Exploring the Noisy Threshold Function in Designing Bayesian Networks*

Rasa Jurgelenaite, Peter Lucas and Tom Heskes
Radboud University Nijmegen, Nijmegen, The Netherlands
E-mail: {rasa, peterl, tomh}@cs.ru.nl

Abstract

Causal independence modelling is a well-known method both for reducing the size of probability tables and for explaining the underlying mechanisms in Bayesian networks. Many Bayesian network models incorporate causal independence assumptions; however, only the noisy OR and noisy AND, two examples of causal independence models, are used in practice. Their underlying assumption that either at least one cause, or all causes together, give rise to an effect, however, seems unnecessarily restrictive. In the present paper a new, more flexible, causal independence model is proposed, based on the Boolean threshold function. A connection is established between conditional probability distributions based on the noisy threshold model and Poisson binomial distributions, and the basic properties of this probability distribution are studied in some depth. The successful application of the noisy threshold model in the refinement of a Bayesian network for the diagnosis and treatment of ventilator-associated pneumonia demonstrates the practical value of the presented theory.

1 Introduction

Bayesian networks offer an appealing language for building models of domains with inherent uncertainty. However, the assessment of a probability distribution in Bayesian networks is a challenging task, even if its topology is sparse. This task becomes even more complex if the model has to integrate expert knowledge. While learning algorithms can be forced to take into account an expert's view, for the best possible results the experts must be willing to reconsider their ideas in light of the model's 'discovered' structure. This requires a clear understanding of the model by the domain expert. Causal independence models can both limit the number of conditional probabilities to be assessed and provide the ability for models to be understood by domain experts in the field. The concept of *causal independence* refers to a situation where multiple causes independently influence a common effect.

Many actual Bayesian network models use causal independence assumptions. However, only the logical OR and AND operators are used in practice in defining the interaction among causes; their underlying assumption is that the presence of either at least one cause or all causes at the same time give

*This research was supported by the Netherlands Organization for Scientific Research (NWO).

rise to the effect. The resulting probabilistic submodels are called *noisy OR* and *noisy AND*, respectively. Our feeling is that in building Bayesian-network models, the expressiveness of the noisy OR and noisy AND is too restrictive. In this paper, we discuss a way to expand the space of causal independence models using symmetric Boolean functions. It is known that any symmetric Boolean function can be decomposed into threshold functions [15]. Thus, threshold functions offer a natural basis for the analysis of causal independence models. Causal independence models with the threshold interaction function are the main topic of this paper. They will be referred to as the *noisy threshold models*.

The remainder of this paper is organised as follows. In the following section, the basic properties of Bayesian networks are reviewed. Causal independence models and Boolean functions are introduced in Section 3 as is the noisy threshold model. In Section 4, we establish a connection between the noisy threshold model and Poisson binomial distribution, and provide an interpretation of the relevant properties of this distribution. Section 6 offers results on the application of the presented theory to the refinement of an existing medical Bayesian network model. Finally, in Section 7, we summarise what has been achieved by this research.

2 Review of Bayesian Networks

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

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

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

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

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

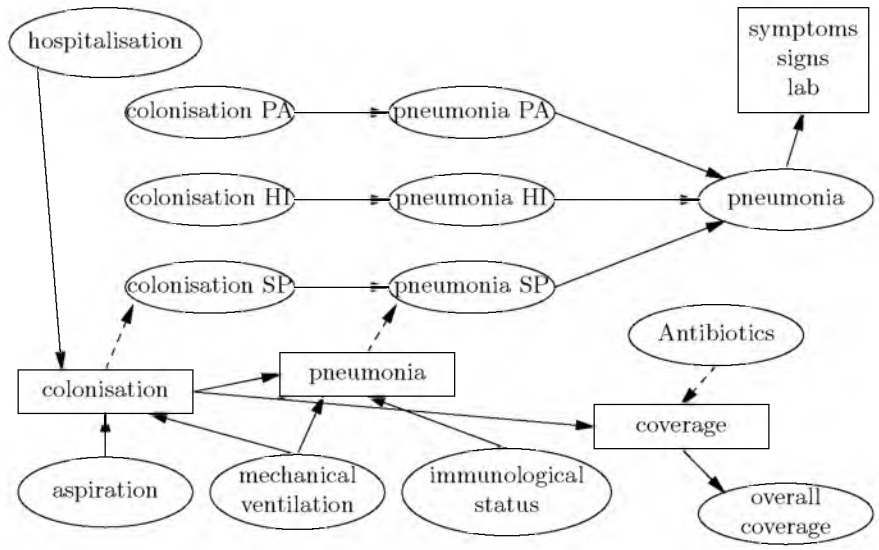

Figure 1: Detailed structure of part of the VAP model. Only three of the seven bacteria included in the model are shown. Boxes stand for collections of similar vertices. Solid arcs stand for atemporal stochastic influences, whereas dashed arcs indicate temporal influences. Abbreviations of names of bacteria: PA = Pseudomonas aeruginosa, HI = Haemophilus influenzae, SP = Streptococcus pneumoniae.

