Exploiting Causal Independence in Large Bayesian Networks

Rasa Jurgelenaite and Peter Lucas
Radboud University Nijmegen, Nijmegen, The Netherlands
Email: {rasa, peterl}@cs.ru.nl

Abstract

The assessment of a probability distribution associated with a Bayesian network is a challenging task, even if its topology is sparse. Special probability distributions based on the notion of causal independence have therefore been proposed, as these allow defining a probability distribution in terms of Boolean combinations of local distributions. However, for very large networks even this approach becomes infeasible: in Bayesian networks which need to model a large number of interactions among causal mechanisms, such as in fields like genetics or immunology, it is necessary to further reduce the number of parameters that need to be assessed. In this paper, we propose using equivalence classes of binomial distributions as a means to define very large Bayesian networks. We analyse the behaviours obtained by using different symmetric Boolean functions with these probability distributions as a means to model joint interactions. Some surprisingly complicated behaviours are obtained in this fashion, and their intuitive basis is examined.

1 Introduction

Bayesian networks offer an appealing language with associated set of tools for building models of domains with inherent uncertainty. However, a significant bottleneck in constructing Bayesian networks, whether done manually or by learning from data, is the size of their underlying probability tables. Even though adopting a sound design methodology may render the resulting graph representation of the Bayesian network relatively sparse, typically, real-world Bayesian networks include some probability tables which are really large. There are several proposals in the literature which may help reducing the size of the tables. One of the more systematic ways to cope with large probability tables is offered by the theory of causal independence; it allows decomposing a probability distribution in terms of Boolean interactions among local parameters.

As a consequence of the success of Bayesian networks in solving realistic problems, increasingly complicated situations are being tackled. We are in particular interested in the modelling of biomedical knowledge, for example in fields such as genetics and immunology; in these fields hundreds to thousands of interactions between variables may need to be captured in a probabilistic model. Clearly, such models cannot be handled without exploiting (potentially

hypothetical) knowledge about underlying causal mechanisms and associated simplifying assumptions.

The aim of the present work was to develop a theory on top of the theory of causal independence which allows defining interactions between a huge number of causal factors. This is done by assuming that the parameters in terms of which the probability distribution is defined are members of an equivalence class. We apply symmetric Boolean functions to combine first the causal factors inside an equivalence class and subsequently the effects of the equivalence classes. The probabilistic behaviour obtained in this fashion is analysed in detail.

The practical significance of such an analysis becomes clear if one realises that many practical Bayesian network models use causal independence assumptions. A well-known example is the probabilistic reformulation of the Quick Medical Reference (QMR), called QMR-DT, a very large diagnostic knowledge-based system in the area of internal medicine, in which causal independence was used to manage the complexity of the underlying Bayesian network model [8].

The remainder of this paper is organised as follows. In the following section, the basic properties of Bayesian networks, Boolean functions, and the notion of causal independence are introduced. A mathematical analysis of the behaviour of various models is given in Sections 3 and 4. The paper is rounded off by a summary of what has been achieved and by plans for future research.

2 Preliminaries

2.1 Bayesian Networks and Causal Modelling

A *Bayesian network* $\mathcal{B} = (G, \Pr)$ represents a factorised joint probability distribution on a set of variables $\mathbf{V}$. It consists of two parts: (1) a qualitative part, represented as an acyclic directed graph (ADG) $G = (\mathbf{V}(G), \mathbf{A}(G))$, whose vertices $\mathbf{V}(G)$ correspond to the random variables in $\mathbf{V}$, and arcs $\mathbf{A}(G)$ represent the conditional (in)dependencies between the variables; (2) a quantitative part Pr consisting of local probability distributions $\Pr(V \mid \pi(V))$, for each vertex $V \in \mathbf{V}(G)$ given its parents $\pi(V)$. The joint probability distribution Pr is factorised according to the structure of the graph, as follows:

$$\Pr(\mathbf{V}(G)) = \prod_{V \in \mathbf{V}(G)} \Pr(V \mid \pi(V)).$$

Each variable $V \in \mathbf{V}$ has a finite set of mutually exclusive states. In this paper, we assume all variables to be binary; as an abbreviation, we will often use v to denote $V = \top$ (true) and $\bar{v}$ to denote $V = \bot$ (false). Variables V can either act as free variables, in which case their binding is arbitrary, or they can act as bound variables, where bindings are established by associated operators. Furthermore, an expression such as

