Restoring monotonicity in Bayesian networks

— Master's thesis —

Merel T. Rietbergen

Supervised by prof. dr. ir. L.C. van der Gaag

INF/SCR-09-54

April 26, 2010

Contents

Chapter 1

Introduction

In our society problems of scientific nature abound. These problems involve stochastic variables, which are in some way related to each other. For example, a physician determining whether a patient has a specific disease, needs to take into consideration symptoms and results from diagnostic tests, which have some relation to each other. A Bayesian network is a concise representation of a joint probability distribution on a set of stochastic variables for such a problem. It consists of a qualitative part, which is a directed graph where the nodes represent the variables, and a quantitative part, which is a set of conditional probabilities. The qualitative part captures the independencies between the variables, while the quantitative part captures the strength of the dependencies using parameters, which represent the probabilities. Together the two parts uniquely define a joint probability distribution on the variables of the problem. Inserting observations into the network allows us to compute probabilities for a variable of interest given those observations.

In general, we expect that with worse observations a worse outcome for the variable of interest becomes more likely. For example, if a physician observes that a patient has worse symptoms or worse test results, then he will assume that it is more likely that the patient has a worse disease. When we use a Bayesian network to represent such a problem, we expect the Bayesian network to also exhibit this behavior, which is called monotonicity. This is, however, not always the case. When inserting observations, which are worse than before, we may expect the output of the network to be that a worse outcome is more likely, while in fact its output says that the worse outcome is less likely. In such cases we say that there is a violation of monotonicity. If we want to restore monotonicity to the Bayesian network, then all such violations must be resolved. In this thesis we will investigate one method to do so.

While monotonicity may be restored to a network by changing its qualitative part, we restrict ourselves to changes in the quantitative part. More precisely, we look for a single parameter that may be varied such that there are no violations of monotonicity in the Bayesian network. It is by no means certain that such a parameter exists, since it is not necessarily possible to restore monotonicity to the network by varying a single parameter. If, however, it is possible to do so, then this thesis provides a method to find such a parameter and the value to which to vary it. We use the graphical structure of the network to eliminate variables for which parameter variation cannot restore monotonicity. For the remaining variables we use a method called the intersection-of-intervals approach to determine whether a parameter can be varied to resolve specific violations of monotonicity and, if so, to which values. The intersection-of-intervals approach must be applied repeatedly to determine if a parameter can be varied to restore monotonicity.

The thesis is organised as follows. In Section 2, we introduce notations and concepts from graph theory as well as from Bayesian-network theory necessary in the rest of this thesis. In Section 3, we give a formal definition of monotonicity and a description of the problem. We also include some details of previous studies of monotonicity and the approach we will use to solve our problem. In Section 4, we introduce some concepts that allow us to shrink the Bayesian network in which we want to restore monotonicity. In Section 5, we detail the intersection-of-intervals approach, which is used to determine if a parameter can be varied to resolve specific violations of monotonicity. We also discuss the complexity of this method and how it can be used to restore monotonicity. In Section 6, we discuss the application of the intersection-of-intervals approach, since under specific conditions application is not necessary. Finally, in Section 7, we outline our results and conclusions as well as some ideas for further study.

Chapter 2

Preliminaries

We assume that the reader has basic knowledge of graph theory and of probability theory. Here we will just review some notations and concepts from graph theory. We will review the basic concepts for Bayesian networks more extensively; for a more detailed description of Bayesian networks and related concepts we refer to [1] and [2].

2.1 Graph theory

We begin by introducing some terminology and notations for the graph-theoretical aspects of this thesis. We consider directed acyclic graphs, or *DAGs* for short. The set of vertices of a DAG G is denoted $V(G)$, and its set of arcs is denoted $A(G)$. An arc $(V_i, V_j) \in A(G)$ is written $V_i \to V_j$ or $V_j \leftarrow V_i$. The set of *predecessors* of a vertex, or set of vertices, V is denoted $\pi(V)$ and the set of *ancestors* of V is denoted $\pi^*(V)$. Analogously, the set of *successors* of a vertex, or set of vertices, V is denoted $\sigma(V)$ and the set of *descendants* of V is denoted $\sigma^*(V)$. A *chain* is a sequence of vertices $V_i, V_{i+1}, \ldots, V_{i+j} \in V(G)$ such that for all $k = i, i+1 \ldots, i+j-1$ there is an arc $V_k \to V_{k+1}$ or $V_k \leftarrow V_{k+1}$.

We review the concept of blocking a chain in a DAG.

Definition 2.1.1. *Let G be a DAG. A chain between two vertices V_i and V_j in G is* blocked *by a set of vertices Z if there are three subsequent vertices V_1, V_2, V_3 on the chain such that one of the following three statements holds:*

- $V_1 \to V_2 \to V_3$ *and* $V_2 \in Z$*;*
- $V_1 \leftarrow V_2 \to V_3$ *and* $V_2 \in Z$*;*
- $V_1 \to V_2 \leftarrow V_3$ *and* $\sigma^*(V_2) \cap Z = \emptyset$*.*

Using the concept of blocking we define the *d-separation criterion.*

Definition 2.1.2 (d-separation criterion)**.** *Let G be a DAG. Two sets of vertices $X, Y \subseteq V(G)$ are* d-separated *by a set of vertices $Z \subseteq V(G)$, denoted $\langle X \mid Z \mid Y \rangle^d_G$, if for every $V_i \in X$ and $V_j \in Y$ all chains between V_i and V_j in G are blocked by Z.*

2.2 Bayesian networks

A Bayesian network represents a joint probability distribution on a set of stochastic variables. Each such variable can take a number of values.

Definition 2.2.1. *Let V_i be a stochastic variable. The* domain *of V_i, denoted $\Omega(V_i)$, is the set of all possible values for V_i.*

In the sequel we assume that the domain of a variable V_i has a total ordering $\leq$. For a binary variable V_i, its possible values are denoted $\bar{v}_i$ and v_i, with $\bar{v}_i \leq v_i$. For a non-binary variable V_i, we use v_i^k with $k = 1, 2, \ldots, |\Omega(V_i)|$ to denote its values; we assume that $v_i^k \leq v_i^{k+1}$. The set of all joint value assignments to a set of variables S equals the Cartesian product of the sets of values for each variable from S: $\Omega(S) = \times_{V_i \in S} \Omega(V_i)$. This set of value assignments has a partial ordering $\preceq$ induced by the total orderings $\leq$ of the sets of values for the individual variables.

A Bayesian network consists of a qualitative and a quantitative part. The qualitative part is a directed acyclic graph in which the variables are represented by vertices. The quantitative part is a set of probabilities.

Definition 2.2.2. *A* Bayesian network *is a tuple $B = (G, P)$ where*

- *G is a DAG with vertices $V(G) = \{V_1, \ldots, V_n\}, n \geq 1$, and arcs $A(G)$;*
- *P is a set of conditional probabilities $p\,(V_i \,|\, \pi(V_i))$, for all $V_i \in V(G)$.*

The set of conditional probabilities for a variable $V_i \in V(G)$ is called the conditional probability table, or *CPT*, for V_i. The CPT of a variable V_i contains a *parameter* $p\left(v_i^j \,|\, \pi\right)$ for all $v_i^j \in \Omega(V_i)$ and all $\pi \in \Omega(\pi(V_i))$. The variable of interest of a Bayesian network is denoted C, and its set of observable variables is denoted $\mathbf{E}$.

The d-separation criterion, defined in the previous section, is related to the probabilistic concept of independence as follows.

Definition 2.2.3. *Let* Pr *be a joint probability distribution on a set V of stochastic variables, and let G be a DAG with $V(G) = V$. G is called an* I-map *for* Pr *if, for all sets of variables $X, Y, Z \in V(G)$, $\langle X \,|\, Z | \, Y \rangle_G^d$ in G implies that X and Y are conditionally independent given Z in* Pr.

The set of conditional probabilities P of a Bayesian network uniquely defines a joint probability distribution on the variables which respects the independencies depicted in the DAG G.

Proposition 2.2.4. *Let $B = (G, P)$ be a Bayesian network. Then*

$$\Pr(V(G)) = \prod_{V_i \in V(G)} p\,(V_i \,|\, \pi(V_i))$$

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

We use $\sum_{\Omega(V)}$ to indicate summing over all value assignments to the set of variables V and the notation $|_{X\,=\,x}$ to indicate that the variables in X take the combination of values x in the preceding formula.

In the sequel, we will change values in the CPTs of a Bayesian network, that is we will be varying parameters in the network like in sensitivity analysis. We denote the effect of varying a parameter $p\,(u \,|\, \pi)$ on a probability $\Pr(X)$ by $\Pr(X)\,(p\,(u \,|\, \pi) = x)$, where x is an algebraic variable or a constant. If x is a constant, then we replace the original value of $p\,(u \,|\, \pi)$ by x. If x is an algebraic variable, then $\Pr(X)\,(p\,(u \,|\, \pi) = x)$ is a formula in terms of x. Note that when we vary $p\,(u \,|\, \pi)$ other parameters must also be changed. To ensure that we change these parameters in the same way every time and that their relative weights are maintained, we use a scheme called proportional scaling.

Definition 2.2.5. *Let $B = (G, P)$ be a Bayesian network, and let $\Omega(V) = \{v^1, \ldots, v^n\}$, $n \geq 1$, be the domain of a variable $V \in V(G)$. Let $p\left(v^1 \mid \pi\right)$ be a parameter from the CPT of V. Then,* proportional scaling *is the scheme of parameter variation such that*

- *$p\left(v^1 \mid \pi\right)$ is varied to x;*
- *for all $k = 2, \ldots, n$, $p\left(v^k \mid \pi\right)$ is varied to $\dfrac{1-x}{1-p\left(v^1 \mid \pi\right)} \cdot p\left(v^k \mid \pi\right)$;*
- *for all $k = 1, \ldots, n$ and $\pi' \neq \pi$, $p\left(v^k \mid \pi'\right)$ remains unchanged.*

Note that if a variable V can only take two values $v, \bar{v}$ then there is no difference between varying $p\left(v \mid \pi\right)$ and varying $p\left(\bar{v} \mid \pi\right)$, since $p\left(\bar{v} \mid \pi\right) = 1 - p\left(v \mid \pi\right)$.

We also use a well-known property from sensitivity analysis.

Proposition 2.2.6. *Let $B = (G, P)$ be a Bayesian network with a variable of interest C, a set of observable variables $\mathbf{E}$ and a joint probability distribution $\Pr$. Then, for $c \in \Omega(C)$ and $\mathbf{e} \in \Omega(\mathbf{E})$ we have that*

$$\Pr\left(c \mid \mathbf{e}\right)\left(p\left(u \mid \pi\right) = x\right) = \frac{\Pr\left(c, \mathbf{e}\right)}{\Pr\left(\mathbf{e}\right)} = \frac{\alpha x + \beta}{\gamma x + \delta},$$

where α, β, γ and δ are constants dependent of u, π, c and $\mathbf{e}$.

For more information on sensitivity analysis we refer to [3] and [4].

Chapter 3

Problem description

In this chapter we formally describe the problem of restoring monotonicity in Bayesian networks. We also discuss some results from previous studies of the subject, and outline the approach we will use to solve the problem.

3.1 The problem

In general, when people make worse observations, they expect a worse outcome or cause to become more likely. For example, if a physician observes worse symptoms for a patient, then he assumes that the patient's condition is more likely to be worse than that of a patient with less severe symptoms. When a Bayesian network is used to model the physician's problem, it has a set of observable variables $\mathbf{E}$, which represent the symptoms, and a variable of interest, which represents the condition of the patient. The physician will then expect that if he enters observations $\mathbf{e}' \in \Omega(\mathbf{E})$ which are worse than his earlier observations $\mathbf{e} \in \Omega(\mathbf{E})$, then the probability that the condition of the patient is no worse than $c \in \Omega(C)$ becomes smaller. This property is called monotonicity and is formally defined as follows.

Definition 3.1.1 (Monotonicity)**.** *Let $B = (G, P)$ be a Bayesian network with a variable of interest C, a set of observable variables $\mathbf{E}$ and a joint probability distribution* $\Pr$*. B is* isotone in distribution *for the variables $\mathbf{E}$ if $\mathbf{e} \preceq \mathbf{e}'$ implies that* $\Pr(C \leq c \mid \mathbf{e}') \leq \Pr(C \leq c \mid \mathbf{e})$ *for all $c \in \Omega(C)$ and $\mathbf{e}, \mathbf{e}' \in \Omega(\mathbf{E})$. If $\mathbf{e} \preceq \mathbf{e}'$ implies that* $\Pr(C \leq c \mid \mathbf{e}) \leq \Pr(C \leq c \mid \mathbf{e}')$ *for all $c \in \Omega(C)$ and $\mathbf{e}, \mathbf{e}' \in \Omega(\mathbf{E})$, then B is* antitone in distribution *for $\mathbf{E}$.*

In the sequel, when we say that the Bayesian network B exhibits monotonicity, we mean that B is isotone in distribution for the variables $\mathbf{E}$. If a Bayesian network B does not exhibit monotonicity, then there are one or more violations of the property of monotonicity in the network.

Definition 3.1.2. *Let $B = (G, P)$ be a Bayesian network with C, $\mathbf{E}$ and* $\Pr$ *as before. A* violation of monotonicity *is an inequality* $\Pr(C \leq c \mid \mathbf{e}) < \Pr(C \leq c \mid \mathbf{e}')$ *with $\mathbf{e} \preceq \mathbf{e}'$, for some $c \in \Omega(C)$ and $\mathbf{e}, \mathbf{e}' \in \Omega(\mathbf{E})$. Varying a parameter $p(u \mid \pi) \in P$ is said to* resolve *this violation of monotonicity if there is a value $y \in [0, 1]$ such that* $\Pr(C \leq c \mid \mathbf{e})\big(p(u \mid \pi) = y\big) \geq \Pr(C \leq c \mid \mathbf{e}')\big(p(u \mid \pi) = y\big)$.

Given a Bayesian network $B = (G, P)$ with a set of observable variables $\mathbf{E}$ and a variable of interest C, our problem is to find a value $y \in [0, 1]$ for a parameter p in the

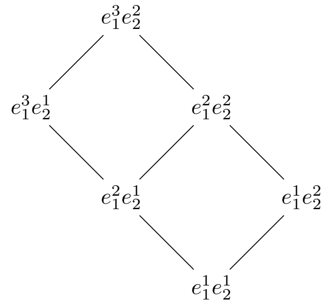

Figure 3.2.1: Assignment lattice of $\mathbf{E} = \{E_1, E_2\}$ with $\Omega(E_1) = \{e_1^1, e_1^2, e_1^3\}$ and $\Omega(E_2) = \{e_2^1, e_2^2\}$.

CPT of a variable $V \in V(G)$ such that there are no violations of monotonicity in B. Note that there is no guarantee that there actually is a variable V with a parameter p such that such a y exists.

3.2 Previous studies

The concept of monotonicity in Bayesian networks and the complexity of determining whether a network exhibits the property of monotonicity have been studied in [5] and [6]. In [5] the problem of verifying monotonicity in a Bayesian network was shown to be of a highly unfavourable computational complexity; more specifically, the problem was shown to be co-NP$^{\mathrm{PP}}$-complete. Note that since we could use a method for restoring monotonicity to verify monotonicity, we cannot expect to find a computationally favourable solution to our problem.

In [6] the concept of lattices was used in a method to verify monotonicity in Bayesian networks in time exponential in the number of observable variables. The *assignment lattice* of the set of observable variables $\mathbf{E}$ captures all joint value assignments to $\mathbf{E}$ and the partial ordering on those assignments; Figure 3.2.1 shows a graphical representation of such an assignment lattice. Note that in the lattice two distinct value assignments $\mathbf{e}, \mathbf{e}''$ to $\mathbf{E}$ are only directly connected if there is no third distinct value assignment $\mathbf{e}' \in \Omega(\mathbf{E})$ such that $\mathbf{e} \preceq \mathbf{e}' \preceq \mathbf{e}''$. Since it follows from $\Pr(C \leq c \mid \mathbf{e}'') \leq \Pr(C \leq c \mid \mathbf{e}')$ and $\Pr(C \leq c \mid \mathbf{e}') \leq \Pr(C \leq c \mid \mathbf{e})$ that $\Pr(C \leq c \mid \mathbf{e}'') \leq \Pr(C \leq c \mid \mathbf{e})$, we need only resolve all violations for value assignments which are directly connected in the assignment lattice. If two value assignments $\mathbf{e}, \mathbf{e}'$ to $\mathbf{E}$ are directly connected in the lattice, then they have a different value in a single observable variable E_i only. The value assignment to the other observable variables is the same in both $\mathbf{e}$ and $\mathbf{e}'$ and is called the *context*, denoted by $\mathbf{e}^-$; we use $\mathbf{E}^-$ to denote the set $\mathbf{E}\backslash\{E_i\}$. We denote a violation of monotonicity $\Pr\left(C \leq c \,\middle|\, e_i^k, \mathbf{e}^-\right) < \Pr\left(C \leq c \,\middle|\, e_i^{k+1}, \mathbf{e}^-\right)$ by $viol\left(c, e_i^k, \mathbf{e}^-\right)$.

3.3 Our approach

Let $B = (G, P)$ be a Bayesian network as before. To limit the number of variables in $V(G)$ for which we must investigate whether there is a parameter p in its CPT which can be varied to a value $y \in [0, 1]$ such that monotonicity is restored, we begin by eliminating all variables for which varying parameters cannot resolve any violations of monotonicity in B and can, therefore, also not restore monotonicity. We will perform these eliminations based on the graphical structure of G.

Once we have thus restricted the set of variables, we continue with a method which checks per parameter p from the CPT of a variable V whether varying it to some value $y \in [0, 1]$ can resolve all violations. Finally, to ensure that we don't apply the method needlessly, we apply the method to the variables in an order which may allow us to further restrict the set of variables under study.

Chapter 4

Eliminating variables based on graphical structure

We begin by eliminating variables from a Bayesian network for which parameter variation can never restore monotonicity based on the structure of the network. To this end we will use a concept already used in sensitivity analysis of Bayesian networks, namely the concept of sensitivity set. Furthermore, we will introduce the concept of resolution set. Both concepts restrict the set of variables for which varying a parameter from the CPT may restore monotonicity.

4.1 The sensitivity set

To illustrate how a sensitivity set can be used to restrict the set of variables for which parameter variation may restore monotonicity, we begin with its definition.

Definition 4.1.1. *Let G be the DAG of a Bayesian network with a variable of interest C and a set of observable variables $\mathbf{E}$. Now, let G^* be the DAG constructed from G by adding an auxiliary predecessor X_i to every vertex $V_i \in V(G)$. The set of all vertices $V_i \in V(G)$ for which $\neg\langle\{X_i\}|\mathbf{E}|\{C\}\rangle^d_{G^*}$ is called the* sensitivity set *of C given $\mathbf{E}$, denoted $Sen(C, \mathbf{E})$.*

In Definition 4.1.1 we have added an auxiliary predecessor X_i to every vertex $V_i \in V(G)$. This auxiliary vertex X_i may be considered a representation of the CPT of V_i.

