# Converting Graphic Relationships into Conditional Probabilities in Bayesian Network

Prof. Loc Nguyen PhD, MD, MBA
An Giang University, Vietnam
Email: ng_phloc@yahoo.com
Homepage: www.locnguyen.net

# Abstract

Bayesian network (BN) is a powerful mathematical tool for prediction and diagnosis applications. A large BN can be constituted of 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 BN 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 is essential feature of the research.

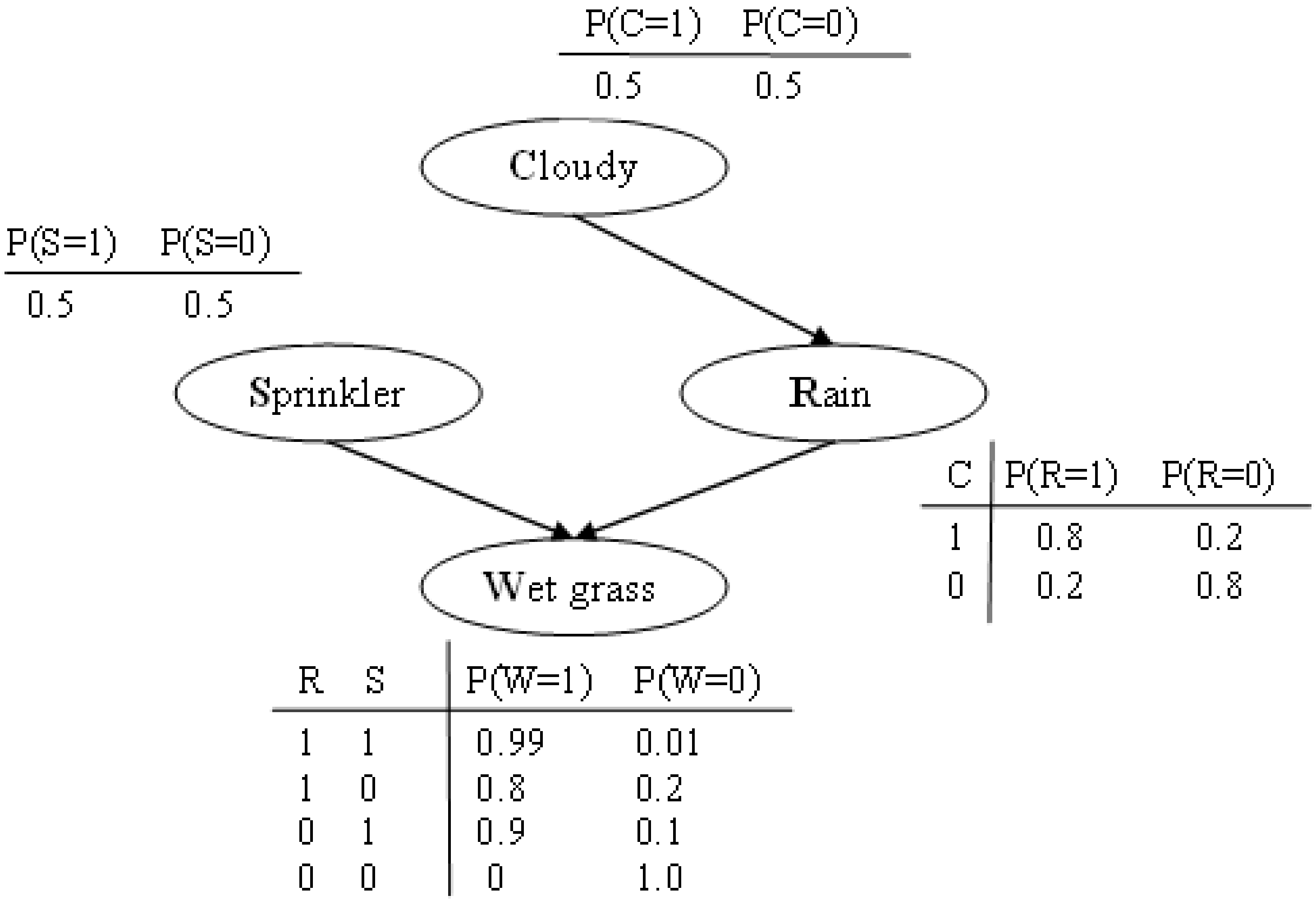
# Table of contents

# 1. Introduction

- In general, Bayesian network (BN) is a directed acyclic graph (DAG) consisting 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 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:
  - 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 CPT (s). **This research focuses on such second approach**.

# 1. Introduction

- **An example of BN**, event "cloudy" is cause of event "rain" which in turn is cause of "grass is wet" (Murphy, 1998).
- Binary random variables (nodes) *C*, *S*, *R*, and *W* denote events such as cloudy, sprinkler, rain, and wet grass. Each node is associated with a CPT.
- Given "wet grass" evidence *W*=1, due to $P(S=1|W=1) = 0.7 > P(R=1|W=1) = 0.67$, which leads to conclusion that sprinkler is the most likely cause of wet grass.



# 1. Introduction

- *Relationship conversion* aims to determined conditional probabilities based on weights and meanings of relationships. Especially, these relationships are adhered to logic X-gates (Wikipedia, 2016) 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" by author (Neapolitan, 2003, pp. 157-159).
- Authors (Díez & Druzdzel, 2007) also researched OR/MAX, AND/MIN, noisy XOR inferences but they focused on canonical models, deterministic models, ICI models whereas I focus on logic gate and graphic relationships.
- Factor graph (Wikipedia, Factor graph, 2015) represents factorization of a global function into many partial functions. Pearl's propagation algorithm (Pearl, 1986), a variant of factor graph, is very successful in BN inference. I did not use factor graph for constructing BN. The concept "X-gate inference" only implies how to convert simple graph into BN.