$$\sum_{\psi(I_1,\ldots,I_n)=e} g(I_1, \ldots, I_n)$$

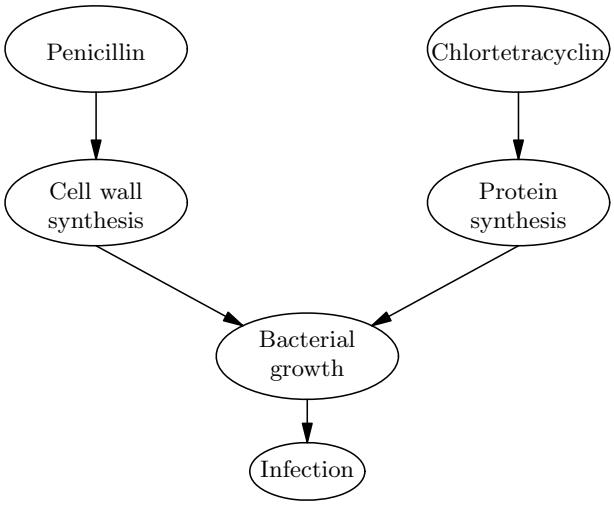

Figure 1: Example Bayesian network, modelling the interaction between the antimicrobial agents penicillin and chlortetracyclin on infection.

stands for summing over all possible values of $g(I_1, \ldots, I_n)$ for all possible values of the variables I_k for which the constraint $\psi(I_1, \ldots, I_n) = e$ holds.

Even though it is acknowledged by researchers that Bayesian networks are excellent tools for the modelling of uncertain causal mechanisms, the question remains in what way different causal mechanisms can be best modelled. Let us look at two real-world examples, which provide motivation for the approach developed in this paper.

Consider the interaction between bactericidal antimicrobial agents, i.e. drugs that kill bacteria by interference with their metabolism, resulting, for example, in fragile cell walls, and bacteriostatic antimicrobial agents, i.e. drugs that inhibit the multiplication of bacteria, for example by suppressing the production of necessary proteins. Penicillin is an example of a bactericidal drug, whereas chlortetracyclin is an example of a bacteriostatic drug. It is well known among medical doctors that the interaction between bactericidal and bacteriostatic drugs can have antagonistic effects; e.g. the drug combination penicillin and chlortetracyclin may have as little effect against an infection as prescribing no antimicrobial agent at all, even if the bacteria are susceptible to each of these drugs. The depiction of the causal interaction of the relevant variables is shown in Figure 1.

As a second example, consider the administration of chemotherapy to patients. If a patient has cancer, chemotherapy increases the chances of survival; however, if the patient does not have cancer, chemotherapy reduces the chances of survival. Clearly, the causal interaction between chemotherapy, cancer and survival has some underlying logic. This is shown schematically in Figure 2.

Although the Bayesian networks shown in figures 1 and 2 have a very similar structure, their underlying interaction semantics is very different as we will see.

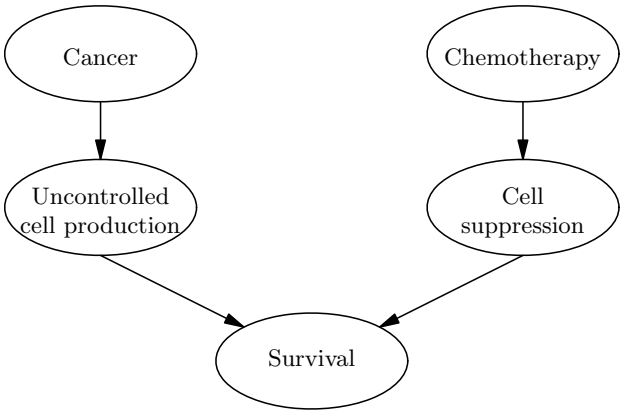

Figure 2: Example Bayesian network, modelling the interaction among cancer, chemotherapy and survival.

2.2 Probabilistic Representation

Causal independence [10], also called *noisy functional dependence* [5], is a popular way to specify interactions among cause variables. The global structure of a causal independence model is shown in Figure 3; it expresses the idea that causes $C_1, \ldots, C_n$ influence a given common effect E through intermediate variables $I_1, \ldots, I_n$ and a deterministic function f, called the *interaction function*. The impact of each cause C_k on the common effect E is independent of each other cause $C_j, j \neq k$. The function f represents in which way the intermediate effects I_k, and indirectly also the causes C_k, interact to yield the final effect E. Hence, the function f is defined in such a way that when a relationship, as modelled by the function f, between $I_k, k = 1, \ldots, n$, and $E = \top$ is satisfied, then it holds that $e = f(I_1, \ldots, I_n)$. It is assumed that $\Pr(e \mid I_1, \ldots, I_n) = 1$ if $f(I_1, \ldots, I_n) = e$, and $\Pr(e \mid I_1, \ldots, I_n) = 0$ if $f(I_1, \ldots, I_n) = \bar{e}$.

