Chapter 6

Converting Graphic Relationships into Conditional Probabilities in Bayesian Network

Loc Nguyen

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.70057

Abstract

Bayesian network (BN) is a powerful mathematical tool for prediction and diagnosis applications. A large Bayesian network can constitute many simple networks, which in turn are constructed from simple graphs. A simple graph consists of one child node and many parent nodes. The strength of each relationship between a child node and a parent node is quantified by a weight and all relationships share the same semantics such as prerequisite, diagnostic, and aggregation. The research focuses on converting graphic relationships into conditional probabilities in order to construct a simple Bayesian network from a graph. Diagnostic relationship is the main research object, in which sufficient diagnostic proposition is proposed for validating diagnostic relationship. Relationship conversion is adhered to logic gates such as AND, OR, and XOR, which are essential features of the research.

Keywords: diagnostic relationship, Bayesian network, transformation coefficient

1. Introduction

Bayesian network (BN) is a directed acyclic graph (DAG) consists of a set of nodes and a set of arcs. Each node is a random variable. Each arc represents a relationship between two nodes. The strength of a relationship in a graph can be quantified by a number called *weight*. There are some important relationships such as prerequisite, diagnostic, and aggregation. The difference between BN and normal graph is that the strength of every relationship in BN is represented by a conditional probability table (CPT) whose entries are conditional probabilities of a child node given parent nodes. There are two main approaches to construct a BN, which are as follows

- The first approach aims to learn BN from training data by learning machine algorithms.
- The second approach is that experts define some graph patterns according to specific relationships and then, BN is constructed based on such patterns along with determined CPTs.

This research focuses on the second approach in which relationships are converted into CPTs. Essentially, relationship conversion aims to determine conditional probabilities based on weights and meanings of relationships. We will have different ways to convert graphic weights into CPTs for different relationships. It is impossible to convert all relationships but some of them such as diagnostic, aggregation, and prerequisite are mandatory ones that we must specify as computable CPTs of BN. Especially, these relationships are adhered to logic X-gates [1] such as AND-gate, OR-gate, and SIGMA-gate. The X-gate inference in this research is derived and inspired from noisy OR-gate described in the book "Learning Bayesian Networks" Neapolitan ([2], pp. 157–159). Díez and Druzdzel [3] also researched OR/MAX, AND/MIN, and noisy XOR inferences but they focused on canonical models, deterministic models, and ICI models whereas I focused on logic gate and graphic relationships. So, their research is different from mine but we share the same result that is AND-gate model. In general, my research focuses on applied probability adhered to Bayesian network, logic gates, and Bayesian user modeling [4]. The scientific results are shared with Millán and Pérez-de-la-Cruz [4].

Factor graph [5] represents factorization of a global function into many partial functions. If joint distribution of BN is considered as the global function and CPTs are considered as partial functions, the sumproduct algorithm [6] of factor graph is applied into calculating posterior probabilities of variables in BN. Pearl's propagation algorithm [7] is very successful in BN inference. The application of factor graph into BN is only realized if all CPT (s) of BN are already determined whereas this research focuses on defining such CPTs firstly. I did not use factor graph for constructing BN. The concept "X-gate inference" only implies how to convert simple graph into BN. However, the arrange sum with a fixed variable mentioned in this research is the "*not-sum*" ([6], p. 499) of factor graph. Essentially, X-gate probability shown in Eq. (10) is as same as λ message in the Pearl's algorithm ([6], p. 518) but I use the most basic way to prove the X-gate probability.

As default, the research is applied in learning context in which BN is used to assess students' knowledge. Evidences are tests, exams, exercises, etc. and hypotheses are learning concepts, knowledge items, etc. Note that diagnostic relationship is very important to Bayesian evaluation in learning context because it is used to evaluate student's mastery of concepts (knowledge items) over entire BN. Now, we start relationship conversion with a research on diagnostic relationship in the next section.

2. Diagnostic relationship

In some opinions like mine, the diagnostic relationship should be from hypothesis to evidence. For example, disease is hypothesis and symptom is evidence. The symptom must be conditionally dependent on disease. Given a symptom, calculating the posterior probability of

http://dx.doi.org/10.5772/intechopen.70057

disease is essentially to diagnose likelihood of such disease ([8], p. 1666). Inversely, the arc from evidence to hypothesis implies prediction where evidence and hypothesis represent observation and event, respectively. Given an observation, calculating the posterior probability of the event is essentially to predict/assert such event ([8], p. 1666). **Figure 1** shows diagnosis and prediction.

The weight w of the relationship between X and D is 1. **Figure 1** depicts simplest graph with two random variables. We need to convert diagnostic relationship into conditional probabilities in order to construct a simplest BN from the simplest graph. Note that hypothesis is binary but evidence can be numerical. In learning context, evidence D can be test, exam, exercise, etc. The conditional probability of D given X (likelihood function) is $P(D|X)$. The posterior probability of X is $P(X|D)$, which is used to evaluate student's mastery over concept (hypothesis) X given evidence D. Eq. (1) specifies CPT of D when D is binary (0 and 1)

$$P(D|X) = \begin{cases} D \text{ if } X = 1 \\ 1 - D \text{ if } X = 0 \end{cases} \tag{1}$$

Eq (1) is our first relationship conversion. It implies

$$P(D|X = 0) + P(D|X = 1) = D + 1 - D = 1$$

Evidence D can be used to diagnose hypothesis X if the so-called *sufficient diagnostic proposition* is satisfied, as seen in **Table 1**.

The concept of sufficient evidence is borrowed from the concept of sufficient statistics and it is inspired from equivalence of variables T and T' in the research ([4], pp. 292-295). The proposition can be restated that evidence D is only used to assess hypotheses if it is sufficient evidence. As a convention, the proposition is called *diagnostic condition* and hypotheses have uniform distribution. The assumption of hypothetic uniform distribution ($P(X = 1) = P(X = 0)$) implies that we cannot assert whether or not given hypothesis is true before we observe its evidence.

In learning context, D can be totally used to assess student's mastery of X if diagnostic condition is satisfied. Derived from such condition, Eq. (2) specifies transformation coefficient k given uniform distribution of X.

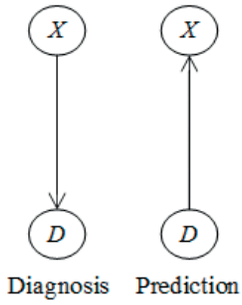

Figure 1. Diagnosis and prediction with hypothesis X and evidence D.

D is equivalent to *X* in diagnostic relationship if $P(X|D) = kP(D|X)$ given uniform distribution of *X* and the *transformation coefficient k* is independent from *D*. In other words, *k* is constant with regards to *D* and so *D* is called *sufficient evidence*.

Table 1. Sufficient diagnostic proposition.

$$k = \frac{P(X|D)}{P(D|X)} \tag{2}$$

We need to prove that Eq. (1) satisfies diagnostic condition. Suppose the prior probability of *X* is uniform.

$$P(X = 0) = P(X = 1)$$

we have

$$P(X|D) = \frac{P(D|X)P(X)}{P(D)} = \frac{P(D|X)P(X)}{P(D|X = 0)P(X = 0) + P(D|X = 1)P(X = 1)}$$

$$(\text{due to Bayes'rule})$$

$$= \frac{P(D|X)P(X)}{P(X)\Big(P(D|X = 0) + P(D|X = 1)\Big)}$$

$$\Big(\text{due to } P(X = 0) = P(X = 1)\Big)$$

$$= \frac{P(D|X)}{P(D|X = 0) + P(D|X = 1)} = 1 * P(D|X)$$

$$\Big(\text{due to } P(D|X = 0) + P(D|X = 1) = 1\Big) \blacksquare$$

It is easy to infer that the transformation coefficient *k* is 1, if *D* is binary. In practice, evidence *D* is often a test whose grade ranges within an interval {0, 1, 2,…, *η*}. Eq. (3) specifies CPT of *D* in this case

$$P(D|X) = \begin{cases} \dfrac{D}{S} \text{if } X = 1 \\ \dfrac{\eta}{S} - \dfrac{D}{S} \text{if } X = 0 \end{cases} \tag{3}$$

Where

$$D \in \{0, 1, 2, \ldots, \eta\}$$

$$S = \sum_{D=0}^{n} D = \frac{\eta(\eta + 1)}{2}$$

http://dx.doi.org/10.5772/intechopen.70057

As a convention, $P(D|X) = 0, \forall D \notin \{0, 1, 2, \ldots, \eta\}$. Eq. (3) implies that if student has mastered concept ($X = 1$), the probability that she/he completes the exercise/test D is proportional to her/his mark on D $\left(P(D|X) = \frac{D}{S}\right)$. We also have

$$P(D|X = 0) + P(D|X = 1) = \frac{D}{S} + \frac{\eta - D}{S} = \frac{\eta}{S} = \frac{2}{(\eta + 1)}$$

$$\sum_{D=0}^{\eta} P(D|X = 1) = \sum_{D=0}^{\eta} \frac{D}{S} = \frac{\sum_{D=0}^{\eta} D}{S} = \frac{S}{S} = 1$$

$$\sum_{D=0}^{\eta} P(D|X = 0) = \sum_{D=0}^{\eta} \frac{\eta - D}{S} = \frac{\sum_{D=0}^{\eta} (\eta - D)}{S} = \frac{\sum_{D=0}^{\eta} \eta - \sum_{D=0}^{\eta} D}{S} = \frac{\eta(\eta + 1) - S}{S} = \frac{2S - S}{S} = 1$$

We need to prove that Eq. (3) satisfies diagnostic condition. Suppose the prior probability of X is uniform.

$$P(X = 0) = P(X = 1)$$

The assumption of prior uniform distribution of X implies that we do not determine if student has mastered X yet. Similarly, we have

$$P(X|D) = \frac{P(D|X)P(X)}{P(D)} = \frac{P(D|X)}{P(D|X = 0) + P(D|X = 1)} = \frac{\eta + 1}{2} P(D|X) \blacksquare$$

So, the transformation coefficient k is $\frac{\eta+1}{2}$ if D ranges in {0, 1, 2,…, η}.

In the most general case, discrete evidence D ranges within an arbitrary integer interval $\{a, a + 1, a + 2, \ldots, b\}$. In other words, D is bounded integer variable whose lower bound and upper bound are a and b, respectively. Eq. (4) specifies CPT of D, where $D \in \{a, a + 1, a + 2, \ldots, b\}$.

$$P(D|X) = \begin{cases} \frac{D}{S} \text{ if } X = 1 \\ \frac{b + a}{S} - \frac{D}{S} \text{ if } X = 0 \end{cases} \tag{4}$$

Where

$$D \in \{a, a + 1, a + 2, \ldots, b\}$$

$$S = a + (a + 1) + (a + 2) + \ldots + b = \frac{(b + a)(b - a + 1)}{2}$$

Note, $P(D|X) = 0, \forall D \notin \{a, a + 1, a + 2, \ldots, b\}$. According to the diagnostic condition, we need to prove the equality $P(X|D) = kP(D|X)$, where

$$k = \frac{b - a + 1}{2}$$

Similarly, we have

$$P(X|D) = \frac{P(D|X)P(X)}{P(D)} = \frac{P(D|X)}{P(D|X=0) + P(D|X=1)} = \frac{b-a+1}{2} P(D|X) \blacksquare$$

If evidence D is continuous in the real interval $[a, b]$ with note that a and b are real numbers, Eq. (5) specifies probability density function (PDF) of continuous evidence $D \in [a, b]$. The PDF $p(D|X)$ replaces CPT in case of continuous random variable.

$$p(D|X) = \begin{cases} \dfrac{2D}{b^2 - a^2} \text{ if } X = 1 \\ \dfrac{2}{b-a} - \dfrac{2D}{b^2 - a^2} \text{ if } X = 0 \end{cases}$$

where

$$D \in [a, b] \text{ where } a \text{ and } b \text{ are real numbers}$$

$$S = \int_a^b D \mathrm{d}D = \frac{b^2 - a^2}{2} \tag{5}$$

As a convention, $[a, b]$ is called domain of continuous evidence, which can be replaced by open or half-open intervals such as (a, b), $(a, b]$, and $[a, b)$. Of course we have $p(D|X) = 0$, $\forall D \notin [a, b]$. In learning context, evidence D is often a test whose grade ranges within real interval $[a, b]$.

Functions $p(D \mid X = 1)$ and $p(D \mid X = 0)$ are valid PDFs due to

$$\int_D p(D|X=1)\mathrm{d}D = \int_a^b \frac{2D}{b^2 - a^2} \mathrm{d}D = \frac{1}{b^2 - a^2} \int_a^b 2D \mathrm{d}D = 1$$

$$\int_D p(D|X=0)\mathrm{d}D = \frac{2}{b-a} \int_a^b \mathrm{d}D - \frac{1}{b^2 - a^2} \int_a^b 2D \mathrm{d}D = 1.$$

According to the diagnostic condition, we need to prove the equality

$$P(X|D) = kp(D|X)$$

where,

$$k = \frac{b-a}{2}$$

When D is continuous, its probability is calculated in ε-vicinity where ε is very small number. As usual, ε is bias if D is measure values produced from equipment. The probability of D given X, where $D + \varepsilon \in$ [a, b] and $D - \varepsilon \in$ [a, b] is

http://dx.doi.org/10.5772/intechopen.70057

$$P(D|X) = \int_{D-\varepsilon}^{D+\varepsilon} p(D|X)\mathrm{d}D = \begin{cases} \int_{D-\varepsilon}^{D+\varepsilon} \frac{2D}{b^2-a^2}\mathrm{d}D \text{ if } X=1 \\ \int_{D-\varepsilon}^{D+\varepsilon} \left(\frac{2}{b-a} - \frac{2D}{b^2-a^2}\right)\mathrm{d}D \text{ if } X=0 \end{cases}$$

$$= \begin{cases} \frac{4\varepsilon D}{b^2-a^2} \text{if } X=1 \\ \frac{4\varepsilon}{b-a} - \frac{4\varepsilon D}{b^2-a^2} \text{if } X=0 \end{cases} = 2\varepsilon p(D|X)$$

In fact, we have

$$P(X|D) = \frac{P(D|X)P(X)}{P(D|X=0)P(X=0)+P(D|X=1)P(X=1)} = \frac{P(D|X)}{P(D|X=0)+P(D|X=1)}$$

$$\left(\text{due to Bayes}^{'}\text{rule and the assumption } P(X=0)=P(X=1)\right)$$

$$= \frac{b-a}{4\varepsilon}P(D|X) = kp(D|X) \blacksquare$$

In general, Eq. (6) summarizes CPT of evidence of single diagnostic relationship.

$$P(D|X) = \begin{cases} \frac{D}{S} \text{ if } X=1 \\ \frac{M}{S} - \frac{D}{S} \text{ if } X=0 \end{cases}$$

$$k = \frac{N}{2}$$

Where,

$$N = \begin{cases} 2 \text{ if } D \in \{0,1\} \\ \eta+1 \text{ if } D \in \{0,1,2,\ldots,\eta\} \\ b-a+1 \text{ if } D \in \{a,a+1,a+2,\ldots,b\} \\ b-a \text{ if } D \text{ continuous and } D \in [a,b] \end{cases}$$

$$M = \begin{cases} 1 \text{ if } D \in \{0,1\} \\ \eta \text{ if } D \in \{0,1,2,\ldots,\eta\} \\ b+a \text{ if } D \in \{a,a+1,a+2,\ldots,b\} \\ b+a \text{ if } D \text{ continuous and } D \in [a,b] \end{cases}$$

$$S = \sum_D D = \frac{NM}{2} = \begin{cases} 1 \text{ if } D \in \{0,1\} \\ \frac{\eta(\eta+1)}{2} \text{ if } D \in \{0,1,2,\ldots,\eta\} \\ \frac{(b+a)(b-a+1)}{2} \text{ if } D \in \{a,a+1,a+2,\ldots,b\} \\ \frac{b^2-a^2}{2} \text{ if } D \text{ continuous and } D \in [a,b] \end{cases} \quad (6)$$

In general, if the conditional probability $P(D|X)$ is specified by Eq. (6), the diagnostic condition will be satisfied. Note that the CPT $P(D|X)$ is the PDF $p(D|X)$ in case of continuous evidence. The diagnostic relationship will be extended with more than one hypothesis. The next section will mention how to determine CPTs of a simple graph with one child node and many parent nodes based on X-gate inferences.

