# Certainty-factor-like Structures in Bayesian Belief Networks

Peter Lucas

Department of Computer Science, Utrecht University
Padualaan 14, 3584 CH Utrecht, The Netherlands
E-mail: lucas@cs.uu.nl

### Abstract

The certainty-factor model was one of the most popular models for the representation and manipulation of uncertain knowledge in the early rule-based expert systems of the 1980s. After the model was criticised by researchers in artificial intelligence and statistics as being ad-hoc in nature, researchers and developers have stopped looking at the model. Nowadays, it is often stated that the model is merely interesting from a historical point of view. Its place has been taken over by more expressive formalisms for the representation and manipulation of uncertain knowledge, in particular by the formalism of Bayesian belief networks. In this paper, it is shown that this view underestimates the importance of the principles underlying the certainty-factor model. In particular, it is shown that certainty-factor-like structures occur frequently in practical Bayesian network models as causal independence assumptions. In fact, the noisy-OR and noisy-AND models, two probabilistic models frequently employed, appear to be reinventions of combination functions previously introduced as part of the certainty-factor model. This insight may lead to a reappraisal of the certainty-factor model.

*Keywords*: Bayesian networks, certainty-factor model, causal independence, noisy-OR model, noisy-AND model.

## 1   Introduction

In the early rule-based expert systems as developed in the 1980s, the representation and manipulation of uncertain knowledge was accomplished by various ad-hoc schemes. Typical examples of such schemes are the certainty-factor calculus of Shortliffe and Buchanan [4, 20] and the subjective Bayesian method [7]. At the time in particular the certainty-factor model enjoyed much popularity, possibly due to its mathematical and computational simplicity.

After the introduction of more expressive, and mathematically sound, probabilistic methods for the representation and manipulation of uncertainty the early methods have been criticised, sometimes severely, by researchers. Examples of such criticism are easily found in the literature; a selection is shown below (explanations added for the purpose of this paper are written in parentheses). For example, Heckerman states ([9], page 309):

> "... it is recommended that those who intend to build a system incorporating the certainty factor model consider these more general techniques (i.e. Bayesian belief networks)."

whereas Neapolitan states ([17], pages 70–71):

> "The calculus for combining certainty factors ... is of interest primarily for its historical significance."

Jensen's opinion is even stronger ([14], page 3):

> "... it is not possible to capture reasoning with uncertainty with inference rules for production rules."

In contrast, the criticism of Pearl is more moderate ([18], page 6):

> "The price attached to extensional systems (e.g. rule-based systems) is that they often yield updating that is incoherent, i.e. subject to surprise and counterintuitive conclusions."

Nowadays, the majority of AI and uncertainty in AI researchers probably agrees with the opinions summarised above. They believe that the framework of Bayesian belief networks offers a coherent, expressive, and flexible formalism for the representation and manipulation of uncertain knowledge, and there seems to be little reason to revert to the use of the early models of uncertainty. The author of this paper shares this opinion. Yet, the situation is not as clear-cut as it appears, and researchers should be aware of this.

In this paper it is shown that the certainty-factor model has in fact been reintroduced by Bayesian network theoreticians without knowing, and is used as an essential ingredient of many practical network models. It even appears that fragments of the certainty-factor model are currently very popular for the representation and manipulation of uncertainty. We think the results of this study to be important, because it shows that particular probabilistic models have been introduced by various research traditions independently, indicating their general significance. Furthermore, this study also sheds some light on current Bayesian belief network modelling practice. As we have not been able to find similar results in the literature, we believe this result to be new.

The structure of the remainder of this paper is as follows. In the next section, the certainty-factor model is introduced. Next, we study various probabilistic models that seem good candidates for their mapping to fragments of the certainty-factor model. To what extent such mappings are possible is also investigated. The practical significance of the various probabilistic models selected is next illustrated by a number of real-life applications as found in the literature. The paper is rounded off by a discussion of the practical consequences of the results achieved in this paper.

## 2    The certainty-factor model

In this section, the basic principles of the certainty-factor model, so far as needed for the reading of this paper, are briefly reviewed.
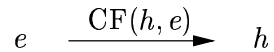
### 2.1    Certainty factors

The certainty-factor model was introduced by Shortliffe and Buchanan as a method for the representation and manipulation of uncertain knowledge in the rule-based expert system

MYCIN [20, 4], and later incorporated, in slightly modified form, in the prototypical rule-based expert-system shell EMYCIN [4]. The basic idea underlying the method is that when representing knowledge as production rules of the form **if** $e$ **then** $h_x$ **fi**, a measure of uncertainty $x$ is associated with the hypothesis $h$, expressing the degree to which the observation of evidence $e$ influences the confidence in $h$. In developing the certainty-factor model Shortliffe and Buchanan have chosen two basic measures of uncertainty: the *measure of belief* $\mathrm{MB}(h, e)$ expressing the degree to which an observed piece of evidence $e$ increases the belief in a hypothesis $h$, and the *measure of disbelief* $\mathrm{MD}(h, e)$, expressing the degree to which an observed piece of evidence $e$ decreases the belief in a hypothesis $h$. Each of these measures lie in the closed interval [0,1].

