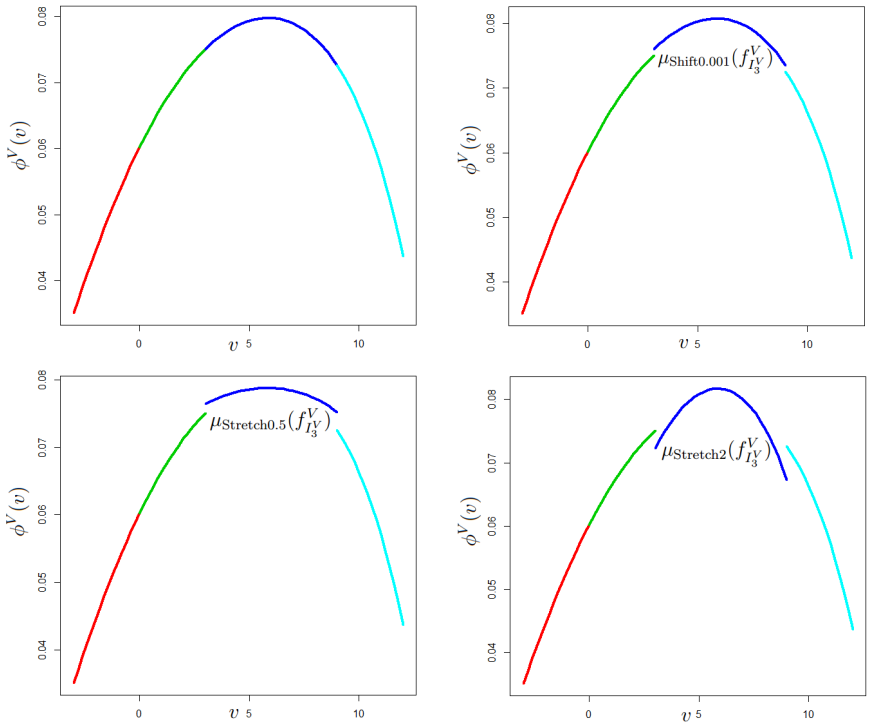
Utrecht University

Faculty of Science

Department of Information and Computing sciences



Master Thesis - Artificial Intelligence

Timo Koppenberg - 3697258

Sensitivity Analysis in Bayesian networks with Mixtures of Truncated Base Functions

Daily Advisor & First Examiner: dr. S. Renooij
Second Examiner: prof. dr. ir. L.C. van der Gaag

November 2, 2017

Abstract

Sensitivity analysis is a technique used to determine the robustness of the output of a mathematical model to inaccuracies in the assessments of its parameters. An existing method of sensitivity analysis for discrete Bayesian networks, where the effect of varying quantitative parameters on the output is analysed, is generalised towards a type of hybrid Bayesian network, namely the Bayesian network with Mixtures of Truncated Base Functions. The generalisation offers multiple ways of varying the parameter functions, such as by shifting and stretching, and gives multiple ways of co-varying the other parameters, where proportional co-variation is deemed best.

Contents

1 Introduction **4**
- 1.1 Bayesian Network . . . 4
- 1.2 Sensitivity Analysis . . . 5
- 1.3 Hybrid Bayesian Network . . . 6
- 1.4 Research Question and Structure of Thesis . . . 6

2 Preliminaries **8**
- 2.1 Notation . . . 8
  - 2.1.1 Variables . . . 8
  - 2.1.2 Graph . . . 8
  - 2.1.3 Intervals . . . 8
- 2.2 Mixture of Truncated Exponentials . . . 9
- 2.3 MTE-network . . . 10
  - 2.3.1 Examples . . . 14
- 2.4 Mixture of Truncated Base Functions . . . 16
  - 2.4.1 Uniform Subinterval MoTBF-network . . . 17
  - 2.4.2 Single Parameter MoTBF-network . . . 17

3 Parameter Co-variation **20**
- 3.1 Conditions . . . 22
- 3.2 Proportional Co-variation . . . 24
  - 3.2.1 Analysis . . . 24
- 3.3 Shift & Normalise . . . 26
  - 3.3.1 Analysis . . . 28
- 3.4 Shift & Linear . . . 29
  - 3.4.1 Analysis . . . 30
- 3.5 Conclusion . . . 31

4 Variation Expression **33**
- 4.1 US MoTBF . . . 33
  - 4.1.1 Examples . . . 36
- 4.2 General MoTBF . . . 39
  - 4.2.1 Subinterval probabilities . . . 39
  - 4.2.2 Marginal probabilities . . . 42
  - 4.2.3 Example . . . 43

5 Sensitivity Analysis **46**
5.1 Introduction . . . 46
5.2 Shift . . . 47
5.2.1 Sensitivity Function . . . 47
5.3 Stretch . . . 49
5.3.1 Sensitivity Function . . . 51
5.4 Extra Base Functions . . . 52
5.4.1 Sensitivity Function . . . 53
5.5 Experiments . . . 53

6 Conclusion and Future Research **56**
6.1 Conclusion . . . 56
6.2 Future Research . . . 57

A Mathematics **59**
A.1 Subinterval Sets . . . 59
A.2 Variation Expression . . . 60
A.2.1 General . . . 60
A.2.2 Prior marginal . . . 61
A.2.3 Posterior marginal . . . 62
A.3 Sensitivity Analysis . . . 63
A.3.1 Stretch . . . 63
A.3.2 Sensitivity Function for Stretch . . . 64

Appendices **59**

List of Figures

2.1 A potential of a root variable V divided into three truncated functions. 12
2.2 An example of a root variable A with two different subinterval sets $\mathbf{I}^A$ and $\mathbf{I}^{A\to B}$ 13
2.3 The set $\mathbf{I}^{A+}$ for the example in Figure 2.2, with $\mathbf{I}^{A+} = \{I_1^{A+}, I_2^{A+}, I_3^{A+}\}$ 14
2.4 The qualitative part of Example 2.3.1 (made using Elvira [15]) 14
2.5 An example of a root variable A with a single child B in a US MoTBF-network 18
2.6 An example of a root variable A with a single child B in an SP MoTBF-network 19

3.1 The original potential of our running example. 21
3.2 The potential of our running example, with parameter variation. $g(v) = \mu(f^V_{I_3^V})$. No co-variation scheme has been applied yet. 21
3.3 The potential of our running example, co-varied with proportional co-variation. 25
3.4 The potential of our running example, with variation and the Shift action. . . . 27
3.5 The potential of our running example, co-varied with the Shift&Normalise scheme. 27
3.6 The Shift&Normalise scheme on an example potential, showing the scheme can lead to a full reset. 29
3.7 The potential of our running example, co-varied with the Shift&Linear scheme. 30

4.1 An example of the conditional potential $\phi^{A|\rho(A)}$ varied on $J^{\rho(A)}$ and J^A. 34
4.2 Top image: An example of the elements of subinterval sets $\mathbf{I}^A$ and $\mathbf{I}^{A+}$ for variable A on the interval $\Omega_A = [0, 1)$. Bottom image: The set $\mathbf{J}^{A+}$ for the example in the top image, for $J^A = I_2^A$. 40

5.1 Plots of the effect of the Shift and Stretch variations. The original potential (top-left) has its parameter $f^V_{I_3^V}$ varied. In the top-right corner it is shifted 0.001 upwards, in the bottom-left corner it is stretched(/shrunk) by a factor 0.5 and in the bottom-right corner it is stretched by a factor 2. No co-variation is applied. . 48
5.2 A plot of the sensitivity function of Example 5.2.2. 50
5.3 Left: The output probability for the shift variation for several values of x. Right: The same image, with the addition of a plotted line that fits through these points. 55
5.4 Left: The output probability for the stretch variation for several values of x. Right: The same image, with the addition of a plotted line that fits through these points. 55

Chapter 1

Introduction

One of the main themes in Artificial Intelligence (AI) is reasoning. Although the definition of intelligence is much debated, most agree that it involves the ability to reason about the world around us. So in any AI must then be able to perform some form reasoning to show intelligence. The earliest methods in AI used hard logic to reason. While this works for some domains, hard logic fails if the AI is applied to a domains that are very complex or domains that have a lot of uncertainty. For example, an AI that monitors a security system might receive as input that an alarm is triggered. Using hard logic, it would then conclude that there is a burglar in the house. But there might be many other reasons the alarm is triggered (e.g. an earthquake, an accident by a resident, a software fault). So the interactions in the domain are quite complex, and there is uncertainty about the actual presence of burglars. In this domain it might be best to model the probability of burglars, given the evidence of the triggered alarm. Such probabilistic methods for AI can be used on many complex and uncertain domains. The prototypical example of probabilistic methods in AI is the Bayesian Network.

The Bayesian Network can be applied to many domains, and is in practice applied to do AI tasks such as military threat evaluation [12] and medical diagnosis [13]. In such domains, it is not hard to understand why it is of vital importance that the output of the network is accurate. One way of increasing the accuracy of the output, is by analysing the sensitivity of the output to the initial assessment of the parameters of the network. More work can then be spend on improving the assessments of the parameters the output is sensitive to.

This thesis concerns the adaptation of such a sensitivity analysis for Bayesian Networks, so it can be applied to a generalisation of Bayesian Networks. In this chapter some terms will be explained so the central question of this thesis can be stated in Section 1.4. That section also provides a road map of the thesis by shortly describing each chapter.

1.1 Bayesian Network

A Bayesian Network (BN) is used to model a probability distribution over a set of random variables, by modelling and using the independencies between those variables. A BN consists of a qualitative part, and an associated quantitative part. The qualitative part of the network is an acyclic directed graph. The vertices of these graph are the random variables of the network, and the edges between the vertices determine the independencies between the variables

according to the so-called d-separation criterion [22]. The associated quantitative part of the network specifies the modelled probability distribution. The complete probability distribution is not directly defined, but is rather factorised into several probability distributions, where each probability distribution is a conditional distribution for a variable of the network given values of its parents. A BN can be used to perform probabilistic inference, i.e. to find probabilities over the probability distribution it models [8, 27, 18].

The most widely used type of BNs is the discrete BN [2]. In a discrete BN, each variable has a discrete (and finite) state space (i.e. set of possible values). The conditional distributions that define the quantitative part of the network are discrete conditional distributions. This means that each of these distributions gives a probability for each possible configuration of values of its parents.

1.2 Sensitivity Analysis

A BN is mathematical model that can be constructed using a variety of data and domain expertise. In any such mathematical model, there may be inaccuracies in the initial assessments of the parameters of the model. It is important to be aware of the effects of these inaccuracies on the output. If inaccuracies of parts of the network can be of huge consequence to the output, it might be best to put more effort into getting reliable parameter estimates. Whereas if these inaccuracies are hardly of influence at all, no such extra research is necessary. After the model has been built, it may also turn out that the model gives inaccurate predictions based on real life instances. It will then be useful to be aware of which parts of the network might cause this unexpected output, so it can be found which parameters require more reliable estimates. For both these problems, we can use the technique of *Sensitivity Analysis* [7].

Sensitivity analysis is a technique that studies the effects of inaccuracies of the parameters of a mathematical model on the output of the model. A parameter of the network is something that makes up the instance of the model. With BNs, we identify two types of parameters we might do sensitivity analysis on. First of all there are the qualitative parameters, which are the vertices (which correspond with variables and their values) and the edges making up the qualitative part of the network. Then there are the quantitative parameters. These are the parameters that specify the individual conditional distributions of a variable given values of its parents. In this thesis, we will focus on these quantitative parameters.

One method for performing sensitivity analysis on discrete BNs is described in an overview paper by van der Gaag et al. [26]. This method analyses the sensitivity through the quantitative parameters of the network. A quantitative parameter is studied by varying it, and then studying the effect on a predefined (prior or posterior) output probability of the network. The variation is done by adapting the parameter to a new value. The function that maps this new value to the new output probability is called the *sensitivity function* of the output probability to the parameter. If the output probability varies much for small changes in the parameter, then the output probability is deemed sensitive to it.

In the method of sensitivity analysis described by van der Gaag et al. [26], upon varying a

parameter some other parameters of the network, namely those from the same conditional distribution, must then also be changed. This is because the probabilities of the individual distribution must sum to 1. The process of changing these additional parameters is called *co-variation*. The most widely used type of co-variation for discrete BNs is called proportional co-variation, where the ratio between each pair of the co-varied parameters is preserved.

1.3 Hybrid Bayesian Network

BNs are often applied to domains that work with both discrete and continuous random variables. This poses a problem for discrete BNs, as they can not deal with continuous random variables. One solution to this problem is to discretise the continuous variables so we can model them as discrete variables[14, 21]. This, however, often results in a loss of information. A better solution might be to model the domain as a Hybrid Bayesian Network (HBN). In a HBN, a variable can have either continuous or discrete conditional distributions over values of its parents. We do need some conditions on our continuous distributions. For example, we still want to be able to combine all individual conditional distributions into a joint distribution over all variables, and we want to be able to find the marginal distribution from a joint one. The first type of HBN developed for this purpose is the Conditional Gaussian model [17]. This model imposes restrictions on both the quantitative part and the qualitative part of the network: The continuous variables need to be assumed to be distributed normally, and discrete variables are not allowed to be children of continuous variables in the graph. Some adaptations have been suggested to overcome the shortcoming of the Conditional Gaussian model. Lerner et al. [19] made it possible for continuous variables to have discrete variables as children, but as discussed by Fernández et al. [9] their model does not allow for exact inference (i.e. finding the exact prior or posterior probabilities of some variables in the network).

A more versatile proposal for a HBN is found using a mixture of truncated functions. This approach was first suggested in the form of using a Mixture of Truncated Exponentials (MTE) in Moral et al. [20]. More recently an approach has been suggested using a Mixture of Polynomials (MOP) [24], and both have been generalised in a global framework using a Mixture of Truncated Base Functions (MoTBF) [16]. In all these models, the domain of the conditional probability distributions is truncated into subdomains and the conditional probability distributions are defined as a combination of *truncated* functions on each of these subdomains. A truncated function has a standard form in each of these types of HBNs. In the MTE it is a summation of weighted exponentials, in the MOP a sum of weighted polynomials. In the MoTBF this is generalised and a truncated function is a sum of weighted base functions, where multiple legal sets of base functions are possible. Each of these models can be used to approximate any type of continuous probability distribution, for example by using Taylor Series [6] or Generalised Fourier Series [25].

1.4 Research Question and Structure of Thesis

A HBN has obvious advantages over the discrete BN, in its capacity to model distributions over both continuous and discrete variables. Therefore it is important that techniques of sensitivity analysis are devised for types of HBNs. Whereas there have been techniques developed for the

Conditional Gaussian model [11, 10, 3], no research has been done into sensitivity analysis of the more versatile MoTBF-networks. The central question of this thesis is then: Is it possible to generalise the method of sensitivity analysis as described by van der Gaag et al. [26] towards the MoTBF-network? This thesis aims to generalise an existing method of sensitivity for discrete BNs, so a first technique for sensitivity analysis of MoTBF-networks can be introduced.

Chapter 2 will start of with some preliminaries we will use throughout the thesis. Some notation needs to be introduced, mostly on the many intervals that are relevant due to the truncation used in the MoTBF-network. The MoTBF-network itself will also be formally introduced, with the definition of the MTE-network used as a stepping stone.

The topic of Chapter 3 is a necessary element of our sensitivity analysis: co-variation. Three methods of performing co-variation in an MoTBF-network are defined and discussed according to conditions that are either necessary or would be nice to have.

Chapters 4 and Chapter 5 together then develop the core of the sensitivity analysis. Chapter 4 develops expressions of probabilities in the network given an arbitrary variation of a truncated function. Chapter 5 gives three specific variations that might be useful for sensitivity analysis and develops specific expressions according to the general expressions of Chapter 4.

The thesis ends with a conclusion in Chapter 6, concluding that, with some reservations in mind, the central question of this thesis can be answered positively. Some suggestion for future research are also given.

Chapter 2

Preliminaries

This chapter will introduce the necessary preliminaries for the thesis. First some notation will be introduced. Then the MTE-network and the more general MoTBF-network will be defined. The original work of this chapter pertains several definitions concerning intervals, which will be necessary to develop a sensitivity analysis of MoTBF-networks later in this thesis.

2.1 Notation

In this section we will work out some notation we will use throughout the thesis.

2.1.1 Variables

We use upper case letters to denote a single variable V. We use lower case to indicate specific values of the upper case letter, so v would be used to denote a specific value of variable V. We use bold case upper letters to denote a set of variables $\mathbf{V}$ and we write $\mathbf{v}$ to indicate a set of specific values for these variables.

2.1.2 Graph

Because the graph of a Bayesian Network constitutes its qualitative part and in the thesis we will mostly concern ourselves with the quantitative part, not much graph theory is needed. Given a directed graph $G = \{\mathbf{V}, \mathbf{A}\}$. For any $V, W \in \mathbf{V}$, we say there is an arc from V to W whenever $(V, W) \in \mathbf{E}$. In this case V is called a *parent* of W, and W is called a *child* of V.

2.1.3 Intervals

In MoTBF-networks we will work with truncated functions. Different truncated functions are used on different parts of the interval of a random variable and the intervals of its parents. Because we need to do much calculations with these truncated functions, we will introduce some notation on these different parts of an interval.

Consider a continuous random variable V. The interval of possible values of V is written as Ω_V. We can divide this continuous interval Ω_V into a set of subintervals $\mathbf{I}^V$. We call this $\mathbf{I}^V$ a *subinterval set* of V if its elements are mutually disjoint and any element in Ω_V is in an

interval of $\mathbf{I}^V$. We call the intervals in a subinterval set *subintervals*. We write the elements of $\mathbf{I}^V$ as subsequent subintervals $\mathbf{I}^V = \{I_1^V, \cdots, I_{n_V}^V\}$, so $|\mathbf{I}^V| = n_V$. When we take an arbitrary element of $\mathbf{I}^V$, we write I^V.

In many cases we are given a set of variables, say this set is $\mathbf{V} = \{V_1, \cdots, V_l\}$. If we want to indicate the Cartesian product of all domains we write $\Omega_{\mathbf{V}} = \Omega_{V_1} \times \cdots \times \Omega_{V_l}$ and for the subinterval sets we write $\mathbf{I}^{\mathbf{V}} = \mathbf{I}^{V_1} \times \cdots \times \mathbf{I}^{V_l}$. We also call this a *subinterval set* of $\mathbf{V}$. We once again write $I^{\mathbf{V}} \in \mathbf{I}^{\mathbf{V}}$ to indicate an element of the Cartesian product of subinterval sets, so one subinterval per variable in $\mathbf{V}$.

2.2 Mixture of Truncated Exponentials

Before we will introduce the Mixture of Truncated Base Functions, we will first introduce the Mixture of Truncated Exponentials [20]. This special case is a good starting point as it more intuitive than the general case. We are largely inspired by Fernández et al. [9] in our style of definition. We define an MTE-potential as follows.

Definition 2.2.1. Let $\mathbf{X} = \mathbf{Y} \cup \mathbf{Z}$ with $\mathbf{Y} = \{Y_1, \cdots, Y_c\}$ and $\mathbf{Z} = \{Z_1, \cdots, Z_d\}$ as a set of discrete and continuous variables, respectively. Now a function $\phi^{\mathbf{X}} : X \to \mathbb{R}$ is a *Mixture of Truncated Exponentials Potential (MTE-potential)* over $\mathbf{X}$, if and only if for some subinterval set $\mathbf{I}^{\mathbf{Z}}$, it holds that for any $I^{\mathbf{Z}} \in \mathbf{I}^{\mathbf{Z}}$, we can write $\phi^{\mathbf{X}}$ as:

$$\phi^{\mathbf{X}}(\mathbf{x}) = \phi^{\mathbf{X}}(\mathbf{y}, \mathbf{z}) = a_0 + \sum_{i=1}^{m} a_i \exp(\sum_{j=1}^{n} b_i^j \cdot z_i) \qquad \text{if } \mathbf{z} \in I^{\mathbf{Z}} \tag{2.1}$$

with $m \in \mathbb{N}$, $a_i \in \mathbb{R}$ for $i \in \{0, \cdots, m\}$, and $b_i^j \in \mathbb{R}$ for $i \in \{1, \cdots, m\}$ and $j \in \{1, \cdots, n\}$.

In other words, the potential is defined in terms of multiple *truncated functions*. We have one truncated function per value set of the discrete variables, and interval set of the continuous variables. A truncated function is a summation of weighted exponentials, where the exponent of each exponential is a weighted sum of values of the continuous variables. It is called a truncated function because the domain of the potential is cut up ('truncated') into smaller subintervals, where on each of these subintervals a different truncated function determines the value.

Example 2.2.2. We consider a set of variables $\mathbf{X} = \{A, B, C\}$ with a binary discrete variable A and continuous variables B, C. We have that $\Omega_A = \{\text{true}, \text{false}\}$ and $\Omega_B = \Omega_C = [2, 7)$. We have that $\mathbf{I}^B = \{[2, 4), [4, 7)\}$ and $\mathbf{I}^C = \{[2, 7)\}$. This means that the domain of variable B is divided into two subsequent subintervals $[2, 4)$ and $[4, 7)$, and the domain of variable C is divided into a single subinterval $[2, 7)$. We define an MTE-potential over $\mathbf{X}$ as follows:

$$\phi^{\mathbf{X}}(a, b, c) = \begin{cases} 0.121e^{0.3b-0.23c} + 0.23, & \text{if } a = \text{ false and } 2 \leq b < 4 \\ 0.34e^{2.4b} + 0.12e^{-0.23c} + 0.06, & \text{if } a = \text{ false and } 4 \leq b < 7 \\ 0.2e^{1.6b+0.4c} + 0.1e^{-0.2b+0.4c}, & \text{if } a = \text{ true and } 2 \leq b < 4 \\ 0.3e^{-2.5b+0.2c} + 2, & \text{if } a = \text{ true and } 4 \leq b < 7 \end{cases}$$

Now we have seen an example MTE-potential, we are ready to define what it means for a random variable to follow an MTE-Distribution:

Definition 2.2.3. Consider a mixed variable set $\mathbf{X} = \mathbf{Y} \cup \mathbf{Z}$. We say that $\mathbf{X}$ follows an *MTE-Distribution* if its MTE-potential $\phi^{\mathbf{X}}$ is a *density*, i.e.:

$$P(\mathbf{X} \in \Omega_{\mathbf{X}}) = \sum_{\mathbf{y} \in \Omega_Y} \int_{\Omega_Z} \phi^{\mathbf{X}}(\mathbf{y}, \mathbf{z}) d\mathbf{z} = 1 \tag{2.2}$$

In Section 2.3 we will define the distribution of a network through the conditional distributions of each variable conditioned on its parents. We require these individual distributions to be conditional densities. This is defined as follows:

Definition 2.2.4. Consider two disjoint mixed variable sets $\mathbf{X}_1$ and $\mathbf{X}_2$. We say an MTE-potential $\phi^{\mathbf{X}_1 \cup \mathbf{X}_2}$ is a *conditional density* over $\mathbf{X}_1$ conditioned on $\mathbf{X}_2$ if, for every $c_{\mathbf{X}_2} \in \Omega_{\mathbf{X}_2}$ we have that $\phi^{\mathbf{X}_1}$ defined as

$$\phi^{\mathbf{X}_1}(\mathbf{x}_1) =_{\text{def}} \phi^{\mathbf{X}_1 \cup \mathbf{X}_2}(\mathbf{x}_1, c_{\mathbf{X}_2}) \tag{2.3}$$

is a density. We will then write $\phi^{\mathbf{X}_1 \cup \mathbf{X}_2}$ as $\phi^{\mathbf{X}_1 | \mathbf{X}_2}$ to clearly indicate we are discussing a conditional density.

