

PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a preprint version which may differ from the publisher's version.

For additional information about this publication click this link.

<http://hdl.handle.net/2066/35578>

Please be advised that this information was generated on 2017-12-06 and may be subject to change.

Noisy Threshold Functions for Modelling Causal Independence in Bayesian Networks*

Rasa Jurgelenaite, Tom Heskes and Peter Lucas
Institute for Computing and Information Sciences
Radboud University Nijmegen
Toernooiveld 1, 6525 ED Nijmegen, The Netherlands
E-mail: {rasa, tomh, peter}@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. We present and analyse recursive methods as well as approximation and bounding techniques to assess the conditional probabilities in the noisy threshold models.

Keywords: Bayesian networks, causal independence, parameter assessment, knowledge representation, probability theory.

1 Introduction

Bayesian networks [22] 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 [6, 11, 27, 31] 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.

*This research, carried out in the TimeBayes project, was supported by the Netherlands Organization for Scientific Research (NWO) under project number FN4556.

Causal independence assumptions are often used in practical Bayesian network models [18, 26]. 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 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 [28]. 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*. We present a theoretical basis for the introduced models and study in some depth different ways to assess their conditional probability distributions. The presented theory is illustrated by examples which motivate the use of threshold functions.

The structure of this paper is 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. Section 5 explains two recursive methods to compute the Poisson binomial distribution while Section 6 presents and investigates the approximation and bounding techniques that assess the probabilities in linear number of operations. Finally, in Section 7, we summarise what has been achieved by this research.

2 Review of Bayesian Networks

A *Bayesian network* $\mathcal{B} = (G, \text{Pr})$ 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 $\text{Pr}(V \mid \pi(V))$, 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:

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

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 $V = \perp$ (false). An expression such as

$$\sum_{\psi(H_1, \dots, H_n) = e} g(H_1, \dots, H_n)$$

stands for summing over all possible values of $g(H_1, \dots, H_n)$ for all possible values of the variables H_k for which the constraint $\psi(H_1, \dots, H_n) = e$ holds.

Let us look at an example that provides motivation for this paper. Rheumatoid arthritis is a chronic, systemic, inflammatory disease that mainly affects the synovial membranes

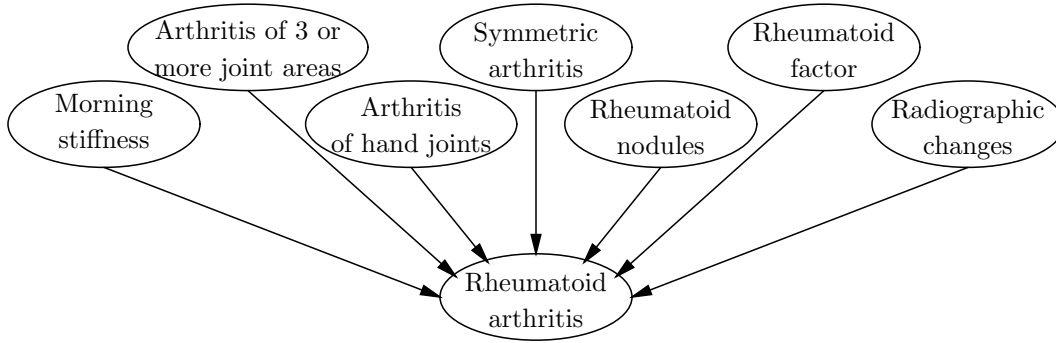


Figure 1: Example Bayesian network, modelling the interaction between the diagnostic criteria and rheumatoid arthritis.

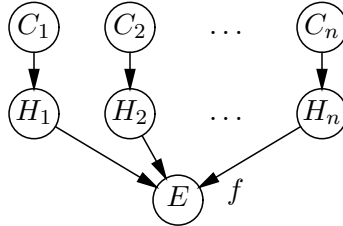


Figure 2: Causal independence model.

of multiple joints in the body. Although in most cases diagnosis of rheumatoid arthritis is generally made without difficulty, some persons have atypical clinical and radiological features. As a consequence, certain diagnostic criteria have been proposed: (1) morning stiffness in and around joints lasting at least 1 hour before maximal improvement; (2) soft tissue swelling of three or more joint areas observed by a physician; (3) swelling of the proximal interphalangeal, metacarpophalangeal or wrist joints; (4) symmetric swelling; (5) presence of rheumatoid nodules; (6) presence of rheumatoid factor; (7) radiographic erosions with or without periarticular osteopenia in hand and/or wrist joints. Rheumatoid arthritis is defined by the presence of four or more of the foregoing criteria [1]. A Bayesian network modelling the described interaction is shown in Figure 1.

The given example as well as many similar examples for diagnosis of other diseases, in particular mental disorders, has a simple underlying logic which suggests that the state of the diagnosis variable depends on the number of diagnostic criteria that are present.

As the reader can notice, the discussed examples are concerned with diagnosis of the disease and thus do not follow cause-effect interpretation which is characteristic for causal independence models. Our feeling is that the discussed underlying logic should not be limited to diagnostic problems and can be successfully used for the prognosis of the disease. Therefore, another medical example which solves the problem of prognosis of the disease and for which we have quantitative information will be presented in Section 6.

3 Causal Modelling and Boolean Functions

3.1 Causal Independence

Causal independence (also known as independence of causal influence) 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, \dots, C_n influence a given common effect E through hidden variables H_1, \dots, H_n and a deterministic function f , called the *interaction function*. The impact of each cause C_i on the common effect E is independent of each other cause $C_j, j \neq i$. The hidden variable H_i is considered to be a contribution of the cause variable C_i to the common effect E . The function f represents in which way the hidden effects H_i , and indirectly also the causes C_i , 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 $H_i, i = 1, \dots, n$, and $E = \top$ is satisfied, then it holds that $e = f(H_1, \dots, H_n)$. It is assumed that $\Pr(e \mid H_1, \dots, H_n) = 1$ if $f(H_1, \dots, H_n) = e$, and $\Pr(e \mid H_1, \dots, H_n) = 0$ if $f(H_1, \dots, H_n) = \bar{e}$.