The measure of belief $\mathrm{MB}(h, e)$ and the measure of disbelief $\mathrm{MD}(h, e)$ are defined in terms of probability theory as relative changes with respect to the prior probability $\mathrm{Pr}(h)$, based on available evidence $e$. Although intuitively attractive at first sight, Heckerman [9] and Van der Gaag [22] showed that this choice renders the certainty-factor model inconsistent with the basic axioms of probability theory. Heckerman, however, has been able to find alternative definitions for these measures in terms of likelihood ratios, yielding mathematically sound probabilistic interpretations of the model [9].

A *certainty factor* $\mathrm{CF}(h, e)$ is just a numerical measure between $-1$ and $+1$, defined in terms of measures of belief and disbelief. The actual definition is not relevant for this paper (cf. [15]). A negative certainty factor indicates that the hypothesis $h$ is disconfirmed by the evidence $e$; a positive certainty factor indicates that the hypothesis $h$ is confirmed by the evidence $e$. A certainty factor equal to zero indicates that the evidence $e$ does not influence the belief in the hypothesis $h$. In most implementations of the certainty factor model, the measures of belief $\mathrm{MB}(h, e)$ and disbelief $\mathrm{MD}(h, e)$ are no longer used; only the certainty factor is employed. Consequently, with each production rule **if** $e$ **then** $h$ **fi** is now associated a certainty factor $\mathrm{CF}(h, e)$; this can also be depicted as a directed graph, as shown below:

$$
e \quad \xrightarrow{\ \mathrm{CF}(h, e)\ } \quad h
$$

## 2.2 Combination functions

For the manipulation of certainty factors, Shortliffe and Buchanan have defined a number of combination functions, expressed in terms of certainty factors. For an extensive motivation underlying their design, the reader is referred to [4].

The combination function for the propagation of uncertain evidence from the antecedent of a production rule to its consequences, as shown in the following directed graph

$$
e' \quad \xrightarrow{\ \mathrm{CF}(e, e')\ } \quad e \quad \xrightarrow{\ \mathrm{CF}(h, e)\ } \quad h
$$
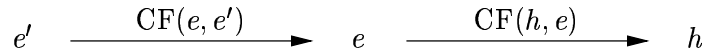
is the following:

$$
\mathrm{CF}(h, e') = \mathrm{CF}(h, e) \cdot \max\{0, \mathrm{CF}(e, e')\} \tag{1}
$$

Here, $\mathrm{CF}(h, e)$ is the certainty factor associated with the hypothesis $h$ by the production rule **if** $e$ **then** $h$ **fi** if the evidence $e$ has been observed with absolute certainty; $\mathrm{CF}(e, e')$ indicates the actual confidence in $e$ based on some prior evidence $e'$, and acts as a weighting factor to

$\text{CF}(h, e)$. If the rule's antecedent is false, the resulting weighting factor will be 0, as indicated in formula (1).

The function for combining two certainty factors $\text{CF}(e_1, e')$ and $\text{CF}(e_2, e')$ of two constituting pieces of evidence $e_1$ and $e_2$ to obtain a certainty factor for the conjunction $e_1$ **and** $e_2$ of these pieces of evidence, as shown in the directed graph below
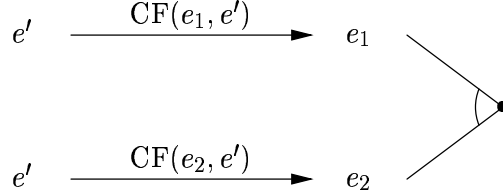


is the following:

$$\text{CF}(e_1 \textbf{ and } e_2, e') = \min\{\text{CF}(e_1, e'), \text{CF}(e_2, e')\} \tag{2}$$

For the disjunction of two pieces of evidence, we have the following formula:

$$\text{CF}(e_1 \textbf{ or } e_2, e') = \max\{\text{CF}(e_1, e'), \text{CF}(e_2, e')\} \tag{3}$$

The combination functions (2) and (3) are commutative and associative in their first argument; so the order in which conjunctions and disjunctions are evaluated has no effect on the resulting certainty factor.

Finally, the combination function for combining two certainty factors $\text{CF}(h, e'_1)$ and $\text{CF}(h, e'_2)$ which have been derived from two co-concluding production rules **if** $e_i$ **then** $h$ **fi**, $i = 1, 2$, as shown in the following directed graph
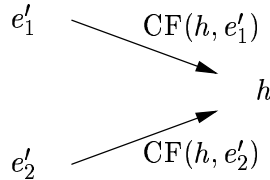


is as follows:

$$\text{CF}(h, e'_1 \textbf{ co } e'_2) = \begin{cases} x + y(1 - x) & \text{if } x, y > 0 \\ \frac{x+y}{1-\min\{|x|,|y|\}} & \text{if } -1 < xy \leq 0 \\ x + y(1 + x) & \text{if } x, y < 0 \end{cases} \tag{4}$$

where $\text{CF}(h, e'_1) = x$ and $\text{CF}(h, e'_2) = y$. Combination function (4) is commutative and associative in its second argument; so, the order in which production rules are applied has no effect on the final result.