2.3 MTE-network

In this section we will introduce the MTE-network. An MTE-network is much like a discrete Bayesian network, except the quantitative part of the network is an MTE-potential instead of a discrete probability distribution. There are also some special conditions that need apply for an MTE-network. This section will give these special conditions and simultaneously introduce much notation. Much of the subject matter of this section is vital to the understanding of the thesis, but it can also be perceived as quite difficult. If the reader struggles with the mathematics of this section, note that the example at the end of the section might be of help.

A *Mixture of Truncated Exponentials network* (MTE-network) is defined as $(G, \phi^{\mathbf{V}})$ where $G = (\mathbf{V}, \mathbf{A})$ an acyclic directed graph, $\mathbf{V}$ a set of continuous and discrete random variables, and $\phi^{\mathbf{V}}$ an MTE-potential over $\mathbf{V}$. We will only consider networks where all variables in $\mathbf{V}$ are continuous, so no variable is discrete. The generalisation towards a mixed set of variables is trivial, and focusing on only continuous variables makes the mathematics clearer to the reader.

As with discrete Bayesian Networks, an MTE-network defines the joint distribution $\phi^{\mathbf{V}}$ in terms of the conditional distributions of the children conditioned upon their parents in G, that is[9]:

$$\phi^{\mathbf{V}}(\mathbf{v}) = \prod_{V \in \mathbf{V}} \phi^{V|\rho(V)}(v, \rho(v)) \tag{2.4}$$

where $\rho(V)$ indicates the parents of V and $\rho(v)$ the values for the parents of V consistent with the set $\mathbf{v}$.

Consider a node V in the network, and its given potential $\phi^{V|\rho(V)}$ with associated interval sets $\mathbf{I}^V$ and $\mathbf{I}^{\rho(V)}$. It is argued by Langseth et al. [16] that it is hard to make the potential $\phi^{V|\rho(V)}$

a conditional density. Roughly their argument states that there are an infinite constraints to fulfil, since for all the values of $\rho(V)$ we want $\phi^{V|\rho(V)}$ fixed on those values to be a potential as defined in Definition 2.2.4. But there are only finitely many arguments. The way they solved this is by making $\rho(V)$ influence the value of $\phi^{V|\rho(V)}$ through the way the function is truncated, but not through the truncated functions themselves. In other words, if we consider arbitrary intervals $I^V \in \mathbf{I}^V$ and $I^{\rho(V)} \in \mathbf{I}^{\rho(v)}$, we have that

$$\phi^{V|\rho(V)}(v, \rho(v)) = a_0 + \sum_{i=1}^{m} a_i \exp(b_i \cdot v) \qquad \text{if } v \in I^V \text{ and } \rho(v) \in I^{\rho(V)} \tag{2.5}$$

with $m \in \mathbb{N}$, $a_i \in \mathbb{R}$ for $i \in \{0, \cdots, m\}$ and $b_i \in \mathbb{R}$ for $i \in \{1, \cdots, m\}$. We see this definition is a special case of Equation 2.1. It is both special in the sense that all variables involved are continuous, and in that only the value of v has any influence on the truncated function itself (but none of the values in $\rho(v)$). The values in $\rho(v)$ may not influence the truncated function itself. All they do determine which truncated function is selected based on the intervals the values are in.

Since $\rho(v)$ only influences the potential through the truncation of the functions, it is the case that for any $I^{\rho(V)} \in \mathbf{I}^{\rho(v)}$, we get a (non-conditional) potential over v given by these subintervals. To indicate this potential, we write $\phi^{V|I^{\rho(V)}}(v)$. Note that if we have two, possibly different, sets of values for the parents $\rho(v), \rho'(v)$ such they are both part of the same subintervals $I^{\rho(V)}$ (so $\rho(v) \in I^{\rho(V)}$ and $\rho'(v) \in I^{\rho(V)}$) then they will select the same potential over v, since the values of the parents only influence the selection of a potential through the subintervals they belong to. This potential is referred to as $\phi^{V|I^{\rho(V)}}(v)$.

When we refer to a specific truncated function of the MTE-network, we write it is as a function $f^V_{I^V, I^{\rho(V)}}$. So this is the potential $\phi^{V|I^{\rho(V)}}$ on the subinterval I^V. Since we define our MTE-network in terms of these truncated functions, we will sometimes call them the (quantitative) *parameters* of the network. We see an example of a partition of a potential in truncated functions in Figure 2.1.

Of course we require that each potential $\phi^{V|\rho(V)}$ is a conditional density, for any value of the parents. This means, for each $\rho(v) \in \Omega_{\rho(V)}$, the following holds:

$$\int_{\Omega_V} \phi^{V|\rho(V)}(v, \rho(v))dv = 1 \tag{2.6}$$

In terms of subintervals this means that for every $I^{\rho(V)} \in \mathbf{I}^{\rho(V)}$ the following holds:

$$\sum_{I^V \in \mathbf{I}^V} \int_{I^V} f^V_{I^V, I^{\rho(V)}}(v)dv = 1 \tag{2.7}$$

Often, when we want to indicate a specific function in $\mathbf{f}$, we have a set of subintervals for more variables than a variable V and its parents. In this case the selection of correct subintervals is left implicit. For example, if we have a variable $V \in \mathbf{V}$ and a selection of intervals over all the variables $I^{\mathbf{V}}$, with $I^V \in I^{\mathbf{V}}$ and $I^{\rho(V)} \subset I^{\mathbf{V}}$, then we can write:

$$f^V_{I^{\mathbf{V}}} = f^V_{I^V, I^{\rho(V)}} \tag{2.8}$$

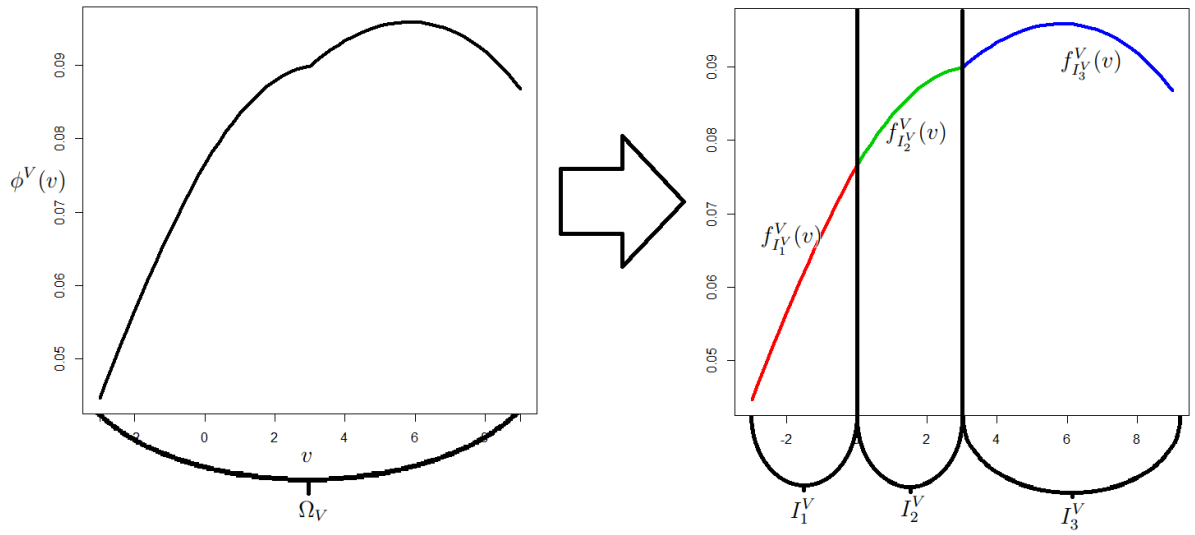


Figure 2.1: A potential of a root variable V divided into three truncated functions.

Different Subinterval Sets

It is important to note that in the definition of a MTE Network, we are allowed to have different subinterval sets for the same variable. If we have two variables A and B and in the network B is a child of A, then there could be a different subinterval set for variable A in the definition of the condition density of variable A than that of variable B. The first subinterval set we will still refer to as $\mathbf{I}^A$, but the other subinterval set we will indicate as $\mathbf{I}^{A\to B}$. Note that the latter is a subinterval set of A, but is used in the partition of B. An example of this is shown in Figure 2.2.

In Chapter 4 we will have to make calculations using the overlap between these different subintervals. So say we have a variable V with children $C_1, \cdots, C_n$, and we have associated subinterval sets $\mathbf{I}^V, \mathbf{I}^{V\to C_1}, \cdots, \mathbf{I}^{V\to C_n}$. Now we are interested in all the nonempty intersections of subintervals in these sets, defined as such:

$$\mathbf{I}^{V+} =_{\text{def}} \tag{2.9}$$

$$\{I^V \cap I^{V\to C_1} \cap \cdots \cap I^{V\to C_n} | I^V \in \mathbf{I}^V, I^{V\to C_1} \in \mathbf{I}^{V\to C_1}, \cdots, I^{V\to C_n} \in \mathbf{I}^{V\to C_n}\} - \{\emptyset\} \tag{2.10}$$

We note that $\mathbf{I}^{V+}$ is also a subinterval set, as proven in Theorem A.1.1. Figure 2.3 gives a nice example of such a set. We notice that for any $I^{V+} \in \mathbf{I}^{V+}$ there is exactly one element in each of $\mathbf{I}^V, \mathbf{I}^{V\to C_1}$ etc. such that it is a superset of I^{V+}. This is proven in Theorem A.1.2. It will be useful in Chapter 4 to be able to refer to these supersets, since we can find associated truncated functions using these. So we introduce the function d subscripted with the relevant variable to refer to specific supersets. For example, if we look at Figure 2.2 and Figure 2.3 we have that $d_A(I_2^{A+}) = I_1^A$ and $d_B(I_2^{A+}) = I_2^{A\to B}$.

Now if we have a variable B with a single parent A, we see that a subinterval I^{A+} of $\mathbf{I}^{A+}$ is the subset of exactly one subinterval in $\mathbf{I}^{A\to B}$, namely $d_B(I^{A+})$. So we can write $\phi^{B|I^{A+}}$ as a shorthand for the potential $\phi^{B|d_B(I^{A+})}$. This of course generalised trivially to multiple parents. We will use this often in Chapter 4.

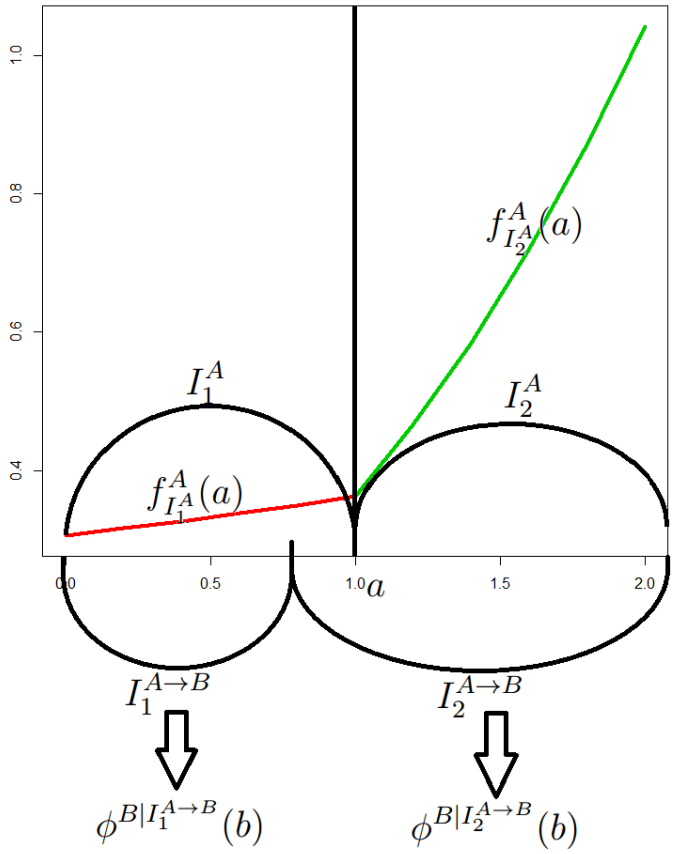


Figure 2.2: An example of a root variable A with two different subinterval sets $\mathbf{I}^A$ and $\mathbf{I}^{A\to B}$

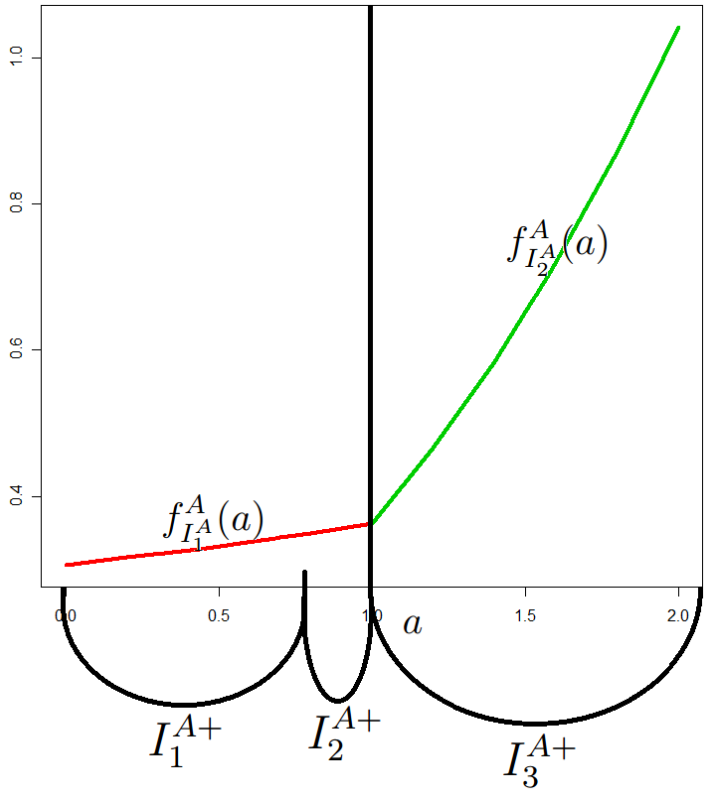


Figure 2.3: The set $\mathbf{I}^{A+}$ for the example in Figure 2.2, with $\mathbf{I}^{A+} = \{I_1^{A+}, I_2^{A+}, I_3^{A+}\}$

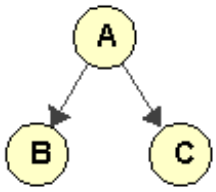


Figure 2.4: The qualitative part of Example 2.3.1 (made using Elvira [15])

2.3.1 Examples

We will first look at an example MTE-network, and then we will verify for some of the potentials that they are indeed conditional densities.

Example 2.3.1. For our example we look at an MTE-network with three variables: A, B, and C. The qualitative part of the network is shown in Figure 2.4. We first look at the potential of variable A. Since A does not have any parents, the potential is not conditioned on any variable:

$$\phi^A(a) = \begin{cases} 0.152e^{-2.3a} + 0.228, & \text{if } 0 \leq a < 1 \\ 0.152e^{1.2a} - 0.261, & \text{if } 1 \leq a < 2 \end{cases}$$

Now we look at the potentials for variables B and C. Since B and C both have A as their single

parent, both potentials are conditioned on A.

$$\phi^{B|A}(b,a) = \begin{cases} 0.121e^{1.3b} + 0.121e^{-0.23b} + 0.06, & \text{if } 0 \leq b < 1 \text{ and } 0 \leq a < 1 \\ 0.121e^{-0.5b} + 0.526, & \text{if } 1 \leq b < 2 \text{ and } 0 \leq a < 1 \\ 0.116e^{1.3b} + 0.174, & \text{if } 0 \leq b < 1 \text{ and } 1 \leq a < 2 \\ 0.116e^{-0.33b} + 0.517, & \text{if } 1 \leq b < 2 \text{ and } 1 \leq a < 2 \end{cases}$$

$$\phi^{C|A}(c,a) = \begin{cases} -0.773e^{-2c} + 1.546, & \text{if } 0 \leq c < 0.5 \text{ and } 0 \leq a < 1.3 \\ -0.773e^{0.9c} + 2.474, & \text{if } 0.5 \leq c < 1 \text{ and } 0 \leq a < 1.3 \\ 0.256e^{2.3c} + 0.359, & \text{if } 0 \leq c < 0.5 \text{ and } 1.3 \leq a < 2 \\ 0.256e^{-6.3c} + 1.157, & \text{if } 0.5 \leq c < 1 \text{ and } 1.3 \leq a < 2 \end{cases}$$

Note that in the conditional potentials of variable B and C, the parent variable A only influences the value of the potential through the way the potential is truncated but not through the truncated functions themselves. I.e. a is included in the right-hand conditional side of the definitions, but not on the left-hand side which specify the truncated functions.

We will look at the domains and the subinterval sets of all variables. Firstly we see that $\Omega_A = \Omega_B = [0,2)$ and $\Omega_C = [0,1)$. The variable A has associated subinterval set $\mathbf{I}^A = \{I_1^A, I_2^A\}$ with $I_1^A = [0,1)$ and $I_2^A = [1,2)$. Now since B and C are both children of A, they too have associated subinterval sets for A. We have in this case that $\mathbf{I}^{A\to B}$ is equal to $\mathbf{I}^A$ but $\mathbf{I}^{A\to C}$ is not. We have that $\mathbf{I}^{A\to C} = \{I_1^{A\to C}, I_2^{A\to C}\}$ with $I_1^{A\to C} = [0, 1.3)$ and $I_2^{A\to C} = [1.3, 2)$. Now variables B and C also have respective subinterval sets $\mathbf{I}^B = \{[0,1),[1,2)\}$ and $\mathbf{I}^C = \{[0,0.5),[0.5,1)\}$.

In relation to the different subinterval sets, we note that $\mathbf{I}^{A+} = \{[0,1),[1,1.3),[1.3,2)\}$. Note that these are the intersections of the truncation of A in the different subinterval sets $\mathbf{I}^A$, $\mathbf{I}^{A\to B}$ and $\mathbf{I}^{A\to C}$. Now we write $I_2^{A+} = [1, 1.3)$ and note that $d_A(I_2^{A+}) = I_2^A$ and $d_C(I_2^{A+}) = I_1^{A\to C}$.

Example 2.3.2. In Example 2.3.1, we have seen three potentials defined. All these potentials are conditional densities, save for some rounding error (smaller than 0.01). Note that the potential for A is a conditional density conditioned on an empty set of parents, which means it is a density. We see that, because of the following property, the potential over A is a density:

$$\begin{aligned} & P(A \in \Omega_A) \\ =& \int_{\Omega_A} \phi^A(a)da \\ =& \int_0^1 (0.152e^{-2.3a} + 0.228)da + \int_1^2 (0.152e^{1.2a} - 0.261da) \approx 1 \end{aligned}$$

The potential over C conditioned on A is a conditional density. To see this we need to consider all values of A. We can do this easily, by noting that for any $a \in \Omega_A$, we have that either $a \in I_1^{A\to C}$ or $a \in I_2^{A\to C}$. If the former holds then:

$$\begin{aligned} & P(C \in \Omega_C | A = a) \\ =& \int_{\Omega_C} \phi^{C|I_1^{A\to C}}(c)dc \\ =& \int_0^{0.5} (-0.773e^{-2c} + 1.546)dc + \int_{0.5}^1 (-0.773e^{0.9c} + 2.474dc) \approx 1 \end{aligned}$$

and if the latter holds (i.e. $a \in I_2^{A \to C}$) then:

$$\begin{aligned} & P(C \in \Omega_C | A = a) \\ = & \int_{\Omega_C} \phi^{C|I_2^{A \to C}}(c) dc \\ = & \int_0^{0.5} (0.256e^{2.3c} + 0.359) dc + \int_{0.5}^{1} (0.256e^{-6.3c} + 1.157 dc) \approx 1 \end{aligned}$$

2.4 Mixture of Truncated Base Functions

In this section, we will define the notion of Mixture of Truncated Base Functions, largely following the work of Langseth et al. [16]. As a general form of truncated functions, we want to abstract away from exponential functions and allow any set of functions as long as some criteria are met. These criteria ensure that in the resulting network, any function can be approximated with arbitrarily good precision[16, 25]. The criteria on the set of functions are not further used in the thesis, so they do not have to be understood. But we will include them for completeness.

Definition 2.4.1. Consider a set of functions $\psi = \{\psi_0, \psi_1, \cdots\}$ with $\psi_i : \mathbb{R} \to \mathbb{R}$ for $i \in \mathbb{N}$. Let Q be the set of all linear combinations of the members of ψ, i.e. all members of the type $\sum_{i=0}^{\infty} a_i \psi_i$ with $a_i \in \mathbb{R}$ for $i \in \mathbb{N}$. Now ψ is a *legal set of base functions* if the following conditions hold:

1. The first function ψ_0 is constant in its argument.
2. If we have $f \in Q$ and $g \in Q$, then their composition $(f \circ g) \in Q$.
3. For any pair of real number $s, t \in \mathbb{R}$, there exists a function $f \in Q$ such that $f(s) \neq f(t)$.

Now we know what a legal set of base functions constitutes, we can define an MoTBF-potential. Notice how this definition is a more general version of Definition 2.2.1.

Definition 2.4.2. Let $\mathbf{X} = \mathbf{Y} \cup \mathbf{Z}$ with $\mathbf{Y} = \{Y_1, \cdots, Y_c\}$ and $\mathbf{Z} = \{Z_1, \cdots, Z_d\}$ as a set of discrete and continuous variables respectively. Now $\phi^{\mathbf{X}}$ is a *Mixture of Truncated Base Functions Potential (MoTBF-potential)* over $\mathbf{X}$, if and only if for some $\mathbf{I}^{\mathbf{Z}}$ it holds that for any $I^{\mathbf{Z}} \in \mathbf{I}^{\mathbf{Z}}$, we can write $\phi^{\mathbf{X}}$ as:

$$\phi^{\mathbf{X}}(\mathbf{x}) = \phi^{\mathbf{X}}(\mathbf{y}, \mathbf{z}) = \sum_{i=0}^{m} \prod_{j=1}^{n} a_i^j \psi_i(z_j) \qquad \text{if } \mathbf{z} \in I^{\mathbf{Z}} \tag{2.11}$$

for $m \in \mathbb{N}$, $a_i^j \in \mathbb{R}$ for $i \in \{0, \cdots, m\}$ and $j \in \{1, \cdots, n\}$, and an arbitrary legal set of base functions ψ.

