Parameter tuning

Thomas Jellema: INF/SCR-05-80

August 9, 2006

Contents

Chapter 1

Introduction

Probabilistic networks are used in many knowledge-based systems and for many tasks. For example, we can use probabilistic networks to model the weather or to model diseases. Probabilistic networks consist of two parts, namely a qualitative part and a quantitative part. The qualitative part of the probabilistic network describes the variables with their independency assumptions. The quantitative part of the probabilistic network is a set of conditional probabilities. Probabilistic networks are generally constructed with the help of experts from the domain of application [3].

A probabilistic network allows for computing any probability of interest. After creating a probabilistic network, you might find out that the network is not working properly. For example, you might find out that an output probability calculated by the network is not as high as it should. This can be caused either by a wrong assignment of one of the probabilities in the quantitative part or by a wrong assignment of the independency assumptions in the qualitative part.

In this thesis, we formulate a *constraint* on such an output probability. Then we identify those parameters, for which a change in value serves to enforce the constraint. Although assessing the probabilities of the network seems easy, little deviations of the ’real’ values can generate many wrong outcomes. The question is what you should do when you encounter such a wrong outcome. Trying to change the parameters by hand will often lead to disturbing the whole probability distribution, as results from changing one parameter are hard to predict. Adjusting the parameters in order to improve the results in the network is called parameter tuning. We will not discuss how to adjust the qualitative part of the network.

In this thesis, we have investigated the constraints that are useful to enforce when you want to tune the parameter in the network. We have found out that all these constraints can be enforced using *sensitivity functions*. These sensitivity functions describe the sensitivity of an output probability to a parameter and thus can be used to find the parameter value that enforces the constraint.

If we want to enforce a constraint, we will have to change some parameters. It is clear that changing a parameter can disturb probabilities in the network. In this thesis, we investigate how we can select the optimal parameter change. We formulate this *optimal parameter change* as the parameter change or parameter changes that maintain most of the output probabilities. We discuss the possible selection criteria for selecting this optimal parameter change from a set of parameter changes. We also discuss the selection criteria to select the optimal multiple parameter change when we want to change multiple parameters in one CPT. We have tested these selection criteria in an experiment, these selection criteria performed reasonable. However, the selection criteria did not find the optimal parameter change.

The structure of the rest of this thesis is as follows; the first chapter recapitulates the prerequisites we need to be able to understand this thesis. We will describe these prerequisites in a summarized fashion. In the third chapter, we investigate how we can use constraints to tune the network. We will explain how these constraints can be enforced using sensitivity functions and we will discuss how we can use the algorithms for finding the required sensitivity functions. In the fourth chapter, we discuss how to select the optimal parameter change(s). We investigate how we can select the parameter change that enforces the constraint and maintains most of the current output probabilities. In the final chapter, we present some conclusions of this thesis and directions for further research.

Chapter 2

Preliminaries

In this chapter, we will describe the prerequisites for reading this thesis. In this chapter, we will first discuss the probabilistic network in general. Then we will discuss the sensitivity functions.

2.1 What is a probabilistic network

A probabilistic network consists of a qualitative part and a quantitative part. The qualitative part of a belief network describes the variables with their independency assumptions. The quantitative part of the network is a set of conditional probabilities that describe the strengths of the dependences between the variables represented in the qualitative part [3].

Probabilistic networks predict the outcome given some observations as input. For example, a probabilistic network can be used to calculate the probability of having a disease, when having some symptoms.

More formally, we use the following definition when we discuss a probabilistic network.

Definition 1 *A probabilistic network is a tupe B=(G,Γ), where:*

- *G is an acyclic digraph with vertices V(G) = $\{V_1,...,V_n\}$, $n \geq 1$, and arcs A(G);*
- Γ *is a set of conditional probabilities* $\tau_{x|u} = Pr(x|u)$ *where* x *is a value associated with a variable of the digraph and* u *denotes a combination of*

values for the parents of that variable in the diagraph. We will call these conditional probabilities the parameters of the network.

The conditional probabilities together define a unique joint probability distribution Pr on $V(G)$ that respects the independences portrayed in G [7]. These parameters are stored in conditional probability tables (CPTs).

2.2 Sensitivity functions

Sensitivity analysis can be used to find the one-way sensitivity function that describes the relation between an output probability and a single parameter. In this thesis, we use this sensitivity function to find the change in a parameter that enforces a constraint. As described by Coupe and van der Gaag [3], this sensitivity function is always of the following form:

Theorem 1 *Let $Pr(a|e)$ denote an output probability, where a is a value of the variable A and e denotes the observations. Let $\tau_{x|u}$ denote a parameter of the network as before. Then:*

$$Pr(a|e)(\tau_{x|u}) = \frac{Pr(a \wedge e)(\tau_{x|u})}{Pr(e)(\tau_{x|u})} = \frac{\alpha \cdot \tau_{x|u} + \beta}{\gamma \cdot \tau_{x|u} + \sigma},$$

Where α ,β, γ and σ are constants of the sensitivity function.

We will use $Pr(a|e)(\tau_{x|u})$ to denote the sensitivity function that describes the sensitivity of the probability $Pr(a|e)$ to the parameter $\tau_{x|u}$. This sensitivity function is either a linear function or a fragment of a rectangular hyperbola, Renooij and van der Gaag [8].

As described by the axioms of the probability theory the values of the output probabilities $Pr(a|e)$ have to be between 0 and 1; the same holds for the values of a parameter $\tau_{x|u}$. Moreover, $Pr(a|e)(\tau_{x|u})$ must be well-defined for each value of $\tau_{x|u}$. We have that in the domain between 0 and 1 there cannot be an asymptote. We conclude that in the domain between 0 and 1 the sensitivity function is either monotonically non-increasing or monotonically non-decreasing.

2.3 Obtaining the sensitivity functions

Kjærulff and van der Gaag [5] have described two algorithms for finding sensitivity functions.

The first algorithm is OneOutAllIn, which takes a probability $Pr(a|e)$ as input and returns for each network parameter $\tau_{x|u}$ the constants α, β, γ and σ of the sensitivity function $Pr(a|e)(\tau_{x|u})$. A call to the algorithm OneOutAllIn with input $Pr(a|e)$ will be denoted by $OneOutAllIn : Pr(a|e)$

The second algorithm is AllOutOneIn, which takes the parameter $\tau_{x|u}$ and the observation e as input and returns for each probability $Pr(a|e)$ the costants α, β, γ and σ of the sensitivity function $Pr(a|e)(\tau_{x|u})$. A call to the algorithm AllOutOneIn with input $\tau_{x|u}$ and e will be denoted by $AllOutOneIn : (e)(\tau_{x|u})$

Computing the probabilities of the network requires one inward and two outward propagations of the network. The algorithms, $OneOutAllIn$ and $AllOutOneIn$ can compute the values of the constants of these sensitivity functions with one inward an two outward propagations of the network which is thus an additional cost of one outward propagation. The computing cost of these inward and outward propagations are not linear with the size of the network. With propagation, we will denote one inward and one outward propagation of the network. The cost of the algorithms, $OneOutAllIn$ and $AllOutOneIn$ can thus be seen as requiring 1.5 propagations of the network.

At this moment there are no efficient algorithms to calculate for each of the parameters and for each of the output probabilities in the network the values of the constants of the sensitivity functions that describes the sensitivity of the output probability with respect to a parameter, $AllOutAllIn$. An N-way sensitivity analysis as described in [5] is computationally very demanding.

Chapter 3

Possible constraints and how to enforce them

Upon reasoning with a probabilistic network, we may find out that the network answers a certain query incorrectly. Parameter tuning can be used to improve the results returned by a probabilistic network. To this end, we formulate a *constraint* on the possible values of one or more of the network's output probabilities. After we have formulated the constraint, we identify those parameters, for which a change in value serves to enforce the constraint.

In this chapter, we first consider all the constraints that may be interesting when tuning a network. Then we will show how to determine the parameter changes that serves to enforce these constraints.

3.1 Constraints of interest

In this section, we will list the important constraints we would like to be able to enforce when tuning a network. We will give some intuition about the constraints.

In the remainder of this section we will assume that a and b are values of variables A and B in the network. Variables A and B may be the same, unless stated otherwise. In addition, we assume that e_1 and e_2 are observations for two subsets E_1 and E_2, respectively, of network variables, where E_1 and E_2 do not contain the variables A and B. Sets E_1 and E_2 may be equivalent and so may the observations e_1 and e_2, unless stated otherwise. If e_1 and e_2 are necessarily equivalent and hence $E_1 = E_2$, we remove the subscripts and use e and E. Finally, we assume that η is a constant.

1 *Constraining a single probability*

When testing a network, a domain expert may find that an outcome probability that the network is giving is unexpected. When the expert has an idea of a more realistic value for the outcome, then we want to enforce a constraint of the form $Pr(a|e) = \eta$, where η is considered the correct outcome probability.

Suppose we are investigating a network that is used to predict the probability of having Glandular fever, also know as the Kissing disease. The expert, in this case a physician, may know the exact probability of having the Glandular fever given an inflammation of the throat and fever. If this presumed probability is different from the output probability of the network, the physician may find that the latter has to change. To change this output probability we need to change one or multiple network parameters.

Another possibility is that the physician only knows that the probability computed by the network should be higher or lower than the current probability or that the physician knows that the probability computed should be higher than a presumed value. For example, the physician could know that the probability of Glandular fever when having an inflammation of the throat and fever is much higher than the current output value, but he does not know what it should be exactly. Then we can use a slight variation of the constraint: $Pr(a|e) < \eta$ (1a) or $Pr(a|e) > \eta$ (1b).

We will investigate how to determine the parameter changes that serve to enforce this type of constraint in Section 3.2.

2 *Constraining the probability of the occurrence of two simultaneous events*

When testing a network, we could check whether the network returns a realistic probability of the occurrence of two values of two (different) output variables at the same time. If we know the exact probability, then we want to enforce a constraint of the form $Pr(a \wedge b|e) = \eta$.

We could, for example investigate a network modelling the flu and its symptoms. If a physician knows that the probability of having a fever and coughing when having a flu should be 0.2 instead of the current probability 0.6, then you would want to enforce the constraint: $Pr(fever \wedge coughing|flu) = 0.2$.

Instead, the expert may want to enforce the less strict constraints $Pr(a \wedge b|e) < \eta$ (2a) or $Pr(a \wedge b|e) > \eta$ (2b). For example, the physician may want to make the probability of both coughing and having a fever when having the flu lower than 0.2.

When we want to change the probability of one of the two different output values occurring, we need a constraint of the form $Pr(a \vee b|e) = \eta$. For example, the expert may know the probability of having a fever or coughing (or both). Since $Pr(a \vee b|e)$= $1 - Pr(\neg a \wedge \neg b|e)$ we can establish this by enforcing the constraint $Pr(\neg a \wedge \neg b|e) = 1 - \eta$.

We will investigate how to determine the parameter changes that serve to enforce this type of constraint in Section 3.3.

3 *Constraining the ratio of two output probabilities*

The constraint $Pr(a|e_1) = \eta \cdot Pr(b|e_2)$ can be used to constrain the ratio of two output probabilities. For example, you may want to change the probabilities of two outcome values in such a way that one outcome value is twice as likely as the other outcome value.

Suppose we are still investigating a network modelling the flu and its symptoms. If the physician knows that the probability of coughing when having the flu is four times as likely as the probability of having a fever when having the flu, then you could impose the following constraint $Pr(coughing|flu) = 4 \cdot Pr(fever|flu)$.

If we know that two outcome values have to have equal probabilities, we will have to impose the constraint: $Pr(a|e_1) = \eta \cdot Pr(b|e_2)$ with $\eta = 1$.

Off course when we only know that one outcome probability has to be higher or lower than another outcome probability then we would need the constraint $Pr(a|e_1) < \eta \cdot Pr(b|e_2)$ (3a) or $Pr(a|e_1) > \eta \cdot Pr(b|e_2)$ (3b).

We will investigate how to determine the parameter changes that serve to enforce this type of constraint in Section 3.4.

4 *Constraining the difference between two output probabilities*

An expert might know the exact value of the difference between two probabilities. If this difference is unequal to the difference between the two corresponding probabilities computed from the network, then we want to enforce a constraint of the form $Pr(a|e_1) - Pr(b|e_2) = \eta$.

Suppose we are investigating a probabilistic network that contains an alarm system that is able to detect fire and burglars. Suppose that the manufacturer of the alarm system knows that the alarm system detects a fire slightly better than it detects a burglar. He knows the difference has to be around 0.1. Then we could impose the constraint $Pr(alarm|fire) - Pr(alarm|burglar) = 0.1$.

Sometimes we do not know the exact value of the difference between two probabilities, but we know that the difference between two probabilities has to be higher than a certain value. For example, the probability of having the Glandular fever when having an inflammation of the throat and a fever has to be higher than the probability of the Glandular fever when having only an inflammation of the throat. In this case we can use a constraint of the form $Pr(a|e_1) - Pr(b|e_2) < \eta$ (4a) or $Pr(a|e_1) - Pr(b|e_2) > \eta$.

We will investigate how to determine the parameter changes that serve to enforce this type of constraint in Section 3.5.

Combining different types of constraints

It is also possible to enforce any combination of multiple constants. For example, the physician may know that the probability of having a flu when coughing, $Pr(flu|coughing)$, has to be higher than the current probability $Pr(flu|coughing)^o$, but the probability of having a flu in general $Pr(flu)$, cannot be higher than 0.1. In this case you would like to enforce $Pr(flu|coughing) > Pr(flu|coughing)^o$ and $Pr(flu) \leq 0.1$ at the same time.

We will briefly discuss combining different types of constraint in Section 3.6.

3.2 Constraining a single probability

As we have seen, if we want to change a single output probability we have to enforce a constraint of the form $Pr(a|e) = \eta$. In this section, we will investigate whether or not this constraint can be enforced and if so, how. We are going to use the sensitivity functions relating the output probability of interest to the different network parameters, to determine if there are parameters that can be changed such that the constraint is enforced.

3.2.1 Calculating the new parameter values

With the constants of the sensitivity function for the output probability $Pr(a|e)$ with respect to a single parameter $\tau_{x|u}$, we can calculate the value of the parameter which enforces the output probability $Pr(a|e)$ to be equal to η.

Recall that the sensitivity function, $Pr(a|e)(\tau_{x|u})$, has the following form:

$$Pr(a|e) = \frac{\alpha \cdot \tau_{x|u} + \beta}{\gamma \cdot \tau_{x|u} + \sigma}$$

To determine the value of the parameter $\tau_{x|u}$ for which it holds that $Pr(a|e) = \eta$, we have to solve:

$$\frac{\alpha \cdot \tau_{x|u} + \beta}{\gamma \cdot \tau_{x|u} + \sigma} = \eta$$

which gives:

$$\tau_{x|u} = \frac{\sigma \cdot \eta - \beta}{\alpha - \gamma \cdot \eta}$$

It is obvious from this formula, that once we obtain the constants of the sensitivity functions for all the parameters, we can determine for each parameter the value that enforces the constraint. If the formula returns a value of the parameter that is higher than one or lower than zero, then the constraint cannot be enforced using this parameter.

Instead of changing a single parameter, a constraint of this form can also be enforced by changing multiple parameters simultaneously. The advantage of changing multiple parameters is that in some cases this will cause less changes in the probability distribution. We will cover this subject in Chapter 4. Another advantage is that there are constraints that can be enforced changing multiple parameters, which could not be enforced changing a single parameter.

Recall that enforcing constraint of this form using multiple parameter changes would require us to determine the constants of all n-way sensitivity functions for each combination of parameters. This is known to be computationally very demanding. If we only allow multiple parameter changes in a single conditional probability table (CPT) then this is computationally still doable.

To investigate the sensitivity function for multiple parameters, we will first examine the sensitivity function for two parameters, the 2-way function. Under the assumption that both parameters are taken from the same CPT, we now know that we can obtain this 2-way function from the sensitivity functions for the single parameter changes.

Let $\tau_i, i = 1, 2$, denote the two parameters that we want to change and let α_i, β_i,γ_i and σ_i, $i = 1, 2$, denote the values of the constants of the sensitivity function that describes the sensitivity of the output probability of interest $Pr(a|e)$ to parameter τ_i, then we have the following sensitivity functions for the single parameter changes:

$$Pr(a|e) = \frac{\alpha_1 \cdot \tau_1 + \beta_1}{\gamma_1 \cdot \tau_1 + \sigma_1}$$

$$Pr(a|e) = \frac{\alpha_2 \cdot \tau_2 + \beta_2}{\gamma_2 \cdot \tau_2 + \sigma_2}$$

The general form of the 2-way sensitivity function is.

$$Pr(a|e) = \frac{\delta \cdot \tau_1 \cdot \tau_2 + \alpha_1 \cdot \tau_1 + \alpha_2 \cdot \tau_2 + \kappa}{\epsilon \cdot \tau_1 \cdot \tau_2 + \gamma_1 \cdot \tau_1 + \gamma_2 \cdot \tau_2 + \mu}$$

This is a complex function to use because it requires the computation of a larger number of constants. However, in the specific case that the parameters in the function are from the same CPT, we can simplify this function. Because the parameters are from the same CPT, we know that there are no interaction terms. We know this, as the probabilities corresponding the parameters have observations that are contradictory. This means that the constants δ and ϵ are zero and that the sensitivity function now reduces to the following form:

$$Pr(a|e) = \frac{\alpha_1 \cdot \tau_1 + \alpha_2 \cdot \tau_2 + \kappa}{\gamma_1 \cdot \tau_1 + \gamma_2 \cdot \tau_2 + \mu}$$

Where the constants κ and μ are the values of $Pr(a \wedge e)$ and respectively. $Pr(e)$ if the parameters τ_1 and τ_2 are zero.

If we want to change all the parameters in the CPT, we have to sum over all these parameter.

$$Pr(a|e) = \frac{\sum_i \alpha_i \cdot \tau_i + \kappa}{\sum_i \gamma i \cdot \tau_i + \mu}$$

We can determine the values of κ and μ by calculating the probabilities $Pr(a \wedge e)$ and $Pr(e)$ after setting all the parameters in the CPT to zero:

Let τ_i^o denote the current value of parameter τ_i, then we know that the difference in the probability $Pr(a \wedge e)$ after changing parameter τ_i^o to zero is equal to $\alpha_i \cdot \tau_i^o$. We know that the values of $Pr(a \wedge e)$ and $Pr(e)$ are, for example, equal to the values of β_1 and σ_1 if the parameter τ_1 is zero and the other parameters have the original values.

Thus κ has the following value:

$$\kappa = \beta_1 - \sum_{i \neq 1} \alpha_i \cdot \tau_i^o$$

Where $\sum_{i \neq 1} \alpha_i \cdot \tau_i^o$ equals the difference in the probability $Pr(a \wedge e)$ when we change the current probabilities of the parameters, except the first, to zero.

We can calculate μ using the same method:

$$\mu = \sigma_1 - \sum_{i \neq 1} \gamma_i \cdot \tau_i^o$$

Where $\sum_{i \neq 1} \gamma_i \cdot \tau_i^o$ equals the difference in the probability $Pr(e)$ when we change the current probabilities of the parameters, except the first, to zero.

We conclude that after obtaining the sensitivity functions for each parameter in the network, we can determine which single parameter change or multiple parameter changes in one CPT, serve to impose the given constraint. In the next section, we discuss how to compute the constants required.

3.2.2 Which parameters enforce the constraint

In the previous section we have obtained the functions for finding the parameter changes that can enforce the constraint $Pr(a|e) = \eta$. To be able to use these functions we will need the values of the constants of the sensitivity function $Pr(a|e)(\tau_{x|u})$. There are two algorithms available that serve to directly compute the values of the constants.

