shunt}, \texttt{pulm\_art}, \texttt{LV\_failure}, \texttt{diast\_}P_{AP}/P_{Ao}, \texttt{syst\_}P_{RV}/P_{LV}, \texttt{syst\_murmur}, \texttt{RV\_failure}, \texttt{diast\_flow\_murmur}, \texttt{paleness}, \texttt{sweating}, \texttt{thrill}\} \quad (3.9)$$

The variables are listed in decreasing order of influence on the respective variables.

The quantifications were stepwise refined by replacing, at step i, all parameter estimates pertaining to the i^{th} variable in the above sets by the corresponding estimates from quantification Q3. So, to improve the prediction for shunt, a total of seven and twelve refinement steps were made for Q1 and Q2, respectively. For resis_ratio, a total of six and twelve refinement steps were made for Q1 and Q2. With each of the refined quantifications, the posteriors $\Pr(\texttt{shunt} \geq 2:1 \mid \mathit{profile}_i)$ and $\Pr(1:10 \leq \texttt{resis\_ratio} \leq 1:8 \mid \mathit{profile}_i)$ were computed for the various case profile; these posteriors are plotted in Figures 3.4 and 3.5.

First consider the refinements for the posterior probability $\Pr(\texttt{shunt} > 2:1 \mid \mathit{profile}_i)$. The plots indicate that for both quantifications, the posteriors rapidly shift towards the posteriors of quantification Q3. For quantification Q1, for example, two refinement steps suffice to approach Q3's predictions. For profile 5, even a single refinement of quantification Q1 (replacing the estimates of the variable diast_flow_murmur) is enough to nearly reach Q3's prediction. However, the difference between posteriors

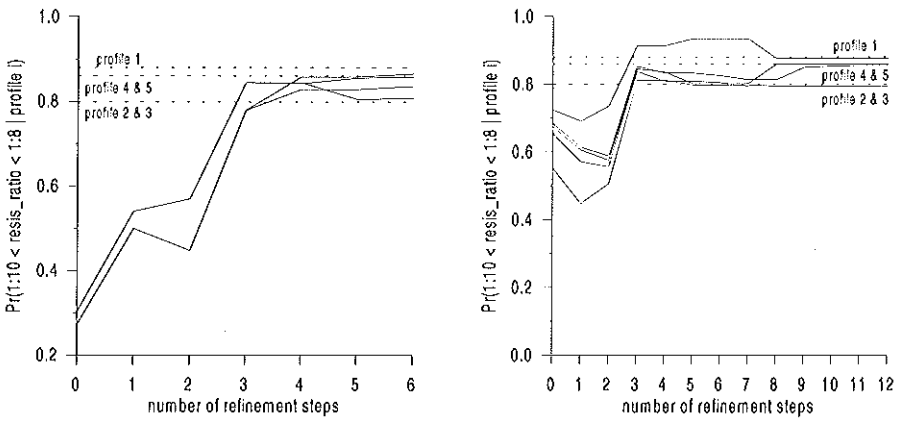


Figure 3.5: The predictions of (a) quantification Q1 and (b) quantification Q2 for the posterior probability $\Pr(1:10 \leq \texttt{resis\_ratio} \leq 1:8 \mid \mathit{profile}_i)$, after successive refinement steps. The dotted horizontal lines correspond to the predictions of quantification Q3.

does generally not decrease monotonically. This is most notable with the refinements of quantification Q2: the posteriors for profiles 3 and 4 show considerable fluctuations in the first seven steps. Thereafter, they quickly converge to the desired level. Unfortunately, this convergence is not obtained for quantification Q1: even after replacing the distributions of all seven variables with influential parameters, the posteriors still deviate from the posteriors of Q3.

Now, we turn to the refinements for the posterior $\Pr(1:10 \leq \texttt{resis\_ratio} \leq 1:8 \mid \mathit{profile}_i)$. As before, the posteriors quickly approach the posteriors of quantification Q3. For both quantification Q1 and Q2, three refinement steps suffice to reduce the difference with Q3's predictions considerably. After the maximum of six refinement steps, however, still no convergence is reached for Q1. For quantification Q2, the refined posteriors lie very close to the predictions for Q3 after the maximum of eight steps.

So far, we have seen that the results of refinements are encouraging. However, these results pertain to the five profiles that were also used in the sensitivity analyses. In order to investigate whether the results generalise over more cases, the effects of refinements were also tested on clinical data. Thirty-six cases were selected from a database of VSD patients collected at the Leiden University Medical Centre in The Netherlands. These cases correspond to patients aged 3–6 months having VSD as their

quantification	step	mean	max	min
Q1	0	0.2920	0.6396	0.1178
	2	0.1999	0.8381	0.0002
	4	0.1905	0.8080	0.0000
	6	0.0927	0.3859	0.0001
Q2	0	0.1313	0.5501	0.0042
	2	0.1497	0.5501	0.0022
	4	0.0686	0.3285	0.0000
	6	0.1032	0.3859	0.0001

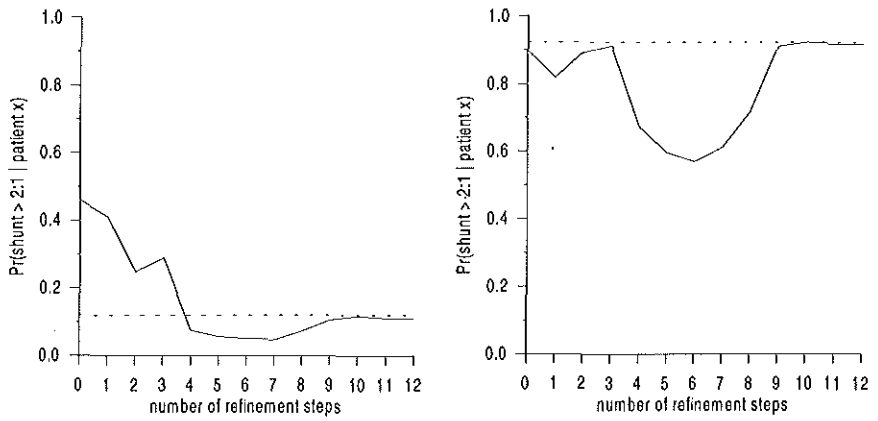
Table 3.5: The mean, maximum and minimum difference with Q3 of the refined predictions of Q1 and Q2 for the variable `shunt`, for all cases in the database whose predictions are not determined by a consistency constraint.

primary diagnosis. For each case, the predictions of quantifications Q1 and Q2 both before and after the refinements were compared with the predictions of quantification Q3. The average, maximum and minimum difference between the predictions from Q3 and both the original and refined predictions from Q1 and Q2 for the thirty-six cases were computed. In Tables 3.5 and 3.6, these differences for refinement with two, four and six variables, respectively, are shown.

For quantification Q1, the results show that stepwise refining this quantification indeed steadily reduces the differences with the predictions from quantification Q3. Note, that we do not claim that the predictions for Q3 are reliable. Since no reliable outcome measurements were available for these thirty-six patients no validation of the various quantifications was performed. Therefore, we only compare the refined quantifications with Q3 and aim to obtain a network quantification giving similar predictions as Q3, without using all parameter estimates from Q3.

quantification	step	mean	max	min
Q1	0	0.5186	0.6058	0.0616
	2	0.2999	0.3527	0.2252
	4	0.0412	0.0868	0.0083
	6	0.0303	0.0697	0.0002
Q2	0	0.3881	0.7789	0.0410
	2	0.3829	0.7664	0.0223
	4	0.4426	0.9113	0.0024
	6	0.4407	0.8992	0.0035

Table 3.6: The mean, maximum and minimum difference with Q3 of the refined predictions of Q1 and Q2 for the variable `resis_ratio`, for all cases in the database whose predictions are not determined by a consistency constraint.



(a) (b)

Figure 3.6: The effect of refining quantification Q2 for (a) a patient for whom the procedure is effective and (b) a patient for whom the procedure gives large fluctuations.

For quantification Q2, the results of the refinement procedure are less clear. For the variable `shunt`, the mean difference between Q3 and both Q2 and refinements of Q2 indeed show a decreasing trend. However, the original differences between the predictions from Q2 and Q3 are quite small, making drastic changes impossible. For `resis_ratio`, no significant effect of the refinements can be observed. This is an unsatisfying result; future research will have to uncover its causes.

As an example, in Figures 3.6a and b, the detailed results of the refinement procedure are given for two patients, one patient for whom Q3 predicts a low shunt and one patient for whom a high shunt is expected. Figures 3.6a illustrates that the refinement procedure may work very effectively. After four refinement steps, the predictions from Q2 have approached Q3 considerably. Convergence, however, is reached only after nine steps. Figures 3.6b shows a patient for whom the refinement procedure originally only worsens the predictions. This suggests that in applying the refinement procedure to quantify a network efficiently, it may be worthwhile to use real patient data for the sensitivity analyses.

3.5 Discussion and conclusions

Quantifying a Bayesian belief network is a difficult and time-consuming task, precluding easy application of belief-network technology in practice. However, it has been claimed that, once the graphical part of the network correctly models the independence relations

in the domain of application, then the behaviour of the network is insensitive to the quality of the majority of quantification parameters, [Pradhan *et al.*, 1996]. If this is true, then a satisfactory network quantification can be obtained by only estimating a small set of highly influential parameters from well-informed sources, and taking rough estimates for the others. We have presented an empirical investigation into this claim, by comparing the predictions of a well-informed quantification with poorly-informed quantifications, in which only influential parameters were reconsidered. These influential parameters were identified by performing one-way sensitivity analyses.

The results of our investigation suggest that using a two-step procedure of only reconsidering influential parameters works: partially refined, poorly-informed quantifications give predictions that are comparable to a well-informed quantification. The procedure could be repeated several times depending on the level of informedness of the quantification at hand. However, the procedure shows better results for the case profiles that were used to identify influential parameters, than for cases from a real-world, clinical database. We conclude that it is preferable to use a large set of cases when refining a network quantification. With respect to sensitivity analysis, it was found that the structure of the belief network considerably affects the influence of parameter variation to posterior probabilities; the quantification that is used in the analysis is of secondary importance. Furthermore, our results confirm that there is no relation between sensitivity of posteriors to parameter variation and the level of informedness of the quantification at hand.

Throughout the investigation, we have assumed that the structure of the belief network, as elicited from the field expert, is correct; the same assumption was made for functional relations and consistency constraints on network variables that were identified prior to quantification. We believe that these assumptions have had little or no impact on the results that were found. When building a real-world application, however, critical evaluation of these parts of the network model is also necessary: sensitivity analyses should not be restricted to numerical information. Furthermore, the re-quantifications concerned influential parameters that were found by performing one-way sensitivity analyses. These analyses measure the effects of individual parameter variations. Therefore, synergetic effects of varying multiple parameters are not detected, although they may have an important effect on the network's predictions. To reveal such effects, higher-order sensitivity analyses are required.

The method proposed here is not limited to elicitation of network parameters from field experts; it is also applicable to parameter elicitation from other sources such as clinical data sets or frequencies reported in the literature. In fact, we believe that the usage of objective statistical sources is indispensable to obtain a network quantification of sufficient quality. Subjective probability estimates are known to suffer from several forms of calibration and bias, [Kahneman *et al.*, 1982], and their reliability is therefore not beyond dispute. Moreover, in the medical field it is often possible to collect datasets

of reasonable size. The basic procedure investigated, however, applies equally well when a combination of quantification sources is employed.

To conclude, we believe that sensitivity analysis provides a promising addition to the methods that exist to facilitate belief network quantification. In future research, we plan to investigate more sophisticated procedures than the one described here. For instance, instead of making a final network quantification on the basis of a single sensitivity analysis, it is probably better to have a few alternating steps of sensitivity analyses and improvements of the quantification. This approach takes into consideration that by each refinement, the set of highly influential parameters is changed. Furthermore, it may be worthwhile to additionally perform higher-order sensitivity analyses, or even *uncertainty analyses*, which investigate the joint effect of varying all network parameters simultaneously. And finally, the expected accuracy of parameter estimates, as expressed by confidence intervals, can be involved in the analysis. This is accomplished by taking the variation of a sensitivity function over a confidence interval instead of over the whole range as a measure of the parameter's influence. Then, parameter estimates with high expected accuracy (i.e., having small confidence intervals) will only be reconsidered when the sensitivity function is extremely steep over the confidence interval.

Chapter 4

Properties of sensitivity analysis of Bayesian belief networks

Abstract

The assessments obtained for the various conditional probabilities of a Bayesian belief network inevitably are inaccurate. The inaccuracies involved influence the reliability of the network's output. By subjecting the belief network to a *sensitivity analysis* with respect to its conditional probabilities, the reliability of the output can be investigated. Unfortunately, straightforward sensitivity analysis of a Bayesian belief network is highly time-consuming. In this chapter, we show that, by qualitative considerations, several analyses can be identified as being uninformative as the conditional probabilities under study cannot affect the network's output. In addition, we show that the analyses that are informative comply with simple mathematical functions; more specifically, we show that the network's output can be expressed as a quotient of two functions that are linear in a conditional probability under study. The constants in this fractional function can be determined by solving the set of linear equations that results from only a small number of network computations. These properties allow for considerably reducing the computational burden of sensitivity analysis of Bayesian belief networks, as will be illustrated by means of various examples and experiments.

4.1 Introduction

During the last decades much effort in artificial-intelligence research has focused on modelling and reasoning with uncertainty in knowledge-based systems. As the oldest, well-founded mathematical theory of uncertainty, probability theory plays a prominent role in this research effort. Unfortunately, straightforward application of probability theory in a knowledge-based system leads to prohibitively high computational costs. Over the years, various attempts have been made to settle this problem, leading, in

the late 1980s, to the framework of *Bayesian belief networks*. Bayesian belief networks by now have become widely accepted as intuitively appealing probabilistic models that are highly valuable in addressing real-life problems in complex domains. Practical applications of the framework of belief networks are being developed for various problem domains, most notably in the field of medical diagnosis and prognostic assessment [Andreassen *et al.*, 1987, Heckerman & Nathwani, 1992].

A Bayesian belief network basically is a concise representation of a joint probability distribution on a set of statistical variables [Pearl, 1988]. It consists of a qualitative part and an associated quantitative part. The qualitative part of a belief network encodes, in a directed graph, the variables under study, along with their probabilistic interrelationships. The nodes in the digraph represent the statistical variables. The digraph's arcs with each other serve to capture the independences among these variables: absence of an arc between two nodes indicates that the corresponding variables do not influence each other directly and, hence, are (conditionally) independent. The quantitative part of the belief network is a set of conditional probabilities that describe the strengths of the dependences between the variables represented in the qualitative part: with each node are associated conditional probabilities describing the joint influence of values of the node's predecessors on the probabilities of the values of the node itself. A belief network's qualitative and quantitative part with each other provide enough information to uniquely define a joint probability distribution on the statistical variables under study. A Bayesian belief network thus allows for computing any (prior or posterior) probability of interest [Pearl, 1988].

Bayesian belief networks are generally constructed with the help of experts from the domain of application. Experience shows that, although it may require considerable effort, building the qualitative part of a belief network is quite practicable. In fact, as it has parallels to designing a domain model for a more traditional knowledge-based system, well-known knowledge-engineering techniques can be employed. Assessing the conditional probabilities for the quantitative part of a Bayesian belief network, however, is generally found to be a much harder task, not in the least because of the large number of assessments required [Druzdzel & Van der Gaag, 1995]. In general, various different sources of information can be exploited for probability assessment, ranging from databases and literature to human experts. The assessments obtained from these sources, however, are inevitably inaccurate, due to incompleteness of data and partial knowledge of the problem under study. Particularly assessments obtained from experts are known to be highly inaccurate [Kahneman *et al.*, 1982].

The inaccuracies in the probability assessments for a Bayesian belief network influence the reliability of the network's output. In a medical application, for example, erroneous diagnoses or non-optimal treatment recommendations may result from building upon inaccurate assessments. The reliability of the output of a belief network can be investigated by studying its robustness. Robustness pertains to the extent to which

the network's conditional probabilities influence the output when deviations from the specified assessments are assumed. For gaining detailed insight in output robustness, a Bayesian belief network can be subjected to a sensitivity analysis. In general, *sensitivity analysis* of a mathematical model amounts to investigating the effects of the inaccuracies in the model's parameters on its output; to this end, the values of the model's parameters are varied systematically [Morgan & Henrion, 1990, Habbema *et al.*, 1990]. For a belief network, sensitivity analysis amounts to varying the assessments for one or more conditional probabilities of the network's quantitative part simultaneously and investigating the effects on a probability of interest or, for example, on a diagnosis or decision based upon this probability of interest [Laskey, 1995] (see also Chapter 2). Upon such an analysis, some conditional probabilities will show a considerable effect, while others will hardly reveal any influence.

Straightforward sensitivity analysis of a Bayesian belief network, unfortunately, is highly time-consuming. In the simplest type of sensitivity analysis, for example, for every single conditional probability of the network's quantitative part, a number of deviations from the specified assessment are investigated. For every value under study, the probability of interest is computed from the network. Even for a rather small belief network, the analysis thus easily requires tens of thousands of network computations. By restricting the sensitivity analysis to the conditional probabilities that are expected to be influential, as indicated for example by a domain expert, the computational effort required can be reduced. The computational burden still remains considerable, however, and, in fact, is prohibitive when sensitivity analysis is to be used for verifying the robustness of a network's output in, for example, daily medical practice. To be of practical use, therefore, more efficient methods for sensitivity analysis of belief networks are indispensable.

In this chapter, we present an efficient method for sensitivity analysis of Bayesian belief networks that requires considerably less computational effort than straightforward variation of conditional probabilities. Our method builds to a large extent on the qualitative part of a belief network. As the digraph of a network represents the independences among the statistical variables involved, it allows for identifying conditional probabilities that upon variation cannot influence the probability of interest. Analyses with respect to these conditional probabilities are uninformative and can therefore be excluded from the overall analysis. Experiments on randomly generated belief networks indicate that the number of analyses that can be thus excluded may be considerable. In addition, we show that the analyses that are informative comply with simple mathematical functions. More in specific, we show that the probability of interest of a belief network can be expressed as a quotient of two functions that are linear in a conditional probability under study. The constants in this fractional function determine the sensitivity of the probability of interest to the conditional probability concerned. We show that computing the constants from the network requires

just a small number of network computations. These properties with each other allow for considerably reducing the computational burden and thus for improving upon the practicability of sensitivity analysis of Bayesian belief networks.

The chapter is organised as follows. In Section 4.2 we briefly review the framework of Bayesian belief networks and detail some of the concepts that will be used throughout the chapter. We then present the various properties of sensitivity analysis of belief networks outlined above. In doing so, we focus on a one-way sensitivity analysis, that is, an analysis in which a network's conditional probabilities are investigated one at a time. In Section 4.3, we discuss the identification of a belief network's conditional probabilities that upon variation cannot influence the probability of interest. In Section 4.4, we detail the functional relation that holds between a network's probability of interest and a single conditional probability under study. In Section 4.5, we comment on results obtained from experiments with one-way sensitivity analysis of randomly generated belief networks. In Section 4.6, we compare our results with previous work on sensitivity analysis of Bayesian belief networks. The chapter ends with our conclusions and directions for further research in Section 4.7.

4.2 The belief-network framework

A *Bayesian belief network* basically is a concise representation of a joint probability distribution on a set of statistical variables. In a belief network, information about the independences holding among the variables is explicitly separated from the numerical quantities involved in the distribution. To this end, the network comprises a qualitative part and an associated quantitative part. In this section, we briefly review the formalism of belief networks; for further details, we refer the reader to [Pearl, 1988].

The qualitative part of a Bayesian belief network is a graphical representation of the independences holding among the variables in the probability distribution that is being represented. It takes the form of an *acyclic directed graph*, or *digraph*, for short. In this digraph G, each node represents a statistical variable that can take one of a finite set of values. Informally speaking, the digraph's arcs model the dependences among the represented variables. An arc $V_i \rightarrow V_j$ to is interpreted as a direct influential or causal relationship between the variables V_i and V_j; the arc's direction designates V_j as the effect or consequence of the cause V_i. More precisely, the absent arcs represent independencies. Absence of an arc between two nodes means that the corresponding variables do not influence each other directly and, hence, are (conditionally) independent. In the sequel, we will use $\pi_G(V_i)$ to denote the set of (immediate) predecessors, or causes, of node V_i in G and use $\pi^*_G(V_i)$ to denote the set of nodes composed of V_i and all its ancestors; we will use $\sigma_G(V_i)$ to denote the set of (immediate) successors, or effects, of node V_i in G and use $\sigma^*_G(V_i)$ to denote the set of nodes composed of V_i and

all its descendants. The following definitions review the probabilistic meaning that is assigned to the digraph of a Bayesian belief network more formally.

Definition 4.2.1 *Let $G = (V(G), A(G))$ be an acyclic digraph and let s be a chain in G between the nodes V_i and V_j. We say that s is* blocked *by the set of nodes $Y \subseteq V(G)$, if either V_i or V_j is included in Y, or s contains three consecutive nodes X_1, X_2, X_3, for which one of the following conditions holds:*

1. *arcs $X_1 \leftarrow X_2$ and $X_2 \rightarrow X_3$ are on the chain s, and $X_2 \in Y$;*
2. *arcs $X_1 \rightarrow X_2$ and $X_2 \rightarrow X_3$ are on the chain s, and $X_2 \in Y$;*
3. *arcs $X_1 \rightarrow X_2$ and $X_2 \leftarrow X_3$ are on the chain s and $\sigma^*_G(X_2) \cap Y = \varnothing$.*

In reviewing the concept of a blocked chain, we have distinguished between three conditions. Figure 4.1 serves as a reference for these conditions; in the two chains representing the conditions 1 and 2, node X_2 is drawn with shading to indicate that it is comprised in the blocking set Y for the chain at hand.

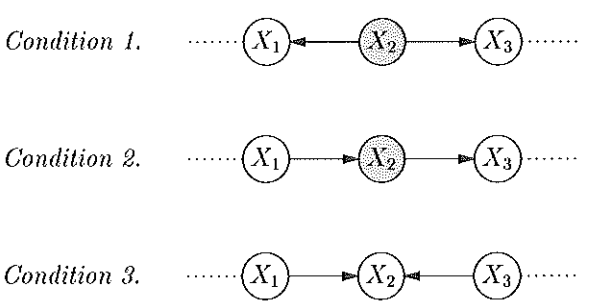


Figure 4.1: The three conditions for chain blocking.

Building upon the concept of blocking, we review the well-known *d-separation criterion* for sets of chains.

Definition 4.2.2 *Let $G = (V(G), A(G))$ be an acyclic digraph and let $X, Y, Z \subseteq V(G)$. The set of nodes Y is said to* d-separate *the sets of nodes X and Z in G, denoted $\langle X \mid Y \mid Z \rangle^d_G$, if for each node $V_i \in X$ and each node $V_j \in Z$, every chain from V_i to V_j in G is blocked by Y.*

The following definition relates the d-separation criterion to the concept of independence.

Definition 4.2.3 *Let $G = (V(G), A(G))$ be an acyclic digraph and let* Pr *be a joint probability distribution on $V(G)$. Then, G is called an* I-map *(independence map) for* Pr *if for all sets of variables $X, Y, Z \subseteq V(G)$, we have: if $\langle X \mid Y \mid Z \rangle^d_G$, then X and Z are conditionally independent given Y in* Pr.

The d-separation criterion thus provides for reading independences from a belief network's digraph without having to resort to probabilistic computations. We would like to note that the criterion of d-separation generally is defined for mutually exclusive sets of nodes only. We have extended the definition to apply to overlapping set of nodes as well, to provide for reading from a digraph independences for instantiated (observed) nodes. We take an instantiated node to be d-separated from any other node. Our extension has been inspired by previous work on informational independence [Van der Gaag & Meyer, 1998].

Associated with the qualitative part of a Bayesian belief network are numerical quantities that describe the strengths of the dependences among the represented variables. With each node V_i of the network's digraph G is associated a set of conditional probabilities $p(V_i \mid \pi_G(V_i))$ describing the joint influence of the various values for the node's (immediate) predecessors $\pi_G(V_i)$ on the probabilities of the values of V_i itself. These probabilities with each other constitute the quantitative part of the belief network.

We review the concept of Bayesian belief network more formally.

Definition 4.2.4 *A* Bayesian belief network *is a tuple* $B = (G, P)$ *where*

- $G = (V(G), A(G))$ *is an acyclic digraph with nodes* $V(G) = \{V_1, \ldots, V_n\}$, $n \geq 1$, *and arcs* $A(G)$;
- P *is a set of conditional probabilities* $p(V_i \mid \pi_G(V_i))$, *for all* $V_i \in V(G)$.

We illustrate the concept of Bayesian belief network by means of an example that will be used for our running example throughout the chapter.

Example 4.2.5 We consider the well-known ALARM-network [Beinlich *et al.*, 1989]. ALARM, which stands for A Logical Alarm Reduction Mechanism, simulates an anesthesia monitor. The digraph of the network is reproduced in Figure 4.2; for the examples in the remainder of the chapter, we have indicated the node of interest, *LV failure* (left ventricular failure), by a double circle and the network's observable nodes by shading. From the network's digraph, various independences are read. For example, the variable *LV failure* is independent of the variable *Insuff anest*, if no information is available yet; the two variables become dependent, however, when, for example, the value of the variable *Blood press* becomes available. The variables *Pulm emb* and *Heart rate*, on the other hand, are dependent, but become independent once a value for *SaCO2* is observed. Associated with the nodes of the network are conditional probabilities. For example, for the node *Stroke vol*, the following conditional probabilities are specified:

$$\begin{aligned} p(\textit{Stroke vol} = \textit{low} \mid \textit{Hypovolemia} = \textit{false} \wedge \textit{LV failure} = \textit{false}) &= 0.05 \\ p(\textit{Stroke vol} = \textit{normal} \mid \textit{Hypovolemia} = \textit{false} \wedge \textit{LV failure} = \textit{false}) &= 0.90 \\ p(\textit{Stroke vol} = \textit{high} \mid \textit{Hypovolemia} = \textit{false} \wedge \textit{LV failure} = \textit{false}) &= 0.05 \end{aligned}$$

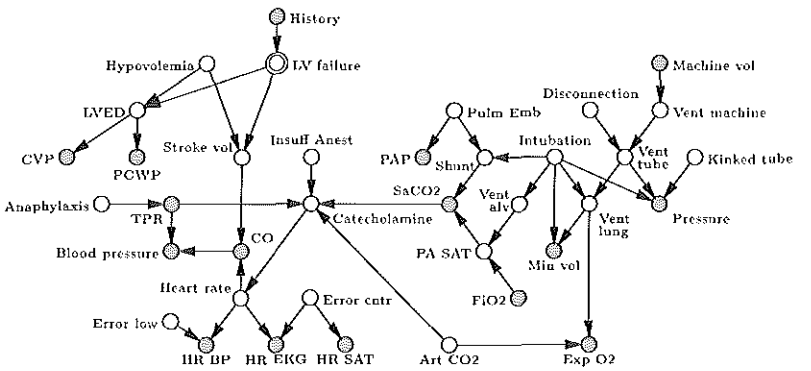


Figure 4.2: The digraph of the ALARM belief network.

$$
\begin{aligned}
p(\mathit{Stroke\ vol} = \mathit{low} \mid \mathit{Hypovolemia} = \mathit{true} \wedge \mathit{LV\ failure} = \mathit{false}) &= 0.5 \\
p(\mathit{Stroke\ vol} = \mathit{normal} \mid \mathit{Hypovolemia} = \mathit{true} \wedge \mathit{LV\ failure} = \mathit{false}) &= 0.49 \\
p(\mathit{Stroke\ vol} = \mathit{high} \mid \mathit{Hypovolemia} = \mathit{true} \wedge \mathit{LV\ failure} = \mathit{false}) &= 0.01
\end{aligned}
$$

$$
\begin{aligned}
p(\mathit{Stroke\ vol} = \mathit{low} \mid \mathit{Hypovolemia} = \mathit{false} \wedge \mathit{LV\ failure} = \mathit{true}) &= 0.95 \\
p(\mathit{Stroke\ vol} = \mathit{normal} \mid \mathit{Hypovolemia} = \mathit{false} \wedge \mathit{LV\ failure} = \mathit{true}) &= 0.04 \\
p(\mathit{Stroke\ vol} = \mathit{high} \mid \mathit{Hypovolemia} = \mathit{false} \wedge \mathit{LV\ failure} = \mathit{true}) &= 0.01
\end{aligned}
$$

$$
\begin{aligned}
p(\mathit{Stroke\ vol} = \mathit{low} \mid \mathit{Hypovolemia} = \mathit{true} \wedge \mathit{LV\ failure} = \mathit{true}) &= 0.98 \\
p(\mathit{Stroke\ vol} = \mathit{normal} \mid \mathit{Hypovolemia} = \mathit{true} \wedge \mathit{LV\ failure} = \mathit{true}) &= 0.01 \\
p(\mathit{Stroke\ vol} = \mathit{high} \mid \mathit{Hypovolemia} = \mathit{true} \wedge \mathit{LV\ failure} = \mathit{true}) &= 0.01
\end{aligned}
$$

As for this chapter, the specific assessments for the various conditional probabilities are not of interest, we refrain from further detailing them. □

The following proposition states that the conditional probabilities of a Bayesian belief network provide all information necessary for uniquely defining a joint probability distribution on the variables discerned that respects the independences portrayed by the network's qualitative part; henceforth, we will call this distribution the joint probability distribution *defined* by the network.

Proposition 4.2.6 *Let $B = (G, P)$ be a Bayesian belief network. Then,*

$$\Pr(V(G)) = \prod_{V_i \in V(G)} p(V_i \mid \pi_G(V_i))$$

defines a joint probability distribution Pr *on $V(G)$ such that G is an I-map for* Pr.

Since the digraph of a Bayesian belief network and its associated conditional probabilities with each other define a unique joint probability distribution on the variables discerned, any (prior or posterior) probability of interest can be computed from the network. For this purpose various algorithms are available [Pearl, 1988, Lauritzen & Spiegelhalter, 1988].

4.3 Uninfluential probabilities in a sensitivity analysis

Sensitivity analysis is a general technique for studying the effects of the inaccuracies in the parameters of a mathematical model on this model's output [Habbema *et al.*, 1990, Morgan & Henrion, 1990]. Sensitivity analysis basically amounts to systematically varying the values of the parameters of the model under study. In a one-way sensitivity analysis, the values of the parameters are varied one at a time while keeping the values of all other parameters fixed. For a Bayesian belief network, a one-way sensitivity analysis amounts to varying the assessment for a single conditional probability of the network's quantitative part. As discussed in Chapter 2, the analysis provides for studying the effects of the inaccuracy in the specified assessment on a probability of interest.

In essence, in a one-way sensitivity analysis of a Bayesian belief network, the sensitivity of the network's probability of interest is investigated with respect to every single conditional probability. Various conditional probabilities of a belief network, however, are known beforehand not to affect the probability of interest upon variation, for example because this probability of interest is shielded from their influence by available observations. These uninfluential probabilities can be readily identified by inspection of the network's digraph, that is, without any probabilistic computations. We say that the probability of interest is *algebraically independent* of these uninfluential conditional probabilities. For abbreviation, we will write $p \propto q$ to denote that the probability p is algebraically independent of the probability q. We would like to note that the phrase algebraic independence is used to refer to the absence of any effect of varying the assessment for a conditional probability under study on a probability of interest, *as induced by the network's digraph.* Also note that the phrase applies to probabilities whereas the phrase probabilistic independence pertains to variables. Now, in a one-way sensitivity analysis of a Bayesian belief network, for a conditional probability of which the network's probability of interest is algebraically independent, no further investigation is required. The sensitivity analysis of the network can therefore be restricted to the conditional probabilities of which the probability of interest is algebraically *dependent.* The nodes to which these conditional probabilities refer constitute the *sensitivity set* for the node of interest.

We define the concept of sensitivity set more formally.

Definition 4.3.1 *Let B be a Bayesian belief network with the digraph $G = (V(G), A(G))$. Let $V_r \in V(G)$ be the network's node of interest and let $O \subseteq V(G)$ be the set of observed nodes in G. Now, let G^* be the digraph that is constructed from G by adding an auxiliary predecessor X_i to every node $V_i \in V(G)$. Then, the* sensitivity set *for V_r given O, denoted $Sen(V_r, O)$, is the set of all nodes $V_i \in V(G)$ for which $\neg\langle\{X_i\}|O|\{V_r\}\rangle^d_{G^*}$.*

From the previous definition we have that the sensitivity set for a node of interest V_r is computed from the digraph of a belief network under consideration by adding an auxiliary predecessor X_i to every node V_i and thereupon exploiting the d-separation criterion. The auxiliary predecessor X_i of node V_i can be looked upon as capturing the presence of inaccuracy in the probability assessments for V_i. If the presence of inaccuracy in V_i's assessments is not d-separated from the node of interest or, in other words, if V_r is not shielded from the inaccuracy by the available evidence, then varying the assessments for V_i's conditional probabilities may influence the probabilities of the values of V_r. V_i is therefore included in V_r's sensitivity set. We would like to note that the basic idea of capturing the presence of inaccuracy by means of auxiliary nodes has been exploited before [Spiegelhalter, 1989]. We further note that we capture the *presence* of inaccuracy rather than the inaccuracy itself by auxiliary nodes.

The following example illustrates our concept of sensitivity set.

Example 4.3.2 We consider once again the ALARM-network, the digraph of which is shown in Figure 4.2. We are interested in the diagnostic variable *LV failure*; our probability of interest is the probability that *LV failure* = *true*. We consider the sensitivity set for the node *LV failure* given various different sets of observed nodes.

If the set of observed nodes is empty, that is, when no patient observations are available, the sensitivity set for the node *LV failure* equals

$$Sen(LV\ failure, \varnothing) = \{LV\ failure, History\}$$

Upon performing a one-way sensitivity analysis of the a priori belief network, only the conditional probabilities of these two nodes need be investigated; the conditional probabilities of all other nodes in the network upon variation cannot influence the probability of interest.

Now, suppose that we would like to evaluate the sensitivity of the network's probability of interest in view of observations for the nodes in the set $O_1 = \{History, CVP, TPR, Blood\ press, CO\}$. The sensitivity set for *LV failure* given O_1 equals

$$\begin{aligned} Sen(LV\ failure, O_1) = \{&LV\ failure, Hypovolemia, LVED, CVP, Stroke\ vol,\\ &CO, Insuff, anest, Catecholamine, Heart\ rate,\\ &Art\ CO2, SaCO2, PA\ SAT, FiO2, Vent\ alv, Shunt,\\ &Intubation, Pulm\ emb\} \end{aligned}$$

From the 37 nodes included in the belief network, the conditional probabilities of only 17 nodes need be investigated in the analysis. We would like to note that, in general, a sensitivity set does *not* coincide with the set of non-d-separated nodes for the node of interest. From the sensitivity set for the node *LV failure* given O_1, for example, it is readily seen that a sensitivity set can include both non-d-separated nodes (such as the node *Stroke vol*) and d-separated nodes (such as the node *CO*); also, the set of nodes that are not comprised in the sensitivity set can include non-d-separated nodes (such as the node *PCWP*) as well as d-separated nodes (such as *Blood press*).

Now, if in addition to observations for the nodes in the set O_1 an observation is assumed for the node *SaCO2*, yielding O_2 for the new set of observed nodes, the sensitivity set for *LV failure* reduces in size from 17 nodes to 10 nodes:

$$\begin{aligned} Sen(LV\ failure, O_2) \;=\; & \{LV\ failure, Hypovolemia, LVED, CVP, Stroke\ vol, \\ & CO, Insuff\ anest, Catecholamine, Heart\ rate, \\ & Art\ CO2\} \end{aligned}$$

Note that, when a value for the node of interest *LV failure* is available, every node in the auxiliary network for determining the sensitivity set is d-separated from *LV failure*. The sensitivity set then is empty. □

In order to prove the claims we have made so far with respect to a sensitivity set, we will partition a belief network's set of nodes that are not included in a sensitivity set under study into three mutually exclusive sets of nodes. We will then show that, for various different reasons, the conditional probabilities for the nodes included in these sets upon variation have no effect on the network's probability of interest.

Definition 4.3.3 *Let B be a Bayesian belief network with the digraph G, let V_r be the network's node of interest, and let O the set of observed nodes, as before. We define the sets of nodes $Insen_1(V_r, O)$, $Insen_2(V_r, O)$, and $Insen_3(V_r, O)$, respectively, as*

- *for every node $V_i \in \pi^*_G(V_r)$, if $\langle(\{V_i\} \cup \pi_G(V_i)) \mid O \mid \{V_r\}\rangle^d_G$, then $V_i \in Insen_1(V_r, O)$;*
- *for every node $V_i \in V(G) \setminus \pi^*_G(V_r)$, if $\langle(\{V_i\} \cup \pi_G(V_i)) \mid O \mid \{V_r\}\rangle^d_G$ and $\sigma^*_G(V_i) \cap O \neq \varnothing$, then $V_i \in Insen_2(V_r, O)$;*
- *for every node $V_i \in V(G) \setminus \pi^*_G(V_r)$, if $\sigma^*_G(V_i) \cap O = \varnothing$, then $V_i \in Insen_3(V_r, O)$.*

The sets $Insen_1(V_r, O)$, $Insen_2(V_r, O)$, and $Insen_3(V_r, O)$ include nodes to whose conditional probabilities a belief network's probability of interest is insensitive. Before illustrating the three sets of nodes for our running example, we informally address their meaning. In doing so, we begin by considering the ancestors V_i of the node of

interest V_r. We observe that any unblocked chain from V_i to V_r, be it a direct chain or a chain via a predecessor of V_i, provides for conveying an influence from V_i's probability assessments to V_r. If no such chain is present, therefore, varying the assessments for node V_i can have no influence on the probability of interest. The set $Insen_1(V_r, O)$ now includes all ancestors V_i of V_r such that V_r is d-separated by the available observations from both V_i and V_i's predecessors. We now consider the non-ancestors of the node of interest. We observe that the probability assessments for a non-ancestor V_i of V_r cannot influence the probability of interest if there are no observations available. Only an observed descendant of V_i that induces an influence on V_r through V_i, can cause varying V_i's probability assessments to affect the probability of interest. The set $Insen_3(V_r, O)$ now includes all non-ancestors of V_r that do not have any observed descendants. The set $Insen_2(V_r, O)$, to conclude, includes the non-ancestors of V_r that happen to have observed descendants yet whose influence is shielded from V_r by the available observations: the set includes all non-ancestors V_i of V_r with at least one observed descendant such that V_r is d-separated from both V_i and V_i's predecessors.

We illustrate the various sets of nodes defined above by means of our running example.

Example 4.3.4 We consider again the ALARM-network, the digraph of which is shown in Figure 4.2. We are once more interested in the variable *LV failure*; for our probability of interest, we take the probability that *LV failure* is *true*. We recall from Example 4.3.2 that, if the set of observed nodes is empty, the sensitivity set for the node of interest *LV failure* equals $Sen(LV\ failure, \varnothing) = \{LV\ failure, History\}$. The set of all remaining nodes, that is, the set of all nodes, *LV failure* and *History* excluded, is partitioned into three sets as defined above. Of these, the sets $Insen_1(LV\ failure, \varnothing)$ and $Insen_2(LV\ failure, \varnothing)$ are empty; the set $Insen_3(LV\ failure, \varnothing)$ includes any node that is not comprised in the sensitivity set. We consider, as an example, the node *Stroke vol*. From Figure 4.2, we see that *Stroke vol* is not an ancestor of the node of interest *LV failure*; furthermore, it does not have any observed descendants. From Definition 4.3.3, we conclude that the node *Stroke vol* belongs to the set $Insen_3(LV\ failure, \varnothing)$. Informally speaking, as the node *Stroke vol* is not observed and does not have any observed descendants, it cannot exert nor pass on any diagnostic influence on the probabilities for *LV failure*. The probability of interest $\Pr(LV\ failure = true)$ therefore is algebraically independent of the conditional probabilities for *Stroke vol*. A similar observation applies to any other node from the set $Insen_3(LV\ failure, \varnothing)$.

We now assume that observations are obtained for the nodes in the set $O_1 = \{History, CVP, TPR, Blood\ press, CO\}$. The sets $Insen_1(LV\ failure, O_1)$, $Insen_2(LV\ failure, O_1)$, and $Insen_3(LV\ failure, O_1)$ equal

$$Insen_1(LV\ failure, O_1) = \{History\}$$

$$Insen_2(LV\ failure, O_1) = \{Blood\ press, TPR, Anaphylaxis\}$$

$$Insen_3(LV\ failure, O_1) = \{PCWP, Error\ low, HR\ BP, HR\ EKG, HR\ SAT, Error\ cntr, Exp\ O2, Min\ vol, Vent\ lung, Pressure, Vent\ tube, Kinked\ tube, Disconnection, Vent\ machine, Machine\ vol, PAP\}$$

We consider, as an example, the node *History*. This node is a predecessor of the node of interest *LV failure*. It is d-separated from *LV failure* and does not have any immediate predecessors that are not d-separated from *LV failure*. From Definition 4.3.3, therefore, we have that *History* is included in the set $Insen_1(LV\ failure, O_1)$. Informally speaking, as a value for the node *History* is available, its prior probabilities are irrelevant to the probabilities for its successor *LV failure*. The probability of *LV failure* = *true* given the available observations for O_1 therefore is algebraically independent of the prior probabilities for the node *History*. To conclude our example, we consider the node *TPR*. From Figure 4.2, we observe that *TPR* is not an ancestor of the node of interest *LV failure*. The node *TPR* itself as well as its immediate predecessor *Anaphylaxis* are d-separated from *LV failure* given the available observations. Furthermore, the descendant *Blood press* of *TPR* is observed. By definition, we have that the node *TPR* is included in the set $Insen_2(LV\ failure, O_1)$. Informally speaking, from *TPR* and its predecessor *Anaphylaxis* being d-separated from the node of interest *LV failure*, we find that any diagnostic influence originating from *TPR* is shielded from *LV failure* by the available observations. Therefore, the probability of interest is algebraically independent of the conditional probabilities for the node *TPR*. A similar observation applies to any other node from the set $Insen_2(LV\ failure, O_1)$. □

We would like to note that for a node of interest V_r and any set of observed nodes O, the three sets $Insen_1(V_r, O)$, $Insen_2(V_r, O)$, and $Insen_3(V_r, O)$, and the sensitivity set $Sen(V_r, O)$ are mutually exclusive and collectively exhaustive; for a formal proof of this property, the reader is referred to the appendix.

In the remainder of this section, we will show that the probability of interest of a Bayesian belief network is indeed algebraically independent of the conditional probabilities of any node that is not included in the sensitivity set $Sen(V_r, O)$ under study. To this end, we investigate the three sets $Insen_1(V_r, O)$, $Insen_2(V_r, O)$, and $Insen_3(V_r, O)$ separately and provide for each of these sets a lemma stating algebraic independence of the probability of interest for the conditional probabilities of the nodes in the set at hand. Our main result is then stated in Proposition 4.3.11, building upon these lemmas. The proofs of the three lemmas, although not complicated, are rather elaborate; the full proofs therefore are deferred to the appendix.

In the first lemma, we state that a belief network's probability of interest for a node V_r given observations for nodes O is algebraically independent of the conditional probabilities of any node from the set $Insen_3(V_r, O)$.

Lemma 4.3.5 *Let B be a Bayesian belief network and let* Pr *be the joint probability distribution defined by B. Let O be the set of observed nodes and let o denote the corresponding observations. Let V_r be the network's node of interest. Then, for any value v_r of V_r, we have that* $\Pr(v_r \mid o) \propto p(V_i \mid \pi(V_i))$ *for every node* $V_i \in \mathit{Insen}_3(V_r, O)$.

Proof (Sketch). The probability of interest $\Pr(v_r \mid o)$ for the belief network B equals

$$\Pr(v_r \mid o) = \frac{\Pr(v_r \wedge o)}{\Pr(o)}$$

We recall from Section 4.2 that the joint probability distribution Pr that is defined by B, can be written as a product of the network's conditional probabilities. From the basic property of marginalisation, we now have that both the numerator and the denominator can be written as a sum of products of conditional probabilities. In these sums, for every unobserved leaf node, there appear as many products as there are values for this node that differ in this node's probability only. Summing over these products amounts to summing out the leaf node by marginalisation. The same argument applies recursively to all unobserved non-ancestors of V_r that do not have any observed descendants, that is, the argument applies to every node from the set $\mathit{Insen}_3(V_r, O)$. We conclude that the probability of interest is algebraically independent of the conditional probabilities of any node from this set. □

We illustrate the property stated in the previous lemma by means of an example.

Example 4.3.6 We consider the belief network from Figure 4.3, which is a small

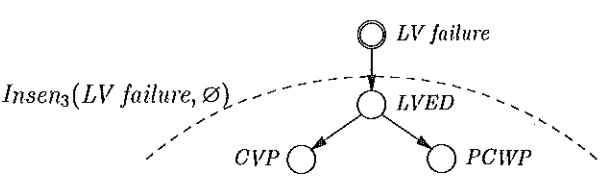


Figure 4.3: An example belief network, illustrating the property stated in Lemma 4.3.5 for the node of interest *LV failure* and the empty set of observed nodes; the set $\mathit{Insen}_3(\mathit{LV\ failure}, \varnothing)$ consists of the nodes *LVED*, *CVP*, and *PCWP*.

fragment of the ALARM-network. The possible values of the node *LV failure* are *fail* and *no fail*; the possible values for each of the nodes *LVED*, *CVP*, and *PCWP* are *low*, *normal*, and *high*. Our node of interest once again is the node *LV failure*, indicated in the figure by a double circle. We now address the situation where no observations are available yet and investigate the probability of interest $\Pr(\mathit{fail})$. From Definition 4.3.3, we find that the set $\mathit{Insen}_3(\mathit{LV\ failure}, \varnothing)$ consists of the three nodes *LVED*, *CVP*, and

PCWP. For the probability of interest, we find that

$$
\begin{aligned}
&\Pr(\mathit{fail}) = \\
&= \sum_{\substack{\{LVED,\, CVP,\\ PCWP\}}} p(PCWP \mid LVED) \cdot p(CVP \mid LVED) \cdot p(LVED \mid \mathit{fail}) \cdot p(\mathit{fail}) \\
&= \sum_{\{LVED,CVP\}} \Big(p(\mathit{low}\ PCWP \mid LVED) + p(\mathit{normal}\ PCWP \mid LVED) + \\
&\qquad + p(\mathit{high}\ PCWP \mid LVED) \Big) \cdot p(CVP \mid LVED) \cdot p(LVED \mid \mathit{fail}) \cdot p(\mathit{fail}) \\
&= \sum_{\{LVED,CVP\}} p(CVP \mid LVED) \cdot p(LVED \mid \mathit{fail}) \cdot p(\mathit{fail}) \\
&= \sum_{\{LVED\}} \Big(p(\mathit{low}\ CVP \mid LVED) + p(\mathit{normal}\ CVP \mid LVED) + p(\mathit{high}\ CVP \mid LVED) \Big) \cdot \\
&\qquad \cdot p(LVED \mid \mathit{fail}) \cdot p(\mathit{fail}) \\
&= \sum_{\{LVED\}} p(LVED \mid \mathit{fail}) \cdot p(\mathit{fail}) \\
&= \Big(p(\mathit{low}\ LVED \mid \mathit{fail}) + p(\mathit{normal}\ LVED \mid \mathit{fail}) + p(\mathit{high}\ LVED \mid \mathit{fail}) \Big) \cdot p(\mathit{fail}) \\
&= p(\mathit{fail})
\end{aligned}
$$

From this derivation, it is readily seen that the probability of interest $\Pr(\mathit{fail})$ is algebraically independent of the conditional probabilities of the three nodes included in the set $\mathit{Insen}_3(\mathit{LV\ failure}, \varnothing)$.

We now address the situation where the value *high* is observed for the node *PCWP*. This situation is depicted in Figure 4.4, where the node *PCWP* is drawn with shading to indicate that its value has been observed. The set $\mathit{Insen}_3(\mathit{LV failure}, \{PCWP\})$ is

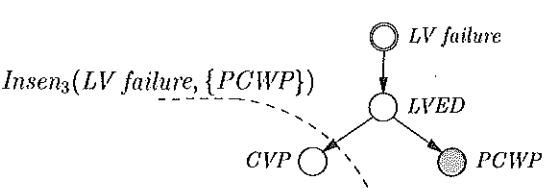


Figure 4.4: An example belief network, illustrating the property stated in Lemma 4.3.5 for the node of interest *LV failure* and the set of observed nodes $\{PCWP\}$; the set $\mathit{Insen}_3(\mathit{LV\ failure}, \{PCWP\})$ consists of the single node *CVP*.

composed of the node *CVP* only. For our probability of interest $\Pr(fail \mid high\ PCWP)$, we now find that

$$\Pr(fail \mid high\ PCWP) = \frac{\Pr(fail \wedge high\ PCWP)}{\Pr(high\ PCWP)}$$

The numerator in this equation equals

$$\begin{aligned}
&\Pr(fail \wedge high\ PCWP) = \\
&= \sum_{\{LVED, CVP\}} p(CVP \mid LVED) \cdot p(high\ PCWP \mid LVED) \cdot p(LVED \mid fail) \cdot p(fail) \\
&= \sum_{\{LVED\}} \Big(p(low\ CVP \mid LVED) + p(normal\ CVP \mid LVED) + p(high\ CVP \mid LVED) \Big) \cdot \\
&\qquad \cdot p(high\ PCWP \mid LVED) \cdot p(LVED \mid fail) \cdot p(fail) \\
&= \sum_{\{LVED\}} p(high\ PCWP \mid LVED) \cdot p(LVED \mid fail) \cdot p(fail)
\end{aligned}$$

The denominator equals

$$\begin{aligned}
&\Pr(high\ PCWP) = \\
&= \sum_{\substack{\{LV\ failure,\\ LVED, CVP\}}} p(CVP \mid LVED) \cdot p(high\ PCWP \mid LVED) \cdot p(LVED \mid LV\ failure) \cdot \\
&\qquad \cdot p(LV\ failure) \\
&= \sum_{\substack{\{LV\ failure,\\ LVED\}}} \Big(p(low\ CVP \mid LVED) + p(normal\ CVP \mid LVED) + p(high\ CVP \mid LVED) \Big) \cdot \\
&\qquad \cdot p(high\ PCWP \mid LVED) \cdot p(LVED \mid LV failure) \cdot p(LV failure) \\
&= \sum_{\substack{\{LV\ failure,\\ LVED\}}} p(high\ PCWP \mid LVED) \cdot p(LVED \mid LV\ failure) \cdot p(LV\ failure)
\end{aligned}$$

We conclude that the probability of interest equals

$$\begin{aligned}
&\Pr(fail \mid high\ PCWP) = \\
&= \frac{\displaystyle\sum_{\{LVED\}} p(high\ PCWP \mid LVED) \cdot p(LVED \mid fail) \cdot p(fail)}{\displaystyle\sum_{\substack{\{LV\ failure,\\ LVED\}}} p(high\ PCWP \mid LVED) \cdot p(LVED \mid LV\ failure) \cdot p(LV\ failure)}
\end{aligned}$$

From this derivation, it is readily seen that the probability of interest $\Pr(\mathit{fail} \mid \mathit{high\ PCWP})$ is algebraically independent of the conditional probabilities of CVP, the only node included in the set $\mathit{Insen}_3(\mathit{LV\ failure}, \{\mathit{PCWP}\})$. □

So far, we have shown that a belief network's probability of interest for a node V_r given observations for nodes O is algebraically independent of the conditional probabilities of any node from the set $\mathit{Insen}_3(V_r, O)$. We now proceed by observing that this probability of interest is also algebraically independent of the conditional probabilities of the nodes from the set $\mathit{Insen}_2(V_r, O)$.

Lemma 4.3.7 *Let B be a Bayesian belief network and let* Pr *be its joint probability distribution. Let O be the set of observed nodes with observations o, as before. Let V_r be the network's node of interest. Then, for any value v_r of V_r, we have that* $\Pr(v_r \mid o) \propto p(V_i \mid \pi(V_i))$ *for every node* $V_i \in \mathit{Insen}_2(V_r, O)$.

Proof (Sketch). The probability of interest $\Pr(v_r \mid o)$ for the belief network B equals

$$\Pr(v_r \mid o) = \frac{\Pr(v_r \wedge o)}{\Pr(o)}$$

Both the numerator and the denominator of this equation can be written as a sum of products of conditional probabilities from the network. From Definition 4.3.3, we know that the nodes from the set $\mathit{Insen}_2(V_r, O)$ as well as their immediate predecessors are d-separated from the node of interest V_r by the available observations; more in specific, we know that a predecessor of any node from $\mathit{Insen}_2(V_r, O)$ either belongs to $\mathit{Insen}_2(V_r, O)$ itself or is an observed node. In both the numerator and the denominator of the above equation, therefore, a term can be isolated that includes all the nodes from $\mathit{Insen}_2(V_r, O)$ and no other nodes that are not observed. As this term appears in the numerator as well as in the denominator, it cancels out. The conditional probabilities of the nodes from $\mathit{Insen}_2(V_r, O)$ upon variation therefore do not affect the probability of interest. □

We illustrate the property stated in Lemma 4.3.7 by means of an example.

Example 4.3.8 We consider the belief network from Figure 4.5, which again is a small fragment of the ALARM-network. The possible values of the node *LV failure* are *fail* and *no fail*; the possible values for each of the other nodes are *low*, *normal*, and *high*. Our node of interest once again is the node *LV failure*. We now address the situation where the value *low* has been observed for both the nodes *CO* and *Blood press*, and investigate the probability of interest $\Pr(\mathit{fail} \mid \mathit{low\ CO} \wedge \mathit{low\ Blood\ press})$. From Definition 4.3.3, we have that the set $\mathit{Insen}_2(\mathit{LV\ failure}, \{\mathit{CO}, \mathit{Blood\ press}\})$ comprises the node *Blood press* only. Note that $\mathit{Insen}_3(\mathit{LV\ failure}, \{\mathit{CO}, \mathit{Blood\ press}\}) = \varnothing$. For the probability of interest, we find that

$$\Pr(\mathit{fail} \mid \mathit{low\ CO} \wedge \mathit{low\ Blood\ press}) = \frac{\Pr(\mathit{fail} \wedge \mathit{low\ CO} \wedge \mathit{low\ Blood\ press})}{\Pr(\mathit{low\ CO} \wedge \mathit{low\ Blood\ press})}$$

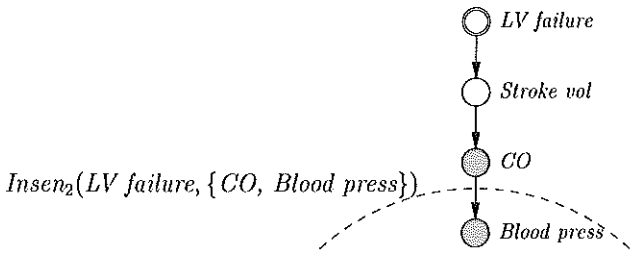


Figure 4.5: An example belief network, illustrating the property stated in Lemma 4.3.7 for the node of interest *LV failure* and the set of observed nodes $\{CO, Blood\ press\}$; the set $Insen_2(LV\ failure, \{CO, Blood\ press\})$ consists of the node *Blood press* only.

The numerator in this equation equals

$$\Pr(fail \wedge low\ CO \wedge low\ Blood\ press) =$$

$$= \sum_{\{Stroke\ vol\}} p(low\ Blood\ press \mid low\ CO) \cdot p(low\ CO \mid Stroke\ vol) \cdot \cdot p(Stroke\ vol \mid fail) \cdot p(fail)$$

$$= p(low\ Blood\ press \mid low\ CO) \cdot \cdot \left(\sum_{\{Stroke\ vol\}} p(low\ CO \mid Stroke\ vol) \cdot p(Stroke\ vol \mid fail) \cdot p(fail) \right)$$

The denominator in the equation equals

$$\Pr(low\ CO \wedge low\ Blood\ press) =$$

$$= \sum_{\substack{\{LV\ failure,\\ Stroke\ vol\}}} p(low\ Blood\ press \mid low\ CO) \cdot p(low\ CO \mid Stroke\ vol) \cdot \cdot p(Stroke\ vol \mid LV\ failure) \cdot p(LV\ failure)$$

$$= p(low\ Blood\ press \mid low\ CO) \cdot \cdot \left(\sum_{\substack{\{LV\ failure,\\ Stroke\ vol\}}} p(low\ CO \mid Stroke\ vol) \cdot p(Stroke\ vol \mid LV\ failure) \cdot p(LV\ failure) \right)$$

We now conclude that the probability of interest equals

$$\Pr(\mathit{fail} \mid \mathit{low\ CO} \wedge \mathit{low\ Blood\ press}) =$$

$$= \frac{\sum_{\{\mathit{Stroke\ vol}\}} p(\mathit{low\ CO} \mid \mathit{Stroke\ vol}) \cdot p(\mathit{Stroke\ vol} \mid \mathit{fail}) \cdot p(\mathit{fail})}{\sum_{\substack{\{\mathit{LV\ failure},\\ \mathit{Stroke\ vol}\}}} p(\mathit{low\ CO} \mid \mathit{Stroke\ vol}) \cdot p(\mathit{Stroke\ vol} \mid \mathit{LV\ failure}) \cdot p(\mathit{LV\ failure})}$$

From this derivation, it is readily seen that the probability of interest is algebraically independent of the conditional probabilities of *Blood press*, the only node that is included in the set $\mathit{Insen}_2(\mathit{LV\ failure}, \{\mathit{CO}, \mathit{Blood\ press}\})$. □

So far, we have shown that a belief network's probability of interest for a node V_r given observations for nodes O is algebraically independent of the conditional probabilities of any node from the sets $\mathit{Insen}_3(V_r, O)$ and $\mathit{Insen}_2(V_r, O)$. To conclude, we now state that this probability of interest is also algebraically independent of the conditional probabilities of the nodes from the set $\mathit{Insen}_1(V_r, O)$.

Lemma 4.3.9 *Let B be a Bayesian belief network and let* Pr *be its joint probability distribution. Let O be the set of observed nodes with observations o, as before. Let V_r be the network's node of interest. Then, for any value v_r of V_r, we have that* $\Pr(v_r \mid o) \propto p(V_i \mid \pi(V_i))$ *for every node $V_i \in \mathit{Insen}_1(V_r, O)$.*

Proof (Sketch). The probability of interest $\Pr(v_r \mid o)$ for the belief network B once again equals

$$\Pr(v_r \mid o) = \frac{\Pr(v_r \wedge o)}{\Pr(o)}$$

As before, both the numerator and the denominator of this equation can be written as a sum of products of conditional probabilities from the network. The proof is now based on canceling out terms from the numerator and the denominator as in the proof of the previous lemma. From Definition 4.3.3, we know that the nodes from the set $\mathit{Insen}_1(V_r, O)$ as well as their immediate predecessors are d-separated from the node of interest V_r by the available observations; more in specific, we know that a predecessor of any node from $\mathit{Insen}_1(V_r, O)$ either belongs to $\mathit{Insen}_1(V_r, O)$ itself or is an observed node. In both the numerator and the denominator of the above equation, therefore, a term can be isolated that includes all the nodes from $\mathit{Insen}_1(V_r, O)$ and no other nodes that are not observed. As this term appears in the numerator as well as in the denominator, it once again cancels out. The conditional probabilities of the nodes from $\mathit{Insen}_1(V_r, O)$ upon variation therefore do not affect the probability of interest. □

We illustrate the property stated in the previous lemma by means of an example.

Example 4.3.10 We consider the belief network from Figure 4.6, which again is a small fragment of the ALARM-network. The possible values of the node *LV failure*

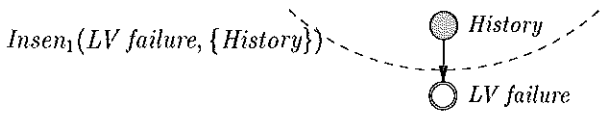


Figure 4.6: An example belief network, illustrating the property stated in Lemma 4.3.9 for the node of interest *LV failure* and the set of observed nodes $\{History\}$; the set $Insen_1(LV\ failure, \{History\})$ consists of just the node *History*.

once more are *fail* and *no fail*; the values of the node *History* are *history* and *no history*. Our node of interest again is the node *LV failure*. We now address the situation where the value *history* is observed for the node *History* and investigate the probability of interest $\Pr(fail \mid history)$. From Definition 4.3.3, we have that the set $Insen_1(LV\ failure, \{History\})$ consists of the node *History* only. Note that both $Insen_2(LV\ failure, \{History\})$ and $Insen_3(LV\ failure, \{History\})$ are empty. For the probability of interest, we find that

$$
\begin{aligned}
&\Pr(fail \mid history) = \\
&\quad = \frac{\Pr(fail \wedge history)}{Pr(history)} \\
&\quad = \frac{p(fail \mid history) \cdot p(history)}{\sum\limits_{\{LV\ failure\}} \Big(p(LV\ failure \mid history) \cdot p(history) \Big)} \\
&\quad = \frac{p(fail \mid history) \cdot p(history)}{p(fail \mid history) \cdot p(history) + p(no\ fail \mid history) \cdot p(history)} \\
&\quad = \frac{p(fail \mid history) \cdot p(history)}{\Big(p(fail \mid history) + p(no\ fail \mid history) \Big) \cdot p(history)} \\
&\quad = p(fail \mid history)
\end{aligned}
$$

From this derivation, it is readily seen that the probability of interest $\Pr(fail \mid history)$ is algebraically independent of the prior probabilities of *History*, the only node that is included in the set $Insen_1(LV\ failure, \{History\})$. □

Building upon the three preceding lemmas, we now state our main result.