Consider the Bayesian network, shown in Figure 1, that provides motivation for the methods developed in this paper. *Ventilator-associated pneumonia*, or VAP for short, is a low-prevalence disease occurring in mechanically-ventilated patients in critical care and involves infection of the lower respiratory tract. VAP is associated with signs and symptoms such as fever, sputum production, abnormal chest X-ray and high numbers of white blood cells. As diagnosing and treating a disorder in medicine involves reasoning with uncertainty, a Bayesian network was constructed as the primary tool for building a decision-support system to support clinicians in the diagnosis and treatment of VAP [11]. The Bayesian network models the temporal process of colonisation of the mechanically ventilated patient by bacteria during stay in the critical care unit, which may, but need not, give rise to VAP with its associated signs and symptoms. This process is represented in the left part of the network. In addition, the effects of particular antimicrobial drugs, represented by the vertex 'antibiotics', is modelled in the network in terms of coverage of the bacteria by these antibiotics (each bacterium is only susceptible to some antibiotics and not too all). If a particular antibiotic covers many bacteria it is said to have a *broad spectrum*; otherwise, its spectrum is *narrow*. Prescription of broad spectrum antibiotics promotes the creation of resistance of bacteria to antibiotics, and should therefore be avoided if possible. Thus, the problem for the clinician is to ensure that the spectrum of antibiotic treatment is as narrow as possible, in

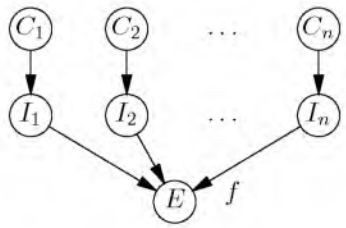

Figure 2: Causal independence model.

order to prevent the occurrence of resistance, while still covering as many of the bacteria as possible.

The Bayesian network model shown in Figure 1 includes two vertices where probabilistic information has been expressed in terms of logical operators; the conditional probability distribution defined for the variable 'pneumonia' was defined in terms of a logical OR, whereas the distribution for the variable 'coverage' was defined in terms of the logical AND. We will return to the meaning of these definitions below.

3 Causal Modelling and Boolean Functions

3.1 Causal independence

Causal independence (also known as independence of causal influence) [6, 16] is a popular way to specify interactions among cause variables. The global structure of a causal independence model is shown in Figure 2; it expresses the idea that causes $C_1, \ldots, C_n$ influence a given common effect E through intermediate variables $I_1, \ldots, I_n$ and a deterministic function f, called the *interaction function*. The impact of each cause C_l on the common effect E is independent of each other cause $C_j, j \neq l$. The intermediate variable I_l is considered to be a contribution of the cause variable C_l to the common effect E. The function f represents in which way the intermediate effects I_l, and indirectly also the causes C_l, interact to yield the final effect E. Hence, the function f is defined in such a way that when a relationship, as modelled by the function f, between $I_l, l = 1, \ldots, n$, and $E = \top$ is satisfied, then it holds that $e = f(I_1, \ldots, I_n)$. It is assumed that $\Pr(e \mid I_1, \ldots, I_n) = 1$ if $f(I_1, \ldots, I_n) = e$, and $\Pr(e \mid I_1, \ldots, I_n) = 0$ if $f(I_1, \ldots, I_n) = \bar{e}$. The smallest possible causal independence model has two cause variables.

A causal independence model is defined in terms of the causal parameters $\Pr(I_l \mid C_l)$, for $l = 1, \ldots, n$. An assumption underlying the causal independence models introduced in [12] is that absent causes do not contribute to the effect. In terms of probability theory this implies that it holds that $\Pr(i_l \mid \bar{c}_l) = 0$. As a consequence, it holds that $\Pr(\bar{i}_l \mid \bar{c}_l) = 1$. In this paper we make the same assumption.