The conditional probability of the occurrence of the effect E given the causes $C_1, \ldots, C_n$, i.e. $\Pr(e \mid C_1, \ldots, C_n)$, can be obtained from the conditional probabilities $\Pr(I_k \mid C_k)$ as follows [7, 10]:

$$\Pr(e \mid C_1, \ldots, C_n) = \sum_{f(I_1,\ldots,I_n)=e} \prod_{k=1}^{n} \Pr(I_k \mid C_k). \tag{1}$$

It is assumed that absent causes do not contribute to the effect, i.e. $\Pr(i_k \mid \bar{c}_k) = 0$.

An important subclass of causal independence models is formed by models in which the deterministic function f can be defined in terms of separate binary functions g_k, also denoted by $g_k(I_k, I_{k+1})$. Such causal independence models have been called *decomposable* causal independence models [4]; these models are of significant practical importance. Usually, all functions $g_k(I_k, I_{k+1})$ are identical for each k; a function $g_k(I_k, I_{k+1})$ may therefore be simply denoted by $g(I, I')$.

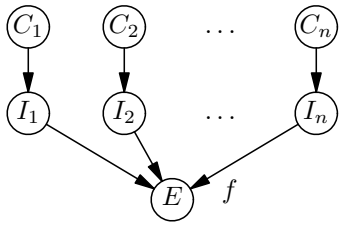

Figure 3: Causal independence model.

Well-known examples of causal independence models are the noisy-OR and noisy-AND models, where the function f represents a logical OR and a logical AND function, respectively.

2.3 Symmetric Boolean Functions

The function f in equation (1) is actually a Boolean function. However, there are 2^{2^n} different n-ary Boolean functions [2, 9]. Consequently, the potential number of causal interaction models is huge. However, in the case of causal independence it is usually assumed that the function f is decomposable to identical, binary functions. In addition, it is attractive to assume that the order of the cause variables does not matter; thus, it makes sense to restrict causal independence models to symmetric Boolean functions, where the order of arguments is irrelevant [9].

There are 8 symmetric binary Boolean functions, of which 6 commutative and associative, which we will take as a basis for defining Boolean functions of n arguments [7]. The advantage of this choice is that the order of arguments is irrelevant, as for any symmetric Boolean function, and that the resulting functions are also decomposable. Logical truth and falsity are constants, and act as the global extremes in a partial order among Boolean functions. As such they give rise to trivial causal independence models. The remaining four causal independence models are defined in terms of the logical OR, AND, XOR and bi-implication.

We use $*$ to denote a commutative, associative binary operator. A Boolean function can then also be expressed as an expression: $f_*(I_1, \ldots, I_n) = I_1 * \cdots * I_n$. Table 1 gives the truth tables for the n-nary Boolean functions of interest. From now one, the following notation is adopted: $\vee$ (OR), $\wedge$ (AND), $\otimes$ (exclusive OR), $\leftrightarrow$ (bi-implication).

Table 1: The truth tables for the n-ary symmetric Boolean functions; here we have that $k = \sum_{j=1}^{n} \nu(I_j)$, with $\nu(I_j)$ equal to 1 if I_j is equal to true and 0 otherwise.

$I_1 \vee \cdots \vee I_n$	$I_1 \wedge \cdots \wedge I_n$	$I_1 \otimes \cdots \otimes I_n$	$I_1 \leftrightarrow \cdots \leftrightarrow I_n$
$k \geq 1$	$k = n$	$odd(k)$	$even(n-k)$

We return to our example Bayesian-network models shown in figures 1 and 2. The interaction between penicillin and chlortetracyclin as depicted in Figure 1 can be described by means of an exclusive OR, $\otimes$, as presence of either of these in the patient's body tissues leads to a decrease in bacterial growth, whereas if both are present or absent, there will be little or no effect on bacterial growth. The interaction between cancer and chemotherapy as shown in Figure 2 can be described by means of a bi-implication, $\leftrightarrow$, as chances of survival are large in the case of cancer if it is being treated by chemotherapeutics, and also in the absence of cancer without treatment.

2.4 Symmetric Causal Independence Models

Recall that the function $f_\vee(I_1, \ldots, I_n)$ yields the value *true* if there is at least one variable I_j with the value *true*. Therefore, the probability distribution for the OR causal independence model is defined as follows:

$$\begin{aligned} \Pr_\vee(e \mid C_1, \ldots, C_n) &= 1 - \left(1 - \sum_{I_1 \vee \cdots \vee I_n} \prod_{k=1}^{n} \Pr(I_k|C_k)\right) \\ &= 1 - \prod_{k=1}^{n} \Pr(\bar{\imath}_k \mid C_k). \end{aligned} \tag{2}$$

The probability distribution for the AND causal independence model is defined similarly:

$$\Pr_\wedge(e \mid C_1, \ldots, C_n) = \prod_{k=1}^{n} \Pr(i_k \mid C_k). \tag{3}$$