3. X-gate inferences

Given a simple graph consisting of one child variable Y and n parent variables X_i, as shown in **Figure 2**, each relationship from X_i to Y is quantified by normalized weight w_i where $0 \le w_i \le 1$. A large graph is an integration of many simple graphs. **Figure 2** shows the DAG of a simple BN. As aforementioned, the essence of constructing simple BN is to convert graphic relationships of simple graph into CPTs of simple BN.

Child variable Y is called target and parent variables X_is are called sources. Especially, these relationships are adhered to X-gates such as AND-gate, OR-gate, and SIGMA-gate. These gates are originated from logic gate [1]. For instance, AND-gate and OR-gate represent prerequisite relationship. SIGMA-gate represents aggregation relationship. Therefore, relationship conversion is to determined X-gate inference. The simple graph shown in **Figure 2** is also called X-gate graph or X-gate network. Please distinguish the letter "X" in the term "X-gate inference" which implies logic operators (AND, OR, XOR, etc.) from the "variable X".

All variables are binary and they represent events. The probability $P(X)$ indicates event X occurs. Thus, $P(X)$ implicates $P(X = 1)$ and $P(\text{not}(X))$ implicates $P(X = 0)$. Eq. (7) specifies the simple NOT-gate inference.

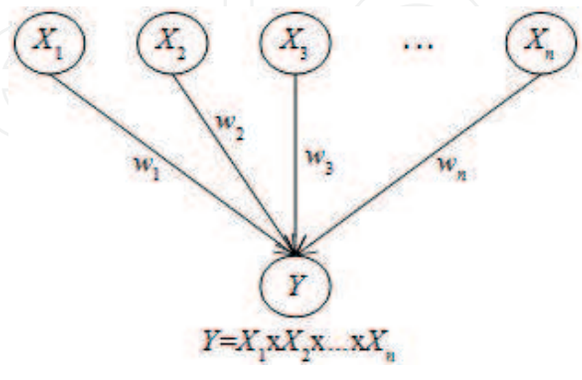

Figure 2. Simple graph or simple network.

http://dx.doi.org/10.5772/intechopen.70057

$$P\Big(\text{not}(X)\Big) = P(\overline{X}) = P(X=0) = 1 - P(X=1) = 1 - P(X)$$
$$P\Big(\text{not}\Big(\text{not}(X)\Big)\Big) = P(X) \tag{7}$$

X-gate inference is based on three assumptions mentioned in Ref. ([2], p. 157), which are as follows

- *X-gate inhibition*: Given a relationship from source X_i to target Y, there is a factor I_i that inhibits X_i from being integrated into Y. Factor I_i is called inhibition of X_i. That the inhibition I_i is turned off is prerequisite of X_i integrated into Y.
- *Inhibition independence*: Inhibitions are mutually independent. For example, inhibition I_1 of X_1 is independent from inhibition I_2 of X_2.
- *Accountability*: X-gate network is established by accountable variables A_i for X_i and I_i. Each X-gate inference owns particular combination of A_is.

Figure 3 shows the extended X-gate network with accountable variables A_is ([2], p. 158).

The strength of each relationship from source X_i to target Y is quantified by a weight $0 \le w_i \le 1$. According to the assumption of inhibition, probability of $I_i = OFF$ is p_i, which is set to be the weight w_i.

$$p_i = w_i$$

If notation w_i is used, we focus on the strength of relationship. If notation p_i is used, we focus on probability of *OFF* inhibition. In probabilistic inference, p_i is also prior probability of $X_i = 1$. However, we will assume each X_i has uniform distribution later on. Eq. (8) specifies probabilities of inhibitions I_is and accountable variables A_is.

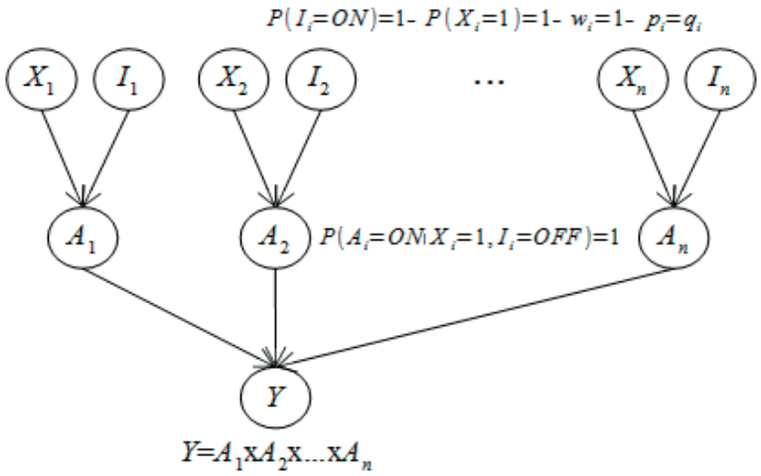

Figure 3. Extended X-gate network with accountable variables A_is.

$$\begin{gathered}
P(I_i = OFF) = p_i = w_i \\
P(I_i = ON) = 1 - p_i = 1 - w_i \\
P(A_i = ON|X_i = 1, I_i = OFF) = 1 \\
P(A_i = ON|X_i = 1, I_i = ON) = 0 \\
P(A_i = ON|X_i = 0, I_i = OFF) = 0 \\
P(A_i = ON|X_i = 0, I_i = ON) = 0 \\
P(A_i = OFF|X_i = 1, I_i = OFF) = 0 \\
P(A_i = OFF|X_i = 1, I_i = ON) = 1 \\
P(A_i = OFF|X_i = 0, I_i = OFF) = 1 \\
P(A_i = OFF|X_i = 0, I_i = ON) = 1
\end{gathered} \tag{8}$$

According to Eq. (8), given probability $P(A_i$=$ON \mid X_i$=1, I_i=$OFF)$, it is assured 100% confident that accountable variables A_i is turned on if source X_i is 1 and inhibition I_i is turned off. Eq. (9) specifies conditional probability of accountable variables A_i (s) given X_i (s), which is corollary of Eq. (8).

$$\begin{gathered}
P(A_i = ON|X_i = 1) = p_i = w_i \\
P(A_i = ON|X_i = 0) = 0 \\
P(A_i = OFF|X_i = 1) = 1 - p_i = 1 - w_i \\
P(A_i = OFF|X_i = 0) = 1
\end{gathered} \tag{9}$$

Appendix A1 is the proof of Eq. (9). As a definition, the set of all X_is is complete if and only if

$$P(X_1 \cup X_2 \cup \cdots \cup X_n) = P(\Omega) = \sum_{i=1}^{n} w_i = 1$$

The set of all X_is is mutually exclusive if and only if

$$X_i \cap X_j = \varnothing, \forall i \neq j$$

For each X_i, there is only one A_i and vice versa, which establishes a bijection between X_is and A_is. Obviously, the fact that the set of all X_is is complete is equivalent to the fact that the set of all A_i (s) is complete. We will prove by contradiction that "the fact that the set of all X_i (s) is mutually exclusive is equivalent to the fact that the set of all A_i (s) is mutually exclusive." Suppose $X_i \cap X_j = \varnothing, \forall i \neq j$ but $\exists i \neq j$: $A_i \cap A_j = B \neq \varnothing$. Let $B^{-1} \neq \varnothing$ be preimage of B. Due to $B \subseteq A_i$ and $B \subseteq A_j$, we have $B^{-1} \subseteq X_i$ and $B^{-1} \subseteq X_j$, which causes that $X_i \cap X_j = B^{-1} \neq \varnothing$. There is a contradiction and so we have

$$X_i \cap X_j = \varnothing, \forall i \neq j \Rightarrow A_i \cap A_j = \varnothing, \forall i \neq j$$

By similar proof, we have

http://dx.doi.org/10.5772/intechopen.70057

$$A_i \cap A_j = \varnothing, \forall i \neq j \Rightarrow X_i \cap X_j = \varnothing, \forall i \neq j \blacksquare$$

The extended X-gate network shown in **Figure 3** is interpretation of simple network shown in **Figure 2**. Specifying CPT of the simple network is to determine the conditional probability $P(Y = 1 \mid X_1, X_2, \ldots, X_n)$ based on extended X-gate network. The X-gate inference is represented by such probability $P(Y = 1 \mid X_1, X_2, \ldots, X_n)$ specified by Eq. (10) ([2], p. 159).

$$P(Y|X_1, X_2, \ldots, X_n) = \sum\nolimits_{A_1, A_2, \ldots, A_n} P(Y|A_1, A_2, \ldots, A_n) \prod\nolimits_{i=1}^{n} P(A_i|X_i) \tag{10}$$

Appendix A2 is the proof of Eq. (10). It is necessary to make some mathematical notations because Eq. (10) is complicated, which is relevant to arrangements of X_i (s). Given the set $\Omega = \{X_1, X_2, \ldots, X_n\}$ where all variables are binary, **Table 2** specifies binary arrangements of Ω.

Given $\Omega = \{X_1, X_2, \ldots, X_n\}$ where $|\Omega| = n$ is cardinality of Ω.
Let $a(\Omega)$ be an arrangement of Ω which is a set of n instances $\{X_1{=}x_1, X_2{=}x_2, \ldots, X_n{=}x_n\}$ where x_i is 1 or 0. The number of all $a(\Omega)$ is $2^{|\Omega|}$. For instance, given $\Omega = \{X_1, X_2\}$, there are 2^2=4 arrangements as follows:

$$\begin{aligned} a(\Omega) &= \{X_1 = 1, X_2 = 1\}, a(\Omega) = \{X_1 = 1, X_2 = 0\}, a(\Omega) = \{X_1 = 0, X_2 = 1\}, a(\Omega) \\ &= \{X_1 = 0, X_2 = 0\}. \end{aligned}$$

Let $a(\Omega{:}\{X_i\})$ be the arrangement of Ω with fixed X_i. The number of all $a(\Omega{:}\{X_i\})$ is $2^{|\Omega|-1}$. Similarly, for instance, $a(\Omega{:}\{X_1, X_2, X_3\})$ is an arrangement of Ω with fixed X_1, X_2, X_3. The number of all $a(\Omega{:}\{X_1, X_2, X_3\})$ is $2^{|\Omega|-3}$.
Let $c(\Omega)$ and $c(\Omega{:}\{X_i\})$ be the number of arrangements $a(\Omega)$ and $a(\Omega{:}\{X_i\})$, respectively. Such $c(\Omega)$ and $c(\Omega{:}\{X_i\})$ are called arrangement counters. As usual, counters $c(\Omega)$ and $c(\Omega{:}\{X_i\})$ are equal to $2^{|\Omega|}$ and $2^{|\Omega|-1}$, respectively but they will vary according to specific cases.
Let $\sum_a F\Big(a(\Omega)\Big)$ and $\prod_a F\Big(a(\Omega)\Big)$ denote sum and product of values generated from function F acting on every $a(\Omega)$. The number of arrangements on which F acts is $c(\Omega)$.

Let x denote the X-gate operator, for instance, x = $\odot$ for AND-gate, x = $\oplus$ for OR-gate, x = not $\odot$ for NAND-gate, x = not $\oplus$ for NOR-gate, x = $\otimes$ for XOR-gate, x = not $\otimes$ for XNOR-gate, x = $\uplus$ for U-gate, x = $+$ for SIGMA-gate. Given an x-operator, let $s(\Omega{:}\{X_i\})$ and $s(\Omega)$ be sum of all $P(X_1 \mathrm{x} X_2 \mathrm{x} \ldots \mathrm{x} X_n)$ through every arrangement of Ω with and without fixed X_i, respectively.

$$\begin{aligned} s(\Omega) &= \sum_a P\Big(X_1 \mathrm{x} X_2 \mathrm{x} \ldots \mathrm{x} X_n | a(\Omega)\Big) = \sum_a P\Big(Y = 1 | a(\Omega)\Big) \\ s(\Omega : \{X_i\}) &= \sum_a P\Big(X_1 \mathrm{x} X_2 \mathrm{x} \ldots \mathrm{x} X_n | a(\Omega : \{X_i\})\Big) = \sum_a P\Big(Y = 1 | a(\Omega : \{X_i\})\Big) \end{aligned}$$

For example, $s(\Omega)$ and $s(\Omega{:}\{X_i\})$ for OR-gate are:

$$\begin{aligned} s(\Omega) &= \sum_a P\Big(X_1 \oplus X_2 \oplus \ldots \oplus X_n | a(\Omega)\Big) \\ s(\Omega : \{X_i\}) &= \sum_a P\Big(X_1 \oplus X_2 \oplus \ldots \oplus X_n | a(\Omega : \{X_i\})\Big) \end{aligned}$$

Such $s(\Omega)$ and $s(\Omega{:}\{X_i\})$ are called arrangement sum. They are acting function F.
Note that Ω can be any set of binary variables.

Table 2. Binary arrangements.

It is not easy to produce all binary arrangements of Ω. **Table 3** shows a code snippet written by Java programming language for producing such all arrangements.

Each element of the list "*arrangements*" is a binary arrangement $a(\Omega)$ presented by an array of bits (0 and 1). The method "*create*(*int*[] *a, int i*)" which is recursive method, is the main one that generates arrangements. The method call "*ArrangementGenerator.parse*(2, *n*)" will list all possible binary arrangements.

Eq. (11) specifies the connection between $s(\Omega{:}\{X_i = 1\})$ and $s(\Omega{:}\{X_i = 0\})$, between $c(\Omega{:}\{X_i = 1\})$ and $c(\Omega{:}\{X_i = 0\})$.

$$\begin{aligned} s(\Omega : \{X_i = 1\}) + s(\Omega : \{X_i = 0\}) &= s(\Omega) \\ c(\Omega : \{X_i = 1\}) + c(\Omega : \{X_i = 0\}) &= c(\Omega) \end{aligned} \tag{11}$$

It is easy to draw Eq. (11) when the set of all arrangements $a(\Omega{:}\{X_i = 1)$ is complement of the set of all arrangements $a(\Omega{:}\{X_i = 0)$.

