

MODELLING EXPERT JUDGEMENT INTO A BAYESIAN BELIEF NETWORK

A METHOD FOR CONSISTENT AND ROBUST
DETERMINATION OF CONDITIONAL PROBABILITY
TABLES

FRIDA HANSSON AND STINA SJÖKVIST

Master's thesis
2013:E37



LUND UNIVERSITY

Faculty of Engineering
Centre for Mathematical Sciences
Mathematical Statistics

TYP AV DOKUMENT <input checked="" type="checkbox"/> Examensarbete <input type="checkbox"/> Delrapport	DOKUMENTBETECKNING LUTFMS—3219—2013
<input type="checkbox"/> Kompendium <input type="checkbox"/> Rapport	

INSTITUTION
Matematikcentrum. Matematisk statistik, Lunds universitet, Box 118, 221 00 LUND

FÖRFATTARE
 Frida Hansson and Stina Sjökvist

DOKUMENTTITEL OCH UNDERTITEL
 Modelling Expert Judgement into a Bayesian Belief Network - A Method for Consistent and Robust Determination of Conditional Probability Tables

SAMMANFATTNING

In project RASTEP (RApid Source TErm Prediction) a computerized tool for real time prediction of source terms at a nuclear power plant is developed. The tool consists of two modules where one is a Bayesian belief network (BBN). A BBN consists of connected nodes and each node has a defined conditional probability table (CPT), which contains the probabilities that a node is in its different states given the states of the node's immediate predecessors. Due to the lack of data the CPTs for some nodes are subjectively determined by experts in the field. Expert judgment may induce uncertainties in the network and it is desirable to know how a relevant and defensible set of conditional probabilities in a BBN can be defined.

This Master Thesis is part of a R&D project run by Scandpower on behalf of the Nordic Nuclear Safety Research (NKS). The aim of the thesis is to develop a general method where experts' beliefs can be included in a systematic way when defining the CPTs in the BBN. The proposed method consists of four parts; Network structure, Probability estimation, Sensitivity analysis and Verification and validation. These parts are performed iteratively until the network is robust and reliable. The main focus of the thesis is on the two parts *Probability estimation* and *Sensitivity analysis*.

From literature different elicitation methods to help the experts assess probabilities in a CPT were found. Two types of elicitation methods were studied; elicitation of a single probability and elicitation of a full CPT. The method preferred when eliciting a single probability was *Probability scale* since it is an easy and straightforward method for the expert to use. Understanding and implementing methods for generating full CPTs required more attention and were tested both on nodes for example networks and for a network developed in RASTEP. The *Likelihood method* showed the best result for elicitation of a full CPT and this method is beneficial to use when the expert is uncomfortable at expressing his beliefs as probabilities.

An important outcome of the work performed, is that rough estimates of the probabilities are sufficient as a first assignment since a sensitivity analysis will reveal which probabilities have significant effect on the network's output and thus need to be more accurately assessed. The sensitivity analysis also shows the constructor of the BBN how observable nodes, given evidence, influence the network and may lead to modifications in the network's structure.

NYCKELORD
DOKUMENTTITEL OCH UNDERTITEL - SVENSK ÖVERSÄTTNING AV UTLÄNDSK ORIGINALTITEL

UTGIVNINGSDATUM	ANTAL SID	SPRÅK
år 2013 mån 6		<input type="checkbox"/> svenska <input checked="" type="checkbox"/> engelska <input type="checkbox"/> annat

ÖVRIGA BIBLIOGRAFISKA UPPGIFTER	ISSN 1404-6342
	ISBN
	2013:E37

I, the undersigned, being the copyright owner of the abstract, hereby grant to all reference source permission to publish and disseminate the abstract.

Signature _____ Date _____

POPULÄRVETENSKAPLIG SAMMANFATTNING

Risken för en olycka i ett svenskt kärnkraftverk är mycket låg, tack vare de höga säkerhetskrav som myndigheterna ställer. Skulle en olycka ändå inträffa kan det få mycket allvarliga konsekvenser för allmänheten och därför måste anläggningar ha beredskapsplaner. Vid en olycka är det viktigt att man snabbt kan skapa sig en uppfattning om olyckans omfattning och möjliga konsekvenser, detta så att räddningspersonal kan vidta nödvändiga åtgärder.

I projekt RASTEP (RAPid Source TErM Prediction) utvecklas ett datoriserat verktyg för att i realtid prediktera potentiella radioaktiva utsläpp i händelse av en olycka. En viktig komponent i verktyget är ett Bayesianskt nätverk (BBN) vars uppgift är att prediktera möjliga tillstånd i ett kärnkraftverk och sannolikheterna för olika utsläpp.

Ett BBN består av noder som representerar uppsättningar av tillstånd samt länkar som representerar orsakssamband mellan noderna. Fördelen med ett BBN är att observerade tillstånd i kärnkraftverket kan ges som insignal till nätverket och ju fler observationer man har desto tillförlitligare blir prediktionen av utsläppssannolikheterna.

För varje nod i ett BBN finns en sannolikhetstabell som ger sannolikheterna för tillstånden i noden givet tillstånd i andra noder. Ansättningen av sannolikhetstabellerna är av stor vikt vid konstruktion av ett nätverk. Detta görs enklast genom att använda data. Om data saknas, vilket ofta är fallet för kärnkraftsanläggningar, görs ansättningarna av experter inom området. Användningen av expertbedömningar kan dock leda till osäkerheter i nätverket.

I examensarbetet utvecklades en iterativ metod som består av fyra delar: Nätverksstruktur, Sannolikhetsskattning, Känslighetsanalys samt Verifiering och validering. Genom att använda denna metod kan relevanta och försvarbara sannolikhetstabeller definieras trots att expertbedömningar används för ansättning av sannolikheterna. Metoden går ut på att först definiera nätverkets struktur och sedan ansätta grova skattningar av sannolikheterna. Därefter utförs en känslighetsanalys som visar hur nätverkets utsignal påverkas av variationer i insignalerna. Känslighetsanalysen identifierar också vilka sannolikheter som har störst påverkan och därför behöver ansättas mer noggrant. Även fel i nätverkets struktur kan upptäckas. Till sist bör nätverket verifieras och valideras, även i detta steg kan felaktigheter i nätverket identifieras. Under utvecklingen av metoden har fokus legat på sannolikhetsskattning och känslighetsanalys.

Om inte data finns tillgänglig vid skattning av sannolikheter kan expertbedömningar inkluderas systematiskt med hjälp av elicitationsmetoder. Några metoder går ut på att experten får ansätta sannolikheterna en och en i tabellerna och det kan ske med hjälp av grafiska hjälpmedel. Det finns även metoder som genererar hela sannolikhetstabeller baserat på olika typer av ansättningar gjorda av experten. Ansättningarna kan t.ex. vara enstaka sannolikheter eller vikter för noder och tillstånd. Dessa metoder är fördelaktiga att använda om experten har svårigheter att ansätta alla sannolikheterna i en tabell.

I en känslighetsanalys studeras hur nätverkets utsignal påverkas av olika observationer och variationer i sannolikhetstabellernas värden. Resultatet av en känslighetsanalys kopplad till observationer kan leda till att strukturen i nätverket behöver modifieras eller att sannolikheter behöver skattas om. Hur känslig en utsignal är för variationer i sannolikheterna kan kvantifieras. De sannolikheter i tabellerna för vilka utsignalen är mest känslig är de sannolikheter som kräver större noggrannhet för att ett tillförlitligt BBN ska erhållas.

Om denna iterativa metod används vid konstruktionen av ett BBN så kan ett robust och pålitligt nätverk uppnås trots att expertbedömningar har använts vid ansättningen av sannolikhetstabellerna.

ABSTRACT

In project RASTEP (Rapid Source TErm Prediction) a computerized tool for real time prediction of source terms at a nuclear power plant is developed. The tool consists of two modules where one is a Bayesian belief network (BBN). A BBN consists of connected nodes and each node has a defined conditional probability table (CPT), which contains the probabilities that a node is in its different states given the states of the node's immediate predecessors. Due to the lack of data the CPTs for some nodes are subjectively determined by experts in the field. Expert judgment may induce uncertainties in the network and it is desirable to know how a relevant and defensible set of conditional probabilities in a BBN can be defined.

This Master Thesis is part of a R&D project run by Scandpower on behalf of the Nordic Nuclear Safety Research (NKS). The aim of the thesis is to develop a general method where experts' beliefs can be included in a systematic way when defining the CPTs in the BBN. The proposed method consists of four parts; Network structure, Probability estimation, Sensitivity analysis and Verification and validation. These parts are performed iteratively until the network is robust and reliable. The main focus of the thesis is on the two parts *Probability estimation* and *Sensitivity analysis*.

From literature different elicitation methods to help the experts assess probabilities in a CPT were found. Two types of elicitation methods were studied; elicitation of a single probability and elicitation of a full CPT. The method preferred when eliciting a single probability was *Probability scale* since it is an easy and straightforward method for the expert to use. Understanding and implementing methods for generating full CPTs required more attention and were tested both on nodes for example networks and for a network developed in RASTEP. The *Likelihood method* showed the best result for elicitation of a full CPT and this method is beneficial to use when the expert is uncomfortable at expressing his beliefs as probabilities.

An important outcome of the work performed, is that rough estimates of the probabilities are sufficient as a first assignment since a sensitivity analysis will reveal which probabilities have significant effect on the network's output and thus need to be more accurately assessed. The sensitivity analysis also shows the constructor of the BBN how observable nodes, given evidence, influence the network and may lead to modifications in the network's structure.

ACKNOWLEDGEMENTS

We would like to take this opportunity to express our thanks to those who helped us with various aspects of writing this thesis. First of all, our supervisor Vidar Hedtj rn Swaling at Scandpower for providing us with an interesting problem and for his feedback and support throughout the work of the thesis. Johan Lindstr m at the Department of Mathematical Statistics at Lund University for very helpful inputs on the report. Michael Knochenhauer at Scandpower for providing us with the background and over-all context of RASTEP. Anders Enerholm at Scandpower for his valuable expert knowledge about nuclear power plants. Last but not least we would like to thank everyone at the office in Malm  for including us in the team and giving us a pleasant experience at Scandpower.

TABLE OF CONTENTS

	<u>Page</u>
1. INTRODUCTION	1
1.1 Aim	1
1.2 Outline of work	1
1.3 Outline of the report	2
2. BACKGROUND	3
2.1 Bayesian Statistics	3
2.2 Subjectivity	3
2.3 Bayesian belief networks	3
2.3.1 d-separation	5
2.3.2 Inference	6
2.3.3 Illustrative example	6
2.4 RASTEP	8
2.4.1 Background and over-all context of the project	8
2.4.2 The Bayesian belief network in RASTEP	9
3. STATING THE PROBLEM	11
4. NETWORK STRUCTURE	12
5. PROBABILITY ESTIMATION	13
5.1 Elicitation of a single probability	13
5.1.1 Probability scale	13
5.1.2 Gamble-like methods	14
5.1.3 Probability wheel	15
5.2 Elicitation of a full conditional probability table	15
5.2.1 Likelihood method	15
5.2.2 EBBN method	17
5.2.3 Weighted sum algorithm	20
5.3 More examples of elicitation of a conditional probability table	21
5.4 Discussion	24
6. SENSITIVITY ANALYSIS	25
6.1 Sensitivity to findings	25
6.2 Sensitivity to parameters	25
6.2.1 Sensitivity set	26
6.2.2 Sensitivity functions	27
6.2.3 Sensitivity value	28
6.2.4 Vertex proximity	30
6.3 An illustrative example of sensitivity analysis	30
6.3.1 Problem set up	31
6.3.2 Sensitivity to findings	31
6.3.3 Sensitivity to parameters	32
6.3.4 Discussion	36
7. VERIFICATION AND VALIDATION	38
8. A BBN IN RASTEP	39
8.1 Probability estimation	39
8.1.1 Elicitation of a single probability	39
8.1.2 Elicitation of a full CPT	42

8.2	Sensitivity analysis	44
8.2.1	Implementation	44
8.2.2	Result and conclusions	45
9.	PROPOSED METHOD	48
9.1	Network structure	48
9.2	Probability estimation	49
9.3	Sensitivity analysis	49
9.4	Verification and validation	50
10.	DISCUSSION	51
11.	FURTHER WORK	52
12.	REFERENCES	53
13.	APPENDIX	55
	Appendix A: Scripts for elicitation methods	55
	Appendix B: Scripts for sensitivity analysis	60

List of notation

X_i	node i in a BBN
x_i^j	state j in node X_i
X_c	a child node
X_p	a parent node
$pa(X_c)$	the set of parent nodes of X_c
a	a combination of states of the parents of a node
m_{abs}	mean of the absolute difference between an original CPT and a CPT generated with any of the elicitation methods.
X_r	node of interest
p	output probability of interest
y	parameter under study
SV	sensitivity value
y_v	vertex of the hyperbola branch

List of abbreviations

BBN	Bayesian Belief Network
CPT	Conditional Probability Table
MAAP	Modular Accident Analysis Program
MELCOR	An integrated, engineering-level computer code used to model the progression of postulated accidents in light-water reactors, as well as non-reactor systems.
NKS	Nordic Nuclear Safety Research
NPP	Nuclear Power Plant
O3	Reactor no.3 in Oskarshamn
PSA	Probabilistic Safety Assessment
QA	Quality Assurance
RASTEP	RApid Source TErm Prediction
SSM	Swedish Radiation Safety Authority
STERPS	Source Term Indicator Based on Plant Status

1. INTRODUCTION

Development of tools for use in the fast, online event and accident diagnosis and subsequent source term forecasting at nuclear power plants is increasingly desired by off-site emergency planning and response personnel. In project RASTEP (Rapid Source Term Prediction) such a computerized tool for real time prediction of source terms (expressing crucial properties of radioactive releases) is being developed. The tool consists of two modules, each with a specific aim:

- A Bayesian belief network (BBN) to predict plant states and the probability of different releases. The more information that is fed into the network, the more precise the prediction.
- A source term module that characterizes the amount, chemical composition, release height, and timing of the predicted radioactive release.

A BBN is built up of connected nodes and for each node a conditional probability table (CPT) is defined, which contains the probabilities that a node is in its different states given the states of the parent nodes. The determination of a CPT depends on the characteristics of the node with some nodes requiring special attention since the lack of data requires a subjective determination of the conditional probabilities. Handling those nodes is time consuming and with probabilities based mainly on expert beliefs it is a challenge to achieve a consistent and quality assured assignment of probabilities in CPTs.

1.1 Aim

This Master Thesis is part of a R&D project run by Scandpower on behalf of the Nordic Nuclear Safety Research (NKS). The aim of the thesis is to develop a general method where beliefs can be included in a systematic way to define CPTs in the BBN. This semi-quantitative method should have qualities which enable the BBN constructor to find uncertainties in the network and study the importance of different parts of the BBN along with determining CPTs. Therefore, sensitivity analysis regarding parameters and model structure of the BBN must be included in the method. This analysis should also show how parameters and model structure influence the prediction of source terms.

The intention of the general method is to give a consistent approach of combining the different parts in building a reliable BBN, where the main parts are:

- *Network structure* – provide a graphical representation of the relationship between events.
- *Probability estimation* – define probabilities in the CPTs with a focus on methods for expert elicitation.
- *Sensitivity analysis* – identify how observations and probability uncertainties affect the network's output.
- *Verification and validation* – make sure that the network's output is reliable and makes sense.

This thesis focuses on the *Probability estimation* and *Sensitivity analysis* parts.

1.2 Outline of work

Bayesian statistics and discrete BBNs will first be studied. In addition a BBN developed in the RASTEP project will be examined in order for us to become familiar with the model structure. The next step is to perform a literature study to identify different expert elicitation methods

used for determining CPTs and the application of sensitivity analysis. With previous approaches as a base, a general method will be developed and then implemented on an example network. Furthermore, it will be decided which of the elicitation methods that can be applied on CPTs in a BBN developed within RASTEP. As a last step a sensitivity analysis will be performed on the same network.

1.3 Outline of the report

Section 2 *Background* introduces the theory needed to understand the concept of Bayesian belief networks and gives a brief introduction to project RASTEP. The problem is stated in Section 3. The network structure is briefly discussed in Section 4. In Section 5 different expert elicitation methods are described. The details of the methods in Section 5.2 are primarily aimed for the specially interested reader. Section 6 presents *Sensitivity analysis* and an illustrative example is given in Section 6.3. Verification and validation of a BBN is briefly discussed in Section 7. Section 8 analyses how probability estimation and sensitivity analysis can be applied on a BBN developed in RASTEP. The *Proposed method* is presented in Section 9. The final sections, Section 10 and 11, contain a discussion and recommendations for further work.

References are given according to the Vancouver system. A reference placed in a sentence only applies to that sentence. A reference placed after a sentence is valid from the previous reference in the same section.

2. BACKGROUND

2.1 Bayesian Statistics

In probability theory the definition of conditional probability is

$$P(Y|X) = \frac{P(X \cap Y)}{P(X)},$$

where X and Y are random variables. Using the above definition leads to Bayes' theorem

$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)},$$

which defines the posterior probability, $P(X|Y)$, in terms of the prior probability, $P(X)$, a normalizing factor, $P(Y)$, and the probability of Y given X , $P(Y|X)$. The probability $P(Y|X)$ is also called the likelihood of X given Y .

Two of the basic axioms in probability theory are

1. $0 \leq P(X) \leq 1$
2. $P(X) = 1$ if and only if X is certain. [1]

2.2 Subjectivity

In probability theory there are two distinct views, *objective* and *subjective* probability. Objective probability is usually calculated based on a mathematical model. [2] Subjective probability on the other hand is based on a person's individual assessment of a probability, for example an expert's judgement or belief [3]. When new information is gained, a probability can be modified through Bayes' theorem [2].

2.3 Bayesian belief networks

BBNs are established in a wide variety of domains, including medicine, ecology and engineering, to model cause and effect [4] [5]. This is because of their ability to capture the probabilistic relationship between events and to combine different sources of data. Another advantage of a BBN is that the result is often convincing and conclusive even if the data is uncertain, which is very common in many applications. [6] [7]

A BBN gives a graphical representation of events that occur in reality and is built as a directed acyclic graph, i.e. a directed graph without cycles. Each node in a BBN represents an event described by a number of possible states, which can be either continuous or discrete. A causal relationship between two nodes is represented by a directed arc, leading from the parent node (cause node) to the child node (effect node), see Figure 2.1. The absence of an arc between two nodes represents conditional independence assumptions. [8]

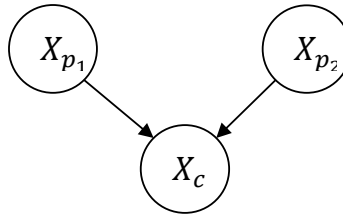


Figure 2.1 Simple Bayesian Network where node X_c has the two parent nodes X_{p_1} and X_{p_2} .