The two algorithms can actually exploited to answer two different types of tuning questions. The first type of question is the type we have addressed so far and relates to which of a large number of parameters serve(s) to enforce a constraint on a single, or a small number of, output probabilities. We will refer to such tuning questions as *parameter questions*. We could for example want to know the parameter changes for each parameter in the network that enforce a constraint, or we could want to know the smallest absolute parameter change that would enforce a constraint. For this type of question, we employ the algorithm OneOutAllIn mentioned in Section 2.

The second type of question addresses situations concerning a small number of parameters and a large number of output probabilities. We will refer to these questions as *output questions* and employ the algorithm AllOutOneIn to answer them. For example, when evaluating the network, we might find out that many constraints are not satisfied. If we are uncertain about the value of a particular parameter, we could examine whether there is a value of that parameter that enforces all constraints. Another output question could be that we want to increase a probability, without changing the other probabilities containing the same observations more than 5

In the next section, we are going to investigate these questions. First, we are going to consider the parameter questions. We will start with parameter questions involving equality constraints. Then we will consider parameter questions involving inequality constraints. Subsequently, we will give an example where we need to enforce a constraint using multiple parameter changes in one CPT. Finally, we will consider the output questions.

We will use the example network of Figure 3.1 in the examples. This network is a modified version of the network described in [7].

The network has the following parameters:

$\tau_{B=burgling}$ = 0.2

$\tau_{E=active}$ = 0.6

$\tau_{W=broken|B=burgling}$ = 0.4
$\tau_{W=broken|B=sleeping}$ = 0.9

$\tau_{A=ringing|B=burgling \wedge E=active}$ = 0.1
$\tau_{A=ringing|B=burgling \wedge E=active}$ = 0.4
$\tau_{A=ringing|B=sleeping \wedge E=inactive}$ = 0.8
$\tau_{A=ringing|B=sleeping \wedge E=inactive}$ = 0.3

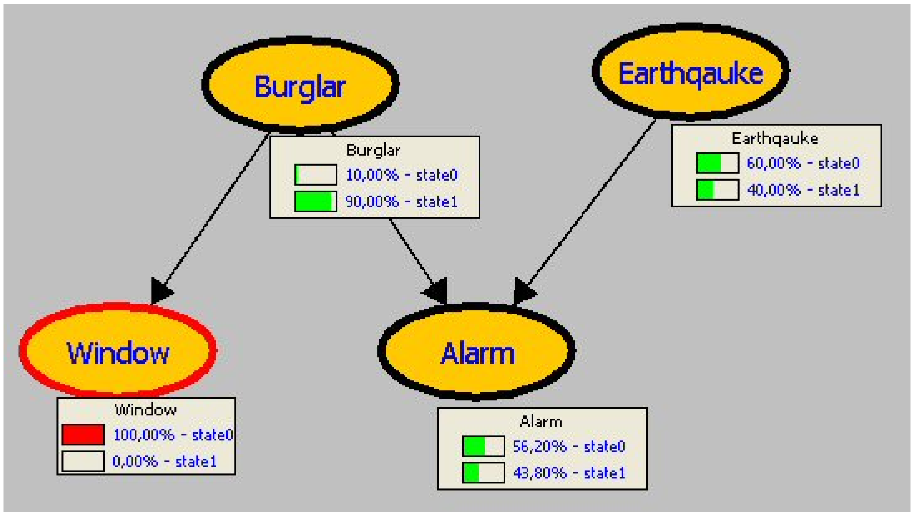

Figure 3.1: An example network

Parameter questions

$OneOutAllIn : Pr(a|e)$ returns the values of the constants of all the sensitivity functions $Pr(a|e)(\tau_{x|u})$ for one output probability $Pr(a|e)$ with respect to all parameters in the network. Using these sensitivity functions we can determine for each parameter which value, if any, serves to enforce the constraint $Pr(a|e) = \eta$. It is likely that more than one parameter upon changing enforces the constraint. In that case, we need to choose one. Criteria for doing so, for example the parameter that requires the smallest absolute change, are discussed in Chapter 4. In the next example, we will try to answer such a question.

Example 1 *Suppose we want to change the probability $Pr(A = ringing|W = broken)$ for the network in Figure 3.1. We want to know for each parameter, the parameter change that enforces the constraint. Of course, it could also be the case that for some parameter the constraint cannot be enforced at all. As we can see, the probability $Pr(A = ringing|W = broken)$ currently is 0.562. Suppose we want to change this to 0.60. A call to $OneOutAllIn : Pr(A = ringing|W = broken)$,returns the constants of the sensitivity functions, which are summarised in Table 3.1. We can use these constants to find which parameters can, upon change, enforce the given constraint, and to determine by how much they should be changed.*

parameter	α	β	γ	σ
$\tau_{B=burgling}$	-0.452	0.540	-0.500	0.900
$\tau_{E=active}$	0.336	0.248	0.000	0.800
$\tau_{W=broken\|B=burgling}$	0.044	0.432	0.200	0.720
$\tau_{W=broken\|B=sleeping}$	0.480	0.018	0.800	0.080
$\tau_{A=ringing\|B=burgling\wedge E=active}$	0.048	0.444	0.000	0.800
$\tau_{A=ringing\|B=burgling\wedge E=inactive}$	0.032	0.437	0.000	0.800
$\tau_{A=ringing\|B=sleeping\wedge E=active}$	0.432	0.104	0.000	0.800
$\tau_{A=ringing\|B=sleeping\wedge E=inactive}$	0.288	0.363	0.000	0.800

Table 3.1: Values of the constants of the sensitivity functions for the outcome probability $Pr(A = ringing|W = broken)$ for each parameter.

For each parameter we compute from the constants of its corresponding sensitivity function

$$\tau_{x|u} = \frac{0.60 \cdot \sigma - \beta}{\alpha - 0.60 \cdot \gamma}$$

For example, for the parameter $\tau_{B=burgling}$ *we find*

$$\tau_{B=burgling} = \frac{0.60 \cdot 0.900 - 0.540}{-0.452 + 0.60 \cdot 0.500} = 0.000$$

Similarly, for parameter $\tau_{E=active}$ *we get:*

$$\tau_{E=active} = \frac{0.60 \cdot 0.800 - 0.248}{0.336 - 0.60 \cdot 0.000} = 0.690$$

So changing the parameter $\tau_{B=burgling}$ *from its original value of 0.200 to 0.000, or changing* $\tau_{E=active}$ *from 0.600 to 0.690 will result in a change in the output probability* $Pr(A = ringing|W = broken)$ *from 0.560 to 0.600.*

Other parameter changes can also accomplish this, as can be seen from Table 3.2, which lists the parameter values that enforce the constraint. Notice the difference of change necessary for the different parameters. Furthermore, note that changing some of the parameters cannot give the desired effect at all. This is indicated in the table as np (not possible).

The above method can also be applied to answer another parameter question. It can be used to find all the parameter changes that enforce constraints of the type $Pr(a|e) > \eta$ or $Pr(a|e) < \eta$. However, additional information is required besides the constants obtained using the above method:

parameter $\tau_{x\|u}$	**current** $\tau^o_{x\|u}$	**new** $\tau_{x\|u}$	$\|\tau_{x\|u} - \tau^o_{x\|u}\|$
$\tau_{B=burgling}$	0.200	0.000	0.200
$\tau_{E=active}$	0.600	0.690	0.090
$\tau_{W=broken\|B=burgling}$	0.400	0.000	0.400
$\tau_{W=broken\|B=sleeping}$	0.900	np	np
$\tau_{A=ringing\|B=burgling \wedge E=active}$	0.100	0.750	0.650
$\tau_{A=ringing\|B=burgling \wedge E=active}$	0.400	np	np
$\tau_{A=ringing\|B=sleeping \wedge E=inactive}$	0.800	0.870	0.070
$\tau_{A=ringing\|B=sleeping \wedge E=inactive}$	0.300	0.406	0.106

Table 3.2: Parameter changes that serve to enforce the constraint $Pr(A = ringing|W = broken) = 0.60$.

we need to know whether the sensitivity function is increasing or decreasing. As argued in Section 2.2, the sensitivity function is either monotonically increasing or monotonically decreasing. We know that the sensitivity function is increasing when you need a higher value of the parameter to increase the output probability, or when we need a lower parameter value to decrease the output probability.

In Figure 3.2 we have shown an example graph of a monotonically increasing sensitivity function. Suppose with the above method we find that to enforce $Pr(a|e) = \eta$ the parameter $\tau_{x|u}$ should have the new value $\tau^1_{x|u}$. We now know that if the sensitivity function is monotonically increasing then:

$Pr(a|e) > \eta$ when $\tau_{x|u} > \tau^1_{x|u}$.

If the sensitivity function is monotonically decreasing then:

$Pr(a|e) > \eta$ when $\tau_{x|u} < \tau^1_{x|u}$.

In the previous example, we have found the parameter values that enforce the constraint $Pr(a|e) = \eta$. In the next example we are going to use these parameter values to enforce the constraint $Pr(a|e) > \eta$.

Example 2 *Suppose we are using the same network from Figure 3.1 but now we want the output probability $Pr(A = ringing|W = broken)$, which is 0.562, to be higher than 0.60. In other words, we have that $Pr(A = ringing|W = broken)^o$=0.562 and $Pr(A = ringing|W = broken)^1$=0.60. To enforce the mentioned constraint we can use the results from the previous example.*

In the previous example we found that if we want to change the value of $Pr(A = ringing|W = broken)$ from 0.562 to 0.60 using $\tau_{E=active}$, then we have to change

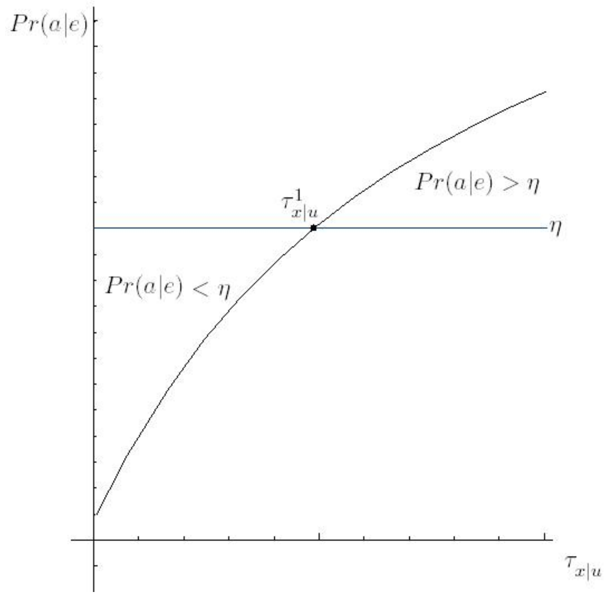

Figure 3.2: If the sensitivity function is monotonically increasing, then $Pr(a|e) > \eta$ when $\tau_{x|u} > \tau^1_{x|u}$.

this parameter from $\tau^o_{E=active} = 0.60$ *to* $\tau^1_{E=active} = 0.690 = 0.690$.

Since $\tau^o_{E=active} < \tau^1_{E=active}$ *and* $Pr(A = ringing|W = broken)^o < Pr(A = ringing|W = broken)^1$ *we know that the sensitivity function relating* $Pr(A = ringing|W = broken)$ *to parameter* $\tau_{E=active}$ *is monotonically increasing. As the sensitivity function is monotonically increasing we know, as seen in Figure 3.2, that the value of* $\tau_{E=active}$ *has to be higher than 0.690 in order for the probability of interest to become higher than 0.60. We can repeat this argument for each parameter in the network. The parameter changes that enforce the constraint* $Pr(A = ringing|W = broken) > 0.6$ *are summarised in Table 3.3.*

Upon tuning, the constraint $Pr(a|e) > \eta$ can also be used to ensure that the value of $Pr(a|e)$ becomes higher than its current value. In that case, we only have to know whether the sensitivity function is monotonically increasing or monotonically decreasing to enforce the constraint.

There are also parameter questions for which you would like to know the values of multiple parameter changes in one CPT to enforce a constraint. When we are uncertain about one parameter in the CPT, we are probably uncertain about the other parameters in that CPT as well. This approach,

parameter	**current** $\tau^o_{x\|u}$	**new** $\tau_{x\|u}$	$\|\tau_{x\|u} - \tau^o_{x\|u}\|$
$\tau_{B=burgling}$	0.200	np	np
$\tau_{E=active}$	0.600	> 0.690	> 0.090
$\tau_{W=broken\|B=burgling}$	0.400	np	np
$\tau_{W=broken\|B=sleeping}$	0.900	np	np
$\tau_{A=ringing\|B=burgling \wedge E=active}$	0.100	> 0.750	> 0.650
$\tau_{A=ringing\|B=burgling \wedge E=active}$	0.400	np	np
$\tau_{A=ringing\|B=sleeping \wedge E=inactive}$	0.800	> 0.870	> 0.070
$\tau_{A=ringing\|B=sleeping \wedge E=inactive}$	0.300	> 0.406	> 0.106

Table 3.3: Parameter changes that serve to enforce the constraint $Pr(A = ringing|W = broken) > 0.60$.

changing multiple parameters in one CPT, can be more convenient than changing only one of them. Finding the optimal solution for the parameter changes in one CPT is very difficult and depends on the definition of optimal. We will consider finding optimal solutions in the next chapter.

In the following example, we will show how we can find these multiple parameter changes. However, using this method we find many solutions. In the next chapter, we will discuss how to select the optimal multiple parameter changes from a set of solutions.

Example 3 *Suppose we want to change the parameters in the CPT of Alarm of the network of Figure 3.1 in such a way that they enforce the constraint* $Pr(A = ringing|W = Broken) = 0.5$. *To find the sensitivity functions that are employed to compute the changes required for the parameters in the CPT, we only have to combine the constants of the sensitivity functions summarised in Table 3.1.*

For ease of reference we call:

$\tau_{A=ringing|B=burgling \wedge E=active} = \tau_1$

$\tau_{A=ringing|B=Burgling \wedge E=inactive} = \tau_2$

$\tau_{A=ringing|B=sleeping \wedge E=active} = \tau_3$

$\tau_{A=ringing|B=sleeping \wedge E=inactive} = \tau_4$

Recall that the sensitivity function for changing multiple parameters in one CPT is as follows:

$$Pr(a|e) = \frac{\sum_i \alpha_i \cdot \tau_i + \kappa}{\sum_i \gamma_i \cdot \tau_i + \mu}$$

Where κ is the value of $Pr(a \wedge e)$ and μ the value of $Pr(e)$ when the value of all the parameters in the CPT are zero.

Using the values of the constants of the sensitivity functions of the single parameter changes results in:

$$Pr(A = ringing|W = broken) = \frac{0.048 \cdot \tau_1 + 0.032 \cdot \tau_2 + 0.432 \cdot \tau_3 + 0.288 \cdot \tau_4 + \kappa}{0.000 \cdot \tau_1 + 0.000 \cdot \tau_2 + 0.000 \cdot \tau_3 + 0.000 \cdot \tau_4 + \mu}$$

Recall that the constant κ can be calculated from:

$$\kappa = \beta_1 - \sum_{i \neq 1} \alpha_i \cdot \tau_i^o$$

Using the values of the constants in Table 3.1 we calculate:

$$\kappa = 0.4496 - 0.048 \cdot 0.1 - 0.032 \cdot 0.4 - 0.432 \cdot 0.8 - 0.288 \cdot 0.3 = 0$$

The constant μ can be calculated using this formula:

$$\mu = \sigma_1 - \sum_{1.N} \gamma_u \cdot \tau_u^o$$

Where $\sum_{1.N} \gamma_u \cdot \tau_u^o$ equals the difference in the probability $Pr(e)$ when we change the current probabilities of the parameters, except the first, to zero.

Using the values of the constants in Table 3.1 we calculate:

$$\mu = 0.8 - 0.000 \cdot 0.1 - 0.000 \cdot 0.4 - 0.000 \cdot 0.8 - 0.000 \cdot 0.3 = 0.8$$

To determine the parameter changes that can enforce the constraint we now have to solve:

$$\frac{0.048 \cdot \tau_1 + 0.032 \cdot \tau_2 + 0.432 \cdot \tau_3 + 0.288 \cdot \tau_4 + 0.000}{0.000 \cdot \tau_1 + 0.000 \cdot \tau_2 + 0.000 \cdot \tau_3 + 0.000 \cdot \tau_4 + 0.800} = \eta$$

This equation has many solutions. In the next chapter, we will investigate how to choose the best combination of parameter changes. One easy way to find a solution is by taking the values of the parameters of the CPT to be equivalent. Of course, this solution is obvious; however, we will use this method to illustrate how to use the sensitivity function.

If all the parameter values are to be equivalent, we have that, $\tau_1 = \tau_2 = \tau_3 = \tau_4$*. We can replace these parameters with one variable* τ_{1-4}*, which now has the same value as the four parameters. This makes the equation much easier to solve:*

$$\frac{0.048 \cdot \tau_{1-4} + 0.032 \cdot \tau_{1-4} + 0.432 \cdot \tau_{1-4} + 0.288 \cdot \tau_{1-4} + 0.00}{0.000 \cdot \tau_{1-4} + 0.000 \cdot \tau_{1-4} + 0.000 \cdot \tau_{1-4} + 0.000 \cdot \tau_{1-4} + 0.8} = \frac{0.8 \cdot \tau_{1-4}}{0.8} = \eta$$

To determine the parameter changes that can enforce the constraint $Pr(A = ringing|W = Broken) = 0.5$ *with equal parameter values in the CPT of Alarm we have that:*

$$\frac{0.8 \cdot \tau_{1-4}}{0.8} = 0.5$$

When solving this equation we have that the value of τ_{1-4} *has to be 0.5, that is, we have to change all parameter values to 0.5 to enforce the constraint.*

In this paragraph we have investigated how we can answer parameter questions with respect to the constraint $Pr(a|e) = \eta$. We used the OneOutAllIn algorithm to find the values of the constants needed to find the sensitivity functions for all the parameters. If we want to answer output questions where the expert wants to enforce many constraints of the form $Pr(a|e) = \eta$ then using OneOutAllIn to solve this would be infeasible, since for each constraint, we would need to use one call to the algorithm. We will now discuss how to handle output questions more efficiently.

Output questions

Instead of using OneOutAllIn to enforce multiple constraints, which would require applying the algorithm multiple times, we can use the AllOutOneIn algorithm. With one run of $AllOutOneIn : (e)(\tau_{x|u})$ we get the sensitivity functions we can use to determine the parameter change of one parameter $\tau_{x|u}$ that serves to enforce the constraint $Pr(a|e) = \eta$ for all probabilities in the network that contain the evidence e. Using this algorithm, we can answer questions containing multiple constraints if these constraints involve probabilities containing the same observations.

With the sensitivity functions found from one run of AllOutOneIn, you could, for example, find the value of a single parameter that enforces the constraint $Pr(a|e) = 0$, but simultaneously enforce the constraint $Pr(b|e) = 1$, for $a \neq b$.

We can of course also want to answer output questions relating to inequality constraints. These questions are more interesting, because now we can find the parameter value for which multiple constraints containing inequalities hold. For example, for a certain parameter we could find the value of the parameter that enforces the constraint $Pr(a|e) < 0.1$, but also enforces the constraint $Pr(b|e) > 0.2$.

If we know between which values the output probabilities are likely to be, then we could define constraints for each output probability in the network containing the same observations, like $0.4 < Pr(a|e) < 0.6$ and $0.2 < Pr(b|e) < 0.3$. Using AllOutOneIn, we could find the values for the constants for the sensitivity function we need to determine the parameter changes to enforce these constraints. We will illustrate this in the following example.