Proposition 4.3.11 *Let B be a Bayesian belief network and let* Pr *be its joint probability distribution. Let O be the set of observed nodes with the observations o, as before.*

Let V_r be the network's node of interest and let $Sen(V_r, O)$ be the sensitivity set for V_r given O. Then, for any value v_r of V_r, we have that $\Pr(v_r \mid o) \propto p(V_i \mid \pi(V_i))$ for every node $V_i \notin Sen(V_r, O)$.

Proposition 4.3.11 states that a belief network's probability of interest is algebraically independent of the conditional probabilities of any node that is not included in the sensitivity set under study. From this property, we have that sensitivity analyses with respect to these conditional probabilities are uninformative as they will reveal no effect whatsoever on the probability of interest. These sensitivity analyses, therefore, are to no avail and can be excluded from the overall analysis. The number of analyses that can be thus excluded may be considerable, as will be demonstrated in Section 4.5.

4.4 Functional relations in a sensitivity analysis

In the previous section, we have argued that a sensitivity analysis of a Bayesian belief network can be restricted to the conditional probabilities of the nodes in a sensitivity set under study: we know that the conditional probabilities of any other node do not contribute to the probability of interest and upon variation will not show any effect on this probability. To gain insight into the sensitivity of the probability of interest to the various conditional probabilities of the nodes that *are* included in the sensitivity set, further analysis is required. In essence, for every such conditional probability, the effect on the probability of interest can be studied by investigating a number of deviations from the specified assessment. Now, the curve yielded by such an analysis is not arbitrarily shaped, but instead is strongly constrained by the independences that are portrayed by the digraph of the network. In fact, the network's probability of interest relates as a quotient of two linear functions to a conditional probability under study. As we will argue presently, knowledge of this mathematical function renders systematic variation of conditional probabilities in a sensitivity analysis unnecessary.

Proposition 4.4.1 *Let B be a Bayesian belief network with the digraph $G = (V(G), A(G))$ and let $\Pr$ be the joint probability distribution defined by B. Let $O \subseteq V(G)$ be the set of observed nodes in G and let o denote the corresponding observations. Let V_r be the network's node of interest and let $Sen(V_r, O)$ be the sensitivity set for V_r given O. Then, for any value v_r of V_r, we have that*

$$\Pr(v_r \mid o) = \frac{a \cdot x + b}{c \cdot x + d}$$

for every conditional probability $x = p(v_s \mid \pi')$ of every node $V_s \in Sen(V_r, O)$, where a, b, c, and d are constants that are dependent upon the values v_s of V_s and π' of $\pi_G(V_s)$.

Proof (Sketch). The probability of interest $\Pr(v_r \mid o)$ for the belief network B equals

$$\Pr(v_r \mid o) = \frac{\Pr(v_r \wedge o)}{\Pr(o)}$$

We recall that the joint probability distribution Pr, that is defined by the network, can be written as a product of the network's conditional probabilities. From the basic property of marginalisation, we further have that both the numerator and the denominator can be written as a sum of products of conditional probabilities. By separating, in these sums, the terms that specify the conditional probability x under study and those that do not, it is readily seen that $\Pr(v_r \wedge o)$ as well as $\Pr(o)$ relate linearly to x. □

We illustrate the property stated in Proposition 4.4.1 by means of an example.

Example 4.4.2 We consider the Bayesian belief network from Figure 4.7, which again is a small fragment of the ALARM-network. The possible values of the node *Shunt* are *normal* and *high*; the possible values of the node *Pulm emb* are *pulm emb* and *no pulm emb*, and the possible values of the node *PAP* are *low*, *normal*, and *high*. Our node of interest is the node *Shunt*, indicated in the figure by a double circle. We address

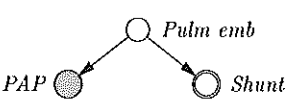


Figure 4.7: An example belief network, illustrating the property stated in Proposition 4.4.1 for the probability of interest $\Pr(\textit{normal Shunt} \mid \textit{high PAP})$ and the conditional probability under study $p(\textit{high PAP} \mid \textit{no pulm emb})$.

the situation where the value *high* has been observed for the node *PAP*, indicated by shading, and consider the probability of interest $\Pr(\textit{normal Shunt} \mid \textit{high PAP})$. From Definition 4.3.1, we have that the sensitivity set $\textit{Sen}(\textit{Shunt}, \{\textit{PAP}\})$ comprises all three nodes from the network. We now investigate the functional relation between the probability of interest and the conditional probability $x = p(\textit{high PAP} \mid \textit{no pulm emb})$ for the node $\textit{PAP} \in \textit{Sen}(\textit{Shunt}, \{\textit{PAP}\})$. For our probability of interest, we find that

$$\Pr(\textit{normal Shunt} \mid \textit{high PAP}) = \frac{\Pr(\textit{normal Shunt} \wedge \textit{high PAP})}{\Pr(\textit{high PAP})}$$

The numerator in this equation equals

$$\Pr(normal\ Shunt \wedge high\ PAP) =$$

$$= \sum_{\{Pulm\ emb\}} p(high\ PAP \mid Pulm\ emb) \cdot p(normal\ Shunt \mid Pulm\ emb) \cdot p(Pulm\ emb)$$

$$= p(high\ PAP \mid no\ pulm\ emb) \cdot p(normal\ Shunt \mid no\ pulm\ emb) \cdot p(no\ pulm\ emb)$$

$$+ p(high\ PAP \mid pulm\ emb) \cdot p(normal\ Shunt \mid pulm\ emb) \cdot p(pulm\ emb)$$

$$= a \cdot x + b$$

where a equals

$$a = p(normal\ Shunt \mid no\ pulm\ emb) \cdot p(no\ pulm\ emb)$$

and b equals

$$b = p(high\ PAP \mid pulm\ emb) \cdot p(normal\ Shunt \mid pulm\ emb) \cdot p(pulm\ emb)$$

The denominator of the probability of interest equals

$$\Pr(high\ PAP) =$$

$$= \sum_{\substack{\{Shunt,\\ Pulm\ emb\}}} p(high\ PAP \mid Pulm\ emb) \cdot p(Shunt \mid Pulm\ emb) \cdot p(Pulm\ emb)$$

$$= \sum_{\{Pulm\ emb\}} p(high\ PAP \mid Pulm\ emb) \cdot p(Pulm\ emb)$$

$$= p(high\ PAP \mid no\ pulm\ emb) \cdot p(no\ pulm\ emb) +$$

$$+ p(high\ PAP \mid pulm\ emb) \cdot p(pulm\ emb)$$

$$= c \cdot x + d$$

where c equals

$$c = p(no\ pulm\ emb)$$

and d equals

$$d = p(high\ PAP \mid pulm\ emb) \cdot p(pulm\ emb)$$

From the previous derivations, it is readily seen that both the numerator $\Pr(normal\ Shunt \wedge high\ PAP)$ and the denominator $\Pr(high\ PAP)$ of the probability of interest $\Pr(normal\ Shunt \mid high\ PAP)$ relate linearly to the conditional probability $p(high\ PAP \mid no\ pulm\ emb)$. The probability of interest therefore relates as a quotient of two linear functions to this conditional probability. The sensitivity of the probability of interest with regard to the conditional probability under study is now uniquely determined by the values of the constants a, b, c, and d. These values are computed from the assessments for the appropriate conditional probabilities in the network:

$$\begin{aligned} p(high\ PAP \mid pulm\ emb) &= 0.8 \\ p(high\ PAP \mid no\ pulm\ emb) &= 0.05 \\ \\ p(normal Shunt \mid pulm\ emb) &= 0.096 \\ p(normal Shunt \mid no\ pulm\ emb) &= 0.905 \\ \\ p(pulm\ emb) &= 0.01 \end{aligned}$$

We find that

$$\begin{aligned} a &= 0.896 \\ b &= 0.00076 \\ c &= 0.99 \\ d &= 0.008 \end{aligned}$$

The mathematical function relating the probability of interest $Pr(normal\ Shunt \mid high\ PAP)$ to the conditional probability $x = p(high\ PAP \mid no\ pulm\ emb)$ therefore equals

$$\Pr(normal\ Shunt \mid high\ PAP) = \frac{0.896 \cdot x + 0.00076}{0.99 \cdot x + 0.008}$$

The function is depicted in Figure 4.8. Note that the probability of interest shows a high sensitivity for the conditional probability under study at the specified assessment 0.05. □

So far, we have shown that a belief network's probability of interest relates as a quotient of two linear functions to a conditional probability under study. For a conditional probability that pertains to a node from the sensitivity set that does not have any observed descendants, this functional relation reduces to a *linear function*. The following proposition states this property more formally.

Proposition 4.4.3 *Let B be a Bayesian belief network and let* Pr *be its joint probability distribution. Let O be the set of observed nodes with the corresponding observations*

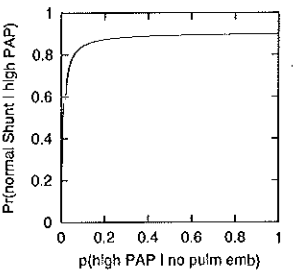


Figure 4.8: The function relating the probability of interest Pr(*normal Shunt* | *high PAP*) to the conditional probability under study p(*high PAP* | *no pulm emb*).

o, as before. Let V_r be the network's node of interest and let $Sen(V_r, O)$ be the sensitivity set for V_r given O. Let $V_s \in Sen(V_r, O)$ with $\sigma^*(V_s) \cap O = \emptyset$. Then, for any value v_r of V_r, we have that

$$\Pr(v_r \mid o) = a \cdot x + b$$

for every conditional probability $x = p(v_s \mid \pi')$ of V_s, where a and b are constants that are dependent upon the values v_s of V_s and π' of $\pi_G(V_s)$.

Proof (Sketch). The probability of interest $\Pr(v_r \mid o)$ for the belief network B once more equals

$$\Pr(v_r \mid o) = \frac{\Pr(v_r \wedge o)}{\Pr(o)}$$

From the proof of Proposition 4.4.1, we have that the numerator $\Pr(v_r \wedge o)$ in this equation relates linearly to the conditional probability x under study. Now, with regard to the probability $\Pr(o)$, we recall from the previous section that, if no observations are available for descendants of a non-ancestor, the probability of interest is algebraically independent of the conditional probabilities of this node. Likewise, the probability of a combination of observations is algebraically independent of the conditional probabilities of any non-ancestor without observed descendants. From this property, we have that the probability $\Pr(o)$ is algebraically independent of the conditional probability x under study. We conclude that $\Pr(o)$ is a constant with respect to x. □

We illustrate the property stated in the previous proposition by means of an example.

Example 4.4.4 We consider again the belief network from Figure 4.7. Once more, we address the situation where the value *high* has been observed for the node *PAP*, and consider the probability of interest Pr(*normal Shunt* | *high PAP*). As mentioned in Example 4.4.2, the sensitivity set $Sen(Shunt, \{PAP\})$ comprises all three nodes from

the network. We now investigate the functional relation between the probability of interest and the conditional probability $p(normal\ Shunt \mid pulm\ emb)$ for the node $Shunt \in Sen(Shunt, \{PAP\})$. Note that the node *Shunt* does not have any observed descendants. For our probability of interest, we once again find that

$$\Pr(normal\ Shunt \mid high\ PAP) = \frac{\Pr(normal\ Shunt \wedge high\ PAP)}{\Pr(high\ PAP)}$$

The numerator in this equation equals

$$\begin{aligned}
&\Pr(normal\ Shunt \wedge high\ PAP) = \\
&= \sum_{\{Pulm\ emb\}} p(high\ PAP \mid Pulm\ emb) \cdot p(normal\ Shunt \mid Pulm\ emb) \cdot p(Pulm\ emb) \\
&= p(high\ PAP \mid pulm\ emb) \cdot p(normal\ Shunt \mid pulm\ emb) \cdot p(pulm\ emb) + \\
&\quad + p(high\ PAP \mid no\ pulm\ emb) \cdot p(normal\ Shunt \mid no\ pulm\ emb) \cdot p(no\ pulm\ emb) \\
&= a' \cdot x + b'
\end{aligned}$$

where a' equals

$$a' = p(high\ PAP \mid pulm\ emb) \cdot p(pulm\ emb)$$

and b' equals

$$b' = p(high\ PAP \mid no\ pulm\ emb) \cdot p(normal\ Shunt \mid no\ pulm\ emb) \cdot p(no\ pulm\ emb)$$

The denominator of the probability of interest equals

$$\begin{aligned}
&\Pr(high\ PAP) = \\
&= \sum_{\substack{\{Shunt,\\ Pulm\ emb\}}} p(high\ PAP \mid Pulm\ emb) \cdot p(Shunt \mid Pulm\ emb) \cdot p(Pulm\ emb) \\
&= \sum_{\{Pulm\ emb\}} p(high\ PAP \mid Pulm\ emb) \cdot p(Pulm\ emb) \\
&= p(high\ PAP \mid pulm\ emb) \cdot p(pulm\ emb) + \\
&\quad + p(high\ PAP \mid no\ pulm\ emb) \cdot p(no\ pulm\ emb) = \\
&= c'
\end{aligned}$$

The previous derivations show that the denominator $\Pr(\textit{high PAP})$ of the probability of interest does not depend on the conditional probability under study $p(\textit{normal Shunt} \mid \textit{pulm emb})$. The numerator $\Pr(\textit{normal Shunt} \wedge \textit{high PAP})$ relates linearly to this conditional probability. We conclude that our probability of interest relates linearly to the conditional probability under study:

$$\Pr(\textit{normal Shunt} \mid \textit{high PAP}) = \frac{a' \cdot x + b'}{c'} = a \cdot x + b$$

The sensitivity of the probability of interest with regard to the conditional probability under study is now uniquely determined by the values of the constants a and b. These values again are computed from the assessments for the appropriate conditional probabilities in the network, as specified in Example 4.4.2. We find that

$$a = 0.139$$
$$b = 0.779$$

The linear function relating the probability of interest $\Pr(\textit{normal Shunt} \mid \textit{high PAP})$ to the conditional probability $p(\textit{normal Shunt} \mid \textit{pulm emb})$, denoted by x, therefore equals

$$\Pr(\textit{normal Shunt} \mid \textit{high PAP}) = 0.139 \cdot x + 0.779$$

The function is depicted in Figure 4.9. □

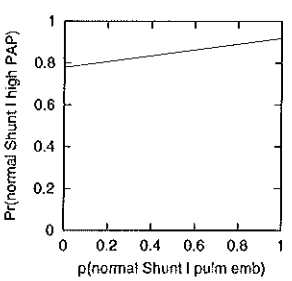


Figure 4.9: The function relating the probability of interest $\Pr(\textit{normal Shunt} \mid \textit{high PAP})$ to the conditional probability under study $p(\textit{normal Shunt} \mid \textit{pulm emb})$.

We would like to note that, in the special case where none of the nodes in a Bayesian belief network are observed, Proposition 4.4.3 implies that the network's probability of interest relates linearly to *every* conditional probability of *every* node from the sensitivity set under study.

Corollary 4.4.5 *Let B be a Bayesian belief network and let* Pr *be its joint probability distribution. Let V_r be the network's node of interest and let $Sen(V_r, \varnothing)$ be the sensitivity set for V_r given the empty set of observed nodes. Let $V_s \in Sen(V_r, \varnothing)$. Then, for any value v_r of V_r, we have that*

$$\Pr(v_r \mid o) = a \cdot x + b$$

for every conditional probability $x = p(v_s \mid \pi')$ *of* V_s*, where* a *and* b *are constants that are dependent upon the values* v_s *of* V_s *and* π' *of* $\pi_G(V_s)$.

In the foregoing, we have argued that a belief network's probability of interest relates to a conditional probability under study by a simple mathematical function. Knowledge of this function allows for considerably reducing the computational burden of a one-way sensitivity analysis of a Bayesian belief network as only the constants in the function need be known. These constants can be determined by computing the probability of interest from the network for a small number of values for a conditional probability under study and solving the resulting system of equations; systematic variation of the conditional probability is then no longer necessary. For a conditional probability that is related linearly to the probability of interest, two network computations suffice; for all other conditional probabilities, three network computations are required. The following example illustrates the basic idea.

Example 4.4.6 We consider again the belief network from Figure 4.7. As in Example 4.4.2, we investigate the functional relation between the probability of interest Pr(*normal Shunt* | *high PAP*) and the conditional probability $x = p(\textit{high PAP} \mid \textit{no pulm emb})$. We recall that this function equals

$$\Pr(\textit{normal Shunt} \mid \textit{high PAP}) = \frac{a \cdot x + b}{c \cdot x + d}$$

In Example 4.4.2, we have determined the values of the constants a, b, c, and d by expressing every constant in terms of conditional probabilities from the network and subsequently filling in the appropriate assessments. The functional relation can be determined more efficiently, however, by computing the probability of interest from the network for three different values of the conditional probability under study. Note that three network computations suffice since the constant c can be eliminated from the above equation, yielding

$$\Pr(\textit{normal Shunt} \mid \textit{high PAP}) = \frac{a' \cdot x + b'}{x + c'}$$

Using the three values $x = 0.2$, $x = 0.4$, and $x = 0.6$ for the conditional probability under study and the assessments for the other conditional probabilities as specified in Example 4.4.2, we find by computing the probability of interest Pr(*normal Shunt* | *high PAP*) from the network, the values

$$\begin{aligned}
\Pr(\textit{normal Shunt} \mid \textit{high PAP})_{x=0.2} &= 0.87356 \\
\Pr(\textit{normal Shunt} \mid \textit{high PAP})_{x=0.4} &= 0.88897 \\
\Pr(\textit{normal Shunt} \mid \textit{high PAP})_{x=0.6} &= 0.89424
\end{aligned}$$

From these values, we now obtain the three linear equations

$$\frac{0.2 \cdot a' + b'}{0.2 + c'} = 0.87356 \quad \Rightarrow \quad 0.2 \cdot a' + b' - 0.87356 \cdot c' - 0.2 \cdot 0.87356 = 0$$

$$\frac{0.4 \cdot a' + b'}{0.4 + c'} = 0.88897 \quad \Rightarrow \quad 0.4 \cdot a' + b' - 0.88897 \cdot c' - 0.4 \cdot 0.88897 = 0$$

$$\frac{0.6 \cdot a' + b'}{0.6 + c'} = 0.89424 \quad \Rightarrow \quad 0.6 \cdot a' + b' - 0.89424 \cdot c' - 0.6 \cdot 0.89424 = 0$$

Solving this system of linear equations gives

$$a' = 0.905$$
$$b' = 0.00061$$
$$c' = 0.00789$$

It is readily verified, by dividing the values of the constants a, b, and d specified in Example 4.4.2 by the value of the constant c, that the mathematical function yielded coincides with the function found in Example 4.4.2. □

4.5 Experimental results

In the previous sections, we have detailed various properties that allow for reducing the computational burden of a one-way sensitivity analysis of a Bayesian belief network. In Section 4.3, we have argued that a belief network's probability of interest is algebraically independent of the conditional probabilities of any node that is not included in the sensitivity set under study. As sensitivity analyses with respect to these conditional probabilities are uninformative, they can be excluded from the overall analysis. In Section 4.4, we have argued that for any conditional probability, that pertains to a node that is included in the sensitivity set, a small number of network computations suffice to determine the sensitivity of the probability of interest with regard to a conditional probability under study. Systematic variation of conditional probabilities then is no longer necessary. Now, to gain insight into the effect of exploiting these properties, we have conducted several experiments on randomly generated Bayesian belief networks. In these experiments, we have investigated, for various different sets of networks, the number of nodes in the sensitivity set under study and the number of nodes whose conditional probabilities are related linearly to the probability of interest, as these numbers reflect the computational burden of a network's sensitivity analysis.

In each experiment, we have generated a set of one thousand connected acyclic digraphs; for details of the graph-generator used, we refer the reader to [Van der Gaag, 1994]. We have generated various sets of digraphs with fifty nodes each, comprising fifty, seventy five, one hundred, one hundred and fifty, two hundred, and two hundred and fifty arcs, respectively. As our investigations are concerned with the digraph of a

Bayesian belief network only, we have refrained from quantifying the generated digraphs with conditional probabilities. For each digraph from every set, we have randomly selected a single node of interest and k observed nodes, where, for the various sets of digraphs, k is varied from zero to thirty by steps of two nodes.

To study the behaviour of our method on Bayesian belief networks that have been developed for different types of application, we have also generated various sets of digraphs for which a *diagnostic* and a *prognostic* bias, respectively, have been used in the selection of the node of interest and of the observed nodes. For diagnostic applications, we have assumed that a belief network's node of interest tends to be located in the upper part of the digraph, whereas the observed nodes are likely to be situated in its lower part. For prognostic applications, on the other hand, we have assumed that the node of interest tends to be located in the lower part of the digraph and the observed nodes in the upper part. The two biases have been realised as a two-stage selection. The selection of a node of interest in the lower part of a digraph, for example, starts with selecting a single auxiliary node in a random fashion. The node of interest is then selected from among the nodes that are assigned a lower number in a topological ordering of the digraph than the auxiliary node. For computational reasons, the maximum number of observed nodes considered with the diagnostic and prognostic biases, respectively, has been limited to sixteen nodes.

In each experiment, we have determined, for every digraph, the number of nodes in the sensitivity set for the selected node of interest given the set of observed nodes under study, and the number of nodes whose conditional probabilities are related linearly to the probability of interest. The results are summarised in Figure 4.10. Figure 4.10(a) and Figure 4.10(b) pertain to the digraphs for which no bias has been used in the selection of the node of interest and of the observed nodes. Figure 4.10(a) shows the average number of nodes in the sensitivity set, plotted against the number of observed nodes; the six curves pertain to the sets of digraphs with different numbers of arcs. Figure 4.10(b) shows the average number of nodes, from the sensitivity set, whose conditional probabilities are related linearly to the probability of interest. Figure 4.10(c) and Figure 4.10(d) depict the same information for the digraphs for which a diagnostic bias has been used in the selection of the node of interest and of the observed nodes; Figure 4.10(e) and Figure 4.10(f) show the information for the digraphs for which a prognostic bias has been used.

To discuss the results obtained from our experiments, we start by considering the average number of nodes in the sensitivity set for digraphs for which the node of interest and the set of observed nodes have been selected randomly. From Figure 4.10(a), we see that the average number of nodes in the sensitivity set increases at first, with an increasing number of observed nodes. This property is readily explained by observing that, initially, only observations for ancestors of the node of interest, that is, only observations for nodes from the sensitivity set, allow for diminishing the set's

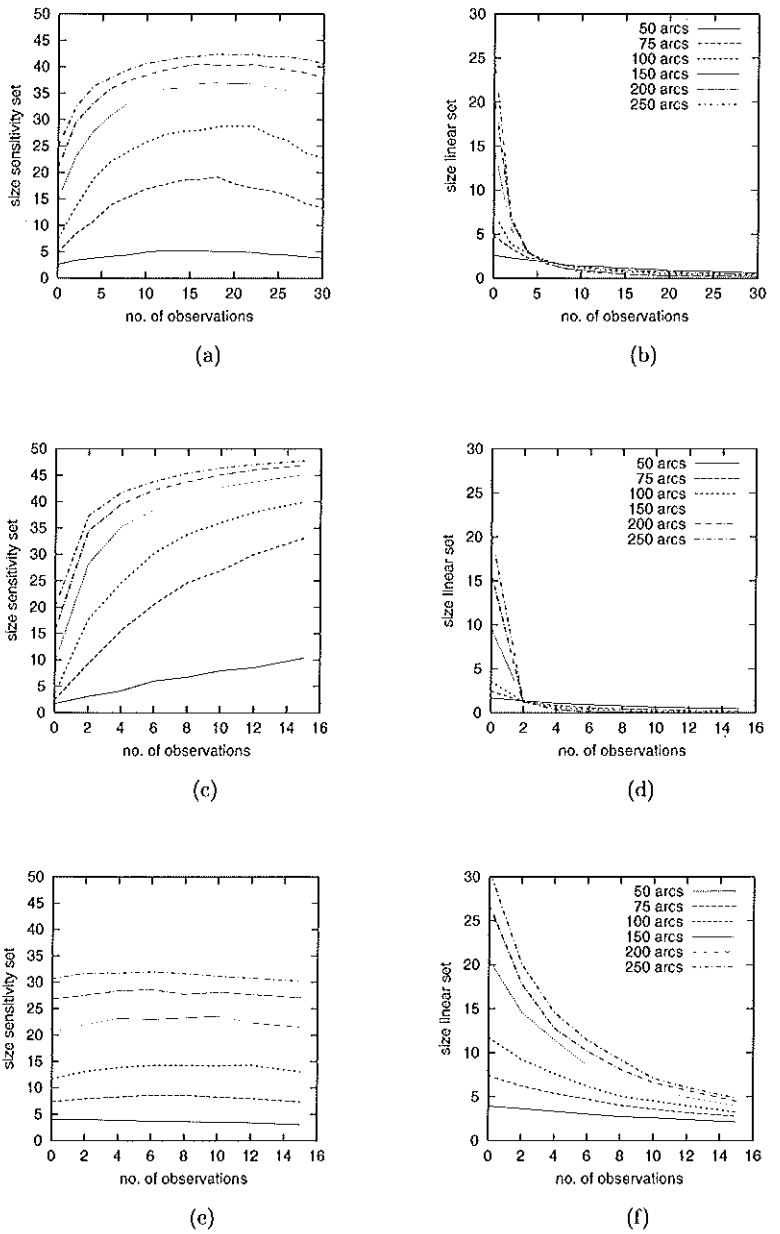


Figure 4.10: The average number of nodes in the sensitivity set under study and the average number of linearly related nodes, for various sets of networks without any bias (figures (a) and (b)), with a diagnostic bias (figures (c) and (d)), and with a prognostic bias (figures (e) and (f)), respectively.

size. For all other nodes in the digraph, an observation will either have no effect or increase the size of the sensitivity set. In the digraphs under consideration, the node of interest will, on average, be located in 'the middle' of the digraph. The number of ancestors of this node will, on average, be smaller than its number of non-ancestors. The tendency of additional observations for the ancestors of the node of interest to decrease the size of the sensitivity set will therefore be outweighed by the tendency of additional observations for its non-ancestors to increase the sensitivity set's size. Now, for a still further increasing number of observed nodes, the increase in size of the sensitivity set diminishes. In fact, when roughly eighteen observed nodes have been selected, additional observations cause the sensitivity set to decrease in size. This property is explained by observing that a new node can only be inserted into the sensitivity set if one of its descendants is selected as an observed node where it had no observed descendants before. The more observed nodes have been selected, however, the fewer nodes remain without observed descendants. On the other hand, additional observations for nodes from the sensitivity set will serve to decrease the set's size. For larger numbers of observed nodes, the sensitivity set will be quite large and the latter tendency will therefore outweigh the former, resulting in an overall decrease in the size of the sensitivity set. Figure 4.10(a) further reveals that a larger number of arcs in a belief network's digraph will result in a larger sensitivity set. This property is explained by observing that, in a digraph with more arcs, the node of interest is likely to have more ancestors, resulting in a larger sensitivity set to begin with. Moreover, a larger number of arcs will, on average, result in a larger number of chains between a node under consideration and the node of interest. To block the influence of this node's conditional probabilities on the probability of interest, that is, to exclude the node from the sensitivity set, on average, a larger number of observations is required. For a fixed number of observed nodes, therefore, an increase in the number of arcs leads to an increase in size of the sensitivity set.

We now consider the average number of nodes, from a sensitivity set under study, whose conditional probabilities are related linearly to the selected probability of interest. From Figure 4.10(b), we see that this number diminishes with an increasing number of observed nodes. This property is readily explained by observing that only the conditional probabilities of ancestors of the node of interest that do not have any observed descendants, are related linearly to the probability of interest. The more observed nodes have been selected, the fewer ancestors of the node of interest remain without observed descendants and, hence, the smaller the number of nodes whose conditional probabilities are related linearly to the probability of interest. Figure 4.10(b) further shows that, for a fixed number of observed nodes, the number of linearly related nodes increases with an increasing number of arcs, which conforms with the tendency of the number of ancestors of the node of interest to increase with the number of arcs.

We proceed with addressing the results from our experiments with digraphs for

which a diagnostic bias has been used in the selection of the node of interest and of the set of observed nodes. From Figure 4.10(c) and Figure 4.10(d), we see that these digraphs show tendencies similar to those shown by digraphs for which no bias has been used. The initial increase in the size of the sensitivity set with an increasing number of observed nodes, however, is stronger and reaches a higher maximum for the digraphs with a diagnostic bias than for the digraphs for which no bias has been used. This property is readily explained by once more observing that, initially, only observations for ancestors of the node of interest allow for diminishing the size of the sensitivity set. Since the node of interest in digraphs with a diagnostic bias is, on average, situated higher in the digraph than in unbiased digraphs, its ratio of the number of ancestors to the number of non-ancestors will, on average, be smaller. As a result, the tendency of additional observations for non-ancestors to increase the size of the sensitivity set is even more dominant in digraphs for which a diagnostic bias has been used than in digraphs without any bias. The smaller number of ancestors further accounts for the stronger decrease of the number of nodes from the sensitivity set whose conditional probabilities are related linearly to the probability of interest, as revealed in Figure 4.10(d).

We now consider the results from our experiments with digraphs for which a prognostic bias has been used in the selection of the node of interest and of the set of observed nodes. Figure 4.10(e) suggests that the size of the sensitivity set for these digraphs remains reasonably constant with an increasing number of observed nodes. The tendency of additional observations for the ancestors of the node of interest to decrease the size of the sensitivity set is therefore balanced, in these digraphs, by the tendency of observations for its non-ancestors to increase the sensitivity set's size. We feel that this property is the coincidental result of the 'degree' of prognostic bias we have used. We expect that a more extreme location of the node of interest and of the observed nodes in the digraphs under study, that is, a larger ratio of the number of ancestors to the number of non-ancestors, will lead to a decrease in the size of the sensitivity set with an increasing number of observations. In fact, further experiments, using a three-stage selection for a prognostic bias, have met this expectation. Figure 4.10(e) further shows that the size of the sensitivity set increases with an increasing number of arcs, as we have seen before for the digraphs without any bias as well as for the digraphs for which a diagnostic bias has been used. Similar tendencies as for unbiased and for diagnostic digraphs are also seen in Figure 4.10(f) with respect to the number of nodes from the sensitivity set whose conditional probabilities are related linearly to the probability of interest.

4.6 Previous work

Sensitivity analysis is a general, well-known technique for studying the effects of the inaccuracies in the parameters of a mathematical model on the model's output; it is widely used in mathematical modelling in various different domains of application [Morgan & Henrion, 1990, Habbema *et al.*, 1990, Dippel *et al.*, 1992, Helton, 1993, Doubilet *et al.*, 1985]. As more and more Bayesian belief networks are being developed for real-life applications, interest in sensitivity analysis of belief networks is increasing. In this section, we review previous work on sensitivity analysis of belief networks. In doing so, we do not intend to give an exhaustive overview of the state of the art. We merely discuss the results from related work and compare it with the results that we have presented in this chapter.

In her work on sensitivity analysis of Bayesian belief networks, K. Blackmond Laskey has been motivated, as in fact we have been, by the observation that straightforward, systematic variation of the assessments of a network's conditional probabilities is too much time-consuming to be of practical use. She has developed an efficient method for analytically computing first-order approximations of exact analyses [Laskey, 1995]. Her method sets out by identifying, in a belief network under study, conditional probabilities that upon variation have no effect on a probability of interest. Laskey suggests two procedures for this purpose. She suggests that the assessment of every single conditional probability be varied over a small number of values, serving to reveal all conditional probabilities that have no influence on the probability of interest. For an alternative procedure, she observes that some uninfluential probabilities can be identified using graphical considerations. For this purpose, she introduces a concept similar to our sensitivity set; in fact, our notion of sensitivity set has been inspired to a large extent by her concept. Laskey's method excludes the identified uninfluential conditional probabilities from further analysis. For the remaining conditional probabilities, the effect of variation on the network's probability of interest is measured by a so-called sensitivity value. A sensitivity value is the partial derivative of the probability of interest with respect to a conditional probability under study. A sensitivity value thus provides an approximation of the effect of small deviations from the probability's assessment on the probability of interest. Laskey presents two procedures for analytically computing sensitivity values; these procedures build upon the propagation algorithm by Lauritzen and Spiegelhalter and upon Monte Carlo sampling, respectively. Compared to straightforward variation of conditional probabilities in a sensitivity analysis, Laskey's method requires considerably less computational effort.

In her method, Laskey has introduced a powerful concept upon which we have built our concept of sensitivity set. She suggests, as we do, to construct, from a belief network's digraph, an auxiliary digraph in which a predecessor X_i is added to every node V_i. She proceeds by observing that, if the auxiliary predecessor X_i of

a node V_i is d-separated from the *auxiliary predecessor* X_r of the node of interest V_r, then sensitivity analyses with respect to the conditional probabilities for node V_i are uninformative as these probabilities cannot influence the probability of interest upon variation. Her observation, unfortunately, is incorrect, as it can declare several conditional probabilities to be not influential while in fact they are. For example, from her observation, we would conclude that, if no observations are available as yet, the conditional probabilities of all ancestors of the node of interest V_r are uninfluential, since $\langle\{X_i\} \mid \varnothing \mid \{X_r\}\rangle^d$ for all $V_i \in \pi^*(V_r) \setminus \{V_r\}$. To show that this conclusion is incorrect, we give an example from the ALARM-network. We are interested in the probability that the node *LV failure* takes the value *true* when no observations are available as yet. For the probability of interest, we have that

$$\Pr(\textit{fail}) = p(\textit{fail} \mid \textit{history}) \cdot p(\textit{history}) + p(\textit{fail} \mid \textit{no history}) \cdot p(\textit{no history})$$

which reveals that the probability of interest is algebraically dependent upon the probabilities of the nodes *History* and *LV failure*. Since, in the absence of observations, the auxiliary predecessor of the node *History* is d-separated from the auxiliary predecessor of *LV failure*, building upon Laskey's observation would incorrectly declare the prior probabilities of the node *History* to be uninfluential. With the various lemmas presented in Section 4.3 of this chapter, we have shown that our concept of sensitivity set provides for correctly identifying uninfluential nodes.

As mentioned before, Laskey's method of computing sensitivity values requires considerably less computational effort than straightforward variation of probability assessments for studying sensitivity. The method, however, provides insight in the effect of small deviations from a probability's assessment only: as Laskey indicates, when larger deviations are considered, the quality of the approximation may break down rapidly. For the ALARM-network, Figure 4.8 illustrates how an approximation may fail to reveal the extent of the sensitivity of a probability of interest to a conditional probability under study. The figure shows the effect of variation of the assessment for the conditional probability $p(\textit{high PAP} \mid \textit{no pulm emb})$ on the probability of interest $\Pr(\textit{normal Shunt} \mid \textit{high PAP})$. The assessment specified for the conditional probability under consideration is 0.05. For variation of this assessment to higher values, the derivative of the sensitivity function does not change rapidly. The derivative at the specified assessment therefore provides a good approximation of the effect on the probability of interest for larger values. However, even a slight shift in the specified assessment to a smaller value has a very large effect on the derivative of the sensitivity function. The approximation therefore does not suffice. We feel that exact sensitivity analysis of a Bayesian belief network is to be preferred to approximate analysis.

We briefly review two other methods for sensitivity analysis of Bayesian belief networks that take a different approach than our method. In [Chang & Fung, 1995] and [Castillo *et al.*, 1997b], the idea of symbolic propagation in belief networks is exploited

for studying sensitivity. Instead of yielding a single number as the standard propagation algorithms do, a symbolic propagation algorithm yields an algebraic expression for a network's probability of interest in terms of all conditional probabilities in the network. From this expression, the sensitivity of the probability of interest to a conditional probability under study is readily computed, basically by filling in the specified assessments for all other conditional probabilities. A disadvantage of building upon symbolic propagation is that it is quite time-consuming. We therefore feel that methods for sensitivity analysis that build upon the faster standard propagation algorithms are preferred. In [Spiegelhalter, 1989], a method for sensitivity analysis of Bayesian belief networks is presented that builds upon an explicit specification of the inaccuracies in a network's conditional probabilities. As in our method, an auxiliary graph is constructed from the digraph of a belief network by adding an auxiliary predecessor to every node. The auxiliary predecessor now captures second-order distributions for the conditional probabilities of its successor. Using standard propagation algorithms, the effects of the specified inaccuracies on a probability of interest are readily computed. A disadvantage of this method is that it requires an explicit specification of the inaccuracies in a belief network's probability assessments. As second-order distributions for the specified assessments often are not available, assumptions on the nature of the inaccuracies have to be made that may not be realistic. We would like to note that with our method for studying sensitivity no assumptions with regard to the inaccuracies involved are necessary.

While in this chapter we have focused on *sensitivity analysis* of Bayesian belief networks, we would like to note that the reliability of a belief network's output can in addition be studied by subjecting the network to an *uncertainty analysis*. In an uncertainty analysis of a belief network, the assessments of *all* conditional probabilities of the network's quantitative part are varied simultaneously. To this end, for each conditional probability, values are drawn from some probability distribution. Uncertainty analysis of a Bayesian belief network serves to reveal the overall reliability of the network's output. Uncertainty analysis, however, yields less insight into the effect of single conditional probabilities than sensitivity analysis does. Previous experiments with uncertainty analysis of Bayesian belief networks have led to the suggestion that belief networks are highly insensitive to inaccuracies in the assessments of their conditional probabilities [Henrion *et al.*, 1996, Pradhan *et al.*, 1996]. In these experiments, performed on belief networks for diagnostic applications, a measure of the reliability of a network's diagnosis is obtained by assuming a log-normal distribution for every conditional probability, having the initially specified assessment for its mean, and subsequently averaging over the probability of the true diagnosis for various diagnostic situations. Unfortunately, when using probability distributions to model inaccuracies in the assessments for a network's conditional probabilities, it is not the *average* of the probabilities of the true diagnosis that reflects the effects of these inaccuracies, but the

variation in these probabilities. In addition, we would like to note that the reported results are based on experience with a single belief network only, in which the conditional probability distributions have been simplified using noisy-OR and noisy-MAX assumptions. From the results reported so far for uncertainty analysis of Bayesian belief networks, therefore, no decisive conclusions can be drawn. We feel that the sensitivity of a network's probability of interest to the various conditional probabilities involved will vary from application to application. In fact, the results of the sensitivity analyses of a Bayesian belief network for congenital heart disease, as presented in Chapter 3, show that a network's conditional probabilities can have a large effect on a probability of interest.

4.7 Conclusions

The assessments obtained for the various conditional probabilities of a Bayesian belief network are inevitably inaccurate, due to incompleteness of data and partial knowledge of the problem under study. The inaccuracies in these probability assessments may severely compromise the reliability of the network's output. To gain insight into the reliability of a probability of interest computed from a belief network, the network can be subjected to a sensitivity analysis. A sensitivity analysis can be performed by systematically varying the assessments for one or more of the network's conditional probabilities simultaneously. We have argued that even for a rather small belief network such a straightforwardly performed analysis is highly time-consuming. In this chapter, we have shown that, by qualitative considerations pertaining to a belief network's digraph, various conditional probabilities can be identified that upon variation cannot influence the network's probability of interest. Analyses with respect to these probabilities are uninformative and can therefore be excluded from the overall analysis. More specifically, we have shown that a sensitivity analysis of a Bayesian belief network can be restricted to the conditional probabilities of the nodes from the sensitivity set for the network's node of interest. Excluding uninformative analyses can lead to a considerable reduction in the computational burden of a sensitivity analysis, as is evidenced by the results from the experiments we have performed on randomly generated belief networks. We have further shown that for sensitivity analyses that are informative, simple mathematical functions exist expressing the network's probability of interest in terms of the conditional probabilities under study. Knowledge of these functions allows for even further reduction of the computational burden of a sensitivity analysis, as only the constants in the functions need be determined, rendering systematic variation of conditional probabilities unnecessary.

In this chapter, we have focused attention on a one-way sensitivity analysis of a Bayesian belief network in which the network's conditional probabilities are investi-

gated one at a time. For such an analysis, we have detailed the mathematical function expressing the network's probability of interest in terms of a single conditional probability. More specifically, we have shown that, in general, the probability of interest relates as a quotient of two linear functions to a conditional probability under study. In essence, it is also possible to investigate the effect of simultaneous variation of two or more conditional probabilities. Such a higher-order sensitivity analysis can, just as a one-way analysis, be restricted to the conditional probabilities of the nodes that are included in the sensitivity set for a belief network's node of interest. Moreover, for higher-order sensitivity analyses also simple mathematical functions exist between a network's probability of interest and the conditional probabilities under study. Although not reported in this chapter, we have detailed the functions that hold in a two-way sensitivity analysis in which conditional probabilities are studied pairwise. These functions comprise terms for the separate effects of each of the two conditional probabilities being investigated as well as terms for their joint effect. More specifically, the probability of interest, in general, relates as a quotient of two bi-linear functions to the probabilities under study. The more conditional probabilities of a belief network are investigated simultaneously, the more involved the mathematical functions will be. We feel that the results of higher-order sensitivity analyses in which three or more conditional probabilities are studied simultaneously will in general be very hard to interpret.

In the near future, we envision further experiments with our method of sensitivity analysis on real-life Bayesian belief networks. In these experiments, we would like to study the reliability of belief network's output in general. Also, we would like to evaluate in more detail the effect of the location of the node of interest and of the observed nodes in a network's digraph. In addition, we envision further investigation of the properties of sensitivity analysis, both from a theoretical and an experimental point of view. Our experiments so far on randomly generated belief networks and on the ALARM-network have shown considerable computational savings. Motivated by these initial results, we hope to be able to arrive at a generally applicable, practicable method for sensitivity analysis of Bayesian belief networks.

Appendix

In the Sections 4.3 and 4.4 of this chapter we have presented various properties of sensitivity analysis of Bayesian belief networks. In Section 4.3, we have introduced the concept of a sensitivity set for a network's node of interest given available observations. We have shown that the conditional probabilities of the nodes that are not included in a sensitivity set under consideration upon variation cannot influence the probability of interest. For the nodes that are included in the sensitivity set, we have shown in

Section 4.4 that the probability of interest relates to the conditional probabilities of these nodes as a quotient of two linear functions. So far, we have presented these properties with short, intuitive proofs. In this appendix, we provide full proofs for the various different properties.

In order to prove that a belief network's probability of interest for a node V_r is algebraically independent of the conditional probabilities of any node that is not included in a sensitivity set $Sen(V_r, O)$ under consideration, we have partitioned, in Definition 4.3.3, the set of remaining nodes into the three sets $Insen_1(V_r, O)$, $Insen_2(V_r, O)$, and $Insen_3(V_r, O)$. In the following lemma, we show that these three sets and the sensitivity set are mutually exclusive and collectively exhaustive.

Lemma A.1 *Let B be a Bayesian belief network with the digraph $G = (V(G), A(G))$. Let $V_r \in V(G)$ be the network's node of interest and let $O \subseteq V(G)$ be the set of observed nodes in G. Let $Sen(V_r, O)$ be the sensitivity set for V_r given O and let $Insen_1(V_r, O)$, $Insen_2(V_r, O)$, $Insen_3(V_r, O)$, and $Sen(V_r, O)$ be defined as in Definition 4.3.3. Then,*

- $Insen_i(V_r, O) \cap Insen_j(V_r, O) = \varnothing$, *for all* $i, j = 1, 2, 3$ *with* $i \neq j$;
- $V(G) \setminus Sen(V_r, O) = \bigcup_{i=1,2,3} Insen_i(V_r, O)$.

Proof. From Definition 4.3.3, it is readily seen that the sets $Insen_1(V_r, O)$, $Insen_2(V_r, O)$, and $Insen_3(V_r, O)$ are mutually exclusive. In our proof, we therefore focus on the second property stated in the lemma.

To prove that $V(G) \setminus Sen(V_r, O) = \bigcup_{i=1,2,3} Insen_i(V_r, O)$, we have to show that any node that is included in one of the sets $Insen_i(V_r, O)$, is not included in $Sen(V_r, O)$, and vice versa. To show that a node V_j is not included in the set $Sen(V_r, O)$, we construct from the belief network's digraph G the auxiliary digraph G^* as defined in Definition 4.3.1 and show that in G^* any chain from V_j's auxiliary predecessor X_j to the node of interest V_r is blocked by O. We now begin by showing that $\bigcup_{i=1,2,3} Insen_i(V_r, O) \subseteq V(G) \setminus Sen(V_r, O)$:

- We assume that $Insen_1(V_r, O) \cup Insen_2(V_r, O) \neq \varnothing$ and consider a node $V_j \in Insen_1(V_r, O) \cup Insen_2(V_r, O)$. We observe that, in the digraph G, any chain from this node V_j to node V_r includes either a predecessor or a successor of V_j; in the auxiliary digraph G^*, therefore, any chain from V_j's auxiliary predecessor X_j to node V_r equally includes either an(other) predecessor or a successor of V_j. Now, for node V_j, we have by definition that $\langle(\{V_j\} \cup \pi_G(V_j)) \mid O \mid \{V_r\}\rangle_G^d$. From $\langle(\{V_j\} \cup \pi_G(V_j)) \mid O \mid \{V_r\}\rangle_G^d$, we have that, in the digraph G, any chain $V_j \rightarrow \cdots V_r$ from V_j to V_r that includes a successor of V_j, is blocked by O. In the auxiliary digraph G^*, therefore, any chain $X_j \rightarrow V_j \rightarrow \cdots V_r$ from X_j to V_r that includes a successor of V_j, is blocked by O. From $\langle(\{V_j\} \cup \pi_G(V_j)) \mid O \mid \{V_r\}\rangle_G^d$, we further have that, in G, any chain $V_k \cdots V_r$ from a node $V_k \in \pi_G(V_j)$ to node

V_r is blocked by O. In G^*, therefore, any chain $X_j \rightarrow V_j \leftarrow V_k \cdots V_r$ from X_j to V_r that includes a node $V_k \in \pi_G(V_j)$, is blocked by O. We conclude that $\langle\{X_j\} \mid O \mid \{V_r\}\rangle^d_{G^*}$. By definition, we have that $V_j \in V(G) \setminus Sen(V_r, O)$.

- We assume that $Insen_3(V_r, O) \neq \varnothing$ and consider a node $V_j \in Insen_3(V_r, O)$. For this node V_j, we have by definition that $V_j \in V(G) \setminus \pi^*_G(V_r)$ and $\sigma^*_G(V_j) \cap O = \varnothing$. From these properties, we have that, in the digraph G, any chain from node V_j to node V_r either includes a predecessor V_k of V_j or includes a descendant $V_m \in \sigma^*_G(V_j)$ with two incoming arcs for which $\sigma^*_G(V_m) \cap O = \varnothing$. In the auxiliary digraph G^*, any chain $X_j \rightarrow V_j \leftarrow V_k \cdots V_r$ from V_j's auxiliary predecessor X_j to node V_r that includes an(other) predecessor V_k of V_j, is blocked by O because $\sigma^*_G(V_j) \cap O = \varnothing$. Furthermore, in G^*, any chain $X_j \rightarrow V_j \rightarrow \cdots \rightarrow V_m \leftarrow \cdots V_r$ from X_j to V_r is blocked by O because $\sigma^*_G(V_m) \cap O = \varnothing$. We conclude that $\langle\{X_j\} \mid O \mid \{V_r\}\rangle^d_{G^*}$. By definition, we have that $V_j \in V(G) \setminus Sen(V_r, O)$.

From the previous observations, we conclude that $\bigcup_{i=1,2,3} Insen_i(V_r, O) \subseteq V(G) \setminus Sen(V_r, O)$; note that the property trivially holds for the case where $Insen_i(V_r, O) = \varnothing$, $i = 1, 2, 3$.

We proceed by showing that $V(G) \setminus Sen(V_r, O) \subseteq \bigcup_{i=1,2,3} Insen_i(V_r, O)$. We assume that $V(G) \setminus Sen(V_r, O) \neq \varnothing$; the property trivially holds otherwise. We now consider a node $V_j \in V(G) \setminus Sen(V_r, O)$. For this node V_j, we have by definition that $\langle\{X_j\} \mid O \mid \{V_r\}\rangle^d_{G^*}$. We distinguish between two cases, the case where $V_j \in \pi^*_G(V_r)$ and the case where $V_j \notin \pi^*_G(V_r)$:

- We assume that $V_j \in \pi^*_G(V_r)$. From $\langle\{X_j\} \mid O \mid \{V_r\}\rangle^d_{G^*}$, we have that, in the auxiliary digraph G^*, any chain $X_j \rightarrow V_j \rightarrow \cdots V_r$ from X_j to V_r that includes a successor of V_j, is blocked by O. We conclude from this observation that, in the digraph G, any chain $V_j \rightarrow \cdots V_r$ from V_j to V_r is blocked by O. Note that from $V_j \in \pi^*_G(V_r)$, we have that there exists at least one (directed) path $V_j \rightarrow \cdots \rightarrow V_r$ from V_j to V_r in G. From this path being blocked, we conclude that $\sigma^*_G(V_j) \cap O \neq \varnothing$. Now, from $\langle\{X_j\} \mid O \mid \{V_r\}\rangle^d_{G^*}$, we further observe that, in the digraph G^* any chain $X_j \rightarrow V_j \leftarrow V_k \cdots V_r$ from X_j to V_r, that includes an(other) predecessor V_k of V_j, is blocked by O. From $\sigma^*_G(V_j) \cap O \neq \varnothing$ and the previous observations, we have that, in G, any chain $V_k \cdots V_r$ from a predecessor V_k of V_j to V_r is blocked by O. We conclude that $\langle(\{V_j\} \cup \pi^*_G(V_j)) \mid O \mid \{V_r\}\rangle^*_G$ and, hence, that $V_j \in Insen_1(V_r, O)$.

- We assume that $V_j \notin \pi^*_G(V_r)$. We once more distinguish between two cases, the case where $\sigma^*_G(V_j) \cap O = \varnothing$ and the case where $\sigma^*_G(V_j) \cap O \neq \varnothing$:

  - We assume that $\sigma^*_G(V_j) \cap O = \varnothing$. By definition, we have that $V_j \in Insen_3(V_r, O)$.

– We assume that $\sigma^*_G(V_j) \cap O \neq \varnothing$. From $\langle\{X_j\} \mid O \mid \{V_r\}\rangle^d_{G^*}$, we have that, in the auxiliary digraph G^*, any chain $X_j \rightarrow V_j \rightarrow \cdots V_r$ from X_j to V_r that includes a successor of V_j, is blocked by O. We conclude from this observation that, in the digraph G, any chain $V_j \rightarrow \cdots V_r$ from V_j to V_r is blocked by O. From $\langle\{X_j\} \mid O \mid \{V_r\}\rangle^d_{G^*}$, we further observe that, in the auxiliary digraph G^*, any chain $X_j \rightarrow V_j \leftarrow V_k \cdots V_r$ from X_j to V_r that includes an(other) predecessor V_k of V_j, is blocked by O. From $\sigma^*_G(V_j) \cap O \neq \varnothing$ and the previous observations, we have that, in the digraph G, any chain $V_k \cdots V_r$ from a predecessor V_k of V_j to V_r is blocked by O. We conclude that $\langle(\{V_j\} \cup \pi^*_G(V_j)) \mid O \mid \{V_r\}\rangle^d_G$ and, hence, that $V_j \in \mathit{Insen}_2(V_r, O)$.

From the previous considerations we conclude that $V_j \in \bigcup_{i=1,2,3} \mathit{Insen}_i(V_r, O)$ and, hence, that $V(G) \setminus \mathit{Sen}(V_r, O) \subseteq \bigcup_{i=1,2,3} \mathit{Insen}_i(V_r, O)$.

From $V(G) \setminus \mathit{Sen}(V_r, O) \subseteq \bigcup_{i=1,2,3} \mathit{Insen}_i(V_r, O)$ and $\bigcup_{i=1,2,3} \mathit{Insen}_i(V_r, O) \subseteq V(G) \setminus \mathit{Sen}(V_r, O)$ we conclude that $V(G) \setminus \mathit{Sen}(V_r, O) = \bigcup_{i=1,2,3} \mathit{Insen}_i(V_r, O)$, as stated in the lemma. □

In Section 4.3, we have provided the three lemmas 4.3.5, 4.3.7, and 4.3.9, stating that the probability of interest of a Bayesian belief network is algebraically independent of the conditional probabilities of the nodes included in the sets $\mathit{Insen}_1(V_r, O)$, $\mathit{Insen}_2(V_r, O)$, and $\mathit{Insen}_3(V_r, O)$. We will provide formal proofs for these lemmas shortly. Before doing so, however, we introduce the concept of a *sensitivity ordering* of the nodes of a belief network's digraph that will be used throughout the proofs.

Definition A.2 *Let B be a Bayesian belief network with the digraph $G = (V(G), A(G))$ where $V(G) = \{V_1, \ldots, V_n\}$, $n \geq 1$. Let $V_r \in V(G)$ be the network's node of interest and let $O \subseteq V(G)$ be the set of observed nodes in G. Let $\mathit{Sen}(V_r, O)$ be the sensitivity set for V_r given O and let $\mathit{Insen}_1(V_r, O)$, $\mathit{Insen}_2(V_r, O)$, and $\mathit{Insen}_3(V_r, O)$ be defined as in Definition 4.3.3. Let $\iota : V(G) \longleftrightarrow \{1, \ldots, n\}$ be a total ordering on $V(G)$, such that*

- *for any two nodes $V_i, V_j \in V(G)$ with $V_i \rightarrow V_j \in A(G)$, we have $\iota(V_i) < \iota(V_j)$;*
- *for any two nodes $V_i \in \mathit{Insen}_1(V_r, O) \cup \mathit{Sen}(V_r, O)$, $V_j \in \mathit{Insen}_2(V_r, O)$, we have $\iota(V_i) < \iota(V_j)$;*
- *for any two nodes $V_i \in \mathit{Insen}_2(V_r, O)$, $V_j \in \mathit{Insen}_3(V_r, O)$, we have $\iota(V_i) < \iota(V_j)$.*

Then, ι is a sensitivity ordering *of G with respect to V_r and O.*

For any node of interest and any set of observed nodes, there exists a sensitivity ordering of a belief network's digraph. Any such sensitivity ordering is a topological ordering of the digraph at hand.

Lemma A.3 (cf. Lemma 4.3.5) *Let B be a Bayesian belief network with the digraph $G = (V(G), A(G))$. Let* Pr *be the joint probability distribution defined by B. Let $O \subseteq V(G)$ be the set of observed nodes in G and let o denote the corresponding observations. Let $V_r \in V(G)$ be the network's node of interest. Then, for any value v_r of V_r, we have that* $\Pr(v_r \mid o) \propto p(V_i \mid \pi_G(V_i))$ *for every node $V_i \in Insen_3(V_r, O)$.*

Proof. Let ι be a sensitivity ordering of G with respect to V_r and O. Without loss of generality, we assume that the nodes in G are indexed by their ordering number, that is, we assume that $\iota(V_i) = i$; we take $n \geq 1$ to be the number of nodes in G. From the rule of marginalisation, we have that the probability of interest $\Pr(v_r \mid o)$ can be written as

$$\Pr(v_r \mid o) = \frac{\sum_{\{V_1,\ldots,V_n\}\setminus(\{V_r\}\cup O)} \Pr((\{V_1,\ldots,V_n\} \setminus (\{V_r\} \cup O)) \wedge v_r \wedge o)}{\sum_{\{V_1,\ldots,V_n\}\setminus O} \Pr((\{V_1,\ldots,V_n\} \setminus O) \wedge o)}$$

In the above equation, we have used the notation $\sum_W$ to indicate summation over all possible values of the variables in the set W. In the following, we will also use the notation $|_{X=x}$; this notation is used to indicate that in the preceding formula the variables in the set X take the combination of values x. Now, using the property stated in Proposition 4.2.6 for the probability distribution Pr defined by the network, we find that

$$\Pr(v_r \mid o) = \frac{\sum_{\{V_1,\ldots,V_n\}\setminus(\{V_r\}\cup O)} \prod_{i=1,\ldots,n} p(V_i \mid \pi_G(V_i)) \Big|_{\substack{V_r = v_r \\ O = o}}}{\sum_{\{V_1,\ldots,V_n\}\setminus O} \prod_{j=1,\ldots,n} p(V_j \mid \pi_G(V_j)) \Big|_{O = o}}$$

From the definition of sensitivity ordering, we have that the nodes in the set $Insen_3(V_r, O)$ have the highest ordering numbers in the network's digraph; without loss of generality, we assume that $Insen_3(V_r, O)$ includes the nodes $V_{m+1}, \ldots, V_n$. Now,

$$\Pr(v_r \mid o) = \frac{\sum_{\substack{\{V_1,\ldots,V_n\}\setminus \\ (\{V_r\}\cup O)}} \left(\left(\prod_{i=m+1,\ldots,n} p(V_i \mid \pi_G(V_i))\right) \cdot \left(\prod_{i=1,\ldots,m} p(V_i \mid \pi_G(V_i))\right)\right)\Big|_{\substack{V_r = v_r \\ O = o}}}{\sum_{\{V_1,\ldots,V_n\}\setminus O} \left(\left(\prod_{j=m+1,\ldots,n} p(V_j \mid \pi_G(V_j))\right) \cdot \left(\prod_{j=1,\ldots,m} p(V_j \mid \pi_G(V_j))\right)\right)\Big|_{O = o}}$$

From Definition 4.3.3, we know that the set $Insen_3(V_r, O)$ does not include any nodes from the set $\{V_r\} \cup O$, that is, $(\{V_r\} \cup O) \cap \{V_{m+1}, \ldots, V_n\} = \varnothing$. Since our sensitivity ordering is a topological ordering, we further know that $(\bigcup_{i=1,\ldots,m} \pi_G(V_i)) \cap \{V_{m+1}, \ldots, V_n\} = \varnothing$. Using these observations, we find that

$$\Pr(v_r \mid o) = \frac{\displaystyle\sum_{\substack{\{V_1,\ldots,V_m\}\setminus \\ (\{V_r\}\cup O)}} \left(\left(\sum_{\{V_{m+1},\ldots,V_n\}} \prod_{i=m+1,\ldots,n} p(V_i \mid \pi_G(V_i)) \right) \cdot \prod_{i=1,\ldots,m} p(V_i \mid \pi_G(V_i)) \right) \Bigg|_{\substack{V_r = v_r \\ O = o}}}{\displaystyle\sum_{\{V_1,\ldots,V_m\}\setminus O} \left(\left(\sum_{\{V_{m+1},\ldots,V_n\}} \prod_{j=m+1,\ldots,n} p(V_j \mid \pi_G(V_j)) \right) \cdot \prod_{j=1,\ldots,m} p(V_j \mid \pi_G(V_j)) \right) \Bigg|_{O = o}}$$

The rule of marginalisation now implies that the sum terms in parentheses in the equation above equal one: for node V_n, marginalisation gives

$$\begin{aligned} &\sum_{\{V_{m+1},\ldots,V_n\}} \prod_{i=m+1,\ldots,n} p(V_i \mid \pi_G(V_i)) = \\ &= \sum_{\{V_{m+1},\ldots,V_{n-1}\}} \left(\left(\sum_{\{V_n\}} p(V_n \mid \pi_G(V_n)) \right) \cdot \prod_{i=m+1,\ldots,n-1} p(V_i \mid \pi_G(V_i)) \right) \\ &= \sum_{\{V_{m+1},\ldots,V_{n-1}\}} \prod_{i=m+1,\ldots,n-1} p(V_i \mid \pi_G(V_i)) \end{aligned}$$

Recursively repeating this argument for the nodes $V_{n-1}, \ldots, V_{m+1}$ results in

$$\sum_{\{V_{m+1},\ldots,V_n\}} \prod_{i=m+1,\ldots,n} p(V_i \mid \pi_G(V_i)) = 1$$

We conclude that

$$\Pr(v_r \mid o) = \frac{\displaystyle\sum_{\{V_1,\ldots,V_m\}\setminus(\{V_r\}\cup O)} \prod_{i=1,\ldots,m} p(V_i \mid \pi_G(V_i)) \Bigg|_{\substack{V_r = v_r \\ O = o}}}{\displaystyle\sum_{\{V_1,\ldots,V_m\}\setminus O} \prod_{j=1,\ldots,m} p(V_j \mid \pi_G(V_j)) \Bigg|_{O = o}}$$

which shows that the probability of interest $\Pr(v_r \mid o)$ is algebraically independent of the conditional probabilities of any node from the set $Insen_3(V_r, O)$, as stated in the lemma. □

So far, we have shown that a belief network's probability of interest for a node V_r given observations for nodes O is algebraically independent of the conditional probabilities of

any node from the set $Insen_3(V_r, O)$. We now proceed by showing that this probability of interest is also algebraically independent of the conditional probabilities of the nodes from the set $Insen_2(V_r, O)$.