As mentioned above, Heckerman has proposed a number of suitable transformations of the certainty-factor model to probability theory [9]; it is therefore known that the model, with slightly altered definitions, permits a probabilistic interpretation. Although important as a result, this work is still in line with the basic ideas of Shortliffe and Buchanan, in the sense that the original model was the central focus of the study; the meaning of certainty factors as changes to belief states was left unchanged. Our goals are different, because we employ the certainty-factor model as a tool in the analysis of Bayesian belief network models. Thus,
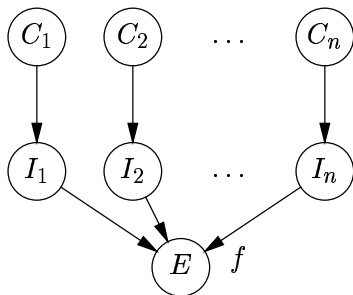
Figure 1: Causal independence model.

we are actually interested in the inverse problem. Taking solving the inverse problem as our aim, it seems unnecessary complicated to maintain the idea of certainty factors as modelling belief changes, and we shall therefore abandon this idea. In facts, although certainty factors were originally defined in terms of probability theory, they have never been used in this way, not even in the MYCIN system for which the model was originally developed.

Below, it is investigated under which conditions Bayesian belief network models correspond to ingredients of the certainty-factor model.

# 3 Certainty-factor interpretation of Bayesian belief networks

Above, we have summarised the principles of the certainty-factor model. In this section, we trace probabilistic models that correspond to fragments of the certainty-factor model.

## 3.1 Employed notation

Stochastic variables will be denoted by upper-case letter, e.g. $X$; values of variables will be denoted by lower-case letters, e.g. $x$. In the case of binary variables, the value $X = true$ will be denoted by $X = x$, or simply $x$; the value $X = false$ is denoted by $X = \neg x$, or simply $\neg x$. All variables are assumed to be discrete. By an expression like

$$\sum_{f(X_1,\ldots,X_n)=c} \psi(X_1,\ldots,X_n)$$

is indicated a summation of function values of a function $\psi$, ranging over all possible values of the variables $X_1,\ldots,X_n$ satisfying the functional constraint $f(X_1,\ldots,X_n) = c$. A probability distribution will be denoted by Pr.

## 3.2 Causal independence

In building Bayesian belief networks for practical purposes it has been argued by several researchers that considering the assumption of *causal independence* may be very fruitful [8, 10, 11, 12]. The global structure of a causal-independence model is shown in Figure 1; 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 $f$. The influence of each cause $C_k$ on the common effect $E$ is independent of each 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 a final effect $E$. Hence, this function $f$ is defined in such way that when a relationship as modelled

5

by the function $f$ between $I_k = i_k$, $k = 1, \ldots, n$, and $E = e$ is satisfied, then it holds that $e = f(i_1, \ldots, i_n)$.

In terms of probability theory, the notion of causal independence can be formalised for a *distinguished value* $e$ of $E$ as follows:

$$\Pr(e|C_1, \ldots, C_n) = \sum_{f(I_1, \ldots, I_n) = e} \Pr(e|I_1, \ldots, I_n) \Pr(I_1, \ldots, I_n|C_1, \ldots, C_n)$$

meaning that the causes $C_1, \ldots, C_n$ influence the common effect $E$ through the intermediate effects $I_1, \ldots, I_n$ only when $e = f(I_1, \ldots, I_n)$ for certain values $I_k = i_k$, $k = 1, \ldots, n$. Under this condition, it is assumed that $\Pr(e|i_1, \ldots, i_n) = 1$; otherwise, when $f(i_1, \ldots, i_n) = e' \neq e$, it holds that $\Pr(e|i_1, \ldots, i_n) = 0$. Note that the effect variable $E$ is conditionally independent of $C_1, \ldots, C_n$ given the intermediate variables $I_1, \ldots, I_n$, and that each variable $I_k$ is only dependent on its associated variable $C_k$; hence, it holds that

$$\Pr(e|I_1, \ldots, I_n, C_1, \ldots, C_n) = \Pr(e|I_1, \ldots, I_n)$$

and

$$\Pr(I_1, \ldots, I_n|C_1, \ldots, C_n) = \prod_{k=1}^{n} \Pr(I_k|C_k)$$

The formula above can now be simplified to:

$$\Pr(e|C_1, \ldots, C_n) = \sum_{f(I_1, \ldots, I_n) = e} \prod_{k=1}^{n} \Pr(I_k|C_k) \tag{5}$$

Based on the assumptions above, it also holds that

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

Formula (5) above is practically spoken not very useful, because the size of the specification of the function $f$ is exponential in the number of its arguments. The resulting probability distribution is therefore in general computationally intractable, both in terms of space and time requirements. An important subclass of causal independence models, however, 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 [11]; 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 $g(I, I')$. Typical examples of decomposable causal independence models are the noisy-OR [5, 8, 13, 18, 21] and noisy-MAX [5, 12, 21] models. Other popular causal independence models are the noisy-AND and noisy-MIN models. These models will be studied below.