Since this thesis aims to be as general as possible, we will prove theorems on MoTBF-potentials. But it is recommended to think of the special case of MTE-potentials throughout this thesis since they are more intuitive. As explained by Langseth et al. [16], we can get a definition for the MTE-potential as a special case of the MoTBF-potential exactly by taking the legal set of base functions $\psi = \{1, exp(-x), exp(x), exp(-2x), \cdots\}$.

The definition of an MoTBF-density and an MoTBF-network are entirely analogous to the definitions of their respective MTE counterpart. When we consider the conditional potentials defining our MoTBF-network, we again want the parent variables to only influence the truncation of the potential, but not the truncated functions themselves. We formalise this condition with an equation analogous to Equation 2.5:

$$\phi^{V|\rho(V)}(v, \rho(v)) = \sum_{i=0}^{m} a_i \psi_i(v) \qquad \text{if } v \in I^V \text{ and } \rho(v) \in I^{\rho(V)} \tag{2.12}$$

for $m \in \mathbb{N}, a_i \in \mathbb{R}$ with $i \in \{0, \cdots, m\}$. We note that Equation 2.12 is a special case of Equation 2.11 with the extra condition that $a_i^j = 0$ when X_i is a parent variable.

We will now look at two special cases of the MoTBF-network, the Uniform Subinterval MoTBF-network and the Single Parameter MoTBF-network. The first will be used as a stepping stone to prove theorems on the general MoTBF-network we have just seen in Chapter 4, and the latter is a special case often used in practice.

2.4.1 Uniform Subinterval MoTBF-network

We call an MoTBF-network a *Uniform Subinterval MoTBF-network (US MoTBF-network)* if all subinterval sets associated with a single variable are equal. Formally, this means that for any variable V in the network with children $C_1, \cdots, C_n$, it holds that:

$$\mathbf{I}^V = \mathbf{I}^{V \to C_1} = \cdots = \mathbf{I}^{V \to C_n} \tag{2.13}$$

An example of a variable in this network is shown in Figure 2.5. We introduce this type of network because it is simpler to work with than the general MoTBF-network. The US MoTBF-network is a good stepping stone towards the general MoTBF-network when developing our sensitivity analysis in Section 4.1.

We see that the network used in Example 2.3.1 is not a US MoTBF-network. It would be if for example $I_1^{A \to C} = [0, 1)$ and $I_2^{A \to C} = [1, 2)$, because then all subinterval sets of A would be equal.

2.4.2 Single Parameter MoTBF-network

We call an MoTBF-network a *Single Parameter MoTBF-Network (SP MoTBF-network)* if no conditional potential is truncated through the values of its associated variable. Formally, this means that for any variable V in the network it holds that:

$$|\mathbf{I}^V| = 1 \tag{2.14}$$

Note that this means that the conditional potential is not truncated through the value of the associated variable itself, but it can be truncated through the values of its parents. An example of a variable in this network is shown in Figure 2.6.

This special case of the MoTBF-network is found in the paper by Langseth et al. [16]. Whereas

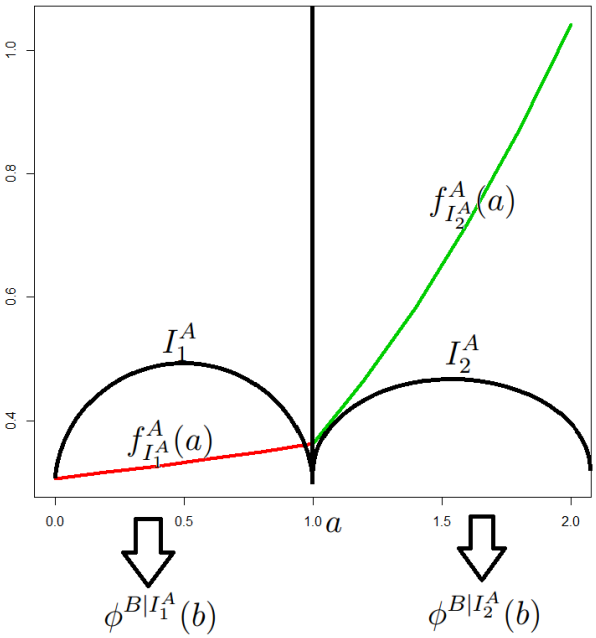


Figure 2.5: An example of a root variable A with a single child B in a US MoTBF-network

the paper starts with the generalised MoTBF-network, it switches to the definition of the SP MoTBF-network when the implementation is discussed. It is argued that one truncated function is enough to approximate any function in its interval, given enough legal basis functions[25]. Therefore only one truncated function is used, presumably to make the representation less complex. The R package for MoTBFs also implements this special case of the network[23].

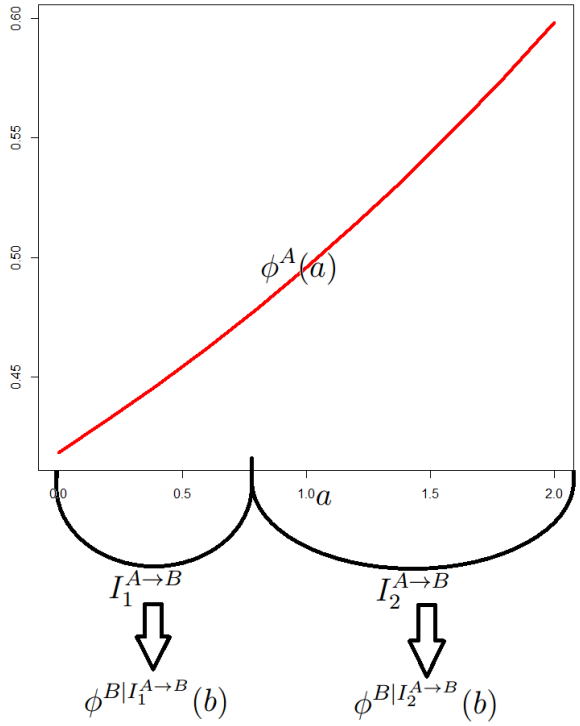


Figure 2.6: An example of a root variable A with a single child B in an SP MoTBF-network

Chapter 3

Parameter Co-variation

This chapter concerns parameter co-variation. In Chapter 4 we will discuss variations of a parameter in a MoTBF-network and the effect of these variations on the output of the network. But to preserve the validity of the network under variation, some additional parameters do need to be co-varied with the varied parameter. This chapter will define and evaluate multiple ways of doing this co-variation.

Suppose we want to vary a parameter of the quantitative part of the network. That is, for a variable $V \in \mathbf{V}$ and intervals $J^V \in \mathbf{I}^{\mathbf{V}}$, $J^{\rho(V)} \in \mathbf{I}^{\rho(V)}$ we want to adapt the truncated function $f^V_{J^V,J^{\rho(V)}}$. We want to adapt it into a new truncated function, an arbitrary truncated function g. Formally, we write:

$$\mu(f^V_{J^V,J^{\rho(V)}})(v) = g(v) \tag{3.1}$$

Here we write μ as the function that maps the original parameter to its varied counterpart. For example, we see an original potential in Figure 3.1, and the variation in Figure 3.2.

Now we need to adapt other parameters of the same variable to ensure that we still have a conditional density. This is called *co-variation*, and a systematic way to co-vary parameters given a varied parameters is called a *co-variation scheme*. We also write μ to show the mapping of the original parameters to the co-varied parameters. And to combine parameter variation and co-variation, we write μ to map the original potential of a variable to its varied counterpart, by application of μ to all its individual parameters.

If we look at Equation 2.7, we see we only need to co-vary the parameters that are in the same parent interval as our varied parameters to ensure our potential stays a density under variation. In other words, upon varying $f^V_{J^V,J^{\rho(V)}}$ we only need to co-vary parameters $f^V_{I^V,J^{\rho(V)}}$, where $I^V \neq J^V$. This means we can keep all our other parameters constant, which is of course something we want to do. So we have that for $I^{\rho(V)} \neq J^{\rho(V)}$ and for any $I^V \in \mathbf{I}^V$, that μ is constant, i.e. $\mu(f^V_{I^V,I^{\rho(V)}}) = f^V_{I^V,I^{\rho(V)}}$. So for this chapter it is only interesting to look at a potential for fixed parent subintervals $J^{\rho(V)}$. We will therefore restrict ourselves in this chapter to looking at a root variable V, since its potential behaves the same as a potential for a variable under fixed parent subintervals.

We consider a root variable V, with potential ϕ^V that is a probability density. We then would

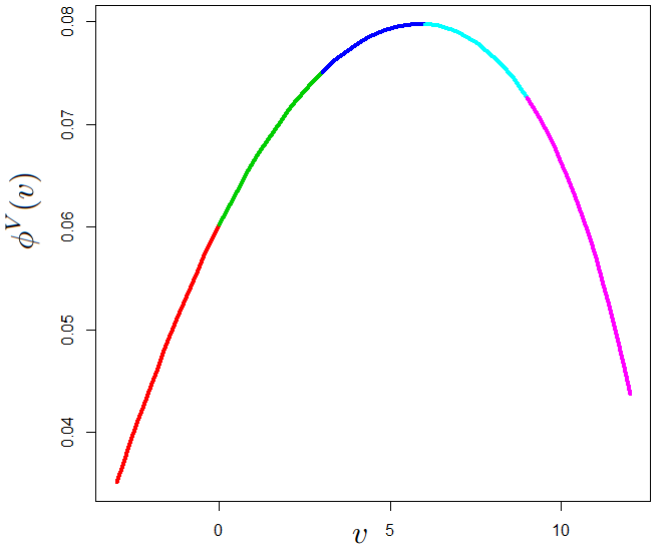


Figure 3.1: The original potential of our running example.

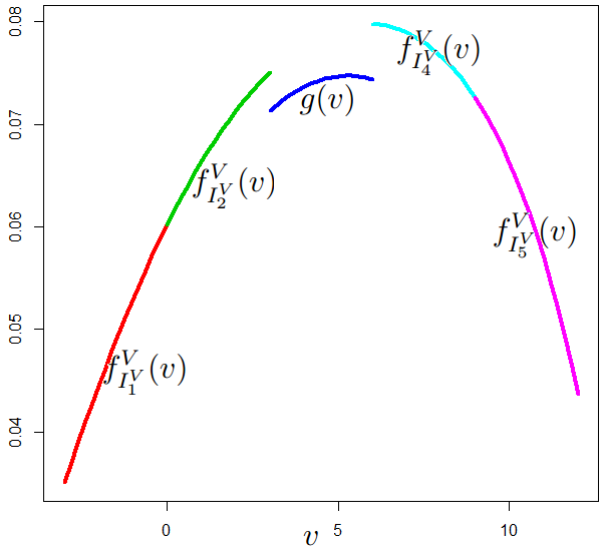


Figure 3.2: The potential of our running example, with parameter variation. $g(v) = \mu(f^V_{I_3^V})$. No co-variation scheme has been applied yet.

want $\mu(\phi^V)$ to be a probability density as well. For this its area must sum to 1, corresponding with the following equation:

$$\begin{aligned} &\int_{\Omega_V} \mu(\phi^V)(v)dv \\ &= \sum_{I^V \in \mathbf{I}^V} \int_{I^V} \mu(f^V_{I^V})(v)dv = 1. \end{aligned} \tag{3.2}$$

And we want the potential to be positive, which means that for every $v \in \Omega_V$ that:

$$\mu(\phi^V)(v) \geq 0 \tag{3.3}$$

Throughout the chapter we will find that Equation 3.2 is easy to fulfil, where Equation 3.3 will be harder to fulfil under more intricate co-variation schemes. In this chapter we will look at several co-variation schemes. The next section will describe some conditions we want for our different co-variation schemes. This will include the two conditions we just discussed, but also include conditions we might impose to save as much of the original potential as possible. The sections thereafter will each define a co-variation scheme and analyse it according to those conditions. The chapter ends with a concluding section that determines which co-variation scheme will be used throughout the rest of the thesis.

3.1 Conditions

In this section we will look at some conditions that might be favourable for a co-variation scheme. The first two conditions are as described in the previous section, and are necessary if we want to preserve densities under variation. The other conditions are nice to have, and all concern maintaining certain properties of the parameters under co-variation, so co-variation saves as much of the original potential as possible. This makes intuitive sense: We want to study the effect of variation, and we only use co-variation as a means to preserve a probability density.

In Figure 3.1, we see an example potential of a root variable V, and in Figure 3.2 we see one of its parameters varied. We see that the parameters we need to co-vary can be either on the left or on the right side of the varied parameter. We will refer to these as the left side and the right side of the potential, which we will often use in the definition of the conditions. We define the following conditions:

Area If the potential has an area summing to 1 under its interval, then this will be maintained under potential variation.

This is the condition we have formalised in Equation 3.2.

Positive Positive parameters remain positive under co-variation.

This is the condition we have formalised in Equation 3.3.[1] Note that this means that every co-varied parameter is a positive function in its subinterval. We will see in Sections

[1] We use 'positive' to indicate anything greater or equal to zero. Otherwise we will use 'strictly positive'.

3.3.1 and 3.4 that this condition is hard to maintain, since many simple transformations could already make the parameters negative. We might check if we can at least fulfil this condition within reasonable variation. It would at least be reasonable to restrict the varied parameter to having an area smaller or equal to 1. If it would have a larger area, it would be impossible make the varied potential a probability density; the area of the rest of the potential must be negative to ensure Condition Area, and therefore this condition can not hold.

Ratio The ratio between the area of the left side and the area of the right side remains constant under potential variation.

When we vary a parameter, the area under its subinterval might change. Since we want the potential to be a density, this means that the area on the left and right side can not always remain constant under co-variation. What is possible however is that the ratio between the area of the left and right side remains constant.

Joining The property of joining parameters is preserved under potential variation.

With the property of joining parameters, we mean that consecutive parameters join each other on the extremes of the intervals. We can see this is the case in Figure 3.1, as all lines of different colours 'touch'. Variation alone might not preserve this property, as we see in Figure 3.2, where the visualisation of the truncated function g does not 'touch' those of $f^V_{I^V_2}$ and $f^V_{I^V_4}$ at respectively its left and right end.

The formalisation of this property is quite hard, mostly because of the mutually disjoint property of subinterval sets. The joining condition corresponds to the following:

For every $i \in \{1, n_V - 1\}$, we have two consecutive subintervals $I^V_i, I^V_{i+1} \in \mathbf{I}^V$. Now if $I^V_i = [a, b)$ for some $a, b \in \mathbb{R}$ then $I^V_{i+1} = [b, c)$ for some $c \in \mathbb{R}$. It then holds that:

$$f^V_{I^V_i}(b) = f^V_{I^V_{i+1}}(b) \tag{3.4}$$

then it holds that

$$\mu(f^V_{I^V_i})(b) = \mu(f^V_{I^V_{i+1}})(b). \tag{3.5}$$

Note that even though $b \notin I^V_i$ there is a value for $f^V_{I^V_i}(b)$ since $f^V_{I^V_i} : \mathbb{R} \to \mathbb{R}$. So this property states that for every two consecutive subintervals, if their truncated functions are equal on their 'overlapping' end point, this will persist under variation. Here 'overlapping' is not entirely correct, since the lower interval is open at that end point, but the property is still correctly defined.

One way of fulfilling this condition is found by shifting the left and the right side vertically to join the varied parameter. We will follow this strategy later in this chapter.

Shape The shape of the co-varied functions remains largely unchanged.

We see that, since variation might change the area of the varied parameter, some co-variation will be unavoidable. Still, we would rather not change the shape of the parameters too much, since we want to preserve a function as much as possible under

Table 3.1: Co-variation Schemes and Conditions

	Proportional	Shift&Normalise	Shift&Linear
Condition Area	Yes	Yes	Yes
Condition Positive	Yes	No	No
Condition Ratio	Yes	No	Yes
Condition Joining	No	Yes	Yes
Condition Shape	Yes	Yes	No
Condition Varied	Yes	No	Yes

co-variation. If we only shift and scale the co-varied parameters, we will consider this condition fulfilled.

Varied The varied parameter remains unchanged under the co-variation scheme.

It seems very reasonable we would want a co-variation scheme to not further vary the already varied function. We include this condition because we will look at a naive scheme that will not fulfil this.

This concludes the conditions we would like to impose upon our co-variation schemes. We will proceed discussing several schemes. In Table 3.1 we summarise how well the different schemes fulfil the conditions.

3.2 Proportional Co-variation

For our first co-variation, we consider proportional co-variation. Proportional co-variation is a method of co-variation in discrete sensitivity analysis. Chan and Darwiche [4] show that proportional co-variation in the discrete case minimises a distance from the original according to a distance measure they argue as very reasonable (and admittedly have invented themselves). With proportional co-variation, we want to multiply each function $f^V_{I^V}$ with a c that is constant given a fixed variation. We want to pick c in such a way that the proportion between each of the co-varied functions remains unchanged. But we want to pick c in a way that we fulfil Condition Area, i.e. the area of the varied potential is 1. We find this to be the case for:

$$c =_{\text{def}} \frac{1 - \int_{J^V} g(v)dv}{1 - \int_{J^V} f^V_{J^V}(v)dv}. \tag{3.6}$$

Wee see this c is constant for a given variation $\mu(f^V_{J^V}) = g$. We will show this c fulfils Condition Area in the analysis. So under proportional co-variation, we have that, for $I^V \neq J^V$:

$$\mu(f^V_{I^V})(v) = f^V_{I^V}(v) \cdot c \tag{3.7}$$

For example, we can see the effect of proportional co-variation on the example variation of Figure 3.2 in Figure 3.3.

3.2.1 Analysis

We will now analyse proportional co-variation using the conditions we have defined in Section 3.1.

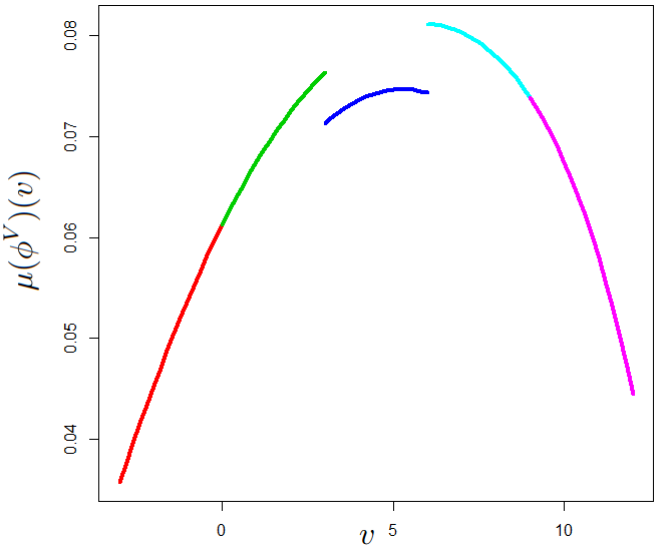


Figure 3.3: The potential of our running example, co-varied with proportional co-variation.

Firstly we will look at Condition Area. We want to show this condition indeed holds. To show this, we first recall a definition of a conditional density given in Equation 2.7. To properly show Condition Area holds, we also consider the following equivalent definition:

$$\sum_{I^V \in \mathbf{I}^V, I^V \neq J^V} \int_{I^V} f^V_{I^V}(t)dt = 1 - \int_{J^V} f^V_{J^V}(t)dt \tag{3.8}$$

Now we will shows the condition holds:

$$\int_{\Omega_V} \mu(\phi^V)(v)dv \tag{3.9}$$

$$= \sum_{I^V \in \mathbf{I}^V} \int_{I^V} \mu(f^V_{I^V})(v)dv \tag{3.10}$$

$$= \int_{J^V} g(v)dv + \sum_{I^V \in \mathbf{I}^V, I^V \neq J^V} \int_{I^V} \mu(f^V_{I^V}(v))dv \tag{3.11}$$

$$= \int_{J^V} g(v)dv + \sum_{I^V \in \mathbf{I}^V, I^V \neq J^V} \int_{I^V} f^V_{I^V}(v) \cdot \frac{1 - \int_{J^V} g(v)dv}{1 - \int_{J^V} f^V_{J^V}(v)dv} dv \tag{3.12}$$

$$= \int_{J^V} g(v)dv + \frac{1 - \int_{J^V} g(v)dv}{1 - \int_{J^V} f^V_{J^V}(v)dv} \cdot \sum_{I^V \in \mathbf{I}^V, I^V \neq J^V} \int_{I^V} f^V_{I^V}(v)dv \tag{3.13}$$

Now by the identity of Equation 3.8, we see that Equation 3.13 is equivalent to

$$\int_{J^V} g(v)dv + \frac{1 - \int_{J^V} g(v)dv}{1 - \int_{J^V} f^V_{J^V}(v)dv} \cdot \left(1 - \int_{J^V} f^V_{J^V}(v)dv\right) \tag{3.14}$$

$$= \int_{J^V} g(v)dv + 1 - \int_{J^V} g(v)dv = 1 \tag{3.15}$$

We therefore see that Condition Area holds for proportional co-variation.

We will continue with Condition Positive. We can assume we are going to vary a potential that is a probability density, so all parameters are positive within their respective subinterval and their areas in those subintervals sum to 1. We see that Condition Positive will be fulfilled if the parameters are multiplied by a positive constant, so if $c \geq 0$. We recall the definition of c:

$$c = \frac{1 - \int_{I^V} g(v)dv}{1 - \int_{J^V} f^V_{J^V}(v)dv}. \tag{3.16}$$

We see that the denominator of the fraction will be positive, since no original parameter can have an area larger than 1 in their respective subinterval. So c will be positive if the numerator is positive. We see this is the case if $\int_{I^V} g(v)dv \leq 1$. As explained in the definition of Condition Positive, this is a very reasonable assumption since otherwise it would be impossible to make the varied potential a probability density. So we consider Condition Positive fulfilled.

We will now look at the rest of the conditions, which are all not necessary but nice to have. Firstly we see that Condition Ratio is fulfilled. If we multiply all the parameters on the left side and on the right side with the same constant, then the ratio between the areas will be preserved since the constant will cancel out in the ratio. Condition Joining, however, is not fulfilled. This is clearly shown in Figure 3.3. We start of with a graph with all sides properly joined, but after proportional co-variation we see that the left side does not join the varied parameter, and the varied parameter does not join the right side. We also see Condition Shape is fulfilled, since we only scale the co-varied parameters by multiplying them with a constant. Finally, Condition Varied is fulfilled, since we do not alter the varied parameter under the co-variation scheme.

In conclusion, we see that only Condition Joining is not fulfilled under proportional co-variation. The next co-variation scheme we will consider will be a naive approach to fulfilling this condition.

3.3 Shift & Normalise

In Section 3.2, we have seen that proportional co-variation fulfils all our conditions except for Condition Joining. This means that if we vary a potential with all its parameters joined at their respective ends, this property is not necessarily preserved under proportional co-variation. We will look at a naive approach to preserve this condition.

