PRACTICABLE SENSITIVITY ANALYSIS OF BAYESIAN BELIEF NETWORKS

Veerle M.H. Coupé

Center for Clinical Decision Sciences, Erasmus University Rotterdam,
P.O. Box 1738, 3000 DR Rotterdam, The Netherlands
e-mail: coupe@mgz.fgg.eur.nl

Linda C. van der Gaag

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

Abstract

By subjecting a Bayesian belief network to a sensitivity analysis with respect to its conditional probabilities, the reliability of its output can be evaluated. Unfortunately, straightforward sensitivity analysis of a belief network is highly time-consuming, as a result of the usually large number of probabilities to be investigated. In this paper, we show that the graphical independence structure of a Bayesian belief network induces various properties that allow for reducing the computational burden of a sensitivity analysis. We show that several analyses can be identified as being uninformative because 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. By exploiting these properties, the practicability of sensitivity analysis of Bayesian belief networks is enhanced considerably.

AMS classification: 60K10, 62A15, 62N05, 68Q25.
Key words: Probabilistic models, Bayesian belief networks, sensitivity analysis, computational efficiency.

1 INTRODUCTION

Bayesian belief networks have become widely accepted as intuitively appealing probabilistic models that are highly valuable in addressing real-life problems in such complex domains as medical diagnosis, weather forecast, and probabilistic information retrieval. A Bayesian belief network basically is a concise representation of a joint probability distribution on a set of statistical variables. It encodes, in a graphical structure, the variables under study, along with their independences; the dependences among the variables are captured by conditional probabilities (Pearl, 1988).

The conditional probabilities of a Bayesian belief network are generally assessed from statistical data or by human experts. As a result of incompleteness of data and partial knowledge of the problem domain being modelled, the assessments obtained are inevitably inaccurate. Inaccuracies in the assessments for the various conditional probabilities 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 on inaccurate assessments.

The reliability of the output of a Bayesian belief network can be evaluated by subjecting the network 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 model's parameters are varied systematically (Morgan and Henrion, 1990; Habbema *et al.*, 1990). For a Bayesian belief network, sensitivity analysis amounts to varying the assessments for one or more of its conditional probabilities simultaneously and investigating the effects of deviation on a probability of interest or, for example, on a diagnosis or decision based upon this probability of interest (Laskey, 1995; Coupé *et al.*, 1998).

Straightforward sensitivity analysis of a Bayesian belief network, unfortunately, is highly time-consuming. In the simplest type of analysis, for example, for every single conditional probability, a number of deviations from its assessment is investigated. For every value under study, the probability of interest is computed from the network.

Even for a rather small belief network, this easily requires tens of thousands of network computations. In fact, the computational burden involved is prohibitive when sensitivity analysis is to be used for verifying the reliability of a belief network's output in, for example, daily medical practice. To be of practical use, therefore, more efficient methods for sensitivity analysis of Bayesian belief networks should be found.

In earlier work on sensitivity analysis of Bayesian belief networks, K. Blackmond Laskey has introduced an efficient method for computing *sensitivity values* from a belief network (Laskey, 1995). A sensitivity value is the partial derivative of the network's probability of interest with respect to a conditional probability under study: it is a first-order *approximation* of the effect of deviation from this probability's assessment. Compared to straightforward variation of conditional probabilities, Laskey's method requires considerably less computational effort. The method, however, provides insight in the effect of small deviations from an assessment only: as Laskey indicates, when larger deviations are considered, the quality of the approximation may break down rapidly. As the assessments for a network's conditional probabilities are often quite inaccurate, we feel that *exact* sensitivity analysis of a Bayesian belief network is preferred to approximate analysis.

In this paper, we show that the graphical independence structure of a Bayesian belief network induces various properties that allow for reducing the computational burden of exact sensitivity analysis. We show that, by inspection of a network's structure, conditional probabilities can be identified that upon variation cannot influence the probability of interest. Analyses with respect to these probabilities can be excluded from the overall analysis as they are uninformative. We further show that analyses that are informative comply with simple mathematical functions relating the network's probability of interest to the conditional probabilities under study. Computing the constants in these functions suffices to determine the sensitivity of the probability of interest to the conditional probabilities at hand. These properties with each other provide for a method for exact sensitivity analysis of Bayesian belief networks that requires considerably less computational effort than straightforward variation of probabilities.

The paper is organised as follows. In Section 2, we provide some preliminaries on Bayesian belief networks. In the subsequent sections, we present the properties of sensitivity analysis outlined above: in Section 3 we address identifying uninformative analyses and in Section 4 we discuss the mathematical functions that informative analyses comply with. In our discussions, we focus on *one-way sensitivity analysis* of Bayesian belief networks, in which probabilities are investigated one at a time. The paper ends with some conclusions and directions for further research in Section 5.