An arc between a parent node X_p and a child node X_c , see Figure 2.1, can be interpreted in different ways and one of the following statements usually holds:

- X_p causes X_c or
- X_p partially causes or predisposes X_c or
- X_c is an imperfect version of X_p or
- X_p and X_c are functionally related or
- X_p and X_c are statistically correlated. [9]

A BBN consists of a qualitative part and a quantitative part and the relationship described by the nodes and arcs is the qualitative part [10]. The strength of influence between a child node and its parent nodes is quantified by conditional probabilities which are represented in a CPT, see Table 2.1. Each probability in a CPT represents the probability of a child node being in a certain state given a set of parent states, e.g. according to Table 2.1

$P(x_c^1 | x_{p_1}^1, x_{p_2}^1) = 0.1$. For nodes that have no parents the table consists of the unconditional probabilities of each state of the node. The probabilities in the CPTs are the quantitative part of the network and they are based upon different types of information, ranging from well-founded theory over frequencies of events in a data base to experts' beliefs. [11]

Table 2.1 A CPT for a child node X_c with the two states x_c^1 and x_c^2 . The node has two parents, X_{p_1} and X_{p_2} , with two states each, $x_{p_1}^1$ and $x_{p_1}^2$ respectively $x_{p_2}^1$ and $x_{p_2}^2$.

		Child node X_c	
Parent node X_{p_1}	Parent node X_{p_2}	x_c^1	x_c^2
$x_{p_1}^1$	$x_{p_2}^1$	0.1	0.9
$x_{p_1}^1$	$x_{p_2}^2$	0.3	0.7
$x_{p_1}^2$	$x_{p_2}^1$	0.8	0.2
$x_{p_1}^2$	$x_{p_2}^2$	0.7	0.3

The following is a more formal definition of a BBN.

Definition 2.1. A Bayesian belief network is a set $B = (G, P)$ where

- G is a directed acyclic graph with nodes $X(G) = \{X_1, \dots, X_n\}$, $n \geq 1$, and arcs $A(G)$;
- P is a set of conditional probabilities $P(X_i | pa(X_i))$, for all $X_i \in X(G)$, where $pa(X_i)$ is the set of immediate predecessors of node X_i . [12]

One valuable property of a BBN is that the product of the conditional probabilities can be used to define the joint probability distribution for all the nodes in a BBN. The joint probability of the nodes X_1, \dots, X_n in a BBN is

$$P(X_1, \dots, X_n) = \prod_{i=1}^n P(X_i | pa(X_i)).$$

Using this product, the conditional probabilities and the structure of the BBN the marginal probability of each node being in one of its states can be determined by marginalisation. [8] The nodes in a BBN can either be observable or hidden. Observable nodes can be updated by the BBN user, i.e. the marginal probability of the observed state is set to 1 and the probabilities of the other states in the node are set to 0. Hidden nodes can not be updated and are entirely based on the states of the other nodes. [13]

2.3.1 d-separation

The d-separation criterion provides a way to determine whether any pair of nodes in a BBN are dependent given a set of observations. Before defining d-separation the concept of blocking has to be introduced.

Definition 2.2. Let $G = (X(G), A(G))$ be an acyclic directed graph with nodes $X(G)$ and arcs $A(G)$, and let s be a chain, i.e. a path in the undirected underlying graph of G , in G between nodes X_i and X_j . We say that s is *blocked* by the set of nodes $O \subseteq X(G)$, if either X_i or X_j is included in O , or s contains three consecutive nodes X_{k-1}, X_k, X_{k+1} for which one of the following conditions holds:

- (1) arcs $X_{k-1} \leftarrow X_k$ and $X_k \rightarrow X_{k+1}$ are on the chain s , and $X_k \in O$;
- (2) arcs $X_{k-1} \rightarrow X_k$ and $X_k \rightarrow X_{k+1}$ are on the chain s , and $X_k \in O$;
- (3) arcs $X_{k-1} \rightarrow X_k$ and $X_k \leftarrow X_{k+1}$ are on the chain s , and $\sigma(X_k) \cap O = \emptyset$, where $\sigma(X_k)$ is the set composed of X_k and its descendants.

Letting O be the set of observed nodes the definition of d-separation between two nodes X_i and X_j is stated as follows.

Definition 2.3. Let G be an acyclic graph and let $X_i, O, X_j \subseteq X(G)$. The set of nodes O is said to *d-separate* the nodes X_i and X_j in G , denoted $\langle X_i | O | X_j \rangle_G^d$, if every chain from X_i to X_j in G is blocked by O . [12]

If two nodes are d-separated they are called structurally independent which means

$$P(X_i, X_j | O) = P(X_i | O)P(X_j | O).$$

The three types of blocking causing a d-separation between two nodes X_i and X_j are illustrated in Figures 2.2-2.4.

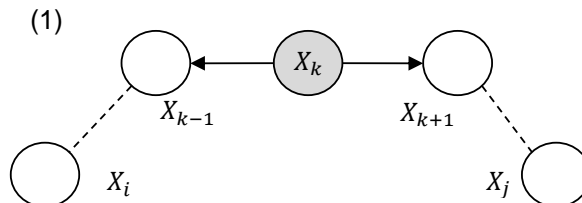


Figure 2.2 An illustration of d-separation. The connection between X_i and X_j is diverging and blocked by the observed node X_k , grey represents an observation.

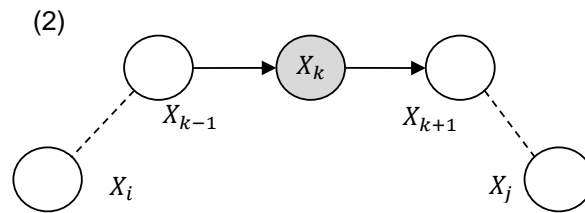


Figure 2.4 An illustration of d-separation. The connection between X_i and X_j is serial and blocked by the observed node X_k , grey representing an observation.

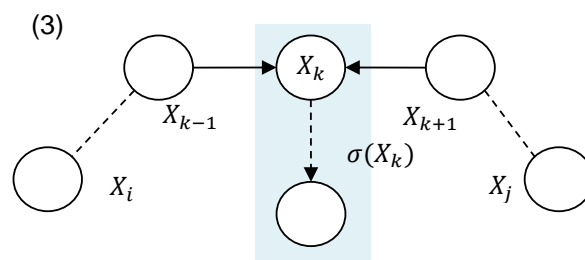


Figure 2.3 An illustration of d-separation. The connection between X_i and X_j is converging and blocked since neither X_k or its descendants are observed (shaded in blue).

2.3.2 Inference

Statistical inference is the calculation of parameters of interest, given a known model and data, and can be characterised as either frequentist inference or Bayesian inference. Frequentist inference calculates the probability as the limit of an event's relative frequency in a large number of trials and can only be done if the experiments are well-defined and random. Bayesian inference is a way to represent probability given evidence or as an individual's degree of belief in a statement. [14]

Applying Bayesian inference on BBNs enables us to compute the probability of each state in a node given observed states in other nodes. When performing inference, beliefs propagate through the BBN, i.e. the probability of each state is updated when evidence is given to some of the nodes. [8] Belief propagation is done automatically in Netica; a program for working with BBNs [15].

2.3.3 Illustrative example

The example *Chest Clinic*, see Figure 2.5, is available in Netica and is a BBN that could be used to diagnose patients arriving at a clinic [15]. In this network each node represents a condition of the patient. The probabilities displayed in Figure 2.5 are the marginal probabilities of each node being in its states given the rest of the network.

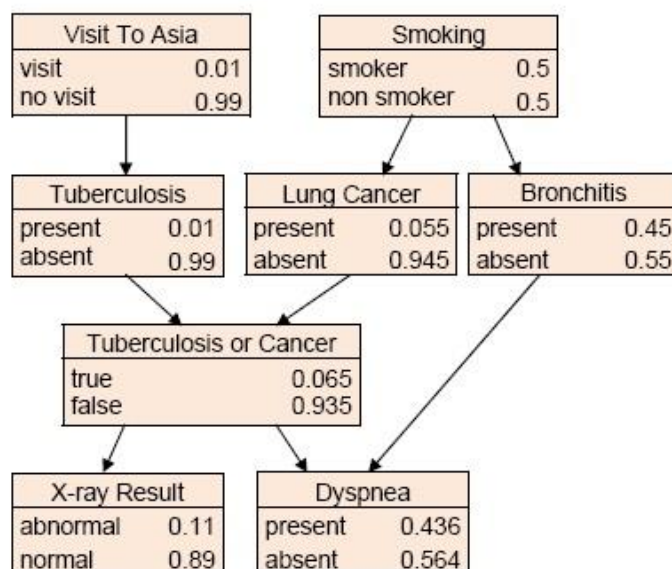


Figure 2.5 The BBN *Chest Clinic*.

All nodes except *Tuberculosis or Cancer*, which is deterministic, are probabilistic and the difference between these two types of nodes can be seen in their CPTs, see Table 2.2 and 2.3. The CPT in Table 2.2 shows the probabilistic node *Dyspnea* where for each parent combination probabilities are assigned to the child states. In the CPT in Table 2.3 the parent combination completely determines the child state.

Table 2.2 CPT of node *Dyspnea*.

		Dyspnea	
Tuberculosis or Cancer	Bronchitis	present	absent
true	present	0.1	0.9
true	absent	0.3	0.7
false	present	0.8	0.2
false	absent	0.7	0.3

Table 2.3 CPT of node *Tuberculosis or Cancer*.

Tuberculosis	Lung Cancer	Tuberculosis or Cancer
present	present	true
present	absent	true
absent	present	true
absent	absent	false

When observables are known, in this case after the patient has been examined, they can be fed into the network as observations and the state probabilities will be updated using Bayesian inference, see Figure 2.6. In this case, the observation of tuberculosis being *present* was given as input to the network and some of the probabilities were updated. The structure of the network determines which probabilities are affected by an observation. It should be noted that the information from observations can propagate in the opposite direction of the arcs which can be seen in the node *Visit to Asia*. This upward propagation is based on Bayes' theorem [16].

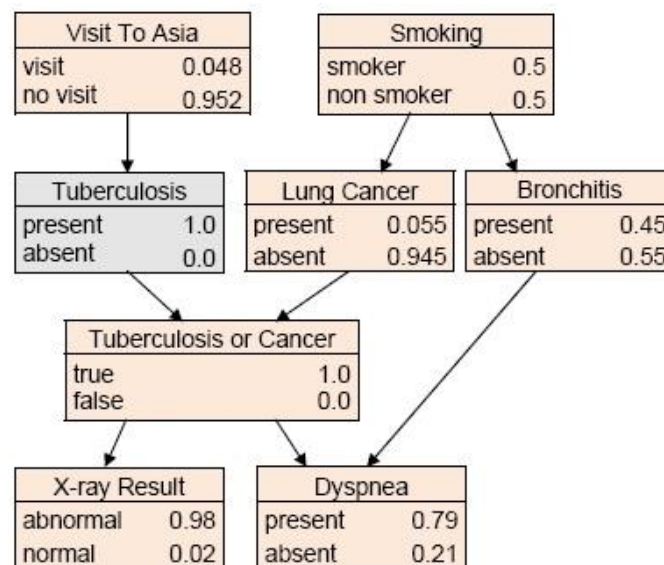


Figure 2.6 *Chest Clinic* with observed node *Tuberculosis*.

2.4 RASTEP

2.4.1 Background and over-all context of the project

This Master Thesis is part of a R&D project run by Scandpower as part of the research programme of NKS, Nordic Nuclear Safety Research (NKS-RASTEP). It is also indirectly related to the previous EU Project STERPS and to a major ongoing project at SSM, the Swedish Radiation Safety Authority (SSM-RASTEP).

Both SSM-RASTEP and NKS-RASTEP take as their starting point the outcome of the EU project STERPS (Source Term Indicator Based on Plant Status). The STERPS project was part of the European Union 5th and 6th Euroatom Framework program, and had the objective to develop for trial use a tool for rapid and early diagnosis of plant status and estimation of likely environmental releases. The EU project showed the feasibility of using BBN technique for modelling of severe accidents, but also identified some issues and challenges related to this. The Swedish contribution to the project (through KTH, Scandpower and OKG) aimed at the development of a first prototype version of a BBN model for the Swedish boiling water reactor Oskarshamn 3 (ASEA-Atom BWR 75).

SSM-RASTEP basically aims at the development of BBN models for all Swedish nuclear power plants, using the basic approach defined by the STERPS project, i.e. a model consisting of two different parts; a BBN model used to predict plant states and release paths and a source term definition part used to characterise the source term (height, composition,

amount and timing). This development is to include the development and documentation of an analysis methodology, including the necessary QA (Quality Assurance) procedures and procedures for validation and verification of developed BBN models, as well as the definition of procedures for update and maintenance of the specific NPP (Nuclear Power Plant) models in RASTEP. During the past four years a basic BBN model (with associated source term definitions) has been developed and largely validated for Oskarshamn 3 (O3). In later stages, further models have been developed for other Swedish plants (generic pressurised water reactor and Oskarshamn 2). The O3 model developed as part of the SSM project is the reference also for NKS-RASTEP.

The basic aim of NKS-RASTEP is to address a number of advanced topics that constitute R&D challenges in the application of BBN to source term predictions during an NPP severe accident. The NKS project has been run in two phases, with phase 1 run in 2011-2012, and the ongoing second phase in 2012-2013. The project has dealt mainly with the following issues:

- Definition of the source terms (ways to improve precision and functionality of the source term module of RASTEP; supported through two M.Sc. theses)
- Comparison of codes for accident sequence and source term calculation (comparison between analysis codes MAAP and MELCOR; supported by one M.Sc. thesis).
- Challenges in BBN structure and quantification (subject of the M.Sc. thesis presented in this report)
 - methods for dealing with sensitivity with respect to parameters and model structure and
 - development of a systematic approach for defining complex CPTs in a BBN.

2.4.2 The Bayesian belief network in RASTEP

The stage when the user should start making use of RASTEP is assumed to be the time of failure of the "first line of defence", i.e. failure of one or more of the systems for fission control, pressure control, core cooling and residual heat removal that are expected to function in case of a normal disturbance.

When creating a RASTEP plant model it is essential to map the plant characteristics. The aim of this task is to give a general understanding of relevant plant characteristics and of systems designed to mitigate severe accidents. Furthermore, key plant parameters and observables to include in the BBN are identified and source terms are defined. The output from RASTEP is a set of possible plant states where each state has an associated source term. The set of plant states are ranked depending on probability and are intended to be used with off-site consequence analysis tools.

One of the modules in the computerized tool RASTEP is a BBN used to model severe accident progression. The BBN predicts plant states and the probability of different environmental source terms.

Nodes in the BBN can be either observable or hidden. Observables are nodes indicating the status of the plant, typically related to pressure and temperature, during a severe accident and can be updated by the user. Hidden nodes are entirely updated by the states of the other nodes in the BBN. In RASTEP an interface will let the user answer questions about the plant observables. Given observations the joint probability of the network is changed and hence the state probabilities are updated. As more questions are answered the accuracy of the predictions increases.

When constructing the BBN in RASTEP the probabilities can be derived from a variety of sources. The CPTs can be based on information from the plant PSA (Probabilistic Safety Assessment), estimates from data or fault tree analyses. When none of these sources are available one has to rely on expert beliefs to define the conditional probabilities. [13]

3. STATING THE PROBLEM

RASTEP is designed to support decisions in case of a severe accident. Thus, it is crucial that the predictions are as credible and reliable as possible, which in turn increases the requirements on validity, robustness, and transparency of the BBN. Therefore it is important to analyse how the choice of model parameters and model structure affects the network's output.

The determination of the CPTs depends on the characteristics of the node. Some nodes require special attention since there is no or very little data to derive the probabilities from and hence the CPTs have to be based on expert judgement. Developing CPTs using experts can be time consuming since the number of probabilities that need to be estimated is usually very large. Furthermore, using expert judgment may induce large uncertainties in the network and it is desirable to know how the uncertainties affect the reliability of the output.

The aim of this thesis is to propose a method for the definition of a relevant and defensible set of conditional probabilities in a BBN that also considers uncertainties in the network.

4. NETWORK STRUCTURE

The qualitative part, i.e. the nodes and arcs, forms the basis of the network and is the first part to be determined when building a BBN. How to set up the network structure will not be discussed in detail in this thesis. However, a way of modifying the structure of the BBN to make the probabilities in CPTs easier to assess is *node divorcing* which is presented briefly below.

The technique of node divorcing reduces the number of combination of parent states for a node by introducing a mediating node. This node becomes the parent of the original child node and the child node of the divorced parents. In Figure 4.1 the nodes X_{p_1} and X_{p_2} have been divorced from the child node, X_c , by the mediating node X_m . If the nodes in Figure 4.1a all have four states the CPT for X_c requires 256 probabilities to be assessed. However, if X_m is introduced, also having four states, the CPTs for both X_m and X_c contain 64 probabilities reducing the total number of probabilities to be assessed to 128. Node divorcing only works for discrete nodes and it is only effective if the number of combinations of states from X_{p_1} and X_{p_2} is greater than the number of states in X_m . [17] [18]

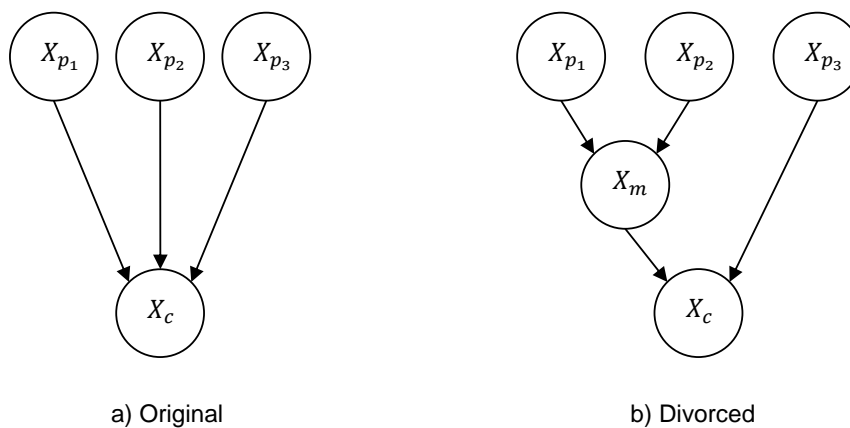


Figure 4.1 Node divorcing. X_m divorces the nodes X_{p_1} and X_{p_2} from X_c .