Lemma 4.1.2. *Let $B = (G, P)$ be a Bayesian network with a variable of interest C, a set of observable variables $\mathbf{E}$ and a joint probability distribution $\Pr$. For all $V_i \in V(G)$, if $\langle\{X_i\}|\mathbf{E}|\{C\}\rangle^d_{G^*}$ then $\Pr(c \mid \mathbf{e}) \not\sim p(V_i \mid \pi(V_i))$ for any $c \in \Omega(C)$ and $\mathbf{e} \in \Omega(\mathbf{E})$, where $\not\sim$ denotes algebraic independence.*

The proof of Lemma 4.1.2 can be extracted from [3]. Intuitively, since G is an I-map of $\Pr$, we have that if $\langle\{X_i\}|\mathbf{E}|\{C\}\rangle^d_{G^*}$, then C is independent of X_i given $\mathbf{E}$. The variable of interest C thus is independent of the representation of the CPT of V_i given the observed variables $\mathbf{E}$, which means that the probability of any outcome $C = c$ given observations $\mathbf{E} = \mathbf{e}$ is not influenced by the parameter values in the CPT of V_i, i.e. $\Pr(c \mid \mathbf{e}) \not\sim p(V_i \mid \pi(V_i))$.

To relate the concept of sensitivity set to the problem of restoring monotonicity, we now first show that varying an arbitrary parameter from the CPT of a variable which is not in the sensitivity set cannot resolve any violation of monotonicity.

Theorem 4.1.3. *Let $B = (G, P)$ be a Bayesian network with C, $\mathbf{E}$ and $\Pr$ as before. Let $V_i \notin Sen(C, \mathbf{E})$ and let p be an arbitrary parameter from the CPT of V_i. Then, varying p cannot resolve any violation of monotonicity.*

Proof. Let $p = p\left(v_i \mid \pi\right)$. Since $V_i \notin Sen(C, \mathbf{E})$, we have that $\langle \{X_i\} | \mathbf{E} | \{C\} \rangle^d_{G^*}$ and hence that $\Pr\left(c \mid \mathbf{e}\right) \not\sim p\left(v_i \mid \pi\right)$ for all $c \in \Omega(C)$ and $\mathbf{e} \in \Omega(\mathbf{E})$. $\Pr\left(c \mid \mathbf{e}\right)$ thus is a constant in terms of the parameter $p\left(v_i \mid \pi\right)$ for all $c \in \Omega(C)$ and $\mathbf{e} \in \Omega(\mathbf{E})$. Varying the parameter thus cannot resolve any violation of monotonicity. □

The reverse of Theorem 4.1.3 does not hold. As shown in the following example, varying a parameter from the CPT of a variable in the sensitivity set can, but does not necessarily, resolve a violation of monotonicity.

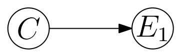

Figure 4.1.1: Directed acyclic graph used in Example 1.

Example 1. *Let B be a Bayesian network with a DAG as depicted in Figure 4.1.1 with a variable of interest C and a set of observable variables $\mathbf{E} = \{E_1\}$. Let $\Omega(E_1) = \{\bar{e}_1, e_1\}$ and $\Omega(C) = \{c^1, c^2, c^3\}$, and let the parameters from the CPTs of C and E_1 be as follows:*

$$p\left(c^1\right) = 0.5 \qquad p\left(c^2\right) = 0.3 \qquad p\left(c^3\right) = 0.2$$
$$p\left(\bar{e}_1 \mid c^1\right) = 0.6 \qquad p\left(\bar{e}_1 \mid c^2\right) = 0.8 \qquad p\left(\bar{e}_1 \mid c^3\right) = 0.7$$

We have that

$$\begin{aligned} \Pr\left(C \preceq c^2 \mid \bar{e}_1\right) &= \frac{0.54}{0.68} \approx 0.79 \quad \textit{and} \\ \Pr\left(C \preceq c^2 \mid e_1\right) &= 0.8125, \end{aligned}$$

which is a violation of monotonicity since $\bar{e}_1 \preceq e_1$. Since $E_1 \in Sen(C, \mathbf{E})$, varying one of the parameters in the CPT of E_1 may resolve the violation. If we vary $p\left(\bar{e}_1 \mid c^1\right)$ to 0.7, for example, then we have that

$$\begin{aligned} \Pr\left(C \preceq c^2 \mid \bar{e}_1\right)\left(p\left(\bar{e}_1 \mid c^1\right) = y\right) &= \frac{0.5 \cdot y + 0.24}{0.5 \cdot y + 0.38} \\ &\Downarrow \\ \Pr\left(C \preceq c^2 \mid \bar{e}_1\right)\left(p\left(\bar{e}_1 \mid c^1\right) = 0.7\right) &= \frac{0.59}{0.73} \approx 0.81, \end{aligned}$$

and

$$\begin{aligned} \Pr\left(C \preceq c^2 \mid e_1\right)\left(p\left(\bar{e}_1 \mid c^1\right) = y\right) &= \frac{-0.5 \cdot y + 0.56}{-0.5 \cdot y + 0.62} \\ &\Downarrow \\ \Pr\left(C \preceq c^2 \mid e_1\right)\left(p\left(\bar{e}_1 \mid c^1\right) = 0.7\right) &= \frac{0.21}{0.27} \approx 0.78. \end{aligned}$$

So, varying a parameter from the CPT of $E_1 \in Sen(C, \mathbf{E})$ indeed resolves the mentioned violation of monotonicity. Note that varying $p\left(\bar{e}_1 \mid c^1\right)$ to 0.7 does not restore monotonicity, since $\Pr\left(C \preceq c^1 \mid \bar{e}_1\right)\left(p\left(\bar{e}_1 \mid c^1\right) = 0.7\right) \leq \Pr\left(C \preceq c^1 \mid e_1\right)\left(p\left(\bar{e}_1 \mid c^1\right) = 0.7\right)$. □

Varying parameters from the CPTs of variables which are not in the sensitivity set of C given $\mathbf{E}$ thus cannot restore monotonicity, while varying parameters from the CPTs of variables which are in $Sen(C, \mathbf{E})$ may do so.

4.2 Restricting the Bayesian network

Using the concept of sensitivity set, we now restrict a Bayesian network to the part which is relevant for restoring monotonicity by removing variables for which varying parameters cannot restore monotonicity. To do so, we cannot simply remove all variables not in the sensitivity set of its variable of interest given its set of observable variables, however.

Lemma 4.2.1. *Let $B = (G, P)$ be a Bayesian network with a variable of interest C and a set of observable variables $\mathbf{E}$. Then, for all $V \in Sen(C, \mathbf{E})$ we have that $\pi(V) \subseteq Sen(C, \mathbf{E}) \cup \mathbf{E}$.*

Proof. Suppose that $V_1 \in Sen(C, \mathbf{E})$ and $V_2 \in \pi(V_1)$, but $V_2 \notin Sen(C, \mathbf{E}) \cup \mathbf{E}$. So, where $\neg\langle\{X_1\}|\mathbf{E}|\{C\}\rangle^d_{G^*}$, we suppose that $\langle\{X_2\}|\mathbf{E}|\{C\}\rangle^d_{G^*}$ or $V_2 \notin \mathbf{E}$. From $\neg\langle\{X_1\}|\mathbf{E}|\{C\}\rangle^d_{G^*}$ we have that there is an unblocked chain between X_1 and C. Then, there is a chain between X_2 and C on which V_2 has one incoming arc $X_2 \to V_2$ and one outgoing arc $V_2 \to V_1$. If $V_2 \notin \mathbf{E}$, then $\neg\langle\{X_2\}|\mathbf{E}|\{C\}\rangle^d_{G^*}$ and so $V_2 \in Sen(C, \mathbf{E})$. Thus if $V_1 \in Sen(C, \mathbf{E})$, then for all $V_2 \in \pi(V_1)$ either $V_2 \in \mathbf{E}$ or $V_2 \in Sen(C, \mathbf{E})$. □

From Lemma 4.2.1 we conclude that to restrict a network to the relevant part, we can only remove all variables not in $Sen(C, \mathbf{E}) \cup \mathbf{E}$, since the variables in $\mathbf{E} \backslash Sen(C, \mathbf{E})$ may still be needed to input observations. This leads to the following method.

Method 1 (Restricting a Bayesian network)**.** *Let $B = (G, P)$ be a Bayesian network with a variable of interest C and a set of observable variables $\mathbf{E}$. To restrict the Bayesian network B, remove from G all vertices not in $Sen(C, \mathbf{E}) \cup \mathbf{E}$ and their incident arcs. The DAG of the restricted Bayesian network consists only of the connected component containing C.*

Every variable V_i removed by Method 1 is removed because $V_i \notin Sen(C, \mathbf{E})$, and therefore, by Theorem 4.1.3, varying parameters from the CPT of V_i cannot resolve any violations of monotonicity. Note that after removal of all vertices not in $Sen(C, \mathbf{E}) \cup \mathbf{E}$, the only variables not in the connected component of C are observable variables. If an observable variable E_i is removed by Method 1, because it is not in the connected component containing C, then there is no violation $viol\big(c, e_i^k, \mathbf{e}^-\big)$, for $c \in \Omega(C)$, $e_i^k \in \Omega(E_i)$ and $\mathbf{e}^- \in \Omega(\mathbf{E}^-)$. Since $E_i \notin Sen(C, \mathbf{E})$, we have that $\Pr(c \mid \mathbf{e}) \nsim p(E_i \mid \pi(E_i))$ for all $c \in \Omega(C)$ and $\mathbf{e} \in \Omega(\mathbf{E})$ and therefore $\Pr(c \mid \mathbf{e})$ is constant in terms of $p(E_i \mid \pi(E_i))$.

We will use the restricted Bayesian network in the remainder of this thesis.

4.3 The resolution set

The other set, which serves to restrict the set of variables for which varying a parameter may restore monotonicity, is the resolution set and is defined as follows.

Definition 4.3.1. *Let G be the DAG of a Bayesian network with a variable of interest C and a set of observable variables $\mathbf{E}$. Let G^- be the DAG constructed from G by removing all outgoing arcs from the variables $\mathbf{E}\backslash\{E_i\}$ and C. The* resolution set *of E_i, denoted R_{E_i}, is the set of vertices of the connected component in G^- containing E_i.*

The following lemma relates the resolution set to resolving a violation of monotonicity.

Lemma 4.3.2. *Let $B = (G, P)$ be a Bayesian network with a binary variable of interest C and a set of observable variables $\mathbf{E}$. Now, let $viol\big(\bar{c}, e_i^k, \mathbf{e}^-\big)$ be a violation of monotinicity for some observable variable $E_i \in \mathbf{E}$, and let R_{E_i} be the resolution set of E_i. Then, varying a parameter from the CPT of a variable $V \notin V(G)\backslash R_{E_i}$ cannot resolve the violation.*

Proof. Let G^- be the DAG constructed from G as in Definition 4.3.1. Let $S_0, \ldots, S_j$ be the connected components of G^-, which do not contain E_i. We now show that the inequality of the violation, $\Pr\left(C = \bar{c} \mid e_i^k, \mathbf{e}^-\right) < \Pr\left(C = \bar{c} \,\middle|\, e_i^{k+1}, \mathbf{e}^-\right)$, can be expressed in terms of parameters from the CPTs of variables in R_{E_i} only.

We have

$$\Pr\left(C = \bar{c}, e_i^k, \mathbf{e}^-\right) = \prod_{j=1}^{n} \left(\sum_{\Omega(V(S_j))} \prod_{V \in V(S_j)} p\left(V \mid \pi(V)\right) \Bigg|_{\substack{\mathbf{E}^- = \mathbf{e}^- \\ C = \bar{c}}} \right) \cdot \left(\sum_{\Omega(R_{E_i})} \prod_{V \in R_{E_i}} p\left(V \mid \pi(V)\right) \Bigg|_{\substack{\mathbf{E}^- = \mathbf{e}^- \\ E_i = e_i^k \\ C = \bar{c}}} \right)$$

$$\Pr\left(e_i^k, \mathbf{e}^-\right) = \sum_{c' \in \Omega(C)} \left(\prod_{j=1}^{n} \left(\sum_{\Omega(V(S_j))} \prod_{V \in V(S_j)} p\left(V \mid \pi(V)\right) \Bigg|_{\substack{\mathbf{E}^- = \mathbf{e}^- \\ C = c'}} \right) \cdot \left(\sum_{\Omega(R_{E_i})} \prod_{V \in R_{E_i}} p\left(V \mid \pi(V)\right) \Bigg|_{\substack{\mathbf{E}^- = \mathbf{e}^- \\ E_i = e_i^k \\ C = c'}} \right) \right)$$

and similarly $\Pr\left(C = \bar{c}, e_i^{k+1}, \mathbf{e}^-\right)$ and $\Pr\left(e_i^{k+1}, \mathbf{e}^-\right)$.

To simplify notations we use the following functions

$$\begin{aligned}
f(x) &= \prod_{j=1}^{n} \left(\sum_{\Omega(V(S_j))} \prod_{V \in V(S_j)} p\left(V \mid \pi(V)\right) \Bigg|_{\substack{\mathbf{E}^- = \mathbf{e}^- \\ C = x}} \right) \\
g(x) &= \sum_{\Omega(R_{E_i})} \prod_{V \in R_{E_i}} p\left(V \mid \pi(V)\right) \Bigg|_{\substack{\mathbf{E}^- = \mathbf{e}^- \\ E_i = e_i^k \\ C = x}} \\
h(x) &= \sum_{\Omega(R_{E_i})} \prod_{V \in R_{E_i}} p\left(V \mid \pi(V)\right) \Bigg|_{\substack{\mathbf{E}^- = \mathbf{e}^- \\ E_i = e_i^{k+1} \\ C = x}}
\end{aligned}$$

Note that $f(x)$ contains parameters from the CPTs of variables not in R_{E_i} only, while $g(x)$ and $h(x)$ contain the parameters from the CPTs of the variables in R_{E_i}.

Furthermore we know that

$$\begin{aligned}
\Pr\left(C = \bar{c} \,\middle|\, e_i^k, \mathbf{e}^-\right) &< \Pr\left(C = \bar{c} \,\middle|\, e_i^{k+1}, \mathbf{e}^-\right) \\
&\Updownarrow \\
\Pr\left(C = \bar{c}, e_i^k, \mathbf{e}^-\right) \cdot \Pr\left(e_i^{k+1}, \mathbf{e}^-\right) &< \Pr\left(C = \bar{c}, e_i^{k+1}, \mathbf{e}^-\right) \cdot \Pr\left(e_i^k, \mathbf{e}^-\right).
\end{aligned}$$

Since

$$\begin{aligned}
\Pr\left(C=\bar{c},e_i^k,\mathbf{e}^-\right)\cdot\Pr\left(e_i^{k+1},\mathbf{e}^-\right) &= f(\bar{c})\cdot g(\bar{c})\cdot(f(\bar{c})\cdot h(\bar{c})+f(c)\cdot h(c)) \\
&= f(\bar{c})\cdot f(\bar{c})\cdot g(\bar{c})\cdot h(\bar{c})+f(\bar{c})\cdot f(c)\cdot g(\bar{c})\cdot h(c)
\end{aligned}$$

and

$$\begin{aligned}
\Pr\left(C=\bar{c},e_i^{k+1},\mathbf{e}^-\right)\cdot\Pr\left(e_i^{k},\mathbf{e}^-\right) &= f(\bar{c})\cdot h(\bar{c})\cdot(f(\bar{c})\cdot g(\bar{c})+f(c)\cdot g(c)) \\
&= f(\bar{c})\cdot f(\bar{c})\cdot h(\bar{c})\cdot g(\bar{c})+f(\bar{c})\cdot f(c)\cdot h(\bar{c})\cdot g(c),
\end{aligned}$$

it follows that

$$\begin{array}{rcl}
\Pr\left(C=\bar{c},e_i^k,\mathbf{e}^-\right)\cdot\Pr\left(e_i^{k+1},\mathbf{e}^-\right) & < & \Pr\left(C=\bar{c},e_i^{k+1},\mathbf{e}^-\right)\cdot\Pr\left(e_i^{k},\mathbf{e}^-\right) \\
 & \Updownarrow & \\
g(\bar{c})\cdot h(c) & < & h(\bar{c})\cdot g(c).
\end{array}$$

Since this inequality does not contain any terms involving $f(x)$, varying parameters from the CPTs of variables not in the resolution set of E_i cannot resolve the violation of monotonicity □

Lemma 4.3.2 is restricted to Bayesian networks with $|\Omega(C)| = 2$ only. The following example demonstrates the necessity of this restriction.

Example 2. *Suppose that* $\Omega(C) = \{c^1, c^2, c^3\}$ *and that we have a violation of monotonicity* *viol*$\left(c_1, e_i^k, \mathbf{e}^-\right)$. *Using the same functions as in the proof of Lemma 4.3.2, we have*

$$\begin{aligned}
&\Pr\left(C=c^1,e_i^k,\mathbf{e}^-\right)\cdot\Pr\left(e_i^{k+1},\mathbf{e}^-\right) \\
&\qquad = f(c^1)\cdot g(c^1)\cdot\left(f(c^1)\cdot h(c^1)+f(c^2)\cdot h(c^2)+f(c^3)\cdot h(c^3)\right)
\end{aligned}$$

and

$$\begin{aligned}
&\Pr\left(C=c^1,e_i^{k+1},\mathbf{e}^-\right)\cdot\Pr\left(e_i^{k},\mathbf{e}^-\right) \\
&\qquad = f(c^1)\cdot h(c^1)\cdot\left(f(c^1)\cdot g(c^1)+f(c^2)\cdot g(c^2)+f(c^3)\cdot g(c^3)\right).
\end{aligned}$$

It follows that

$$\begin{array}{rcl}
\Pr\left(C=c^1 \,\middle|\, e_i^k,\mathbf{e}^-\right) & < & \Pr\left(C=c^1 \,\middle|\, e_i^{k+1},\mathbf{e}^-\right) \\
 & \Updownarrow & \\
g(c^1)\cdot\left(f(c^2)\cdot h(c^2)+f(c^3)\cdot h(c^3)\right) & < & h(c^1)\left(f(c^2)\cdot g(c^2)+f(c^3)\cdot g(c^3)\right).
\end{array}$$

Since this inequality contains $f(x)$, *varying parameters from the CPTs of variables not in the resolution set of* E_i *may resolve the violation of monotonicity.* □

Since we are only varying a single parameter from a single CPT at a time, the set of variables for which varying a parameter may restore monotonicity is restricted by the resolution sets of all $E_i \in \mathbf{E}$ for which there are violations of monotonicity.

Theorem 4.3.3. *Let B be a Bayesian network with a binary variable of interest C and a set of observable variables $\mathbf{E}$. Now, let $\mathbf{E}_V$ be the set of observable variables for which there are violations of monotonicity. Then, only varying a parameter from the CPT of a variable in*

$$\mathbf{R} = \bigcap_{E \in \mathbf{E}_V} R_E$$

may restore monotonicity.

Proof. Suppose that there exists a variable $V \notin \mathbf{R}$ such that varying a parameter for V restores monotonicity. Since $V \notin \mathbf{R}$, there is an $E_i \in \mathbf{E}_V$ such that $V \notin R_{E_i}$. By Lemma 4.3.2 we have, however, that varying a single parameter from the CPT of V cannot resolve any violation of monotonicity $viol\big(c, e_i^k, \mathbf{e}^-\big)$ with $e_i^k \in \Omega(E_i)$. Varying a parameter for V thus cannot restore monotonicity. □

Observe that $\mathbf{R}$ can easily be obtained by removing the outgoing arcs from all observable variable $\mathbf{E}$ and the variable of interest C. Note that Theorem 4.3.3 applies to Bayesian networks with a binary variable of interest only. Such Bayesian networks are fairly common.