Let K be a set of X_is whose values are 1 and let L be a set of X_is whose values are 0. K and L are mutually complementary. Eq. (12) determines sets K and L.

$$\begin{cases} K = \{i{:}\ X_i = 1\} \\ L = \{i{:}\ X_i = 0\} \\ K \cap L = \varnothing \\ K \cup L = \{1, 2, \ldots, n\} \end{cases} \tag{12}$$

The AND-gate inference represents prerequisite relationship satisfying AND-gate condition specified by Eq. (13).

$$P(Y = 1 | A_i = OFF \text{ for some } i) = 0 \tag{13}$$

From Eq. (10), we have

$$\begin{aligned} P(Y = 1 | X_1, X_2, \ldots, X_n) &= \sum_{A_1, A_2, \ldots, A_n} P(Y = 1 | A_1, A_2, \ldots, A_n) \prod_{i=1}^{n} P(A_i | X_i) \\ &= \prod_{i=1}^{n} P(A_i = ON | X_i) \\ &\left(\text{Due to } P(Y = 1 | A_i = OFF \text{ for some } i) = 0\right) \\ &= \left(\prod_{i \in K} P(A_i = ON | X_i = 1)\right)\left(\prod_{i \notin K} P(A_i = ON | X_i = 0)\right) \\ &= \left(\prod_{i \in K} p_i\right)\left(\prod_{i \notin K} 0\right) = \begin{cases} \prod_{i=1}^{n} p_i \text{ if all } X_i(s) \text{ are } 1 \\ 0 \text{ if there exists at least one } X_i = 0 \end{cases} \end{aligned}$$

(Due to Eq. (9))

```
public class ArrangementGenerator {
private ArrayList<int[]> arrangements;
private int n;
private int r;

private ArrangementGenerator(int n, int r) {
        this.n = n;
        this.r = r;
        this.arrangements = new ArrayList();
}
private void create(int[] a, int i) {
       for(int j = 0; j < n; j++) {
             a[i] = j;
            if(i < r - 1)
                  create(a, i + 1);
            else if(i == r -1) {
                  int[] b = new int[a.length];
                  for(int k = 0; k < a.length; k++) b[k] = a[k];
                  arrangements.add(b);
            }
       }
}
public int[] get(int i) {
       return arrangements.get(i);
}
public long size() {
       return arrangements.size();
}
public static ArrangementGenerator parse(int n, int r) {
       ArrangementGenerator arr =
             new ArrangementGenerator(n, r);
       int[] a = new int[r];
       for(int i=0; i<r; i++) a[i] = -1;
       arr.create(a, 0);
       return arr;
}
}
```

Table 3. Code snippet generating all binary arrangements.

In general, Eq. (14) specifies AND-gate inference.

$$P(X_1 \odot X_2 \odot \ldots \odot X_n) = P(Y = 1|X_1, X_2, \ldots, X_n) = \begin{cases} \prod_{i=1}^{n} p_i \text{ if all } X_i(s) \text{ are } 1 \\ 0 \text{ if there exists at least one } X_i = 0 \end{cases}$$
$$P(Y = 0|X_1, X_2, \ldots, X_n) = \begin{cases} 1 - \prod_{i=1}^{n} p_i \text{ if all } X_i(s) \text{ are } 1 \\ 1 \text{ if there exists at least one } X_i = 0 \end{cases} \quad (14)$$

The AND-gate inference was also described in ([3], p. 33). Eq. (14) varies according to two cases whose arrangement counters are listed as follows

$L = \varnothing$

$$c(\Omega : \{X_i = 1\}) = 1, c(\Omega : \{X_i = 0\}) = 0, c(\Omega) = 1.$$

$L \neq \varnothing$

$$c(\Omega : \{X_i = 1\}) = 2^{n-1} - 1, c(\Omega : \{X_i = 0\}) = 2^{n-1}, c(\Omega) = 2^n - 1.$$

The **OR-gate** inference represents prerequisite relationship satisfying OR-gate condition specified by Eq. (15) ([2], p. 157).

$$P(Y = 1|A_i = ON \text{ for some } i) = 1 \tag{15}$$

The OR-gate condition implies

$$P(Y = 0|A_i = ON \text{ for some } i) = 0$$

From Eq. (10), we have ([2], p. 159)

$$\begin{aligned} P(Y = 0|X_1, X_2, \ldots, X_n) &= \sum_{A_1, A_2, \ldots, A_n} P(Y = 1|A_1, A_2, \ldots, A_n) \prod_{i=1}^{n} P(A_i|X_i) \\ &= \prod_{i=1}^{n} P(A_i = OFF|X_i) \\ &\left(\text{due to } P(Y = 1|A_i = ON \text{ for some } i) = 0\right) \\ &= \left(\prod_{i \in K} P(A_i = OFF|X_i = 1)\right)\left(\prod_{i \notin K} P(A_i = OFF|X_i = 0)\right) \\ &= \left(\prod_{i \in K}(1 - p_i)\right)\left(\prod_{i \notin K} 1\right) = \begin{cases} \prod_{i \in K}(1 - p_i) \text{if } K \neq \varnothing \\ 1 \text{ if } K = \varnothing \end{cases} \end{aligned}$$

(Due to Eq. (9))

In general, Eq. (16) specifies OR-gate inference.

$$\begin{aligned} P(X_1 \oplus X_2 \oplus \ldots \oplus X_n) = 1 - P(Y = 0|X_1, X_2, \ldots, X_n) &= \begin{cases} 1 - \prod_{i \in K}(1 - p_i) \text{ if } K \neq \varnothing \\ 0 \text{ if } K = \varnothing \end{cases} \\ P(Y = 0|X_1, X_2, \ldots, X_n) &= \begin{cases} \prod_{i \in K}(1 - p_i) \text{ if } K \neq \varnothing \\ 1 \text{ if } K = \varnothing \end{cases} \end{aligned} \tag{16}$$

where K is the set of X_is whose values are 1. The OR-gate inference was mentioned in Refs. ([2], p. 158) and ([3], p. 20). Eq. (16) varies according to two cases whose arrangement counters are listed as follows

http://dx.doi.org/10.5772/intechopen.70057

$$K \neq \varnothing$$

$$c(\Omega : \{X_i = 1\}) = 2^{n-1}, c(\Omega : \{X_i = 0\}) = 2^{n-1} - 1, c(\Omega) = 2^n - 1.$$

$$K = \varnothing$$

$$c(\Omega : \{X_i = 1\}) = 0, c(\Omega : \{X_i = 0\}) = 1, c(\Omega) = 1.$$

According to De Morgan's rule with regard to AND-gate and OR-gate, we have

$$P\Big(\text{not}(X_1 \odot X_2 \odot \ldots \odot X_n)\Big) = P\Big(\Big(\text{not}(X_1)\Big) \oplus \Big(\text{not}(X_2)\Big) \oplus \ldots \oplus \Big(\text{not}(X_n)\Big)\Big)$$
$$= \begin{cases} 1 - \prod_{i \in L}\Big(1 - (1 - p_i)\Big) \text{if } L \neq \varnothing \\ 0 \text{ if } L = \varnothing \end{cases}$$

(Due to Eq. (16))

According to Eq. (14), we also have

$$P\Big(\text{not}(X_1 \oplus X_2 \oplus \ldots \oplus X_n)\Big) = P\Big(\Big(\text{not}(X_1)\Big) \odot \Big(\text{not}(X_2)\Big) \odot \ldots \odot \Big(\text{not}(X_n)\Big)\Big)$$
$$= \begin{cases} \prod_{i=1}^{n} P\Big(\text{not}(X_i)\Big) \text{ if all not } (X_i)(s) \text{ are } 1 \\ 0 \text{ if there exists at least one not } (X_i) = 0 \end{cases}$$
$$= \begin{cases} \prod_{i=1}^{n} (1 - p_i) \text{ if all } X_i(s) \text{ are } 0 \\ 0 \text{ if there exists at least one } X_i = 1 \end{cases}$$

In general, Eq. (17) specifies NAND-gate inference and NOR-gate inference derived from AND-gate and OR-gate

$$P\Big(\text{not}(X_1 \odot X_2 \odot \ldots \odot X_n)\Big) = \begin{cases} 1 - \prod_{i \in L} p_i \text{ if } L \neq \varnothing \\ 0 \text{ if } L = \varnothing \end{cases}$$
$$P\Big(\text{not}(X_1 \oplus X_2 \oplus \ldots \oplus X_n)\Big) = \begin{cases} \prod_{i=1}^{n} q_i \text{ if } K = \varnothing \\ 0 \text{ if } K \neq \varnothing \end{cases} \quad (17)$$

where K and L are the sets of X_is whose values are 1 and 0, respectively.

Suppose the number of sources X_is is even. Let O be the set of X_is whose indices are odd. Let O_1 and O_2 be subsets of O, in which all X_is are 1 and 0, respectively. Let E be the set of X_is whose indices are even. Let E_1 and E_2 be the subsets of E, in which all X_is are 1 and 0, respectively.

$$\begin{cases} E = \{2, 4, 6, \ldots, n\} \\ E_1 \subseteq E \\ E_2 \subseteq E \\ E_1 \cup E_2 = E \\ E_1 \cap E_2 = \varnothing \\ X_i = 1, \forall i \in E_1 \\ X_i = 0, \forall i \in E_2 \end{cases} \text{ and } \begin{cases} O = \{1, 3, 5, \ldots, n-1\} \\ O_1 \subseteq O \\ O_2 \subseteq O \\ O_1 \cup O_2 = O \\ O_1 \cap O_2 = \varnothing \\ X_i = 1, \forall i \in O_1 \\ X_i = 0, \forall i \in O_2 \end{cases}$$

Thus, O_1 and E_1 are the subsets of K. Sources X_is and target Y follow **XOR-gate** if one of two XOR-gate conditions specified by Eq. (18) is satisfied.

$$\begin{aligned} &P\left(Y=1 \middle| \begin{cases} A_i = ON \text{ for } i \in O \\ A_i = OFF \text{ for } i \notin O \end{cases}\right) = P(Y=1|A_1 = ON, A_2 = OFF, \ldots, A_{n-1} = ON, A_n = OFF) = 1 \\ &P\left(Y=1 \middle| \begin{cases} A_i = ON \text{ for } i \in E \\ A_i = OFF \text{ for } i \notin E \end{cases}\right) = P(Y=1|A_1 = OFF, A_2 = ON, \ldots, A_{n-1} = OFF, A_n = ON) = 1 \end{aligned} \tag{18}$$

From Eq. (10), we have

$$P(Y=1|X_1, X_2, \ldots, X_n) = \sum_{A_1, A_2, \ldots, A_n} P(Y=1|A_1, A_2, \ldots, A_n) \prod_{i=1}^{n} P(A_i|X_i)$$

If both XOR-gate conditions are not satisfied then,

$$P(Y=1|X_1, X_2, \ldots, X_n) = 0$$

If the first XOR-gate condition is satisfied, we have

$$\begin{aligned} &P(Y=1|X_1, X_2, \ldots, X_n) \\ &\quad = P(Y=1|A_1 = ON, A_2 = OFF, \ldots, A_{n-1} = ON, A_n = OFF) \prod_{i=1}^{n} P(A_i|X_i) \\ &\quad = \left(\prod_{i \in O} P(A_i = ON|X_i)\right)\left(\prod_{i \in E} P(A_i = OFF|X_i)\right) \end{aligned}$$

We have

$$\begin{aligned} &\prod_{i \in O} P(A_i = ON|X_i) \\ &\quad = \left(\prod_{i \in O_1} P(A_i = ON|X_i = 1)\right) * \left(\prod_{i \in O_2} P(A_i = ON|X_i = 0)\right) \\ &\quad = \left(\prod_{i \in O_1} p_i\right) * \left(\prod_{i \in O_2} 0\right) = \begin{cases} \prod_{i \in O_1} p_i \text{ if } O_2 = \varnothing \\ 0 \text{ if } O_2 \neq \varnothing \end{cases} \end{aligned}$$

http://dx.doi.org/10.5772/intechopen.70057

(Due to Eq. (9))

We also have

$$\prod_{i\in E} P(A_i = OFF|X_i)$$
$$= \left(\prod_{i\in E_1} P(A_i = OFF|X_i = 1)\right) * \left(\prod_{i\in E_2} P(A_i = OFF|X_i = 0)\right)$$
$$= \left(\prod_{i\in E_1}(1-p_i)\right)\left(\prod_{i\in E_2} 1\right) = \begin{cases} \prod_{i\in E_1}(1-p_i) \text{ if } E_1 \neq \varnothing \\ 1 \text{ if } E_1 = \varnothing \end{cases}$$

(Due to Eq. (9))

Given the first XOR-gate condition, it implies

$$P(Y = 1|X_1, X_2, \ldots, X_n) = \left(\prod_{i\in O} P(A_i = ON|X_i)\right)\left(\prod_{i\in E} P(A_i = OFF|X_i)\right)$$
$$= \begin{cases} \left(\prod_{i\in O_1} p_i\right)\left(\prod_{i\in E_1}(1-p_i)\right) \text{if } O_2 = \varnothing \text{ and } E_1 \neq \varnothing \\ \prod_{i\in O_1} p_i \text{if } O_2 = \varnothing \text{ and } E_1 = \varnothing \\ 0 \text{ if } O_2 \neq \varnothing \end{cases}$$

Similarly, given the second XOR-gate condition, we have

$$P(Y = 1|X_1, X_2, \ldots, X_n) = \left(\prod_{i\in E} P(A_i = ON|X_i)\right)\left(\prod_{i\in O} P(A_i = OFF|X_i)\right)$$
$$= \begin{cases} \left(\prod_{i\in E_1} p_i\right)\left(\prod_{i\in O_1}(1-p_i)\right) \text{if } E_2 = \varnothing \text{ and } O_1 \neq \varnothing \\ \prod_{i\in E_1} p_i \text{ if } E_2 = \varnothing \text{ and } O_1 = \varnothing \\ 0 \text{ if } E_2 \neq \varnothing \end{cases}$$

If one of XOR-gate conditions is satisfied then,

$$P(Y = 1|X_1, X_2, \ldots, X_n)$$
$$= \left(\prod_{i\in O} P(A_i = ON|X_i)\right)\left(\prod_{i\in E} P(A_i = OFF|X_i)\right) + \left(\prod_{i\in E} P(A_i = ON|X_i)\right)\left(\prod_{i\in O} P(A_i = OFF|X_i)\right)$$

This implies Eq. (19) to specify XOR-gate inference.

$$P(X_1 \otimes X_2 \otimes \ldots \otimes X_n) = P(Y = 1|X_1, X_2, \ldots, X_n)$$

$$= \begin{cases} \left(\prod_{i \in O_1} p_i\right)\left(\prod_{i \in E_1}(1-p_i)\right) + \left(\prod_{i \in E_1} p_i\right)\left(\prod_{i \in O_1}(1-p_i)\right) \text{ if } O_2 = \varnothing \text{ and } E_2 = \varnothing \\ \left(\prod_{i \in O_1} p_i\right)\left(\prod_{i \in E_1}(1-p_i)\right) \text{ if } O_2 = \varnothing \text{ and } E_1 \neq \varnothing \text{ and } E_2 \neq \varnothing \\ \prod_{i \in O_1} p_i \text{ if } O_2 = \varnothing \text{ and } E_1 = \varnothing \\ \left(\prod_{i \in E_1} p_i\right)\left(\prod_{i \in O_1}(1-p_i)\right) \text{ if } E_2 = \varnothing \text{ and } O_1 \neq \varnothing \text{ and } O_2 \neq \varnothing \\ \prod_{i \in E_1} p_i \text{ if } E_2 = \varnothing \text{ and } O_1 = \varnothing \\ 0 \text{ if } O_2 \neq \varnothing \text{ and } E_2 \neq \varnothing \\ 0 \text{ if } n < 2 \text{ or } n \text{ is odd} \end{cases} \tag{19}$$

where

$$\begin{cases} O = \{1, 3, 5, \ldots, n-1\} \\ O_1 \subseteq O \\ O_2 \subseteq O \\ O_1 \cup O_2 = O \\ O_1 \cap O_2 = \varnothing \\ X_i = 1, \forall i \in O_1 \\ X_i = 0, \forall i \in O_2 \end{cases} \text{ and } \begin{cases} E = \{2, 4, 6, \ldots, n\} \\ E_1 \subseteq E \\ E_2 \subseteq E \\ E_1 \cup E_2 = E \\ E_1 \cap E_2 = \varnothing \\ X_i = 1, \forall i \in E_1 \\ X_i = 0, \forall i \in E_2 \end{cases}$$