# 1. Introduction

- In general, my research focuses on applied probability adhered to Bayesian network, logic gates, and Bayesian user modeling (Millán & Pérez-de-la-Cruz, 2002). The scientific results are shared with authors **Eva Millán** and **José Luis Pérez-de-la-Cruz**.
- 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.
- *Diagnostic relationship* is very important to Bayesian evaluation in learning context because it is used to evaluate student's mastery of concepts. Now we start relationship conversion with a research on diagnostic relationship in the next section.

# 2. Diagnostic relationship

- In some opinions (Millán, Loboda, & Pérez-de-la-Cruz, 2010, p. 1666) like mine, the *diagnostic relationship* should be from hypothesis to evidence. For example, disease is hypothesis *X* and symptom is evidence *D*.

$$X \xrightarrow{1} D$$

- The weight *w* of relationship between *X* and *D* is 1. Formula 2.1 specifies CPT of *D* when *D* is binary (0 and 1) for converting simplest diagnostic relationship to simplest BN.

$$P(D|X) = \begin{cases} D \text{ if } X = 1 \\ 1 - D \text{ if } X = 0 \end{cases} \quad (2.1)$$

- Evidence *D* can be used to diagnose hypothesis *X* if the so-called sufficient diagnostic proposition is satisfied. Such proposition is called shortly **diagnostic condition** stated that "*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*".

# 2. Diagnostic relationship

- I survey three other common cases of evidence $D$ such as $D \in \{0, 1, 2, \dots, \eta\}$, $D \in \{a, a+1, a+2, \dots, b\}$, and $D \in [a, b]$.
- In general, formula 2.6 summarizes CPT of evidence of single diagnosis relationship, satisfying **diagnostic condition**.

$$P(D|X) = \begin{cases} \dfrac{D}{S} \text{ if } X = 1 \\ \dfrac{M}{S} - \dfrac{D}{S} \text{ if } X = 0 \end{cases} \quad (2.6)$$

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

Where

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

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

# 2. Diagnostic relationship

As an example, we prove that the CPT of evidence $D \in \{0, 1, 2, \ldots, \eta\}$, a special case of formula 2.6, satisfies diagnostic condition.

$$P(D|X) = \begin{cases} \frac{D}{S} \text{ if } X = 1 \\ \frac{\eta}{S} - \frac{D}{S} \text{ if } X = 0 \end{cases}$$ where $D \in \{0,1,2,\ldots,\eta\}$ and $S = \sum_{D=0}^{n} D = \frac{\eta(\eta+1)}{2}$

In fact, we have:

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

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

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

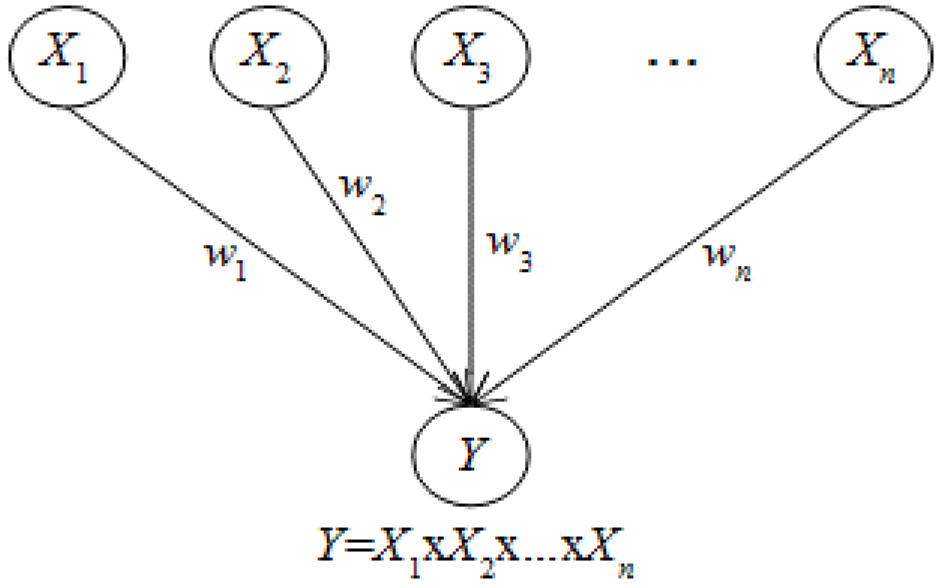
Suppose the prior probability of $X$ is uniform: $P(X=0) = P(X=1)$. We have:

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

So, the transformation coefficient $k$ is $\frac{\eta+1}{2}$.

# 3. X-gate inferences

- The diagnostic relationship is now extended with more than one hypothesis. Given a *simple graph* consisting of one child variable $Y$ and $n$ parent variables $X_i$. Each relationship from $X_i$ to $Y$ is quantified by a normalized weight $w_i$ where $0 \leq w_i \leq 1$.
- Now we convert graphic relationships of simple graph into CPT (s) of simple BN. These relationships are adhere to X-gates such as AND-gate, OR-gate, and SIGMA-gate. Relationship conversion is to determine *X-gate inference*. The simple graph is also called *X-gate graph* or *X-gate network.*