A property of a BBN is that the immediate successors of a node should be conditionally independent given that node. If the BBN does not satisfy this property it indicates that a hidden node should be introduced. [19] The introduction of a hidden node is the same technique as node divorcing.

5. PROBABILITY ESTIMATION

When the network structure has been set up the quantitative part, i.e. the probabilities in the CPTs, of the network need to be defined. If no data is available for the determination of probabilities, the source of probabilistic information is expert judgement [20]. Using experts' beliefs to assign conditional probabilities is called expert elicitation.

The main steps of the elicitation process are:

- determine what the experts need to elicit,
- select experts,
- expert elicitation, where the experts may use an elicitation method to assign probabilities,
- if there are several experts, combine their assessments,
- document the process and the result. [21]

Due to the structure of a BBN the number of probabilities that populates a CPT grows rapidly with the number of parent nodes related to that CPT, e.g. a child node with i states and N parents each having k states demands $i \cdot k^N$ probabilities to populate its CPT. Thus, the elicitation can be very time consuming.

For an expert it can be particularly difficult to assign probabilities for events that are very rare. Therefore, different methods to help the expert and to systematise the elicitation have been developed. There are two types of elicitation methods; *Elicitation of a single probability* and *Elicitation of a full conditional probability table*. The methods presented below are findings from literature studies. Methods for eliciting single probabilities were more frequently used than elicitation of full CPTs.

5.1 Elicitation of a single probability

This type of elicitation can further be divided into two groups; direct methods and indirect methods. In direct methods experts should give their degree of belief as a number directly, e.g. a probability, whilst for indirect methods the expert will make a decision from which his belief is inferred. [23]

5.1.1 Probability scale

A numerical probability scale is a well-known direct method. It is a horizontal or vertical line showing the probability either in the interval 0-1 or 0-100%. The line is divided into several non-overlapping intervals, where 5-7 is the optimal number of intervals. For each probability that the expert assess he puts a mark, which corresponds to his belief, on the scale and the probability can then be determined by measuring the distance between the mark and the 0.

Since not all experts are familiar with probabilities and are more comfortable expressing their beliefs with words there are scales with both verbal and numerical intervals, see Figure 5.1.

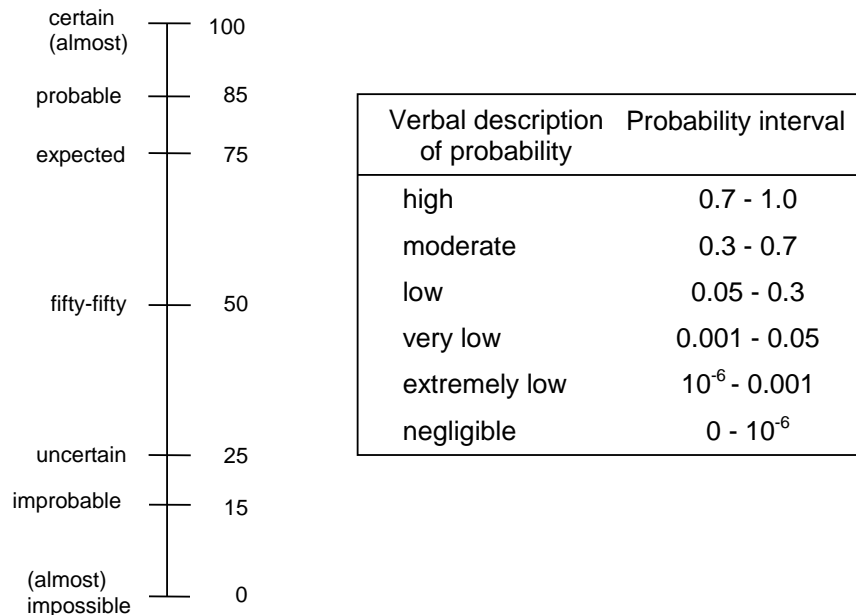


Figure 5.1 Verbal and numerical probability scale, represented as a scale and as a table.

The use of probability scale is advantageous since it is easy for the expert to express his beliefs in a fast and distinct way. However, when using a probability scale the assessments are prone to have scaling biases, such as centring and spacing effects. Centring is the tendency of experts to use the centre of the scale and spacing is the tendency of experts to divide their responses more or less evenly over the scale. [23] [24] [25]

5.1.2 Gamble-like methods

When an expert finds it hard to express his belief of an event as a number a gamble-like method can be used. The basic idea of gamble-like methods is to infer the expert's assessment of a probability from his behaviour in a controlled situation. Gamble-like methods are indirect methods that are designed to represent controlled situations. Two types of gamble-like methods can be distinguished.

In the *certain-equivalent* gamble an expert chooses between an exact reward, x , and a probabilistic alternative where the reward, r , depends on the probability, p , to be assessed. The exact reward is then altered until the expert is indifferent between the two choices. The probability, p , can be calculated as $p=x/r$.

In the *lottery-equivalent* gamble the expert chooses between a lottery where the outcome depends on a given probability and a lottery that depends on the probability, p , to be assessed. Altering p until the expert is indifferent between the two lotteries give the value of p .

Using gamble-like methods can suppress biases as centring and spacing but they are instead influenced by risk attitudes, especially certain-equivalent gamble. Two other drawbacks of gamble-like methods are that they are complicated for an expert to learn and are very time-consuming. [23] [24]

5.1.3 Probability wheel

The probability wheel is an indirect method that is not influenced by risk attitudes. It is usually a circle divided into two sections. The sections are altered until the expert believes he can spin a pointer and the probability that it will stop in a section is equivalent to the probability being assessed. Probability wheel is very similar to a direct method but a drawback is that it cannot elicit small or large probabilities. [23] [24]

5.2 Elicitation of a full conditional probability table

In this section three methods of elicitation of a full CPT, using a reduced number of assessments, will be discussed. Their aim is to generate a full CPT without the expert assigning each probability individually. In the literature other methods exist with this purpose but the advantage of the *Likelihood method*, the *EBBN method* and the *Weighted sum algorithm* is that they are not restricted to binary nodes.

To illustrate how the methods work in practice; guidelines are created as a result from testing the methods on several nodes. Each method will then be used on a node, *Battery voltage*, from the network *Car Diagnosis 2* available in Netica, see Table 5.1 [15]. As a measure of accuracy the mean, m_{abs} , of the absolute difference between the original CPT given in Netica and the CPT generated by an elicitation method is calculated. The methods are used on additional nodes and the mean will be used to compare the methods.

The approach for testing the methods is that, in the absence of a real expert, the original CPT has been used as the knowledge of an expert. Consequently, if a method requires probabilities as assignments, these have been taken from the original CPT. If weights are required; they are based on the characteristics of the original CPT.

Table 5.1 The original CPT for the node *Battery voltage* from the network *Car Diagnosis 2* in Netica.

		Battery voltage		
Charging system	Battery age	strong	weak	dead
okay	new	0.950	0.040	0.010
okay	old	0.800	0.150	0.050
okay	very old	0.600	0.300	0.100
faulty	new	0.008	0.300	0.692
faulty	old	0.004	0.200	0.796
faulty	very old	0.002	0.100	0.898

5.2.1 Likelihood method

This method takes its starting point in Bayes' theorem, which can be written as

$$P(x_c | x_{p_1}, x_{p_2}, \dots, x_{p_N}) \propto L(x_c | x_{p_1}, x_{p_2}, \dots, x_{p_N}) P(x_c),$$

where $P(x_c | x_{p_1}, x_{p_2}, \dots, x_{p_N})$ is the conditional probability and $L(x_c | x_{p_1}, x_{p_2}, \dots, x_{p_N})$ the likelihood of a child state, x_c , given the states for each of the parent nodes, x_{p_i} . The prior probability for x_c is $P(x_c)$ and in this method $P(x_c)$ will be derived from a typical distribution, T_{x_c} . A typical distribution describes the normal state of affairs and its specific form is not of importance. As an example the typical distribution for a node with three discrete states might

be [0.25 0.50 0.25] but it could also take the form of a continuous distribution. By focusing on the likelihood, instead of the conditional probability, the elicitation procedure becomes easier since the expert does not have to remember the relative frequency of different states of the child node and hence the method is called the likelihood method.

Using the log likelihood, instead of the likelihood, is convenient since the log likelihood covers a smaller range of values. Assuming that the parent nodes are independent the log likelihood can be expressed as

$$\log_b L(x_c | x_{p_1}, x_{p_2}, \dots, x_{p_N}) = \sum_{i=1}^N \gamma_{x_c x_{p_i}}$$

$$\Leftrightarrow$$

$$L(x_c | x_{p_1}, x_{p_2}, \dots, x_{p_N}) = b^{(\gamma_{x_c x_{p_1}} + \gamma_{x_c x_{p_2}} + \dots + \gamma_{x_c x_{p_N}})}$$

and a more convenient variant is to define $\gamma_{x_c x_{p_i}} = \beta_{x_c} \alpha_{x_{p_i}}$.

The elicitation procedure is then accomplished by letting the expert provide the following information:

1. a typical distribution, T_{x_c} ,
2. the base, b ,
3. a weighting factor for each state of the child node, β_{x_c} ,
4. a weighting factor for each state of the parent nodes, $\alpha_{x_{p_i}}$.

Given this information the CPT can be calculated using the equations above. It can be seen that if the parameter $\gamma_{x_c x_{p_i}}$ is positive, the likelihood for that combination of x_c and x_{p_i} will increase, and if the parameter is negative the likelihood will decrease. The value of the likelihood determines how the probability of the child node, $P(x_c | x_{p_1}, x_{p_2}, \dots, x_{p_N})$, shifts in relation to T_{x_c} . The base b can be set so that the values of $\gamma_{x_c x_{p_i}}$ are of desired magnitude. Advantages of the likelihood method are that only one value for each state of each parent node needs to be specified and that the expert is asked to give influence weights instead of probabilities. [26] [27]

Some other advantages of this method are that it works even if a node only has one parent and the algorithm that generates the CPT is easy to implement. If the child node has more than three states this method will be very complex and therefore only guidelines for nodes with two and three states are stated.

Two states

- b A suitable start for b is 10. If a smaller/larger range of the probabilities is desired b should be smaller/larger.
- T_{x_c} If possible, use a typical distribution for the states in the child node. If there is none use a uniform distribution.
- β_{x_c} Let the weights of the child states have opposite signs, e.g. [-1 1].
- $\alpha_{x_{p_i}}$ Start with one of the parents and decide for which of its states each child state should have a high probability. The parent state that results in a high probability of a child state should have the same sign as that child state's weight. A suitable assignment for the first parent is [-1...0...1]. The other parents' weights are then assigned in the same way except that the sign and magnitude of the first parent weights must be considered as well. A reasonable interval for the weights are [-3...3].

Three states

- b A suitable start for b is 10. If a smaller/larger range of the probabilities is desired b should be smaller/larger.
- T_{x_c} If it is known that one of the child states, independent of the combination of parents states, always will have approximately the same probability, first assign T_{x_c} so that this state have the maximum value of the probability for all of the parent combinations. Then distribute the probabilities for the two other states uniformly. If all three states can have high probabilities assign a high probability to one of the states and let the two others be uniformly distributed, e.g. $T_{x_c} = [0.9 \ 0.05 \ 0.05]$.
- β_{x_c} If a weight is equal to 0 it does not shift $P(x_c | x_{p_1}, x_{p_2}, \dots, x_{p_N})$ away from T_{x_c} . Hence, assign the state that was first assigned in T_{x_c} with 0 and then assign the other states with -1 and 1.
- $\alpha_{x_{p_i}}$ Use the same reasoning as for two states but the assignments are only based on the weights of the child states that are not equal to 0.

When using this method to generate the CPT for the node *Battery Voltage* the following assignments were made:

$$\begin{aligned}
 b &= 10 \\
 T_{x_c} &= [1/3 \ 1/3 \ 1/3] \\
 \beta_{x_c} &= [-1 \ 0 \ 1] \\
 \alpha_{x_{p_1}} &= [-1 \ 1] \\
 \alpha_{x_{p_2}} &= [-0.5 \ 0 \ 0.5].
 \end{aligned}$$

These assignments resulted in the CPT in Table 5.2 and the mean, m_{abs} , was equal to 0.054.

Table 5.2 Likelihood method used on the node *Battery Voltage*.

		Battery voltage		
Charging system	Battery age	strong	weak	dead
okay	new	0.968	0.031	0.001
okay	old	0.901	0.090	0.009
okay	very old	0.706	0.223	0.071
faulty	new	0.071	0.223	0.706
faulty	old	0.009	0.090	0.901
faulty	very old	0.001	0.031	0.968

5.2.2 EBBN method

The EBBN method (an elicitation method for BBNs) requires only a limited amount of elicited probabilities from an expert to derive a node's CPT. It uses piecewise linear interpolation, based on the ranks of the parent nodes' states, to determine the CPT. The method requires that the states of the child node, X_c , and the states of the parent nodes, $pa(X_c)$, can be ordered on the form low to high. Before the actual method is described, some definitions are introduced.

X_c	The node whose CPT is to be determined, where x_c^{min} and x_c^{max} are the lowest and highest ordered state of X_c , respectively.
$pa(X_c)$	The set of parent nodes of X_c .
a	An assignment of states of the parents, $pa(X_c)$.
a_{neg}	The assignment in which all the parent nodes are in their most favourable state for low ordered states of X_c .
a_{pos}	The assignment in which all the parent nodes are in their most favourable state for high ordered states of X_c .
a_{neg, v_i^+}	The assignment of $pa(X_c)$ in which $X_{p_i} \in pa(X_c)$ is in its most favourable state for high ordered states of X_c and all $X_q \in pa(X_c) \setminus X_{p_i}$ are in their least favourable state for higher ordered states of X_c .
$S^+(X_p, X_c)$	Represents that $X_p \in pa(X_c)$ has a positive influence on X_c , which means that observing a higher ordered state for X_p does not decrease the likelihood of higher ordered states of X_c , regardless of the states of the other nodes $pa(X_c) \setminus X_p$.
$S^-(X_p, X_c)$	Represents that $X_p \in pa(X_c)$ has a negative influence on X_c , which means that observing a lower ordered state for X_p does not decrease the likelihood of lower ordered states of X_c , regardless of the states of the other nodes $pa(X_c) \setminus X_p$.
I	Is a function of parent states that expresses the positiveness (or negativeness) of the joint influence of $pa(X_c)$ on X_c . The function can take on values in the range 0 to 1, i.e. $I \in (0,1)$ with $I(a_{neg}) = 0$ and $I(a_{pos}) = 1$. The method makes use of two types of influence factors, <i>individual influence factor</i> I_{ind} and <i>joint influence factor</i> I_{joint} . The two factors are determined as follows:

$$I_{ind}(x_{p_i}^j) = \begin{cases} \frac{\text{rank}(x_{p_i}^j) - 1}{\text{rank}(x_{p_i}^{max}) - 1}, & \text{if } S^+(X_p, X_c) \\ \frac{\text{rank}(x_{p_i}^{max}) - \text{rank}(x_{p_i}^j)}{\text{rank}(x_{p_i}^{max}) - 1}, & \text{if } S^-(X_p, X_c) \end{cases}$$

$$I_{joint}(a) = \frac{\sum_{\{i: X_{p_i} \in pa(X_c)\}} I_{ind}(x_{p_i}^j) \cdot (\text{rank}(x_{p_i}^j) - 1)}{\sum_{\{i: X_{p_i} \in pa(X_c)\}} (\text{rank}(x_{p_i}^{max}) - 1)},$$

where the rank of the lowest ordered state is set to 1.

The EBBN method requires that the states of $pa(X_c)$ are ordered such that each of these nodes have either a positive or negative influence on X_c . Before the determination of a CPT can be done an expert must make the following assessments:

1. For each of the parent nodes $X_p \in pa(X_c)$, order the states of X_p such that X_p has either a negative or a positive influence on X_c .
2. For each of the states x_c of X_c :
 - Determine the assignment $pa(X_c) = a_{x_c}$ such that the probability $P(X_c = x_c | a_{x_c})$ is as large as possible.
 - Assess the probabilities $P(X_c | a_{x_c})$.

3. For each of the parent nodes $X_{p_i} \in pa(X_c)$, assess $P(X_c = x_c^{max} | a_{neg, p_i^+})$ and

$$P(X_c = x_c^{min} | a_{neg, p_i^+}).$$

The determination of a CPT can be divided in to two steps. In the first step $P(X_c = x_c)$ is estimated as a function of I_{joint} , $f_{x_c}(I_{joint})$, for each state x_c of X_c . This is done by constructing a piecewise linear function $f_{x_c}: [0,1] \rightarrow [0,1]$ through the points $[I_{joint}(a_{x_c}), P(X_c = x_c | a_{x_c})]$. The estimated probabilities $P(X_c = x_c)$ for all x_c sum up to 1. In the second step the conditional probabilities $P(X_c | pa(X_c) = a)$ of the CPT are derived using

$$P(X_c = x_c^m | pa(X_c) = a) = \sum_{i: X_{p_i} \in pa(X_c)} w_i \cdot \frac{\int_{I_{min}(i,j,a)}^{I_{max}(i,j,a)} f_{x_c^m}(I_{joint}(a)) dI_{joint}(a)}{I_{max}(i,j,a) - I_{min}(i,j,a)}$$

where

$$I_{max}(i,j,a) = \max(I_{ind}(x_{p_i}^j), I_{joint}(a)),$$

$$I_{min}(i,j,a) = \min(I_{ind}(x_{p_i}^j), I_{joint}(a)),$$

and w_i is the weight for each parent $X_{p_i} \in pa(X_c)$ calculated as

$$w_i = \frac{1}{2} \frac{\delta_i^+}{\sum_{n: X_{p_n} \in pa(X_c)} \delta_n^+} + \frac{1}{2} \frac{\delta_i^-}{\sum_{n: X_{p_n} \in pa(X_c)} \delta_n^-}$$

where

$$\delta_i^+ = P(X_c = x_c^{max} | a_{neg, p_i^+}) - P(X_c = x_c^{max} | a_{neg}),$$

$$\delta_i^- = P(X_c = x_c^{min} | a_{neg}) - P(X_c = x_c^{min} | a_{neg, p_i^+}).$$

[20]

The EBBN method only requires the expert to assign as many rows of the CPT as there are child states and one weight for each parent node. This method is based on linear interpolation and therefore it has difficulties to produce a large difference between two adjacent probabilities of a state in the CPT. According to Wisse et.al [20], the EBBN method includes the expression *dominance*, which can be a characteristic of a parent node. Applying dominance comes down to, if the expert is certain of a probability, replacing a probability calculated by the method with the expert's belief. This can be done for all three methods and is therefore not discussed in this section.