4.4 Example

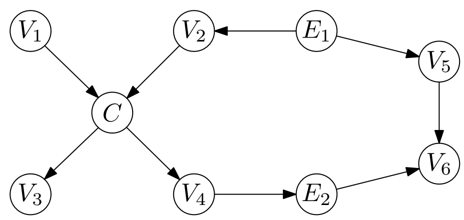

Figure 4.4.2: DAG of the unrestricted Bayesian network of the example.

Let $B = (G, P)$ be a Bayesian network with its DAG G as depicted in Figure 4.4.2. Let C be its variable of interest and let $\mathbf{E} = \{E_1, E_2\}$ be its set of observable variables. Using Method 1 we restrict this Bayesian network by removing all vertices not in $Sen(C, \mathbf{E}) \cup \mathbf{E}$ and their incident arcs. Since $Sen(C, \mathbf{E}) = \{C, V_1, V_2, V_4, E_2\}$, this means we remove V_3, V_5 and V_6 and their incident arcs from G. The resulting DAG is depicted in Figure 4.4.3.

Since E_1 is not in $Sen(C, \mathbf{E})$, we know that we cannot restore monotonicity by varying a parameter from its CPT. If C is a binary variable, then by Theorem 4.3.3 we have that only varying a parameter from the CPT of V_4 or E_2 may restore monotonicity, since $R_{E_2} = \{V_4, E_2\}$.

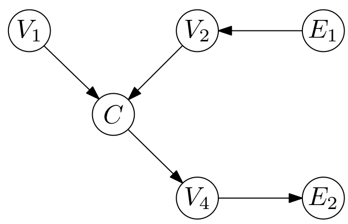

Figure 4.4.3: DAG of the restricted Bayesian network of the example.

Chapter 5

The intersection-of-intervals approach

In the previous chapter we introduced two concepts which serve to restrict the set of variables for which varying a parameter may restore monotonicity. For each of the remaining variables we must still determine whether varying a parameter from its CPT can actually restore monotonicity. To this end we introduce a method which determines whether varying a specific parameter from a variable's CPT can resolve all violations of monotonicity with respect to two given assignments for the observable variables. Using this method we can then determine whether monotonicity can be restored.

5.1 The method

We consider a restricted Bayesian network $B = (G, P)$ with a variable of interest C and a set of observable variables $\mathbf{E}$. Suppose there is a violation of monotonicity $viol\big(c^j, e_i^k, \mathbf{e}^-\big)$ in B. To resolve this violation by varying a single parameter, we must vary a parameter $p\,(u \,|\, \pi)$ from the CPT of a variable $U \in V(G)$ to a value $x \in [0,1]$ such that

$$\Pr\left(C \leq c^j \,\middle|\, e_i^k, \mathbf{e}^-\right)\big(p\,(u \,|\, \pi) = x\big) \geq \Pr\left(C \leq c^j \,\middle|\, e_i^{k+1}, \mathbf{e}^-\right)\big(p\,(u \,|\, \pi) = x\big). \tag{5.1}$$

Note that doing so may cause a new violation $viol\big(c, e_i^k, \mathbf{e}^-\big)$ for some $c \in \Omega(C)\backslash\{c^j\}$. Varying $p\,(u \,|\, \pi)$ can only resolve all violations of monotonicity $viol\big(c, e_i^k, \mathbf{e}^-\big)$, $c \in \Omega(C)$, if there is a value $x \in [0,1]$ for $p\,(u \,|\, \pi)$ such that

$$\Pr\left(C \leq c \,\middle|\, e_i^k, \mathbf{e}^-\right)\big(p\,(u \,|\, \pi) = x\big) \geq \Pr\left(C \leq c \,\middle|\, e_i^{k+1}, \mathbf{e}^-\right)\big(p\,(u \,|\, \pi) = x\big)$$

for all $c \in \Omega(C)$. These observations give rise to the following method, called the *intersection-of-intervals approach*. This method determines whether varying a parameter $p\,(u \,|\, \pi)$ from the CPT of a variable U can resolve all violations of monotonicity with respect to the two assignments $(e_i^k, \mathbf{e}^-), (e_i^{k+1}, \mathbf{e}^-) \in \Omega(E_i) \times \Omega(\mathbf{E}^-)$.

Method 2 (Intersection-of-intervals approach)**.** *Let $I = [0,1]$ and $l = 1$. Repeat the following steps while $l < |\Omega(C)|$ and $I \neq \emptyset$:*

1. *Compute I_l, which is the set of all intervals of values x for $p\,(u \,|\, \pi)$ for which*

$$0 \leq \sum_{j=1}^{l} \left(\Pr\left(c^j \,\middle|\, e_i^k, \mathbf{e}^-\right)\big(p\,(u \,|\, \pi) = x\big) - \Pr\left(c^j \,\middle|\, e_i^{k+1}, \mathbf{e}^-\right)\big(p\,(u \,|\, \pi) = x\big)\right).$$

2. *Compute* $I = I \cap I_l$ *and* $l = l + 1$.

The result I of Method 2 is a set of intervals of values for $p\,(u \mid \pi)$ for which there is no violation $viol\big(c, e_i^k, \mathbf{e}^-\big)$ for any $c \in \Omega(C)$. Note that for $j = |\Omega(C)|$ equation (5.1) is always an equality, since the probabilities on both sides equal 1.

Lemma 5.1.1. *Let* $B = (G, P)$ *be a Bayesian network with a variable of interest* C *and a set of observable variables* $\mathbf{E}$. *Let* I *be the result of the intersection-of-intervals approach when applied to* $e_i^k, e_i^{k+1} \in \Omega(E_i)$ *and* $\mathbf{e}^- \in \Omega(\mathbf{E}^-)$. *Then,* $I \neq \emptyset$ *if and only if varying* $p\,(u \mid \pi)$ *can resolve all violations of monotonicity with respect to the value assignments* $(e_i^k, \mathbf{e}^-), (e_i^{k+1}, \mathbf{e}^-) \in \Omega(E_i) \times \Omega(\mathbf{E}^-)$.

Proof. Since $\Pr\big(C \leq c^l \mid e_i, \mathbf{e}^-\big)\,(p\,(u \mid \pi) = x) = \sum_{j=1}^{l} \Pr\big(c^j \mid e_i, \mathbf{e}^-\big)\,(p\,(u \mid \pi) = x)$ for all $e_i \in \Omega(E_i)$ for all $(e_i, \mathbf{e}^-) \in \Omega(E_i) \times \Omega(\mathbf{E}^-)$, the result I_l contains the values x for which inequality (5.1) holds.

At initialization, I equals the probability interval $[0, 1]$ containing all possible values x for $p\,(u \mid \pi)$. Suppose that after $l - 1$ iterations of steps 1 and 2 of Method 2, I contains all values x for which inequality (5.1) holds for all $c^j \in \{c^1, \ldots, c^{l-1}\}$. Since I_l contains all values x for which inequality (5.1) holds for $c^j = c^l$ and I contains all values x for which inequality (5.1) holds for all $c^j \in \{c^1, \ldots, c^{l-1}\}$, the intersection of I and I_l contains only values x for which inequality (5.1) holds for all $c^j \in \{c^1, \ldots, c^l\}$.

If $I \neq \emptyset$, then inequality (5.1) holds for all $c^j \in \Omega(C)$ for all values $x \in I$, which means that varying $p\,(u \mid \pi)$ can resolve all violations of monotonicity with respect to the value assignments $(e_i^k, \mathbf{e}^-), (e_i^{k+1}, \mathbf{e}^-) \in \Omega(E_i) \times \Omega(\mathbf{E}^-)$. If $I = \emptyset$, then there is no value x such that inequality (5.1) holds for all $c^j \in \Omega(C)$, which means that varying $p\,(u \mid \pi)$ cannot resolve all violations of monotonicity with respect to the value assignments $(e_i^k, \mathbf{e}^-), (e_i^{k+1}, \mathbf{e}^-) \in \Omega(E_i) \times \Omega(\mathbf{E}^-)$. □

5.2 Computation and complexity

We study the computations involved and the complexity of the two steps of the intersection-of-intervals approach separately. We recall that the first step of Method 2 requires the computation of intervals, while the second step requires the computation of intersections of intervals.

5.2.1 Computing the intervals

To compute the set of intervals I_l we compute the real values x for which

$$\sum_{j=1}^{l} \left(\Pr\left(c^j \;\middle|\; e_i^k, \mathbf{e}^-\right)\big(p\,(u \mid \pi) = x\big) - \Pr\left(c^j \;\middle|\; e_i^{k+1}, \mathbf{e}^-\right)\big(p\,(u \mid \pi) = x\big) \right) = 0, \qquad (5.2)$$

since these are endpoints of the intervals in I_l. As proposed in Proposition 2.2.6, we have that

$$\begin{aligned} \Pr\left(c^j \mid e_i^k, \mathbf{e}^-\right)\big(p\,(u \mid \pi) = x\big) &= \frac{\alpha_j x + \beta_j}{\gamma x + \delta} \quad \text{and} \\ \Pr\left(c^j \;\middle|\; e_i^{k+1}, \mathbf{e}^-\right)\big(p\,(u \mid \pi) = x\big) &= \frac{\alpha'_j x + \beta'_j}{\gamma' x + \delta'}, \end{aligned} \qquad (5.3)$$

where $\alpha_j, \alpha'_j, \beta_j, \beta'_j, \gamma, \gamma', \delta$ and δ' are constants, which can be computed with the algorithm from [7]. Using this notation, we rewrite equality (5.2) to

$$\frac{\sum_{j=1}^{l}(\alpha_j x + \beta_j)}{\gamma x + \delta} - \frac{\sum_{j=1}^{l}(\alpha'_j x + \beta'_j)}{\gamma' x + \delta'} = 0.$$

This equality only holds if

$$\sum_{j=1}^{l}(\alpha_j x + \beta_j)(\gamma' x + \delta') - \sum_{j=1}^{l}(\alpha'_j x + \beta'_j)(\gamma x + \delta) = 0. \tag{5.4}$$

Computing the intervals in I_l thus consists of solving a quadratic equation and determining on which side of the real solutions inequality (5.1) holds; this can be done in constant time using the quadratic formula. However to obtain the quadratic equation we must compute the constants $\alpha_j, \alpha'_j, \beta_j$ and β'_j, for $j = 1, \ldots, l$, as well as the constants γ, γ', δ and δ'. An algorithm to compute these constants is given in [7]. The complexity of step 1 of the intersection-of-intervals approach is therefore bound by the complexity of that algorithm. Note that for each iteration of the intersection-of-intervals approach we need only compute the set of constants for $\Pr\left(c^l \mid e_i^k, \mathbf{e}^-\right)\left(p\left(u \mid \pi\right) = x\right)$ and $\Pr\left(c^l \,\middle|\, e_i^{k+1}, \mathbf{e}^-\right)\left(p\left(u \mid \pi\right) = x\right)$; all other constants were already computed in the previous iterations. Since the quadratic equation (5.4) can have two solutions within the interval $[0, 1]$, I_l consists of up to two intervals.

5.2.2 Computing the intersection of intervals

In general, computing the intersection of two intervals is simple. It can be done in constant time by comparing the endpoints of the two intervals and taking the correct ones as endpoints for the resulting interval, which may be the empty set. As argued above, however, the quadratic equation (5.4) may have two real solutions in the interval $[0, 1]$. In step 2 of the intersection-of-intervals approach, therefore the set of intervals I_l may consist of two intervals in which inequality (5.1) holds. I itself may already consist of up to l intervals, since each iteration can only increase the number of intervals in I by 1, as is illustrated in Figure 5.2.1. Therefore, taking the intersection of I and I_l consists of comparing the endpoints of each of the intervals in I with the endpoints of each of the intervals in I_l, which amounts to computing at most $2l$ intersections of two intervals in each iteration of the intersection-of-intervals approach. Since l is incremented by 1 in each iteration, the whole method requires taking at most $m(m + 1)$ intersections, where m is the number of possible values for the variable of interest. Thus, in the worst case the complexity of step 2 of the intersection-of-intervals approach is quadratic in $|\Omega(C)|$.

There are cases, however, in which the set of intervals I_l consists of at most one interval, namely if the coefficient of the quadratic term in equation (5.4) equals 0. Note that the coefficient of the quadratic term is

$$\sum_{j=1}^{l}\alpha_j\gamma' - \sum_{j=1}^{l}\alpha'_j\gamma,$$

which is equal to 0 in the following five cases:

1. γ and γ' are both equal to 0;

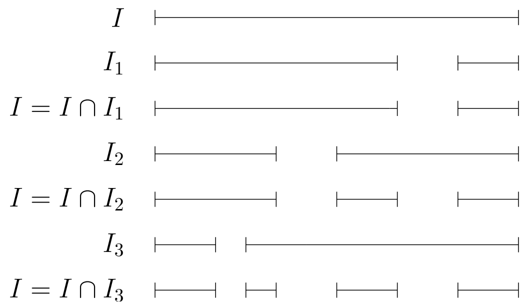

Figure 5.2.1: Example of intersections of intervals.

2. α_j and α'_j are both equal to 0 for all $j = 1, \ldots, l$;

3. there is an $m \in \{1, \ldots, l\}$ such that $\alpha_m = \gamma$ and $\alpha_j = 0$ for all $j \neq m$, and there is an $m' \in \{1, \ldots, l\}$ such that $\alpha'_{m'} = \gamma'$ and $\alpha'_j = 0$ for all $j \neq m'$;

4. for every m there is a distinct m' such that $\alpha_m \gamma' = \alpha'_{m'} \gamma$;

5. $\sum_{j=1}^{l} \alpha_j \gamma' = \sum_{j=1}^{l} \alpha'_j \gamma$ by coincidence, due to the values in the CPTs.

The fifth case overlaps with the first four cases, since the values in the CPTs can be such that these situations occur. However, there are also graphical condtions under which the first four cases occur. We will examine these conditions in the following lemmas, where we assume that we have a restricted Bayesian network B with a variable of interest C and a set of observable variables $\mathbf{E}$. Furthermore, we assume that we are varying a parameter $p\,(u \mid \pi)$ from the CPT of a variable U.

The first lemma applies to case 1.

Lemma 5.2.1. *If $\sigma^*(U) \cap \mathbf{E} = \emptyset$, then $\gamma = \gamma' = 0$.*

Proof. From equations (5.3) we have that $\Pr(e_i, \mathbf{e}^-)$ is linear in the parameter $p\,(u \mid \pi)$ for every $(e_i, \mathbf{e}^-) \in \Omega(E_i) \times \Omega(\mathbf{E}^-)$. We now observe that, if $\sigma^*(U) \cap \mathbf{E} = \emptyset$, then $\sigma^*(X_U) \cap \mathbf{E} = \emptyset$, where X_U is the auxiliary predecessor of U in G^*. It follows that $\langle \{X_U\} | \, \emptyset \, | \mathbf{E} \rangle^d_{G^*}$, which, by Lemma 4.1.2, means that $\Pr(e_i, \mathbf{e}^-) \nsim p\,(u \mid \pi)$ for all $(e_i, \mathbf{e}^-) \in \Omega(E_i) \times \Omega(\mathbf{E}^-)$. Thus if $\sigma^*(U) \cap \mathbf{E} = \emptyset$, then $\Pr\left(e_i^k, \mathbf{e}^-\right) \nsim p\,(u \mid \pi)$ and $\Pr\left(e_i^{k+1}, \mathbf{e}^-\right) \nsim p\,(u \mid \pi)$. We conclude that γ and γ' are equal to 0. $\square$

There are no graphical conditions under which the second case occurs in the restricted Bayesian network. The graphical condition under which α_j and α'_j are both equal to 0 is $\sigma^*(U) \cap \mathbf{E} \cap \{C\} = \emptyset$. We then have that $\sigma^*(X_U) \cap \mathbf{E} \cap \{C\} = \emptyset$ in G^*, which means that $\langle \{X_U\} | \, \emptyset \, | \mathbf{E} \cup \{C\} \rangle^d_{G^*}$. By Lemma 4.1.2, that means that $\Pr\left(c^j, e_i, \mathbf{e}^-\right) \nsim p\,(u \mid \pi)$ for all $(e_i, \mathbf{e}^-) \in \Omega(E_i) \times \Omega(\mathbf{E}^-)$ and, therefore, that α_j and α'_j are both equal to 0. However, by Definition 4.1.1, $\langle \{X_U\} | \, \emptyset \, | \mathbf{E} \cup \{C\} \rangle^d_{G^*}$ also means that $U \notin Sen(C, \mathbf{E})$. Thus $\sigma^*(U) \cap \mathbf{E} \cap \{C\} \neq \emptyset$ for all variables U in the restricted network.

The following lemma pertains to the third case.

Lemma 5.2.2. *If $C \in \pi(U)$, then there is an $m \in \{1, \ldots, l\}$ such that $\alpha_m = \gamma$ and $\alpha_j = 0$ for all $j \neq m$, and there is an $m' \in \{1, \ldots, l\}$ such that $\alpha'_{m'} = \gamma'$ and $\alpha'_j = 0$ for all $j \neq m'$.*

Proof. The constants α_j and γ are defined in [3]. We have adapted them to non-binary variables as follows.

$$\alpha_j = \sum_{\Omega(V(G))} \prod_{\substack{V\in V(G)\\ V\neq U}} p\left(V \mid \pi(V)\right)\Bigg|_{\substack{\mathbf{E}^-=\mathbf{e}^-\\ E_i=e_i^k\\ C=c^j\\ \pi(U)=\pi\\ U=u}} - \sum_{\substack{\Omega(V(G))\\ U\neq u}} \frac{p\left(U \mid \pi\right)}{1-p\left(u \mid \pi\right)} \prod_{\substack{V\in V(G)\\ V\neq U}} p\left(V \mid \pi(V)\right)\Bigg|_{\substack{\mathbf{E}^-=\mathbf{e}^-\\ E_i=e_i^k\\ C=c^j\\ \pi(U)=\pi}} \tag{5.5}$$

$$\gamma = \sum_{\Omega(V(G))} \prod_{\substack{V\in V(G)\\ V\neq U}} p\left(V \mid \pi(V)\right)\Bigg|_{\substack{\mathbf{E}^-=\mathbf{e}^-\\ E_i=e_i^k\\ \pi(U)=\pi\\ U=u}} - \sum_{\substack{\Omega(V(G))\\ U\neq u}} \frac{p\left(U \mid \pi\right)}{1-p\left(u \mid \pi\right)} \prod_{\substack{V\in V(G)\\ V\neq U}} p\left(V \mid \pi(V)\right)\Bigg|_{\substack{\mathbf{E}^-=\mathbf{e}^-\\ E_i=e_i^k\\ \pi(U)=\pi}} \tag{5.6}$$

The constants α'_j and γ' are defined similarly. Now, suppose that $C \in \pi(U)$. If c^j is the value assigned to C in π, then $\alpha_j = \gamma$. If c^j is not the value assigned to C in π, then the conditions $C = c^j$ and $\pi(U) = \pi$ contradict each other, and therefore $\alpha_j = 0$. The same arguments hold for α'_j and γ'. Since C has a single specific value $c^j \in \Omega(C)$ assigned to it in π, there is an $m \in \{1, \dots, l\}$ such that $\alpha_m = \gamma$ and $\alpha_j = 0$ for all $j \neq m$, and there is an $m' \in \{1, \dots, l\}$ such that $\alpha'_{m'} = \gamma'$ and $\alpha'_j = 0$ for all $j \neq m'$. Moreover, $m = m' = j$. □

In the last lemma we study the conditions under which the fourth case occurs.

