Bayesian Networks: the Parental Synergy

Janneke H. Bolt
Department of Information and Computing Sciences, Utrecht University
P.O. Box 80.089, 3508 TB Utrecht, The Netherlands

Abstract

In a Bayesian network, for any node its conditional probabilities given all possible combinations of values for its parent nodes are specified. In this paper a new notion, the *parental synergy*, is introduced which can be computed from these probabilities. This paper also conjectures a general expression for the error which is found in the marginal prior probabilities computed for a node when the parents of this node are assumed to be independent. The parental synergy is an important factor of this expression; it determines to what extend the actual dependency between the parent nodes can affect the computed probabilities. This role in the expression of the prior convergence error indicates that the parental synergy is a fundamental feature of a Bayesian network.

1 Introduction

A Bayesian network is a concise representation of a joint probability distribution over a set of stochastic variables, consisting of a directed acyclic graph and a set of conditional probability distributions (Pearl, 1988). The nodes of the graph represent the variables of the distribution. From a Bayesian network, in theory, any probability of the represented distribution can be inferred. Inference, however, is NP-hard in general (Cooper, 1990) and may be infeasible for large, densely connected networks. For those networks, approximate algorithms have been designed. A widely used algorithm for approximate inference with a Bayesian network is the loopy-propagation algorithm (Pearl, 1988).

In Bolt and van der Gaag (2004), we studied the performance of the loopy-propagation algorithm from a theoretical point of view. We observed that in a network in its prior state, a prior convergence error may arise in the marginal probabilities computed for a node with two or more incoming arcs and noted that such an error may arise because the algorithm assumes the parents of a node to be independent, while, in fact, they may be dependent. Thereafter, for binary networks, we derived an expression for the prior convergence error found in a node with two incoming arcs. This expression is composed of some factors that capture the degree of dependency between the parents of this node, and of a weighting factor w that determines to what extent this degree of dependency can contribute to the prior convergence error. The factor w is composed of the conditional probabilities specified for the node.

In this paper, the notion of parental synergy is introduced. This notion is a generalisation of the factor w and can be computed for each node, irrespective its number of parents and irrespective of the cardinality of the involved nodes. Thereafter, the expression for the prior convergence error is generalised to nodes with an arbitrary number of parents and to network with nodes of arbitrary cardinality. In this generalised expression, the parental synergy fulfils the role of weighting factor. The role of the parental synergy in the expression of the prior convergence error indicates that it captures a fundamental feature of the probability landscape in a Bayesian network.

More details about the research described in this paper can be found in Bolt (2008).

2 General Preliminaries

A *Bayesian network* is a model of a joint probability distribution Pr over a set of stochastic variables $\mathbf{V}$, consisting of a directed acyclic graph and a set of conditional probability distributions[1]. Each variable A is represented by a node A in the network's digraph[2]. (Conditional) independency between the variables is captured by the digraph's set of arcs according to the d-separation criterion (Pearl 1988). The strength of the probabilistic relationships between the variables is captured by the conditional probability distributions $\Pr(A \mid \mathbf{p}(\mathbf{A}))$, where $\mathbf{p}(\mathbf{A})$ denotes the instantiations of the parents of A. The joint probability distribution is presented by:

$$\Pr(\mathbf{V}) = \prod_{A \in \mathbf{V}} \Pr(A \mid \mathbf{p}(\mathbf{A}))$$

Figure 1 depicts the graph of an example Bayesian network. The network includes a node C with n parents $A^1, \ldots, A^n$, $n \geq 0$. The nodes $A^1, \ldots, A^n$ in turn have a common parent D. For $n = 0$ and $n = 1$, no loop is included in the network. For $n = 2$, the graph consists of a simple loop. For $n > 2$, the graph consists of a compound loop; for $n = 3$, this compound loop will be termed a double loop.

For a Bayesian network with a graph as depicted in Figure 1, the marginal probability $\Pr(c_i)$ equals:

$$\Pr(c_i) = \sum_{\mathbf{A}} \Pr(c_i \mid \mathbf{A}) \cdot \Pr(\mathbf{A})$$

Wrongfully assuming independence of the parents $A^1, \ldots, A^n$ would give the approximation $\widetilde{\Pr}(c_i)$:

$$\widetilde{\Pr}(c_i) = \sum_{\mathbf{A}} \Pr(c_i \mid \mathbf{A}) \cdot \Pr(A^1) \cdot \ldots \cdot \Pr(A^n)$$

In the loopy-propagation algorithm (Pearl, 1988), a widely used algorithm for approximate inference, indeed the parents of a node are always considered to be independent. With this algorithm, in the network from Figure 1 for node C the probabilities $\widetilde{\Pr}(c_i)$ would be yielded.