Lemma A.4 (cf. Lemma 4.3.7) *Let B be a Bayesian belief network with the digraph $G = (V(G), A(G))$. Let* Pr *be the joint probability distribution defined by B. Let $O \subseteq V(G)$ be the set of observed nodes in G and let o denote the corresponding observations. Let $V_r \in V(G)$ be the network's node of interest. Then, for any value v_r of V_r, we have that* $\Pr(v_r \mid o) \propto p(V_i \mid \pi_G(V_i))$ *for every node $V_i \in Insen_2(V_r, O)$.*

Proof. Let ι be a sensitivity ordering of G with respect to V_r and O. Without loss of generality, we assume that $Insen_3(V_r, O) = \varnothing$. Also without loss of generality, we assume that the nodes in G are indexed by their ordering number, that is, we assume that $\iota(V_i) = i$. We take $n \geq 1$ to be the number of nodes in G. From the definition of sensitivity ordering, we have that the nodes in $Insen_2(V_r, O)$ have the highest ordering numbers in the digraph; we assume that $Insen_2(V_r, O)$ consists of the nodes $V_{m+1}, \ldots, V_n$. For our probability of interest $\Pr(v_r \mid o)$, we find that

$$\Pr(v_r \mid o) = \frac{\displaystyle\sum_{\substack{\{V_1,\ldots,V_n\}\setminus \\ (\{V_r\}\cup O)}} \left(\left(\prod_{i=m+1,\ldots,n} p(V_i \mid \pi_G(V_i)) \right) \cdot \left(\prod_{i=1,\ldots,m} p(V_i \mid \pi_G(V_i)) \right) \right) \Bigg|_{\substack{V_r = v_r \\ O = o}}}{\displaystyle\sum_{\{V_1,\ldots,V_n\}\setminus O} \left(\left(\prod_{j=m+1,\ldots,n} p(V_j \mid \pi_G(V_j)) \right) \cdot \left(\prod_{j=1,\ldots,m} p(V_j \mid \pi_G(V_j)) \right) \right) \Bigg|_{O = o}}$$

Since our sensitivity ordering is a topological ordering of G, we know that $(\bigcup_{i=1,\ldots,m} \pi_G(V_i)) \cap \{V_{m+1}, \ldots, V_n\} = \varnothing$. Since $V_r \notin Insen_2(V_r, O)$ by definition, we also have that $V_r \notin \{V_{m+1}, \ldots, V_n\}$. Using these observations, we find that

$$\Pr(v_r \mid o) = \frac{\displaystyle\sum_{\substack{\{V_1,\ldots,V_m\}\setminus \\ (\{V_r\}\cup O)}} \left(\left(\sum_{\{V_{m+1},\ldots,V_n\}\setminus O} \prod_{i=m+1,\ldots,n} p(V_i \mid \pi_G(V_i)) \right) \cdot \prod_{i=1,\ldots,m} p(V_i \mid \pi_G(V_i)) \right) \Bigg|_{\substack{V_r = v_r \\ O = o}}}{\displaystyle\sum_{\{V_1,\ldots,V_m\}\setminus O} \left(\left(\sum_{\{V_{m+1},\ldots,V_n\}\setminus O} \prod_{j=m+1,\ldots,n} p(V_j \mid \pi_G(V_j)) \right) \cdot \prod_{j=1,\ldots,m} p(V_j \mid \pi_G(V_j)) \right) \Bigg|_{O = o}}$$

Now, from Definition 4.3.3, we have that the nodes $V_{m+1}, \ldots, V_n$ from the set $Insen_2(V_r, O)$ and their predecessors are d-separated from the node of interest V_r. Any

predecessor of a node V_i, $i = m+1, \ldots, n$, therefore, is either included in $Insen_2(V_r, O)$ itself or is an observed node. We conclude that $(\bigcup_{i=m+1,\ldots,n} \pi_G(V_i)) \cap (\{V_1, \ldots, V_m\} \setminus (\{V_r\} \cup O)) = \varnothing$. The probability of interest can now be written as

$$
\begin{aligned}
&\Pr(v_r \mid o) = \\
&= \frac{\left(\sum_{\{V_{m+1},\ldots,V_n\}\setminus O} \prod_{i=m+1,\ldots,n} p(V_i \mid \pi_G(V_i))\right) \cdot \left(\sum_{\substack{\{V_1,\ldots,V_m\}\setminus \\ (\{V_r\}\cup O)}} \prod_{i=1,\ldots,m} p(V_i \mid \pi_G(V_i))\right)\Bigg|_{\substack{V_r = v_r \\ O = o}}}{\left(\sum_{\{V_{m+1},\ldots,V_n\}\setminus O} \prod_{j=m+1,\ldots,n} p(V_j \mid \pi_G(V_j))\right) \cdot \left(\sum_{\{V_1,\ldots,V_m\}\setminus O} \prod_{j=1,\ldots,m} p(V_j \mid \pi_G(V_j))\right)\Bigg|_{O = o}} \\
&= \frac{\sum_{\{V_1,\ldots,V_m\}\setminus(\{V_r\}\cup O)} \prod_{i=1,\ldots,m} p(V_i \mid \pi_G(V_i))\Big|_{\substack{V_r = v_r \\ O = o}}}{\sum_{\{V_1,\ldots,V_m\}\setminus O} \prod_{j=1,\ldots,m} p(V_j \mid \pi_G(V_j))\Big|_{O = o}}
\end{aligned}
$$

which shows that the probability of interest $\Pr(v_r \mid o)$ is algebraically independent of the conditional probabilities of any node from the set $Insen_2(V_r, O)$, as stated in the lemma. □

So far, we have shown that a belief network's probability of interest for a node V_r given observations for nodes O is algebraically independent of the conditional probabilities of any node from the sets $Insen_3(V_r, O)$ and $Insen_2(V_r, O)$. To conclude, we now prove that this probability of interest is also algebraically independent of the conditional probabilities of the nodes from the set $Insen_1(V_r, O)$.

Lemma A.5 (cf. Lemma 4.3.9) *Let B be a Bayesian belief network with the digraph $G = (V(G), A(G))$. Let* Pr *be the joint probability distribution defined by B. Let $O \subseteq V(G)$ be the set of observed nodes in G and let o denote the corresponding observations. Let $V_r \in V(G)$ be the network's node of interest. Then, for any value v_r of V_r, we have that* $\Pr(v_r \mid o) \propto p(V_i \mid \pi_G(V_i))$ *for every node $V_i \in Insen_1(V_r, O)$.*

Proof. Without loss of generality, we assume that $Insen_2(V_r, O) = \varnothing$ and $Insen_3(V_r, O) = \varnothing$. From these assumptions and Lemma A.1, we have that $V(G) = Sen(V_r, O) \cup Insen_1(V_r, O)$. Let the nodes from $Sen(V_r, O)$ be called $V_1, \ldots, V_m$ and let the nodes from $Insen_1(V_r, O)$ be called $V_{m+1}, \ldots, V_n$, $n \geq 1$; note that, in contrast with the proofs of the previous lemmas, the nodes are *not* indexed by their ordering number according to some sensitivity ordering of G. For our probability of interest $\Pr(v_r \mid o)$, we now

find that

$$\Pr(v_r \mid o) = \frac{\sum_{\{V_1,\ldots,V_n\}\setminus(\{V_r\}\cup O)} \left(\left(\prod_{i=m+1,\ldots,n} p(V_i \mid \pi_G(V_i))\right)\cdot\left(\prod_{i=1,\ldots,m} p(V_i \mid \pi_G(V_i))\right)\right)\Big|_{\substack{V_r = v_r \\ O = o}}}{\sum_{\{V_1,\ldots,V_n\}\setminus O} \left(\left(\prod_{j=m+1,\ldots,n} p(V_j \mid \pi_G(V_j))\right)\cdot\left(\prod_{j=1,\ldots,m} p(V_j \mid \pi_G(V_j))\right)\right)\Big|_{O=o}}$$

From Definition 4.3.3, we have that the nodes $V_{m+1},\ldots,V_n$ from the set $Insen_1(V_r,O)$ and their predecessors are d-separated from the node of interest V_r. Any predecessor of a node V_i, $i = m+1,\ldots,n$, therefore, is either included in $Insen_1(V_r,O)$ itself or is an observed node. We conclude that $(\bigcup_{i=m+1,\ldots,n} \pi_G(V_i)) \cap (\{V_1,\ldots,V_m\}\setminus(\{V_r\}\cup O)) = \varnothing$. In addition, for every node V_i, $i = 1,\ldots,m$, from $Sen(V_r,O)$, we have that any predecessor that is included in the set $Insen_1(V_r,O)$ is an observed node. Hence, $\bigcup_{i=1,\ldots,m} \pi_G(V_i)) \cap (\{V_{m+1},\ldots,V_n\}\setminus O) = \varnothing$. Building upon these observations, the probability of interest can be written as

$$\begin{aligned}
&\Pr(v_r \mid o) = \\
&= \frac{\left(\sum_{\{V_{m+1},\ldots,V_n\}\setminus O} \prod_{i=m+1,\ldots,n} p(V_i \mid \pi_G(V_i))\right)\cdot\left(\sum_{\substack{\{V_1,\ldots,V_m\}\setminus \\ (\{V_r\}\cup O)}} \prod_{i=1,\ldots,m} p(V_i \mid \pi_G(V_i))\right)\Big|_{\substack{V_r = v_r \\ O = o}}}{\left(\sum_{\{V_{m+1},\ldots,V_n\}\setminus O} \prod_{j=m+1,\ldots,n} p(V_j \mid \pi_G(V_j))\right)\cdot\left(\sum_{\{V_1,\ldots,V_m\}\setminus O} \prod_{j=1,\ldots,m} p(V_j \mid \pi_G(V_j))\right)\Big|_{O=o}} \\
&= \frac{\sum_{\{V_1,\ldots,V_m\}\setminus(\{V_r\}\cup O)} \prod_{i=1,\ldots,m} p(V_i \mid \pi_G(V_i))\Big|_{\substack{V_r = v_r \\ O = o}}}{\sum_{\{V_1,\ldots,V_m\}\setminus O} \prod_{j=1,\ldots,m} p(V_j \mid \pi_G(V_j))\Big|_{O=o}}
\end{aligned}$$

which shows that the probability of interest $\Pr(v_r \mid o)$ is algebraically independent of the conditional probabilities of any node from the set $Insen_1(V_r,O)$, as stated in the lemma. □

In the foregoing, we have shown that a belief network's probability of interest is algebraically independent of the conditional probabilities of any node that is not included in the sensitivity set under consideration. We now show that the probability of interest relates to any conditional probability for a node from the sensitivity set as a quotient of two linear functions.

Proposition A.6 (cf. Proposition 4.4.1) *Let B be a Bayesian belief network with the digraph $G = (V(G), A(G))$ and let* Pr *be the joint probability distribution defined by B. Let $O \subseteq V(G)$ be the set of observed nodes in G and let o denote the corresponding observations. Let V_r be the network's node of interest and let $Sen(V_r, O)$ be the sensitivity set for V_r given O. Then, for any value v_r of V_r, we have that*

$$\Pr(v_r \mid o) = \frac{a \cdot x + b}{c \cdot x + d}$$

for every conditional probability $x = p(v_s \mid \pi')$ of every node $V_s \in Sen(V_r, O)$, where a, b, c, and d are constants that are dependent upon the values v_s of V_s and π' of $\pi_G(V_s)$.

Proof. The probability of interest $\Pr(v_r \mid o)$ for the belief network B equals

$$\Pr(v_r \mid o) = \frac{\Pr(v_r \wedge o)}{\Pr(o)}$$

Without loss of generality, we take the nodes of the belief network B to be $V_1, \ldots, V_n$, $n \geq 1$. For ease of exposition, we assume all variables in the network to be binary, taking one of the truth values *true* and *false*. We will use v_i to denote the proposition that the variable V_i takes the value *true*; $V_i = \mathit{false}$ will be denoted as $\neg v_i$. We will return to our assumption of binary variables at the end of the proof. We now consider a node V_s from the sensitivity set $Sen(V_r, O)$ under study. Without loss of generality, we investigate the sensitivity of the probability of interest with regard to the conditional probability $p(v_s \mid \pi')$ for this node, where π' is a specific combination of values for the nodes from the set $\pi_G(V_s)$. For the numerator $\Pr(v_r \wedge o)$ of the probability of interest, we find that

$$\Pr(v_r \wedge o) =$$

$$= \sum_{\substack{\{V_1, \ldots, V_n\} \setminus \\ (\{V_r\} \cup O)}} \prod_{i=1,\ldots,n} p(V_i \mid \pi_G(V_i)) \Bigg|_{\substack{V_r = v_r \\ O = o}}$$

$$= \sum_{\substack{\{V_1, \ldots, V_n\} \setminus \\ (\{V_r\} \cup O)}} \left(p(V_s \mid \pi_G(V_s)) \cdot \prod_{\substack{i = 1, \ldots, n, \\ i \neq s}} p(V_i \mid \pi_G(V_i)) \right) \Bigg|_{\substack{V_r = v_r \\ O = o}}$$

$$= \sum_{\substack{\{V_1,\ldots,V_n\}\setminus \\ (\{V_r,V_s\}\cup\pi_G(V_s)\cup O)}} \left(p(v_s \mid \pi') \cdot \prod_{\substack{i=1,\ldots,n,\\ i\neq s}} p(V_i \mid \pi_G(V_i)) \right) \Bigg|_{\substack{V_r=v_r\\ O=o\\ V_s=v_s\\ \pi_G(V_s)=\pi'}}$$

$$+ \sum_{\substack{\{V_1,\ldots,V_n\}\setminus \\ (\{V_r,V_s\}\cup\pi_G(V_s)\cup O)}} \left((1-p(v_s \mid \pi')) \cdot \prod_{\substack{i=1,\ldots,n,\\ i\neq s}} p(V_i \mid \pi_G(V_i)) \right) \Bigg|_{\substack{V_r=v_r\\ O=o\\ V_s=\neg v_s\\ \pi_G(V_s)=\pi'}}$$

$$+ \sum_{\substack{\{V_1,\ldots,V_n\}\setminus \\ (\{V_r\}\cup O),\\ \pi_G(V_s)\neq\pi'}} \left(p(V_s \mid \pi_G(V_s)) \cdot \prod_{\substack{i=1,\ldots,n,\\ i\neq s}} p(V_i \mid \pi_G(V_i)) \right) \Bigg|_{\substack{V_r=v_r\\ O=o}}$$

The first term in the above sum of three assembles all products that specify the conditional probability $p(v_s \mid \pi')$. The second term gathers all products specifying the complement, $p(\neg v_s \mid \pi')$, of the conditional probability under study. Note that this term, as the first one, depends on the value of $p(v_s \mid \pi')$. The third term, to conclude, collects the remaining products; these products specify for the node V_s a conditional probability that has another combination of values than π' for its conditioning part. Note that the third term does not depend on the value of the conditional probability under study. Writing x for $p(v_s \mid \pi')$, we find that

$$\Pr(v_r \wedge o) = a \cdot x + b$$

where

$$a = \sum_{\substack{\{V_1,\ldots,V_n\}\setminus \\ (\{V_r,V_s\}\cup\pi_G(V_s)\cup O)}} \left(\prod_{\substack{i=1,\ldots,n,\\ i\neq s}} p(V_i \mid \pi_G(V_i)) \right) \Bigg|_{\substack{V_r=v_r\\ O=o\\ V_s=v_s\\ \pi_G(V_s)=\pi'}}$$

$$- \sum_{\substack{\{V_1,\ldots,V_n\}\setminus \\ (\{V_r,V_s\}\cup\pi_G(V_s)\cup O)}} \left(\prod_{\substack{i=1,\ldots,n,\\ i\neq s}} p(V_i \mid \pi_G(V_i)) \right) \Bigg|_{\substack{V_r=v_r\\ O=o\\ V_s=\neg v_s\\ \pi_G(V_s)=\pi'}}$$

and

$$b = \sum_{\substack{\{V_1,\ldots,V_n\}\setminus \\ (\{V_r,V_s\}\cup\pi_G(V_s)\cup O)}} \left(\prod_{\substack{i=1,\ldots,n,\\ i\neq s}} p(V_i \mid \pi_G(V_i)) \right)\Bigg|_{\substack{V_r=v_r\\ O=o\\ V_s=\neg v_s\\ \pi_G(V_s)=\pi'}} + \sum_{\substack{\{V_1,\ldots,V_n\}\setminus \\ (\{V_r\}\cup O),\\ \pi_G(V_s)\neq\pi'}} \left(\prod_{i=1,\ldots,n} p(V_i \mid \pi_G(V_i)) \right)\Bigg|_{\substack{V_r=v_r\\ O=o}}$$

Note that the constants a and b are related to the conditional probability under study but are not dependent upon its value.

For the denominator $\Pr(o)$ of the probability of interest, we analogously find that

$$\begin{aligned}\Pr(o) &= \\ &= \sum_{\substack{\{V_1,\ldots,V_n\}\setminus \\ (\{V_s\}\cup\pi_G(V_s)\cup O)}} \left(p(v_s \mid \pi') \cdot \prod_{\substack{i=1,\ldots,n,\\ i\neq s}} p(V_i \mid \pi_G(V_i)) \right)\Bigg|_{\substack{O=o\\ V_s=v_s\\ \pi_G(V_s)=\pi'}} \\ &\quad + \sum_{\substack{\{V_1,\ldots,V_n\}\setminus \\ (\{V_s\}\cup\pi_G(V_s)\cup O)}} \left((1-p(v_s \mid \pi')) \cdot \prod_{\substack{i=1,\ldots,n,\\ i\neq s}} p(V_i \mid \pi_G(V_i)) \right)\Bigg|_{\substack{O=o\\ V_s=\neg v_s\\ \pi_G(V_s)=\pi'}} \\ &\quad + \sum_{\substack{\{V_1,\ldots,V_n\}\setminus O,\\ \pi_G(V_s)\neq\pi'}} \left(p(V_s \mid \pi_G(V_s)) \cdot \prod_{\substack{i=1,\ldots,n,\\ i\neq s}} p(V_i \mid \pi_G(V_i)) \right)\Bigg|_{O=o} \\ &= c\cdot x + d\end{aligned}$$

once more writing x for the conditional probability under study. For the constants c

and d, we have that

$$c = \sum_{\substack{\{V_1,\ldots,V_n\}\setminus \\ (\{V_s\}\cup\pi_G(V_s)\cup O)}} \left(\prod_{\substack{i=1,\ldots,n,\\ i\neq s}} p(V_i \mid \pi_G(V_i)) \right)\Bigg|_{\substack{O=o\\ V_s=v_s\\ \pi_G(V_s)=\pi'}}$$

$$- \sum_{\substack{\{V_1,\ldots,V_n\}\setminus \\ (\{V_s\}\cup\pi_G(V_s)\cup O)}} \left(\prod_{\substack{i=1,\ldots,n,\\ i\neq s}} p(V_i \mid \pi_G(V_i)) \right)\Bigg|_{\substack{O=o\\ V_s=\neg v_s\\ \pi_G(V_s)=\pi'}}$$

and

$$d = \sum_{\substack{\{V_1,\ldots,V_n\}\setminus \\ (\{V_s\}\cup\pi_G(V_s)\cup O)}} \left(\prod_{\substack{i=1,\ldots,n,\\ i\neq s}} p(V_i \mid \pi_G(V_i)) \right)\Bigg|_{\substack{O=o\\ V_s=\neg v_s\\ \pi_G(V_s)=\pi'}}$$

$$+ \sum_{\substack{\{V_1,\ldots,V_n\}\setminus O,\\ \pi_G(V_s)\neq\pi'}} \left(\prod_{i=1,\ldots,n} p(V_i \mid \pi_G(V_i)) \right)\Bigg|_{O=o}$$

From the previous observations, we conclude that the probability of interest $\Pr(v_r \mid o)$ equals

$$\Pr(v_r \mid o) = \frac{a\cdot x + b}{c\cdot x + d}$$

where x, a, b, c, and d are as above.

In our proof so far, we have assumed all variables in the belief network B to be binary. We would like to note that the proof can be generalised to non-binary variables, provided that for varying the value of a conditional probability $p(v_s \mid \pi')$ for a node V_s from the sensitivity set under study, the ratio of any pair of complementary probabilities $p(v_s' \mid \pi')$ and $p(v_s'' \mid \pi')$ for this node is kept fixed. □

So far, we have shown that a belief network's probability of interest relates as a quotient of two linear functions to a conditional probability under study. For a conditional probability that pertains to a node from the sensitivity set that does not have any observed descendants, this functional relation reduces to a *linear function.*

Proposition A.7 (cf. Proposition 4.4.3) *Let B be a Bayesian belief network with the digraph $G = (V(G), A(G))$ and let* Pr *be the joint probability distribution defined by B. Let $O \subseteq V(G)$ be the set of observed nodes in G and let o denote the corresponding*

observations. Let V_r be the network's node of interest and let $Sen(V_r, O)$ be the sensitivity set for V_r given O. Let $V_s \in Sen(V_r, O)$ with $\sigma^*(V_s) \cap O = \varnothing$. Then, for any value v_r of V_r, we have that

$$\Pr(v_r \mid o) = a \cdot x + b$$

for every conditional probability $x = p(v_s \mid \pi')$ of V_s, where a and b are constants that are dependent upon the values v_s of V_s and π' of $\pi_G(V_s)$.

Proof. The probability of interest $\Pr(v_r \mid o)$ for the belief network B once more equals

$$\Pr(v_r \mid o) = \frac{\Pr(v_r \wedge o)}{\Pr(o)}$$

From the proof of Proposition A.6, we have that the numerator $\Pr(v_r \wedge o)$ in this equation relates linearly to the conditional probability x under study. More formally, we have that

$$\Pr(v_r \wedge o) = a' \cdot x + b'$$

where a' and b' are constants as specified in the proof of the proposition.

Let ι be a sensitivity ordering of G with respect to V_r and O. Without loss of generality, we assume that the nodes in G are indexed by their ordering number, that is, we assume that $\iota(V_i) = i$; we take $n \geq 1$ to be the number of nodes in G. For ease of exposition, we further assume all variables in the network to be binary, taking one of the truth values *true* and *false*. We will once more use v_i to denote the proposition that the variable V_i takes the value *true*; $V_i = \mathit{false}$ will be denoted as $\neg v_i$. Our proof can be generalised to non-binary variables as indicated in the proof of Proposition A.6. We now consider a node V_s from the sensitivity set $Sen(V_r, O)$ under study. Without loss of generality, we assume that the set $\sigma^*(V_s)$ consists of the nodes $V_s, \ldots, V_n$. We investigate the sensitivity of the probability of interest with regard to the conditional probability $p(v_s \mid \pi')$ for the node V_s, where π' is a specific combination of values for the nodes from $\pi_G(V_s)$. For the denominator $\Pr(o)$ of the probability of interest, we find that

$$\Pr(o) = \sum_{\{V_1,\ldots,V_n\}\setminus O} \prod_{i=1,\ldots,n} p(V_i \mid \pi_G(V_i)) \Bigg|_{O=o} =$$

$$= \sum_{\{V_1,\ldots,V_n\}\setminus O} \left(\left(\prod_{i=s,\ldots,n} p(V_i \mid \pi_G(V_i)) \right) \cdot \left(\prod_{i=1,\ldots,s-1} p(V_i \mid \pi_G(V_i)) \right) \right) \Bigg|_{O=o}$$

Since our sensitivity ordering ι is a topological ordering, we know that $(\bigcup_{i=1,\ldots,s-1} \pi_G(V_i)) \cap \{V_s, \ldots, V_n\} = \varnothing$. In addition, we have that $\sigma^*(V_s) \cap O = \varnothing$ and,

hence, that $\{V_s, \ldots, V_n\} \cap O = \varnothing$. Building upon these observations, we find that

$$\Pr(o) = \sum_{\{V_1,\ldots,V_{s-1}\}\setminus O} \left(\left(\sum_{\{V_s,\ldots,V_n\}} \prod_{i=s,\ldots,n} p(V_i \mid \pi_G(V_i)) \right) \cdot \prod_{i=1,\ldots,s-1} p(V_i \mid \pi_G(V_i)) \right) \Bigg|_{O=o}$$

The rule of marginalisation now implies that the sum term in parentheses in the equation above equals one. We conclude that

$$\begin{aligned} \Pr(o) &= \sum_{\{V_1,\ldots,V_{s-1}\}\setminus O} \prod_{i=1,\ldots,s-1} p(V_i \mid \pi_G(V_i)) \Bigg|_{O=o} \\ &= c' \end{aligned}$$

From this derivation, we have that $\Pr(o)$ is a constant with respect to the conditional probability under study x. For our probability of interest, we now find that

$$\begin{aligned} \Pr(v_r \mid o) &= \frac{a' \cdot x + b'}{c'} \\ &= a \cdot x + b \end{aligned}$$

where $a = \frac{a'}{c'}$ and $b = \frac{b'}{c'}$. □

Chapter 5

A computational architecture for n-way sensitivity analysis of Bayesian networks

Abstract

A probability computed from a Bayesian network relates to the parameters of the network by a simple mathematical function. A prior probability can be expressed as a multilinear function in the network's parameters; a posterior probability is a quotient of two such functions. These functions serve to yield insight into the robustness of a Bayesian network and thus constitute the basis for a sensitivity analysis. Sensitivity analysis amounts to establishing the coefficients in the functions under study. In the past, various methods for sensitivity analysis have been suggested. Most of these methods are very much demanding from a computational point of view. In this chapter, we present a new and efficient method. The method builds upon a junction-tree representation of a Bayesian network for its computational architecture. It computes the coefficients of the sensitivity functions under study by propagating and combining vectors of (partially computed) coefficients through the junction tree.

5.1 Introduction

A Bayesian network is a concise representation of a joint probability distribution on a set of statistical variables. It can be used for computing the prior or posterior probability of any variable in the network in view of the currently available evidence. To that end, various methods for propagating evidence in a Bayesian network have been developed [Pearl, 1988, Lauritzen & Spiegelhalter, 1988].

A Bayesian network of realistic size contains a considerable number of conditional probabilities; these probabilities are the network's parameters. The parameters of a

Bayesian network are typically assessed by experts in the domain of application or may be estimated from data. Due to, for example, problems of bias and poor calibration or lack of sufficient, reliable, data, the estimates obtained tend to be inaccurate. Inaccuracies in a network's parameters may influence the reliability of the probabilities of interest computed from the network. An integral part of investigating a network's reliability is to study its sensitivity. The sensitivity of a Bayesian network refers to the effect on a probability of interest of changes in the estimates for one or more of the network's parameters; to study network sensitivity, a *sensitivity analysis* is carried out.

Basically, sensitivity analysis of a Bayesian network amounts to systematically varying the estimates for one or more parameters in the network simultaneously and investigating the effects on a probability of interest. Varying one parameter estimate at a time is called a *one-way sensitivity analysis*. It serves to reveal the independent effect of the parameter under study on a probability of interest. The term *n-way sensitivity analysis* is used to indicate that n, $n \geq 1$, network parameters are varied simultaneously. It reveals how the n parameters interact in their effect on a probability of interest. In clinical decision analyses [Habbema *et al.*, 1990, Dippel *et al.*, 1992], it is customary to perform a two-way or three-way sensitivity analysis as well as a one-way sensitivity analysis, since it yields additional insight in a model's robustness. The results of such an analysis are relatively easy to interpret and can be represented graphically. Higher order sensitivity analyses are uncommon since their results are harder to interpret. However, for generality, we consider in this chapter the situation in which an arbitrary number of n parameters is varied.

The most straightforward way of performing a sensitivity analysis of a Bayesian network is to vary the parameters under study stepwise and in a systematic way. The effect of these stepwise variations on the probability of interest is evaluated by using any standard propagation algorithm. For larger networks and for sensitivity analyses involving more than one parameter, such a straightforward approach is computationally unfeasible. In recent years, various researchers have addressed the computational complexity of performing a sensitivity analysis [Laskey, 1995, Castillo *et al.*, 1997a, Coupé & Van der Gaag, 1998, Kjærulff & Van der Gaag, 2000]. Most currently available methods exploit the property that a probability of interest computed from a Bayesian network relates to the parameters of the network by a simple mathematical function; a prior probability can be expressed as a multilinear function in the network's parameters and a posterior probability is a quotient of two such functions. Performing a sensitivity analysis then amounts to establishing the coefficients in these sensitivity functions. At this moment, the method developed by E. Castillo *et al.* (1997) to compute the required coefficients is the most efficient method available. For each coefficient, a different combination of values for the parameters under study is asssumed and subsequently the probability of interest is computed from the network. As a result, a system of linear equations is obtained which is solved to give the required coefficients.

In this chapter, we present a new method for n-way sensitivity analysis of Bayesian networks. Our method builds upon the computational architecture of a junction tree derived from a Bayesian network and is closely related to standard junction-tree propagation. In fact, the algorithm presented is an adaptation of a standard propagation algorithm. Instead of potential functions as in standard propagation, the messages sent between the cliques in the junction tree are vectors of (partially computed) coefficients. These coefficients are processed locally per clique and are combined and accumulated to yield the coefficients in the required n-way sensitivity function that describes the probability of interest in terms of the n parameters under study.

Our algorithm for n-way sensitivity analysis is more efficient than currently available methods. So far, the most efficient method for n-way sensitivity analysis has been presented by Castillo *et al.* (1997). In its simplest form, our method is comparable to theirs with respect to computational efficiency. However, an advantage of our method is that it is integrated in an existing propagation scheme, thereby providing a framework that can be easily extended and further optimized. In this chapter, we discuss some optimizations of the basic algorithm, such as the determination of the optimal root to start the algorithm. Furthermore, with our method it is not necessary to solve systems of linear equations, as is the case in the method by Castillo *et al.*.

The chapter is structured as follows. In Section 5.2, some preliminaries of Bayesian networks, junction trees, and propagation in junction trees are presented. The properties of sensitivity analysis of a Bayesian network are reviewed in Section 5.3. Section 5.4, then, describes our algorithm for computing the coefficients in an n-way sensitivity function. Section 5.5 deals with the determination of the optimal clique in the junction tree to start the computation of the required coefficients. Some optimizations of our method are briefly discussed in Section 5.6. In Section 5.7, we review related work on sensitivity analysis of Bayesian networks, compare our method to [Castillo *et al.*, 1997b] and discuss possible optimizations of both our method and the method by Castillo *et al.*. The chapter ends with some conclusions in Section 5.8.

5.2 Bayesian networks and junction trees

A *Bayesian network* basically is a concise representation of a joint probability distribution on a set of statistical variables [Pearl, 1988]. Information about the independences holding among the variables is explicitly separated from the numerical quantities involved in the distribution. To this end, the network comprises a qualitative part and an associated quantitative part.

The qualitative part of a Bayesian network BN is a graphical representation of the independences holding among the variables in the probability distribution that is being represented. It takes the form of an *acyclic directed graph* G with nodes $V(G)$. In this

digraph, each node V_i, $i = 1, \ldots, r$, $r \geq 1$, represents a statistical variable that can take one of a finite set of values. In the sequel, the set of possible values of variable V_i, called the *universe* of V_i, will be indicated by Ω_{V_i}. Any subset of variables $W \subseteq V(G)$ has associated a universe Ω_W that is defined as the Cartesian product of the universes of all variables from W, that is, $\Omega_W = \times_{V_i \in W} \Omega_{V_i}$. An element from a universe will be termed a *configuration*. In the remainder of the chapter, formula will generally be stated as schemata involving variables. From such schemata multiple instantiations can be obtained by filling in configurations for the variables involved.

The set of arcs in the digraph of a Bayesian network models the independences among the represented variables. Informally speaking, we take an arc $V_i \rightarrow V_j$ to represent a direct influential or causal relationship between the variables V_i and V_j; the arc's direction designates V_j as the effect or consequence of the cause V_i. Absence of an arc between two nodes means that the corresponding variables do not influence each other directly and, hence, are (conditionally) independent.

Associated with the qualitative part of a Bayesian network are numbers that describe the strengths of the influential relationships among the represented variables. With each node V_i of the network's digraph is associated a set of conditional probability distributions describing the joint influence of the various values for the node's (immediate) predecessors $\pi(V_i)$ on the probabilities of the values of this node itself, that is, with each node V_i are associated conditional probability distributions $p(V_i \mid \pi(V_i))$. A conditional probability $p(v_i \mid \pi')$, for a specific value $v_i \in \Omega_{V_i}$ and a configuration $\pi' \in \Omega_{\pi(V_i)}$, is termed a *parameter* of the network. The conditional probability distributions jointly constitute the quantitative part of the network.

We illustrate the concept of Bayesian network by means of an example that will be used as a running example throughout the chapter.

Example 5.2.1 Consider the Bayesian network shown in Figure 5.1a. The network pertains to the statistical variables V_i, $i = 1, \ldots, 8$. From the network's digraph various independences are read [Pearl, 1988]. For example, variable V_5 is dependent on variable V_1; when evidence on the value of variable V_3 becomes available, however, V_1 and V_5 become independent. If the value of variable V_8 is not observed, the variables V_6 and V_7 are independent of each other. Observing V_8, however, induces a dependence between V_6 and V_7. Associated with the variables in the network are the conditional probability distributions $p(V_i \mid \pi(V_i))$. We assume that the variables in our example network are binary. That is, they either take the value *true*, denoted by v_i, or *false*, denoted by $\neg v_i$. In Figure 5.1a, for each variable the estimates for the required conditional probabilities are shown. From the conditional probabilities for the variable V_6 we see, for example, that the probability that V_6 takes the value *true* is high, unless both parents V_3 and V_4 are *false*. □

The conditional probability distributions of a Bayesian network provide all informa-

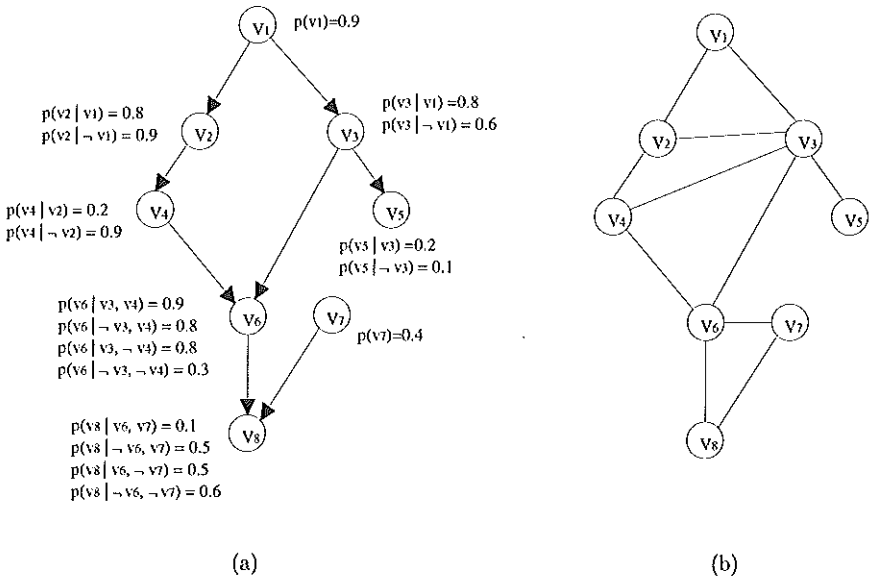


Figure 5.1: (a) An example Bayesian network and (b) the triangulated moral graph obtained from this network.

tion necessary for uniquely defining a joint probability distribution on the variables discerned that respects the independences portrayed by the network's qualitative part. Hence, from the network, any (prior or posterior) probability of interest can be computed. For this purpose various algorithms are available [Pearl, 1988, Lauritzen & Spiegelhalter, 1988, Shachter, 1986, Shafer & Shenoy, 1990, Jensen *et al.*, 1990].

The algorithm for evidence propagation developed by S.L. Lauritzen and D.J. Spiegelhalter forms the basis of the method for sensitivity analysis that is presented in this chapter. This algorithm and and its underlying computational architecture, the *junction tree* [Jensen *et al.*, 1990], are therefore reviewed briefly. The algorithm transforms a Bayesian network into an equivalent undirected representation. To this end, the network's digraph is transformed into a *triangulated moral graph*. A triangulated graph is an undirected graph in which no cycle of length four or more exists without a shortcut. The transformation of a Bayesian network's graph G into a triangulated moral graph H involves three steps. First, arcs are added to G such that the (original) predecessors of each node in $V(G)$ are connected. As a result, all pairs of variables that may influence each other directly are connected. In the second step, the direction of the arcs are dropped. Now, an undirected representation of the independences in

the probability distribution is obtained. The third, and last, step consists of cutting short each cycle of length four or more by adding an edge. Note that, in general, an acyclic digraph allows several different triangulated moral graphs. In the following, an illustration of this transformation is given.

Example 5.2.2 Consider once more the digraph G depicted in Figure 5.1a. The undirected graph in Figure 5.1b is a triangulated moral graph for G. □

A triangulated moral graph H has associated local *potential functions* on small sets of variables to arrive at a representation of the joint probability distribution on the problem domain. These potential functions will be detailed shortly. The triangulated moral graph together with the potential functions allows for an efficient propagation algorithm, in which the computations to be performed are local to these small sets of variables. For that purpose, the computational architecture of the *junction tree* is used [Jensen *et al.*, 1990].

A junction tree T for a triangulated moral graph H includes for its nodes the *maximal cliques* U_i , $i = 1, \ldots k$. A *clique* of a graph H is a subgraph I in H, such that any two variables in I are connected by an edge. Clique I in H is a maximal clique if there is no clique in H larger than I that properly contains I. In the sequel, we will use the term clique to denote a maximal clique. For ease of exposition, furthermore, we will write U_i to denote both the clique itself and the set of variables in U_i. The intersections between the cliques U_i, $i = 1, \ldots k$, in T give rise to the tree's edges. These edges satisfy the following property: for any two cliques $U_i, U_j \in T$ and each clique U_h on the (unique) path from U_i to U_j in T, we have that $U_i \cap U_j \subseteq U_h$. Associated with each edge (U_i, U_j) is the clique intersection of U_i and U_j. These clique intersections are called *separators* and will be denoted by S_l, $l < k$. Usually, a triangulated graph allows various different junction trees.

With each clique U_i in the junction tree, a potential function ϕ_{U_i} is associated. These potential functions capture probabilistic information about the variables involved and are obtained from the conditional probability distributions from the original Bayesian network BN. The potential function ϕ_{U_i} of U_i is a product of probability distributions. To compose this product, the probability distributions of each variable in the original network BN are assigned to a clique under the following conditions; a probability distribution can only be assigned to a clique that includes both the variable to which this distribition pertains as well as the predecessors of this variable and every probability distribution can be assigned to one clique only. Consider a clique U_i to which the conditional probability distributions $p(V_j \mid \pi(V_j))$ of the nodes V_j from a set $W \subseteq U_i$ are assigned. The potential function ϕ_{U_i} then equals

$$\phi_{U_i}(U_i) = \prod_{V_j \in W} p(V_j \mid \pi(V_j))$$

We illustrate this with our running example.

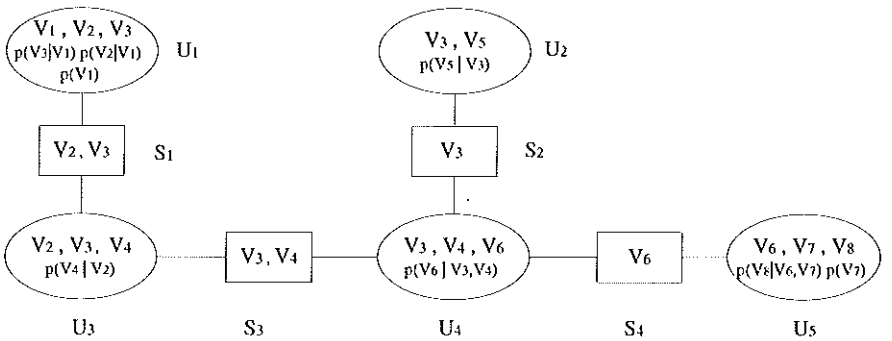


Figure 5.2: The junction tree obtained from the network in Figure 5.1a.

Example 5.2.3 Consider once more the digraph from Figure 5.1a. Figure 5.2 shows a junction tree for this network, constructed from the triangulated graph in Figure 5.1b. The ovals represent the cliques and the boxes indicate the separators. The conditional probability distributions for the variables in the Bayesian network are assigned to the cliques in the junction tree. The conditional probability distributions for node V_6, for example, are attached to clique U_4. Therefore, all the individual parameters from these conditional probability distributions are located in clique U_4 only. □

The junction tree is taken as a computational architecture for processing evidence and for computing an updated probability distribution from a Bayesian network. The cliques in the tree are viewed as autonomous objects and the tree's edges are looked upon as bi-directional communication channels. Through the communication channels, the cliques send each other messages providing information about the represented joint probability distribution and about the evidence entered in the tree. Each clique is able to compute the (updated) marginal probability distribution on its variables from its local marginal distribution and the information it receives from its neighbours [Jensen, 1996]. A clique is allowed to send a message over the edge to a neighbouring clique only if it has received a message from all its other neighbours; it then is said to be triggered. Initially, all leaf cliques are triggered and there is always a triggered clique until a message has been passed in both directions over all edges. The computations can be organised as *rooted propagation*: a clique is chosen as the root of the junction tree and the message passing is directed to and from the root. First, there is an *inward propagation* in which all messages are directed at the root; subsequently, messages are sent from the root towards the leaves, called *outward propagation*. Before propagating messages through the tree, first the evidence available is entered in the tree. To that end, the values of the observed variables are kept fixed in the potential functions associated with each clique.

To further detail the propagation of information in the junction tree, we focus on *Shafer-Shenoy propagation* [Shafer & Shenoy, 1990]. All separators in the junction tree are looked upon as having two mail boxes, one for each direction. Consider a clique U_i with neighbouring separators $S_{i_1}, \cdots, S_{i_{t+1}}$, $t \geq 0$, and a clique U_j that is connected with U_i through the separator $S_{i_{t+1}}$. Clique U_i computes its message to U_j via $S_{i_{t+1}}$, written $\hat{\phi}_{U_i \to S_{i_{t+1}}}$, by multiplying, for each configuration $u \in \Omega_{U_i}$, its marginal potential $\phi_{U_i}(u)$ with the message $\hat{\phi}_{S_{i_j} \to U_i}(s_j)$, where $s_j \in \Omega_{S_{i_j}}$, s_j and u specify the same values for the variables in S_j, that is, $s_j \wedge u = u$, and $j = 1, \ldots, t$:

$$\phi'_{U_i}(u) = \phi_{U_i}(u) \cdot \prod_{j=1,\ldots,t} \hat{\phi}_{S_{i_j} \to U_i}(s_j)$$

The resulting function ϕ'_{U_i} is *projected* onto $S_{i_{t+1}}$, that is, ϕ'_{U_i} is marginalized over all variables that are not in $S_{i_{t+1}}$. The result is the message $\hat{\phi}_{U_i \to S_{i_{t+1}}}$ to U_j via $S_{i_{t+1}}$;

$$\hat{\phi}_{U_i \to S_{i_{t+1}}}(s_{t+1}) = \sum_{u : s_{t+1} \wedge u = u} \phi'_{U_i}(u)$$

The message, a function over the variables in $S_{i_{t+1}}$ is placed in the appropriate mail box. When message passing has stopped, the two mail boxes of each separator contain two such functions. The marginal distribution over the variables in any clique U_i can now be computed by multiplying, for each configuration $u \in \Omega_{U_i}$, the potential $\phi_{U_i}(u)$ with all incoming messages. To obtain the probability distribution over one specific variable V_k, the marginal distribution of a clique containing V_k is projected onto V_k. To obtain the probability of the evidence, the marginal probability distribution of any clique U_i is marginalised over all variables in that clique. To obtain the marginal probability distribution over the variables in a separator S, for each configuration $s \in \Omega_S$, the messages in the two mail boxes are multiplied. In the following, inward propagation towards a chosen root, clique R, is described in pseudocode. At root R, the probability of the evidence $\Pr(e)$ is computed by marginalising the marginal distribution of R over all variables in R.

The propagation algorithm amounts to the following:

```
        procedure propagate (T, e, R)
        begin
step 1      enter evidence e into the clique tree T;
            until root R in T is triggered do
               for each triggered clique U_i do
step 2            for each configuration u ∈ Ω_{U_i} do
                     for each configuration s_j ∈ Ω_{S_{i_j}} with s_j ∧ u = u do
                        compute φ'_{U_i}(u) = φ_{U_i}(u) · ∏_{j=1,...,t} φ̂_{S_{i_j}→U_i}(s_j)
                     od;
                  od;
step 3            project φ'_{U_i}(U_i) onto S_{i_{t+1}}, giving φ̂_{U_i→S_{i_{t+1}}}(S_{i_{t+1}});
step 4            send φ̂_{U_i→S_{i_{t+1}}} to S_{i_{t+1}}
               od
            od;
step 5      for each configuration r ∈ Ω_R do
               for each configuration s_j ∈ Ω_{S_{R_j}} with s_j ∧ r = r do
                  compute Pr(r, e) = φ_R(r) · ∏_{j=1,...,k} φ̂_{S_{R_j}→R}(s_j)
               od
            od;
step 6      compute Pr(e) = Σ_{r∈Ω_R} Pr(r, e);
        end
```

In the following, we illustrate evidence propagation in a junction tree with our running example.

Example 5.2.4 Consider again the junction tree in Figure 5.2. Suppose the variable V_8 is observed to take the value v_8. Furthermore, assume that we are interested in the posterior probability that the variable V_6 takes the value v_6, that is, we want to compute $\Pr(v_6 \mid v_8)$. To that end, we perform an inward propagation towards clique U_4 in Figure 5.2. The outward propagation from U_4 towards the leaves is omitted. We only need the incoming messages in U_4 to compute $\Pr(v_6 \mid v_8)$ because V_6 is located in U_4. For the computation of several posterior probabilities simultaneously, an outward propagation should be performed too.

First, the evidence $V_8 = v_8$ is entered in the junction tree. This accords with `step 1` of procedure `propagate`. Then `step 2` to `step 4` of `propagate` are carried out for all triggered cliques until no triggered cliques remain. Initially, only leaf cliques are

triggered. We start by computing the message $\hat{\phi}_{U_1 \to S_1}$ that leaf U_1 sends to separator S_1. `Step 2` is omitted since U_1 is a leaf node for which there are no incoming messages. In `step 3`, the local potential function ϕ_{U_1} in clique U_1 is projected onto separator S_1, that is, we sum over the values of the variable V_1:

$$\hat{\phi}_{U_1 \to S_1}(V_2, V_3) = p(V_3 \mid v_1) \cdot p(V_2 \mid v_1) \cdot p(v_1) + p(V_3 \mid \neg v_1) \cdot p(V_2 \mid \neg v_1) \cdot p(\neg v_1)$$

For the four combinations of values for V_2 and V_3, the function values $\hat{\phi}_{U_1 \to S_1}(v_2, v_3)$, $\hat{\phi}_{U_1 \to S_1}(\neg v_2, v_3)$, $\hat{\phi}_{U_1 \to S_1}(v_2, \neg v_3)$, and $\hat{\phi}_{U_1 \to S_1}(\neg v_2, \neg v_3)$ are sent to clique U_3 via S_1 (`step 4`). Represented as a table, separator S_1 receives

	v_3	$\neg v_3$
v_2	0.63	0.18
$\neg v_2$	0.15	0.04

Note that, since $\hat{\phi}_{U_1 \to S_1}$ is sent on to clique U_3, $\hat{\phi}_{U_1 \to S_1} = \hat{\phi}_{S_1 \to U_3}$. For the message from clique U_2 to S_2 we find,

$$\hat{\phi}_{U_2 \to S_2}(V_3) = p(v_5 \mid V_3) + p(\neg v_5 \mid V_3) = 1$$

Represented as a table, $\hat{\phi}_{U_2 \to S_2}$ equals

v_3	1
$\neg v_3$	1

The message $\hat{\phi}_{U_5 \to S_4}$ from U_5 to S_4 is computed in the same way as $\hat{\phi}_{U_1 \to S_1}$ and $\hat{\phi}_{U_2 \to S_2}$. Recall that the variable V_8 is observed to take the value v_8. $\hat{\phi}_{U_5 \to S_4}(V_6)$ then equals

v_6	0.34
$\neg v_6$	0.56

To compute the message from U_3 to S_3, the marginal potential function $\phi_{U_3}(V_2, V_3, V_4) = p(V_4 \mid V_2)$ of clique U_3 is multiplied with the incoming messages $\hat{\phi}_{U_1 \to S_1}(V_2, V_3)$ from S_1 (`step 2`) and subsequently projected onto S_3 (`step 3`). That is,

$$\hat{\phi}_{U_3 \to S_3}(V_3, V_4) = p(V_4 \mid v_2) \cdot \hat{\phi}_{U_1 \to S_1}(v_2, V_3) + p(V_4 \mid \neg v_2) \cdot \hat{\phi}_{U_1 \to S_1}(\neg v_2, V_3)$$

Represented as a table, the message is

	v_3	$\neg v_3$
v_4	0.261	0.072
$\neg v_4$	0.519	0.148

At clique U_4 the messages from the three separators S_2, S_3, and S_4 are available. The probability distribution over the variables V_3, V_4 and V_6 from U_4 with the evidence $V_8 = v_8$ is now computed by carrying out `step 5` of procedure `propagate`;

$$\begin{aligned}\Pr(V_3 \wedge V_4 \wedge V_6 \wedge v_8) &= \phi_{U_4}(V_3, V_4, V_6) \cdot \hat{\phi}_{U_2 \to S_2}(V_3) \cdot \hat{\phi}_{U_3 \to S_3}(V_3, V_4) \cdot \hat{\phi}_{U_5 \to S_4}(V_6) \\ &= p(V_6 \mid V_3, V_4) \cdot \hat{\phi}_{U_2 \to S_2}(V_3) \cdot \hat{\phi}_{U_3 \to S_3}(V_3, V_4) \cdot \hat{\phi}_{U_5 \to S_4}(V_6)\end{aligned}$$

Represented as a table $\Pr(V_3 \wedge V_4 \wedge V_6 \wedge v_8)$ equals

	v_3		$\neg v_3$	
	v_4	$\neg v_4$	v_4	$\neg v_4$
v_6	0.0799	0.1412	0.0196	0.0151
$\neg v_6$	0.0146	0.0581	0.0081	0.0580

As we are interested in $\Pr(v_6 \mid v_8)$, we marginalize over the variables V_3 and V_4, to give

$\Pr(v_6, v_8) = 0.2557$

and

$\Pr(\neg v_6, v_8) = 0.1388$

The posterior probability of interest $\Pr(v_6 \mid v_8)$, now, is the quotient of $\Pr(v_6 \wedge v_8)$ and the probability of the evidence $\Pr(v_8)$. This latter probability is obtained by marginalization of $\Pr(V_6 \wedge v_8)$ over V_6. So,

$$\Pr(v_6 \mid v_8) = \frac{\Pr(v_6, v_8)}{\Pr(v_8)} = \frac{\Pr(v_6, v_8)}{\Pr(v_6, v_8) + \Pr(\neg v_6, v_8)} = \frac{0.2557}{0.3945} = 0.6482$$

□

5.3 Sensitivity analysis of a Bayesian network

Sensitivity analysis is a technique to systematically study the effects of variations in the parameters of a mathematical model on this model's outcome. The technique is widely used in the fields of decision theory and mathematical modelling to investigate the possible consequences of inaccuracies in a model's parameters [Habbema *et al.*, 1990, Morgan & Henrion, 1990, Von Winterfeldt & Edwards, 1986]. For a Bayesian network, sensitivity analysis provides for studying the effects of variations in the estimates for an arbitrary subset of parameters in the network on a prior or posterior probability computed from the network. As such, sensitivity analysis of a Bayesian network allows for identifying network parameters that independently or jointly have a large effect on a probability of interest. The results of such an analysis can be used to measure the robustness of the network to parameter variation and to guide refinement of the network by pointing out highly influential parameters. In this chapter, we present a

method for performing sensitivity analysis of a Bayesian network in an efficient way. To that end, we review here relevant work on sensitivity analysis.

In its most general form, in a sensitivity analysis the estimates for n network parameters are varied stepwise in a systematic way, while the estimates for the remaining network parameters are kept fixed. This is termed an n-way sensitivity analysis. It reveals the joint effect of variation of the n parameters under study.

When varying the estimate for a parameter under study, the estimates for the network's parameters that pertain to the same conditional probability distribution have to be adjusted such that the sum of all parameter estimates in this distribution again equals one. The parameters that have to be adjusted are called *co-varying* parameters for the parameter under study. Assume, that we have a parameter under study $x = p(b_i \mid \pi')$ that pertains to the value b_i of the variable B and the configuration π' for the parents of B. The possible values for the variable B are $b_1, \ldots, b_k$, $k \geq 1$. The co-varying parameters for x then are the conditional probabilities $p(b_j \mid \pi')$, $j \neq i$. The sum of the estimates for the co-varying parameters, here called the *residual probability*, equals $1 - x$. Now, in varying the estimate for x, the estimates for these co-varying parameters are adjusted by the ratio of the new residual probability $1 - x$ and the original residual probability $1 - p(b_i \mid \pi')$. Considering $p(b_j \mid \pi')$ as a function of x, denoted $p(b_j \mid \pi')(x)$, we have that

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

If more than one parameter from the same probability distribution is taken as a parameter under study, the residual probability equals one minus the sum of the estimates for these parameters.

Under the assumption that systematic variation of network parameters is carried out as described above, any prior probability can be expressed as a multilinear function in the n network parameters under study; a multilinear function in n parameters is a sum of products of all possible subsets of those parameters. Any posterior probability is a quotient of two multilinear functions [Castillo *et al.*, 1995, Coupé & Van der Gaag, 1998]. The functional relation describing a probability of interest in terms of n parameters under study is termed an n-way sensitivity function.

Suppose, we are interested in the posterior probability $\Pr(a \mid e)$ for some specific value a of variable A and evidence e. Assume, furthermore, that we take for the parameters under study the n network parameters $X = \{x_0, \ldots, x_{n-1}\}$, $n \geq 1$. Recall that, by definition, $\Pr(a \mid e) = \Pr(a \wedge e)/\Pr(e)$. Both the numerator, $\Pr(a \wedge e)$, and the denominator, $\Pr(e)$, in this equation can be expressed as a multilinear relation in the parameters $x_0, \ldots, x_{n-1}$. For each multilinear relation 2^{n-1} coefficients are required. In studying the robustness of $\Pr(a \mid e)$ with respect to variations in the n parameters under

study, it suffices to establish for $\Pr(a \wedge e)$ and $\Pr(e)$ the values of these 2^{n-1} coefficients. In the following, we illustrate determining these coefficients with our running example.

Example 5.3.1 Consider the junction tree of Figure 5.2. As in Example 5.2.4, we assume that for the variable V_8 the value v_8 is observed. The probability of interest again is the probability that the variable V_6 takes the value v_6. In the following, the sensitivity of $\Pr(v_6 \mid v_8)$ to variations in the estimates for the network parameters $x_0 = p(v_8 \mid v_6, v_7)$, $x_1 = p(v_7)$, and $x_2 = p(v_4 \mid \neg v_2)$ is investigated. In the remainder of the chapter, the notation $\Pr(v_6 \mid v_8)(x_0, x_1, x_2)$, refering to the expression for $\Pr(v_6 \mid v_8)$ in terms of x_0, x_1, and x_2, will be omitted if there is no doubt that the functional form rather than the function value is meant. The general form of the relationship between $\Pr(v_6 \mid v_8)$ and x_0, x_1 and x_2 then is,

$$\Pr(v_6 \mid v_8) = \frac{\Pr(v_6, v_8)}{\Pr(v_8)} =$$

$$= \frac{c_0 + c_1 \cdot x_0 + c_2 \cdot x_1 + c_3 \cdot x_0 x_1 + c_4 \cdot x_2 + c_5 \cdot x_0 x_2 + c_6 \cdot x_1 x_2 + c_7 \cdot x_0 x_1 x_2}{d_0 + d_1 \cdot x_0 + d_2 \cdot x_1 + d_3 \cdot x_0 x_1 + d_4 \cdot x_2 + d_5 \cdot x_0 x_2 + d_6 \cdot x_1 x_2 + d_7 \cdot x_0 x_1 x_2}$$

In studying the robustness of $\Pr(v_6 \mid v_8)$ to variations in the estimates for x_0, x_1 and x_2, it suffices to establish the value of all coefficients in the functional relation of $\Pr(v_6 \mid v_8)$. We start with the computation of the coefficients d_j, $j = 0, \ldots, 7$. To that end, we compute the probability $\Pr(v_8)$ from the junction tree for eight different combinations of values for x_0, x_1 and x_2; for the remaining network parameters, the estimates specified in Figure 5.1a are taken. Computing for each selected combination of values for $\{x_0, x_1, x_2\}$ the probability $\Pr(v_8)_{\{x_0,x_1,x_2\}}$, we find for example

$$\begin{aligned}
\Pr(v_8)_{\{0.4,0.4,0.4\}} &= 0.4865 \\
\Pr(v_8)_{\{0.6,0.4,0.4\}} &= 0.5453 \\
\Pr(v_8)_{\{0.4,0.6,0.4\}} &= 0.4665 \\
\Pr(v_8)_{\{0.4,0.4,0.6\}} &= 0.4858 \\
\Pr(v_8)_{\{0.6,0.6,0.4\}} &= 0.5547 \\
\Pr(v_8)_{\{0.6,0.4,0.6\}} &= 0.5452 \\
\Pr(v_8)_{\{0.4,0.6,0.6\}} &= 0.4658 \\
\Pr(v_8)_{\{0.6,0.6,0.6\}} &= 0.5548
\end{aligned}$$

Substituting the left hand side in each of the equations above with $d_0 + d_1 \cdot x_0 + d_2 \cdot x_1 + d_3 \cdot x_0 \cdot x_1 + d_4 \cdot x_2 + d_5 \cdot x_0 \cdot x_2 + d_6 \cdot x_1 \cdot x_2 + d_7 \cdot x_0 \cdot x_1 \cdot x_2$ and filling in the appropriate values for x_0, x_1, and x_2 gives a system of multi-linear equations. Solving

this system, yields the coefficients

$$\begin{aligned} d_0 &= 0.5279 \\ d_1 &= 0 \\ d_2 &= -0.3882 \\ d_3 &= 0.7206 \\ d_4 &= -0.0035 \\ d_5 &= 0 \\ d_6 &= -0.0140 \\ d_7 &= 0.0350 \end{aligned}$$

The coefficients c_i, $i = 0, \ldots, 7$ are determined analogously. We find for c_0, c_1, c_2, c_3, c_4, c_5, c_6, and c_7 the values 0.3603, 0, -0.3603, 0.7206, 0.0175, 0, -0.0175, and 0.0350, respectively. The relationship of $\Pr(v_6 \mid v_8)$ with x_0, x_1 and x_2 thus equals

$$\Pr(v_6 \mid v_8) =$$

$$= \frac{0.3603 - 0.3603 \cdot x_1 + 0.7206 \cdot x_0 x_1 + 0.0175 \cdot x_2 - 0.0175 \cdot x_1 x_2 + 0.0350 \cdot x_0 x_1 x_2}{0.5279 - 0.3882 \cdot x_1 + 0.7206 \cdot x_0 x_1 - 0.0035 \cdot x_2 - 0.0140 \cdot x_1 x_2 + 0.0350 \cdot x_0 x_1 x_2}$$

□

From the example, it is readily seen that a multilinear function in $X = \{x_0, \ldots, x_{n-1}\}$, in general, contains a term for every possible subset $X' \subseteq X$ of these parameters. This term is the product of the parameters in X' and a coefficient c. In the following, we define a coding for the coefficients in an n-way sensitivity function that serves to uniquely identify the subset X' of parameters to which it pertains. In the method for sensitivity analysis, presented in Section 4, this coding is required to trace the parameters under study that are taken into account at any time during propagation in a junction tree. Now, the product of the parameters from the subset X' can be written as a product of *all* parameters in X, by raising the parameters $x_i \in X'$ to the power one and the parameters $x_j \in X \setminus X'$ to the power zero, that is,

$$\prod_{x_i \in X'} x_i = \Big(\prod_{x_i \in X'} x_i^1 \Big) \cdot \Big(\prod_{x_j \in X \setminus X'} x_j^0 \Big)$$

With every possible subset $X' \subseteq X$ thus is associated a set of exponents, one for each parameter in X. The exponent for a parameter x_i is indicated by ϵ_i. To identify that the coefficient c pertains to subset X', a subscript is given to c. To that end, the exponents ϵ_i, $i = 0, \ldots, n-1$, encoding X' in terms of X are placed one after the other, starting from ϵ_{n-1} upto ϵ_0. A binary number, $bin(\epsilon_{n-1}, \ldots, \epsilon_0)$, results. This binary number provides a unique code for the subset X'. The same holds for the decimal number $dec(\epsilon_{n-1}, \ldots, \epsilon_0)$ into which $bin(\epsilon_{n-1}, \ldots, \epsilon_0)$ can be translated. Consider, for example, the subscripts of the coefficients c_i and d_j, $i, j = 0, \ldots, 7$, in Example 5.3.1.