The conditional probability of the occurrence of the effect E given the causes $C_1, \ldots, C_n$, i.e., $\Pr(e \mid C_1, \ldots, C_n)$, can be obtained from the causal parameters

$\Pr(I_l \mid C_l)$ as follows [12, 16]:

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

In this paper we assume that the function f in Equation (1) is a Boolean function. Systematic analyses of the global probabilistic patterns in causal independence models based on restricted Boolean functions was presented in [12] and [8]. However, there are 2^{2^n} different n-ary Boolean functions [4, 15]; thus, the potential number of causal interaction models is huge. However, if we assume that the order of the cause variables does not matter, the Boolean functions become *symmetric* [15] and the number reduces to 2^{n+1}.

An important symmetric Boolean function is the *exact* Boolean function ϵ_l, which has function value true, i.e. $\epsilon_l(I_1, \ldots, I_n) = \top$, if $\sum_{j=1}^{n} \nu(I_j) = l$ with $\nu(I_j)$ equal to 1, if I_j is equal to true and 0 otherwise. Symmetric Boolean function can be decomposed in terms of the exact functions ϵ_l as follows [15]:

$$f(I_1, \ldots, I_n) = \bigvee_{l=0}^{n} \epsilon_l(I_1, \ldots, I_n) \wedge \gamma_l \tag{2}$$

where γ_l are Boolean constants only dependent on the function f. For example, for the Boolean function defined in terms of the OR operator we have $\gamma_0 = \bot$ and $\gamma_1 = \ldots = \gamma_n = \top$.

Another useful symmetric Boolean function is the *threshold* function τ_k, which simply checks whether there are at least k trues among the arguments, i.e. $\tau_k(I_1, \ldots, I_n) = \top$, if $\sum_{j=1}^{n} \nu(I_j) \geq k$ with $\nu(I_j)$ equal to 1, if I_j is equal to true and 0 otherwise. To express it in the Boolean constants we have: $\gamma_0 = \cdots = \gamma_{k-1} = \bot$ and $\gamma_k = \cdots = \gamma_\rho = \top$. Obviously, any exact function can be written as the subtraction of two threshold functions and thus any symmetric Boolean function can be decomposed into threshold functions.

The interaction among variables modelled by the pneumonia and coverage variables, as shown in Figure 1, was modelled by assuming f to be an OR and an AND, respectively. This corresponds to threshold functions τ_k with $k = 1$ for the OR, and $k = n$ for the AND. Hence, these two Boolean functions can be taken as two extremes of a spectrum of Boolean functions based on the threshold function. As by definition a patient has pneumonia independent of the specific bacterium causing pneumonia, the logical OR appeared to be the right way to model the interactions between the pneumonia variables. However, as argued before, clinicians need to be careful in the prescription of antibiotics as they have a tendency to prescribe antibiotics with a spectrum that is too broad. This casts doubts about the appropriateness of the logical AND for the modelling of interactions concerning coverage of bacteria by antibiotics. Using the threshold function τ_k with $k \neq 1, n$, may result in a better model. In the following we therefore investigate properties of the threshold function, and subsequently study its use in improving the Bayesian network model shown in Figure 1.

3.2 The noisy threshold model

Using the property of Equation (2) of the symmetric Boolean functions, the conditional probability of the occurrence of the effect E given the causes $C_1, \ldots, C_n$ can be decomposed in terms of probabilities that exactly l intermediate variables $I_1, \ldots, I_n$ are true, as follows:

$$\Pr(e \mid C_1, \ldots, C_n) = \sum_{\substack{0 \leq l \leq n \\ \gamma_l}} \sum_{\epsilon_l(I_1,\ldots,I_n)} \prod_{i=1}^{n} \Pr(I_i \mid C_i). \tag{3}$$

Thus, Equation (3) yields a general formula to compute the probability of the effect in terms of exact functions in any causal independence model where an interaction function f is a symmetric Boolean function.

Let us denote a conditional probability of the effect E given causes $C_1, \ldots, C_n$ in a noisy threshold model with interaction function τ_k as $\Pr_{\tau_k}(e \mid C_1, \ldots, C_n)$. Then, from Equation (3) it follows that:

$$\Pr_{\tau_k}(e \mid C_1, \ldots, C_n) = \sum_{k \leq l \leq n} \sum_{\epsilon_l(I_1,\ldots,I_n)} \prod_{i=1}^{n} \Pr(I_i \mid C_i). \tag{4}$$

4 The Poisson Binomial Distribution

It turns out that causal independence models defined in terms of the Boolean threshold function, as discussed above, are closely connected to the so-called Poisson binomial distribution known from statistics. Before establishing this connection, we review some its its relevant properties and discuss what these properties mean.