In Bolt and van der Gaag (2004), we termed the error which arises in the marginal prior probabilities computed for a child node under assumption of independence of its parent nodes, a prior convergence error. Moreover, we analysed the prior convergence error found in a binary network with a graph consisting of a node C with just the parents A and B with the common parent D, as the network depicted in Figure 2.

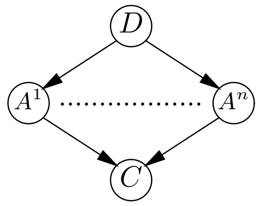

Figure 1: An example graph of a Bayesian network with a node C with the dependent parents $A^1, \ldots, A^n$.

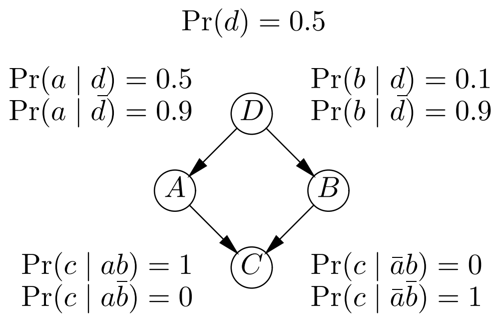

Figure 2: An example Bayesian network with a node C with the dependent parents A and B.

For the prior convergence error $v_i = \Pr(c_i) - \widetilde{\Pr}(c_i)$ in such a network the following expression was found:

$$v_i = l \cdot m \cdot n \cdot w$$

[1] Variables will be denoted by upper-case letters (A, A^i), and their values by indexed lower-case letters (a_i); sets of variables by bold-face upper-case letters ($\mathbf{A}$) and their instantiations by bold-face lower-case letters ($\mathbf{a}$). The upper-case letter is also used to indicate the whole range of values of a variable or a set of variables. Given binary variables, $A = a_1$ is often written as a and $A = a_2$ is often written as $\bar{a}$.

[2] The terms node and variable will be used interchangeably.

where

$$
\begin{aligned}
l &= \Pr(d) - \Pr(d)^2 \\
m &= \Pr(a \mid d) - \Pr(a \mid \bar{d}) \\
n &= \Pr(b \mid d) - \Pr(b \mid \bar{d}) \\
w &= \Pr(c_i \mid ab) - \Pr(c_i \mid a\bar{b}) \\
&\quad - \Pr(c_i \mid \bar{a}b) + \Pr(c_i \mid \bar{a}\bar{b})
\end{aligned}
$$

The factors were illustrated graphically with Figure 3. The line segment in this figure captures the exact probability $\Pr(c)$ as a function of $\Pr(d)$, given the conditional probabilities for the nodes A, B and C from Figure 2. $\Pr(d)$ itself is not indicated in the figure, note however, that each particular $\Pr(d)$ has a corresponding $\Pr(a)$ and $\Pr(b)$. The end points of the line segment, for example, are found at $\Pr(d) = 1$ with the corresponding $\Pr(a) = 0.5$ and $\Pr(b) = 0.1$ and at $\Pr(d) = 0$ with the corresponding $\Pr(a) = 0.9$ and $\Pr(b) = 0.9$. The surface captures $\widetilde{\Pr}(c)$ as a function of $\Pr(a)$ and $\Pr(b)$, given the conditional probabilities for node C. The convergence error equals the distance between the point on the line segment that matches the probability $\Pr(d)$ from the network and its orthogonal projection on the surface. For $\Pr(d) = 0.5$ the difference between $\Pr(c)$ and $\widetilde{\Pr}(c)$ is indicated by the vertical dotted line segment and equals $0.66 - 0.5 = 0.16$. The factor l now reflects the location of the point with the exact probability on the line segment and the factors m and n reflect the location of the line segment. The factor w, to conclude, reflects the curvature of the surface with the approximate probabilities. This curvature is determined by the change of the influence of one of the parent nodes on C occasioned by the change of the value of the other parent node. We argued that the factors l, m and n capture the degree of dependency between the parent nodes A and B and that the factor w acts as a weighting factor, determining to what extent the dependence between A and B can affect the computed probabilities.