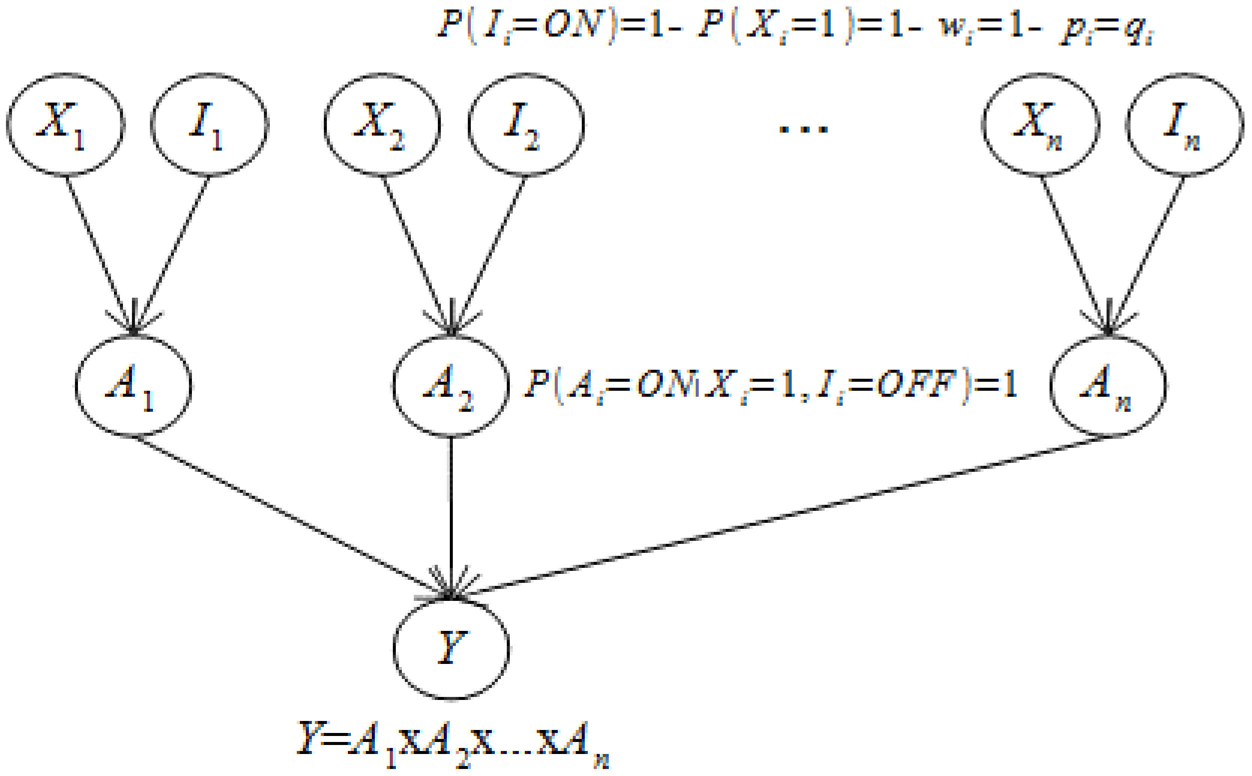



**X-gate network**

# 3. X-gate inferences

X-gate inference is based on three following assumptions mentioned in (Neapolitan, 2003, p. 157)

- *X-gate inhibition*: Given a relationship from source $X_i$ to target $Y$, there is a factor $I_i$ that inhibits $X_i$ from integrated into $Y$.
- *Inhibition independence*: Inhibitions are mutually independent.
- *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 $A_i$ (s).