When we want to fulfil Condition Joining, the easiest way to do this is to shift the left side and the right side to meet the varied parameter. We call this the *Shift action*. We for example see the result of parameter variation in Figure 3.2. The effect of Shift action is then seen in Figure 3.4. We see all parameters to the left of the varied parameter are all increased (or decreased) by the same constant, and all parameters on the right side all by another constant.

We want all of our schemes to at least fulfil Condition Area. For this scheme, we solve this using the *Normalise action*. This action means that we multiply the entire potential by one over its area, so that the resulting area is exactly equal to 1. The result of the Normalise action

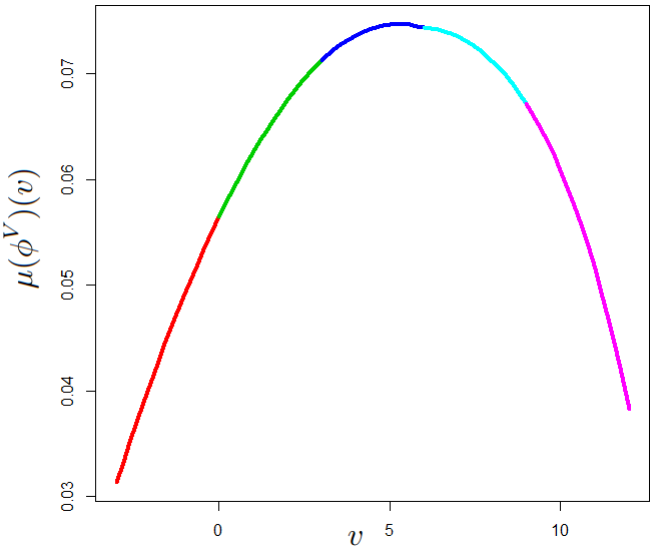


Figure 3.4: The potential of our running example, with variation and the Shift action.

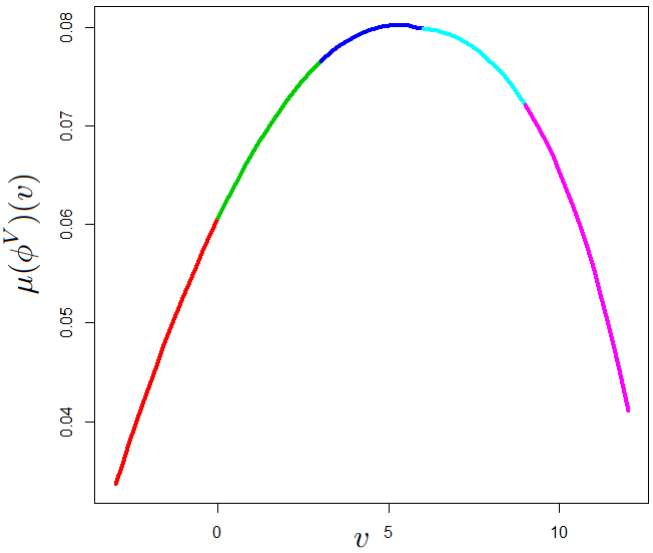


Figure 3.5: The potential of our running example, co-varied with the Shift&Normalise scheme.

applied to Figure 3.4 is shown in Figure 3.5.

The combination of the Shift action and the Normalise action is called the Shift&Normalise scheme.

3.3.1 Analysis

We will again analyse our co-variation scheme according to the conditions defined in Section 3.1.

Firstly we see Condition Area is met. The area under the co-varied potential is equal to one, since we end with a Normalise action. We note that the Normalise action will fail if, after the Shift action, the potential has an area of 0. This is quite a degenerate case however, and we will not see this in practice because it would almost certainly only occur if Condition Positive is violated.

We see Condition Positive is not necessarily fulfilled with this scheme. The Shift action could shift parts of the potential below the x-axis. Luckily it is easy to check for which variations the condition will be violated. We can look at the lowest point of the left side of the potential, and we can reason we can only shift the left part so that the lowest point meets the x-axis. So this gives us a direct limit to how low the leftmost point of the varied potential can be so we can join the left side with a Shift action while preserving the positivity of the left side of the potential. Likewise we can find a limit to how low the right most point of the varied potential can be so we preserve positivity of the right side of the potential.

We see Condition Ratio is also not met. While the Normalise action does preserve ratios, the Shift action does not. Condition Joining is preserved, however, since the Shifting action corrects the joining of left side to varied potential and varied potential to right side, and the Normalise action preserves the joining. We see Condition Shape is met, since all co-varied potentials are first shifted by a constant (Shift action) and then multiplied by a constant (Normalise action).

Lastly, we have Condition Varied. We see this condition is not met. While the Normalise action seems like a good and direct way to fulfil Condition Area, it does cause us to further change our varied parameter under co-variation. In some cases this might lead to our co-variation scheme 'resetting' the variation, as demonstrated in Figure 3.6.

In conclusion we see that a naive approach to fulfilling Condition Joining causes us to violate several other conditions. The violation of Condition Positive will not be too negative since we can at least easily verify the space of variations we can do without violating it. Condition Ratio is not maintained, which is unfortunate. More problematic is Condition Varied. We see our scheme can invalidate our original variation. We note that the scheme works well for variations that change the shape of the parameter, but works horribly for variations that shift the parameter. All in all, the Shift&Normalise scheme has too many problems to properly work with. In the next section, we will attempt to develop a scheme that addresses some of the issues.

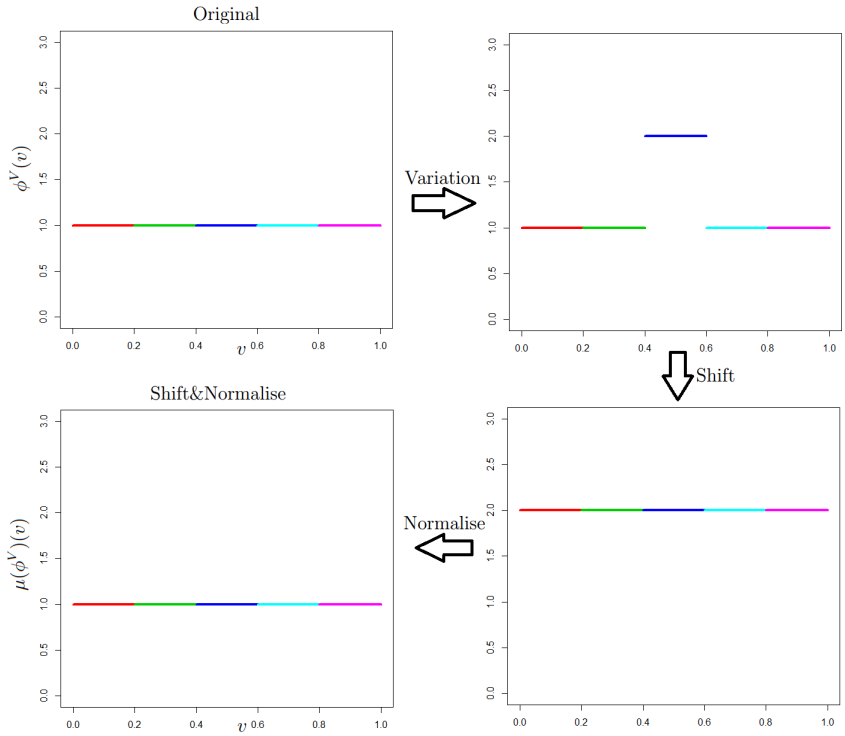


Figure 3.6: The Shift&Normalise scheme on an example potential, showing the scheme can lead to a full reset.

3.4 Shift & Linear

In the previous section, we have seen the Shift&Normalise scheme. We have seen that where this scheme does fulfil Condition Joining, it does quite badly at many of the other conditions. In this section, we will develop a scheme based on the Shift&Normalise scheme, that addresses some of the shortcomings.

First of all, we want our new scheme to fulfil Condition Ratio. Let us refer to the original area of the left and right side as L and R respectively. If we have the co-varied area of the left and right side equal to L' and R', with

$$L' = \frac{L \cdot (1 - \int_{J^V} g(v)dv)}{L + R}, R' = \frac{R \cdot (1 - \int_{J^V} g(v)dv)}{L + R}, \tag{3.17}$$

then we will fulfil both Conditions Area and Ratio. We will show this in the analysis.

The first step of our scheme is again going to be the Shift action as defined in Section 3.3.1. For the second step we want to add a linear function to the left side of the potential, and a linear function to the right side of the potential. We have two conditions we want to place on both linear functions. We will give the conditions for the linear function added to the left side, the right side is analogous:

- The area under the linear function plus the area under the shifted function should total L'. This way, we fulfil Conditions Area and Ratio as previously explained.

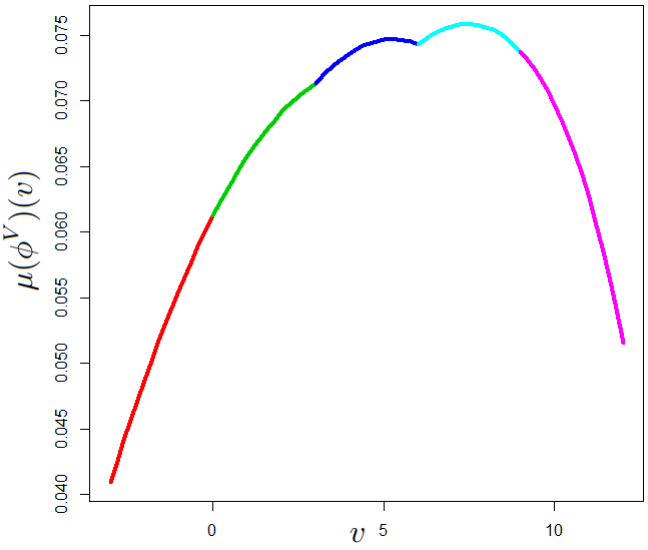


Figure 3.7: The potential of our running example, co-varied with the Shift&Linear scheme.

- The linear function should be 0 at the joining point with the varied parameter. This way we maintain Condition Joining, since the Shifting action has shifted the left side to the right place and the linear function will not shift it further.

Now we have a target area and a target value at one point, so we have exactly one linear function that fulfils both. We add this linear function to all the parameters on the left side, and add an analogous linear function to all the parameters on the right side. We call adding these linear functions the *Linear action*. We see the effect of the Linear action on the example of Figure 3.4 in Figure 3.7.

The combination of the Shift action and the Linear action is called the Shift&Linear scheme.

3.4.1 Analysis

We again will analyse this scheme through the conditions defined in Section 3.1.

Firstly we have Condition Area. In the Linear action we have made exactly sure that the new area of the left and right side is L' and R' respectively. We therefore observe that:

$$\int_{\Omega_V} \phi^V(v)dv = L' + R' + \int_{J^V} g(v)dv \tag{3.18}$$

$$= \frac{L \cdot (1 - \int_{J^V} g(v)dv)}{L+R} + \frac{R \cdot (1 - \int_{J^V} g(v)dv)}{L+R} + \int_{J^V} g(v)dv \tag{3.19}$$

$$= \frac{L+R}{L+R} \cdot (1 - \int_{J^V} g(v)dv) + \int_{J^V} g(v)dv \tag{3.20}$$

$$= 1 - \int_{J^V} g(v)dv + \int_{J^V} g(v)dv = 1 \tag{3.21}$$

So Condition Area is met.

Condition Positive is quite problematic in this scheme. Not only is it possibly violated in this scheme, it is also hard to check if it is violated as both the Shift action and the Linear action might be the cause. Because the Linear action is more intricate than the Shift action, it is harder to check for legal variations with this scheme than it was with the Shift&Normalise scheme.

For Condition Ratio, we will again look at the mathematics. We see:

$$\frac{L'}{R'} = \frac{L \cdot (1 - \int_{JV} g(v)dv)/(L+R)}{R \cdot (1 - \int_{JV} g(v)dv)/(L+R)} = \frac{L}{R} \tag{3.22}$$

so the ratio is kept constant under co-variation.

Condition Joining is fulfilled, as the Shift action joins the left and right side with the varied parameter and the Linear action does not disturb these join points. Condition Varied is also fulfilled, as both actions leave the varied potential undisturbed. Condition Shape is violated however, as we add a linear function to the left and right side. This is not just a shifting or scaling of the co-varied parameters, and if we compare Figure 3.1 with Figure 3.7, we see scheme disturbs the parameters quite a lot in their shape.

In conclusion, we find that this scheme does a lot better than the Shift&Normalise we tried to improve upon, so we have done good on that part. While it is unfortunate that Condition Shape is violated, we do not think this invalidates our scheme and do think one could come up with likewise schemes that disturb the original shape less. The big problem, however, is Condition Positive. It seems impossible to fulfil Condition Joining without some form of shifting, but any shifting will possibly make the new potential negative at some points. Even worse, the more intricate we make our scheme the harder it is to check for which variations we will violate Condition Positive. So we see it is extremely hard to combine the 'nice to have' Condition Joining with the necessary Condition Positive.

3.5 Conclusion

In conclusion, we have introduced the concept of co-variation and have set some conditions we would want our co-variation schemes to meet. The first scheme we have seen, Proportional Co-variation, has done great for almost all conditions. The big problem with the scheme was that, if the original potential was a nice function without gaps, then after co-variation there might be gaps between the varied parameter and the left and right side of the potential. A naive solution to this problem was found in the Shift&Normalise scheme. This solution violated many important conditions however. Many of these violations were solved in the Shift&Linear scheme. This scheme had a minor problem in that not the entire original shape was preserved, and the major problem that the varied potential is not necessarily a positive function. It was argued that it might be impossible to find a good co-variation scheme that gives a positive potential where the varied parameter joins the left and right side.

So Proportional Co-variation might be the best scheme we can find for co-variation in the MoTBF context. While it does not fulfil a condition that would be nice to have, it might be

no scheme can fulfil this condition without violating a necessary condition, and if the necessary conditions are violated we do not necessarily have a probability density under co-variation. Therefore we will use Proportional Co-variation as our selected method of Co-variation through the rest of the thesis.

Chapter 4

Variation Expression

For our sensitivity analysis, we want to know the effect of variations. This means that if we vary a parameter of the MoTBF-network, we want to know what the effect is on the prior and posterior probabilities of events in the network. So we would want to find an expression for those probabilities in terms of the variation function μ we introduced in Chapter 3. We call these expressions the *variation expressions* of those probabilities. We want to find these expressions for arbitrary variations of the parameters. In the next chapter we will look at families of specific variations which will be the sensitivity functions and will be the core of our sensitivity analysis, and we will do this by using the variation expressions we will develop here for general variation.

The variation expressions of subintervals in the US MoTBF-Networks (as defined in Section 2.4.1) will first be developed in Section 4.1. This is a good stepping stone to then develop the variation expressions of subintervals in the general MoTBF-Networks in Section 4.2. For the general MoTBF-Network we will then continue to find variation expressions for arbitrary marginal probabilities in the network.

4.1 US MoTBF

We consider a US MoTBF-network $(G, \phi^{\mathbf{V}})$. We recall that a US MoTBF-network is a special case of the general MoTBF-network in that it has only one subinterval set associated with each variable. We want to vary a certain parameter of the network, say $f^A_{J^A, J^{\rho(A)}}$. We remind ourselves that A is a variable, J^A is a subinterval of variable A, and $J^{\rho(A)}$ is a set of subintervals of the parents $\rho(A)$ of A. We adapt it according to Equation 3.1. To reiterate:

$$\mu(f^A_{J^A, J^{\rho(A)}})(a) = g(a)$$

where μ is the function mapping all parameters to their counterpart under variation of the single parameter $f^A_{J^A, J^{\rho(A)}}$, and $g(a)$ is exactly this single varied parameter. For co-variation we use proportional co-variation, as defined in Section 3.2. We will use proportional co-variation throughout the chapter without specific mention. As in the previous chapter, we write μ to indicate the effect of variation on the different truncated functions of the network. Since the potentials and conditional potentials of the network are defined in terms of those truncated functions, we now also use μ to indicate the variation of these potentials and conditional potentials. So for example, the conditional potential $\mu(\phi^{A|\rho(A)})$ is the conditional potential $\phi^{A|\rho(A)}$

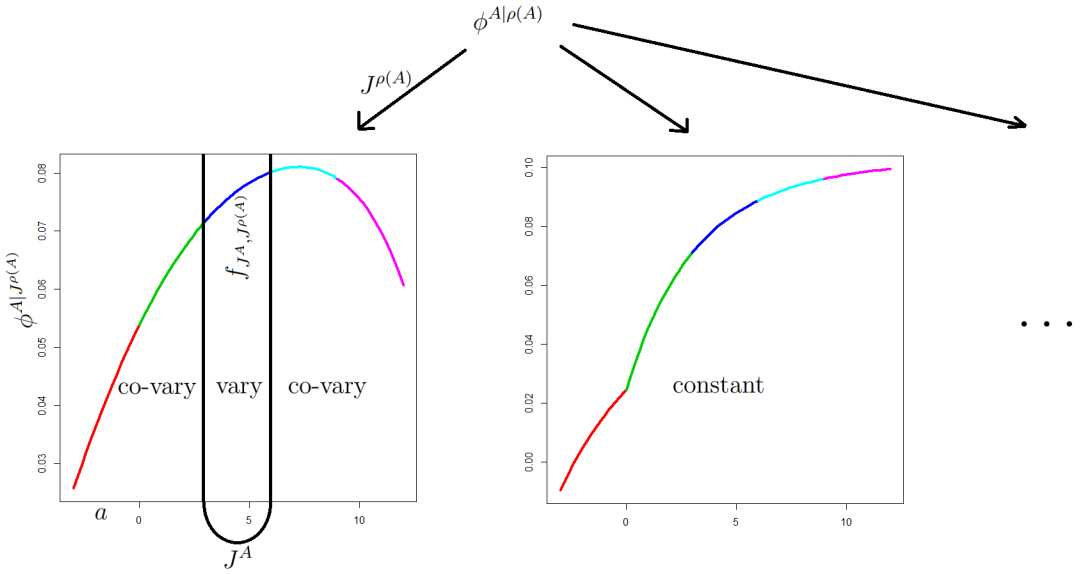


Figure 4.1: An example of the conditional potential $\phi^{A|\rho(A)}$ varied on $J^{\rho(A)}$ and J^A.

with each of its truncated functions mapped by μ, i.e. it is $\phi^{A|\rho(A)}$ under the specified variation. Likewise, since we write probabilities in terms of potentials, we write $\mu(P)$ to indicate the probability function under the specified variation.

The effect of variation on an example conditional potential $\phi^{A|\rho(A)}$ is shown in Figure 4.1. The potential $\phi^{A|J^{\rho(A)}}$, which is the conditional potential $\phi^{A|\rho(A)}$ conditioned on the set of parent subintervals $J^{\rho(A)}$, is varied in the subinterval J^A and co-varied on the other parts of the domain Ω_A. It is also shown in the figure that for any other $I^{\rho(A)} \in \mathbf{I}^{\rho(A)}, I^{\rho(A)} \neq J^{\rho(A)}$, the potential $\phi^{A|I^{\rho(A)}}$ is constant under variation. Suppose we want to know, for certain target variables $\mathbf{T} \subseteq \mathbf{V}$, what the prior probability is that our variables fall into subintervals $I^{\mathbf{T}}$ (i.e. that each variable $T \in \mathbf{T}$ falls into some subinterval $I^T \in I^{\mathbf{T}}$). We want to write $P(\mathbf{T} \in I^{\mathbf{T}})$ in terms of the variation μ.

We use $\mathbf{N} = \mathbf{V} - \mathbf{T}$ to indicate all non-target variables, and $\mathbf{V}' = \mathbf{V} - \{A\}$ to indicate all variables minus the variable that has its parameter varied. When we have subintervals for the target variables $I^{\mathbf{T}}$ and subintervals for the non-target variables $I^{\mathbf{N}}$ we write $I^{\mathbf{V}} = I^{\mathbf{T}} \times I^{\mathbf{N}}$. In the proofs we will use the notation $O(1)$ to indicate any constant under μ; this makes the proofs easier to read than if we would have used a variety of different constants.

Theorem 4.1.1. *Given a US MoTBF-network $(G, \phi^{\mathbf{V}})$ and a set of target variables $\mathbf{T} \subseteq \mathbf{V}$. Suppose we vary a network parameter $f^A_{J^A,J^{\rho(A)}}$ for some $A \in \mathbf{V}$, $J^A \in \mathbf{I}^A$ and $J^{\rho A} \in \mathbf{I}^{\rho(A)}$ as:*

$$\mu(f^A_{J^A,J^{\rho(A)}})(a) = g(a).$$

We can then write the prior probability of the variables in $\mathbf{T}$ falling into subintervals $I^{\mathbf{T}} \in \mathbf{I}^{\mathbf{T}}$ as:

$$\mu(P)(\mathbf{T} \in I^{\mathbf{T}}) = c_1 + c_2 \int_{J^A} g(a)da \tag{4.1}$$

where $c_1, c_2 \in \mathbb{R}$ are constants with respect to μ.

Proof: We will first break down the prior probability into a summation of multiple parts, and then show that each of these parts can be written down in a certain general form. So firstly:

$$\mu(P)(\mathbf{T} \in I^{\mathbf{T}}) \tag{4.2}$$

(we write the probability as a potential)

$$= \int_{I^{\mathbf{T}}} \mu(\phi^{\mathbf{T}})(\mathbf{t}) d\mathbf{t} \tag{4.3}$$

(we marginalise out the non-target variables)

$$= \int_{I^{\mathbf{T}}} \int_{\Omega^{\mathbf{N}}} \mu(\phi^{\mathbf{V}})(\mathbf{t}, \mathbf{n}) d\mathbf{n} d\mathbf{t} \tag{4.4}$$

(we rewrite using Equation 2.4)

$$= \int_{I^{\mathbf{T}}} \int_{\Omega^{\mathbf{N}}} \prod_{V \in \mathbf{V}} \mu(\phi^{V|\rho(V)})(v|\rho(v)) d\mathbf{n} d\mathbf{t} \tag{4.5}$$

(we split the domain of $\mathbf{N}$ into subintervals, and split ϕ into the truncated functions)

$$= \int_{I^{\mathbf{T}}} \sum_{I^{\mathbf{N}} \in \mathbf{I}^{\mathbf{N}}} \int_{I^{\mathbf{N}}} \prod_{V \in \mathbf{V}} \mu(f^{V}_{I^{V}, I^{\rho(V)}})(v) d\mathbf{n} d\mathbf{t} \tag{4.6}$$

(we take the summation out, which is independent of $\mathbf{t}$)

$$= \sum_{I^{\mathbf{N}} \in \mathbf{I}^{\mathbf{N}}} \int_{I^{\mathbf{T}}} \int_{I^{\mathbf{N}}} \prod_{V \in \mathbf{V}} \mu(f^{V}_{I^{V}, I^{\rho(V)}})(v) d\mathbf{n} d\mathbf{t} \tag{4.7}$$

(we merge $\mathbf{T}$ and $\mathbf{N}$ into $\mathbf{V}$)

$$= \sum_{I^{\mathbf{N}} \in \mathbf{I}^{\mathbf{N}}} \int_{I^{\mathbf{V}}} \prod_{V \in \mathbf{V}} \mu(f^{V}_{I^{V}, I^{\rho(V)}})(v) d\mathbf{v} \tag{4.8}$$

(we split $\mathbf{V}$ into A and $\mathbf{V}'$)

$$= \sum_{I^{\mathbf{N}} \in \mathbf{I}^{\mathbf{N}}} \int_{I^{A}} \int_{I^{\mathbf{V}'}} \mu(f^{A}_{I^{A}, I^{\rho(A)}})(a) \prod_{V' \in \mathbf{V}'} f^{V'}_{I^{V}, I^{\rho(V)}}(v') d\mathbf{v}' da \tag{4.9}$$

(we move the truncated function out of the integral)

$$= \sum_{I^{\mathbf{N}} \in \mathbf{I}^{\mathbf{N}}} \int_{I^{A}} \mu(f^{A}_{I^{A}, I^{\rho(A)}})(a) \int_{I^{\mathbf{V}'}} \prod_{V' \in \mathbf{V}'} f^{V'}_{I^{V}, I^{\rho(V)}}(v') d\mathbf{v}' da \tag{4.10}$$

(the integral over $I^{\mathbf{V}'}$ is constant in both μ and a. Because of the former we can make it $O(1)$, and because of the latter we can move it out of the integral over I^{A})

$$= \sum_{I^{\mathbf{N}} \in \mathbf{I}^{\mathbf{N}}} O(1) \int_{I^{A}} \mu(f^{A}_{I^{A}, I^{\rho(A)}})(a) da \tag{4.11}$$

This leaves us a sum of integrals over functions $\mu(f^{A}_{I^{A}, I^{\rho(A)}})$, multiplied by a respective constant factor. In each component of the summation we can have one of three cases:

Case 1 The function $\mu(f^{A}_{I^{A}, I^{\rho(A)}})$ under consideration is exactly the one we are varying, i.e.