4.1 Its definition and relationship to the noisy threshold model

Let l denote the number of successes in n independent trials, where p_i is a probability of success in the ith trial, $i = 1, \ldots, n$; let $\mathbf{p} = (p_1, \ldots, p_n)$. The trials are then called *Poisson trials* [5], and $\mathrm{B}(l; \mathbf{p})$ denotes the *Poisson binomial distribution* [3, 10]:

$$\mathrm{B}(l; \mathbf{p}) = \left\{ \prod_{i=1}^{n} (1 - p_i) \right\} \sum_{1 \leq j_1 < \ldots < j_l \leq n} \prod_{z=1}^{l} \frac{p_{j_z}}{1 - p_{j_z}} \tag{5}$$

The Poisson trials have mean, defined as $\mu = \frac{1}{n} \sum_{i=1}^{n} p_i$, and variance, defined as $\sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (p_i - \mu)^2$. When the variance $\sigma^2 = 0$, i.e., the success probability p_i is a constant p, the trials are called Bernoulli trials and $\mathrm{B}(l; \mathbf{p})$ reduces to the binomial distribution: $\mathrm{B}(l; p) = \binom{n}{l} p^l (1 - p)^{n-l}$.

As it was assumed that absent causes do not contribute to the effect it follows that the conditional probabilities $\Pr_{\tau_k}(e \mid C_1, \ldots, C_n)$ depend only on

the 'active' causes, i.e., causes C_l that are equal to $\top$. Let $L = \{l \mid C_l = \top, l = 1, \ldots, n\}$, and let r be a bijective renumbering function, $r : L \leftrightarrow \{1, \ldots, |L|\}$, that respects the total order $<$ on the natural numbers, i.e., if $l < l'$, $l, l' \in L$, then $r(l) < r(l')$. Then, $\mathbf{p}(C_1, \ldots, C_n) = \{P(i_l \mid c_l) \mid l \in L\} = \{p_1, \ldots, p_{|L|}\}$, where $\Pr(i_l \mid c_l) = p_{r(l)}$, for each $l \in L$.

Then, the connection between the Poisson binomial distribution and the causal independence model using the noisy threshold function is as follows.

Proposition 1 *It holds that:*

$$\Pr_{\tau_k}(e \mid C_1, \ldots, C_n) = \sum_{k \leq l \leq |\mathbf{p}(C_1, \ldots, C_n)|} \mathrm{B}(l; \mathbf{p}(C_1, \ldots, C_n)) \tag{6}$$

Proof: Note that in Section 3.2 $\sum_{\epsilon_l(I_1, \ldots, I_n)} \prod_{i=1}^{n} \Pr(I_i \mid C_i)$ was defined as the probability that exactly l intermediate variables $I_1, \ldots, I_n$ are true. An intermediate variable I_l can be seen as an independent trial which has a probability of success $p_l = \Pr(i_l \mid C_l)$, which is equal to 0 if $C_l = \bot$, and otherwise equal to $\Pr(i_l \mid c_l)$. Thus, in order to find the probability that exactly l intermediate variables are true it is enough to look only at those intermediate variables that have a corresponding active cause. The set of the probabilities of such intermediate variables has been defined as $\mathbf{p}(C_1, \ldots, C_n)$. Considering the definition of the Poisson binomial distribution in Equation (5), Equation (4) yields what is stated in the premise of this proposition. □

If the number of active cause variables is smaller than the threshold k the conditional probability of the effect equals zero as it is shown in the following corollary.

Corollary 1 *Let* $|\mathbf{p}(C_1, \ldots, C_n)| < k, 1 \leq k \leq n$, *then* $\Pr_{\tau_k}(e \mid C_1, \ldots, C_n) = 0$.

Proof: This follows directly from Equation (6). □

From Proposition 1 it follows that in a noisy threshold model with interaction function τ_k and n cause variables, $\sum_{i=0}^{k-1} \binom{n}{i}$ of the probabilities $\Pr_{\tau_k}(e \mid C_1, \ldots, C_n)$ are set to 0, while the other $\sum_{i=k}^{n} \binom{n}{i}$ conditional probabilities of the effect such that $\rho(C_1, \ldots, C_n) \geq k$ are computed from the corresponding Poisson binomial distributions.