Example 4 *Suppose we are uncertain about the parameter $\tau_{E=active}$. From interviewing the manufacturer of the alarm system, the police and a geographer, we know that the outputs of the network should satisfy the following constraints:*

- $Pr(B = burgling|W = broken) < 0.80$
- $Pr(A = ringing|W = broken) > 0.50$
- $Pr(E = active|W = broken) > 0.60$

First we have to solve the equation $Pr(a|e) = \eta$ for the three output probabilities. To find the parameter changes that enforce these constraints we need the sensitivity functions that describe the sensitivity of the three output probabilities to the parameter. With one run of $AllOutOneIn : (W = broken)(\tau_{E=active})$ we find the values of the constants of the sensitivity functions $Pr(a|e)(\tau_{E=active})$ for all output probabilities in the network with the observation $W = broken$. The values of the constants of these sensitivity functions are summarised in Table 3.4.

To find the parameter change that enforces $Pr(a|e) = \eta$ we compute:

$$\tau_{E=active} = \frac{0.60 \cdot \sigma - \beta}{\alpha - 0.60 \cdot \gamma}$$

For $Pr(B = burgling|W = broken)$ this gives:

$$\tau^1_{E=active} = \frac{0.800 \cdot 0.80 - 0.080}{0.000 - 0.00 \cdot 0.80} = \infty$$

output probability	α	β	γ	σ
$Pr(B = burgling \mid W = broken)$	0.000	0.080	0.000	0.800
$Pr(B = sleeping \mid W = broken)$	0.000	0.072	0.000	0.800
$Pr(E = active \mid W = broken)$	0.800	0.000	0.000	0.800
$Pr(E = inactive \mid W = broken)$	-0.800	0.800	0.000	0.800
$Pr(A = ringing \mid W = broken)$	0.336	0.248	0.000	0.800
$Pr(A = silent \mid W = broken)$	-0.336	0.552	0.000	0.800

Table 3.4: Values of the constants for the sensitivity functions for different outcome probabilities with respect to the parameter $\tau_{E=active}$.

For $Pr(E = active|W = broken)$ this gives:

$$\tau^1_{E=active} = \frac{0.800 \cdot 0.50 - 0.000}{0.800 - 0.00 \cdot 0.50} = 0.500$$

For $Pr(A = ringing|W = broken)$ this gives:

$$\tau^1_{E=active} = \frac{0.800 \cdot 0.60 - 0.248}{0.336 - 0.00 \cdot 0.60} = 0.690$$

When changing $\tau_{E=active}$, the probability $Pr(B = burgling|W = broken)$ can never become 0.8. This is because the probability $Pr(B = burgling|W = broken)$ is independent of the parameter $\tau_{E=active}$. We could have excluded this probability before running the analysis. We know that the probability of $Pr(B = burgling|W = broken)$ is 0.1, which is lower than 0.8, and will remain to be so for any value of this parameter.

For the second constraint, we know that the original probability is 0.6 and the original value of the parameter is also 0.600. So $Pr(E = active|W = broken)^o = 0.6$ and $\tau^o_{E=active} = 0.600$. The probability $Pr(E = active|W = broken)^1$ is 0.50 when $\tau^1_{E=active}$ is 0.500. Because $\tau^0_{E=active} > \tau^1_{E=active}$ and $Pr(E = active|W = broken)^0 > Pr(E = active|W = broken)^1$ we know that the sensitivity function is monotonically increasing. Thus we know that $Pr(E = active|W = broken) > 0.5$ when $\tau_{E=active} > 0.5$.

For the third constraint, we know that the original value of the probability is 0.562 and the original value of the parameter is 0.600. So we know that $Pr(E = active|W = broken)^o = 0.562$ and $\tau^o_{E=active} = 0.600$. The probability $Pr(E = active|W = broken)^1$ is 0.60 when $\tau^1_{E=active}$ is 0.690. Because $\tau^0_{E=active} < \tau^1_{E=active}$ and $Pr(E = active|W = broken)^0 < Pr(E =$

$active|W = broken)^1$ we know that the sensitivity function is monotonically increasing. Thus we know that $Pr(E = active|W = broken) > 0.6$ whenever $\tau_{E=active} > 0.690$.

With the original value of 0.600, only the first constraint was satisfied. The value of $\tau_{E=active}$ has to be higher than 0.690 to enforce all three constraints.

3.3 Constraining the probability of the occurrence of two simultaneous events

In this section, we investigate how we can constrain the probability of the occurrence of two events. We will consider the constraint $Pr(a \wedge b) = \eta$ for the situation where there are no observations and we will consider the constraint $Pr(a \wedge b|e) = \eta$ when there are observations.

3.3.1 Calculating the new parameter values

First, we will discuss the situation when we want to change the probability of occurrence of two events when there are no observations. We start with this situation, because this can be solved in an easier way than when there are observations.

With the constants of the sensitivity functions for the output probability $Pr(a|b)$ with respect to the parameter $\tau_{x|u}$ we can calculate the new value of the parameter which enforces the output probability $Pr(a \wedge b)$ to have the value η. This is because the sensitivity function of $Pr(a|b)$ is made out of the sensitivity function of two other probabilities, namely $Pr(a \wedge b)$ and $Pr(b)$:

$$\frac{Pr(a \wedge b)}{Pr(b)} = Pr(a|b)$$

In other words, we can use the numerator of the sensitivity function for $Pr(a|b)$ to get the sensitivity function of the probability $Pr(a \wedge b)$. We can find the values of the sensitivity function of the probability $Pr(a \wedge b)$ by using the α and β values of the sensitivity function of the probability $Pr(a|b)$:

$$Pr(a \wedge b) = \alpha \cdot \tau_{x|u} + \beta$$

Solving the equation $\alpha\tau_{x|u} + \beta = \eta$ gives:

$$\tau_{x|u} = \frac{\eta - \beta}{\alpha}$$

This indicates that if we have the values of the constants α and β of the sensitivity function that describes the relation between a parameter $\tau_{x|u}$ with respect to a output probability $Pr(a|b)$, then we can find the parameter value of $\tau_{x|u}$ that enforces the constraint $Pr(a \wedge b) = \eta$. This approach can only be used when there are no observations in the constraint.

In the case that we do have observations, we need two sensitivity functions. We need the sensitivity function that describes the relation between a parameter $\tau_{x|u}$ and the probabilities $Pr(a|b \wedge e)$ and $Pr(b|e)$ respectively. We will now show that using these sensitivity functions we can calculate the value of the parameter $\tau_{x|u}$ for which it holds that $Pr(a \wedge b|e) = \eta$. We first observe that from the definition of conditional probability, we have that.

$$Pr(a \wedge b|e)(\tau_{x|u}) = (Pr(a|b \wedge e) \cdot Pr(b|e))(\tau_{x|u})$$

$$\left[\frac{Pr(a \wedge b \wedge e)}{Pr(b \wedge e)} \cdot \frac{Pr(b \wedge e)}{Pr(e)}\right](\tau_{x|u}) = \left[\frac{Pr(a \wedge b \wedge e)}{Pr(b \wedge e)}\right](\tau_{x|u}) \cdot \left[\frac{Pr(b \wedge e)}{Pr(e)}\right](\tau_{x|u})$$

If we have the following sensitivity function $Pr(a|b \wedge e)(\tau_{x|u})$:

$$Pr(a|b \wedge e)(\tau_{x|u}) = \frac{\alpha_1 \cdot \tau_{x|u} + \beta_1}{\gamma_1 \cdot \tau_{x|u} + \sigma_1}$$

And the following sensitivity function $Pr(b|e)(\tau_{x|u})$:

$$Pr(b|e)(\tau_{x|u}) = \frac{\alpha_2 \cdot \tau_{x|u} + \beta_2}{\gamma_2 \cdot \tau_{x|u} + \sigma_2}$$

Then $\gamma_1 \cdot \tau_{x|u} + \sigma_1$ should be equivalent to $\alpha_2 \cdot \tau_{x|u} + \beta_2$ and multiplying both sensitivity functions, results in:

$$Pr(a \wedge b|e) = \frac{\alpha_1 \cdot \tau_{x|u} + \beta_1}{\gamma_2 \cdot \tau_{x|u} + \sigma_2}$$

We can find the values of the constants for the sensitivity function $Pr(a \wedge b|e)(\tau_{x|u})$ by using the values of the constants α and β of the sensitivity function $Pr(a|b \wedge e)(\tau_{x|u})$ and the values of the constants γ and σ of the sensitivity function $Pr(b|e)(\tau_{x|u})$.

the next section, we are going to investigate how to most efficiently find the constants of the sensitivity functions needed.

3.3.2 Which parameters enforce the constraint

We now know how to enforce the constraint, if we have the needed sensitivity functions. In this section, we investigate how to find these sensitivity functions. As in the previous section, we first discuss the parameter questions. These questions can be answered using the algorithm OneOutAllIn. Then we will discuss the output questions, which can be answered using the algorithm AllOutOneIn.

Parameter questions

We will first consider the constraint $Pr(a \wedge b) = \eta$. If we want to find the parameter change of parameter $\tau_{x|u}$ that enforces this constraint, we have to obtain the values of the constants α and β of the sensitivity function that describes the relation between the parameter and the output probability $Pr(a|e)$. If we want to find these parameter changes for all parameters in the network, we will have to obtain the values of the constants α and β for all these sensitivity functions. Recall that using the algorithm OneOutAllIn, we can find these values of the constants of the sensitivity functions.

Example 5 *Suppose we want to change the probability $Pr(A = ringing \wedge W = broken)$ to 0.4 in the network of Figure 3.1 and suppose we want to find the smallest parameter change that enforces the constraint. Choosing the smallest parameter change is one of the selection criteria that we will discuss in Chapter 4. To find these parameter changes we need to run the algorithm on the output probability $Pr(A = ringing|W = broken)$, because:*

$$Pr(A = ringing|W = broken) = \frac{Pr(A = ringing \wedge W = broken)}{Pr(W = broken)}$$

In this way, we can use the constants in the numerators of the sensitivity functions that are returned by the algorithm. The values of the constants from the sensitivity functions are summarised in Table 3.1. We have to use the values of α and β and fill in the formula:

$$\tau_{x|u} = \frac{0.4 - \beta}{\alpha}$$

If we do this for $\tau_{B=burgling}$ this gives:

$$\tau_{x|u} = \frac{0.4 - 0.540}{-0.452} = 0.3097$$

The results of doing this for all the parameters are summarised in Table 3.5. We can see that parameter $\tau_{W=broken|B=sleeping}$ *needs the smallest change to enforce the constraint.*

parameter	**current** $\tau^o_{x\|u}$	**new** $\tau_{x\|u}$	$\|\tau_{x\|u} - \tau^o_{x\|u}\|$
$\tau_{B=burgling}$	0.200	0.310	0.110
$\tau_{E=active}$	0.600	0.452	0.148
$\tau_{W=broken\|B=burgling}$	0.400	np	np
$\tau_{W=broken\|B=sleeping}$	0.900	0.796	0.104
$\tau_{A=ringing\|B=burgling \wedge E=active}$	0.100	np	np
$\tau_{A=ringing\|B=burgling \wedge E=active}$	0.400	np	np
$\tau_{A=ringing\|B=sleeping \wedge E=inactive}$	0.800	0.685	0.115
$\tau_{A=ringing\|B=sleeping \wedge E=inactive}$	0.300	0.128	0.172

Table 3.5: Parameter changes that serve to enforce the constraint $Pr(A = ringing \wedge W = broken) = 0.4$

The above method can also used to find all the parameter changes that enforces constraints of the type $Pr(a \wedge b) > \eta$ or $Pr(a \wedge b) < \eta$. We know that the sensitivity function of a parameter with respect to $Pr(a \wedge b)$ is linear, because it is of the form $\alpha \cdot \tau_{x|u} + \beta$. Because the function is linear, if the value of $alpha$ is higher than zero we know that the function is increasing and if the value is lower than zero we know that the function is decreasing. If the value of the constant α is zero, we know that the parameter cannot change the probability.

Example 6 *Suppose that now we want to change the probability* $Pr(A = ringing \wedge W = broken)$ *to be higher than 0.5. The current value of the probability is 0.450. We still want to find the smallest parameter change that enforces this constraint so we will need to calculate the parameter changes for all the parameters in the network.*

We can use the same function that we used to find the parameter changes that enforce the constraint $Pr(A = ringing \wedge W = broken) = 0.4$*, only now with the value 0.5. We use the values of the constants of the sensitivity function summarised in Table 3.1.*

$$\tau_{x|u} = \frac{0.5 - \beta}{\alpha}$$

We have summarised all these parameter changes in Table 3.6. We can see that the constraint can be enforced using the parameters $\tau_{B=burgling}$ and $\tau_{E=active}$. Because the parameter $\tau_{B=burgling}$ has to decrease to enforce the constraint $Pr(A = ringing \wedge W = broken) = 0.5$, we know that to enforce the constraint $Pr(A = ringing \wedge W = broken) > 0.5$ the parameter value has to be lower than 0.089. The parameter value of $\tau_{E=active}$ has to be higher than 0.750 to enforce the constraint.

parameter	**current** $\tau_{x\|u}$	**new** $\tau_{x\|u}$	$\|\tau_{x\|u} - \tau^o_{x\|u}\|$
$\tau_{B=burgling}$	0.200	0.089	0.111
$\tau_{E=active}$	0.600	0.750	0.150
$\tau_{W=broken\|B=burgling}$	0.400	np	np
$\tau_{W=broken\|B=sleeping}$	0.900	np	np
$\tau_{A=ringing\|B=burgling\wedge E=active}$	0.100	np	np
$\tau_{A=ringing\|B=burgling\wedge E=active}$	0.400	np	np
$\tau_{A=ringing\|B=sleeping\wedge E=inactive}$	0.800	np	np
$\tau_{A=ringing\|B=sleeping\wedge E=inactive}$	0.300	np	np

Table 3.6: Parameter changes that serve to enforce the constraint $Pr(A = ringing \wedge W = broken) = 0.5$

If we want to find the parameter change of parameter $\tau_{x|u}$ that enforces the constraint $Pr(a \wedge b|e) = \eta$ we need the values of the constants of two sensitivity functions, namely the sensitivity function that describes the relation between parameter $\tau_{x|u}$ and the output probability $Pr(a|b \wedge e)$ as well as the function describing the relation between the parameter and t the probability $Pr(b|e)$.

Output questions

For one parameter, but for all combinations of outcome probabilities conditioned on the same observations, we can get the value for which it holds that $Pr(a \wedge b) = \eta$.

At first we run AllOutOneIn once on the output probability $Pr(a|b)$. As explained in Section 3.3.1 we can use the numerator of the sensitivity function to get the constants for the sensitivity function for $Pr(a \wedge b)$. We can find the values of the numerator by using the α and β values, that are returned by the algorithm.

With one call to AllOutOneIn on the parameter $\tau_{x|u}$ and evidence e we can enforce the constraint $Pr(a \wedge b) = \eta$, for all values of a, b or η. Using the sensitivity functions of one run of the algorithm, we can enforce the constraint $Pr(a \wedge b) = \eta$, but also the constraint $Pr(c \wedge b) = \eta_2$ or even $Pr(d \wedge b) = \eta_3$, all only with respect to the parameter $\tau_{x|u}$.

Recall that the sensitivity function $Pr(a \wedge b)(\tau_{x|u})$ is of the form $\alpha\tau_{x|u} \cdot \beta$. This is a linear function. As seen before, we can also use this sensitivity function to enforce the constraints $Pr(a \wedge b) < \eta$ or $Pr(a \wedge b) > \eta$. This can be useful. If we obtain the sensitivity functions of the parameter $\tau_{x|u}$ with respect to each of the output probabilities in the network containing the same observations using the algorithm AllOutOneIn, we can enforce the constraint $Pr(a \wedge b) < \eta$, but also the constraint $Pr(c \wedge b) > \eta_2$ or even $Pr(d \wedge b) < \eta_3$, all only with respect to the parameter $\tau_{x|u}$.

Example 7 *Suppose we want to enforce the following constraints using the parameter $\tau_{E=active}$:*

- $Pr(A = ringing \wedge W = broken) > 0.3$
- $Pr(A = silent \wedge W = broken) > 0.3$
- $Pr(E = active \wedge W = broken) < 0.7$

We only need one run of the algorithm AllOutOneIn. If we use this algorithm with the parameter $\tau_{E=active}$ and the evidence $W = broken$, we obtain all the sensitivity functions needed. The constants of these sensitivity functions are summarised in Table 3.4. To get the sensitivity function for these probabilities we can use the sensitivity functions from the table. For example, to get the sensitivity function for $Pr(A = ringing \wedge W = broken)$, we use the values of the constants α and β of the sensitivity function of $Pr(A = ringing|W = broken)$.

This sensitivity function of $Pr(A = ringing \wedge W = broken)$ is:

$$Pr(A = ringing \wedge W = broken) = 0.336 \cdot \tau_{E=active} + 0.248$$

$Pr(A = ringing \wedge W = broken)$ is higher than 0.3 when $\tau_{E=active} > 0.155$.

The sensitivity function of $Pr(A = silent \wedge W = broken)$ is:

$$Pr(A = silent \wedge W = broken) = -0.336 \cdot \tau_{E=active} + 0.552$$

$Pr(A = silent \wedge W = broken)$ *is higher than 0.3 when* $\tau_{E=active} < 0.750$.

The sensitivity function of $Pr(E = active \wedge W = broken)$ *is:*

$$Pr(A = ringing \wedge W = broken) = 0.800 \cdot \tau_{E=active} + 0.000$$

$Pr(E = active \wedge W = broken)$ *is lower than 0.7 when* $\tau_{E=active} < 0.875$.

To enforce the three constraints, the value of $\tau_{E=active}$ *has to be between* 0.155 *and* 0.875.

Recall that using the sensitivity function $Pr(b|e)(\tau_{x|u})$ and the sensitivity function $Pr(a|b \wedge e)(\tau_{x|u})$ we can calculate the sensitivity function $Pr(a \wedge b|e)(\tau_{x|u})$.

We can use AllOutOneIn to get the values of the constants of the sensitivity functions $Pr(b|e)(\tau_{x|u})$ and $Pr(a|b \wedge e)(\tau_{x|u})$ for all parameters in the network. Using these sensitivity functions we can calculate the values of the constants of the sensitivity function $Pr(a \wedge b|e)(\tau_{x|u})$ for all the parameters in the network. We can use these sensitivity functions to calculate the parameter value that enforces the constraints $Pr(a \wedge b|e) = \eta$, $Pr(a \wedge b|e) < \eta$ or $Pr(a \wedge b|e) > \eta$ for every parameter in the network.

Example 8 *Suppose we want to enforce the following constraints using the parameter* $\tau_{E=active}$*:*

- $Pr(A = ringing \wedge E = active|W = broken) > 0.5$
- $Pr(B = burgling \wedge E = active|W = broken) > 0.05$

We need two runs of the algorithm AllOutOneIn. If we deploy $AllOutOneIn : (E = Active \wedge W = broken)(\tau_{E=active})$ *and* $AllOutOneIn : (W = broken)(\tau_{E=active})$ *we obtain all the sensitivity functions needed.*

The values of the constants of the sensitivity functions we obtained using $AllOutOneIn : (E = Active \wedge W = broken)(\tau_{E=active})$ *are summarised in Table 3.7. The values of the constants of the sensitivity functions obtained using* $AllOutOneIn : (W = broken)(\tau_{E=active})$ *are summarised in Table 3.4.*

To calculate the sensitivity function $Pr(A = ringing \wedge E = active|W = broken)(\tau_{E=active})$ *we use the values of the constants* α *and* γ *of the sensitivity function* $Pr(A = ringing|E = active \wedge W = broken)(\tau_{E=active})$ *and the*

output probability	α	β	γ	σ
$Pr(B = burgling \mid E = active \wedge W = broken)$	0.080	0.000	0.800	0.000
$Pr(B = sleeping \mid E = active \wedge W = broken)$ 0.720	0.000	0.800	0.000	
$Pr(A = ringing \mid E = active \wedge W = broken)$	0.584	0.000	0.800	0.000
$Pr(A = silent \mid E = active \wedge W = broken)$	0.216	0.000	0.800	0.000

Table 3.7: Values of the constants for the sensitivity functions for different outcome probabilities with respect to the parameter $\tau_{E=active}$.

values of the constants γ *and* σ *of the sensitivity function* $Pr(E = active \mid W = broken)(\tau_{E=active})$.