Where,

Given $n \geq 2$ and n is even, Eq. (19) varies according to six cases whose arrangement counters are listed as follows

$$O_2 = \varnothing \text{ and } E_2 = \varnothing$$

$$c(\Omega : \{X_i = 1\}) = 1, c(\Omega : \{X_i = 0\}) = 0, c(\Omega) = 1.$$

$$O_2 = \varnothing \text{ and } E_1 \neq \varnothing \text{ and } E_2 \neq \varnothing$$

$$c(\Omega : \{X_i = 1\}) = 2^{\frac{n}{2}} - 2, c(\Omega : \{X_i = 0\}) = 0, c(\Omega) = 2^{\frac{n}{2}} - 2.$$

$$O_2 = \varnothing \text{ and } E_1 = \varnothing$$

$$c(\Omega : \{X_i = 1\}) = 1, c(\Omega : \{X_i = 0\}) = 0, c(\Omega) = 1.$$

$$E_2 = \varnothing \text{ and } O_1 \neq \varnothing \text{ and } O_2 \neq \varnothing$$

$$c(\Omega : \{X_i = 1\}) = 2^{\frac{n}{2}-1} - 1, c(\Omega : \{X_i = 0\}) = 2^{\frac{n}{2}-1} - 1, c(\Omega) = 2^{\frac{n}{2}} - 2.$$

$$E_2 = \varnothing \text{ and } O_1 = \varnothing$$

$$c(\Omega : \{X_i = 1\}) = 0, c(\Omega : \{X_i = 0\}) = 1, c(\Omega) = 1.$$

$$O_2 \neq \varnothing \text{ and } E_2 \neq \varnothing$$

$$c(\Omega : \{X_i = 1\}) = \left(2^{\frac{n}{2}-1} - 1\right)\left(2^{\frac{n}{2}} - 1\right), c(\Omega : \{X_i = 0\}) = 2^{\frac{n}{2}-1}\left(2^{\frac{n}{2}} - 1\right), c(\Omega) = \left(2^{\frac{n}{2}} - 1\right)^2.$$

Suppose the number of sources X_is is even. According to **XNOR-gate** inference [1], the output is on if all inputs get the same value 1 (or 0). Sources X_i (s) and target Y follow XNOR-gate if one of two XNOR-gate conditions specified by Eq. (20) is satisfied.

$$\begin{array}{l} P(Y = 1|A_i = ON, \forall i) = 1 \\ P(Y = 1|A_i = OFF, \forall i) = 1 \end{array} \tag{20}$$

From Eq. (10), we have

$$P(Y = 1|X_1, X_2, \ldots, X_n) = \sum_{A_1, A_2, \ldots, A_n} P(Y = 1|A_1, A_2, \ldots, A_n)\prod_{i=1}^{n} P(A_i|X_i)$$

If both XNOR-gate conditions are not satisfied then,

$$P(Y = 1|X_1, X_2, \ldots, X_n) = 0$$

If A_i = *ON* for all i, we have

$$\begin{aligned} P(Y = 1|X_1, X_2, \ldots, X_n) &= P(Y = 1|A_i = ON, \forall i)\prod_{i=1}^{n} P(A_i = ON|X_i) \\ &= \prod_{i=1}^{n} P(A_i = ON|X_i) = \begin{cases} \prod_{i=1}^{n} p_i \text{if } L = \varnothing \\ 0 \text{ if } L \neq \varnothing \end{cases} \end{aligned}$$

(Please see similar proof in AND-gate inference)

If A_i = *OFF* for all i, we have

$$P(Y = 1|X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P(A_i = OFF|X_i) = \begin{cases} \prod_{i \in K}(1 - p_i) \text{ if } K \neq \varnothing \\ 1 \text{ if } K = \varnothing \end{cases}$$

(Please see similar proof in OR-gate inference)

If one of XNOR-gate conditions is satisfied then,

$$P(Y = 1|X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P(A_i = ON|X_i) + \prod_{i=1}^{n} P(A_i = OFF|X_i)$$

This implies Eq. (21) to specify XNOR-gate inference.

$$P\Big(\text{not}(X_1 \otimes X_2 \otimes \ldots \otimes X_n)\Big) = P(Y = 1|X_1, X_2, \ldots, X_n) = \begin{cases} \prod_{i=1}^{n} p_i + \prod_{i=1}^{n}(1 - p_i) \text{ if } L = \varnothing \\ \prod_{i \in K}(1 - p_i) \text{ if } L \neq \varnothing \text{ and } K \neq \varnothing \\ 1 \text{ if } L \neq \varnothing \text{ and } K = \varnothing \end{cases} \tag{21}$$

where K and L are the sets of X_is whose values are 1 and 0, respectively. Eq. (21) varies according to three cases whose arrangement counters are listed as follows

$$L = \varnothing$$

$$c(\Omega : \{X_i = 1\}) = 1, c(\Omega : \{X_i = 0\}) = 0, c(\Omega) = 1.$$

$$L \neq \varnothing \text{ and } K \neq \varnothing$$

$$c(\Omega : \{X_i = 1\}) = 2^{n-1} - 1, c(\Omega : \{X_i = 0\}) = 2^{n-1} - 1, c(\Omega) = 2^n - 2.$$

$$L \neq \varnothing \text{ and } K = \varnothing$$

$$c(\Omega : \{X_i = 1\}) = 0, c(\Omega : \{X_i = 0\}) = 1, c(\Omega) = 1.$$

Let U be a set of indices such that $A_i = ON$ and let $\alpha \geq 0$ and $\beta \geq 0$ be predefined numbers. The U-gate inference is defined based on α, β and cardinality of U. **Table 4** specifies four common U-gate conditions.

Note that U-gate condition on $|U|$ can be arbitrary and it is only relevant to A_is (ON or OFF) and the way to combine A_is. For example, AND-gate and OR-gate are specific cases of U-gate with $|U| = n$ and $|U| \geq 1$, respectively. XOR-gate and XNOR-gate are also specific cases of U-gate with specific conditions on A_i (s). However, it must be assured that there is at least one combination of A_is satisfying the predefined U-gate condition, which causes that U-gate probability is not always equal to 0. In this research, U-gate is the most general nonlinear gate where U-gate probability contains products of weights (see **Table 5**). Later on, we will research a so-called SIGMA-gate that contains only linear combination of weights (sum of weights, see Eq. (23)). Shortly, each X-gate is a pattern owning a particular X-gate inference that is X-gate probability $P(X_1 \times X_2 \times \ldots \times X_n)$. Each X-gate inference is based on particular X-gate condition (s) relevant to only variables A_is.

From Eq. (10), we have

$$P(Y = 1|X_1, X_2, \ldots, X_n) = \sum_{A_1, A_2, \ldots, A_n} P(Y = 1|A_1, A_2, \ldots, A_n) \prod_{i=1}^{n} P(A_i|X_i)$$

Let $\mathcal{U}$ be the set of all possible U (s), we have

$\|U\|=\alpha$ $P(Y=1\|A_1, A_2, \ldots, A_n)=1$ if there are exactly α variables A_i = *ON* (s). Otherwise, $P(Y=1\|A_1, A_2, \ldots, A_n)=0$.
$\|U\|\geq\alpha$ $P(Y=1\|A_1, A_2, \ldots, A_n)=1$ if there are at least α variables A_i = *ON* (s). Otherwise, $P(Y=1\|A_1, A_2, \ldots, A_n)=0$.
$\|U\|\leq\beta$ $P(Y=1\|A_1, A_2, \ldots, A_n)=1$ if there are at most β variables A_i = *ON* (s). Otherwise, $P(Y=1\|A_1, A_2, \ldots, A_n)=0$.
$\alpha\leq\|U\|\leq\beta$ $P(Y=1\|A_1, A_2, \ldots, A_n)=1$ if the number of A_i = *ON* (s) is from α to β. Otherwise, $P(Y=1\|A_1, A_2, \ldots, A_n)=0$.

Table 4. U-gate conditions.

$$P(Y=1|X_1, X_2, \ldots, X_n)=\sum_{U\in\mathcal{U}}P(Y=1|A_1, A_2, \ldots, A_n)\prod_{i=1}^{n}P(A_i|X_i)$$
$$=\sum_{U\in\mathcal{U}}\prod_{i\in U}P(A_i=ON|X_i)\prod_{j\notin U}P(A_j=OFF|X_j)$$

If $X_i=0, \forall i\in U$ then,

$$P(Y=1|X_1, X_2, \ldots, X_n)=\sum_{U\in\mathcal{U}}\prod_{i\in U}0\prod_{j\notin U}P(A_j=OFF|X_j)=0$$

This implies all sets U (s) must be subsets of K. The U-gate probability is rewritten as follows

$$P(Y=1|X_1, X_2, \ldots, X_n)=\sum_{U\in\mathcal{U}}\prod_{i\in U}P(A_i=ON|X_i=1)\prod_{j\notin U}P(A_j=OFF|X_j)$$
$$=\sum_{U\in\mathcal{U}}\prod_{i\in U}p_i\prod_{j\notin U}P(A_j=OFF|X_j)$$
$$=\sum_{U\in\mathcal{U}}\prod_{i\in U}p_i\prod_{j\in K\setminus U}P(A_j=OFF|X_j=1)\prod_{j\notin K}P(A_j=OFF|X_j=0)$$
$$=\sum_{U\in\mathcal{U}}\prod_{i\in U}p_i\prod_{j\in K\setminus U}(1-p_j)\prod_{j\notin K}1=\sum_{U\in\mathcal{U}}\prod_{i\in U}p_i\prod_{j\in K\setminus U}(1-p_j)$$

(Due to Eq. (9))

Let P_U be the U-gate probability; **Table 5** specifies U-gate inference and cardinality of $\mathcal{U}$ where $\mathcal{U}$ is the set of subsets (U) of K.

Note that the notation $\binom{n}{j}$ denotes the number of combinations of j elements taken from n elements.

$$\binom{n}{j}=\frac{n!}{j!(n-j)!}$$

Arrangement counters relevant to U-gate inference and the set K are listed as follows

Let,	$S_U = \sum_{U \in \mathcal{U}} \prod_{i \in U} p_i \prod_{j \in K \setminus U} (1 - p_j)$ $P_U = P(X_1 \uplus X_2 \uplus \ldots \uplus X_n) = P(Y = 1 \mid X_1, X_2, \ldots, X_n)$
As a convention,	$\prod_{i \in U} p_i = 1 \text{ if} \lvert U \rvert = 0$ $\prod_{j \in K \setminus U} (1 - p_j) = 1 \text{ if} \lvert U \rvert = \lvert K \rvert$
$\lvert U \rvert$=0	$P_U = \begin{cases} \prod_{j=1}^{n} (1 - p_j) \text{ if } \lvert K \rvert > 0 \\ 1 \text{ if } \lvert K \rvert = 0 \end{cases}$ $\lvert \mathcal{U} \rvert = 1$
$\lvert U \rvert \geq 0$	$P_U = \begin{cases} S_U \text{ if } \lvert K \rvert > 0 \\ 1 \text{ if } \lvert K \rvert = 0 \end{cases}$ $\lvert \mathcal{U} \rvert = 2^{\lvert K \rvert}$ The case $\lvert U \rvert \geq 0$ is the same to the case $\lvert U \rvert \leq n$
$\lvert U \rvert = n$	$P_U = \begin{cases} \prod_{i=1}^{n} p_i \text{if } \lvert K \rvert = n \\ 0 \text{ if } \lvert K \rvert < n \end{cases}$ $\lvert \mathcal{U} \rvert = \begin{cases} 1 \text{ if } \lvert K \rvert = n \\ 0 \text{ if } \lvert K \rvert < n \end{cases}$
$\lvert U \rvert = \alpha$ $0 < \alpha < n$	$P_U = \begin{cases} S_U \text{ if } \lvert K \rvert \geq \alpha \\ 0 \text{ if } \lvert K \rvert < \alpha \end{cases}$ $\lvert \mathcal{U} \rvert = \begin{cases} \binom{\lvert K \rvert}{\alpha} \text{ if } \lvert K \rvert \geq \alpha \\ 0 \text{ if } \lvert K \rvert < \alpha \end{cases}$
$\lvert U \rvert \geq \alpha$ $0 < \alpha < n$	$P_U = \begin{cases} S_U \text{ if } \lvert K \rvert \geq \alpha \\ 0 \text{ if } \lvert K \rvert < \alpha \end{cases}$ $\lvert \mathcal{U} \rvert = \begin{cases} \sum_{j=\alpha}^{\lvert K \rvert} \binom{\lvert K \rvert}{j} \text{ if } \lvert K \rvert \geq \alpha \\ 0 \text{ if } \lvert K \rvert < \alpha \end{cases}$
$\lvert U \rvert \leq \beta$ $0 < \beta < n$	$P_U = \begin{cases} S_U \text{ if } \lvert K \rvert > 0 \\ 1 \text{ if } \lvert K \rvert = 0 \end{cases}$ $\lvert \mathcal{U} \rvert = \begin{cases} \sum_{j=0}^{\min(\beta, \lvert K \rvert)} \binom{\lvert K \rvert}{j} \text{ if } \lvert K \rvert > 0 \\ 1 \text{ if } \lvert K \rvert = 0 \end{cases}$
$\alpha \leq \lvert U \rvert \leq \beta$ $0 < \alpha < n$ $0 < \beta < n$	$P_U = \begin{cases} S_U \text{ if } \lvert K \rvert \geq \alpha \\ 0 \text{ if } \lvert K \rvert < \alpha \end{cases}$ $\lvert \mathcal{U} \rvert = \begin{cases} \sum_{j=\alpha}^{\min(\beta, \lvert K \rvert)} \binom{\lvert K \rvert}{j} \text{ if } \lvert K \rvert \geq \alpha \\ 0 \text{ if } \lvert K \rvert < \alpha \end{cases}$

Table 5. U-gate inference.

http://dx.doi.org/10.5772/intechopen.70057

$$|K| = 0$$

$$c(\Omega : \{X_i = 1\}) = 0, c(\Omega : \{X_i = 0\}) = 1, c(\Omega) = 1.$$

$$|K| = 1$$

$$c(\Omega : \{X_i = 1\}) = 1, c(\Omega : \{X_i = 0\}) = 0, c(\Omega) = 1.$$

$$|K| = \alpha \text{ and } \alpha > 0$$

$$c(\Omega : \{X_i = 1\}) = \binom{n-1}{\alpha-1}, c(\Omega : \{X_i = 0\}) = \binom{n-1}{\alpha}, c(\Omega) = \binom{n}{\alpha}.$$

$$|K| \leq \alpha \text{ and } \alpha > 0$$

$$c(\Omega : \{X_i = 1\}) = \sum_{j=1}^{\alpha} \binom{n-1}{j-1}, c(\Omega : \{X_i = 0\}) = \sum_{j=0}^{\alpha} \binom{n-1}{j}, c(\Omega) = \sum_{j=0}^{\alpha} \binom{n}{j}.$$

$$|K| \geq \alpha \text{ and } \alpha > 0$$

$$c(\Omega : \{X_i = 1\}) = \sum_{j=\alpha}^{n} \binom{n-1}{j-1}, c(\Omega : \{X_i = 0\}) = \sum_{j=\alpha}^{n-1} \binom{n-1}{j}, c(\Omega) = \sum_{j=\alpha}^{n} \binom{n}{j}.$$

The **SIGMA-gate** inference [9] represents aggregation relationship satisfying SIGMA-gate condition specified by Eq. (22).

$$P(Y) = P\left(\sum_{i=1}^{n} A_i\right)$$

where the set of A_i is complete and mutually exclusive

$$\begin{gathered} \sum_{i=1}^{n} w_i = 1 \\ A_i \cap A_j = \emptyset, \forall i \neq j \end{gathered} \tag{22}$$

The sigma sum $\sum_{i=1}^{n} A_i$ indicates that Y is exclusive union of A_is and here, it does not express arithmetical additions.

$$Y = \sum_{i=1}^{n} A_i = \bigcup_{i=1}^{n} A_i$$

This implies

$$P(Y) = P\left(\sum_{i=1}^{n} A_i\right) = P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$$

The sigma sum $\sum_{i=1}^{n} P(A_i)$ now expresses arithmetical additions of probabilities $P(A_i)$.