In comparison, the noisy AND model has only one conditional probability of the effect that is computed, i.e. $\Pr(e \mid C_1, \ldots, C_n)$ with $\rho(C_1, \ldots, C_n) = n$, while the other conditional probabilities are set 0. In the noisy OR model only the conditional probability $\Pr(e \mid C_1, \ldots, C_n)$ with $\rho(C_1, \ldots, C_n) = 0$ is set to 0 and the other conditional probabilities in the model are computed.

In the remainder of the paper, we review some probabilistic results for the Poisson binomial distribution. We also present examples illustrating the discussed properties. We use both n and ρ to define the cardinality of the set $\mathbf{p}$: n is used while discussing the properties of the Poisson binomial distribution and ρ is employed to analyse these properties in the context of noisy threshold models.

4.2 Statistical characterisation

4.2.1 Mean and variance

The *mean* $\mu_{\mathbf{p}}$ of the distribution $\mathrm{B}(i;\mathbf{p})$ is by definition equal to

$$\mu_{\mathbf{p}} = \sum_{i=0}^{n} i\,\mathrm{B}(i;\mathbf{p}),$$

and the *variance* $\sigma^2_{\mathbf{p}}$ is equal to

$$\sigma^2_{\mathbf{p}} = \sum_{i=0}^{n} (i-\mu_{\mathbf{p}})^2 \mathrm{B}(i;\mathbf{p}).$$

By means of some algebraic manipulation it can be shown that the mean $\mu_{\mathbf{p}}$ and variance $\sigma^2_{\mathbf{p}}$ obey the following equations: $\mu_{\mathbf{p}} = n\mu$ and $\sigma^2_{\mathbf{p}} = n\mu(1-\mu) - n\sigma^2$ [5]. In words: the mean of the Poisson binomial distribution $\mu_{\mathbf{p}}$ is equal to the sum of the probabilities $p_1, \ldots, p_n$. The variance $\sigma^2_{\mathbf{p}}$ increases as the set of probabilities $(p_1, \ldots, p_n)$ tends to be more and more homogeneous and attains its maximum as they become identical. Therefore, in the noisy threshold models a larger difference between the conditional probabilities $\Pr(i_l \mid C_l)$ causes a smaller variability of the success probability $\mathrm{B}(l;\mathbf{p})$.

4.2.2 Mode

The *mode* $m_{\mathbf{p}}$ of the Poisson binomial distribution $\mathrm{B}(l;\mathbf{p})$ is defined as a local maximum. Darroch has shown that [2]:

$$m_{\mathbf{p}} = \begin{cases} l & \text{if } l \le \mu_{\mathbf{p}} < l + \frac{1}{l+2} \\ l \text{ or } l+1 \text{ or both} & \text{if } l + \frac{1}{l+2} \le \mu_{\mathbf{p}} \le l + 1 - \frac{1}{n-l+1} \\ l+1 & \text{if } l + 1 - \frac{1}{n-l+1} < \mu_{\mathbf{p}} \le l+1 \end{cases} \tag{7}$$

where $0 \le l \le n$. Thus, the most probable number of successes $m_{\mathbf{p}}$ differs from the mean $\mu_{\mathbf{p}}$ by less than 1.

4.2.3 Shape of the Poisson binomial distribution

The Poisson binomial distribution is 'bell-shaped' [2]:

- the probabilities $\mathrm{B}(-1;\mathbf{p}), \mathrm{B}(0;\mathbf{p}), \mathrm{B}(1;\mathbf{p}), \ldots, \mathrm{B}(n;\mathbf{p}), \mathrm{B}(n+1;\mathbf{p})$ strictly increase and then strictly decrease, except that there may be at most two equal maxima;
- the probabilities $\mathrm{B}(-2;\mathbf{p}), \mathrm{B}(-1;\mathbf{p}), \mathrm{B}(0;\mathbf{p}), \ldots, \mathrm{B}(n;\mathbf{p}), \mathrm{B}(n+1;\mathbf{p}), \mathrm{B}(n+2;\mathbf{p})$ first increase convexly and then concavely, and then decrease concavely and then convexly.

The largest probabilities $\mathrm{B}(l;\mathbf{p})$ are concentrated around the mode $m_{\mathbf{p}}$ of the Poisson binomial distribution and the probabilities $\mathrm{B}(l;\mathbf{p})$ decline in the right and left tails. From this property it follows that combined knowledge of the

mode of the Poisson binomial distribution and the Boolean constants $\gamma_0, \ldots, \gamma_\rho$ can give some insight into the conditional probabilities of the effect E in the noisy threshold models.

