# The Probabilistic Interpretation of Model-based Diagnosis

Ildikó Flesch
MICC, Maastricht University,
Maastricht, the Netherlands
Email: ildiko@cs.ru.nl

Peter J.F. Lucas
Institute for Computing and Information Sciences,
Radboud University Nijmegen,
Nijmegen, the Netherlands
Email: peterl@cs.ru.nl

**Abstract**

Model-based diagnosis is the field of research concerned with the problem of finding faults in systems by reasoning with abstract models of the systems. Typically, such models offer a description of the structure of the system in terms of a collection of interacting components. For each of these components it is described how they are expected to behave when functioning normally or abnormally. The model can then be used to determine which combination of components is possibly faulty in the face of observations derived from the actual system. There have been various proposals in literature to incorporate uncertainty into the diagnostic reasoning process about the structure and behaviour of systems, since much of what goes on in a system cannot be observed. This paper proposes a method to decompose the probability distribution in probabilistic model-based diagnosis, partly in terms of the Poisson-binomial probability distribution.

## 1 Introduction

Almost from the inception of the field of probabilistic graphical models, Bayesian networks have been popular as formalisms to built *model-based*, diagnostic systems (Pearl, 1988). An alternative theory of model-based diagnosis was developed at approximately the same time, founded on techniques from logical reasoning (Reiter, 1987; de Kleer et al., 1992). The General Diagnostic Engine, GDE for short, is a well-known implementation of the logical theory; however, it also includes a restricted form of uncertainty reasoning to focus the diagnostic reasoning process (de Kleer and Williams, 1987). Previous research by Geffner and Pearl proved that the GDE approach to model-based diagnosis can be equally well dealt with by Bayesian networks (Geffner and Pearl, 1987; Pearl, 1988). Geffner and Pearl's results is basically a mapping from the logical representation as traditionally used within the logical diagnosis community to a specific Bayesian-network representation. The theory of model-based diagnosis supports multiple-fault diagnoses, which are similar to maximum a posteriori hypotheses, MAP hypotheses for short, in Bayesian networks (Gámez, 2004). Thus, although the logical and the probabilistic theory of model-based diagnosis have different origins, they are closely related.

Any theory of model-based diagnosis should consist of two parts: (*i*) a theory of how to *construct* models that can be used for diagnosing problems; (*ii*) a theory of how to *use* models to compute diagnoses. Whereas in both logic-based and probabilistic diagnosis issues concerning representation are clear—basically, logical expressions versus Bayesian networks—there is less clarity with regard to computing diagnoses. In logical model-based diagnosis there is a huge amount of literature investigating logical properties of diagnoses. In contrast, in literature on probabilistic diagnosis the emphasis is mostly on algorithmic properties of computing MAP diagnoses, rather then on probabilistic properties.

This paper proposes a new way to look at model-based diagnosis, taking the Bayesian-network representation by Geffner and Pearl as

the point of departure (Geffner and Pearl, 1987; Pearl, 1988). It is shown that by adding probabilistic information to a model of a system, the predictions that can be made by the model can be decomposed into a logical and a probabilistic part. The logical specifications are determined by the system components that are assumed to behave normally, constituting part of the system behaviour. This is complemented by uncertainty about behaviour for components that are assumed to behave abnormally. In addition, it is shown that the Poisson-binomial distribution plays a central role in determining model-based diagnoses. The results of this paper establish new links between traditional logic-based diagnosis, Bayesian networks and probability theory.

## 2 Poisson-binomial Distribution

First, we begin by summarising some of the relevant theory of discrete probability distributions (cf. (Cam, 1960; Darroch, 1964)).

Let $s = (s_1, \ldots, s_n)$ be a Boolean vector with elements $s_k \in \{0, 1\}$, $k = 1, \ldots, n$, where $s_k$ is a Bernoulli discrete random variable and is equal to the outcome of trial $k$ being either success (1) or failure (0). Let the probability of success of trial $k$ be indicated by $p_k \in [0, 1]$ and the probability of failure be set to $1 - p_k$. Then, the probability of obtaining vector $s$ as outcome is equal to

$$P(s) = \prod_{k=1}^{n} p_k^{s_k}(1 - p_k)^{1-s_k}. \quad (1)$$

This probability distribution acts as the basis for the *Poisson-binomial distribution*. The Poisson binomial distribution is employed to describe the outcomes of $n$ independent Bernoulli distributed random variables, where only the number of success and failure are counted. The probability that there are $m$, $m \leq n$, successful outcomes amongst the $n$ trials performed is then defined as:

$$f(m; n) = \sum_{s_1+\cdots+s_n=m} \prod_{k=1}^{n} p_k^{s_k}(1 - p_k)^{1-s_k}, \quad (2)$$

where $f$ is a probability function. Here, the summation means that we sum over all the possible values of elements of the vector $s$, where the sum of the values of the elements must be equal to $m$.