$J^{\rho(A)} = I^{\rho(A)}$ and also $J^A = I^A$. So per definition:

$$O(1) \int_{I^A} \mu(f^A_{I^A, I^{\rho(A)}})(a)da \tag{4.12}$$

$$= O(1) \int_{J^A} g(a)da \tag{4.13}$$

Case 2 Our function $\mu(f^A_{I^A, I^{\rho(A)}})$ is on the same parent subinterval as the function we are varying, but it is not equal to it. So we have that $J^{\rho(A)} = I^{\rho(A)}$ but $J^A \neq I^A$. In this case we are co-varying the function. As argued in Section 3.5, we have decided to use the Proportional Co-variation scheme as defined in Section 3.2. Using Equation 3.6, we then see:

$$O(1) \int_{I^A} \mu(f^A_{I^A, I^{\rho(A)}})(a)da = O(1) \int_{I^A} f^A_{I^A, I^{\rho(A)}}(a) \cdot c \ \ da \tag{4.14}$$

$$= O(1) \int_{I^A} f^A_{I^A, I^{\rho(A)}}(a) \cdot \frac{1 - \int_{J^A} g(a)da}{1 - \int_{J^A} f^A_{J^A, I^{\rho(A)}}(a)da} da \tag{4.15}$$

$$= O(1) \frac{1 - \int_{J^A} g(a)da}{1 - \int_{J^A} f^A_{J^A, I^{\rho(A)}}(a)da} \int_{I^A} f^A_{I^A, I^{\rho(A)}}(a)da \tag{4.16}$$

$$= O(1) \frac{1 - \int_{J^A} g(a)da}{O(1)} \tag{4.17}$$

$$= O(1) + O(1) \int_{J^A} g(a)da \tag{4.18}$$

Case 3 We have that our function $f^A_{I^A, I^{\rho(A)}}$ is not on the same parent subinterval as the parameter we are varying. In this case we are neither varying nor co-varying the truncated function, so the function is constant under μ. So $J^{\rho(A)} \neq I^{\rho(A)}$. In this case we see:

$$O(1) \int_{I^A} \mu(f^A_{I^A, I^{\rho(A)}})(a)da \tag{4.19}$$

$$= O(1) \int_{I^A} f^A_{I^A, I^{\rho(A)}}(a)da = O(1) \tag{4.20}$$

Any of these final forms can be written as $O(1) + O(1) \int_{J^A} g(a)da$. So we sum over parts that we can write as $O(1) + O(1) \int_{J^A} g(a)da$. The summation of these parts can then also be written as $O(1) + O(1) \int_{J^A} g(a)da$, i.e. $c_1 + c_2 \int_{J^A} g(a)da$ for some $c_1, c_2 \in \mathbb{R}$. □

4.1.1 Examples

We will calculate an entire variation expression in an example network. Then we will show part of the calculation of a different variation expression, chosen so we will see examples of all the cases of the proof of Theorem 4.1.1.

Example 4.1.2. We will use the network specified in Example 2.3.1, with the alteration that $I_1^{A \to C} = [0, 1)$ and $I_2^{A \to C} = [1, 2)$. This makes all the subinterval sets of A equal, and since

the other two variables, B and C, are leaves so only have one associated subinterval set, the resulting network is a US MoTBF-Network.

Say we vary $\mu(f^A_{I^A_1}(a)) = g(a)$, and we are interested in the variation expression of $P(B \in I^B_2)$ in terms of this variation. We will write out this probability, just like in the proof of Theorem 4.1.1.

$$\begin{aligned}
&\mu(P)(B \in I^B_2) \\
=&\sum_{I^{A,C} \in \mathbf{I}^{A,C}} \int_{I^A} \mu(f^A_{I^A})(a) \int_{I^B_2} \int_{I^C} f^B_{I^B_2, I^A}(b) f^C_{I^C, I^A}(c) dc\ db\ da
\end{aligned}$$

We note that this summation consists of four parts, for each element in $\mathbf{I}^{A,C} = \{\{I^A_1, I^C_1\}, \{I^A_1, I^C_2\}, \{I^A_2, I^C_1\}, \{I^A_2, I^C_2\}\}$. We will look at all four parts individually. Firstly we consider I^A_1 and I^C_1. We then see:

$$\begin{aligned}
&\int_{I^A_1} \mu(f^A_{I^A_1})(a) \int_{I^B_2} \int_{I^C_1} f^B_{I^B_2, I^A_1}(b) f^C_{I^C_1, I^A_1}(c) dc\ db\ da \\
=&0.308 \cdot \int_{I^A_1} \mu(f^A_{I^A_1})(a) da \\
=&0.308 \cdot \int_{I^A_1} g(a, x) da
\end{aligned}$$

This corresponds to Case 1 of the cases in our proof of Theorem 4.1.1. Likewise, if we consider I^A_1 and I^C_2, we get:

$$\begin{aligned}
&\int_{I^A_1} \mu(f^A_{I^A_1})(a) \int_{I^B_2} \int_{I^C_2} f^B_{I^B_2, I^A_1}(b) f^C_{I^C_2, I^A_1}(c) dc\ db\ da \\
=&0.275 \cdot \int_{I^A_1} g(a) da
\end{aligned}$$

For a more complicated expression, we consider I_2^A and I_1^C. We see:

$$\begin{aligned}
&\int_{I_2^A} \mu(f^A_{I_2^A})(a_1) \int_{I_2^B} \int_{I_1^C} f^B_{I_2^B, I_2^A}(b) f^C_{I_1^C, I_2^A}(c) dc\ db\ da_1 \\
=&0.247 \cdot \int_{I_2^A} \mu(f^A_{I_2^A})(a_1) da_1 \\
=&0.247 \cdot \int_{I_2^A} f^A_{I_2^A}(a_1) \cdot \frac{1 - \int_{I_1^A} g(a_2) da_2}{1 - \int_{I_1^A} f^A_{A_1}(a_3) da_3} da_1 \\
=&0.247 \cdot \frac{1 - \int_{I_1^A} g(a_2) da_2}{1 - \int_{I_1^A} f^A_{A_1}(a_3) da_3} \int_{I_2^A} f^A_{I_2^A}(a_1) da_1 \\
=&0.247 \cdot 0.287 \cdot \frac{1 - \int_{I_1^A} g(a_2) da_2}{1 - \int_{I_1^A} f^A_{A_1}(a_3) da_3} \\
=&0.07 \cdot \frac{1 - \int_{I_1^A} g(a_2) da_2}{0.713} \\
=&0.1 + -0.1 \cdot \int_{I_1^A} g(a) da
\end{aligned}$$

This corresponds to Case 2. Likewise, if we consider I_2^A and I_2^C, we see:

$$\begin{aligned}
&\int_{I_2^A} \mu(f^A_{I_2^A})(a) \int_{I_2^B} \int_{I_1^C} f^B_{I_2^B, I_2^A}(b) f^C_{I_1^C, I_2^A}(c) dc\ db\ da \\
=&0.14 + -0.14 \cdot \int_{I_1^A} g(a) da
\end{aligned}$$

Now that we have all the parts of the summation, we can sum them together. We then get:

$$\begin{aligned}
&\mu(P)(B \in I_2^B) \\
=&0.308 \cdot \int_{I_1^A} g(a) da + 0.275 \cdot \int_{I_1^A} g(a) da + 0.1 \\
&+ -0.1 \cdot \int_{I_1^A} g(a) da + 0.14 + -0.14 \cdot \int_{I_1^A} g(a) da \\
=&0.24 + 0.083 \int_{I_1^A} g(a) da
\end{aligned}$$

This result is in the general form of Theorem 4.1.1.

Example 4.1.3. We have seen an example of Case 1 and Case 2 of Theorem 4.1.1, but no example of Case 3. For this we consider a new example. We vary $\mu(f^C_{I_1^C, I_2^A})(c) = g(c)$, and we are again interested in the variation expression of $P(B \in I_2^B)$ in terms of this variation. We

find:

$$\begin{aligned}&\mu(P)(B \in I_2^B)\\ =&\sum_{I^{A,C} \in \mathbf{I}^{A,C}} \int_{I^C} \mu(f^C_{I^C,I^A})(c) \int_{I_2^B} \int_{I^A} f^B_{I_2^B,I^A}(b) f^A_{I^A}(a) da\ db\ dc\end{aligned}$$

Now we consider only one part of the summation, namely where the subintervals of the summation are I_1^A and I_1^C. We find:

$$\begin{aligned}&\int_{I_1^C} \mu(f^C_{I_1^C,I_1^A})(c) \int_{I_2^B} \int_{I_1^A} f^B_{I_2^B,I_1^A}(b) f^A_{I_1^A}(a) da\ db\ dc\\ =&\int_{I_1^C} f^C_{I_1^C,I_1^A}(c) \int_{I_2^B} \int_{I_1^A} f^B_{I_2^B,I_1^A}(b) f^A_{I_1^A}(a) da\ db\ dc\\ =&0.287 \cdot 0.584 \cdot 0.527 = 0.088\end{aligned}$$

We see this corresponds with Case 3. We leave our example at this, because we have now seen all cases.

4.2 General MoTBF

In this section, we will develop the variation expression for a general MoTBF-network. The main difference between the general MoTBF-network and the US MoTBF-network we have seen previously, is that different subinterval sets may be associated with the variable A of which we vary a parameter. This leads us to work with the intersection of the different subinterval sets, i.e. with elements of the set I^{A+} (as defined in Section 2.3).

We will first develop variation expressions for prior probabilities of subintervals, and then generalise these results so we can find a variation expression for any marginal probability of the network.

4.2.1 Subinterval probabilities

Consider a MoTBF-Network $(G, \phi^{\mathbf{V}})$. We vary a parameter of the network, say $f^A_{J^A,J^{\rho(A)}}$. The main difference between the US MoTBF-Network and the general MoTBF-Network is that we have a subinterval set $\mathbf{I}^{A+}$ in the general network that is not equal to $\mathbf{I}^A$. This set was defined in Equation 2.9. Therefore we are now interested in the pieces of $\mathbf{I}^{A+}$ that make up J^A. We define:

$$\mathbf{J}^{A+} =_{\text{def}} \{I^{A+} | I^{A+} \in \mathbf{I}^{A+} \text{ and } d(I^{A+}) = J^A\} \tag{4.21}$$

An example of this set is shown in Figure 4.2. We note that $\mathbf{J}^{A+}$ is a subinterval set over the interval J^A.

Suppose we want to know, for certain target variables $\mathbf{T} \subseteq \mathbf{V}$, the variation expression of the prior probability that our variables fall into subintervals $I^{\mathbf{T}+}$ for $I^{\mathbf{T}+} \in \mathbf{I}^{\mathbf{T}+}$.

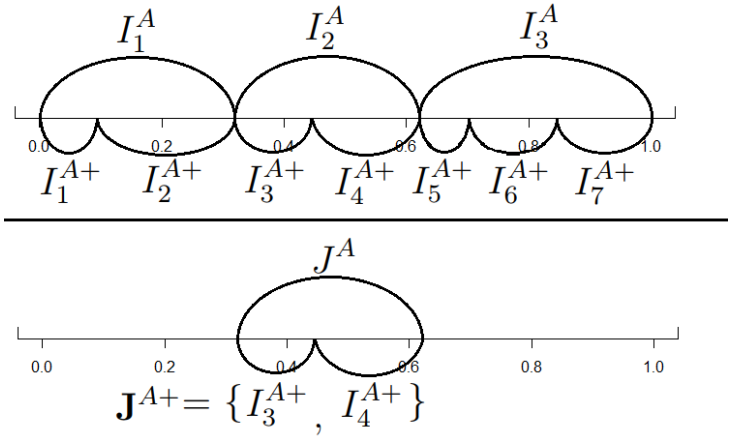


Figure 4.2: Top image: An example of the elements of subinterval sets $\mathbf{I}^A$ and $\mathbf{I}^{A+}$ for variable A on the interval $\Omega_A = [0, 1)$. Bottom image: The set $\mathbf{J}^{A+}$ for the example in the top image, for $J^A = I_2^A$.

Theorem 4.2.1. *Given an MoTBF-network $(G, \phi^{\mathbf{V}})$ and a set of target variables $\mathbf{T} \subseteq \mathbf{V}$. Suppose we vary a network parameter $f^A_{J^A, J^{\rho(A)}}$ for some $A \in \mathbf{V}$, $J^A \in \mathbf{I}^A$ and $J^{\rho(A)} \in \mathbf{I}^{\rho(A)}$ as:*

$$\mu(f^A_{J^A, J^{\rho(A)}})(a) = g(a).$$

We co-vary the other parameters with proportional co-variation. We can then write the prior probability of the variables in $\mathbf{T}$ falling into subintervals $I^{\mathbf{T}+} \in \mathbf{I}^{\mathbf{T}+}$ as a variation expression of μ as:

$$\mu(P)(\mathbf{T} \in I^{\mathbf{T}+}) \tag{4.22}$$

$$= c_0 + \sum_{J^{A+} \in \mathbf{J}^{A+}} c_{J^{A+}} \int_{J^{A+}} g(a)da \tag{4.23}$$

where $c_0 \in \mathbb{R}$ and $c_{J^{A+}} \in \mathbb{R}$ (for all $J^{A+} \in \mathbf{J}^{A+}$) are constants under μ.

Proof: Our proof will greatly resemble the proof of Theorem 4.1.1. Firstly we will break down the prior probability in a summation. The steps in between are omitted and can be found in Appendix A.2.1.

$$\mu(P)(\mathbf{T} \in I^{\mathbf{T}+}) \tag{4.24}$$

$$= \sum_{I^{\mathbf{N}+} \in \mathbf{I}^{\mathbf{N}+}} O(1) \int_{I^{A+}} \mu(\phi^{A|I^{\mathbf{V}+}})(a)da \tag{4.25}$$

This leaves us a sum of integrals over functions $\mu(\phi^{A|I^{\mathbf{V}+}})$, multiplied by a respective constant factor.

Inside the summation, we sum over some $I^{N+} \in \mathbf{I}^{N+}$ and we are considering some target subintervals $I^{T+} \in \mathbf{I}^{T+}$. We consider the combination of these, $I^{V+} =_{\text{def}} I^{N+} \cup I^{T+}$. In the proof it is shown that $I^{A+} \in I^{V+}$, this is the subinterval of the varied potential. Since the

potential is varied on a subinterval of $\mathbf{I}^A$, not $\mathbf{I}^{A+}$, we are interested in the subinterval in $\mathbf{I}^A$ of which I^{A+} is a subset. This is the subinterval $I^A =_{\text{def}} d_A(I^{A+})$. Likewise, we also have subintervals of each of the parents of A in the set $I^{\rho(A)+} \subset I^{V+}$, but we are interested in the subintervals in $\mathbf{I}^{\rho(A)}$ of which the subintervals $I^{\rho A+}$ are subsets. This is set of parent subintervals $I^{\rho(A)} =_{\text{def}} \{d_A(I^{P+}) | I^{P+} \in I^{\rho(A)+}\}$. Here each P is a parent variable of A, i.e. $P \in \rho(A)$.

In each component of the summation we can have one of three cases:

Case 1 The function $\mu(\phi^{A|I^{\mathbf{V}+}})$ under consideration is part of the domain of the function we are varying, i.e. $J^{\rho(A)} = I^{\rho(A)}$ and also $J^A = I^A$. So:

$$O(1) \int_{I^{A+}} \mu(\phi^{A|I^{\mathbf{V}+}})(a)da \tag{4.26}$$

$$=O(1) \int_{I^{A+}} g(a)da \tag{4.27}$$

Also, we have that $I^{A+} \in \mathbf{J}^{A+}$.

Case 2 The domain of our function $\mu(\phi^{A|I^{\mathbf{V}+}})$ is defined for the same parents as our varied function, but the domain is not a sub-domain of the domain of the varied function. So we have that $J^{\rho(A)} = I^{\rho(A)}$ but $J^A \neq I^A$. In this case we are co-varying the function. We again use proportional co-variation with the constant c as defined in Equation 3.6.

$$O(1) \int_{I^{A+}} \mu(\phi^{A|I^{\mathbf{V}+}})(a)da = O(1) \int_{I^{A+}} \phi^{A|I^{\mathbf{V}+}}(a) \cdot c \;\; da \tag{4.28}$$

$$=O(1) \int_{I^{A+}} \phi^{A|I^{\mathbf{V}+}}(a) \cdot \frac{1 - \int_{J^A} g(a)da}{1 - \int_{J^A} \phi^{A|J^V, J^{\rho(v)}}(a)da} da \tag{4.29}$$

$$=O(1) \frac{1 - \int_{J^A} g(a)da}{1 - \int_{J^A} \phi^{A|J^V, J^{\rho(v)}}(a)da} \int_{I^{A+}} \phi^{A|I^{\mathbf{V}+}}(a)da \tag{4.30}$$

$$=O(1) \frac{1 - \int_{J^A} g(a)da}{O(1)} \tag{4.31}$$

$$=O(1) + O(1) \int_{J^A} g(a)da \tag{4.32}$$

$$=O(1) + O(1) \left(\sum_{J^{A+} \in \mathbf{J}^{A+}} \int_{J^{A+}} g(a)da \right) \tag{4.33}$$

$$=O(1) + \sum_{J^{A+} \in \mathbf{J}^{A+}} O(1) \int_{J^{A+}} g(a)da \tag{4.34}$$

$$\tag{4.35}$$

Case 3 We have that our function $\mu(\phi^{A|I^{\mathbf{V}+}})$ is not on the same parent subinterval as the parameter we are varying. In this case we are neither varying nor co-varying the truncated function. So $J^{\rho(A)} \neq I^{\rho(A)}$. In this case we see:

$$O(1) \int_{I^{A+}} \mu(\phi^{A|I^{\mathbf{V}+}})(a)da \tag{4.36}$$

$$=O(1) \int_{I^{A+}} \phi^{A|I^{\mathbf{V}+}}(a)da = O(1) \tag{4.37}$$

So we sum over parts that we can write, in any case, as equations of the form of Equation 4.23. The summation of these parts can then also be written as an equation of the form of Equation 4.23. □

4.2.2 Marginal probabilities

For our sensitivity analysis, we want to know the probabilities of the values of target variables falling into intervals, as expressions of the variation in the network. We have seen such expressions for prior probabilities of a specific type of intervals (i.e. the probabilities $\mu(P)(\mathbf{T} \in I^{\mathbf{T}+})$). But we want to find variation expressions for probabilities of arbitrary intervals, and we want to find both prior and posterior probabilities. We will state some corollaries for these expressions, and will sketch the idea for proofs heavily based on the proof of Theorem 4.2.1.

We again consider an MoTBF-Network $(G, \phi^{\mathbf{V}})$, and vary a network parameter $f^A_{J^A, J^{\rho(A)}}$ for some $A \in \mathbf{V}$, $J^A \in \mathbf{I}^A$, $J^{\rho(A)} \in \mathbf{I}^{\rho(A)}$, and we use co-variation to co-vary the other parameters.

Corollary 4.2.2. Given a target variable $T \in \mathbf{V}$ such that $A \neq T$. Given some interval U^T such that $U^T \subseteq I^{T+}$ for some $I^{T+} \in \mathbf{I}^{T+}$. Then we can write the probability

$$\mu(P)(T \in U^T) \tag{4.38}$$

as a variation expression of the form of Equation 4.23.

The proof of this corollary is given in Appendix A.2.2.

Corollary 4.2.2 gives us variation expressions for prior probabilities of intervals U^T such that $U^T \subseteq I^{T+}$ for some $I^{T+} \in \mathbf{I}^{T+}$. The next corollary will give us prior probabilities of arbitrary intervals $U^T \subseteq \Omega_T$.

Corollary 4.2.3. Let A, T be as before, with $A \neq T$. Given some interval $U^T \subseteq \Omega_T$. Then we can write the probability

$$\mu(P)(T \in U^T) \tag{4.39}$$

as a variation expression of the form of Equation 4.23.

This follows directly from Corollary 4.2.2. The interval U^T can be split into smaller intervals such that each is a subset of some $I^{T+} \in \mathbf{I}^{T+}$, since the subinterval set $\mathbf{I}^{T+}$ covers the interval Ω_T. The probabilities for each of these intervals can then be written as a variation expression of the form of Equation 4.23 by Corollary 4.2.2, and the sum of these variation expressions will then also be of the form of Equation 4.23. □

Corollary 4.2.4. Let A, T be as before, with $A \neq T$. Given some observed variables $\mathbf{E} \subseteq \mathbf{V}$, such that $A \notin \mathbf{E}$, and their observed values $\mathbf{e} \in \Omega_{\mathbf{E}}$. Given some interval U^T such that $U^T \subseteq I^{T+}$ for some $I^{T+} \in \mathbf{I}^{T+}$. Then we can write the probability

$$\mu(P)(T \in U^T | \mathbf{E} = \mathbf{e}) \tag{4.40}$$

as a variation expression of the following form:

$$\frac{c_1 + \sum_{J^{A+} \in \mathbf{J}^{A+}} c^1_{J^{A+}} \int_{J^{A+}} g(a)da}{c_2 + \sum_{J^{A+} \in \mathbf{J}^{A+}} c^2_{J^{A+}} \int_{J^{A+}} g(a)da} \tag{4.41}$$

where $c_1, c_2 \in \mathbb{R}$ and $c^x_{J^{A+}} \in \mathbb{R}$ (for all $J^{A+} \in \mathbf{J}^{A+}$) are constants under μ.