5 Approximations of the Poisson Binomial Distribution

Since the Poisson binomial distribution has a complicated structure it is often approximated by other distributions that have well-known properties.

5.1 Poisson approximation

Let

$$\mathrm{P}(l; \mu_{\mathbf{p}}) = \frac{e^{-\mu_{\mathbf{p}}} \mu_{\mathbf{p}}^{l}}{l!}$$

denote the Poisson distribution. The following bound on the total variation distance between the Poisson binomial distribution and the Poisson distribution was established in [10]:

$$\sum_{l=0}^{\infty} |\mathrm{B}(l; \mathbf{p}) - \mathrm{P}(l; \mu_{\mathbf{p}})| < 2 \sum_{i=1}^{n} p_i^2. \tag{8}$$

Thus, the Poisson approximation will be accurate whenever the probabilities $p_1, \ldots, p_n$ are small.

We used the Kullback-Leibler divergence [9],

$$\mathrm{K}(\mathrm{B}, \mathrm{P}) = \sum_{i=0}^{n} \mathrm{B}(i; \mathbf{p}) \lg \frac{\mathrm{B}(i; \mathbf{p})}{\mathrm{P}(i; \mu_{\mathbf{p}})}, \tag{9}$$

to measure the distance between the Poisson binomial distribution and the Poisson approximation. Figure 3 plots the Kullback-Leibler divergence between the Poisson binomial distribution with mean $\mu_{\mathbf{p}} = 0.1, 0.2, 0.3, 0.4, 0.5$ and the Poisson approximation. It is not surprising that the approximation becomes more accurate as the number of probabilities n increases, i.e., the value of $\mu = \frac{\mu_{\mathbf{p}}}{n}$ decreases.

Figure 4 illustrates how accurate the Poisson approximation is for the Poisson binomial distribution $\mathrm{B}(l; \mathbf{p})$ when probabilities are small:
$\mathbf{p} = (0.01,\ 0.06,\ 0.09,\ 0.11,\ 0.14,\ 0.19)$.

5.2 Normal approximation

Another approximation for the Poisson binomial distribution found in the probabilistic literature is the approximation by the standard normal distribution [1, 13]. Let

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \tag{10}$$

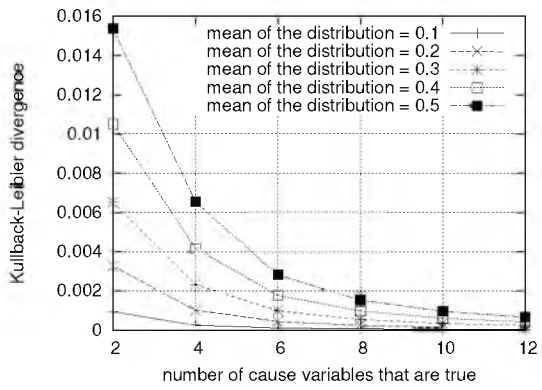

Figure 3: Kullback-Leibler divergence between the Poisson binomial distribution and the Poisson approximation.

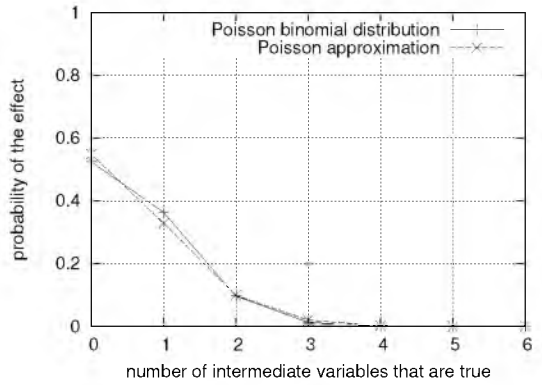

Figure 4: Poisson binomial distribution approximated by the Poisson distribution.

denote the normal density function, and let

$$\Phi(z) = \int_{-\infty}^{z} \phi(x)\mathrm{d}x. \tag{11}$$

Then for every Poisson binomial distribution B with mean $\mu_{\mathbf{p}}$, variance $\sigma_{\mathbf{p}}^2$,

$$\max_{0 \le l \le n} \left| \sum_{i=0}^{l} \mathrm{B}(i; \mathbf{p}) - \Phi\left(\frac{l - \mu_{\mathbf{p}}}{\sigma_{\mathbf{p}}} \right) \right| < \frac{0.7975}{\sigma_{\mathbf{p}}}. \tag{12}$$

Thus, we see that the normal approximation is accurate when the standard deviation of the Poisson binomial distribution

$$\sigma_{\mathbf{p}} = \sqrt{n\left(\mu(1-\mu) - \sigma^2\right)}$$

is large, i.e., when $n \to \infty$.