## 3.3   Noisy-OR model

As shown above, it is the precise definition of the function $f$ of a causal independence model that distinguishes one model from the other. Hence, for simplicity's sake, it is allowed to discard intermediate variables. Also, since only the use of decomposable causal independence
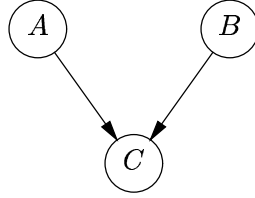
6

Figure 2: Three-node model.

models is practically feasible, it seems justified to merely consider three-node models as shown in Figure 2, without loss of generality.

Now, suppose that the three variables $A$, $B$ and $C$ are binary variables, and assume that node $C$ represents a logical OR, i.e.:

$$\Pr(c|A, B) = \begin{cases} 0 & \text{if } A = \neg a \text{ and } B = \neg b \\ 1 & \text{otherwise} \end{cases}$$

The binary functions $g_k$ as defined above all correspond to a logical OR, i.e. $g(i, i') = g_k(i, i') = (i \vee i')$, for each $k$. The marginal probability $\Pr(C)$ is now obtained as follows:

$$\begin{aligned} \Pr(c) &= \sum_{A,B} \Pr(c|A, B)\,\Pr(A, B) \\ &= \Pr(a)\,\Pr(b) + \Pr(\neg a)\,\Pr(b) + \Pr(a)\,\Pr(\neg b) \end{aligned}$$

since the variables $A$ and $B$ are independent. We can rewrite this result in two different ways:

1. $\Pr(a)\,\Pr(b) + \Pr(\neg a)\,\Pr(b) + \Pr(a)\,\Pr(\neg b) = \Pr(a) + \Pr(b)(1 - \Pr(a))$, which corresponds to the combination function for co-concluding production rule (4) in the certainty-factor model for positive certainty factors;

2. $\Pr(a)\,\Pr(b) + \Pr(\neg a)\,\Pr(b) + \Pr(a)\,\Pr(\neg b) = 1 - \Pr(\neg a, \neg b) = 1 - \Pr(\neg a)\,\Pr(\neg b)$ (because, $\Pr(\neg a, \neg b)$ was discarded in the sum above), which is a well-known formula used to define the noisy-OR model [18].

Note that $\Pr(\neg c) = 1 - \Pr(c) = \Pr(\neg a)\,\Pr(\neg b)$ is unspecified in the certainty-factor model, because certainty factors only concern positive literals. In principle, however, it might be dealt with by the method discussed below in Section 3.5. Actually, this limitation is not too important, because we take probability theory as our starting point. Hence, it appears that the noisy-OR model has exactly the same mathematical structure as combination function (4) for co-concluding production rules of the certainty-factor model for binary variables.

## 3.4   Noisy-MAX model

When we drop the restriction that the variables in the three-node model must be binary, the probabilistic interpretation changes. Without loss of generality, let us assume that $A$, $B$ and $C$ are ternary variables. Furthermore, assume that the values of the given variables satisfy the following linear order: $a_1 > a_2 > a_3$; $a_1 = b_1 = c_1$, $a_2 = b_2 = c_2$; $a_3 = b_3 = c_3$. Also assume that the function $g(A, B) = C$ that underlies a particular causal independence model

7

|       | $a_1$       | $a_2$       | $\cdots$ | $a_n$       |
|-------|-------------|-------------|----------|-------------|
| $b_1$ | $a_1, b_1$  | $a_2, b_1$  | $\cdots$ | $a_n, b_1$  |
| $b_2$ | $a_1, b_2$  | $a_2, b_2$  | $\cdots$ | $a_n, b_2$  |
| $\vdots$ | $\vdots$ | $\vdots$  | $\vdots$ | $\vdots$    |
| $b_n$ | $a_1, b_n$  | $a_2, b_n$  | $\cdots$ | $a_n, b_n$  |

Figure 3: Enumeration of elements from which is selected for the computation of the probability distribution $\Pr(C)$.

defines a MAX function in accordance with the given linear order. The function $g$ is thus defined as follows:

$$g(A, B) = \begin{cases} c_1 & \text{if } A = a_1 \text{ or } B = b_1 \\ c_2 & \text{if } A = a_2 \text{ or } B = b_2, \text{ and} \\ & \quad A \neq a_1 \text{ and } B \neq b_1 \\ c_3 & \text{otherwise} \end{cases}$$

The resulting probabilistic model is known as the *noisy-MAX* model [5, 12].