The factor w, as used in the expression for the convergence error above, only applies to nodes with two binary parents. In the next Section, this factor is extended to a general notion which will be called the *parental synergy*. Subsequently, in Sections 4 and 5, the expression of the prior convergence error is generalised to a node C with an arbitrary number of dependent parents and with parents of arbitrary cardinality. The parental synergy is an important factor of these expressions and, analogous to the factor w, has the function of weighting factor.

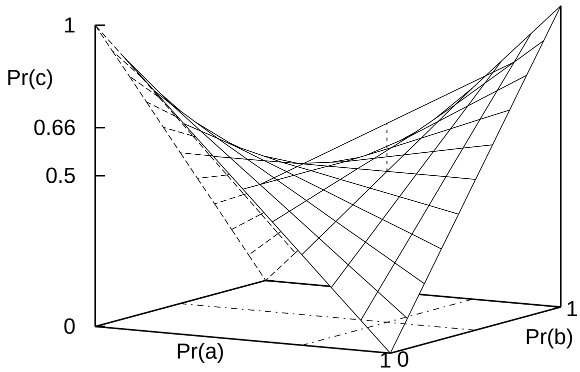

Figure 3: The line segment capturing $\Pr(c)$ and the surface capturing $\widetilde{\Pr}(c)$, for the example network from Figure 2.

3 The Parental Synergy

Before formally defining the parental synergy, the indicator function δ on the joint value assignments $a^1_{i1}, \ldots, a^n_{in}$ to a set of variables $A^1, \ldots, A^n$, $n \geq 0$, given a specific assignment $a^1_{s1}, \ldots, a^n_{sn}$ to these variables is introduced:

$$
\delta(a^1_{i1}, \ldots, a^n_{in} \mid a^1_{s1}, \ldots, a^n_{sn}) =
\begin{cases} 1 & \text{if } \sum_{k=1,\ldots,n} a^k_{ik} \neq a^k_{sk} \text{ is even} \\ -1 & \text{if } \sum_{k=1,\ldots,n} a^k_{ik} \neq a^k_{sk} \text{ is odd} \end{cases}
$$

where true $\equiv 1$ and false $\equiv 0$. The indicator function compares the joint value assignment $a^1_{i1}, \ldots, a^n_{in}$ with the joint assignment $a^1_{s1}, \ldots, a^n_{sn}$, and counts the number of differences: the assignment $a^1_{i1}, \ldots, a^n_{in}$ is mapped to the value 1 if the number of differences is even and is mapped to -1 if the number of differences is odd. For the binary variables A and B, for example, $\delta(ab \mid ab) = 1$, $\delta(a\bar{b} \mid ab) = -1$, $\delta(\bar{a}b \mid ab) = -1$ and $\delta(\bar{a}\bar{b} \mid ab) = 1$.

Building upon the indicator function δ, the notion of parental synergy is defined as follows:

Definition 3.1. Let $\mathbf{B}$ be a Bayesian network, representing a joint probability distribution Pr over a set of variables $\mathbf{V}$. Let $\mathbf{A} = \{A^1, \ldots, A^n\} \subseteq \mathbf{V}, n \geq 0$, and let $C \in \mathbf{V}$ such that C is a child of all variables in the set $\mathbf{A}$, that is, $A^j \rightarrow C, j = 1, \ldots, n$. Let $\mathbf{a}$ be a joint value assignment to $\mathbf{A}$ and let c_i be a value of C. Furthermore, let $\mathbf{X} \subseteq \rho(C) \backslash \mathbf{A}$, where $\rho(C)$ denotes the parents of C, and let $\mathbf{x}$ be a value assignment to $\mathbf{X}$. The *parental synergy* of $\mathbf{a}$ with respect to c_i given $\mathbf{X} = \mathbf{x}$, denoted as $Y^\star_{\mathbf{x}}(\mathbf{a}, c_i)$, is

$$Y^\star_{\mathbf{x}}(\mathbf{a}, c_i) = \sum_{\mathbf{A}} \delta(\mathbf{A} \mid \mathbf{a}) \cdot \Pr(c_i \mid \mathbf{Ax})$$

□

Given an an empty value assignment to the nodes $\mathbf{X}$, the parental synergy is denoted by $Y^\star(\mathbf{a}, c_i)$.