It is easy to check that when all probabilities $p_k$ are equal, i.e. $p_1 = \cdots = p_n = p$, where $p$ denotes this identical probability, then the probability function $f(m; n)$ becomes that of the well-known *binomial distribution*:

$$g(m; n) = \binom{n}{m} p^m (1 - p)^{n-m}. \quad (3)$$

Finally, suppose that we model interactions between the outcomes of the trials by means of a Boolean function $b$. This means that we have an oracle that is able to observe the outcomes, and then gives a verdict whether the overall outcome is successful. The *expectation* or *mean* of this Boolean function is then equal to:

$$\mathcal{E}_P(b(S)) = \sum_{s} b(s)P(s). \quad (4)$$

with $P$ defined according to Equation (1). This expectation also acts as the basis for the theory of causal independence, where a causal process is modelled in terms of interacting independent processes (cf. (Lucas, 2005)). Note that for $b(s) = b_m(s) \equiv s_1 + \cdots + s_n = m$ (i.e., the Boolean function that checks whether the number of successful trials is equal to $m$), we have that $\mathcal{E}_P(b_m(S)) = f(m; n)$. Thus, Equation (4) can be looked on as a generic way to combine the outcome of independent trials.

In the theory of model-based diagnosis, it is common to represent models of systems by means of logical specifications, which are equivalent to Boolean functions. Below, it will become clear that if we interpret the success probabilities $p_k$ as the probability of observing the expected output of a system's component under the assumption that the component is faulty, then the theory of Poisson-binomial distributions can be used to describe part of probabilistic model-based diagnosis. However, first the necessary background to model-based diagnosis research is reviewed.

## 3 Uncertainty in Model-based Diagnosis

### 3.1 Model-based Diagnosis

In the theory of model-based diagnosis (Reiter, 1987), the structure and behaviour of a system is represented by a *logical diagnostic system* $\mathcal{S}_L = (\text{SD}, \text{COMPS})$, where

- SD denotes the *system description*, which is a finite set of logical formulae, specifying structure and behaviour;
- COMPS is a finite set of constants, corresponding to the *components* of the system; these components can be faulty.

The system description consists of *behaviour descriptions* and *connections*. A behavioural description is a formula specifying *normal* and *abnormal* (faulty) functionalities of the components. A connection is a formula of the form $i_c = o_{c'}$, where $i_c$ and $o_{c'}$ denote the input and output of components $c$ and $c'$, respectively. This way an equivalence relation on the inputs and outputs is defined, denoted by $\text{IO}_{\setminus\equiv}$. The class representatives from this set are denoted by $[r]$.

A *logical diagnostic problem* is defined as a pair $\mathcal{P}_L = (\mathcal{S}_L, \text{OBS})$, where $\mathcal{S}_L$ is a logical diagnostic system and OBS is a finite set of logical formulae, representing *observations*.

Adopting the definition from (de Kleer et al., 1992), a diagnosis in the theory of consistency-based diagnosis is defined as follows. Let $\Delta$ be the assignment of either a normal or an abnormal behavioural assumption to *each* component. Then, $\Delta$ is a *consistency-based diagnosis* of the logical diagnostic problem $\mathcal{P}_L$ iff the observations are consistent with both the system description and the diagnosis:

$$\text{SD} \cup \Delta \cup \text{OBS} \nvDash \bot. \tag{5}$$

Here, $\nvDash$ stands for the negation of the logical entailment relation, and $\bot$ represents a contradiction.