The function $f_\otimes(I_1, \ldots, I_n)$ yields the value *true* if there are an odd number of variables I_j with the value *true*. Therefore, in order to determine the probability of the effect variable E, $\Pr(e \mid C_1, \ldots, C_n)$, the probabilities for all cause variable combinations with an odd number of present causes have to be added. We have:

$$\begin{aligned} &\Pr_\otimes(e|C_1, ..., C_n) = \sum_{I_1 \otimes \cdots \otimes I_n} \prod_{k=1}^{n} \Pr(I_k|C_k) \\ &= \Pr(\bar{\imath}_1|C_1) \cdots \Pr(\bar{\imath}_n|C_n) \sum_{\substack{1 \le k \le n \\ odd(k)}} \left(\sum_j \frac{\Pr(i_j|C_j)}{\Pr(\bar{\imath}_j|C_j)}\right)^k \\ &= \Pr(\bar{\imath}_1|C_1) \cdots \Pr(\bar{\imath}_n|C_n) \cdot \\ &\sum_{\substack{1 \le k \le n \\ odd(k)}} \sum_{j_1=1}^{n-k+1} \cdots \sum_{j_t=j_{t-1}+1}^{n-k+t} \cdots \sum_{j_k=j_{k-1}+1}^{n} \frac{\Pr(i_{j_1}|C_{j_1})}{\Pr(\bar{\imath}_{j_1}|C_{j_1})} \cdots \frac{\Pr(i_{j_k}|C_{j_k})}{\Pr(\bar{\imath}_{j_k}|C_{j_k})}. \end{aligned} \tag{4}$$

The function $f_\leftrightarrow(I_1, \ldots, I_n)$ gives the value *true* if there are an even number of variables I_j with the value *false*. Thus, to determine $\Pr(e \mid C_1, \ldots, C_n)$ the

probabilities for all cause variable combinations with an even number of absent causes have to be added:

$$\begin{aligned}
&\Pr_{\leftrightarrow}(e \mid C_1, \ldots, C_n) = \sum_{I_1 \leftrightarrow \cdots \leftrightarrow I_n} \prod_{k=1}^{n} \Pr(I_k \mid C_k) \\
&= \Pr(i_1 \mid C_1) \cdots \Pr(i_n | C_n) \cdot \\
&\left(1 + \sum_{\substack{1 \leq k \leq n \\ even(k)}} \left(\sum_j \frac{\Pr(\bar{\imath}_j \mid C_j)}{\Pr(i_j \mid C_j)} \right)^k \right) \\
&= \Pr(i_1 \mid C_1) \cdots \Pr(i_n \mid C_n) \cdot \\
&\left(1 + \sum_{\substack{1 \leq k \leq n \\ even(k)}} \sum_{j_1=1}^{n-k+1} \cdots \sum_{j_t=j_{t-1}+1}^{n-k+t} \cdots \sum_{j_k=j_{k-1}+1}^{n} \frac{\Pr(\bar{\imath}_{j_1} | C_{j_1})}{\Pr(i_{j_1} | C_{j_1})} \cdots \frac{\Pr(\bar{\imath}_{j_k} | C_{j_k})}{\Pr(i_{j_k} | C_{j_k})} \right). \quad (5)
\end{aligned}$$

The following proposition establishes the relationship between the probability distribution obtained when taking the XOR and the bi-implication, respectively, as a basis for a causal interaction model:

Proposition 1 $\Pr_{\otimes}(e \mid C_1, \ldots, C_n) = \Pr_{\leftrightarrow}(e \mid C_1, \ldots, C_n)$*, for odd*$(n)$*, and* $\Pr_{\otimes}(e \mid C_1, \ldots, C_n) = \Pr_{\leftrightarrow}(\bar{e} \mid C_1, \ldots, C_n)$*, for even*$(n)$*.*

The XOR and bi-implication causal interaction models are very sensitive to changes in the probabilities of the cause variables. If at least one cause variable is equally likely to be absent or present, the probability of the effect variable E is also equally likely to be absent or present, as is shown in the following proposition:

Proposition 2 *Let XOR and bi-implication be the Boolean functions of two causal independence models. If at least one cause variable is equally likely to be present or absent, i.e.* $\Pr(i_k \mid C_k) = \Pr(\bar{\imath}_k \mid C_k)$*, the probabilities of the effect* E *to be present or absent are also equal:*

$$\Pr_*(e \mid C_1, \ldots, C_n) = \Pr_*(\bar{e} \mid C_1, \ldots, C_n) = \frac{1}{2}$$

where $* \in \{\otimes, \leftrightarrow\}$.