Using this definition of $g$, it holds that

$$\Pr(c_1) = \sum_{g(A,B)=c_1} \Pr(A)\Pr(B)$$
$$= \sum_{A=a_1 \vee B=b_1} \Pr(A)\Pr(B)$$

Now, since

$$\sum_{A,B} \Pr(A)\Pr(B) = 1$$

it follows that

$$\Pr(c_1) = 1 - \sum_{A \neq a_1, B \neq b_1} \Pr(A)\Pr(B)$$
$$= 1 - \Pr(\neg a_1)\Pr(\neg b_1)$$

The last expression can be rewritten to $\Pr(c_1) = \Pr(a_1) + \Pr(b_1)(1 - \Pr(a_1))$, which corresponds to the combination function for co-concluding production rules for positive certainty factors. Similarly, applying the definition of the noisy-MAX model to $\Pr(c_2)$ yields: $\Pr(c_2) = \Pr(a_2)(\Pr(b_2) + \Pr(b_3)) + \Pr(b_2)\Pr(a_3)$. However, according to the combination function for co-concluding production rules, we should have obtained: $\Pr'(c_2) = \Pr(a_2) + \Pr(b_2)(1 - \Pr(a_2)) = \Pr(c_2) + \Pr(a_1)\Pr(b_2) + \Pr(a_2)\Pr(b_1)$; the two results differ.

The difference between the two results can be explained in terms of the matrix shown in Figure 3. Using the noisy-MAX model, the probability $\Pr(c_i)$ is defined as the sum of the probability $\Pr(a_i, b_i)$ on the diagonal of the matrix, supplemented with elements below and to the right of the element $(a_i, b_i)$, i.e.

$$\Pr(c_i) = \Pr(a_i, b_i) + \sum_{j=i+1}^{n} (\Pr(a_i, b_j) + \Pr(a_j, b_i))$$

8

In the case of the combination function for co-concluding production rules, all elements on the $i$th row and column are added, with the diagonal element $\Pr(a_i, b_i)$ just added once, resulting in:

$$\Pr'(c_i) = \sum_{j=1}^{n} (\Pr(a_i, b_j) + \Pr(a_j, b_i)) - \Pr(a_i, b_i)$$

As a consequence, particular numbers $\Pr(a_i, b_j)$ will not only contribute to $\Pr'(c_i)$, but also to $\Pr'(c_j)$, $i \neq j$, i.e. they are counted twice as part of the probability distribution $\Pr'(C)$. The resulting probability distribution $\Pr'$ is therefore inconsistent.

An interesting question is whether the probability distribution $\Pr'$ can be made consistent, and, if so, in which way. The causal independence model that corresponds to the certainty-factor model appears to be defined by the following function $g'$:

$$g'(A, B) = \begin{cases} c_1 & \text{if } A = a_1 \text{ or } B = b_1 \\ c_2 & \text{if } A = a_2 \text{ or } B = b_2 \\ c_3 & \text{if } A = a_3 \text{ or } B = b_3 \end{cases}$$

which is only identical to $g$ when $g'(A, B) = c_1$. Although the function $g'$ does not take the order of the values of the variables $C$, $B$ and $A$ into account, it is still very much alike the noisy-MAX model.

The probability distribution $\Pr'$ can be rendered consistent by uniformly distributing the contribution of probabilities $\Pr(a_i, b_j)$ among $\Pr'(c_i)$ and $\Pr'(c_j)$, $i \neq j$. This manipulation results in:

$$\Pr'(c_i) = \Pr(a_i)\Pr(b_i) + \frac{1}{2} \sum_{A = a_i \oplus B = b_i} \Pr(A)\Pr(B)$$

where $\oplus$ denotes the exclusive OR operator. This equation is equal to $\frac{1}{2}(\Pr(a_i) + \Pr(b_i))$, but not in general equal to $\frac{1}{2}(\Pr(a_i) + \Pr(b_i)(1 - \Pr(a_i)))$. Note that it holds that

$$\begin{aligned} \sum_{i=1}^{n} \Pr'(c_i) &= \frac{1}{2} \sum_{i=1}^{n} (\Pr(a_i) + \Pr(b_i)) \\ &= 1 \end{aligned}$$

Hence, $\Pr'$ is now consistent. We thus have shown that it is in general not possible to map the noisy-MAX model for non-binary variables to the combination functions of the certainty-factor model by a similarity transformation, i.e. a transformation $T : U \to V$, such that $T(u) = cv + b$, $b, c \in \mathbb{R}$.

It appears that the causal independence model for non-binary variables using combination function (4) for co-concluding production rules is related to the noisy-MAX model, but nevertheless different for all but the probability of the value of the variable $C$ that is highest in the given linear order. The resulting model was shown to be probabilistically inconsistent. It is not difficult to resolve the inconsistency, yielding a model that is still related to the noisy-MAX model. However, the resulting probabilistic model cannot be represented in terms of the certainty-factor model.