This gives the following sensitivity function:

$$Pr(A = ringing \wedge E = active \mid W = broken) = \frac{0.584 \cdot \tau_{E=active} + 0.00}{0.00 \cdot \tau_{E=active} + 0.800}$$

Using the method described in the previous section we calculate the parameter value that enforces the constraint $Pr(A = ringing \wedge E = active \mid W = broken) = 0.5$. *This parameter value is* 0.685. *The current value of the parameter is* 0.6 *and the current probability* $Pr(A = ringing \wedge E = active \mid W = broken)$ *is* 0.438. *This indicates that the sensitivity function is increasing. To enforce the constraint* $Pr(A = ringing \wedge E = active \mid W = broken) = 0.5$ *the parameter value has to be higher than* 0.685.

To calculate the sensitivity function $Pr(B = burgling \wedge E = active \mid W = broken)(\tau_{E=active})$ *we use the values of the constants* α *and* γ *of the sensitivity function* $Pr(B = burgling \mid E = active \mid W = broken)(\tau_{E=active})$ *and the values of the constants* γ *and* σ *of the sensitivity function* $Pr(E = active \mid W = broken)(\tau_{E=active})$.

This gives the following sensitivity function:

$$Pr(A = ringing \wedge E = active \mid W = broken) = \frac{0.08 \cdot \tau_{E=active} + 0.00}{0.00 \cdot \tau_{E=active} + 0.800}$$

To enforce the constraint $Pr(B = burgling \wedge E = active \mid W = broken) > 0.05$, *the value of* $\tau_{E=active}$ *has to be higher than 0.5.*

To enforce both constraints the value of $\tau_{E=active}$ *has to be higher than* 0.685.

3.4 Constraining the ratio of two probabilities

In this section, we are going to investigate how we can constrain the ratio of two probabilities. In this section, we assume that these probabilities are of the from $Pr(a|e)$ and we want to enforce the following constraint $Pr(a|e_1)(\tau_{x|u}) = \eta \cdot Pr(b|e_2)(\tau_{x|u})$. However, these probabilities can also be of the form $Pr(a \wedge b|e)$.

3.4.1 Calculating the new parameter values

Suppose we have the following sensitivity functions:

Sensitivity function $Pr(a|e_1)(\tau_{x|u})$:

$$Pr(a|e_1)(\tau_{x|u}) = \frac{\alpha_1 \cdot \tau_{x|u} + \beta_1}{\gamma_1 \cdot \tau_{x|u} + \sigma_1}$$

Sensitivity function $Pr(b|e_2)(\tau_{x|u})$:

$$Pr(b|e_2)(\tau_{x|u}) = \frac{\alpha_2 \cdot \tau_{x|u} + \beta_2}{\gamma_2 \cdot \tau_{x|u} + \sigma_2}$$

Then we can use these sensitivity functions to calculate the ratio between the two probabilities.

If the evidence is the same in both cases, that is $e_1 = e_2$, then $\gamma_1 = \gamma_2$ and $\sigma_1 = \sigma_2$, and therefore in order to enforce the constraint, we only have to ensure that:

$$\alpha_1 \cdot \tau_{x|u} + \beta_1 = \eta \cdot \Big(\alpha_2 \cdot \tau_{x|u} + \beta_2\Big)$$

Solving this equation gives:

$$\tau_{x|u} = \frac{\beta_1 - \beta_2 \cdot \eta}{\alpha_2 \cdot \eta - \alpha_1}$$

This indicates that if we have the sensitivity function $Pr(a|e)(\tau_{x|u})$ and the sensitivity function $Pr(b|e)(\tau_{x|u})$, we can calculate the sensitivity of the ratio η to the parameter $\tau_{x|u}$. If we want to constrain a ratio $Pr(a|e)(\tau_{x|u}) =$

$\eta \cdot Pr(b|e)(\tau_{x|u})$, we can calculate the needed change in the parameter to enforce this constraint.

If we use different observations, that is $e_1 \neq e_2$ then we find two solutions:

solution 1:

$$\tau_{x|u} = \frac{-\beta_1 \cdot \gamma_2 - \alpha_1 \cdot \sigma_2 + \eta \cdot \beta_2 \cdot \gamma_1 + \eta \cdot \alpha_2 \cdot \sigma_1 - \sqrt{\zeta}}{2(\alpha_1 \cdot \gamma_2 - \eta \cdot \alpha_2 \cdot \gamma_1)}$$

solution 2:

$$\tau_{x|u} = \frac{-\beta_1 \cdot \gamma_2 - \alpha_1 \cdot \sigma_2 + \eta \cdot \beta_2 \cdot \gamma_1 + \eta \cdot \alpha_2 \cdot \sigma_1 + \sqrt{\zeta}}{2(\alpha_1 \cdot \gamma_2 - \eta \cdot \alpha_2 \cdot \gamma_1)}$$

Where

$$\zeta = (\beta_1 \cdot \gamma_2 + \alpha_1 \cdot \sigma_2 - \eta \cdot \beta_2 \cdot \gamma_1 - \eta \cdot \alpha_2 \cdot \sigma_1)^2 - 4(\alpha_1 \cdot \gamma_2 - \eta \cdot \alpha_2 \cdot \gamma_1)(\beta_1 \cdot \sigma_2 - \eta \cdot \beta_2 \cdot \sigma_1)$$

This indicates that the two sensitivity functions are also sufficient to calculate the parameter change that enforces the constraint $Pr(a|e_1)(\tau_{x|u}) = \eta \cdot Pr(b|e_2)(\tau_{x|u})$ when the two probabilities contain different observations. Although, in this case, we have to use a different, larger, formula. Furthermore, when using different observations, we have two possible solutions. The reason that we have two possible solutions is that when we have different observations, we have different denominators in both functions. If the dominators are equal, the asymptotes of the functions are at the same position. Since the functions are monotonically non-increasing or non-decreasing this is the reason that we only have one solution. If these dominators are different, the asymptotes will be on a different position. Then we can have two solutions.

These two solutions do not have to be valid solutions. It can be the case that we have zero, one or even two valid solutions. We have a valid solution when the suggested parameter value is a value between zero and one. If there are two valid solutions, then we can enforce the constraint with two parameter values.

3.4.2 Which parameters enforce the constraint

In the next section, we are going to investigate how we can obtain the value of the constants for the required sensitivity functions.

Parameters questions

Using two runs of the algorithm OneOutAllIn, we can enforce this constraint for all the parameters. For probabilities with equal observations, we need the same sensitivity functions as for the probabilities with different observations. The difference is that, as described above, we need another function to find the parameter changes that upon changing enforce the constraint.

Example 9 *Suppose we want to tune the network in Figure 3.1. We want to enforce $Pr(A = ringing|W = broken) = Pr(E = active|W = broken)$, that is $\eta = 1$. We want to know the new parameter values that enforce the constraint. To do this, we have to run the algorithm twice. First for $Pr(A = ringing|W = broken)$, as we already have done in a previous example. The results are summarised in Table 3.1. Second for $Pr(E = active|W = broken)$. We can find the values of the constants of the sensitivity functions in Table 3.8.*

parameter	α	β	γ	σ
$\tau_{B=burgling}$	-0.300	0.540	-0.500	0.900
$\tau_{E=active}$	0.800	0.000	0.000	0.800
$\tau_{W=broken\|B=burgling}$	0.120	0.432	0.200	0.720
$\tau_{W=broken\|B=sleeping}$	0.480	0.048	0.800	0.080
$\tau_{A=ringing\|B=burgling\wedge E=active}$	0.000	0.480	0.000	0.800
$\tau_{A=ringing\|B=burgling\wedge E=active}$	0.000	0.480	0.000	0.800
$\tau_{A=ringing\|B=sleeping\wedge E=inactive}$	0.000	0.480	0.000	0.800
$\tau_{A=ringing\|B=sleeping\wedge E=inactive}$	0.000	0.480	0.000	0.800

Table 3.8: Values of the constants of the sensitivity functions for the outcome probability $Pr(E = active|W = broken)$ for each parameter.

Because we condition on the same set of observations in both probabilities, we can use the following formula:

$$\tau_{x|u} = \frac{\beta_2 - \beta_1}{\alpha_1 - \alpha_2}$$

For the parameter $\tau_{B=burgling}$ we find:

$$\tau_{B=burgling} = \frac{0.540 - 0.540}{0.240 - -0.452} = 0.000$$

So we know that if $\tau_{B=burgling}$ *= 0.000 then* $Pr(A = ringing|W = broken) = Pr(E = active|W = broken)$*.*

We have done this for all the parameters in the network. The results can be found in Table 3.9.

parameter	**current** $\tau^o_{x\|u}$	**new** $\tau_{x\|u}$	$\|\tau_{x\|u} - \tau^o_{x\|u}\|$
$\tau_{B=burgling}$	0.200	0.000	0.200
$\tau_{E=active}$	0.600	0.534	0.066
$\tau_{W=broken\|B=burgling}$	0.400	0.000	0.400
$\tau_{W=broken\|B=sleeping}$	0.900	np	np
$\tau_{A=ringing\|B=burgling\wedge E=active}$	0.100	0.750	0.650
$\tau_{A=ringing\|B=burgling\wedge E=active}$	0.400	np	np
$\tau_{A=ringing\|B=sleeping\wedge E=inactive}$	0.800	0.870	0.070
$\tau_{A=ringing\|B=sleeping\wedge E=inactive}$	0.300	0.406	0.106

Table 3.9: Parameter changes that serve to enforce the constraint $Pr(A = ringing|W = broken) = Pr(E = active|W = broken)$.

The above approach also serves to answer constraints of the type $Pr(a|e_1) < \eta \cdot Pr(b|e_2)$ or $Pr(a|e_1) > \eta \cdot Pr(b|e_2)$. First we have to enforce constraints $Pr(a|e_1) = \eta \cdot Pr(b|e_2)$ like we have done before. If we are using the same set of observations for both probabilities, then we know that the function $Pr(a|e)(\tau_{x|u}) = \eta \cdot Pr(b|e)(\tau_{x|u})$ has at most one intersection as we just have seen. Because it has one intersection we only have to know on which side of the intersection the constraint holds. If we call the solution we found $\tau^1_{x|u}$ and the current paramter value $\tau^1_{x|u}$ then we know that

if $\tau^o_{x|u} < \tau^1_{x|u}$ then the parameter has to be higher than or equal to $\tau^1_{x|u}$;

if $\tau^o_{x|u} > \tau^1_{x|u}$ then the parameter has to be lower than or equal to $\tau^1_{x|u}$.

Example 10 *Suppose we want to impose the following constraint:* $Pr(A = ringing|W = broken) > Pr(E = active|W = broken)$*. We know from the previous example that* $Pr(A = ringing|W = broken) = Pr(E = active|W = broken)$ *when* $\tau_{E=active} = 0.534$*. We know that the sensitivity functions are either monotonically non-increasing or monotonically non-decreasing. Moreover as the probabilities contain the same observations, we know that the asymptotes of these sensitivity functions are equivalent.*

We know that with the current values the constraint is not satisfied. If $\tau^o_{x|u} < \tau^1_{x|u}$ *then the parameter value has to be higher than or equal to the calculated parameter value to enforce the constraint. If* $\tau^o_{x|u} > \tau^1_{x|u}$ *then the parameter value has to be lower than or equal to the calculated parameter.*

We have done this for all the parameters in the network. The results can be found in Table 3.10.

parameter	**current** $\tau^o_{x\|u}$	**new** $\tau_{x\|u}$	$\|\tau_{x\|u} - \tau^o_{x\|u}\|$
$\tau_{B=burgling}$	0.200	np	np
$\tau_{E=active}$	0.600	< 0.534	¿0.066
$\tau_{W=broken\|B=burgling}$	0.400	np	np
$\tau_{W=broken\|B=sleeping}$	0.900	np	np
$\tau_{A=ringing\|B=burgling \wedge E=active}$	0.100	> 0.750	¿ 0.650
$\tau_{A=ringing\|B=burgling \wedge E=active}$	0.400	np	np
$\tau_{A=ringing\|B=sleeping \wedge E=inactive}$	0.800	> 0.870	¿ 0.070
$\tau_{A=ringing\|B=sleeping \wedge E=inactive}$	0.300	> 0.406	¿ 0.106

Table 3.10: Parameter values that serve to enforce the constraint $Pr(A = ringing|W = broken) > Pr(E = active|W = broken)$.

If we use different observations, then we know that the sensitivity functions, $(\tau_{x|u})(Pr(a|e_1))$ and $(\tau_{x|u}(Pr(b|e_2))$ can have different asymptotes. These asymptotes cannot be on the domain between zero and one. This implies that we can use the two solutions of the constraint $Pr(a|e_1) = \eta \cdot Pr(b|e_2)$ to enforce the constraint $Pr(a|e_1) > \eta \cdot Pr(b|e_2)$ or $Pr(a|e_1) < \eta \cdot Pr(b|e_2)$.

Let $\tau^1_{x|u}$ and $\tau^2_{x|u}$ denote the parameter values for which holds that $Pr(a|e_1) = \eta \cdot Pr(b|e_2)$, where $\tau^1_{x|u} < \tau^2_{x|u}$. Suppose that we want to enforce $Pr(a|e_1) < \eta \cdot Pr(b|e_2)$. Because there are two solutions, there are two values for which it holds that $Pr(a|e_1) = \eta \cdot Pr(b|e_2)$. This means that one of the following equations has to hold:

1: $Pr(a|e_1) < \eta \cdot Pr(b|e_2)$ if $\tau^1_{x|u} < \tau_{x|u} < \tau^2_{x|u}$
2: $Pr(a|e_1) < \eta \cdot Pr(b|e_2)$ if $\tau_{x|u} < \tau^1_{x|u}$ and if $\tau_{x|u} > \tau^2_{x|u}$.

Always one of these two situations holds. We can find out which one of these two situations holds using the current values of the parameters and the current probabilities. If we know that one of the situations does not hold, we know that the other one holds.

The constraint $Pr(a|e_1) > \eta Pr(b|e_2)$ can be handled analogously.

Output questions

Using AllOutOneIn we can solve another type of question. With only one run of algorithm AllOutOneIn we can find the constants of the values of the sensitivity functions we can use to calculate the parameter values that enforce the constraints $Pr(a|e) = \eta \cdot Pr(b|e)$ for all the probabilities in the network containing the same observations. If we want to enforce constraints on probabilities with different observations, we need one run for each combination of observations. This method is especially useful if we want to constrain the ratio between probabilities with the same observations, as in that case we only need one run.

Using these sensitivity functions we can also find the parameter values that enforce the constraints $Pr(a|e) < \eta \cdot Pr(b|e)$ or $Pr(a|e) > \eta \cdot Pr(b|e)$.

Example 11 *Suppose we want to enforce the following constraints using the parameter $\tau_{E=active}$.*

- $Pr(A = ringing|W = broken) < 2 \cdot Pr(E = active|W = broken)$
- $Pr(B = burgling|W = broken) < 2 \cdot Pr(E = active|W = broken)$

With one run of $AllOutOneIn : (W = broken)(\tau_{E=active})$ we obtain the values of the constants of the sensitivity functions summarised in Table 3.4. We can use these values to find the parameter change in parameter $\tau_{E=active}$ that enforces the constraints.

From this table we get the following sensitivity functions

$Pr(A = ringing|W = broken)(\tau_{E=active})$*:*

$$Pr(A = ringing|W = broken) = \frac{0.336 \cdot \tau_{E=active} + 0.248}{0.000 \cdot \tau_{E=active} + 0.800}$$

$Pr(E = active|W = broken)(\tau_{E=active})$*:*

$$Pr(E = active|W = broken) = \frac{0.800 \cdot \tau_{E=active} + 0.000}{0.000 \cdot \tau_{E=active} + 0.800}$$

$Pr(B = burgling|W = broken)(\tau_{E=active})$:

$$Pr(B = burgling|W = broken) = \frac{0.000 \cdot \tau_{E=active} + 0.080}{0.000 \cdot \tau_{E=active} + 0.800}$$

To find the parameter change that enforces $Pr(A = ringing|W = broken) < 2 \cdot Pr(E = active|W = broken)$, we first compute the parameter change that enforces $Pr(A = ringing|W = broken) = 2 \cdot Pr(E = active|W = broken)$:

$$\frac{0.336 \cdot \tau_{E=active} + 0.248}{0.000 \cdot \tau_{E=active} + 0.800} = 2 \cdot \frac{0.800 \cdot \tau_{E=active} + 0.000}{0.000 \cdot \tau_{E=active} + 0.800}$$

We establish that the parameter value has to be 0.1962 to solve this equation. With the current parameter value of 0.6 the constraint $Pr(A = ringing|W = broken) < 2 \cdot Pr(E = active|W = broken)$ is not enforced, so we know that the parameter has to be lower than 0.1962 to enforce the constraint.

To find the parameter change that enforces the second constraint $Pr(B = burgling|W = broken) < 2 \cdot Pr(E = active|W = broken)$, we first compute the parameter change that enforces $Pr(B = burgling|W = broken) = 2 \cdot Pr(E = active|W = broken)$.

$$\frac{0.000 \cdot \tau_{E=active} + 0.080}{0.000 \cdot \tau_{E=active} + 0.800} = 2 \cdot \frac{0.800 \cdot \tau_{E=active} + 0.000}{0.000 \cdot \tau_{E=active} + 0.800}$$

With this equation, we establish that the parameter value has to be 0.05 to enforce this equation. As the constraint $Pr(B = burgling|W = broken) < 2 \cdot Pr(E = active|W = broken)$ is not satisfied with the current values of the parameter, we know that the parameter value has to be lower than 0.05 to enforce the constraint.

Combining these result, the parameter value has to be lower than 0.05 to enforce both constraints.

3.5 Constraining the difference between two output probabilities

In this section, we investigate how we can constrain the difference between two output probabilities. In this section, we assume that these probabilities are of the form $Pr(a|e)$ and we want to enforce the following constraint $Pr(a|e_1)(\tau_{x|u}) - Pr(b|e_2)(\tau x|u) = \eta$. However, these probabilities can also be of the form $Pr(a \wedge b|e)$.

3.5.1 Calculating the new parameter values

If we want to constrain the difference $Pr(a|e_1)(\tau_{x|u}) - Pr(b|e_2)(\tau x|u)$ we need the values of the constants of the sensitivity function of $Pr(a|e_1)$ and the values of the constants of the sensitivity function of $Pr(b|e_2)$.