**EXAMPLE 1** Consider the logical circuit depicted in Figure 1, which represents a full adder,

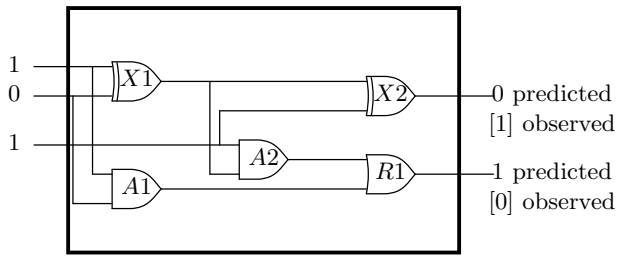


Figure 1: Full adder with inputs $\{i_1, \bar{i}_2, i_3\}$ and observed and predicted outputs.

i.e. a circuit that can be used for the addition of two bits with carry-in and carry-out bits. It is an example frequently used to illustrate concepts from model-based diagnosis. This circuit consists of two AND gates ($A1$ and $A2$), one OR gate ($R1$) and two exclusive-OR (XOR) gates ($X1$ and $X2$). These are the components that can be either faulty (abnormal) or normal. □

### 3.2 Probabilistic Model-based Diagnosis

In this section, we will map logical diagnostic problems onto probabilistic representations, called *Bayesian diagnostic problems*, using the Bayesian-network representation proposed by Geffner and Pearl (Geffner and Pearl, 1987; Pearl, 1988). As will become clear, a Bayesian diagnostic problem is defined as $(i)$ a Bayesian diagnostic system representing the components, including their behaviour and interaction, based on information from the logical diagnostic system of concern, and $(ii)$ a set of observations.

#### 3.2.1 Graphical Representation

First the graphical structure used to represent the structural information from a logical diagnostic system is defined. It has the form of an acyclic directed graph $G = (V, E)$, where $V$ is the set of *vertices* and $E \subseteq (V \times V)$ the set of directed edges, or *arcs* as we shall say.

**Definition 1** ***(diagnostic mapping)*** *Let $\mathcal{S}_L = (\text{SD}, \text{COMPS})$ be a logical diagnostic system. The* diagnostic mapping *$m_d$ maps $\mathcal{S}_L$ onto an acyclic directed graph $G = m_d(\mathcal{S}_L)$, as follows (see Figure 2):*

- *The vertices $V$ of graph $G$ are created according to the following rules:*

– *Each component* $c \in \text{COMPS}$ *yields a vertex* $A_c$ *used to represent its normal and abnormal behaviour;*
– *Each class representative of an input or output* $[r] \in \text{IO}_{\setminus\equiv}$ *yields an associated vertex* $[r]$.

*The set of all abnormality vertices* $A_c$ *is denoted by* $\Delta$*, i.e.* $\Delta = \{A_c \mid c \in \text{COMPS}\}$. *The vertices of graph* $G$ *are, thus, obtained as follows:*

$$V = \Delta \cup \text{IO}_{\setminus\equiv},$$

*where* $\text{IO}_{\setminus\equiv} = I \cup O$*, with disjoint sets* $I$ *of* input vertices *and* $O$ *of* output vertices.

- *The arcs* $E$ *of* $G$ *are constructed as follows:*
  – *There is an arc from each each input of a component* $c$ *to each output of the component;*
  – *There is an arc for each component* $c$ *from* $A_c \in V$ *to the corresponding output of the component.*

An example of this diagnostic mapping is shown in the following example.

**EXAMPLE 2** Figure 3 shows the graphical representation of the full-adder circuit from Figure 1. The set $V$ of vertices is:

$$\begin{aligned} V &= \Delta \cup O \cup I \\ &= \{A_{X1}, A_{X2}, A_{A1}, A_{A2}, A_{R1}\} \\ &\quad \cup \{O_{X1}, O_{X2}, O_{A1}, O_{A2}, O_{R1}\} \\ &\quad \cup \{I_1, I_2, I_3\}. \end{aligned}$$

The arcs from $E$ connect $(i)$ outputs of two components such as $O_{X1} \rightarrow O_{X2}$, $(ii)$ an abnormality vertex with an output vertex such as $A_{A2} \rightarrow O_{A2}$ and $(iii)$ an input vertex with an output vertex such as $I_3 \rightarrow O_{X2}$. □

### 3.2.2 Bayesian Diagnostic Problems

Recall that Bayesian networks that act as the basis for diagnostic Bayesian networks consist of two parts: a joint probability distribution and a graphical representation of the relations among the random variables defined by the joint probability distribution. Based on the definition of

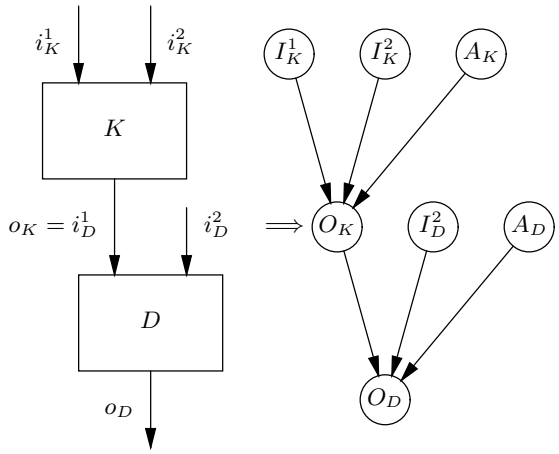


Figure 2: The diagnostic mapping.