2 BELIEF NETWORKS

A *Bayesian belief network* is a concise representation of a joint probability distribution on a set of statistical variables, consisting of a graphical structure and an associated set of probabilities.

The graphical structure of a Bayesian belief network encodes the independences holding among the variables in the represented probability distribution. It takes the form of an *acyclic directed graph* $G = (V(G), A(G))$, where $V(G)$ is a finite set of nodes and $A(G)$ is a set of arcs. Each node V_i in G represents a statistical variable that takes one of a finite set of values. The digraph's set of arcs $A(G)$ models the independences among the represented variables. Informally speaking, we take an arc $V_i \rightarrow V_j$ to represent an influential relationship between the variables V_i and V_j; the arc's direction marks V_j as the effect 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 graphical structure of a Bayesian belief network are numerical quantities from the represented probability distribution. With each node V_i in 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 V_i's (immediate) predecessors $\pi_G(V_i)$ on the probabilities of the values of V_i itself.

We define the semantics of a Bayesian belief network more formally (Pearl, 1988). In doing so, we assign a probabilistic meaning to a network's digraph.

Definition 2.1 *Let $G = (V(G), A(G))$ be an acyclic digraph and let t be a trail in G between the nodes V_i and V_j. We say that t is* blocked *by the set of nodes $Y \subseteq V(G)$, if either V_i or V_j is included in Y, or t contains three consecutive nodes X_1, X_2, X_3 for which one of the following conditions holds:*

- *arcs $X_1 \leftarrow X_2$ and $X_2 \rightarrow X_3$ are on the trail and $X_2 \in Y$;*
- *arcs $X_1 \rightarrow X_2$ and $X_2 \rightarrow X_3$ are on the trail and $X_2 \in Y$;*
- *arcs $X_1 \rightarrow X_2$ and $X_2 \leftarrow X_3$ are on the trail, and $\sigma^*_G(X_2) \cap Y = \varnothing$, where $\sigma^*_G(X_2)$ is the set of nodes composed of X_2 itself and all its descendants.*

Building on the notion of blocking, we define the *d-separation criterion* for sets of trails.

Definition 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 X and Z, denoted $\langle X \mid Y \mid Z \rangle_G^d$, if for each $V_i \in X$ and $V_j \in Z$ every trail from V_i to V_j in G is blocked by Y.*

The d-separation criterion provides for reading independences from a belief network's graphical structure, as indicated in the following theorem.

Theorem 2.3 *Let B be a Bayesian belief network with the digraph $G = (V(G), A(G))$ and the conditional probabilities $p(V_i \mid \pi_G(V_i))$, $V_i \in V(G)$. 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 for all sets $X, Y, Z \subseteq V(G)$, if $\langle X \mid Y \mid Z \rangle_G^d$, then X and Z are conditionally independent given Y in* Pr.

From the previous theorem, we have that a Bayesian belief network provides all information necessary for uniquely defining a joint probability distribution on the variables discerned, that respects the independences portrayed by its graphical structure. A belief network therefore provides for computing any (prior or posterior) probability of interest, for which purpose various algorithms are available (Pearl, 1988; Lauritzen and Spiegelhalter, 1988).

3 UNINFORMATIVE ANALYSES

In a one-way sensitivity analysis of a Bayesian belief network, the sensitivity of the network's probability of interest in essence is investigated with respect to every single conditional probability. By inspection of the network's graphical structure, however, various conditional probabilities can be identified that upon variation cannot influence the probability of interest. We say that the probability of interest is *algebraically independent* of these conditional probabilities; for abbreviation, we will write $p \not\sim q$ to denote that the probabilities p and q are algebraically independent. For a conditional probability of which the probability of interest is algebraically independent, no further analysis is required. The sensitivity analysis of the belief network under study can therefore be restricted to the conditional probabilities of which the probability of interest is algebraically *dependent*. The nodes to which these probabilities refer constitute the *sensitivity set* for the node of interest. We define the concept of sensitivity set more formally.

Definition 3.1 *Let B be a Bayesian belief network with the digraph $G = (V(G), A(G))$. Let $O \subseteq V(G)$ be the set of observed nodes in G and let $V_r \in V(G)$ be the network's node of interest. Now, let G^* be the digraph that is constructed from G by adding to each node $V_i \in V(G)$ an auxiliary predecessor X_i. 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 $\neg\langle \{X_i\} \mid O \mid \{V_r\} \rangle_{G^*}^d$.*

The following proposition now asserts that the probability of interest of a Bayesian belief network indeed is algebraically independent of the conditional probabilities that are specified for any node that is not included in the sensitivity set under study.