SIGMA-gate inference requires the set of A_is is complete and mutually exclusive, which means that the set of X_is is complete and mutually exclusive too. The SIGMA-gate probability is [9]

$$P(Y|X_1, X_2, \ldots, X_n) = P\left(\sum_{i=1}^{n} A_i \middle| X_1, X_2, \ldots, X_n\right)$$
$$(\text{due to SIGMA} - \text{gate condition})$$
$$= \sum_{i=1}^{n} P(A_i|X_1, X_2, \ldots, X_n)$$
$$\left(\text{because } A_i(\text{s}) \text{ are mutually exclusive}\right)$$
$$= \sum_{i=1}^{n} P(A_i|X_i)$$
$$(\text{because } A_i \text{ is only dependent on } X_i)$$

It implies

$$P(Y = 1|X_1, X_2, \ldots, X_n) = \sum_{i=1}^{n} P(A_i = ON|X_i)$$
$$= \left(\sum_{i \in K} P(A_i = ON|X_i = 1)\right) + \left(\sum_{i \notin K} P(A_i = ON|X_i = 0)\right)$$
$$= \sum_{i \in K} w_i + \sum_{i \notin K} 0 = \sum_{i \in K} w_i$$

(Due to Eq. (9))

In general, Eq. (23) specifies the theorem of SIGMA-gate inference [9]. The base of this theorem was mentioned by Millán and Pérez-de-la-Cruz ([4], pp. 292-295).

$$P(X_1 + X_2 + \ldots + X_n) = P\left(\sum_{i=1}^{n} X_i\right) = P(Y = 1|X_1, X_2, \ldots, X_n) = \sum_{i \in K} w_i$$
$$P(Y = 0|X_1, X_2, \ldots, X_n) = 1 - \sum_{i \in K} w_i = \sum_{i \in L} w_i$$

where the set of X_is is complete and mutually exclusive.

$$\sum_{i=1}^{n} w_i = 1$$
$$X_i \cap X_j = \varnothing, \forall i \neq j \tag{23}$$

The arrangement counters of SIGMA-gate inference are $c(\Omega{:}\{X_i = 1\}) = c(\Omega{:}\{X_i = 0\}) = 2^{n-1}$, $c(\Omega) = 2^n$.

Eq. (9) specifies the "clockwise" strength of relationship between X_i and Y. Event $X_i = 1$ causes event $A_i = ON$ with "clockwise" weight w_i. There is a question "given $X_i = 0$, how likely the event $A_i = OFF$ is". In order to solve this problem, I define a so-called "counterclockwise" strength of relationship between X_i and Y denoted ω_i. Event $X_i = 0$ causes event $A_i = OFF$ with

"counterclockwise" weight ω_i. In other words, each arc in simple graph is associated with a clockwise weight w_i and a counterclockwise weight ω_i. Such graph is called *bi-weight simple graph* shown in **Figure 4**.

With bi-weight simple graph, all X-gate inferences are extended as so-called X-gate bi-inferences. Derived from Eq. (9), Eq. (24) specifies conditional probability of accountable variables with regard to bi-weight graph.

$$\begin{aligned} P(A_i = ON|X_i = 1) &= p_i = w_i \\ P(A_i = ON|X_i = 0) &= 1 - \rho_i = 1 - \omega_i \\ P(A_i = OFF|X_i = 1) &= 1 - p_i = 1 - w_i \\ P(A_i = OFF|X_i = 0) &= \rho_i = \omega_i \end{aligned} \tag{24}$$

The probabilities $P(A_i = ON \mid X_i = 0)$ and $P(A_i = OFF \mid X_i = 1)$ are called clockwise adder d_i and counterclockwise adder δ_i. As usual, d_i and δ_i are smaller than w_i and ω_i. When $d_i = 0$, bi-weight graph becomes normal simple graph.

$$\begin{aligned} d_i &= P(A_i = ON|X_i = 0) = 1 - \rho_i = 1 - \omega_i \\ \delta_i &= P(A_i = OFF|X_i = 1) = 1 - p_i = 1 - w_i \end{aligned}$$

The total clockwise weight or total counterclockwise weight is defined as sum of clockwise weight and clockwise adder or sum of counterclockwise weight and counterclockwise adder. Eq. (25) specifies such total weights W_i and $\mathcal{W}_i$. These weights are also called relationship powers.

$$\begin{aligned} W_i &= w_i + d_i \\ \mathcal{W}_i &= \omega_i + \delta_i \end{aligned}$$

where

$$\begin{aligned} d_i &= 1 - \rho_i = 1 - \omega_i \\ \delta_i &= 1 - p_i = 1 - w_i \end{aligned} \tag{25}$$

Given Eq. (25), the set of all A_is is complete if and only if $\sum_{i=1}^{n} w_i = 1$.

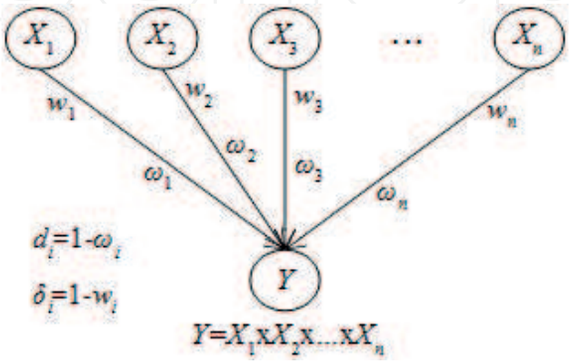

Figure 4. Bi-weight simple graph.

By extending aforementioned X-gate inferences, we get bi-inferences for AND-gate, OR-gate, NAND-gate, NOR-gate, XOR-gate, XNOR-gate, and U-gate as shown in **Table 6**.

The largest cardinalities of K (L) are 2^{n-1} and 2^n with and without fixed X_i. Thus, it is possible to calculate arrangement counters. As a convention, the product of probabilities is 1 if indices set is empty.

$$\prod_{i \in I} f_i = 1 \text{ if } I = \varnothing$$

With regard to SIGMA-gate bi-inference, the sum of all total clockwise weights must be 1 as follows

$$\sum_{i=1}^{n} W_i = \sum_{i=1}^{n} (w_i + d_i) = \sum_{i=1}^{n} (w_i + 1 - \omega_i) = 1$$

Derived from Eq. (23), the SIGMA-gate probability for bi-weight graph is

$$\begin{aligned} &P(X_1 + X_2 + \ldots + X_n) = \sum_{i=1}^{n} P(A_i = ON | X_i) \\ &= \sum_{i \in K} P(A_i = ON | X_i = 1) + \sum_{i \in L} P(A_i = ON | X_i = 0) \\ &= \sum_{i \in K} w_i + \sum_{i \in L} d_i \end{aligned}$$

Shortly, Eq. (26) specifies SIGMA-gate bi-inference.

$$P(X_1 + X_2 + \ldots + X_n) = \sum_{i \in K} w_i + \sum_{i \in L} d_i$$

where the set of X_i(s) is complete and mutually exclusive.

$$\begin{gathered} \sum_{i=1}^{n} W_i = 1 \\ X_i \cap X_j = \varnothing, \forall i \neq j \end{gathered} \tag{26}$$

The next section will research diagnostic relationship which adheres to X-gate inference.

4. Multihypothesis diagnostic relationship

Given a simple graph shown in **Figure 2**, if we replace the target source Y by an evidence D, we get a so-called *multihypothesis diagnostic relationship* whose property adheres to X-gate inference. Maybe there are other diagnostic relationships in which X-gate inference is not concerned. However, this research focuses on X-gate inference and so multi-hypothesis diagnostic relationship is called *X-gate diagnostic relationship*. Sources X_1, X_2,…, X_n become hypotheses. As a convention, these hypotheses have prior uniform distribution.

http://dx.doi.org/10.5772/intechopen.70057

According to aforementioned X-gate network shown in **Figures 2** and **3**, the target variable must be binary whereas evidence D can be numeric. It is impossible to establish the evidence D as direct target variable. Thus, the solution of this problem is to add an augmented target binary variable Y and then, the evidence D is connected directly to Y. In other words, the *X-gate diagnostic network* have n sources $\{X_1, X_2,\ldots, X_n\}$, one augmented hypothesis Y, and one evidence D. As a convention, X-gate diagnostic network is called *X-D network*. The CPTs of the entire network are determined based on combination of diagnostic relationship and X-gate inference mentioned in previous sections. **Figure 5** depicts the augmented X-D network. Note that variables $X_1, X_2,\ldots, X_n$, and Y are always binary.

Appendix A3 is the proof that the augmented X-D network is equivalent to X-D network with regard to variables $X_1, X_2,\ldots, X_n$ and D. As a convention, augmented X-D network is considered as same as X-D network.

The simplest case of X-D network is NOT-D network having one hypothesis X_1 and one evidence D, equipped with NOT-gate inference. NOT-D network satisfies diagnostic condition because it essentially represents the single diagnostic relationship. Inferred from Eqs. (1) and (7), the conditional probability $P(D|X_1)$ and posterior probability $P(X_1|D)$ of NOT-D network are

$$P(D|X_1) = \begin{cases} 1 - D \text{ if } X_1 = 1 \\ D \text{ if } X_1 = 0 \end{cases}$$

$$P(X_1|D) = \frac{P(D|X_1)P(X_1)}{P(X_1)\Big(P(D|X_1 = 0) + P(D|X_1 = 1)\Big)}$$

(Due to Bayes' rule and uniform distribution of X_1)

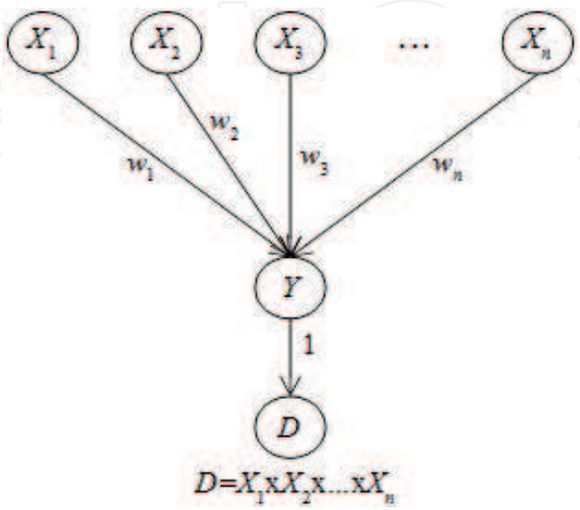

Figure 5. Augmented X-D network.

$$= \frac{P(D|X_1)}{P(D|X_1 = 0) + P(D|X_1 = 1)} = 1 * P(D|X_1)$$

$$\Big(\text{due to } P(D|X_1 = 0) + P(D|X_1 = 1) = 1\Big)$$

It implies NOT-D network satisfies diagnostic condition. Let

$$\Omega = \{X_1, X_2, \ldots, X_n\}$$
$$n = |\Omega|$$

We will validate whether the CPT of diagnostic relationship, $P(D|X)$ specified by Eq. (6), still satisfies diagnostic condition within general case, X-D network. In other words, X-D network is general case of single diagnostic relationship.

Recall from dependencies shown in **Figure 5**, Eq. (27) specifies the joint probability of X-D network.

$$P(\Omega, Y, D) = P(X_1, X_2, \ldots, X_n, Y, D) = P(D|Y)P(Y|X_1, X_2, \ldots, X_n)\prod_{i=1}^{n} P(X_i) \quad (27)$$
$$\text{where } \Omega = \{X1, X2, \ldots, Xn\}.$$

Eq. (28) specifies the conditional probability of D given X_i (likelihood function) and the posterior probability of X_i given D.

$$P(D|X_i) = \frac{P(X_i, D)}{P(X_i)} = \frac{\sum_{\{\Omega, Y, D\}\setminus\{X_i, D\}} P(\Omega, Y, D)}{\sum_{\{\Omega, Y, D\}\setminus\{X_i\}} P(\Omega, Y, D)}$$
$$P(X_i|D) = \frac{P(X_i, D)}{P(D)} = \frac{\sum_{\{\Omega, Y, D\}\setminus\{X_i, D\}} P(\Omega, Y, D)}{\sum_{\{\Omega, Y, D\}\setminus\{D\}} P(\Omega, Y, D)} \quad (28)$$

where $\Omega = \{X_1, X_2, \ldots, X_n\}$ and the sign "\" denotes the subtraction (excluding) operator in set theory [10]. Eq. (29) specifies the joint probability $P(X_i, D)$ and the marginal probability $P(D)$ given uniform distribution of all sources. Appendix A4 is the proof of Eq. (29).

$$P(X_i, D) = \frac{1}{2^n S}\Big((2D - M)s(\Omega : \{X_i\}) + 2^{n-1}(M - D)\Big)$$
$$P(D) = \frac{1}{2^n S}\Big((2D - M)s(\Omega) + 2^n(M - D)\Big) \quad (29)$$

where $s(\Omega)$ and $s(\Omega{:}\{X_i\})$ are specified in **Table 2**. From Eqs. (28–30) specifies conditional probability $P(D|X_i)$, posterior probability $P(X_i|D)$, and transformation coefficient for X-gate inference.

http://dx.doi.org/10.5772/intechopen.70057

$$
\begin{aligned}
P(D|X_i=1) &= \frac{P(X_i=1,D)}{P(X_i=1)} = \frac{(2D-M)s(\Omega:\{X_i=1\})+2^{n-1}(M-D)}{2^{n-1}S} \\
P(D|X_i=0) &= \frac{P(X_i=0,D)}{P(X_i=0)} = \frac{(2D-M)s(\Omega:\{X_i=0\})+2^{n-1}(M-D)}{2^{n-1}S} \\
P(X_i=1|D) &= \frac{P(X_i=1,D)}{P(D)} = \frac{(2D-M)s(\Omega:\{X_i=1\})+2^{n-1}(M-D)}{(2D-M)s(\Omega)+2^{n}(M-D)} \\
P(X_i=0|D) &= 1-P(X_i=1|D) = \frac{(2D-M)s(\Omega:\{X_i=0\})+2^{n-1}(M-D)}{(2D-M)s(\Omega)+2^{n}(M-D)} \\
k &= \frac{P(X_i|D)}{P(D|X_i)} = \frac{2^{n-1}S}{(2D-M)s(\Omega)+2^{n}(M-D)}
\end{aligned}
\tag{30}
$$

The transformation coefficient is rewritten as follows

$$
k = \frac{2^{n-1}S}{2D\left(s(\Omega)-2^{n-1}\right)+M\left(2^{n}-s(\Omega)\right)}
$$

Note that S, D, and M are abstract symbols and there is no proportional connection between $2^{n-1}S$ and D for all D, specified by Eq. (6). Assuming that such proportional connection $2^{n-1}S = aD^j$ exists for all D where a is arbitrary constant. Given binary case when $D = 0$ and $S = 1$, we have

$$
2^{n-1} = 2^{n-1} * 1 = 2^{n-1}S = aD^j = a * 0^j = 0
$$

There is a contradiction, which implies that it is impossible to reduce k into the following form

$$
k = \frac{aD^j}{bD^j}
$$

Therefore, if k is constant with regard to D then,

$$
2D\left(s(\Omega)-2^{n-1}\right)+M\left(2^{n}-s(\Omega)\right) = C \neq 0, \forall D
$$

where C is constant. We have

$$
\begin{aligned}
&\sum_{D}\left(2D\left(s(\Omega)-2^{n-1}\right)+M\left(2^{n}-s(\Omega)\right)\right) = \sum_{D} C \\
&\Rightarrow 2S\left(s(\Omega)-2^{n-1}\right)+NM\left(2^{n}-s(\Omega)\right) = NC \\
&\Rightarrow 2^{n}S = NC
\end{aligned}
$$