A causal independence model is defined in terms of the causal parameters $\Pr(H_i \mid C_i)$, for $i = 1, \dots, n$ and the function $f(H_1, \dots, H_n)$. Most papers on causal independence models assume that absent causes do not contribute to the effect [11, 22]. In terms of probability theory this implies that it holds that $\Pr(h_i \mid \bar{c}_i) = 0$; as a consequence, it holds that $\Pr(\bar{h}_i \mid \bar{c}_i) = 1$. In this paper we make the same assumption.

In situations in which model does not capture all possible causes it is useful to introduce a *leaky cause* which summarizes the unidentified causes contributing to the effect and is assumed to be always present [12]. We will not separate the leaky cause from the other causes as in an arithmetic context they do not differ.

The conditional probability of the occurrence of the effect E given the causes C_1, \dots, C_n , i.e., $\Pr(e \mid C_1, \dots, C_n)$, can be obtained from the causal parameters $\Pr(H_i \mid C_i)$ as follows [21, 31]:

$$\Pr(e \mid C_1, \dots, C_n) = \sum_{f(H_1, \dots, H_n) = e} \prod_{i=1}^n \Pr(H_i \mid C_i). \quad (1)$$

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 were presented in [21] and [17]. However, there are 2^{2^n} different n -ary Boolean functions [8, 28]; 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* [28] 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. $\epsilon_l(H_1, \dots, H_n) = \top$, if $\sum_{i=1}^n \nu(H_i) = l$ with $\nu(H_i)$ equal to 1, if H_i is equal to true and 0 otherwise. A symmetric Boolean function can be decomposed in terms of the exact functions ϵ_l as follows [28]:

$$f(H_1, \dots, H_n) = \bigvee_{i=0}^n \epsilon_i(H_1, \dots, H_n) \wedge \gamma_i \quad (2)$$

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

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. $\tau_k(H_1, \dots, H_n) = \top$, if $\sum_{i=1}^n \nu(H_i) \geq k$ with $\nu(H_i)$ equal to 1, if H_i is equal to true and 0 otherwise. To express it in the Boolean constants we have: $\gamma_0 = \dots = \gamma_{k-1} = \perp$ and $\gamma_k = \dots = \gamma_n = \top$. Note that the noisy OR corresponds to a threshold function τ_k with $k = 1$ and the noisy AND corresponds to a threshold function τ_k with $k = n$. Hence, these two commonly used Boolean functions are the extremes of a spectrum of Boolean functions based on the threshold function. 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.

Modelling the interaction among the diagnostic factors and presence or absence of rheumatoid arthritis, as shown in Figure 1, by means of a causal independence model with a threshold function τ_k where $k = 4$ would preserve the underlying logic. However, even though the noisy threshold models have been applied in a medical Bayesian network [29], the conditional probability distributions of these models have not been investigated. In the following we therefore explore properties of the noisy threshold models, and look at different ways to compute their conditional probability distributions.

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, \dots, C_n can be decomposed in terms of probabilities that exactly i hidden variables H_1, \dots, H_n are true, as follows:

$$\Pr(e \mid C_1, \dots, C_n) = \sum_{\substack{0 \leq i \leq n \\ \gamma_i}} \sum_{\epsilon_i(H_1, \dots, H_n)} \prod_{j=1}^n \Pr(H_j \mid C_j). \quad (3)$$

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, \dots, C_n in a noisy threshold model with interaction function τ_k as $\Pr_{\tau_k}(e \mid C_1, \dots, C_n)$. Then, from Equation (3) it follows that:

$$\Pr_{\tau_k}(e \mid C_1, \dots, C_n) = \sum_{k \leq i \leq n} \sum_{\epsilon_i(H_1, \dots, H_n)} \prod_{j=1}^n \Pr(H_j \mid C_j). \quad (4)$$

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. In this section we establish this connection.

Let l denote the number of successes in n independent trials, where p_i is a probability of success in the i th trial, $i = 1, \dots, n$; let $\mathbf{p} = (p_1, \dots, p_n)$. The trials are then called *Poisson trials* [9], and $B(l; \mathbf{p})$ denotes the *Poisson binomial distribution* (also known as distribution

of the number of successes of independent trials) [7, 19]:

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

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

As it was assumed that absent causes do not contribute to the effect it follows that the conditional probabilities $\Pr_{\tau_k}(e \mid C_1, \dots, C_n)$ depend only on the ‘active’ causes, i.e., causes C_i that are equal to \top . Let $L = \{i \mid C_i = \top, i = 1, \dots, n\}$, and let r be a bijective renumbering function, $r : L \leftrightarrow \{1, \dots, |L|\}$, that respects the total order $<$ on the natural numbers, i.e., if $i < i'$, $i, i' \in L$, then $r(i) < r(i')$. Then, $\mathbf{p}(C_1, \dots, C_n) = \{P(h_i \mid c_i) \mid i \in L\} = \{p_1, \dots, p_{|L|}\}$, where $\Pr(h_i \mid c_i) = p_{r(i)}$, for each $i \in L$.

For example, to compute the conditional probability $\Pr_{\tau_k}(e \mid c_1, \bar{c}_2, \bar{c}_3, c_4)$, the set of probabilistic parameters is $\mathbf{p}(c_1, \bar{c}_2, \bar{c}_3, c_4) = \{p_1 = \Pr(h_1 \mid c_1), p_2 = \Pr(h_4 \mid c_4)\}$.

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:*

$$\Pr_{\tau_k}(e \mid C_1, \dots, C_n) = \sum_{k \leq i \leq |\mathbf{p}(C_1, \dots, C_n)|} B(i; \mathbf{p}(C_1, \dots, C_n)). \quad (6)$$

Proof: Note that in Section 3.2 $\sum_{e \mid (H_1, \dots, H_n)} \prod_{j=1}^n \Pr(H_j \mid C_j)$ was defined as the probability that exactly l hidden variables H_1, \dots, H_n are true. A hidden variable H_i can be seen as an independent trial which has a probability of success $\Pr(h_i \mid C_i)$, which is equal to 0 if $C_i = \perp$, and otherwise equal to $\Pr(h_i \mid c_i)$. Thus, in order to find the probability that exactly l hidden variables are true it is enough to look only at those hidden variables that have a corresponding active cause. Such a set of the probabilities $\Pr(h_i \mid c_i)$ has been defined as $\mathbf{p}(C_1, \dots, C_n)$. Considering the definition of the Poisson binomial distribution, Equation (4) yields what is stated in the premise of this proposition. \square

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

Proof: This follows directly from Equation (6). \square

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 $\Pr_{\tau_k}(e \mid C_1, \dots, C_n)$ are set to 0, while the other $\sum_{i=k}^n \binom{n}{i}$ conditional probabilities of the effect such that $|\mathbf{p}(C_1, \dots, C_n)| \geq k$ 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. $\Pr(e \mid C_1, \dots, C_n)$ with $|\mathbf{p}(C_1, \dots, C_n)| = n$, while the other conditional probabilities are set 0. In the noisy OR model only the conditional probability $\Pr(e \mid C_1, \dots, C_n)$

with $|\mathbf{p}(C_1, \dots, C_n)| = 0$ is set to 0 and the other conditional probabilities in the model are computed.