Bayesian networks, particular parts of a logical diagnostic system will be related to the graphical structure of a diagnostic Bayesian network, whereas other parts will have a bearing on the content of the probability table of the Bayesian network.

Having introduced the mapping of a logical diagnostic system to its associated graph structure, we next introduce the full concept of a Bayesian diagnostic system.

**Definition 2** ***(Bayesian diagnostic system)*** *Let* $\mathcal{S}_L = (\text{SD}, \text{COMPS})$ *be a logical diagnostic system, and* $G = m_d(\mathcal{S}_L)$ *be obtained by applying the diagnostic mapping. Let* $P$ *be a joint probability distribution of the vertices of* $G$*, interpreted as random variables. Then,* $\mathcal{S}_B = (G, P)$ *is the associated* Bayesian diagnostic system.

Recall that by the definition of a Bayesian network, the joint probability distribution $P$ of a Bayesian diagnostic system can be factorised as follows:

$$P(I, O, \Delta) = \prod_c P(O_c \mid \pi(O_c)) P(I) P(\Delta), \quad (6)$$

where $O_c$ is an output variable associated to component $c \in \text{COMPS}$, and $\pi(O_c)$ are the random variables corresponding to the *parents* of $O_c$. The parents will normally consist of inputs $I_c$ and an abnormality variable $A_c$.

To simplify notation, in the following, (sets of) random variables of a Bayesian diagnos-

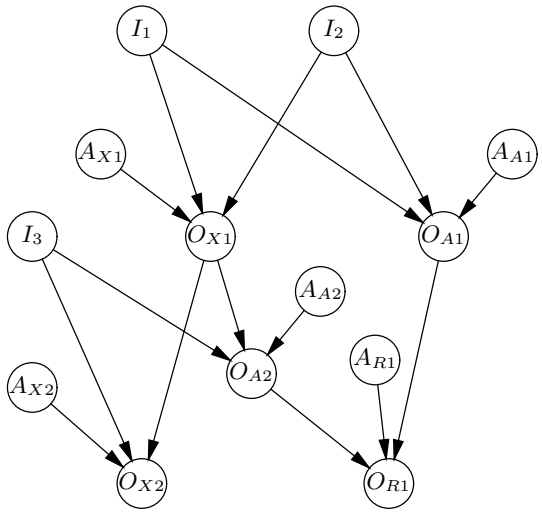


Figure 3: A Bayesian diagnostic system corresponding to the circuit in Figure 1.

tic problem have the same names as the corresponding vertices. By $a_c$ is indicated that abnormality variable $A_c$ takes the value 'true', whereas by $\bar{a}_c$ it is indicated that $A_c$ takes the value 'false'. A similar notation will be used for the other random variables. Finally, a specific abnormality assumption concerning all abnormality variables is denoted by $\delta_C$, which is defined as follows:

$$\delta_C = \{a_c \mid c \in C\} \cup \{\bar{a}_c \mid c \in \text{COMPS} - C\},$$

with $C \subseteq$ COMPS. There are some sensible constraints on the joint probability distribution $P$ of a Bayesian diagnosis system that can be derived from the specification of the logical diagnostic system. These will be discussed later.

As with logical diagnostic problems, we need to add observations to Bayesian diagnostic systems in order to be able to solve diagnostic problems. In logical diagnostic systems, observations are the inputs and outputs of a system. It is generally not the case that the entire set of inputs and outputs of a system is observed. The set of input and output variables that have been observed, are referred to by $I_\omega$ and $O_\omega$, respectively. The unobserved input and output variables will be referred to as $I_u$ and $O_u$, respectively. We will use the notation $i_\omega$ to denote the values of the observed inputs, and $o_\omega$ for the observed output values. The set of *observations* is then denoted as $\omega = i_\omega \cup o_\omega$.

Now, we are ready to define the notion of Bayesian diagnostic problem, which is a Bayesian diagnostic system augmented by a set of observations.