For the product of parameters x_1x_2, we can write $x_0^0x_1^1x_2^1$. Placing the exponents ϵ_2, ϵ_1, and ϵ_0 for x_2, x_1 and x_0, respectively, one after the other, we obtain the binary number 110, which stands for the decimal number 6. The coefficients corresponding with $x_1 \cdot x_2$, thus are indicated by c_6 and d_6.

Using the notations introduced above, the relationship between a posterior probability of interest $\Pr(a \mid e)$ and the parameters $x_0, \ldots, x_{n-1}$ can be written as

$$\Pr(a \mid e) = \frac{\sum_{k:k=dec(\epsilon_{n-1},\ldots,\epsilon_0)} \left(c_k \cdot \prod_{i=0,\ldots,n-1} x_i^{\epsilon_i} \right)}{\sum_{l:l=dec(\epsilon_{n-1},\ldots,\epsilon_0)} \left(c_l \cdot \prod_{j=0,\ldots,n-1} x_j^{\epsilon_j} \right)}$$

5.4 An efficient method for computing n-way sensitivity

In this section, we outline our method for obtaining the coefficients in a sensitivity function expressing the prior probability of the evidence in a Bayesian network in terms of an arbitrary subset of network parameters. As was seen in Example 5.3.1, the computation of the coefficients in the n-way sensitivity function of a posterior probability proceeds analogously; it is built from the n-way sensitivity functions of two prior probabilities. We start, in Section 5.4.1, by giving an example of the method, using the junction tree introduced in Section 5.2. In Section 5.4.2, then, a formal description of the method is given.

5.4.1 An illustration

The method that we present in this section is closely related to standard propagation methods for evidence propagation in a junction tree. Our method can in fact be seen as a variant of these methods. In a junction tree, the probability of the available evidence is computed by performing one inward propagation to some arbitrary clique and subsequent marginalization over all variables from that clique (see Section 5.2). For this inward propagation, standard algorithms send messages taking the form of potential functions. These messages can be represented as tables, containing in each entry a single numerical value. The algorithm we propose in this chapter closely follows the basic idea of these propagation algorithms. In our method, instead of numerical numbers, expressions are being propagated through the junction tree. The idea resembles symbolic propagation; in our method, however, the use of a specific encoding renders the propagation of symbols unnecessary.

The potential function used in standard evidence propagation are tables of reals; in our method, these tables are replaced by tables of vectors of coefficients. For each value of a potential function, the corresponding vector table has an entry that contains a vector of coefficients which expresses this function value in terms of the parameters under study in the junction tree. The messages in our method are the *vector tables* corresponding to the messages in standard evidence propagation. Following the notation introduced in Section 5.2, a local potential function $\phi_{U_i}(U_i)$ for clique U_i has an associated vector table $\psi_{U_i}(U_i)$. A message $\hat{\phi}_{U_i \rightarrow S_{i_j}}(S_{i_j})$ from clique U_i to separator S_{i_j} in standard evidence propagation corresponds to a message $\hat{\psi}_{U_i \rightarrow S_{i_j}}(S_{i_j})$ in our method.

Example 5.4.1 We illustrate the idea of propagating vector tables with the example junction tree in Figure 5.2. We again assume that for the variable V_8 the value v_8 is observed. The prior probability of interest is the probability of the evidence, $\Pr(v_8)$. We are interested in the sensitivity of $\Pr(v_8)$ to variations in the estimates for the three network parameters $x_0 = p(v_8 \mid v_6, v_7)$, $x_1 = p(v_7)$, and $x_2 = p(v_4 \mid \neg v_2)$. To that end, the coefficients in the multilinear function expressing $\Pr(v_8)$ in terms of x_0, x_1 and x_2 are determined. The form of this multilinear function is

$$\begin{aligned} \Pr(e) &= c_0 + c_1 \cdot x_0 + c_2 \cdot x_1 + c_3 \cdot x_0 x_1 + c_4 \cdot x_2 + c_5 \cdot x_0 x_2 \\ &\quad + c_6 \cdot x_1 x_2 + c_7 \cdot x_0 x_1 x_2 \end{aligned} \tag{5.1}$$

Our method starts with an initialization step; for each clique U_i in the junction tree, a local vector table $\psi_{U_i}(U_i)$ is derived from the potential function $\phi_{U_i}(U_i)$. For cliques U_i that do not contain any of the parameters under study, $\psi_{U_i}(U_i)$, in essence, equals $\phi_{U_i}(U_i)$; each function value in $\phi_{U_i}(U_i)$, however, is taken as a vector with one element in $\psi_{U_i}(U_i)$. Since cliques U_1, U_2 and U_4 do not contain any of the three parameters under study, we thus basically have that $\psi_{U_1}(U_1) = \phi_{U_1}(U_1)$, $\psi_{U_2}(U_2) = \phi_{U_2}(U_2)$, and $\psi_{U_4}(U_4) = \phi_{U_4}(U_4)$. Clique U_5 contains the parameters x_0 and x_1. Under the instantiation $V_8 = v_8$, potential function $\phi_{U_5}(U_5)$ in terms of x_0 and x_1 is:

v_8	v_7	$\neg v_7$
v_6	$x_0 \cdot x_1$	$0.5 \cdot (1 - x_1)$
$\neg v_6$	$0.5 \cdot x_1$	$0.6 \cdot (1 - x_1)$

Each entry in this table is a multilinear function in x_0 and x_1. The general shape of the various different entries is $c_0(U_5) + c_1(U_5) \cdot x_0 + c_2(U_5) \cdot x_1 + c_3(U_5) \cdot x_0 x_1$ or, alternatively, $\sum_{k:k=dec(\epsilon_1,\epsilon_0)} \left(c_k(U_5) \cdot \prod_{i=0,1} x_i^{\epsilon_i} \right)$. The notation $c_k(U_5)$, $k = 0, \ldots, 3$, is used to indicate that the coefficient c_k, which is a constant with respect to the parameters x_0 and x_1, may differ for the clique U_5 from configuration to configuration. Each entry in the table can be encoded by a vector of coefficients, $(c_0(U_5), c_1(U_5), c_2(U_5), c_3(U_5))$, where each coefficient pertains to a specific product of parameters from $\{x_0, x_1\}$. The vector table $\psi_{U_5}(U_5)$ corresponding to $\phi_{U_5}(U_5)$ then becomes

v_8	v_7	$\neg v_7$
v_6	$(0, 0, 0, 1)$	$(0.5, 0, -0.5, 0)$
$\neg v_6$	$(0, 0, 0.5, 0)$	$(0.6, 0, -0.6, 0)$

Clique U_3, finally, contains the parameter x_2 under study. Writing the potential function $\phi_{U_3}(U_3)$ in terms of x_2 gives:

	v_2	$\neg v_2$
v_4	0.2	x_2
$\neg v_4$	0.8	$1 - x_2$

Note that, as the potential function $\phi_{U_3}(U_3)$ does not depend on the values of V_3 in U_3, this table holds for both $V_3 = v_3$ and $V_3 = \neg v_3$. Each value of $\phi_{U_3}(U_3)$ is a linear function in x_2, with the general shape $c_0(U_3) + c_1(U_3) \cdot x_2$ or $\sum_{k:k=dec(\epsilon_2)} \left(c_k(U_3) \cdot x_2^{\epsilon_2} \right)$. Encoding each value of $\phi_{U_3}(U_3)$ as a vector of coefficients $(c_0(U_3), c_1(U_3))$ gives the following vector table $\psi_{U_3}(U_3)$:

	v_2	$\neg v_2$
v_4	$(0.2, 0)$	$(0, 1)$
$\neg v_4$	$(0.8, 0)$	$(1, -1)$

Note, that for clique U_3, the local numbering of the coefficients does not coincide with the global numbering of coefficients in Equation 5.1. The subscript 1 of $c_1(U_3)$, for example, would refer to the product $x_0^1 x_1^0 x_2^0 = x_0$ in the global numbering. During propagation of vector tables in the junction tree, therefore, the order number of the highest ordered parameter to which the vectors pertain should be sent along with the table. It serves to identify, at each time during propagation, which parameters of the set of all parameters under study are considered.

Initialization of the junction tree has now taken place. Propagation of vector tables follows. We assume that, from the leaves of the junction tree, an inward propagation towards clique U_4 is performed, that is, clique U_4 is taken as the root of the tree. Each leaf computes from its local vector table the message it should send to its neighbouring separator(s). If this vector table equals the clique's potential function, because it does not contain any parameter under study, this message will basically be the same as in standard evidence propagation. Neither leaf U_1 nor U_2 contain any of the parameters x_0, x_1 or x_2. The messages $\hat{\psi}_{U_1 \to S_1}$ from U_1 to S_1 and $\hat{\psi}_{U_2 \to S_2}$ from U_2 to S_2, therefore, equal $\hat{\phi}_{U_1 \to S_1}$ and $\hat{\phi}_{U_2 \to S_2}$, respectively. So $\hat{\psi}_{U_1 \to S_1}$ equals

	v_3	$\neg v_3$
v_2	(0.63)	(0.18)
$\neg v_2$	(0.15)	(0.04)

and $\hat{\psi}_{U_2 \rightarrow S_2}$ equals

v_3	(1)
$\neg v_3$	(1)

Leaf U_5 contains the parameters x_0 and x_1. To compute the message from clique U_5 to separator S_4 in standard evidence propagation, the potential function of U_5 is projected onto S_4 (cf. `step 3` of the procedure `propagate`). From a conceptual point of view, our method is equivalent to standard propagation in a junction tree. Therefore, to compute the message $\hat{\psi}_{U_5 \rightarrow S_4}$ from U_5 to S_4, the vector table ψ_{U_5} is projected onto S_4. Projection of a vector table amounts to vector summation of the appropriate vectors of coefficients. From the table ψ_{U_5}, the vectors in the entries that specify for V_6 the value v_6 are added and similarly for value $\neg v_6$. The result is

v_6	$(0.5, 0, -0.5, 1)$
$\neg v_6$	$(0.6, 0, -0.1, 0)$

The coefficients in the vector table $\hat{\psi}_{U_5 \rightarrow S_4}(S_4)$ are $c_0(S_4)$, $c_1(S_4)$, $c_2(S_4)$, and $c_3(S_4)$ refering to the multilinear relation $c_0(S_4) + c_1(S_4) \cdot x_0 + c_2(S_4) \cdot x_1 + c_3(S_4) \cdot x_0 x_1$.

For cliques in the junction tree that are not leaves, the local vector table is multiplied with the incoming messages (conform `step 2` of the procedure `propagate`) before projection onto the next separator. Multiplication of a local vector table $\psi_{U_i}(U_i)$ with an incoming message $\hat{\psi}_{S_{i_j} \rightarrow U_i}(S_{i_j})$ proceeds as follows. Assume that we have that $\psi_{U_i}(U_i) = (c_0(U_i), \ldots, c_k(U_i))$ and $\hat{\psi}_{S_{i_j} \rightarrow U_i}(S_{i_j}) = (c_0(S_{i_j}), \ldots, c_m(S_{i_j}))$. For each configuration $u \in \Omega_{U_i}$ and $s \in \Omega_{S_{i_j}}$ with $s \wedge u = u$, the vector product of $\psi_{U_i}(u)$ and $\hat{\psi}_{S_{i_j} \rightarrow U_i}(s)$, indicated by $\psi_{U_i}(u) * \hat{\psi}_{S_{i_j} \rightarrow U_i}(s)$, equals

$$\begin{aligned} \psi_{U_i}(u) * \hat{\psi}_{S_{i_j} \rightarrow U_i}(s) &= (c_0(u), \ldots, c_k(u)) * (c_0(s), \ldots, c_m(s)) = \\ &= (c_0(u) \cdot c_0(s), \ldots, c_0(u) \cdot c_m(s), c_1(u) \cdot c_0(s), \ldots c_k(u) \cdot c_m(s)) \end{aligned}$$

Note that the multiplication is not commutative, that is, $\psi_{U_i}(u) * \hat{\psi}_{S_{i_j} \rightarrow U_i}(s)$ is not necessarily equal to $\hat{\psi}_{S_{i_j} \rightarrow U_i}(s) * \psi_{U_i}(u)$. To obtain the global ordering on coefficients at the root of the junction tree, the following order of multiplying a vector table $\psi_{U_i}(U_i)$ with incoming messages $\hat{\psi}_{S_{i_j} \rightarrow U_i}$, $j = 1, \ldots, t$, is taken. The messages $\hat{\psi}_{S_{i_j} \rightarrow U_i}$ are ordered from $j = 1$ to $j = t$, according to the descending number of their highest ordered parameter. For each configuration $u \in \Omega_{U_i}$, and $s_j \in \Omega_{S_{i_j}}$, $j = 1, \ldots, t$, with $s_j \wedge u = u$,

the multiplication is defined as follows:

$$\psi_{U_i}(u) * \prod_{j=1,\ldots,t}{}^{*} \hat{\psi}_{S_{i_j} \to U_i}(s_j) =$$

$$= (\cdots((\psi_{U_i}(u) * \hat{\psi}_{S_{i_1} \to U_i}(s_1)) * \hat{\psi}_{S_{i_2} \to U_i}(s_2)) * \cdots * \hat{\psi}_{S_{i_{t-1}} \to U_i}(s_t))$$

Consider now clique U_3 in our running example. The local vector table ψ_{U_3} is multiplied with the message $\hat{\psi}_{S_1 \to U_3}$ coming from S_1. Since the message $\hat{\psi}_{S_1 \to U_3}$ comes from a part of the junction tree where no parameters under study reside, this vector table contains in each entry a single coefficient $c(S_1)$. For each configuration $u \in \Omega_{U_3}$ and $s \in \Omega_{S_1}$ with $u \wedge s = u$ equals:

$$\begin{aligned} \psi_{U_3}(u) * \hat{\psi}_{S_1 \to U_3}(s) &= \psi_{U_3}(u) * \hat{\psi}_{U_1 \to S_1}(s) \\ &= (c_0(u), c_1(u)) * (c(s)) = \\ &= (c_0(u) \cdot c(s), c_1(u) \cdot c(s)) = \\ &= (c'_0(u), c'_1(u)) \end{aligned}$$

For example, the vector $(0.15, -0.15)$ in the product for the configuration $\neg v_2, v_3, \neg v_4$ is obtained after multiplication of the vector $(1, -1)$ in $\psi_{U_3}(\neg v_2, v_3, \neg v_4)$ with the coefficient 0.15 in $\hat{\psi}_{U_1 \to S_1}(\neg v_2, v_3)$; note that this is equivalent to multiplying the linear function $1 - x_2$ with the constant 0.15. The vector table $\psi_{U_3} * \hat{\psi}_{U_1 \to S_1}$ equals:

	v_2		$\neg v_2$	
	v_3	$\neg v_3$	v_3	$\neg v_3$
v_4	$(0.126, 0)$	$(0.036, 0)$	$(0, 0.15)$	$(0, 0.04)$
$\neg v_4$	$(0.504, 0)$	$(0.144, 0)$	$(0.15, -0.15)$	$(0.04, -0.04)$

To obtain the message $\hat{\psi}_{U_3 \to S_3}$, finally, this vector table is projected onto S_3, giving the following table with in each entry a vector of coefficients $(c_0(S_3), c_1(S_3))$:

	v_3	$\neg v_3$
v_4	$(0.126, 0.15)$	$(0.036, 0.04)$
$\neg v_4$	$(0.654, -0.15)$	$(0.184, -0.04)$

At clique U_4, now, all messages from neighbouring separators are available. This clique can therefore compute the coefficients in the n-way sensitivity function expressing $\Pr(e)$ in terms of the parameters x_0, x_1, and x_2 under study. To that end, the local vector table ψ_{U_4} is multiplied, subsequently, with the messages $\hat{\psi}_{S_2 \to U_4}$, $\hat{\psi}_{S_3 \to U_4}$, and $\hat{\psi}_{S_4 \to U_4}$. As the message $\hat{\psi}_{U_2 \to S_2}$ equals unity, it has no effect on ψ_{U_4}. Multiplication with the messages $\hat{\psi}_{S_3 \to U_4}$ and $\hat{\psi}_{S_4 \to U_4}$ is carried out in the order of decreasing number of the highest ordered parameter. Thus, ψ_{U_4} is first multiplied by $\hat{\psi}_{S_3 \to U_4}$. For each

configuration $u \in \Omega_{U_4}$ and $s \in \Omega_{S_3}$ with $s \wedge u = u$, $\psi_{U_4}(u) = (c(u))$ is multiplied by $\hat{\psi}_{S_3 \to U_4}(s) = (c_0(s), c_1(s))$ giving $\psi'_{U_4}(u) = (c'_0(u), c'_1(u)) = (c(u) \cdot c_0(s), c(u) \cdot c_1(s))$

	v_4		$\neg v_4$	
	v_3	$\neg v_3$	v_3	$\neg v_3$
v_6	(0.1134, 0.135)	(0.0288, 0.032)	(0.5232, −0.12)	(0.0552, −0.012)
$\neg v_6$	(0.0126, 0.015)	(0.0072, 0.008)	(0.1308, −0.03)	(0.1288, −0.028)

Note that the coefficients in $\psi'_{U_4}(U_4)$ refer to a linear relation in x_2.

Multiplying the vector table ψ'_{U_4} with $\hat{\psi}_{S_4 \to U_4}$, finally, proceeds as follows. For each configuration $u \in \Omega_{U_4}$ and $s \in \Omega_{S_4}$ with $s \wedge u = u$, we compute

$$\begin{aligned}
\psi''_{U_4}(u) &= \psi'_{U_4}(u) * \hat{\psi}_{S_4 \to U_4}(s) = \\
&= (c'_0(u) \cdot c_0(s)\ ,\ c'_0(u) \cdot c_1(s)\ ,\ c'_0(u) \cdot c_2(s)\ ,\ c'_0(u) \cdot c_3(s)\ , \\
&\qquad c'_1(u) \cdot c_0(s)\ ,\ c'_1(u) \cdot c'_1(s)\ ,\ c'_1(u) \cdot c_2(s)\ ,\ c'_1(u) \cdot c_3(s)) \\
&= (c''_0(u)\ ,\ c''_1(u)\ ,\ c''_2(u)\ ,\ c''_3(u)\ ,\ c''_4(u)\ ,\ c''_5(u)\ ,\ c''_6(u)\ ,\ c''_7(u))
\end{aligned}$$

This vector refers to the coefficients yielded by multiplication of the two multilinear functions

$$\sum_{k:k=dec(\epsilon_2)} \left(c'_k(u) \cdot x_i^{\epsilon_2}\right) \cdot \sum_{l:l=dec(\epsilon_1,\epsilon_0)} \left(c_l(s) \cdot \prod_{i=0,1} x_i^{\epsilon_i}\right) =$$

$$= \sum_{\substack{k\,:\,k\,=\,dec(\epsilon_2) \\ l\ :\ l\ = \\ dec(\epsilon_1,\epsilon_0)}} \left(c'_k(u) \cdot c_l(s) \cdot \prod_{i=0,1,2} x_i^{\epsilon_i}\right) =$$

$$= \sum_{k:k=dec(\epsilon_2,\epsilon_1,\epsilon_0)} \left(c''_k(u) \cdot \prod_{i=0,1,2} x_i^{\epsilon_i}\right)$$

Note that each subscript for the new coefficients $c''_k(u)$, $k = 0, \ldots 7$, refers to the product of a specific subset of parameters in $\{x_0, x_1, x_2\}$. This is achieved by placing the binary number of a coefficient in ψ'_{U_4} to the right of the binary number of the coefficient in $\hat{\psi}_{U_5, \to S_4}$ with which it is multiplied. The vector table $\psi''_{U_4}(U_4)$ resulting from this multiplication now equals

	v_4		$\neg v_4$	
	v_3	$\neg v_3$	v_3	$\neg v_3$
v_6	(0.057, 0, −0.057, 0.113, 0.068, 0, −0.068, 0.135)	(0.014, 0, −0.014, 0.029, 0.016, 0, −0.016, 0.032)	(0.262, 0, −0.262, 0.523, −0.06, 0, 0.06, −0.12)	(0.028, 0, −0.028, 0.055, −0.006, 0, 0.006, −0.012)
$\neg v_6$	(0.0076, 0, −0.0013, 0, 0.009, 0, −0.0015, 0)	(0.0013, 0, −0.0007, 0, 0.0048, 0, −0.0008, 0)	(0.078, 0, −0.013, 0, −0.018, 0, 0.003, 0)	(0.077, 0, −0.0013, 0, −0.017, 0, 0.0028, 0)

The vector table $\psi''_{U_4}(U_4)$ now contains the coefficients in the multilinear function expressing $\Pr(V_3, V_4, V_6, v_8)$ in terms of the parameters x_0, x_1, and x_2. To obtain from

$\psi''_{U_4}(U_4)$ the function expressing $\Pr(v_8)$ in terms of x_0, x_1, and x_2, all entries of $\psi''_{U_4}(U_4)$ are added. This gives the vector $(0.5279, 0, -0.3882, 0.7206, -0.0035, 0, -0.0140, 0.0350)$ which indicates that

$$\Pr(v_8) = 0.5279 - 0.3882 \cdot x_1 + 0.7206 \cdot x_0x_1 - 0.0035 \cdot x_2 \\ -0.0140 \cdot x_1x_2 + 0.0350 \cdot x_0x_1x_2$$

□

5.4.2 Formal description of the method

In this section, we formally detail our method for computing the coefficients in the n-way sensitivity function expressing a prior probability of evidence in terms of the parameters under study $x_0, \ldots, x_{n-1}$. We begin by introducing some concepts and notational conventions. Then, we sketch the method in pseudocode. From this sketch, it will be evident that the method closely resembles standard evidence propagation. We therefore focus attention on the steps in our method that differ from standard evidence propagation. We show that this difference basically lies in the representation of information. At any time during propagation, a translation of potential functions into vector tables and vice versa can be made.

The following notation is used. As in Section 5.2, for a clique U_i in a junction tree T, the neighbouring separators are indicated by S_{i_j}, $j = 1, \ldots, t+1$; we assume, without loss of generality that $S_{i_{t+1}}$ is the separator to which U_i is to send a message and that the incoming messages $\hat{\psi}_{S_{i_j} \to U_i}$, $j = 1, \ldots, t$, are ordered according to the descending number of their highest ordered parameter. For root R of junction tree T, we indicate the neighbouring separators by S_{R_j}, $j = 1, \ldots, k$. The coefficients in the n-way sensitivity function to be computed are indicated by $(c_0, \ldots, c_{2^n-1})$. R is used to indicate the root of T. The notation T_{U_i} indicates the subtree of T rooted at U_i; that is, U_i lies on the (unique) path from R to any clique $U_j \in T_{U_i}$.

The parameters under study in T are ordered according to a *parameter ordering.*

Definition 5.4.2 *Let T be a junction tree with the cliques $U(T)$ and let $R \in U(T)$ be the root of T. Let $X = \{x_0, \ldots, x_{n-1}\}$ be the set of parameters under study. Let $o : X \longleftrightarrow \{0, \ldots, n-1\}$ be a total ordering on X, such that*

- *for any two parameters $x_i, x_j \in X$ with x_i in U and x_j in U', for which it holds that U' lies on the (unique) path from R to U, we have $o(x_i) < o(x_j)$;*
- *for any clique U, for the subset of parameters $X' \subseteq X$ located in T_U, we have that $\max\{o(x_i) \mid x_i \in X'\} - \min\{o(x_i) \mid x_i \in X'\} = |X'|$;*

Then, o is a parameter ordering *of X with respect to R.*

Informally speaking, the parameters are ordered in such a way that their order numbers increase from the leaves towards the root and that at any clique the total set of parameters in the subtree rooted at this clique is consecutively numbered. A parameter ordering of the parameters under study in a junction tree is not unique; within a clique, the parameters under study can be ordered arbitrarily and between cliques in general also various different possibilites exist. The purpose of a parameter ordering is to allow for identifying to which subset of parameters the coefficients in a vector table pertain at each point during propagation. When successively multiplying a local vector table with incoming vector tables, the binary subscripts of the incoming coefficients are placed to the right of the binary subscript of each coefficient resulting from the multiplication so far. For every coefficient at the root then, automatically, a subscript is obtained whose binary number indicates to which subset of parameters under study it pertains. This idea was also presented and applied in Example 5.4.1.

Now, in pseudocode our method amounts to the following.

```
        procedure vector-propagation (T, e, R, X)
        begin
step 1    enter evidence e into junction tree T;
step 2    order the set X of n parameters under study from 0 to n − 1;
step 3    for each node U_i in T do   (initialization)
            derive ψ_{U_i} from φ_{U_i}
          od;
          until root R in T is triggered do
            for each triggered clique U_i do
step 4        for each configuration u ∈ Ω_{U_i} do
                for each configuration s_j ∈ Ω_{S_{i_j}}
                with s_j ∧ u = u, j = 1,...t, do
                  compute ψ'_{U_i}(u) = ψ_{U_i}(u_i) * ⊛_{j=1,...,t} ψ̂_{S_{i_j}→U_i}(s_j)
                od
              od;
step 5        project ψ'_{U_i}(U_i) onto S_{i_{t+1}}, giving ψ̂_{U_i→S_{i_{t+1}}}(S_{i_{t+1}});
step 6        send ψ̂_{U_i→S_{i_{t+1}}} to S_{i_{t+1}},
                with the number of the highest ordered parameter
            od
          od;
step 7    for each configuration r ∈ Ω_R do
            for each configuration s_j ∈ Ω_{S_{R_j}}
            with s_j ∧ r = r, j = 1,...k, do
              compute ψ'_R(r) = ψ_R(r) * ⊛_{j=1,...,k} ψ̂_{S_{R_j}→R}(s_j)
            od
          od;
step 8    compute (c_0, ..., c_{2^n−1}) = Σ_{r∈Ω_R} ψ'_R(r)
        end
```

Using the parameter ordering as introduced in Definition 5.4.2, we first focus on the initialization phase (conform step 3 of the procedure vector-propagation). In the initialization phase, the vector table ψ_{U_i} for clique U_i is derived from the local potential function ϕ_{U_i}. We formally define ψ_{U_i} and then show that ϕ_{U_i} and ψ_{U_i} basically contain the same information; filling in the values for the parameters under study in U_i in the multilinear relation represented by the coefficients in ψ_{U_i} gives ϕ_{U_i}.

Definition 5.4.3 *Let T be a junction tree and let U_i be a clique in T. Let ϕ_{U_i} be U_i's local potential function. Let $X = \{x_0, \ldots, x_{m-1}\}$ be the set of parameters under study that are located in U_i, $m \geq 1$; for each parameter x_j, $j = 0, \ldots, m-1$, let p_j be the value for x_j as specified in T. Let $\{\epsilon_0, \ldots, \epsilon_{m-1}\}$ be the set of exponents that is associated with $\{x_0, \ldots, x_{m-1}\}$, as introduced in Section 5.3. For each configuration $u \in \Omega_{U_i}$, the set X is partitioned into three subsets of parameters X^u_{var}, X^u_{co-var}, and X^u_{other}, where for each $x_j = p(v \mid \pi)$, $j = 0, \ldots, m-1$, we have that*

$$
\begin{array}{lll}
x_j \in X^u_{var} & if & v \wedge \pi \wedge u = u \\
x_j \in X^u_{co-var} & if & v \wedge u = false \text{ and } \pi \wedge u = u \\
X^u_{other} = X \setminus (X^u_{var} \cup X^u_{co-var}) & &
\end{array}
$$

Then, the vector table ψ_{U_i} of U_i is defined, for each configuration $u \in \Omega_{U_i}$, by

$$\psi_{U_i}(u) = (c_0(u), \ldots, c_{2^m-1}(u))$$

where for each k with $k = dec(\epsilon_{m-1}, \ldots, \epsilon_0)$ we have that

$$
c_k(u) = \begin{cases} 0 & \text{, if there is an } x_i \in X^u_{var} \text{ with } \epsilon_i = 0 \\ & \quad \text{or an } x_i \in X^u_{other} \text{ with } \epsilon_i = 1 \end{cases}
$$

and

$$
c_k(u) = (-1)^{D(k)} \cdot \frac{\phi_{U_i}(u)}{(\prod_{x_j \in X^u_{var}} p_j) \cdot (\prod_{x_l \in X^u_{co-var}} (1 - p_l))} \quad \text{, otherwise}
$$

where $D(k) = \sum_{j: x_j \in X^u_{co-var}} \epsilon_j$.

We recall that the coefficients in a vector $\psi_{U_i}(u)$ specify the multilinear relation of the potential function value $\phi_{U_i}(u)$ in terms of the parameters $x_0, \ldots, x_{m-1}$. Informally speaking Definition 5.4.3 now states that a coefficient $c_k(u)$ in $\psi_{U_i}(u)$ equals zero if it pertains to a subset of parameters that cannot co-occur in the configuration u of U_i. As the configuration u matches all the parameters from X^u_{var}, any coefficient indicating a subset of parameters that does not contain all the parameters from X^u_{var} equals zero. Similarly, as configuration u does *not* match any parameter from X^u_{other}, any coefficient for a subset of parameters that does include a parameter from X^u_{other} equals zero. The remaining coefficients are obtained from the potential function value $\phi_{U_i}(u)$ by factorizing out the parameter values of the parameters under study that concord with u. Since the values of all parameters from X^u_{var} are contained in $\phi_{U_i}(u)$, these parameter values are factorized out by simply dividing $\phi_{U_i}(u)$ by $\prod_{x_j \in X^u_{var}} p_j$. For each parameter from X^u_{co-var}, the parameter value of a co-varying parameter is contained in $\phi_{U_i}(u)$. Recall from Section 5.3 that for a parameter x_l taking value p_l, the value p'_l

of a specific co-varying parameter equals $p_l'(x) = p_l' \cdot \frac{1-x}{1-p} = f_l \cdot (1-x)$. To factorize out of $\phi_{U_i}(u)$ the parameter values for parameters x_l from X^u_{co-var}, we thus divide by $\prod_{x_l \in X^u_{co-var}} (1-p_l)$. Note that in doing this, the fraction f_l for the co-varying parameter of $x_l \in X^u_{co-var}$ matching u is contained in $c_k(u)$. The sign of a coefficient c_k depends on the number of parameters from X^u_{co-var} in the subset of parameters to which it pertains.

In the following proposition, we will show that ψ_{U_i} in essence is equivalent to ϕ_{U_i}. Filling in, for a specific configuration $u \in \Omega_{U_i}$, the *values* of the parameters under study residing in U_i in the multilinear relation represented by $\psi_{U_i}(u)$ yields the potential function value $\phi_{U_i}(u)$. Therefore, ϕ_{U_i} and $\psi_{U_i}(u)$ basically give the same information.

Proposition 5.4.4 *Let T be a junction tree and let U_i be a clique in T. Let ϕ_{U_i} be U_i's local potential function. Let $X = \{x_0, \ldots, x_{m-1}\}$ be the set of parameters under study that are located in U_i, $m \geq 1$; for each parameter x_j, $j = 0, \ldots, m-1$, let p_j be the value for x_j as specified in T. Let $\psi_{U_i}(U_i) = (c_0(U_i), \ldots, c_{2^m-1}(U_i))$ be U_i's vector table as defined above. Then, for each configuration $u \in \Omega_{U_i}$, we have that*

$$\phi_{U_i}(u) = \sum_{k:k=dec(\epsilon_{m-1},\ldots,\epsilon_0)} \left(c_k(u) \cdot \prod_{j=0,\ldots,m-1} p_j^{\epsilon_j} \right)$$

Proof. Without loss of generality, we assume that the subsets X^u_{var}, X^u_{co-var}, and X^u_{other} of X equal $\{x_0, \ldots x_{f-1}\}$, $\{x_f, \ldots x_{h-1}\}$, and $\{x_h, \ldots x_{m-1}\}$, for some $f, h \geq 1$, respectively. For each configuration $u \in \Omega_{U_i}$, we use the notation $f_{U_i}(u)(x_0, \ldots, x_{m-1})$ to indicate the multilinear function in of $x_0, \ldots, x_{m-1}$ specified by $\psi_{U_i}(u)$; that is,

$$f_{U_i}(u)(x_0, \ldots, x_{m-1}) = \sum_{k:k=dec(\epsilon_{m-1},\ldots,\epsilon_0)} \left(c_k(u) \cdot \prod_{j=0,\ldots,m-1} x_j^{\epsilon_j} \right)$$

We will now show that $f_{U_i}(u)(p_0, \ldots, p_{m-1}) = \phi_{U_i}(u)$. By substituting in the function $f_{U_i}(u)$ every coefficient $c_k(u)$ with the expression for $c_k(u)$ as defined in Definition 5.4.3 and subsequently filling in the values p_j for the parameters x_j, $j = 0, \ldots, m-1$, we find

$$f_{U_i}(u)(p_0, \ldots, p_{m-1}) =$$

$$= \sum_{\substack{k\,:\,k=dec(\epsilon_{m-1},\ldots,\epsilon_0),\\ \epsilon_0,\ldots,\epsilon_{f-1}=1,\\ \epsilon_h,\ldots,\epsilon_{m-1}=0}} \left((-1)^{D(k)} \cdot \frac{\phi_{U_i}(u)}{\left(\prod_{x_r \in X^u_{var}} p_r\right) \cdot \left(\prod_{x_s \in X^u_{co-var}} (1-p_s)\right)} \cdot \prod_{j=0,\ldots,m-1} p_j^{\epsilon_j} \right)$$

where $D(k) = \sum_{j:x_j \in X^u_{co-var}} \epsilon_j = \epsilon_f + \cdots + \epsilon_{h-1}$.

Now, since the exponents $\epsilon_h, \ldots, \epsilon_{m-1}$ equal zero, the product $\prod_{j=h,\ldots,m-1} p_j^{\epsilon_j}$ equals one. Multiplication with one has no effect, so this product can be left out from the

expression for $f_{U_i}(u)(p_0,\ldots,p_{m-1})$. Substituting, furthermore, X^u_{var} and X^u_{co-var} with $\{x_0,\ldots x_{f-1}\}$, and $\{x_f,\ldots x_{h-1}\}$, respectively, gives

$$f_{U_i}(u)(p_0,\ldots,p_{m-1}) =$$

$$= \sum_{\substack{k\,:\,k=dec(\epsilon_{h-1},\ldots,\epsilon_0),\\ \epsilon_0,\ldots,\epsilon_{f-1}=1}} \left(\frac{(-1)^{(\epsilon_f+\cdots+\epsilon_{h-1})}\cdot\phi_{U_i}(u)}{(\prod_{r=0,\ldots,f-1} p_r)\cdot(\prod_{s=f,\ldots,h-1}(1-p_s))} \cdot \prod_{j=0,\ldots,h-1} p_j^{\epsilon_j} \right)$$

Since the exponents $\epsilon_0,\ldots,\epsilon_{f-1}$ equal one, we have that $\prod_{j=0,\ldots,f-1} p_j^{\epsilon_j} = \prod_{r=0,\ldots,f-1} p_j$. So,

$$f_{U_i}(u)(p_0,\ldots,p_{m-1}) =$$

$$= \left(\sum_{k:k=dec(\epsilon_{h-1},\ldots,\epsilon_f)} \left((-1)^{(\epsilon_f+\cdots+\epsilon_{h-1})} \cdot \frac{\phi_{U_i}(u)}{(\prod_{s=f,\ldots,h-1}(1-p_s))} \cdot \prod_{j=f,\ldots,h-1} p_j^{\epsilon_j} \right) \right)$$

$$= \frac{\phi_{U_i}(u)}{(\prod_{s=f,\ldots,h-1}(1-p_s))} \cdot \left(\sum_{k:k=\mathcal{N}(\epsilon_{h-1},\ldots,\epsilon_f)} \left((-1)^{(\epsilon_f+\cdots+\epsilon_{h-1})} \cdot \prod_{j=f,\ldots,h-1} p_j^{\epsilon_j} \right) \right)$$

We focus attention on the term $\sum_{k:k=dec(\epsilon_{h-1},\ldots,\epsilon_f)} \left((-1)^{(\epsilon_f+\cdots+\epsilon_{h-1})} \cdot \prod_{j=f,\ldots,h-1} p_j^{\epsilon_j} \right)$ and show that it equals $\prod_{s=f,\ldots,h-1}(1-p_s)$. The product $\prod_{s=f,\ldots,h-1}(1-p_s)$ can be written as a sum of products. Writing P for the set $\{p_f,\ldots,p_{h-1}\}$, each product in this sum is composed from parameter values from P and a constant which equals either one or minus one. The constant pertaining to a specific subset $P' \subseteq P$ equals one if the cardinality of P', denoted $\|P'\|$, is even; it equals minus one if $\|P'\|$ is odd. So, the constant equals $(-1)^{\|P'\|}$. As was introduced in Section 5.3, the product of a subset of parameter values P' can be written as a product of all parameter values in P where each $p_l \in P'$ is raised to the power one, $\epsilon_l = 1$, and each $p_l \notin P'$ is raised to the power zero, $\epsilon_l = 0$. Writing $\|P'\|$ in terms of the exponents, that is, $\|P'\| = \epsilon_f + \cdots + \epsilon_{h-1}$, we find

$$\prod_{l=f,\ldots,h-1}(1-p_l) = \sum_{k:k=dec(\epsilon_{h-1},\ldots,\epsilon_f)} \left((-1)^{(\epsilon_f+\cdots+\epsilon_{h-1})} \cdot \prod_{j=f,\ldots,h-1} p_j^{\epsilon_j} \right)$$

We conclude that

$$f_{U_i}(u)(p_0,\ldots,p_{m-1}) = \phi_{U_i}(u)$$

□

In the proof of Proposition 5.4.4, we have shown that, for each configuration $u \in \Omega_{U_i}$, the vector table $\psi_{U_i}(u)$ for clique U_i contains the coefficients describing $\phi_{U_i}(u)$ in terms of the parameters under study in U_i. During propagation, the vector table ψ_{U_i} is multiplied by various other vector tables. Before proceeding, we formally define the multiplication of vector tables, called the *vector product*; the vector product was informally introduced in Example 5.4.1.

Definition 5.4.5 *Let each W_i, $i = 1, \ldots, t$, $t \geq 1$, be a set of variables and let each $\psi_i(W_i)$, $i = 1, \ldots, t$, be a vector table over the variables in W_i. The vector product $\prod\!\!*_{i=1,\ldots,t} \psi_i(W_i)$ of these vector tables equals*

$$\prod\!\!*_{i=1,\ldots,t} \psi_i(W_i) = (\cdots((\psi_1(W_1) * \psi_2(W_2)) * \psi_3(W_3)) * \cdots * \psi_t(W_t))$$

where for every two vector tables $\psi_i(W_i)$ and $\psi_j(W_j)$ with $\psi_i(W_i) = (c_{i_0}(W_i), \ldots, c_{i_k}(W_i))$ and $\psi_j(W_j) = (c_{j_0}(W_j), \ldots, c_{j_l}(W_j))$, for every configuration $w_i \in \Omega_{W_i}$ and $w_j \in \Omega_{W_j}$ the vector product $\psi_i(w_i) * \psi_j(w_j)$ equals

$$\begin{aligned}
&\psi_i(w_i) * \psi_j(w_j) = \\
&= (c_{i_0}(w_i), \ldots, c_{i_k}(w_i)) * (c_{j_0}(w_j), \ldots, c_{j_l}(w_j)) \\
&= (c_{i_0}(w_i) \cdot c_{j_0}(w_j), \ldots, c_{i_0}(w_i) \cdot c_{j_l}(w_j), c_{i_1}(w_i) \cdot c_{j_0}(w_j), \ldots, c_{i_k}(w_i) \cdot c_{j_l}(w_j))
\end{aligned}$$

This vector product is used in multiplying vector table ψ_{U_i} for clique U_i with the incoming messages $\hat{\psi}_{S_{i_j} \to U_i}$ from cliques S_{i_j}, $j = 1, \ldots, t$ (conform `step` 4 of procedure `vector-propagate`). The resulting vector table ψ'_{U_i} basically contains the same information as the potential function ϕ'_{U_i} resulting from standard propagation; for each configuration $u \in \Omega_{U_i}$, $\psi'_{U_i}(u)$ contains precisely the coefficients in the multilinear function describing the potential function value $\phi'_{U_i}(u)$ in terms of the parameters under study in subtree T_{U_i}. Filling in the specified values for the parameters in T_{U_i} in the multilinear function represented by $\psi'_{U_i}(u)$ gives $\phi'_{U_i}(u)$. Below, this is detailed for a clique U_i receiving one incoming message.

Proposition 5.4.6 *Let T be a junction tree and let U_i be a clique in T. Let ϕ_{U_i} be U_i's local potential function. Let $X = \{x_0, \ldots, x_{m-1}\}$, $m \geq 1$, be the set of parameters under study that are located in T_{U_i}, ordered as in Definition 5.4.2; for each parameter x_j, $j = 0, \ldots, m-1$, let p_j be the value for x_j as specified in T. Let ψ_{U_i} be the vector table associated with ϕ_{U_i}. Let S be a neighbouring separator for U_i, let $\hat{\phi}_{S \to U_i}(S)$ be the message from S to U_i in standard propagation and let $\hat{\psi}_{S \to U_i}(S)$ be the vector table associated with $\hat{\phi}_{S \to U_i}(S)$. For each configuration $u \in \Omega_{U_i}$ and $s \in \Omega_S$ with $s \wedge u = u$, let $\phi'_{U_i}(u) = \phi_{U_i}(u) \cdot \hat{\phi}_{S \to U_i}(s)$ and let $\psi'_{U_i}(u) = \psi_{U_i}(u) * \hat{\psi}_{S \to U_i}(s)$ Then,*

$$\phi'_{U_i}(u) = \sum_{k : k = dec(\epsilon_{m-1}, \ldots, \epsilon_0)} \left(c'_k(u) \cdot \prod_{j=0,\ldots,m-1} p_j^{\epsilon_j} \right)$$

where

$(c'_0(u), \ldots, c'_{2^m-1}(u)) = \psi'_{U_i}(u)$

Proof. Without loss of generality, we assume the parameters under study in clique U_i to be $x_f, \ldots, x_{m-1}$ and the parameters in $T_{U_i} \backslash U_i$ to be $x_0, \ldots, x_{f-1}$. We assume that, for each configuration $u \in \Omega_{U_i}$, $\psi_{U_i}(u) = (c_0(u), \ldots, c_{2^{m-f}-1}(u))$ and, for each configuration $s \in \Omega_S$, $\hat{\psi}_{S \to U_i}(s) = (c_0(s), \ldots, c_{2^f-1}(s))$. Furthermore, for each configuration $u \in \Omega_{U_i}$, we use $f_{U_i}(u)(x_0, \ldots, x_{m-1})$ to indicate the multilinear function in $x_0, \ldots, x_{m-1}$ specified by $\psi'_{U_i}(u)$, that is,

$$f_{U_i}(u)(x_0, \ldots, x_{m-1}) = \sum_{l : l = dec(\epsilon_{m-1}, \ldots, \epsilon_0)} \left(c'_l(u) \cdot \prod_{j=0,\ldots,m-1} x_j^{\epsilon_j} \right)$$

where $(c'_0(u), \ldots, c'_{2^m-1}(u)) = \psi'_{U_i}(u)$. We will now show that $f_{U_i}(u)(p_0, \ldots, p_{m-1}) = \phi'_{U_i}(u)$. From $\psi'_{U_i}(u) = \psi_{U_i}(u) * \hat{\psi}_{S \to U_i}(s)$ we find that

$$\begin{aligned} \psi'_{U_i}(u) &= (c'_0(u), \ldots, c'_{2^m-1}(u)) = \\ &= (c_0(u), \ldots, c_{2^{m-f}-1}(u)) * (c_0(s), \ldots, c_{2^f-1}(s)) = \\ &= (c_0(u) \cdot c_0(s), \ldots, c_0(u) \cdot c_{2^f-1}(s), c_1(u) \cdot c_0(s), \ldots, c_{2^{m-f}-1}(u) \cdot c_{2^f-1}(s)) \end{aligned}$$

or alternatively,

$c'_l(u) = c_k(u) \cdot c_{k'}(s)$

where for each $k = dec(\epsilon_{m-1}, \ldots, \epsilon_f)$ and $k' = dec(\epsilon_{f-1}, \ldots, \epsilon_0)$, l equals $dec(\epsilon_{m-1}, \ldots, \epsilon_0)$. The expression $f_{U_i}(u)(x_0, \ldots, x_{m-1})$ can therefore be written as

$$f_{U_i}(u)(x_0, \ldots, x_{m-1}) =$$

$$= \sum_{k : k = dec(\epsilon_{m-1}, \ldots, \epsilon_f)} \sum_{k' : k' = dec(\epsilon_{f-1}, \ldots, \epsilon_0)} \left(c_k(u) \cdot c_{k'}(s) \cdot \left(\prod_{j=f,\ldots,m-1} x_j^{\epsilon_j} \right) \cdot \left(\prod_{l=0,\ldots,f-1} x_l^{\epsilon_l} \right) \right)$$

$$= \left(\sum_{k : k = dec(\epsilon_{m-1}, \ldots, \epsilon_f)} \left(c_k(u) \cdot \prod_{j=f,\ldots,m-1} x_j^{\epsilon_j} \right) \right) \cdot \left(\sum_{k' : k' = dec(\epsilon_{f-1}, \ldots, \epsilon_0)} \left(c_{k'}(s) \cdot \prod_{l=0,\ldots,f-1} x_l^{\epsilon_l} \right) \right)$$

Filling in the value p_j for parameter x_j, $j = 0, \ldots m-1$, we find for $f_{U_i}(u)(p_0, \ldots, p_{m-1})$ that

$f_{U_i}(u)(p_0, \ldots, p_{m-1}) =$

$$\left(\sum_{k : k = dec(\epsilon_{m-1}, \ldots, \epsilon_f)} \left(c_k(u) \cdot \prod_{j=f,\ldots,m-1} p_j^{\epsilon_j} \right) \right) \cdot \left(\sum_{k' : k' = dec(\epsilon_{f-1}, \ldots, \epsilon_0)} \left(c_{k'}(s) \cdot \prod_{l=0,\ldots,f-1} p_l^{\epsilon_l} \right) \right)$$

From proposition 5.4.4, we know that the two factors of this multiplication are the two mulilinear relations expressing $\phi_{U_i}(u)$ and $\hat{\phi}_{S\to U_i}(s)$ in terms of the parameters $x_f, \ldots, x_{m-1}$ and $x_0, \ldots, x_{f-1}$, respectively. So,

$$\begin{aligned} f_{U_i}(u)(p_0, \ldots, p_{m-1}) &= \phi_{U_i}(u) \cdot \hat{\phi}_{S\to U_i}(s) \\ &= \phi'_{U_i}(u) \end{aligned}$$

□

In Proposition 5.4.6, we have shown that for a clique U_i receiving an incoming message $\hat{\psi}_{S\to U_i}$, the vector product of U_i's vector table ψ_{U_i} with $\hat{\psi}_{S\to U_i}$, that is ψ'_{U_i}, gives for each configuration $u \in \Omega_{U_i}$ the coefficients describing ϕ_{U_i} in terms of the parameters under study in subtree T_{U_i}. Now, proposition 5.4.6 can be extended to apply to multiple incoming messages. For a clique U_i receiving the messages $\hat{\psi}_{S_{i_j}\to U_i}$, $j = 1, \ldots t$, for each configuration $u \in \Omega_{U_i}$ and $s_j \in \Omega_{S_{i_j}}$, with $s_j \wedge u = u$ and $j = 1, \ldots, t$, the vector product $\psi_{U_i}(u) * \prod_{j=1,\ldots,t} \hat{\psi}_{S_{i_j}\to U_i}(s_j)$ gives ψ'_{U_i} describing ϕ'_{U_i} in terms of the parameters in T_{U_i}. Since this extension is straightforward, we will not detail this any further.

In step 5 of the procedure `vector-propagation`, the vector table ψ'_{U_i} is projected onto the separator $S_{i_{t+1}}$ to which a message is to be sent. This step corresponds to step 3 of the standard propagation procedure. In standard propagation, the potential function ψ'_{U_i} is projected onto separator $S_{i_{t+1}}$ by marginalizing over all variables that are not contained in $S_{i_{t+1}}$. Similarly, ψ'_{U_i} is projected onto $S_{i_{t+1}}$ by marginalizing over all variables that are not in $S_{i_{t+1}}$. For ψ'_{U_i}, this amounts to adding the vectors of coefficients in the entries specifying the same configuration for $S_{i_{t+1}}$. Adding vectors of coefficients is done by straightforward vector summation. The result of projection onto $S_{i_{t+1}}$ of ϕ'_{U_i} and ψ'_{U_i} is the message $\hat{\phi}_{U_i\to S_{i_{t+1}}}(s)$ and $\hat{\psi}_{U_i\to S_{i_{t+1}}}(s)$, respectively. Below, we show that, for a specific configuration $s \in \Omega_{S_{i_{t+1}}}$, $\hat{\psi}_{U_i\to S_{i_{t+1}}}(s)$ contains the coefficients that describe $\hat{\phi}_{U_i\to S_{i_{t+1}}}(s)$ in terms of the parameters under study in T_{U_i}; filling in the specified values for the parameters in T_{U_i} in the multilinear function represented by $\hat{\psi}_{U_i\to S_{i_{t+1}}}(s)$ gives $\hat{\phi}_{U_i\to S_{i_{t+1}}}(s)$.

Proposition 5.4.7 *Let T be a junction tree and let U_i be a clique in T. Let $X = \{x_0, \ldots, x_{m-1}\}$, $m \geq 1$, be the parameters under study in subtree T_{U_i} ordered as in Definition 5.4.2; for each parameter x_j, $j = 0, \ldots, m-1$, let p_j be the value for x_j as specified in T. Let $\phi'_{U_i}(U_i)$ be a potential function for U_i and let $\psi'_{U_i}(U_i)$ be the vector table associated with $\phi'_{U_i}(U_i)$. Let S be a neighbouring separator of U_i and let $\hat{\phi}_{U_i\to S}(S) = \sum_{U_i\setminus S} \phi'_{U_i}(U_i)$ be the messages from U_i to S. Let $\hat{\psi}_{U_i\to S}(S)$ be $\sum_{U_i\setminus S} \psi'_{U_i}(U_i)$. Then, for each configuration $s \in \Omega_S$, we have that*

$$\hat{\phi}_{U_i\to S}(s) = \sum_{k:k=dec(\epsilon_{m-1},\ldots,\epsilon_0)} \left(c_k(s) \cdot \prod_{j=0,\ldots,m-1} p_j^{\epsilon_j} \right)$$

where

$(c_0(s), \ldots, c_{2^m-1}(s)) = \hat{\psi}_{U_i \to S}(s)$

Proof. For each configuration $s \in \Omega_S$, we use $f_S(s)(x_0, \ldots, x_{m-1})$ to indicate the multilinear function in terms of $x_0, \ldots, x_{m-1}$ specified by the coefficients in $\hat{\psi}_{U_i \to S}(s)$, that is

$$f_S(s)(x_0, \ldots, x_{m-1}) = \sum_{k:k=dec(\epsilon_{m-1},\ldots,\epsilon_0)} \left(c_k(s) \cdot \prod_{j=0,\ldots,m-1} x_j^{\epsilon_j} \right) \tag{5.2}$$

where $(c_0(s), \ldots, c_{2^m-1}(s)) = \hat{\psi}_{U_i \to S}(s)$. We will now show that $f_S(s)(p_0, \ldots, p_{m-1}) = \hat{\phi}_{U_i \to S}(s)$. By definition, the vector table $\hat{\psi}_{U_i \to S}(s)$ equals $\sum_{u' \in \Omega_{U_i \setminus S}} \psi'_{U_i}(u' \wedge s)$ Without loss of generality, we assume that $\psi'_{U_i}(u) = (c_0(u), \ldots, c_{2^m-1}(u))$. The coefficients $c_k(s)$, $k = 0, \ldots 2^m - 1$ can be written as

$$\begin{aligned}(c_0(s), \ldots, c_{2^m-1}(s)) &= \sum_{u' \in \Omega_{U_i \setminus S}} (c_0(u' \wedge s), \ldots, c_{2^m-1}(u' \wedge s)) \\ &= \left(\sum_{u' \in \Omega_{U_i \setminus S}} c_0(u' \wedge s), \ldots, \sum_{u' \in \Omega_{U_i \setminus S}} c_{2^m-1}(u' \wedge s) \right)\end{aligned}$$

that is

$$c_k(s) = \sum_{u' \in \Omega_{U_i \setminus S}} c_k(u' \wedge s)$$

for $k = 0, \ldots 2^m - 1$. Substituting this expression for $c_k(s)$ in Equation 5.2, we find for $f_S(s)(x_0, \ldots, x_{m-1})$ that

$$\begin{aligned}f_S(s)(x_0, \ldots, x_{m-1}) &= \sum_{k:k=dec(\epsilon_{m-1},\ldots,\epsilon_0)} \left(\left(\sum_{u' \in \Omega_{U_i \setminus S}} c_k(u' \wedge s) \right) \cdot \left(\prod_{j=0,\ldots,m-1} x_j^{\epsilon_j} \right) \right) \\ &= \sum_{k:k=dec(\epsilon_{m-1},\ldots,\epsilon_0)} \left(\sum_{u' \in \Omega_{U_i \setminus S}} \left(c_k(u' \wedge s) \cdot \prod_{j=0,\ldots,m-1} x_j^{\epsilon_j} \right) \right) \\ &= \sum_{u' \in \Omega_{U_i \setminus S}} \left(\sum_{k:k=dec(\epsilon_{m-1},\ldots,\epsilon_0)} \left(c_k(u' \wedge s) \cdot \prod_{j=0,\ldots,m-1} x_j^{\epsilon_j} \right) \right)\end{aligned}$$

By definition, the multilinear relation between brackets is the expression of $\phi'_{U_i}(u)$ in terms of the parameters $x_0, \ldots x_{m-1}$. Filling in the value p_j for parameter x_j, $j = 0, \ldots m-1$, we find for $f_S(s)(p_0, \ldots, p_{m-1})$

$$\begin{aligned}f_S(s)(p_0, \ldots, p_{m-1}) &= \sum_{u' \in \Omega_{U_i \setminus S}} \phi'_{U_i}(u' \wedge s) \\ &= \hat{\phi}_{U_i \to S}(s)\end{aligned}$$

□

After one inward propagation of vector tables from the leaves to the root R of a junction tree T, at root R all incoming messages are available. Now, in step 7 of procedure `vector-propagation`, the vector table ψ_R of R is multiplied with these messages. This corresponds to step 5 in procedure `propagate`. For a specific configuration $r \in \Omega_R$, the resulting vector $\psi'_R(r)$ contains the coefficients that describe $\Pr(r, e)$ in terms of all parameters under study in T. Since this step is very similar to step 4 in procedure `vector-propagation`, we will omit a formal description of this multiplication.

Finally, in the last step, step 8, of procedure `vector-propagation`, all vectors in ψ'_R are added. This corresponds, in standard evidence propagation, to the computation of the probability $\Pr(e)$ from the marginal probability distribution $\Pr(R, e)$ over the variables in R. The result of the summation $\sum_{r \in \Omega_R} \psi'_R(r)$ in procedure `vector-propagation` thus is the vector of coefficients $(c_0, \ldots, c_{2^n-1})$ describing $\Pr(e)$ in terms of all parameters $x_0, \ldots, x_{n-1}$ under study in the junction tree. Technically, this last step is the same as step 5 in `vector-propagation`; it will not be further detailed.

5.5 Finding the optimal root

In the previous section, it is shown that the coefficients in an n-way sensitivity function expressing the probability of the evidence in terms of n of the network's parameters can be established by performing one inward propagation of vector tables towards any root. The location in the junction tree of the clique that is chosen as the root of procedure `vector-propagate` determines the number of computations that is required to obtain these coefficients. We call that clique the *optimal root*, for which the number of computations to be performed is minimized. The optimal root is determined with respect to a specific computation scheme for establishing the required coefficients. In Section 5.4, we have presented the scheme `vector-propagate`. More efficient computation schemes exist; in particular with respect to the order of multiplying incoming messages and marginalizing over variables, optimizations of the scheme `vector propagate` can be achieved. In Section 5.6, possibilities for optimization are discussed briefly. The root which is optimal for our computation scheme `vector-propagate` is not necessarily optimal for other computation schemes. In this section, we discuss the identification of the optimal root for procedure `vector-propagate`. We first introduce the basic idea; subsequently, the idea is illustrated using our running example.

For procedure `vector-propagate`, the number of computations that is required to obtain the coefficients in a sensitivity function from a root R in junction tree T equals the sum of the number of computations required for each step in this procedure. In establishing the total number of computations for a root R, we assume that initialization of the junction tree has already taken place; the number of computations required for the initialization is independent of the chosen root and is therefore not relevant in

determining the optimal root. The total number of computations $N_R(T)$ required if R is the root of `vector-propagate` then is the sum of the number of computations performed at each clique U_i in `steps` 4 and 5 of procedure `vector-propagate` and the computations at root R in `steps` 7 and 8 of `vector-propagate`.

The number of computations $N(U_i)$ for a clique U_i consist of the number of multplications $M(U_i)$ required for multiplying ψ_{U_i} with the messages $\hat{\psi}_{S_{ij} \to U_i}$, $j = 1, \ldots, t$, (`step` 4) and the number of additions $A(U_i)$ required for projecting ψ'_{U_i} onto $S_{i_{t+1}}$ (`step` 5). Both concern floating point operations, in the sequel called *flops*. From Example 5.4.1, it will be clear that the number of flops $N(U_i)$ depends on the size of clique U_i, denoted $size(U_i)$, the number of parameters under study in U_i, denoted $n(U_i)$, and the number of parameters under study in the incoming vector tables $\hat{\psi}_{S_{ij} \to U_i}$, denoted $n(S_{ij}, U_i)$, $j = 1, \ldots, t$. This latter information is not locally available at clique U_i; separators S_{ij}, $j = 1, \ldots, t$, provide this information. For every clique U_i in T, where $U_i \neq R$, $N(U_i)$ is determined. In a similar way, for root R the number of flops $N(R)$ is determined. $N(R)$ again depends on $size(R)$, $n(R)$ and $n(S_{R_j}, R)$, $j = 1, \ldots, k$. The total number of flops $N_R(T)$ to obtain the required coefficients from R then equals $N(R) + \sum_{U_i \in T \setminus R} N(U_i)$.