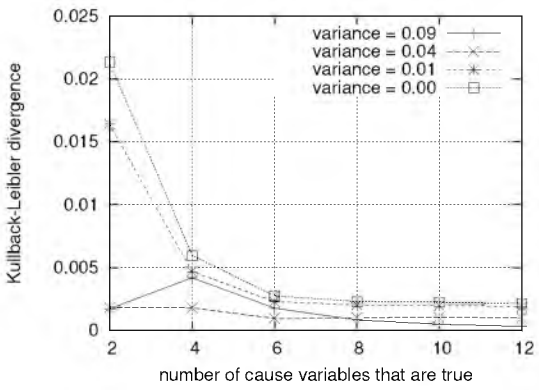

Figure 5: Kullback-Leibler divergence between the Poisson binomial distribution and the normal approximation.

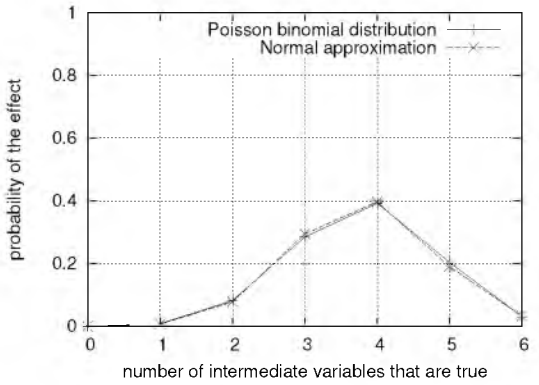

Figure 6: Poisson binomial distribution approximated by the normal distribution.

Let

$$\mathrm{N}(i; \mu_{\mathbf{p}}; \sigma_{\mathbf{p}}) = \Phi\left(\frac{i - \frac{1}{2} - \mu_{\mathbf{p}}}{\sigma_{\mathbf{p}}}\right) - \Phi\left(\frac{i + \frac{1}{2} - \mu_{\mathbf{p}}}{\sigma_{\mathbf{p}}}\right)$$

be an approximation of $\mathrm{B}(i; \mathbf{p})$. We use the Kullback-Leibler divergence $\mathrm{K}(\mathrm{B}, \mathrm{N})$ to measure the distance between the Poisson binomial distribution and the normal approximation.

Figure 5 shows a plot of the Kullback-Leibler divergence between the Poisson binomial distribution and the normal approximation for the probabilities $(p_1, \ldots, p_n)$ that have the same mean $\mu = 0.6$ but differ in variance σ^2 and number of probabilities n. We can see that the accuracy of the approximation improves when either the variance σ^2 or the number of probabilities n increases.

Figure 6 shows the accuracy of the normal approximation for the Poisson binomial distribution by an example: $\mathbf{p} = (0.2,\ 0.4,\ 0.6,\ 0.8,\ 0.8,\ 0.99)$, i.e.

the standard deviation of the distribution is high $\sigma_{\mathbf{p}} = 0.985$.

The Poisson binomial distribution can also be approximated by the binomial distribution [14]. The binomial approximation is accurate whenever the variance σ^2 is small.

6 Experimental Results

In Section 2, we have described a Bayesian network, shown in Figure 1, that is aimed at assisting clinicians in the diagnosis and treatment of patient with VAP in the critical-care unit. It was noted that a limitation of the present Bayesian network model is that it attempts to cover every possible bacterium by which the patient is being colonised, even if it is normally unlikely that VAP is caused by every possible bacterium included in the model at the same time. Such behaviour was accomplished by defining the interaction between the individual coverage vertices, each of them with an incoming arc from a colonisation vertex and the antibiotic vertex, by means of a logical AND. This results in the prescription of antibiotic treatment with a spectrum that is very often too broad. The hypothesis we investigated was, therefore, whether any noisy threshold model with function τ_k, where $1 < k < 7$, would yield a performance superior to that of the noisy AND model.

Initial experimentation with various Bayesian networks obtained by replacing the noisy AND by a noisy threshold model showed that for $k = 1, 2, \ldots, 7$ the noisy threshold model yielded posterior probability distributions where for $k = 1, 2$ the antibiotics prescribed were always very narrow, even when the patient was assumed to have an infection caused by 5 different bacteria. For $k = 5, 6$ the antibiotics prescribed were always broad-spectrum antibiotics.