## 3.5  Noisy-AND model

The noisy-OR appears to be one of the most popular probabilistic models used in building practical Bayesian belief networks. However, this model is not suitable when one is primarily interested in modelling the *conjunctive* effect of particular causes. This idea naturally leads to the concept of the noisy-AND model. Using again the topology of the graph depicted in Figure 2, the noisy-AND model can be defined in terms of a probabilistic representation of the logical AND:

$$\Pr(c|A, B) = \begin{cases} 1 & \text{if } A = a \text{ and } B = b \\ 0 & \text{otherwise} \end{cases}$$

where $A$, $B$ and $C$ are binary stochastic variables. It now follows that:

$$\begin{aligned} \Pr(c) &= \sum_{A,B} \Pr(c|A, B) \Pr(A, B) \\ &= \Pr(a) \Pr(b) \end{aligned}$$

because $A$ and $B$ are independent. This result does not correspond to the combination function for co-concluding production rules (4). However, it is possible to represent the noisy-AND model in terms of the combination function for the propagation of evidence (1), using two production rules as follows:

> **if** $a$ **then** $b_{\Pr(b)}$ **fi**
> **if** $b$ **then** $c_{1.0}$ **fi**

given the probability distributions $\Pr(A)$ and $\Pr(B)$, and assuming that the subscript 1.0 attached to $c$ represents a certainty factor. Using combination function (1) twice yields the required result for $C$.

### 3.5.1  Noisy-MIN model

As with the noisy-OR model, it is possible to generalise the noisy-AND model for non-binary variables; the result is known as the *noisy-MIN model*. Assume that the stochastic variables $A$, $B$ and $C$ are ternary variables, with values ordered as in Section 3.4. Also assume that the functions $g(A, B) = C$ underlying the causal independence model respects the linear order of values, yielding the minimum of its arguments as a result:

$$g(A, B) = \begin{cases} c_3 & \text{if } A = a_3 \text{ or } B = b_3 \\ c_2 & \text{if } A = a_2 \text{ or } B = b_2, \text{ and} \\ & \quad A \neq a_3 \text{ and } B \neq b_3 \\ c_1 & \text{otherwise} \end{cases}$$

Using this definition of $g$ yields the following result for $\Pr(c_1)$:

$$\begin{aligned} \Pr(c_1) &= \sum_{g(A,B)=c_1} \Pr(A) \Pr(B) \\ &= \sum_{A=a_1 \wedge B=b_1} \Pr(A) \Pr(B) \\ &= \Pr(a_1) \Pr(b_1) \end{aligned}$$
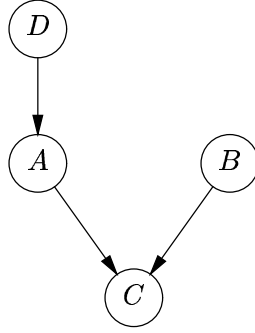
Figure 4: Propagation of evidence.

This probability can be mapped to the certainty-factor model in the same way as done for the noisy-AND model discussed above.

In general we have that

$$\Pr(c_i) = \Pr(a_i, b_i) + \sum_{j=1}^{i-1}(\Pr(a_i, b_j) + \Pr(a_j, b_i))$$

which expresses that $\Pr(c_i)$ is obtained as the sum of $\Pr(a_i, b_i)$ on the diagonal of Figure 3, supplemented with elements $\Pr(a_i, b_j)$ on the $i$th row to the left of element $(a_i, b_i)$ and elements $\Pr(a_j, b_i)$ on the $i$th column above element $(a_i, b_i)$ of Figure 3.

For $\Pr(c_2)$ we thus get:

$$\Pr(c_2) = \Pr(a_2)\Pr(b_2) + \Pr(a_2)\Pr(b_1) + \Pr(a_1)\Pr(b_2)$$

No mapping exists in this case, and the same is true for $\Pr(c_3)$.

Similar to the noisy-MAX model, a mapping of the full noisy-MIN model to the certainty-factor model is only possible for binary variables, in which case the noisy-MIN model and the noisy-AND model are identical.

## 3.6 Propagation of evidence

Consider the probabilistic network shown in Figure 4; it is identical to the network model in Figure 2, except that a node $D$ is added. Let us assume that the corresponding variable $D$ only influences the uncertainty with respect to $A$ for the distinguished value $d$ ($D$ is present); if $D$ is absent, i.e. $\neg d$, it holds that $A$ cannot occur, formally: $\Pr(a|\neg d) = 0$. We only consider the noisy-OR probabilistic model of causal independence in this section.

Now, when assuming that evidence with respect to $D$ has been observed with certainty, it holds that:

$$
\begin{aligned}
\Pr(C|D) &= \sum_{A,B}\Pr(C, A, B|D) \\
&= \sum_{A,B}\Pr(C|A, B)\Pr(A|B, D)\Pr(B|D) \\
&= \sum_{A,B}\Pr(C|A, B)\Pr(A|D)\Pr(B)
\end{aligned}
$$

by the (conditional) independence information represented in the network. Let node $C$ again model a logical OR, as defined in Section 3.3. Under this condition, it holds that when $D = d$:

$$\Pr(c|d) = \Pr(a|d)\Pr(b) + \Pr(a|d)\Pr(\neg b) + \Pr(\neg a|d)\Pr(b)$$

which is equal to $\Pr(a|d) + \Pr(b)(1 - \Pr(a|d))$. This equation combines the effects of combination function (1), the propagation of evidence from $D$ to $A$, and combination function (4) for co-concluding production rules. Note that $\Pr(c|\neg d) = \Pr(\neg a|\neg d)\Pr(b) = \Pr(b)$, which result corresponds to the situation when the rule **if** $d$ **then** $a$ **fi** fails; only $B$ contributes to $C$. This result is again in accordance with the certainty-factor model.