In [30] it was shown that the threshold function is a monotonic function, i.e., for all k it holds:

$$\sum_{i=1}^n \nu(H_i) \leq \sum_{i=1}^{n'} \nu(H'_i) \Rightarrow \nu(\tau_k(H_1, \dots, H_n)) \leq \nu(\tau_k(H'_1, \dots, H'_{n'})).$$

In words, if the number of hidden variables that are true increases then the output of the threshold function goes from false to true or remains the same.

We can also show that the probability function $\Pr_{\tau_k}(e | C_1, \dots, C_n)$ is monotonic with respect to the probability $\Pr(h_t | c_t)$ when the other causes and their corresponding probabilities are fixed.

Proposition 2 *The probability function $\Pr_{\tau_k}(e | C_1, \dots, C_n)$ is monotonically increasing with respect to any conditional probability $\Pr(h_t | c_t)$, $1 \leq t \leq n$.*

Proof: The Poisson Binomial probabilities have the following property [5]:

$$B(l; \mathbf{p}) = B(l; \mathbf{p}_{\setminus m})(1 - p_m) + B(l - 1; \mathbf{p}_{\setminus m})p_m.$$

Let $p_m = \Pr(h_t | c_t) \in \mathbf{p}(C_1, \dots, C_n)$ and $\rho = |\mathbf{p}(C_1, \dots, C_n)|$, then:

$$\begin{aligned} \Pr_{\tau_k}(e | C_1, \dots, C_n) &= (1 - p_m) \sum_{k \leq i \leq \rho} B(i; \mathbf{p}(C_1, \dots, C_{t-1}, C_{t+1}, \dots, C_n)) \\ &\quad + p_m \sum_{k-1 \leq i \leq \rho-1} B(i; \mathbf{p}(C_1, \dots, C_{t-1}, C_{t+1}, \dots, C_n)). \end{aligned}$$

As $B(\rho; \mathbf{p}(C_1, \dots, C_{t-1}, C_{t+1}, \dots, C_n)) = 0$, the equation becomes:

$$\begin{aligned} \Pr_{\tau_k}(e | C_1, \dots, C_n) &= \sum_{k \leq i \leq \rho-1} B(i; \mathbf{p}(C_1, \dots, C_{t-1}, C_{t+1}, \dots, C_n)) \\ &\quad + p_m B(k - 1; \mathbf{p}(C_1, \dots, C_{t-1}, C_{t+1}, \dots, C_n)). \end{aligned}$$

The Poisson binomial probabilities in the equation are independent of p_m , thus the probability function $\Pr_{\tau_k}(e | C_1, \dots, C_n)$ is monotonically increasing with respect to $p_m = \Pr(h_t | c_t)$. \square

In the remainder of the paper, we review the exact, approximation and bounding methods to compute the conditional probabilities in the noisy threshold models. We also present examples illustrating the discussed methods. We use both n and $\rho = |\mathbf{p}(C_1, \dots, C_n)|$ 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.

5 Recursive Computation of the Poisson Binomial Distribution

Computing the probability $B(l; \mathbf{p})$ naively, e.g. through (5), one needs to sum $\frac{n!}{l!(n-l)!}$ terms, which is impractical even when l and n are of moderate sizes. Recursive formulas require a

smaller number of operations for computing this sum. We review two recursive methods to compute the Poisson binomial probabilities in a recursive way.

The first method was presented by Chen et al. [4]. Let $B(l; \mathbf{p})$ denote l successes in n Poisson trials, then

$$B(l; \mathbf{p}) = R(l) \prod_{i=1}^n (1 - p_i)$$

where $R(l)$ can be computed recursively from

$$R(l) = \frac{1}{l} \sum_{i=1}^l (-1)^{i+1} T(i) R(l - i)$$

with $T(i) = \sum_{j=1}^n \left(\frac{p_j}{1-p_j} \right)^i$ for any $i \geq 1$.

The second recursive method to calculate the Poisson binomial distribution is very similar to the method presented by Howard [14]. The Poisson binomial probability $B(l; \mathbf{p})$ can be computed recursively from

$$B(l; (p_1, \dots, p_n)) = B(l; (p_1, \dots, p_{n-1}))(1 - p_n) + B(l - 1; (p_1, \dots, p_{n-1}))p_n.$$

Both presented methods require the order $l \times n$ operations to compute the Poisson binomial probability $B(l; \mathbf{p})$, $|\mathbf{p}| = n$.

Using

$$B(l; (p_1, \dots, p_n)) = B(n - l; (1 - p_1, \dots, 1 - p_n))$$

and

$$\sum_{i=k}^n B(i; (p_1, \dots, p_n)) = 1 - \sum_{i=0}^{k-1} B(i; (p_1, \dots, p_n))$$

we can always rewrite Equation (6) so that the probability $\Pr_{\tau_k}(e | C_1, \dots, C_n)$ can be computed in order $\rho \times \min(k, \rho - k)$ steps applying both methods.

6 Approximations and Bounds for the Poisson Binomial Distribution

In Section 5 we presented a recursive method to compute the Poisson Binomial probabilities in a quadratic number of operations. However, it can still be too expensive, thus in this section we present approximations and bounds that can be computed in a linear number of operations.