To know the number of parameters under study $n(S_{R_j}, R)$, $j = 1, \ldots, k$, in the incoming vector tables of R, each neighbouring clique U_j of R should send the number of parameters under study in subtree T_{U_j} to R. This can be formulated recursively; each clique U_i in T sends the number of parameters under study in subtree T_{U_i}, denoted $n(U_i, S_{i_{t+1}})$, to $S_{i_{t+1}}$. As such, the computation of $N_R(T)$ itself can be formulated recursively, that is, the idea of propagation in a junction tree is again exploited. Starting from the leaves of T, each triggered clique U_i sends to $S_{i_{t+1}}$ the number of parameters under study in subtree T_{U_i} and the number of flops done so far. The number of parameters $n(U_i, S_{i_{t+1}})$ is obtained by adding $n(U_i)$ to the incoming numbers of parameters $n(S_{ij}, U_i)$, $j = 1, \ldots, t$. The number of flops done so far, denoted $N(U_i, S_{i_{t+1}})$, is the sum of $N(U_i)$ and the incoming numbers of flops $N(S_{ij}, U_i)$, $j = 1, \ldots, t$. In going from the leaves of T to R, now, the total number of flops $N_R(T)$ is accumulated. In pseudocode, the computation of $N_R(T)$ amount to the following.

```
procedure compute-flops (T, e, R, X)
begin
   until root R in T is triggered do
      for each triggered clique U_i do
step 1      compute N(U_i) = M(U_i) + A(U_i);
step 2      compute N(U_i, S_{i_{t+1}}) = N(U_i) + Σ_{j=1,...t} N(S_{i_j}, U_i);
step 3      compute n(U_i, S_{i_{t+1}}) = n(U_i) + Σ_{j=1,...t} n(S_{i_j}, U_i);
step 4      send N(U_i, S_{i_{t+1}}) and n(U_i, S_{i_{t+1}}) to S_{i_{t+1}}
      od
   od;
step 5   compute N(R) = M(R) + A(R);
step 6   compute N_R(T) = N(R) + Σ_{j=1,...k} N(S_{R_j}, R);
end.
```

By taking every clique in T as the root of procedure `compute-flops`, the clique that minimizes the number of computations to establish the coefficients in the required sensitivity function is found. This clique is the optimal root for procedure `vector propagate`. Note that, instead of repeating `compute-flops` for every clique in T, it is more efficient to extend the procedure such that $N_{U_i}(T)$ is computed for all cliques U_i simultaneously. To this end, the propagation of numbers of flops and numbers of parameters is not stopped when the root of `compute-flops` is reached, but continues until a full inward and outward propagation has been performed. Then, at each clique U_i in T the incoming numbers of flops and parameters are given and this clique can compute the total number of flops $N_{U_i}(T)$ required to establish all coefficients from U_i. Since the idea is relatively straightforward, it will not be detailed further. Below, we illustrate the computation of the optimal root for our running example.

Example 5.5.1 Consider again our running example in Figure 5.2. As in Example 5.4.1, we focus on the probability $\Pr(v_8)$ of the evidence $V_8 = v_8$ and we are interested in the sensitivity of $\Pr(v_8)$ to variations in the estimates for the three network parameters $x_0 = p(v_8 \mid v_6, v_7)$, $x_1 = p(v_7)$, and $x_2 = p(v_4 \mid \neg v_2)$. To that end, the coefficients in the multilinear function expressing $\Pr(v_8)$ in terms of x_0, x_1 and x_2 are determined from procedure `vector-propagate`. Before applying procedure `vector-propagate`, however, the optimal root to start the procedure is established from procedure `compute-flops`. First, we compute the number of flops if the required coefficients are established using clique U_4 as root of T; we perform `compute-flops` $(T, v_8, U_4, \{x_0, x_1, x_2\})$.

From the leaves of the junction tree, an inward propagation towards clique U_4 is

performed. Each leaf U_i computes from its size and its number of parameters under study the number of flops required to compute the vector table it should send to its neighbouring separator. Leaf U_1 does not contain any parameter under study; $n(U_1)$ equals zero. The size of U_1, $size(U_1)$, is the product of the number of values for the variables in U_1; $size(U_1) = 2 \cdot 2 \cdot 2 = 8$. From $n(U_1)$ and $size(U_1)$, it is known that vector table ψ_{U_1} consists of eight entries each containing one coefficient ($2^{n(U_1)}$). Projecting ψ_{U_1} onto S_1 amounts to marginalizing over variable V_1 with $size(V_1) = 2$, giving the vector table $\hat{\psi}_{U_1 \to S_1}$ with $size(S_1) = 4$. The number of additions $A(U_1)$ required for this operation equals four. Leaf U_1 thus sends to separator S_1 the messages;

$$N(U_1, S_1) = A(U_1) = 4$$

and

$$n(U_1, S_1) = n(U_1) = 0$$

Similarly, clique U_2 sends to S_2 the information that

$$N(U_2, S_2) = A(U_1) = 2$$

and

$$n(U_2, S_2) = n(U_1) = 0$$

Leaf U_5 contains the two parameters x_0 and x_1. Since V_8 is observed, the size of clique U_5 is four. The vector table ψ_{U_5} thus consists of four entries each containing four coefficients. Projecting ψ_{U_5} onto separator S_4, with $size(S_4) = 2$, then requires eight additions; $A(U_5) = 8$. We thus find

$$N(U_5, S_4) = A(U_5) = 8$$

and

$$n(U_5, S_4) = n(U_5) = 2$$

For a clique U_i that is not a leaf, the vector table ψ_{U_i} should be multiplied with the incoming vector tables before projection onto the separator to which a message is to be sent. The number of flops performed by U_i then is the sum of the number of multiplications $M(U_i)$ required for this first step and the number of additions $A(U_i)$ for the latter. Consider clique U_3 containing the parameter x_2; $n(U_3) = 1$. The size of clique U_3 is eight. The vector table ψ_{U_3} thus consist of eight entries each containing two ($2^{n(U_3)}$) coefficients. ψ_{U_3} is to be multiplied with the incoming vector table $\hat{\psi}_{S_1 \to U_3}$, which equals $\hat{\psi}_{U_1 \to S_1}$ and contains $n(U_1, S_1) = 0$ parameters under study. The two coefficients in each of the eight entries in ψ_{U_3} have to be multiplied with the one coefficient in the appropriate entry in $\hat{\psi}_{S_1 \to U_3}$. As such, $M(U_3)$ equals $8 \cdot 2 \cdot 1 = 16$.

The resulting table ψ'_{U_3}, again consisting of eight entries with each two coefficients, is projected onto S_3, with $size(S_3) = 4$. This requires eight additions; $A(U_3) = 8$. For the total number of local computations at clique U_3 we thus find

$$N(U_3) = M(U_3) + A(U_3) = 16 + 8 = 24$$

Adding this to the incoming number of flops $N(S_1, U_3) = N(U_1, S_1) = 4$ gives

$$N(U_3, S_3) = N(U_3) + N(S_1, U_3) = 28$$

The total number of parameters under study accumulated so far is

$$n(U_3, S_3) = n(U_3) + n(S_1, U_3) = 1$$

At clique U_4, the root of procedure `compute-flops`, now, for each neighbouring separator S_2, S_3, and S_4, the number of flops required to compute the incoming messages $\hat{\psi}_{S_2 \to U_4}$, $\hat{\psi}_{S_3 \to U_4}$, and $\hat{\psi}_{S_4 \to U_4}$ in procedure `vector-propagate` is available. Furthermore, U_4 receives the number of parameters $n(S_2, U_4)$, $n(S_3, U_4)$, and $n(S_4, U_4)$ in each of these messages. At root U_4 then the number of multiplications $M(U_4)$ required to multiply the vector table ψ_{U_4} with these incoming vector tables can be determined. We assume that ψ_{U_4} is multiplied with the incoming vector tables in order of increasing total number of parameters in each table. That is, we consider first multiplication with $\hat{\psi}_{S_2 \to U_4}$, then with $\hat{\psi}_{S_3 \to U_4}$, and finally with $\hat{\psi}_{S_4 \to U_4}$. Recall that $size(U_4) = 8$ and $n(U_4) = 0$. The message $\hat{\psi}_{S_2 \to U_4}$ contains no parameters under study. As such, for each of the eight entries in ψ_{U_4}, each single coefficient is multiplied with the one coefficient in the appropriate entry in $\hat{\psi}_{S_2 \to U_4}$. This requires $8 \cdot 1 \cdot 1 = 8$ multiplications. Message $\hat{\psi}_{S_3 \to U_4}$ contains one parameter under study, corresponding to two coefficients in each entry. Subsequent multiplication with this message thus requires $8 \cdot 1 \cdot 2 = 16$ multiplications. The vector table resulting from the multiplication so far has eight entries each containing two coefficients; it is multiplied with $\hat{\psi}_{S_4 \to U_4}$. This requires $8 \cdot 2 \cdot 4 = 64$ multiplications. In total thus $M(U_4) = 88$. After the multiplication of ψ_{U_4} with all incoming vector tables, again a vector table with eight entries results. Each entry now contains $2^{n(U_4)} \cdot 2^{n(S_2,U_4)} \cdot 2^{n(S_3,U_4)} \cdot 2^{n(S_4,U_4)} = 8$ coefficients. Marginalizing over all variables in U_4 thus requires $(size(U_4) - 1) \cdot 8 = 56$ additions; $A(U_4) = 56$. For root U_4 thus a total of $N(U_4) = M(U_4) + A(U_4) = 88 + 56 = 144$ flops are required. Adding this to the incoming number of flops gives

$$N_{U_4}(T) = N(U_4) + N(S_2, U_4) + N(S_3, U_4) + N(S_4, U_4) = 144 + 2 + 28 + 8 = 182$$

The total number of parameters under study in T evidently equals

$$n(T) = n(U_4) + +n(S_2, U_4) + n(S_3, U_4) + n(S_4, U_4) = 0 + 0 + 1 + 2 = 3$$

Now, procedure `compute-flops` has been carried out with clique U_4 as root.

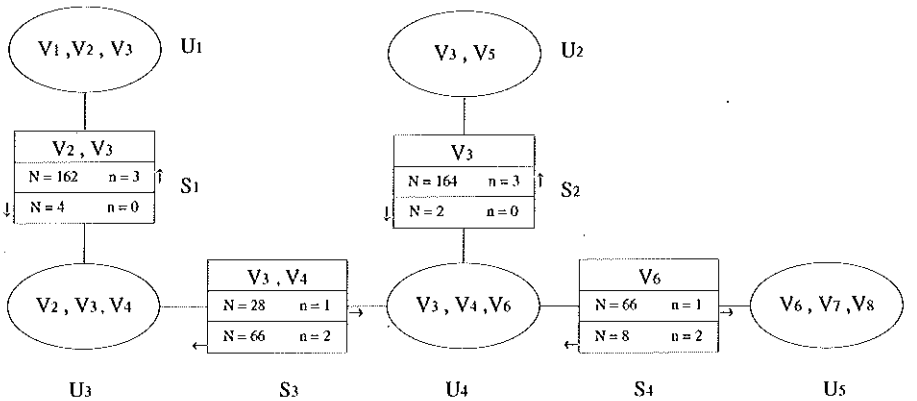


Figure 5.3: The example junction tree from Figure 5.2, in which one full propagation of numbers of floating point operations (flops) and numbers of parameters has been performed. At each clique, then, the number of flops required to compute each incoming vector table is available.

By repeating the procedure for every clique in the junction tree T in Figure 5.2, for each clique U_i, $i = 1, \ldots, 5$ the number of flops $N_{U_i}(T)$ required to compute the coefficients in the three-way sensivity function of $\Pr(v_8)$ in terms of the parameters $x_0 = p(v_8 \mid v_6, v_7)$, $x_1 = p(v_7)$, and $x_2 = p(v_4 \mid \neg v_2)$ is obtained. As was noted earlier, instead of repeating `compute-flops` for every clique U_i in T, the number of flops can also be computed for all cliques simultaneously. This is done by performing a full (inward and outward) propagation of numbers of flops and numbers of parameters. In Figure 5.3, the result of such a full propagation is shown. At each clique, now, the number of flops required to compute all incoming messages is available. With this information each clique U_i computes the total number of flops $N_{U_i}(T)$ required to compute all coefficients in the three-way sensitivity function. In Table 5.1, the numbers $N_{U_i}(T)$, $i = 1, \ldots 5$, are shown .

Table 5.1 shows that clique U_5 requires the minimum number of flops for computing all coefficients in the three-way sensitivity function expressing $Pr(v_8)$ in terms of x_0, x_1, and x_2, when the computation scheme of procedure `vector-propagate` is used.

Now, we show that by optimizing procedure `vector-propagate` slightly, another optimal root results. In procedure `vector-propagate`, the vector table ψ_{U_i} of U_i is first multiplied with all incoming messages before marginalization takes place. It is more efficient, however, to alternate the multiplication by incoming messages with marginalization over variables as soon as this is possible. For clique U_4, for example, it is possible to marginalize over variables V_3 and V_4 after ψ_{U_4} is multiplied with $\hat{\psi}_{S_2 \to U_4}$ and $\hat{\psi}_{S_3 \to U_4}$. Marginalization over these two variables can be carried out be-

clique U_i	$N_{U_i}(T)$
U_1	282
U_2	220
U_3	206
U_4	182
U_5	122

Table 5.1: The number of floating point operations (flops) required to compute the coefficients in the three-way sensitivity function of $\Pr(v_8)$ in terms of the parameters $x_0 = p(v_8 \mid v_6, v_7)$, $x_1 = p(v_7)$, and $x_2 = p(v_4 \mid \neg v_2)$. The computation scheme used is procedure `vector-propagate`.

fore multiplication with $\hat{\psi}_{S_4 \to U_4}$ since this message is not a function of these variables. Multiplication of ψ_{U_4} with $\hat{\psi}_{S_2 \to U_4}$ and $\hat{\psi}_{S_3 \to U_4}$ requires $M_1(U_4) = 24$ multiplications. Subsequent marginalization over V_3 and V_4 requires $A_1(U_4) = 12$ additions. Now, the resulting vector table with two entries each containing two coeffcients is multiplied with $\hat{\psi}_{S_4 \to U_4}$ and marginalized over V_6. This requires $M_2(U_4) = 16$ multiplications and $A_1(U_4) = 8$ additions. In alternating multiplications and additions at U_4 thus a total of $N'(U_4) = 60$ flops are required. In that case, $N'_{U_4}(T) = 98$ compared to $N'_{U_4}(T) = 182$ in Table 5.1. Assuming that multiplication by incoming messages and marginalization over variables is carried out alternatingly, we find for each clique U_i, $i = 1, \ldots, 5$, the number of flops $N'_{U_i}(T)$ as given in Table 5.2. The minimum number of flops is now found for clique U_4. □

clique U_i	$N'_{U_i}(T)$
U_1	282
U_2	164
U_3	150
U_4	98
U_5	122

Table 5.2: The number of floating point operations (flops) required to compute the coefficients in the three-way sensitivity function of $\Pr(v_8)$ in terms of the parameters $x_0 = p(v_8 \mid v_6, v_7)$, $x_1 = p(v_7)$, and $x_2 = p(v_4 \mid \neg v_2)$. The computation scheme used is procedure `vector-propagate` in which order of multiplying incoming messages and marginalizing over variables is optimized.

5.6 Optimizations

In Section 5.4, we have presented the basic idea of computing the coefficients in a sensitivity function by propagating vector tables through the junction tree. In pseudocode, procedure `vector-propagate` presents a computation scheme for obtaining these coefficients. Although it reflects the basic idea, this scheme is not optimized with respect to computation time. In this section, we will discuss several possibilities for increasing the efficiency of the basic computation scheme. First, we will focus on efficient computation of the messages that are propagated through the junction tree. It concerns optimising step 4 and 5 in procedure `vector-propagate`. This was also briefly touched upon in Section 5.5. Subsequently, we discuss the acquisition of a factorized form of the sensitivity function under study. A factorized form of the sensitivity function can, for example, be obtained by omitting the computations at the root of the junction tree (step 7 and 8 in procedure `vector-propagate`). The number of computations required for a factorized sensitivity function is smaller than for the unfactorized form and the information contained in the sensivity function is equivalent. Finally, we focus on exploiting independence relations induced by evidence to reduce the number of computations. Firstly, if there is evidence that causes the junction tree to fall apart in two or more independent subgraphs, the sensitivity function is a product of the functions obtained from each subgraph separately. Secondly, evidence for separator variables may allow for postponing the multiplication with incoming vector tables. In the following, the above mentioned possibilities for improvement will be discussed briefly without going into technical details.

In procedure `vector-propagate`, the message of a clique U_i to separator $S_{i_{t+1}}$ is computed by first multiplying ψ_{U_i} with each incoming message $\hat{\psi}_{S_{i_j} \to U_i}$ for separator S_{i_j}, $j = 1, \ldots, t$, and subsequently projecting the resulting vector table onto $S_{i_{t+1}}$. In Section 5.5, it was already noted that first carrying out all multiplications and after this all marginalizations is not efficient. It is far more efficient to marginalize over a variable as soon as this is possible. We can marginalize over a variable when ψ_{U_i} has been multiplied with all incoming messages that are a function of this variable. In doing this, the total number of entries in a vector table that is stepwise multiplied with incoming messages is kept at a minimum, thereby reducing the number of multiplications required for subsequent incoming messages. Note that the idea of alternating multiplications and marginalizations is not limited to its use in propagating vector tables. It also increases the efficiency of standard propagation. This has been shown by [Shenoy, 1996]. To alternate multiplications and marginalizations in an efficient way, P. Shenoy proposes to construct from a junction tree a binary junction tree in which each clique receives at most three incoming messages.

A second way to increase the efficiency of computing message $\hat{\psi}_{U_i \to S_{i_{t+1}}}$ from U_i to $S_{i_{t+1}}$ is to order the incoming messages according to increasing total amount of

their parameters; ψ_{U_i} is first multiplied with the incoming vector table that contains the smallest total number of parameters and for subsequent incoming vector tables this number increases. Using this constraint on the ordering of incoming vector tables decreases the number of multiplications in step 4 of procedure `vector-propagate`. Basically, incorporating this constraint means that the ordering of incoming messages according to descending number of their highest ordered parameter, as assumed in Section 5.4, corresponds to increasing total number of their parameters. In the parameter ordering o defined in Section 5.4, this additional constaint can be incorporated by enforcing that for each two neighbouring cliques of U, denoted U' and U'', for which it holds that U lies on the path from U' and U'' to R and for which hold that X' and X'' denote the subsets of parameters contained in $T_{U'}$ and $T_{U''}$, respectively, we have that $\max\{o(x_i) \mid x_i \in X'\} > \max\{o(x_i) \mid x_i \in X''\}$ if $|X'| < |X''|$. Now, using this latter constraint in combination with the use of binary junction trees allows for computing $\hat{\psi}_{U_i \rightarrow S_{i_{t+1}}}$ efficiently.

Until now, we have focussed on computing the 2^n coefficients in an unfactorized n-way sensitivity function. Instead of computing these 2^n coefficients, it is also possible to obtain the coefficients for a factorized form of the sensitivity function. First, we discuss the possibility to establish a factorized n-way sensitivity function from the root of procedure `vector-propagate`. Basically, a factorized n-way sensitivity function at root R is obtained by omitting any computations at R; as such, the vector table ψ_R and the incoming messages $\hat{\psi}_{S_{R_j} \rightarrow U_i}$, $j = 1, \ldots, k$, are kept as they are. The factors in the factorized n-way sensitivity function are the multilinear relations established by the coefficient in each entry of these vector tables. The sensitivity function is a sum of products of the appropriate factors. We will illustrate this using Example 5.4.1 in Section 5.4.

Example 5.6.1 Consider again the three-way sensitivity analysis discussed in Example 5.4.1. We are interested in the sensitivity of the probability of the evidence $\Pr(v_8)$ to variations in the estimates for the three network parameters $x_0 = p(v_8 \mid v_6, v_7)$, $x_1 = p(v_7)$, and $x_2 = p(v_4 \mid \neg v_2)$. To that end, in Example 5.4.1, the coefficients in the unfactorized multilinear function expressing $\Pr(v_8)$ in terms of x_0, x_1 and x_2 are determined from root U_4. Here, we focus on the factorized form of the sensitivity function that is built from the multilinear functions established by the coefficients in vector table ψ_{U_4} and the incoming vector tables $\hat{\psi}_{S_2 \rightarrow U_4}$, $\hat{\psi}_{S_3 \rightarrow U_4}$, and $\hat{\psi}_{S_4 \rightarrow U_4}$. For each configuration $u \in \Omega_{U_4}$, the multilinear function expressing the probability $\Pr(u, v_8)$ in terms of x_0, x_1, and x_2 is the product of the multilinear functions established by the coefficients in $\psi_{U_4}(u)$, $\hat{\psi}_{S_2 \rightarrow U_4}(s)$, $\hat{\psi}_{S_3 \rightarrow U_4}(s')$ and $\hat{\psi}_{S_3 \rightarrow U_4}(s'')$ where $u \wedge s \wedge s' \wedge s'' = u$. Using the vector tables given in Example 5.4.1, we find for $u = v_3, v_4, v_6$

$$\Pr(v_3, v_4, v_6, v_8) = 0.9 \cdot (0.126 + 0.15x_2) \cdot (0.5 - 0.5x_1 + x_0x_1)$$

Adding the factorized expressions of $\Pr(u, v_8)$ for all $u \in \Omega_{U_4}$, we find the factorized multilinear function expressing $\Pr(v_8)$ in terms of x_0, x_1, and x_2;

$$\begin{aligned}
\Pr(v_8) &= \Pr(v_3, v_4, v_6, v_8) + \Pr(\neg v_3, v_4, v_6, v_8) + \Pr(v_3, \neg v_4, v_6, v_8) \\
&\quad + \Pr(\neg v_3, \neg v_4, v_6, v_8) + \Pr(v_3, v_4, \neg v_6, v_8) + \Pr(\neg v_3, v_4, \neg v_6, v_8) \\
&\quad + \Pr(v_3, \neg v_4, \neg v_6, v_8) + \Pr(\neg v_3, \neg v_4, \neg v_6, v_8) \\
&= 0.9 \cdot (0.126 + 0.15x_2) \cdot (0.5 - 0.5x_1 + x_0x_1) \\
&\quad + 0.8 \cdot (0.036 + 0.04x_2) \cdot (0.5 - 0.5x_1 + x_0x_1) \\
&\quad + 0.8 \cdot (0.654 - 0.15x_2) \cdot (0.5 - 0.5x_1 + x_0x_1) \\
&\quad + 0.3 \cdot (0.184 - 0.04x_2) \cdot (0.5 - 0.5x_1 + x_0x_1) \\
&\quad + 0.1 \cdot (0.126 + 0.15x_2) \cdot (0.6 - 0.1x_1) \\
&\quad + 0.2 \cdot (0.036 + 0.04x_2) \cdot (0.6 - 0.1x_1) \\
&\quad + 0.2 \cdot (0.654 - 0.15x_2) \cdot (0.6 - 0.1x_1) \\
&\quad + 0.7 \cdot (0.184 - 0.04x_2) \cdot (0.6 - 0.1x_1)
\end{aligned}$$

□

From the example, we see that a factorized form of our three-way sensitivity function contains more coefficients than the unfactorized three-way sensitivity function. However, for a larger number of parameters under study, in general, the factorized form of the n-way sensitivity function will contain less coeffcients. Moreover, the number of computations required to obtain the factorized form is smaller without any loss of information.

In general, for a factorized sensitivity function to be useful for interpretation it should not contain too many different factors. This can be established by taking as a root a clique with a limited number of incoming messages and with a limited size. Another possibility is to take a sufficiently small separator as root. In that case, we perform an inward propagation to this separator and compose the factorized sensitivity function from the messages coming from both directions. We again illustrate this with the sensitivity analysis presented in Example 5.4.1.

Example 5.6.2 Consider again the three-way sensitivity analysis of $\Pr(v_8)$ with respect to variations in the estimates for the three network parameters $x_0 = p(v_8 \mid v_6, v_7)$, $x_1 = p(v_7)$, and $x_2 = p(v_4 \mid \neg v_2)$, as discussed in Example 5.4.1. In Example 5.4.1, an inward propagation to U_4 is performed. With the incoming messages $\hat{\psi}_{S_2 \to U_4}$, and $\hat{\psi}_{S_3 \to U_4}$ at U_4, this clique can compute its message $\hat{\psi}_{U_4 \to S_4}$ to separator S_4. This is done by marginalizing the product $\psi'_{U_4} = (\psi_{U_4} * \hat{\psi}_{S_2 \to U_4}) * \hat{\psi}_{S_3 \to U_4}$ over the variables V_3 and V_4. From the vector table ψ'_{U_4} given in Example 5.4.1, it can be seen that $\hat{\psi}_{U_4 \to S_4}$ equals

$V_6 = v_6$	$(0.7206, 0.035)$
$V_6 = \neg v_6$	$(0.2794, -0.035)$

Now, from this message and the message $\hat{\psi}_{U_5 \to S_4}$, at S_4 a factorized form of the three-way sensitivity function expressing $\Pr(v_8)$ in terms of x_0, x_1 and x_2 can be composed;

$$\begin{aligned}\Pr(v_8) &= \Pr(v_6, v_8) + \Pr(\neg v_6, v_8) \\ &= (0.7206 + 0.035x_2) \cdot (0.5 - 0.5x_1 + x_0x_1) + \\ &\quad + (0.2794 - 0.035x_2) \cdot (0.6 - 0.1x_1)\end{aligned}$$

□

Obtaining a factorized form of the sensitivity function by omitting multiplication with incoming messages, as was done above at the root of a junction tree, can be used recursively. That is, recursively the junction tree is split in subgraphs each containing their own root at which a factorized function of the parameters in that subgraph is computed. Further research is required to establish the benefits of this approach. We expect that this approach saves a considerable number of computations.

Finally, we discuss the use of independence relations induced by evidence to improve the efficiency of computing an n-way sensitivity function. By entering evidence in the junction tree, it may occur that several subgraphs of the junction tree become independent. By independence of two subgraphs H_1 and H_2, we mean that the probability distribution over variables in H_1 is independent of any of the variable values or parameter values in H_2, and the other way around. A junction tree falls apart in several independent subgraphs when all variables in one or more separators are observed. Now, the n-way sensitivity function is a product of the sensitivity functions in these independent subgraphs. Since propagation of vector tables can now be restricted to a set of smaller subgraphs, the total number of computations needed to obtain all coefficients is much smaller than by applying procedure `vector-propagate` on the whole junction tree.

Evidence in the junction tree may also be useful to postpone or avoid multiplications. Carrying out the multiplication of a vector table associated with clique U by an incoming message is only needed if this message is a function of a variable over which we have to marginalize in order to compute the message of clique U to the next separator. Otherwise, the multiplication can be postponed. In that case, the message from U to the next separator is not a single vector table but a set of vector tables. This idea is known as *lazy evaluation* and is presented in detail in [Madsen & Jensen, 1998, Madsen, 1999]. Although postponing the multiplication of vector tables leads to the propagation of a set of vector tables, it saves computations, mainly by keeping the lengths of the vectors of coefficients in each vector table at a minimum.

5.7 Related research

K. Blackmond Laskey (1995) focuses on one-way sensitivity analysis of a Bayesian network. She presents an efficient method for analytically computing *sensitivity values*, that is, first-order approximations of one-way sensitivity analyses. A sensitivity value captures the approximate effect of small variations in the parameter's estimate on the probability of interest. Compared to straightforward variation, Laskey's method requires considerably less computational effort. The method, however, provides insight in the effect of small variations only: as Laskey indicates, when larger variations are considered, the quality of the approximation may break down rapidly. As the estimates for a Bayesian network's parameters often are highly inaccurate, we prefer exact sensitivity analysis.

In [Chang & Fung, 1995] and [Castillo *et al.*, 1997a], a different approach is taken to sensitivity analysis of Bayesian networks. The idea of symbolic propagation for sensitivity analysis is exploited. Instead of yielding a single number as the standard propagation algorithms do, a symbolic propagation algorithm yields an algebraic expression for a network's probability of interest in terms of all parameters in the network. From this expression, the sensitivity of the probability of interest to a parameter under study is readily computed, basically by filling in the specified estimates for all other parameters. This method can be used for one-way sensitivity analysis as well as for higher order sensitivity analyses. A major disadvantage of building upon symbolic propagation is that it is time-consuming. Methods for sensitivity analysis that build upon the faster standard propagation algorithms are therefore preferable.

The use of symbolic propagation for sensitivity analysis, however, led E. Castillo et al. to recognize the mathematical form of the function relating a probability of interest to the network's parameters [Castillo *et al.*, 1995]; later, this functional form was analytically derived for a one-way sensitivity analysis of a Bayesian network [Coupé & Van der Gaag, 1998] (see also Chapter 4). In general, the function expressing a probability of interest in terms of n parameters is a quotient of two multilinear functions. Exploiting this knowledge, standard propagation algorithms [Pearl, 1988, Jensen *et al.*, 1990, Shafer & Shenoy, 1990] can be used to obtain the coefficients in this function. Castillo *et al.* (1997) were the first to propose a method for this purpose. For the parameters under study, various different combinations of values are assumed; one different combination of values for each coefficient is required. For each combination of values, the probability of interest is computed from network using a standard propagation method. As a result, a system of linear equations is obtained which is solved to give the required coefficients. This method is far more efficient than straightforward sensitivity analysis. However, the number of network computations required increases exponentially with the number of parameters considered in the analysis.

For a one-way sensitivity analysis, U. Kjærulff and L.C. van der Gaag (2000) present a substantial computational improvement of the methods discussed above. They suggest using the computational architecture of a junction tree to obtain the coefficients in the functions expressing a probability of interest in terms of every single parameter in the network. Just one inward and two outward propagations in the junction tree are needed to obtain the one-way sensitivity analyses with respect to *all* parameters in the network. Basically, the idea is to compute the coefficients for each parameter under study within the clique in the junction tree in which the parameter resides. Currently, this is the most efficient method for performing a one-way sensitivity analysis of a Bayesian network.

In [Kjærulff & Van der Gaag, 2000], it is noted that their method for one-way sensitivity analysis can be extended to higher order sensitivity analyses if the parameters considered are located in the same clique. Even if the parameters do not satisfy this restriction, though, the method can be exploited for higher order sensitivity analyses. A two-way sensitivity analysis with respect to two parameters located in different cliques, for example, can be directly derived from the functions describing the one-way sensitivity analyses with respect to each parameter independently. So, one propagation suffices for all one- and two-way sensitivity analyses. For four-way sensitivity analysis, the function describing the relation between the probability of the evidence and the parameters under consideration contains sixteen coefficients. One propagation gives eight coefficients derived from the two coefficients in the one-way sensitivity functions of each parameter. From a second propagation with different values for all four parameters, the remaining eight coefficients are obtained. In general, then, $2^{n-1}/n$ propagations are needed for an n-way sensitivity analysis. The method is efficient with respect to the number of propagations required. Note, however, that the task of solving equations for determining the coefficients is not a local matter, whereas the computation of the coefficients is local in the method by U. Kjærulff and L.C. van der Gaag and in the method presented in this chapter.

As in [Kjærulff & Van der Gaag, 2000], the method we present in this chapter builds upon the computational architecture of a junction tree. We focus on n-way sensitivity analysis of a Bayesian network using local computations only. Above, we have described two methods for n-way sensitivity analysis. In the following, we will compare the efficiency of the method by E Castillo et al. to ours.

In the method presented by Castillo *et al.*, basically one inward propagation is needed for every combination of values assumed for the n parameters under study. The more parameters are considered simultaneously, the more involved the system of linear equations that should be solved will be. To slightly increase the efficiency of this method, the system of equations can be simplified by taking the values for each parameter under study to be either zero or one. To further evolve this relatively straightforward way of carrying out the method by Castillo *et al.*, the *order* in which

parameter values are varied can be optimized. When a parameter value is varied, this change needs to be propagated only through the relevant subgraph of the junction tree. Therefore, there exists an optimal order to vary parameter values. This order depends on the location of the root of the junction tree to which we perform an inward propagation and on the location of the parameters under study. For a given root, the optimal order of varying parameter values roughly corresponds to the order in which parameters are encountered in going from this root to the leaves of the tree. Alternatively, but in essence equivalent, varying parameter values and performing inward propagations can be formulated as *parallel propagation*. With parallel propagation, we mean that each separator and clique receive messages (tables) for each combination of values for all parameters that reside in the the subtree 'outwards' from that separator or clique. When preparing a message from this separator or clique further 'inwards' to the root of the junction tree, the local table has to be combined with every combination of tables coming from the outside.

This 'parallel' way of carrying out the method proposed by Castillo *et al.* closely resembles the method proposed in this chapter. Instead of sending 2^m tables from a separator S inwards, where m is the number of parameters outside S, in our method we send one table with in every entry a vector containing 2^m coefficients. A difference between parallel propagation and our method is that, in our method, we do not need to solve systems of linear equations. As our framework for sensitivity analysis is based on local computations only, further optimizations in the computation of the coefficients are easily incorporating. In Section 5.6, several possibilities for optimizations were presented. Some of these optimizations, though, can also be used to optimize parallel propagation. At this point, it is difficult to draw decisive conclusions with respect to the benefits and drawbacks of the three approaches for n-way sensitivity analysis that are now available. As a next step, extensive experimentation is required.

5.8 Discussion and conclusions

In building Bayesian networks, sensitivity analysis plays an important role, as it allows for investigating the robustness of the output of the network for variations in one or more network parameters. Until now, however, techniques for performing sensitivity analysis of a Bayesian network are computationally expensive. This fact is the basic motivation for the research presented here.

This chapter describes a new method for efficient n-way sensitivity analysis; we focus on efficiently computing the coefficients in an n-way sensitivity function expressing a prior probability in the network in terms of n network parameters under study. Our method exploits the computational architecture of a junction tree derived from the Bayesian network under consideration. The method is closely related to standard

propagation algorithms. In standard propagation, messages taking the form of potential functions are propagated through the junction tree. A potential function value is a single numerical number. In our method, vector tables of partially computed coefficients are propagated through the junction tree. With each potential function in standard propagation a vector table in our method is associated. For a specific value of this potential function, the corresponding vector table contains the coefficients expressing this potential function value in terms of the parameters under study in the junction tree. By performing a single inward propagation of vector tables towards a chosen root in the junction tree, all coefficients required for the sensitivity function of interest are accumulated.

In comparing our method to the method currently available for sensitivity analysis [Castillo *et al.*, 1997b], we have argued that our basic algorithm performs equally well as a slightly optimized version of their method. Implementing the optimizations suggested in this chapter increases the efficiency of our method. In particular, using the optimal root as a starting point of our algorithm reduces the number of computations to be performed. On the other hand, we have argued that the method by Castillo *et al.* can also be optimized considerably. However, in our method it is not necessary to solve systems of linear equations while this is required in the method by Castillo *et al.*. Furthermore, we briefly introduced an extension of the method by U. Kjærulff and L.C. van der Gaag (2000). This method needs only a limited number of propagations in a junction tree to obtain an n-way sensitivity function. However, it is also necessary to to solve systems of linear equations. To compare the various methods with respect to computational efficiency, now thorough experimentation is needed.

In this chapter, we have focussed on the multilinear relation of a prior probability in a Bayesian network in terms of a number of network parameters under study. This can be easily extended to obtain the function describing a posterior probability in terms of the network parameters under study. From the definition of a conditional probability, we know that any posterior probability can be written as a quotient of two prior probabilities. Therefore, basically, our method is carried out twice. The second run of the method will generally be more efficient than the first run, since the propagation has to be carried out only in those parts of the junction tree in which new evidence is entered or evidence is removed.

Note, that our method closely resembles symbolic propagation in a junction tree. However, due to a specific encoding of the symbols, that is, the parameters under study, we do not need to propagate the symbols themselves. Our method provides for performing symbolic propagation using numerical propagation algorithms only. In the past, symbolic propagation has also been used to perform sensitivity analysis of a belief network [Chang & Fung, 1995, Castillo *et al.*, 1997b]. These approaches, however, required symbolic manipulations, which are much slower than numerical manipulations. Therefore, our method is much more efficient than these approaches.

Until now, our method has not been extensively tested. Therefore, we do not yet have a clear picture of the strengths and weaknesses of the method. For this purpose, experimental investigation is required which will be undertaken in the near future. We believe, however, that our method has the potential of a valuable tool for performing sensitivity analysis of a Bayesian network.

Part III

Sensitivity analysis of influence diagrams

Chapter 6

Sensitivity analysis for threshold decision making with Bayesian belief networks

Abstract

The probability assessments of a Bayesian belief network generally include inaccuracies. These inaccuracies influence the reliability of the network's output. An integral part of investigating the output's reliability is to study its robustness. Robustness pertains to the extent to which varying the probability assessments of the network influences the output. It is studied by subjecting the network to a sensitivity analysis. In this chapter, we address the issue of robustness of a belief network's output in view of the threshold model for decision making. We present a method for sensitivity analysis that provides for the computation of bounds between which a network's assessments can be varied without inducing a change in recommended decision.

6.1 Introduction

Bayesian belief networks are widely accepted in artificial intelligence as intuitively appealing representations of domain knowledge. A Bayesian belief network basically is a concise representation of a joint probability distribution. It encodes, in a qualitative, graphical part, the variables of importance in the domain that is being represented, along with their probabilistic interrelationships; the strengths of these relationships are quantified by conditional probabilities, that with each other constitute the network's quantitative part. The increasing number of knowledge-based systems that build upon the framework of Bayesian belief networks for knowledge representation and inference, clearly demonstrate its usefulness for addressing complex real-life problem domains in which uncertainty is predominant. Most notably, applications are being realised in the

medical domain, for diagnosis, prognostic assessment, and treatment planning.

Bayesian belief networks are generally constructed with the help of experts from the domain of application. Experience shows that, although it may require considerable effort, building the qualitative part of a belief network is quite practicable. In fact, as it has parallels to designing a domain model for a more traditional knowledge-based system, well-known knowledge-engineering techniques can be employed. Constructing a belief network's quantitative part is generally considered a far harder task, not in the least because it tends to be much more time-consuming. It amounts to assessing various conditional probabilities for the variables represented in the network's qualitative part. Although, for most application domains, probabilistic information is available from literature or from statistical data, it often turns out that this information does not provide for estimating all probabilities required [Druzdzel & Van der Gaag, 1995]. For most domains, therefore, many probabilities remain to be assessed by domain experts. Upon eliciting judgemental probabilities from experts, various problems of bias and poor calibration are typically encountered [Kahneman *et al.*, 1982]. The probability assessments obtained for a belief network as a consequence tend to be inaccurate.

The inaccuracies in the probability assessments of a Bayesian belief network influence the reliability of its output. An integral part of investigating the reliability of a network's output is to study its robustness. Robustness pertains to the extent to which varying the probability assessments of the network influences its output. For gaining detailed insight in output robustness, a Bayesian belief network can be subjected to a sensitivity analysis [Coupé & Van der Gaag, 1998]. Sensitivity analysis is a general technique for investigating the effects of the inaccuracies in the parameters of a mathematical model on the model's output [Morgan & Henrion, 1990]. The basic idea of performing a sensitivity analysis of a belief network is to systematically vary the assessments for the network's conditional probabilities and study the effects on the output. Upon such an analysis, some conditional probabilities will show a considerable impact, while others will hardly reveal any influence.

In this chapter, we address the issue of output robustness of a Bayesian belief network in view of applications in which the output is used for decision making. To this end, we focus on the threshold model for decision making. Although generally applicable, this model is used most notably for patient management in medical applications [Pauker & Kassirer, 1980]. With the threshold model, an attending physician decides whether or not to gather additional information from diagnostic tests and whether or not to give treatment based upon the probability of disease for a patient under consideration. The robustness of the output of a belief network now pertains not just to the probability of disease computed from the network, but also to the decision for patient management based upon it. For some conditional probabilities, varying their assessment may have a considerable effect on the probability of disease and yet not induce a change in patient management; for other probabilities, variation may have little effect

on the probability of disease and nonetheless result in a different management decision. Studying the effects of varying the assessments for the network's conditional probabilities on the probability of disease therefore no longer suffices for establishing robustness: the effects on the recommended decision need also be taken into consideration.

To provide for studying output robustness of a Bayesian belief network in view of the threshold model for decision making, we enhance the basic method of sensitivity analysis with the computation of upper and lower bounds between which a belief network's assessments can be varied without inducing a change in recommended decision. Informally speaking, the more a belief network's probability assessments can be varied, the more robust the decision based upon the network is.

The chapter is organised as follows. In Section 6.2, we briefly review the formalism of Bayesian belief networks. In Section 6.3, we outline the threshold model for decision making. In Section 6.4, we detail the basic method of sensitivity analysis and its enhancement for threshold decision making. The chapter ends with some conclusions and directions for further research in Section 6.5.

6.2 Bayesian belief networks

A *Bayesian belief network* basically is a representation of a joint probability distribution on a set of statistical variables [Pearl, 1988]. 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 statistical variable that takes its value from a finite set of discrete values. In this chapter we will restrict the discussion to binary variables, taking one of the values *true* and *false*. 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 digraph represent the 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 and, hence, are conditionally independent.

For our running example we consider the following fragment of (fictitious and incomplete) medical information, adapted from [Cooper, 1984]:

> Consider a primary tumour with an uncertain prognosis in an arbitrary patient. The cancer can metastasize to the brain and to other sites. We are interested in the course of the cancer within the next few years, especially with regard to the development of a brain tumour and its associated problems. Metastatic cancer (denoted MC) may be detected by an increased level of serum calcium (ISC). The presence of a brain tumour (B) may be established from a CT scan (CT). Severe headaches (SH) are indicative of the presence of a brain tumour. Both a brain

tumour and an increased level of serum calcium are likely to cause a patient to fall into a coma (C) in due course.

In this fragment of information, six statistical variables are identified. The influential relationships among these variables are encoded in the digraph depicted in Figure 6.1. The digraph for example reflects, by means of the arc $B \rightarrow SH$, that the presence of a

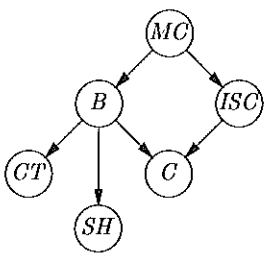


Figure 6.1: The digraph of an example belief network; it expresses information concerning the presence of a brain tumour and its associated problems in an arbitrary patient.

brain tumour is a possible cause of severe headaches.

The relationships among the variables that are represented in the qualitative part of a Bayesian belief network basically are probabilistic dependences. The strengths of these dependences are described by conditional probabilities: for each variable, the probabilities of its values are specified conditional on the various possible combinations of values for its parents in the digraph. For our running example, we assume the following probabilities:

$p(b \mid mc) =$	0.20	$p(mc) =$	0.20	$p(ct \mid b) =$	0.95
$p(b \mid \neg mc) =$	0.05			$p(ct \mid \neg b) =$	0.10
		$p(c \mid b, isc) =$	0.80		
$p(isc \mid mc) =$	0.80	$p(c \mid \neg b, isc) =$	0.80	$p(sh \mid b) =$	0.80
$p(isc \mid \neg mc) =$	0.20	$p(c \mid b, \neg isc) =$	0.80	$p(sh \mid \neg b) =$	0.60
		$p(c \mid \neg b, \neg isc) =$	0.05		

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 an increased level of serum calcium in an arbitrary patient. The relationship between metastatic cancer and increased total serum calcium therefore is a strong dependence. On the other hand, severe headaches are expressed as quite common in both patients with and without a brain tumour. Severe headaches thus have a low predictive value for a brain tumour. The probabilities with each other constitute the network's quantitative part.

The qualitative and quantitative parts of a Bayesian belief network with each other uniquely define a joint probability distribution. A belief network therefore allows for the computation of any probability pertaining to its variables. For this purpose, various algorithms are available, that provide for computing probabilities of interest and for processing evidence, that is, for entering observations into the network and subsequently computing the revised probability distribution given these observations [Pearl, 1988, Lauritzen & Spiegelhalter, 1988]. The details of these algorithms are not relevant for the present chapter.

6.3 Threshold decision making

In the medical domain, Bayesian belief networks are often used for diagnostic purposes. A diagnostic belief network typically comprises one or more variables modeling the presence or absence of disease, various variables modeling findings and results from diagnostic tests, and a number of intermediate variables modeling unobservable pathophysiological states. In our example network, for instance, the variable B models the disease of interest, being the presence or absence of a brain tumour; the variable MC models an unobservable state and the remaining variables capture findings and test results. A medical diagnostic belief network is used for computing a most likely diagnosis for a patient given his or her presentation findings and test results.

The most likely diagnosis for a patient, along with its uncertainty, is generally taken by an attending physician to decide upon management of the patient. The physician may decide, for example, to start treatment rightaway. For our running example, the physician may decide to perform neurosurgery if a brain tumour is indicated. Alternatively, the physician may defer the decision whether or not to treat the patient until additional diagnostic information has become available, for example from a CT scan. Or, the physician may decide to withhold treatment altogether. To support choosing among these decision alternatives, the threshold model for patient management can be used.

The *threshold model for patient management*, or for decision making more in general, builds upon various threshold probabilities of disease [Pauker & Kassirer, 1980]. The *treatment threshold probability* of disease, written $P^*(d)$ for disease d, is the probability at which an attending physician is indifferent between giving treatment and withholding treatment. If, for a specific patient, the probability of disease $\Pr(d)$ exceeds the treatment threshold probability, that is, if $\Pr(d) > P^*(d)$, then the physician will decide to treat the patient as if the disease were known to be present with certainty. Alternatively, if $\Pr(d) \leq P^*(d)$, the physician will basically withhold treatment from the patient.

As a consequence of the uncertainty concerning the presence of disease in a patient,

additional information from a diagnostic test may affect an attending physician's basic management decision. If the probability of disease exceeds the treatment threshold probability, then interpreting a negative test result may result in an updated probability of disease *below* the threshold probability. Alternatively, if the pretest probability of disease falls below the treatment threshold probability, a positive test result may raise the probability of disease to a value *above* the threshold probability. To reckon with such effects, the threshold model for patient management includes another two threshold probabilities. The *no treatment-test threshold probability* of disease, written $P^-(d)$, is the probability at which the attending physician is indifferent between the decision to withhold treatment and the decision to obtain additional diagnostic information. The *test-treatment threshold probability* of disease, written $P^+(d)$, is the probability at which the physician is indifferent between obtaining additional information and starting treatment rightaway.

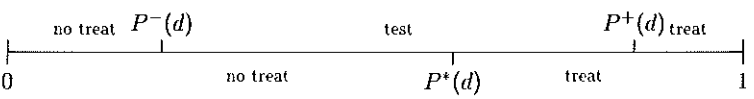


Figure 6.2: The threshold model for patient management, indicating three threshold probabilities and the various decision alternatives at a physician's disposal.

Figure 6.2 summarises the basic idea of the threshold model for patient management. As long as the diagnostic test under consideration has not been performed, a physician has three decision alternatives at his or her disposal. If the probability of disease $\Pr(d)$ for a patient falls below the no treatment-test threshold probability, that is, if $\Pr(d) < P^-(d)$, then the physician will withhold treatment from the patient without gathering additional diagnostic information. If the probability of disease exceeds the test-treatment threshold probability, that is, if $\Pr(d) > P^+(d)$, then the physician will start treatment rightaway. Otherwise, that is, if $P^-(d) \leq \Pr(d) \leq P^+(d)$, the physician will perform the diagnostic test. After testing, there are only two decision alternatives left. If the updated probability of disease for the patient exceeds the treatment threshold probability, the physician will start treatment; otherwise, treatment will be withheld from the patient.

The treatment threshold probability of disease $P^*(d)$ used in the threshold model is typically established by a physician after carefully weighing the various utilities involved. These utilities pertain to the presence or absence of disease on the one hand and giving or withholding treatment on the other hand. From the expected utilities for giving and withholding treatment in view of the uncertainty concerning the presence of disease, the probability of disease at which the physician is indifferent between the two decision alternatives is readily determined; the basic idea is illustrated in Figure 6.3(a). For our running example, the physician will typically take into consideration the life

expectancy for a patient, with and without a brain tumour, and the patient's attitude towards impaired health states; we assume that the physician sets the treatment threshold probability of a brain tumour at 0.15. The two threshold probabilities $P^-(d)$ and $P^+(d)$ for deciding whether or not to perform a diagnostic test are established from the test's characteristics. For our example, the physician will typically weigh the discomfort of a CT scan for a patient against the additional diagnostic information

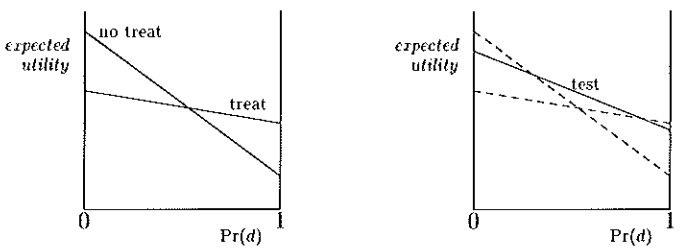


Figure 6.3: The basic idea of establishing the treatment threshold probability of disease, (a), and the no treatment-test and test-treatment threshold probabilities, (b).

yielded by the scan; we assume that the physician sets the no treatment-test threshold probability of a brain tumour at 0.045 and the test-treatment threshold probability at 0.56. The basic idea of establishing these two threshold probabilities is illustrated in Figure 6.3(b).

Although we have discussed the threshold model for decision making in a medical context, we would like to note that the model's use is not restricted to the medical domain but in fact is broadly applicable.

6.4 Sensitivity analysis for threshold decision making

In the introduction, we have argued that the various probability assessments of a Bayesian belief network tend to be inaccurate. To gain insight into the effects of the inaccuracies involved, a belief network can be subjected to a sensitivity analysis. In Section 6.4.1, we outline sensitivity analysis of a Bayesian belief network with regard to a probability of interest. In Section 6.4.2, we address sensitivity analysis of a belief network in view of threshold decision making.

6.4.1 Sensitivity analysis of a Bayesian belief network

For a Bayesian belief network, sensitivity analysis amounts to systematically varying the assessments for the network's conditional probabilities and investigating the effects on a probability of interest [Coupé & Van der Gaag, 1998]. In essence, for every conditional probability of the network, a number of deviations from the original assessment are investigated. For every investigated value, the probability of interest is computed from the network. The results thus obtained reflect the probability of interest as a function of the conditional probability under study.

We illustrate performing a sensitivity analysis of our example belief network, taking the prior probability of the presence of a brain tumour in an arbitrary patient for the probability of interest. The effects of varying the assessments for the probabilities $p(mc)$ and $p(b \mid \neg mc)$ on this probability of interest are shown in Figure 6.4. Figure

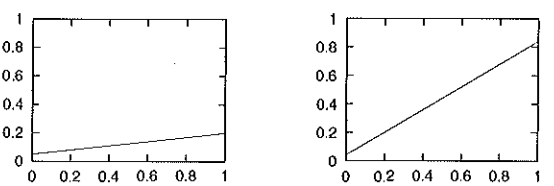


Figure 6.4: A sensitivity analysis of the example belief network; the effects of varying the assessments for the probabilities $p(mc)$, (a), and $p(b \mid \neg mc)$, (b), on the prior probability of disease $\Pr(b)$ are shown.

6.4(a) shows that systematically varying, from 0 to 1, the assessment for the probability $p(mc)$ of the presence of metastatic cancer has a rather small effect on the probability of interest: $\Pr(b)$ increases from 0.05 to 0.20. Figure 6.4(b) shows that varying the assessment for the conditional probability $p(b \mid \neg mc)$ of the presence of a brain tumour in the absence of metastatic cancer has a much stronger effect: $\Pr(b)$ now ranges from 0.04 to 0.84. As long as no further information is available about the degrees of inaccuracy in the assessments for the two probabilities under study, we conclude that the probability of interest is more robust with regard to the assessment for the probability $p(mc)$ than with regard to the assessment for $p(b \mid \neg mc)$.

A sensitivity analysis of a Bayesian belief network with regard to a prior probability of interest allows for assessing the robustness of the 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 very well show different sensitivities. To reveal these, a sensitivity analysis can be performed with regard to a *posterior* probability of interest. Such an analysis allows for investigating the robustness of the network's output for specific cases or for case profiles.

We once again perform a sensitivity analysis of our example belief network, 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 severe headaches. The effects of varying the assessments for the conditional probabilities $p(b \mid \neg mc)$ and $p(sh \mid \neg b)$ on the posterior probability of interest are shown in Figure 6.5. Figure 6.5(a), when compared with Figure 6.4(b), reveals that the observation that a patient suffers from severe headaches has little impact on the robustness of the probability of disease with regard to the assessment for $p(b \mid \neg mc)$. Figure 6.5(b) shows that varying the assessment for the conditional probability $p(sh \mid \neg b)$ can have a considerable effect on the posterior probability of disease; small deviations from the original assessment 0.60 have little effect, however.

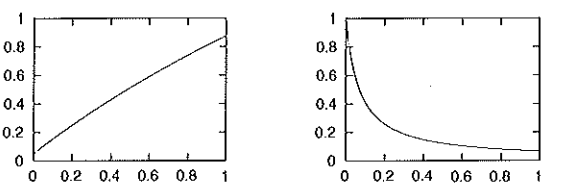


Figure 6.5: A sensitivity analysis of the example belief network; the effects of varying the assessments for the conditional probabilities $p(b \mid \neg mc)$, (a), and $p(sh \mid \neg b)$, (b), on the posterior probability of disease $\Pr(b \mid sh)$ are shown.

In a sensitivity analysis of a Bayesian belief network, the relation between a probability of interest and a conditional probability under study is a simple mathematical function. In general, the probability of interest relates to a conditional probability under study as a quotient of two linear functions. For the posterior probability of interest $\Pr(d \mid o)$, given some observations o, and a conditional probability x, we have that

$$\Pr(d \mid o) = \frac{a \cdot x + b}{e \cdot x + f}$$

where a, b, e, and f are constants. For an example, we reconsider Figure 6.5(b) showing for our belief network the posterior probability of interest $\Pr(b \mid sh)$ as a function of the conditional probability $x = p(sh \mid \neg b)$; this function equals

$$\Pr(b \mid sh) = \frac{0.06957}{x + 0.06957}$$

If the conditional probability under study pertains to a variable without any observed descendants, that is an ancestor of the variable of interest in the network's qualitative part, the mathematical function reduces to a linear function. For the probability of

interest and a conditional probability x as indicated, we then have that

$$\Pr(d \mid o) = a \cdot x + b$$

where a and b are constants. In particular, a prior probability of interest relates linearly to any conditional probability from the network. For an example, we reconsider Figure 6.4(b) showing for our belief network the prior probability of interest $\Pr(b)$ as a function of the conditional probability $x = p(b \mid \neg mc)$; the function equals

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

The constants in the mathematical functions involved in a sensitivity analysis of a Bayesian belief network are readily determined by computing the probability of interest from the network for a small number of values for the conditional probability under study and solving the resulting system of linear equations. For further technical details, we refer the reader to [Coupé & Van der Gaag, 1998].

6.4.2 Sensitivity analysis in view of threshold decision making

Sensitivity analysis of a Bayesian belief network with regard to a probability of interest yields a functional relation between this probability of interest and every single conditional probability from the network. These relations indicate how the probability of interest will shift upon varying the assessments for the various conditional probabilities. For a probability of interest that is used in the threshold model for decision making, not every shift is significant. In fact, only a shift that results in a different decision recommendation is of interest. In sensitivity analysis in view of threshold decision making, therefore, we have to take the various threshold probabilities employed into consideration. To this end, we enhance sensitivity analysis of a Bayesian belief network with the computation of upper and lower bounds between which the network's assessments can be varied without inducing a change in decision.

The computation of bounds on variation of a belief network's probability assessments builds upon the mathematical functions that we have detailed before, relating a probability of interest to the network's conditional probabilities. Once again focusing on patient management, we begin by considering a probability of disease $\Pr(d \mid o)$ and a conditional probability x to which it is linearly related, that is, we have

$$\Pr(d \mid o) = a \cdot x + b$$

for some constants a and b. For case of exposition, we assume that $\Pr(d \mid o)$ increases with increasing values for x; we will return to this assumption presently. If, in view of the threshold model, the probability of disease indicates withholding treatment, that is, if $\Pr(d \mid o) < P^{-}(d)$, then the decision will remain unaltered as long as the conditional probability x is smaller than the value x^{-} that is computed from

$$a \cdot x^{-} + b = P^{-}(d)$$

More precisely, the decision to withhold treatment remains unaltered for any value of the conditional probability x within the interval $(-\infty, x^-) \cap [0, 1]$. If the probability of disease on the other hand indicates starting treatment rightaway, that is, if $\Pr(d \mid o) > P^+(d)$, then the decision will remain unaltered as long as the conditional probability x is greater than the value x^+ that is computed from

$$a \cdot x^+ + b = P^+(d)$$

More precisely, the decision to start treatment rightaway remains unaltered for any value of the conditional probability x within the interval $(x^+, +\infty) \cap [0, 1]$. If the probability of disease indicates testing, that is, if $P^-(d) \leq \Pr(d \mid o) \leq P^+(d)$, then this decision will be the same for any value of the conditional probability x within the interval $[x^-, x^+] \cap [0, 1]$.

So far, we have addressed the computation of bounds on the variation of a conditional probability that is related linearly to the probability of disease. For a conditional probability that is related to the probability of disease by a quotient of two linear functions, bounds on variation are computed in a similar fashion. We have further assumed so far that the probability of disease increases with increasing values of the conditional probability x under study. With this assumption, we have implicitly assumed that $x^- \leq x^+$. For a conditional probability x of which increasing values serve to decrease the probability of disease, we have that $x^- \geq x^+$. Using this observation, the bounds derived above are readily adjusted.

We illustrate the computation of bounds on variation for our example belief network; we recall from Section 6.3 that the three threshold probabilities of disease have been set at $P^*(b) = 0.15$, $P^-(b) = 0.045$, and $P^+(b) = 0.56$. We begin by addressing the robustness of the decision for management of an arbitrary patient. From our belief network, the prior probability of disease is computed to be $\Pr(b) = 0.08$. For this probability, we have that $P^-(b) \leq \Pr(b) \leq P^+(b)$. Using the threshold model for patient management, therefore, the physician will decide to gather additional information from a CT scan. We investigate the robustness of this decision by computing an upper and lower bound on variation of the assessment for the conditional probability $x = p(b \mid \neg mc)$. The lower bound x^- on variation is computed from

$$\Pr(b) = 0.8 \cdot x^- + 0.04 = 0.045$$

yielding $x^- = 0.00625$; the upper bound x^+ on variation is computed from

$$\Pr(b) = 0.8 \cdot x^+ + 0.04 = 0.56$$

yielding $x^+ = 0.65$. For any value of the conditional probability $p(b \mid \neg mc)$ within the interval $[0.00625, 0.65]$, therefore, the decision to gather additional diagnostic information will remain unaltered. Since the conditional probability under study has been assessed at 0.05, we conclude that the decision is fairly robust with regard to this

assessment; variation of the assessment to smaller values, however, may change the decision to the recommendation to withhold treatment without gathering additional diagnostic information.

To conclude, we address the robustness of the management decision for a patient with a primary tumour who is known to suffer from severe headaches. From our belief network, the posterior probability of disease is computed to be $\Pr(b \mid sh) = 0.1039$. For this probability, we observe that $P^-(b) \leq \Pr(b \mid sh) \leq P^+(b)$. The physician will therefore order a CT scan for the patient. We investigate the robustness of this decision by computing the upper and lower bound on variation of the assessment of the conditional probability $x = p(sh \mid \neg b)$. Note that the probability of disease decreases with increasing values for this conditional probability. The lower bound x^+ on variation is computed from

$$\Pr(b \mid sh) = \frac{0.06957}{x^+ + 0.06957} = 0.56$$

yielding $x^+ = 0.1938$. Upon computing the upper bound x^- on variation, we find a value greater than one. For any value of the conditional probability $p(sh \mid \neg b)$ within the interval $[0.1938, 1]$, therefore, the decision to gather additional diagnostic information for the patient will remain unaltered. Since the conditional probability under study has been assessed at 0.60, we conclude that the decision is quite robust with regard to this assessment.

6.5 Conclusions

The probability assessments of a Bayesian belief network tend to be inaccurate. The belief network as a consequence will yield inaccurate output. If the network's output is used for decision making, its inaccuracy influences the reliability of a decision that is based upon it. An integral part of investigating reliability is to study output robustness. To investigate the robustness of a belief network's output in view of threshold decision making, we have presented an enhanced method for sensitivity analysis that provides for the computation of upper and lower bounds between which a network's assessments can be varied without inducing a change in decision.

We have addressed the issue of robustness in view of a simplified threshold model for decision making, involving binary variables and a single diagnostic test. The more general threshold model addresses variables that have multiple discrete values and provides for selecting among multiple tests. Our method of sensitivity analysis will be further elaborated upon for use with this more general model. Although often used in practice, the threshold model is a simple model for decision making. With a Bayesian belief network, more complex models can be used. More specifically, a belief network can be extended to an influence diagram to provide for addressing more elaborate

trade-offs in decision making. The results put forward in this chapter hold unabatedly for influence diagrams. In the near future, we hope to extend our method of sensitivity analysis for decision making to apply to influence diagrams.