Example 3.2. Consider an arbitrary-valued node C with the two ternary parents A and B; the conditional probabilities for the value c_i of C given A and B, are listed in Table 1. The parental synergy $Y^\star(a_2b_2, c_i)$ of a_2 and b_2 with respect to c_i, for example, is computed from $\Pr(c_i \mid a_1b_1) - \Pr(c_i \mid a_1b_2) + \Pr(c_i \mid a_1b_3) - \Pr(c_i \mid a_2b_1) + \Pr(c_i \mid a_2b_2) - \Pr(c_i \mid a_2b_3) + \Pr(c_i \mid a_3b_1) - \Pr(c_i \mid a_3b_2) + \Pr(c_i \mid a_3b_3) = 2.0$. Table 2 lists all parental synergies $Y^\star(a_jb_k, c_i), j, k = 1, 2, 3$. □

Table 1: The conditional probabilities $\Pr(c_i \mid AB)$ for a variable C with the ternary parents A and B.

$\Pr(c_i \mid AB)$	a_1	a_2	a_3
b_1	0.7	0.2	0.3
b_2	0.2	1.0	0.8
b_3	0.4	0.1	0.9

Table 2: The parental synergies matching the conditional probabilities $\Pr(c_i \mid AB)$ from Table 1.

$Y^\star(AB, c_i)$	a_1	a_2	a_3
b_1	2.4	0.4	−0.6
b_2	−1.2	2.0	−0.2
b_3	0.8	−0.4	1.4

From the definition of parental synergy, it is readily seen that for a binary parent A^k of C, we have that $Y^\star_{\mathbf{x}}(\mathbf{a}a^k, c_i) = -Y^\star_{\mathbf{x}}(\mathbf{a}\bar{a}^k, c_i)$ for any value assignments $\mathbf{a}$ and $\mathbf{x}$. For a node C with binary parents only, therefore, for a given $\mathbf{x}$ the parental synergies with respect to some value c_i can only differ in sign. From the definition it further follows that given a binary parent A^k of C, with the values a^k_m and a^k_n, that $Y^\star_{\mathbf{x}}(\mathbf{a}a^k_m, c_i) = Y^\star_{\mathbf{x}a^k_m}(\mathbf{a}, c_i) - Y^\star_{\mathbf{x}a^k_n}(\mathbf{a}, c_i)$.

Example 3.3. Consider an arbitrary-valued node C with the ternary parent A and the binary parent B; the conditional probabilities for the value c_i of C given A and B, are listed in Table 3; the matching parental synergies are listed in Table 4. It is easily verified that $Y^\star(Ab, c_i)$ equals $-Y^\star(A\bar{b}, c_i)$ for all possible values of A. Furthermore from, for example, $Y^\star(a_1b, c_i) = (0.8 - 0.3 - 0.5) - (0 - 0.4 - 0.9) = 1.3$, $Y^\star_b(a_1, c_i) = 0.8 - 0.3 - 0.5 = 0$ and $Y^\star_{\bar{b}}(a_1, c_i) = 0 - 0.4 - 0.9 = -1.3$, it is readily verified that $Y^\star(a_1b, c_i) = Y^\star_b(a_1, c_i) - Y^\star_{\bar{b}}(a_1, c_i)$. □

Table 3: The conditional probabilities $\Pr(c_i \mid AB)$ for a node C with the ternary parent A and the binary parent B.

$\Pr(c_i \mid AB)$	a_1	a_2	a_3
b	0.8	0.3	0.5
$\bar{b}$	0.0	0.4	0.9

For a node with only binary parents, the parental synergies can be thought of as a measure of the feasible changes in its probability landscape, given a change in the value of one of its parents. This is explained in more detail below. When no parents are involved, the parental

Table 4: The parental synergies matching the conditional probabilities $\Pr(c_i \mid AB)$ from Table 3.