To illustrate the quality of the presented results we will use a noisy threshold model shown in Figure 3 that represents a real-world medical problem of prognosis in patients with gastric non-Hodgkin's lymphoma. Gastric non-Hodgkin's lymphoma is a type of cancer of the lymphatic system, the disease-fighting network spread throughout the body, which originates in the stomach. The following pretreatment variables are used as the prognostic factors in the model: (1) age; (2) general health status; (3) bulky disease; (4) histological classification;

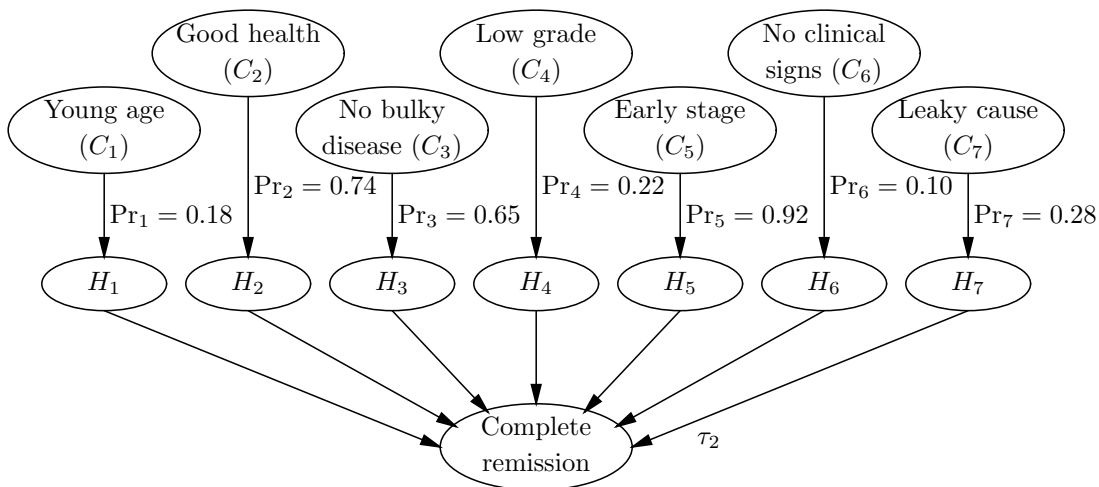


Figure 3: Noisy threshold model modelling complete remission following the treatment of gastric Non-Hodgkin lymphoma. Pr_k is a shortening for $\text{Pr}(h_k | c_k)$, and τ_2 stands for Boolean threshold function with $k = 2$. Leaky cause C_7 is assumed to be always active.

(5) stage of the cancer; (6) clinical signs (hemorrhage, perforation, obstruction) due to the disease; (7) leaky cause that stands for unidentified prognostic factors. The prognosis stands for endoscopically verified result of the treatment, six to eight weeks after treatment with complete remission defining a situation in which all clinical signs of disease disappear with the treatment. A more elaborate description of the domain can be found in [20].

6.1 Noisy Threshold Models for Classification

Classification is one of the ways to use the noisy threshold models. In this case, the causes can be interpreted as feature variables, the effect as the class variable, and the conditional probability $\text{Pr}_{\tau_k}(e | C_1, \dots, C_n)$ as the class probability. For binary classifiers the default *classification threshold* (not to be confused with the threshold function) typically is set to $\frac{1}{2}$. To classify a data instance using a noisy threshold classifier as defined there is no need to compute the exact probability $\text{Pr}_{\tau_k}(e | C_1, \dots, C_n)$, it is enough to know whether $\text{Pr}_{\tau_k}(e | C_1, \dots, C_n) \leq \frac{1}{2}$ or $\text{Pr}_{\tau_k}(e | C_1, \dots, C_n) \geq \frac{1}{2}$. We will show that in many cases there is a simple way to determine which state of the effect/class variable is more likely to occur.

We start by introducing some properties of the Poisson binomial distribution that are needed to derive this result.

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

$$m_{\mathbf{p}} = \sum_{i=0}^n i B(i; \mathbf{p}).$$

By means of some algebraic manipulation it can be shown that for the Poisson binomial distribution $B(l; \mathbf{p})$ the mean $m_{\mathbf{p}}$ is equal to the sum of the probabilities p_1, \dots, p_n [9]:

$$m_{\mathbf{p}} = \sum_{i=1}^n p_i.$$

The *median* $M_{\mathbf{p}}$ of the discrete probability distribution $\mathbf{B}(l; \mathbf{p})$ is the integer number such that:

$$(1) \quad \sum_{i=0}^{M_{\mathbf{p}}} \mathbf{B}(i; \mathbf{p}) \geq \frac{1}{2},$$

$$(2) \quad \sum_{i=M_{\mathbf{p}}}^n \mathbf{B}(i; \mathbf{p}) \geq \frac{1}{2}.$$

Jogdeo and Samuels [16] established a connection between the mean $m_{\mathbf{p}}$ and the median $M_{\mathbf{p}}$ of the Poisson binomial distribution:

$$M_{\mathbf{p}} = \begin{cases} l & \text{if } m_{\mathbf{p}} = l \\ l \text{ or } l + 1 & \text{if } l < m_{\mathbf{p}} < l + 1 \end{cases} \quad (7)$$

where $0 \leq l \leq n$ is an integer.

Knowing the connection between the median and the mean we can distinguish between the conditional probabilities where the effect E is more likely to be present and the conditional probabilities where the effect E is more likely to be absent.

Proposition 3 *Let $|\mathbf{p}(C_1, \dots, C_n)| \geq k, 1 \leq k \leq n$, then*

- $\Pr_{\tau_k}(e \mid C_1, \dots, C_n) \geq \frac{1}{2}$ if $k \leq m_{\mathbf{p}}$,
- $\Pr_{\tau_k}(e \mid C_1, \dots, C_n) \leq \frac{1}{2}$ if $k \geq m_{\mathbf{p}} + 1$.