The proposition indicates that the probability for one cause variable can completely dominate the probability of the effect variable E. However, the situation changes if this particular cause variable is instantiated. This property is invalid for OR and AND causal interaction models: in these models one cause variable cannot completely dominate the probability distribution for the effect variable E.

3 Grouping Probabilistic Information

Even if we use the theory of causal independence as a tool to simplify estimating a conditional probability distributions $\Pr(E \mid C_1, \ldots, C_n)$ if n is very large, the entire process becomes rapidly infeasible. However, the larger n becomes, the more likely it becomes that parameters $\Pr(I_k \mid C_k)$ of a causal independence model become arbitrary close to each other. Hence, one way to simplify the estimation of the probability distribution is to group parameters in particular equivalence classes, and to assume that the class representative $\Pr(I_k \mid C_k)$ follows a particular statistical law. In the remainder of the paper, we study the various probability distributions that are obtained in this fashion. In the case of a Bayesian network with discrete variables, taking the binomial distribution as a basis for estimation purposes seems to offer a good starting point.

3.1 The Binomial Distribution

The binomial distribution is one of the most commonly used discrete probability distributions. In an experiment which follows a binomial distribution, trials are independent and identical, with possible outcomes 'success' and 'failure', and with a probability of success that is constant.

The probability distributions of a causal independence model can be interpreted as representing a sequence of results of an experiment of n identical trials, where n is equal to the number of cause variables. From the definition above we can see that cause variables can be treated as trials of an experiment satisfying the requirements of a binomial distribution, as the number of cause variables n is known in advance, all cause variables have two states, are independent, and the probability of occurrence of each cause is the same.

3.2 Equivalence Classes of Binomial Distributions

We organise the intermediate variables $I_1, \ldots, I_n$ and their associated variables $C_1, \ldots, C_n$ by their influence on the common effect E, in accordance to the increasing order of the associated probabilistic parameters $\Pr(I_k \mid C_k)$. Next, we choose a small positive number $\varepsilon \in \mathbb{R}^+$, which determines how much the probabilities may vary inside an equivalence class. An intermediate variable I_k belongs to the t-th equivalence class if its probability of success $\Pr(i_k \mid C_k)$ falls into the interval $[2(t-1)\varepsilon, 2t\varepsilon)$. The number of equivalence classes is equal to $r = \frac{1}{2\varepsilon}$. Further, we assume that all intermediate variables from the same equivalence class have the same probability of success $\Pr(i_t \mid C_t)$ and apply the concepts of the binomial distribution to estimate the probability distribution of the t-th equivalence class $\sum_{I_{m_t} * \cdots * I_{m_t+n_t-1}} \prod_{k=m_t}^{m_t+n_t-1} \Pr(I_k \mid C_k)$, where $C_{m_t}, \ldots, C_{m_t+n_t-1}$ are the cause variables that belong to the t-th equivalence class, m_t and n_t respectively are the index of the first variable and the number of variables in the equivalence class where $\sum_{k=1}^{r} n_k = n$. In this paper we assume the class representative to be $\Pr(i_t \mid C_t) = (2t-1)\varepsilon$; however, there are

other possible ways to define the probability of success inside an equivalence class, e.g. $\Pr(i_t \mid C_t) = \frac{1}{n_t} \sum_{k=m_t}^{m_t+n_t-1} \Pr(i_k \mid C_k)$.

To determine the probability distribution of the effect variable E based on the probability distributions of contributing equivalence classes, exactly the same combining functions are employed as when combining single probability distributions $\Pr(I_k \mid C_k)$ associated with cause variables C_k.

Dependent on the Boolean function employed, the probability distribution inside an equivalence class is then determined by one of the following equations:

$$\Pr_{\vee}(e \mid C_1, \ldots, C_n) = 1 - \Pr(\bar{i}_t \mid C_t)^n \tag{6}$$

$$\Pr_{\wedge}(e \mid C_1, \ldots, C_n) = \Pr(i_t \mid C_t)^n \tag{7}$$

$$\Pr_{\otimes}(e \mid C_1, \ldots, C_n) = \sum_{\substack{1 \leq k \leq n \\ odd(k)}} \binom{n}{k} \Pr(i_t \mid C_t)^k \Pr(\bar{i}_t \mid C_t)^{n-k} \tag{8}$$

$$\Pr_{\leftrightarrow}(e \mid C_1, \ldots, C_n) = \sum_{\substack{0 \leq k \leq n \\ even(k)}} \binom{n}{k} \Pr(\bar{i}_t \mid C_t)^k \Pr(i_t|C_t)^{n-k} \tag{9}$$

4 Analysis of Probabilistic Behaviour

In this section, we study the properties of the causal independence models introduced above, and in particular we examine patterns in the resulting probability distribution as a function of the number of contributing causes. This will give us insight into the global probabilistic characteristics of large causal independence models.