Lemma 5.2.3. *Let $B = (G, P)$ be a Bayesian network with a variable of interest C and a set of observable variables $\mathbf{E}$, and let U be the variable in the CPT of which we are varying. Let G' be the DAG constructed from G by removing all outgoing arcs from the variables $\mathbf{E}\backslash\{E_i\}$ and $\pi(U)$. Let S_C be the connected component in G' containing C, and let S_{E_i} be the connected component in G' containing E_i. If $S_C \neq S_{E_i}$, then for every m there is a distinct $m' \in \{1, \dots, l\}$ such that $\alpha_m\gamma' = \alpha'_{m'}\gamma$.*

Proof. Let $S_1, \dots, S_q$ be the connected components in G', which do not contain E_i or C. We assume that $S_C \neq S_{E_i}$. We can compute $\alpha_j, \alpha'_j, \gamma$ and γ' from G using the sets of variables $V(S_1), \dots, V(S_q)$. Note that $\cup_{n=1}^{q} V(S_n) = V(G)$. We can rewrite α_j from equation (5.5) and γ from equation (5.6) as follows, where

$$\delta(S_x, U) = \begin{cases} 1 & \text{if } U \notin V(S_x), \\ 1 & \text{if } U = u, \\ -\frac{p(U \mid \pi)}{1-p(u \mid \pi)} & \text{otherwise.} \end{cases}$$

$$\alpha_j = \prod_{n=1}^{q} \sum_{\Omega(V(S_n))} \delta(S_n, U) \prod_{\substack{V \in V(S_n) \\ V \neq U}} p\left(V \mid \pi(V)\right)\Bigg|_{\substack{\mathbf{E}^- = \mathbf{e}^- \\ \pi(U) = \pi}} \cdot$$
$$\sum_{\Omega(V(S_C))} \delta(S_C, U) \prod_{\substack{V \in V(S_C) \\ V \neq U}} p\left(V \mid \pi(V)\right)\Bigg|_{\substack{\mathbf{E}^- = \mathbf{e}^- \\ \pi(U) = \pi \\ C = c^j}} \cdot$$
$$\sum_{\Omega(V(S_{E_i}))} \delta(S_{E_i}, U) \prod_{\substack{V \in V(S_{E_i}) \\ V \neq U}} p\left(V \mid \pi(V)\right)\Bigg|_{\substack{\mathbf{E}^- = \mathbf{e}^- \\ \pi(U) = \pi \\ E_i = e_i^k}} \tag{5.7}$$

$$\gamma = \prod_{n=1}^{q} \sum_{\Omega(V(S_n))} \delta(S_n, U) \prod_{\substack{V \in V(S_n) \\ V \neq U}} p\left(V \mid \pi(V)\right)\Bigg|_{\substack{\mathbf{E}^- = \mathbf{e}^- \\ \pi(U) = \pi}} \cdot$$
$$\sum_{\Omega(V(S_C))} \delta(S_C, U) \prod_{\substack{V \in V(S_C) \\ V \neq U}} p\left(V \mid \pi(V)\right)\Bigg|_{\substack{\mathbf{E}^- = \mathbf{e}^- \\ \pi(U) = \pi}} \cdot$$
$$\sum_{\Omega(V(S_{E_i}))} \delta(S_{E_i}, U) \prod_{\substack{V \in V(S_{E_i}) \\ V \neq U}} p\left(V \mid \pi(V)\right)\Bigg|_{\substack{\mathbf{E}^- = \mathbf{e}^- \\ \pi(U) = \pi \\ E_i = e_i^k}} \tag{5.8}$$

The constants α'_j and γ' are defined similarly.

Observe that α_j and γ only differ in the terms regarding S_C. The same also holds for α'_j and γ'. Furthermore, α_j and α'_j only differ in the terms regarding S_{E_i}, and the same holds for γ and γ'. Therefore $\alpha_j\gamma'$ contains exactly the same terms as $\alpha'_j\gamma$, but in a different order. Thus if $S_C \neq S_{E_i}$, then for every m there is a distinct m' such that $\alpha_m\gamma' = \alpha'_{m'}\gamma$, namely $m' = m$. □

The following theorem combines Lemmas 5.2.1, 5.2.2 and 5.2.3 to describe graphical conditions under which a set of intervals I_i from Method 2 will consist of a single interval if any.

Theorem 5.2.4. *Let $B = (G, P)$ be a Bayesian network with a variable of interest C and a set of observable variables $\mathbf{E}$, and let U be the variable in the CPT of which we are varying. For all $j \in \{1, \ldots, |\Omega(C)|\}$, the set of intervals I_j in Method 2 consist of at most one interval if one of the following conditions holds:*

- $\sigma^*(U) \cap \mathbf{E} = \emptyset$;
- $C \in \pi(U)$;

- *$S_C \neq S_{E_i}$, where S_C and S_{E_i} are as defined in Lemma 5.2.3.*

Proof. By Lemmas 5.2.1, 5.2.2 and 5.2.3 the conditions described in the theorem are conditions under which the coefficient of the quadratic term in the quadratic equation (5.4) equals 0. If the coefficient of the quadratic term equals 0, then I_l consists of at most one interval. □

Note that if the conditions in Theorem 5.2.4 hold, then in each iteration of the intersection-of-intervals approach step 2 takes constant time. Thus for variables for which the conditions in Theorem 5.2.4 holds, the complexity of step 2 of the intersection-of-intervals approach is linear in $|\Omega(C)|$.

5.3 Using the intersection-of-intervals approach to restore monotonicity

To determine whether monotonicity can be restored by varying a single parameter, we must find a parameter $p\,(u \mid \pi)$ for which there is a value x such that $x \in I$ by the intersection-of-intervals approach for all combinations of value assignments $(e_i^k, \mathbf{e}^-), (e_i^{k+1}, \mathbf{e}^-) \in \Omega(E_i) \times \Omega(\mathbf{E}^-)$ for all $E_i \in \mathbf{E}$. The number of parameters in the restricted Bayesian network $B = (G, P)$ is exponential in $|V(G)|$. The number of combinations $(e_i^k, \mathbf{e}^-), (e_i^{k+1}, \mathbf{e}^-)$ in $\Omega(\mathbf{E})$ is $\sum_{E_i \in \mathbf{E}}(|\Omega(E_i)| - 1)|\Omega(\mathbf{E} \backslash \{E_i\})|$, which is $O(|\mathbf{E}| \cdot |\Omega(\mathbf{E})|)$. Since $|\Omega(\mathbf{E})|$ is exponential in $|\mathbf{E}|$ and $|\mathbf{E}| < |V(G)|$, we have that worst case the intersection-of-intervals approach must be applied an exponential number of times in $|V(G)|$. In practice, however, we need only continue applying the intersection-of-intervals approach for a parameter so long as the resulting intervals overlap.

In general, if monotonicity cannot be restored by varying a single parameter, then it may still be possible to do so by varying multiple parameters. One possibility is varying a sequence of parameters until monotonicity is restored. We might, for instance, vary the parameter which resolves the most violations of monotonicity and then try to restore monotonicity in the Bayesian network with the new value for that parameter. This method, however, doesn't necessarily result in the optimal sequence of parameters to vary. Another possibility is varying multiple parameters at the same time. These possibilities are, however, beyond the scope of this thesis.

5.4 Example

To illustrate using the intersection of intervals approach we use the same Bayesian network as in Example 1. This Bayesian network $B = (G, P)$ has a DAG as depicted in Figure 5.4.2, a variable of interest C and a set of observable variables $\mathbf{E} = \{E_1\}$.

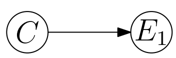

Figure 5.4.2: Directed acyclic graph used in Example 1.

Again we let $\Omega(E_1) = \{\bar{e}_1, e_1\}$ and $\Omega(C) = \{c^1, c^2, c^3\}$, and let the parameters from the CPTs of C and E_1 be as follows:

$$p\left(c^1\right) = 0.5 \qquad p\left(c^2\right) = 0.3 \qquad p\left(c^3\right) = 0.2$$
$$p\left(\bar{e}_1 \mid c^1\right) = 0.6 \qquad p\left(\bar{e}_1 \mid c^2\right) = 0.8 \qquad p\left(\bar{e}_1 \mid c^3\right) = 0.7$$

We first use the intersection-of-intervals approach to investigate if varying parameter $p\left(\bar{e}_1 \,|\, c^1\right)$ can restore monotonicity. We begin with $I = [0,1]$, and we have that

$$\Pr\left(C \leq c^1 \mid \bar{e}_1\right)\left(p\left(\bar{e}_1 \,|\, c^1\right) = x\right) = \frac{0.5x}{0.5x + 0.38}$$

and

$$\Pr\left(C \leq c^1 \mid e_1\right)\left(p\left(\bar{e}_1 \,|\, c^1\right) = x\right) = \frac{-0.5x + 0.5}{-0.5x + 0.62}.$$

It follows that

$$0 \leq \sum_{j=1}^{1}\left(\Pr\left(c^j \mid \bar{e}_1\right)\left(p\left(\bar{e}_1 \,|\, c^1\right) = x\right) - \Pr\left(c^j \mid e_1\right)\left(p\left(\bar{e}_1 \,|\, c^1\right) = x\right)\right)$$

only for $x \notin [0,1]$. Thus after the first iteration $I = \emptyset$, which means that varying parameter $p\left(\bar{e}_1 \,|\, c^1\right)$ cannot restore monotonicity.

Next we use the intersection-of-intervals approach to investigate if varying parameter $p\left(\bar{e}_1 \,|\, c^2\right)$ can restore monotonicity. We again begin with $I = [0,1]$, and we have that

$$\Pr\left(C \leq c^1 \mid \bar{e}_1\right)\left(p\left(\bar{e}_1 \,|\, c^2\right) = x\right) = \frac{0.3}{0.3x + 0.44}$$

and

$$\Pr\left(C \leq c^1 \mid e_1\right)\left(p\left(\bar{e}_1 \,|\, c^2\right) = x\right) = \frac{0.2}{-0.3x + 0.56}.$$

It follows that

$$0 \leq \sum_{j=1}^{1}\left(\Pr\left(c^j \mid \bar{e}_1\right)\left(p\left(\bar{e}_1 \,|\, c^2\right) = x\right) - \Pr\left(c^j \mid e_1\right)\left(p\left(\bar{e}_1 \,|\, c^2\right) = x\right)\right)$$

for all $x \in [0,1]$. Thus $I_1 = [0,1]$ and after the first iteration I is still $[0,1]$. For the following iteration we have that

$$\Pr\left(C \leq c^2 \mid \bar{e}_1\right)\left(p\left(\bar{e}_1 \,|\, c^2\right) = x\right) = \frac{0.3x + 0.3}{0.3x + 0.44}$$

and

$$\Pr\left(C \leq c^2 \mid e_1\right)\left(p\left(\bar{e}_1 \,|\, c^2\right) = x\right) = \frac{-0.3x + 0.5}{-0.3x + 0.56}.$$

It follows that

$$0 \leq \sum_{j=1}^{2}\left(\Pr\left(c^j \mid \bar{e}_1\right)\left(p\left(\bar{e}_1 \,|\, c^2\right) = x\right) - \Pr\left(c^j \mid e_1\right)\left(p\left(\bar{e}_1 \,|\, c^2\right) = x\right)\right)$$

for all $x \in \left[\frac{13}{15}, 1\right]$. Thus $I_2 = \left[\frac{13}{15}, 1\right]$ and after the second, and final, iteration I is $\left[\frac{13}{15}, 1\right]$. Since $\bar{e}_1, e_1$ is the only combination of value assignments in $\mathbf{E}$, we have used the intersection-of-intervals approach on all combinations of value assignments. Thus varying parameter $p\left(\bar{e}_1 \,|\, c^2\right)$ to a value $x \in \left[\frac{13}{15}, 1\right]$ will restore monotonicity.

Chapter 6

Application of the intersection-of-intervals approach

There are cases in which we need not apply the intersection-of-intervals approach to any parameter from the CPT of a variable W since we already know that parameter variation cannot resolve a violation of monotonicity based upon investigation of some other variable U cannot resolve that violation of monotonicity. By identifying these cases we can ensure that the intersection-of-intervals approach is not applied needlessly. In this chapter we investigate these cases, culminating in a general case. For each case we will show that if we cannot resolve a violation of monotonicity by varying any parameter from the CPT of some variable U, then we also cannot resolve it by varying any parameter from the CPT of a specific variable W.

6.1 A simple case

To illustrate the above considerations, we begin with a simple case, which will also allow us to make some assumptions for the general case.

We consider an arbitrary (restricted) Bayesian network $B = (G, P)$ in which there are two vertices $U, W \in V(G)$ such that W has no predecessors, U is its only successor, and W is the only predecessor of U; a graphical representation of such a Bayesian network is depicted in Figure 6.1.1. We assume that neither U nor W equals the variable of interest C or contains any observation, that is, we assume that $\{U, W\} \cap (\{C\} \cup \mathbf{E}) = \emptyset$.

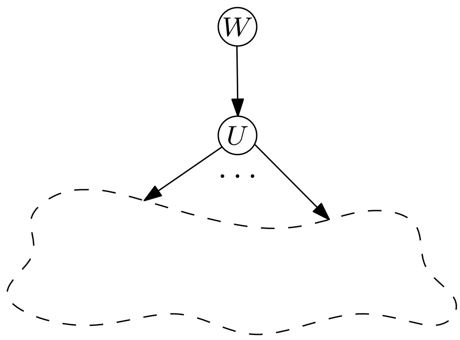

Figure 6.1.1: An arbitrary restricted Bayesian network in which there are two vertices U, W such that W has no predecessors, U is its only successor, and W is the only predecessor of U.

Now suppose that there is a violation of monotonicity $viol\left(c', e_i^k, \mathbf{e}^-\right)$ in this Bayesian network. Thus we have that

$$\Pr\left(C \leq c' \;\middle|\; e_i^k, \mathbf{e}^-\right) < \Pr\left(C \leq c' \;\middle|\; e_i^{k+1}, \mathbf{e}^-\right),$$

which can also be written as

$$\sum_{\substack{c \in \Omega(C) \\ c \leq c'}} \frac{\Pr\left(c, e_i^k, \mathbf{e}^-\right)}{\Pr\left(e_i^k, \mathbf{e}^-\right)} < \sum_{\substack{c \in \Omega(C) \\ c \leq c'}} \frac{\Pr\left(c, e_i^{k+1}, \mathbf{e}^-\right)}{\Pr\left(e_i^{k+1}, \mathbf{e}^-\right)}.$$

For the Bayesian network B we have for all $c \in \Omega(C)$ and $\mathbf{e} \in \Omega(\mathbf{E})$ that

$$\begin{aligned}
\Pr(c, \mathbf{e}) &= \sum_{\Omega(V(G))} \prod_{V \in V(G)} p\left(V \mid \pi(V)\right)\Bigg|_{\substack{C = c \\ \mathbf{E} = \mathbf{e}}} \\
&= \sum_{\Omega(U)} \left(\Pr(U) \cdot \sum_{\Omega(V(G) \setminus \{U\})} \prod_{\substack{V \in V(G) \\ V \neq U, V \neq W}} p\left(V \mid \pi(V)\right)\Bigg|_{\substack{C = c \\ \mathbf{E} = \mathbf{e}}} \right) \quad \text{and} \\
\Pr(\mathbf{e}) &= \sum_{\Omega(V(G))} \prod_{V \in V(G)} p\left(V \mid \pi(V)\right)\Bigg|_{\mathbf{E} = \mathbf{e}} \\
&= \sum_{\Omega(U)} \left(\Pr(U) \cdot \sum_{\Omega(V(G)) \setminus \{U\}} \prod_{\substack{V \in V(G) \\ V \neq U, V \neq W}} p\left(V \mid \pi(V)\right)\Bigg|_{\mathbf{E} = \mathbf{e}} \right),
\end{aligned}$$

where

$$\Pr(U) = \sum_{\Omega(W)} p\left(U \mid W\right) p\left(W\right).$$

We observe that varying a parameter from the CPTs of U and W can affect the probabilities $\Pr\left(C \leq c' \mid e_i^k, \mathbf{e}^-\right)$ and $\Pr\left(C \leq c' \;\middle|\; e_i^{k+1}, \mathbf{e}^-\right)$ only through $\Pr(u)$ and $\Pr(\bar{u})$.

Theorem 6.1.1. *Let B be as before and let U, W be binary variables and as despicted in Figure 6.1.1. Let $viol\left(c', e_i^k, \mathbf{e}^-\right)$ be a violation of monotonicity. If $viol\left(c', e_i^k, \mathbf{e}^-\right)$ cannot be resolved by varying a parameter from the CPT of U, then it can also not be resolved by varying a parameter from the CPT of W.*

Proof. Suppose that the violation of monotonicity cannot be resolved by varying a parameter from the CPT of U. Then, varying a parameter from the CPT of W can resolve the violation of monotonicity only if there are values for $\Pr\left(C \leq c' \mid e_i^k, \mathbf{e}^-\right)$ and $\Pr\left(C \leq c' \;\middle|\; e_i^{k+1}, \mathbf{e}^-\right)$ which can be obtained by varying a parameter from the CPT of W, but cannot be obtained by varying a parameter from the CPT of U. To prove the property stated in the theorem, we now show that all values for $\Pr\left(C \leq c' \mid e_i^k, \mathbf{e}^-\right)$ and $\Pr\left(C \leq c' \;\middle|\; e_i^{k+1}, \mathbf{e}^-\right)$ that can be obtained by varying any parameter $p\left(w''\right)$, $w'' \in \Omega(W)$, can also be obtained by varying a parameter $p\left(u' \mid w'\right)$, $u' \in \Omega(U), w' \in \Omega(W)$.

Since the parameters from the CPTs of U and W can only affect the probabilities $\Pr\left(C \leq c' \mid e_i^k, \mathbf{e}^-\right)$ and $\Pr\left(C \leq c' \,\middle|\, e_i^{k+1}, \mathbf{e}^-\right)$ through $\Pr(u)$ and $\Pr(\bar{u})$, it is sufficient to show that for every value $p(w'') = x$, $w'' \in \Omega(W)$, there is a $p(u' \mid w')$, $u' \in \Omega(U)$, $w' \in \Omega(W)$, and a value y such that

$$\Pr(u)\big(p(w'') = x\big) = \Pr(u)\big(p(u' \mid w') = y\big). \tag{6.1}$$

Note that since the variable U is binary, we then also have that

$$\Pr(\bar{u})\big(p(w'') = x\big) = \Pr(\bar{u})\big(p(u' \mid w') = y\big).$$

To find such a value y for every value $p(w'') = x$, we investigate for all $p(u' \mid w')$, $u' \in \Omega(U)$, $w' \in \Omega(W)$, the interval of values for x for which there is a value y such that equality (6.1) holds. Observe that our theorem only holds if the union of these intervals is $[0, 1]$.

We will begin by determining for which values $p(w) = x$ the parameter $p(u \mid w)$ can be varied to a value y such that

$$\Pr(u)\big(p(w) = x\big) = \Pr(u)\big(p(u \mid w) = y\big). \tag{6.2}$$

Using equality (6.2) we can express x in terms of y from

$$y \cdot p(w) + p(u \mid \bar{w})\, p(\bar{w}) = p(u \mid w) \cdot x + p(u \mid \bar{w})(1 - x).$$

It follows that

$$x = \frac{(y - p(u \mid \bar{w}))p(w)}{p(u \mid w) - p(u \mid \bar{w})},$$

unless $p(u \mid w) - p(u \mid \bar{w}) = 0$. In that case equality (6.2) holds for all $x \in [0, 1]$ by taking the original value of $p(u \mid w)$ for y.