**Definition 3** ***(Bayesian diagnostic problem)*** *A* Bayesian diagnostic problem*, denoted by $\mathcal{P}_B$, is defined as the pair $\mathcal{P}_B = (\mathcal{S}_B, \omega)$, where $\mathcal{S}_B$ is a Bayesian diagnostic system and $\omega$ the set of observations of this system.*

Determining the diagnoses of a Bayesian diagnostic problem amounts to computing $P(\delta_C \mid \omega)$, and then finding the $\delta_C$ which maximises $P(\delta_C \mid \omega)$, i.e.

$$\delta_C^* = \arg\max_{\delta_C} P(\delta_C \mid \omega).$$

This problem is NP-hard (Gámez, 2004). The probability $P(\delta_C \mid \omega)$ can be computed by Bayes' rule, using the probabilities from the specification of a Bayesian diagnostic system:

$$P(\delta_C \mid \omega) = \frac{P(\omega \mid \delta_C)P(\delta_C)}{P(\omega)}. \tag{7}$$

As a consequence of the independences that hold for a Bayesian diagnostic system, it is possible to simplify the computation of the conditional probability distribution $P(\omega \mid \delta_C)$. According to the definition of a Bayesian diagnostic system it holds that

$$P(i \mid \delta_C) = P(i) \ ,$$

for each $i \subseteq (i_\omega \cup i_u)$, as the input variables and abnormality variables are independent. In addition, it is assumed that the input variables are independent.

Using these results, basic probability theory and the definition of a Bayesian diagnostic problem yields the following derivation:

$$\begin{aligned} P(\omega \mid \delta_C) &= P(i_\omega, o_\omega \mid \delta_C) \\ &= \sum_{i_u} P(i_u) P(i_\omega, o_\omega \mid i_u, \delta_C) \\ &= P(i_\omega) \sum_{i_u} P(i_u) \\ &\quad \times \sum_{o_u} \prod_{c} P(O_c \mid \pi(O_c)), \end{aligned} \tag{8}$$

since it holds by the axioms of probability theory that

$$P(i_\omega, o_\omega \mid i_u, \delta_C) = \sum_{o_u} P(i_\omega) \prod_c P(O_c \mid \pi(O_c)) \,.$$

To emphasise that the set of parents $\pi(O_c)$ includes an abnormality variable that is assumed to be true, i.e. the component is assumed to behave abnormally, the following notation is used $P(O_c \mid \pi(O_c) : a_c)$; similar, for the situation where the component $c$ is assumed to behave normally the notation $P(O_c \mid \pi(O_c) : \bar{a}_c)$ is employed. Finally, the following assumptions, which will be used in the remainder of this paper, are made:

- $P(O_c \mid \pi(O_c) : a_c) = P(O_c \mid a_c)$, i.e. the probabilistic behaviour of a component that is faulty is independent of its inputs;
- $P(O_c \mid \pi(O_c) : \bar{a}_c) \in \{0, 1\}$, i.e. normal components behave deterministically.

The probability $P(o_c \mid a_c)$ will be abbreviated in the following section as $p_c$; thus $P(\bar{o}_c \mid a_c) = 1 - p_c$ These are realistic assumptions, as it is unlikely that detailed functional behaviour will be known for a component that is faulty, whereas when the component is not faulty, it is certain it will behave as intended.

## 4 Decomposition of Probability Distribution

To establish that probabilistic model-based diagnosis can be partly interpreted in terms of a Poisson-binomial distribution, it is necessary to decompose Equation (8) into various parts. The first part will represent the probabilities that components $c$ produce the right, $o_c$, or wrong, $\bar{o}_c$, output, which correspond to the success and failure probabilities, respectively, of a Poisson-binomial distribution. The second part represents a normally functioning system fragment, which will be represented by a Boolean function. There is also a third part, which concerns the observed and unobserved inputs. We start by distinguishing between various types of components, inputs and outputs, in order to make the necessary distinction:

- The sets of components assumed to function *normally* and *abnormally* will be denoted by $C^{\bar{a}}$ and $C^a$, respectively, with $C^{\bar{a}}, C^a \subseteq \text{COMPS}$;
- The sets $C^{\bar{a}}$ and $C^a$ are partitioned into sets of components, for *observed* and *unobserved* outputs, indicated by the sets $C^{\bar{a}}_\omega$, $C^{\bar{a}}_u$, $C^a_\omega$ and $C^a_u$, respectively.