Since the EBBN method only requires assignments of probabilities the method only needs two guidelines. First, order the states both for the child node and for the parent nodes. Second, if the child node has few parents, each having few states, consider to let the expert set the parent weights, w , direct instead of calculating them as in the method. This could also be done if the node has many parents and the expert knows how the parent nodes are weighted.

Assignments that were made when using the EBBN method on the node *Battery voltage* were:

$a_{x_c^{strong}} = [\text{okay new}]$

$a_{x_c^{weak}} = [\text{faulty new}]$

$a_{x_c^{dead}} = [\text{faulty very old}]$

$x_c^{max} = \text{dead}$

$x_c^{min} = \text{strong}$

$a_{neg} = [\text{okay new}]$

$a_{neg, p_1^+} = [\text{faulty new}]$

$a_{neg, p_2^+} = [\text{okay very old}]$

$$\begin{aligned}
P(X_c | a_{x_c^{strong}}) &= [0.95 \ 0.04 \ 0.01] \\
P(X_c | a_{x_c^{weak}}) &= [0.008 \ 0.3 \ 0.692] \\
P(X_c | a_{x_c^{dead}}) &= [0.002 \ 0.1 \ 0.898] \\
\left. \begin{aligned}
P(X_c = x_c^{max} | a_{neg}) &= 0.01 \\
P(X_c = x_c^{min} | a_{neg}) &= 0.95 \\
P(X_c = x_c^{max} | a_{neg,p_1^+}) &= 0.692 \\
P(X_c = x_c^{min} | a_{neg,p_1^+}) &= 0.008 \\
P(X_c = x_c^{max} | a_{neg,p_2^+}) &= 0.1 \\
P(X_c = x_c^{min} | a_{neg,p_2^+}) &= 0.6
\end{aligned} \right\} \text{Assignments required for the weights, } w_1 \text{ and } w_2.
\end{aligned}$$

This resulted in the CPT in Table 5.3, where the assigned rows of probabilities have replaced the generated, and the mean, m_{abs} , was 0.085.

Table 5.3 EBBN method used on the node *Battery Voltage*. Bold numbers indicate assigned probabilities.

		Battery voltage		
Charging system	Battery age	strong	weak	dead
okay	new	0.950	0.040	0.010
okay	old	0.600	0.134	0.266
okay	very old	0.197	0.198	0.605
faulty	new	0.008	0.300	0.692
faulty	old	0.004	0.160	0.836
faulty	very old	0.002	0.100	0.898

5.2.3 Weighted sum algorithm

This method consists of an algorithm that estimates the $k_1 \times \dots \times k_N$ conditional probabilities, $P(X_c = x_c^m | X_{p_1} = x_{p_1}^{j_1}, X_{p_2} = x_{p_2}^{j_2}, \dots, X_{p_N} = x_{p_N}^{j_N})$, that populate a CPT. With X_c as the child node with l states and $\{X_{p_i}\}_{i=1}^N$ as the parent nodes with k_i states each the algorithm takes the following form

$$P(x_c^m | x_{p_1}^{j_1}, x_{p_2}^{j_2}, \dots, x_{p_N}^{j_N}) = \sum_{i=1}^N w_i P(x_c^m | \{Comp(X_{p_i} = x_{p_i}^{j_i})\})$$

where $m = 1, 2, \dots, l$ and $j_i = 1, 2, \dots, k_i$. The method requires the expert to elicit two things;

1. the relative weights w_1, \dots, w_N for the parent nodes, where $0 \leq w_i \leq 1$ and $\sum_{i=1}^N w_i = 1$
2. the $k_1 + \dots + k_N$ probability distributions, $P(x_c^m | \{Comp(X_{p_i} = x_{p_i}^{j_i})\})$, over X for compatible parental configurations.

Compatible parental configurations refer to the term $\{Comp(X_{p_i} = x_{p_i}^{j_i})\}$ which has the following definition.

Definition 5.1. The state $X_{p_n} = x_{p_n}^{j_n}$, for the parent X_{p_n} , is compatible with the state $X_{p_i} = x_{p_i}^{j_i}$, if according to the expert's mental model the state $X_{p_n} = x_{p_n}^{j_n}$ is most likely to coexist with the state $X_{p_i} = x_{p_i}^{j_i}$. Then $\{Comp(X_{p_i} = x_{p_i}^{j_i})\}$ denotes the compatible parental configuration

where X_{p_i} is in the state $x_{p_i}^{j_i}$ and the rest of the parents are in states compatible with $X_{p_i} = x_{p_i}^{j_i}$.

Using the weighted sum algorithm will make the number of assessments of a CPT linear instead of exponential. [22]

The weighted sum algorithm is a simple elicitation method that doesn't need any specific guidelines before it is used. The expert only has to assign as many rows of the CPT as there are states in the parent nodes. However, the method is based on the concept of compatible parental configuration, which can be hard for an expert to determine.

The following assignments were made when using this method on the node *Battery Voltage*,

$$w = [0.9 \ 0.1]$$

$$P(x_c^m | \{Comp(X_{p_1} = x_{p_1}^{okay})\}) = P(x_c^m | x_{p_1}^{okay}, x_{p_2}^{new}) = [0.95 \ 0.04 \ 0.01]$$

$$P(x_c^m | \{Comp(X_{p_1} = x_{p_1}^{faulty})\}) = P(x_c^m | x_{p_1}^{faulty}, x_{p_2}^{very\ old}) = [0.002 \ 0.1 \ 0.898]$$

$$P(x_c^m | \{Comp(X_{p_2} = x_{p_2}^{new})\}) = P(x_c^m | x_{p_1}^{okay}, x_{p_2}^{new}) = [0.95 \ 0.04 \ 0.01]$$

$$P(x_c^m | \{Comp(X_{p_2} = x_{p_2}^{old})\}) = P(x_c^m | x_{p_1}^{faulty}, x_{p_2}^{old}) = [0.004 \ 0.2 \ 0.796]$$

$$P(x_c^m | \{Comp(X_{p_2} = x_{p_2}^{very\ old})\}) = P(x_c^m | x_{p_1}^{faulty}, x_{p_2}^{very\ old}) = [0.002 \ 0.1 \ 0.898].$$

The assignments for the weighted sum algorithm gave the CPT in Table 5.4, where the assigned rows of probabilities have replaced the generated, with a mean, m_{abs} , calculated to 0.062.

Table 5.4 Weighted sum algorithm used on the node *Battery Voltage*. Bold numbers indicate assigned probabilities.

		Battery voltage		
Charging system	Battery age	strong	weak	dead
okay	new	0.950	0.040	0.010
okay	old	0.855	0.056	0.089
okay	very old	0.855	0.046	0.099
faulty	new	0.097	0.094	0.809
faulty	old	0.004	0.200	0.796
faulty	very old	0.002	0.100	0.898

5.3 More examples of elicitation of a conditional probability table

The three methods were also used on two other nodes, both found in Netica; *Voltage at plug* from the network *Car Diagnosis 2*, see Table 5.5, and *Cardiac output* from the network *ALARM*, see Table 5.7 [15]. For the node *Voltage at plug* the methods were only used for the upper half of the CPT, i.e. when the parent node *Main fuse* is in state *okay*. This is done because the probabilities in the rows where *Main fuse* is in state *blown* are all 0 or 1 and therefore these distributions are considered to be known. The result from the three methods for each of the nodes can be seen in Table 5.6 and 5.8, where the assigned rows of probabilities have replaced the generated. The mean values, m_{abs} , for all three methods and nodes are presented in Table 5.9.

Table 5.7 The original CPT for the node Cardiac output from the network *ALARM*.

		Cardiac output		
Heart rate	Stroke volume	low	normal	high
low	low	0.98	0.01	0.01
low	normal	0.95	0.04	0.01
low	high	0.30	0.69	0.01
normal	low	0.95	0.04	0.01
normal	normal	0.04	0.95	0.01
normal	high	0.01	0.30	0.69
high	low	0.80	0.19	0.01
high	normal	0.01	0.04	0.95
high	high	0.01	0.01	0.98

Table 5.9 The resulting CPTs from the three methods used on the node *Cardiac output*.
Bold numbers indicate assigned probabilities.

Likelihood			EBBN			Weighted sum		
Low	Normal	High	Low	Normal	High	Low	Normal	High
0.999	0.001	0.000	0.980	0.010	0.010	0.755	0.235	0.010
0.945	0.055	0.000	0.304	0.611	0.085	0.118	0.872	0.010
0.342	0.636	0.022	0.096	0.417	0.487	0.300	0.690	0.010
0.871	0.129	0.000	0.556	0.402	0.042	0.950	0.040	0.010
0.113	0.662	0.225	0.040	0.950	0.010	0.040	0.950	0.010
0.000	0.003	0.997	0.020	0.339	0.641	0.010	0.300	0.690
0.681	0.318	0.001	0.197	0.543	0.260	0.800	0.190	0.010
0.012	0.225	0.763	0.025	0.464	0.511	0.268	0.722	0.010
0.000	0.000	1.000	0.010	0.010	0.980	0.450	0.540	0.010

Table 5.8 The mean of the absolute difference between the node's original CPT and the method's generated CPT.

	Likelihood	EBBN	Weighted sum
Battery voltage	0.054	0.085	0.062
Voltage at plug	0.045	0.157	0.131
Cardiac output	0.082	0.193	0.220

5.4 Discussion

Of the three methods used for elicitation of a single probability the probability scale is considered the best method. This method helps with systemising the elicitation process and the scale can be designed according to the type of probabilities in the network. Although it may be difficult to design a suitable scale for some networks, the other two methods have even greater disadvantages; the gamble-like methods because they are hard to learn for an expert and the probability wheel because of its lack of eliciting high and low probabilities.

A general conclusion for the three methods that elicit full CPTs is that they generate more correct CPTs if the probabilities in each child state do not shift too much between high and low probabilities. Another finding for these methods is that an expert must check to see if there are any obvious errors in the generated CPTs and if there are, change these according to his opinion. This replacement also has to take place if the generated probabilities are not the same as the assigned probabilities. For small CPTs these methods might not shorten the elicitation procedure but they can still be useful if:

- the expert is uncertain of assessing specific probabilities and feels more confident in assessing weights for the child and parent states, then he could use the likelihood method.
- the expert is certain of some probabilities and uncertain of others, then the EBBN method or the weighted sum algorithm can be used.

A last conclusion is that the person that performs the elicitation, with the help of an expert, must be well versed in the methods he chooses to use in the elicitation process.

To determine which of the three methods that is best at generating full CPTs the mean of the absolute difference is compared and each generated probability is compared with the original probability. In Table 5.9 it can be seen that the likelihood method produced the lowest mean for all three nodes, which implies that it is the best one to use. Another advantage of the likelihood method is that it handles situations where probabilities in each child state shift between high and low probabilities better than the other two methods. The EBBN method's inability to handle these situations can clearly be seen when comparing the probabilities in Table 5.1 and 5.3. A comparison between the EBBN method and the weighted sum algorithm shows that the latter gives lower mean. This might be due to the fact that the expert sometimes has to assign more probabilities in the weighted sum algorithm and therefore more generated probabilities have been replaced resulting in a lower mean.

6. SENSITIVITY ANALYSIS

Sensitivity analysis is a technique to investigate how much a mathematical model's inputs, when varied, affects the model's output. In a BBN, the inputs can be seen either as states of observable nodes or as the probabilities in the conditional probability tables and the output is the probability of the state of interest. Consequently, one can distinguish between two different sensitivities:

- sensitivities oriented to findings, i.e. observations,
- sensitivities oriented to network parameters, i.e. the probabilities in the CPTs.

By studying these two kinds of sensitivities in a BBN the constructor of the network will be aware of which observable nodes and parameters that are critical for the network. The two sensitivity analyses are often done separately and when investigating the network's reliability both need to be considered. [28]

6.1 Sensitivity to findings

Based on the concept of d-separation, see Section 2.3.1, it can be determined by only looking at the structure of the network if an observable node, given evidence, has any influence on a node of interest. Together with an expert one can check if this information is consistent with his intuition of how the network should work.

If an observable node, X_o , is not d-separated from the node of interest, X_r , the influence can be quantified through *entropy reduction*. [5] Entropy, $H(X_r)$, is a measure of how scattered the probability mass is between the states in X_r and is calculated as

$$H(X_r) = - \sum_{x_r \in X_r} P(x_r) \log_2 P(x_r)$$

[3]. $H(X_r)$ is the entropy of X_r before any new findings and the effect of an observation in X_o can be measured through entropy reduction, I ,

$$I = H(X_r) - H(X_r|X_o) = \sum_{x_r} \sum_{x_o} \frac{P(x_r, x_o) \log_2 P(x_r, x_o)}{P(x_r)P(x_o)},$$

where $H(X_r|X_o)$ is the entropy of X_r given new findings from X_o [29]. Entropy reduction can be calculated in Netica which also ranks the nodes based on their value of I . By performing this analysis it is possible to identify if a node is either too sensitive or insensitive to other nodes given different sets of observations. The result of the analysis could assure the constructor that the network's structure is correct or highlight errors in the network structure or CPTs. [5] It may also be worth considering the usefulness of nodes with less important inputs, even contemplating if they should be rejected from the BBN [28].

6.2 Sensitivity to parameters

Using experts' beliefs to define probabilities will inevitably result in inaccurate assessments and in a BBN these inaccuracies affect the reliability of the network's output. Sensitivity to parameters can be used to identify how uncertainties in parameters influence the output probabilities, i.e. the output robustness. [12]

If uncertainty in a parameter has large impact on the network this represents a probability where the accuracy is important. Consequently, rough estimates of all probabilities in the CPTs are sufficient as first assignments and sensitivity analysis will reveal probabilities that need to be assigned more accurately. [30]

Depending on the number of inputs varied, a sensitivity analysis to parameters can be either a one-way analysis or an n-way analysis. A one-way analysis implies that only one of the model inputs is varied whilst the others are fixed. In an n-way analysis the effect of varying n inputs is considered. [31]

When performing a one-way sensitivity analysis on the network's parameters one can study how a probability of interest (output value) is affected by varying a parameter. How sensitive an output is to uncertainties in a parameter depends on the certain case, i.e. current states of all observable nodes. [28] This means that in order to make a full scale sensitivity analysis for each output probability of interest; for every possible case the sensitivity is calculated for each parameter in the network. However, this straightforward sensitivity analysis is very time-consuming. One way to reduce the computations is to first ask experts in the field which parameters that are expected to be influential and focus primarily on these parameters. To further limit the computations, for each specific case, only the nodes in the *sensitivity set*, Section 6.2.1, are studied. [10]

For each parameter belonging to a node in the sensitivity set the probability of interest is related as a quotient, called *sensitivity function*, of two linear functions to the parameter under study, Section 6.2.2 [12]. By using the sensitivity function the influence a parameter has on a probability of interest can be quantified by calculating the *sensitivity value* and the *vertex proximity*, Section 6.2.3 and 6.2.4 [31].

6.2.1 Sensitivity set

A sensitivity set consists of those, and only those, nodes where variation in the parameters may influence the network's probability of interest. The definition of a sensitivity set is based on the concept of d-separation, see Section 2.4.1, and is stated below.

Definition 6.1. Let $B = (G, P)$ be a BBN. Let $X_r \in X(G)$ be the network's node of interest and let $O \subseteq X(G)$ be the set of observed nodes. Now, let G^* be the digraph that is constructed from G by adding an auxiliary predecessor Z_i to every node $X_i \in X(G)$. Then, the *sensitivity set* for X_r given O , denoted $Sen(X_r, O)$, is the set of all nodes X_i for which $\neg \langle \{Z_i\} \mid O \mid X_r \rangle_G^d$.

The definition of sensitivity set may need some clarification. The auxiliary predecessor Z_i is used to represent inaccuracy in X_i 's assessment. If Z_i is not d-separated from the node of interest, X_r , then X_i belongs to the sensitivity set for that specific case of observed nodes. [12] Figure 6.1 shows an example when X_1 and X_2 don't belong to the sensitivity set for the node of interest, X_r , since their auxiliary predecessors Z_1 and Z_2 are d-separated from X_r .

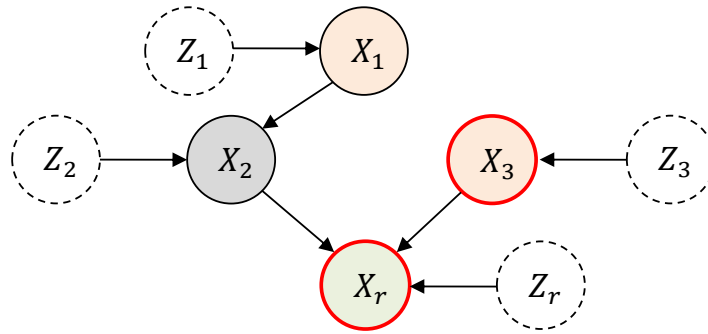


Figure 6.1 A BBN where X_2 is observed (grey), auxiliary predecessors are added to the network and X_r is the node of interest. A red circle represents a node that belongs to the sensitivity set.

6.2.2 Sensitivity functions

A sensitivity function provides a way of studying how variation in a parameter, y , affects a probability of interest, p . The parameter under study is $y = P(x^i | p a(X) = a)$, where x^i is a state of a node X and a is a combination of states of the parents of X . The probability of interest is $p = P(x_r | O)$, where x_r is a specific state of the node of interest and O is the set of observations.

When varying a parameter y the other parameters of the same distribution, $P(x^j | a), j \neq i$, must be co-varied so that the parameters sum to one. Keeping the sum to one is accomplished by, if y with an original value of y_0 is varied to the value y_1 , letting each of the other parameters $P(x^j | a)$ vary as follows

$$\hat{P}(x^j | a) = P(x^j | a) \cdot \frac{1 - y_1}{1 - y_0},$$

where $\hat{P}(x^j | a)$ are the co-varied parameters and this is called proportional co-variation.

If the parameters are proportionally co-varied the sensitivity function, $f_p(y)$, which describes how p varies with y takes the form

$$f_p(y) = \frac{c_1 y + c_2}{c_3 y + c_4} = \frac{d_1 y + d_2}{y + d_3}.$$

[31] The sensitivity function reduces to a *linear* function, i.e. $c_3 = 0$, if the parameter under study belongs to a node that is an ancestor of the node of interest and the parameter's node has no observed descendants. Also if there are no observations; the probability of interest is linearly related to all parameters in the sensitivity set. The constants in a linear sensitivity function can be determined by computing the probability of interest for two different values of the parameter y and then solving a simple linear equation system.