Proposition 3.2 *Let B be a Bayesian belief network with the digraph $G = (V(G), A(G))$ and the conditional probabilities $p(V_i \mid \pi_G(V_i))$, $V_i \in V(G)$; 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 be the network's node of interest. Then, for every node $V_i \notin Sen(V_r, O)$, we have that $\Pr(v_r \mid o) \not\sim p(V_i \mid \pi_G(V_i))$, for any value v_r of V_r.*

From the previous proposition, we find that a belief network's probability of interest $\Pr(v_r \mid o)$ is algebraically independent of the conditional probabilities of any node V_i for which one of the following properties holds:

- $V_i \in \pi_G^*(V_r)$ and $\langle (\{V_i\} \cup \pi_G(V_i)) \mid O \mid \{V_r\} \rangle_G^d$;
- $V_i \notin \pi_G^*(V_r)$, $\langle (\{V_i\} \cup \pi_G(V_i)) \mid O \mid \{V_r\} \rangle_G^d$, and $\sigma_G^*(V_i) \cap O \neq \varnothing$;
- $V_i \notin \pi_G^*(V_r)$ and $\sigma_G^*(V_i) \cap O = \varnothing$;

where $\pi_G^*(V_r)$ is used to denote the set of nodes composed of V_r along with all its ancestors and $\sigma_G^*(V_i)$ once more denotes the set of nodes including V_i and all its descendants. Informally speaking, any causal influence originating from an ancestor V_i of the node V_r is shielded from the probability of interest $\Pr(v_r \mid o)$ if both V_i and its immediate predecessors are d-separated from V_r by the set of observed nodes. Under similar conditions, any diagnostic influence from a non-ancestor V_i of V_r is shielded from the probability of interest. A non-ancestor without any observed descendants, to conclude, cannot exert nor pass on any diagnostic influence on the probability of interest. Further details are provided in an extended technical paper (Coupé and Van der Gaag, 1998).

We illustrate the concept of sensitivity set and its use by means of an example.

Example 3.3 We consider a one-way sensitivity analysis of the well-known ALARM-network (Beinlich *et al.*, 1989).

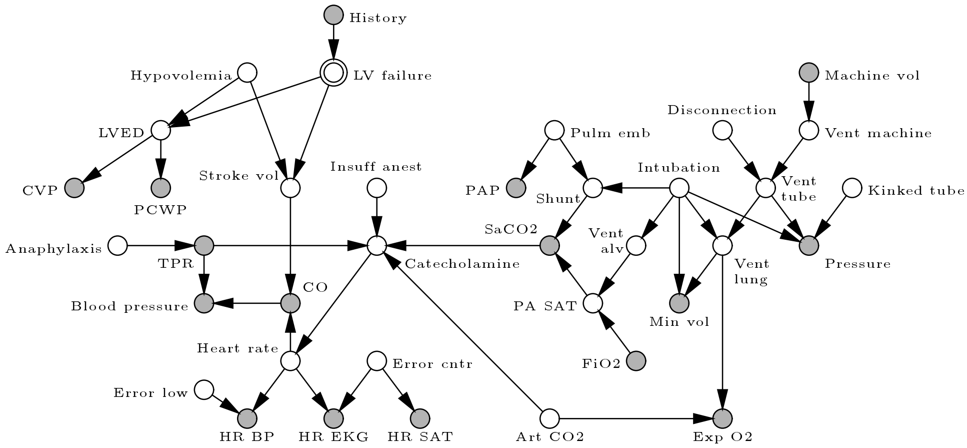

Figure 1: The digraph of the ALARM-belief network.

For ease of reference, the digraph of the network is reproduced in Figure 1; the network's conditional probabilities are not given here, because of space limitations. For our probability of interest, we take the probability that *LV failure* is *true*; the node of interest is indicated by a double circle in the figure. The network's observable nodes are drawn with shading. We now consider the sensitivity set for the node *LV failure* given different sets of observed nodes. If the set of observed nodes is empty, that is, if no observations are available, the sensitivity set for *LV failure* equals

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

Upon performing a one-way sensitivity analysis of the a priori network, therefore, only the conditional probabilities of these two nodes need be investigated. 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 $O_1 = \{History, CVP, TPR, Blood\ pressure, CO\}$. The sensitivity set for *LV failure* given O_1 equals

$$Sen(LV\ failure, O_1) = \{LV\ failure, Hypovolemia, LVED, CVP, Stroke\ vol, CO, Insuff\ anest, Cate-cholamine, Heart\ rate, Art\ CO2, SaCO2, PA\ SAT, FiO2, Vent\ alv, Shunt, Intubation, Pulm\ Emb\}$$