By varying y in the interval $[0, 1]$ we obtain the interval of values for x for which there is a y such that equality (6.2) holds. We find that x lies between the minimum and the maximum of $\frac{-p(u \mid \bar{w})p(w)}{p(u \mid w) - p(u \mid \bar{w})}$ and $\frac{p(\bar{u} \mid \bar{w})p(w)}{p(u \mid w) - p(u \mid \bar{w})}$. Since

$$\min\left(\frac{-p(u \mid \bar{w})\, p(w)}{p(u \mid w) - p(u \mid \bar{w})}, \frac{p(\bar{u} \mid \bar{w})\, p(w)}{p(u \mid w) - p(u \mid \bar{w})}\right) < 0$$

and x is a probability, we have that the parameter $p(u \mid w)$ can be varied to a value y such that equality (6.2) holds, for all values x with

$$x \in \left[0, \min\left(1, \max\left(\frac{-p(u \mid \bar{w})\, p(w)}{p(u \mid w) - p(u \mid \bar{w})}, \frac{p(\bar{u} \mid \bar{w})\, p(w)}{p(u \mid w) - p(u \mid \bar{w})}\right)\right)\right]. \tag{6.3}$$

Next we determine for which values $p(w) = x$ the parameter $p(u \mid \bar{w})$ can be varied to a value y such that

$$\Pr(u)\big(p(w) = x\big) = \Pr(u)\big(p(u \mid \bar{w}) = y\big). \tag{6.4}$$

Using equality (6.4) we can again express x in terms of y from

$$p(u \mid w)\, p(w) + y \cdot p(\bar{w}) = p(u \mid w) \cdot x + p(u \mid \bar{w})(1 - x).$$

It follows that

$$x = \frac{p(u \mid w)\, p(w) + y \cdot p(\bar{w}) - p(u \mid \bar{w})}{p(u \mid w) - p(u \mid \bar{w})} = 1 + \frac{(y - p(u \mid w))p(\bar{w})}{p(u \mid w) - p(u \mid \bar{w})},$$

unless $p\,(u\,|\,w) - p\,(u\,|\,\bar{w}) = 0$. In that case equality (6.4) holds for all $x \in [0,1]$ by taking the original value of $p\,(u\,|\,\bar{w})$ for y.

By varying y in the interval $[0,1]$ we obtain the interval of values for x for which there is a y such that equality (6.4) holds. We have that x must lie between the minimum and the maximum of $1 - \frac{p(u\,|\,w)p(\bar{w})}{p(u\,|\,w)-p(u\,|\,\bar{w})}$ and $1 + \frac{p(\bar{u}\,|\,w)p(\bar{w})}{p(u\,|\,w)-p(u\,|\,\bar{w})}$. Since

$$\max\left(1 - \frac{p\,(u\,|\,w)\,p\,(\bar{w})}{p\,(u\,|\,w) - p\,(u\,|\,\bar{w})}, 1 + \frac{p\,(\bar{u}\,|\,w)\,p\,(\bar{w})}{p\,(u\,|\,w) - p\,(u\,|\,\bar{w})}\right) > 1$$

and x is a probability, we have that the parameter $p\,(u\,|\,\bar{w})$ can be varied to values y such that equality (6.4) holds, for all values x such that

$$x \in \left[\max\left(0, \min\left(1 - \frac{p\,(u\,|\,w)\,p\,(\bar{w})}{p\,(u\,|\,w) - p\,(u\,|\,\bar{w})}, 1 + \frac{p\,(\bar{u}\,|\,w)\,p\,(\bar{w})}{p\,(u\,|\,w) - p\,(u\,|\,\bar{w})}\right)\right), 1\right]. \tag{6.5}$$

We now have two intervals, namely (6.3) and (6.5), for which we have to show that the union is $[0,1]$. The endpoints of these intervals depend on the sign of $p\,(u\,|\,w) - p\,(u\,|\,\bar{w})$. We distinguish between these cases.

If $p\,(u\,|\,w) - p\,(u\,|\,\bar{w}) > 0$ then

$$\begin{aligned}
\max\left(\frac{-p\,(u\,|\,\bar{w})\,p\,(w)}{p\,(u\,|\,w) - p\,(u\,|\,\bar{w})}, \frac{p\,(\bar{u}\,|\,\bar{w})\,p\,(w)}{p\,(u\,|\,w) - p\,(u\,|\,\bar{w})}\right) &= \frac{p\,(\bar{u}\,|\,\bar{w})\,p\,(w)}{p\,(u\,|\,w) - p\,(u\,|\,\bar{w})} \\
\min\left(1 - \frac{p\,(u\,|\,w)\,p\,(\bar{w})}{p\,(u\,|\,w) - p\,(u\,|\,\bar{w})}, 1 + \frac{p\,(\bar{u}\,|\,w)\,p\,(\bar{w})}{p\,(u\,|\,w) - p\,(u\,|\,\bar{w})}\right) &= 1 - \frac{p\,(u\,|\,w)\,p\,(\bar{w})}{p\,(u\,|\,w) - p\,(u\,|\,\bar{w})},
\end{aligned}$$

and

$$1 - \frac{p\,(u\,|\,w)\,p\,(\bar{w})}{p\,(u\,|\,w) - p\,(u\,|\,\bar{w})} \le \frac{p\,(\bar{u}\,|\,\bar{w})\,p\,(w)}{p\,(u\,|\,w) - p\,(u\,|\,\bar{w})}.$$

If $p\,(u\,|\,w) - p\,(u\,|\,\bar{w}) < 0$ then

$$\begin{aligned}
\max\left(\frac{-p\,(u\,|\,\bar{w})\,p\,(w)}{p\,(u\,|\,w) - p\,(u\,|\,\bar{w})}, \frac{p\,(\bar{u}\,|\,\bar{w})\,p\,(w)}{p\,(u\,|\,w) - p\,(u\,|\,\bar{w})}\right) &= \frac{-p\,(u\,|\,\bar{w})\,p\,(w)}{p\,(u\,|\,w) - p\,(u\,|\,\bar{w})} \\
\min\left(1 - \frac{p\,(u\,|\,w)\,p\,(\bar{w})}{p\,(u\,|\,w) - p\,(u\,|\,\bar{w})}, 1 + \frac{p\,(\bar{u}\,|\,w)\,p\,(\bar{w})}{p\,(u\,|\,w) - p\,(u\,|\,\bar{w})}\right) &= 1 + \frac{p\,(\bar{u}\,|\,w)\,p\,(\bar{w})}{p\,(u\,|\,w) - p\,(u\,|\,\bar{w})}.
\end{aligned}$$

and

$$1 + \frac{p\,(\bar{u}\,|\,w)\,p\,(\bar{w})}{p\,(u\,|\,w) - p\,(u\,|\,\bar{w})} \le \frac{-p\,(u\,|\,\bar{w})\,p\,(w)}{p\,(u\,|\,w) - p\,(u\,|\,\bar{w})}.$$

Therefore, regardless of the sign of $p\,(u\,|\,w) - p\,(u\,|\,\bar{w})$, we have that

$$\begin{aligned}
&\left[0, \min\left(1, \max\left(\frac{-p\,(u\,|\,\bar{w})\,p\,(w)}{p\,(u\,|\,w) - p\,(u\,|\,\bar{w})}, \frac{p\,(\bar{u}\,|\,\bar{w})\,p\,(w)}{p\,(u\,|\,w) - p\,(u\,|\,\bar{w})}\right)\right)\right] \cup \\
&\left[\max\left(0, \min\left(1 - \frac{p\,(u\,|\,w)\,p\,(\bar{w})}{p\,(u\,|\,w) - p\,(u\,|\,\bar{w})}, 1 + \frac{p\,(\bar{u}\,|\,w)\,p\,(\bar{w})}{p\,(u\,|\,w) - p\,(u\,|\,\bar{w})}\right)\right), 1\right] = [0,1].
\end{aligned}$$

Thus for every $x \in [0,1]$ there is a value y for a parameter $p\,(u'\,|\,w')$, $u' \in \Omega(U)$, $w' \in \Omega(W)$, from the CPT of U such that

$$\Pr\,(u)\,\big(p\,(w) = x\big) = \Pr\,(u)\,\big(p\,(u'\,|\,w') = y\big).$$

Since there is no conceptual difference between varying $p\left(w\right)$ and $p\left(\bar{w}\right)$, we also have that for every $x = p\left(\bar{w}\right) \in [0,1]$ there is a value y for a parameter $p\left(u' \mid w'\right)$, $u' \in \Omega(U)$, $w' \in \Omega(W)$, from the CPT of U such that

$$\Pr\left(u\right)\left(p\left(\bar{w}\right) = x\right) = \Pr\left(u\right)\left(p\left(u' \mid w'\right) = y\right).$$

Therefore for every possible value x for a parameter $p\left(w''\right)$, $w'' \in \Omega(W)$, there is a value y for a parameter $p\left(u' \mid w'\right)$, $u' \in \Omega(U)$, $w' \in \Omega(W)$, such that

$$\Pr\left(C \leq c' \,\middle|\, e_i^k, \mathbf{e}^-\right)\left(p\left(w''\right) = x\right) = \Pr\left(C \leq c \,\middle|\, e_i^k, \mathbf{e}^-\right)\left(p\left(u' \mid w'\right) = y\right) \qquad \text{and}$$
$$\Pr\left(C \leq c' \,\middle|\, e_i^{k+1}, \mathbf{e}^-\right)\left(p\left(w''\right) = x\right) = \Pr\left(C \leq c' \,\middle|\, e_i^{k+1}, \mathbf{e}^-\right)\left(p\left(u' \mid w'\right) = y\right).$$

Thus if the violation of monotonicity cannot be resolved by varying a parameter from the CPT of U, then it can also not be resolved by varying a parameter from the CPT of W. □

From Theorem 6.1.1 we have for binary variables U and W that if the violation of monotonicity cannot be resolved by varying a parameter $p\left(u' \mid w'\right)$, $u' \in \Omega(U)$, $w' \in \Omega(W)$, of U, then it can also not be resolved by varying a parameter $p\left(w''\right)$, $w'' \in \Omega(W)$ of W. From this result it might be expected that this also holds in general. However the following examples will show that this is not the case, not even if one of the variables is binary.

Example 3. *Let $\Omega(U) = \{u^1, u^2, u^3\}$ and $\Omega(W) = \{w, \bar{w}\}$. Let the parameters from the CPTs of U and W be as follows:*

$$\begin{array}{llll}
p\left(w\right) = 0.2 & p\left(u^1 \mid w\right) = 0.1 & p\left(u^2 \mid w\right) = 0.3 & p\left(u^3 \mid w\right) = 0.6 \\
p\left(\bar{w}\right) = 0.8 & p\left(u^1 \mid \bar{w}\right) = 0.4 & p\left(u^2 \mid \bar{w}\right) = 0.1 & p\left(u^3 \mid \bar{w}\right) = 0.5
\end{array}$$

Suppose that varying $p\left(w\right)$ to 0.95 resolves the violation of monotonicity. We compute that $\Pr\left(u^1\right) = 0.115$, $\Pr\left(u^2\right) = 0.29$ and $\Pr\left(u^3\right) = 0.595$. If there is no value y for any $p\left(u' \mid w'\right)$ from the CPT of U such that

$$\begin{array}{ll}
\Pr\left(u^1\right)\left(p\left(u' \mid w'\right) = y\right) = 0.115, & \\
\Pr\left(u^2\right)\left(p\left(u' \mid w'\right) = y\right) = 0.29 & \textit{and} \\
\Pr\left(u^3\right)\left(p\left(u' \mid w'\right) = y\right) = 0.595, &
\end{array} \tag{6.6}$$

then there is no value y for any $p\left(u' \mid w'\right)$ from the CPT of U such that

$$\Pr\left(C \leq c' \,\middle|\, e_i^k, \mathbf{e}\right)\left(p\left(w\right) = 0.95\right) = \Pr\left(C \leq c' \,\middle|\, e_i^k, \mathbf{e}^-\right)\left(p\left(u' \mid w'\right) = y\right) \qquad \textit{and}$$
$$\Pr\left(C \leq c' \,\middle|\, e_i^{k+1}, \mathbf{e}^-\right)\left(p\left(w\right) = 0.95\right) = \Pr\left(C \leq c' \,\middle|\, e_i^{k+1}, \mathbf{e}^-\right)\left(p\left(u' \mid w'\right) = y\right).$$

The value y must be a probability and must therefore lie in the interval $[0,1]$.

We show that no such value y exists for any $p\left(u' \mid w'\right)$ from the CPT of U. We first investigate the parameters $p\left(U \mid w\right)$:

$$\begin{array}{lll}
\Pr\left(u^1\right)\left(p\left(u^1 \mid w\right) = y\right) = 0.32 + 0.2 \cdot y = 0.115 & \Rightarrow & y < 0 \\
\Pr\left(u^1\right)\left(p\left(u^2 \mid w\right) = y\right) = \dfrac{61}{175} - \dfrac{5}{175} \cdot y = 0.115 & \Rightarrow & y > 1 \\
\Pr\left(u^1\right)\left(p\left(u^3 \mid w\right) = y\right) = 0.37 - 0.05 \cdot y = 0.115 & \Rightarrow & y > 1
\end{array}$$

Since y does not lie in the interval $[0, 1]$ for any of the values for U, we now investigate $p\left(U \mid \bar{w}\right)$:

$$\Pr\left(u^1\right)\left(p\left(u^1 \mid \bar{w}\right) = y\right) = 0.02 + 0.8 \cdot y = 0.115 \Rightarrow \quad y = 0.11875$$

Since we have found a value for y, we verify if $\Pr\left(u^2\right)\left(p\left(u^1 \mid \bar{w}\right) = 0.11875\right) = 0.29$:

$$\begin{aligned}
\Pr\left(u^2\right)\left(p\left(u^1 \mid \bar{w}\right) = y\right) &= p\left(u^2 \mid w\right) p\left(w\right) + \frac{1-y}{1 - p\left(u^1 \mid \bar{w}\right)} p\left(u^2 \mid \bar{w}\right) \\
&= \frac{29}{150} - \frac{20}{150} \cdot y \\
&\Downarrow \\
\Pr\left(u^2\right)\left(p\left(u^1 \mid \bar{w}\right) = 0.11875\right) &= \frac{29}{150} - \frac{20}{150} \cdot 0.11875 \\
&= 0.1775 \neq 0.29
\end{aligned}$$

Thus equalities (6.6) do not hold for any value y for $p\left(u^1 \mid \bar{w}\right)$. The same arguments also hold for the parameters $p\left(u^2 \mid \bar{w}\right)$ and $p\left(u^3 \mid \bar{w}\right)$:

$$\begin{aligned}
&\Pr\left(u^1\right)\left(p\left(u^2 \mid \bar{w}\right) = y\right) = \frac{169}{450} - \frac{160}{450} \cdot y = 0.115 \qquad \Rightarrow \quad y = 0.7328125 \\
&\Pr\left(u^2\right)\left(p\left(u^2 \mid \bar{w}\right) = 0.7328125\right) = 0.06 + 0.8 \cdot 0.7328125 = 0.64625 \neq 0.29 \\
&\Pr\left(u^1\right)\left(p\left(u^3 \mid \bar{w}\right) = y\right) = 0.66 - 0.64 \cdot y = 0.115 \qquad \Rightarrow \quad y = 0.8515625 \\
&\Pr\left(u^2\right)\left(p\left(u^3 \mid \bar{w}\right) = 0.8515625\right) = 0.22 - 0.16 \cdot 0.8515625 = 0.08375 \neq 0.29
\end{aligned}$$

Thus for each $p\left(u' \mid w'\right)$ we find that either y does not lie in the interval $[0, 1]$ or the values y for which $\Pr\left(u^1\right)\left(p\left(u' \mid w'\right) = y\right) = 0.115$ and $\Pr\left(u^2\right)\left(p\left(u' \mid w'\right) = y\right) = 0.29$ differ. Therefore, even if the violation of monotonicity cannot be resolved by varying any $p\left(u' \mid w'\right)$ from the CPT of U, then it is possible to resolve the violation by varying $p\left(w\right)$ to 0.95. □

The behaviour we observe in Example 3 is caused by the fact that when we vary a parameter from the CPT of U, we apply proportional scaling. A result of this proportional scaling is that, for example, varying $p\left(u^1 \mid w\right)$ cannot change the sign of $\Pr\left(u^2\right) - \Pr\left(u^3\right)$, while varying in the CPT of W can. Thus we cannot obtain the same result by varying in the CPT of U as we can obtain by varying in the CPT of W.

Example 4. *Let $\Omega(U) = \{u, \bar{u}\}$ and $\Omega(W) = \{w^1, w^2, w^3\}$. Let the parameters from the CPTs of U and W be as follows:*

$$\begin{array}{ccc}
p\left(w^1\right) = 0.6 & p\left(w^2\right) = 0.3 & p\left(w^3\right) = 0.1 \\
p\left(u \mid w^1\right) = 0.9 & p\left(u \mid w^2\right) = 0.4 & p\left(u \mid w^3\right) = 0.2
\end{array}$$

Suppose that varying $p\left(w^1\right)$ to 0.95 resolves the violation of monotonicity. Then, by proportional scaling, $p\left(w^2\right) = 0.025$ and $p\left(w^3\right) = \frac{1}{120}$. We compute that $\Pr\left(u\right) = 0.8725$ and $\Pr\left(\bar{u}\right) = 0.1275$. If there is no value y for any $p\left(u' \mid w'\right)$ from the CPT of U such that

$$\begin{aligned}
\Pr\left(u\right)\left(p\left(u' \mid w'\right) = y\right) &= 0.8725 \quad \textit{and} \\
\Pr\left(\bar{u}\right)\left(p\left(u' \mid w'\right) = y\right) &= 0.1275,
\end{aligned}$$

then there is no value y for any $p(u' \mid w')$ from the CPT of U such that

$$\Pr\left(C \leq c' \,\middle|\, e_i^k, \mathbf{e}\right)\left(p\left(w^1\right) = 0.95\right) = \Pr\left(C \leq c' \,\middle|\, e_i^k, \mathbf{e}^-\right)\left(p\left(u' \mid w'\right) = y\right) \qquad \textit{and}$$
$$\Pr\left(C \leq c' \,\middle|\, e_i^{k+1}, \mathbf{e}^-\right)\left(p\left(w^1\right) = 0.95\right) = \Pr\left(C \leq c' \,\middle|\, e_i^{k+1}, \mathbf{e}^-\right)\left(p\left(u' \mid w'\right) = y\right).$$

The value y must be a probability and must therefore lie in the interval $[0,1]$.

We show that no such value y exists for any $p(u' \mid w')$ from the CPT of U. For the parameters $p\left(u \mid w^1\right)$, $p\left(u \mid w^2\right)$ and $p\left(u \mid w^3\right)$ we find that y must be larger than 1 and therefore does not lie in the interval $[0,1]$.

$$\begin{aligned}
\Pr(u)\left(p\left(u \mid w^1\right) = y\right) &= 0.14 + 0.6 \cdot y = 0.8725 &\Rightarrow\quad & y > 1\\
\Pr(u)\left(p\left(u \mid w^2\right) = y\right) &= 0.56 + 0.3 \cdot y = 0.8725 &\Rightarrow\quad & y > 1\\
\Pr(u)\left(p\left(u \mid w^3\right) = y\right) &= 0.66 + 0.1 \cdot y = 0.8725 &\Rightarrow\quad & y > 1
\end{aligned}$$

It follows that for the parameters $p\left(\bar{u} \mid w^1\right)$, $p\left(\bar{u} \mid w^2\right)$ and $p\left(\bar{u} \mid w^3\right)$ we find that y must be smaller than 0 and therefore does not lie in the interval $[0,1]$. Since no y lies in the interval $[0,1]$ for any $p(u' \mid w')$ from the CPT of U, even if the violation of monotonicity cannot be resolved by varying any $p(u' \mid w')$ from the CPT of U, then it is possible to resolve the violation by varying $p\left(w^1\right)$ to 0.95. □

6.2 The general case

In the simple case we have seen that we can only say that if we cannot resolve a violation of monotonicity by varying a parameter from the CPT of U, then we can also not resolve it by varying a parameter from the CPT of W, if U and W are binary. In the general case we will therefore assume that U and W are always binary.