If the sensitivity function is not linear it can be observed as a fragment of a *rectangular hyperbola*, see Figure 6.2. The constants in a hyperbolic sensitivity function can be determined by computing the probability of interest for three different values of y . A rectangular hyperbola takes the general form

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

where for a hyperbolic sensitivity function, the constants are

$$s = -d_3, t = d_1 \text{ and } r = d_2 + s \cdot t.$$

[12]

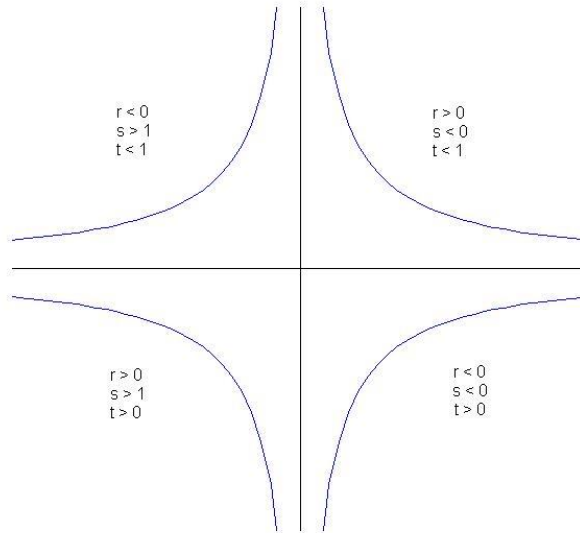


Figure 6.2 A rectangular hyperbola where in each quadrant a sensitivity function is illustrated.

6.2.3 Sensitivity value

The derivative of the sensitivity function can be used as a measure of how infinitesimally small shifts in the parameter under study affects the probability of interest. This measure is called the *sensitivity value* of y and p and it is defined as the absolute value of the first derivative of the sensitivity function at the original value, y_0 , of y .

$$SV = |f'(y_0)| = \frac{d_1 d_3 - d_2}{(y_0 + d_3)^2}$$

is the sensitivity value of a hyperbolic sensitivity function and

$$SV = |f'(y_0)| = \frac{c_1}{c_4}$$

is the sensitivity value of a linear sensitivity function. According to the literature; parameters with a sensitivity value larger than one need further attention and the accuracy of these parameters are of importance for the network [32].

When studying a probability of interest the sensitivity value can be calculated for all the parameters in the sensitivity set for all possible cases. However, this can be very time consuming and thus an upper bound on the sensitivity value can be used to reduce the number of computations. The upper bound for y_0 and the original value of the output probability, p_0 , is

$$|f'(y_0)| \leq \frac{p_0(1-p_0)}{y_0(1-y_0)}$$

and it is based on the so called bounding functions, see [31] for more details. In Figure 6.3 the upper bound as a function of y_0 and p_0 is plotted. Figure 6.3 can be used for identification of which combinations of y_0 and p_0 that could result in high sensitivity values. Hence, it is also possible to identify parameters that will have small sensitivity values for all cases, i.e. all possible values of p_0 . This means that it is not necessary to calculate the actual sensitivity value of these parameters and thus the number of computations is reduced. [31]

In Figure 6.4 the upper bound as a function of y_0 and p_0 can be seen from above and the black parts indicate functional values lower than 1.01. This plot implies that parameters close to 0.5 will have a maximum sensitivity value smaller than 1.01 for all possible output probabilities p_0 . The black shape in the in Figure 6.4 would be even more narrow for a lower limit than 1.01 and a limit lower than one would induce a gap in the middle of the black shape.

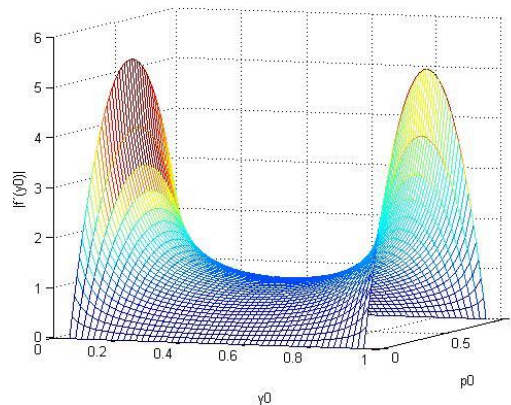


Figure 6.4 The upper bound as a function of x_0 and p_0 plotted for $0.05 < y_0 < 0.95$.

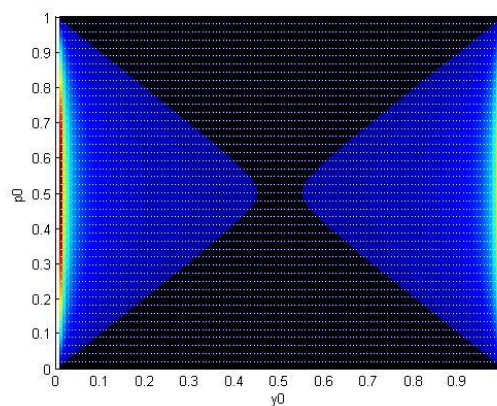


Figure 6.3 The upper bound as a function of y_0 and p_0 seen from above. The black parts indicate functional values lower than 1.01.

If the sensitivity function is linear, the maximum sensitivity value is one [31]. Worth noting is that for a linear sensitivity function, if $SV = 1$, it directly reflects the uncertainty of the parameter in the output. This means that independently of the size of the shift of a parameter the output probability is shifted the same amount, i.e. if the parameter is shifted 0.2 so is the output. For a linear function, a sensitivity value equal to one also implies that either $y_0 = p_0$ or $y_0 = 1 - p_0$.

6.2.4 Vertex proximity

The assessed probabilities in a CPT may be very uncertain and to study only infinitesimally small shifts in the parameters is then not enough. If the relation between the output and the parameter is a linear function the sensitivity value remains the same for larger parameter shifts. But if the sensitivity function is hyperbolic the sensitivity value could change significantly for larger shifts. Thus, another measure, which looks at larger shifts of parameters, is needed to quantify the sensitivity of the output probabilities. A technique for this is to find the y for which the sensitivity value is equal to one and use that as a measure of the shift from large sensitivity values to small ones and vice versa. The point is called the *vertex* of the hyperbola branch under study and can easily be computed from the constants of the sensitivity function using

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

The original value of the parameter, y_0 , and y_v can be compared and the conclusion that can be made depends on the type of hyperbolic function that the sensitivity function represents.

In Figure 6.2 one can see that the absolute value of the derivative of the functions in the second- and third-quadrant, i.e. $s > 1$, is increasing. Which means that if y_0 is smaller than y_v but quite close to it, it can be indicative of possibly significant effects of variation of the parameter to larger values. With the same reasoning, for functions where $s < 0$, and if y_0 is larger than y_v but quite close to it, it can be an indication that, if the parameter is shifted to a smaller value, it affects the output probability significantly. [31]

Further, it is not interesting to calculate the vertex if the sensitivity value is larger than one, since the output is already considered to be sensitive to the value of the parameter.

6.3 An illustrative example of sensitivity analysis

A quite thorough sensitivity analysis is done on a BBN, available through Netica [15], called *Car Diagnosis 2* and the network can be seen in Figure 6.5. This example is meant to show how a sensitivity analysis can be done and the usefulness of the analysis. First, rough estimates are assigned in the CPTs. Throughout the sensitivity analysis some of the estimates are replaced by the probabilities given by the original network, which are assumed to be the correct probabilities. After a replacement, the parameter is excluded from the rest of the sensitivity analysis. This procedure, showing how a sensitivity analysis works, is inspired by Coupé et al [30].

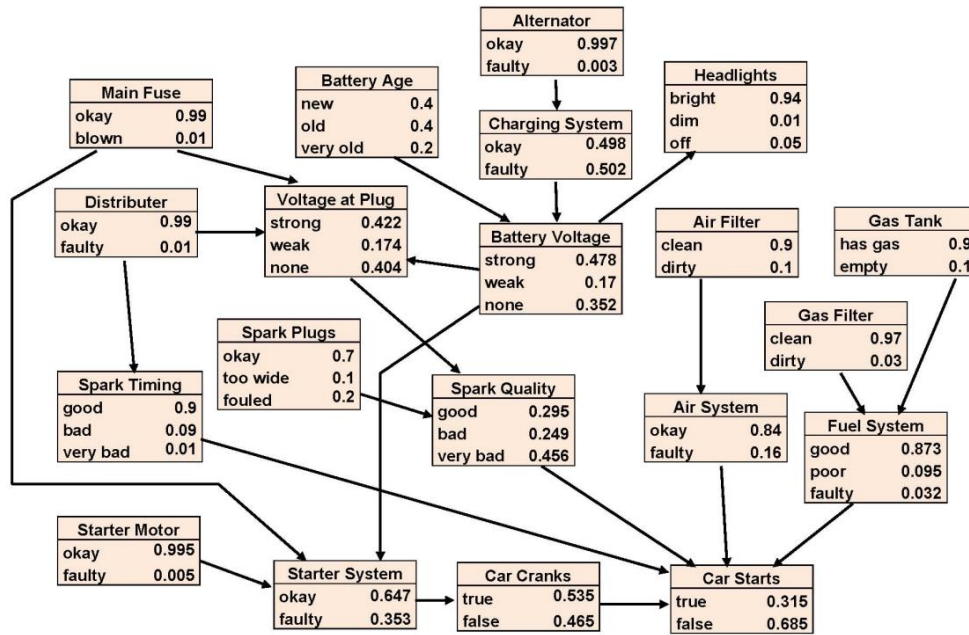


Figure 6.5 The network *Car Diagnosis 2*.

6.3.1 Problem set up

The likelihood method is considered to be the best method of the ones studied in Section 5.2 and it is used to generate rough estimates of the parameters in the two nodes *Voltage at Plug* and *Battery Voltage*. For the rest of the nodes, a random number is generated between 0 and 0.2 for each row in each original CPT. The first parameter in a row is then changed as much as the generated number and the parameter is increased if the original value is below 0.5 and decreased otherwise. Remaining parameters in the row are decreased respectively increased uniformly.

If the original parameters in a row are all either 0 or 1, this is considered to represent a situation when the expert is certain of a probability distribution. These probabilities are kept and the row is excluded from the sensitivity analysis. For the same reason all parameters with a value of 0 are kept and excluded from the analysis.

The state probabilities in three nodes; *Car Starts*, *Starter System* and *Voltage at Plug*, are used as output of interest for this sensitivity analysis. *Car Starts* is chosen on the basis that it's the final output of the network and the main interest. The other two nodes are chosen since they both connect different parts of the network. The selection of the nodes has been inspired by the nodes of interest in the BBN developed within RASTEP.

6.3.2 Sensitivity to findings

Sensitivity to findings is done in Netica and the result, for all three nodes of interest, using no observations is that the *entropy reduction*, see Section 6.1, is equal to zero for the nodes *Headlights* and *Gas Tank*. Studying the CPTs of those nodes and the CPT of the node *Fuel System* confirms that *Headlights* and *Gas Tank* have no impact on the rest of the network. Therefore these two nodes are removed from the network. If an expert is at hand when performing sensitivity to findings a more thorough analysis can be done.

6.3.3 Sensitivity to parameters

At the beginning of this analysis the network consists of 57 rows with estimated parameters making it a total of 127 parameters. If a row has two parameters they will result in the same sensitivity value since the parameters will be varied with the same amount and therefore only half of the parameters in the binary nodes are studied. There is no symmetry in the sensitivity values for nodes with more than two states and all parameters in these nodes need to be included in the sensitivity analysis. This results in a sensitivity analysis where 80 parameters are studied.

The analysis is done in three stages, one for each node of interest, and in each stage five cases with three observed states and one case with no observations are studied. The cases are both randomly generated and chosen with the aim to produce interesting types of sensitivity sets. The node *Spark quality* is excluded from the analysis since it is a deterministic node.

For each case with observed nodes the number of studied probabilities is reduced to the number of parameters in the nodes in the sensitivity set. The number of parameters is further reduced if a node in the sensitivity set has an observed parent since only the parameters belonging to a parental combination with the observed state need to be studied.

In a first stage of sensitivity analysis to parameters the probability of the state *true* in node *Car Starts* is used as an output probability of interest. For each case the sensitivity set is found and the sensitivity values, SV , and vertexes, x_v , are computed. After all six cases have been analysed the maximum sensitivity value of each parameter is found. Very few sensitivity values are larger than one and therefore it is decided that all parameters with $SV \geq 0.3$ should be replaced with the original probabilities. Since the sum of a row in a CPT must sum to one the entire row of a parameter with $SV \geq 0.3$ must be replaced and the total number of changed parameters is 13, see Table 6.1.

Table 6.1 Result of the first stage of sensitivity analysis to parameters.

Number of parameters with $SV \geq 0.3$	7
Number of rows changed	6
Total number of changed parameters	13

In the second stage the state *okay* in *Starter system* was used as output probability of interest. Sensitivity values larger than 0.3 are again considered to represent parameters that should be replaced and another 14 parameters are changed, see Table 6.2.

Table 6.2 Result of the second stage of sensitivity analysis to parameters.

Number of parameters with $SV \geq 0.3$	8
Number of rows changed	6
Total number of changed parameters	14

In the third stage all the three states in the node *Voltage at plug* are used as output probabilities and studied at the same time. Demonstrations of two different cases, Case 1 and Case 2, can be seen in Figures 6.6 and 6.7. In the figures the node of interest, the observations and the sensitivity sets are shown in different colours. The sensitivity function with the output probability $p = P(\text{Voltage at plug} = \text{strong})$ as a function of the parameter $y = P(\text{Voltage at plug} = \text{strong} \mid \text{Main fuse} = \text{okay}, \text{Distributer} = \text{okay}, \text{Battery voltage} = \text{strong})$ is plotted for both cases in Figure 6.8. In Case 1 the sensitivity function is linear since the node *Voltage at plug* has no observed descendants. Since *Car starts* is observed in Case 2 the sensitivity function for the same output and parameter becomes hyperbolic.

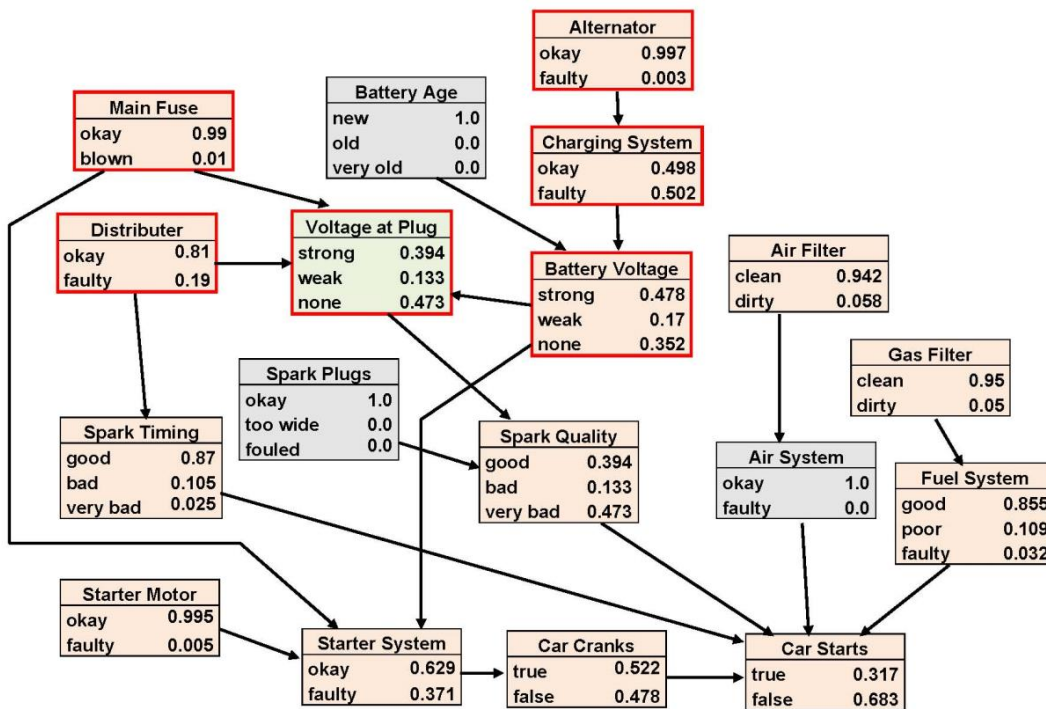


Figure 6.6 The network *Car Diagnosis 2 - Case 1*. Green represents the node of interest, grey represents an observed node and a red border indicates a node that belongs to the sensitivity set.

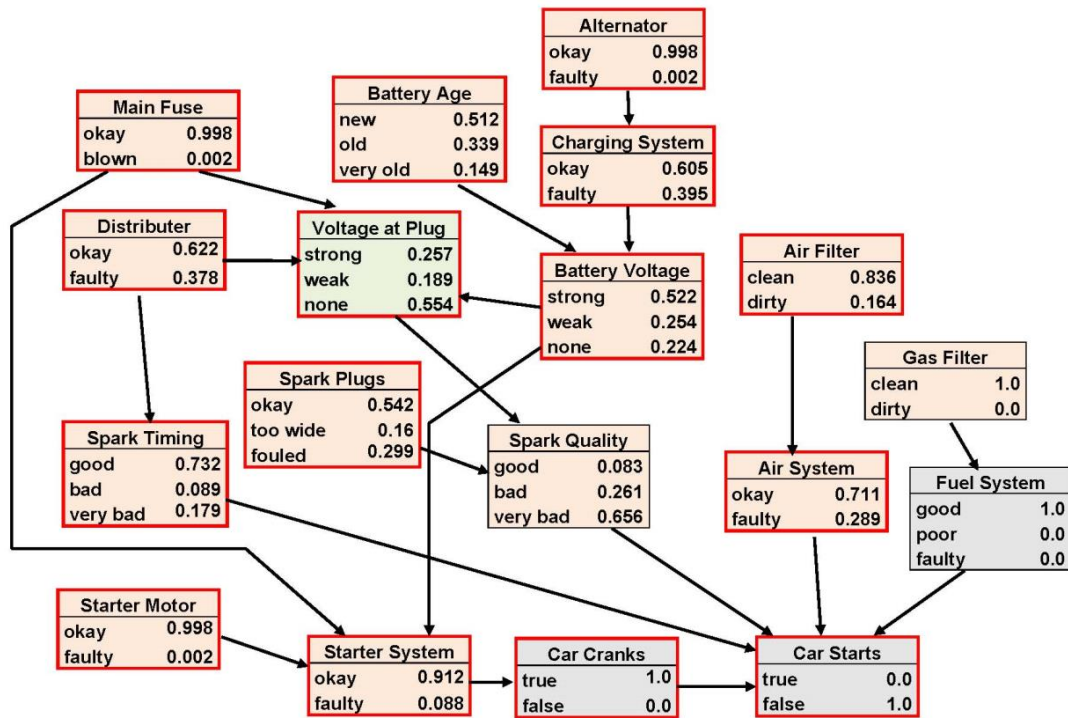


Figure 6.7 The network *Car Diagnosis 2 - Case 2*. Green represents the node of interest, grey represents an observed node and a red border indicates a node that belongs to the sensitivity set.