Thus, $C^{\bar{a}} = C^{\bar{a}}_\omega \cup C^{\bar{a}}_u$ and $C^a = C^a_\omega \cup C^a_u$. In addition, we will sometimes make a distinction between components $c$ for which $o_c$ has been observed, and components $c$ for which $\bar{o}_c$ has been observed. These sets will be denoted by $C^o_\omega$ and $C^{\bar{o}}_\omega$, respectively. It holds that $C^o_\omega$ and $C^{\bar{o}}_\omega$ constitute a partition of $C_\omega$. The notations can also be combined, e.g., as $C^{a,o}_\omega$ and $C^{a,\bar{o}}_\omega$. Furthermore, we will sometimes use a similar notation for sets of output variables, e.g., $O^{\bar{a}}_u = \{O_c \mid c \in C^{\bar{a}}_u\}$ and $O^{\bar{a}}_\omega = \{O_c \mid c \in C^{\bar{a}}_\omega\}$, and input variables, e.g., $I^{\bar{a}}_u = \bigcup_{c \in C^{\bar{a}}_u} I_c$ indicates unobserved inputs of components that are assumed to behave normally and $I^{\bar{a}}_\omega = \bigcup_{c \in C^{\bar{a}}_\omega} I_c$ are observed inputs of components that are assumed to behave normally, with $I_c$ the set of input variables of component $c \in \text{COMPS}$ and $I^{\bar{a}} = I^{\bar{a}}_\omega \cup I^{\bar{a}}_u$.

The following lemma shows that it is possible to decompose part of the joint probability distribution of Equation (6) using the component sets defined above.

**Lemma 1** *The following statements hold:*

- *The joint probability distribution of the outputs of the set of assumed* normally *functioning components $C^{\bar{a}}$, can be decomposed into two products as follows:*

$$\prod_{c \in C^{\bar{a}}} P(O_c \mid \pi(O_c) : \bar{a}_c) = \prod_{c \in C^{\bar{a}}_u} P(O_c \mid \pi(O_c) : \bar{a}_c) \times \prod_{c \in C^{\bar{a}}_\omega} P(O_c \mid \pi(O_c) : \bar{a}_c).$$

- *Similarly, the joint probability distribution of the outputs of the set of assumed* abnormally *functioning components $C^a$, can be*

*decomposed into two products as follows:*

$$\prod_{c\in C^a} P(O_c \mid \pi(O_c) : a_c) = \prod_{c\in C^a_u} P(O_c \mid a_c) \prod_{c\in C^a_\omega} P(O_c \mid a_c).$$

**Proof**: The decompositions follows from the definitions of the sets $C^a$, $C^a_\omega$, $C^a_u$, $C^{\bar{a}}_u$ and $C^{\bar{a}}_\omega$, and the independence assumptions underlying the distribution $P$. □

Now, based on Lemma 1, we can also decompose the product of the *entire* set of components, as follows:

$$\begin{aligned}&\prod_c P(O_c \mid \pi(O_c))\\ &= \prod_{c\in C^{\bar{a}}_u} P(O_c \mid \pi(O_c) : \bar{a}_c) \prod_{c\in C^{\bar{a}}_\omega} P(O_c \mid \pi(O_c) : \bar{a}_c)\\ &\quad \times \prod_{c\in C^a_u} P(O_c \mid a_c) \prod_{c\in C^a_\omega} P(O_c \mid a_c).\end{aligned}$$

Next, we show that the outputs of the set of observed abnormal components $C^a_\omega$ only depend on probabilities $p_c = P(o_c \mid a_c)$, $c \in C^a_\omega$.

**Lemma 2** *The joint probability of observed outputs of the abnormally assumed components can be written as:*

$$\prod_{c\in C^a_\omega} P(O_c \mid \pi(O_c) : a_c) = \prod_{c\in C^{a,o}_\omega} p_c \prod_{c\in C^{a,\bar{o}}_\omega} (1 - p_c).$$

**Proof**: This follows straight from the definitions of $C^a_\omega$, $C^{a,o}_\omega$ and $C^{a,\bar{o}}_\omega$. □

Recall that the probability of an output of a normally functioning component was assumed to be either 0 or 1, i.e. $P(O_c \mid \pi(O_c) : \bar{a}_c) \in \{0, 1\}$. Clearly, these probabilities yield, when multiplied, Boolean functions. One of these Boolean functions, denoted by $\varphi$, is defined as follows: $\varphi(o^{\bar{a}}_u, o^a_u, i^{\bar{a}}) = \prod_{c\in C^{\bar{a}}_u} P(O_c \mid \pi(O_c) : \bar{a}_c)$, where the set of parents $\pi(O_c)$ may, but need not, contain variables from the sets of variables $O^a_u$ and $I^{\bar{a}}$. However, $\pi(O_c)$ does not contain variables from the set $I^a$, as these only condition variables that are assumed to behave abnormally and are then ignored, as mentioned at the end of the previous section. Similarly, we define Boolean functions $\psi(o_u, o^{\bar{a}}_\omega, i^{\bar{a}}) = \prod_{c\in C^{\bar{a}}_\omega} P(O_c \mid \pi(O_c) : \bar{a}_c)$.