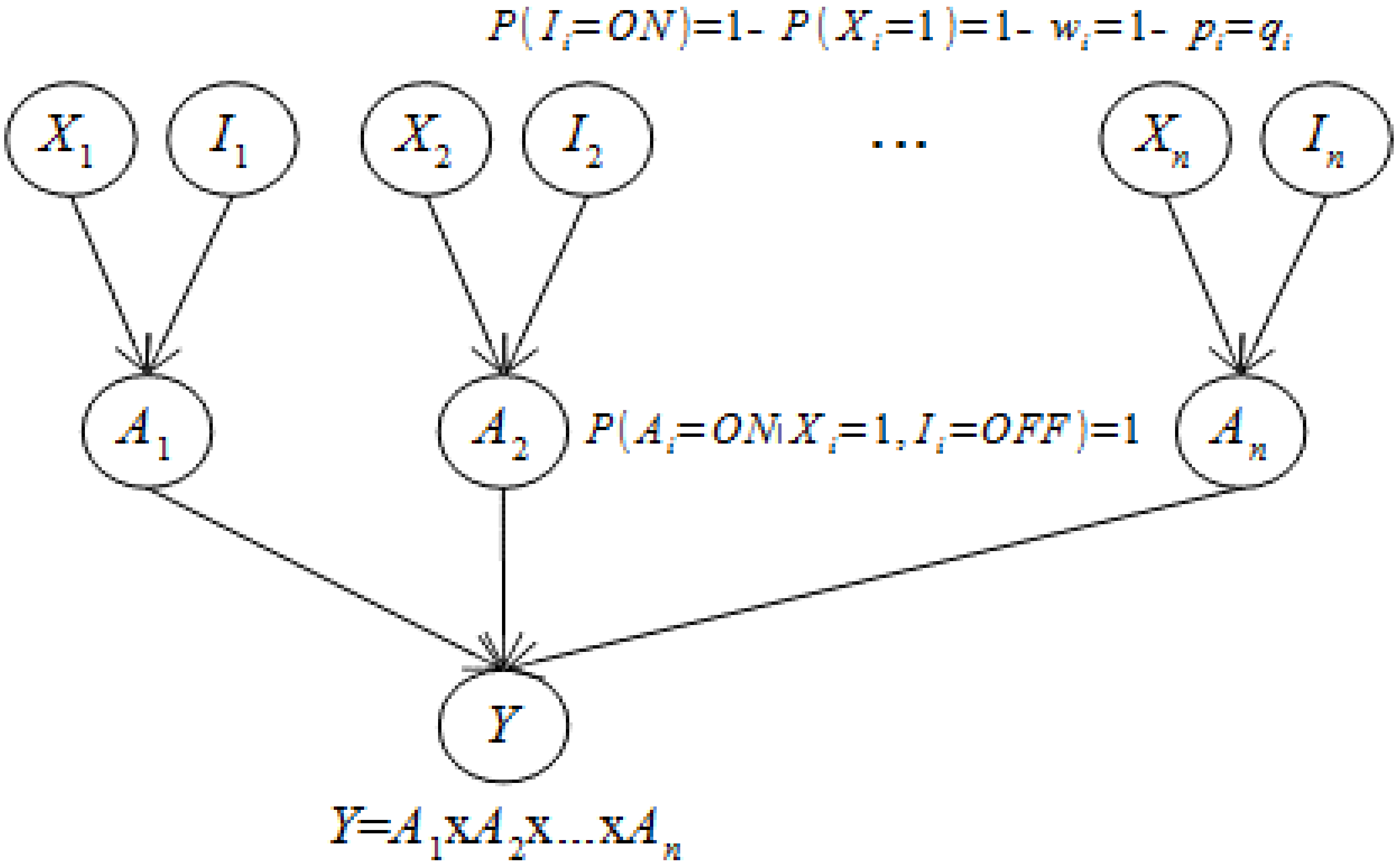

**extended X-gate network with accountable variables**

# 3. X-gate inferences

**Probabilities of inhibitions $I_i$ (s) and accountable variables $A_i$ with formulas 3.2, 3.3**

$P(I_i = OFF) = p_i = w_i$
$P(I_i = ON) = 1 - p_i = 1 - w_i$
$P(A_i = ON|X_i = 1, I_i = OFF) = 1$
$P(A_i = ON|X_i = 1, I_i = ON) = 0$
$P(A_i = ON|X_i = 0, I_i = OFF) = 0$
$P(A_i = ON|X_i = 0, I_i = ON) = 0$
$P(A_i = OFF|X_i = 1, I_i = OFF) = 0$
$P(A_i = OFF|X_i = 1, I_i = ON) = 1$
$P(A_i = OFF|X_i = 0, I_i = OFF) = 1$
$P(A_i = OFF|X_i = 0, I_i = ON) = 1$

$P(A_i = ON|X_i = 1) = p_i = w_i$
$P(A_i = ON|X_i = 0) = 0$
$P(A_i = OFF|X_i = 1) = 1 - p_i = 1 - w_i$
$P(A_i = OFF|X_i = 0) = 1$



# 3. X-gate inferences

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

(Due to Bayes' rule)

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

(Due to total probability rule)

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

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

(Because $Y$ is conditionally independent from $X_i$ (s) given $A_i$ (s))

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

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

(Due to Bayes' rule)

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

(Because $A_i$ (s) are mutually independent)

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

(Because each $A_i$ is only dependent on $X_i$)

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

**X-gate probability (3.4)**

The X-gate inference is represented by X-gate probability $P(Y=1 \mid X_1, X_2, \dots, X_n)$ specified by (Neapolitan, 2003, p. 159)

# 3. X-gate inferences

- Given $\Omega = \{X_1, X_2, \ldots, X_n\}$ where $|\Omega|=n$ is cardinality of $\Omega$.
- Let $a(\Omega)$ be an arrangement of $\Omega$ 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: $a(\Omega) = \{X_1=1, X_2=1\}$, $a(\Omega) = \{X_1=1, X_2=0\}$, $a(\Omega) = \{X_1=0, X_2=1\}$, $a(\Omega) = \{X_1=0, X_2=0\}$. Let $a(\Omega:\{X_i\})$ be the arrangement of $\Omega$ with fixed $X_i$.
- Let $c(\Omega)$ and $c(\Omega:\{X_i\})$ be the number of arrangements $a(\Omega)$ and $a(\Omega:\{X_i\})$, respectively.
- Let x denote the X-gate operator, for instance, $\text{x} = \odot$ for AND-gate, $\text{x} = \oplus$ for OR-gate, $\text{x} = \text{not}\odot$ for NAND-gate, $\text{x} = \text{not}\oplus$ for NOR-gate, $\text{x} = \otimes$ for XOR-gate, $\text{x} = \text{not}\otimes$ for XNOR-gate, $\text{x} = \uplus$ for U-gate, $\text{x} = +$ for SIGMA-gate. Given an x-operator, let $s(\Omega:\{X_i\})$ and $s(\Omega)$ be sum of all $P(X_1 \text{x} X_2 \text{x} \ldots \text{x} X_n)$ through every arrangement of $\Omega$ with and without fixed $X_i$, respectively.

$$s(\Omega) = \sum_a P(X_1 \oplus X_2 \oplus \ldots \oplus X_n | a(\Omega))$$

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

# 3. X-gate inferences

```
public class ArrangementGenerator {
    private ArrayList<int[]> arrangements; private int n, 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);
            }
        }
    }

}
```

*code for producing all arrangements*

# 3. X-gate inferences

- Connection between $s(\Omega:\{X_i=1\})$ and $s(\Omega:\{X_i=0\})$, between $c(\Omega:\{X_i=1\})$ and $c(\Omega:\{X_i=0\})$.

$$s(\Omega:\{X_i = 1\}) + s(\Omega:\{X_i = 0\}) = s(\Omega)$$

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

- Let $K$ be a set of $X_i$ (s) whose values are 1 and let $L$ be a set of $X_i$ (s) whose values are 0. $K$ and $L$ are mutually complementary.

$$\begin{cases} K = \{i: X_i = 1\} \\ L = \{i: X_i = 0\} \\ K \cap L = \emptyset \\ K \cup L = \{1,2,\dots,n\} \end{cases}$$

# 3. X-gate inferences

- AND-gate condition (3.7)

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

- **AND-gate inference** (3.8)

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

$$=\begin{cases} \prod_{i=1}^{n} p_i \text{ if all } X_i\ (s) \text{ are } 1 \\ 0 \text{ if there exists at least one } X_i=0 \end{cases}$$

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

# 3. X-gate inferences

- OR-gate condition (3.9)

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

- **OR-gate inference** (3.10)

$$P(X_1 \oplus X_2 \oplus \dots \oplus X_n) = 1 - P(Y = 0 | X_1, X_2, \dots, X_n)$$

$$= \begin{cases} 1 - \prod_{i \in K} (1 - p_i) & \text{if } K \neq \emptyset \\ 0 & \text{if } K = \emptyset \end{cases}$$

$$P(Y = 0 | X_1, X_2, \dots, X_n) = \begin{cases} \prod_{i \in K} (1 - p_i) & \text{if } K \neq \emptyset \\ 1 & \text{if } K = \emptyset \end{cases}$$