Suppose we have these sensitivity functions:

Sensitivity function $Pr(a|e_1)(\tau_{x|u})$:

$$Pr(a|e_1)(\tau_{x|u}) = \frac{\alpha_1 \cdot \tau_{x|u} + \beta_1}{\gamma_1 \cdot \tau x|u + \sigma_1}$$

Sensitivity function $Pr(b|e_2)(\tau_{x|u})$:

$$Pr(b|e_2)(\tau_{x|u}) = \frac{\alpha_2 \cdot \tau_{x|u} + \beta_2}{\gamma_2 \cdot \tau x|u + \sigma_2}$$

Then we want to enforce:

$$\frac{\alpha_1 \cdot \tau_{x|u} + \beta_1}{\gamma_1 \cdot \tau_{x|u} + \sigma_1} - \frac{\alpha_2 \cdot \tau_{x|u} + \beta_2}{\gamma_2 \cdot \tau_{x|u} + \sigma_2} = \eta$$

Using the same set of observations, that is $e_1 = e_2$, we have that $\gamma_1 = \gamma_2$ and $\sigma_2 = \sigma_2$, and therefore:

$$\tau_{x|u} = \frac{\beta_1 - \beta_2 - \sigma_1 \cdot \eta}{-\alpha_1 + \alpha_2 + \gamma_1 \cdot \eta}$$

If we are using a different set of observation for both outcome probabilities, it gives two solutions:

solution one:

$$\tau_{x|u} = \frac{\gamma_2 \cdot \eta \cdot \sigma_2 - \alpha_2 \cdot \sigma_1 - \gamma_1 \cdot \beta_2 - \beta_1 \cdot \gamma_2 - \alpha_1 \cdot \sigma_2 + \gamma_1 \cdot \sigma_2 \cdot \eta - \sqrt{\zeta}}{2(-\gamma_2 \cdot \eta \cdot \gamma_1 + \alpha_2 \cdot \gamma_1 + \alpha_1 \cdot \gamma_2)}$$

solution two:

$$\tau_{x|u} = \frac{\gamma_2 \cdot \eta \cdot \sigma_2 - \alpha_2 \cdot \sigma_1 - \gamma_1 \cdot \beta_2 - \beta_1 \cdot \gamma_2 - \alpha_1 \cdot \sigma_2 + \gamma_1 \cdot \sigma_2 \cdot \eta + \sqrt{\zeta}}{2(-\gamma_2 \cdot \eta \cdot \gamma_1 + \alpha_2 \cdot \gamma_1 + \alpha_1 \cdot \gamma_2)}$$

where

$$\zeta = (-\gamma_2 \cdot \eta \cdot \sigma_1 + \alpha_2 \cdot \sigma_1 + \gamma_1 \cdot \beta_2 + \beta_1 \cdot \gamma_2 + \alpha_1 \cdot \sigma_2 - \gamma_1 \cdot \sigma_2 \cdot \eta)^2$$
$$-4(-\gamma_2 \cdot \eta \cdot \gamma_1 + \alpha_2 \cdot \gamma_2 + \alpha_1 \cdot \gamma_2)(\sigma_1 \cdot \beta_2 + \beta_1 \cdot \sigma_2 - \sigma_1 \cdot \sigma_2 \cdot \eta)$$

Because in this case the dominators of the sensitivity functions are different, we again have two possible valid solutions. We can have zero, one or two parameter values that enforce the constraint. The parameter values are valid if the values are between zero and one.

3.5.2 Which parameters enforce the constraint

To calculate the parameter changes that enforce $Pr(a|e_1)(\tau_{x|u}) - Pr(b|e_2)(\tau_{x|u}) = \eta$, we need the values of the constants of the sensitivity function $Pr(a|e_1)(\tau x|u)$ and the values of the constants of the sensitivity function $Pr(b|e_1)(\tau x|u)$. In this section, we are going to investigate how we can obtain these values.

Parameter questions

To enforce the constraint $Pr(a|e_1)(\tau_{x|u}) - Pr(b|e_2)(\tau_{x|u}) = \eta$ we need two runs of the algorithm OneOutAllIn; $OneOutAllIn : Pr(a|e_1)$ to get for each parameter the sensitivity function $Pr(a|e_1)(\tau_{x|u})$ and $OneOutAllIn : Pr(b|e_2)$ to get for each parameter the sensitivity function $Pr(b|e_2)(\tau_{x|u})$. If the observations of the probabilities are equal, we need the same sensitivity functions. However, in that case, we can use the smaller formula to get the parameter value that enforces the constraint.

Example 12 *Suppose we want to enforce the following constraint:* $Pr(A = ringing|W = broken) - Pr(E = active|W = broken) = 0.1$. *We can use the sensitivity functions we have found earlier summarised in Table 3.1 and those summarised in Table 3.11 to enforce this constraint.*

The two probabilities in the constraint contain the same observations; therefore, we can use the formula:

$$\tau_{x|u} = \frac{\beta_1 - \beta_2 - \sigma_1 \cdot \eta}{-\alpha_1 + \alpha_2 + \gamma_1 \cdot \eta}$$

parameter	α	β	γ	σ
$\tau_{B=burgling}$	-0.300	0.540	-0.500	0.900
$\tau_{E=active}$	0.800	0.000	0.000	0.800
$\tau_{W=broken\|B=burgling}$	0.120	0.432	0.200	0.720
$\tau_{W=broken\|B=sleeping}$	0.480	0.048	0.800	0.080
$\tau_{A=ringing\|B=burgling \wedge E=active}$	0.000	0.480	0.000	0.800
$\tau_{A=ringing\|B=burgling \wedge E=active}$	0.000	0.480	0.000	0.800
$\tau_{A=ringing\|B=sleeping \wedge E=inactive}$	0.000	0.480	0.000	0.800
$\tau_{A=ringing\|B=sleeping \wedge E=inactive}$	0.000	0.480	0.000	0.800

Table 3.11: Values of the constants of the sensitivity function for the outcome probability $Pr(E = active|W = broken)$ for each parameter.

Filling in the constants of the sensitivity functions for $\tau_{E=active}$ *this gives:*

$$\tau_{E=active} = \frac{0.248 - 0.000 - 0.800 \cdot 0.100}{-0.336 + 0.800 + 0.000 \cdot 0.100} = 0.362.$$

So when $\tau_{E=active} = 0.362$ *then* $Pr(A = ringing|W = broken) - Pr(E = active|W = broken) = 0.1$ *We have done this for all parameters, the results are summarised in Table 3.12.*

parameter	**current** $\tau^o_{x\|u}$	**new** $\tau_{x\|u}$	$\|\tau_{x\|u} - \tau^o_{x\|u}\|$
$\tau_{B=burgling}$	0.200	np	np
$\tau_{E=active}$	0.600	0.362	0.238
$\tau_{W=broken\|B=burgling}$	0.400	np	np
$\tau_{W=broken\|B=sleeping}$	0.900	np	np
$\tau_{A=ringing\|B=burgling \wedge E=active}$	0.100	np	np
$\tau_{A=ringing\|B=burgling \wedge E=active}$	0.400	np	np
$\tau_{A=ringing\|B=sleeping \wedge E=inactive}$	0.800	np	np
$\tau_{A=ringing\|B=sleeping \wedge E=inactive}$	0.300	np	np

Table 3.12: Parameter values serve to enforce the constraint $Pr(A = ringing|W = broken) - Pr(E = active|W = broken) = 0.1$.

The above approach also serves to answer constraints of the type $Pr(a|e_1) - Pr(b|e_2) < \eta$ or $Pr(a|e_1) - Pr(b|e_2) > \eta$. First we have to enforce constraints $Pr(a|e_1) - Pr(b|e_2) = \eta$. If we are using the same set of observations for both probabilities, then we know that the function $Pr(a|e)(\tau_{x|u}) - Pr(b|e)(\tau_{x|u}) =$

η has only one intersection as we just have seen. Because it has one intersection we only have to know on which side of the intersection the constraint holds. If we call the solution we found $\tau^1_{x|u}$ and the current paramter value $\tau^1_{x|u}$ then we know that:

If $\tau^o_{x|u} < \tau^1_{x|u}$ then the parameter has to be higher than or equal to $\tau^1_{x|u}$.

If $\tau^o_{x|u} > \tau^1_{x|u}$ then the parameter has to be lower than or equal to $\tau^1_{x|u}$.

If we use different observations then we know that the sensitivity functions $Pr(a|e_1)(\tau_{x|u})$ and $Pr(b|e_2)(\tau_{x|u})$, cannot have any asymptotes on the domain between zero and one. This implies that we can use the two solutions of the constraint $Pr(a|e_1) - Pr(b|e_2) = \eta$ to enforce the constraint $Pr(a|e_1) - Pr(b|e_2) < \eta$ or $Pr(a|e_1) - Pr(b|e_2) > \eta$.

Let $\tau^1_{x|u}$ and $\tau^2_{x|u}$ denote the parameter values for which holds that $Pr(a|e_1) - Pr(b|e_2) = \eta$, where $\tau^1_{x|u} < \tau^2_{x|u}$. Suppose that we want to enforce $Pr(a|e_1) - Pr(b|e_2) < \eta$.

Because there are two solutions, there are two values for which holds that $Pr(a|e_1) - Pr(b|e_2) = \eta$. This means that one of the following equations has to hold

1: $Pr(a|e_1) - Pr(b|e_2) < \eta$ if $\tau^1_{x|u} < \tau_{x|u} < \tau^2_{x|u}$
2: $Pr(a|e_1) - Pr(b|e_2) < \eta$ if $\tau_{x|u} < \tau^1_{x|u}$ and if $\tau_{x|u} > \tau^2_{x|u}$.

Always one of these two situations holds when changing the parameter values of one parameter. We can find out which one of these two situations holds using the current values of the parameters and the probabilities. If we know that one of the situations holds in the current situation, we know it always holds when changing the parameter.

The constraint $Pr(a|e_1) - Pr(b|e_2) > \eta$ can be handled analogously.

Example 13 *Suppose we want to know when $Pr(A = ringing|W = broken) - Pr(E = active|W = broken) < 0.1$. We can get the values of the sensitivity function as we have done before. We know that with the current parameter values the constraint is not enforced. We have found in the previous example that the constraint $Pr(A = ringing|W = broken) - Pr(E = active|W = broken) = 0.1$ is enforced if we change $\tau_{E=active}$ to 0.362. This means that we have to change the value of the parameter $\tau_{E=active}$ to lower than or equal to 0.362 to enforce the constraint $Pr(A = ringing|W = broken) - Pr(E = active|W = broken) < 0.1$.*

Output questions

With only one run of algorithm AllOutOneIn we can find the constants of the values of the sensitivity functions we can use to calculate the parameter values that enforce the constraints $Pr(a|e) - Pr(b|e) = \eta$ for all the probabilities in the network containing the same observations. If we want to enforce constraints on probabilities with different observations, we need one run for each combination of observations. This method is especially useful if we want to constrain the difference between probabilities with the same observations, as in that case we only need one run.

Using these sensitivity functions we can also find the parameter values that enforce the constraints $Pr(a|e) - Pr(b|e) < \eta$ or $Pr(a|e) - Pr(b|e) > \eta$

Example 14 *Suppose we want to enforce the following constraints using the parameter* $\tau_{E=active}$*:*

- $Pr(A = ringing|W = broken) - Pr(E = active|W = broken) > 0.2$
- $Pr(A = ringing|W = broken) - Pr(B = burgling|W = broken) < 0.4$

With one iteration of $AllOutOneIn : (W = broken)(\tau_{E=active})$ *we obtain the values of the constants of the sensitivity functions summarised in Table 3.4. We can use these values to find the parameter change in parameter* $\tau_{E=active}$ *that enforces the constraints.*

Recall that from this table we can obtain the following sensitivity functions

$Pr(A = ringing|W = broken)(\tau_{E=active})$*:*

$$Pr(A = ringing|W = broken) = \frac{0.336 \cdot \tau_{E=active} + 0.248}{0.000 \cdot \tau_{E=active} + 0.800}$$

$Pr(E = active|W = broken)(\tau_{E=active})$*:*

$$Pr(E = active|W = broken) = \frac{0.800 \cdot \tau_{E=active} + 0.000}{0.000 \cdot \tau_{E=active} + 0.800}$$

$Pr(E = active|W = broken)(\tau_{E=active})$*:*

$$Pr(B = burgling|W = broken) = \frac{0.000 \cdot \tau_{E=active} + 0.080}{0.000 \cdot \tau_{E=active} + 0.800}$$

To find the parameter change that enforces $Pr(A = ringing|W = broken) - Pr(E = active|W = broken) > 0.2$*, we first compute the parameter change that enforces* $Pr(A = ringing|W = broken) - Pr(E = active|W = broken) = 0.2$*:*

$$\frac{0.336 \cdot \tau_{E=active} + 0.248}{0.000 \cdot \tau_{E=active} + 0.800} - \frac{0.800 \cdot \tau_{E=active} + 0.000}{0.000 \cdot \tau_{E=active} + 0.800} = 0.2$$

We compute that the parameter value has to be 0.190 to solve this equation. With the current parameter value of 0.6 the constraint $Pr(A = ringing|W = broken) - Pr(E = active|W = broken) > 0.2$ *is not enforced, so we know that the parameter has to be lower than 0.190 to enforce the constraint.*

To find the parameter change that enforces the second constraint $Pr(A = ringing|W = broken) - Pr(B = burgling|W = broken) < 0.4$*, we first compute the parameter change that enforces* $Pr(A = ringing|W = broken) - Pr(B = burgling|W = broken) = 0.4$*:*

$$\frac{0.336 \cdot \tau_{E=active} + 0.248}{0.000 \cdot \tau_{E=active} + 0.800} - \frac{0.000 \cdot \tau_{E=active} + 0.080}{0.000 \cdot \tau_{E=active} + 0.800} = 0.4$$

With this equation, we compute that the parameter value has to be 0.214 to enforce this equation. As the constraint $Pr(A = ringing|W = broken) - Pr(B = burgling|W = broken) < 0.4$ *is not satisfied with the current values of the parameter, we know that the parameter value has to be lower than 0.214 to enforce the constraint.*

Combining these results, the parameter value has to be lower than 0.190 to enforce both constraints.

3.6 Combining different types of constraints

In the previous sections, we have showed that we can enforce multiple constraints of the same type simultaneously. However, we can also combine the solutions of multiple different constraints. This is only useful for constraints that contain inequalities. If we, for example, want to enforce two constraints and we know that a parameter value has to be higher than 0.6 to enforce the first constraint and the parameter value has to be lower than 0.8 to enforce the second constraint than we know that the value has two be between 0.6 and 0.8 to enforce both constraints.

Example 15 *Suppose we want to know for which value of the parameter $\tau_{B=burgling}$ the following constraints are enforced:*

- $Pr(A = ringing|W = broken) > 0.55$
- $Pr(A = ringing|W = broken \wedge E = active) > 0.21$
- $Pr(E = active \wedge A = ringing) < 0.2$

Using the methods of the previous section, we can find out that:

- $Pr(A = ringing|W = broken) > 0.55$ *if* $\tau_{B=burgling} > 0.254$
- $Pr(A = ringing|W = broken \wedge E = active) > 0.21$ *if* $\tau_{B=burgling} < 0.931$
- $Pr(E = active \wedge A = ringing) < 0.2$ *if* $\tau_{B=burgling} < 0.666$

Combining these results, we know that the parameter $\tau_{B=burgling}$ has to be between 0.254 and 0.666 to enforce all constraints.

3.7 Related work

Chan and Darwiche [1] also proposed an approach for determining the parameter changes that can enforce certain constraints for the purpose of tuning. In this section, we will investigate the difference between our approach and theirs. First, we will examine how to enforce constraints using the approach by Chan and Darwiche. Then we will consider the constraints that can be enforced using this approach. Finally, we will compare both approaches. We will limit the discussion to the situation where only a single parameter is changed at time.

3.7.1 The d-sensitivity function

Whereas we calculate all constants of the sensitivity function which relates a probability $Pr(a|e)$ to a parameter $\tau_{x|u}$, Chan and Darwiche use a different interpretation of the sensitivity function which requires calculating only two constants. Their function does not relate an output probability to a parameter, but directly to a change in the parameter value. We will call this function the d-sensitivity function.

In formulating the d-sensitivity function for an output probability $Pr(a|e)$, we require the current original probabilities $Pr(a \wedge e)^o$ and $Pr(e)^o$, and the change in probabilities $\Delta Pr(a \wedge e)$ and $\Delta Pr(e)$, resulting from changing the parameter under consideration:

$$Pr(a|e) = \frac{Pr(a \wedge e)}{Pr(e)} = \frac{Pr(a \wedge e)^o + \Delta Pr(a \wedge e)}{Pr(e)^o + \Delta Pr(e)}$$

The difference in the probabilities, $\Delta Pr(a \wedge e)$ and $\Delta Pr(e)$ can be calculated using the first derivative of the linear functions for $Pr(a \wedge e)$ and $Pr(e)$ as a function of $\tau_{x|u}$. These first derivatives correspond to the values of our constants α and γ:

$$\Delta Pr(a \wedge e) = \alpha \cdot \Delta\tau_{x|u}$$

$$\Delta Pr(e) = \gamma \cdot \Delta\tau_{x|u}$$

This gives the following d-sensitivity function:

$$Pr(a|e) = \frac{Pr(a \wedge e)^o + \alpha \cdot \Delta\tau_{x|u}}{Pr(e)^o + \gamma \cdot \Delta\tau_{x|u}}$$

The difference between this function and the sensitivity function is that, instead of the constants β and σ, which depend on the actual parameter under consideration, we now need the current original probabilities $Pr(a \wedge e)^o$ and $Pr(e)^o$, which can be calculated prior to parameter variation and do not depend on the actual parameter under consideration. Using the d-sensitivity function, we thus do not need to compute the value of all constants for all parameters.

3.7.2 Enforcing constraints

Chan and Darwiche [1] employ the d-sensitivity function to determine the parameter changes required to enforce the following types of constraint:

- $Pr(a|e) - Pr(b|e) = \eta$
- $Pr(a|e) = \eta \cdot Pr(b|e)$

Although not discussed in the mentioned paper, their approach serves to enforce all the other constraints as well, even with probabilities with different observations. We can enforce these constraints by replacing the sensitivity function with the d-sensitivity function. For example, the formula used to describe parameter changes that could enforce the constraint $Pr(a|e) = \eta$ then becomes:

$$\frac{Pr(a \wedge e)^o + \alpha \cdot \Delta\tau_{x|u}}{Pr(e)^o + \gamma \cdot \Delta\tau_{x|u}} = \eta$$

This equation can be solved after obtaining the values of the constants. The difference between the d-sensitivity function and the sensitivity function is that using the d-sensitivity function we have to calculate the change in the parameter $\tau_{x|u}$ required to enforce the constraint, instead of just its new value.

3.7.3 Sensitivity function vs. d-sensitivity function

Using the d-sensitivity function, we need to obtain the values of fewer constants than using the sensitivity function. This is a benefit only if the algorithms that we use to compute the constants can also exploit the fact that fewer constants are required. In this section, we compare the d-sensitivity function with the sensitivity function in terms of the amount of network propagations needed to calculate the values of the constants needed.

Recall that both algorithms OneOutAllIn and AllOutOneIn return all the constants of the sensitivity function and therefore, also the two constants required by the d-sensitivity function. When using these algorithms therefore, there will not be a difference whether you use the sensitivity function or the d-sensitivity function. Both techniques will need an equal amount of network propagations to get the required constants.

If we change the algorithms in such a way that the constants β and σ are not computed, that is, the algorithms have to solve less equations, then using the d-sensitivity function could be slightly faster. Chan and Darwiche, however, propose to use a different algorithm for computing the constants required for the d-sensitivity function. They propose using the *differential approach*[4] which allows for computing the value of a single constant in the sensitivity functions for all parameters in the network. Using the sensitivity function requires running the differential approach four times, whereas the d-sensitivity function requires only two runs.

Recall that the algorithms OneOutAllIn and AllOutOneIn each basically require 1.5 network propagations. The differential approach require one network propagation [4]. It may be apparent that using the differential approach is only sensible in combination with the d-sensitivity function. In addition, the differential approach is only useful for answering parameter questions and not for output questions.

Table 3.13 lists the amount of network propagations required to answer different parameter questions using the sensitivity function and the d-sensitivity functions, and the related algorithms. We will give some examples as illustration.