Proof: First we assume that $T \notin \mathbf{E}$. (If $T \in \mathbf{E}$ then the probability of T is fixed so constant under variation, so the corollary holds.) We can write the probability as follows:

$$\mu(P)(T \in U^T | \mathbf{E} = \mathbf{e}) = \tag{4.42}$$

$$= \int_{U^T} \mu(\phi^{T|E})(t, e)dt \tag{4.43}$$

$$= \int_{U^T} \frac{\mu(\phi^{T,E})(t, e)}{\mu(\phi^E)(e)}dt \tag{4.44}$$

$$= \int_{U^T} \mu(\phi^{T,E})(t, e)dt \cdot \frac{1}{\mu(\phi^E(e))} \tag{4.45}$$

We note that we can write both $\int_{U^T} \mu(\phi^{T,E})(t, e)dt$ and $\mu(\phi^E(e))$ as variation expressions of the form of Equation 4.23. This is shown for $\int_{U^T} \mu(\phi^{T,E})(t, e)dt$ in Appendix A.2.3, the rewriting for $\mu(\phi^E(e))$ goes entirely analogous except with no variable T. This means that we can write Equation 4.45 as a fraction with a nominator and a denominator both of the form of Equation 4.23. This fraction has the form of Equation 4.41. □

Corollary 4.2.5. Let A, T be as before, with $A \neq T$. Given some observed variables $\mathbf{E} \subseteq \mathbf{V}$, such that $A \notin \mathbf{E}$, and their observed values $\mathbf{e} \in \Omega_{\mathbf{E}}$. Given some interval $U^T \in \Omega_T$. Then we can write the probability

$$\mu(P)(T \in U^T | \mathbf{E} = \mathbf{e}) \tag{4.46}$$

as a variation expression of the following form:

$$\frac{c_1 + \sum_{J^{A+} \in \mathbf{J}^{A+}} c^1_{J^{A+}} \int_{J^{A+}} g(a)da}{c_2 + \sum_{J^{A+} \in \mathbf{J}^{A+}} c^2_{J^{A+}} \int_{J^{A+}} g(a)da}$$

where $c_1, c_2 \in \mathbb{R}$ and $c^x_{J^{A+}} \in \mathbb{R}$ (for all $J^{A+} \in \mathbf{J}^{A+}$) are constants under μ.

This corollary follows directly from Corollary 4.2.4 in the same way that Corollary 4.2.3 follows from Corollary 4.2.2.

4.2.3 Example

We will calculate a variation expression in an example network.

Example 4.2.6. We will use the network specified in Example 2.3.1. We note that for this example it holds that $\mathbf{I}^{B+} = \mathbf{I}^B$ and $\mathbf{I}^{C+} = \mathbf{I}^C$ since both B and C are leaves of the network. We have that $\mathbf{I}^{A+} = \{I^{A+}_1, I^{A+}_2, I^{A+}_3\}$ with $I^{A+}_1 = [0, 1)$, $I^{A+}_2 = [1, 1.3)$, $I^{A+}_3 = [1.3, 2)$.

Now say we vary $\mu(f^A_{I_2^A}(a)) = g(a)$, and we are interested in the variation expression of the probability $P(C \in I_2^{C+})$ in terms of this variation. We will write out this probability, directly rewriting an equation of the form of Equation 4.24 into an equation of the form of Equation 4.25:

$$\begin{aligned}
&\mu(P)(C \in I_2^C) \\
=&\sum_{I^{A+},B+ \in \mathbf{I}^{A+},B+} \int_{I^{A+}} \mu(\phi^A)(a) \int_{I^{B+}} \int_{I_2^{C+}} \phi^{B|I^{A+}}(b)\phi^{C|I^{A+}}(c)dc\ db\ da
\end{aligned}$$

We will go over all individual parts of the summation. Firstly, for I_1^{A+} and I_1^{B+}:

$$\begin{aligned}
&\int_{I_1^{A+}} \mu(\phi^A)(a) \int_{I_1^{B+}} \int_{I_2^{C+}} \phi^{B|I_1^{A+}}(b)\phi^{C|I_1^{A+}}(c)dc\ db\ da \\
=&0.196 \cdot \int_{I_1^{A+}} \mu(\phi^A)(a)da \\
=&0.196 \cdot \int_{I_1^{A+}} \phi^A(a_1) \cdot \frac{1 - \int_{I_2^A} g(a_2)da_2}{1 - \int_{I_2^A} \phi^A(a_3)da_3} da_1 \\
=&0.196 \cdot \frac{1 - \int_{I_2^A} g(a_2)da_2}{1 - \int_{I_2^A} \phi^A(a_3)da_3} \cdot \int_{I_1^{A+}} \phi^A(a_1)da_1 \\
=&0.196 \cdot 0.287 \cdot \frac{1 - \int_{I_2^A} g(a_2)da_2}{0.287} \\
=&0.196 - 0.196 \cdot \int_{I_2^A} g(a)da \\
=&0.196 - 0.196 \cdot \int_{I_2^{A+}} g(a)da - 0.196 \cdot \int_{I_3^{A+}} g(a)da
\end{aligned}$$

We see this corresponds with Case 2, i.e. co-variation. Now the part of the summation for I_2^{A+} and I_1^{B+}. We see:

$$\begin{aligned}
&\int_{I_2^{A+}} \mu(\phi^A)(a) \int_{I_1^{B+}} \int_{I_2^{C+}} \phi^{B|I_2^{A+}}(b)\phi^{C|I_2^{A+}}(c)dc\ db\ da \\
=&0.194 \cdot \int_{I_2^{A+}} \mu(\phi^A)(a)da \\
=&0.194 \cdot \int_{I_2^{A+}} g(a)da
\end{aligned}$$

We see this corresponds with Case 1: The subinterval we are considering (I_2^{A+}) is a subset of the subinterval we are varying on (I_2^A) so we have that the varied potential $\mu(\phi^A)$ is equal to g on the subinterval under consideration. The same holds for the subinterval I_3^{A+}, so we have

that:

$$\begin{aligned}
&\int_{I_3^{A+}} \mu(\phi^A)(a) \int_{I_1^{B+}} \int_{I_2^{C+}} \phi^{B|I_3^{A+}}(b)\phi^{C|I_3^{A+}}(c)dc\ db\ da \\
=&0.239 \cdot \int_{I_3^{A+}} \mu(\phi^A)(a)da \\
=&0.239 \cdot \int_{I_3^{A+}} g(a)da
\end{aligned}$$

The calculation for the other parts of the summation (with I_2^{B+}) is performed the same way, so we will not go over these individually. Now we combine all these 6 parts of the summation[1].

$$\begin{aligned}
&\mu(P)(C \in I_2^C) \\
=&\sum_{I^{A+},B+\in \mathbf{I}^{A+},B+} \int_{I^{A+}} \mu(\phi^A)(a) \int_{I^{B+}} \int_{I_2^{C+}} \phi^{B|I^{A+}}(b)\phi^{C|I^{A+}}(c)dc\ db\ da \\
=&0.196 - 0.196 \cdot \int_{I_2^{A+}} g(a)da - 0.196 \cdot \int_{I_3^{A+}} g(a)da \\
&+0.194 \cdot \int_{I_2^{A+}} g(a)da + 0.239 \cdot \int_{I_3^{A+}} g(a)da \\
&+0.275 - 0.275 \cdot \int_{I_2^{A+}} g(a)da - 0.275 \cdot \int_{I_3^{A+}} g(a)da \\
&+0.277 \cdot \int_{I_2^{A+}} g(a)da + 0.341 \cdot \int_{I_3^{A+}} g(a)da \\
=&0.471 + 0 \cdot \int_{I_2^{A+}} g(a)da + 0.109 \cdot \int_{I_3^{A+}} g(a)da
\end{aligned}$$

[1]The 0 coefficient for $\int_{I_2^{A+}} g(a)da$ results from some special properties of this example. There will not be a 0 coefficient in the general case.

Chapter 5

Sensitivity Analysis

5.1 Introduction

In this chapter, this method of sensitivity analysis for discrete BNs will be generalised to MoTBF-networks. Note that qualitative parameters are of the same types for both type of networks, but quantitative parameters are different. Whereas the discrete BN has numerical probabilities as its quantitative parameters, the MoTBF-network has truncated functions. The major distinction between these two types of parameters is that truncated functions allow for many ways of variation. With the probabilities of the discrete BN the only method of variation is to increase or decrease the probability by a variation quantity. The MoTBF-network also allows to shift the truncated function (as we shall see in Section 5.2) but since we have a mixture of basis functions much more variations are possible. There is not one right method of varying the truncated functions, a method can be useful as long as it provides insight in the effect of inaccuracies of the parameter on the output probability.

In this chapter we will develop a sensitivity analysis for MoTBF-networks by using several variations of the truncated functions. We will specify each variation, motivate its use for sensitivity analysis, and then develop the sensitivity function for arbitrary output probabilities of the network.

We end this introduction with some notes on the connection between Chapter 4 and this chapter. In Chapter 4 we developed variation expressions for arbitrary probabilities in the network given general variations. For sensitivity analysis, we are interested in the variation expressions for specific variations. So we can use the theorems and corollaries we have developed in Chapter 4 directly. Moreover, we want to study a set of variations indexed by some variation quantity x, which denotes how 'large' the variation is. A sensitivity function maps this x to a variation expression. So for each specific variation we will introduce in this chapter, we will study a family of variation expressions. On these variation expressions we can apply the theory developed in Chapter 4.

5.2 Shift

One method of parameter variation for an MoTBF-network is shifting the parameter in its entirety upward or downward. This method of variation is directly inspired by the sensitivity analysis of van der Gaag et al. [26] for discrete BNs. By shifting the parameter we change the probability mass of the parameter. For a positive(/negative) shift we increase(/decrease) the probability mass in the subinterval of the parameter while decreasing(/increasing) the probability mass in the rest of the interval (by proportional co-variation). This gives an insight in the sensitivity of the output probability to overestimation or underestimation of the probability mass in the subinterval of the parameter.

Definition 5.2.1. Given an MoTBF-network $(G, \phi^{\mathbf{V}})$. Given a variable in the network $A \in \mathbf{V}$, and a parameter $f^A_{J^A, J^{\rho(A)}}$ of A. The Shift variation of a parameter by a variation quantity $x \in \mathbb{R}$, denoted as $\mu_{\text{Shift}x}$, is defined as follows:

$$\mu_{\text{Shift}x}(f^A_{J^A, J^{\rho(A)}})(a) = f^A_{J^A, J^{\rho(A)}}(a) + x \tag{5.1}$$

Note that the right hand side of this equation is an instance of the function $g(a)$ we used in Chapter 4, but for the specific variation of shifting the parameter by x. An example of the Shift variation can be found in the top-right corner of Figure 5.1.

5.2.1 Sensitivity Function

For sensitivity analysis using the Shift variation, we need the sensitivity function for probabilities in the network. We will first look at prior probabilities. So if we have a target variable $T \in \mathbf{V}$ and some interval $U^T \subseteq \Omega_T$ then we want to know the probability that the value of T falls within U^T. We consider a $T \neq A$ so we can use the results of Chapter 4. We want to express this probability in terms of x, since a sensitivity function maps the magnitude of the specific variation to an output probability. We note that x is only used in the variation, so we can use the results of Corollary 4.2.3. This means that:

$$\mu_{\text{Shift}x}(P)(T \in U^T) = \tag{5.2}$$

$$= c_0 + \sum_{J^{A+} \in \mathbf{J}^{A+}} c_{J^{A+}} \int_{J^{A+}} g(a, x) da \tag{5.3}$$

$$= c_0 + \sum_{J^{A+} \in \mathbf{J}^{A+}} c_{J^{A+}} \int_{J^{A+}} f^A_{J^A, J^{\rho(A)}}(a) + x \; da \tag{5.4}$$

$$= c_0 + \sum_{J^{A+} \in \mathbf{J}^{A+}} \left(c_{J^{A+}} \int_{J^{A+}} f^A_{J^A, J^{\rho(A)}}(a) da \right) + \left| \mathbf{J}^{A+} \right| \cdot x \tag{5.5}$$

$$= c_1 + c_2 \cdot x \tag{5.6}$$

for some $c_1, c_2 \in \mathbb{R}$ which are constant in x. Likewise, if we have some variables $\mathbf{E} \subseteq \mathbf{V}$ (with $A \notin \mathbf{E}$, again so we can use the results of Chapter 4) with observed values $\mathbf{e} \in \Omega_{\mathbf{E}}$ then we can

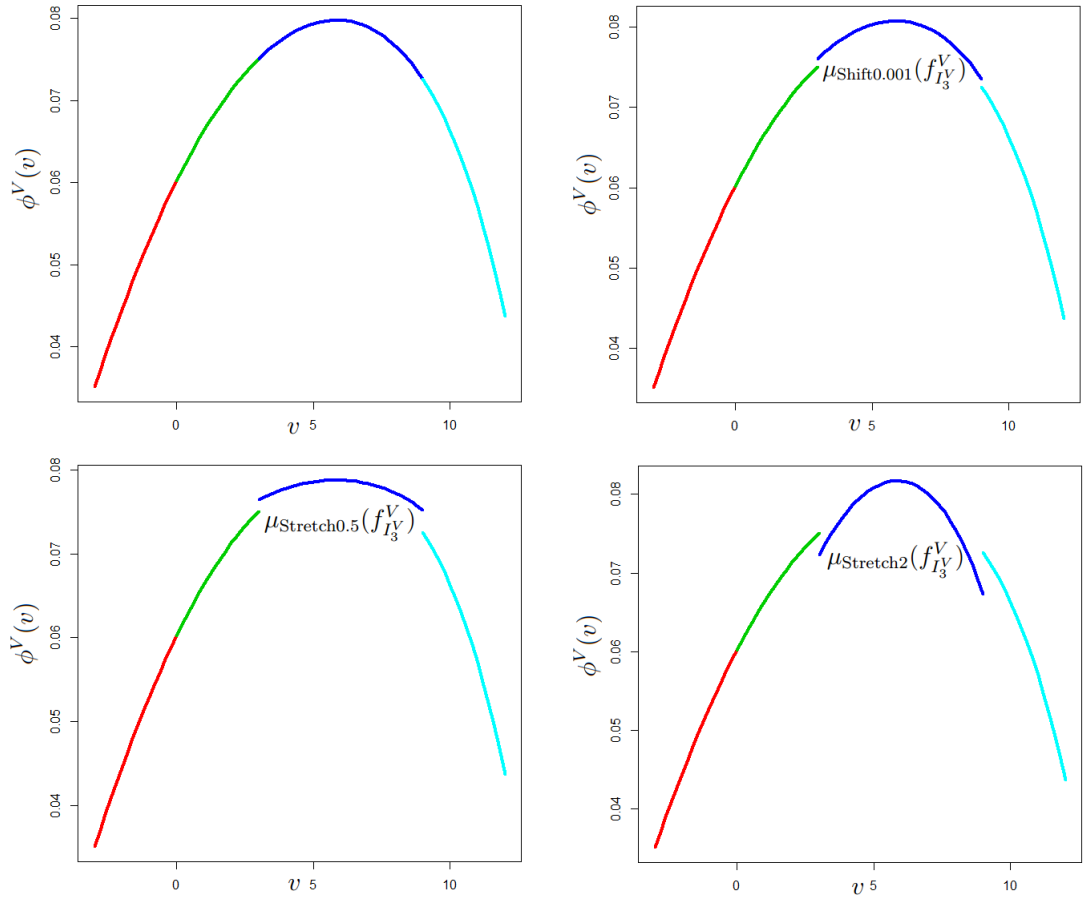


Figure 5.1: Plots of the effect of the Shift and Stretch variations. The original potential (top-left) has its parameter $f^V_{I_3^V}$ varied. In the top-right corner it is shifted 0.001 upwards, in the bottom-left corner it is stretched(/shrunk) by a factor 0.5 and in the bottom-right corner it is stretched by a factor 2. No co-variation is applied.

find the sensitivity function for the posterior probability the same way using Corollary 4.2.5, so:

$$\mu_{\text{Shift}x}(P)(T \in U^T | \mathbf{E} = \mathbf{e}) = \tag{5.7}$$

$$= \frac{c_1 + c_2 \cdot x}{c_3 + c_4 \cdot x} \tag{5.8}$$

Example 5.2.2. We will give an example of a sensitivity function using the Shift variation with a prior probability as an output probability. This example is a direct continuation of Example 4.2.6. In that example, we were interested in the probability that C fell within I_2^C as an expression of the variation of $f^A_{I_2^C}$, so $\mu f^A_{I_2^C}(a) = g(a)$ for some general function g. For this example, we consider the variation of $f^A_{I_2^C}$ as a Shift variation of x, so we determine $g(a) =_{\text{def}} f^A_{J^A, J^{\rho(A)}}(a) + x$. We can then find the sensitivity function as follows:

$$\begin{aligned}
&\mu_{\text{Shift}x}(P)(C \in I_2^C) \\
=&0.471 + 0 \cdot \int_{I_2^{A+}} g(a)da + 0.109 \cdot \int_{I_3^{A+}} g(a)da \\
=&0.471 + 0.109 \cdot \int_{I_3^{A+}} f^A_{I_2^A}(a) + x \; da \\
=&0.471 + 0.109 \cdot \int_{I_3^{A+}} f^A_{I_2^A}(a)da + 0.109 \cdot |I_3^{A+}| \cdot x \\
=&0.471 + 0.109 * 0.61 + 0.7 * 0.109 * x \\
=&0.537 + 0.076 \cdot x
\end{aligned}$$

This sensitivity function is plotted in Figure 5.2. At $x = 0$ no variation has taken place so we find that $P(C \in I_2^C) = 0.537$. The upwards slope tells us there is a positive relationship between x and $\mu_{\text{Shift}x}(P)(C \in I_2^C)$. So if the parameter $f^A_{I_2^C}$ is estimated too low in our model, then this will affect our estimation of $P(C \in I_2^C)$ to be too low as well (at least for a Shift of $f^A_{I_2^C}$). The absolute value of the slope, here 0.076, tells us how sensitive the output probability $P(C \in I_2^C)$ is to a shift in the parameter. A larger slope means the output probability is more affected by a Shift of the parameter. The sensitivity analysis does not determine what a large slope is objectively, just as in discrete sensitivity analysis, but it does determine that some output probabilities are more sensitive to Shifts in a certain parameter.

5.3 Stretch

Another method of variation is by stretching (or shrinking) the parameter. We want to do this while keeping the average value of the parameter constant in its associated subinterval. This way the probability mass under the subinterval of the parameter also stays constant, as will be shown later in this section.

Since the variation keeps the probability mass of the parameter constant, no co-variation is necessary and proportional co-variation will not affect the co-varied parameters. This also means that this variation is applicable to an SP MoTBF-network, since the probability mass of

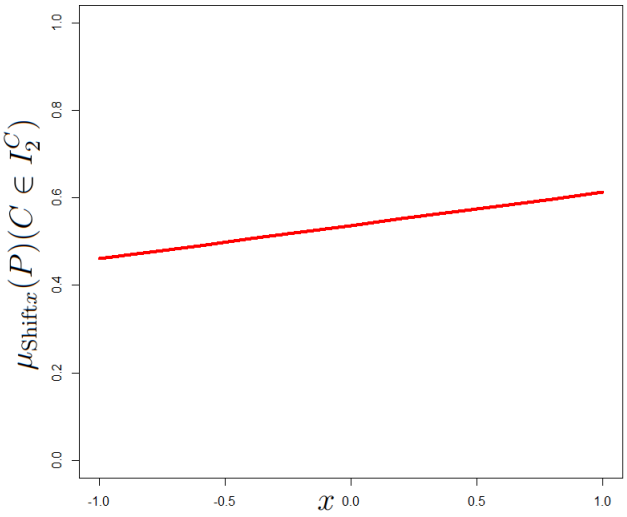


Figure 5.2: A plot of the sensitivity function of Example 5.2.2.

the varied parameter then remains 1. This variation however is not suited for the US MoTBF-network, as the parameter then only affects the output probability through its probability mass within the entire subinterval, as can be seen in the result of Theorem 4.1.1, so the output probability will be constant under variation.

We will first define the average of a parameter before we move on to the definition of the Stretch variation.

Definition 5.3.1. Given A and $f^A_{J^A,J^{\rho(A)}}$ as before. The average of the parameter $f^A_{J^A,J^{\rho(A)}}$, denoted as $avg(f^A_{J^A,J^{\rho(A)}})$, is defined as follows:

$$avg(f^A_{J^A,J^{\rho(A)}}) =_{\text{def}} \frac{\int_{J^A} f^A_{J^A,J^{\rho(A)}}(a)da}{|J^{\rho(A)}|} \tag{5.9}$$

By stretching the parameter, we want to stretch out the points at the values furthest from the average the most, while keeping the points at the average constant. We do this by shifting the parameter so its average value is on the x-axis, then multiplying the parameter with a variation quantity, and then shifting it back up.

Definition 5.3.2. Given A and $f^A_{J^A,J^{\rho(A)}}$ as before. The Shift variation of a parameter by a variation quantity $x \in \mathbb{R}^+$, denoted as $\mu_{\text{Stretch}x}$, is defined as follows:

$$\mu_{\text{Stretch}x}(f^A_{J^A,J^{\rho(A)}})(a) \tag{5.10}$$

$$=(f^A_{J^A,J^{\rho(A)}}(a) - avg(f^A_{J^A,J^{\rho(A)}})) \cdot x + avg(f^A_{J^A,J^{\rho(A)}}) \tag{5.11}$$

Note that we only consider $x \in \mathbb{R}^+$ since any negative x will flip the potential vertically, and we do not deem it interesting to analyse the sensitivity to such a variation. We note that the set of feasible variation quantities x can differ for different variations.

We will show that the stretch variation preserves the probability mass of the varied parameter. We note:

$$\begin{aligned}
&\int_{J^A} avg(f^A_{J^A,J^{\rho(A)}})da \\
=&\int_{J^A} \frac{\int_{J^A} f^A_{J^A,J^{\rho(A)}}(a)da}{|J^A|}da \\
=&|J^A| \cdot \frac{\int_{J^A} f^A_{J^A,J^{\rho(A)}}(a)da}{|J^A|} \\
=&\int_{J^A} f^A_{J^A,J^{\rho(A)}}(a)da
\end{aligned}$$

so it follows that:

$$\begin{aligned}
&\int_{J^A} \mu_{\text{Stretch}x}(f^A_{J^A,J^{\rho(A)}})(a)da \\
=&\int_{J^A} (f^A_{J^A,J^{\rho(A)}}(a) - avg(f^A_{J^A,J^{\rho(A)}})) \cdot x + avg(f^A_{J^A,J^{\rho(A)}})da \\
=&(\int_{J^A} f^A_{J^A,J^{\rho(A)}}(a)da - \int_{J^A} avg(f^A_{J^A,J^{\rho(A)}})da) \cdot x + \int_{J^A} avg(f^A_{J^A,J^{\rho(A)}})da \\
=&(\int_{J^A} f^A_{J^A,J^{\rho(A)}}(a)da - \int_{J^A} f^A_{J^A,J^{\rho(A)}}(a)da) \cdot x + \int_{J^A} f^A_{J^A,J^{\rho(A)}}(a)da \\
=&\int_{J^A} f^A_{J^A,J^{\rho(A)}}(a)da
\end{aligned}$$

So the probability mass is preserved.