$Y^\star(AB, c_i)$	a_1	a_2	a_3
b	1.3	−0.5	−1.1
$\bar{b}$	−1.3	0.5	1.1

synergy with respect to a value c_i of a node C equals zero, reflecting that no change is possible. For a node C with a parent A, the parental synergy of, for example, a with respect to c_i equals $\Pr(c_i \mid a) - \Pr(c_i \mid \bar{a})$ and thus gives the feasible change in the probability of c_i, given a change of the value of A from $\bar{a}$ to a. For a node C with parents A and B, the parental synergy of, for example, ab with respect to c_i equals $\big(\Pr(c_i \mid ab) - \Pr(c_i \mid \bar{a}b)\big) - \big(\Pr(c_i \mid a\bar{b}) - \Pr(c_i) \mid \bar{a}\bar{b})\big)^3$. It thus gives the difference between the feasible change of the probability of c_i given a change of the value of A from $\bar{a}$ to a when the value of B is b and when the value of B is $\bar{b}^4$. And so on. Note that the number of parent nodes involved in the computation of the parental synergy can be considered to determine a kind of 'dimensionality' of the synergy. Note furthermore that for multiple-valued variables, the interpretation of the parental synergy as a measure of feasible change does not hold straightforwardly. Consider, for example, a three valued parent A with a child C. Given, for example, the conditional probabilities $\Pr(c \mid a_1) = 0.7$, $\Pr(c \mid a_2) = 0.7$ and $\Pr(c \mid a_3) = 0.7$, the parental synergy of a_i with respect to c equals $0.7 - 0.7 - 0.7 = -0.7$, whereas no change in the probability of c can be occasioned by a change in the value of A. Note, to conclude, that the parental synergy is related to the concepts of qualitative influence and additive synergy as defined for qualitative probabilistic networks by Wellman (1990). Most obviously, in a binary network, given a node C with a single parent A, the sign of the qualitative influence between A and C is computed from $\Pr(c \mid a) - \Pr(c \mid \bar{a})$, which equals $Y^\star_{\mathbf{x}}(a, c)$; given a node C with just the parents A and B the sign of the additive synergy of A and B with respect to C is computed from $\Pr(c \mid ab) - \Pr(c \mid a\bar{b}) - \Pr(c \mid \bar{a}b) + \Pr(c \mid \bar{a}\bar{b})$, which equals $Y^\star_{\mathbf{x}}(ab, c)$.

[3] Or equally $\big(\Pr(c_i \mid ab) - \Pr(c_i \mid a\bar{b})\big) - \big(\Pr(c_i \mid \bar{a}b) - \Pr(c_i \mid \bar{a}\bar{b})\big)$.

[4] Or equally: it gives the difference between the feasible change of the probability of c_i given a change of the value of B from $\bar{b}$ to b when the value of A is a and when the value of A is $\bar{a}$

4 The Convergence Error given Binary Nodes

In section 2 an expression for the prior convergence error found in the marginal prior probabilities of a node C with the two binary parent nodes A and B with a common binary parent D was given. In Section 4.1 an alternative for this expression is stated. This alternative expression is more apt for generalisation. It will be extended to apply to convergence nodes with more than two binary parents in Section 4.2 and it will be extended to multiple valued nodes in Section 5.

4.1 Two Parent Nodes; an Alternative Expression

The expression for the prior convergence error from Section 2 can also be written as

$$v_i = (s - t) \cdot w$$

where w is as before and

$$\begin{aligned} s &= \sum_D \Pr(a \mid D) \cdot \Pr(b \mid D) \cdot \Pr(D) \\ t &= \left(\sum_D \Pr(a \mid D) \cdot \Pr(D)\right) \cdot \\ &\qquad\qquad \left(\sum_D \Pr(b \mid D) \cdot \Pr(D)\right) \end{aligned}$$

The degree of dependency between the nodes A and B now is captured by $s - t$ instead of by $l \cdot m \cdot n$. Note that the term s equals $\Pr(ab)$ and that the term t equals $\Pr(a) \cdot \Pr(b)$. Note furthermore that $w = \Pr(c_i \mid ab) - \Pr(c_i \mid a\bar{b}) - \Pr(c_i \mid \bar{a}b) + \Pr(c_i \mid \bar{a}\bar{b})$ equals $Y^\star(ab, c_i)$.

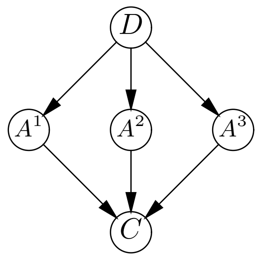

Figure 4: A node C with the dependent parents A^1, A^2 and A^3.