# 3. X-gate inferences

- **NAND-gate inference** and **NOR-gate inference** (3.11)

$$P\big(\text{not}(X_1 \odot X_2 \odot \dots \odot X_n)\big) = \begin{cases} 1 - \prod_{i \in L} p_i & \text{if } L \neq \emptyset \\ 0 & \text{if } L = \emptyset \end{cases}$$

$$P\big(\text{not}(X_1 \oplus X_2 \oplus \dots \oplus X_n)\big) = \begin{cases} \prod_{i=1}^{n} q_i & \text{if } K = \emptyset \\ 0 & \text{if } K \neq \emptyset \end{cases}$$

# 3. X-gate inferences

- Two XOR-gate conditions (3.12)

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

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

- Let $O$ be the set of $X_i$ (s) whose indices are odd. Let $O_1$ and $O_2$ be subsets of $O$, in which all $X_i$ (s) are 1 and 0, respectively. Let $E$ be the set of $X_i$ (s) whose indices are even. Let $E_1$ and $E_2$ be subsets of $E$, in which all $X_i$ (s) are 1 and 0, respectively. **XOR-gate inference** (3.13) is:

$$P(X_1\otimes X_2\otimes\dots\otimes X_n)=P(Y=1|X_1,X_2,\dots,X_n)=\begin{cases}\left(\prod_{i\in O_1}p_i\right)\left(\prod_{i\in E_1}(1-p_i)\right)+\left(\prod_{i\in E_1}p_i\right)\left(\prod_{i\in O_1}(1-p_i)\right)\text{ if }O_2=\emptyset\text{ and }E_2=\emptyset\\\left(\prod_{i\in O_1}p_i\right)\left(\prod_{i\in E_1}(1-p_i)\right)\text{ if }O_2=\emptyset\text{ and }E_1\neq\emptyset\text{ and }E_2\neq\emptyset\\\prod_{i\in O_1}p_i\text{ if }O_2=\emptyset\text{ and }E_1=\emptyset\\\left(\prod_{i\in E_1}p_i\right)\left(\prod_{i\in O_1}(1-p_i)\right)\text{ if }E_2=\emptyset\text{ and }O_1\neq\emptyset\text{ and }O_2\neq\emptyset\\\prod_{i\in E_1}p_i\text{ if }E_2=\emptyset\text{ and }O_1=\emptyset\\0\text{ if }O_2\neq\emptyset\text{ and }E_2\neq\emptyset\\0\text{ if }n<2\text{ or }n\text{ is odd}\end{cases}$$

# 3. X-gate inferences

- Two XNOR-gate conditions (3.14)

$$P(Y = 1 | A_i = ON, \forall i) = 1$$
$$P(Y = 1 | A_i = OFF, \forall i) = 1$$

- **XNOR-gate inference** (3.15)

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

# 3. X-gate inferences

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 $\alpha$, $\beta$ and cardinality of $U$. Formula 3.16 specifies three common **U-gate conditions**.

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

# 3. X-gate inferences

Let $P_U$ be the U-gate probability, following formula 3.17 specifies **U-gate inference** and cardinality of $\mathcal{U}$ where $\mathcal{U}$ is the set of subsets (*U*) of *K*

- Let $S_U = \sum_{U \in \mathcal{U}} \prod_{i \in U} p_i \prod_{j \in K \setminus U} \left(1 - p_j\right)$ and $P_U = P(X_1 \uplus X_2 \uplus \ldots \uplus X_n) = P(Y = 1 | X_1, X_2, \ldots, X_n)$
- As a convention, $\prod_{i \in U} p_i = 1$ if $|U| = 0$ and $\prod_{j \in K \setminus U} \left(1 - p_j\right) = 1$ if $|U| = |K|$
- $|U|$=0: we have $P_U = \begin{cases} \prod_{j=1}^{n} \left(1 - p_j\right) \text{ if } |K| > 0 \\ 1 \text{ if } |K| = 0 \end{cases}$ and $|\mathcal{U}| = 1$
- $|U|$≥0: we have $P_U = \begin{cases} S_U \text{ if } |K| > 0 \\ 1 \text{ if } |K| = 0 \end{cases}$ and $|\mathcal{U}| = 2^{|K|}$. The case $|U|$≥0 is the same to the case $|U|$≤*n*.
- $|U|$=*n*: we have $P_U = \begin{cases} \prod_{i=1}^{n} p_i \text{ if } |K| = n \\ 0 \text{ if } |K| < n \end{cases}$ and $|\mathcal{U}| = \begin{cases} 1 \text{ if } |K| = n \\ 0 \text{ if } |K| < n \end{cases}$

# 3. X-gate inferences

**U-gate inference (continue)**

- $|U|=\alpha$ and $0<\alpha<n$: we have $P_U = \begin{cases} S_U \text{ if } |K| \geq \alpha \\ 0 \text{ if } |K| < \alpha \end{cases}$ and $|\mathcal{U}| = \begin{cases} \binom{|K|}{\alpha} \text{ if } |K| \geq \alpha \\ 0 \text{ if } |K| < \alpha \end{cases}$
- $|U|\geq\alpha$ and $0<\alpha<n$: we have $P_U = \begin{cases} S_U \text{ if } |K| \geq \alpha \\ 0 \text{ if } |K| < \alpha \end{cases}$ and $|\mathcal{U}| = \begin{cases} \sum_{j=\alpha}^{|K|} \binom{|K|}{j} \text{ if } |K| \geq \alpha \\ 0 \text{ if } |K| < \alpha \end{cases}$
- $|U|\leq\beta$ and $0<\beta<n$: we have $P_U = \begin{cases} S_U \text{ if } |K| > 0 \\ 1 \text{ if } |K| = 0 \end{cases}$ and $|\mathcal{U}| = \begin{cases} \sum_{j=0}^{\min(\beta,|K|)} \binom{|K|}{j} \text{ if } |K| > 0 \\ 1 \text{ if } |K| = 0 \end{cases}$
- $\alpha\leq|U|\leq\beta$ and $0<\alpha<n$ and $0<\beta<n$: we have $P_U = \begin{cases} S_U \text{ if } |K| \geq \alpha \\ 0 \text{ if } |K| < \alpha \end{cases}$ and $|\mathcal{U}| = \begin{cases} \sum_{j=\alpha}^{\min(\beta,|K|)} \binom{|K|}{j} \text{ if } |K| \geq \alpha \\ 0 \text{ if } |K| < \alpha \end{cases}$

# 3. X-gate inferences

- U-gate condition on $|U|$ can be arbitrary and it is only relevant to $A_i$ (s) (*ON* or *OFF*) and the way to combine $A_i$ (s). For example, AND-gate and OR-gate are specific cases of U-gate with $|U|=n$ and $|U|\geq1$, respectively. XOR-gate and XNOR-gate are also specific cases of U-gate with specific conditions on $A_i$ (s).
- In this research, U-gate is the most general nonlinear gate where U-gate probability contains products of weights.
- Next, we will research a so-called SIGMA-gate that contains only linear combination of weights.