We consider an arbitrary restricted Bayesian network $B = (G, P)$ in which there are two vertices $U, W \in V(G)$ such that U is the only successor of W, and the only predecessors of U are W and predecessors of W, i.e. $\pi(U) \subseteq \pi(W) \cup \{W\}$ and $\sigma(W) = \{U\}$. We use $\mathbf{Q}$ to denote $\pi(U) \backslash \{W\}$ and $\mathbf{R}$ to denote $\pi(W) \backslash \pi(U)$. Furthermore, if the outgoing arcs of U are removed, then U is in a connected component S_u which is distinct from the connected component containing the variable of interest; a graphical representation of such a Bayesian network is depicted in Figure 6.2.2. We again assume that neither U nor W equals the variable of interest C or contains any observation, i.e. $\{U, W\} \cap (\{C\} \cup \mathbf{E}) = \emptyset$.

Now suppose that there is a violation of monotonicity $\mathit{viol}\left(c', e_i^k, \mathbf{e}^-\right)$ in this Bayesian network. Thus we have that

$$\Pr\left(C \leq c' \,\middle|\, e_i^k, \mathbf{e}^-\right) < \Pr\left(C \leq c' \,\middle|\, e_i^{k+1}, \mathbf{e}^-\right).$$

This violation of monotonicity can again be written as

$$\sum_{\substack{c \in \Omega(C)\\ c \leq c'}} \frac{\Pr\left(c, e_i^k, \mathbf{e}^-\right)}{\Pr\left(e_i^k, \mathbf{e}^-\right)} < \sum_{\substack{c \in \Omega(C)\\ c \leq c'}} \frac{\Pr\left(c, e_i^{k+1}, \mathbf{e}^-\right)}{\Pr\left(e_i^{k+1}, \mathbf{e}^-\right)}.$$

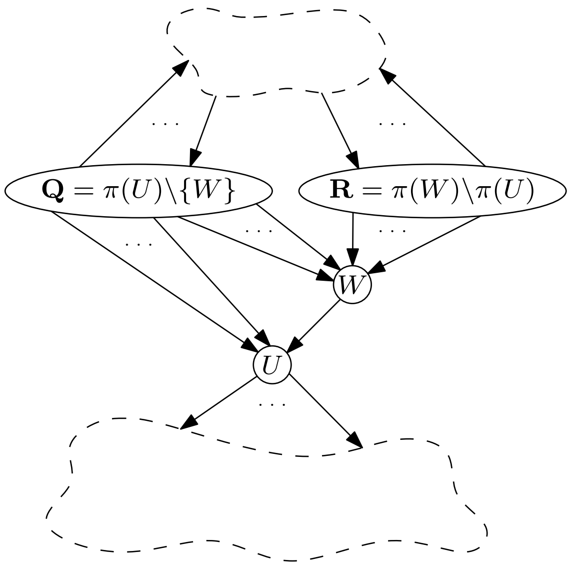

Figure 6.2.2: An arbitrary restricted Bayesian network in which there are two vertices U, W such that $\pi(U) \subseteq \pi(W) \cup \{W\}$ and $\sigma(W) = \{U\}$, and U lies on all chains between W and the variable of interest C.

Until stated otherwise we assume that $V(S_u) \cap \mathbf{E} = \emptyset$. Then, for the Bayesian network B we have for all $c \in \Omega(C)$ and $\mathbf{e} \in \Omega(\mathbf{E})$ that

$$\Pr(c, \mathbf{e}) = \sum_{\Omega(U)} \left(\Pr(U) \cdot \sum_{\Omega(V(G)\backslash V(S_u))} \prod_{\substack{V \in V(G) \\ V \neq U, V \neq W}} p(V \mid \pi(V)) \Bigg|_{\substack{C = c \\ \mathbf{E} = \mathbf{e}}} \right) \quad \text{and}$$

$$\Pr(\mathbf{e}) = \sum_{\Omega(U)} \left(\Pr(U) \cdot \sum_{\Omega(V(G)\backslash V(S_u))} \prod_{\substack{V \in V(G) \\ V \neq U, V \neq W}} p(V \mid \pi(V)) \Bigg|_{\mathbf{E} = \mathbf{e}} \right),$$

where

$$\Pr(U) = \sum_{\Omega(W) \times \Omega(\mathbf{Q}) \times \Omega(\mathbf{R})} p(U \mid W, \mathbf{Q})\, p(W \mid \mathbf{Q}, \mathbf{R}) \Pr(\mathbf{Q}, \mathbf{R}).$$

Therefore varying parameters from the CPTs of U and W can only affect the probabilities $\Pr\left(C \leq c' \mid e_i^k, \mathbf{e}^-\right)$ and $\Pr\left(C \leq c' \,\middle|\, e_i^{k+1}, \mathbf{e}^-\right)$ through $\Pr(u)$ and $\Pr(\bar{u})$.

Theorem 6.2.1. *Let B be as before and let U, W be binary and as depicted in Figure 6.2.2. If viol$\left(c', e_i^k, \mathbf{e}^-\right)$ cannot be resolved by varying a parameter from the CPT of U, then it can also not be resolved by varying a parameter from the CPT of W.*

Proof. Suppose that the violation of monotonicity cannot be resolved by varying a parameter from the CPT of U. Then, varying a parameter from the CPT of W can resolve the violation of monotonicity only if there are values for $\Pr\left(C \leq c' \mid e_i^k, \mathbf{e}^-\right)$ and $\Pr\left(C \leq c'' \,\middle|\, e_i^{k+1}, \mathbf{e}^-\right)$ which can be obtained by varying a parameter from the CPT of W, but cannot be obtained by varying a parameter from the CPT of U. To prove the property stated in the theorem, we show that all values for $\Pr\left(C \leq c' \mid e_i^k, \mathbf{e}^-\right)$ and

$\Pr\left(C \leq c' \,\middle|\, e_i^{k+1}, \mathbf{e}^-\right)$ that can obtained by varying any parameter $p\,(w'' \,|\, \mathbf{q}, \mathbf{r})$, $w'' \in \Omega(W)$, $\mathbf{q} \in \Omega(\mathbf{Q})$, $\mathbf{r} \in \Omega(R)$, can also be obtained by a parameter $p\,(u' \,|\, w', \mathbf{q}')$, $u' \in \Omega(U)$, $w' \in \Omega(W)$, $\mathbf{q}' \in \Omega(\mathbf{Q})$.

Since the parameters from the CPTs of U and W can only affect the probabilities $\Pr\left(C \leq c' \,\middle|\, e_i^k, \mathbf{e}^-\right)$ and $\Pr\left(C \leq c' \,\middle|\, e_i^{k+1}, \mathbf{e}^-\right)$ through $\Pr(u)$ and $\Pr(\bar{u})$, it is sufficient to show that for every value $p\,(w'' \,|\, \mathbf{q}, \mathbf{r}) = x$, $w'' \in \Omega(W)$, $\mathbf{q} \in \Omega(\mathbf{Q})$, $\mathbf{r} \in \Omega(\mathbf{R})$, there is a $p\,(u' \,|\, w', \mathbf{q}')$, $u' \in \Omega(U)$, $w' \in \Omega(W)$, $\mathbf{q}' \in \Omega(\mathbf{Q})$, and a value y such that

$$\Pr(u)\left(p\left(w'' \,|\, \mathbf{q}, \mathbf{r}\right) = x\right) = \Pr(u)\left(p\left(u' \,|\, w', \mathbf{q}'\right) = y\right). \tag{6.7}$$

Note that since the variable U is binary, we then also have that

$$\Pr(\bar{u})\left(p\left(w'' \,|\, \mathbf{q}, \mathbf{r}\right) = x\right) = \Pr(\bar{u})\left(p\left(u' \,|\, w', \mathbf{q}'\right) = y\right)$$

To find such a value y for every value $p\,(w'' \,|\, \mathbf{q}, \mathbf{r}) = x$, we investigate for all $p\,(u' \,|\, w', \mathbf{q}')$, $u' \in \Omega(U)$, $w' \in \Omega(W)$, $\mathbf{q}' \in \Omega(Q)$, the interval of values for x for which there is a value y such that equality (6.7) holds. Observe that our theorem only holds if the union of these intervals is $[0, 1]$.

We will begin by determining for which values $p\,(w \,|\, \mathbf{q}, \mathbf{r}) = x$ the parameter $p\,(u \,|\, w, \mathbf{q})$ can be varied to a value y such that

$$\Pr(u)\left(p\,(w \,|\, \mathbf{q}, \mathbf{r}) = x\right) = \Pr(u)\left(p\,(u \,|\, w, \mathbf{q}) = y\right). \tag{6.8}$$

Using equality (6.8) we can express x in terms of y from

$$\begin{gathered}
\big(p\,(u \,|\, w, \mathbf{q}) - p\,(u \,|\, \bar{w}, \mathbf{q})\big) \cdot x \cdot \Pr(\mathbf{q}, \mathbf{r}) + p\,(u \,|\, \bar{w}, \mathbf{q}) \Pr(\mathbf{q}, \mathbf{r}) + \\
\sum_{\mathbf{r}' \in \Omega(\mathbf{R}) \setminus \{\mathbf{r}\}} \Big(p\,(u \,|\, w, \mathbf{q})\, p\left(w \,|\, \mathbf{q}, \mathbf{r}'\right) + p\,(u \,|\, \bar{w}, \mathbf{q})\, p\left(\bar{w} \,|\, \mathbf{q}, \mathbf{r}'\right)\Big) \Pr\left(\mathbf{q}, \mathbf{r}'\right) + \\
\sum_{\substack{\mathbf{r}' \in \Omega(\mathbf{R}) \\ \mathbf{q}' \in \Omega(\mathbf{Q}) \setminus \{\mathbf{q}\}}} \Big(p\left(u \,|\, w, \mathbf{q}'\right) p\left(w \,|\, \mathbf{q}', \mathbf{r}'\right) + p\left(u \,|\, \bar{w}, \mathbf{q}'\right) p\left(\bar{w} \,|\, \mathbf{q}', \mathbf{r}'\right)\Big) \Pr\left(\mathbf{q}', \mathbf{r}'\right) \\
= \\
\sum_{\mathbf{r}' \in \Omega(\mathbf{R})} p\left(w \,|\, \mathbf{q}, \mathbf{r}'\right) \Pr\left(\mathbf{q}, \mathbf{r}'\right) \cdot y + \sum_{\mathbf{r}' \in \Omega(\mathbf{R})} p\,(u \,|\, \bar{w}, \mathbf{q})\, p\left(\bar{w} \,|\, \mathbf{q}, \mathbf{r}'\right) \Pr\left(\mathbf{q}, \mathbf{r}'\right) + \\
\sum_{\substack{\mathbf{r}' \in \Omega(\mathbf{R}) \\ \mathbf{q}' \in \Omega(\mathbf{Q}) \setminus \{\mathbf{q}\}}} \Big(p\left(u \,|\, w, \mathbf{q}'\right) p\left(w \,|\, \mathbf{q}', \mathbf{r}'\right) + p\left(u \,|\, \bar{w}, \mathbf{q}'\right) p\left(\bar{w} \,|\, \mathbf{q}', \mathbf{r}'\right)\Big) \Pr\left(\mathbf{q}', \mathbf{r}'\right).
\end{gathered}$$

It follows that

$$x = \frac{\big(y - p\,(u \,|\, \bar{w}, \mathbf{q})\big) p\,(w \,|\, \mathbf{q}, \mathbf{r}) \Pr(\mathbf{q}, \mathbf{r}) + \big(y - p\,(u \,|\, w, \mathbf{q})\big) \sum_{\mathbf{r}' \in \Omega(\mathbf{R}) \setminus \{\mathbf{r}\}} p\,(w \,|\, \mathbf{q}, \mathbf{r}') \Pr(\mathbf{q}, \mathbf{r}')}{\big(p\,(u \,|\, w, \mathbf{q}) - p\,(u \,|\, \bar{w}, \mathbf{q})\big) \Pr(\mathbf{q}, \mathbf{r})}$$

unless $p\,(u \,|\, w, \mathbf{q}) - p\,(u \,|\, \bar{w}, \mathbf{q}) = 0$. In that case equality (6.8) holds for all $x \in [0, 1]$ by taking the original value of $p\,(u \,|\, w, \mathbf{q})$ for y.

By varying y in the interval $[0,1]$ we obtain the interval of values for x for which there is a y such that equality (6.8) holds. We find that x lies between the minimum and the maximum of L_1 and L_2, where

$$L_1 = -\frac{p\left(u \mid \bar{w}, \mathbf{q}\right) p\left(w \mid \mathbf{q}, \mathbf{r}\right) \Pr\left(\mathbf{q}, \mathbf{r}\right) + p\left(u \mid w, \mathbf{q}\right) \sum_{\mathbf{r}' \in \Omega(\mathbf{R})} p\left(w \mid \mathbf{q}, \mathbf{r}'\right) \Pr\left(\mathbf{q}, \mathbf{r}'\right)}{\big(p\left(u \mid w, \mathbf{q}\right) - p\left(u \mid \bar{w}, \mathbf{q}\right)\big) \Pr\left(\mathbf{q}, \mathbf{r}\right)} \quad \text{and}$$

$$L_2 = \frac{p\left(\bar{u} \mid \bar{w}, \mathbf{q}\right) p\left(w \mid \mathbf{q}, \mathbf{r}\right) \Pr\left(\mathbf{q}, \mathbf{r}\right) + p\left(\bar{u} \mid w, \mathbf{q}\right) \sum_{\mathbf{r}' \in \Omega(\mathbf{R})} p\left(w \mid \mathbf{q}, \mathbf{r}'\right) \Pr\left(\mathbf{q}, \mathbf{r}'\right)}{\big(p\left(u \mid w, \mathbf{q}\right) - p\left(u \mid \bar{w}, \mathbf{q}\right)\big) \Pr\left(\mathbf{q}, \mathbf{r}\right)}.$$

Since

$$\min\left(L_1, L_2\right) < 0$$

and x is a probability, we have that the parameter $p\left(u \mid w, \mathbf{q}\right)$ can be varied to values y such that equality (6.8) holds, for all values x with

$$x \in \Big[0, \min\big(1, \max\left(L_1, L_2\right)\big)\Big]. \tag{6.9}$$

Next we determine for which values $p\left(w \mid \mathbf{q}, \mathbf{r}\right) = x$ the parameter $p\left(u \mid \bar{w}, \mathbf{q}\right)$ can be varied to a value y such that

$$\Pr\left(u\right)\left(p\left(w \mid \mathbf{q}, \mathbf{r}\right) = x\right) = \Pr\left(u\right)\left(p\left(u \mid \bar{w}, \mathbf{q}\right) = y\right). \tag{6.10}$$

Using equality (6.10) we can express x in terms of y from

$$\begin{aligned}
&\big(p\left(u \mid w, \mathbf{q}\right) - p\left(u \mid \bar{w}, \mathbf{q}\right)\big) \cdot x \cdot \Pr\left(\mathbf{q}, \mathbf{r}\right) + p\left(u \mid \bar{w}, \mathbf{q}\right) \Pr\left(\mathbf{q}, \mathbf{r}\right) + \\
&\quad \sum_{\mathbf{r}' \in \Omega(\mathbf{R}) \setminus \{\mathbf{r}\}} \Big(p\left(u \mid w, \mathbf{q}\right) p\left(w \mid \mathbf{q}, \mathbf{r}'\right) + p\left(u \mid \bar{w}, \mathbf{q}\right) p\left(\bar{w} \mid \mathbf{q}, \mathbf{r}'\right)\Big) \Pr\left(\mathbf{q}, \mathbf{r}\right) + \\
&\quad \sum_{\substack{\mathbf{r}' \in \Omega(\mathbf{R}) \\ \mathbf{q}' \in \Omega(\mathbf{Q}) \setminus \{\mathbf{q}\}}} \Big(p\left(u \mid w, \mathbf{q}'\right) p\left(w \mid \mathbf{q}'\right) + p\left(u \mid \bar{w}, \mathbf{q}'\right) p\left(\bar{w} \mid \mathbf{q}', \mathbf{r}'\right)\Big) \Pr\left(\mathbf{q}', \mathbf{r}'\right) \\
&= \\
&\sum_{\mathbf{r}' \in \Omega(\mathbf{R})} p\left(u \mid w, \mathbf{q}\right) p\left(w \mid \mathbf{q}, \mathbf{r}'\right) \Pr\left(\mathbf{q}, \mathbf{r}'\right) + \sum_{\mathbf{r}' \in \Omega(\mathbf{R})} p\left(\bar{w} \mid \mathbf{q}, \mathbf{r}'\right) \Pr\left(\mathbf{q}, \mathbf{r}'\right) \cdot y + \\
&\quad \sum_{\substack{\mathbf{r}' \in \Omega(\mathbf{R}) \\ \mathbf{q}' \in \Omega(\mathbf{Q}) \setminus \{\mathbf{q}\}}} \Big(p\left(u \mid w, \mathbf{q}'\right) p\left(w \mid \mathbf{q}', \mathbf{r}'\right) + p\left(u \mid \bar{w}, \mathbf{q}'\right) p\left(\bar{w} \mid \mathbf{q}', \mathbf{r}'\right)\Big) \Pr\left(\mathbf{q}', \mathbf{r}'\right).
\end{aligned}$$

It follows that

$$x = 1 + \frac{\big(y - p\left(u \mid \bar{w}, \mathbf{q}\right)\big) \displaystyle\sum_{\mathbf{r}' \in \Omega(\mathbf{R}) \setminus \{\mathbf{r}\}} p\left(\bar{w} \mid \mathbf{q}\right) \Pr\left(\mathbf{q}, \mathbf{r}'\right) + \big(y - p\left(u \mid w, \mathbf{q}\right)\big) p\left(\bar{w} \mid \mathbf{q}, \mathbf{r}\right) \Pr\left(\mathbf{q}, \mathbf{r}\right)}{\big(p\left(u \mid w, \mathbf{q}\right) - p\left(u \mid \bar{w}, \mathbf{q}\right)\big) \Pr\left(\mathbf{q}, \mathbf{r}\right)}$$

unless $p\left(u \mid w, \mathbf{q}\right) - p\left(u \mid \bar{w}, \mathbf{q}\right) = 0$. In that case equality (6.10) holds for all $x \in [0,1]$ by taking the original value of $p\left(u \mid \bar{w}, \mathbf{q}\right)$ for y.