It is implied that

$$P(X_1 \odot X_2 \odot \ldots \odot X_n) = \prod_{i \in K} p_i \prod_{i \in L} d_i$$

$$P(X_1 \oplus X_2 \oplus \ldots \oplus X_n) = 1 - \prod_{i \in K} \delta_i \prod_{i \in L} \rho_i$$

$$P\Big(\text{not}(X_1 \odot X_2 \odot \ldots \odot X_n)\Big) = 1 - \prod_{i \in L} \rho_i \prod_{i \in K} \delta_i$$

$$P\Big(\text{not}(X_1 \oplus X_2 \oplus \ldots \oplus X_n)\Big) = \prod_{i \in L} d_i \prod_{i \in K} p_i$$

$$P(X_1 \otimes X_2 \otimes \ldots \otimes X_n) = \prod_{i \in O_1} p_i \prod_{i \in O_2} d_i \prod_{i \in E_1} \delta_i \prod_{i \in E_2} \rho_i + \prod_{i \in E_1} p_i \prod_{i \in E_2} d_i \prod_{i \in O_1} \delta_i \prod_{i \in O_2} \rho_i$$

$$P\Big(\text{not}(X_1 \otimes X_2 \otimes \ldots \otimes X_n)\Big) = \prod_{i \in K} p_i \prod_{i \in L} d_i + \prod_{i \in K} \delta_i \prod_{i \in L} \rho_i$$

$$P(X_1 \uplus X_2 \uplus \ldots \uplus X_n) = \sum_{U \in \mathcal{U}} \left(\prod_{i \in U \cap K} p_i \prod_{i \in U \cap L} d_i \right) \left(\prod_{i \in \overline{U} \cap K} \delta_i \prod_{i \in \overline{U} \cap L} \rho_i \right)$$

There are four common conditions of U: $|U|=\alpha$, $|U|\geq\alpha$, $|U|\leq\beta$, and $\alpha\leq|U|\leq\beta$. Note that $\overline{U}$ is the complement of U,

$$\overline{U} = \{1, 2, \ldots, n\} \setminus U$$

The largest cardinality of $\mathcal{U}$ is:

$$|\mathcal{U}| = 2^n$$

Table 6. Bi-inferences for AND-gate, OR-gate, NAND-gate, NOR-gate, XOR-gate, XNOR-gate, and U-gate.

$$k = \frac{2^{n-1}S}{2D\Big(s(\Omega) - 2^{n-1}\Big) + M\Big(2^n - s(\Omega)\Big)} = \frac{NC}{2C} = \frac{N}{2}$$

This holds

$$2^n S = N\Big(2D\Big(s(\Omega) - 2^{n-1}\Big) + M\Big(2^n - s(\Omega)\Big)\Big) = 2ND\Big(s(\Omega) - 2^{n-1}\Big) + 2S\Big(2^n - s(\Omega)\Big)$$
$$\Rightarrow 2ND\Big(s(\Omega) - 2^{n-1}\Big) - 2S\Big(s(\Omega) - 2^{n-1}\Big) = 0$$
$$\Rightarrow (ND - S)\Big(s(\Omega) - 2^{n-1}\Big) = 0$$

Assuming $ND = S$ we have

$$ND = S = 2NM \Rightarrow D = 2M$$

There is a contradiction because M is maximum value of D. Therefore, if k is constant with regard to D then $s(\Omega) = 2^{n-1}$. Inversely, if $s(\Omega) = 2^{n-1}$ then k is

$$k = \frac{2^{n-1}S}{2D(2^{n-1} - 2^{n-1}) + M(2^n - 2^{n-1})} = \frac{N}{2}$$

Given X-D network is combination of diagnostic relationship and X-gate inference:

$$P(Y=1|X_1, X_2, \ldots, X_n) = P(X_1 \mathrm{x} X_2 \mathrm{x} \ldots \mathrm{x} X_n)$$

$$P(D|Y) = \begin{cases} \dfrac{D}{S} \text{if } Y=1 \\ \dfrac{M}{S} - \dfrac{D}{S} \text{if } Y=0 \end{cases}$$

The diagnostic condition of X-D network is satisfied if and only if

$$s(\Omega) = \sum_a P\Big(Y=1|a(\Omega)\Big) = 2^{|\Omega|-1}, \forall \Omega \neq \varnothing$$

At that time, the transformation coefficient becomes:

$$k = \frac{N}{2}$$

Note that weights $p_i = w_i$ and $\rho_i = \omega_i$, which are inputs of $s(\Omega)$, are abstract variables. Thus, the equality $s(\Omega) = 2^{|\Omega|-1}$ implies all abstract variables are removed and so $s(\Omega)$ does not depend on weights.

Table 7. Diagnostic theorem.

In general, the event that k is constant with regard to D is equivalent to the event $s(\Omega) = 2^{n-1}$. This implies *diagnostic theorem* stated in **Table 7**.

The diagnostic theorem is the optimal way to validate the diagnostic condition.

The Eq. (30) becomes simple with AND-gate inference. Recall that Eq. (14) specified AND-gate inference as follows

$$P(X_1 \odot X_2 \odot \ldots \odot X_n) = P(Y=1|X_1, X_2, \ldots, X_n) = \begin{cases} \prod_{i=1}^{n} p_i \text{ if all } X_i(s) \text{ are } 1 \\ 0 \text{ if there exists at least one } X_i = 0 \end{cases}$$

Due to only one case $X_1 = X_2 = \ldots = X_n = 1$, we have

$$s(\Omega) = s(\Omega : \{X_i = 1\}) = \prod_{i=1}^{n} p_i$$

Due to $X_i = 0$, we have

$$s(\Omega : \{X_i = 0\}) = 0$$

Derived from Eq. (30), Eq. (31) specifies conditional probability $P(D|X_i)$, posterior probability $P(X_i|D)$, and transformation coefficient according to X-D network with AND-gate reference called *AND-D network*.

$$P(D|X_i=1)=\frac{(2D-M)\prod_{i=1}^{n}p_i+2^{n-1}(M-D)}{2^{n-1}S}$$

$$P(D|X_i=0)=\frac{M-D}{S}$$

$$P(X_i=1|D)=\frac{(2D-M)\prod_{i=1}^{n}p_i+2^{n-1}(M-D)}{(2D-M)\prod_{i=1}^{n}p_i+2^{n}(M-D)} \tag{31}$$

$$P(X_i=0|D)=\frac{2^{n-1}(M-D)}{(2D-M)\prod_{i=1}^{n}p_i+2^{n}(M-D)}$$

$$k=\frac{2^{n-1}S}{(2D-M)\prod_{i=1}^{n}p_i+2^{n}(M-D)}$$

For convenience, we validate diagnostic condition with a case of two sources $\Omega = \{X_1, X_2\}$, $p_1 = p_2 = w_1 = w_2 = 0.5$, $D \in \{0, 1, 2, 3\}$. According to diagnostic theorem stated in **Table 7**, if $s(\Omega) \neq 2$ for given X-gate then, such X-gate does not satisfy diagnostic condition.

Given AND-gate inference, by applying Eq. (14), we have

$$s(\Omega)=(0.5*0.5)+0+0+0=0.25$$

Given OR-gate inference, by applying Eq. (16), we have

$$s(\Omega)=(1-0.5*0.5)+(1-0.5)+(1-0.5)+0=3-3*0.5*0.5=1.75$$

Given XOR-gate inference, by applying Eq. (19), we have

$$s(\Omega)=(0.5*0.5+0.5*0.5)+0.5+0.5+0=1.5$$

Given XNOR-gate inference, by applying Eq. (21), we have

$$s(\Omega)=(0.5*0.5+0.5*0.5)+0.5+0.5+1=2.5$$

Given SIGMA-gate inference, by applying Eq. (23), we have

$$s(\Omega)=(0.5+0.5)+0.5+0.5+0=2$$

It is asserted that AND-gate, OR-gate, XOR-gate, and XNOR-gate do not satisfy diagnostic condition and so they should not be used to assess hypotheses. However, it is not asserted if U-gate and SIGMA-gate satisfy such diagnostic condition. It is necessary to expend equation for SIGMA-gate diagnostic network (called *SIGMA-D network*) in order to validate it.

In case of SIGMA-gate inference, by applying Eq. (23), we have

http://dx.doi.org/10.5772/intechopen.70057

$$\sum_i w_i = 1$$

$$s(\Omega) = 2^{n-1}\sum_i w_i = 2^{n-1}$$

$$s(\Omega : \{X_i = 1\}) = 2^{n-1}w_i + 2^{n-2}\sum_{j\neq i} w_j = 2^{n-1}w_i + 2^{n-2}(1 - w_i) = 2^{n-2}(1 + w_i)$$

$$s(\Omega : \{X_i = 0\}) = s(\Omega) - s(\Omega : \{X_i = 1\}) = 2^{n-2}(1 - w_i)$$

It is necessary to validate SIGMA-D network with SIGMA-gate bi-inference. By applying Eq. (26), we recalculate these quantities as follows

$$s(\Omega) = 2^{n-1}\sum_i w_i + 2^{n-1}\sum_i d_i = 2^{n-1}\sum_i (w_i + d_i) = 2^{n-1}$$

$$\left(\text{due to} \sum_i (w_i + d_i) = 1\right)$$

$$s(\Omega : \{X_i = 1\}) = 2^{n-1}w_i + 2^{n-2}\sum_{j\neq i} w_j + 2^{n-2}\sum_i d_i = 2^{n-2}w_i + 2^{n-2}\sum_i (w_i + d_i) = 2^{n-2}(1 + w_i)$$

$$s(\Omega : \{X_i = 0\}) = s(\Omega) - s(\Omega : \{X_i = 1\}) = 2^{n-2}(1 - w_i)$$

Obviously, quantities $s(\Omega)$, $s(\Omega{:}\{X_i{=}1\})$, and $s(\Omega{:}\{X_i = 0\})$ are kept intact. According to diagnostic theorem, we conclude that SIGMA-D network does satisfy diagnostic condition due to $s(\Omega){=}2^{n-1}$. Thus, SIGMA-D network can be used to assess hypotheses.

Eq. (32), an immediate consequence of Eq. (30), specifies conditional probability $P(D|X_i)$, posterior probability $P(X_i|D)$, and transformation coefficient for SIGMA-D network.

$$\begin{aligned} P(D|X_i = 1) &= \frac{(2D - M)w_i + M}{2S} \\ P(D|X_i = 0) &= \frac{(M - 2D)w_i + M}{2S} \\ P(X_i = 1|D) &= \frac{(2D - M)w_i + M}{2M} \\ P(X_i = 0|D) &= \frac{(M - 2D)w_i + M}{2M} \\ k &= \frac{N}{2} \end{aligned} \tag{32}$$

In case of SIGMA-gate, the augmented variable Y can be removed from X-D network. The evidence D is now established as direct target variable. **Figure 6** shows a so-called *direct SIGMA-gate diagnostic network* (direct SIGMA-D network).

Derived from Eq. (23), the CPT of direct SIGMA-D network is determined by Eq. (33).

$$P(D|X_1, X_2, \ldots, X_n) = \sum_{i \in K} \frac{D}{S} w_i + \sum_{j \in L} \frac{M - D}{S} w_j$$

where the set of X_i (s) is complete and mutually exclusive.

$$\begin{gathered} \sum_{i=1}^{n} w_i = 1 \\ X_i \cap X_j = \varnothing, \forall i \neq j \end{gathered} \tag{33}$$

Eq. (33) specifies valid CPT due to

$$\begin{aligned} \sum_D P(D|X_1, X_2, \ldots, X_n) &= \frac{1}{S} \sum_{i \in K} w_i \sum_D D + \frac{1}{S} \sum_{j \in L} w_j \sum_D (M - D) \\ &= \frac{1}{S} \sum_{i \in K} S w_i + \frac{1}{S} \sum_{j \in L} w_j (NM - S) = \frac{1}{S} \sum_{i \in K} S w_i + \frac{1}{S} \sum_{j \in L} S w_j = \sum_{i=1}^{n} w_i = 1 \end{aligned}$$

From dependencies shown in **Figure 6**, Eq. (34) specifies the joint probability of direct SIGMA-D network.

$$P(X_1, X_2, \ldots, X_n, Y, D) = P(D|X_1, X_2, \ldots, X_n) \prod_n^{i=1} P(X_i) \tag{34}$$

Inferred from Eq. (29), Eq. (35) specifies the joint probability $P(X_i, D)$ and the marginal probability $P(D)$ of direct SIGMA-D network, given uniform distribution of all sources.

$$\begin{gathered} P(X_i, D) = \frac{1}{2^n} s(\Omega : \{X_i\}) \\ P(D) = \frac{1}{2^n} s(\Omega) \end{gathered} \tag{35}$$

where $s(\Omega)$ and $s(\Omega{:}\{X_i\})$ are specified in **Table 2**.

By browsing all variables of direct SIGMA-D network, we have

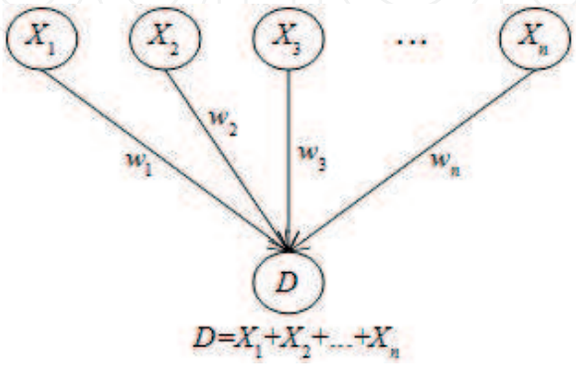

Figure 6. Direct SIGMA-gate diagnostic network (direct SIGMA-D network).

$$s(\Omega:\{X_i=1\}) = 2^{n-1}\frac{D}{S}w_i + 2^{n-2}\sum_{j\neq i}\frac{D}{S}w_j + 2^{n-2}\sum_{j\neq i}\frac{M-D}{S}w_j$$

$$= \frac{2^{n-2}}{S}(2Dw_i + M\sum_{j\neq i}w_j) = \frac{2^{n-2}}{S}\Big(2Dw_i + M(1-w_i)\Big)$$

$$\left(\text{Due to } \sum_{i=1}^{n} w_i = 1\right)$$

$$= \frac{2^{n-2}}{S}\Big((2D-M)w_i + M\Big)$$

Similarly, we have

$$s(\Omega:\{X_i=0\}) = 2^{n-1}\frac{M-D}{S}w_i + 2^{n-2}\sum_{j\neq i}\frac{M-D}{S}w_j + 2^{n-2}\sum_{j\neq i}\frac{D}{S}w_j = \frac{2^{n-2}}{S}\Big((M-2D)w_i + M\Big)$$

$$s(\Omega) = 2^{n-1}\sum_{i}\frac{D}{S}w_i + 2^{n-1}\sum_{i}\frac{M-D}{S}w_i = \frac{2^{n-1}M}{S}$$

By applying Eq. (35), $s(\Omega:\{X_i=0\})$, $s(\Omega:\{X_i=1\})$, and $s(\Omega)$, we get the same result with Eq. (32).

$$P(D|X_i=1) = \frac{(2D-M)w_i + M}{2S}$$

$$P(D|X_i=0) = \frac{(M-2D)w_i + M}{2S}$$

$$P(X_i=1|D) = \frac{(2D-M)w_i + M}{2M}$$

$$P(X_i=0|D) = \frac{(M-2D)w_i + M}{2M}$$

$$k = \frac{N}{2}$$

Therefore, it is possible to use direct SIGMA-D network to assess hypotheses. It is asserted that SIGMA-D network satisfy diagnostic condition when single relationship, NOT-D network, direct SIGMA-D network are specific cases of SIGMA-D network. There is a question: does an X-D network that is different from SIGMA-D network and not aforementioned exist such that it satisfies diagnostic condition?