# 3. X-gate inferences

- The SIGMA-gate inference (Nguyen, 2016) represents aggregation relationship satisfying **SIGMA-gate condition** specified by formula 3.18.

$$P(Y) = P\left(\sum_{i=1}^{n} A_i\right) \quad (3.18)$$

Where the set of $A_i$ (s) is complete and mutually exclusive.

$$\sum_{i=1}^{n} w_i = 1 \text{ and } A_i \cap A_j = \emptyset, \forall i \neq j$$

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

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

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

# 3. X-gate inferences

- In general, formula 3.19 specifies the theorem of **SIGMA-gate inference** (Nguyen, 2016). The base of this theorem was mentioned by authors (Millán & Pérez-de-la-Cruz, 2002, pp. 292-295).

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

$$P(Y = 0|X_1, X_2, \dots, X_n) = 1 - \sum_{i \in K} w_i = \sum_{i \in L} w_i$$

Where the set of $A_i$ (s) is complete and mutually exclusive.

$$\sum_{i=1}^{n} w_i = 1 \text{ and } A_i \cap A_j = \emptyset, \forall i \neq j$$

- Next slide is the proof of SIGMA-gate theorem.

# 3. X-gate inferences

## Proof of SIGMA-gate theorem

$$(Y|X_1, X_2, \ldots, X_n) = P\left(\sum_{i=1}^{n} A_i \,|X_1, X_2, \ldots, X_n\right)$$

(due to SIGMA – gate condition)

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

(because $A_i$ (s) are mutually exclusive)

$$= \sum_{i=1}^{n} P(A_i|X_i)$$

(because $A_i$ is only dependent on $X_i$)

It implies

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

$$= \left(\sum_{i \in K} P(A_i = ON|X_i = 1)\right)$$

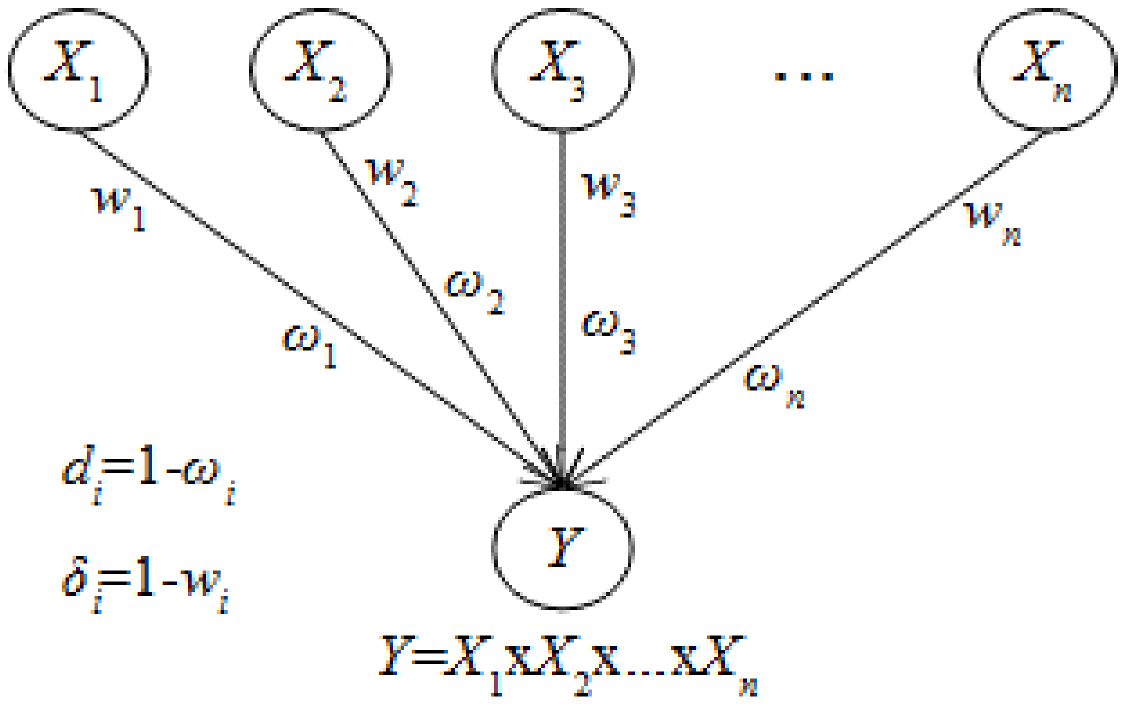
$$+ \left(\sum_{i \notin K} P(A_i = ON|X_i = 0)\right)$$

$$= \sum_{i \in K} w_i + \sum_{i \notin K} 0 = \sum_{i \in K} w_i$$

(Due to formula 3.3)

# 3. X-gate inferences

- As usual, each arc in simple graph is associated with a "*clockwise*" strength of relationship between $X_i$ and $Y$. Event $X_i$=1 causes event $A_i$=$ON$ with "clockwise" weight $w_i$.
- I define a so-called "*counterclockwise*" strength of relationship between $X_i$ and $Y$ denoted $\omega_i$. Event $X_i$=0 causes event $A_i$=$OFF$ with "counterclockwise" weight $\omega_i$.
- In other words, each arc in simple graph is associated with a clockwise weight $w_i$ and a counterclockwise weight $\omega_i$. Such graph is called *bi-weight simple graph*.