By varying y in the interval $[0,1]$ we obtain the interval of values for x for which there is a y such that equality (6.10) holds. We have that x must lie between the minimum and the maximum of L_3 and L_4, where

$$L_3 = 1 - \frac{p\left(u \,|\, \bar{w}, \mathbf{q}\right) \sum_{\mathbf{r}' \in \Omega(\mathbf{R}) \setminus \{\mathbf{r}\}} p\left(\bar{w} \,|\, \mathbf{q}, \mathbf{r}'\right) \Pr\left(\mathbf{q}, \mathbf{r}'\right) + p\left(u \,|\, w, \mathbf{q}\right) p\left(\bar{w} \,|\, \mathbf{q}, \mathbf{r}\right) \Pr\left(\mathbf{q}, \mathbf{r}\right)}{\big(p\left(u \,|\, w, \mathbf{q}\right) - p\left(u \,|\, \bar{w}, \mathbf{q}\right)\big) \Pr\left(\mathbf{q}, \mathbf{r}\right)} \quad \text{and}$$
$$L_4 = 1 + \frac{p\left(\bar{u} \,|\, \bar{w}, \mathbf{q}\right) \sum_{\mathbf{r}' \in \Omega(\mathbf{R}) \setminus \{\mathbf{r}\}} p\left(\bar{w} \,|\, \mathbf{q}, \mathbf{r}'\right) \Pr\left(\mathbf{q}, \mathbf{r}'\right) + p\left(\bar{u} \,|\, w, \mathbf{q}\right) p\left(\bar{w} \,|\, \mathbf{q}, \mathbf{r}\right) \Pr\left(\mathbf{q}, \mathbf{r}\right)}{\big(p\left(u \,|\, w, \mathbf{q}\right) - p\left(u \,|\, \bar{w}, \mathbf{q}\right)\big) \Pr\left(\mathbf{q}, \mathbf{r}\right)}.$$

Since

$$\max\left(L_3, L_4\right) > 1$$

and x is a probability, we have that the parameter $p\left(u \,|\, w, \mathbf{q}\right)$ can be varied to values y such that equality (6.8) holds, for all values x such that

$$x \in \Big[\max\big(0, \min\left(L_3, L_4\right)\big), 1\Big]. \tag{6.11}$$

We now have two intervals, namely (6.9) and (6.11), for which we have to show that the union is $[0, 1]$. The endpoints of these intervals depend on the sign of $p\left(u \,|\, w, \mathbf{q}\right) - p\left(u \,|\, \bar{w}, \mathbf{q}\right)$. We distinguish between these cases.

If $p\left(u \,|\, w, \mathbf{q}\right) - p\left(u \,|\, \bar{w}, \mathbf{q}\right) > 0$ then

$$\begin{aligned} \max\left(L_1, L_2\right) &= L_2 \\ \min\left(L_3, L_4\right) &= L_3, \end{aligned}$$

and $L_3 \leq L_2$.

If $p\left(u \,|\, w, \mathbf{q}\right) - p\left(u \,|\, \bar{w}, \mathbf{q}\right) < 0$ then

$$\begin{aligned} \max\left(L_1, L_2\right) &= L_1 \\ \min\left(L_3, L_4\right) &= L_4, \end{aligned}$$

and $L_4 \leq L_1$.

Therefore, regardless of the sign of $p\left(u \,|\, w, \mathbf{q}\right) - p\left(u \,|\, \bar{w}, \mathbf{q}\right)$, we have that

$$\Big[0, \min\big(1, \max\left(L_1, L_2\right)\big)\Big] \cup \Big[\max\big(0, \min\left(L_3, L_4\right)\big), 1\Big] = [0, 1].$$

Thus for every $x \in [0, 1]$ for a parameter $p\left(w \,|\, \mathbf{q}, \mathbf{r}\right)$, $\mathbf{q} \in \Omega(\mathbf{Q})$, $\mathbf{r} \in \Omega(\mathbf{R})$, there is a value y for a parameter $p\left(u' \,|\, w', \mathbf{q}'\right)$, $u' \in \Omega(U)$, $w' \in \Omega(W)$, $\mathbf{q}' \in \Omega(\mathbf{Q})$, from the CPT of U such that

$$\Pr\left(u\right)\big(p\left(w \,|\, \mathbf{q}, \mathbf{r}\right) = x\big) = \Pr\left(u\right)\big(p\left(u' \,|\, w', \mathbf{q}'\right) = y\big).$$

Since there is no conceptual difference between varying $p\left(w \,|\, \mathbf{q}, \mathbf{r}\right)$ and $p\left(\bar{w} \,|\, \mathbf{q}, \mathbf{r}\right)$, we also have that for every $x = p\left(\bar{w} \,|\, \mathbf{q}, \mathbf{r}\right) \in [0, 1]$ there is a value y for a parameter $p\left(u' \,|\, w', \mathbf{q}'\right)$, $u' \in \Omega(U)$, $w' \in \Omega(W)$, $\mathbf{q}' \in \Omega(\mathbf{Q})$, from the CPT of U such that

$$\Pr\left(u\right)\big(p\left(\bar{w} \,|\, \mathbf{q}, \mathbf{r}\right) = x\big) = \Pr\left(u\right)\big(p\left(u' \,|\, w', \mathbf{q}'\right) = y\big).$$

Therefore for every possible value x, namely $x \in [0, 1]$, of parameter $p\left(w'' \,|\, \mathbf{q}, \mathbf{r}\right)$, $w'' \in \Omega(W)$, $\mathbf{q} \in \Omega(\mathbf{Q})$, $\mathbf{r} \in \Omega(\mathbf{R})$, there is a value y for a parameter $p\left(u' \,|\, w', \mathbf{q}'\right)$, $u' \in \Omega(U)$, $w' \in \Omega(W)$, $\mathbf{q}' \in \Omega(\mathbf{Q})$, such that

$$\Pr\left(C \leq c' \,\middle|\, e_i^k, \mathbf{e}^-\right)\big(p\left(w'' \,|\, \mathbf{q}, \mathbf{r}\right) = x\big) = \Pr\left(C \leq c \,\middle|\, e_i^k, \mathbf{e}^-\right)\big(p\left(u' \,|\, w', \mathbf{q}'\right) = y\big) \quad \text{and}$$
$$\Pr\left(C \leq c' \,\middle|\, e_i^{k+1}, \mathbf{e}^-\right)\big(p\left(w'' \,|\, \mathbf{q}, \mathbf{r}\right) = x\big) = \Pr\left(C \leq c' \,\middle|\, e_i^{k+1}, \mathbf{e}^-\right)\big(p\left(u' \,|\, w', \mathbf{q}'\right) = y\big).$$

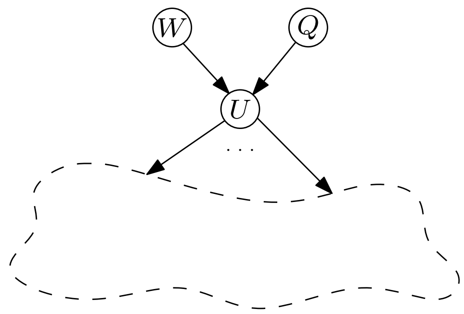

Figure 6.2.3: Directed acyclic graph used in Example 5.

Thus if the violation of monotonicity cannot be resolved by varying a parameter from the CPT of U, then it can also not be resolved by varying a parameter from the CPT of W. $\square$

So far, we assumed that $V(S_u) \cap \mathbf{E} = \emptyset$. However, if $(V(S_u)\backslash\{U, W\}) \cap \mathbf{E} = \mathbf{E}_S \neq \emptyset$ and $E_i \notin \mathbf{E}_S$, then Theorem 6.2.1 still holds. We can substitute $\Pr(U \mid \mathbf{e}^-)$ for $\Pr(U)$ and $\Pr(\mathbf{Q}, \mathbf{R} \mid \mathbf{e}^-)$ for $\Pr(\mathbf{Q}, \mathbf{R})$ in the proof of Theorem 6.2.1 without invalidating any of its arguments. Note that if $\mathbf{q} \in \Omega(\mathbf{Q}), \mathbf{r} \in \Omega(\mathbf{R})$ and $\mathbf{e}^-$ assign conflicting values to a variable, then $\Pr(\mathbf{q}, \mathbf{r} \mid \mathbf{e}^-) = 0$.

We discuss examples of (restricted) Bayesian networks which do not conform to the general case by violating one of the conditions to the associated DAG. These examples suggest that we cannot generalize further.

Our first example violates the condition that $\pi(U) \subseteq \pi(W) \cup \{W\}$.

Example 5. *Let B be a Bayesian network with a DAG as depicted in Figure 6.2.3 with $\Omega(U) = \{\bar{u}, u\}$, $\Omega(W) = \{\bar{w}, w\}$ and $\Omega(Q) = \{\bar{q}, q\}$. Suppose there is a violation of monotonicity $viol\big(c', e_i^k, \mathbf{e}^-\big)$. Like before, varying parameters from the CPTs of U and W can only affect $\Pr\left(C \le c' \mid e_i^k, \mathbf{e}^-\right)$ and $\Pr\left(C \le c' \,\middle|\, e_i^{k+1}, \mathbf{e}^-\right)$ through $\Pr(u)$ and $\Pr(\bar{u})$, where*

$$\Pr(U) = \sum_{\Omega(\{W,Q\})} p(U \mid W, Q)\, p(W)\, p(Q)\,.$$

The following value assignments for the CPTs of U, W and Q illustrate that varying a parameter from the CPT of W may resolve the violation of monotonicity, while varying a parameter from the CPT of U does not.

Let the parameters from the CPTs of U and W be as follows:

$$\begin{array}{lll} p(w) = 0.6 & p(u \mid w, q) = 0.5 & p(u \mid \bar{w}, q) = 0.1 \\ p(q) = 0.7 & p(u \mid w, \bar{q}) = 0.8 & p(u \mid \bar{w}, \bar{q}) = 0.2 \end{array}$$

Suppose that varying $p(w)$ to 0.1 resolves the violation of monotonicity. We compute that $\Pr(u) = 0.176$. If there is no value y for any $p(u' \mid w', q')$, $u' \in \Omega(U)$, $w' \in \Omega(W)$, $q' \in \Omega(Q)$, from the CPT of U such that

$$\Pr(u)\,(p(u' \mid w', q') = y) = 0.176, \tag{6.12}$$

then there is no value y for any $p(u' \mid w', q')$ from the CPT of U such that

$$\Pr\left(C \le c' \;\middle|\; e_i^k, \mathbf{e}^-\right)\left(p(w) = 0.1\right) = \Pr\left(C \le c' \;\middle|\; e_i^k, \mathbf{e}^-\right)\left(p(u' \mid w', q') = y\right) \quad \textit{and}$$
$$\Pr\left(C \le c' \;\middle|\; e_i^{k+1}, \mathbf{e}^-\right)\left(p(w) = 0.1\right) = \Pr\left(C \le c' \;\middle|\; e_i^{k+1}, \mathbf{e}^-\right)\left(p(u' \mid w', q') = y\right).$$

The value y must be a probability and must therefore lie in the interval $[0,1]$. We show that no such value y exists for any $p(u' \mid w', q')$ from the CPT of U.

$$\begin{aligned}
\Pr(u)\left(p(u \mid w, q) = y\right) &= 0.42 \cdot y + 0.196 = 0.176 && \Rightarrow \quad y < 0\\
\Pr(u)\left(p(u \mid w, \bar{q}) = y\right) &= 0.18 \cdot y + 0.262 = 0.176 && \Rightarrow \quad y < 0\\
\Pr(u)\left(p(u \mid \bar{w}, q) = y\right) &= 0.28 \cdot y + 0.378 = 0.176 && \Rightarrow \quad y < 0\\
\Pr(u)\left(p(u \mid \bar{w}, \bar{q}) = y\right) &= 0.12 \cdot y + 0.382 = 0.176 && \Rightarrow \quad y < 0
\end{aligned}$$

Thus we find that for all $p(u \mid w', q')$, $w' \in \Omega(W)$, $q' \in \Omega(Q)$, from the CPT of U equation (6.12) only holds if $y < 0$, and it follows that for all $p(\bar{u} \mid w', q')$ equation (6.12) only holds if $y > 1$. Therefore there is no $y \in [0,1]$ for any $p(u' \mid w', q')$ from the CPT of U such that equation (6.12) holds, which means that even if the violation of monotonicity cannot be resolved by varying any parameter from the CPT of U, then it is possible to resolve the violation by varying $p(w)$ to 0.1. □

Our next example violates the condition that $\sigma(W) = \{U\}$.

Example 6. *Let B be a restricted Bayesian network with a DAG as depicted in Figure 6.2.4 with $\Omega(U) = \{\bar{u}, u\}$, $\Omega(W) = \{\bar{w}, w\}$ and $\Omega(Q) = \{\bar{q}, q\}$. Suppose there is a violation of monotonicity $viol(c', e_i^k, \mathbf{e}^-)$, where $\mathbf{e}^-$ assigns the observation q to the variable Q. Like before, varying parameters from the CPTs of U and W can only affect $\Pr\left(C \le c' \mid e_i^k, \mathbf{e}^-\right)$ and $\Pr\left(C \le c' \;\middle|\; e_i^{k+1}, \mathbf{e}^-\right)$ through $\Pr(u \mid q)$ and $\Pr(\bar{u} \mid q)$, where*

$$\Pr(U \mid q) = \sum_{\Omega(W)} p(U \mid W)\, p(q \mid W)\, p(W).$$

The following value assignments for the CPTs of U, W and Q illustrate that varying in the CPT of W may resolve the violation of monotonicity, while varying in the CPT of U does not.

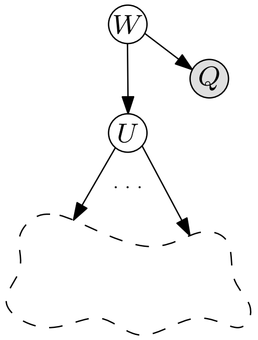

Figure 6.2.4: Directed acyclic graph used in Example 6.

Let the parameters from the CPTs of U and W be as follows:

$$p\,(w) = 0.3 \qquad p\,(q \,|\, w) = 0.7 \qquad p\,(u \,|\, w) = 0.6$$
$$p\,(q \,|\, \bar{w}) = 0.1 \qquad p\,(u \,|\, \bar{w}) = 0.8$$

Suppose that varying $p\,(w)$ to 0.55 resolves the violation of monotonicity. We compute that $\Pr\,(u \mid q) = 0.267$. If there is no value y for any $p\,(u' \,|\, w')$, $u' \in \Omega(U)$, $w' \in \Omega(W)$, from the CPT of U such that

$$\Pr\,(u \mid q)\,(p\,(u' \,|\, w') = y) = 0.267, \tag{6.13}$$

then there is no value y for any $p\,(u' \,|\, w')$ from the CPT of U such that

$$\Pr\left(C \leq c' \;\middle|\; e_i^k, \mathbf{e}^-\right)\left(p\,(w) = 0.55\right) = \Pr\left(C \leq c' \;\middle|\; e_i^k, \mathbf{e}^-\right)\left(p\,(u' \,|\, w') = y\right) \quad \textit{and}$$
$$\Pr\left(C \leq c' \;\middle|\; e_i^{k+1}, \mathbf{e}^-\right)\left(p\,(w) = 0.55\right) = \Pr\left(C \leq c' \;\middle|\; e_i^{k+1}, \mathbf{e}^-\right)\left(p\,(u' \,|\, w') = y\right).$$

The value y must be a probability and must therefore lie in the interval $[0,1]$. We show that no such value y exists for any $p\,(u' \,|\, w')$, $u' \in \Omega(U)$, $w' \in \Omega(W)$ from the CPT of U.

$$\Pr\,(u \mid q)\,(p\,(u \,|\, w) = y) = 0.21 \cdot y + 0.056 = 0.267 \qquad \Rightarrow \qquad y > 1$$
$$\Pr\,(u \mid q)\,(p\,(u \,|\, \bar{w}) = y) = 0.07 \cdot y + 0.126 = 0.267 \qquad \Rightarrow \qquad y > 1$$

Thus we find that for all $p\,(u \,|\, w')$, $w' \in \Omega(W)$, from the CPT of U equation (6.13) only holds if $y > 1$, and it follows that for all $p\,(\bar{u} \,|\, w')$ equation (6.13) only holds if $y < 0$. Therefore there is no $y \in [0,1]$ for any parameter from the CPT of U such that equation (6.13) holds, which means that even if the violation of monotonicity cannot be resolved by varying any parameter from the CPT of U, then it is possible to resolve the violation by varying $p\,(w)$ to 0.55. □

The following example violates the condition that U lies on all chains between W and C.

Example 7. *Let B be a restricted Bayesian network with a DAG as depicted in Figure 6.2.5 with $\Omega(U) = \{\bar{u}, u\}$, $\Omega(W) = \{\bar{w}, w\}$ and $\Omega(Q) = \{\bar{q}, q\}$. Suppose there is a violation of monotonicity $viol\big(c', e_i^k, \mathbf{e}^-\big)$. Varying parameters from the CPTs of U and W can only*

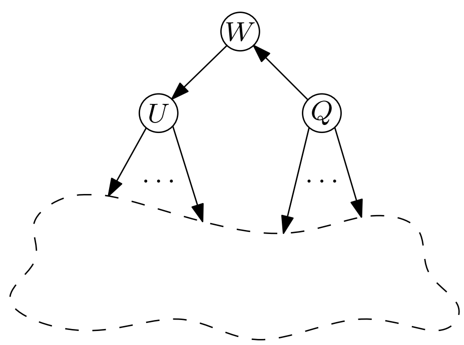

Figure 6.2.5: Directed acyclic graph used in Example 7.

affect $\Pr\left(C \le c' \mid e_i^k, \mathbf{e}^-\right)$ *and* $\Pr\left(C \le c' \,\middle|\, e_i^{k+1}, \mathbf{e}^-\right)$ *through* $\Pr(u,q)$, $\Pr(u,\bar{q})$, $\Pr(\bar{u},q)$ *and* $\Pr(\bar{u},\bar{q})$, *where*

$$\Pr(U,Q) = \sum_{\Omega(W)} p(U \mid W)\, p(W \mid Q)\, p(Q)\,.$$

The following value assignments for the CPTs of U, W and Q illustrate that varying in the CPT of W may resolve the violation of monotonicity, while varying in the CPT of U does not.

Let the parameters from the CPTs of U, W and Q be as follows:

$$\begin{array}{lll} p(u \mid w) = 0.3 & p(w \mid q) = 0.7 & p(q) = 0.2 \\ p(u \mid \bar{w}) = 0.8 & p(w \mid \bar{q}) = 0.4 & \end{array}$$

Suppose that varying $p(w \mid q)$ to 0.6 resolves the violation of monotonicity. We compute that $\Pr(u,q) = 0.1$ and $\Pr(u,\bar{q}) = 0.48$. Note that $\Pr(u,\bar{q})$ is not changed by varying $p(w \mid q)$. If there is no value y for any $p(u' \mid w')$, $u' \in \Omega(U)$, $w' \in \Omega(W)$, from the CPT of U such that

$$\begin{array}{l} \Pr(u,q)\left(p\left(u' \mid w'\right) = y\right) = 0.1 \quad \textit{and} \\ \Pr(u,\bar{q})\left(p\left(u' \mid w'\right) = y\right) = 0.48 \end{array} \tag{6.14}$$

then there is no value y for any $p(u' \mid w')$ from the CPT of U such that

$$\begin{array}{c} \Pr\left(C \le c' \,\middle|\, e_i^k, \mathbf{e}^-\right)\left(p(w) = 0.6\right) = \Pr\left(C \le c' \,\middle|\, e_i^k, \mathbf{e}^-\right)\left(p\left(u' \mid w'\right) = y\right) \quad \textit{and} \\ \Pr\left(C \le c' \,\middle|\, e_i^{k+1}, \mathbf{e}^-\right)\left(p(w) = 0.6\right) = \Pr\left(C \le c' \,\middle|\, e_i^{k+1}, \mathbf{e}^-\right)\left(p\left(u' \mid w'\right) = y\right). \end{array}$$

The value y must be a probability and must therefore lie in the interval $[0,1]$. We show that no such value y exists for any $p(u' \mid w')$, $u' \in \Omega(U)$, $w' \in \Omega(W)$, from the CPT of U.

$$\begin{array}{lll} \Pr(u,q)\left(p(u \mid w) = y\right) = 0.14 \cdot y + 0.048 = 0.1 & \Rightarrow & y = \dfrac{13}{35} \\ \Pr(u,\bar{q})\left(p(u \mid w) = y\right) = 0.32 \cdot y + 0.384 = 0.48 & \Rightarrow & y = 0.3 \end{array}$$

Since these values for y are different, the equations (6.14) do not hold for any value $y \in [0,1]$ for $p(u \mid w)$.

$$\begin{array}{lll} \Pr(u,q)\left(p(u \mid \bar{w}) = y\right) = 0.06 \cdot y + 0.042 = 0.1 & \Rightarrow & y = \dfrac{29}{30} \\ \Pr(u,\bar{q})\left(p(u \mid \bar{w}) = y\right) = 0.48 \cdot y + 0.096 = 0.48 & \Rightarrow & y = 0.8 \end{array}$$

Since these values for y are different, the equations (6.14) also do not hold for any value $y \in [0,1]$ for $p(u \mid \bar{w})$.