Chapter 7

Properties of sensitivity analysis of influence diagrams

Abstract

For an influence diagram, simple mathematical functions exist that express expected utility in terms of one or more conditional probabilities or utilities from the diagram. These functions provide for performing efficient sensitivity analysis of an influence diagram. The sensitivity of the expected utility of each decision alternative to variations in one or two conditional probabilities or utilities under study is established by computing the required constants in the mathematical functions. To obtain these constants only a limited number of evaluations of the diagram is required. Now, relating the mathematical functions for all decision alternatives to each other, provides for investigating the sensitivity of the recommended decision, that is, the decision alternative with maximum expected utility. As a measure of the sensitivity of the recommended decision, we propose the *minimum deviation*. The *minimum deviation* is the smallest change in the assessment(s) for one or two probabilities or utilities under study that leads to a change in the recommended decision.

7.1 Introduction

The field of decision making under uncertainty has been receiving considerable attention within artificial-intelligence research over the past decade. The field is concerned with the basic problem of determining, from among various different decision alternatives, the alternative that best meets a decision maker's preferences in view of uncertainty. Decision making under uncertainty builds for its mathematical foundation on probability theory and utility theory for describing and reasoning with the uncertainties and preferences involved in a decision problem [Von Neumann & Morgenstern, 1944, Raiffa & Schlaifer, 1961]. For formally representing decision problems, various for-

malisms have been designed. One of these formalisms is the formalism of influence diagrams [Howard & Matheson, 1981].

An *influence diagram* is a concise representation of a decision problem involving uncertainty. It consists of a qualitative part and an associated quantitative part. The qualitative part of an influence diagram takes the form of a directed graph composed of various types of node and arc. The chance nodes of an influence diagram model the statistical variables involved in the represented problem; the arcs among these nodes with each other capture the independences among the corresponding variables. The decision nodes of the diagram represent the decision variables from the problem under study, each representing a choice among various decision alternatives; the arcs between these nodes indicate the order in which decisions are taken. The utility node, to conclude, summarises the possible outcome scenarios of the represented problem; its incoming arcs indicate the decision and chance variables whose values influence the problem's outcome. The quantitative part of an influence diagram is composed of conditional probabilities and utilities. The conditional probabilities basically describe the strengths of the dependences between the statistical variables. The utilities express the decision maker's preferences for the various possible outcome scenarios. An influence diagram's qualitative part and quantitative part together provide for computing, for each decision node, the most preferred decision alternative, that is, the decision alternative with maximum expected utility [Shachter, 1986].

Influence diagrams are generally constructed with the help of experts from the domain of application. Experience shows that building the qualitative part of an influence diagram is relatively straightforward. Assessing the various different conditional probabilities and utilities for the quantitative part, however, is generally considered a much more difficult and time-consuming task. For most applications, available data collections do not provide for assessing all probabilities required [Druzdzel & Van der Gaag, 1995]. Many probabilities have to be assessed therefore by domain experts. Although various elicitation techniques have been designed to avert the problems of bias and poor calibration typically found in judgemental probabilities, elicited from humans, the large number of probabilities required for an influence diagram often prohibits the use of these rather time-consuming methods. The assessment of utilities is generally found to be an even harder task than probability assessment. Often, utilities have to be associated with outcome scenarios with which the assessor is not acquainted. In contrast with probability assessments, moreover, evaluating the calibration of utility assessments is very difficult, if not practically impossible. From the above observations, we conclude that both the probability assessments and the utility assessments obtained for an influence diagram are prone to inaccuracy.

The inaccuracies in the various probability and utility assessments of an influence diagram may influence the reliability of the diagram's output. In a medical application, for example, non-optimal treatment recommendations may result from building on in-

accurate assessments. An integral part of investigating the reliability of an influence diagram's output is studying its sensitivity [Morgan & Henrion, 1990]. In general, *sensitivity analysis* of a mathematical model amounts to investigating the effects of the inaccuracies in the model's parameters by varying the values of the parameters in a systematic way [Morgan & Henrion, 1990, Habbema *et al.*, 1990]. For an influence diagram, sensitivity analysis amounts to varying the assessments for one or more conditional probabilities or utilities simultaneously and investigating the effects on the preferred decision alternative, that is, the decision alternative with maximum expected utility. Informally speaking, the more the assessments can be varied without inducing a change in the most preferred decision alternative, the less sensitive the outcome of the diagram is to inaccuracies.

Unfortunately, as a result of the usually large number of conditional probabilities and utilities involved, performing a sensitivity analysis of an influence diagram in a straightforward manner is computationally unfeasible. Even the simplest type of sensitivity analysis in which one assessment at a time is investigated, easily requires tens of thousands of computations: for every single conditional probability and every utility, a number of deviations from the specified assessment is investigated and for every such deviation the most preferred decision alternative needs to be computed. To be of practical use, more efficient methods for sensitivity analysis are required.

In this chapter, we present an efficient method for sensitivity analysis of influence diagrams. Our method builds upon properties of influence diagrams that render straightforward variation of conditional probabilities and utilities unnecessary. We show that there exist simple mathematical functions expressing the expected utility of a decision alternative in terms of one or more conditional probabilities and utilities from the diagram. Computing the constants in these mathematical functions suffices to establish the sensitivity of the expected utility of a decision alternative to variations in the assessments of the influence diagram under study. We further show that these functions provide for the computation of the minimum deviation from the original assessments required to induce a change in the most preferred decision alternative. The minimum deviation required is indicative of the sensitivity of the diagram's output to inaccuracies in its various assessments.

The chapter is organised as follows. In Section 7.2, we review the formalism of influence diagrams and introduce our running example. In Section 7.3, we provide some background on sensitivity analysis. In Section 7.4, we discuss various properties of an influence diagram that serve to reduce the computational burden of a sensitivity analysis with respect to its conditional probabilities; in Section 7.5, we extend these properties to include the utilities of a diagram. In Section 7.6, we briefly discuss sensitivity analysis with respect to both conditional probabilities and utilities. The chapter ends with our conclusions and directions for further research in Section 7.7.

7.2 Influence diagrams

An influence diagram is a concise representation of a decision problem involving uncertainty. In this section, we review the basic formalism of influence diagrams; for a more elaborate introduction, we refer the reader to [Howard & Matheson, 1981, Shachter, 1986].

In an influence diagram, qualitative information describing the structure of the represented decision problem is explicitly separated from the quantitative information involved. The qualitative part of an influence diagram is an acyclic directed graph $G = (V(G), A(G))$ with nodes $V(G)$ and arcs $A(G)$. The set of nodes $V(G)$ is partitioned into three disjoint sets of nodes that capture variables having different meanings in the represented problem. A *chance node* is a node representing a statistical variable that takes a value from among a finite set of values. The set of all chance nodes will be denoted $C(G)$. A *decision node* is a node that models a decision variable or moment of choice for the decision maker. A decision variable is a variable that represents the various different decision alternatives or actions at the decision maker's disposal; the value of a decision variable is under direct control of the decision maker. In the sequel, we will restrict the discussion to influence diagrams with just a single decision node D. The third type of node in an influence diagram is the *utility node*. The utility node represents the various different outcome scenarios that may arise from the decision alternatives and serves to encode the desirability of these scenarios to the decision maker. The utility node is unique and will be denoted U.

The set $A(G)$ of arcs in the digraph of an influence diagram is equally partitioned into disjoint subsets. The arcs between the chance nodes with each other encode the independences among the represented statistical variables. Informally speaking, we take an arc $C_i \rightarrow C_j$ to represent a direct influential or causal relationship between the statistical variables C_i and C_j; the arc's direction designates C_j as the effect or consequence of the cause C_i. Absence of an arc between two chance nodes means that the corresponding variables do not influence each other directly and, hence, are (conditionally) independent. An arc from the decision node into a chance node expresses an influence on the represented statistical variable, exerted by the decision maker through his choice for the decision variable at hand. The incoming arcs of the decision node are generally referred to as *informational arcs*. An informational arc from a chance node into the decision node indicates that the corresponding statistical variable must have been observed at the time the decision is made. Absence of an arc from a chance node into the decision node means that the decision maker need not necessarily have observed the value of the corresponding variable. An incoming arc for the utility node indicates that the variable or decision represented by the node at the tail of the arc directly affects expected utility. The utility node does not have any outgoing arcs.

The quantitative part of an influence diagram is composed of conditional probabil-

ities and utilities. With each chance node $C_i \in C(G)$, having the set $\pi_G(C_i)$ for its (immediate) predecessors, is associated a set of *conditional probabilities* $p(C_i \mid \pi_G(C_i))$ describing the joint influence of the values of the nodes in $\pi_G(C_i)$ on the probabilities of the values of C_i. With the utility node U is associated a set of *utilities* $u(\pi_G(U))$ specifying for each combination of values for the set $\pi_G(U)$ of predecessors of U, a number expressing the desirability of this value combination to the decision maker.

We define the concept of an influence diagram more formally; we would like to recall that we restrict our discussion to influence diagrams involving a single decision node only.

Definition 7.2.1 *An* influence diagram *is a tuple* $ID = (G, p, u)$ *where*

- $G = (V(G), A(G))$ *is an acyclic digraph with nodes* $V(G) = C(G) \cup \{D\} \cup \{U\}$ *where*
  - $C(G)$ *is the diagram's set of* chance nodes*;*
  - D *is the diagram's* decision node*;*
  - U *is the* utility node*;*

  and arcs $A(G)$ *with* $\sigma_G(U) = \varnothing$*;*
- p *is a set of conditional probabilities* $p(C_i \mid \pi_G(C_i))$*, for all* $C_i \in C(G)$*, where* $\pi_G(C_i)$ *is the set of predecessors of* C_i *in* G*;*
- u *is a set of utilities* $u(\pi_G(U))$.

We illustrate the concept of influence diagram by means of an example that will be used for our running example throughout the chapter.

Example 7.2.2 We consider a small, highly simplified influence diagram for patient-specific therapy selection for oesophageal carcinoma.

As a consequence of a lesion of the oesophageal wall, for example as a result of frequent reflux, a carcinoma may develop in a patient's oesophagus. An oesophageal carcinoma has various characteristics that influence its prospective growth and possible infiltration into neighbouring structures. These characteristics include the *Location* of the carcinoma in the oesophagus, the *Length* of the carcinoma, and its macroscopic *Shape*; the location and length of the carcinoma are established from a *Gastro*scopic examination of the oesophagus. Oesophageal carcinoma, unfortunately, is associated with a poor *Life expectancy*. The presence of an oesophageal carcinoma in addition influences a patient's quality of life. For example, dependent upon its length and macroscopic shape, an oesophageal carcinoma may block the *Passage* of food through the oesophagus, with considerable *Weightloss* as a consequence. In the presence of

Necrosis (serious decay of tissue), moreover, may an oesophageal carcinoma give rise to a *Fistula* (an infiltration, resulting in an open connection) into the trachea and bronchi. In the presence of a fistula, food, upon swallowing, may enter into the patient's trachea, causing choking and extensive coughing.

While establishing the presence of an oesophageal carcinoma in a patient is relatively easy, the selection of an appropriate therapy is a far harder task. The effects aimed at by instilling a therapy include removal or reduction of the patient's primary tumour and an improved *Passage after therapy* of food through the oesophagus. The various different therapeutic alternatives available differ in the extent to which these effects can be attained. Instillation of a therapy further is expected to be accompanied not just by beneficial effects but also by various, possibly serious, complications. The effects and complications to be expected from a therapy depend on the characteristics of the patient's carcinoma and need be weighed carefully before deciding upon a specific therapy for the patient.

In our simplified influence diagram, we consider the possible effects and complications of just three therapeutic alternatives; these are a surgical removal of the oesophagus, positioning an endoprosthesis or stent into the oesophagus, and withholding treatment. A surgical procedure is aimed at removal of the primary tumour and, thus, at a prolonged life expectancy. Only if upon surgery all tumour cells are removed and no *Residual* cancer remains, will a prolonged life expectancy be attained for the patient. The procedure brings a high mortality risk, especially for older patients. Positioning a prosthesis in the oesophagus is aimed primarily at an improved quality of life after therapy. The prosthesis serves to improve the passage of food through the oesphagus and, in the presence of a fistula, in addition serves to cover the fistula, thereby alleviating associated problems. Positioning a prosthesis, however, is not without risk. Especially if a patient's carcinoma is associated with necrosis can positioning the prosthesis cause a *Perforation* of the oesophageal wall; also, a patient may suffer from *Migration* of the prosthesis.

The digraph of the simplified influence diagram representing the decision problem of therapy selection for oesophageal carcinoma is shown in Figure 7.1. The chance nodes are depicted as circles, the decision node is represented by a box, and the utility node is drawn as a hexagon.

Associated with each chance node in the influence diagram is a set of conditional probabilities describing the joint influence of values for its predecessors on the probabilities of its own values. For example, for the chance node *Necrosis*, the diagram specifies the following conditional probabilities for the presence of serious decay of tissue in the oesophageal wall, indicated by the value *yes*:

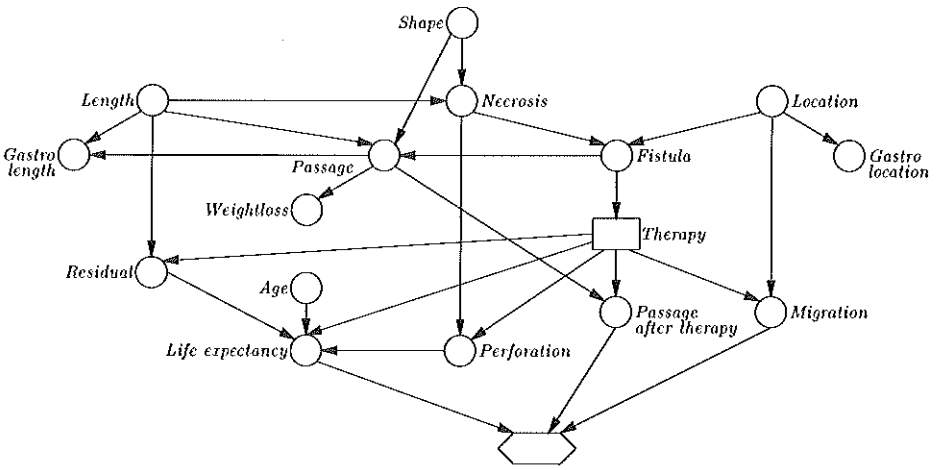


Figure 7.1: The digraph of the *Oesophagus* influence diagram.

$$
\begin{aligned}
\Pr(Necrosis = yes \mid Shape = polypoid, Length < 6cm) &= 0 \\
\Pr(Necrosis = yes \mid Shape = polypoid, Length \geq 6cm) &= 0.25 \\
\Pr(Necrosis = yes \mid Shape = scirrheus, Length < 6cm) &= 0 \\
\Pr(Necrosis = yes \mid Shape = scirrheus, Length \geq 6cm) &= 0.10 \\
\Pr(Necrosis = yes \mid Shape = ulcerating, Length < 6cm) &= 0 \\
\Pr(Necrosis = yes \mid Shape = ulcerating, Length \geq 6cm) &= 0.85
\end{aligned}
$$

For the influence diagram, a total number of 217 probabilities have been specified.

With the utility node of the influence diagram is associated a set of utilities, specifying for each combination of values for the node's predecessors a number indicating desirability. For example, for a patient in whom treatment has not resulted in a prolonged life expectancy the diagram specifies the following utilities:

$$
\begin{aligned}
u(Life\ expectancy = same, Passage\ after\ therapy = improved, Migration = no) &= 0.80 \\
u(Life\ expectancy = same, Passage\ after\ therapy = improved, Migration = yes) &= 0.75 \\
u(Life\ expectancy = same, Passage\ after\ therapy = same, Migration = no) &= 0.60 \\
u(Life\ expectancy = same, Passage\ after\ therapy = same, Migration = yes) &= 0.35 \\
u(Life\ expectancy = same, Passage\ after\ therapy = worse, Migration = no) &= 0.20 \\
u(Life\ expectancy = same, Passage\ after\ therapy = worse, Migration = yes) &= 0.05
\end{aligned}
$$

For the diagram, a total number of 18 utilities have been specified. □

The conditional probabilities of an influence diagram provide all information necessary for uniquely defining a joint probability distribution on the statistical variables in the diagram that respects the independences portrayed by the diagram's qualitative part. This property is stated more formally in the following proposition. The notation $|_{X=x}$ is used to indicate that, in the preceding formula, the variables in the set X take the combination of values x.

Proposition 7.2.3 *Let $ID = (G, p, u)$ be an influence diagram with the chance nodes $C(G)$ and the decision node D. Then, for each decision alternative d_j of D,*

$$\Pr(C(G), D = d_j) = \prod_{C_i \in C(G)} p(C_i \mid \pi_G(C_i))\Big|_{D = d_j}$$

defines a joint probability distribution Pr *on $C(G)$ that respects the independences portrayed by the digraph G.*

From the previous proposition, we have that the part of an influence diagram that is induced by its chance nodes provides for probabilistic inference conditional on any decision alternative. By taking the various utilities associated with the utility node into consideration, the diagram provides for computing the decision alternative with maximum expected utility. This property is stated more formally in the following proposition.

Proposition 7.2.4 *Let $ID = (G, p, u)$ be an influence diagram with the chance nodes $C(G)$, the decision node D, taking one of the values $d_1, \ldots, d_m$, $m \geq 1$, and the utility node U. Let $O \subseteq V(G)$ be the set of observed chance nodes in G and let o denote the corresponding observations. Then, the decision alternative with maximum expected utility, denoted d_{max}, equals*

$$d_{max} = \arg \max_{d_j \in \{d_1, \ldots, d_m\}} \Big(eu(D = d_j \mid O = o) \Big)$$

where, for each decision alternative d_j of D, we have that

$$eu(D = d_j \mid O = o) = \sum_{\pi_G(U)} \Pr(\pi_G(U) \mid D, O) \cdot u(\pi_G(U)) \Big|_{\substack{D = d_j \\ O = o}}$$

In the following, we illustrate computing the preferred decision alternative for our example influence diagram.

Example 7.2.5 We consider the *Oesophagus* influence diagram from Figure 7.1. We are interested in the decision alternative with maximum expected utility for a ninety-two year old patient who has not suffered from any weightloss and in whom the absence of necrosis and a fistel have been established; the observations for the set of observed variables $O = \{Age, Weightloss, Necrosis, Fistula\}$ will be indicated by o. We start by computing the expected utility of positioning an *endoprosthesis* in the patient's oesophagus. From Proposition 7.2.4, we find that

$$eu(endoprosthesis \mid o) = \sum_{\pi(U)} \Pr(\pi(U) \mid endoprosthesis, o) \cdot u(\pi(U))$$

where the set $\pi(U)$ comprises the chance nodes *Life expectancy*, *Passage after therapy*, and *Migration*. The posterior probabilities Pr(*Life expectancy*, *Passage after therapy*,

Migration | *endoprosthesis*, *o*) are computed from the diagram for all values of the three variables involved. The variables *Life expectancy* and *Passage after therapy* take the values *improved*, *same*, and *worse*; the variable *Migration* takes the values *yes* and *no*. Positioning an endoprosthesis in a patient's oesphagus only serves to improve the passage of food and has no effect on life expectancy. The posterior probabilities of the outcome scenarios involving an improved or worsened life expectancy therefore equal zero. Now, inserting the computed posterior probabilities and the various utilities involved in the equation above, we find

$$
\begin{aligned}
&eu(endoprosthesis \mid o) = \\
&= \Pr(LE{=}s, Pass = i, Migr = no \mid endoprosthesis, o) \cdot u(LE{=}s, Pass = i, Migr = no) \\
&\quad + \Pr(LE{=}s, Pass = i, Migr = yes \mid endoprosthesis, o) \cdot u(LE{=}s, Pass = i, Migr = yes) \\
&\quad + \Pr(LE{=}s, Pass = s, Migr = no \mid endoprosthesis, o) \cdot u(LE{=}s, Pass = s, Migr = no) \\
&\quad + \Pr(LE{=}s, Pass = s, Migr = yes \mid endoprosthesis, o) \cdot u(LE{=}s, Pass = s, Migr = yes) \\
&\quad + \Pr(LE{=}s, Pass = w, Migr = no \mid endoprosthesis, o) \cdot u(LE{=}s, Pass = w, Migr = no) \\
&\quad + \Pr(LE{=}s, Pass = w, Migr = yes \mid endoprosthesis, o) \cdot u(LE{=}s, Pass = w, Migr = yes) \\
&= 0.283 \cdot 0.8 + 0.0230 \cdot 0.75 + 0.458 \cdot 0.6 + 0.0372 \cdot 0.35 + 0.175 \cdot 0.2 + 0.0142 \cdot 0.05 \\
&= 0.567
\end{aligned}
$$

abbreviating the names of the variables and values for readability. For the two decision alternatives *surgery* and *no treatment*, we analogously find

$$eu(surgery \mid o) = 0.495$$

and

$$eu(no\ treatment \mid o) = 0.6$$

From the expected utilities for the three decision alternatives, we have that for our ninety-two year old patient the decision alternative to refrain from treatment has maximum expected utility and, hence, is the preferred decision alternative. □

For computing from an influence diagram the decision alternative with maximum expected utility, various different algorithms are available [Shachter, 1986, Cooper, 1988]. The algorithm by R.D. Shachter operates directly on an influence diagram by recursively reducing the diagram and combining probabilities and utilities [Shachter, 1986]. The algorithm by G.F. Cooper, on the other hand, transforms an influence diagram into a Bayesian belief network and subsequently performs probabilistic inference [Cooper, 1988]. Since in the sequel we will build upon Cooper's algorithm, we briefly summarize its essence.

The basic idea of Cooper's algorithm is to transform an influence diagram into a Bayesian belief network by converting both the decision node and the utility node

into a chance node and transforming the utility function to a probability function. In converting the decision node D into a chance node, its incoming arcs are removed. The various decision alternatives $d_1, \ldots, d_m$, $m \geq 1$, of the decision node are taken for the new chance node's values. These values are assigned an even probability distribution. The utility node is converted into a binary chance node taking one of the truth values *true* and *false*. Without loss of generality, we assume that the utilities for the diagram's utility node lie between zero and one, where zero is the utility for the worst possible outcome scenario and one is the utility for the best outcome scenario. For the chance node U modeling the original utility node, a set of conditional probabilities $p(U \mid \pi_G(U))$ is defined with

$$p(U = true \mid \pi_G(U)) = u(\pi_G(U))$$

For any decision alternative d_j, $j = 1, \ldots, m$, the posterior probability $\Pr(D = d_j \mid U = true, O = o)$ given the observations o and given the instantiation *true* for U can now be shown to be equal to the expected utility $eu(D = d_j \mid O = o)$ of d_j. Determining the decision alternative that maximises expected utility therefore amounts to establishing the decision alternative that yields the maximum posterior probability given the available observations and the value *true* for U.

7.3 Sensitivity analysis

Sensitivity analysis is a general technique for investigating the effects of the inaccuracies in the parameters of a mathematical model on this model's output [Habbema *et al.*, 1990, Morgan & Henrion, 1990]. The analysis basically amounts to systematically varying the values of a number of parameters of the model under study and computing, for each of the combinations of values considered, the outcome of the model. For an influence diagram, sensitivity analysis amounts to varying the assessments for one or more conditional probabilities or utilities, termed parameters, of the diagram's quantitative part simultaneously. The effect of these variations on the expected utility of each decision alternative and, in particular, on the decision alternative with maximum expected utility is studied. In fact, the effect of varying assessments on the decision alternative with maximum expected utility can be established from the observed effect of these variations on the expected utility of each alternative. Having established the relation between the expected utility of each alternative and the parameters under study, it can be computed how much the original assessments for these parameters should be changed to alter the decision alternative with maximum expected utility from the formerly preferred alternative to any other alternative. We are interested in determining the *minimum deviation* in the parameter assessments that causes this change in preferred decision. This minimum deviation is a measure of the robustness

of the influence diagram to inaccuracies in its parameters; it gives an indication of the diagram's reliability.

The simplest type of sensitivity analysis is a *one-way sensitivity analysis*. In a one-way sensitivity analysis of an influence diagram, the assessments for the conditional probabilities and utilities in the diagram are varied one at a time, keeping all other assessments fixed. A one-way sensitivity analysis thus reveals the independent effect of deviation from the assessment of a single conditional probability or utility. To assess the joint effect of various conditional probabilities and utilities together, higher order sensitivity analyses are required. In this chapter, we discuss *two-way sensitivity analysis*, in addition to one-way sensitivity analysis. In a two-way sensitivity analysis, the assessments for two conditional probabilities or utilities are varied simultaneously, keeping all others fixed. Such an analysis reveals how the two parameters under study interact in their effect on expected utility. In principle, a three-way or higher order sensitivity analysis is also possible. However, the results of such an analysis are hard to interpret. In clinical decision analysis, it is customary to perform one- and two-way sensitivity analyses; the results of these analyses can be represented graphically and are, therefore, easier to use.

In principle, in a one-way and two-way sensitivity analysis of an influence diagram, every single assessment and every pair of assessments in the diagram is investigated. The analysis can be carried out by simply varying the assessments stepwise using a sufficiently large number of steps and, subsequently, evaluating the influence diagram for each step. However, such a straightforward sensitivity analysis is far too time-consuming. Even for rather small influence diagrams, easily tens of thousands of evaluations would be required. Until now, to our knowledge, no efforts have been undertaken to increase the efficiency of performing a sensitivity analysis of an influence diagram. For belief networks, however, an efficient method is available [Castillo *et al.*, 1997b, Coupé & Van der Gaag, 1998] (see also Chapter 4). As influence diagrams are closely related to belief networks, we briefly review this method. It forms the basis of the research presented in this chapter.

In a sensitivity analysis of a belief network, the effect of varying parameter assessments on a prior or posterior probability computed from the network is investigated. With respect to the results of such an analysis, Coupé and Van der Gaag, and Castillo *et al.* identified two important properties. Firstly, the independences reflected by the qualitative part of a belief network allow for identifying, by simple visual inspection, conditional probabilities that cannot affect the probability of interest in the network. These conditional probabilities can be excluded from the sensitivity analysis. Secondly, the remaining conditional probabilities relate to the probability of interest as a quotient of two multilinear functions. We will detail this shortly.

Coupé and Van der Gaag use the term *sensitivity set* to refer to the set nodes whose conditional probabilities may affect the probability of interest. To identify the sensi-

tivity set, they propose constructing an auxiliary graph. To that end, to every node in the network an auxiliary parent is added. Intuitively said, an auxiliary node represents the inaccuracy in the conditional probabilities of its child. Then, the sensitivity set for the node of interest, that is, the node to which the probability of interest relates, consists of those nodes in the network for which the auxiliary parent is not d-separated from the auxiliary parent of the node of interest. For details in d-separation, we refer the reader to [Pearl, 1988]. The remaining auxiliary nodes are d-separated from the auxiliary parent of the node of interest and, as such, their conditional probabilities do not affect the probability of interest. Note that, in addition to these uninfluential probabilities that are identified by qualitative considerations only, there are conditional probabilities that cannot affect the probability of interest because they specify values that do not correspond with the observations.

As the conditional probabilities for nodes in the sensitivity set may influence the probability of interest, it is useful to perform a sensitivity analysis with respect to these probabilities. When varying the assessment for a conditional probability under study in a sensitivity analysis, the assessments for the network's probabilities that pertain to the same conditional probability distribution have to be adjusted such that the sum of all probability assessments in this distribution again equals one. The assessments that have to be adjusted are called *co-varying* probabilities for the probability under study. Assume, that we have a probability under study $x = p(a_i \mid \pi')$ that pertains to the value a_i of the variable A and the configuration π' for the parents of A. The possible values for the variable A are $a_1, \ldots, a_k$, $k \geq 1$. The co-varying probabilities for x then are the conditional probabilities $p(a_j \mid \pi')$, $j \neq i$. The sum of the assessments for the co-varying probabilities, here called the *residual probability*, equals $1 - x$. Now, in varying the assessment for x, the assessments for these co-varying parameters are adjusted by the ratio of the new residual probability $1 - x$ and the original residual probability $1 - p(a_i \mid \pi')$. Considering $p(a_j \mid \pi')$ as a function of x, denoted $p(a_j \mid \pi')(x)$, we have that

$$p(a_j \mid \pi')(x) = p(a_j \mid \pi') \cdot \frac{1 - x}{1 - p(a_i \mid \pi')}$$

for all $j = 1, \ldots, k$, $j \neq i$. If more than one probability from the same probability distribution is investigated in a sensitivity analysis, the residual probability equals one minus the sum of the assessments for these probabilities.

Under the assumption that systematic variation of the network's probabilities is carried out as described above, in general, a probability of interest can be expressed as a quotient of two multilinear functions in an arbitrary number of probabilities under study [Castillo *et al.*, 1997b, Coupé & Van der Gaag, 1998]. For one-way sensitivity analysis, we then find a quotient of two functions linear in the single probability under study. Writing $Pr(V|o)$ for the posterior probability of the node of interest V given

the observations o and x for the conditional probability under study, we have that

$$Pr(V|o) = \frac{a \cdot x + b}{c \cdot x + d}$$

For a two-way sensitivity analysis with respect to the conditional probabilities x and y, the relation is

$$Pr(V|o) = \frac{a \cdot xy + b \cdot x + c \cdot y + d}{e \cdot xy + f \cdot x + g \cdot y + h}$$

Coupé *et al.* and van der Gaag show that, for nodes in the network that are not observed and have no observed descendants, the relation between the probability of interest and one or two of their conditional probabilities is simply linear or bilinear, respectively.

The two properties of sensitivity analysis of belief networks reviewed above allow for considerably more efficient sensitivity analysis than by performing straightforward variation of assessments. Nodes that are not in the sensitivity set are excluded from the analysis and for the remaining conditional probabilities it suffices to establish the coefficients in the functional relation. To establish these coefficients only a limited number of network evaluations is required; one for each coefficient. This gives a system of linear equations which is easily solved to give the required coefficients.

As belief networks are closely related to influence diagrams, the properties described for belief networks can be used in developing an efficient method for sensitivity analysis of influence diagrams. In the previous section is has been mentioned that an influence diagram can be transformed to a belief network using Cooper's transformation. Of the resulting belief network only the probability distribution of the utility node is meaningful. The expected utility of a specific decision alternative in the influence diagram equals the posterior probability of the value *true* for the utility node conditional on this decision alternative. The concept of the sensitivity set now allows for identifying conditional probabilities that do not affect expected utility. Furthermore, for one- and two-way sensitivity analysis of an influence diagram, equivalent functional relations hold as given above. This allows us to efficiently compute the effect of varying probability assessments on the preferred decision alternative, that is, it allows for efficiently computing the minimum deviation required to change this initially preferred decision alternative to any other alternative. In the next section, we focus on computing the minimum deviation in one or two of a diagram's conditional probabilities to change the preferred decision alternative. In section 5, this is done for sensitivity analysis with respect to the utilities in an influence diagram. As for the utilities in an influence diagram, the functional relations observed obey additional constraints compared to the functional relations in terms of conditional probabilities, we will discuss sensitivity analysis with respect to utilities in more detail.

7.4 Sensitivity analysis with respect to probabilities

As indicated in the previous section, the concept of sensitivity set can be used to identify the conditional probabilities in an influence diagram that cannot affect expected utility. More precisely, in the belief network derived from this influence diagram, these are the conditional probabilities pertaining to the nodes that are not in the sensitivity set of the utility node. We use the notation $Sen(U, \{D, O\})$ to indicate the sensitivity set of the utility node given the set of observed nodes O and the instantiated decision node D. The following example illustrates the concept of the sensitivity set in an influence diagram.

Example 7.4.1 We consider once again the *Oesophagus* diagram. We are interested in the set of nodes whose conditional probabilities may influence expected utility of any of the decision alternatives. To that end, we first construct the belief-network representation of the *Oesophagus* diagram.

Now, consider a patient aged ninety-two who has no fistel and an impaired passage. This patient has also been observed to have no necrosis. The sensitivity set for the node *Utility* given the set of observed nodes $O = \{Age, Fistula, Necrosis, Passage\}$ and the decision node *Therapy* equals

$$Sen(Utility, \{O, Therapy\}) = \{Migration, Life\ expectancy, Passage\ after\ therapy, Perforation, Residual, Passage, Length, Shape, Fistula, Necrosis, Location\}$$

Upon performing a sensitivity analysis of the *Oesophagus* diagram, therefore, the conditional probabilities for these eleven nodes need be investigated. The probabilities pertaining to the nodes *Age*, *Gastro length*, *Gastro location*, and *Weight loss* have no influence on expected utility. Note, that the nodes that are not in the sensitivity set are not necessarily d-separated from the utility node. Given the set of observed nodes O, for example, we see that the uninfluential node *Weight loss* is d-separated from the node *Utility*, but the node *Gastro location* is not d-separated. Since *Gastro location* is not observed and has no observed descendants, however, this node cannot exert any diagnostic influence via its parent *Location* on utility. □

For the conditional probabilities of an influence diagram, there exist simple functions describing the relation between expected utility and a single conditional probability that is varied in a one-way sensitivity analysis. These functions are the same as in a belief network. In a one-way sensitivity analysis, expected utility is a quotient of two functions depending linearly on a conditional probability under study. We illustrate this with an example.

Example 7.4.2 We consider again the *Oesophagus* diagram. For the patient introduced in Example 7.4.1, we have investigated the sensitivity of the recommended therapy. From Example 7.4.1, we have that the probabilities for the nodes *Age*, *Gastro length*, *Gastro location*, and *Weight loss* have no effect on expected utility. The conditional probabilities for the remaining nodes may be influential. Now, upon performing a sensitivity analysis of the *Oesophagus* diagram, it became clear that expected utility and the recommended decision are hardly effected by varying the conditional probabilities in the diagram independently. Carefully studying both the structure of the diagram and its probability assessments revealed that the strength of the observations blocks the influence of many conditional probabilities. For conditional probabilities that do affect expected utility, the effective change in the expected utility of each decision alternative does mostly not lead to a change in recommended decision. These findings result from the fact that we chose to simplify the original *Oesophagus* diagram considerably in order to obtain a practical example with a limited number of variables. In this small network, shown in Figure 7.1, observations easily cause blocking of influences. Sensitivity analyses of the original network, however, did reveal the typical curves that describe expected utility in terms of a conditional probability. To illistrate these typical effects here, we modified our patient profile; we assume the variable *passage* to be unobserved. Instead, we take the observation *no* for the variable *Weight loss*. Although the resulting profile is not a realistic patient profile, we will use it throughout the example to illustrate our theoretical results.

Now, for a patient aged ninety-two who has no fistel, no weight loss and no necrosis, only the three nodes *Age*, *Gastro length*, *Gastro location* are not included in the sensitivity set of the utility node given the set of observed nodes $O = \{Age, Fistula, Necrosis, Passage\}$. The conditional probabilities for the remaining nodes in the diagram may affect expected utility. We now investigate the functional relation between expected utility and the conditional probability $x = p(Weight\ loss = no \mid Passage = impaired)$ for the node $Weightloss \in Sen(Utility, \{O', Therapy\})$. We will write o for the combination of values for the observed nodes. For each treatment alternative, in Figure 7.2, the relation between expected utility and the conditional probability under study is shown. From Figure 7.2, we see that the expected utility of positioning an endoprosthesis in the patients oesophagus shows a relatively high sensitivity for the conditional probability $p(Weight\ loss = no \mid Passage = impaired)$ at the original point estimate of 0.1. Varying this estimates with less than +0.1 changes the decision alternative with maximum expected utility from no treatment to endoprosthesis. As an endoprosthesis is positioned in the oesophagus with the primary objective to improve the passage of food through the oesophagus, this result is not surprising. Increasing $p(Weight\ loss = no \mid Passage = impaired)$ not only increases the chance for a patient with an impaired passage to keep his normal weight, but also increases the chance that a patient with normal weight, that is, a patient who is apparently not seriously affected

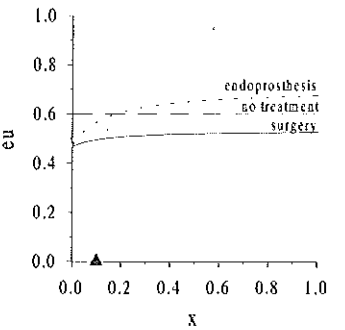


Figure 7.2: The functions relating the expected utilities for the decision alternatives *endoprosthesis*, *surgery*, and *no treatment* to the conditional probability under study $x = p(\textit{Weight loss} = \textit{no} \mid \textit{Passage} = \textit{impaired})$.

by the tumour, has indeed an impaired passage. As such, *no* weight loss is no longer a contraindication for positioning an endoprosthesis.

In formula, the functional relation between the expected utility of positioning an endoprosthesis and the probability x is

$$eu(\textit{endoprosthesis} \mid o) = \frac{0.7050 \cdot x + 0.0696}{x + 0.1470}$$

For the decision alternatives *surgery* and *no treatment*, the functional relations are

$$eu(\textit{surgery} \mid o) = \frac{0.5353 \cdot x + 0.0688}{x + 0.1470}$$

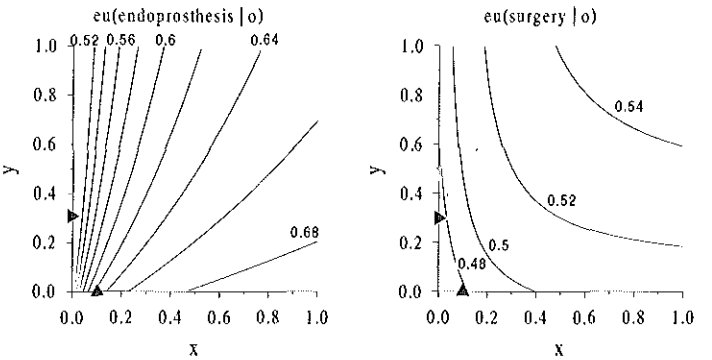
and

$$eu(\textit{no treatment} \mid o) = \frac{0.6 \cdot x + 0.0882}{x + 0.1470} = 0.6$$

The constants in these relations can be determined by evaluating the diagram three times for three different values of the probability under study x. The resulting system of linear equations is solved to give the required constants [Coupé & Van der Gaag, 1998]. □

In a two-way sensitivity analysis, expected utility is a quotient of bilinear functions in two conditional probabilities under study. The following example illustrates two-way sensitivity analysis of the *Oesophagus* diagram.

Example 7.4.3 We discuss a two-way sensitivity analysis of the *Oesophagus* diagram, that is, we investigate the effect of varying two conditional probabilities simultaneously on expected utility. We take the same observations o for the set of observed nodes O'



(a) (b)

Figure 7.3: The functions relating the expected utility of the decision alternatives *endoprosthesis* (a) and *surgery* (b) to the conditional probabilities under study $x = p(Weight\ loss = no \mid Passage = impaired)$ and $y = p(Length < 6\ cm)$. The original assessments for x and y, indicated by triangles, are 0.1 and 0.3, respectively.

as in Example 7.4.2. The two probabilities under study are $p(Weight\ loss = no \mid Passage = impaired)$ and $y = p(Length < 6\ cm)$.

In Figures 7.3a and b, the relation between expected utility and the two probabilities under study is shown for the decision alternative to position an endoprosthesis in the oesophagus and for the alternative to perform surgery. The expected utility of no treatment is not affected by varying any of the two probabilities under study; it invariantly takes the value 0.6. The contour lines in the figures connect combinations of values for the two probabilities under study that result in the same value for the expected utility. Note that the distances between the contour lines differ, indicating that varying the two probabilities simultaneously has a joint effect on expected utility beyond the effects of their separate variation.

Comparing the two figures reveals that, whatever combination of values the two conditional probabilities take, the expected utility of positioning an endoprosthesis is always larger than the expected value of operating. Using furthermore the knowledge that the expected utility of no treatment is 0.6, we see from Figure 7.3a that changing the value of $p(Length < 6\ cm)$ to values lower than about 0.08 changes the decision alternative with maximum expected utility from no treatment to endoprosthesis. Similarly, changing the value of $p(Weight\ loss = no \mid Passage = impaired)$ to values higher than about 0.2 changes the preferred decision from no treatment to endoprosthesis. This latter effect was also observed in Example 7.4.2, Figure 7.2. It is related to the predictive value of observing *no* weight loss for a patient with a tumour in the oesophagus. The former effect simply describes that, as the chance of having a large

tumour increases, it becomes more favourable to position an endoprosthesis compared to no treatment. This is explained as follows; the larger a tumour in the oesophagus is, the larger the chance becomes that it will cause problems for the passage of food. An endoprosthesis thus becomes more desirable.

In formula, the relations between expected utility and x and y are

$$eu(\textit{endoprosthesis} \mid o) = \frac{0.7050 \cdot x \cdot y + 1.7565 \cdot x + 0.3974 \cdot y + 0.0751}{x \cdot y + 2.4915 \cdot x + 0.8588 \cdot y + 0.1528}$$

$$eu(\textit{surgery} \mid o) = \frac{0.7216 \cdot x \cdot y + 1.2778 \cdot x + 0.4288 \cdot y + 0.0635}{x \cdot y + 2.4915 \cdot x + 0.8588 \cdot y + 0.1528}$$

and

$$eu(\textit{no treatment} \mid o) = \frac{0.6 \cdot x \cdot y + 1.4949 \cdot x + 0.5153 \cdot y + 0.0917}{x \cdot y + 2.4915 \cdot x + 0.8588 \cdot y + 0.1528} = 0.6$$

□

So far, we have focused on the functional relations that hold between expected utility and each single conditional probability or each pair of conditional probabilities in an influence diagram. These functional relations describe the effect of one or two conditional probabilities on each decision alternative in isolation. As the outcome of interest in an influence diagram is the decision alternative with highest expected utility, studying the effect of varying one or two probabilities on each decision alternative in isolation does not suffice. The effects of variations on all decision alternatives are to be considered in relation to each other. For this purpose, we focus on the *minimum deviation*. The *minimum deviation* is the smallest change in the assessment(s) for one or two probabilities under study that leads to a change in the preferred decision alternative.

We start with the computation of the minimum deviation in a one-way sensitivity analysis. We observe that the expected utility of each decision alternative can be represented as a line or curve. Now, exploiting the functional form of these curves, their points of intersection can be computed. The values of the probability under study at the intersections are called *critical values*. The minimum deviation now is the smallest change in the original assessment for the probability under study that is required to reach a critical value. If the probability under study surpasses this critical value, the preferred decision alternative changes to the alternative to which the line pertains that intersects with the line for the formerly preferred alternative. The following proposition details the computation of the minimum deviation more formally.

Proposition 7.4.4 *Let ID be an influence diagram as defined in Definition 7.2.1. Let D be the diagram's decision node taking one of the values $\{d_1, \ldots, d_k\}$. Let O be the set of observed nodes and let o denote the corresponding observations. Let $Sen(U, \{D, O\})$*

be the sensitivity set for utility node U given D and O, and let $x = p(c_s \mid \pi')$ be a conditional probability pertaining to node $C_s \in Sen(U, \{D, O\})$. Let a, b, c, and e be constants such that the expected utility of the preferred decision alternative d_{max} equals

$$eu(d_{max} \mid o) = \frac{a \cdot x + b}{c \cdot x + e}$$

For each $d_i \in \{d_1, \ldots, d_k\}$, $d_i \neq d_{max}$, let a_i, b_i, c_i, and e_i be constants such that the expected utility of d_i equals

$$eu(d_i \mid o) = \frac{a_i \cdot x + b_i}{c_i \cdot x + e_i}$$

Then, the minimum deviation Δ_{min} required to change the decision alternative with maximum expected utility is

$$\Delta_{min} = \min_{\substack{d_i \in D \\ d_i \neq d_{max}}} \left(-\frac{b - b_i}{a - a_i} - x_0\right)$$

where x_0 is the original assessment of x and $0 \leq x_0 + \Delta_{min} \leq 1$.

Proof. In the following, we first show that the denominators in the expressions for $eu(d_{max} \mid o)$ and $eu(d_i \mid o)$ in terms of x are equal, up to a constant factor. The minimum deviation in the parameter x required to change the decision alternative with maximum expected utility then is given by the difference between the original assessment for x and the point of intersection of the two curves for $eu(d_{max} \mid o)$ and $eu(d_i \mid o)$.

Using Cooper's transformation, the expected utilities of the decision alternative d_{max} and any other alternative d_i can be written as

$$eu(d_{max} \mid o) = \Pr(U = true \mid d_{max}, o) = \frac{\Pr(U = true, d_{max}, o)}{\Pr(d_{max}, o)} \tag{7.1}$$

and

$$eu(d_i \mid o) = \Pr(U = true \mid d_i, o) = \frac{\Pr(U = true, d_i, o)}{\Pr(d_i, o)} \tag{7.2}$$

where Pr is the joint probability distribution defined by the belief network that results from Cooper's transformation of ID. From Proposition 7.2.3 and the property of marginalization, we have that both the numerator and the denominator in the above equations can be written as a sum of products of conditional probabilities. In the two denominators, most conditional probabilities in the products are the same. Only those probabilities differ that pertain to the chance node D or the direct descendants of D. For computing $\Pr(d_{max}, o)$ and $\Pr(d_i, o)$, the joint probability distribution of the belief network is marginalized over all variables except those contained in $\{D, O\}$. Since, by convention, none of the descendants of D is observed, marginalizing over these

variables amounts to multiplying by one. Conditional probabilities pertaining to these unobserved descendants therefore have no effect on $\Pr(d_{max}, o)$ and $\Pr(d_i, o)$. As for the prior probabilities pertaining to node D itself, it is easily seen that all terms in both the numerator and the denominator in each of the two equations above contain the prior probability of the chosen value for D. This prior probability therefore cancels out in the division. We conclude that the denominators in the functional relation expressing $eu(d_{max} \mid o)$ and $eu(d_i \mid o)$ in terms of x are the same up to a constant factor; as such the constants in the denominator can be taken the same, that is, $c = c_i$ and $e = e_i$.

Now, the critical value of x, denoted x_{crit}, for which the expected utilities of d_{max} and d_i are equal can be computed from

$$\frac{a \cdot x_{crit} + b}{c \cdot x_{crit} + e} = \frac{a_i \cdot x_{crit} + b_i}{c \cdot x_{crit} + e}$$

We find that

$$x_{crit} = -\frac{b - b_i}{a - a_i}$$

In order to change the decision alternative with maximum expected utility from d_{max} to d_i, the original assessment x_0 for parameter x should be varied more than the deviation Δ_i required to obtain equal expected utilities for both alternatives,

$$\Delta_i = x_{crit} - x_0 = -\frac{b - b_i}{a - a_i} - x_0$$

The minimum deviation Δ_{min} that changes the decision alternative with maximum expected utility from d_{max} to an other decision alternative, now, is the minimum of the deviations Δ_i required for each decision alternative d_i, $d_i \neq d_{max}$.

$$\Delta_{min} = \min_{\substack{d_i \in D \\ d_i \neq d_{max}}} \left(-\frac{b - b_i}{a - a_i} - x_0\right)$$

□

We illustrate computing the minimum deviation in one-way sensitivity analysis of an influence diagram with our running example.

Example 7.4.5 Consider the one-way sensitivity analysis of the *Oesophagus* diagram that is presented in Example 7.4.2. In the example, the functional relations expressing expected utility in terms of the conditional probability $x = p(\textit{Weight loss} = \textit{no} \mid \textit{Passage} = \textit{impaired})$ are given. In Figure 7.2 these functions are represented graphically; the figure shows that increasing the assessment of x can change the preferred decision alternative from *no treatment* to *endoprosthesis*.

Now, the critical value x_{crit} for which the expected utilities of endoprosthesis and no treatment are equal can be computed from the functional relations describing $eu(endoprosthesis \mid o)$ and $eu(no\ treatment \mid o)$ in terms of x.

$$\frac{0.7050 \cdot x_{crit} + 0.0696}{x_{crit} + 0.1470} = \frac{0.6 \cdot x_{crit} + 0.0882}{x_{crit} + 0.1470}$$

Solving this for x_{crit} yields $x_{crit} = -(0.0882 - 0.0696)/(0.6 - 0.705) = 0.177$. In other words, if the original assessment of 0.1 for x is increased by at least 0.077, then the preferred decision alternative changes from *no treatment* to *endoprosthesis*. The recommended decision is thus quite sensitive to changes in the assessment for x. □

To determine the minimum deviation for two probabilities under study in a two-way sensitivity analysis, a similar procedure is adopted. The functional relation between the expected utility of each decision alternative and the two conditional probabilities under study can be represented as a surface. Now, exploiting the functional form of these surfaces, their intersection lines can be computed. These intersection lines, called *critical lines*, connect the combination of values of the probabilities under study for which the expected utilities of two decision alternatives are equal. The minimum deviation then is the smallest change in the original assessments of the probabilities under study that is required to reach a critical line. The minimum deviation thus is a vector whose length is the distance from the point indicating the original assessments perpendicular to that line. If two probabilities under study are varied by less than the minimum deviation, the preferred decision alternative remains the same. For variations larger than the minimum deviation, the preferred decision may change. However, whether or not the preferred decision alternative actually changes depends not only on the length of the vector representing the minimum deviation but also on the direction of this vector.

Proposition 7.4.6 *Let ID be an influence diagram as defined in Definition 7.2.1. Let D be the diagram's decision node taking one of the values $\{d_1, \ldots, d_k\}$. Let O be the set of observed nodes and let o denote the corresponding observations. Let $Sen(U, \{D, O\})$ be the sensitivity set for utility node U given D and O, and let $x = p(c_s \mid \pi')$ and $y = p(c_t \mid \pi'')$ be conditional probabilities pertaining to nodes $C_s, C_t \in Sen(U, \{D, O\})$. Let a, b, c, e, f, g, h, and l be constants such that the expected utility of the preferred decision alternative d_{max} is*

$$eu(d_{max} \mid o) = \frac{a \cdot x \cdot y + b \cdot x + c \cdot y + e}{f \cdot x \cdot y + g \cdot x + h \cdot y + l}$$

For each $d_i \in \{d_1, \ldots, d_k\}$, $d_i \neq d_{max}$, let a_i, b_i, c_i, e_i, f_i, g_i, h_i, and l_i be constants such that the expected utility of d_i equals

$$eu(d_i \mid o) = \frac{a_i \cdot x \cdot y + b_i \cdot x + c_i \cdot y + e_i}{f_i \cdot x \cdot y + g_i \cdot x + h_i \cdot y + l_i}$$

Then, the minimum deviation required to change the preferred decision alternative from d_{max} to any other decision alternative equals Δ_{min} where

$$\|\Delta_{min}\| = \min_{\substack{d_i \in D \\ d_i \neq d_{max}}} \|\Delta_i\|$$

and where, for each d_i, we have that

$$\Delta_i = \begin{pmatrix} x_{crit} - x_0 \\ y_{crit} - y_0 \end{pmatrix}$$

where x_{crit} is a solution to

$$y_0 + (x_{crit} - x_0) \cdot \frac{-1}{f'(x_{crit})} = f(x_{crit})$$

with

$$f(x) = -\frac{(b - b_i) \cdot x + (e - e_i)}{(a - a_i) \cdot x + (c - c_i)}$$

and y_{crit} equals $f(x_{crit})$, and x_0 and y_0 are the original assessments for x and y, respectively.

Proof. Similarly to the proof of Proposition 7.4.4, it can be shown that the constants f, g, h, and l are equal, up to a constant factor, to the constants f_i, g_i, h_i, and l_i. As such, the critical line at which the expected utilities $eu(d_{max} \mid o)$ and $eu(d_i \mid o)$ are equal is expressed by

$$a \cdot x \cdot y + b \cdot x + c \cdot y + e = a_i \cdot x \cdot y + b_i \cdot x + c_i \cdot y + e_i$$

Writing y in terms of x, we find

$$y = -\frac{(b - b_i) \cdot x - (e - e_i)}{(a - a_i) \cdot x - (c - c_i)}$$

For ease of exposition, we write $f(x)$ to denote this function. Now, in order to change the decision alternative with maximum expected utility from d_{max} to d_i, the original assessments for x and y should be varied until they lie on this critical line. The minimum change in x and y required is given by the distance perpendicular to the critical line. The line through (x_0, y_0) perpendicular to the critical line is

$$y = y_0 + (x - x_0) \cdot \frac{-1}{f'(x)}$$

where $f'(x)$ is the derivative of $f(x)$. This perpendicular line and the critical line intersect at the critical point (x_{crit}, y_{crit}), where x_{crit} is a solution to

$$y_0 + (x_{crit} - x_0) \cdot \frac{-1}{f'(x_{crit})} = f(x_{crit})$$

and $y_{crit} = f(x_{crit})$. Now, to change the preferred decision from d_{max} to d_i, the original assessments for the probabilities x and y thus need to be varied by at least $x_{crit} - x_0$ for probability x and $y_{crit} - y_0$ for probability y. The minimum deviation for decision alternative d_i is the vector

$$\Delta_i = \begin{pmatrix} x_{crit} - x_0 \\ y_{crit} - y_0 \end{pmatrix}$$

The minimum deviation required to change the preferred decision alternative from d_{max} to any other alternative then is given by Δ_j, where Δ_j pertains to the decision alternative d_j, $j = 1, \ldots, k$, $d_j \neq d_{max}$, for which the length $\|\Delta_j\|$ of vector Δ_j is minimized. □

In the following, we illustrate the computation of the minimum deviation for a two-way sensitivity analysis of the *Oesophagus* diagram.

Example 7.4.7 Consider the two-way sensitivity analysis of the *Oesophagus* diagram discussed in Example 7.4.3. The probabilities under study are $x = p(\textit{Weight loss} = \textit{no} \mid \textit{Passage} = \textit{impaired})$ and $z = p(\textit{Length} < 6\ \textit{cm})$. From the example, we saw that *no treatment* is the prefered decision alternative at the original assessments of 0.1 and 0.3 for x and y, respectively. By varying the assessments, the preferred decision alternative may change to *endoprosthesis*, but never to *operation*. Now, we are interested in the minimum deviation in x and y required to change the preferred decision from *no treatment* to *endoprosthesis*. The expected utilities for these two decision alternatives are equal if

$$\frac{0.7050 \cdot x \cdot y + 1.7565 \cdot x + 0.3974 \cdot y + 0.0751}{x \cdot y + 2.4915 \cdot x + 0.8588 \cdot y + 0.1528} = \frac{0.6 \cdot x \cdot y + 1.4949 \cdot x + 0.5153 \cdot y + 0.0917}{x \cdot y + 2.4915 \cdot x + 0.8588 \cdot y + 0.1528}$$

or

$$y = -\frac{0.262 \cdot x - 0.0166}{0.105 \cdot x - 0.118}$$

This expression represents the critical line between *no treatment* and *endoprosthesis*. It is shown in Figure 7.4.

Now, the minimum distance from (x_0, y_0) to the critical line between *no treatment* and *endoprosthesis* is in the direction perpendicular to this critical line. Thus, we construct a line perpendicular to the critical line, through the original assessments x_0 and y_0:

$$y = 0.3 + (x - 0.1) \cdot -\frac{(0.105 \cdot x - 0.118)^2}{0.0292}$$

This perpendicular line and the critical line intersect at the critical point (x_{crit}, y_{crit}) where x_{crit} is a solution to

$$0.3 + (x - 0.1) \cdot -\frac{(0.105 \cdot x_{crit} - 0.118)^2}{0.0292} = -\frac{0.262 \cdot x_{crit} - 0.0165}{0.105 \cdot x_{crit} - 0.118}$$

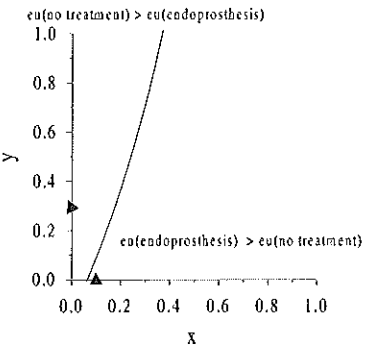


Figure 7.4: The critical line where the expected utility of positioning an endoprosthesis and no treatment are equal. Left from this line, the optimal decision is to treat. At the right of this line the optimal decision is to position an endoprosthesis.

Solving this quartic equation gives two imaginary and two real solutions. Only one real solution lies between zero and one, namely $x_{crit} = 0.176$. Filling in x_{crit} in the equation for the critical line gives $y_{crit} = 0.298$. The minimum deviation then equals

$$\Delta_{min} = (x_{crit} - x_0, y_{crit} - y_0) = (0.076, 0.002)$$

The minimum distance from (x_0, y_0) to the critical line thus is $\sqrt{(0.076^2 + 0.002^2)} = 0.076$. Recall from Example 7.4.5 that the minimum deviation when varying only x is 0.078. □

The properties of one- and two-way sensitivity analysis with respect to the probabilities in an influence diagram, as presented in this section, allow for considerably reducing the computational burden of such an analysis. The functional relations between expected utility and one or two conditional probabilities in the diagram can be established with a limited number of diagram evaluations. Furthermore, these functions can be exploited to efficiently compute the minimum deviation in one or two conditional probabilities for which the decision alternative with maximum expected utility changes from the formerly preferred decision alternative to another alternative. This minimum deviation is a measure of the sensitivity of the diagram to variation of the assessments for its conditional probabilities. From the example of the computation of the minimum deviation in two-way sensitivity analysis, however, it will be obvious that the computation of the minimum deviation becomes very complicated if several conditional probabilities are considered simultaneously in a sensitivity analysis. In general, the problem amounts to determining the minimum distance of a point in an n-dimensional space to a surface in this space that satisfies a number of constraints. In Chapter 5, we address computing the constants that establish the shape of the surface efficiently. To determine the minimum distance, efficient methods are still being sought.

7.5 Sensitivity analysis with respect to utilities

In a one- and two-way sensitivity analysis with respect to the utilities of an influence diagram, in principle, the effect on expected utility of varying every single utility and every pair of utilities is investigated. As the utilities in an influence diagram correspond with the conditional probabilities of the utility node in the belief network derived from the diagram by Cooper's transformation, sensitivity analysis with respect to utilities, in principle, is equivalent to sensitivity analysis with respect to probabilities. However, there are several differences. Firstly, unlike for the conditional probabilities in an influence diagram, it is not possible to exclude utilities from the analysis on the basis of the digraph of the diagram. Since the utility node is the node of interest in the belief network derived from an influence diagram and is uninstantiated, it is comprised in the sensitivity whatever observations are considered. As all utilities pertain to the utility node, each single utility may thus affect expected utility as well as the preferred decision alternative. Secondly, the functional relation expressing expected utility in terms of one or two utilities of the diagram is considerably simpler than this relation for conditional probabilities. Sensitivity analysis with respect to the utilities of an influence diagram can therefore be carried out very efficiently. Finally, and most importantly, unlike the diagram's probabilities, the set of utilities usually contains an underlying structure. This is called a preference structure on the possible outcomes of the decision problem; in a sensitivity analysis with respect to utilities, the various assumptions concerning this preference structure may be studied in detail.

This section is structured as follows. In Section 7.5.1, the essentials of multi-attribute utility theory (MAUT) are given. Multi-attribute utility theory is concerned with the specification of a utility function establishing preferences over outcomes that are described in terms of more than one attribute. The simplest type of utility function, in which *holistic assessments* are used, is presented. Furthermore, it is discussed how the utility function can be obtained from combining component utility functions. One specific form of such a combined utility function, the additive utility function, is detailed. In Section 7.5.2, then, one- and two-way sensitivity analysis with respect to holistic utility assessments is discussed. In Section 7.5.3, we discuss one- and two-way sensitivity analysis with respect to the components of the additive utility function.

7.5.1 Multi-attribute utility theory

Multi-attribute utility theory provides a framework for establishing a utility function. A utility function reflects preferences over outcomes. The possible outcomes of a decision problem can usually be expressed in terms of a number of attributes that are directly related to the objectives to be achieved. In a complex medical decision problem, for example, often the quality of life has to be weighed against the gain in life

expectancy that can be obtained. The attributes that measure the different objectives have to be combined into a single-valued utility function. This single-valued utility function provides for computing the preferred decision alternative; it is the alternative that maximizes the utility function [Von Neumann & Morgenstern, 1944, Chernoff & Moses, 1959, Raiffa & Schlaifer, 1961].

In principle, it is possible to specify the utility of each possible combination of values for all attributes that jointly describe an outcome, as a whole. A utility for a combination of values for all attributes is called a *holistic utility* [Keeney & Raiffa, 1976]. A set of holistic utilities, jointly representing the utility function for the problem under study, reflects the preferences of the decision maker. However, in a holistic utility function, the structure of these preferences is only implicitly given. Often, it is possible to explicitly detail the *preference structure* of a decision maker. To that end, a utility function is constructed in which the preferences over values are detailed for each attribute separately. A utility function for a single attribute is called a *component utility function*. The overall utility function, called a *multi-attribute utility function*, now, is a combination of these component utility functions. Such a multi-attribute utility function gives insight into the preferences for each attribute separately, in the importance of each attribute to the overall utilities, and in possible interactions between attributes. In general, a multi-attribute utility function $u(W_1, \ldots, W_q)$ takes the form

$$u(W_1, \ldots, W_q) = f(u(W_1), \ldots, u(W_q))$$

where W_i, $i = 1, \cdots, q$, are the attributes that jointly describe the possible outcomes and $u(W_i)$ are the corresponding component utility functions. In [Keeney & Raiffa, 1976], various types of combination function are discussed together with the conditions under which they are valid. Here, we will discuss the form that is used most, that is, the *additive utility function*. The additive utility function assumes no interaction between the various attributes. For a specification of the required conditions and other details, we refer the reader to [Keeney & Raiffa, 1976].