Constraint	**OneOutAllIn**	**Difference Approach**
$Pr(a\|e) = \eta$	$1 \cdot 1.5$ propagations	$2 \cdot 1$ propagations
$Pr(a \wedge b) = \eta$	$1 \cdot 1.5$ propagations	$1 \cdot 1$ propagations
$Pr(a \wedge b\|e) = \eta$	$2 \cdot 1.5$ propagations	$3 \cdot 1$ propagations
$Pr(a\|e) = \eta \cdot Pr(b\|e)$	$2 \cdot 1.5$ propagations	$2 \cdot 1$ propagations
$Pr(a\|e_1) = \eta \cdot Pr(b\|e_2)$	$2 \cdot 1.5$ propagations	$4 \cdot 1$ propagations
$Pr(a\|e) - Pr(b\|e) = \eta$	$2 \cdot 1.5$ propagations	$3 \cdot 1$ propagations
$Pr(a\|e_1) - Pr(b\|e_2) = \eta$	$2 \cdot 1.5$ propagations	$4 \cdot 1$ propagations

Table 3.13: Amount of network propagations required for establishing the constants of the sensitivity function with OneOutAllIn and of the d-sensitivity function with the differential apporach in order to identify the single parameter changes that serve to enforce different constraints.

When we use the differential approach, we should also use the d-sensitivity function; otherwise, we will need more propagation, to calculate the values of the constants. For this reason, we will use the d-sensitivity function when we are using the differential approach. When using the OneOutAllIn algorithm, the amount of propagations is independent of whether we use the sensitivity function or the d-sensitivity function we will use the sensitivity function in the remainder of this section.

The amount of network propagations we need is dependant on the constraint that we want to enforce, because the amount of constants needed depends on the constraint. We first consider the constraint $Pr(a|e_1) - Pr(a|e_2) = \eta$. Using the sensitivity function and OneOutAllIn, we will need two runs as explained in the previous section. We will need one run for the values of the four constants relating the probability $Pr(a|e_1)$ to the parameters and one run for the four constants relating the parameters to the probability $Pr(b|e_2)$. Recall that one run of OneOutAllIn needs one and a half propagation, so two iterations will need three propagations. One the other hand, using the differential approach, we will need the constants of α and γ for the probability $Pr(a|e_1)$ and the values of the constants of α and γ for the probability $Pr(b|e_2)$. We will need to run the algorithm for every constant. The algorithm needs one propagation, so we will need four propagations. When we want to enforce this constraint, using OneOutAllIn we will need less propagations.

The algorithm OneOutAllIn does not always need a smaller amount of propagations. When we want to enforce the constraint $Pr(a|e) = \eta \cdot Pr(b|e)$, using OneOutAllIn, we will need again two runs, although we know beforehand that the values of the constants of γ and σ are equal, because the probabilities contain the same observations. Algorithm OneOutAllIn needs again three propagations. The d-sensitivity function needs only two constants. We can use the fact that the values of the constant γ are the same for both probabilities. Because we know that the values are the same, we do not need the values at all.

Recall that we have the following d-sensitivity functions.

$$Pr(a|e) = \frac{Pr(a \wedge e)^o + \alpha_1 \cdot \Delta\tau_{x|y}}{Pr(e)^o + \gamma_1 \cdot \Delta\tau_{x|y}}$$

$$Pr(b|e) = \frac{Pr(b \wedge e)^o + \alpha_1 \cdot \Delta\tau_{x|y}}{Pr(e)^o + \gamma_1 \cdot \Delta\tau_{x|y}}$$

Solving the equation $Pr(a|e) = \eta \cdot Pr(b|e)$ gives:

$$\tau_{x|u} = \frac{Pr(a \wedge e)^o - Pr(b \wedge e)^o \cdot +\eta}{\alpha_2 \cdot \eta - \alpha_1}$$

To get the value of $\tau_{x|u}$ that enforces the constraint, we will only need the values of the constants α_1 and α_2. We will have to run the differential approach twice and this needs two propagations.

In their article [1], Chan and Darwiche proposed enforcing the constraint $Pr(a|e) = \eta$, by using the constraint $Pr(a|e) - Pr(b|e) = \eta_2$ with $Pr(b|e) = Pr(\neg a|e)$ and $\eta_2 = \eta - (1 - \eta)$. Using this technique, we need to use the differential approach three times. One run for the value of the constant α for the probability $Pr(a|e)$, one run for the value of the constant α for the probability $Pr(b|e)$ and one run for the value of the constant for $Pr(e)$, which is the value of the constant γ for both the probabilities $Pr(a|e)$ and $Pr(b|e)$; the constant γ is equal for both probabilities.

If we, instead, use the d-sensitivity function of $Pr(a|e)$ to calculate the parameter change needed to enforce the constraint $Pr(a|e) = \eta$ we would need only two runs of the differential approach. This way we only need the values of the constants α and γ. Instead of three propagations of the network, we now only need two propagations.

3.8 Conclusion

In this chapter, we have investigated the constraints that could be interesting to enforce when we are tuning a network. Then we have investigated whether we are able to enforce these constraints.

In this chapter, we have seen that we can find for each parameter in the network, the parameter change that enforces one of the constraints. We can use this method to compare different parameters. We have also seen that we can find the parameter change that enforces multiple constraints on different probabilities simultaneously. However, we cannot constrain many parameters containing different observations.

In this chapter, we have enforced the constraints using a single parameter change or multiple parameter changes in one CPT. When we would like to enforce the constraints using multiple parameter changes from different CPTs calculating the needed change in the parameters becomes computationally too complex. N-way sensitivity analyses as described in [5] does not only need more propagations, there are also many combinations of multiple parameters that can be changed and for every combination there can be multiple valid solutions.

We have also compared the algorithms that can be used for the parameter questions. These algorithms seemed to have comparable running times.

When we are asking single parameter questions or multiple parameter questions concerning one CPT, we most of the times will end up with a list of potential parameter changes. In the next chapter, we are going to look at those possible parameter changes. Although all the parameter changes enforce the constraint, some of the parameter changes turn out to be better than other parameter changes.

Chapter 4

How to select the optimal parameter change

The methods described in the previous chapter allow us to compute the (possibly empty) set of parameters that can be changed to enforce a given constraint. If we consider changing only one parameter from this set, then the methods described in the previous section also prescribe how to change a selected parameter. If we consider changing multiple parameters from this set then establishing the exact changes required and their joint effect quickly becomes unfeasible.

In this chapter, we discuss and compare a number of selection criteria that can be used to choose between the parameters in the set. First, we will investigate these selection criteria when we consider changing only one parameter in this set. Second, we will observe how to choose between the solutions when we consider changing multiple parameters in a single conditional probability table.

4.1 Selecting the optimal single parameter change

When using the methods described in Chapter 3 for single parameter changes, the obtained set of parameters will include the needed parameter change to enforce the constraint. From this set, we want to select the parameter change that would improve the network performance most. Although all the parameters in the set enforce the constraint, they can also impose (unwanted) changes in other output probabilities.

Changing parameters to enforce a constraint without knowing the influence of the parameters on the other output probabilities will probably disturb a great part of the network. In this chapter we assume that, besides the output for which we want to enforce a constraint, we want to maintain the network's output as much as possible. Unfortunately, we cannot calculate for each parameter change the exact difference in all the output probabilities. This would be computationally too demanding, because there are too many output probabilities.

In this section, we discuss how we can use selection criteria to choose the parameter that disturbs our network the least. We will study the performance of the selection criteria in an experiment.

4.1.1 Selection criteria for single parameter change

In this section, we discuss the selection criteria that can be used to determine the parameter change that would disturb our network the least. These selection criteria try to predict the additional effects of the parameter changes based on the amount of change in the parameter and on the current value of the parameter. We will study these selection criteria and explain the underlying reasoning.

The smallest absolute change

From the fact that the sensitivity function is monotonically increasing or monotonically decreasing, we know that a small absolute change in a parameter in the network will lead to a smaller absolute change in the output probabilities than a larger change in the parameter. So in general, we would like to apply a small change to one of the parameters to enforce the constraint. To find such a small change, we could use the absolute change, referred to as $\Delta A(\tau_{x|u})$, as a selection criterion.

To use this selection criterion, we should obtain all possible parameter changes that enforce the constraint. From these parameter changes, we can then select the parameter that requires the smallest absolute change, and use this parameter to enforce the constraint.

The smallest relative change

When we are using the absolute change as a selection criteria, a change from 0.00 to 0.09 would be selected over a change from 0.50 to 0.60. However, the former would impose something that is assumed impossible to be possible. Because of this we could consider parameter changes of a parameter with an extreme value to be more vital. The output of the network is found to be more sensitive to parameters with values closer to zero or one [1].

We can use the relative change $\Delta R(\tau_{x|u})$ to penalise changes in parameters having extreme values. We calculate the change in the parameter relative to the closest distance to zero or one. We use the distance to zero if the current value, $\tau^o_{x|u}$, of the parameter is lower than, or equal to 0.5, and the distance to one if the current value of the parameter is higher than 0.5.

$$\Delta R(\tau_{x|u}) = \frac{\Delta A(\tau_{x|u})}{\min\{|0 - \tau^o_{x|u}|\ ,\ |1 - \tau^o_{x|u}|\}}$$

This selection criterion would select a parameter change of a parameter with a value close to 0.5 over a parameter change of a parameter close to zero or one, but it would also prefer a small change over a large change.

The smallest log-odds change

Chan and Darwiche [1] proposed a measure to bound the change in any output probability $Pr(a|e)$ when changing a parameter $\tau_{x|u}$. This bound can also be used to indicate the maximum change in any of the output probabilities in the network. This bound can be used as a selection criterion. If we can select a parameter for which this bound is low, we know that the maximum change in any of the output probabilities has to be small.

The bound is formulated in terms of log-odds. Let $O(\tau_{x|u})$ denote the odds of the parameter $\tau_{x|u}$: $O(\tau_{x|u}) = \tau_{x|u}/(1 - \tau_{x|u})$ and let $O(a|e)$ denote the odds of a given e: $O(a|e) = Pr(a|e)/(1 - Pr(a|e))$. Let $O(\tau_{x|u})^o$ and $O(a|e)^o$ denote the current values of these odds and let $O(\tau_{x|u})^1$ and $O(a|e)^1$ denote these odds after having applied an arbitrary change to parameter $\tau_{x|u}$. Let $\Delta O(a|e)$ denote the log-odds change in the probability of interest: $\Delta O(a|e) = |O(a|e)^o - O(a|e)^1|$ and let $\Delta O(\tau_{x|u})$ denote the log-odds change in the parameter of interest: $\Delta O(\tau_{x|u}) = |O(\tau_{x|u})^o - O(\tau_{x|u})^1|$. Then the log-odds change in the probability of interest, $\Delta O(a|e)$, is bounded by the log-odds change in parameter $\Delta O(\tau_{x|u})$:

$$\Delta O(a|e) \leq \Delta O(\tau_{x|u})$$

This bound implies that the log-odds change in any output probability is smaller or equal to the log-odds change in the parameter.

The given bound is a tight upper bound:

Theorem 2 *Let a,e,x, and u be as before, then for each parameter $\tau_{x|u}$:*

$$\exists a, e \; \Delta O(a|e) = \Delta O(\tau_{x|u})$$

This result immediately follows from the observation that a parameter $\tau_{x|u}$ corresponds to the output probability $Pr(x|u)$.

Chan and Darwiche [1], proposed the following selection method: when selecting a single parameter to change, choose the parameter $\tau_{x|u}$ which needs the smallest log-odds change to enforce the wanted constraint. This minimizes the maximum log-odds change in the output probabilities.

4.1.2 Experiment

We will illustrate the performance of the three selection criteria by means of an experiment. In this section, we will first explain about the set-up of the experiment. Then we will discuss and explain the results of the experiment for each selection criterion.

Calculating the changes in the output

For this experiment we have used the network in Figure 4.1. This network is a modified version of the network described in [6]. We have adjusted the parameter values of the network. For the experiment, we have used randomly created parameter values. These values are as follows:

$\gamma_{V_1}(v_1) = 0.78287$

$\gamma_{V_2}(v_2|v_1) = 0.25229$
$\gamma_{V_2}(v_2|\neg v_1) = 0.76101$

$\gamma_{V_3}(v_3|v_2 \wedge v_5) = 0.43430$
$\gamma_{V_3}(v_3|v_2 \wedge \neg v_5) = 0.13837$
$\gamma_{V_3}(v_3|\neg v_2 \wedge v_5) = 0.53719$
$\gamma_{V_3}(v_3|\neg v_2 \wedge \neg v_5) = 0.67183$

$\gamma_{V_4}(v_4|v_3) = 0.85489$
$\gamma_{V_4}(v_4|\neg v_3) = 0.67861$

$\gamma_{V_5}(v_5|v_6) = 0.39859$
$\gamma_{V_5}(v_5|\neg v_6) = 0.24058$

$\gamma_{V_6}(v_6) = 0.58836$

$\gamma_{V_7}(v_7|v_6) = 0.30294$
$\gamma_{V_7}(v_7|\neg v_6) = 0.58181$

$\gamma_{V_8}(v_8|v_3 \wedge v_7) = 0.54046$
$\gamma_{V_8}(v_8|v_3 \wedge \neg v_7) = 0.06450$
$\gamma_{V_8}(v_8|\neg v_3 \wedge v_7) = 0.37461$
$\gamma_{V_8}(v_8|\neg v_3 \wedge \neg v_7) = 0.43342$

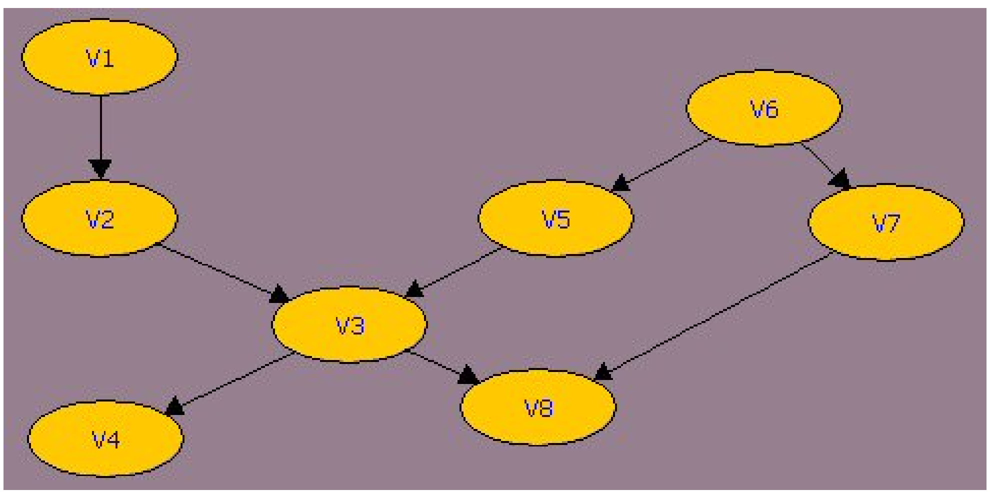

Figure 4.1: An example network

Using the methods of Chapter 3, we calculated the single parameter changes that can be used to increase the output $Pr(v_3|v_1 \wedge v_8)$ by 0.05, that is, from 0.4195 to 0.4695.

We have summarized the resulting parameter changes in Table 4.1. Next to these parameter changes are the values that are calculated by the different selection criteria. This table is sorted on the log-odds difference. Although these selection criteria select the same parameter as being the optimal parameter, there are some differences. For example, the log-odds difference of

$\tau_{v_2|v_1}$ is higher than the log-odds difference of $\tau_{v_7|v_6}$. However, the absolute and relative change in $\tau_{v_2|v_1}$ is higher. We will investigate which selection criterion has the best prediction of the disturbance.

Parameter	$\tau^o_{x\|u}$	$\tau^1_{x\|u}$	$\Delta A(\tau_{x\|u})$	$\Delta R(\tau_{x\|u})$	$\Delta O(\tau_{x\|u})$
$\tau_{v_3\|\neg v_2 \wedge \neg v_5}$	0.6718	0.7712	0.0994	0.1480	0.4987
$\tau_{v_8\|\neg v_3 \wedge \neg v_7}$	0.4334	0.3047	0.1287	0.2970	0.5570
$\tau_{v_8\|v_3 \wedge v_7}$	0.5405	0.6820	0.1415	0.3079	0.6010
$\tau_{v_7\|v_6}$	0.3029	0.4961	0.1932	0.6378	0.8178
$\tau_{v_2\|v_1}$	0.2523	0.1249	0.1274	0.5050	0.8605
$\tau_{v_8\|\neg v_3 \wedge v_7}$	0.3746	0.1948	0.1798	0.4700	0.9067
$\tau_{v_3\|\neg v_2 \wedge v_5}$	0.5372	0.7424	0.2052	0.4434	0.9094
$\tau_{v_8\|v_3 \wedge \neg v_7}$	0.0645	0.1663	0.1018	0.6121	1.0623

Table 4.1: The parameter changes that enforce the output $Pr(v_3|v_1 \wedge v_8)$ to increase by 0.05. Next to this are the values calculated by the selection criteria; $\Delta A(\tau_{x|u})$ is the absolute change, $\Delta R(\tau_{x|u})$ is the relative change, and $\Delta O(\tau_{x|u})$ is the log-odds change in the parameter.

Recall that although all these parameter changes enforce the constraint, we want to maintain as much from the output probabilities as possible. There are several ways to measure the disturbance of the network. In this section, we are going to compare the selection criteria by using three measures for the disturbance. We use three measures to give a good idea of the disturbance, and thus about the performance of the selection criteria.

First we have calculated the average change in all the output probabilities. To measure this disturbance we have calculated all the current output probabilities. Then we have calculated the difference in the output probabilities after changing one of the parameters. For this network, we already have about $35,000$ different output probabilities. Calculating the difference in all the probabilities for this small network already was time consuming. Calculating the differences in a moderate size network will be impracticable as the number of output probabilities increases exponentially. We will abbreviate this average absolute change in the output probabilities as av. $\Delta A(a|e)$.

The second measure we have calculated is the average relative change in the output probabilities. To calculate this relative difference we have used the method from the previous section, this way changes in the probabilities that are near zero and one had a higher relative value than changes in the probabilities near 0.5. We will abbreviate this average relative change in the output probabilities as av. $\Delta R(a|e)$.

The third measure we have calculated is the average log-odds change. For each parameter change, we have calculated the average difference in the log-odds for all the output probabilities. We will abbreviate this average log-odds change in the output probabilities as av. $\Delta O(a|e)$.

Using three measures, we have a better idea of the disturbance. We will inspect whether these measures have different values depending on the selection criteria used, as we expect that for example the selection criterion using the absolute parameter change might have a low absolute disturbance, but a high relative disturbance.

The values of these measures after changing each parameter are summarized in Table 4.2. These values indicate the disturbance of the network. For example, we can see that we would like to select the parameter $\tau_{v_8|v_3 \wedge v_7}$ or the parameter $\tau_{v_3|\neg v_2 \wedge \neg v_5}$, we indicate this with a $+$. We do not want to select the parameter $\tau_{v_7|v_6}$ or the parameter $\tau_{v_2|v_1}$. This is indicated with a $-$. In the next paragraphs, we are going to investigate whether the selection criteria give a good prediction of this disturbance.

Parameter	**av.** $\Delta A(a\|e)$	**av.** $\Delta R(a\|e)$	**av.** $\Delta O(a\|e)$
$\tau_{v_8\|v_3 \wedge v_7}$ $+$	0.0092	0.0290	0.0449
$\tau_{v_3\|\neg v_2 \wedge \neg v_5}$ $+$	0.0094	0.0308	0.0466
$\tau_{v_8\|\neg v_3 \wedge \neg v_7}$	0.0103	0.0328	0.0500
$\tau_{v_8\|v_3 \wedge \neg v_7}$	0.0100	0.0708	0.0640
$\tau_{v_8\|\neg v_3 \wedge v_7}$	0.0146	0.0417	0.0690
$\tau_{v_3\|\neg v_2 \wedge v_5}$	0.0158	0.0506	0.0763
$\tau_{v_2\|v_1}$ $-$	0.0189	0.0613	0.1016
$\tau_{v_7\|v_6}$ $-$	0.0229	0.0730	0.1031

Table 4.2: The three measures of the disturbance caused by the parameter changes that enforce $Pr(v_3|v_1 \wedge v_8)$ to increase with 0.05; av. $\Delta A(a|e)$, summarises the average absolute change, av. $\Delta R(a|e)$, summarises the average relative change and av. $\Delta O(a|e)$ the average log-odds change in the output probabilities.