Next, assume that $D$ is not known with certainty. We then get:

$$
\begin{aligned}
\Pr(C) &= \sum_{A,B,D} \Pr(C, A, B, D) \\
&= \sum_{A,B,D} \Pr(C, A, B|D)\Pr(D)
\end{aligned}
$$

Substituting results from the derivation of $\Pr(C|D)$ above in this equality, we obtain:

$$\Pr(C) = \sum_{A,B,D} \Pr(C|A, B)\Pr(A|D)\Pr(B)\Pr(D)$$

Again, using the causal independence assumption of the logical OR, the following result is obtained:

$$
\begin{aligned}
\Pr(c) &= \Pr(c|d)\Pr(d) + \Pr(c|\neg d)\Pr(\neg d) \\
&= (\Pr(a|d) + \Pr(b)(1 - \Pr(a|d))\Pr(d) + \Pr(b)\Pr(\neg d) \\
&= \Pr(a|d)\Pr(d) + \Pr(b)(1 - \Pr(a|d)\Pr(d))
\end{aligned}
$$

This result corresponds again to the successive application of combination functions (1) and (4) in the certainty-factor model.

## 4 Practical significance

Above we have seen that important Bayesian belief-network models, or parts of such models, can be mapped to fragments of the certainty-factor model. However, the results of this paper would have little significance, when in almost all practical belief-network models the assumptions underlying decomposable, causal independence would not be satisfied. But the opposite seems to be the case: in many practical Bayesian belief network models, a lot of causal independence assumptions are made. This is to be expected, because the technology of Bayesian belief networks is only practically useful when a large amount of information concerning independence, with causal independence as a special case, is available in a domain. A number of actual network models, as described in the literature, is briefly discussed to illustrate the point.

Heckerman and colleagues have described a probabilistic network for printer trouble shooting [12]. The structure of this network seems quite typical for networks used for detecting hardware faults. Assumptions of causal independence are rather essential ingredients of the network, and the arguments developed in this paper seems to fully apply to this network.

Díez and colleagues [6] have developed a Bayesian belief network for the diagnosis of heart disease, in which both the noisy-OR and noisy-MAX models are used to represent interactions among causes. Another, well-known example, is the probabilistic reformulation of Internist-1/QMR, often referred to as QMR-DT (Decision-Theoretic reformulation of QMR), which uses the same assumptions [16].

Another interesting consequence of the results of this paper is that it at least partially explains the similarity in the conclusions of the assessment of the sensitivity of the MYCIN system to changes in certainty factors that was carried out in the 1970s [4], and a recent major study of the sensitivity of Bayesian belief networks to changes in their underlying probability distribution [19]. In both studies, it was concluded that the advice produced by the systems was rather insensitive to changes in the underlying numbers. However, the Bayesian belief networks that were studied incorporated noisy-OR and noisy-MAX models for the representation of interactions. Hence, the structural assumptions of these two studies were quite similar, which has not been recognised before.

Of course, causal independence does not play such an important role in all Bayesian belief networks. However, even in networks in which the notions of causal independence has not been adopted as a central modelling paradigm, e.g. the MUNIN network [2], such assumptions appear to underly a large part of the probabilistic assessments.

Although many Bayesian belief networks heavily rely on the assumption of causal independence, this does not mean that the certainty-factor calculus would be sufficient to manipulate such networks. This only holds when it is prevented that stochastic dependencies are introduced due to the entering of evidence. In networks purely used for diagnosis, and not for prediction purposes, causes of problems usually reside in the upper part of the network; findings that may be observed for a given case reside in the lower part of the network. When particular findings concerning a problem case are entered into the lower part of the network, new probabilistic dependencies would be induced. This could be handled by dynamically adapting the structure of the network, in such way that it explicitly reflects the new dependencies. Although this conclusion limits the practical usefulness of the certainty-factor *calculus* for such applications, it still holds that the underlying probabilistic model would be similar or even identical to fragments of the certainty-factor *model*.

In applications of prediction and planning it is in principle possible to use the certainty-factor calculus as a method for probabilistic inference if the structure of the network follows the principles discussed in this paper. An example of such a network, meant to assist medical specialists in the treatment of infectious diseases, is described in [3]. This observation is interesting from the perspective of the design of efficient algorithms for probabilistic inference [1, 5, 23, 24].

## 5 Discussion

As has been said at the beginning of this paper, it is not our intention to promote renewed popularity of the use of the certainty-factor model. Given the current state of research, this would be absurd. Still, the conclusion that the certainty-factor model is more important than most researcher would think possible seems inescapable. There can be learnt something from the early models of uncertainty in terms of probabilistic structures that have general significance.