Proof: Equation (6) can be written in the form

$$\Pr_{\tau_k}(e \mid C_1, \dots, C_n) = \sum_{i=k}^{M_{\mathbf{p}}-1} \mathbf{B}(i; \mathbf{p}) + \sum_{i=M_{\mathbf{p}}}^{\rho} \mathbf{B}(i; \mathbf{p}) \quad \text{if } k \leq M_{\mathbf{p}},$$

$$\Pr_{\tau_k}(e \mid C_1, \dots, C_n) = 1 - \sum_{i=0}^{M_{\mathbf{p}}} \mathbf{B}(i; \mathbf{p}) - \sum_{i=M_{\mathbf{p}}+1}^{k-1} \mathbf{B}(i; \mathbf{p}) \quad \text{if } k \geq M_{\mathbf{p}} + 1.$$

Then from the definition of the median $M_{\mathbf{p}}$ we get the following inequalities:

$$\Pr_{\tau_k}(e \mid C_1, \dots, C_n) \geq \frac{1}{2} \quad \text{if } k \leq M_{\mathbf{p}},$$

$$\Pr_{\tau_k}(e \mid C_1, \dots, C_n) \leq \frac{1}{2} \quad \text{if } k \geq M_{\mathbf{p}} + 1.$$

From Equation (7) it follows that $M_{\mathbf{p}}$ equals $\lfloor m_{\mathbf{p}} \rfloor$ or $\lceil m_{\mathbf{p}} \rceil$, thus the inequalities above can be written as:

$$\Pr_{\tau_k}(e \mid C_1, \dots, C_n) \geq \frac{1}{2} \quad \text{if } k \leq m_{\mathbf{p}},$$

$$\Pr_{\tau_k}(e \mid C_1, \dots, C_n) \leq \frac{1}{2} \quad \text{if } k \geq m_{\mathbf{p}} + 1.$$

□

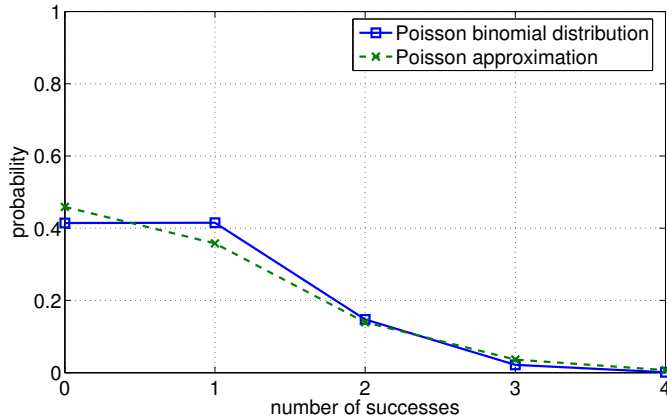


Figure 4: Example of the Poisson approximation for the Poisson binomial distribution.

In the noisy threshold model modelling complete remission of gastric non-Hodgkin lymphoma there are $2^6 = 64$ conditional probabilities of the effect. From Proposition 3 we find that 23 of the probabilities are equal or larger than $\frac{1}{2}$ and 9 probabilities are equal or smaller than $\frac{1}{2}$. Unfortunately, the other 32 probabilities cannot be determined as their $m_{\mathbf{p}}$ value falls into the interval $(1; 2)$. However, one should notice that this result strongly depends on the size of the model, i.e. in a bigger noisy threshold model where $m_{\mathbf{p}}$ values vary more, a bigger percentage of the conditional probabilities of the effect can be classified in the described way.

6.2 Approximations for the Poisson Binomial Distribution

The Poisson binomial distribution can be approximated by other distributions that are computed in linear rather than quadratic time.

6.2.1 Poisson Approximation

Let

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

denote the Poisson distribution. The following bound on the total variation distance between the Poisson binomial distribution and the Poisson distribution was established in [19]:

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

Thus, the Poisson approximation will be accurate whenever the probabilities (p_1, \dots, p_n) are small.

Let us take an example from the noisy threshold model for the gastric non-Hodgkin lymphoma. To compute the conditional probability $\Pr(e \mid c_1, \bar{c}_2, \bar{c}_3, c_4, \bar{c}_5, c_6)$ we have to compute the Poisson binomial distribution $B(l; (0.18, 0.22, 0.10, 0.28))$. Figure 4 shows the quality of the Poisson approximation for this distribution.

6.2.2 Normal Approximation

Another approximation for the Poisson binomial distribution found in the probabilistic literature is the approximation by the standard normal distribution [2], [24]. Let

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

denote the normal density function, and let

$$\Phi(z) = \int_{-\infty}^z \phi(x)dx.$$

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

$$\max_{0 \leq i \leq n} \left| \sum_{j=0}^i B(j; \mathbf{p}) - \Phi\left(\frac{i - m_{\mathbf{p}}}{\sigma_{\mathbf{p}}}\right) \right| < \frac{0.7975}{\sigma_{\mathbf{p}}}.$$

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

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

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

Let

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

be a normal approximation of $B(i; \mathbf{p})$.

To illustrate the quality of this approximation once again we use an example from the noisy threshold model for the gastric non-Hodgkin lymphoma. We chose a conditional probability with the biggest number of active causes $n = 7$. To assess the conditional probability $\Pr(e \mid c_1, c_2, c_3, c_4, c_5, c_6)$ we need to compute the Poisson binomial distribution $B(l; (0.18, 0.74, 0.65, 0.22, 0.92, 0.10, 0.28))$. Figure 5 shows the normal approximation of this distribution.

The Poisson binomial distribution can also be approximated by the binomial distribution [25]. The binomial approximation is accurate whenever the variance σ^2 is small.

6.3 Bounds for the Conditional Probabilities of Effect in the Noisy Threshold Models

Even though the presented approximations for the Poisson binomial distribution can be very handy, in some cases none of them is accurate enough to be employed. In such cases the bounds for the Poisson binomial distribution can provide information on the conditional probabilities of the noisy threshold models.

A number of bounds for the Poisson binomial distribution based on various characteristics of the underlying set of probabilities p_1, \dots, p_n can be found in the literature [13], [10], [3], [15], [23]. All bounds for the Poisson binomial distribution concern bounds for cumulative probabilities, i.e. they are well suited to bound the conditional probabilities in the noisy threshold models. We see Hoeffding's inequalities and the Percus and Percus bounds as most suitable for the domain of noisy threshold models.