4.2 Multiple Parent Nodes

Consider the network from Figure 4. For three parent nodes, the expression for the prior convergence error v_i is found by subtracting the approximate probability

$$\widetilde{\Pr}(c_i) = \sum_{A^1,A^2,A^3} \Pr(c_i \mid A^1A^2A^3) \cdot \Pr(A^1) \cdot \Pr(A^2) \cdot \Pr(A^3)$$

from the exact probability

$$\Pr(c_i) = \sum_{A^1,A^2,A^3,D} \Pr(c_i \mid A^1A^2A^3) \cdot \Pr(A^1 \mid D) \cdot \Pr(A^2 \mid D) \cdot \Pr(A^3 \mid D) \cdot \Pr(D)$$

and manipulating the resulting terms. This results in the following expression

$$\begin{aligned} v_i \quad = \quad & (s_{a^1a^2a^3} - t_{a^1a^2a^3}) \cdot w + \\ & (s_{a^2a^3} - t_{a^2a^3}) \cdot w_{\bar{a}^1}(a^2a^3) + \\ & (s_{a^1a^3} - t_{a^1a^3}) \cdot w_{\bar{a}^2}(a^1a^3) + \\ & (s_{a^1a^2} - t_{a^1a^2}) \cdot w_{\bar{a}^3}(a^1a^2) \end{aligned}$$

where

$$\begin{aligned} s_{a^1a^2a^3} &= \sum_D \prod_{i=1,2,3} \Pr(a^i \mid D) \cdot \Pr(D) \\ t_{a^1a^2a^3} &= \prod_{i=1,2,3} \sum_D \Pr(a^i \mid D) \cdot \Pr(D) \\ w &= Y^\star(a^1a^2a^3, c_i) \end{aligned}$$

$$\begin{aligned} s_{a^ma^n} &= \sum_D \Pr(a^m \mid D) \cdot \Pr(a^n \mid D) \cdot \Pr(D) \\ t_{a^ma^n} &= \left(\sum_D \Pr(a^m \mid D) \cdot \Pr(D) \right) \cdot \left(\sum_D \Pr(a^n \mid D) \cdot \Pr(D) \right) \\ w_{\bar{a}^l}(a^ma^n) &= Y^\star_{\bar{a}^l}(a^ma^n, c_i) \end{aligned}$$

The convergence error is composed of the term $(s_{a^1a^2a^3} - t_{a^1a^2a^3}) \cdot w$ which pertains to the entire double loop, and the three terms $(s_{a^ma^n} - t_{a^ma^n}) \cdot w_{\bar{a}^l}(a^ma^n)$ which pertain to the three simple loops that are included within the double loop. Note that $s_{a^1a^2a^3}$ equals $\Pr(a^1a^2a^3)$; $t_{a^1a^2a^3}$ equals $\Pr(a^1) \cdot \Pr(a^2) \cdot \Pr(a^3)$; $s_{a^ma^n}$ equals $\Pr(a^ma^n)$ and $t_{a^ma^n}$ equals $\Pr(a^m) \cdot \Pr(a^m)$.

Now consider a convergence node C with the binary parents $A^1, \ldots, A^n$ and the common parent D of $A^1, \ldots, A^n$. It is posed as a conjecture that the following expression captures the prior convergence error v_i for the value c_i of C:

$$v_i \quad = \quad \sum_{\mathbf{m}} (s_{\mathbf{a^m}} - t_{\mathbf{a^m}}) \cdot w_{\bar{a}^1 \ldots \bar{a}^n \setminus \mathbf{a^m}}(\mathbf{a^m})$$

where

$$\begin{aligned} &\mathbf{m} \in \mathcal{P}(\{1, \ldots, n\}) \\ &\mathbf{a^m} = a^x \ldots a^y \quad \text{for} \quad \mathbf{m} = \{x, \ldots, y\} \\ &s_{\mathbf{a^m}} = \sum_D \prod_{i \in \mathbf{m}} \Pr(a^i \mid D) \cdot \Pr(D) \\ &t_{\mathbf{a^m}} = \prod_{i \in \mathbf{m}} \sum_D \Pr(a^i \mid D) \cdot \Pr(D) \\ &w_{\bar{a}^1 \ldots \bar{a}^n \setminus \mathbf{a^m}}(\mathbf{a^m}) = Y^\star_{\bar{a}^1 \ldots \bar{a}^n \setminus \mathbf{a^m}}(\mathbf{a^m}, c_i) \end{aligned}$$