Thus we find that for all $p(u \mid w')$, $w' \in \Omega(W)$, from the CPT of U equations (6.14) do not hold for any $y \in [0,1]$, and it follows that for all $p(\bar{u} \mid w')$ equations (6.14) also do not hold for any $y \in [0,1]$. Therefore there is no $y \in [0,1]$ for any parameter from the CPT of U such that equations (6.14) hold, which means that even if the violation of monotonicity cannot be resolved by varying any parameter from the CPT of U, then it is possible to resolve the violation of monotonicity by varying $p(w \mid q)$ to 0.6. □

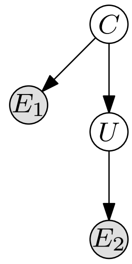

Figure 6.2.6: Directed acyclic graph used in Example 8.

Our final example deals with a fairly different situation. One might expect that, for situations like this, we can make a similar statement to the one we have been discussing, namely that if we cannot resolve a violation of monotonicity by varying parameters from the CPT of a variable U, then we also cannot resolve it by varying parameters from the CPT of a variable W. The example, however, shows that we cannot make such a statement in this situation.

Example 8. *Let B be a restricted Bayesian network with a DAG as depicted in Figure 6.2.6 with $\Omega(C) = \{c^1, c^2, c^3\}$, $\Omega(U) = \{u, \bar{u}\}$, $\Omega(E_1) = \{e_2^1, e_2^2\}$ and $\Omega(E_2) = \{e_2^1, e_2^2\}$.*

Suppose there is a violation of monotonicity $viol\left(c', e_1^1, e_2^2\right)$. We can rewrite this violation as follows.

$$
\begin{gathered}
\Pr\left(C \leq c^1 \;\middle|\; e_1^1, e_2^2\right) < \Pr\left(C \leq c^1 \;\middle|\; e_1^2, e_2^2\right) \\
\Downarrow \\
\Pr\left(c^1 \;\middle|\; e_1^1, e_2^2\right) < \Pr\left(c^1 \;\middle|\; e_1^2, e_2^2\right) \\
\Downarrow \\
\frac{\Pr\left(c^1, e_1^1, e_2^2\right)}{\Pr\left(e_1^1, e_2^2\right)} < \frac{\Pr\left(c^1, e_1^2, e_2^2\right)}{\Pr\left(e_1^2, e_2^2\right)} \\
\Downarrow \\
\Pr\left(c^1, e_1^1, e_2^2\right) \cdot \Pr\left(e_1^2, e_2^2\right) < \Pr\left(c^1, e_1^2, e_2^2\right) \cdot \Pr\left(e_1^1, e_2^2\right)
\end{gathered}
$$

If we write $\Pr\left(e_2^2 \;\middle|\; c^j\right)$ for $\sum_{\Omega(U)} p\left(e_2^2 \,|\, U\right) p\left(U \,|\, c^j\right)$, then we have that

$$
\begin{aligned}
\Pr\left(c^1, e_1^1, e_2^2\right) &= \Pr\left(e_2^2 \;\middle|\; c^1\right) p\left(e_1^1 \,|\, c^1\right) p\left(c^1\right) \\
\Pr\left(c^1, e_1^2, e_2^2\right) &= \Pr\left(e_2^2 \;\middle|\; c^1\right) p\left(e_1^2 \,|\, c^1\right) p\left(c^1\right) \\
\Pr\left(e_1^1, e_2^2\right) &= \Pr\left(e_2^2 \;\middle|\; c^1\right) p\left(e_1^1 \,|\, c^1\right) p\left(c^1\right) + \Pr\left(e_2^2 \;\middle|\; c^2\right) p\left(e_1^1 \,|\, c^2\right) p\left(c^2\right) \\
&\quad + \Pr\left(e_2^2 \;\middle|\; c^3\right) p\left(e_1^1 \,|\, c^3\right) p\left(c^3\right) \\
\Pr\left(e_1^2, e_2^2\right) &= \Pr\left(e_2^2 \;\middle|\; c^1\right) p\left(e_1^2 \,|\, c^1\right) p\left(c^1\right) + \Pr\left(e_2^2 \;\middle|\; c^2\right) p\left(e_1^2 \,|\, c^2\right) p\left(c^2\right) \\
&\quad + \Pr\left(e_2^2 \;\middle|\; c^3\right) p\left(e_1^2 \,|\, c^3\right) p\left(c^3\right)
\end{aligned}
$$

We can again rewrite the violation by substituting these probabilities.

$$\Pr\left(c^1, e_1^1, e_2^2\right) \cdot \Pr\left(e_1^2, e_2^2\right) < \Pr\left(c^1, e_1^2, e_2^2\right) \cdot \Pr\left(e_1^1, e_2^2\right)$$
$$\Downarrow$$
$$\Pr\left(e_2^2 \mid c^1\right) p\left(e_1^1 \mid c^1\right) p\left(c^1\right) \Big(\Pr\left(e_2^2 \mid c^1\right) p\left(e_1^2 \mid c^1\right) p\left(c^1\right)$$
$$+ \Pr\left(e_2^2 \mid c^2\right) p\left(e_1^2 \mid c^2\right) p\left(c^2\right) + \Pr\left(e_2^2 \mid c^3\right) p\left(e_1^2 \mid c^3\right) p\left(c^3\right)\Big)$$
$$<$$
$$\Pr\left(e_2^2 \mid c^1\right) p\left(e_1^2 \mid c^1\right) p\left(c^1\right) \Big(\Pr\left(e_2^2 \mid c^1\right) p\left(e_1^1 \mid c^1\right) p\left(c^1\right)$$
$$+ \Pr\left(e_2^2 \mid c^2\right) p\left(e_1^1 \mid c_2\right) p\left(c^2\right) + \Pr\left(e_2^2 \mid c^3\right) p\left(e_1^1 \mid c^3\right) p\left(c^3\right)\Big)$$
$$\Downarrow$$
$$\Pr\left(e_2^2 \mid c^2\right) p\left(e_1^2 \mid c^2\right) p\left(c^2\right) p\left(e_1^1 \mid c^1\right) + \Pr\left(e_2^2 \mid c^3\right) p\left(e_1^2 \mid c^3\right) p\left(c^3\right) p\left(e_1^1 \mid c^1\right)$$
$$<$$
$$\Pr\left(e_2^2 \mid c^2\right) p\left(e_1^1 \mid c^2\right) p\left(c^2\right) p\left(e_1^2 \mid c^1\right) + \Pr\left(e_2^2 \mid c^3\right) p\left(e_1^1 \mid c^3\right) p\left(c^3\right) p\left(e_1^2 \mid c^1\right)$$
$$\Downarrow$$
$$\Pr\left(e_2^2 \mid c^2\right) p\left(c^2\right) \Big(p\left(e_1^2 \mid c^2\right) p\left(e_1^1 \mid c^1\right) - p\left(e_1^1 \mid c^2\right) p\left(e_1^2 \mid c^1\right)\Big)$$
$$<$$
$$\Pr\left(e_2^2 \mid c^3\right) p\left(c^3\right) \Big(p\left(e_1^1 \mid c^3\right) p\left(e_1^2 \mid c^1\right) - p\left(e_1^2 \mid c^3\right) p\left(e_1^1 \mid c^1\right)\Big)$$
$$\Downarrow$$
$$\Pr\left(e_2^2 \mid c^2\right) p\left(c^2\right) \Big(p\left(e_1^2 \mid c^2\right) - p\left(e_1^2 \mid c^1\right)\Big) < \Pr\left(e_2^2 \mid c^3\right) p\left(c^3\right) \Big(p\left(e_1^2 \mid c^1\right) - p\left(e_1^2 \mid c^3\right)\Big)$$

Writing out $\Pr\left(e_2^2 \mid c^2\right)$ *and* $\Pr\left(e_2^2 \mid c^3\right)$*, we find that the violation of monotonicity is equivalent to the following inequality.*

$$\begin{gathered}\Big(p\left(e_2^2 \mid u\right) p\left(u \mid c^2\right) + p\left(e_2^2 \mid \bar{u}\right) p\left(\bar{u} \mid c^2\right)\Big) p\left(c^2\right) \Big(p\left(e_1^2 \mid c^2\right) - p\left(e_1^2 \mid c^1\right)\Big) \\ < \\ \Big(p\left(e_2^2 \mid u\right) p\left(u \mid c^3\right) + p\left(e_2^2 \mid \bar{u}\right) p\left(\bar{u} \mid c^3\right)\Big) p\left(c^3\right) \Big(p\left(e_1^2 \mid c^1\right) - p\left(e_1^2 \mid c^3\right)\Big)\end{gathered} \tag{6.15}$$

Using inequality (6.15) the following value assignments for the CPTs in B *illustrate that if varying a parameter from the CPT of* U *does not resolve the violation of monotonicity, then varying a parameter from the CPT of* E_2 *may still resolve it, and vice versa.*

Let the parameters from the CPTs in G *be as follows.*

$$\begin{array}{llll} p\left(e_1^2 \mid c^1\right) = 0.4 & p\left(e_2^2 \mid u\right) = 0.3 & p\left(u \mid c^2\right) = 0.4 & p\left(c^2\right) = 0.1 \\ p\left(e_1^2 \mid c^2\right) = 0.7 & p\left(e_2^2 \mid \bar{u}\right) = 0.8 & p\left(u \mid c^3\right) = 0.9 & p\left(c^3\right) = 0.7 \\ p\left(e_1^2 \mid c^3\right) = 0.3 & & & \end{array}$$

Then inequality (6.15) becomes:

$$(0.3 \cdot 0.4 + 0.8 \cdot 0.6) \cdot 0.1 \cdot (0.7 - 0.4) < (0.3 \cdot 0.9 + 0.8 \cdot 0.1) \cdot 0.7 \cdot (0.4 - 0.3)$$
$$\Downarrow$$
$$0.018 < 0.0245$$

Suppose we vary $x = p\left(e_2^2 \,|\, u\right)$ *then inequality (6.15) becomes:*

$$(x \cdot 0.4 + 0.8 \cdot 0.6) \cdot 0.1 \cdot (0.7 - 0.4) < (x \cdot 0.9 + 0.8 \cdot 0.1) \cdot 0.7 \cdot (0.4 - 0.3)$$
$$\Downarrow$$
$$0.012x + 0.0144 < 0.063 \cdot x + 0.0056$$
$$\Downarrow$$
$$0.0088 < 0.051 \cdot x$$
$$\Downarrow$$
$$\tfrac{44}{255} < x$$

Thus by varying x *such that* $x \leq \frac{44}{255}$ *the violation of monotonicity can be resolved.*

Next, suppose we vary $y = p\left(u \,|\, c^2\right)$ *then inequality (6.15) becomes:*

$$(0.3 \cdot y + 0.8 \cdot (1 - y)) \cdot 0.1 \cdot (0.7 - 0.4) < 0.0245$$
$$\Downarrow$$
$$-0.015 \cdot y + 0.024 < 0.0245$$
$$\Downarrow$$
$$-\tfrac{1}{30} < y$$

Thus the violation of monotonicity cannot be resolved by varying y *in* $[0, 1]$*. To resolve the violation we would need that* $y \leq -\frac{1}{30}$*, however this is impossible, since* y *is a probability.*

Now, suppose we vary $y = p\left(u \,|\, c^3\right)$ *then inequality (6.15) becomes:*

$$0.018 < (0.3 \cdot y + 0.8 \cdot (1 - y)) \cdot 0.7 \cdot (0.4 - 0.3)$$
$$\Downarrow$$
$$0.018 < -0.035 \cdot y + 0.056$$
$$\Downarrow$$
$$y < \tfrac{38}{35}$$

Thus the violation of monotonicity cannot be resolved by varying y *in* $[0, 1]$*. To resolve the violation we would need that* $y \geq \frac{38}{35}$*, however this is impossible, since* y *is a probability.*

Therefore the violation of monotonicity can be resolved by varying in the CPT of E_2*, while it cannot be resolved by varying in the CPT of* U*.*

Now, let the parameters in B *be as follows.*

$$\begin{array}{llll} p\left(e_1^2 \,|\, c^1\right) = 0.4 & p\left(e_2^2 \,|\, u\right) = 0.7 & p\left(u \,|\, c^2\right) = 0.9 & p\left(c^2\right) = 0.2 \\ p\left(e_1^2 \,|\, c^2\right) = 0.7 & p\left(e_2^2 \,|\, \bar{u}\right) = 0.1 & p\left(u \,|\, c^3\right) = 0.8 & p\left(c^3\right) = 0.7 \\ p\left(e_1^2 \,|\, c^3\right) = 0.3 & & & \end{array}$$

Then inequality (6.15) becomes:

$$(0.7 \cdot 0.9 + 0.1 \cdot 0.1) \cdot 0.2 \cdot (0.7 - 0.4) < (0.7 \cdot 0.8 + 0.1 \cdot 0.2) \cdot 0.7 \cdot (0.4 - 0.3)$$
$$\Downarrow$$
$$0.0384 < 0.0406$$

Suppose we vary $y = p\left(u \mid c^3\right)$ *then inequality (6.15) becomes:*

$$0.0384 < (0.7 \cdot y + 0.1 \cdot (1 - y)) \cdot 0.7 \cdot (0.4 - 0.3)$$
$$\Downarrow$$
$$0.0384 < 0.042 \cdot y + 0.007$$
$$\Downarrow$$
$$\tfrac{157}{210} < y$$

Thus by varying y *such that* $y < \frac{157}{210}$ *the violation of monotonicity can be resolved.*

Next, suppose we vary $x = p\left(e_2^2 \mid u\right)$ *then inequality (6.15) becomes:*

$$(x \cdot 0.9 + 0.1 \cdot 0.1) \cdot 0.2 \cdot (0.7 - 0.4) < (x \cdot 0.8 + 0.1 \cdot 0.2) \cdot 0.7 \cdot (0.4 - 0.3)$$
$$\Downarrow$$
$$0.054 \cdot x + 0.0006 < 0.056 \cdot x + 0.0014$$
$$\Downarrow$$
$$-0.4 < x$$

Thus the violation of monotonicity cannot be resolved by varying x *in* $[0, 1]$*. To resolve the violation we would need that* $x \leq -0.4$*, however this is impossible, since* x *is a probability.*

Finally, suppose we vary $x = p\left(e_2^2 \mid \bar{u}\right)$ *then inequality (6.15) becomes:*

$$(0.7 \cdot 0.9 + x \cdot 0.1) \cdot 0.2 \cdot (0.7 - 0.4) < (0.7 \cdot 0.8 + x \cdot 0.2) \cdot 0.7 \cdot (0.4 - 0.3)$$
$$\Downarrow$$
$$0.0378 + 0.006 \cdot x < 0.0392 + 0.014 \cdot x$$
$$\Downarrow$$
$$-0.175 < x$$

Thus the violation of monotonicity cannot be resolved by varying x *in* $[0, 1]$*. To resolve the violation we would need that* $x \leq -0.175$*, however this is impossible, since* x *is a probability.*

Therefore the violation of monotonicity can be resolved by varying in the CPT of U*, while it cannot be resolved by varying in the CPT of* E_2*.* $\square$

Chapter 7

Conclusion

In this thesis, we studied the problem of restoring monotonicity in a Bayesian network. We have provided a method to find, if it exists, a single parameter which can be varied to do so. This method, the intersection-of-intervals approach, must be applied to all relevant pairs of observations and for every parameter in the network. There may however be variables in the network for which varying a parameter cannot restore monotonicity. We have determined that this is the case for all variables which are not in the sensitivity set of the variable of interest given the observable variables. Using this information we have therefore restricted the Bayesian network to only include the variables in the sensitivity set and the observable variables. If the variable of interest is binary, then we can eliminate even more variables. To that end we introduced the concept of resolution set for an observable variable. Furthermore, we also found that if applying the intersection-of-intervals approach to one variable does not yield possible parameter variation to restore monotonicity, then there may be other variables to which we need no longer apply the intersection-of-intervals approach. Due to the graphical structure of the Bayesian network, these variables then also do not have a parameter which can be varied to restore monotonicity.

The application of our restrictions to the Bayesian network and the intersection-of-intervals approach can yield one or more parameters which can each individually be varied to restore monotonicity. We can then choose the parameter which requires the smallest amount of variation and, by doing so, minimize the changes that must be made to the Bayesian network to make it exhibit monotonicity. However, it is also possible that the application of our restrictions and intersection-of-intervals approach does not yield any parameter which may be varied to restore monotonicity to the network.

While we have limited ourselves to investigating restoring monotonicity by varying a single parameter, we expect that it would be interesting to investigate restoring monotonicity by varying multiple parameters, one after the other or simultaneously. We surmise that our methods could be used to obtain a sequence of parameters which can be varied in order to restore monotonicity, although the result may not be optimal. Another possibility would be to vary several parameters simultaneously. We expect that a fairly different method will be required to investigate this option, but that the results could be promising if such a method could be found. Finally, instead of attempting to restore monotonicity by varying one or more parameters from the quantitative part of the Bayesian network, it may be done by applying changes to the qualitative part of the network, the directed acyclic graph.

Bibliography

[1] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems. Networks of Plausible Inference*. Morgan Kaufmann, 1988.

[2] Robert G. Cowell, A. Philip Dawid, Steffen L. Lauritzen, and David J. Spiegelhalter. *Probabilistic Networks and Expert Systems: Exact Computational Methods for Bayesian Networks*. Springer Publishing Company, Incorporated, 2007.

[3] Veerle M. H. Coupé and Linda C. van der Gaag. Properties of sensitivity analysis of bayesian belief networks. *Annals of Mathematics and Artificial Intelligence*, 36(4):323–356, 2002.

[4] Hei Chan and Adnan Darwiche. When do numbers really matter? *Journal of Artificial Intelligence Research*, 17:265–287, 2001.

[5] Linda C. van der Gaag, Hans L. Bodlaender, and Ad Feelders. Monotonicity in bayesian networks. In *UAI '04: Proceedings of the 20th conference on Uncertainty in artificial intelligence*, pages 569–576, 2004.

[6] Linda C. van der Gaag, Silja Renooij, and Petra L. Geenen. Lattices for studying monotonicity of bayesian networks. In *Probabilistic Graphical Models*, pages 99–106, 2006.

[7] Linda C. van der Gaag and Uffe Kjærulff. Making sensitivity analysis computationally efficient. In *In Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 317–325, 2000.