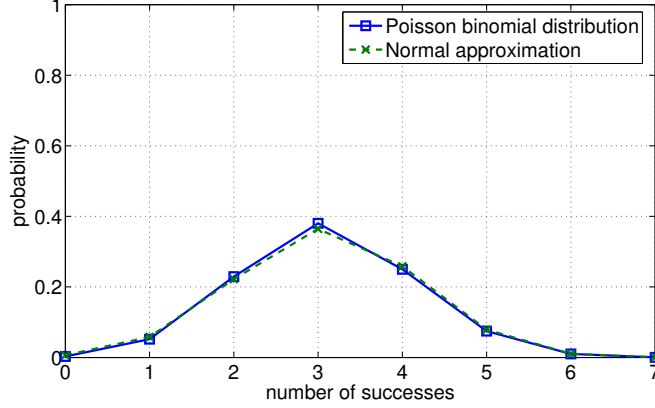


Figure 5: Example of the normal approximation for the Poisson binomial distribution.

6.3.1 Hoeffding's inequalities

Let $B(l; \mu; n) = \binom{n}{l} \mu^l (1 - \mu)^{n-l}$ be a binomial distribution. Hoeffding [13] presents the following bounds for the probabilities $\sum_{i=0}^c B(i; \mathbf{p})$ and $\sum_{i=b}^c B(i; \mathbf{p})$ given the mean $m_{\mathbf{p}}$ of the Poisson binomial distribution:

$$0 \leq \sum_{i=0}^c B(i; \mathbf{p}) \leq \sum_{i=0}^c B(i; \mu; n) \quad \text{if } 0 \leq c \leq m_{\mathbf{p}} - 1 \quad (8)$$

$$\sum_{i=0}^c B(i; \mu; n) \leq \sum_{i=0}^c B(i; \mathbf{p}) \leq 1 \quad \text{if } m_{\mathbf{p}} \leq c \leq n \quad (9)$$

$$\sum_{i=b}^c B(i; \mu; n) \leq \sum_{i=b}^c B(i; \mathbf{p}) \leq 1 \quad \text{if } 0 \leq b \leq m_{\mathbf{p}} \leq c \leq n \quad (10)$$

where b and c are integers.

All bounds are attained. The upper bound for $0 \leq c \leq m_{\mathbf{p}} - 1$ and the lower bound for $m_{\mathbf{p}} \leq c \leq n$ are attained if and only if $p_1 = p_2 = \dots = p_n = \mu$. In Equation (10) both bounds are attained. The lower bound is attained if and only if $p_1 = p_2 = \dots = p_n = \mu$ unless $b = 0$ and $c = n$.

From Hoeffding's inequalities we find the following bounds for the probabilities $\Pr_{\tau_k}(e \mid C_1, \dots, C_n)$.

Proposition 4 *Let $\rho = |\mathbf{p}(C_1, \dots, C_n)| \geq k, 1 \leq k \leq n$, then*

- $\Pr_{\tau_k}(e \mid C_1, \dots, C_n) \geq \sum_{i=k}^{\rho} B(i; \mu; \rho) \quad \text{if } k \leq m_{\mathbf{p}},$
- $\Pr_{\tau_k}(e \mid C_1, \dots, C_n) \leq \sum_{i=k}^{\rho} B(i; \mu; \rho) \quad \text{if } k \geq m_{\mathbf{p}} + 1.$

Proof: If $b = k$ and $c = \rho$ then Equation (10) becomes:

$$\sum_{i=k}^{\rho} B(i; \mu; \rho) \leq \sum_{i=k}^{\rho} B(i; \mathbf{p}) \leq 1 \quad \text{if } k \leq m_{\mathbf{p}}.$$

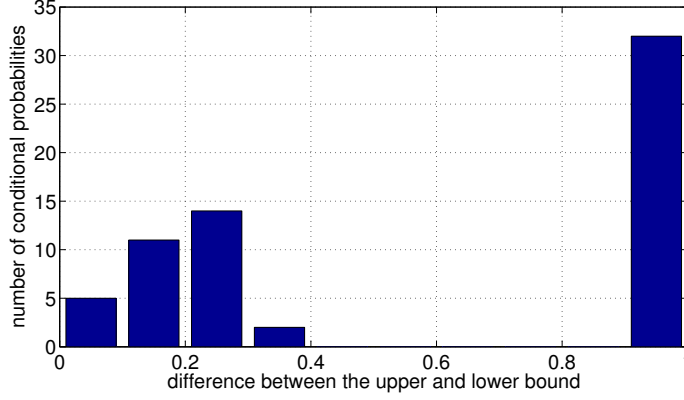


Figure 6: Hoeffding’s bounds for the conditional probabilities of the gastric non-Hodgkin lymphoma model.

If $c = k - 1$ then Equation (9) becomes:

$$\sum_{i=0}^{k-1} B(i; \mu; \rho) \leq \sum_{i=0}^{k-1} B(i; \mathbf{p}) \leq 1 \quad \text{if } m_{\mathbf{p}} \leq k - 1.$$

As $\sum_{i=0}^{k-1} B(i; \mathbf{p}) = 1 - \sum_{i=k}^{\rho} B(i; \mathbf{p})$, we can write the last equation as:

$$0 \leq \sum_{i=k}^{\rho} B(i; \mathbf{p}) \leq \sum_{i=k}^{\rho} B(i; \mu; \rho) \quad \text{if } m_{\mathbf{p}} + 1 \leq k.$$

Finally, using Proposition 1 we get

$$\begin{aligned} \Pr_{\tau_k}(e \mid C_1, \dots, C_n) &\geq \sum_{i=k}^{\rho} B(i; \mu; \rho) && \text{if } k \leq m_{\mathbf{p}}, \\ \Pr_{\tau_k}(e \mid C_1, \dots, C_n) &\leq \sum_{i=k}^{\rho} B(i; \mu; \rho) && \text{if } k \geq m_{\mathbf{p}} + 1. \end{aligned}$$

□

Hoeffding’s inequalities do not provide a bound when $m_{\mathbf{p}} < k < m_{\mathbf{p}} + 1$.

Since μ can be computed in linear time, the Hoeffding’s bounds are computable in linear time as well.

We have examined the tightness of the Hoeffding’s bounds for the conditional probabilities of the noisy threshold model of non-Hodgkin lymphoma. The obtained results, represented as the difference between the bounds, are shown in Figure 6. As the model we use is a small model, 32 probabilities cannot be determined as their $m_{\mathbf{p}}$ value falls into the interval $(1; 2)$. The other half of the conditional probabilities are bounded by intervals smaller than 0.4.