Recall that each X-D network is a pattern owning a particular X-gate inference which in turn is based on particular X-gate condition (s) relevant to only variables A_is. The most general nonlinear X-D network is U-D network whereas SIGMA-D network is linear one. The U-gate inference given arbitrary condition on U is

$$P(X_1 \uplus X_2 \uplus \ldots \uplus X_n) = \sum_{U \in \mathcal{U}} \left(\prod_{i \in U \cap K} p_i \prod_{i \in U \cap L} (1 - \rho_i) \right) \left(\prod_{i \in \overline{U} \cap K} (1 - p_i) \prod_{i \in \overline{U} \cap L} \rho_i \right)$$

Let f be the arrangement sum of U-gate inference.

$$f(p_i, \rho_i) = \sum_{a(\Omega)} \sum_{U \in \mathcal{U}} \left(\prod_{i \in U \cap K} p_i \prod_{i \in U \cap L} (1 - \rho_i) \right) \left(\prod_{i \in \overline{U} \cap K} (1 - p_i) \prod_{i \in \overline{U} \cap L} \rho_i \right)$$

The function f is sum of many large expressions and each expression is product of four possible sub-products (Π) as follows

$$Expr = \prod_{i \in U \cap K} p_i \prod_{i \in U \cap L} (1 - \rho_i) \prod_{i \in \overline{U} \cap K} (1 - p_i) \prod_{i \in \overline{U} \cap L} \rho_i$$

In any case of degradation, there always exist expression *Expr* (s) having at least 2 sub-products (Π), for example,

$$Expr = \prod_{i \in U \cap K} p_i \prod_{i \in U \cap L} (1 - \rho_i)$$

Consequently, there always exist *Expr* (s) having at least 5 terms relevant to p_i and ρ_i if $n \geq 5$, for example,

$$Expr = p_1 p_2 p_3 (1 - \rho_4)(1 - \rho_5)$$

Thus, degree of f will be larger than or equal to 5 given $n \geq 5$. According to diagnostic theorem, U-gate network satisfies diagnostic condition if and only if $f(p_i, \rho_i) = 2^{n-1}$ for all $n \geq 1$ and for all abstract variables p_i and ρ_i. Without loss of generality, each p_i or ρ_i is sum of variable x and a variable a_i or b_i, respectively. Note that all p_i, ρ_i, a_i are b_i are abstract variables.

$$p_i = x + a_i$$

$$\rho_i = x + b_i$$

The equation $f - 2^{n-1} = 0$ becomes equation $g(x) = 0$ whose degree is $m \geq 5$ if $n \geq 5$.

$$g(x) = \pm x^m + C_1 x^{m-1} + \ldots + C_{m-1} x + C_m - 2^{n-1} = 0$$

where coefficients C_i s are functions of a_i and b_is. According to Abel-Ruffini theorem [11], equation $g(x) = 0$ has no algebraic solution when $m \geq 5$. Thus, abstract variables p_i and ρ_i cannot be eliminated entirely from $g(x) = 0$, which causes that there is no specification of U-gate inference $P(X_1 x X_2 x \ldots x X_n)$ so that diagnostic condition is satisfied.

http://dx.doi.org/10.5772/intechopen.70057

It is concluded that there is no nonlinear X-D network satisfying diagnostic condition, but a new question is raised: does there exist the general linear X-D network satisfying diagnostic condition? Such linear network is called GL-D network and SIGMA-D network is specific case of GL-D network. The GL-gate probability must be linear combination of weights.

$$P(X_1 x X_2 x \ldots x X_n) = C + \sum_{i=1}^{n} \alpha_i w_i + \sum_{i=1}^{n} \beta_i d_i$$

where C is arbitrary constant.

The GL-gate inference is singular if α_i and β_i are functions of only X_i as follows

$$P(X_1 x X_2 x \ldots x X_n) = C + \sum_{i=1}^{n} h_i(X_i) w_i + \sum_{i=1}^{n} g_i(X_i) d_i$$

The functions h_i and g_i are not relevant to A_i because the final equation of GL-gate inference is only relevant to X_i (s) and weights (s). Because GL-D network is a pattern, we only survey singular GL-gate. Mentioned GL-gate is singular by default and it is dependent on how to define functions h_i and g_i. The arrangement sum with regard to GL-gate is

$$s(\Omega) = \sum_{a} \left(C + \sum_{i=1}^{n} h_i(X_i) w_i + \sum_{i=1}^{n} g_i(X_i) d_i \right)$$

$$= 2^n C + 2^{n-1} \sum_{i=1}^{n} \Big(h_i(X_i = 1) + h_i(X_i = 0) \Big) w_i + 2^{n-1} \sum_{i=1}^{n} \Big(g_i(X_i = 1) + g_i(X_i = 0) \Big) d_i$$

Suppose h_i and g_i are probability mass functions with regard to X_i. For all i, we have

$$0 \le h_i(X_i) \le 1$$

$$0 \le g_i(X_i) \le 1$$

$$h_i(X_i = 1) + h_i(X_i = 0) = 1$$

$$g_i(X_i = 1) + g_i(X_i = 0) = 1$$

The arrangement sum becomes

$$s(\Omega) = 2^n C + 2^{n-1} \sum_{i=1}^{n} (w_i + d_i)$$

GL-D network satisfies diagnostic condition if

$$s(\Omega) = 2^n C + 2^{n-1}\sum_{i=1}^{n}(w_i + d_i) = 2^{n-1}$$
$$\Rightarrow 2C + \sum_{i=1}^{n}(w_i + d_i) = 1$$

Suppose the set of X_is is complete.

$$\sum_{i=1}^{n}(w_i + d_i) = 1$$

This implies C = 0. Shortly, Eq. (36) specifies the singular GL-gate inference so that GL-D network satisfies diagnostic condition.

$$P(X_1 x X_2 x \ldots x X_n) = \sum_{i=1}^{n} h_i(X_i) w_i + \sum_{i=1}^{n} g_i(X_i) d_i$$
where h_i and g_i are probability mass functions and the set of X_i(s) is complete. (36)
$$\sum_{i=1}^{n} W_i = 1$$

Functions $h_i(X_i)$ and $g_i(X_i)$ are always linear due to $X_i^m = X_i$ for all $m \geq 1$ when X_i is binary. It is easy to infer that SIGMA-D network is GL-D network with following definition of functions h_i and g_i.

$$h_i(X_i) = 1 - g_i(X_i) = X_i, \forall i$$

According to Millán and Pérez-de-la-Cruz [4], a hypothesis can have multiple evidences as seen in **Figure 7**. This is *multi-evidence diagnostic relationship* opposite to aforementioned multi-hypothesis diagnostic relationship.

Figure 7 depicts the multi-evidence diagnostic network called M-E-D network in which there are m evidences $D_1, D_2, \ldots, D_m$ and one hypothesis Y. Note that Y has uniform distribution.

In simplest case where all evidences are binary, the joint probability of M-E-D network is

$$P(Y, D_1, D_2, \ldots, D_m) = P(Y)\prod_{j=1}^{m} P(D_j|Y) = P(Y)P(D_1, D_2, \ldots, D_m|Y)$$

The product $\prod_{j=1}^{m} P(D_j|Y)$ is denoted as likelihood function as follows

$$P(D_1, D_2, \ldots, D_m|Y) = \prod_{j=1}^{m} P(D_j|Y)$$

The posterior probability $P(Y \mid D_1, D_2, \ldots, D_m)$ given uniform distribution of Y is

http://dx.doi.org/10.5772/intechopen.70057

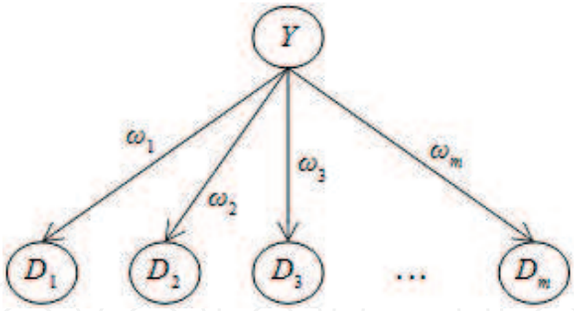

Figure 7. Diagnostic relationship with multiple evidences (M-E-D network).

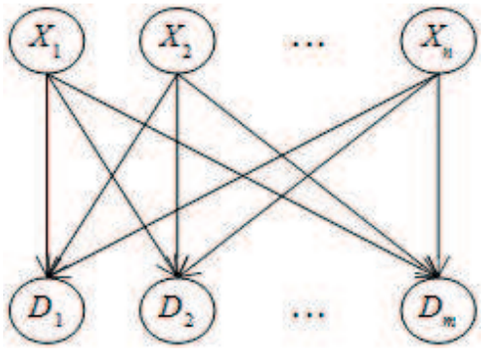

Figure 8. M-HE-D network.

$$P(Y|D_1, D_2, \ldots, D_m) = \frac{P(Y, D_1, D_2, \ldots, D_m)}{P(Y=1, D_1, D_2, \ldots, D_m) + P(Y=0, D_1, D_2, \ldots, D_m)}$$
$$= \frac{1}{\prod_{j=1}^{m} P(D_j|Y=1) + \prod_{j=1}^{m} P(D_j|Y=0)} * P(D_1, D_2, \ldots, D_m|Y)$$

The possible transformation coefficient is

$$\frac{1}{k} = \prod_{j=1}^{m} P(D_j|Y=1) + \prod_{j=1}^{m} P(D_j|Y=0)$$

M-E-D network will satisfy diagnostic condition if $k = 1$ because all hypotheses and evidence are binary, which leads that following equation specified by Eq. (37) has $2m$ real roots $P(D_j|Y)$ for all $m \geq 2$.

$$\prod_{j=1}^{m} P(D_j|Y=1) + \prod_{j=1}^{m} P(D_j|Y=0) = 1 \tag{37}$$

Eq. (37) has no real root given $m = 2$ according to following proof. Suppose Eq. (37) has 4 real roots as follows

$$a_1 = P(D_1 = 1|Y = 1)$$

$$a_2 = P(D_2 = 1|Y = 1)$$

$$b_1 = P(D_1 = 1|Y = 0)$$

$$b_2 = P(D_2 = 1|Y = 0)$$

From Eq. (37), it holds

$$\begin{cases} a_1a_2 + b_1b_2 = 1 \\ a_1(1-a_2) + b_1b_2 = 1 \\ (1-a_1)a_2 + b_1b_2 = 1 \\ a_1a_2 + b_1(1-b_2) = 1 \\ a_1a_2 + (1-b_1)b_2 = 1 \end{cases} \Rightarrow \begin{cases} a_1 = a_2 \\ b_1 = b_2 \\ a_1^2 + b_1^2 = 1 \\ a_1 + 2b_1^2 = 2 \\ b_1 + 2a_1^2 = 2 \end{cases} \Leftrightarrow \begin{cases} a_1 = a_2 = 0 \\ b_1 = b_2 \\ a_1^2 + b_1^2 = 1 \\ b_1 = 2 \end{cases} \text{ or } \begin{cases} a_1 = a_2 = 0.5 \\ b_1 = b_2 \\ a_1^2 + b_1^2 = 1 \\ b_1 = 1.5 \end{cases}$$

The final equation leads a contradiction (b_1 = 2 or b_1 = 1.5) and so it is impossible to apply the sufficient diagnostic proposition into M-E-D network. Such proposition is only used for one-evidence network. Moreover, X-gate inference absorbs many sources and then produces out of one targeted result whereas the M-E-D network essentially splits one source into many results. It is impossible to model M-E-D network by X-gates. The potential solution for this problem is to group many evidences D_1, D_2,…, D_m into one representative evidence D which in turn is dependent on hypothesis Y but this solution will be inaccurate in specifying conditional probabilities because directions of dependencies become inconsistent (relationships from D_j to D and from Y to D) except that all D_js are removed and D becomes a vector. However, evidence vector does not simplify the hazardous problem and it changes the current problem into a new problem.

Another solution is to reverse the direction of relationship, in which the hypothesis is dependent on evidences so as to take advantages of X-gate inference as usual. However, the reversion method violates the viewpoint in this research where diagnostic relationship must be from hypothesis to evidence. In other words, we should change the viewpoint.

Another solution is based on a so-called *partial diagnostic condition* that is a loose case of diagnostic condition for M-E-D network, which is defined as follows

$$P(Y|D_j) = kP(D_j|Y)$$

where k is constant with regard to D_j. The joint probability is

$$P(Y, D_1, D_2, \ldots, D_m) = P(Y)\prod_{j=1}^{m} P(D_j|Y)$$

M-E-D network satisfies partial diagnostic condition. In fact, given all variables are binary, we have

$$P(Y|D_j) = \frac{\sum_{\Psi \setminus \{Y, D_j\}} P(Y, D_1, D_2, \ldots, D_m)}{\sum_{\Psi \setminus \{D_j\}} P(Y, D_1, D_2, \ldots, D_m)}$$

(Let $\Psi = \{D_1, D_2, \ldots, D_m\}$)

$$= \frac{P(D_j|Y)\prod_{k=1, k\neq j}^{m}\left(\sum_{D_k} P(D_k|Y)\right)}{\prod_{k=1, k\neq j}^{m}\left(\sum_{D_k} P(D_k|Y=1)\right) + \prod_{k=1, k\neq j}^{m}\left(\sum_{D_k} P(D_k|Y=0)\right)}$$

(Due to uniform distribution of Y)

$$= \frac{P(D_j|Y)\prod_{k=1, k\neq j}^{m} 1}{\prod_{k=1, k\neq j}^{m} 1 + \prod_{k=1, k\neq j}^{m} 1} = \frac{1}{2}P(D_j|Y)$$

$$\left(\text{Due to } \sum_{D_k} P(D_k|Y) = P(D_k = 0|Y) + P(D_k = 1|Y) = 1\right)$$

Partial diagnostic condition expresses a different viewpoint. It is not an optimal solution because we cannot test a disease based on only one symptom while ignoring other obvious symptoms, for example. The equality $P(Y \mid D_j) = 0.5P(D_j \mid Y)$ indicates the accuracy is decreased two times. However, Bayesian network provides inference mechanism based on personal belief. It is subjective. You can use partial diagnostic condition if you think that such condition is appropriate to your application.

If we are successful in specifying conditional probabilities of M-E-D network, it is possible to define an extended network which is constituted of n hypotheses $X_1, X_2, \ldots, X_n$ and m evidences $D_1, D_2, \ldots, D_m$. Such extended network represents *multi-hypothesis multi-evidence diagnostic relationship*, called M-HE-D network. **Figure 8** depicts M-HE-D network.

The M-HE-D network is the most general case of diagnostic network, which was mentioned in Ref. ([4], p. 297). We can construct any large diagnostic BN from M-HE-D networks and so the research is still open.

5. Conclusion

In short, relationship conversion is to determine conditional probabilities based on logic gates that are adhered to semantics of relationships. The weak point of logic gates is to require that all variables must be binary. For example, in learning context, it is inconvenient for expert to create an assessment BN with studying exercises (evidences) whose marks are only 0 and 1. In order to lessen the impact of such weak point, the numeric evidence is used for extending capacity of simple Bayesian network. However, combination of binary hypothesis and

numeric evidence leads to errors or biases in inference. For example, given a student gets maximum grade for an exercise but the built-in inference results out that she/he has not mastered fully the associated learning concept (hypothesis). Therefore, I propose the sufficient diagnostic proposition so as to confirm that numeric evidence is adequate to make complicated inference tasks in BN. The probabilistic reasoning based on evidence is always accurate. Application of the research can go beyond learning context whenever probabilistic deduction relevant to constraints of semantic relationships is required. A large BN can be constituted of many simple BN (s). Inference in large BN is hazardous problem and there are many optimal algorithms for solving such problem. In future, I will research effective inference methods for the special BN that is constituted of X-gate BN (s) mentioned in this research because X-gate BN (s) have precise and useful features of which we should take advantages. For instance, their CPT (s) are simple in some cases and the meanings of their relationships are mandatory in many applications. Moreover, I try my best to research deeply M-E-D network and M-HE-D network whose problems I cannot solve absolutely now.