From the 37 nodes included in the network, the conditional probabilities of only 17 nodes need now be investigated. If in addition an observation is assumed for the node *SaCO2*, rendering O_2 for the set of observed nodes, the sensitivity set reduces in size from 17 nodes to 10 nodes:

$$Sen(LV\ failure, O_2) = \{LV\ failure, Hypovolemia, LVED, CVP, Stroke\ vol, CO, Insuff\ anest, Cate-cholamine, Heart\ rate, Art\ CO2\}$$

□

From the previous discussion, we conclude that in a one-way sensitivity analysis of a Bayesian belief network, analyses with respect to the conditional probabilities of nodes that are not included in a sensitivity set under study are uninformative and can therefore be excluded from the overall analysis. Also, analyses with respect to conditional probabilities that are incompatible with any of the observations can be excluded as being uninformative. In recent experiments on randomly generated belief networks, we have found that excluding the indicated analyses may lead to a considerable reduction of the computational burden involved (Coupé and Van der Gaag, 1998).

4 INFORMATIVE ANALYSES

In the previous section, we have shown that the graphical structure of a Bayesian belief network induces algebraic independence of the probability of interest with regard to several of the network's conditional probabilities. For the other conditional probabilities, the graphical structure strongly constrains the shape of the associated sensitivity analyses' curves: the network's probability of interest relates as a *quotient* of two linear functions to a conditional probability under study. We state this property more formally.

Proposition 4.1 *Let B be a Bayesian belief network with the digraph $G = (V(G), A(G))$ and the conditional probabilities $p(V_i \mid \pi_G(V_i))$, $V_i \in V(G)$; 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 be the network's node of interest and let $Sen(V_r, O)$ be the sensitivity set for V_r given O. Then, for every conditional probability x for every node $V_i \in Sen(V_r, O)$, we have that*

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

for any value v_r of V_r, where a, b, c, and d are constants.

Proof (*Sketch*). The probability of interest equals

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

We recall from Section 2 that the joint probability distribution Pr that is defined by the belief network under study, can be written as a product of its conditional probabilities. From the basic property of marginalisation, we further have that both $\Pr(v_r \wedge o)$ and $\Pr(o)$ can be written as a sum of products of conditional probabilities. By separating, in these sums, the terms that specify the conditional probability x and those that do not, it is readily seen that $\Pr(v_r \wedge o)$ as well as $\Pr(o)$ relate linearly to x. In addition, it will be evident that the constants a, b, c, and d in the quotient stated above are built from the network's conditional probabilities. For a more detailed proof, we refer the reader to our extended technical paper (Coupé and Van der Gaag, 1998). □

The mathematical function stated in the previous proposition reduces to a *linear* function for a conditional probability that pertains to a node from the sensitivity set under study that does not have any observed descendants.

Proposition 4.2 *Let B be a Bayesian belief network as before. Let $O \subseteq V(G)$ once again be the set of observed nodes in G, with the corresponding observations o. 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 as before. Then, for every conditional probability x for every node $V_i \in Sen(V_r, O)$ with $\sigma^*(V_i) \cap O = \varnothing$, we have that*

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

for any value v_r of V_r, where a and b are constants.

Proof (*Sketch*). The probability of interest once again equals

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

As we have argued before, the probability $\Pr(v_r \wedge o)$ relates linearly to the conditional probability x under study. We now observe that 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. The reader is referred once more to our extended paper for a detailed proof (Coupé and Van der Gaag, 1998). □

Note that from the previous proposition we have that, if no observations are available, a belief network's probability of interest relates linearly to the conditional probabilities of all nodes in a sensitivity set under study.

We illustrate the basic idea of the two propositions by means of our running example.

Example 4.3 We consider once more a one-way sensitivity analysis of the ALARM-network, taking the probability that *LV failure* is *true* for the probability of interest. In Example 3.3 we have seen that, if no observations are available, the sensitivity set for the node of interest *LV failure* comprises the two nodes *History* and *LV failure* only. As these nodes do not have any observed descendants, we have from Proposition 4.2 that the probability of interest relates linearly to the prior probabilities of *History* and to the conditional probabilities of *LV failure* given *History*. As an example, Figure 2 depicts the probability of interest as a function of the conditional probability $p(LV\ failure = true \mid History = false)$. Now