Parameter	$\Delta A(\tau_{x\|u})$	**av.** $\Delta A(a\|e)$	**av.** $\Delta R(a\|e)$	**av.** $\Delta O(a\|e)$
$\tau_{v_3\|\neg v_2 \wedge \neg v_5}$ +	0.0994	0.0094	0.0308	0.0466
$\tau_{v_8\|v_3 \wedge \neg v_7}$	0.1018	0.0100	0.0708	0.0640
$\tau_{v_2\|v_1}$ −	0.1274	0.0189	0.0613	0.1016
$\tau_{v_8\|\neg v_3 \wedge \neg v_7}$	0.1287	0.0103	0.0328	0.0497
$\tau_{v_8\|v_3 \wedge v_7}$ +	0.1415	0.0092	0.0290	0.0449
$\tau_{v_7\|v_6}$ −	0.1932	0.0229	0.0730	0.1031
$\tau_{v_8\|\neg v_3 \wedge v_7}$	0.1948	0.0147	0.0417	0.0690
$\tau_{v_3\|\neg v_2 \wedge v_5}$	0.2052	0.0158	0.0417	0.0763

Table 4.3: The three measures of the disturbance caused by the parameter changes that enforce $Pr(v_3|v_1 \wedge v_8)$ to increase with 0.05, sorted on the selection criterion $\Delta A(\tau_{x|u})$, which indicates the absolute change in the parameter.

The smallest absolute change

In Table 4.3 we have sorted the table on the absolute change in the parameters. We see that a small absolute change does not always imply a small disturbance of the network. Although a small absolute change in a parameter leads to a smaller disturbance than a large absolute change in the same parameter, the disturbance is also dependent on the parameter. The output probabilities are not equal sensitive to each of the parameters, because the sensitivity functions of the parameters differ.

Although this selection criterion is using the absolute change in the parameter, it does not predict the average absolute disturbance better than the relative disturbance or the log-odds disturbance.

We would like the selection criteria to select the parameter that has the lowest disturbance. The two parameters with the lowest disturbance are $\tau_{v_8|v_3 \wedge v_7}$ and $\tau_{v_3|\neg v_2 \wedge \neg v_5}$. The selection criterion selects these parameters as fifth and as first. The parameters that are selected as third and sixth parameter have a very high disturbance. We certainly do not want to change these parameters to enforce the constraint.

The smallest relative change

If we sort the table on the relative change in the parameters, see Table 4.4, we can see that the smallest relative change selection methods performs

Parameter	$\Delta R(\tau_{x\|u})$	**av.** $\Delta A(a\|e)$	**av.** $\Delta R(a\|e)$	**av.** $\Delta O(a\|e)$
$\tau_{v_3\|\neg v_2 \wedge \neg v_5}$ +	0.1480	0.0094	0.0308	0.0466
$\tau_{v_8\|\neg v_3 \wedge \neg v_7}$	0.2970	0.0103	0.0328	0.0500
$\tau_{v_8\|v_3 \wedge v_7}$ +	0.3079	0.0092	0.0290	0.0449
$\tau_{v_3\|\neg v_2 \wedge v_5}$	0.4434	0.0158	0.0506	0.0763
$\tau_{v_8\|\neg v_3 \wedge v_7}$	0.4700	0.0146	0.0417	0.0690
$\tau_{v_2\|v_1}$ −	0.5050	0.0189	0.0613	0.1016
$\tau_{v_8\|v_3 \wedge \neg v_7}$	0.6121	0.0100	0.0708	0.0640
$\tau_{v_7\|v_6}$ −	0.6378	0.0229	0.0730	0.1031

Table 4.4: The three measures of the disturbance caused by the parameter changes that enforce $Pr(v_3|v_1 \wedge v_8)$ to increase with 0.05, sorted on the selection criterion $\Delta R(\tau_{x|u})$, which indicates the relative change in the parameter.

slightly better. We can see that the relative change in the parameter is correlated with the disturbance of the network. The parameter selected is still not the parameter that has the lowest disturbance measures. The parameters with the lowest disturbance $\tau_{v_8|v_3 \wedge v_7}$ and $\tau_{v_3|\neg v_2 \wedge \neg v_5}$ are selected as first and as third. The two parameters with the highest disturbance, $\tau_{v_2|v_1}$ and $\tau_{v_7|v_6}$ are selected as sixth and as eighth.

Notice that the parameter that is selected seventh $\tau_{v_8|v_3 \wedge \neg v_7}$ has indeed a high relative disturbance. However, this parameter has a low absolute disturbance.

The smallest log-odds change

Recall that we are using the log-odds change in the parameter as a selection criterion, because the value of the log-odds change in the parameter can be used as a measure to bound the change in any probability $Pr(a|e)$ in the network. We illustrate this bound using our example network. This bound indicates that the log-odds change in any probability of the network is smaller than or equal to the log-odds change of the parameter. We have calculated the largest log-odds change in any of the output probabilities in the network after applying each of the parameter changes. We have excluded the probability that corresponds with the parameter that is changed. The bound implies that these values of the largest log-odds change have to be equal to or smaller than the log-odds change in the parameter. Table 4.5

Parameter	$\Delta O(\tau_{x\|u})$	**max** $\Delta A(a\|e)$	**max** $\Delta R(a\|e)$	**max** $\Delta O(a\|e)$
$\tau_{v_3\|\neg v_2 \wedge \neg v_5}$	0.4987	0.1240	0.5272	0.4987
$\tau_{v_8\|\neg v_3 \wedge \neg v_7}$	0.5570	0.1287	0.4160	0.5570
$\tau_{v_8\|v_3 \wedge v_7}$	0.6010	0.1416	0.3081	0.6010
$\tau_{v_7\|v_6}$	0.8178	0.2016	0.8530	0.8178
$\tau_{v_2\|v_1}$	0.8605	0.1972	0.5606	0.8605
$\tau_{v_8\|\neg v_3 \wedge v_7}$	0.9067	0.1798	0.7686	0.9067
$\tau_{v_3\|\neg v_2 \wedge v_5}$	0.9094	0.2234	1.2422	0.9094
$\tau_{v_8\|v_3 \wedge \neg v_7}$	1.0623	0.2200	1.5782	1.0623

Table 4.5: The maximum absolute change, max $\Delta A(a|e)$, the maximum relative change, max $\Delta R(a|e)$ and the maximum log-odds change, max $\Delta O(a|e)$, after changing the parameter that enforces $Pr(v_3|v_1 \wedge v_8)$ to increase with 0.05. Sorted on the log-odds change in the parameter $\Delta O(\tau_{x|u})$. The value of the log-odds change in the parameter can be used to predict the maximum log-odds change in the output probability.

summarises the log-odds changes in the parameters and the largest log-odds change in the output. As we can see, the bound described by Chan and Darwiche [1] is indeed the maximum log-odds change of $Pr(y|e)$.

For example, consider changing parameter $\tau_{v_3|\neg v_2 \wedge \neg v_5}$. We have to change the parameter with a log-odds difference of 0.4987 to enforce the constraint. We can see in the fifth column of the table the probability that has the highest log-odds change, has a log-odds change of 0.4987. If we change this parameter to the suggested value, all the output probabilities that are changed, are indeed changed with equal or a lower log-odds change than 0.4987. It is also interesting to notice that this bound is tight when changing a single parameter.

By selecting the parameter, that needs the lowest log-odds change, we minimize the maximum difference in log-odds change of the output probabilities. When changing $\tau_{v_3|\neg v_2 \wedge \neg v_5}$, we can be sure that the output probabilities are not changed with a higher log-odds difference than 0.4987, because the parameter is changed with 0.4987. If we instead change parameter $\tau_{v_8|\neg v_3 \wedge \neg v_7}$, the output probabilities could change with a log-odds difference of 0.5570.

Table 4.6 summarises the disturbance measures calculated as before but now sorted on log-odds value of the parameter. It is clear that a low maximum bound does not always indicate a low average log-odds change. Parameter $\tau_{v_8|v_3 \wedge v_7}$ is changed with a log-odds change of 0.6010, which gives a higher bound than parameter $\tau_{v_3|\neg v_2 \wedge \neg v_5}$ which is changed with a log-odds change of 0.4987. However, we can see in the table that the average disturbance in the output when changing parameter $\tau_{v_8|v_3 \wedge v_7}$ is lower.

This selection criterion also selects the parameters with the lowest disturbance $\tau_{v_8|v_3 \wedge v_7}$ and $\tau_{v_3|\neg v_2 \wedge \neg v_5}$ as first and as third. The two parameters with the highest disturbance, $\tau_{v_2|v_1}$ and $\tau_{v_7|v_6}$ are selected as fourth and fifth.

Parameter	$\Delta O(\tau_{x\|u})$	**av.** $\Delta A(a\|e)$	**av.** $\Delta R(a\|e)$	**av.** $\Delta O(a\|e)$
$\tau_{v_3\|\neg v_2 \wedge \neg v_5}$ +	0.4987	0.0094	0.0308	0.0466
$\tau_{v_8\|\neg v_3 \wedge \neg v_7}$	0.5570	0.0103	0.0328	0.0500
$\tau_{v_8\|v_3 \wedge v_7}$ +	0.6010	0.0092	0.0290	0.0449
$\tau_{v_7\|v_6}$ −	0.8178	0.0229	0.0730	0.1031
$\tau_{v_2\|v_1}$ −	0.8605	0.0189	0.0613	0.1016
$\tau_{v_8\|\neg v_3 \wedge v_7}$	0.9067	0.0146	0.0417	0.0690
$\tau_{v_3\|\neg v_2 \wedge v_5}$	0.9094	0.0158	0.0506	0.0763
$\tau_{v_8\|v_3 \wedge \neg v_7}$	1.0623	0.0100	0.0708	0.0640

Table 4.6: The three measures of the disturbance caused by the parameter changes that enforce $Pr(v_3|v_1 \wedge v_8)$ to increase with 0.05, sorted on the selection criterion $\Delta O(\tau_{x|u})$, which indicates the log-odds change in the parameter.

4.1.3 Conclusion

Both the absolute change in the parameter as the closeness to an extreme value of the current value of the parameter have proved to be an indication of the disturbance of the network. However, the absolute and the relative change in the parameter do not provide enough information to accurately predict the amount of disturbance caused by the parameter. Although we can find an increasing trend in the disturbance, due to too much variation, these two selection criteria cannot find the parameter that disturbs this network the least.

The log-odds change in the parameter can be used as a bound to find the maximum log-odds change in the output. This bound indeed found the maximum log-odds change in our experiment for all the parameters. We found the bound to be tight in our example, even when excluding the probability $Pr(a|e)$ that corresponds to the parameter $\tau_{x|u}$. This bound, is proved tight in general when we do not exclude this parameter.

When this bound is used to predict the maximum absolute and the maximum relative disturbance, it does not find the optimal parameter. This bound does not give a better prediction of the average disturbance than the absolute and relative change of the parameter. The three selection criteria all found the same parameter to be the optimal parameter. Although, another parameter turned out to disturb the network the least, the parameter found by the selection criteria has a reasonable low disturbance measure.

If we observe the two parameters that have the lowest disturbance and the two parameters that has the highest disturbance we see that the relative selection criterion predicts the disturbance best. The parameters with the lowest disturbance can be found in the upper region and the parameters with the highest disturbance in the lower region of the table. As we are only investigating the result of one network, further research is needed to see if this is always the case. We can conclude that the three selection criteria do not accurately predict the disturbance of the network, but can be used to give an indication.

4.2 Selecting the optimal multiple parameter changes

In the previous section, we have investigated the selection criteria for single parameter changes. If a constraint can be enforced by small changes to multiple parameters instead of a larger change to a single parameter, then the changing of multiple parameters may be preferred. Unfortunately, as argued in the previous chapter, finding the parameter changes for multiple parameters from different CPTs that enforce a constraint is computationally too demanding. In this section, we will focus on finding the optimal multiple parameter changes in a single conditional probability table. To find the optimal combination of parameter changes we can again use different selection criteria. We will first investigate these different types of selection criteria. Then we will investigate the performance of these selection criteria in an experiment.

4.2.1 Selection criteria for multiple parameter changes

Instead of choosing between the parameters that are going to be changed, in this section we are searching for the optimal combination of parameter changes for the parameter in a single CPT. Of course this CPT has to be selected. An expert, for example, can choose this CPT because he assumes that the values for the parameters of a CPT were inaccurately found. We can also choose the CPT for which we know that the single parameter changes are sensitive to the constraint by first using the selection criteria for single parameter changes.

In this section, we will consider the different selection criteria that can be used to find the optimal multiple parameter change. Recall that using the methods of Chapter 3, we can find the values for the constants of the sensitivity function for multiple parameter changes. This function can be used to find the parameter changes that enforce a constraint. This sensitivity function has the following form:

$$Pr(a|e) = \frac{\sum_i \alpha_i \cdot \tau_i + \kappa}{\sum_i \gamma_i \cdot \tau_i + \mu}$$

Instead of the constants β and σ, this function uses the constants κ and μ, where κ denotes the value of $Pr(a \wedge e)$ if all the parameters in the CPT are changed to zero and μ denotes the value of $Pr(e)$ if all these parameters are changed to zero. These constants are further explained in the previous chapter. The index i is used to specify the parameters.

If we for example want to enforce the constraint $Pr(a|e) = \eta$ we thus only have to make sure that the following equation is satisfied:

$$\frac{\sum_i \alpha_i \cdot \tau_i + \kappa}{\sum_i \gamma_i \cdot \tau_i + \mu} = \eta$$

However, there are many combinations of parameter changes possible for which this equation is satisfied. All these combinations enforce the constraint. Again, we would like to find the optimal parameter changes to enforce a constraint. In this section, we are going to discuss the different selection criteria to choose between these parameter changes.

smallest total absolute change

One way of selecting the optimal parameter changes, could be by selecting from the set of parameter changes that satisfy the equation, the parameter changes that have the smallest total absolute change, $\sum_i \Delta A(\tau_i)$. We know from the single parameter selection criteria, that the absolute change is a indication of the disturbance in the network. A larger absolute change in the parameter will give a larger disturbance than a small change in the same parameter.

To be certain that the parameter changes enforce the constraint, we adjust, after assigning the other parameter values, the value of the last parameter in such a way that the parameter changes satisfy the equation. This can be easily done, because after assigning the other parameter changes, the last parameter is the only unknown variable. We will illustrate this with an example in Section 4.2.2. We have to be careful, that the only valid solutions are the solutions were all the parameters have a value larger than or equal to zero and smaller than or equal to one. From these set of valid solutions, we want to select the solution that has the smallest total absolute change:

$$\min \sum_i \Delta A(\tau_i)$$

Where the index i is used to specify the parameters.

However, we obtain the smallest total absolute change when we change the parameter that needs the smallest absolute change to enforce the constraint and we keep the other parameters on the current values. This is because we are changing only parameters from a single CPT. The sensitivities of the output probabilities to the parameters are independent of the values of the other parameters in the same CPT. Changing two parameters simultaneously cannot make the output probabilities change more. This makes the function $\sum_i \Delta A(\tau_i)$ linear in the parameters. Changing the parameter for which the output probabilities are most sensitive will result in the lowest total absolute change. Therefore, instead of a multiple parameter change this method will only find a single parameter change. Still, we can compare the performance of this single parameter change with the multiple parameter changes found by the other selection criteria.

smallest total relative change

When changing multiple parameters, we can use the total relative change, $\sum_i \Delta R(\tau_i)$ of the parameters selected as a selection criterion. Again, we use the change in the parameter relative to the distance to zero if the current value of the parameter is lower than 0.5 and the relative to the distance to one if the current value of the parameter is higher than 0.5. We can use this selection criterion to penalties the extreme values. As seen in the previous section, the output probabilities are more sensitive to changes in parameters that have extreme values.

Again, we can adjust the value of the last parameter in such a way that the parameter changes enforce the constraint. From this set of parameter changes, we want to find the smallest relative change.

$$\min \sum_i \Delta R(\tau_i)$$

Where the index i is used to specify the parameters. Unfortunately, the function is not linear in the parameters, if we calculate the relative change. We can use a hill-climbing technique to find the minimum of this function.

smallest total log-odds

The log-odds change can also be used to bound the maximum log-odds change when changing multiple parameters. When changing only one parameter, the log-odds change in the parameter bounds the maximum log-odds change of the output. If a parameter is changed with a log-odds of η, each output probability is changed with a log-odds value between zero and η. If we then change another parameter with a log-odds of η, then the output probabilities are again changing with a log-odds value between zero and η. The total log-odds change for each output probability, when changing two parameters with the value of η thus has to be between zero and $2 \cdot \eta$.

So in general, the log-odds change of the output probabilities when changing multiple parameters is between zero and the sum of the log-odds change of all the parameters that are changed: $\sum_i \Delta O(\tau_i)$. Consequently, the maximum log-odds change in the output probabilities is bounded to be smaller than or equal to the total log-odds change of the changed parameters:

$$\max \Delta O(a|e) \leq \sum_i \Delta O(\tau_i)$$

However, this bound can be large and it is not tight. Only in the extreme case that a output probability is sensitive to all the parameters that we are changing, then that probability can be changed with this maximum value.

Furthermore, the output probability has to be changed in the same direction by all the parameters otherwise the log-odds changes are cancelled out.

To find the optimal multiple parameter change, we can minimize the total log-odds change in the parameters, $\sum_i \Delta O(\tau_o)$. If we can find a very small value for the total log-odds change, we know we have a very small maximum log-odds change in the output. One way to find this minimum is using a hill-climbing technique.

equal log-odds

Chan and Darwiche [2] proposed another method for selecting the combination of parameters when changing multiple parameters. They described a method for changing all parameters with equal log-odds changes. We know that the maximum log-odds change in the output is equal to the log-odds change of the parameter that is changed. Thus when changing multiple parameters, we know that the maximum log-odds change in the output is equal to or greater than the log-odds change of every changed parameter. The maximum log-odds change will be equal to or greater than the log-odds change of the parameter that has been changed with the largest log-odds change.

$$\max \Delta O(a|e) \geq \max_i \Delta O(\tau_i)$$

If we want to minimise the log-odds change of the parameter that has the largest log-odds change we have to change the parameters equally. For example, suppose that we want to change four parameters in the same CPT. If the parameters are not changed equally, the log-odds change of the parameter that has the largest log-odds change can be decreased by increasing the log-odds changes of the other parameters. This will result eventually in equal log-odds changes [2]. We will refer to this selection criterion as $eq.\Delta O(\tau_i)$.

4.2.2 Experiment

To compare these selection criteria we have used the same network as the one used for the experiments with the single parameter changes. We have chosen to change the parameters of the CPT of V_8 to increase the probability $Pr(v_3|v_1 \wedge v_8)$ by 0.05, which is the same constraint as the one used in the previous experiment. We also used the same measures to calculate the disturbance of the network. These measures are explained in the previous experiment. In the next paragraphs, we are going to explain how to find the solutions using the selection criteria. Then we will compare the results of the selection criteria.

We will compare these multiple parameter changes with two single parameter changes from the previous experiment. The first single parameter change is the parameter change that was found to have the lowest disturbance of the network, $\tau_{v_8|v_3 \wedge v_7}$. The second single parameter change is the parameter change that gives the smallest total absolute change, $\min_i \sum_i \Delta A(\tau_i)$. This selection criterion always selects a single parameter change.