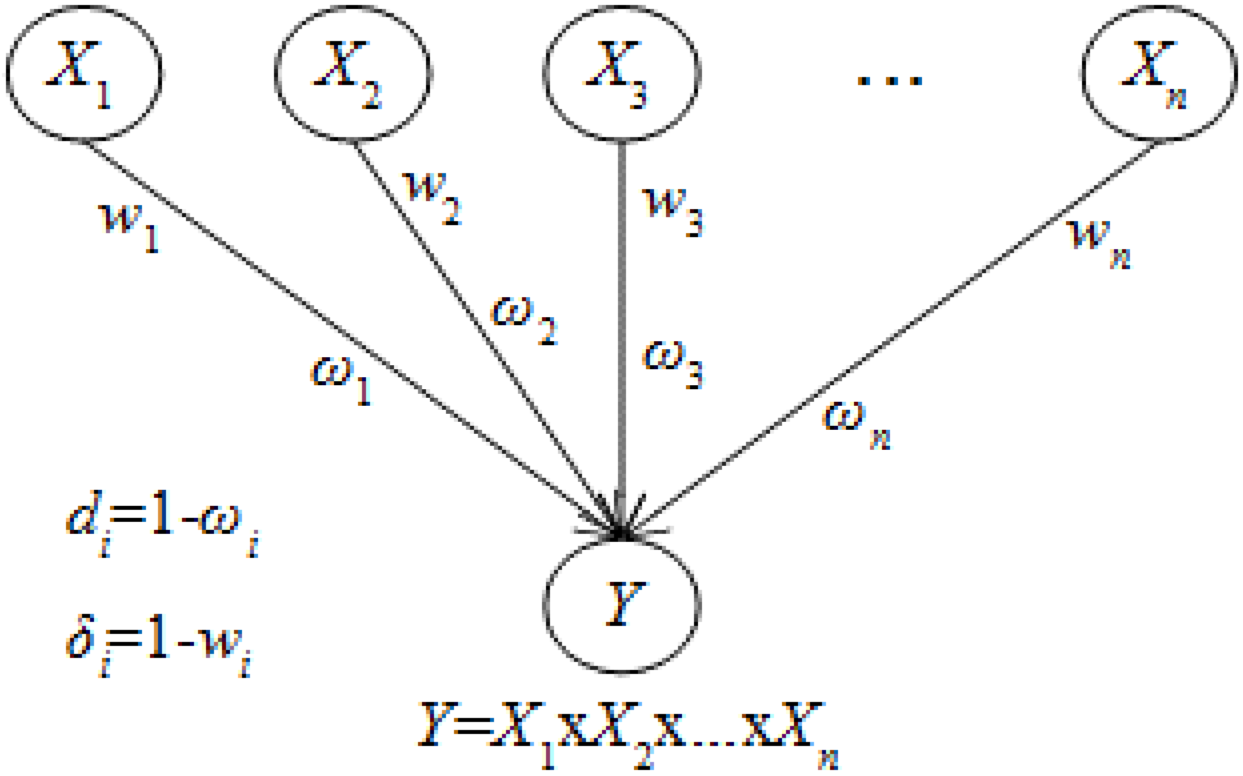

Bi-weight simple graph

# 3. X-gate inferences

$$P(A_i = ON|X_i = 1) = p_i = w_i$$
$$P(A_i = ON|X_i = 0) = 1 - \rho_i = 1 - \omega_i$$
$$P(A_i = OFF|X_i = 1) = 1 - p_i = 1 - w_i$$
$$P(A_i = OFF|X_i = 0) = \rho_i = \omega_i$$

$$d_i = 1 - \omega_i$$
$$\delta_i = 1 - w_i$$
$$W_i = w_i + d_i$$
$$\mathcal{W}_i = \omega_i + \delta_i$$

Bi-weight simple graph



# 3. X-gate inferences

From bi-weight graph, we get bi-inferences for AND-gate, OR-gate, NAND-gate, NOR-gate, XOR-gate, XNOR-gate, and U-gate

- $P(X_1 \odot X_2 \odot \ldots \odot X_n) = \prod_{i \in K} p_i \prod_{i \in L} d_i$
- $P(X_1 \oplus X_2 \oplus \ldots \oplus X_n) = 1 - \prod_{i \in K} \delta_i \prod_{i \in L} \rho_i$
- $P\big(\text{not}(X_1 \odot X_2 \odot \ldots \odot X_n)\big) = 1 - \prod_{i \in L} \rho_i \prod_{i \in K} \delta_i$
- $P\big(\text{not}(X_1 \oplus X_2 \oplus \ldots \oplus X_n)\big) = \prod_{i \in L} d_i \prod_{i \in K} p_i$
- $P(X_1 \otimes X_2 \otimes \ldots \otimes X_n) = \prod_{i \in O_1} p_i \prod_{i \in O_2} d_i \prod_{i \in E_1} \delta_i \prod_{i \in E_2} \rho_i + \prod_{i \in E_1} p_i \prod_{i \in E_2} d_i \prod_{i \in O_1} \delta_i \prod_{i \in O_2} \rho_i$
- $P\big(\text{not}(X_1 \otimes X_2 \otimes \ldots \otimes X_n)\big) = \prod_{i \in K} p_i \prod_{i \in L} d_i + \prod_{i \in K} \delta_i \prod_{i \in L} \rho_i$
- $P(X_1 \uplus X_2 \uplus \ldots \uplus X_n) = \sum_{U \in \mathcal{U}} \left(\prod_{i \in U \cap K} p_i \prod_{i \in U \cap L} d_i\right) \left(\prod_{i \in \bar{U} \cap K} \delta_i \prod_{i \in \bar{U} \cap L} \rho_i\right)$

# 3. X-gate inferences

- Formula 3.22 specifies **SIGMA-gate bi-inference**.