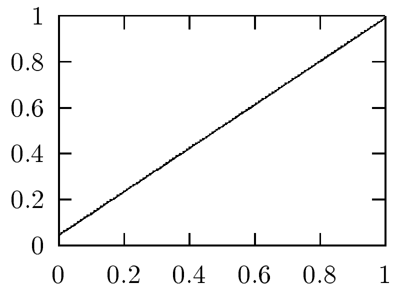

Figure 2: The probability $\Pr(LV\ failure = true)$ as a function of $p(LV\ failure = true \mid History = false)$.

assume that we would like to evaluate the network's probability of interest in view of observations o for the nodes in the set O_2 from Example 3.3. We recall that the sensitivity set for *LV failure* given O_2 includes the nodes *LV failure*, *Hypovolemia*, *LVED*, *CVP*, *Stroke vol*, *CO*, *Insuff anest*, *Catecholamine*, *Heart rate*, and *Art CO2*. We observe that these nodes each have at least one observed descendant. From Proposition 4.1 we now have that the probability of interest relates as a quotient of two linear functions to the conditional probabilities of any of these nodes. As an example, Figure 3 depicts the probability of interest $\Pr(LV\ failure = true \mid o)$ as a function of the prior probability $p(Hypovolemia = true)$. Note that the function is in fact non-linear in the probability under study. To conclude, we would like to note that absence of linearly related conditional probabilities is typically found in Bayesian belief networks for *diagnostic* applications where the nodes of interest are located mainly in the upper part of the network and observable nodes are located in its lower part. □

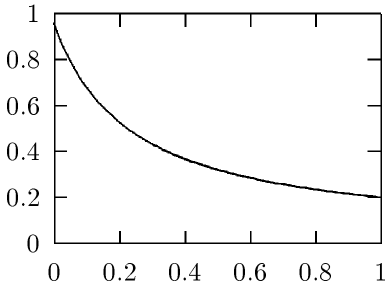

Figure 3: The probability $\Pr(LV\ failure = true \mid o)$ as a function of $p(Hypovolemia = true)$.

Knowledge of the mathematical functions relating a Bayesian belief network's probability of interest to its conditional probabilities allows for considerably reducing the computational burden of a one-way sensitivity analysis of the network. For the conditional probabilities under study, only the constants in these functions need be determined, rendering systematic variation of their assessments unnecessary. 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. For a conditional probability that is related linearly to the probability of interest, two network computations suffice; for all other probabilities, three network computations are required. Alternatively, the constants can be computed directly from the belief network at hand.

5 CONCLUSIONS

Sensitivity analysis of a Bayesian belief network can be performed by systematically varying the assessments for its conditional probabilities. Unfortunately, even for a rather small belief network such a straightforward analysis is too time-consuming to be of any practical use. In this paper, we have shown, however, that the graphical structure of a belief network induces algebraic independence of the network's probability of interest with regard to several of its conditional probabilities. Analyses with respect to these conditional probabilities can be excluded from the overall analysis as they are uninformative. We have further shown that the graphical structure induces simple mathematical functions relating the probability of interest to the conditional probabilities under study. Computing the constants in these functions requires far less computational effort than systematic variation of probabilities.

In this paper, we have focused attention on a one-way sensitivity analysis of a Bayesian belief network, in which the network's conditional probabilities are investigated one at a time. It is also possible to subject a belief network to a two-way sensitivity analysis in which conditional probabilities are investigated pairwise. Such an analysis serves to reveal how two conditional probabilities under study interact to affect the probability of interest. The results that we have presented in this paper are easily generalised to a two-way sensitivity analysis of a Bayesian belief network (Coupé and Van der Gaag, 1998).

In the near future, we envision further investigation of the properties of sensitivity analysis of Bayesian belief networks, 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.

6 BIBLIOGRAPHY

Beinlich I.A., Suermondt H.J., Chavez R.M., Cooper G.F. (1989). The ALARM monitoring system: a case study with two probabilistic inference techniques for belief networks. In: Hunter J., Cookson J., Wyatt J. (Eds). *Proceedings of the Second Conference on Artificial Intelligence in Medicine*, Springer-Verlag, Berlin, pp. 247 – 256.

Coupé V.M.H., Van der Gaag L.C. (1998). Practicable sensitivity analysis of Bayesian belief networks, submitted for publication.

Coupé V.M.H., Van der Gaag L.C., Habbema J.D.F. (1998). Sensitivity analysis: an aid for belief-network quantification, submitted for publication.

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

Laskey K.B. (1995). Sensitivity analysis for probability assessments in Bayesian networks. *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 25, pp. 901 – 909.

Lauritzen S.L., Spiegelhalter D.J. (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.

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

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