The additive utility function takes the following form:

$$u(W_1, \ldots, W_q) = \sum_{i=1,\ldots,q} k_i \cdot u(W_i)$$

Each component utility function $u(W_i)$, $i = 1, \cdots, q$, is normalized such that the utility of the least preferred value of attribute W_i, w_i^o, equals zero, that is $u(w_i^0) = 0$, and the utility of the most preferred value, w_i^*, equals one, that is, $u(w_i^*) = 1$. Furthermore, the multi-attribute utility function is normalized by $u(w_1^0, \ldots, w_q^0) = 0$ and $u(w_1^*, \ldots, w_q^*) = 1$. The scaling constants are given by $k_i = u(w_1^0, \ldots, w_{i-1}^0, w_i^*, w_{i+1}^0, \ldots, w_q^0)$, $i = 1, \cdots, q$; they add up to one to ensure normalisation of the multi-attribute utility function.

In the following, we present an additive utility function for the *Oesophagus* diagram that is based on component utility functions for the various attributes involved.

Example 7.5.1 We consider again the *Oesophagus* diagram from Figure 7.1. From the figure we see that the possible outcomes considered are determined by the three attributes *Life expectancy*, *Passage after therapy*, and *Migration*. Now for each of these three attributes, a component utility function, describing the preferences of each of its values, has been assessed. As the variable *Migration* has only two values, the utility function is straightforward; the utility of the least preferred value *yes* is zero and the utility of the most preferred value *no* is one. For the variable *Life expectancy*, the utility of the value *same* relative to the utilities of the least preferred value *worse* and the most preferred value *improved* has been established at 0.5. For the variable *Passage after therapy*, taking the values *improved*, *same* and *worse*, the utility of the intermediate value *same* is assessed to be 0.8. We thus have

$$
\begin{aligned}
u(\textit{Life expectancy} = \textit{improved}) &= 1\\
u(\textit{Life expectancy} = \textit{same}) &= 0.50\\
u(\textit{Life expectancy} = \textit{worse}) &= 0
\end{aligned}
$$

$$
\begin{aligned}
u(\textit{Passage after therapy} = \textit{improved}) &= 1\\
u(\textit{Passage after therapy} = \textit{same}) &= 0.80\\
u(\textit{Passage after therapy} = \textit{worse}) &= 0
\end{aligned}
$$

$$
\begin{aligned}
u(\textit{Migration} = \textit{no}) &= 1\\
u(\textit{Migration} = \textit{yes}) &= 0
\end{aligned}
$$

For combining these three utility functions into one multi-attribute utility function, three scaling constants, k_{LE}, k_{PAT}, and k_M, describing the importance of the attributes *Life expectancy*, *Passage after therapy*, and *Migration*, respectively, have been assessed at:

$$
\begin{aligned}
k_{LE} &= 0.75\\
k_{PAT} &= 0.24\\
k_M &= 0.01
\end{aligned}
$$

The multi-attribute utility function for the *Oesophagus* diagram has been fully specified by these component utility functions and scaling constants. As an example, we establish the utility of an improved life expectancy with the same passage of food through the oesophagus as prior to therapy and no problems of migration;

$$
\begin{aligned}
&u(\textit{Life expectancy} = \textit{improved}, \textit{Passage after therapy} = \textit{same}, \textit{Migration} = \textit{no}) =\\
&\quad = k_{LE} \cdot u(\textit{Life expectancy} = \textit{improved}) \;+\; k_{PAT} \cdot u(\textit{Passage after therapy} = \textit{same})\\
&\qquad + k_M \cdot u(\textit{Migration} = \textit{no})\\
&\quad = 0.75 \cdot 1 + 0.24 \cdot 0.8 + 0.01 \cdot 1 = 0.9
\end{aligned}
$$

□

A major advantage of the use of a multi-attribute utility function compared to a holistic utility function is the fact that sensitivity analysis can be performed on the separate

components of the function. As such, sensitivity analysis can reveal the effect of varying preference assumptions and can indicate which components of the utility function are responsible for inaccuracies in expected utility and in the preferred decision alternative. For the additive utility function, in a sensitivity analysis the effect of varying assessments for the component utility functions and scaling constants is investigated. If a holistic utility function is specified, it are the holistic utility assessments that will have to be varied in the sensitivity analysis.

7.5.2 Holistic utility estimates

Sensitivity analysis with respect to holistic utilities in an influence diagram amounts to varying the assessments for one or more of these utilities systematically and keeping the other ones fixed. We recall that upon transforming an influence diagram to a belief network, the utilities for the utility node become the conditional probabilities for associated chance node in the belief network. As discussed in Section 7.3, in a belief network, the relation between a probability of interest and a conditional probability pertaining to an unobserved node without observed descendants is simply linear. The relation between expected utility in an influence diagram and a holistic utility is therefore linear. We illustrate this with an example.

Example 7.5.2 We consider again the *Oesophagus* diagram in which the observations o for the ninety-two year old patient also addressed in previous examples are entered. We investigate the functional relation between expected utility and the utility $x = u(\textit{Life expectancy} = \textit{same}, \textit{Passage after} = \textit{same}, \textit{Migration} = \textit{no})$. In Figure 7.5, the effect of varying x on the expected utilities of positioning an endoprosthesis in the oesophagus, of surgery and of refraining from treatment is shown. From the figure, we see that varying the utility x, having as initial assessment the value 0.6, to smaller values can change the decision alternative with maximum expected utility from no treatment to endoprosthesis and even to surgery. This illustrates that, as the value attached by the decision maker to a situation where there is no improvement nor deterioration decreases, it becomes more desirable to treat.

In formula, the relation between expected utility and x is

$$eu(\textit{endoprosthesis} \mid o) = 0.4579 \cdot x + 0.2925$$

$$eu(\textit{surgery} \mid o) = 0.1329 \cdot x + 0.4156$$

and

$$eu(\textit{no treatment} \mid o) = x$$

□

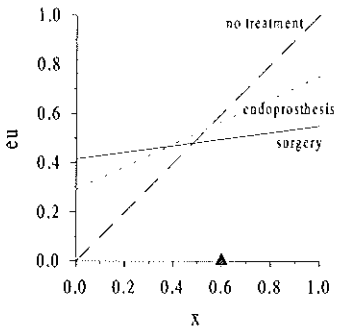


Figure 7.5: The relation between the expected utility of the decision alternatives *endoprosthesis*, *surgery*, and *no treatment* and the holistic utility under study $u(Life\ expectancy = same, Passage\ after = same, Migration = no)$.

As is the case for two conditional probabilities pertaining to an unobserved node without observed descendants, the relation between expected utility and two holistic utility assessments is bilinear. However, two different holistic utilities specify values for the predecessors of the utility node that necessarily exclude each other, that is, two different utilities refer to different outcomes that can never occur simultaneously. As such, the functional relation describing expected utility in terms of two holistic utilities cannot contain an interaction term. In general, thus, for two holistic utilities x and y, a decision alternative d_i, and observations o, we find that

$$eu(d_i \mid o) = a \cdot x + b \cdot y + c$$

We give an example of a two-way sensitivity analysis with respect to two holistic utilities for the *Oesophagus* diagram.

Example 7.5.3 We consider again the *Oesophagus* diagram in which the observations o have been entered. We are interested in the sensitivity of expected utility to the two holistic utilities $x = u(Life\ expectancy = same, Passage\ after = same, Migration = no)$ and $y = u(Life\ expectancy = same, Passage\ after = improved, Migration = no)$. In Figure 7.6, the effect of varying the assessments for the two utilities on the expected utility of positioning an endoprosthesis of surgery and of refraining from treatment is shown. The figures show that both holistic utilities have an (independent) effect on the expected utilities of the decision alternatives *endoprosthesis*, *operation*, and *no treatment*. As there is no interaction between the two holistic utilities under study, the contour lines connecting combinations of values for these utilities that give the same expected utility are parallel. Figure 7.6a shows that varying utility x has a larger effect on the expected utility of positioning an endoprosthesis than utility y. On the expected utility of surgery, the effect of varying x and y is comparable.

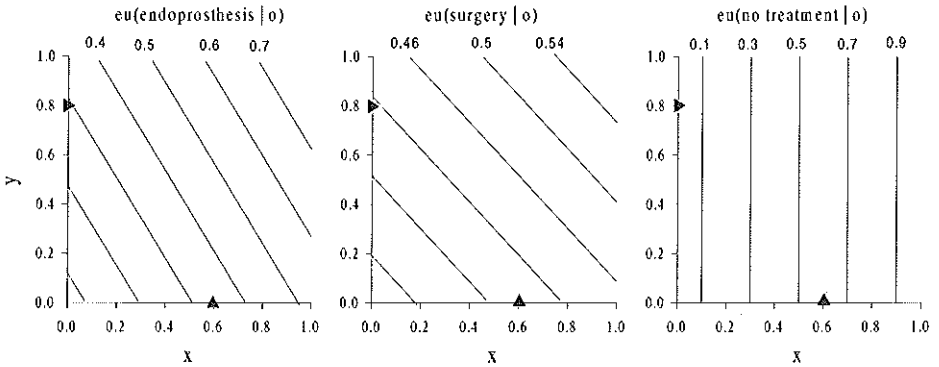


Figure 7.6: The functions relating the expected utility for the decision alternative *endoprosthesis*, *surgery*to the utilities under study $x = u(Life\ expectancy = same, Passage\ after = same, Migration = no)$ and $y = u(Life\ expectancy = same, Passage\ after = improved, Migration = no)$.

Figure 7.6c shows that the holistic utility y has no effect on the expected utility of *no treatment*; it only depends on the value of x. At the original values of the two holistic utilities under study, $x = 0.6$ and $y = 0.8$, the alternative *no treatment* has the highest expected utility. Comparing the three figures, furthermore, reveals that decreasing the assessment for x, increasing the assessment for y, or both decreasing x and increasing y simultaneously may change the preferred decision alternative to *endoprosthesis* and eventually to *surgery*. Increasing the assessments for both x and y may change the decision alternative to *endoprosthesis*, but never to *surgery*.

In formula, the effect of x and y on expected utility is,

$$eu(endoprosthesis \mid o) = 0.4579 \cdot x + 0.2833 \cdot y + 0.0658$$

$$eu(surgery \mid o) = 0.1329 \cdot x + 0.1243 \cdot y + 0.3161$$

and

$$eu(no\ treatment \mid o) = x$$

□

From the example of two-way sensitivity analysis with respect to holistic utilities, we saw that the contour lines connecting combinations of values for the utilities that give the same expected utility are equidistant. This reflects that there is no interaction between two holistic utilities, as is also expressed by the general functional form describing

expected utility in terms of two utilities. As such, a two-way sensitivity analysis with respect to utilities is simply a combination of two one-way sensitivity analyses. In essence, such a two-way sensitivity analysis contains no new information, although it does give a compact representation of the joint effect of two utilities.

Using the functional relations in one- and two-way sensitivity analysis, the *minimum deviation* in one or two holistic utilities under study that causes a change in the preferred decision alternative can be computed. For one-way sensitivity analysis with respect to a utility, the computation of the minimum deviation is completely analogous to the computation of this deviation for one conditional probability under study. The minimum deviation for one conditional probability was detailed in Proposition 7.4.4. We illustrate the computation of the minimum deviation for a utility with an example.

Example 7.5.4 Consider the one-way sensitivity analysis discussed in Example 7.5.2. From Figure 7.5, we see that varying the assessment for the holistic utility $x = u(\textit{Life expectancy} = \textit{same}, \textit{Passage after} = \textit{same}, \textit{Migration} = \textit{no})$ to smaller values causes the preferred decision alternative to change from *no treatment* to *endoprosthesis*. Using the functional relations presented in Example 7.5.2, we find that the expected utilities for both treatment alternatives are equal if

$$0.4579 \cdot x_{crit} + 0.2925 = x_{crit}$$

Solving this equation gives $x_{crit} = 0.54$. The original assessment for x is 0.6. The minimum deviation in x that causes a change in the preferred decision alternative thus is -0.06. As such, we conclude that the recommended decision alternative is quite sensitive to changes in $u(\textit{Life expectancy} = \textit{same}, \textit{Passage after} = \textit{same}, \textit{Migration} = \textit{no})$. □

The computation of the minimum deviation for two conditional probabilities that are varied in a two-way sensitivity analysis is detailed in Proposition 7.4.6. In a two-way sensitivity analysis with respect to holistic utilities, the computation of the minimum deviation is somewhat simpler. This is a result of the fact that holistic utilities cannot interact in their effect on expected utility.

Proposition 7.5.5 *Let ID be an influence diagram as defined in Definition 7.2.1. Let D be the diagram's decision node taking one of the values $\{d_1, \ldots, d_k\}$. Let O be the set of observed nodes and let o denote the corresponding observations. Let $x = u(\pi')$ and $y = u(\pi'')$ be two holistic utilities pertaining to the utility node U. Let a, b, and c be constants such that the expected utility of the preferred decision alternative d_{max} is*

$$eu(d_{max} \mid o) = a \cdot x + b \cdot y + c$$

For each $d_i \in \{d_1, \ldots, d_k\}$, $d_i \neq d_{max}$, let a_i, b_i, and c_i be constants such that the expected utility of d_i equals

$$eu(d_i \mid o) = a_i \cdot x + b_i \cdot y + c_i$$

Then, the minimum deviation required to change the preferred decision alternative from d_{max} to any other decision alternative equals Δ_{min} where

$$\|\Delta_{min}\| = \min_{\substack{d_i \in D \\ d_i \neq d_{max}}} \|\Delta_i\|$$

and where, for each d_i, we have

$$\Delta_i = \lambda \cdot \begin{pmatrix} a - a_i \\ b - b_i \end{pmatrix}$$

with

$$\lambda = -\frac{(a - a_i) \cdot x_0 + (b - b_i) \cdot y_0 + (c - c_i)}{(a - a_i)^2 + (b - b_i)^2}$$

and x_0 and y_0 are the original values for x and y.

Proof. The proof of this proposition is very similar to the proof of Proposition 7.4.6. Using the functional relations expressing expected utility in terms of the two holistic utilities under study, we find for the critical line

$$y = -\frac{(a - a_i) \cdot x + (c - c_i)}{(b - b_i)}$$

Substituting $f(x)$ in the proof of Proposition 7.4.6 with this expression gives the above-mentioned minimum deviation. □

The following example illustrates the computation of the minimum deviation in two holistic utilities that causes a change in the preferred decision alternative.

Example 7.5.6 Consider again the two-way sensitivity analysis of the *Oesophagus* diagram presented in Example 7.5.3. The two utilities under study are $x = u(\textit{Life expectancy} = \textit{same}, \textit{Passage after} = \textit{same}, \textit{Migration} = \textit{no})$ and $y = u(\textit{Life expectancy} = \textit{same}, \textit{Passage after} = \textit{improved}, \textit{Migration} = \textit{no})$. At the original assessments for x and y, that is, at $x_0 = 0.6$ and $y_0 = 0.8$, the expected utilities for *endoprosthesis*, *surgery* and *no treatment* are 0.567, 0.495, and 0.6, respectively. From Example 7.5.3, we saw that varying the assessments for x and y first changes the preferred decision alternative from *no treatment* to *endoprosthesis* and eventually to *surgery*. The minimum deviation thus concerns changing the optimal decision to *endoprosthesis*. The expected utilities of *no treatment* and *endoprosthesis* are equal if

$$0.4579 \cdot x + 0.2833 \cdot y + 0.0658 = x$$

or alternatively

$$-0.5421 \cdot x + 0.2833 \cdot y + 0.0658 = 0$$

This expression represents the critical line between the decision alternatives *no treatment* and *endoprosthesis*. Now, the minimum distance from (x_0, y_0) to this critical line is the distance perpendicular to this line. The line perpendicular to the critical line is given by

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} + \lambda \cdot \begin{pmatrix} -0.5421 \\ 0.2833 \end{pmatrix}$$

The value of λ for which the critical line and the perpendicular line cross is given by

$$-0.5421 \cdot (x_0 + \lambda \cdot -0.5421) + 0.2833 \cdot (y_0 + \lambda \cdot 0.2833) + 0.0658$$

Solving this equation for λ gives -0.088. The minimum deviation thus equals

$$\Delta_{min} = -0.088 \cdot \begin{pmatrix} -0.5421 \\ 0.2833 \end{pmatrix} = \begin{pmatrix} 0.0476 \\ 0.0249 \end{pmatrix}$$

Note that the preferred decision alternative thus is very sensitive to changes in x and y. □

In this section, we have assumed that, when varying one or two holistic utilities, the remainder of the utilities is kept fixed at their original values. This is not very realistic, as the remaining utilities concern outcomes that are closely related to the outcomes for which the utilities are varied. As such, the remaining utilities should co-vary; a higher order sensitivity analysis is required. This lies outside the scope of this chapter. However, we will briefly come back to this in the next section.

7.5.3 The additive utility function

In this section, we assume the utility function in an influence diagram to be an additive utility function. Explicit knowledge of the preference structure allows for varying, in a sensitivity analysis, not the utilities themselves but rather the various component utilities and scaling constants. As such, the results of a sensitivity analysis reveal the effect of varying assumptions on the preferences over the values of each single attribute and the importance of each attribute.

As for a holistic utility, a one-way sensitivity analysis with respect to either a component utility or a scaling constant again shows a linear relation. Although a component utility refers to several outcomes, namely all outcomes that specify the same value for the attribute to which the component utility under study pertains, it affects expected utility linearly. This is a result of the fact that, in general, an

overall utility depends linearly on a component utility under study and that expected utility, in turn, depends linearly on each overall utility. For a scaling constant the same holds. In a two-way sensitivity analysis with respect to two component utilities or two scaling constants, the functional relation is bilinear without an interaction term. As shown for two holistic utilities, two component utilities or two scaling constants cannot interact as they refer to different outcomes that cannot occur simultaneously. For a two-way sensitivity analysis with respect to both a component utility and a scaling constant, however, an interaction term is included; in the additive utility function, the overall utility of an outcome is a sum of the *product* of component utilities and scaling constants. These relation are detailed in the following proposition.

Proposition 7.5.7 *Let ID be an influence diagram as defined in Definition 7.2.1. Let D be the diagram's decision node taking one of the values $\{d_1, \ldots, d_k\}$. Let O be the set of observed nodes and let o denote the corresponding observations. Let the utility function $u(W_1, \ldots, W_q)$ over the variables $\{W_1, \ldots, W_q\} \in \pi(U)$, be additive. Let $u(W_i)$, $i = 1, \ldots, q$, be the component utility function for variable W_i, and let k_i, $i = 1, \ldots, q$, be the corresponding scaling constants. Then, for any value d_r of D, we have that*

- $eu(D = d_r \mid o) = a \cdot x + b$, *for every component utility $x = u(w_j)$ of each variable W_j, where a and b are constants related to the value w_j of W_j; a similar property holds for every scaling constant $x = k_j$ of variable W_j, $j = 1, \ldots, q$;*
- $eu(D = d_r \mid o) = a \cdot x + b \cdot y + c$, *for every pair of component utilities $x = u(w_j)$ and $y = u(w_m)$ for W_j and W_m, $j, m = 1, \ldots, q$, where a, b, and c are constants related to the values w_j of W_j and w_m of W_m; a similar property holds for every pair of scaling constants $x = k_j$ and $y = k_m$ of each pair of variables W_j and W_m;*
- $eu(D = d_r \mid o) = a \cdot xy + b \cdot x + c \cdot y + d$, *for every component utility $x = u(w_j)$ for W_j, $j = 1, \ldots, q$, and scaling constant k_m for W_m, $m = 1, \ldots, q$, where a, b, c, and d are constants related to the values w_j of W_j.*

Proof. In the following, we show that the first of the abovementioned properties holds. The remaining properties are proven analogously.

From Section 7.5.2, we know that expected utility can be expressed as a bilinear function in two holistic utilities. This bilinear function contains no interaction terms. Generalising this functional relation to *all* overall utilities for node U, we find that expected utility can be expressed as a multilinear function without interaction terms in all overall utilities. That is,

$$eu(d_r \mid o) = \sum_{w_1, \ldots, w_q} e_{w_1, \ldots, w_q} \cdot u(w_1, \ldots, w_q) \tag{7.3}$$

where $E_{W_1,\ldots,W_q}$ are constants that depend on the combination of values for all variables $W_1, \ldots, W_q$. Each utility $u(w_1, \ldots, w_q)$ is the weighed sum of the component utilities $u(w_1), \ldots, u(w_q)$, that is,

$$u(w_1, \ldots, w_q) = k_1 u(w_1) + k_2 u(w_2) + \cdots + k_q u(w_q)$$

From this sum, we have that varying component utility $u(w_j)$ linearly affects all overall utilities for which W_j takes the value w_j. That is, $u(W_1, \ldots, W_{j-1}, w_j, W_{j+1}, \ldots, W_q)$ takes the form $A \cdot x + B$ where $x = u(w_j)$, $A = k_j$, and B is built from the component utilities and scaling constants for all parents of U except W_j. Overall utilities for which node W_j takes a value other than w_j are constant with respect to variations in $u(w_j)$. From Equation 7.3, we then see that, as each utility $u(w-1, \ldots, w_q)$, in general, depends linearly on the component utility $u(w_j)$ under study, expected utility also depends linearly on $u(w_j)$. Using a similar argument, it is easily seen that the same holds for a scaling constant k_m under study. □

As noted in the previous section, a two-way sensitivity analysis without an interaction term provides no additional information to two one-way sensitivity analyses with respect to the component utilities or scaling constants under study. Except for the purpose of compactly representing information, we may thus refrain from performing such a two-way sensitivity analysis. A two-way sensitivity analysis with respect to both a component utility and a scaling constant, on the other hand, contains an interaction for the two parameters varied. In that case, a two-way sensitivity analysis is useful to reveal the joint effect of the two parameters in addition to their separate effect.

Using the functional relations that hold between expected utility and one or two component utilities and/or scaling constants, the minimum deviation in the assessments for these component utilities and scaling constants that changes the recommended decision can be easily computed. For the first functional relation in Proposition 7.5.7, the minimum deviation is computed as in Proposition 7.4.4. For the second and third functional relation, the minimum deviation is computed as in Proposition 7.5.5 and Proposition 7.4.6, respectively. In the following, we give an example of one- and two-way sensitivity analysis with respect to component utilities and scaling constants in the additive utility function of the *Oesophagus* diagram.

Example 7.5.8 Consider again the *Oesophagus* diagram, depicted in Figure 7.1. We investigate the sensitivity of expected utility in view of the observations o for the ninety-two year old patient addressed in previous examples. For the utility node of the diagram, an additive utility function has been established; this utility function is given in Example 7.5.1. We first investigate the sensitivity of expected utility for the component utility $x = u(\textit{Life expectancy} = \textit{same})$ for the parent *Life expectancy*

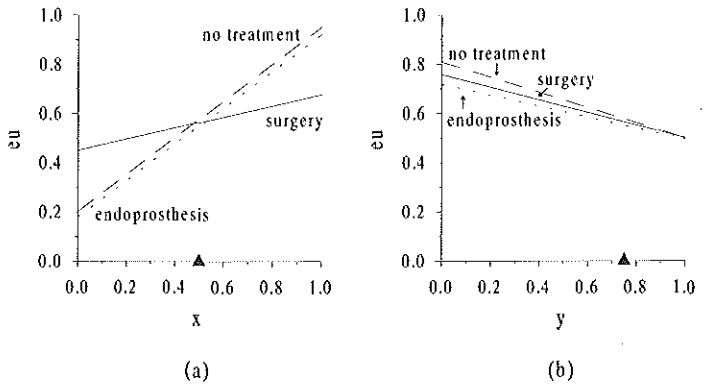


Figure 7.7: The relation between the expected utility of the decision alternatives *endoprosthesis*, *surgery* and *no treatment* and the component utility $x = u(\textit{Life expectancy} = \textit{same})$ (a) and the scaling constant $y = k_{LE}$ (b), respectively.

of *Utility*. In Figure 7.7a, the effect of varying x on the expected utilities of the decision alternatives *endoprosthesis*, *surgery* and *no treatment* is shown. At the original assessment for x, $x_0 = 0.5$, the alternative *no treatment* has the highest expected utility. However, the recommended decision alternative is very sensitive to changes in x. Varying x to only slightly smaller values changes the preferred decision alternative from *no treatment* to *surgery*. This signifies that, as the value attached by the decision maker to an unchanged life expectancy decreases slightly compared to the value attached to an increased life expectancy, the benefits of performing surgery, that is, expected years of life gained, outweigh the risks of an operation.

In formula, the relation between expected utility and x is

$$eu(\textit{endoprosthesis} \mid o) = 0.743 \cdot x + 0.179$$

$$eu(\textit{surgery} \mid o) = 0.226 \cdot x + 0.451$$

and

$$eu(\textit{no treatment} \mid o) = 0.750 \cdot x + 0.202$$

From these linear functions, the minimum deviation in x that changes the preferred decision alternative from *no treatment* to *operation* can be computed. The critical value at which the expected utilities of the two alternatives are equal is $x_{crit} = 0.48$. Thus, the minimum deviation in x is 0.02.

Now, we investigate the sensitivity of expected utility with respect to a scaling constant in the additive utility function. We take $y = k_{LE}$ for the parent *Life expectancy* as the scaling constant under study. In Figure 7.7b, the relation between expected utility

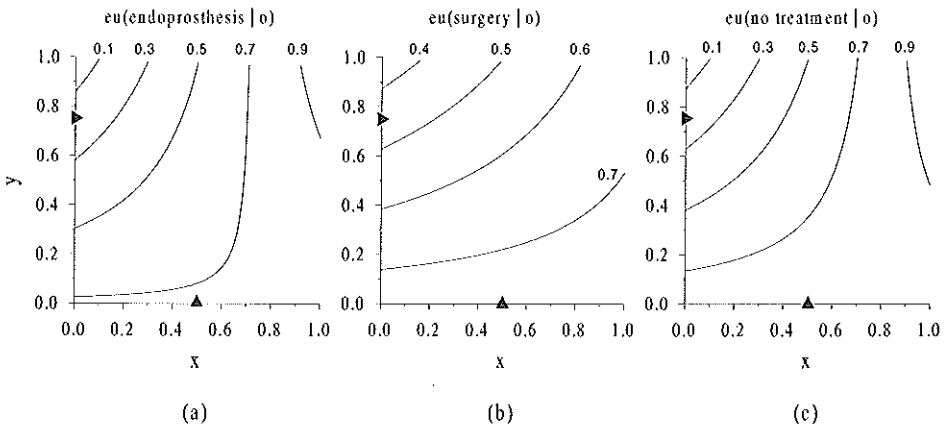


Figure 7.8: The joint effect of the component utility $x = u(Life\ expectancy\ = same)$ and the scaling constant $y = k_{LE}$ on the expected utility of the decision alternatives *endoprosthesis*, *surgery* and *no treatment*.

and y is shown. The figure shows that, whatever value y takes, the decision alternative *no treatment* has the highest expected utility. Although expected utility is affected by the importance attached to the three objectives of the decision problem, namely to improve life expectancy, to improve the passage of food and to avoid migration of the endoprosthesis, the recommended decision remains unaltered whatever values are taken for the scaling constants.

In formula, the relation between expected utility and y is

$$eu(endoprosthesis \mid o) = -0.223 \cdot x + 0.718$$

$$eu(surgery \mid o) = -0.257 \cdot x + 0.757$$

and

$$eu(no\ treatment \mid o) = -0.308 \cdot x + 0.808$$

Finally, we consider a two-way sensitivity analysis in which x and y are varied simultaneously. The effect of the joint variation of component utility x and scaling constant y on expected utility is shown in Figure 7.8. In formula, the depicted relations are given by

$$eu(endoprosthesis \mid o) = 0.99 \cdot x \cdot y - 0.718 \cdot y + 0.718$$

$$eu(surgery \mid o) = 0.301 \cdot x \cdot y - 0.408 \cdot y + 0.757$$

and

$$eu(\textit{no treatment} \mid o) = x \cdot y - 0.808 \cdot y + 0.808$$

At the orginal values for x and y, $x_0 = 0.5$ and $y_0 = 0.75$, the expected utilities of positioning an endoprosthesis in the oesophagus, of performing surgery and of not treating are 0.551, 0.564, and 0.577, respectively. At the original assessments, thus, *no treatment* is the preferred decision. From the figures, now, it is difficult to establish the effect on the preferred decision of jointly varying x and y, as the expected utilities of all three decision alternatives hardly differ. Using the functional relations, we start by computing the minimum deviation in x and y that is required to change the preferred decision from *no treatment* to *surgery*. The expected utilities for these two decision alternatives are equal if

$$0.301 \cdot x \cdot y - 0.408 \cdot y + 0.757 = x \cdot y - 0.808 \cdot y + 0.808$$

or alternatively

$$y = -\frac{-0.0514}{0.699 \cdot x - 0.40}$$

This is the expression for the critical line between *no treatment* to *surgery*. Now, we compute the minimum distance from (x_0, y_0) perpendicular to this critical line. The line through the original assessments x_0 and y_0 perpendicular to the critical line is given by

$$y = 0.75 + (x - 0.5) \cdot -\frac{(0.699 \cdot x - 0.40)^2}{0.0359}$$

This perpendicular line and the critical line intersect at the critical point (x_{crit}, y_{crit}) where x_{crit} is a solution to

$$0.75 + (x_{crit} - 0.5) \cdot -\frac{(0.699 \cdot x_{crit} - 0.40)^2}{0.0359} = -\frac{-0.0514}{0.699 \cdot x_{crit} - 0.40}$$

Solving this quartic equation gives two imaginary and two real solutions. Only one of two real solutions for x_{crit} gives a value of y_{crit} that lies between zero and one, namely $x_{crit} = 0.475$. Filling in $x_{crit} = 0.47$ in the equation for the critical line gives $y_{crit} = 0.753$. The minimum deviation then equals

$$\Delta = (x_{crit} - x_0, y_{crit} - y_0) = (0.025, 0.003)$$

The minimum distance from (x_0, y_0) to the critical line thus is $\sqrt{(0.025^2 + 0.003^2)} = 0.025$.

Computing the intersection of the critical line between *no treatment* and *endoprosthesis* and its perpendicular line through (x_0, y_0) reveals that there is no solution for x and y between zero and one. As such, the preferred decision alternative can never

change to *endoprosthesis* as a result of jointly varying the assessments for x and y. The minimum deviation thus is Δ_{min} is $(0.025, 0.003)$ and it refers to a change in recommended decision alternative from *no treatment* to *surgery*. We see that the preferred decision alternative is very sensitive to variation in the assessments for x and y. □

Recall from the previous section that, in varying holistic utilities in a sensitivity analysis, all remaining utilities are kept fixed at their original values. As this is unrealistic, higher order sensitivity analysis is needed. Varying a component utility or a scaling constant, however, does not only affect one overall utility, but simultaneously affects all utilities for the outcomes to which this component utility or scaling constant refers. As such, a one- or two-way sensitivity with respect to component utilities or scaling constants can, in fact, be seen as a higher order sensitivity analysis with respect to all affected overall utilities. As such, sensitivity analysis with respect to the components of a multi-attribute utility function is a more realistic investigation into the effect of varying preference assumptions on expected utility and on the most preferred decision alternative.

7.6 The joint effect of probabilities and utilities

Until now, we have addressed the sensitivity of an influence diagram to variations in either one or two of the diagram's conditional probabilities or in one or two utilities. We showed that expected utility can be expressed as a quotient of two functions that are linear or bilinear in one or two probabilities, respectively. For one or two utilities, we have shown that this relation is simply linear or bilinear. In addition to considering the diagram's probabilities and utilities in isolation, it may also be worthwhile investigating the joint effect of varying a conditional probability and utility simultaneously. From Proposition 7.2.4, we have that expected utility is a weighed sum of the utilities of the diagram; each utility for a specific outcome is weighed by the posterior probability of that outcome. The posterior probability of a specific outcome can, in general, be expressed as a quotient of two functions linear in a conditional probability under study. The denominator in this expression refers to the probability of the observations in the diagram and is therefore the same whatever outcome is considered. Expected utility thus is a sum of products, where each product is a quotient whose numerator is bilinear in the conditional probability under study and whose denominator is linear this conditional probability. Moreover, the denominators in each product are the same. In general, thus, the relation between expected utility of a decision alternative d_i, given observations o, and a conditional probability x and utility y under study is

$$eu(d_i \mid o) = \frac{a \cdot xy + b \cdot x + c \cdot y + d}{e \cdot x + f}$$

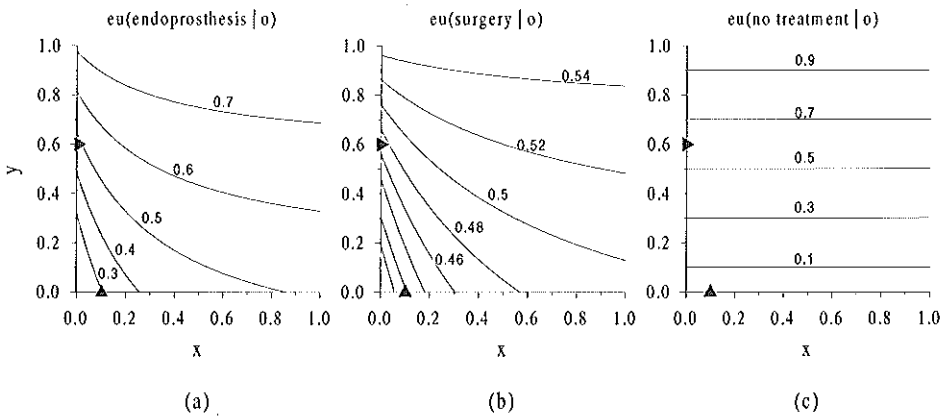


Figure 7.9: The joint effect of the conditional probability $x = p(\textit{Weight loss} = \textit{no} \mid \textit{Passage} = \textit{impaired})$ and the holistic utility $y = u(\textit{Life expectancy} = \textit{same}, \textit{Passage after} = \textit{same}, \textit{Migration} = \textit{no})$ on the expected utility of the decision alternatives *endoprosthesis*, *surgery* and *no treatment*.

Exploiting these functional relations for the expected utility of each decision alternative, in the same way as presented in Section 7.4 the minimum deviation in both the conditional probability and utility under study that leads to a change in the preferred decision alternative can be established. As extending the results from previous sections to sensitivity analysis with respect to both probabilities and utilities is straightforward, we refrain from giving technical details. We restrict the discussion to an illustration taken from the *Oesophagus* diagram.

Example 7.6.1 Consider the *Oesophagus* diagram. We again investigate the sensitivity of the recommended decision for the ninety-two year old patient who was also addressed before. The two parameters under study in the network are the conditional probability $x = p(\textit{Weight loss} = \textit{no} \mid \textit{Passage} = \textit{impaired})$ and the holistic utility $y = u(\textit{Life expectancy} = \textit{same}, \textit{Passage after} = \textit{same}, \textit{Migration} = \textit{no})$. In Figure 7.9, the joint effect of varying probability x and utility y simultaneously on the expected utilities of positioning an endoprosthesis in the oesophagus, performing surgery and not treating the patient is shown. At the original assessments for x and y, $x_0 = 0.1$ and $y_0 = 0.6$, the decision alternative *no treatment* has the highest expected utility. Varying x to larger values while keeping y fixed at its original assessment leads to a change in preferred decision alternative from *no treatment* to *endoprosthesis*. This was also seen from Example 7.4.2, where a one-way sensitivity analysis with respect to this conditional probability is discussed. Likewise, as in the one-way sensitivity analysis presented in Example 7.5.2, Figures 7.9a, b, and c show that varying y to smaller

values at $x = 0.1$ leads to a change in preferred decision from *no treatment* to *endoprosthesis* and finally even to *surgery*. Simultaneously increasing x and decreasing y affects the preferred decision in a similar way. To measure the sensitivity to joint variation of x and y, we compute the minimum deviation required to change the recommended decision from *no treatment* to *endoprosthesis* and from *no treatment* to *surgery*. We use the functional relations describing expected utility in terms of x and y;

$$eu(\textit{endoprosthesis} \mid o) = \frac{0.231 \cdot x \cdot y + 0.567 \cdot x + 0.089 \cdot y + 0.0156}{x + 0.147}$$

$$eu(\textit{surgery} \mid o) = \frac{0.0355 \cdot x \cdot y + 0.514 \cdot x + 0.029 \cdot y + 0.0517}{x + 0.147}$$

$$eu(\textit{no treatment} \mid o) = y$$

After having established the critical line between *no treatment* and *endoprosthesis* and between *no treatment* and *surgery*, the minimum deviation in x and y to reach these critical lines from the original point (x_0, y_0) is computed;

$$\Delta_1 = (x_{crit} - x_0, y_{crit} - y_0) = (0.028, 0.033)$$

for changing the recommended decision from *no treatment* to *endoprosthesis* and

$$\Delta_2 = (x_{crit} - x_0, y_{crit} - y_0) = (0.023, 0.116)$$

for changing the recommended decision from *no treatment* to *surgery*. The length of vector Δ_1 equals 0.043 and the length of vector Δ_2 equals 0.118. The minimum deviation thus is $\Delta_{min} = ()$ and it refers to a change in recommended decision alternative from *no treatment* to *endoprosthesis*. We conclude that the preferred decision alternative is quite sensitive to variation in the assessments for x and y. □

7.7 Conclusions

The assessments for the various parameters in an influence diagram, that is, the diagram's conditional probabilities and utilities, inevitably are inaccurate. These inaccuracies influence the reliability of the diagram's output. An integral part of investigating the reliability of an influence diagram is to study its sensitivity. To this end, a sensitivity analysis of the diagram is carried out. A sensitivity analysis can be performed by varying the assessments for one or more of the diagram's parameters systematically. Unfortunately, for influence diagrams of realistic size, this approach is too time-consuming. In this chapter, we have shown that sensitivity analysis of an influence diagram with one decision node can be carried out more efficiently by exploiting its properties.

We have discussed that, by qualitative considerations, conditional probabilities can be identified that cannot affect expected utility in the diagram. As such, these conditional probabilities also have no effect on the output of the diagram, that is, the optimal decision or decision with maximum expected utility. A sensitivity analysis with respect to these probabilities can thus be omitted. Furthermore, we showed that expected utility can be expressed as a simple mathematical function in the parameters varied in a sensitivity analysis. We detailed these functions for one- and two-way sensitivity analysis in which one and two parameters, respectively, are varied simultaneously. In general, expected utility relates to a single conditional probability under study via a quotient of two functions linear in this probability. For two conditional probabilities under study, expected utility can be expressed as a quotient of two bilinear functions in the two probabilities. For the utilities in an influence diagram, the functional relations are simplified. Expected utility can be expressed as a linear function in one utility under study and as a bilinear function in two utilities under study. These functional relations allows for increasing the efficiency of performing a sensitivity analysis of an influence diagram considerably; to establish the diagram's sensitivity, is suffices to establish the constants in these functions.

As the output of an influence diagram is a recommended decision, studying the effect of varying parameter assessments on the expected utility of each decision alternative in isolation does not suffice. We are interested in how much a parameter or pair of parameter assessments can be varied without inducing a change in recommended decision. In this chapter, we introduced the *minimum deviation* as a measure of the sensitivity of the diagram's output. It is the smallest change in the assessment(s) for one or two parameters under study that leads to a change in recommended decision. We showed that this minimum deviation is easily obtained from the functional relations expressing the expected utility of each decision alternative in terms of the one or two parameters under study.

In this chapter, we have limited ourselves to studying the sensitivity of the output of an influence diagram for variations in a single parameter assessment or in a pair of parameter assessments. As such, the effect of interaction between at most two parameters is covered. To obtain insight into higher order interactions between the parameters in an influence diagram, higher order sensitivity analyses are required. However, the results of a sensitivity analysis in which more than two parameters are varied simultaneously are difficult to interpret; the result are not easily represented graphically. Moreover, the more parameters are considered, the more involved the functional relations expressing expected utility in terms of these parameters are. Although an efficient method to establish the constants in these functions is given in Chapter 5, it is not clear how to use these functions to compute the minimum deviation efficiently. This issue deserves attention in future research.

Finally, in this chapter, we focused on influence diagrams with one decision node.

To obtain a more general method for sensitivity analysis, our results should be extended to influence diagrams containing several decision nodes. In an influence diagram with more than one decision node, the relation between expected utility and one or more parameters in the diagram is not given by a single continuous function but by a continuous function that is defined intervalwise. Methods for sensitivity analysis of influence diagrams with several decision nodes still have to be developed. The results presented in this chapter provide the building blocks for developing such a general method.

Part IV

Discussion

Chapter 8

General discussion

Decision-theoretic networks are a potentially valuable framework for modeling and supporting medical decision making, but methodologies for their construction and evaluation are needed. In this thesis two objectives were pursued; the development of efficient methods for sensitivity analysis of decision-theoretic networks and investigating the applicability of these methods for efficiently quantifying a decision-theoretic network. In Section 8.1, the first objective is addressed. The results achieved with respect to this objective are summarized and discussed in the light of previously published work in this field. Furthermore, with respect to this first objective, we point out some issues that to our opinion require additional research. In Section 8.2, in a similar way, the second objective is addressed. The chapter ends with some broader directions for further research in the field of decision-theoretic networks in Section 8.3.

8.1 Efficient sensitivity analysis of decision-theoretic networks

Sensitivity analysis is an important technique for investigating the robustness of a mathematical model with respect to variations in its parameter assessments. For decision-theoretic networks to be of practical use for clinical decision support, techniques for performing sensitivity analysis are essential. Performing a sensitivity analysis of a decision-theoretic network of realistic size in a straightforward manner, that is, by varying a number of network parameters (conditional probabilities and utilities) systematically over a plausible interval, is computationally not feasible. The main task in developing methods for sensitivity analysis of a decision-theoretic network, therefore, is to limit computation time. In Section 8.1.1, we review the methods that were developed in this thesis for efficient sensitivity analysis of belief networks. In Section 8.1.2, likewise, the methods developed for influence diagrams and belief networks enhanced with the threshold model for decision making, are reviewed.

8.1.1 Sensitivity analysis of belief networks

In Part II of this thesis, efficient methods for sensitivity analysis of belief networks are addressed. Chapter 4 focuses on one-way sensitivity analysis of a belief network; in a one-way sensitivity analysis of a belief network, a single probability assessment is varied at a time while keeping all the other assessments fixed. The effect of these variations on a probability of interest, computed from the network, is investigated. We showed that there are several properties of one-way sensitivity analysis of a belief network that can be exploited to perform such an analysis more efficiently than by straightforward variation of all probability assessments. Firstly, the independences represented by the qualitative part of a belief network allow for identifying conditional probabilities that cannot influence the network's probability of interest. Secondly, conditional probabilities that may affect the probability of interest are found to relate to the probability of interest via a quotient of two functions that are linear in this probability. A one-way sensitivity analysis of a belief network can thus be performed by focusing on the influential probabilities and for these, the coefficients in the function relating the probability of interest to each single network probability can be determined from a small number of network evaluations. The computational gain achieved by using these properties is considerable. In our experience, computation time is reduced by about a factor ten, compared to straightforward variation of probability assessments. The computational gain is discussed in more detail in Chapter 2.

In Chapter 5, n-way sensitivity analysis of a belief network is addressed. In an n-way sensitivity analysis, the joint effect of n, $n \geq 1$, network probabilities on the probability of interest is studied by varying the assessments for these probabilities simultaneously. The relation between a probability of interest in a belief network and n network probabilities again is a simple function. A prior probability can be expressed as a multilinear function in the n network probabilities; any posterior probability is a quotient of such multilinear functions [Castillo *et al.*, 1997b]. To assess n-way sensitivity, it suffices to determine the coefficients in these multilinear functions. An efficient and elegant method for n-way sensitivity analysis has been designed. The method builds upon the junction tree representation of a belief network for its computational architecture. Basically, the method presented is an adaptation of a standard evidence propagation algorithm. The messages sent through the junction tree are vectors of coefficients. These coefficients are processed locally per clique in the junction tree and are accumulated to yield the coefficients in the required n-way sensitivity function, describing the probability of interest in terms of the n probabilities under study. Although at present, our method has not been tested on large real-life belief networks, we believe that it has the potential to become a valuable tool for performing sensitivity analysis. The method provides a framework for sensitivity analysis, elegantly integrated in an existing propagation scheme, that is easily extended and further optimized.

With respect to one-way sensitivity analysis, in the past other researchers have undertaken efforts to develop efficient methods tailored to the belief-network framework. In Chapter 4, most of these methods are reviewed and compared to our own work. Here, we will not repeat this elaborate comparison. However, one issue deserves some more attention. In [Henrion *et al.*, 1996] and [Pradhan *et al.*, 1996], it is stated that the output of a belief network is highly *insensitive* to inaccuracies in the assessments of their conditional probabilities. Their statement is remarkable, in particular since it contradicts experience in classical decision analysis based on decision trees. Moreover, their conviction is a possible cause of the fact that currently little research is done on this topic. However, as a measure of robustness, they used the average probability of true diagnosis, resulting from varying conditional probability assessments by drawing a value from a predefined probability density function (a second-order probability distribution of the inaccurate parameter assessment). It is not the average probability of true diagnosis, though, which measures robustness, but the variation in these probabilities. Therefore, from their investigation, it is not possible to draw conclusions with respect to the robustness of their network. In our experience, in fact, inaccuracies in a belief network's conditional probabilities can have a very large effect on the outcome of the network. The number of conditional probabilities that is highly influential, however, is limited. In Chapter 3, we showed that some eighty percent of the probabilities in the VSD network have little or no effect on the probability of interest. Among the influential probabilities, on the other hand, the gradient of the probability of interest reached a maximum of five. So while, on average, the effect may be small, inaccuracies in the crucial network probabilities can have serious consequences. Similar observations were done for the Oesophagus network [Van der Gaag *et al.*, 1999]. They found that some fifty percent of the probabilities were influential, with a maximum gradient around seven. From their investigation, Pradhan *et al.* concluded that in constructing a belief network, it is important to establish the structure of the network well; the conditional probabilities are only of minor importance. From our experience with the VSD network, we indeed found that the structure of the network is crucial. It determines, for a large part, *which* conditional probabilities are influential. We would thus like to add to the conclusion of Pradhan *et al.* that, in addition to assessing the structure of the network carefully, it is important to accurately assess the most influential probabilities. To our opinion, therefore, sensitivity analysis should be given a prominent role in building belief networks.

In Chapter 2, we analysed the computational gain achieved by exploiting the properties of one-way sensitivity analysis of belief networks, instead of simply varying probability assessments. We found, as also stated above, that an increase in efficiency in the order of a factor ten is obtained. Our method was implemented in Allegro Common Lisp as an extension of Ideal, and network evaluation was carried out using Jensen clustering [Jensen, 1996]. Using this implementation, on a Sun Sparc Ultra 5,

360 MHz, a full one-way sensitivity analysis of the VSD network, with 38 nodes and 738 probabilities, for a specific set of observations required between seven and thirteen minutes. For the Oesophagus network, with 40 nodes and almost 1000 probabilities, between five and twenty minutes were needed for a full one-way sensitivity analysis. Unlike straightforward sensitivity analysis, which may take hours of computation time, our method thus is of practical use, even as part of a belief network application running in a medical setting.

A even more efficient method for one-way sensitivity analysis of belief networks has been developed by now. Building on our idea of exploiting the properties of sensitivity analysis of belief networks, U. Kjærulff and L.C. van der Gaag developed a method for sensitivity analysis that uses the computational architecture of the junction tree [Kjærulff & Van der Gaag, 2000]. Basically, the idea is to compute the coefficients in the functional relation between the probability of interest and each single network probability in the clique in the junction tree in which the conditional probability under study resides. To our knowledge, this is the most efficient method for performing one-way sensitivity analysis of a belief network that is currently available. As one-way sensitivity analysis is now well feasible for belief networks, further research efforts should be directed towards other techniques useful in building belief networks, rather than on developing even more efficient methods for one-way sensitivity analysis. This issue is addressed in Section 8.3.

With respect to n-way sensitivity analysis of belief networks, an extensive comparison with related research is given in Chapter 5. Here, we briefly review the work by E. Castillo *et al.*, as it is closely related to our work. Using symbolic propagation to perform sensitivity analysis, E. Castillo *et al.* recognized the mathematical form of the function expressing a probability of interest in a belief network in terms of the conditional probabilities in the network [Castillo *et al.*, 1995]; the function expressing a probability of interest in terms of n parameters, called an *n-way sensitivity function*, is a quotient of two multilinear functions. Note that, in Chapter 4, we analytically derived this functional form for one-way sensitivity analysis of a belief network. Exploiting this knowledge, in [Castillo *et al.*, 1997b] an algorithm for computing the coefficients in an n-way sensitivity function is proposed. Their method is nearly identical to our method for one-way sensitivity analysis. First, the structure of the belief network is used to identify those conditional probabilities that may affect the probability of interest and those that cannot. Subsequently, for the influential conditional probabilities, they propose computing the coefficients by assuming different combinations of values for these probabilities. One different combination of values for each coefficient is required. For each combination of values the belief network is evaluated using any standard propagation method. As a result, a system of linear equations is obtained which is solved to give the required coefficients.

The method we present in Chapter 5 builds on the ideas put forward by Castillo *et*

al.. As discussed in Chapter 5, in its most simple form, our method performs equally well as the optimal version of the method presented by E. Castillo *et al.*. The substantial advantage of our method, however, is that it is integrated in an existing propagation scheme. It therefore provides a framework for sensitivity analysis that can be easily extended and optimized. Moreover, any optimization of the propagation algorithm for junction trees automatically benefits the efficiency of our method for n-way sensitivity analysis. In Chapter 5, we briefly presented various possibilities for optimizing our method. In the future, these possibilities should be investigated in more detail and integrated in our method. As the computational complexity of n-way sensitivity analysis increases exponentially with the number of conditional probabilities considered in the analysis, such optimizations are particularly useful when larger number of conditional probabilities are considered simultaneously.

Finally, until now, our method for n-way sensitivity analysis has been tested on small artificial belief networks only. Also, only small numbers of probabilities under study were considered. Therefore, it is too early to conclude upon the strengths and weaknesses of our method. In the near future, sensitivity analyses of the VSD network and the Oesophagus network will give more insight into the characteristics of our method.

8.1.2 Sensitivity analysis of influence diagrams

In Part III, the methods presented for sensitivity analysis of a belief network are extended to include decision making under uncertainty. In Chapter 6, sensitivity analysis of a belief network in view of the threshold model for decision making is addressed. Building on the properties presented in Chapter 4, a method for sensitivity analysis is presented that provides for the computation of bounds between which a network's parameter assessments can be varied without inducing a change in recommended decision. Basically, the method amounts to the following. Using, for example, expert opinions, various threshold probabilities of disease are established that divide the probability scale into intervals where a specific decision alternative is optimal. For example, the treatment threshold probability of disease is the probability at which the expert is indifferent between treating and not treating. For higher probabilities of disease, the patient is given treatment and for lower probabilities, treatment is withheld. Now, using the method proposed in Chapter 4, the functional relation between the probability of disease and a specific conditional probability under study is established. Combining this functional relation with the threshold probabilities of disease allows for the computation of the *minimum deviation* in the assessment(s) under study that does induce a change in the currently recommended decision. For smaller deviations, the probability of disease computed from the network does not pass an established threshold probability; the recommended decision thus remains the same as at the original value

of the probability under study. For larger deviations, the preferred decision changes. The minimum deviation, which is easily computed using this method, is a measure of the robustness of the belief network in view of the threshold model for decision making.

In Chapter 7, sensitivity analysis of influence diagrams is considered. In a sensitivity analysis of an influence diagram, the parameters to be considered consist of the conditional probabilities pertaining to the stochastic variables in the influence diagram and the utilities pertaining to the utility node. We showed that, for all parameters in an influence diagram, similar properties hold as for the conditional probabilities in a belief network. In an influence diagram, expected utility can be expressed as a quotient of two functions linear in a single parameter under study. As for belief networks, these properties provide for efficiently performing sensitivity analysis of an influence diagram. Using these functional relations, furthermore, the *minimum deviation* in a single probability assessment (or set of assessments) that would induce a change in recommended decision can be computed. To that end, the functional relation between each possible decision alternative and the probability under study is established. These relations are used to compute at which deviation of the probability under study the expected utility of the currently preferred decision alternative equals the expected utility of any other alternative. If the probability under study is varied more than the minimum of these deviations, the preferred decision changes. Again this minimum deviation provides a measure of the robustness of the influence diagram. Chapter 7 presents, to our knowledge, the first investigation into efficient methods for sensitivity analysis of influence diagrams. Until recently, no such methods were available.

Sensitivity analysis of a belief networks enhanced with the threshold model for decision making closely resembles sensitivity analysis of influence diagrams. Both provide for the computation of bounds between which a parameter assessment can be varied without inducing a change in recommended decision; the measure of robustness is the minimum deviation. In a belief network enhanced with the threshold model for decision making, however, sensitivity analysis is performed with respect to the conditional probabilities in the belief network only. Inaccuracies in the decision model cannot be investigated as they are implicitly integrated in the threshold probabilities of disease. This concerns the risks and benefits of the decision alternatives considered, the utilities by which they are weighed, and the typical risk behaviour of the assessor of the threshold probabilities. In an influence diagram, these elements of the decision model are explicitly represented. Sensitivity analysis of an influence diagram, therefore, allows for investigating the effect on the preferred decision of inaccuracies in all these elements independently.

In the introduction of this thesis, the similarities and differences between influence diagram and decision trees were discussed. We noted that any influence diagram can be translated into a decision tree and vice versa. As such, the functional relation between expected utility and a set of parameters under study in an influence diagram holds

equivalently in a decision tree. To our knowledge, these functional relations in decision trees have never been used in the way we propose for influence diagrams. Clearly, this is a result of the difference in scale between the two representations. Whereas sensitivity analysis of an influence diagram is nearly impossible without efficient methods, sensitivity analysis of decision trees can be carried out by straightforward variation of probabilities and utilities. We would like to stress, however, that the type of relations we found are not unique to influence diagrams.

In this thesis, the study of sensitivity analysis of influence diagrams has been limited to exploiting the functional relations between expected utility and the diagram's parameters in order to compute the minimum deviation for one or two parameters. To compute the functional relations in an influence diagram for higher order sensitivity analyses, the method for n-way sensitivity analysis developed for belief networks can be applied. To that end, the influence diagram is then first transformed to a belief network [Cooper, 1988]. It has not been investigated, however, how to determine the minimum deviation in higher order sensitivity analyses efficiently. The problem amounts to determining the minimum distance of a point in n-dimensional space to a surface in this space. For this problem, probably, a quite different approach from the one presented in Chapter 5 has to be taken. Furthermore, in this thesis, we focused on influence diagrams with one decision node. To obtain a more general method for sensitivity analysis, our results should be extended to influence diagrams containing several decision nodes. In an influence diagram with more than one decision node, the relation between expected utility and one or more parameters in the diagram is not given by a single continuous function but by a continuous function that is defined intervalwise. Methods for sensitivity analysis of influence diagrams with several decision nodes still have to be developed.

8.2 Using sensitivity analysis in building a belief network

In this section, the second objective of this thesis – using sensitivity analysis for efficiently quantifying a decision-theoretic network – is addressed. In Chapter 2, for this purpose, an iterative procedure is suggested. Although the procedure focuses on efficient quantification of a belief network, it applies equally well to influence diagrams. The procedure sets out with initial, probably highly inaccurate, assessments for all probabilities in a belief network under construction. These assessments are, for example, obtained from experts. Subsequently, a sensitivity analysis of the belief network is performed to obtain insight into the possible effects of the inaccuracies involved. Some probabilities are likely to have considerable impact while others will hardly reveal any influence. For the less influential probabilities, the initial assessments may suffice. For

the more influential one, refinement may be required. To that end, more elaborate elicitation techniques to obtain unbiased, sufficiently accurate assessments from experts or statistical data can be employed. We envisage a procedure of iteratively performing sensitivity analyses and refining probabilities until satisfactory behaviour of a belief network in the making is obtained, until the costs of further quantification efforts outweigh the benefits of higher accuracy, or until higher accuracy can no longer be attained due to lack of knowledge. In Chapter 3, an empirical investigation into the viability of such a procedure was presented. The results of this investigation will be reviewed in Section 8.2.1. Subsequently, in Section 8.2.2, it is argued that sensitivity analysis can be also used for detecting errors in the qualitative part of a belief network. This idea originates from experience with performing sensitivity analyses of a belief network developed for the diagnosis of *Wilson's disease*. Finally, in Section 8.2.3 we propose to apply the above-mentioned uses of sensitivity analysis to handle the problems encountered in building a belief network for *ovulatory disorders*. The topics discussed in Sections 8.2.2 and 8.2.3 were not covered before in this thesis.

8.2.1 Efficiently quantifying a belief network

In Chapter 3, the results of a preliminary evaluation of the quantification procedure proposed in Chapter 2 are presented. As a case study, the VSD network was chosen. For the quantification of the VSD network, subjective probability assessments were obtained. The main objective was to establish whether it is possible to reduce the number of probabilities that have to be estimated by experts by using the above-mentioned procedure. To that end, three network quantifications with different levels of 'informedness' were constructed. The term *informedness* refers to the level of expertise of the person supplying the assessments. Two poorly informed quantifications were improved by replacing the most influential probabilities with the corresponding assessments from the well-informed (expert) quantification. The influential probabilities were found by performing one-way sensitivity analyses. The results of the replacements were investigated by comparing network predictions.

The results of our investigation suggest that a two-step procedure of only improving influential probability assessments leads to satisfactory results: partially refined, poorly informed quantifications give predictions that are comparable to a well-informed quantification. Ideally the two-step procedure should be repeated several times. Instead of making a final network quantification on the basis of a single sensitivity analysis, it is probably better to have a few alternating steps of sensitivity analyses and improvements of the quantification. This approach takes into consideration that by each refinement, the set of highly influential probabilities may change.

Furthermore, the accuracy of probability assessments can be involved in the procedure. As a measure of accuracy a confidence interval can be assessed, either by

an expert or from data. To allow for incorporating confidence intervals estimated by experts, the Bayesian interpretation of a confidence interval is taken, that is, the $x\%$ confidence interval of a network probability is the interval within which the real value of probability is supposed to lie with a certainty of $x\%$. These confidence intervals are taken into account in a sensitivity analysis by using the variation of a sensitivity analysis' curve over a confidence interval instead of the gradient at the original assessment as a measure of the probability's influence. Then, probability assessments with high expected accuracy (i.e., having small confidence intervals) will only be reconsidered when the sensitivity analysis' curve is extremely steep over the confidence interval.

We would like to note that in the investigation of Chapter 3 a rather artificial situation was created, in which the best quantification was the one provided by an expert. For real applications, the usage of objective statistical sources is likely to be indispensable to obtain a sufficiently reliable network quantification. Moreover, in the medical field it is often possible to collect databases of considerable size. Note, therefore, that the proposed procedure is not limited to the use of subjective expert assessments.

Currently, a second evaluation of the procedure is being carried out. It concerns the quantification of an influence diagram for treatment planning in patients with oesophageal carcinoma [Van der Gaag *et al.*, 1999]. The influence diagram is developed at the department of Computer Science of Utrecht University in cooperation with the department of Internal Medicine of the Netherlands Cancer Institute. Since the quantification procedure has not yet been completed, it is too early to draw decisive conclusions. However, the results so far are encouraging. In particular, the builders of the model designed a new method for rapid and easy elicitation of the initial probability assessments with which the iterative procedure sets out [Renooij & Witteman, 1999, Van der Gaag *et al.*, 1999]. The method presents the conditional probabilities as fragments of text and provides a scale for marking assessments with both numerical and verbal anchors. Furthermore, the authors are presently working on a method for combining expert assessments in order to refine the network on the basis of the results of the sensitivity analyses. Finally, from their experience with the construction and quantification of an influence diagram for oesophageal carcinoma, L.C. van der Gaag *et al.* found that obtaining the required utilities for an influence diagram is a task at least as difficult as obtaining all probability assessments. It seems therefore worthwhile to focus future research efforts into the development of an efficient elicitation method for utilities.