Further experimental results were obtained by applying Bayesian networks with the noisy AND and noisy threshold model for $k = 3$ and $k = 4$ to the time-series data of 6 different patients. At each time point, each of the these Bayesian networks was used to compute the spectrum of the antibiotics considered optimal. The results are summarised in Table 1. On average, the results obtained by the noisy threshold model with $k = 4$ are best, as it never prescribed broad-spectrum antibiotics; in addition, it did not prescribe antibiotics with very narrow spectrum if the infection was caused by 2 or 3 bacteria. In conclusion, replacing the noisy AND model by a noisy threshold model did give rise to performance improvement of the Bayesian network.

7 Discussion

In this paper, we expanded the space of possible causal independence models by introducing new models based on the Boolean threshold function, which we have called noisy threshold models. It was shown that there is a close connection between the probability distribution of noisy threshold models and the Poisson binomial distribution from statistics. We have investigated what the well-studied properties of the Poisson binomial distribution mean in the

Table 1: Results of the prescription of antibiotics to 6 patients with VAP caused by 1, 2 or 3 bacteria using various causal independence models, in comparison to the clinician. Abbreviations of antibiotic spectrum: v: very narrow; n: narrow; i: intermediate; b: broad. As some of the patients were colonised by different bacteria at different days, there are 10 cases of the prescription of antibiotics.

#Bacteria / Model	1	2	3
Clinician	3n2i	1i3n	1i
Noisy threshold ($k = 3$)	5v	4v	1v
Noisy threshold ($k = 4$)	3v2n	4n	1n
Noisy AND	1i4b	1b3i	1i

context of these newly introduced models. The noisy threshold models can be looked upon as spanning a spectrum of causal independence models with the noisy OR and noisy AND as extremes.

Even though this paper has focused on the conditional probability distributions of noisy threshold models, most of the presented theory can be exploited as a basis for the assessment of probability distributions of causal independence model where the interaction function is defined in terms of symmetric Boolean functions. This is a consequence of the fact that any symmetric Boolean function can be decomposed into a disjunction of Boolean exact functions in conjunction with Boolean constants. This basic property indicates that the theory developed in this paper has an even wider application, which, however, still needs to be explored.

Finally, it was shown that the noisy threshold model is also useful from a practical point of view by using it as a basis for the refinement of an existing real-world Bayesian network.

Acknowledgments

The authors are grateful to Stefan Visscher for his contribution to the experimental results.

References

[1] P. van Beek, *An application of Fourier methods to the problem of sharpening the Berry-Essen inequality.* Z. Wahrscheinlichkeitstheorie verw. Gebiete, 23:187-196, 1972.

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

[3] A.W.P. Edwards, *The Meaning of Binomial Distribution.* Nature 186, 1074, 1960.

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

[5] W. Feller, *An Introduction to Probability Theory and Its Applications*. John Wiley, 1968.

[6] D. Heckerman, *Causal independence for knowledge acquisition and inference*. Proceedings of the Ninth Conference on Uncertainty in Artificial Intelligence, 122-127, 1993.

[7] K. Jogdeo and S.M. Samuels, *Monotone convergence of binomial probabilities and a generalization of Ramanujan's equation*. The Annals of Mathematical Statistics, 39:1191-1195, 1968.

[8] R. Jurgelenaite and P.J.F. Lucas, *Exploiting causal independence in large Bayesian networks*. To appear in: Knowledge Based Systems Journal, 2005.

[9] S. Kullback and R.A. Leibler, *On information and sufficiency*. The Annals of Mathematical Statistics, 22:79-86, 1951.

[10] L. Le Cam, *An approximation theorem for the Poisson binomial distribution*. Pacific Journal of Mathematics, 10: 1181-1197, 1960.

[11] P.J.F. Lucas, N.C. de Bruijn, K. Schurink and I.M. Hoepelman, *A probabilistic and decision-theoretic approach to the management of infectious disease at the ICU*. Artificial Intelligence in Medicine, 19:251-279, 2000.

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

[13] J. Pitman, *Probabilistic bounds on the coefficients of polynomials with only real zeros*. Journal of Combinatorial Theory, Series A, 77:279-303, 1997.

[14] B. Roos, *Binomial approximation to the Poisson binomial distribution: the Krawtchouk expansion*. Theory of Probability and Its Applications, 45:258-272, 2001.

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

[16] N.L. Zhang and D. Poole, *Exploiting causal independence in Bayesian networks inference*. Journal of Artificial Intelligence Research, 5:301-328, 1996.