6.3.2 Percus and Percus bounds

Percus and Percus introduce a lower bound for the sum of the Poisson binomial probabilities $\sum_{i=0}^c B(i; \mathbf{p})$. The bound applies when the probability $B(0; \mathbf{p})$ is given [23]:

$$\sum_{i=0}^c B(i; \mathbf{p}) \geq B(0; \mathbf{p}) \sum_{i=0}^c \binom{n}{i} \left((B(0; \mathbf{p}))^{-\frac{1}{n}} - 1 \right)^i. \quad (11)$$

From Equation (11) we can also derive an upper bound for the sum of Poisson binomial probabilities $\sum_{i=0}^c B(i; \mathbf{p})$.

Let us define the probabilities of failures in Poisson trials $\mathbf{q} = (1 - p_1, \dots, 1 - p_n)$. Then the relation between the two Poisson binomial distributions is as follows:

$$\sum_{i=0}^c B(i; \mathbf{p}) = \sum_{i=n-c}^n B(i; \mathbf{q}) = 1 - \sum_{i=0}^{n-c-1} B(i; \mathbf{q}).$$

We can rewrite Equation (11) as follows:

$$\begin{aligned} \sum_{i=0}^c B(i; \mathbf{q}) &\geq B(0; \mathbf{q}) \sum_{i=0}^c \binom{n}{i} \left((B(0; \mathbf{q}))^{-\frac{1}{n}} - 1 \right)^i \\ &= B(n; \mathbf{p}) \sum_{i=0}^c \binom{n}{i} \left((B(n; \mathbf{p}))^{-\frac{1}{n}} - 1 \right)^i. \end{aligned}$$

Thus, we obtain the upper bound:

$$\sum_{i=0}^c B(i; \mathbf{p}) \leq 1 - B(n; \mathbf{p}) \sum_{i=0}^{n-c-1} \binom{n}{i} \left((B(n; \mathbf{p}))^{-\frac{1}{n}} - 1 \right)^i. \quad (12)$$

Using the Percus and Percus bounds we can find the upper and lower bounds of the conditional probabilities $\Pr_{\tau_k}(e \mid C_1, \dots, C_n)$.

Proposition 5 *Let $\rho = |\mathbf{p}(C_1, \dots, C_n)| \geq k, 1 \leq k \leq n, B(0; \mathbf{p}) = \prod_{i=1}^{\rho} (1 - p_i)$ and $B(\rho; \mathbf{p}) = \prod_{i=1}^{\rho} p_i$. Then,*

- $\Pr_{\tau_k}(e \mid C_1, \dots, C_n) \leq 1 - B(0; \mathbf{p}) \sum_{i=0}^{k-1} \binom{\rho}{i} \left((B(0; \mathbf{p}))^{-\frac{1}{\rho}} - 1 \right)^i,$
- $\Pr_{\tau_k}(e \mid C_1, \dots, C_n) \geq B(\rho; \mathbf{p}) \sum_{i=0}^{\rho-k} \binom{\rho}{i} \left((B(\rho; \mathbf{p}))^{-\frac{1}{\rho}} - 1 \right)^i.$

Proof: Let $c = k - 1$. Then using Proposition 1 Equations (11) and (12) become:

$$\begin{aligned} \Pr_{\tau_k}(e \mid C_1, \dots, C_n) &= 1 - \sum_{i=0}^{k-1} B(i; \mathbf{p}) \leq 1 - B(0; \mathbf{p}) \sum_{i=0}^{k-1} \binom{\rho}{i} \left((B(0; \mathbf{p}))^{-\frac{1}{\rho}} - 1 \right)^i, \\ \Pr_{\tau_k}(e \mid C_1, \dots, C_n) &= 1 - \sum_{i=0}^{k-1} B(i; \mathbf{p}) \geq B(\rho; \mathbf{p}) \sum_{i=0}^{\rho-k} \binom{\rho}{i} \left((B(\rho; \mathbf{p}))^{-\frac{1}{\rho}} - 1 \right)^i. \end{aligned}$$

□

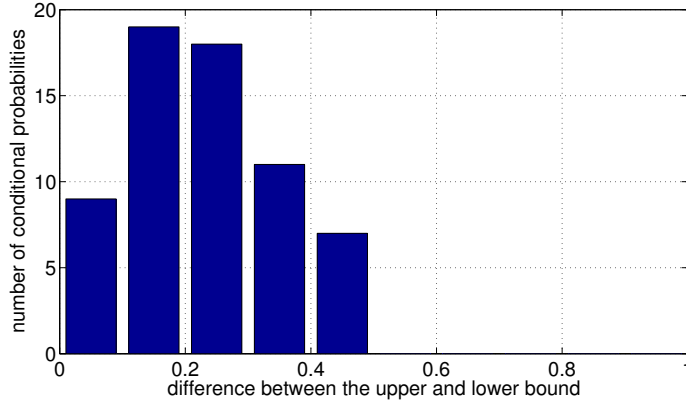


Figure 7: Percus and Percus bounds for the conditional probabilities of the gastric non-Hodgkin lymphoma model.

Since $B(0; \mathbf{p})$ and $B(\rho; \mathbf{p})$ can be computed in linear time, the Percus and Percus bounds can be computed in linear time as well.

We have examined the tightness of the Percus and Percus bounds for the conditional probabilities of the noisy threshold model of non-Hodgkin lymphoma. The obtained results are shown in Figure 7.

The best bounds for the probabilities $\Pr_{\tau_k}(e | C_1, \dots, C_n)$ can be achieved by combining Hoeffding's and Percus and Percus bounds. In our simulations, non-trivial Hoeffding's bounds were always at least as tight as the Percus and Percus bounds. See Figure 8 for the results.

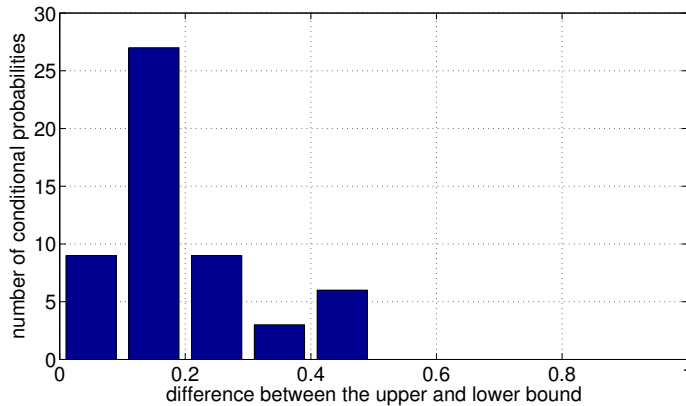


Figure 8: Combined Hoeffding's and Percus and Percus bounds for the conditional probabilities of the gastric non-Hodgkin lymphoma model.