8.2.2 Detecting modeling errors

The results of a sensitivity analysis may reveal errors in both the qualitative and quantitative part of a belief network. This was concluded from elaborate sensitivity

analyses of a belief network for the diagnosis of Wilson's disease. Wilson's disease is a recessively inherited disorder of the liver. It is associated with low levels of serum caeruloplasmin and progressive copper accumulation in the liver. Eventually, the capacity of the hepatocytes to store copper is exceeded and release into the blood and uptake in extrahepatic sites occurs, causing extrahepatic disease. In adult life, Wilson's disease almost invariably presents with neurological manifestations, such as personality change and, if not treated, dementia. In patients who are diagnosed early, treatment can usually improve the liver function and stop further progression of the disease. Patients presenting with serious symptoms, such as hepatic failure or cirrhosis have a poor prognosis. A belief network for the diagnosis of Wilson's disease was derived from a rule-based expert system for the diagnosis of disorders of the liver and the biliary tract [Korver & Lucas, 1993].

Extensive sensitivity analyses of the network, for various different patients, revealed some typical features of the network. Most striking was the fact that for cases that should be easy to diagnose, in particular, patients presenting with *Kayser-Fleischer rings*, varying each probability in the network from zero to one could not raise the posterior probability of Wilson's disease much. The maximum value attained for the posterior probability remained too low considering the strength of the evidence. Acknowledging this phenomenon, the builders of the model dramatically increased the prevalence of Wilson's disease in the network to one in two hundred, which is very high even for an internal medicine clinic. Lower prevalences would result in even lower posterior probabilities for patients presenting with clear evidence of Wilson's disease.

These findings suggest that the independence relations between the relevant variables are not correctly modeled. From careful inspection of the qualitative part of the network, we deduced that a possible reason for the observed phenomenon could be missing arcs between variables in the network. For example, the diagnostic strength of the observation of Kayser-Fleischer rings in the network would increase significantly if there is more than only one mechanism related to Wilson's disease that can cause this observation. Currently, in the network, Kayser-Fleischer rings is modeled to be related to Wilson's disease through elevated levels of copper concentration only. Adding, for example, a direct causal link between the variable Wilson's disease and Kayser-Fleischer rings, circumventing the variable hepatic copper, increases the diagnostic value of observing these rings.

Another reason for the low posterior probabilities found may be missing variables representing other disorders that may cause similar symptoms as Wilson's disease. In particular, adding variables that represent causes of high hepatic copper, other than Wilson's disease, and the possible effects of these causes can increase the impact of observing Kayser-Fleischer rings. Having observed Kayser-Fleischer rings then induces a dependence between the likelihood of having Wilson's disease versus these other causes. If additional evidence renders the other causes unlikely, Kayser-Fleischer rings

become a strong indication for Wilson's disease. As other causes of increased copper levels, it could be considered to add variables for *chronic cholestasis* and *primary cirrhosis* to the network.

A second characteristic of the network that was revealed by the sensitivity analyses is the high number of extreme assessments for many probabilities in the network. With an extreme assessment, a zero or one is meant. As a result of these extreme assessments a large percentage of the probabilities cannot influence the posterior probability of having Wilson's disease. In fact, what happens is that in the presence of observations for leaf variables, these extreme assessments cause informational blocking of intermediate variables. Since most of the variables in the network for Wilson's disease that contain these extreme assessments are naturally continuous variables, it is unlikely that a chosen threshold dividing the range of values into low and high should produce such extreme probabilities. Using more moderate estimates in the network might well result in more acceptable posterior probabilities.

From this experimental investigation, we conclude that sensitivity analysis not only provides an aid in the efficient quantification of a belief network but also in the structuring of the network. We gave an example of the contribution of sensitivity analysis' results in detecting missing arcs or variables and in revealing possible errors in the conditional probabilities of the network. We assume that there are more characteristics of a belief network that can be studied from a sensitivity analysis' results. This requires further investigation. It would be valuable to develop a kind of protocol that structures the various characteristics that can be discovered by performing a sensitivity analysis and that describes how to interpret the results found. Note, that such a protocol would not just apply to belief networks, but to influence diagrams as well.

8.2.3 Using sensitivity analysis to improve a belief network for ovulatory disorders

In Sections 8.2.1 and 8.2.2, it was argued that sensitivity analysis can be used to facilitate the quantification of a decision-theoretic network and to detect modeling errors. In the following, we will discuss how these two usages of sensitivity analysis can be exploited to improve an initial belief network for *ovulatory disorders*. An ovulatory disorder is a disturbance in the menstrual cycle of a woman. It refers to the absence of a cycle, called *amenorrhoea*, or a cycle of 35 days or more, called *oligomenorrhoea*. One subclass of cycle disturbances, referred to as WHOII (according to the criteria of the World Health Organization) is characterized by normal concentrations of the hormones FSH (follicle stimulating hormone) and E_2 (oestrogen). In this group, typical findings are a high concentration of androgens and LH (lutinizing hormone), and large ovaries containing many immature follicles. In order to provide optimal treatment of WHOII patients wishing to have children, it

is important to know the chance of unassisted conception and the expected effect of various treatments. For this purpose, various statistical prediction models have been developed [Fauser *et al.*, 1992, Fauser & Van Heusden, 1997, Imani *et al.*, 1998, Imani *et al.*, 2000, Collins *et al.*, 1995, Snick *et al.*, 1997].

Recognizing that the factors describing the functioning of the hormone system are strongly causally related, it was suggested that the problem domain of ovulatory disorders can be modeled with a belief network. The expected advantage of using a belief network is that prediction is not limited to only one outcome variable of interest at a time. In one belief network, various outcomes of interest can be combined, for example, unassisted conception, ovulation, assisted conception, life birth. A pilot study into the feasibility of building such a belief network was undertaken at the Department of Public Health (Center for Clinical Decision Sciences) of the Erasmus University in Rotterdam in cooperation with the Department of Gynaecology from the Dijkzigt Hospital in Rotterdam. Although currently only preliminary results are available, the modeling efforts undertaken yield useful insights into typical problems encountered.

Initially, the problem of modeling ovulatory disorders in WHOII patients was tackled by graphically representing hormonal changes. Indeed, the interaction between various endocrine features can be described in terms of causal influences. However, modeling hormonal changes in the menstrual cycle, which consists of numerous stimularory and feedback loops, resulted in directed cycles in the network's graph. This lead to the recognition that the problem should be viewed from a different perspective. Instead of modeling the functioning of the hormone system, its dysfunctioning should be the basis for the belief network. For prediction and treatment planning, it is relevant how variables are related in this dysfunctioning hormone system. In the field of artificial intelligence, in fact, this question whether to model proper functioning or dysfunctioning is an ongoing discussion [De Kleer & Williams, 1989, Poole, 1989]. Although the decision to model the dysfunctioning of the hormone system indeed provided progress in the elicitation of the belief network's qualitative part, it also brought a second problem to the fore: a lack of detailed and agreed-upon knowledge, as is the case for many medical problems.

The dysfunctioning of the hormone system is, as yet, only partly understood. Consider for example the *polycystic ovary syndrome* (PCOS), which is by far the largest group within WHOII. As the name indicates, it concerns a syndrome, that is, the clinical picture is described in terms of the clinical features through which it manifests itself, rather than in terms of the underlying disease process. The reason is that there are multiple diseases that may cause the syndrome. Possible causes are mainly found in the ovaries, but also in the brain. This lack of detailed biological knowledge makes it a difficult task to construct a reliable belief network.

At this point, we have the following steps in mind to improve upon the current version of the network. Firstly, the network should be quantified with rough, probably

inaccurate, probability assessments given by experts in the field. For this purpose, the method described in [Renooij & Witteman, 1999] can be employed. Subsequently, with the method described in Chapter 4, the set of conditional probabilities that are uninfluential given a specific patient profile can be established. Feedback of these uninfluential probabilities to the expert will hopefully result in the identification of errors in the qualitative part of the network; probabilities indicated to be uninfluential while, in fact, the experts expect them to be important can be used to identify missing arcs. After adding these missing arcs to the network, a sensitivity analysis can be performed. We suggest to start with the simplest type of sensitivity analysis, a one-way sensitivity analysis, as was also proposed in Chapter 2. The results of this analysis will indicate the most influential probabilities. These again should match the expectations of the experts and can be used, as was discussed in Section 8.2.2, to identify modeling errors in the qualitative part as well as the quantitative part. At a later stage, probably, higher-order sensitivity analysis can be carried out, to identify interactions between variables. Finally, since a considerable amount of data on ovulatory disorders is available, these data can be used to refine influential probabilities and/or to validate the predictions of the belief network. As discussed in Chapter 2, the steps described above can be carried out iteratively, to gradually improve the network.

8.3 Topics for future research

Until now, relatively little research has been done into the development of techniques that support the construction and evaluation of decision-theoretic networks. In clinical decision analysis, a large body of such techniques are available to build more classical decision models, such as decision trees. To support the use of decision-theoretic networks, the concepts and techniques of the field of decision analysis should therefore be translated to the framework of decision-theoretic networks. In this thesis, we have focused attention on the technique of sensitivity analysis. Efficient methods tailored to the framework of decision-theoretic networks have been developed and it has been investigated how sensitivity analysis can support the construction, and in particular the quantification, of a network. We believe that the research in this thesis provides both new and useful methodologies. However, the work only is a first step in making decision-theoretic networks a practical framework for medical decision support. Other concepts deserving attention are, for example, uncertainty analysis, the expected value of perfect information and the expected value of perfect control.

Uncertainty analysis amounts to varying the assessments for *all* parameters of the network's quantitative part simultaneously. To this end, for each parameter, values are drawn from some continuous probability distribution [Morgan & Henrion, 1990]. The result of an uncertainty analysis is a continuous, second-order probability distri-

bution of the network's output, that is, a probability of interest or expected utility. It reveals the overall uncertainty in the output. Apart from this estimate of overall uncertainty in the output, the results of an uncertainty analysis also allow for establishing the percentagewise contribution of the uncertainty (assumed to take the specific form of a chosen probability distribution) in the separate parameters to the overall uncertainty. Uncertainty analysis of a decision-theoretic network is very time-consuming. Usually, tens of thousands of values are drawn for each uncertain parameter; the output thus also has to be computed from the network tens of thousands of times [Pradhan *et al.*, 1996, Henrion *et al.*, 1996]. It would therefore be valuable to have efficient methods for uncertainty analysis of decision-theoretic networks. Possibly, the properties of sensitivity analysis of networks that are presented in this thesis provide a useful basis. It might be possible, for example, to find mathematical relations expressing uncertainty in the output in terms of the uncertainties in the network's parameters. In that way, the number of network computations could be reduced to just those that are required to establish the values of the coefficients describing the continuous, second-order distribution of the output from the coefficients in the distributions of the sampled network parameters.

An approach for investigating uncertainty in the network's output that is related to uncertainty analysis is presented in [Spiegelhalter, 1989]. As in our method, D.J. Spiegelhalter proposes to construct an auxiliary graph from the graph of a belief network by adding an auxiliary parent to every node. The auxiliary parent now captures the uncertainty in the conditional probabilities of its child by discrete second-order distributions. Using standard propagation algorithms, the effects of the specified uncertainties on a probability of interest are readily computed. In fact, the result of this analysis is a discrete, second-order probability distribution of the probability of interest in the network. As the way in which uncertainty is represented and quantified in this method is equivalent to the way influential relationships between stochastic variables are represented and quantified in a belief network, the properties we describe in Chapter 4 apply equally well to these uncertainty nodes. As such, for this manner of investigating uncertainty, the functional relations holding in a belief network between a probability of interest and network probabilities can be exploited to develop an efficient method.

With respect to the concepts of expected value of perfect information and expected value of perfect control, currently, no methods are available. In [Howard, 1990, Matheson, 1990] elaborate descriptions are given of the practical value of these concepts. Associated algorithms tailored to the framework of decision-theoretic networks, however, are still to be developed.

Apart from integrating existing decision-analytic techniques in the framework of decision-theoretic networks, it is important to investigate how statistical data can be used in constructing and quantifying a decision-theoretic network. Until now, the

focus has been on exploiting expert knowledge in building networks. However, in particular in the medical field, large collections of valuable statistical data are available. This data can be exploited to build the structure of the network [Buntine, 1996, Cooper & Herskovits, 1992] or to quantify a network. Using statistical data to quantify a decision-theoretic network has, until now, been proven to be difficult. Usually, essential information is missing to estimate precisely those probabilities that are required in the network. It would be useful to develop methods that allow to incorporate the data without requiring a full specification of the probability distribution; that is, the data are used to formulate constraints on the joint probability distribution represented by the network. This issue has been addressed in [Druzdzel & Van der Gaag, 1995]. Furthermore, it may be possible to use statistical data in parts of a decision-theoretic network for which expert knowledge is lacking. For example, if the structure of a part of the network is difficult to establish because the influential relations between the variables concerned are unclear, but statistical data are available for those variables, then the network can be enhanced with a (logistic) regression model for that specific, problematic part. Looking back at the problems encountered in building the structure of a belief network for ovulatory disorders, this idea may provide help.

Finally, practical experience with building decision-theoretic networks for clinically relevant problems is very important. Firstly, carefully evaluating experience will guide research efforts as it reveals problematic issues for which techniques are still lacking. Secondly, only such experience will give detailed insight into the added value of decision-theoretic networks in clinical practise. It would be interesting to evaluate how decision-theoretic networks perform in clinical practise compared to, for example, decision trees, decision rules or logistic regression models. Another interesting issue is how clinicians experience the use of decision-theoretic networks for medical decision support. In my opinion, especially these latter research questions concerning the applicability of decision-theoretic networks constitute the challenge in future research in this field.

Bibliography

[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.

[Andreassen *et al.*, 1991] S. Andreassen, R. Havorka, J. Benn, K.G. Olesen, and E.R. Carson (1991). A model-based approach to insulin adjustment. *Proceedings of the Third Conference on Artificial Intelligence in Medicine*, Lecture Notes in Medical Informatics 44, M. Stefanelli, A. Hasman, M. Fieschi, J. Talmon (Eds.), pp. 239 – 248.

[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. *Proceedings of the Second Conference on Artificial Intelligence in Medicine*, J. Hunter, J. Cookson, and J. Wyatt (Eds.), Springer-Verlag, Berlin, pp. 247 – 256.

[Breslow & Day, 1987] N.E. Breslow and N.E. Day (1987). *Statistical Methods in Cancer Research, Volume II, the Design and Analysis of Cohort Studies.* Oxford University Press, Oxford, UK.

[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.

[Buntine, 1996] W. Buntine (1996). A guide to the literature on learning graphical models. *IEEE Transactions on Knowledge and Data Engineering*, Vol. 8, pp. 195 – 210.

[Castillo *et al.*, 1995] E. Castillo, J.M. Gutiérrez, and A.S. Hadi (1997). Parametric Structure of Probabilities in Bayesian Networks. *Lecture Notes in Artificial Intelligence: Symbolic and Quantitative Approaches to Reasoning and Uncertainty*, C. Froidevaux and J. Kohlas (Eds.), Springer-Verlag, New York, Vol. 946, pp. 89 – 98.

[Castillo *et al.*, 1997a] E. Castillo, J.M. Gutiérrez, A.S. Hadi, and C. Solares (1997). Symbolic propagation and sensitivity analysis in Gaussian Bayesian networks with application to damage assessment. *Artificial Intelligence in Engineering*, Vol. 11, pp. 173 –181.

[Castillo *et al.*, 1997b] 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.

[Chang & Fung, 1995] K. Chang and R. Fung (1995). Symbolic Probabilistic Inference with Both Discrete and Continuous Variables. *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 25, pp. 910 – 916.

[Chernoff & Moses, 1959] H. Chernoff and L. Moses (1959). *Elementary Decision Theory*. John Wiley & Sons, New York.

[Collins *et al.*, 1995] J.A. Collins, E.A. Burrows, and A.R. Willan (1995). The prognosis for live birth among untreated infertile couples. *Gynecology-endocrinology*, Vol. 64, No. 1, pp. 22 – 28.

[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.

[Cooper, 1988] G.F. Cooper (1988). A Method for Using Belief Networks as Influence Diagrams. *The Fourth Workshop on Uncertainty in Artificial Intelligence*, pp. 55 – 63.

[Cooper & Herskovits, 1992] G.F. Cooper and E. Herskovits (1992). A Bayesian method for the induction of probabilistic networks from data. *Machine Learning*, Vol. 9, pp. 309 – 347.

[Coupé *et al.*, 2000] V.M.H. Coupé, J.D.F. Habbema, N. Peek, and J. Ottenkamp (2000) Structuring and quantifying a belief network for VSD, in preparation.

[Coupé & Van der Gaag, 1998] V.M.H. Coupé and L.C. van der Gaag (1998). Practicable sensitivity analysis of Bayesian belief networks. *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, M. Hušková, P. Lachout, and J.A. Víšek (Eds.), pp. 81 – 86.

[De Kleer & Williams, 1989] J. de Kleer and B.C. Williams (1989). Diagnosis with behavioral modes. *Proceedings of the Eleventh International Joint Conferences on Artificial Intelligence*, Detroit, MI, pp. 1324 – 1330.

[Dippel *et al.*, 1992] D.W.J. Dippel, J.W.M. ter Berg, and J.D.F. Habbema (1992). Screening for unruptured familial intracranial aneurysms. A decision analysis. *Acta Neurologica Scandinavica*, Vol. 86, pp. 381 – 389.

[Doubilet *et al.*, 1985] P. Doubilet, C.B. Begg, M.C. Weinstein, P. Braun, and B.J. McNeil (1985). Probabilistic sensitivity analysis using monte carlo simulation: a practical approach. *Medical Decision Making*, Vol. 5, pp. 157 – 177.

[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.

[Fauser *et al.*, 1992] B.C.J.M. Fauser, T.D. Pache, W.C.J. Hop, F.H. de Jong, and K.D. Dahl (1992). The significance of a single serum LH measurement in woman with cycle disturbances: discrepancies between immunoreactive and bioactive hormone estimates. *Clinical Endocrinology*, Vol. 37, pp. 445 – 452.

[Fauser & Van Heusden, 1997] B.C.J.M. Fauser and A.M. van Heusden (1997). Manipulation of the human ovarian function: physiological concepts and clinical consequences. *Endocrine Reviews*, Vol. 18, No. 1, pp. 71 – 106.

[Goldman & Charniak, 1993] R.P. Goldman and E. Charniak (1993). A language for the construction of belief networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 15, No. 3, pp. 196 – 208.

[Graham & Gutgesell, 1995] T.P. Graham and H.P. Gutgesell (1995). Ventricular septal defects. *Heart Disease in Infants, Children, and Adolescents*, A.J. Moss and F.H. Adams et al. (Eds.), Williams & Wilkins, Baltimore, pp. 724–746.

[Habbema *et al.*, 1981] J.D.F. Habbema, J. Hilden, and B. Bjerregaard (1981). The measurement of performance in probabilistic diagnosis. *Methods of Information in Medicine*, Vol. 20, pp. 97 – 100.

[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.

[Habbema & Hilden, 1979] J.D.F. Habbema and J. Hilden (1979). Some pitfalls and difficulties in evaluating decision aids in in clinical practice. *Evaluation of Efficacy of Medical Action*, Alprovitch, F.T. de Dombal, and F. Grmy (Eds.), North-Holland, Amsterdam, pp. 515 – 521.

[Heckerman *et al.*, 1992] D.E. Heckerman, E.J. Horvitz, and B.N. Nathwani (1992). Toward normative expert systems. Part II: probability-based representations for

efficients knowledge acquisition and inference. *Methods of Information in Medicine*, Vol. 31, pp. 106 – 116.

[Heckerman & Nathwani, 1992] D.E. Heckerman and B.N. Nathwani (1992). Toward normative expert systems. Part I: The Pathfinder project. *Methods of Information in Medicine*, Vol. 31, pp. 90 – 105.

[Helton, 1993] J.C. Helton (1993). Uncertainty and sensitivity analysis techniques for use in performance assessment for radoactive waste disposal. *Reliability Engineering and System Safety*, Vol. 42, pp. 327 – 367.

[Henrion, 1989] M. Henrion (1989). Some practical issues in constructing belief networks. *Uncertainty in Artificial Intelligence 6*, L.N. Kanal, J.F. Lemmer (Eds.), Elsevier Science, Amsterdam, pp. 17 –32.

[Henrion *et al.*, 1996] M. Henrion, M. Pradhan, B. Del Favero, K. Huang, G. Provan, and P. O'Rorke (1996). Why is diagnosis using belief networks insensitive to imprecision in probabilities ? *Proceedings of the Twelfth Conference on Uncertainty in Artificial Intelligence*, pp. 307 – 314.

[Hosmer & Lemeshow, 1989] D.W. Hosmer and S. Lemeshow (1989). *Applied Logistic Regression*. John Wiley & Sons, New York.

[Howard, 1990] R. Howard (1990). From influence to relevance to knowledge. *Influence Diagrams, Belief nets and Decision Analysis*, R.M. Oliver and J.Q. Smith (Eds.), John Wiley & Sons, New York.

[Howard & Matheson, 1981] R. Howard and J. Matheson (1981). Influence Diagrams. *The Principles and Applications of Decision Analysis*, Vol. II, R. Howard and J. Matheson (Eds.), Strategic Decisions Group, Menlo Park, CA, pp. 719 – 762.

[Imani *et al.*, 1998] B. Imani, M.J. Eijkemans, E.R. te Velde, J.D. Habbema, and B.C. Fauser (1998). Predictors of patients remaining anovulatory during clomiphene citrate induction of ovulation in normogonadotrophic oligo-amenorrheic infertility. *Journal of Clinical Endocrinology & Metabolism*, Vol. 83, No. 7, pp. 2361-2365.

[Imani *et al.*, 2000] B. Imani, M.J. Eijkemans, F.H. de Jong, N.N. Payne, P. Bouchard, L.C. Giudice, and B.C. Fauser (2000). Free androgen index and leptin are the most prominent endocrine predictors of ovarian response during clomiphene citrate induction of ovulation in normogonadotrophic oligoamenorrheic infertility. *Journal of Clinical Endocrinology & Metabolism*, Vol. 85, pp. 676-682.

[Jansen *et al.*, 1998] S.L. Jansen, A.M. Stiggelbout, P.P. Wakker, T.P. Vliet Vlieland, J.W. Leer, M.A. Nooy, and J. Kievit (1998). Patients' utilities for cancer treatments:

a study of the chained procedure for the standard gamble and time tradeoff. *Medical Decision Making*, Vol. 18, No. 4, pp. 391-399.

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

[Jensen, 1996] F.V. Jensen (1996). *An Introduction to Bayesian Networks*, UCL Press, London.

[Jensen *et al.*, 1990] F.V. Jensen, S.L. Lauritzen, and K.G. Olesen (1990). Bayesian updating in causal probabilistic networks by local computations. *Computational Statistics Quarterly*, Vol. 4, pp. 269 – 282.

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

[Kalbfleisch & Prentice, 1980] J.D. Kalbfleisch and R.L. Prentice (1980). *The Statistical Analysis of Failure Time Data*. John Wiley & Sons, New York.

[Keeney & Raiffa, 1976] R.L. Keeney and H. Raiffa (1976). *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*. John Wiley & Sons, New York.

[Kjærulff & Van der Gaag, 2000] U. Kjærulff and L.C. van der Gaag (2000). Making sensitivity analysis computationally efficient, submitted for publication.

[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, No. 3, pp. 219 – 241.

[Krovetz, 1998] L.J. Krovetz (1998). Spontaneous closure of ventricular septal defect, *American Journal of Cardiology*, Vol. 81, pp. 100–101.

[Laskey, 1995] K.B. Laskey (1995). Sensitivity analysis for probability assessments in Bayesian networks. *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 25, pp. 901 – 909.

[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.

[Lawless, 1982] J.F. Lawless (1982). *Statistical Models and Methods for Lifetime Data*. John Wiley & Sons, New York.

[Madsen & Jensen, 1998] A.L. Madsen and F.V. Jensen (1998). Lazy propagation in junction trees. *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*, pp. 362 – 369.

[Madsen, 1999] A.L. Madsen (1999). All good things come to those who are lazy; efficient inference in Bayesian networks and influence diagrams based on lazy evaluation. Ph. D. thesis, Department of Computer Science, Aalborg University.

[Matheson, 1990] J.E. Matheson (1990). Using influence diagrams to value information and control. *Influence Diagrams, Belief nets and Decision Analysis*, R.M. Oliver and J.Q. Smith (Eds.), John Wiley & Sons, New York.

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

[Moller *et al.*, 1991] J.H. Moller, C. Patton, R.L. Varco, and C.W. Lillehei (1991). Late results (30 to 35 years) after operative closure of isolated ventricular septal defect from 1954 to 1960, *American Journal of Cardiology*, Vol. 68, pp. 1491–1497.

[Monti & Carenini, 1995] S. Monti and G. Carenini (1995). Dealing with expert inconsistencies: the sooner the better. *Working Notes of the Workshop on Building Probabilistic Networks: Where Do the Numbers Come From?*, M.J. Druzdzel, L.C. van der Gaag, M. Henrion, and F.V. Jensen (Eds.), pp. 33–40.

[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.

[Nease & Owens, 1997] R.F. Nease and D.K. Owens (1997). Use of Influence Diagrams to Structure Medical Decisions. *Medical Decision Making*, Vol. 17, pp. 263 – 275.

[Owens *et al.*, 1997] D.K. Owens, R.D. Shachter, and R.F. Nease (1997). Representation and Analysis of Medical Decision Problems with Influence Diagrams. *Medical Decision Making*, Vol. 17, pp. 241 – 262.

[Pauker & Kassirer, 1980] 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.

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

[Peek, 1999] N. Peek (1999). Explicit temporal models for decision-theoretic planning of clinical management, *Artificial Intelligence in Medicine*, Vol. 15(2), pp. 135–154.

[Peek & Ottenkamp, 1997] N. Peek and J. Ottenkamp, Developing a decision-theoretic network for a congenital heart disease. *AIME '97: Proceedings of the Sixth Conference on Artificial Intelligence in Medicine Europe*, E. Keravnou, C. Garbay, R. Baud, and J. Wyatt (Eds.), Springer Verlag, Berlin, pp. 157–168.

[Poole, 1989] D. Poole (1989). Normality and faults in logic-based diagnosis. *Proceedings of the Eleventh International Joint Conferences on Artificial Intelligence*, Detroit, MI, pp. 1304 – 1310.

[Pradhan *et al.*, 1996] M. Pradhan, M. Henrion, G. Provan, B. Del Favero, and K. Huang (1996). The sensitivity of belief networks to imprecise probabilities: an experimental investigation. *Artificial Intelligence*, Vol. 85, pp. 363 – 397.

[Quaglini *et al.*, 1994] S. Quaglini, R. Bellazzi, F. Locatelli, M. Stefanelli, and C. Salvaneschi (1994). An Influence Diagram for Assessing GVHD Prophylaxis after Bone Marrow Transplantation in Children. *Medical Decision Making*, Vol. 14, pp. 223 – 235.

[Raiffa & Schlaifer, 1961] H. Raiffa and R. Schlaifer (1961). *Applied Statistical Decision Theory.* Addison-Wesley, Reading, MA.

[Renooij & Witteman, 1999] S. Renooij and C.L.M. Witteman (1999). Talking probabilities: communicating probabilistic information with words and numbers. *International Journal of Approximate Reasoning*, Vol. 22, pp. 169 – 194.

[Schreiber *et al.*, 1994] A.T. Schreiber, B. Wielinga, R. de Hoog, H. Akkermans, and W. van de Velde (1994). CommonKADS: a comprehensive methodology for KBS development. *IEEE Expert*, Vol. 9, No. 6, pp. 28 – 37.

[Shachter, 1986] R.D. Shachter (1986). Evaluating influence diagrams, *Operations Research*, Vol. 34, pp. 79 – 90.

[Shafer & Shenoy, 1990] G. Shafer and P.P. Shenoy (1989). Probability propagation. *Annals of Mathematics and Artificial Intelligence*, Vol. 2, pp. 327 – 352.

[Shenoy, 1996] P.P. Shenoy (1996). Binary join trees *Proceedings of the Twelfth Conference on Uncertainty in Artificial Intelligence*, pp. 492 – 499.

[Shwe *et al.*, 1991] M.A. Shwe, B. Middleton, D.E. Heckerman, M. Henrion, E.J. Horvitz, H.P. Lehmann, and G.F. Cooper (1991). Probabilistic diagnosis using a reformulation of the INTERNIST-1/QMR knowledge base I: the probability model and inference algorithms. *Methods of Information in Medicine*, Vol. 30, pp. 241 – 255.

[Snick *et al.*, 1997] H.K.A. Snick, T.S. Snick, J.L.H. Evers, and J.A. Collins (1997). The spontaneous pregnancy prognosis in untreated subfertile couples: the Walcheren primary care study. *Human Reproduction*, Vol. 12, No. 7, pp. 1582 – 1588.

[Sox *et al.*, 1988] H.C. Sox Jr., M.A. Biatt, M.C. Higgins, and K.I. Marton (1988). *Medical Decision Making*, Butterworth, Stoneham.

[Spiegelhalter, 1989] D. Spiegelhalter (1989). A unified approach to imprecision and sensitivity of beliefs in expert systems. *Uncertainty in Artificial Intelligence 3*, L.N. Kanal, T.S. Levitt, and J.F. Lemmer (Eds.), North-Holland, Amsterdam, pp. 199 – 208.

[Stiggelbout, 1995] A.M. Stiggelbout (1995). *Trade-offs between Quality and Quantity of Life; Methodological Aspects of Outcome valuation in Cancer Patients*, Dissertation, Leiden University.

[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.

[Van der Gaag, 1994] L.C. van der Gaag (1994). *Evidence Absorption – Experiments on Different Classes of Randomly Generated Belief Networks*, Technical Report UU-CS-94-42, Utrecht University.

[Van der Gaag & Meyer, 1998] L.C. van der Gaag and J-J.Ch. Meyer (1998). Informational independence: models and normal forms. *International Journal of Intelligent Systems*, Vol. 13, pp. 83 – 109.

[Van der Gaag *et al.*, 1999] L.C. van der Gaag, S. Renooij, C.L.M. Witteman, B.M.P. Aleman, and B.G. Taal (1999). How to elicit many probabilities. *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*, Morgan Kaufmann, San Francisco, pp. 647 – 654.

[Von Neumann & Morgenstern, 1944] J. von Neumann and O. Morgenstern (1944). *The Theory of Games and Economic Behavior*. John Wiley & Sons, New York.

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

[Weinstein & Fineberg, 1980] M.C. Weinstein and H.V. Fineberg (1980). *Clinical Decision Analysis*. Press of W.B. Saunders Company, USA.

[Wellman, 1990] M.P. Wellman (1990). Fundamental concepts of qualitative probabilistic networks, *Artificial Intelligence*, Vol. 44(3), pp. 257–303.

Appendices

Summary

PART I

Decision-theoretic networks

To support medical decision making, a wealth of techniques is available. Recently, there is an increasing interest in decision-theoretic networks. There are two types of decision-theoretic network, termed *belief networks* and *influence diagrams*. Both types of network are used to construct a model of reality. A belief network is a model of the relations between a number of events and states. It is used to compute the probability that a certain event will occur (prognosis) or that a specific state is true (diagnosis). An influence diagram is a model of a decision problem in which various risks and benefits have to be weighed in a situation of uncertainty. An influence diagram is used to compute an optimal decision.

Mathematically, both types of network consist of a graphical structure and a set of parameters. In a belief network, the graphical structure consists of nodes and arrows. Each node in the graph corresponds to a chance variable. When there is an arrow between two nodes, the variable at the head of the arrow is an effect of the variable at the tail of the arrow. For each variable in the graph, the probability of its values are given for every possible combination of values for its 'parents' in the graph. These probabilities are the parameters of a belief network. From a belief network, the probability of any variable can be computed. Suppose, for example, that we have a specific patient for whom the values of some of the variables in the network are observed. From the belief network, then, the probabilities of the values of the remaining variables can be computed for this patient. This gives, as an outcome of the network, a diagnosis or a prognostic assessment.

Influence diagrams resemble belief networks. In addition to chance variables, an influence diagram contains decision nodes and a utility node. The decision nodes represent the various decisions that have to be taken in the decision problem that is addressed. The utility node represents the possible outcomes (in terms of, for example, mortality and morbidity) of the decision problem and their associated utilities. These utilities are the values that the decision maker attaches to the outcomes of the decision problem. Thus, the parameters in an influence diagram are, in addition to

the conditional probabilities for the chance nodes, the utilities belonging to the utility node. From an influence diagram, the expected alue, also called expected utility, of each decision or sequence of decisions can be computed. The outcome of the diagram is the optimal decision for a specific patient; it is the decision for which the expected utility is maximal.

Traditional decision-analytic techniques

In statistical analysis and decision analysis, techniques exist that perform similar tasks as decision-theoretic networks. Regression analysis, for example, provides a formula that is used to compute a probability of interest using a set observations for the patient under consideration. Decision trees are used to compute, for a specific patient, the optimal decision in a decision problem. Regression techniques and decision trees are widely used in decision analysis nowadays. They are practical tools that are relatively easy to understand and to work with in medical practise. However, when large and complicated problems are considered, for example, with complex interactions between variables, decision trees and regression models may fall short. Decision-theoretic networks can handle these problems because of their compact graphical representation.

Problems concerning the use of decision-theoretic networks

Until now, the number of decision-theoretic networks in use for supporting medical decision making is limited. This is mainly due to the fact that building these networks is difficult. The most problematic part is the quantification of the network, that is, finding the required assessments for all parameters. Usually, not enough patient data is available to estimate *all* parameters. Medical experts can also provide the assessments, but this is a time-consuming task and the resulting assessments may be quite inaccurate. To increase the practicability of decision-theoretic networks, therefore, methods for efficient quantification are needed.

A second problem in using decision-theoretic networks for medical applications is the fact that, presently, there are not many techniques that support the analysis of decision-theoretic networks. For example, an important aspect of a model for decision support is the robustness of the results and recommendations of the model to inaccuracies in the model's parameters. Robustness can be investigated with a *sensitivity analysis*. In a sensitivity analysis, the assessments for one or more inaccurate parameters in a model are varied simultaneously over a plausible interval and the effect of these variations on the outcome of the model is studied. If varying a parameter assessment over its plausible interval has a large effect on the outcome of the model, it is important to know its value accurately. For decision trees, techniques for sensitivity analysis are available, but these techniques are too time-consuming for decision-theoretic networks. To make decision-theoretic networks a practical tool for medical decision support, it is thus important to develop an efficient technique for sensitivity analysis.

The rationale for this thesis

The two issues described above provided the motivation for the research presented in this thesis. Efficient methods for sensitivity analysis of decision-theoretic networks were developed and it was investigated how the technique of sensitivity analysis can be used to facilitate the quantification of such networks. In the following, the research presented in this thesis is summarized.

PART II

In Chapters 2 to 5, constituting Part II of this thesis, the focus is on belief networks. Chapters 2 and 3 address the use of sensitivity analysis to facilitate the quantification of a belief network and Chapters 4 and 5 present efficient methods for sensitivity analysis.

Using sensitivity analysis to quantify a belief network efficiently

To quantify a belief network efficiently, we develop in Chapter 2 an iterative procedure. Our procedure sets out with a network that is quantified roughly, that is, only a limited amount of time is spent on the assessment of the parameters and these assessments can be very inaccurate. Subsequently, a sensitivity analysis of the network is performed. The results are used to discriminate between parameters that are highly influential and parameters that are not. Since the assessments for the influential parameters are crucial to the outcome of the network, they should be estimated with care. The following step of the procedure therefore is to refine these assessments. The refinement can, for example, be achieved by using advanced techniques to obtain assessments from experts, to collect data or to conduct a literature study. The resulting refined assessments are supposed to be more accurate than the original assessments. The two steps of sensitivity analysis and parameter refinement should be repeated until satisfactory behaviour of the network is obtained.

In Chapter 3, the quantification procedure of Chapter 2 is tested using a belief network describing the pathophysiology of the congenital heart disease known as *ventricular septal defect*. For the network, three quantifications were obtained. These quantifications differ with respect to the level of expertise of the person who supplies the assessments for the network's parameters. The supposedly best quantifcation was given by an expert cardiologist. The other quantifications were improved upon by applying two steps of the quantification procedure. First, the influential parameters in each network quantification were identified by performing sensitivity analyses. Subsequently, these influential parameters were refined by stepwise replacing the corresponding assessments with the assessments from the expert quantification. The results of the replacements were investigated by comparing network predictions for a number of different patient profiles with the predictions of the network in which *all* parameters

were estimated by the expert. It was found that it may be sufficient to gather a limited number of accurate assessments to obtain an adequate network quantification.

Efficient methods for sensitivity analysis

In Chapter 4 and 5, efficient methods for sensitivity analysis are presented. The naive way to perform a sensitivity analysis of a belief network is to vary parameter assessments stepwise and in a systematic way. For every step, the outcome of the belief network is to be computed. As evaluating a network is computationally expensive, performing sensitivity analysis in this way is far too time-consuming to be of practical use.

Chapter 4 focuses on *one-way sensitivity analysis* of a belief network. In a one-way sensitivity analysis, the effect of inaccuracies in a single parameter assessment is studied. Our method for one-way sensitivity analysis uses the information that is contained in the graphical structure of the network. For a given patient, for whom certain variables are observed, the structure of the network reveals the parameters that cannot influence the network's outcome. These parameters can be excluded from further analysis. Furthermore, we showed that, due to the structure of the network, the outcome is a simple mathematical function of the one parameter that is varied in a one-way sensitivity analysis. Knowing the form of this function, sensitivity analysis simply amounts to establishing the coefficients in this function. These two properties make one-way sensitivity analysis of a belief network computationally far more efficient.

In Chapter 5, our method for one-way sensitivity analysis is extended to sensitivity analysis with respect to an arbitrary number of parameters. Sensitivity analysis in which a number of parameters are considered together reveals possible interactions between the inaccuracies in the assessments for these parameters. For example, it may happen that varying two parameter assessments independently has no effect on the outcome of the network, whereas the outcome changes dramatically if the two parameter assessments are varied at the same time. In a sensitivity analysis in which several parameters are considered simultaneously, it can be shown that the outcome of the network again is a simple function of these parameters. To know the joint effect of these parameters, it is therefore sufficient to establish the coefficients in this function. In Chapter 5, we construct a computational framework that is used to collect all information in the network that is necessary to compile the required coefficients.

PART III

In Part III, consisting of Chapters 6 and 7, the focus is on decision making, that is, on influence diagrams and on belief networks that are used for decision making.

Sensitivity analysis of belief networks enhanced with the threshold model

Belief networks can be used for decision making by using the threshold model. An expert sets, for example, a threshold probability for operation mortality, such that the patient is operated if his operation mortality, computed from the belief network, is below the threshold and not operated if his operation mortality is above the threshold. Sensitivity analysis of a belief network used for decision making, now, shows whether inaccuracies in a parameter assessment can change the decision that is recommended. As a measure of sensitivity, we introduce the concept of the *minimum deviation*; it is the minimum change in a probability assessment under study that does induce a change in the currently recommended decision. In general, the smaller the minimum deviation for a parameter, the more important it is to know the assessment of that parameter accurately. In Chapter 6, we show how to compute the minimum deviation efficiently, using the method for sensitivity analysis presented in Chapter 4.

Sensitivity analysis of influence diagrams

In Chapter 7, decision support with influence diagrams is considered. The outcome of an influence diagram is an optimal decision, that is, the decision that has the maximum expected value (or expected utility). Sensitivity analysis of an influence diagram serves to find out whether changing the value of one or more parameter assessments can change the optimal decision. In Chapter 7, we use again the concept of the minimum deviation as a measure of sensitivity. We present a method that efficiently computes, for each parameter independently, this minimum deviation for changing the recommended decision. The method is closely related to the method in Chapter 4. In fact, similarly to our method for sensitivity analysis of belief networks, we exploit the fact that expected utility in an influence diagram is a simple mathematical function of a parameter that is considered in a sensitivity analysis.

PART IV

The work presented in this thesis focuses on the technique of sensitity analysis. We developed efficient methods that make sensitivity analysis of decision-theoretic networks feasible. Furthermore, it is described how sensitivity analysis can be used to facilitate the difficult task of quantifying a network. As such, this thesis brings decision-theoretic networks one step closer to their practical use in medical decision making. However, much work remains to be done before decision-theoretic networks will be a commonly used decision-analytic tool. Additional techniques supporting their construction and evaluation, such as techniques for uncertainty analysis and for validation, should be designed. And last but not least, there is the great challenge to obtain practical experience with building decision-theoretic networks for clinically relevant problems and to evaluate their merit in daily medical practise.

Samenvatting

DEEL I

Besliskundige netwerken

Om medische beslissingen te ondersteunen is er een scala aan technieken beschikbaar. Sinds kort is er een toenemende interesse in besliskundige netwerken. Er zijn twee typen besliskundige netwerken, het *belief network*, wat ook wel vertaald wordt met *probabilistisch netwerk*, en het *influence diagram*, hier vertaald met beslisdiagram. Beide typen netwerken worden gebruikt om een model van de realiteit te maken. Een probabilistisch netwerk is een model van de relaties tussen een aantal toestanden en gebeurtenissen. Het wordt gebruikt om te berekenen wat de kans is dat een bepaalde toestand werkelijkheid is (diagnose) of dat een bepaalde gebeurtenis zich zal voordoen (prognose). Een beslisdiagram is een model van een beslisprobleem waarin een aantal voors en tegens afgewogen moeten worden in een onzekere situatie. Een beslisdiagram wordt gebruikt om uit te rekenen wat de optimale beslissing is.

Wiskundig gezien bestaan beide typen netwerken uit een grafische structuur en een set parameters. In een probabilistisch netwerk bestaat de grafische structuur uit knopen en pijlen. Elke knoop correspondeert met een kansvariabele. Een pijl tussen twee knopen betekent dat de variabele aan de kop van de pijl een effect is van de variabele aan de staart van de pijl. Voor elke variabele is de kans op elk van zijn waarden gegeven, en wel voor elke combinatie van waarden van de 'ouders' van deze variabele in de grafische structuur. Deze kansen zijn de parameters van een probabilistisch netwerk. Met het netwerk kan de kansverdeling van elke variabele uitgerekend worden. Laten we bijvoorbeeld een specifieke patiënt nemen van wie we de waarden weten voor een aantal variabelen in het netwerk. Uit het netwerk kan dan, voor deze patiënt, de kans uitgerekend worden op de waarden van de overgebleven variabelen. Als uitkomst van het netwerk levert dit dan een diagnostische of prognostische schatting op.

Beslisdiagrammen lijken op probabilistische netwerken. Behalve kansvariabelen bevat een beslisdiagram beslisknopen en een utiliteitsknoop. De beslisknopen vertegenwoordigen de verschillende beslissingen die genomen moeten worden in het desbetreffende beslisprobleem. De utiliteitsknoop vertegenwoordigt de mogelijke uitkomsten

van het beslisprobleem (in termen van, bijvoorbeeld, mortaliteit en morbiditeit) en de daarmee geassocieerde utiliteiten. Deze utiliteiten geven weer welke waarde de besliskundige hecht aan de uitkomsten van het beslisprobleem. De parameters in een beslisdiagram zijn dus, behalve de kansen voor de kansvariabelen, de utiliteiten behorende bij de utiliteitsknoop. Met een beslisdiagram kan de verwachte waarde, ofwel verwachte utiliteit, berekend worden van elke beslissing of elke reeks beslissingen. De uitkomst van een beslisdiagram is de optimale beslissing voor een gegeven patiënt; het is die beslissing waarvoor de utiliteit maximaal is.

Traditionele besliskundige technieken

In de statistiek en de besliskunde bestaan technieken die vergelijkbare taken uitvoeren als besliskundige netwerken. Regressieanalyse levert bijvoorbeeld een formule op die vervolgens gebruikt wordt om de kans waarin men geïnteresseerd is uit te rekenen aan de hand van een set van waarnemingen voor een specifieke patiënt. Beslisbomen worden gebruikt om de optimale beslissing in een beslisprobleem uit te rekenen. Regressietechnieken en beslisbomen worden veel gebruikt in de huidige besliskunde. Het zijn praktische methoden die relatief makkelijk te begrijpen zijn en waarmee in de medische praktijk goed gewerkt kan worden. Wanneer het echter gaat om zeer ingewikkelde problemen, bijvoorbeeld problemen waarbij er een complexe interactie tussen de verschillende variabelen is, schieten beslisbomen en regressiemodellen soms tekort. Besliskundige netwerken zijn in dat geval geschikt vanwege hun compacte grafische weergave van het probleem.

Problemen betreffende het gebruik van besliskundige netwerken

Tot nu toe is het aantal besliskundige netwerken dat gebruikt wordt ter ondersteuning van medische beslissingen zeer beperkt. Dit komt met name doordat deze netwerken moeilijk te bouwen zijn. Het meest lastige is de kwantificatie van een netwerk, dat wil zeggen, het vinden van de benodigde schattingen voor alle parameters. Meestal zijn er niet genoeg patiëntgegevens beschikbaar om alle parameters te schatten. De schattingen kunnen dan gegeven worden door experts. Dit is echter een tijdrovend karwei en de resulterende schattingen kunnen zeer onnauwkeurig zijn. Om de bruikbaarheid van besliskundige netwerken te vergroten zijn er dus methoden nodig voor efficiënte kwantificatie.

Een tweede probleem in het gebruik van besliskundige netwerken voor medische toepassingen is het feit dat er op dit moment niet veel technieken beschikbaar zijn die de analyse van besliskundige netwerken ondersteunen. Een belangrijk aspect is bijvoorbeeld de robuustheid van de resultaten van een beslissingsondersteunend model voor onnauwkeurigheden in de parameters van het model. De robuustheid van een wiskundig model kan onderzocht worden met een *sensitiviteitsanalyse*. In een sensitiviteitsanalyse worden de schattingen voor één of meer parameters gelijktijdig gevariëerd over een aan-

nemelijk interval en wordt het effect van deze variaties op de uitkomst van het netwerk bekeken. Als het variëren van een parameter een groot effect heeft op de uitkomst van het model is het belangrijk de waarde van deze parameter nauwkeurig te kennen. Voor beslisbomen bestaan er technieken voor sensitiviteitsanalyse, maar deze technieken zijn té tijdrovend voor besliskundige netwerken. Om van besliskundige netwerken een bruikbaar instrument te maken voor de ondersteuning van medische beslissingen is het belangrijk een efficiënte techniek voor sensitiviteitsanalyse te ontwikkelen.

De motivering voor dit proefschrift

De twee bovenstaande aspecten vormen de motivatie voor het onderzoek dat in dit proefschrift is beschreven. Enerzijds zijn er efficiënte methoden voor sensitiviteitsanalyse van besliskundige netwerken ontwikkeld en anderzijds is onderzocht hoe de techniek van sensitiviteitsanalyse gebruikt kan worden om de kwantificatie van dergelijke netwerken te vergemakkelijken. Hieronder volgt een samenvatting van het onderzoek dat beschreven is in dit proefschrift.

DEEL II

Deel II, bestaande uit de hoofdstukken 2 tot en met 5, richt zich op probabilistische netwerken. De hoofdstukken 2 en 3 hebben betrekking op het gebruik van sensitiviteitsanalyse om de kwantificatie van een probabilistisch netwerk te vergemakkelijken. In de hoofdstukken 4 en 5 worden efficiënte methoden voor sensitiviteitsanalyse gepresenteerd.

Efficiënte kwantificatie van een probabilistisch netwerk

In hoofdstuk 2 is een iteratieve procedure ontwikkeld om een probabilistisch netwerk efficiënt te kwantificeren. De procedure start met een netwerk dat slechts grof gekwantificeerd is. Dat wil zeggen dat er maar een beperkte hoeveelheid tijd besteed is aan het schatten van de parameters en dat deze schattingen erg onnauwkeurig kunnen zijn. Vervolgens wordt er een sensitiviteitsanalyse van het netwerk uitgevoerd. De resultaten van deze analyse worden gebruikt om onderscheid te maken tussen parameters die erg invloedrijk zijn en parameters die dat niet zijn. Aangezien de schattingen voor de invloedrijke parameters cruciaal zijn voor de uitkomst van het netwerk moeten ze zorgvuldig geschat worden. De volgende stap van de procedure richt zich daarom op het verfijnen van deze schattingen. Deze verfijning kan bijvoorbeeld bereikt worden door het gebruik van geavanceerde technieken voor het verkrijgen van schattingen van experts. Andere mogelijkheden zijn het verzamelen van data of het uitvoeren van een literatuurstudie. De zo verkregen verfijnde schattingen worden verondersteld nauwkeuriger te zijn dan de originele schattingen. De twee stappen van sensitiviteitsanalyse en het verfijnen van parameters moeten herhaald worden totdat een voldoende

goed functionerend netwerk verkregen is.

In hoofdstuk 3 wordt de kwantificatieprocedure van hoofdstuk 2 getest. Hiervoor wordt een netwerk gebruikt dat de fysiologie beschrijft van de congenitale hartafwijking *vertrikelseptumdefect*. Er werden drie kwantificaties van het netwerk verzameld die van elkaar verschillen wat betreft de deskundigheid van de persoon die de parameterschattingen gaf. De best veronderstelde kwantificatie werd gegeven door een expert, een cardioloog. De andere twee kwantificaties werden verbeterd door twee stappen van de kwantificatieprocedure toe te passen. Allereerst werden de invloedrijke parameters in elke kwantificatie geïdentificeerd door het uitvoeren van een sensitiviteitsanalyse. Vervolgens werden deze invloedrijke parameters verfijnd door de desbetreffende schattingen stapsgewijs te vervangen door de schattingen uit de expertkwantificatie. Het resultaat van de vervangingen werd voor een aantal verschillende patiëntprofielen onderzocht. De voorspellingen van elk verfijnd netwerk werden vergeleken met de voorspellingen van het netwerk waarin alle parameters door de expert waren geschat. Uit deze studie werd geconcludeerd dat het voor het verkrijgen van een adekwate netwerkkwantificatie waarschijnlijk voldoende is om een beperkt aantal nauwkeurige schattingen te verzamelen.

Efficiënte methoden voor sensitiviteitsanalyse

In hoofdstuk 4 en 5 worden efficiënte methoden voor sensitiviteitsanalyse gepresenteerd. De meest eenvoudige manier om sensitiviteitsanalyse van een probabilistisch netwerk uit te voeren is door het stapsgewijs en systematisch variëren van parameterschattingen. Bij elke stap wordt de uitkomst van het netwerk berekend. Aangezien het evalueren van een probabilistisch netwerk veel computertijd vergt, is het uitvoeren van een sensitiviteitsanalyse op deze manier te tijdrovend om nog praktisch bruikbaar te zijn.

Hoofdstuk 4 richt zich op *één-wegs sensitiviteitsanalyse* van een probabilistisch netwerk. Met een *één-wegs sensitiviteitsanalyse* wordt het effect van onnauwkeurigheden in één enkele parameterschatting bestudeerd. Onze methode voor *één-wegs sensitiviteitsanalyse* gebruikt de informatie die intrinsiek in de grafische structuur van een probabilistisch netwerk besloten ligt. Uit de structuur van het netwerk kan, voor een gegeven patiënt, afgeleid worden welke parameters de uitkomst van het netwerk niet kunnen beïnvloeden. Deze parameters kunnen verder buiten beschouwing gelaten worden. Voorts hebben we laten zien dat de uitkomst, dankzij de structuur van het netwerk, een simpele wiskundige functie is van de ene parameter die gevariëerd wordt in een één-wegs sensitiviteitsanalyse. Kennis van de vorm van deze functie reduceert het uitvoeren van een sensitiviteitsanalyse tot het bepalen van de coëfficiënten in deze functie. Deze twee eigenschappen maken één-wegs sensitiviteitsanalyse aanzienlijk efficiënter.

In hoofdstuk 5 wordt onze methode voor één-wegs sensitiviteitsanalyse uitgebreid naar sensitiviteitsanalyse met betrekking tot een willekeurig aantal parameters. Een sensitiviteitsanalyse waarin een aantal parameters tegelijk gevariëerd wordt, laat mogelijke interacties zien tussen de onnauwkeurigheden in de schattingen voor deze parameters. Het kan bijvoorbeeld gebeuren dat het variëren van twee parameters onafhankelijk van elkaar geen effect heeft op de uitkomst van het netwerk, terwijl de uitkomst drastisch verandert wanneer de twee parameters tegelijkertijd gevariëerd worden. In een sensitiviteitsanalyse met betrekking tot een willekeurig aantal parameters kan wederom aangetoond worden dat de uitkomst van het netwerk een eenvoudige functie is van deze parameters. Om het gemeenschappelijk effect van deze parameters te meten is het dus voldoende om de coëfficiënten in deze functie te bepalen. In hoofdstuk 5 construeren we een rekenkundig raamwerk dat gebruikt wordt om alle informatie in het netwerk die nodig is om de gewenste coëfficiënten te berekenen te verzamelen.

DEEL III

Deel III, bestaande uit de hoofdstukken 6 en 7, richt zich op beslisdiagrammen en probabilistische netwerken die gebruikt worden voor besliskunde.

Sensitiviteitsanalyse van besliskundige probabilistische netwerken

Probabilistische netwerken kunnen voor besliskunde gebruikt worden door ze uit te breiden met het drempelwaarde model. Een expert bepaalt een drempelwaarde voor, bijvoorbeeld, de operatiemortaliteit. De patiënt wordt geopereerd als zijn operatiemortaliteit, berekend uit het netwerk, lager is dan deze drempelwaarde. Als de operatiemortaliteit hoger is dan deze drempelwaarde wordt de patiënt niet geopereerd. Sensitiviteitsanalyse van een besliskundig probabilistisch netwerk laat zien of onnauwkeurigheid in de schatting van een parameter de aanbevolen beslissing kan veranderen. Als een maat van sensitiviteit introduceren we het concept van de *minimale deviatie*; dit is de minimale verandering in een kansschatting die een verandering in de huidige aanbevolen beslissing veroorzaakt. In het algemeen geldt dat hoe kleiner de minimale deviatie voor een parameter is, des te belangrijker het is de waarde van die parameters nauwkeurig te kennen. In hoofdstuk 6 laten we zien hoe deze minimum deviatie efficiënt berekend kan worden door gebruik te maken van de methode die in hoofdstuk 4 gepresenteerd is.

Sensitiviteitsanalyse van beslisdiagrammen

In hoofdstuk 7 komt beslissingsondersteuning met behulp van beslisdiagrammen aan bod. De uitkomst van een beslisdiagram is een optimale beslissing. De optimale beslissing is die beslissing waarvoor de verwachte waarde (of verwachte utiliteit) maximaal is. Sensitiviteitsanalyse van een beslisdiagram heeft tot doel om erachter te komen of het

variëren van één of meer parameterschattingen de optimale beslissing kan veranderen. In hoofdstuk 7 wordt wederom het concept van de *minimale deviatie* als maat voor de sensitiviteit gebruikt. We presenteren een efficiënte methode die voor elke afzonderlijke parameter uitrekent wat de minimale deviatie is die nodig is om de optimale beslissing te doen veranderen. De methode vertoont veel overeenkomsten met de methode in hoofdstuk 4. We maken gebruik van het feit dat de verwachte utiliteit in een beslisdiagram een eenvoudige wiskundige functie is van de parameter die onderzocht wordt in de sensitiviteitsanalyse.

DEEL IV

In het onderzoek dat beschreven is in dit proefschrift ligt de nadruk op de techniek van sensitiviteitsanalyse. We hebben efficiënte methoden ontwikkeld die sensitiviteitsanalyse van besliskundige netwerken haalbaar maakt. Verder hebben we onderzocht hoe sensitiviteitsanalyse gebruikt kan worden om het kwantificeren van een netwerk te vergemakkelijken. Dit proefschrift brengt besliskundige netwerken hiermee één stap dichter bij het praktische gebruik in de medische besliskunde. Er is echter nog veel werk te verzetten voordat besliskundige netwerken een algemeen gebruikt besliskundig instrument zullen zijn. Aanvullende technieken die hun bouw en evaluatie ondersteunen, zoals technieken voor onzekerheidsanalyse en technieken voor validatie, moeten nog ontwikkeld worden. Tenslotte ligt er de grote uitdaging om praktische ervaring op te doen met het bouwen van besliskundige netwerken voor klinisch relevante problemen en om hun toegevoegde waarde in de medische praktijk te evalueren.

Dankwoord

Beste Tetrade-genoten, collega's, vrienden en familie,

Hier is het dan! Al die jaren hebben jullie, vaak meer dan ik, er het vertrouwen in gehad dat het ook écht zou lukken. Daar wil ik jullie voor bedanken, alsmede voor de steun, de opbouwende kritieken en de welkome afleiding. Allereerst gaat mijn dank uit naar mijn promotores Dik Habbema en Linda van de Gaag. Dik, voor deze lastige pupil was jouw kritische blik niet altijd gemakkelijk; ik heb er veel van geleerd. Op momenten dat de hordes me té hoog werden, was jij een grote steun, met geruststellende woorden en kordaat optreden. Dank je wel. Linda, jij liet me zien hoe wetenschap kan zijn; mensen die gedreven door liefde voor hun vak en warmte voor elkaar op zoek zijn naar antwoorden. Bedankt! Behalve Dik en Linda, wil ik ook Tetrade-genoten Peter en Niels bedanken, voor de interessante inhoudelijke discussies, voor de gezelligheid en de lekkere biertjes. CKB-collega's, ik ben julie dankbaar voor het delen van jullie brede kennis van de medische besliskunde, maar vooral ook voor de leuke contacten. Edith, Sonja en Mona, die paar keer dat ik zelf geprobeerd heb een drukbezet hoogleraar te bereiken heeft me een enorme bewondering gegeven voor jullie secretariële werkzaamheden. Dank jullie wel! Gerard, zonder jouw hulp werd het niets met die lay-out, bedankt. Lieve René, Pieta, Johan en Caroline, mijn *i*MGZ kamergenoten, misschien heb ik jullie met mijn drukte en mijn gezang af en toe tot wanhoop gedreven. Desondanks hoop ik dat jullie met net zo veel plezier als ik terugkijken op de vele uren die we samen doorgebracht hebben. René, vooral met jou heb ik de afgelopen jaren heel wat lief en leed gedeeld. Dank je. Naast mijn wetenschappelijke thuis, het *i*MGZ, wil ik twee vakgroepen bedanken voor hun gastvrijheid. Ten eerste, de vakgroep Informatica van de Universiteit van Utrecht waar ik met groot plezier heel wat dagen heb gewerkt. Secondly, I will never forget the hospitality of the research unit Decision Support Systems of the Aalborg University in Denmark. In particular, I would like to thank Finn Jensen and Uffe Kjærulff. Our intensive discussions and the stimulating atmosphere in your group have made my two months stay very enjoyable. Robert, though you live too far away, you always were there for me. Thank you! Tenslotte, lieve vrienden, zusjes, paps en mams, ik heb jullie de afgelopen jaren verwaarloosd. De knuffels per email lieten, in werkelijkheid, soms lang op zich wachten. En zelfs die emails... Ik hoop het nu allemaal in te halen. Lieve Mart, ga je mee? De wijde wereld lonkt.

Curriculum Vitae

Veerle Coupé was born on March 17, 1972 in Gent, Belgium. She graduated in 1990 at the 'Adelbert College' in Wassenaar (secondary school) and started to study astronomy at the University of Leiden. During her studies, she spent a year at the 'Observatoire de Meudon' (near Paris). There, she worked on the estimation of the mean zonal circulation in Jupiter's atmosphere and on the estimation of the cloud particle scale height of Venus' atmosphere. She obtained observational experience with the 1-meter telescope at the Pic du Midi in the Pyrenees and with the Dutch telescope in la Silla, Chili. After half a year traveling in South America, she graduated in astronomy in January 1996. One month later, she started to work on her Ph.D. at the Department of Public Health at the Erasmus University Rotterdam. In cooperation with the Department of Computer Science at the University of Utrecht, she developed methods for sensitivity analysis of decision-theoretic networks. As part of her Ph.D. research, she spent two months in Aalborg, Denmark, where she worked on a computational architecture for efficiently carrying out a sensitivity analysis of a Bayesian network. Furthermore, she obtained additional insight in techniques to study the robustness and reliability of model output by participating in the European Workshop on Knowledge Acquisition, Modeling and Management, in Barcelona, in the Symposium on Sensitivity analysis of model Output, in Venice, and in the European Conference on Artificial Intelligence in Medicine and Medical Decision Making, in Aalborg. In the summer of 2000, she obtained her MSc-degree in Clinical Epidemiology of the Netherlands Institute of Health Sciences (NIHES).