For $x = 1$ the parameter is kept constant, for $x > 1$ the parameter is stretched and for $x < 1$ the parameter is shrunk. Examples of shrunk and stretched parameters can be found in the bottom-left and bottom-right corner of Figure 5.1 respectively.

5.3.1 Sensitivity Function

The sensitivity function for this method of variation is found in much the same way as in Subsection 5.2.1. For the prior probability it is:

$$\begin{aligned}
&\mu_{\text{Stretch}x}(P)(T \in U^T) && (5.12)\\
=&c_0 + \sum_{J^{A+} \in \mathbf{J}^{A+}} c_{J^{A+}} \int_{J^{A+}} \mu_{\text{Stretch}x}(f^A_{J^A,J^{\rho(A)}})(a)da && (5.13)\\
=&c_1 + c_2 \cdot x && (5.14)
\end{aligned}$$

for some $c_1, c_2 \in \mathbb{R}$ which are constant under x. The rewriting of Equation 5.13 into Equation 5.14 can be found in Appendix A.3.2. We notice this sensitivity function is of the same form as the sensitivity function of the Shift variation in Equation 5.6 (albeit for different constants c_1, c_2). The form of the sensitivity function for a posterior output probability will of course also be equal to the one we found for the Shift variation (i.e. Equation 5.8).

Example 5.3.3. We will again give an example of the sensitivity function with the stretch variation based on Example 4.2.6.

$$
\begin{aligned}
&\mu_{\text{Stretch}x}(P)(C \in I_2^C) \\
=&0.471 + 0 \cdot \int_{I_2^{A+}} g(a)da + 0.109 \cdot \int_{I_3^{A+}} g(a)da \\
=&0.471 + 0.109 \cdot \int_{I_3^{A+}} \left(f^A_{I_2^A}(a) - avg(f^A_{I_2^A}) \right) \cdot x + avg(f^A_{I_2^A})\, da \\
=&0.471 + 0.109 \cdot \left(\int_{I_3^{A+}} f^A_{I_2^A}(a)da - \left(avg(f^A_{I_2^A}) \cdot |I_3^{A+}| \right) \right) \cdot x \\
&+ 0.109 \cdot avg(f^A_{I_2^A}) \cdot |I_3^{A+}| \\
=&0.526 + 0.012 \cdot x
\end{aligned}
$$

5.4 Extra Base Functions

The Shift and Stretch variations are useful for sensitivity analysis of MoTBF-networks without any supplied context. The sensitivity analysis then gives insight in the general inaccuracies of the model. Since MoTBF-networks can be learned from any probability distribution, it can also be interesting to analyse how sensitive the output is in the approximation of the MoTBF-network to the distribution. In this section we develop a variation (with respective sensitivity function) to do just that.

Langseth et al. [16] describe a method for making an approximated MoTBF-network from an arbitrary probability distribution, and predefined graph structure, in their paper as follows. Initially, the conditional potential of each variable is a single truncated function with just a single base function. Then the best improvement is iteratively picked until the MoTBF-network is within permissible error (i.e. a good enough approximation of the distribution). An improvement can be found either by introducing an extra base function to a parameter, or by splitting a subinterval of a continuous parent (e.g. an element in $\mathbf{I}^{V \to C}$) into two subintervals. After such an MoTBF-network is made, it will then be interesting to see how sensitive the output probabilities are to these two improvements. If the output is very sensitive to certain improvements, this means that the MoTBF-network might not be a close enough approximation of the distribution. Alternatively, this sensitivity analysis could be used directly in learning the MoTBF-network, by picking improvements based on the highest sensitivity of the output probability as an adaptation of the original learning algorithm. The first improvement, adding an extra base function, will be used for the variation discussed in this section. The second improvement will be discussed in Section 6.2.

We use the extra base functions variation, $\mu_{\text{ExtraBase}x}$, to introduce extra base functions to a parameter. Equation 2.12 gives us the definition of the original parameter in terms of base

functions:

$$f^A_{J^A,J^{\rho(A)}} = \sum_{i=0}^{m} a_i\psi_i(a) \tag{5.15}$$

Introducing extra base functions then is just increasing this summation. For $x \in \mathbb{N}$, we then have that the extra base functions variation introduces more base functions. This is defined as follows:

Definition 5.4.1. Given A and $f^A_{J^A,J^{\rho(A)}}$ as before. The Shift variation of a parameter by a variation quantity $x \in \mathbb{N}$, denoted as $\mu_{\text{ExtraBase}x}$, is defined as follows:

$$\mu_{\text{ExtraBase}x}(f^A_{J^A,J^{\rho(A)}}) = \sum_{i=0}^{m+x} a_i\psi_i(a) \tag{5.16}$$

So this variation will make truncated function better resemble the potential of the original HBN on the truncated function's associated interval.

5.4.1 Sensitivity Function

For a prior probability, the sensitivity function for the extra base functions variation will be of the form:

$$\mu_{\text{ExtraBase}x}(P)(T \in U^T) \tag{5.17}$$

$$= c_0 + \sum_{J^{A+} \in \mathbf{J}^{A+}} c_{J^{A+}} \int_{J^{A+}} \mu_{\text{ExtraBase}x}(f^A_{J^A,J^{\rho(A)}})(a)da \tag{5.18}$$

$$= c_0 + \sum_{J^{A+} \in \mathbf{J}^{A+}} c_{J^{A+}} \int_{J^{A+}} \sum_{i=0}^{m+x} a_i\psi_i(a)da \tag{5.19}$$

$$= c_0 + \sum_{J^{A+} \in \mathbf{J}^{A+}} c_{J^{A+}} \int_{J^{A+}} \sum_{i=0}^{m} a_i\psi_i(a)da \tag{5.20}$$

$$+ \sum_{J^{A+} \in \mathbf{J}^{A+}} c_{J^{A+}} \int_{J^{A+}} \sum_{i=m+1}^{m+x} a_i\psi_i(a)da \tag{5.21}$$

$$= c_1 + \sum_{J^{A+} \in \mathbf{J}^{A+}} c_{J^{A+}} \int_{J^{A+}} \sum_{i=m+1}^{m+x} a_i\psi_i(a)da \tag{5.22}$$

for some $c_1 \in \mathbb{R}$ and $c_{J^{A+}} \in \mathbb{R}$ for $J^{A+} \in \mathbf{J}^{A+}$ that are constant under x. The sensitivity function for the posterior probability is again a fraction of two of these sensitivity functions.

5.5 Experiments

In this section, we will support the theory of this thesis with two experiments. The goal of these experiments is to test that for an example MoTBF-network the sensitivity functions of the Shfit and Stretch variations for a prior output probability will show the form predicted by the theory.

We use Elvira to do inference with an example MoTBF-network. Elvira is a software package that can be used to model and do inference with BNs [5], and it has an implementation for MoTBF-networks specifically. The example MoTBF-network we will use is based on an example network made by the same author as the Elvira package, called mte_pos [1]. This is a network of three variables, V_1, V_2 and V_3. All three variables are continuous, and the network is an MoTBF-network or more specifically an MTE-network. The variable V_1 is a root of the network, with as only child V_2, and V_2 has as only child V_3 which is a leaf of the network. The subintervalsets of mte_pos are all equal: $\mathbf{I}^{V_1} = \mathbf{I}^{V_1 \to V_2} = \mathbf{I}^{V_2} = \mathbf{I}^{V_2 \to V_3} = \mathbf{I}^{V_3} = \{[0,1),[1,2)\}$.

We will make two small modifications to the mte_pos network to arrive at the network we will use for our experiments. The first is that we redefine $\mathbf{I}^{V_1 \to V_2} = \{[0,0.6),[0.6,2]\}$ and the second is that we redefine $\mathbf{I}^{V_2 \to V_3} = \{[0,1.2),[1.2,2]\}$. The reason we do this is that the mte_pos is a US MoTBF-network. This new network is not a US MoTBF-network. This way our experiments will support the theory of sensitivity analysis for general MoTBF-networks, instead of just the theory of sensitivity analysis for US MoTBF-networks.

For our two experiments, we are interested in the same parameter to vary and the same output probability. The parameter we will be varying is the first parameter of variable V_1, i.e. $f^{V_1}_{I^{V_1}_1}$ with $I^{V_1}_1 = [0,1)$. The output probability of interest will the prior probability $P(0.4 \leq V_3 < 1.5)$.

For the first experiment, we will vary the parameter $f^{V_1}_{I^{V_1}_1}$ according to the shift variation. The parameter $f^{V_1}_{I^{V_1}_2}$ will be co-varied according to proportional co-variation. We will do this shift variation for several variation quantities x, namely $x \in \{-0.2,-0.1,0,0.1,0.2\}$, and then find the output probability under these variations. The theory predicts that

$$\mu_{\text{Shift}x}(P)(0.4 \leq V_3 < 1.5) = c_1 + c_2 \cdot x$$

for some $c_1, c_2 \in \mathbb{R}$ (see Section 5.2.1). So the theory predicts that the output probabilities fall on one line if plotted against the variation quantity x. As can be seen in Figure 5.3, the experiment supports this prediction. The points in the plot fall on the line $0.615 + 0.03 \cdot x$.

For the second experiment, we will vary the parameter $f^{V_1}_{I^{V_1}_1}$ according to the shift variation. This time we will do the variation for the variation quantities $x \in \{0,0.5,1,1.5,2\}$. The theory predicts that

$$\mu_{\text{Shift}x}(P)(0.4 \leq V_3 < 1.5) = c_1 + c_2 \cdot x$$

for some $c_1, c_2 \in \mathbb{R}$ (see Section 5.3.1). The output probabilities for the specified variation quantity again fall on a line as predicted by the theory, as seen in Figure 5.4.

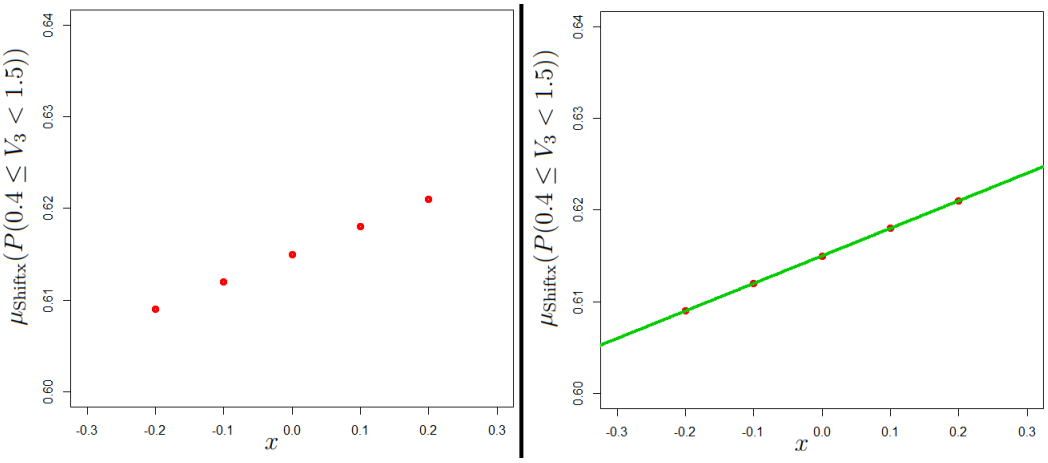


Figure 5.3: Left: The output probability for the shift variation for several values of x. Right: The same image, with the addition of a plotted line that fits through these points.

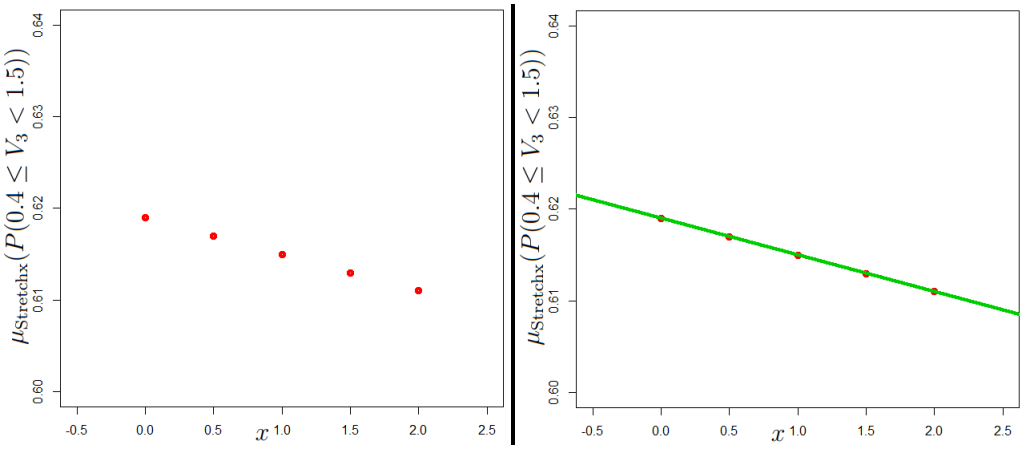


Figure 5.4: Left: The output probability for the stretch variation for several values of x. Right: The same image, with the addition of a plotted line that fits through these points.

Chapter 6

Conclusion and Future Research

6.1 Conclusion

The central question posed in the introduction was: It is possible to generalise the method of sensitivity analysis as described by van der Gaag et al. [26] towards the MoTBF-network? In this thesis, an effort was made to answer this question by developing such a method. The answer to this question is therefore a positive one. The contributions of this thesis will be summarised per chapter.

Firstly some preliminaries were introduced in Chapter 2. The original work of this chapter concerned intervals. The intervals of random variables in the MoTBF-network were divided into subintervals used in the definition of the truncated functions. The notion of a subinterval set was then introduced. This is a set of mutually exclusive subintervals that exactly cover an interval, usually the interval of a random variable. In an MoTBF a random variable can have multiple subinterval sets associated with it, and a new subinterval set was introduced that cuts the interval in even smaller pieces by taking the intersections of all the subintervals in the other sets.

Then, the notion of co-variation was extended towards truncated functions in the MoTBF-network in Chapter 3. Some necessary and optional conditions were imposed on co-variation schemes. Then some schemes were defined and analysed through these conditions. The first of these was proportional co-variation, a direct generalisation of the most widely used scheme for discrete BNs. This scheme fulfilled all conditions save for one: The optional condition that the joining of neighbouring parameters should be preserved under co-variation. The other schemes considered did fulfil this condition but fared worse on other conditions, most critically they all failed the necessary condition that all parameters should be positive on their entire subinterval. Therefore proportional co-variation was suggested as the best co-variation scheme and was used throughout the rest of the thesis.

In Chapter 4, probabilities in the network were expressed in terms of a general variation of a truncated function in the network. These expressions are called variation expressions. Firstly variation expressions were developed for a special kind of MoTBF-network, the US MoTBF-network, where each variable only has a single associated subinterval set. This was then used as a stepping stone for developing variation expressions for the general MoTBF-network, which

was more labourious since more subinterval sets were involved. For both these types of network the initial variation expressions only were developed for a fixed type of probability of the network, namely probabilities of target variables falling into subintervals of some subinterval set. The theory was therefore extended to more general probabilities, namely the probabilities that a certain target variable falls into an arbitrary interval given some observed evidence for some variables. For this generalisation to work, it was assumed that the varied variable (i.e. the variable with the varied parameter) did not equal the target variable and was not part of the observed variables.

In Chapter 5, a sensitivity analysis for MoTBF-networks was developed. Where in Chapter 4 variation expressions for a general variation were developed, this chapter introduced some specific methods of variation. For each of these a sensitivity function was developed: A function that gives for a variation quantity (i.e. how large the variation is) an expression of probabilities in the network in terms of the variation. This was done by directly applying the theory of Chapter 4. Three methods of variation were introduced and had sensitivity functions developed. Firstly the Shift variation, that is the most direct generalisation of the variation in the discrete case, and gives insight in the sensitivity of probabilities to the varied parameter being too low or high overall. Then the Stretch variation, which give insight in the sensitivity of probabilities to the values of the parameter being closer or further away from the average value of the parameter. Lastly the Extra Base Functions variation, which gave insight in the sensitivity of probabilities to the parameter being a better approximation of a probability distribution the MoTBF-network models.

In conclusion it has been shown in this thesis that it is possible to generalise the method of sensitivity analysis as described by van der Gaag et al. [26] towards the MoTBF-network. This is an important achievement, since no method for sensitivity analysis of MoTBF-networks had been developed thus far. Much groundwork was necessary to eventually develop the method in Chapter 5. It is regretful though that this generalisation is not entirely complete, since in Chapter 4 there were some assumptions made on the probabilities that variation expressions can be found for (and thus sensitivity analysis can be performed on), namely only probabilities where the varied variable does not equal the target variable and the varied variable is not part of the observed evidence. But with this note in mind, sensitivity analysis can still be performed on all these probabilities. By doing sensitivity analysis using the Shift variation and using proportional co-variation, the most direct generalisation is made. Proportional co-variation seems to make the most sensible scheme, but many variations next to the Shift variation are possible to which the theory in this thesis can be directly applied, which makes that in this generalisation, sensitivity analysis comes in even more flavours.

6.2 Future Research

The major future research should be looking into the assumption made in Chapter 4. Here the assumption was made that the varied variable should not be the target variable and not be part of the set of observed variables. Without these assumptions the general form of the variation expressions will no longer be valid, but we do think that the general form of the sensitivity functions for the Shift and Stretch variations will still be valid. A proof of this did not fit in the

structure of this thesis, but it should be possible. This allows the sensitivity analysis to work on Naive Bayesian networks as well for example, as they only have evidence and target variables.

A variation considered in Chapter 5 was the More Base Functions variation, where a parameter was made a better estimation of an original potential by adding more base functions to it. It is also possible to make the parameter a better estimation by splitting its subinterval in two pieces. Subinterval sets without associated parameters, the parent subinterval sets, can also have their subintervals split into two pieces. This can all give insight into the size of subinterval sets, and if the output probability is sensitive to increasing this size. The focus of this thesis has been entirely on changing the parameters of the network, but not on changing the subinterval sets of the network. So new theory would be needed to do sensitivity analysis on subinterval sets as well. This branch of sensitivity analysis seems less interesting to us however, since it only gives insight in the sensitivity of output probabilities to how good the approximation of the MoTBF is to the modelled probability distribution, but not to the validity of the probability distribution itself.

Appendix A

Mathematics

A.1 Subinterval Sets

Theorem A.1.1. *Consider subinterval sets* $\mathbf{I}_1^V, \cdots, \mathbf{I}_n^V$ *over the domain* Ω_V*. Then the set*

$$\mathbf{I}^{V+} =_{def} \{I_1^V \cap \cdots \cap I_n^V | I_1^V \in \mathbf{I}_1^V, \cdots, I_n^V \in \mathbf{I}_n^V\} - \{\emptyset\}$$

is also a subinterval set over the domain Ω_V*.*

Proof: We recall a set $\mathbf{I}^V$ is a subinterval set over the Ω_V if:

1. Every element $I^V \in \mathbf{I}^V$ is a subinterval of Ω_V, i.e. $I^V \subseteq \Omega_V$
2. The elements in $\mathbf{I}^V$ are mutually exclusive, i.e. for any two $I^V, J^V \in \mathbf{I}^V$ such that $I^V \neq J^V$, it holds that $I^V \cap J^V = \emptyset$.
3. The domain Ω_V is entirely covered by the subintervals in $\mathbf{I}^V$, i.e., for any $v \in \Omega_V$, there is some $I^V \in \mathbf{I}^V$ such that $v \in I^V$.

We will proof all of these requirements for the set $\mathbf{I}^{V+}$.

1. Consider an arbitrary element $I^{V+} \in \mathbf{I}^{V+}$. From the definition of $\mathbf{I}^{V+}$ we see that $I^{V+} = I_1^V \cap \cdots \cap I_n^V$ for some $I_1^V \in \mathbf{I}_1^V, \cdots, I_n^V \in \mathbf{I}_n^V$. Now for any I_i^V with $i \in \{1, \cdots, n\}$ it holds that $I_i^V \subseteq \Omega_V$ since $\mathbf{I}_i^V$ is a subinterval set over the domain Ω_V. Since any I_i^V is a subinterval of Ω_V, their intersection I^{V+} is too. Since I^{V+} is an arbitrary element of $\mathbf{I}^{V+}$, this proofs that every element of I^{V+} is a subinterval of Ω_V.
2. Consider two arbitrary elements $I^{V+}, J^{V+} \in \mathbf{I}^{V+}$ such that $I^{V+} \neq J^{V+}$. From the definition of $\mathbf{I}^{V+}$ we see that $I^{V+} = I_1^V \cap \cdots \cap I_n^V$ for some $I_1^V \in \mathbf{I}_1^V, \cdots, I_n^V \in \mathbf{I}_n^V$, and $J^{V+} = J_1^V \cap \cdots \cap J_n^V$ for some $J_1^V \in \mathbf{J}_1^V, \cdots, J_n^V \in \mathbf{J}_n^V$. We can see that there must be an $i \in \{1, \cdots, n\}$ such that $I_i^V \neq J_i^V$, otherwise we would have that $I^{V+} = J^{V+}$. Since $I_i^V, J_i^V \in \mathbf{I}_i^V$ and $I_i^V \neq J_i^V$ it follows that $I_i^V \cap J_i^V = \emptyset$, since $\mathbf{I}_i^V$ is a subinterval set so its elements are mutually exclusive. It is also clear that $I^{V+} \subseteq I_i^V$ and likewise $J^{V+} \subseteq J_i^V$ since I_i^V and J_i^V are part of the intersections that define respectively I^{V+} and J^{V+}. If two elements have an empty intersection, then so do their subsets. So $I^{V+} \cap J^{V+} = \emptyset$. Since we haven taken two arbitrary unequal elements of $\mathbf{I}^{V+}$ and have shown their intersection empty, this shows that the elements of $\mathbf{I}^{V+}$ are mutually exclusive.

3. Consider an arbitrary value $v \in \Omega_V$. Now for any $i\{1, \cdots, n\}$ the set $\mathbf{I}_i^V$ is a subinterval set over the domain Ω_V, so there is an element $I_i^V \in \mathbf{I}_i^V$ such that $v \in I_i^V$. Now consider the intersection of all these sets, $I^{V+} =_{\text{def}} I_1^V \cap \cdots \cap I_n^V$. Since v is in all the intersected elements, it is also part of the intersection, so $v \in I^{V+}$. It holds that $I^{V+} \in \{I_1^V \cap \cdots \cap I_n^V | I_1^V \in \mathbf{I}_1^V, \cdots, I_n^V \in \mathbf{I}_n^V\}$. Also since $v \in I^{V+}$ so $I^{V+} \neq \emptyset$. This shows that $I^{V+} \in \mathbf{I}^{V+}$. So there is an element $I^{V+} \in \mathbf{I}^{V+}$ such that $v \in I^{V+}$. Since v is an arbitrary value in Ω_V, this shows that the subintervals in $\mathbf{I}^{V+}$ cover the entire domain Ω_V.