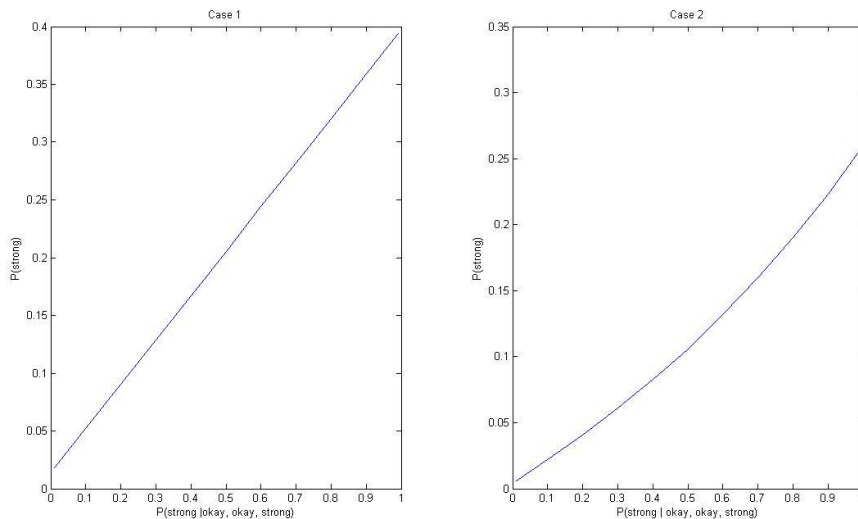


Figure 6.8 The sensitivity function with the output probability $p = P(\text{Voltage at plug} = \text{strong})$ as a function of the parameter $y = P(\text{Voltage at plug} = \text{strong} \mid \text{Main fuse} = \text{okay}, \text{Distributer} = \text{okay}, \text{Battery voltage} = \text{strong})$. Case 1 to the left and Case 2 to the right.

For hyperbolic sensitivity functions the vertexes are calculated, see Section 6.2.4, in order to study what happens if larger parameter shifts are made. Throughout the sensitivity analysis

of this network only a handful calculated vertexes are in the range $0 \leq y_v \leq 1$ which are the only ones of interest since a parameter only can take values in that range. In Case 2 when the parameter $y = P(\text{Distributor} = \text{okay})$ is studied together with the output probability $p = P(\text{Voltage at plug} = \text{none})$ the vertex is $y_v = 0.88$. In Figure 6.9 the hyperbolic sensitivity function can be seen where the red dot is $y_0 = 0.81$ and the green line represents the vertex. In this plot it can be seen that if the original value of the parameter would increase the sensitivity value, i.e. the derivative of the sensitivity function, would be larger and for a parameter value larger than the vertex the sensitivity value becomes larger than 1. The conclusion is that, although $SV = 0.85$ for y_0 , the sensitivity value for this parameter and output probability should be considered to be larger than 1.

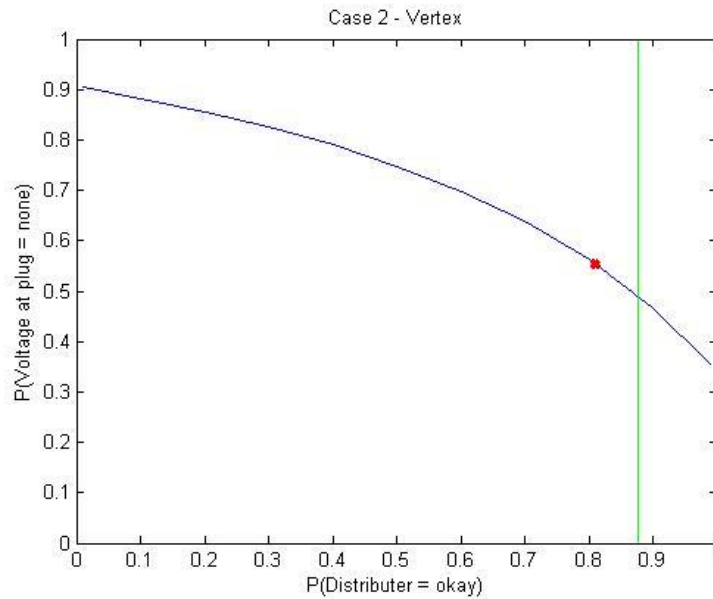


Figure 6.9 The sensitivity function with the output probability $p = P(\text{Voltage at plug} = \text{none})$ as a function of the parameter $y = P(\text{Distributor} = \text{okay})$. Red dot is $y_0 = 0.81$ and the green line represents the vertex $y_v = 0.88$.

The resulting changes after sensitivity analysis with respect to the output probabilities in the node *Voltage at plug* can be seen in Table 6.3. It should be noted that the maximum sensitivity value of all output probabilities is considered.

Table 6.3 Result of the third stage of sensitivity analysis to parameters.

Number of parameters with $SV \geq 0.3$	10
Number of rows changed	6
Total number of changed parameters	14

Table 6.5 shows the resulting output probabilities after each stage. The results after stage three are compared with the original values and one can see that there is still room for improvement. Therefore a fourth stage of sensitivity analysis is carried out with *Car Starts* as node of interest since it is considered to be the most interesting node in this network. Three cases are studied in this stage and as expected the sensitivity values are relatively lower in this stage compared to the other stages. The reason for this is that many parameters have already been changed and are not included in the further analysis. Although only the six parameters with a $SV \geq 0.1$ are replaced, see Table 6.4, the resulting output probabilities are closer to the original, see the result of stage three and four in Table 6.5.

Table 6.4 Result of the fourth stage of sensitivity analysis to parameters.

Number of parameters with $SV \geq 0.1$	6
Number of rows changed	6
Total number of changed parameters	14

A total of 24 rows with estimated parameters are replaced with the original probabilities in the sensitivity analysis. This means that even though more than half of the rows have rough estimates of the parameters the output probabilities are fairly close to the original.

Table 6.5 Output probabilities of the nodes of interest with different amounts of estimated parameters. Values in parenthesis indicates the number of estimated rows in the CPTs of the network.

Output probability	Only rough estimates (57)	Result of stage 1 (51)	Result of stage 2 (45)	Result of stage 3 (39)	Result of stage 4 (33)	Original (0)
$P(\text{Car Starts} = \text{True})$	0.168	0.192	0.229	0.265	0.298	0.315
$P(\text{Starter system} = \text{okay})$	0.490	0.491	0.580	0.580	0.607	0.647
$P(\text{Voltage at plug} = \text{strong})$	0.415	0.358	0.368	0.410	0.420	0.422

6.3.4 Discussion

As mentioned in Section 6.2.3 parameters with a sensitivity value larger than one need further attention. In the sensitivity analysis of the network *Car Diagnosis 2* very few sensitivity values were larger than 1 and therefore a lower limit was used to identify parameters that had significant impact on the output. Another way of determining the parameters that need further attention is to simply replace a specific number of probabilities in each stage. It is difficult to give an exact limit or a specific number since it depends on the network under study and the magnitude of sensitivity values generated. Regardless of the method to determine the parameters whose accuracy is important, the total number of re-estimated

parameters comes down to weigh the time and money to be invested against the benefits of higher accuracy [10].

7. VERIFICATION AND VALIDATION

Before a BBN can be used its ability to model a system and give a reliable output must be evaluated. Errors in some parts of the network can be identified both when the structure is reviewed and during the sensitivity analysis. However, the network should be further verified and validated, preferably with the help of several experts in the field.

The network can be verified by testing if the BBN predictions are in line with results from other available sources of information [13]. Experts can help to verify if the behaviour of the network is as expected when different observations are given as inputs to the network.

During sensitivity analysis parameters with high influence on the network are identified and the probabilities should be more accurately estimated. If for some of these parameters it is only possible to give rough estimates it should be evaluated if it is accepted or if the structure of the network needs to be revised.

8. A BBN IN RASTEP

A BBN for O3, reactor no. 3 in the nuclear power plant in Oskarshamn, has been developed within RASTEP, see Section 2.4. This network contains approximately 90 nodes and it is divided into connected segments, see Figure 8.1. The primary output of the network is a node that contains a set of possible source terms. Another important node has states with possible initiating events whose probabilities are derived from a PSA model.

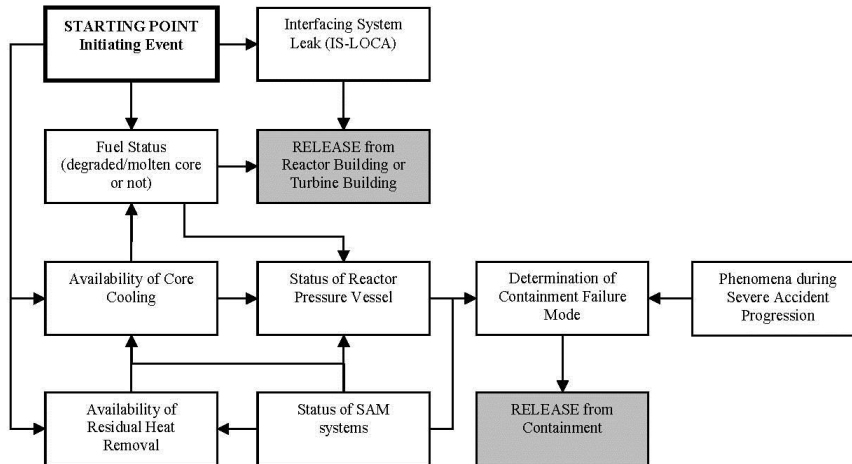


Figure 8.1 Basic structure of the O3 BBN. [13]

Although several nodes in the network are deterministic and many nodes have deterministic parts in their CPTs, there are more than 200 parameters that need to be assigned. Data is not available for all nodes and a large part of these parameters have to be based on expert judgements.

Based on the CPTs in the O3 network, recommendations for the probability estimations using expert judgements in networks like O3 and similar are given in Section 8.1. A simplified sensitivity analysis is done on the developed network and in Section 8.2 it is discussed how to perform and continue the analysis.

8.1 Probability estimation

8.1.1 Elicitation of a single probability

Of the three elicitation methods, described in Section 5.1, probability wheel and gamble-like methods are not suitable for the O3 network. This is because probability wheel can't elicit small and large probabilities which are common in the CPTs in the O3 network and since gamble-like methods are needlessly complicated for this type of elicitation. Even though probability scale can enter bias in the assessments, see Section 5.1.1, it is considered the most suitable method for the O3 network. It is an easy and straightforward method for the expert to use and it can be designed according to the type of probabilities in the CPTs.

Given the probabilities in the CPTs in the O3 network a scale with appropriate intervals is created, see Figure 8.2. The approach the expert should use is to first pick an interval from the table and then mark his belief in the corresponding interval on the scale. If the expert believes that the probability is in the interval "Almost certain" or "Almost impossible" he is asked to mark his belief in a new scale for that specific interval. This is done since these

intervals are small on the original scale and it can therefore be difficult for the expert to mark his assessment. How to determine the probability given the expert's mark on the scale is done in the same way for both the original scale and the subscales, see Section 5.1.1 for more details. If the expert needs to assess very small probabilities the subscales can be enlarged and hence make it easier for the expert to mark his beliefs and for the constructor to determine the probability.

The rows in a CPT must sum to 1 (100%) and this is something the constructor of the BBN must have in mind when the expert assess probabilities. A way to assure that a row sum to 1 is to let the expert assess the probabilities of $k-1$ states of a CPT with k states. The probability of the k th state is then calculated as the difference between 1 and the sum of the probabilities of the other states. However, this probability must be confirmed by the expert. If he disagrees with the result the constructor must ask the expert to rethink the assessments that were made for the probabilities of the other states.

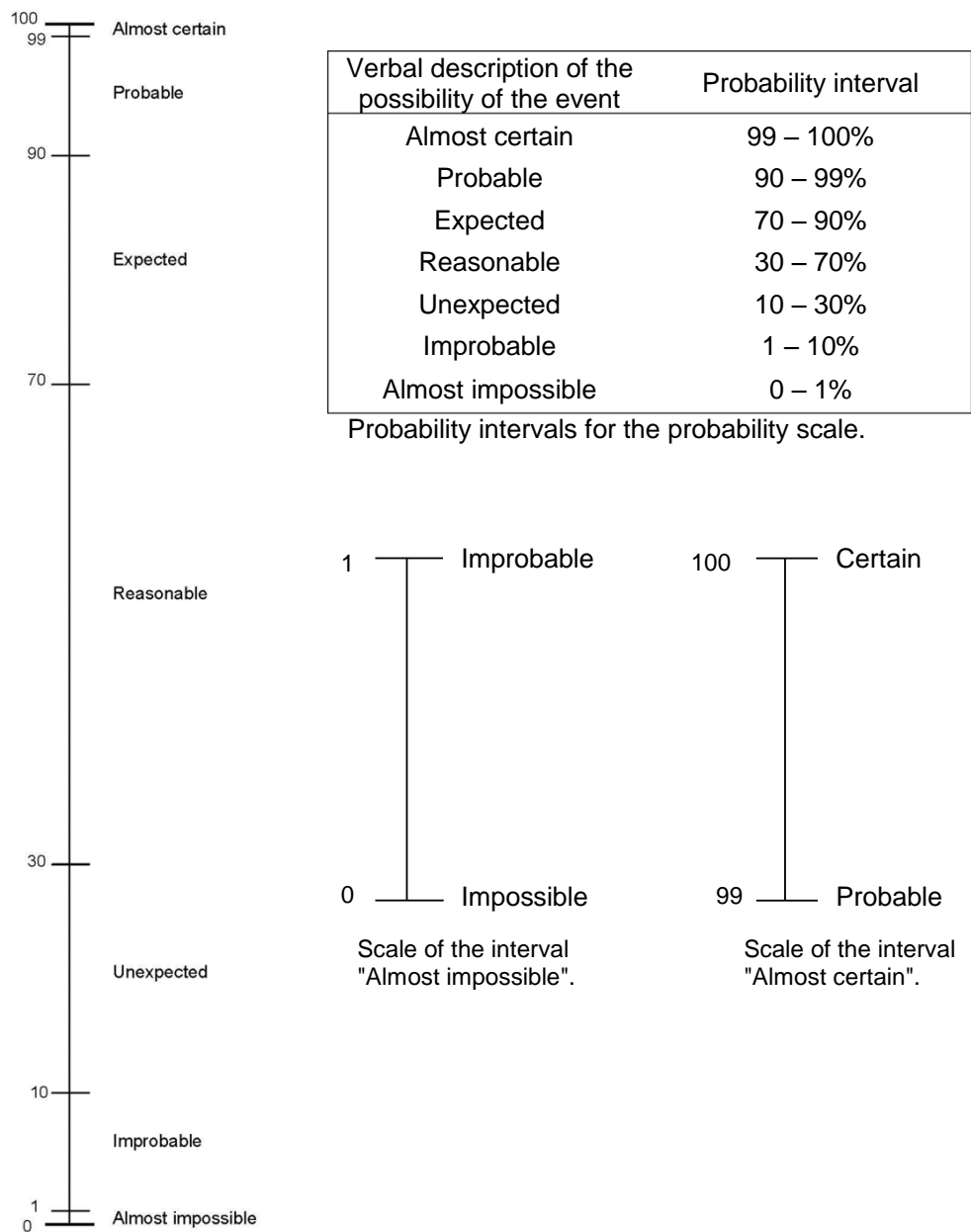


Figure 8.2 Suggested probability scale for the O3 network, table with probability intervals and subscales for the intervals "Almost impossible" and "Almost certain".

8.1.2 Elicitation of a full CPT

By studying the CPTs in the O3 network there are some nodes where the elicitation methods in Section 5.2 may be suitable. Although the CPTs in O3 are relatively small the methods may be useful if the expert is certain of some probabilities and uncertain of others or if the expert doesn't feel confident in assessing specific probabilities.

The likelihood method, see Section 5.2.1, which was considered the best method, is applied on two of the nodes in O3; X_{c_1} and X_{c_2} . Descriptions of the two nodes and their parent nodes are seen in Table 8.1.

Table 8.1 Description of child and parent nodes from O3 used for elicitation of full CPTs.

Node	Description
X_{c_1}	Identification of failure to reclose SRVs after initial depressurisation
X_{c_2}	Indication of the likely fuel status
X_{p_1}	Designated initiating events
X_{p_2}	Indication of core cooling recovery
X_{p_3}	Identification of LOCA size based on indicators

The original CPT of the node X_{c_1} can be seen in Table 8.2 and since the EBBN method and the weighted sum algorithm both require at least two parents they can't be used. The assignments made, according to the guidelines for two child states in Section 5.2.1, and the resulting CPT when applying the likelihood method on the CPT in node X_{c_1} can be seen in Table 8.3. Comparing the probabilities in the original CPT and the generated CPT shows that this method captures the pattern of the CPT and this is foremost due to the weights of the parent node's states in $\alpha_{x_{p_1}}$. Moreover since all generated probabilities are similar to the original the likelihood method works well on these types of CPTs.

Table 8.2 The CPT of the node X_{c_1} in the O3 network.

X_{p_1}	X_{c_1}	
	$x_{c_1}^1$	$x_{c_1}^2$
$x_{p_1}^1$	0.055	0.945
$x_{p_1}^2$	0.055	0.945
$x_{p_1}^3$	0.055	0.945
$x_{p_1}^4$	0.055	0.945
$x_{p_1}^5$	0.055	0.945
$x_{p_1}^6$	0.055	0.945
$x_{p_1}^7$	0.999	0.001

Table 8.3 The resulting CPT of the node X_{c_1} and assignments for the likelihood method.

X_{c_1}		Assignments:
$x_{c_1}^1$	$x_{c_1}^2$	
0.038	0.962	$b = 10$
0.038	0.962	$T_{x_{c_1}} = [0.5 \ 0.5]$
0.038	0.962	$\beta_{x_{c_1}} = [-1 \ 1]$
0.038	0.962	$\alpha_{x_{p_1}} = [0.7 \ 0.7 \ 0.7 \ 0.7 \ 0.7 \ 0.7 \ -1.4]$
0.038	0.962	
0.038	0.962	
0.998	0.002	

For the node X_{c_2} the rows with a parent combination including the state $x_{p_2}^2$ are not estimated since the probabilities in these rows are all 0 or 1 and therefore these distributions are considered to be known, see Table 8.4. The generated CPT and the assignments made, according to the guidelines for three child states in Section 5.2.1, when using the likelihood method on this node can be seen in Table 8.5.