Section 3 mentioned a scheme to combine the effects of the individual equivalence classes. Here it is therefore permitted to restrict the mathematical analysis to one equivalence class of binomial distributions only as the analysis for the other equivalence classes is identical. The basis of the analysis is provided by the mathematical theory of sequences and series.

Let $S_1^*, S_2^*, \ldots$ be a sequence, abbreviated to $\langle S_n^* \rangle$; throughout this section, a member S_n^* of this sequence represents a sum of products of probability distribution in an equivalence class of binomial distributions, i.e.:

$$S_n^* = \sum_{I_1 * \cdots * I_n} \prod_{t=1}^{n} \Pr(I_t \mid C_t).$$

We assume the probability $\Pr(i_t \mid C_t)$ to be constant, i.e. $p = \Pr(i_t \mid C_t)$.

In our treatment we combine various causal independence models based on similarity in behaviour. For example, the OR and AND causal independence models possess similar behaviours, which in most cases appear to be each other opposites. Analogous remarks can be made for the two other types of causal independence models. The following propositions show that OR and AND

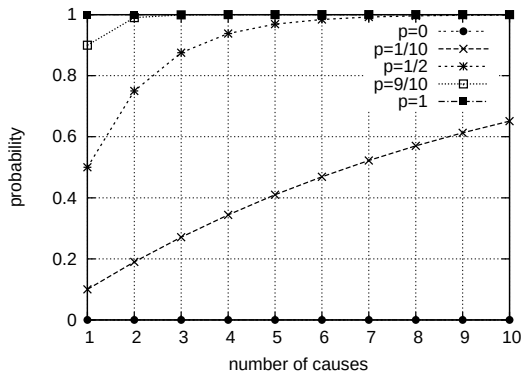

Figure 4: The patterns of the OR causal independence model.

causal independence models yield first-order behaviour, which is monotonic for any probability p with the exception of the bounds $p \in \{0, 1\}$. The proofs are omitted because of lack of space (cf. [6]).

Proposition 3 *Let $\langle S_n^* \rangle$ be a sequence as defined above. For each member S_n^* of the sequence it holds that:*

- *if $p \in (0, 1)$ then*
  - *$S_n^* \in [p, 1)$ for $* = \vee$, and*
  - *$S_n^* \in (0, p]$ for $* = \wedge$.*
- *otherwise, if $p \in \{0, 1\}$ then $S_n^* = p$ for both $* = \vee$ and $* = \wedge$.*

Proposition 4 *If $p \in (0, 1)$ then a sequence $\langle S_n^* \rangle$ is*

- *strictly monotonically increasing for $* = \vee$, and*
- *strictly monotonically decreasing for $* = \wedge$.*

It appears that the sequences converge to one of their bounds. As we try to understand the behaviour of large causal independence models, the rate of convergence is clearly also relevant. The first derivative of the function F, used to generate the sequence $S_{n+1}^* = F(S_n^*)$, can serve as a basis for this. If $* = \vee$ then $F(S_n^*) = 1 - (1 - p)S_n^*$; thus the larger the value of p, the faster the sequence converges to 1. If $* = \wedge$ then $F(S_n^*) = pS_n^*$; thus the smaller the value of p, the faster the sequence converges to 0. Figures 4 and 5 illustrate the results above by means of plots.

So far the study of the OR and AND causal models; the nature of the monotonic behaviour revealed by the propositions above and the associated plots are presumably consistent with the expectations of the reader. However, the study of the properties of the causal independence models with XOR and bi-implication interactions revealed surprisingly complicated behaviours. In addition to the expected bounds of 0 and 1, the sequences have an additional bound at $\frac{1}{2}$.

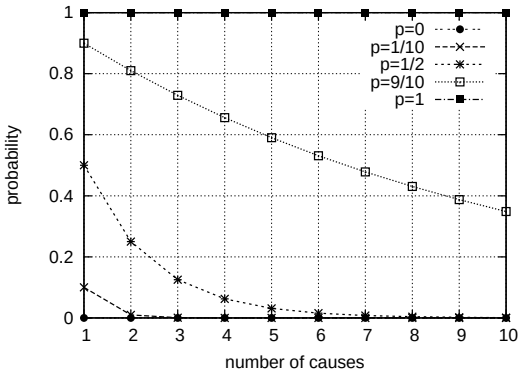

Figure 5: The patterns of the AND causal independence model.