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

It was shown that there is a close connection between the probability distribution of noisy threshold models and the Poisson binomial distribution. We have investigated what the well-studied properties of the Poisson binomial distribution mean in the context of these newly introduced models.

We presented recursive methods to compute the exact Poisson binomial probabilities as well as approximation and bounding techniques that can assess the probabilities in a linear number of operations. To illustrate the quality of the approximations and bounds we have used a noisy threshold model that represents a real-world medical problem. More experimental results can be found in a follow-up paper. The follow-up paper introduces the EM algorithm to learn the hidden parameters of the model and presents the classification results using the noisy threshold models that show their competitive performance in comparison with the noisy OR classifier as well as widely used classifiers such as naive Bayes, logistic regression and decision trees.

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 models where the interaction function is defined in terms of any symmetric Boolean function. 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.

References

- [1] F.C. Arnett, S.M. Edworthy, D.A. Bloch, D.J. McShane, J.F. Fries, N.S. Cooper, et al, *The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis*. Arthritis Rheum, 31:315-324, 1988.
- [2] 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.
- [3] G. Bennett, *Probability inequalities for the sum of independent random variables*. Journal of the American Statistical Association, 57:33-45, 1962.
- [4] X.H. Chen, A.P. Dempster and J.S. Liu, *Weighted finite population sampling to maximize entropy*. Biometrika, 81:457-469, 1994.
- [5] J. Darroch, *On the distribution of the number of successes in independent trials*. The Annals of Mathematical Statistics, 35:1317-1321, 1964.
- [6] F.J. Díez, *Parameter adjustment in Bayes networks. The generalized noisy OR-gate*. Proceedings of the Ninth Conference on Uncertainty in Artificial Intelligence, 99-105, 1993.

- [7] A.W.P. Edwards, *The Meaning of Binomial Distribution*. Nature 186, 1074, 1960.
- [8] H.B. Enderton, *A Mathematical Introduction to Logic*. Academic Press, San Diego, 1972.
- [9] W. Feller, *An Introduction to Probability Theory and Its Applications*. John Wiley, 1968.
- [10] J.L. Gastwirth, *A probability model of a pyramid scheme*. The American Statistician, 31:79-82, 1977.
- [11] D. Heckerman and J.S. Breese, A new look at causal independence, *Proc. UAI-94*, pp. 286–292, 1994.
- [12] M. Henrion, *Some practical issues in constructing belief networks*. Uncertainty in Artificial Intelligence, 3:161–173, 1989.
- [13] W. Hoeffding, *On the distribution of the number of successes in independent trials*. The Annals of Mathematical Statistics, 27:713-721, 1956.
- [14] S. Howard, *Discussion on Professor Cox's paper*. Journal of the Royal Statistical Society, Series B, 34:210-211, 1972.
- [15] S. Janson, *Large deviation inequalities for sums of indicator variables*. Technical Report, 1994.
- [16] 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.
- [17] R. Jurgelenaite and P.J.F. Lucas, *Exploiting Causal Independence in Large Bayesian Networks*. Knowledge-Based Systems Journal, 18:153-162, 2005.
- [18] H.J. Kappen, J.P. Neijt, *Promedas, a probabilistic decision support system for medical diagnosis*. Technical report, SNN - UMCU, 2002.
- [19] L. Le Cam, *An approximation theorem for the Poisson binomial distribution*. Pacific Journal of Mathematics, 10: 1181-1197, 1960.
- [20] P.J.F. Lucas, H. Boot and B. Taal, *Computer-based Decision Support in Management of Primary Gastric non-Hodgkin Lymphoma*. Methods of Information in Medicine, 37: 206-219, 1998.
- [21] P.J.F. Lucas, *Bayesian network modelling through qualitative patterns*. Artificial Intelligence, 163: 233-263, 2005.
- [22] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kauffman Publishers, 1988.
- [23] O.E. Percus and J.K. Percus, *Probability bounds on the sum of independent nonidentically distributed binomial random variables*. SIAM Journal on Applied Mathematics, 45:621-640, 1985.
- [24] J. Pitman, *Probabilistic bounds on the coefficients of polynomials with only real zeros*. Journal of Combinatorial Theory, Series A, 77:279-303, 1997.

- [25] B. Roos, *Binomial approximation to the Poisson binomial distribution: the Krawtchouk expansion*. Theory of Probability and Its Applications, 45:258-272, 2001.
- [26] M.A. Shwe, B. Middleton, D.E. Heckerman, M. Henrion, E.J. Horvitz, H.P. Lehmann and G.F. Cooper, *Probabilistic diagnosis using a reformulation of the INTERNIST-1/QMR knowledge base, I – The probabilistic model and inference algorithms*. Methods of Information in Medicine, 30: 241-255, 1991.
- [27] S. Srinivas, *A generalization of the noisy-or model*. Proceedings of the Ninth Conference on Uncertainty in Artificial Intelligence, 208-215, 1993.
- [28] I. Wegener, *The Complexity of Boolean Functions*. John Wiley & Sons, New York, 1987.
- [29] S. Visscher, Peter J.F. Lucas, M. Bonten and K. Schurink, *Improving the Therapeutic Performance of a Medical Bayesian Network Using Noisy Threshold Models*. Proceedings of ISBMDA 2005, the 6th International Symposium on Biological and Medical Data Analysis, 161-172, 2005.
- [30] J. Vomlel, *Decomposition of Probability Tables Representing Boolean Functions*. Proceedings of the Eighth Czech-Japan Seminar on Data Analysis and Decision Making under Uncertainty, 18 - 21, 2005.
- [31] N.L. Zhang and D. Poole, *Exploiting causal independence in Bayesian networks inference*. Journal of Artificial Intelligence Research, 5:301-328, 1996.