Table 8.4 The CPT of the node X_{c_2} in the O3 network.

		X_{c_2}		
X_{p_2}	X_{p_3}	$x_{c_2}^1$	$x_{c_2}^2$	$x_{c_2}^3$
$x_{p_2}^1$	$x_{p_3}^1$	0.990	0.010	0.000
$x_{p_2}^1$	$x_{p_3}^2$	0.950	0.010	0.040
$x_{p_2}^1$	$x_{p_3}^3$	0.500	0.001	0.499
$x_{p_2}^2$	$x_{p_3}^1$	0.000	0.000	1.000
$x_{p_2}^2$	$x_{p_3}^2$	0.000	0.000	1.000
$x_{p_2}^2$	$x_{p_3}^3$	0.000	0.000	1.000
$x_{p_2}^3$	$x_{p_3}^1$	0.990	0.010	0.000
$x_{p_2}^3$	$x_{p_3}^2$	0.990	0.010	0.000
$x_{p_2}^3$	$x_{p_3}^3$	0.990	0.010	0.000

Since X_{c_2} has two parents the EBBN method can also be used to generate the CPT of the node. The method was applied by using the guidelines in Section 5.2.2 and the assignments with resulting CPT can be seen in Table 8.6.

Table 8.5 The resulting CPT of the node X_{c_2} and assignments for the likelihood method.

X_{c_2}		
$x_{c_2}^1$	$x_{c_2}^2$	$x_{c_2}^3$
0.9580	0.0040	0.0380
0.9040	0.0060	0.0900
0.4950	0.0100	0.4950
0.0000	0.0000	1.0000
0.0000	0.0000	1.0000
0.0000	0.0000	1.0000
0.9996	0.0003	0.0001
0.9992	0.0004	0.0004
0.9948	0.0013	0.0039

Assignments:

$$b = 10$$

$$T_{x_{c_2}} = [0.495 \ 0.01 \ 0.495]$$

$$\beta_{x_{c_2}} = [1 \ 0 \ -1]$$

$$\alpha_{x_{p_2}} = [0.8 \ 2]$$

$$\alpha_{x_{p_3}} = [-0.1 \ -0.3 \ -0.8]$$

Table 8.6 The resulting CPT of the node X_{c_2} and assignments for the EBBN method. Bold numbers indicate assigned probabilities.

X_{c_2}		
$x_{c_2}^1$	$x_{c_2}^2$	$x_{c_2}^3$
0.942	0.009	0.048
0.950	0.010	0.040
0.500	0.001	0.499
0.000	0.000	1.000
0.000	0.000	1.000
0.000	0.000	1.000
0.990	0.010	0.000
0.984	0.010	0.006
0.907	0.009	0.084

Assignments:

$$P(X_{c_2} | a_{x_{c_2}^1}) = [0.99 \ 0.01 \ 0.0]$$

$$P(X_{c_2} | a_{x_{c_2}^2}) = [0.95 \ 0.01 \ 0.04]$$

$$P(X_{c_2} | a_{x_{c_2}^3}) = [0.5 \ 0.001 \ 0.499]$$

$$w = [0.5 \ 0.5]$$

The probabilities generated, by both elicitation methods, for the CPT in X_{c_2} are all roughly the same as the original and hence the likelihood method and the EBBN method works well for this type of CPT.

8.2 Sensitivity analysis

8.2.1 Implementation

For the O3 network, sensitivity to findings requires knowledge about the power plant and is therefore recommended to be done by the constructor of the network together with an expert who can perform a thorough analysis of the network structure and how the nodes interact. However, a short version of sensitivity to parameters is done on the O3 network.

The seven most important output nodes in the O3 network have been given together with four realistic cases of observed states. One of the cases is a design case and the other three are beyond design cases which are used worldwide within risk analysis. All cases are known sequences from PSA (Probabilistic Safety Assessment) executions. It should be noted that realistic accidental cases are difficult to define for a nuclear power plant due to the lack of

data. When defining cases for a nuclear power plant one also has to consider which time point, after failure of one or more systems, the cases should represent.

Despite the size of the network, since it is divided into segments, it is rather easy to determine the sensitivity set, see Section 6.2.1, for each case and output node. Many of the observable nodes are not given findings in these four cases and it is noted that several of these nodes only could be included in a sensitivity set if they are observed. Therefore additional realistic cases are created by adding more findings to the original four cases by the help of an expert. This is done by letting the expert consider the original cases with its specific set of observations and for each case a time point after failure is chosen. Based on the time point and the original case the expert gives his opinion about what state the unobserved nodes should be in. For the nodes of which the expert is certain about their states, additional observations are added. With this information it is possible to create more cases either with all of the observations given by the expert or with just some of them added.

One way of reducing the number of calculations in the sensitivity analysis is to start by excluding parameters with specific probabilities which could be done by studying the upper bound of the sensitivity value for a parameter, see Figure 6.3 and 6.4 in Section 6.2.3. However, since most of the parameters are outside the range of 0.45-0.55 in the O3 network it is not useful for this kind of network.

8.2.2 Result and conclusions

The result from performing a sensitivity analysis with seven output nodes and the four original cases is that 52 parameters out of 211 have sensitivity values larger than zero. Out of these parameters 13 have a maximum sensitivity value larger than or equal to one and the largest value found is 1156. It is concluded based on literature, see Section 6.2.3, that these parameters are considered as parameters where the accuracy is important and it results in eight probability distributions that need further attention.

For a further analysis of the primary output node, five additional cases are created by adding observations to the original cases. The output node with initiating events as states is also studied for two additional cases. In this analysis another four parameters have sensitivity values larger than one and with the same reasoning as above this results in four additional distributions where the accuracy is important.

This short analysis shows that at least twelve probability distributions in the O3 network require a high level of accuracy in order for the network to be robust. If a further analysis were to be done by creating more cases, it could possibly give indications of more parameters with sensitivity values larger than one.

From this analysis it is concluded that the node with different source terms as states is most sensitive to the parameters in nodes with the following description; designated initiating events, availability of system 327, water level in the condensation pool, indication if drywell sprinkling with the system "322 independent" is available and the hydrogen content in the containment. For the node with initiating events as states the values of its own parameters and in the node describing the availability of system 327 have a significant impact on the output.

It is difficult to say if distributions belonging to parameters with SV lower than one should be further studied or not, as mentioned before some of the literature point out $SV > 1$ as limit. But as confirmed in the example in Section 6.3, it is worth to improve the accuracy of parameters with SV lower than 1 as well since as a total it affected the output probability significantly. Nevertheless, it depends on the specific network and the range of the sensitivity

values calculated. In the view of the fact that the O3 network has 211 parameters and less than 10% of them have $SV \geq 1$ it seems reasonable to consider the accuracy of more parameters. If one want to consider 15% of the parameters with the highest sensitivity value it would result in the parameters with $SV \geq 0.7$ which still is a quite large value.

For the O3 network it can be useful to study parameters with low sensitivity values and a probability value close to vertex, see Section 6.2.4. In this analysis of the network this situation appeared a few times and one example, where $SV = 0.3$, can be seen in Figure 8.1. The figure shows the hyperbolic sensitivity function for a parameter with an original value of 0.69 and a vertex with value 0.84. In this specific situation it means that a shift of about 0.15 to a larger parameter value implies that the parameter actually has significant impact on the output probability.

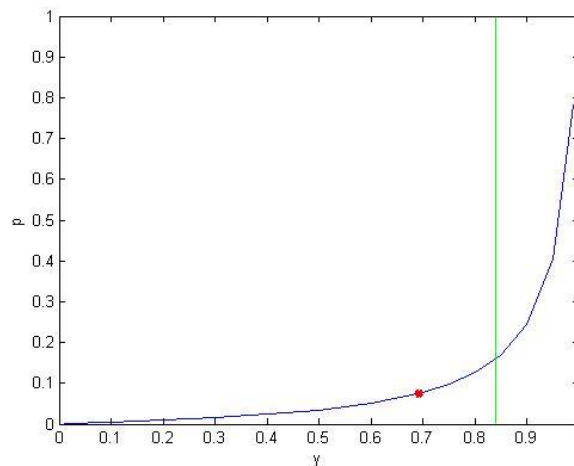


Figure 8.2 The sensitivity function for one of the output probabilities as a function of one of the parameters. Red dot is $y_0 = 0.69$ and the green line represents the vertex $y_v = 0.84$.

It is worth noting that for some parameters and output probabilities in the O3 network, where the sensitivity functions are linear, the sensitivity values are very close to or equal to one. This means that as much as the parameter is shifted the same amount is the output probability shifted, which is an obvious measure of how the parameter affects the output.

A property of many of the used cases is that although some nodes are part of the sensitivity set the sensitivity values of their parameters are zero. This may be due to the fact that some unobserved nodes, particularly deterministic nodes, are largely affected by the observed nodes and the probability of one of the states in the unobserved node becomes very close to or equal to 1. This could prevent changes in the network parameters to propagate through the network to the node of interest and hence the sensitivity value becomes very small or equal to zero. This indicates that more cases with different observations may be needed for a full analysis of the network.

For the cases that have been studied it is more or less the same parameters that have sensitivity values larger than zero for all cases. The exception is nodes that have no child nodes which only are part of the sensitivity set if they are observed. This is another reason that additional cases should be studied in order to find the parameters that have the highest impact on the output probabilities in the O3 network. Exactly how many cases one should use is hard to say and no specific guidelines are found in the literature. In the sensitivity analysis done in [28], where a similar network to O3 is analysed, 117 realistic but randomly chosen cases were used. In [30] however, only five cases were used to get a descent result. The more cases used increase the chance of finding all the most important parts of the

network but it is time consuming. So the conclusion is that the constructor of the network should continue with the analysis until he is confident with the reliability of the network.

For the further analysis of the O3 network it is recommended to only study two of the most important output nodes; the node that contains a set of possible source terms and the node that has states with possible initiating events. The reason for this is that from the sensitivity analysis done it is concluded that the other output nodes do not provide any additional information about the impact of different parameters.

9. PROPOSED METHOD

The proposed method for including experts' beliefs in the CPTs of a BBN can be seen in Figure 9.1. The aim has been to create a method that results in a reliable network even though expert beliefs, which are uncertain sources, have been used. First the structure of the network should be studied to see if it can be modified in order to make the assessments of the probabilities easier. Estimation of probabilities can be divided into three groups and regardless of which method that is used rough estimates of the probabilities are sufficient as a first assignment. Rough estimates are good enough since running a sensitivity analysis on the network will reveal which parameters have large affect on output probabilities of interest and thus need to be more accurately estimated. Sensitivity analysis may also lead to modification of the network's structure. Verification and validation of the network is the last part of the method. The four parts of the method are performed iteratively until the network is robust and reliable.

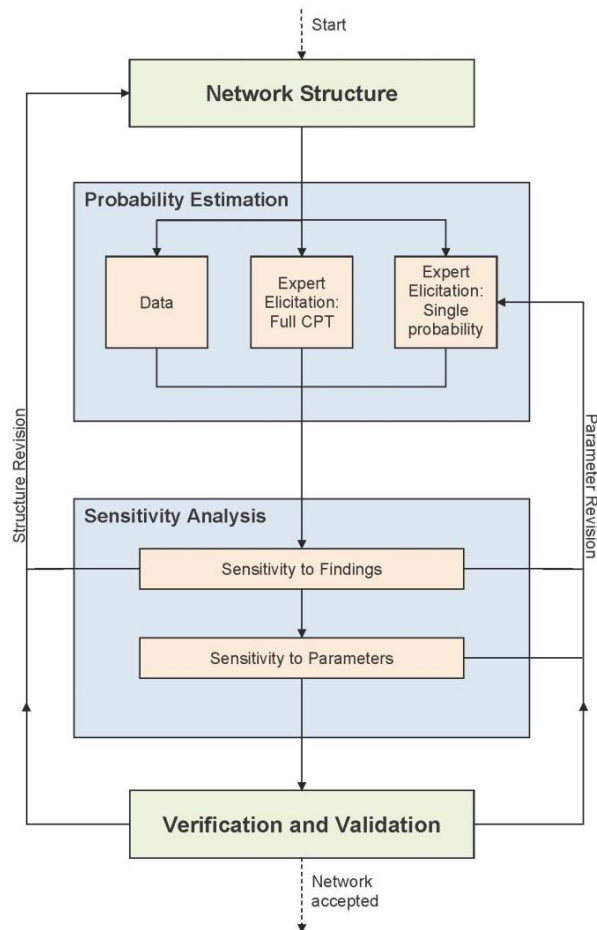


Figure 9.1 Proposed method.

9.1 Network structure

After the structure of a BBN has been constructed it can be studied if some modifications can be made to ease the assignment of probabilities. The network can be modified through the technique of node divorcing, which implies that a node is introduced between a child node and some of its parents, resulting in fewer probabilities to assign, see Section 4.

9.2 Probability estimation

In the second part of the method probabilities in the network's CPTs are estimated. These assignments can be done in three different ways, as is shown in Figure 9.1. The probabilities can be estimated from data or if there are no known data experts' beliefs are used to assess the probabilities.

Assessing probabilities with the help of experts can be done through elicitation and for the proposed method there are two kinds of elicitation procedures; elicitation of a single probability and elicitation of a full CPT. Procedures for elicitation of a single probability are foremost ways of systemising the assessment of probabilities and helping the expert to express his beliefs in probabilities. Single probabilities can be elicited either by letting the expert assign the probabilities directly or by an indirect method where the expert make a decision from which his belief is inferred, see Section 5.1.

Elicitation methods for generating a full CPT are the other kind of elicitation procedures and their purpose is to use fewer assignments to assess a full CPT, thus making it more time efficient. Three different elicitation methods that are not restricted to binary nodes are likelihood method, EBBN method and weighted sum algorithm and have been discussed in Section 5.2. These elicitation methods require different types of assignments. The likelihood method requires the expert to assess weights for the states in the child node of interest and the states of its parent nodes based on a typical probability distribution of the child states. The other two methods require the expert to assess probabilities for some of the rows in the CPT and weights for the parent nodes. The elicitation methods of a full CPT can be used if the expert feels more confident determining these types of different assignments instead of single probabilities although the time of the elicitation procedure won't be significantly reduced.

The choice of elicitation method depends on the structure of the CPT, e.g. how many parent states, and on the expert's ability to assign probabilities or weights. Guidelines for when and how to use each method is found in Section 5.2.1-5.2.3.

9.3 Sensitivity analysis

In this part of the proposed method a sensitivity analysis is performed. This is a technique to study how states of observable nodes and probabilities in CPTs affect the probability of a state of interest. The analysis can be distinguished between two different sensitivities; sensitivity to findings and sensitivity to parameters.

Sensitivity to findings is based on the concept of d-separation, see Section 2.3.1, and can determine if a node, given an observation, has influence on a node of interest. If it has, entropy reduction can be used to quantify the influence. The result of this analysis may require the structure of the network to be reconsidered and imply that single probabilities should be re-estimated, see Section 6.1.

Sensitivity to parameters is used to identify how uncertainties in the parameters, i.e. the probabilities in a CPT, affect an output probability. The influence a parameter has on an output probability can be quantified by using the sensitivity function to calculate the sensitivity value, see Section 6.2. A high sensitivity value indicates that the parameter has a high influence on the output probability and thus it is of importance that the parameter is assessed with high accuracy. The demand of high certainty in the parameters may lead to re-estimation of some parameters after sensitivity to parameters has been performed. After re-estimating parameters a new sensitivity analysis should be performed to see if there are other parameters that result in high sensitivity values. The determination of when the

parameters are sufficiently accurate is a weigh between the time and money to be invested and the benefits of higher accuracy.

By studying these two kinds of sensitivities in a BBN the constructor of the network will be aware of which observable nodes and parameters are critical for the network. The two sensitivity analyses are often done separately but when investigating the network's reliability both should be considered.

9.4 Verification and validation

Before a BBN can be used its ability to model a system and give a reliable output must be evaluated. The network can be verified either by comparison with data or by the help of experts. If the result is not satisfying a revision of either the network structure or the probability estimation may be required.

10. DISCUSSION

The iterative method that is proposed has the ability to produce a relevant and defensible set of conditional probabilities in a BBN even though expert judgement is included in the assessments.

The use of the elicitation methods described, either elicitation of single probabilities or of full CPTs, gives a systematic way to include expert beliefs in the network. If the expert is uncomfortable in assigning specific probabilities a method that generates a full CPT is recommended. The advantage is that it better reflects the expert's knowledge of the causalities in the network and therefore is more defensible to use. A drawback with the full CPT methods is that the complexity of the methods requires the constructor to be well versed in their implementation.

The gain of performing sensitivity analysis, although it may be time consuming, is that in a large part of the parameters rough estimates are acceptable to use. Since sensitivity analysis puts focus on the most essential parts of the network, where accuracy in the probabilities is important, the resulting BBN is reliable although rough estimates are included. If the sensitivity analysis indicates that parameters, for which only rough estimates are possible to assess, have high sensitivity values the constructor has to consider the required reliability of the network. The constructor may accept these rough estimates and otherwise the structure of the network has to be revised.

It has been a challenge to reach a conclusion of which limit for the sensitivity values that determines the parameters that needs to be accurately assessed. Parameters with sensitivity values larger than one have to be re-estimated if possible. However, it depends on the network under study if parameters with lower sensitivity values also should be considered as probabilities with high impact on the network. The number of parameters and the range of the sensitivity values in the network are of great importance in this decision.

The benefit of following this method is that along the way errors in the network can be detected and corrected and in that way prevent the use of an incorrect network. At the same time it can be confirmed that the network is correct and strengthen the reliability of the network.

For the O3 network it is concluded that the elicitation methods for a full CPT may be applied and a probability scale suitable for the O3 network is proposed. Performing sensitivity analysis on the O3 network gives information about which parameters that influence the prediction of source terms and also the ability of the model to predict initiating events. Since only a small percentage of the parameters in O3 have sensitivity values larger than one it seems reasonable to consider the accuracy of more parameters. Further it is concluded that only two of the most important output nodes are needed to be studied. This since performing sensitivity analysis on the other output nodes doesn't give any new valuable information.

11. FURTHER WORK

When investigating a network's reliability both sensitivity to findings and to parameters need to be considered. Sensitivity to findings have not been performed on the O3 network and it needs to be done to reveal how observations in the observable nodes affect the network's output. In the short sensitivity to parameters analysis performed, 10% of the parameters in the O3 network were identified as parameters where the accuracy is important. However, it is recommended to continue the sensitivity analysis of the network with more cases, which may reveal more parameters with significant impact on the network.