We have studied the mapping of probabilistic structures to fragments of the certainty-

factor model. As was to be expected, only very specific probabilistic models can be dealt with in this way. For some probabilistic models a mapping was shown to exist; for other models, the mapping was only partial. Nevertheless, these structures appear to be of major practical importance to developers of Bayesian belief networks for specific problem domains. This explains why we believe that the results of this paper ought to be common knowledge to artificial-intelligence researchers.

We finally would like to plea for a more balanced view of the certainty-factor model: much too often researchers have expressed opinions about the model that are not supported by scientific facts.

# References

[1] J.M. Agosta. Conditional inter-causally independent node distributions, a property of noisy-or models. In: *Proceedings of the 7th Conference on Uncertainty in Artificial Intelligence*, 1991, 9–19.

[2] S. Andreassen, M. Woldbye, B. Falck and S.K. Andersen. MUNIN — A causal probabilistic network for interpretation of electromyographic findings. In: *Proceedings of the 10th International Joint Conference on Artificial Intelligence*, 1987, 366–372.

[3] S. Andreassen, C. Riekehr, B. Kristensen, H.C. Schønheyder and L. Leibovici. Using probabilistic and decision-theoretic methods in treatment and prognosis modeling, *Artificial Intelligence in Medicine* 15, 1999, 121–134.

[4] B.G. Buchanan and E.H. Shortliffe (Eds.). *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project* (Addison-Wesley Publishing Company, Reading, Massachusetts, 1984).

[5] F.J. Díez. Parameter adjustment in Bayes networks: the generalized noisy OR-gate. In: *Proc of the 9th Conference on Uncertainty in Artificial Intelligence*, 1993, 99–105.

[6] F.J. Díez, J. Mira, E. Iturralde and S. Zubillage. DIAVAL: a Bayesian expert system for echocardiography, *Artificial Intelligence in Medicine*, 1997, 10, 59–73.

[7] R.O. Duda, P.E. Hart and N.J. Nilsson. Subjective Bayesian methods for rule-based inference systems. In: *AFIPS Conference Proceedings of the 1976 National Computer Conference*, 45, 1075–82.

[8] I.J. Good. *Good Thinking: the Foundations of Probability and its Applications* (University of Minnesota Press: Mineapolis, MN, 1983).

[9] D. Heckerman. Probabilistic interpretations for MYCIN's certainty factors. In: L.N. Kanal and J.F. Lemmer (eds.), *Uncertainty in Artificial Intelligence* (North-Holland, Amsterdam, 1986) 298–312.

[10] D. Heckerman. Causal independence for knowledge acquisition and inference. In: *Proc of the 9th Conference on Uncertainty in Artificial Intelligence*, 1993, 122–127.

[11] D. Heckerman and J.S. Breese. A new look at causal independence. In: *Proc of the 10th Conference on Uncertainty in Artificial Intelligence*, 1994, 286–292.

[12] D. Heckerman and J.S. Breese. Causal independence for probabilistic assessment and inference using Bayesian networks, *IEEE Transactions on Systems, Man and Cybernetics*, 1996.

[13] M. Henrion. Some practical issues in constructing belief networks. In: J.F. Lemmer and L.N. Kanal (eds.), *Uncertainty in Artificial Intelligence 3* (Elsevier, Amsterdam, 1989) 161–173.

[14] F.V. Jensen. An Introduction to Bayesian Networks (UCL Press, London, 1996).

[15] P.J.F. Lucas and L.C. van der Gaag. *Principles of Expert Systems* (Addison-Wesley, Wokingham, 1991).

[16] B. Middleton, M.A. Shwe, 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, II – Evaluation of diagnostic performance, *Methods of Information in Medicine* 30, 1991, 256–267.

[17] R.E. Neapolitan. *Probabilistic Reasoning in Expert Systems — Theory and Algorithms* (John Wiley & Sons, New York, 1990).

[18] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference* (Morgan Kaufmann Publishers, San Mateo, California, 1988).

[19] M. Pradhan, M. Henrion, G. Provan, B. Del Favero, K. Huang. The sensitivity of belief networks to imprecise probabilities: an experimental investigation, *Artificial Intelligence* 84 (1-2), 1996, 363–397.

[20] E.H. Shortliffe. *Computer-Based Medical Consultations: MYCIN* (Elsevier, New York, 1976).

[21] S. Srinivas. A generalization of the noisy-OR model. In: *Proceedings of the 9th Conference on Uncertainty in Artificial Intelligence*, 1993, 208–215.

[22] L.C. van der Gaag. *Probability-based Models for Plausible Reasoning*, PhD Thesis, University of Amsterdam, 1990.

[23] N.L. Zhang. Inference with causal independence in the CPSC network. In: *Proc of the 11th Conference on Uncertainty in Artificial Intelligence*, 1995, 582–589.

[24] N.L. Zhang and D. Poole. Exploiting causal independence in Bayesian network inference, *Journal of Artificial Intelligence Research* 5, 1996, 301–328.