Proposition 5 *Let $\langle S_n^* \rangle$ be a sequence as defined above. For each member S_n^* of the sequence it holds that:*

- *if $p \in [0, \frac{1}{2})$ then*
  - *$S_n^* \in [p, \frac{1}{2})$ for $* = \otimes$, and*
  - *$S_n^* \in [p, \frac{1}{2}) \cup (\frac{1}{2}, p^2 + (1-p)^2]$ for $* = \leftrightarrow$.*
- *otherwise, if $p \in (\frac{1}{2}, 1]$ then*
  - *$S_n^* \in [2p(1-p), \frac{1}{2}) \cup (\frac{1}{2}, p]$ for $* = \otimes$, and*
  - *$S_n^* \in (\frac{1}{2}, p]$ for $* = \leftrightarrow$.*

Proof: (Sketch) As the proof is by induction, we express $S_{n+1}^{\otimes}$ in terms of $S_n^{\otimes}$. Using the theory of binomial coefficients it follows that

$$\begin{aligned} S_{n+1}^{\otimes} &= \sum_{1 \leq k \leq n+1, odd(k)} \binom{n+1}{k} p^k (1-p)^{n+1-k} \\ &= (1-p)S_n^{\otimes} + p(1 - S_n^{\otimes}) = S_n^{\otimes}(1-2p) + p \end{aligned} \tag{10}$$

We have used the fact that

$$1 - S_n^{\otimes} = \sum_{0 \leq k \leq n, even(k)} \binom{n}{k} p^k (1-p)^{n-k}$$

In a similar way, we obtain the results for the bi-implication:

$$S_{n+1}^{\leftrightarrow} = S_n^{\leftrightarrow}(2p-1) + 1 - p \tag{11}$$

□

Proposition 6 *A sequence $\langle S_n^* \rangle$ is*

- *strictly monotonically increasing if $p \in (0, \frac{1}{2})$ and $* = \otimes$,*

- *strictly monotonically decreasing if* $p \in (\frac{1}{2}, 1)$ *and* $* = \leftrightarrow$,
- *constant* $S^*_n = p$ *if*
  - $p \in \{0, \frac{1}{2}\}$ *and* $* = \otimes$,
  - $p \in \{\frac{1}{2}, 1\}$ *and* $* = \leftrightarrow$,
- *non-monotonic if*
  - $p \in (\frac{1}{2}, 1]$ *and* $* = \otimes$,
  - $p \in [0, \frac{1}{2})$ *and* $* = \leftrightarrow$.

The propositions above yield insight into the behaviour of the sequences but raise questions about the behaviour non-monotonic sequences will show, i.e. when $p \in (\frac{1}{2}, 1]$, $* = \otimes$, and $p \in [0, \frac{1}{2})$, $* = \leftrightarrow$. Let the sequence $\langle S^*_n \rangle$ be divided into two sequences: $S^*_1, S^*_3, \ldots$, denoted by $\langle S^*_{odd(n)} \rangle$, and $S^*_2, S^*_4, \ldots$, denoted by $\langle S^*_{even(n)} \rangle$. We have the following proposition:

Proposition 7 *Let* $\langle S^*_{odd(n)} \rangle$ *and* $\langle S^*_{even(n)} \rangle$ *be sequences as defined above. For each member of the sequences it holds that:*

- *if* $* = \otimes$ *and* $p \in (\frac{1}{2}, 1]$ *then*
  - $S^*_{odd(n)} \in (\frac{1}{2}, p]$,
  - $S^*_{even(n)} \in [2p(1-p), \frac{1}{2})$,
- *if* $* = \leftrightarrow$ *and* $p \in [0, \frac{1}{2})$ *then*
  - $S^*_{odd(n)} \in [p, \frac{1}{2})$,
  - $S^*_{even(n)} \in (\frac{1}{2}, p^2 + (1-p)^2]$.

Proposition 8 *Let* $\langle S^*_{odd(n)} \rangle$ *and* $\langle S^*_{even(n)} \rangle$ *be sequences as defined above. Then it holds that:*

- *if* $p \in (\frac{1}{2}, 1]$ *and* $* = \otimes$
  - $\langle S^*_{odd(n)} \rangle$ *is strictly monotonically decreasing,*
  - $\langle S^*_{even(n)} \rangle$ *is strictly monotonically increasing*
- *if* $p \in [0, \frac{1}{2})$ *and* $* = \leftrightarrow$
  - $\langle S^*_{odd(n)} \rangle$ *is strictly monotonically increasing,*
  - $\langle S^*_{even(n)} \rangle$ *is strictly monotonically decreasing*

From the propositions above we conclude that despite their complicated behaviours, the sequences converge to $\frac{1}{2}$. Once again we will employ the first derivative of the function F, with $S^*_{n+1} = F(S^*_n)$, to determine the convergence rate. From the previous results (10) and (11) we know that $F(S^*_n) = (1-2p)S^*_n + p$ if $* = \otimes$ and $F(S^*_n) = S^*_n(2p-1) + 1 - p$ if $* = \leftrightarrow$. As $F'(S^*_n) = |1-2p|$ for $* \in \{\otimes, \leftrightarrow\}$ the rate of convergence depends on the value of p; the closer the value of p is to $\frac{1}{2}$, the faster the sequence converges to $\frac{1}{2}$. Figures 6 and 7 illustrate this behaviour.