For the BBN in RASTEP it should be possible to implement interfaces for the probability estimation part. For the probability scale the interface should allow the expert to make his marks on the scale and then calculate the probabilities. For the elicitation methods of a full CPT an interface lets the expert give his assignments as inputs and it quickly generates the CPT so that the expert can confirm the result.

In the real case scenario the user of the BBN in RASTEP may be uncertain about the observables that are used as inputs to the network. This situation might call for an interface that enables the user to receive information about the effect of the input that he is about to give the network. Another suggestion is to let the user assign a probability for an observation, i.e. his certainty for this observation being true.

It would be valuable to connect an interface to Netica in order to simplify the sensitivity analysis. Given the output of interest the interface would have the ability to calculate the sensitivity values for all parameters in the sensitivity set.

12. REFERENCES

- [1]. Blom, G., Enger, J., Englund, G., Grandell, J. & Holst, L., 2005, *Sannolikhetsteori och statistikteori med tillämpningar*, 5:5 ed, Studentlitteratur.
- [2]. Rychlik, I. & Rydén, J., 2006, *Probability and Risk Analysis: An Introduction for Engineers*, Berlin: Springer.
- [3]. Jensen, F.V., 2001, *Bayesian Networks and Decision Graphs*, New York: Springer.
- [4]. Pourret, O., Naïm, P. & Marcot, B. (Eds.), 2008, *Bayesian networks: a practical guide to applications*, 73, West Sussex: Wiley.
- [5]. Woodberry, O., Nicholson, A.E., Korb, K.B. & Pollino, C., 2005, Parameterising Bayesian Networks: A Case Study in Ecological Risk Assessment, *AI 2004: Advances in Artificial Intelligence*, 1101-1107, Berlin Heidelberg: Springer.
- [6]. Charles River Analytics, 2008, *About Bayesian Belief Networks*, Cambridge: Charles River Analytics, Inc.
- [7]. Langseth, H. & Portinale, L., 2007, Bayesian networks in reliability, *Reliability Engineering and System Safety*, 92, 92-108.
- [8]. Mihajlovic, V. & Petkovic, M., 2001, Dynamic Bayesian Networks: A State of the Art, *Europhysics Letters*.
- [9]. Norsys, 1997, *NETICA Application for Belief Networks and Influence Diagrams- User's guide*, Vancouver: Norsys Software Corp.
- [10]. Coupé, V.H.M., Gaag, L.C. van der & Habbema, J.D.F., 2000, Sensitivity Analysis: an aid for belief-network quantification, *The Knowledge Engineering Review*, 15, 215-232.
- [11]. Jensen, F.V., 1996, *An introduction to Bayesian Networks*, London: UCL Press.
- [12]. Coupé, V.H.M. & Gaag, L.C. van der, 2002, Properties of sensitivity analysis of Bayesian belief networks, *Annals of Mathematics and Artificial Intelligence*, 36, 323-356.
- [13]. Hedtjärn Swaling, V., Frid, W., Knochenhauer, M. & Lundtofte, C., 2011, Using PSA to Develop a Tool for Rapid Source Term Prediction Based on Bayesian Belief Networks, *Nordic PSA Conference "Castle Meeting 2011"*.
- [14]. Statisticat, LLC., 2013, *Bayesian Inference*, Farmington, CT: Statisticat, LLC.
- [15]. NETICA, Bayesian network development software, Norsys Software Corp.
www.norsys.com
- [16]. Pollino, C.A & Henderson, C., 2010, *Technical Report No. 14: Bayesian networks: A guide for their application in natural resource management and policy*, Landscape Logic: Linking land and water management to resource condition targets.
- [17]. Krieg, M.L., 2001, *A Tutorial on Bayesian Belief Networks*, Edinburgh: DSTO Electronics and Surveillance Research Laboratory.

- [18]. Rajabally, E., Sen, P., Whittle, S. & Dalton, J., 2004, Aids to Bayesian Belief Network Construction, *Second IEEE International Conference on Intelligent Systems*.
- [19]. Henrion, M., 1987, Practical issues in constructing a Bayes' belief network, *Proceedings of Third Workshop on Uncertainty in AI*, AAAI, Seattle, Wash, 132-139.
- [20]. Wisse, B. W., Van Gosliga, S. P., Van Elst, N. P. & Barros, A. I., 2008, Relieving the elicitation burden of Bayesian Belief Networks, *Sixth Bayesian Modelling Applications Workshop on UAI*, 26.
- [21]. Cooke, R.M. & Goossens, H.J., 2000, Procedures guide for structured expert judgement in accident consequence modelling, *Radiation Protection Dosimetry*, 90, 303-309, Nuclear Technology Publishing.
- [22]. Das, B., 2004, *Generating conditional probabilities for Bayesian networks: Easing the knowledge acquisition problem*, Edinburgh: DSTO Command and Control Division.
- [23]. Renooij, S., 2001, Probability elicitation for belief networks: issues to consider, *The Knowledge Engineering Review*, 16, 255-269.
- [24]. Burgman, M., Fidler, F., McBride, M., Walshe, T. & Wintle, B., 2006, *ACERA Project 0611: Eliciting Expert Judgment*, Melbourne: Australian Centre for Excellence in Risk Analysis.
- [25]. Gaag, L.C. van der, Renooij, S., Witteman, C.L.M., Aleman, B. & Taal, B., 1999, How to elicit many probabilities, *Uncertainty in artificial intelligence*, 647-654, San Francisco, CA: Kaufman.
- [26]. Kemp-Benedict, E., 2008, *Elicitation Techniques for Bayesian Network Models*, Working Paper WP-US-0804, Stockholm: Stockholm Environment Institute.
- [27]. Kemp-Benedict, E., Bharwani, S., de la Rosa, E., Krittasudthacheewa, C. & Matin, N., 2009, *Assessing Water-related Poverty Using the Sustainable Livelihoods Framework*, Stockholm Environment Institute Working Paper, Stockholm: Stockholm Environment Institute.
- [28]. Bednarski, M., Cholewa, W. & Frid, W., 2004, Identification of sensitivities in Bayesian networks, *Engineering Applications of Artificial Intelligence*, 17, 327-335.
- [29]. Marcot, B.G., 2012, Metrics for evaluating performance and uncertainty of Bayesian network models, *Ecological Modelling*, 230, 50-62.
- [30]. Coupé, V.M.H., Peek, N., Ottenkamp, J. & Habbema, J.D.F., 1999, Using sensitivity analysis for efficient quantification of a belief network, *Artificial Intelligence in Medicine*, 17, 223-247.
- [31]. Gaag, L.C. van der, Renooij, S. & Coupé, V.M.H., 2007, Sensitivity Analysis of Probabilistic Networks, *Advances in Probabilistic Graphical Models: Studies in Fuzziness and Soft Computing*, 213, 103-124, Berlin: Springer.
- [32]. Gaag, L.C. van der, 2001, Analysing Sensitivity Data from Probabilistic Networks, *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*, 530-537.

13. APPENDIX

Appendix A: Scripts for elicitation methods

The following Matlab scripts are implementations of the three elicitation methods; Likelihood method, EBBN method and Weighted algorithm sum, when applied on the node *Battery Voltage* in the network *Car Diagnosis 2*.

A.1

```
% Likelihood method - Battery Voltage

Tx = [1/3 1/3 1/3];
beta = [-1 0 1];
alpha1 = [-1 1];
alpha2 = [-.5 0 .5];
b = 10;
CPT_BV = [95 4 1; 80 15 5; 60 30 10; 0.8 30 69.2; 0.4 20 79.6;...
          0.2 10 89.8]; %Original

s = 0;
cpt = zeros(length(alpha1)*length(alpha2),length(beta));
i = 1;

for j = 1: length(alpha1)
    for k =1:length(alpha2)
        for l =1:length(beta)
            cpt(i,l) = b^((alpha1(j)+alpha2(k))*beta(l))*Tx(l);
        end
        s = sum(cpt(i,:));
        cpt(i,:) = cpt(i,+)/s;
        i = i + 1;
    end
end
CPT_est = cpt*100
diff = abs(CPT_est-CPT_BV); %Difference between generated and original
m = mean(diff(:))
```

A.2

```
% EBBN - Battery Voltage

% Assessments

% 1. S+(CS,BV); S+(BA,BV)
% Charging system: Okay, Faulty
% Battery Age: New, Old, Very Old
% Battery Voltage: Strong, Weak, Dead

% 2.
Pstrong = [0.95 .04 .01]; %a_strong = okay,new
Pweak = [0.008 0.3 0.692]; % a_weak = faulty,new
Pdead = [0.002 0.1 0.898]; % a_dead = faulty, very old

% xc_min = strong, xc_max = dead
```

```

% a_neg = [okay, new], a_pos = [faulty, very old];

% 3.
%P(Xc=xc_max | aneg)
P_dead_okaynew = 0.01;
%P(Xc=xc_min | aneg)
P_strong_okaynew = 0.95;

%P(Xc=xc_max | aneg,k+)
P_dead_faultynew = 0.692; %Xk=CS
P_dead_okayvery = 0.1; %Xk=BA

%P(Xc=xc_min | aneg,k+)
P_strong_faultynew = 0.008; %Xk=CS
P_strong_okayvery = 0.6; %Xk=BA

%Compute delta and weights
dCSp=P_dead_faultynew-P_dead_okaynew;
dCSm=P_strong_okaynew-P_strong_faultynew;
dBAp=P_dead_okayvery-P_dead_okaynew;
dBAm=P_strong_okaynew-P_strong_okayvery;

wCS = 0.5*(dCSp/(dCSp+dBAp)+dCSm/(dCSm+dBAm));
wBA = 0.5*(dBAp/(dCSp+dBAp)+dBAm/(dCSm+dBAm));

% Individual Influence
i_CS = zeros(2,1);
i_BA = zeros(3,1);

for i =1: length(i_CS)
    i_CS(i) = (i-1)/(length(i_CS)-1);
end
for i =1: length(i_BA)
    i_BA(i) = (i-1)/(length(i_BA)-1);
end

% Joint Influence
%ij_a for all combinations of parents
s = (length(i_CS)-1)+(length(i_BA)-1);
ij = zeros(6,1);
l = 1;
for i = 1:length(i_CS)
    for j = 1:length(i_BA)
        ij(l) = (i_CS(i)*(i-1)+i_BA(j)*(j-1))/s;
        l = l+1;
    end
end

%ij_a_xc
ij_okaynew = (i_CS(1)*(1-1)+i_BA(1)*(1-1))/s;
ij_okayvery = (i_CS(2)*(2-1)+i_BA(1)*(1-1))/s;
ij_faultyvery = (i_CS(2)*(2-1)+i_BA(3)*(3-1))/s;

% STEP 1
f_strong = [Pstrong(1) Pweak(1) Pdead(1)]; %f_strong
f_weak = [Pstrong(2) Pweak(2) Pdead(2)]; %f_weak
f_dead = [Pstrong(3) Pweak(3) Pdead(3)]; %f_dead
ij_x = [ij_okaynew ij_okayvery ij_faultyvery];
figure
plot(ij_x,f_strong)
hold on

```

```

plot(ij_x,f_weak,'r')
plot(ij_x,f_dead,'g')
hold off
legend('strong','weak','dead')

%k-values for function, calculate the slopes of the function
for i = 1:length(f_strong)-1
    k_strong(i) = (f_strong(i+1)-f_strong(i))/(ij_x(i+1)-ij_x(i));
    k_weak(i) = (f_weak(i+1)-f_weak(i))/(ij_x(i+1)-ij_x(i));
    k_dead(i) = (f_dead(i+1)-f_dead(i))/(ij_x(i+1)-ij_x(i));
end

% STEP 2
w = [wCS wBA];

ICS = zeros(6,2); %i_min,i_max for parent CS
IBA = zeros(6,2); %i_min,i_max for parent BA
l = 1;
for i = 1:length(i_CS)
    for j = 1:length(i_BA)
        ICS(l,1) = min(i_CS(i),ij(1));
        ICS(l,2) = max(i_CS(i),ij(1));
        IBA(l,1) = min(i_BA(j),ij(1));
        IBA(l,2) = max(i_BA(j),ij(1));
        l = l+1;
    end
end

CPT = zeros(6,3);
S1 = 0;
S2 = 0;

%Integral
for i = 1:6
    S1 = f_area_general(ICS(i,:),f_strong,k_strong,ij_x);
    S2 = f_area_general(IBA(i,:),f_strong,k_strong,ij_x);
    CPT(i,1) = w(1)*S1 + w(2)*S2;

    S1 = f_area_general(ICS(i,:),f_weak,k_weak,ij_x);
    S2 = f_area_general(IBA(i,:),f_weak,k_weak,ij_x);
    CPT(i,2) = w(1)*S1 + w(2)*S2;

    S1 = f_area_general(ICS(i,:),f_dead,k_dead,ij_x);
    S2 = f_area_general(IBA(i,:),f_dead,k_dead,ij_x);
    CPT(i,3) = w(1)*S1 + w(2)*S2;
end

CPT_BVest = CPT*100;
CPT_BVest(4,:) = Pweak*100;
CPT_BVest
CPT_BV = [95 4 1; 80 15 5; 60 30 10; 0.8 30 69.2; 0.4 20 79.6;...
    0.2 10 89.8]; %Original
diff = abs(CPT_BV-CPT_BVest);%Difference between generated and original
m = mean(diff(:))

% f_area_general

```

```

% Calculates the area under the function f between the endpoints in
% interval and x consists of the extreme points in f
function s = f_area_general(interval,f,k,x)
s = zeros(length(f)-1,1);

if interval(2)==interval(1)
    if interval(1) == x(1)
        s(1) = f(1);
    else
        s(end) = f(end);
    end
else
    for m = 1:length(f)-1
        if interval(1) >= x(m) && interval(2) < x(m+1)
            f1 = k(m)*(interval(1)-x(m)) + f(m);
            f2 = k(m)*(interval(2)-x(m)) + f(m);
            b = interval(2)-interval(1);

            else if interval(1)>=x(m) && interval(1)<x(m+1) &&
interval(2)>=x(m+1)
                f1 = k(m)*(interval(1)-x(m)) + f(m);
                f2 = f(m+1);
                b = x(m+1) - interval(1);
            else if interval(1)<x(m) && interval(2)>x(m) &&
interval(2)<=x(m+1)
                f1 = f(m);
                f2 = k(m)*(interval(2)-x(m)) + f(m);
                b = interval(2) - x(m);
            else if interval(1)<x(m) && interval(2)>x(m+1)
                f1 = f(m);
                f2 = f(m+1);
                b = x(m+1) - x(m);
            else
                f1 = 0;
                f2 = 0;
                b = 1;
            end
        end
        end
        end
        h1 = abs(f2-f1);
        h2 = min(f1,f2);
        s(m) = (b*h1/2 + b*h2)/(interval(2)-interval(1));
    end
end

s = sum(s);
end

```

A.3

%Weighted sum algorithm - Battery Voltage

```
w = [0.9 0.1];
```

```

Comp = [0.95 0.04 0.01;... %Comp(Okay) = okay,new
        0.002 0.1 0.898;... %Comp(Faulty) = faulty, very old
        0.95 0.04 0.01;... %Comp(New) = okay, new
        0.004 0.2 0.796;... %Comp(Old) = faulty, old
        0.002 0.1 0.898]; %Comp(VeryOld) = faulty, very old

```

```
CPT = zeros(6,3);
k = 1;

% one loop for each parent node and the child node
for i = 1:2      %2 = number of states in parent 1
    for j = 3:5  %5-2 = number of states in parent 2
        for m = 1:3 %3 = number of states in child node
            CPT(k,m) = w(1)*Comp(i,m) + w(2)*Comp(j,m);
        end
        k = k + 1;
    end
end
end
CPT_BVest = CPT*100;
CPT_BV = [95 4 1; 80 15 5; 60 30 10; 0.8 30 69.2; 0.4 20 79.6;...
          0.2 10 89.8]; %Original

CPT_BVest(5,:) = Comp(4,:) * 100;
CPT_BVest
diff = abs(CPT_BV - CPT_BVest); %Difference between generated and original
mean(diff(:))
```

Appendix B: Scripts for sensitivity analysis

Matlab scripts for the calculation of sensitivity values for a linear function and a hyperbolic function.

B.1

```
%Sensitivity value for a linear function
%In this case SV is calculated for (yi0,p0) and yi's row contains three
%parameters

%% Calculation of the proportional co-variation
yi = [yi0 yi1];      %yi0 initial value, yi1 is the varied value
yj = [yj_1 yj_2];   %yj contains the rest of the parameters in yi0s row
yj_hat = (yj.*(1-yi(2))/(1-yi(1)))*100 %Proportional co-variation of
                                       %yj is needed for p

%% Calculate sv
p = [p0 p1];      %p is the output probability for yi0 and yi1

%Solving the equation system
Y = [yi(1) 1;yi(2) 1 ];
A = Y\p';      %A contains the constants c1/c4 and c2/c4

%Calculate |c1/c4|
sv = abs(A(1))
```

B.2

```
%Sensitivity value for a hyperbolic function
%In this case SV is calculated for (yi0,p0) and yi's row contains three
%parameters

%% Calculation of the proportional co-variation
yi = [yi0 yi1_1 yi1_2];      %yi0 initial value, yi1_1 & yi1_2 are the
                               %varied values
yj = [yj_1 yj_2];   %yj contains the rest of the parameters in yi0s row

yj_hat1 = (yj.*(1-yi(2))/(1-yi(1)))*100 %Proportional co-variation of
yj_hat2 = (yj.*(1-yi(3))/(1-yi(1)))*100 %yj is needed for p

%% Calculate sv
p = [p0 p1_1 p1_2]; %p is the output probability for yi0, yi1_1 & yi1_2

%Solving the equation system
Y = [yi(1) 1 -p(1); yi(2) 1 -p(2); yi(3) 1 -p(3)];
A = Y \ (P.*yi)';
d1 = A(1);
d2 = A(2);
d3 = A(3);
s = -A(3);
t = A(1);
r = A(2)+s*t;

%Sensitivity value
sv = abs((d1*d3-d2)/(yi(1)+d3)^2);

%Vertex
if s<0
```



```
        xv = s + sqrt(abs(r));
else if s>1
        xv = s - sqrt(abs(r));
    else
        's has wrong value'
        xv = s;
    end
end
end
```


Master's Theses in Mathematical Sciences 2013:E37
ISSN 1404-6342

LUTFMS-3219-2013

Mathematical Statistics
Centre for Mathematical Sciences
Lund University
Box 118, SE-221 00 Lund, Sweden

<http://www.maths.lth.se/>