**Lemma 3** *For each value $o^a_u$ and $i^{\bar{a}}$, there exists exactly* one *value $o^{\bar{a}}_u$ of the set of variables $O^{\bar{a}}_u = \{O_c \mid c \in C^{\bar{a}}_u\}$ for which it holds that $\varphi(o^a_u, o^{\bar{a}}_u, i^{\bar{a}}) = 1$; similarly, for each value $o_u$ and $i^{\bar{a}}$ there exists one value $o^{\bar{a}}_\omega$ of the set of variables $O^{\bar{a}}_\omega = \{O_c \mid c \in C^{\bar{a}}_\omega\}$ for which it holds that $\psi(o_u, o^{\bar{a}}_\omega, i^{\bar{a}}) = 1$.*

**Proof**: As both the functions $\varphi$ and $\psi$ are defined as products of conditional probability distributions $P(O_c \mid \pi(O_c) : \bar{a}_c)$, for which we have that $P(o_c \mid \pi(O_c) : \bar{a}_c) \in \{0, 1\}$, there is, due to the axioms of probability theory, for any value of the variables corresponding to the parents of the variables $O_c$ at most one value for each $O_c$ for which the joint probability $\prod_c P(O_c \mid \pi(O_c) : \bar{a}_c) = 1$. □

The following lemma, which is used later, is a consequence of the definition of these Boolean functions.

**Lemma 4** *Let the Boolean functions $\varphi$ and $\psi$ be as defined above, then:*

$$\sum_{o_u} \varphi(o^a_u, o^{\bar{a}}_u, i^{\bar{a}})\psi(o_u, o^{\bar{a}}_\omega, i^{\bar{a}}) \prod_{c\in C^a} P(O_c \mid a_c) = \sum_{o^a_u} b(o^a_u, i^{\bar{a}}) \prod_{c\in C^{a,o}} p_c \prod_{c\in C^{a,\bar{o}}} (1 - p_c),$$

*with Boolean function $b$ and $p_c = P(o_c \mid a_c)$.*

**Proof**: For a fixed set of observed outputs $o_\omega$ let $b(o_u, i^{\bar{a}}) = \varphi(o^a_u, o^{\bar{a}}_u, i^{\bar{a}})\psi(o_u, o^{\bar{a}}_\omega, i^{\bar{a}})$, then,

$$\sum_{o_u} \varphi(o^a_u, o^{\bar{a}}_u, i^{\bar{a}})\psi(o_u, o^{\bar{a}}_\omega, i^{\bar{a}}) \prod_{c\in C^a} P(O_c \mid a_c) = \sum_{o_u} b(o_u, i^{\bar{a}}) \prod_{c\in C^a} P(O_c \mid a_c).$$

Furthermore, due to Lemma 3, it suffices to only consider the restriction of the function $b$ to the variables $O^a_u$ and $I^{\bar{a}}$, as for given values $o^a_u$ and $i^{\bar{a}}$, $b(o^a_u, o^{\bar{a}}_u, i^{\bar{a}}) = 0$ for all but one value of $O^{\bar{a}}_u$. This function is denoted by $b(o^a_u, i^{\bar{a}})$. The product term results from application of a slight generalisation of Lemma 2. □

We are now ready to establish that $P(\omega \mid \delta_C)$ can be written as the sum of weighted products of the form $\prod_c p_c \prod_{c'}(1 - p_{c'})$.