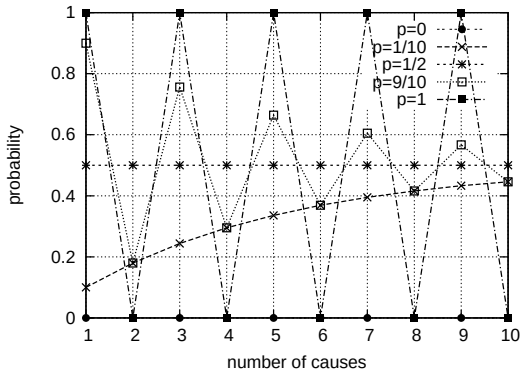

Figure 6: The patterns of the XOR causal independence model.

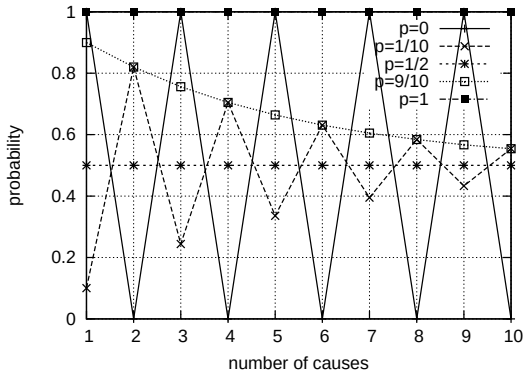

Figure 7: The patterns of the bi-implication causal independence model.

5 Discussion

In this paper, we addressed the problem of probability distribution estimation in very large Bayesian networks. Quite naturally, the theory of causal independence served as a starting point for such networks. As was argued, even if resorting to this theory, it quickly becomes infeasible to assess probability distributions for such networks. Our solution was to group local probability distributions into equivalence classes using probability intervals, and to use a suitably defined probability distribution as a basis for assessment.

The basic tools used for probability estimation were symmetric Boolean functions, which appeared to offer a natural choice as they provide a logical description of interactions between cause variables where the order between variables does not matter, and the binomial distribution, which is a standard choice in discrete probability distributions. As far as we know, this is the first paper offering a systematic analysis of the global probabilistic patterns that occur in large Bayesian networks based on the theory of causal independence.

As was shown, these types of Bayesian networks reveal surprisingly rich probabilistic patterns.

Even though the results achieved in this paper are theoretical, it should be stressed that the theory of causal independence is being used in practice in building Bayesian networks. The theory developed in this paper can be used as a basis for the construction of very large Bayesian networks, for example, in fields such as medicine, in particular internal medicine, and genetics. Although Bayesian networks have been explored in the early 1990s in such fields as part of research projects, it is only now that Bayesian networks are being adopted as tools for solving biomedical problems (cf. [3]). The theory developed in this paper could enhance the practical usefulness of the formalism.

References

[1] F.J. Díez, Parameter adjustment in Bayes networks. The generalized noisy or-gate, *Proc. UAI-93*, pp. 99–105, 1993.

[2] H.B. Enderton, *A Mathematical Introduction to Logic*, Academic Press, San Diego, 1972.

[3] N. Friedman, Inferring cellular networks using probabilistic graphical models, *Science*, 3003, pp. 799–805, 2004.

[4] D. Heckerman and J.S. Breese, A new look at causal independence, *Proc. UAI-94*, pp. 286–292, 1994.

[5] F.V. Jensen, *Bayesian Networks and Decision Graphs*, Springer-Verlag, Berlin, 2001.

[6] R. Jurgelenaite and P.J.F. Lucas, *Parameter Estimation in Large Causal Independence Models*, Technical Report, NIII, Radboud University Nijmegen, NIII-R0414, 2004.

[7] P.J.F. Lucas, Bayesian network modelling by qualitative patterns, *Proc. ECAI-2002*, pp. 690–694, 2002.

[8] M.A. Shwe, B. Middleton, D.E. Heckerman, M. Henrion, E.J. Horvitz, H.P. Lehmann and G.F. Cooper. Probabilistic diagnosis using a reformulation of the INTERNIST-1/QMR knowledge base, I – The probabilistic model and inference algorithms, *Methods Inf Med*, 30, pp. 241–255, 1991.

[9] I. Wegener, *The Complexity of Boolean Functions*, John Wiley, New York, 1987.

[10] N.L. Zhang and D. Poole, Exploiting causal independence in Bayesian networks inference, *JAIR*, 5, pp. 301–328, 1996.