Two main documents that I referred to do this research are the book "Learning Bayesian Networks" [2] by the author Richard E. Neapolitan and the article "A Bayesian Diagnostic Algorithm for Student Modeling and its Evaluation" [4] by authors Eva Millán and José Luis Pérez-de-la-Cruz. Especially, the SIGMA-gate inference is based on and derived from the work of the Eva Millán and José Luis Pérez-de-la-Cruz. This research is originated from my PhD research "A User Modeling System for Adaptive Learning" [12]. Other references relevant to user modeling, overlay model, and Bayesian network are [13–16]. Please concern these references.

Appendices

A1. Following is the proof of Eq. (9)

$$\begin{gathered} P(A_i = ON|X_i) \\ = P(A_i = ON|X_i, I_i = ON)P(I_i = ON) + P(A_i = ON|X_i, I_i = OFF)P(I_i = OFF) \\ = 0 * (1 - p_i) + P(A_i = ON|X_i, I_i = OFF)p_i \\ \text{(By applying Eq. (8))} \\ = p_i P(A_i = ON|X_i, I_i = OFF) \end{gathered}$$

It implies

$$P(A_i = ON|X_i = 1) = p_i P(A_i = ON|X_i = 1, I_i = OFF) = p_i$$

$$P(A_i = ON|X_i = 0) = p_i P(A_i = ON|X_i = 0, I_i = OFF) = 0$$

$$P(A_i = OFF|X_i = 1) = 1 - P(A_i = ON|X_i = 1) = 1 - p_i$$

$$P(A_i = OFF|X_i = 0) = 1 - P(A_i = ON|X_i = 0) = 1 \blacksquare$$

A2. Following is the proof of Eq. (10)

$$P(Y|X_1, X_2, \ldots, X_n) = \frac{P(Y, X_1, X_2, \ldots, X_n)}{P(X_1, X_2, \ldots, X_n)}$$

(Due to Bayes' rule)

$$= \frac{\sum_{A_1, A_2, \ldots, A_n} P(Y, X_1, X_2, \ldots, X_n|A_1, A_2, \ldots, A_n) * P(A_1, A_2, \ldots, A_n)}{P(X_1, X_2, \ldots, X_n)}$$

(Due to total probability rule)

$$= \sum_{A_1, A_2, \ldots, A_n} P(Y, X_1, X_2, \ldots, X_n|A_1, A_2, \ldots, A_n) * \frac{P(A_1, A_2, \ldots, A_n)}{P(X_1, X_2, \ldots, X_n)}$$

$$= \sum_{A_1, A_2, \ldots, A_n} P(Y|A_1, A_2, \ldots, A_n) * P(X_1, X_2, \ldots, X_n|A_1, A_2, \ldots, A_n) * \frac{P(A_1, A_2, \ldots, A_n)}{P(X_1, X_2, \ldots, X_n)}$$

(Because Y is conditionally independent from X_is given A_is)

$$= \sum_{A_1, A_2, \ldots, A_n} P(Y|A_1, A_2, \ldots, A_n) * \frac{P(X_1, X_2, \ldots, X_n, A_1, A_2, \ldots, A_n)}{P(X_1, X_2, \ldots, X_n)}$$

$$= \sum_{A_1, A_2, \ldots, A_n} P(Y|A_1, A_2, \ldots, A_n) * P(A_1, A_2, \ldots, A_n|X_1, X_2, \ldots, X_n)$$

(Due to Bayes' rule)

$$= \sum_{A_1, A_2, \ldots, A_n} P(Y|A_1, A_2, \ldots, A_n) \prod_{i=1}^{n} P(A_i|X_1, X_2, \ldots, X_n)$$

(Because A_is are mutually independent)

$$= \sum_{A_1, A_2, \ldots, A_n} P(Y|A_1, A_2, \ldots, A_n) \prod_{i=1}^{n} P(A_i|X_i)$$

(Because each A_i is only dependent on X_i) ■

A3. Following is the proof that the augmented X-D network (shown in **Figure 5**) is equivalent to X-D network (shown in shown in **Figures 2** and **3**) with regard to variables X_1, X_2,…, X_n, and D.

The joint probability of augmented X-D network shown in **Figure 5** is

$$P(X_1, X_2, \ldots, X_n, Y, D) = P(D|Y)P(Y|X_1, X_2, \ldots, X_n) \prod_{i=1}^{n} P(X_i)$$

The joint probability of X-D network is

$$P(X_1, X_2, \ldots, X_n, D) = P(D|X_1, X_2, \ldots, X_n)\prod_{i=1}^{n} P(X_i)$$

By applying total probability rule into X-D network, we have

$$P(X_1, X_2, \ldots, X_n, D) = \frac{P(D, X_1, X_2, \ldots, X_n)}{P(X_1, X_2, \ldots, X_n)}\prod_{i=1}^{n} P(X_i)$$

(Due to Bayes' rule)

$$= \frac{\sum_Y P(D, X_1, X_2, \ldots, X_n|Y)P(Y)}{P(X_1, X_2, \ldots, X_n)}\prod_{i=1}^{n} P(X_i)$$

(Due to total probability rule)

$$= \frac{\sum_Y P(D, X_1, X_2, \ldots, X_n|Y)P(Y)}{P(X_1, X_2, \ldots, X_n)}\prod_{i=1}^{n} P(X_i)$$

$$= \left(\sum_Y P(D, X_1, X_2, \ldots, X_n|Y) * \frac{P(Y)}{P(X_1, X_2, \ldots, X_n)}\right) * \prod_{i=1}^{n} P(X_i)$$

$$= \left(\sum_Y P(D|Y) * \frac{P(X_1, X_2, \ldots, X_n|Y)P(Y)}{P(X_1, X_2, \ldots, X_n)}\right) * \prod_{i=1}^{n} P(X_i)$$

(Because D is conditionally independent from all X_i (s) given Y)

$$= \left(\sum_Y P(D|Y) * \frac{P(Y, X_1, X_2, \ldots, X_n)}{P(X_1, X_2, \ldots, X_n)}\right) * \prod_{i=1}^{n} P(X_i)$$

$$= \sum_Y P(D|Y)P(Y|X_1, X_2, \ldots, X_n)\prod_{i=1}^{n} P(X_i)$$

(Due to Bayes' rule)

$$= \sum_Y P(X_1, X_2, \ldots, X_n, Y, D) \blacksquare$$

A4. Following is the proof of Eq. (29)

Given uniform distribution of X_i (s), we have

$$P(X_1) = P(X_2) = \cdots = P(X_n) = \frac{1}{2}$$

The joint probability becomes

$$P(\Omega, Y, D) = \frac{1}{2^n}P(Y|X_1, X_2, \ldots, X_n)P(D|Y)$$

The joint probability of X_i and D is

$$
\begin{aligned}
P(X_i, D) &= \sum_{\{\Omega, Y, D\}\setminus\{X_i, D\}} P(\Omega, Y, D) \\
&= P(X_1 = 1, X_2 = 1, \ldots, X_i, \ldots, X_{n-1} = 1, X_n = 1, Y = 1, D) \\
&+ P(X_1 = 1, X_2 = 1, \ldots, X_i, \ldots, X_{n-1} = 1, X_n = 0, Y = 1, D) + \cdots \\
&+ P(X_1 = 0, X_2 = 0, \ldots, X_i, \ldots, X_{n-1} = 0, X_n = 1, Y = 1, D) \\
&+ P(X_1 = 0, X_2 = 0, \ldots, X_i, \ldots, X_{n-1} = 0, X_n = 0, Y = 1, D) \\
&+ P(X_1 = 1, X_2 = 1, \ldots, X_i, \ldots, X_{n-1} = 1, X_n = 1, Y = 0, D) \\
&+ P(X_1 = 1, X_2 = 1, \ldots, X_i, \ldots, X_{n-1} = 1, X_n = 0, Y = 0, D) + \cdots \\
&+ P(X_1 = 0, X_2 = 0, \ldots, X_i, \ldots, X_{n-1} = 0, X_n = 1, Y = 0, D) \\
&+ P(X_1 = 0, X_2 = 0, \ldots, X_i, \ldots, X_{n-1} = 0, X_n = 0, Y = 0, D) \\
&= \frac{1}{2^n}\frac{D}{S}\Big(P(Y = 1|X_1 = 1, X_2 = 1, \ldots, X_i, \ldots, X_{n-1} = 1, X_n = 1) + P(Y = 1|X_1 = 1, X_2 \\
&= 1, \ldots, X_i, \ldots, X_{n-1} = 1, X_n = 0) + \cdots + P(Y = 1|X_1 = 1, X_2 = 1, \ldots, X_i, \ldots, X_{n-1} = 0, X_n = 1) \\
&+ P(Y = 1|X_1 = 1, X_2 = 1, \ldots, X_i, \ldots, X_{n-1} = 0, X_n = 0)\Big) \\
&+ \frac{1}{2^n}\frac{M-D}{S}\Big(P(Y = 0|X_1 = 1, X_2 = 1, \ldots, X_i, \ldots, X_{n-1} = 1, X_n = 1) + P(Y = 0|X_1 = 1, X_2 \\
&= 1, \ldots, X_i, \ldots, X_{n-1} = 1, X_n = 0) + \cdots + P(Y = 0|X_1 = 1, X_2 = 1, \ldots, X_i, \ldots, X_{n-1} = 0, X_n = 1) \\
&+ P(Y = 0|X_1 = 1, X_2 = 1, \ldots, X_i, \ldots, X_{n-1} = 0, X_n = 0)\Big)
\end{aligned}
$$

(Due to Eq. (6))

The marginal probability of D is

$$
\begin{aligned}
P(D) &= \sum_{\{\Omega, Y, D\}\setminus\{D\}} P(\Omega, Y, D) \\
&= P(X_1 = 1, X_2 = 1, \ldots, X_n = 1, Y = 1, D) + P(X_1 = 1, X_2 = 1, \ldots, X_n = 0, Y = 1, D) + \cdots \\
&+ P(X_1 = 0, X_2 = 0, \ldots, X_n = 1, Y = 1, D) + P(X_1 = 0, X_2 = 0, \ldots, X_n = 0, Y = 1, D) \\
&+ P(X_1 = 1, X_2 = 1, \ldots, X_n = 1, Y = 0, D) \\
&= \frac{1}{2^n}\frac{D}{S}\Big(P(Y = 1|X_1 = 1, X_2 = 1, \ldots, X_n = 1) + P(Y = 1|X_1 = 1, X_2 = 1, \ldots, X_n = 0) + \cdots \\
&+ P(Y = 1|X_1 = 1, X_2 = 1, \ldots, X_n = 1) + P(Y = 1|X_1 = 1, X_2 = 1, \ldots, X_n = 0)\Big) \\
&+ \frac{1}{2^n}\frac{M-D}{S}\Big(P(Y = 0|X_1 = 1, X_2 = 1, \ldots, X_n = 1) + P(Y = 0|X_1 = 1, X_2 = 1, \ldots, X_n = 0) + \cdots \\
&+ P(Y = 0|X_1 = 1, X_2 = 1, \ldots, X_n = 1) + P(Y = 0|X_1 = 1, X_2 = 1, \ldots, X_n = 0)\Big) \\
&+ P(X_1 = 1, X_2 = 1, \ldots, X_n = 0, Y = 0, D) + \cdots
\end{aligned}
$$

By applying **Table 2**, the joint probability $P(X_i, D)$ is determined as follows

$$P(X_i, D) = \frac{1}{2^n S}\left(D\sum_a P\Big(Y = 1|a(\Omega : \{X_i\})\Big) + (M - D)\sum_a P\Big(Y = 0|a(\Omega : \{X_i\})\Big)\right)$$
$$= \frac{1}{2^n S}\left(D\sum_a P\Big(Y = 1|a(\Omega : \{X_i\})\Big) + (M - D)\sum_a \Big(1 - P\Big(Y = 1|a(\Omega : \{X_i\})\Big)\Big)\right)$$
$$= \frac{1}{2^n S}\Big((2D - M)s(\Omega : \{X_i\}) + 2^{n-1}(M - D)\Big)$$

Similarly, the marginal probability $P(D)$ is

$$P(D) = \frac{1}{2^n S}\Big((2D - M)s(\Omega) + 2^n(M - D)\Big) \blacksquare$$

Author details

Loc Nguyen

Address all correspondence to: ng_phloc@yahoo.com

Sunflower Soft Company, An Giang, Vietnam

References

[1] Wikipedia. Logic gate. Wikimedia Foundation [Internet]. 2016. [Online]. Available from: https://en.wikipedia.org/wiki/Logic_gate [Accessed June 4, 2016]

[2] Neapolitan RE. Learning Bayesian Networks. Upper Saddle River, New Jersey: Prentice Hall; 2003. p. 674

[3] Díez FJ, Druzdzel MJ. Canonical Probabilistic Models. Madrid: Research Centre on Intelligent Decision-Support Systems; 2007

[4] Millán E, Pérez-de-la-Cruz JL. A bayesian diagnostic algorithm for student modeling and its evaluation. User Modeling and User-Adapted Interaction. 2002;**12**(2-3):281–330

[5] Wikipedia. Factor graph. Wikimedia Foundation [Internet]. 2015. [Online]. Available from: https://en.wikipedia.org/wiki/Factor_graph [Accessed: February 8, 2017]

[6] Kschischang FR, Frey BJ, Loeliger HA. Factor graphs and the sum-product algorithm. IEEE Transactions on Information Theory. 2001;**47**(2):498–519

[7] Pearl J. Fusion, propagation, and structuring in belief networks. Artificial Intelligence. 1986;**29**(3):241–288

[8] Millán E, Loboda T, Pérez-de-la-Cruz JL. Bayesian networks for student model engineering. Computers & Education. 2010:**55**(4):1663–1683

[9] Nguyen L. Theorem of SIGMA-gate inference in Bayesian network. Wulfenia Journal. 2016;**23**(3):280–289

[10] Wikipedia. Set (mathematics), Wikimedia Foundation [Internet]. 2014. [Online]. Available from: http://en.wikipedia.org/wiki/Set_(mathematics) [Accessed: October 11, 2014]

[11] Wikipedia. Abel-Ruffini theorem. Wikimedia Foundation [Internet]. 2016. [Online]. Available from: https://en.wikipedia.org/wiki/Abel%E2%80%93Ruffini_theorem [Accessed: June 26, 2016]

[12] Nguyen L. A User Modeling System for Adaptive Learning. Abuja, Nigeria: Standard Research Journals; 2014

[13] Fröschl C. User modeling and user profiling in adaptive E-learning systems [master thesis]. Graz, Austria: Graz University of Technology; 2005

[14] De Bra P, Smits D, Stash N. The Design of AHA!. In Proceedings of the Seventeenth ACM Hypertext Conference on Hypertext and hypermedia (Hypertext '06); 22-25 August 2006; Odense, Denmark. New York, NY: ACM; 2006. pp. 133–134

[15] Murphy KP. A Brief Introduction to Graphical Models and Bayesian Networks. University of British Columbia; 1998. [Online]. Available from: http://www.cs.ubc.ca/~murphyk/Bayes/bnintro.html [Accessed: 2008]

[16] Heckerman D. A Tutorial on Learning With Bayesian Networks. Redmond: Microsoft Research; 1995

IntechOpen