**Theorem 1** *Let $\mathcal{P}_B = (\mathcal{S}_B, \omega)$ be a Bayesian diagnostic problem. Then, $P(\omega \mid \delta_C)$ can be expressed as follows:*

$$P(\omega \mid \delta_C) = P(i_\omega) \sum_{i_u^{\bar{a}}} P(i_u^{\bar{a}}) \sum_{o_u^a} b(o_u^a, i^{\bar{a}}) \times \prod_{c \in C^{a,o}} p_c \prod_{c \in C^{a,\bar{o}}} (1 - p_c)$$

*where $b(o_u^a, i^{\bar{a}}) \in \{0, 1\}$ and $p_c = P(o_c \mid a_c)$.*

**Proof**: The result follows from the above lemmas and the fact that we sum over (part of) the input variables $I$. Note that only the variables $I^{\bar{a}}$ are used as conditioning variables, which follows from the assumption that $P(O_c \mid \pi(O_c) : a_c) = P(O_c \mid a_c)$. As only the input variables $i_u^{\bar{a}}$ are assumed to be dependent of output variables, we obtain: $\sum_{i_u, o_u^a} P(i_u) \cdots = \sum_{i_u^{\bar{a}}, o_u^a} P(i_u^{\bar{a}}) \cdots$. The Boolean function $b(o_u^a, i^{\bar{a}})$ is as above. □

An alternative version of the theorem can be obtained in terms of expectations using Equation (4) for the Poisson-binomial distribution:

$$P(i_\omega) \sum_{i_u^{\bar{a}}} P(i_u^{\bar{a}}) \sum_{o_u^a} b(o_u^a, i^{\bar{a}}) \prod_{c \in C^{a,o}} p_c \prod_{c \in C^{a,\bar{o}}} (1 - p_c) = P(i_\omega) \prod_{c \in C_\omega^a} P(O_c \mid a_c) \sum_{i_u^{\bar{a}}} P(i_u^{\bar{a}}) \mathcal{E}_P(b_{i^{\bar{a}}}(O_u^a)),$$

i.e. the sum of the mean of the Boolean functions $b_{i^{\bar{a}}}$, which are functions of the unobserved inputs $i_u^{\bar{a}}$, in terms of the probability function $P$ (Equation (4)), weighed by the prior probability of unobserved inputs $i_u^{\bar{a}}$. Combining this with Equation (7) yields $P(\delta_C \mid \omega)$. Thus, to probabilistically rank diagnoses $\delta_C$ it is necessary to compute: $(i)$ $\mathcal{E}_P(b_{i^{\bar{a}}}(O_u^a))$, the Poisson-binomial distribution mean of the behaviour of the normally assumed components, $(ii)$ $P(i_u^{\bar{a}})$, $(iii)$ $\prod_{c \in C_\omega^a} P(O_c \mid a_c)$, the observed abnormal components, and $(iv)$ the prior $P(\delta_c)$. Note that both $P(i_\omega)$ and $P(\omega)$ can be ignored.

## 5 Conclusions

We have shown that probabilistic model-based diagnosis, which is an extension of traditional GDE-like model-based diagnosis, can be decomposed into computation of various probabilities, in which a central role is played by the Poisson-binomial distribution. When all probabilities $p_c = P(o_c \mid a_c)$ are assumed to be equal, a common simplifying assumption in model-based diagnosis, the analysis reduces to the use of the standard binomial distribution.

So far, most other research on integrating probabilistic reasoning with logic-based model-based diagnosis took probabilistic reasoning as adding some sort of uncertain, abductive reasoning to logical reasoning. No attempts were made in related research to look inside what happens in the diagnostic process, as was done in this paper. We expect that it becomes thus possible to investigate further variations in probabilistic model-based diagnosis, for example, by adopting assumptions different from those in this paper with regard to fault behaviour in systems.

## References


L. Le Cam. 1960. An approximation theorem for the poisson binomial distribution. *Pacific Journal of Mathematics*, 10:1181–1197.

J. Darroch. 1964. On the distribution of the number of successes in independent trials. *The Annals of Mathematical Statistics*, 35:1317–1321.

J. de Kleer and B. C. Williams. 1987. Diagnosing multiple faults. *Artificial Intelligence*, 32:97–130.

J. de Kleer, A. K. Mackworth, and R. Reiter. 1992. Characterizing diagnoses and systems. *Artificial Intelligence*, 52:197–222.

J.A. Gámez, 2004. *Abductive inference in Bayesian Networks: a review*, pages 101–120. Springer, Berlin.

H. Geffner and J. Pearl. 1987. Distributed diagnosis of systems with multiple faults. In *Proc. of the 3rd IEEE Conference on AI Applications*, pages 156–162. IEEE.

P.J.F. Lucas. 2005. Bayesian network modelling through qualitative patterns. *Artificial Intelligence*, 163:233–263.

J. Pearl. 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kauffman, San Francisco, CA.

R. Reiter. 1987. A theory of diagnosis from first principles. *Artificial Intelligence*, 32:57–95.