in which $\bar{a}^1 \ldots \bar{a}^n \setminus \mathbf{a^m}$ denotes the value assignment 'False' to the nodes included in the set $\{A^1, \ldots, A^n\} \setminus \{A^x, \ldots, A^y\}$. This expression is a straightforward generalising of the expression for the prior convergence error given three parent nodes. Note that, analogous to before, the term $s_{\mathbf{a^m}}$ equals $\Pr(a^x \ldots a^y)$ and the term $t_{\mathbf{a^m}}$ equals $\Pr(a^x) \cdot \ldots \cdot \Pr(a^y)$. Further note that the expression includes terms for all possible loops included in the compound loop. The term with $\mathbf{m} = 1, \ldots, n$, pertains to the entire compound loop. With $|\mathbf{m}| = n - 1$, the n compound loops with a single incoming arc of C deleted are considered, and so on. Note also that, if the number of elements of $\mathbf{m}$ is smaller than two, just one parent or no parents are left; the term $s_{\mathbf{a}^\mathbf{m}}$ then equals the term $t_{\mathbf{a}^\mathbf{m}}$ and $(s_{\mathbf{a}^\mathbf{m}} - t_{\mathbf{a}^\mathbf{m}}) \cdot w_{\bar{a}^1 \ldots \bar{a}^n \setminus \mathbf{a}^\mathbf{m}}(\mathbf{a}^\mathbf{m})$ equals zero.

5 The Convergence Error given Multiple-valued Nodes

In the generalisation of the expression of the prior convergence error to multiple-valued nodes, a preliminary observation is that the expressions for this error from Section 4 involve just a single value c_i of the convergence node and therefore are valid for multiple-valued convergence nodes as well. Furthermore is observed that these expressions also provide for a multiple-valued node D. In this section expressions for the convergence error given parent nodes with an arbitrary number of values are proposed.

5.1 Two Parent Nodes

Consider a Bayesian network with a graph as the graph of network from Figure 2. It is posed as a conjecture that the following expression captures the prior convergence error for C.

$$v_i = \sum_{A,B} (s_{AB} - t_{AB}) \cdot w(AB)/4$$

where

$$\begin{aligned}
s_{AB} &= \sum_{D} \Pr(A \mid D) \cdot \Pr(B \mid D) \cdot \Pr(D) \\
t_{AB} &= \left(\sum_{D} \Pr(A \mid D) \cdot \Pr(D) \right) \cdot \\
&\qquad \left(\sum_{D} \Pr(A \mid D) \cdot \Pr(D) \right) \\
w(AB) &= Y^{\star}(AB, c_i)
\end{aligned}$$

This conjecture was supported by the fact that for several example networks the expression indeed yielded the prior convergence error.

Note that, analogous to the binary case, the term s_{AB} equals $\Pr(AB)$ and the term t_{AB} equals $\Pr(A) \cdot \Pr(B)$. In contrast with the binary case, now all different value combinations of the nodes A and B are considered. The impact of the dependency between a specific combination of values for the nodes A and B on the convergence error is determined by the parental synergy of this combination with respect to the value c_i of node C. Further note that the expression for the convergence error now includes a division by a constant. This constant equals 2^n, where n is the number of loop parents of the convergence node.

5.2 Multiple Parent Nodes

In Section 4.2, the expression for the convergence error was extended to convergence nodes with an arbitrary number of binary parent nodes and in Section 5.1 the expression was extended to convergence nodes with two multiple valued parent nodes. Now, it is posed as a conjecture, that these expressions combine into the following general expression for the prior convergence error. Given a network with a graph, consisting of a convergence node C with the parents $A^1, \ldots, A^n$ and the common parent D of $A^1, \ldots, A^n$, as the graph depicted in Figure 1, the convergence error equals

$$\begin{aligned}
v_i = \sum_{\mathbf{m}} \Big[\sum_{\mathbf{A}^\mathbf{m}} \Big((s_{\mathbf{A}^\mathbf{m}} - t_{\mathbf{A}^\mathbf{m}}) \cdot \\
\sum_{A^1, \ldots, A^n \setminus \mathbf{A}^\mathbf{m}} w_{A^1, \ldots, A^n \setminus \mathbf{A}^\mathbf{m}}(\mathbf{A}^\mathbf{m}) \Big) \Big] / 2^n
\end{aligned}$$