Finding the solutions for the selection criteria

To find the solutions of the multiple parameter changes we have to use the sensitivity function of multiple parameter changes. As we have explained in the previous chapter, we can find this sensitivity function for multiple parameter changes by using the values of the constants of the sensitivity functions for single parameter changes.

If we want to change the parameters of the CPT for V_8, we can use the values of the constants we have found for the single parameter changes. When using the method of the previous chapter to find these single parameter changes, we receive the values of the constants summarised in Table 4.7:

parameter	α	β	γ	σ
$\tau_{v_8\|v_3 \wedge v_7}$	0.1731	0.0155	0.1731	0.1664
$\tau_{v_8\|v_3 \wedge \neg v_7}$	0.2407	0.0935	0.2407	0.2444
$\tau_{v_8\|\neg v_3 \wedge v_7}$	0.0000	0.1091	0.1540	0.2023
$\tau_{v_8\|\neg v_3 \wedge \neg v_7}$	0.0000	0.1091	0.2151	0.1667

Table 4.7: Values of the constants of the sensitivity functions for the outcome probability $Pr(v_3|v_1 \wedge v_8)$ for all parameters from the CPT for V_8.

Recall that the sensitivity function for multiple parameter changes is as follows:

$$Pr(a|e) = \frac{\sum_i \alpha_i \cdot \tau_i + \kappa}{\sum_i \gamma_i \cdot \tau_i + \mu}$$

For ease of reference we call:
$\tau_{v_8|v_3 \wedge v_7} = \tau_1$
$\tau_{v_8|v_3 \wedge \neg v_7} = \tau_2$
$\tau_{v_8|\neg v_3 \wedge v_7} = \tau_3$
$\tau_{v_8|\neg v_3 \wedge \neg v_7} = \tau_4$

Filling in the values of the constants of α and γ results in:

$$Pr(v_3|v_1 \wedge v_8) = \frac{0.1731 \cdot \tau_1 + 0.2407 \cdot \tau_2 + \kappa}{0.1731 \cdot \tau_1 + 0.2407 \cdot \tau_2 + 0.1540 \cdot \tau_3 + 0.2151 \cdot \tau_4 + \mu}$$

Recall that the value of the constant κ can be calculated using the following formula:

$$\kappa = \beta_1 - \sum_{i \neq 1} \alpha_i \cdot \tau_i^o$$

In our case, for κ this results in:

$$\kappa = 0.0155 - (0.2407 \cdot 0.0645 + 0.000 \cdot 0.3746 + 0.00 \cdot 0.4334) = 0.000$$

The constant μ can be calculated using this formula:

$$\mu = \sigma_1 - \sum_{i \neq 1} \gamma_i \cdot \tau_i^o$$

For μ this results in:

$$\mu = 0.1664 - (0.2407 \cdot 0.0645 + 0.1540 \cdot 0.3746 + 0.2151 \cdot 0.4334) = 0.000$$

We now have the sensitivity function for the parameter changes in the CPT for V_8. To enforce the constraint $Pr(v_3|v_1 \wedge v_8) = 0.4695$, we have to satisfy:

$$\frac{0.1731 \cdot \tau_1 + 0.2407 \cdot \tau_2 + 0.000}{0.1731 \cdot \tau_1 + 0.2407 \cdot \tau_2 + 0.1540 \cdot \tau_3 + 0.2151 \cdot \tau_4 + 0.000} = 0.4695$$

There are many different multiple parameter changes for which this equation holds. For example, suppose we want the first three parameters to have the value of 0.3, we can calculate the value of the fourth parameter:

$$Pr(v_3|v_1 \wedge v_8) = \frac{0.1241}{0.1703 + 0.2151 \cdot \tau_4} = 0.4695$$

The fourth parameter $\tau_{v_8|v_3 \wedge \neg v_7}$ has to have a value of 0.4372. These parameter changes will enforce the constraint. However, it is easy to see that if we increase the parameter change of one parameter, we can adjust the parameter changes of the other parameters and still enforce the constraint. All parameter changes that satisfy the equation enforce the constraint. We use the selection methods described earlier to select a multiple parameter change that is considered optimal.

We have used this equation to find the smallest total relative change and the smallest total log-odds change in the parameters possible. The solution had to satisfy the equation. We used a heuristic to find these parameter values, because these values are hard to find. First, we assigned a random change to three parameters. Then we calculated the value of the fourth parameter using the equation. In this way, we could be sure that our solution enforces the constraint. The value of the fourth parameter had to be between zero and one, otherwise the solution was considered wrong. We searched for the optimal parameters by using a hill climbing algorithm that finds a local optimum and ran this algorithm multiple times to find the global optimum.

Calculating the equal log-odds change that enforces the constraint is easier. Because we now only have one variable, namely the log-odds change of all the parameters. There is only one value of this log-odds change for which the constraint holds. Chan and Darwiche [2] proposed an algorithm for finding the equal log-odds change for which the equation holds. For each value of the log-odds change, we can calculate the output probability. The higher the value of the log-odds changes in the parameters, the higher the value of the log-odds change in the probability. We can use the difference between this probability and the probability we want it to have, to see if we have to increase or decrease the log-odds change. They proposed to increase or decrease this value until the difference between the calculated probability and the probability wanted is small enough. We have found that an equal log-odds change of 0.1808 enforces the constraint in our example.

Results of the experiment

Using the same measures from the previous experiment, we have calculated the disturbance in the network after enforcing the multiple parameter changes proposed by the selection criteria. The results of this experiment is summarised in Table 4.8.The $\min \sum_i \Delta A(\tau_i)$ is the same as changing the parameter that needs the lowest absolute change to enforce the constraint, which is $\tau_{v_8|v_3 \wedge \neg v_7}$, and thus not a multiple parameter change. We also included the single parameter change that gave the best performance in the previous experiment, $\tau_{v_8|v_3 \wedge v_7}$.

Selection Criteria	**av.** $\Delta A(a\|e)$	**av.** $\Delta R(a\|e)$	**av.** $\Delta O(a\|e)$
eq. $\Delta O(\tau_i)$	0.0075	0.0278	0.0390
$\tau_{v_8\|v_3 \wedge v_7}$	0.0092	0.0290	0.0449
$\min_i \sum_i \Delta O(\tau_i)$	0.0098	0.0313	0.0473
$\min_i \sum_i \Delta R(\tau_i)$	0.0102	0.0325	0.0491
$\min_i \sum_i \Delta A(\tau_i)*$	0.0100	0.0708	0.0640

Table 4.8: The average disturbance in the output probability $Pr(a|e)$ when using each selection criterion for multiple parameter changes.

From the table we see that changing the parameters with an equal log-odds change results in the lowest disturbance of the network, as indicated by the three measures. The other selection criteria for multiple parameter changes performed badly. We will investigate why this was the case.

Selection Criteria	$\max \Delta A(a\|e)$	$\max \Delta R(a\|e)$	$\max \Delta O(a\|e)$
eq. $\Delta O(\tau_i)$	0.0582	0.3116	0.2746
$\min_i \sum_i \Delta O(\tau_i)$	0.1103	0.3663	0.4384
$\min_i \sum_i \Delta R(\tau_i)$	0.1250	0.3992	0.5396
$\min_i \sum_i \Delta A(\tau_i)*$	0.2200	1.5782	1.0623

Table 4.9: The maximum disturbance of the output when using each selection criterion for multiple parameter changes.

We have summarised the maximum disturbance of the parameter changes in Table 4.9. Changing the parameters equal also resulted in a lower maximum disturbance.

When discussing the selection criteria we have described two bounds to predict the value of the maximum log-odds change in the output. Recall that these were the following bounds:

$$\max \Delta O(a|e) \leq \sum_i \Delta O(\tau_i)$$

$$\max \Delta O(a|e) \geq \max_i \Delta O(\tau_i)$$

Combining these bounds results in:

$$\max_i \Delta O(\tau_i) \leq \max \Delta O(a|e) \leq \sum_i \Delta O(\tau_i)$$

This means that the maximum log-odds change in the output is a value between the log-odds change of the parameter that has the largest log-odds change and the sum of the log-odds change of all the changed parameters. In Table 4.10 we can see that these bounds are satisfied. The maximum log-odds change in the output is indeed between the two values.

Selection Criteria	$\max \Delta O(\tau_i)$	$\sum_i \Delta O(\tau_i)$	$\max \Delta O(a\|e)$
eq. $\Delta O(\tau_i)$	0.1808	0.7232	0.2746
$\min_i \sum \Delta O(\tau_i)$	0.4384	0.5545	0.4384
$\min_i \sum \Delta R(\tau_i)$	0.5369	0.5568	0.5396
$\min_i \sum \Delta A(\tau_i)*$	1.0623	1.0623	1.0623

Table 4.10: The maximum disturbance of the output when using each selection criterion for multiple parameter changes.

When we change the parameter that needs the smallest log-odds difference to enforce the constraint much more than the other parameters to create a low total log-odds change like when using the selection criterion: $\min_i \sum_i \Delta O(\tau_i)$, one parameter has to be changed much. This will cause a high maximum log-odds change in the output, as the maximum log-odds change is equal to the log-odds change of the parameter that has the largest log-odds change. This is the reason that using the $\min_i \sum_i \Delta O(\tau_i)$ we have such a high disturbance in the network.

Not all the parameters that we are changing have an equal effect on the output probability. If we change all the parameters with equal log-odds, we are changing parameters, which hardly contribute to enforcing the constraint. This does not seem clever. The maximum disturbance of the output when changing the parameters equally is much higher than the log-odds change of the parameters. This is because when all the parameters are changed equally, all the parameters influence the output probabilities in the network much. Even parameters that have a low contribution to enforcing the constraint are disturbing probabilities.

The maximum log-odds change can be seen as to come from two values. First, we have the value of the largest log-odds change. We are sure that the maximum disturbance is at least as high as this value; we can call this the starting value of the maximum log-odds change in the output. Furthermore, the log-odds change of the other parameters can cause the maximum log-odds change to be higher. We can call this an additional log-odds change. This additional value is less effective.

maximum log-odds disturbance = δ + additional value

To illustrate this, we have used another selection criterion. It finds the smallest total log-odds change, but we restrict the log-odds changes of the parameters to be lower than a certain value δ:

$$\min_i \sum_i \Delta O(\tau_i) \text{ while } \Delta O(\tau_i) \leq \delta$$

If $\delta = 0.1808$, this selection criterion will find equal parameter changes. If $\delta = \infty$, this selection criterion is the same as $\min_i \sum_i \Delta O(\tau_i)$. If $0.1808 \leq \delta \leq 0.4384$ then it finds a hybrid solution. For example, when using $\delta = 0.2000$ this selection criterion restricts the log-odds change of the parameters to be larger than 0.2. It searches for the combination of parameter changes that finds the smallest total log-odds change in the parameters.

If we look at the value of the maximum log-odds disturbance as having a starting value and an additional value then δ equals the starting value. This is because the value of δ is equal to the parameter that has the largest log-odds change and this is equal to or lower than the maximum log-odds disturbance of the network. However, a low value of δ has a higher additional value.

Figure 3.1 illustrates this. We can see that the maximum log-odds disturbance is always higher than the value of δ. The value of δ can be seen as the starting value of the maximum disturbance. We can also see that the additional value decreases as δ increases, because the other parameters are changed less. To find the optimal value of δ we have to find a balance between the starting value and the additional value.

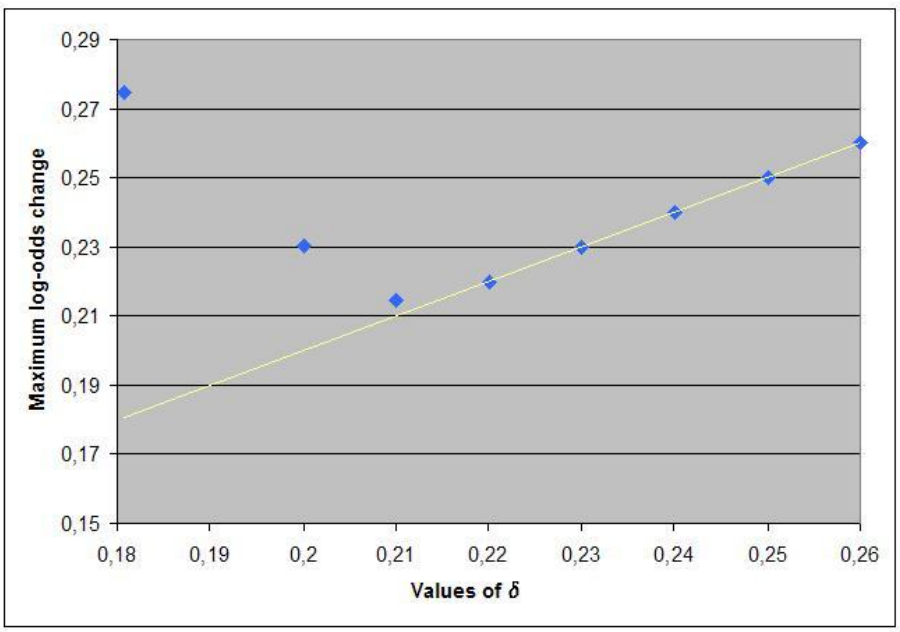

Figure 4.2: The maximum log-odds change in the output for different values of δ. By points on the line, δ is equal to the maximum log-odds change.

Table 4.11 also illustrates this balance. Having a higher value of δ indeed lowers the total log-odds change and thus the values of the other parameters are changed less. The maximum disturbance is indeed always higher than the value of δ. A value of δ between 0.20 and 0.21 does not only result in the lowest maximum disturbance but also in a lower average disturbance.

Sel. Criteria	$\sum_i \Delta O(\tau_i)$	$\max_i \Delta O(\tau_i)$	$\max \Delta O(a\|e)$	**av.** $\Delta O(a\|e)$
$\delta = 0.1808$*	0.7234	0.1808	0.2746	0.0390
$\delta = 0.20$	0.6199	0.2000	0.2301	0.0381
$\delta = 0.21$	0.6110	0.2100	0.2147	0.0381
$\delta = 0.22$	0.6024	0.2200	0.2200	0.0389
$\delta = 0.4384$**	0.5545	0.4384	0.4384	0.0473

Table 4.11: The disturbance in the network for different values of δ. *The selection method eq. $\Delta O(\tau_i)$ can in this example be seen as to have $\delta = 0.1808$. **The selection method $\min_i \sum_i \Delta O(\tau_i)$ can be seen as to have $\delta = 0.4384$.

In this experiment, the multiple parameter changes gave a lower maximum log-odds change than the single parameter changes, but this is not always the case. When we want to change the output $Pr(a|e)$ and we can change a parameter that corresponds directly to this output probability, then changing this single parameter is more effective then changing multiple parameters.

4.2.3 Conclusion

Using equal log-odds change, we find a reasonable good solution. However, to find the optimal solution we have to change the parameters that need a smaller log-odds change to enforce a constraint more then the parameters that need a large change. We cannot change a single parameter too much, because changing a parameter with a large log-odds difference will give a large disturbance. To find the optimum parameter changes, we need to find a balance between changing all the parameters equally and changing single parameters that have a high contribution to enforcing the constraint more.

In the experiment, we have found this balance by using different values for the maximum log-odds difference that the parameters are allowed to have.

The optimum value of this maximum is dependent on the network. Further research is needed to investigate how to find this balance.

4.3 Summary and conclusion

In this chapter, we have investigated different selection criteria. If we want to change only a single parameter, then we can use selection criteria to choose between the different parameter changes that are obtained by the methods described in Chapter 3. However, in our experiment, these selection criteria failed to find the optimal parameter change.

One of the selection criteria selects the parameter which needs the lowest log-odds change to enforce the constraint. Although, this does not always indicate the lowest average disturbance of the output probabilities in the network, it does give us a maximum bound of the log-odds change in the output. We know that the output probabilities are changed with an equal or a lower log-odd change than the log-odds change of the parameter. Moreover, we know that this bound is tight when changing a single parameter.

Sometimes changing multiple parameters in the same CPT can give a better result than changing a single parameter. When changing multiple parameters in one CPT, we will obtain many possible solutions. We have to use a selection criterion to find the optimal values for the changes in the parameter. That is why choosing between combinations of multiple parameter changes is more difficult.

In this chapter, we have introduced a bound that indicates the value of the most changed output probability when changing multiple parameters. However, this bound is not very tight.

Chan and Darwiche [2] proposed changing the parameters equally. Although this gives reasonable results, it is not always giving the best combination. In our experiment, changing parameters that have a high contribution to enforcing the constraint a bit more than the other parameters gave the best results.

To find the optimal solution when changing multiple parameters we would both like to have a small total log-odds value of the parameter changes as well as a small log-odds change in each of the parameters. We have to find a balance between these two values.

Chapter 5

Conclusions and future work

In this thesis, we investigated the constraints that can be useful to enforce when tuning the network. We have explained how these constraints can be enforced using sensitivity functions and how these can be enforced using only the derivatives and we compared these two approaches. We have also discussed the algorithms that can be used to find these sensitivity functions. Using these algorithms, we were able to find the parameter change for each parameter in the network that enforces a constraint.

In this thesis, we have discussed changing a single parameter or changing multiple parameters from a single CPT to enforce a constraint. We have found the parameter change(s) that enforce(s) such a constraint using sensitivity functions. How to find the parameter changes for multiple parameters from different CPTs that enforce a constraint has to be investigated in future research.

If we want to enforce a constraint, we will have to change some parameters. It is clear that changing a parameter can disturb many probabilities of the network. If a little change in a parameter is enough to enforce a constraint, it is probably also enough to disturb the network. In this thesis, we have investigated how we can select the parameter change that maintains most of the output probabilities. We have discussed some selection criteria to select the optimal parameter change, which is the parameter change that maintains most of the output probabilities from a set of parameter changes that enforces a constraint and the selection criteria to select the optimal multiple parameter change when we want to change multiple parameters in one CPT. We have tested the performance of these selection criteria in an experiment. The selection criteria found a reasonable solution, but did not find the optimal solution.

In an experiment, we where able to find the parameter changes that maintain more of the probabilities of the network by using a combination of two selection criteria. How this combination of two selection criteria performs in different networks has to be investigated in future research.

Bibliography

[1] Hei Chan and Adnan Darwiche. When do numbers really matter? In *UAI '01: Proceedings of the 17th Conference in Uncertainty in Artificial Intelligence*, pages 65–74, San Francisco, CA, USA, 2001. Morgan Kaufmann Publishers Inc.

[2] Hei Chan and Adnan Darwiche. Sensitivity analysis in bayesian networks: from single to multiple parameters. In *UAI '04: Proceedings of the 20th conference on Uncertainty in Artificial Intelligence*, pages 67–75, Arlington, VA, USA, 2004. UAI Press.

[3] Veerle M.H. Coupé and Linda C. van der Gaag. Properties of sensitivity analysis of bayesian belief networks. In *Annals of Mathematics and Artificial Intelligence*, pages 323–356. Kluwer Academic Publisher, 2002.

[4] Adnan Darwiche. A differential approach to inference in bayesian networks. In *Uncertainty in Artificial Intelligence*, pages 123–132, 2000.

[5] Uffe Kjærulff and Linda C. van der Gaag. Making sensitivity analysis computationally efficient. In *UAI '00: Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence*, pages 317–325, San Francisco, CA, USA, 2000. Morgan Kaufmann Publishers Inc.

[6] Richard E. Neapolitan. In *Probabilistic reasoning in expert systems : theory and algorithms*. New York [etc.] Wiley, 1990.

[7] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1988.

[8] Silja Renooij and Linda van der Gaag. Exploiting evidence-dependent sensitivity bounds. In *Proceedings of the 21th Annual Conference on Un-*

certainty in Artificial Intelligence (UAI-05), page 485, Arlington, Virginia, 2005. AUAI Press.