□

Theorem A.1.2. *Consider subinterval sets $\mathbf{I}_1^V, \cdots, \mathbf{I}_n^V$ over the domain Ω_V, and the set*

$$\mathbf{I}^{V+} =_{def} \{I_1^V \cap \cdots \cap I_n^V | I_1^V \in \mathbf{I}_1^V, \cdots, I_n^V \in \mathbf{I}_n^V\} - \{\emptyset\}.$$

For any $I^{V+} \in \mathbf{I}^{V+}$ and for any $i \in \{1, \cdots, n\}$, there is exactly one $I_i^V \in \mathbf{I}_i^V$ such that $I^{V+} \subseteq I_i^V$.

Proof: Consider an arbitrary $I^{V+} \in \mathbf{I}^{V+}$ and an arbitrary $i \in \{1, \cdots, n\}$. From the definition of $\mathbf{I}^{V+}$ we see that $I^{V+} = I_1^V \cap \cdots \cap I_i^V \cap \cdots \cap I_n^V$ for some $I_1^V \in \mathbf{I}_1^V, \cdots, I_i^V \in \mathbf{I}_i^V, \cdots, I_n^V \in \mathbf{I}_n^V$. Since I^{V+} is an intersection of some sets including I_i^V, we know that $I^{V+} \subseteq \mathbf{I}_i^V$. So I^{V+} is a subset of at least one element in $\mathbf{I}_i^V$, namely I_i^V.

Consider an arbitrary $J_i^V \in \mathbf{I}_i^V$ such that $J_i^V \neq I_i^V$. First we note that I^{V+} is not empty, since the set $\mathbf{I}^{V+}$ is defined with the empty set subtracted. So there is some element $v \in I^{V+}$. Now since $I^{V+} \subseteq I_i^V$ it follows that $v \in I_i^V$. Since $\mathbf{I}_i^V$ is a subinterval set it follows that its elements are mutually exclusive, so $I_i^V \cap J_i^V = \emptyset$ and so $v \notin J_i^V$. This means that $I^{V+} \not\subseteq J_i^V$. Since J_i^V was an arbitrary element of $\mathbf{I}_i^V$ such that $J_i^V \neq I_i^V$, we know that I^{V+} is a subset of no other element of $\mathbf{I}_i^V$ than I_i^V. So I^{V+} is a subset of exactly one element of $\mathbf{I}_i^V$, namely the set I_i^V. □

A.2 Variation Expression

The following shows how to rewrite Equation 4.24 into Equation 4.25.

A.2.1 General

$$\mu(P)(\mathbf{T} \in I^{\mathbf{T}+}) \tag{A.1}$$

(we write the probability as a potential)

$$= \int_{I^{\mathbf{T}+}} \mu(\phi^{\mathbf{T}})(\mathbf{t}) d\mathbf{t} \tag{A.2}$$

(we marginalise out the non-target variables)

$$= \int_{I^{\mathbf{T}+}} \int_{\Omega^{\mathbf{N}}} \mu(\phi^{\mathbf{V}})(\mathbf{t}, \mathbf{n}) d\mathbf{n} d\mathbf{t} \tag{A.3}$$

(we rewrite using Equation 2.4)

$$= \int_{I^{\mathbf{T}+}} \int_{\Omega^{\mathbf{N}}} \prod_{V \in \mathbf{V}} \mu(\phi^{V|\rho(V)})(v|\rho(v)) d\mathbf{n} d\mathbf{t} \tag{A.4}$$

(we split the domain of $\mathbf{N}$ into truncated functions)

$$= \int_{I^{\mathbf{T}+}} \sum_{I^{\mathbf{N}+} \in \mathbf{I}^{\mathbf{N}+}} \int_{I^{\mathbf{N}+}} \prod_{V \in \mathbf{V}} \mu(\phi^{V|I^{\mathbf{V}+}})(v) d\mathbf{n} d\mathbf{t} \tag{A.5}$$

(we take the summation out, which is independent of $\mathbf{t}$)

$$= \sum_{I^{\mathbf{N}+} \in \mathbf{I}^{\mathbf{N}+}} \int_{I^{\mathbf{T}+}} \int_{I^{\mathbf{N}+}} \prod_{V \in \mathbf{V}} \mu(\phi^{V|I^{\mathbf{V}+}})(v) d\mathbf{n} d\mathbf{t} \tag{A.6}$$

(we merge $\mathbf{T}$ and $\mathbf{N}$ into $\mathbf{V}$)

$$= \sum_{I^{\mathbf{N}+} \in \mathbf{I}^{\mathbf{N}+}} \int_{I^{\mathbf{V}+}} \prod_{V \in \mathbf{V}} \mu(\phi^{V|I^{\mathbf{V}+}})(v) d\mathbf{v} \tag{A.7}$$

(we split $\mathbf{V}$ into A and $\mathbf{V}'$)

$$= \sum_{I^{\mathbf{N}+} \in \mathbf{I}^{\mathbf{N}+}} \int_{I^{A+}} \int_{I^{\mathbf{V}'+}} \mu(\phi^{A|I^{\mathbf{V}+}})(a) \prod_{V' \in \mathbf{V}'} \phi^{V'|I^{\mathbf{V}+}}(v') d\mathbf{v}' da \tag{A.8}$$

(we move the truncated function out of the integral)

$$= \sum_{I^{\mathbf{N}+} \in \mathbf{I}^{\mathbf{N}+}} \int_{I^{A+}} \mu(\phi^{A|I^{\mathbf{V}+}})(a) \int_{I^{\mathbf{V}'+}} \prod_{V' \in \mathbf{V}'} \phi^{V'|I^{\mathbf{V}+}}(v') d\mathbf{v}' da \tag{A.9}$$

(the integral over $I^{\mathbf{V}'}$ is constant in both μ and a. Because of the former we can make it $O(1)$, and because of the latter we can move it out of the integral over I^A)

$$= \sum_{I^{\mathbf{N}+} \in \mathbf{I}^{\mathbf{N}+}} O(1) \int_{I^{A+}} \mu(\phi^{A|I^{\mathbf{V}+}})(a) da \tag{A.10}$$

A.2.2 Prior marginal

The rewriting rules in this proof of Corollary 4.2.2 heavily resemble those in Appendix A.2.1. The notable differences are that we are considering a single variable T here as opposed to a set of variables $\mathbf{T}$, and that we are considering an integral over some $U^T \subseteq I^{T+}$ instead of an integral over I^{T+}. Note that this does not matter in the way the conditional potentials are conditioned, as the same truncated functions are selected by U^T and I^{T+} (since the subintervals in I^{T+} are 'the smallest units' in selecting truncated functions).

Proof of Corollary 4.2.2: We do some familiar rewriting.

$$\mu(P)(\mathbf{T} \in U^{\mathbf{T}}) \tag{A.11}$$

(we write the probability as a potential)

$$= \int_{U^T} \mu(\phi^{\mathbf{T}})(\mathbf{t}) dt \tag{A.12}$$

(we marginalise out the non-target variables)

$$= \int_{U^T} \int_{\Omega^{\mathbf{N}}} \mu(\phi^{\mathbf{V}})(\mathbf{t}, \mathbf{n}) d\mathbf{n} dt \tag{A.13}$$

(we rewrite using Equation 2.4)

$$= \int_{U^T} \int_{\Omega^{\mathbf{N}}} \prod_{V \in \mathbf{V}} \mu(\phi^{V|\rho(V)})(v|\rho(v)) d\mathbf{n} dt \tag{A.14}$$

(we split the domain of $\mathbf{N}$ into truncated functions. Here $I^{\mathbf{V}+} = I^{\mathbf{N}+} \cap \{I^{T+}\}$)

$$= \int_{U^T} \sum_{I^{\mathbf{N}+} \in \mathbf{I}^{\mathbf{N}+}} \int_{I^{\mathbf{N}+}} \prod_{V \in \mathbf{V}} \mu(\phi^{V|I^{\mathbf{V}+}})(v) d\mathbf{n} dt \tag{A.15}$$

(we take the summation out, which is independent of t)

$$= \sum_{I^{\mathbf{N}+} \in \mathbf{I}^{\mathbf{N}+}} \int_{U^T} \int_{I^{\mathbf{N}+}} \prod_{V \in \mathbf{V}} \mu(\phi^{V|I^{\mathbf{V}+}})(v) d\mathbf{n} t \tag{A.16}$$

(we split $\mathbf{N}$ into A and $\mathbf{N}'$ (since $A \neq T$) and we split $\mathbf{V}$ into A and $\mathbf{V}'$

$$= \sum_{I^{\mathbf{N}+} \in \mathbf{I}^{\mathbf{N}+}} \int_{U^T} \int_{I^{A+}} \int_{I^{\mathbf{N}'+}} \mu(\phi^{A|I^{\mathbf{V}+}})(a) \prod_{V' \in \mathbf{V}'} \phi^{V'|I^{\mathbf{V}+}}(v') d\mathbf{n}' \, da \, dt \tag{A.17}$$

(both the truncated function and the integral over a are independent of both $\mathbf{n}'$ and t so we move them out of those integrals)

$$= \sum_{I^{\mathbf{N}+} \in \mathbf{I}^{\mathbf{N}+}} \int_{I^{A+}} \mu(\phi^{A|I^{\mathbf{V}+}})(a) \int_{U^T} \int_{I^{\mathbf{N}'+}} \prod_{V' \in \mathbf{V}'} \phi^{V'|I^{\mathbf{V}+}}(v') d\mathbf{n}' \, dt \, da \tag{A.18}$$

(the integral over U^T and $I^{\mathbf{N}'}$ are constant in both μ and a. Because of the former we can make them $O(1)$, and because of the latter we can move it out of the integral over I^A)

$$= \sum_{I^{\mathbf{N}+} \in \mathbf{I}^{\mathbf{N}+}} O(1) \int_{I^{A+}} \mu(\phi^{A|I^{\mathbf{V}+}})(a) da \tag{A.19}$$

This leaves us with an equation exactly of the form of the one in the proof of Theorem 4.2.1, so the case part of this proof can be applied to this corollary as well. □

A.2.3 Posterior marginal

We do rewriting for $\int_{U^T} \mu(\phi^{T,\mathbf{E}})(t, \mathbf{e}) dt$ for the proof of Corollary 4.2.4. Firstly we write $I^{\mathbf{E}+}$ for the unique set of subintervals for the values of $\mathbf{e}$, i.e. for any $e \in \mathbf{e}$ and $I^{E+} \in I^{\mathbf{E}+}$ we have $e \in I^E$.

$$= \int_{U^T} \mu(\phi^{T,\mathbf{E}})(t, \mathbf{e}) dt \tag{A.20}$$

(we marginalise out the non-target variables)

$$= \int_{U^T} \int_{\Omega^{\mathbf{N}}} \mu(\phi^{\mathbf{V}})(t, \mathbf{e}, \mathbf{n}) dt d\mathbf{n} \tag{A.21}$$

(we rewrite using Equation 2.4)

$$= \int_{U^T} \int_{\Omega^{\mathbf{N}}} \prod_{V \in \mathbf{V}} \mu(\phi^{V|\rho(V)})(v|\rho(v)) d\mathbf{n} dt \tag{A.22}$$

(we split the domain of $\mathbf{N}$ into truncated functions. Here $I^{\mathbf{V}+} = I^{\mathbf{N}+} \cap \{I^{T+}\} \cap I^{\mathbf{E}+}$)

$$= \int_{U^T} \sum_{I^{\mathbf{N}+} \in \mathbf{I}^{\mathbf{N}+}} \int_{I^{\mathbf{N}+}} \prod_{V \in \mathbf{V}} \mu(\phi^{V|I^{\mathbf{V}+}})(v) d\mathbf{n} dt \tag{A.23}$$

(we take the summation out, which is independent of t)

$$= \sum_{I^{\mathbf{N}+} \in \mathbf{I}^{\mathbf{N}+}} \int_{U^T} \int_{I^{\mathbf{N}+}} \prod_{V \in \mathbf{V}} \mu(\phi^{V|I^{\mathbf{V}+}})(v) d\mathbf{n} t \tag{A.24}$$

(we split $\mathbf{N}$ into A and $\mathbf{N}'$ ($A \in \mathbf{N}$ since $A \neq T, A \notin \mathbf{E}$) and we split $\mathbf{V}$ into A and $\mathbf{V}'$

$$= \sum_{I^{\mathbf{N}+} \in \mathbf{I}^{\mathbf{N}+}} \int_{U^T} \int_{I^{A+}} \int_{I^{\mathbf{N}'+}} \mu(\phi^{A|I^{\mathbf{V}+}})(a) \prod_{V' \in \mathbf{V}'} \phi^{V'|I^{\mathbf{V}+}}(v') d\mathbf{n}' \, da \, dt \tag{A.25}$$

(both the truncated function and the integral over a are independent of both $\mathbf{n}'$ and t so we move them out of those integrals)

$$= \sum_{I^{\mathbf{N}+} \in \mathbf{I}^{\mathbf{N}+}} \int_{I^{A+}} \mu(\phi^{A|I^{\mathbf{V}+}})(a) \int_{U^T} \int_{I^{\mathbf{N}'+}} \prod_{V' \in \mathbf{V}'} \phi^{V'|I^{\mathbf{V}+}}(v') d\mathbf{n}' \, dt \, da \tag{A.26}$$

(the integral over U^T and $I^{\mathbf{N}'}$ are constant in both μ and a. Because of the former we can make them $O(1)$, and because of the latter we can move it out of the integral over I^A)

$$= \sum_{I^{\mathbf{N}+} \in \mathbf{I}^{\mathbf{N}+}} O(1) \int_{I^{A+}} \mu(\phi^{A|I^{\mathbf{V}+}})(a) da \tag{A.27}$$

A.3 Sensitivity Analysis

A.3.1 Stretch

We will show that the stretch variation preserves the probability mass of the varied parameter. We note:

$$\begin{aligned}
&\int_{J^A} avg(f^A_{J^A, J^{\rho(A)}}) da \\
=&\int_{} J^A \frac{\int_{J^A} f^A_{J^A, J^{\rho(A)}}(a) da}{|J^A|} da \\
=&|J^A| \cdot \frac{\int_{J^A} f^A_{J^A, J^{\rho(A)}}(a) da}{|J^A|} \\
=&\int_{J^A} f^A_{J^A, J^{\rho(A)}}(a) da
\end{aligned}$$

so it follows that:

$$
\begin{aligned}
&\int_{J^A} \mu_{\text{Stretch}x}(f^A_{J^A,J^{\rho(A)}})(a)da \\
=&\int_{J^A} (f^A_{J^A,J^{\rho(A)}}(a) - avg(f^A_{J^A,J^{\rho(A)}})) \cdot x + avg(f^A_{J^A,J^{\rho(A)}})da \\
=&(\int_{J^A} f^A_{J^A,J^{\rho(A)}}(a)da - \int_{J^A} avg(f^A_{J^A,J^{\rho(A)}})da) \cdot x + \int_{J^A} avg(f^A_{J^A,J^{\rho(A)}})da \\
=&(\int_{J^A} f^A_{J^A,J^{\rho(A)}}(a)da - \int_{J^A} f^A_{J^A,J^{\rho(A)}}(a)da) \cdot x + \int_{J^A} f^A_{J^A,J^{\rho(A)}}(a)da \\
=&\int_{J^A} f^A_{J^A,J^{\rho(A)}}(a)da
\end{aligned}
$$

A.3.2 Sensitivity Function for Stretch

Here we will show the rewriting of Equation 5.13 into Equation 5.14.

$$
\begin{aligned}
&c_0 + \sum_{J^{A+} \in \mathbf{J}^{A+}} c_{J^{A+}} \int_{J^{A+}} \mu_{\text{Stretch}x}(f^A_{J^A,J^{\rho(A)}})(a)da \\
=&c_0 + \sum_{J^{A+} \in \mathbf{J}^{A+}} c_{J^{A+}} \int_{J^{A+}} (f^A_{J^A,J^{\rho(A)}}(a) - avg(f^A_{J^A,J^{\rho(A)}})) \cdot x + avg(f^A_{J^A,J^{\rho(A)}})da \\
=&\sum_{J^{A+} \in \mathbf{J}^{A+}} \left(x \cdot |J^{A+}| \cdot c_{J^{A+}} \int_{J^{A+}} (f^A_{J^A,J^{\rho(A)}}(a) - avg(f^A_{J^A,J^{\rho(A)}}))da \right) + \\
&c_0 + \sum_{J^{A+} \in \mathbf{J}^{A+}} \int_{J^{A+}} avg(f^A_{J^A,J^{\rho(A)}})da \\
=&x \cdot |J^A| \cdot \sum_{J^{A+} \in \mathbf{J}^{A+}} \left(c_{J^{A+}} \int_{J^{A+}} (f^A_{J^A,J^{\rho(A)}}(a) - avg(f^A_{J^A,J^{\rho(A)}}))da \right) + \\
&c_0 + \sum_{J^{A+} \in \mathbf{J}^{A+}} \int_{J^{A+}} avg(f^A_{J^A,J^{\rho(A)}})da \\
=&c_1 + c_2 \cdot x
\end{aligned}
$$

for some $c_1, c_2 \in \mathbb{R}$ independent of x.

Bibliography

[1] mte_pos. `http://leo.ugr.es/elvira/devel/Formato/Networks/Continuous/mte_pos.elv`. Accessed: 2017-10-28.

[2] C. Bielza and P. Larrañaga. Discrete Bayesian network classifiers: a survey. *ACM Computing Surveys (CSUR)*, 47(1):5, 2014.

[3] E. Castillo and U. Kjærulff. Sensitivity analysis in Gaussian Bayesian networks using a symbolic-numerical technique. *Reliability Engineering & System Safety*, 79(2):139–148, 2003.

[4] H. Chan and A. Darwiche. A distance measure for bounding probabilistic belief change. *International Journal of Approximate Reasoning*, 38(2):149–174, 2005.

[5] E. Consortium et al. Elvira: An environment for creating and using probabilistic graphical models. In *Proceedings of the first European workshop on probabilistic graphical models*, pages 222–230, 2002.

[6] G. Corliss and Y. Chang. Solving ordinary differential equations using Taylor series. *ACM Transactions on Mathematical Software (TOMS)*, 8(2):114–144, 1982.

[7] V. M. Coupé, L. C. van der Gaag, and J. D. F. Habbema. Sensitivity analysis: an aid for probability elicitation. *Knowledge Engineering Review*, 15(3):1–18, 2000.

[8] A. Fernández, M. Morales, C. Rodríguez, and A. Salmerón. A system for relevance analysis of performance indicators in higher education using Bayesian networks. *Knowledge and Information Systems*, 27(3):327–344, 2011.

[9] A. Fernández, R. Rumí, and A. Salmerón. Answering queries in hybrid Bayesian networks using importance sampling. *Decision Support Systems*, 53(3):580–590, 2012.

[10] M. A. Gómez-Villegas, P. Maín, and R. Susi. Sensitivity analysis in Gaussian Bayesian networks using a divergence measure. *Communications in Statistics—Theory and Methods*, 36(3):523–539, 2007.

[11] M. A. Gómez-Villegas, P. Main, and R. Susi. The effect of block parameter perturbations in Gaussian Bayesian networks: Sensitivity and robustness. *Information Sciences*, 222: 439–458, 2013.

[12] F. Johansson and G. Falkman. A Bayesian network approach to threat evaluation with application to an air defense scenario. In *Information Fusion, 2008 11th International Conference on*, pages 1–7. IEEE, 2008.

[13] C. E. Kahn, L. M. Roberts, K. A. Shaffer, and P. Haddawy. Construction of a Bayesian network for mammographic diagnosis of breast cancer. *Computers in biology and medicine*, 27(1):19–29, 1997.

[14] A. V. Kozlov and D. Koller. Nonuniform dynamic discretization in hybrid networks. In J. Robins and L. Wasserman, editors, *Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence*, pages 314–325. Morgan Kaufmann Publishers Inc., 1997.

[15] C. Lacave, M. Luque, and F. J. Díez. Explanation of Bayesian networks and influence diagrams in Elvira. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 37(4):952–965, 2007.

[16] H. Langseth, T. D. Nielsen, R. Rumí, and A. Salmerón. Mixtures of truncated basis functions. *International Journal of Approximate Reasoning*, 53(2):212–227, 2012.

[17] S. L. Lauritzen. Propagation of probabilities, means, and variances in mixed graphical association models. *Journal of the American Statistical Association*, 87(420):1098–1108, 1992.

[18] C.-J. Lee and K. J. Lee. Application of Bayesian network to the probabilistic risk assessment of nuclear waste disposal. *Reliability Engineering & System Safety*, 91(5):515–532, 2006.

[19] U. Lerner, E. Segal, and D. Koller. Exact inference in networks with discrete children of continuous parents. In J. Breese and D. Koller, editors, *Proceedings of the Seventeenth conference on Uncertainty in Artificial Intelligence*, pages 319–328, San Francisco, CA, USA, 2001. Morgan Kaufmann Publishers Inc., Morgan Kaufmann Publishers Inc. ISBN 1-55860-800-1.

[20] S. Moral, R. Rumí, and A. Salmerón. Mixtures of truncated exponentials in hybrid Bayesian networks. In S. Benferhat and P. Besnard, editors, *Proceedings of the Sixth European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty*, pages 156–167, London, UK, UK, 2001. Springer, Springer-Verlag. ISBN 3-540-42464-4.

[21] M. Neil, M. Tailor, D. Marquez, N. Fenton, and P. Hearty. Modelling dependable systems using hybrid Bayesian networks. *Reliability Engineering & System Safety*, 93(7):933–939, 2008.

[22] J. Pearl. *Probabilistic reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, 2014.

[23] I. Pérez-Bernabé and A. Salmerón. *MoTBFs: Learning Hybrid Bayesian Networks using Mixtures of Truncated Basis Functions*, 2015. URL `https://CRAN.R-project.org/package=MoTBFs`. R package version 1.0.

[24] P. P. Shenoy and J. C. West. Inference in hybrid Bayesian networks using mixtures of polynomials. *International Journal of Approximate Reasoning*, 52(5):641–657, 2011.

[25] M. H. Stone. Applications of the theory of Boolean rings to general topology. *Transactions of the American Mathematical Society*, 41(3):375–481, 1937.

[26] L. C. van der Gaag, S. Renooij, and V. M. Coupé. Sensitivity analysis of probabilistic networks. In *Advances in Probabilistic Graphical Models*, pages 103–124. Springer, 2007.

[27] C. Van Koten and A. Gray. An application of Bayesian network for predicting object-oriented software maintainability. *Information and Software Technology*, 48(1):59–67, 2006.