where

$$\begin{aligned}
&\mathbf{m} \in \mathcal{P}(\{1, \ldots, n\}) \\
&\mathbf{A}^\mathbf{m} = A^x, \ldots, A^y, \ \mathbf{m} = \{x, \ldots, y\} \\
&s_{\mathbf{A}^\mathbf{m}} = \sum_{D} \prod_{i \in \mathbf{m}} \Pr(A^i \mid D) \cdot \Pr(D) \\
&t_{\mathbf{A}^\mathbf{m}} = \prod_{i \in \mathbf{m}} \sum_{D} \Pr(A^i \mid D) \cdot \Pr(D) \\
&w_{A^1 \ldots A^n \setminus \mathbf{A}^\mathbf{m}}(\mathbf{A}^\mathbf{m}) = Y^{\star}_{A^1 \ldots A^n \setminus \mathbf{A}^\mathbf{m}}(\mathbf{A}^\mathbf{m}, c_i)
\end{aligned}$$

Again, this conjecture was supported by the fact that for several example networks the expression indeed yielded the prior convergence error.

Note that, as before, the term $s_{\mathbf{A^m}}$ equals $\Pr(A^x \dots A^y)$ and the term $t_{\mathbf{A^m}}$ equals $\Pr(A^x) \cdot \ldots \cdot \Pr(A^y)$. As in the binary case given multiple parent nodes, all combinations of parent nodes are considered, now however, for each combination of parent nodes also all combinations of value assignments to these parent nodes have to be taken into account. Again, if the number of elements of $\mathbf{m}$ is smaller than two, that is, if just one parent or zero parents are considered, then the term $s_{\mathbf{A^m}}$ equals the term and $t_{\mathbf{A^m}}$ and thus $\sum_{\mathbf{A^m}} \Big((s_{\mathbf{A^m}} - t_{\mathbf{A^m}}) \cdot \sum_{A^1,\dots,A^n \setminus \mathbf{A^m}} w_{A^1,\dots,A^n \setminus \mathbf{A^m}}(\mathbf{A^m}) \Big)$ equals zero. The general expression, shows that, also in the general case, the parental synergy is the weighting factor that determines the impact of the degree of dependency between the parent nodes for a given value assignment, as reflected by $s_{\mathbf{A^m}} - t_{\mathbf{A^m}}$ on the size of the convergence error.

6 Discussion

In this paper the notion of parental synergy was introduced. This synergy is computed from the conditional probabilities as specified for a node in a Bayesian network. For a node with binary parents, the parental synergies can be thought of as a measure of the feasible changes in its probability landscape, given a change in the value of one of its parents. Moreover, a general expression for the prior convergence error, was proposed. A prior convergence error arises in the prior marginal probabilities computed for a node when its parents are considered to be independent. This type of error arises in the probabilities computed by the loopy-propagation algorithm; a widely used algorithm for approximate probabilistic inference. The expression of the prior convergence error for a node is composed of the parental synergies of this node and of terms that capture the degree of dependence between its parent nodes. The parental synergy acts as weighting factor determining the impact of the degree of dependency between the parent nodes on the size of the convergence error.

In this paper, the parental synergy just features in the expression of the prior convergence error. Its role as weighting factor in this expression, however, indicates that the parental synergy captures a fundamental characteristic of the probability landscape of a Bayesian network. It is conceivable, therefore, that the parental synergy has a wider range of application.

Acknowledgments

The research reported in this thesis was supported by the Netherlands Organisation for Scientific Research (NWO). I would like to thank Linda van der Gaag for her useful comments on earlier drafts.

References

J.H. Bolt, L.C. van der Gaag. 2004. The convergence error in loopy propagation. Paper presented at *the International Conference on Advances in Intelligent Systems: Theory and Applications.*

J.H. Bolt. 2008. *Bayesian Networks: Aspects of Approximate Inference.* PhD thesis, Department of Information and Computing Sciences. Utrecht University.

G.F. Cooper. 1990. The computational complexity of probabilistic inference using Bayesian belief networks. *Artificial Intelligence*, 42:393–405.

J. Pearl. 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufmann Publishers, Palo Alto.

M.P. Wellman. 1990. Fundamental concepts of qualitative probabilistic networks. *Artificial Intelligence*, 44:257–303.