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

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

$$\sum_{i=1}^{n} W_i = 1 \text{ and } X_i \cap X_j = \emptyset, \forall i \neq j$$

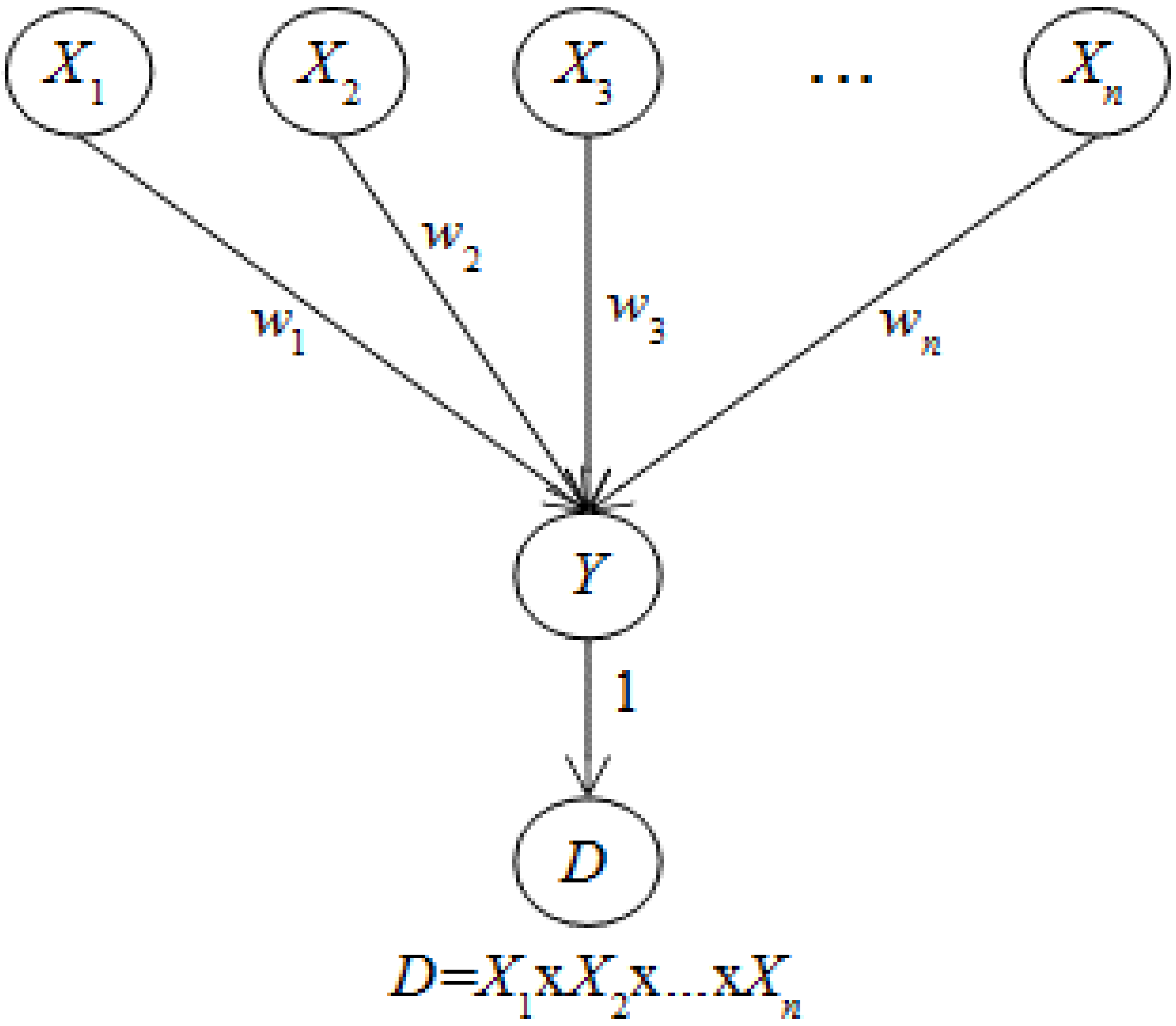
- Following is the proof of SIGMA-gate bi-inference.

$$P(X_1 + X_2 + \cdots + X_n) = \sum_{i=1}^{n} P(A_i = ON | X_i)$$

$$= \sum_{i \in K} P(A_i = ON | X_i = 1) + \sum_{i \in L} P(A_i = ON | X_i = 0) = \sum_{i \in K} w_i + \sum_{i \in L} d_i$$

# 4. Multi-hypothesis diagnostic relationship

- Given a simple graph, if we replace the target source *Y* by an evidence *D*, we get a so-called *multi-hypothesis diagnostic relationship* whose property adheres to X-gate inference. Such relationship is called shortly *X-gate diagnostic relationship*.
- According to aforementioned X-gate network, the target variable must be binary whereas evidence *D* can be numeric. Thus, we add an augmented target binary variable *Y* and then, the evidence *D* is connected directly to *Y*. Finally, we have X-gate diagnostic network or **X-D network**.



# 4. Multi-hypothesis diagnostic relationship

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

(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)

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

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

Joint probability of X-D network (4.1)

# 4. Multi-hypothesis diagnostic relationship

Basic probabilities relevant to X-D network with uniform distribution

- $P(\Omega, Y, D) = P(X_1, X_2, \ldots, X_n, Y, D) = P(D|Y)P(Y|X_1, X_2, \ldots, X_n) \prod_{i=1}^{n} P(X_i)$
- $P(D|X_i) = \frac{P(X_i, D)}{P(X_i)} = \frac{\sum_{\{\Omega, Y, D\} \backslash \{X_i, D\}} P(\Omega, Y, D)}{\sum_{\{\Omega, Y, D\} \backslash \{X_i\}} P(\Omega, Y, D)}$
- $P(X_i|D) = \frac{P(X_i, D)}{P(D)} = \frac{\sum_{\{\Omega, Y, D\} \backslash \{X_i, D\}} P(\Omega, Y, D)}{\sum_{\{\Omega, Y, D\} \backslash \{D\}} P(\Omega, Y, D)}$
- $P(X_i, D) = \frac{1}{2^n S}\Big((2D - M)s(\Omega{:}\{X_i\}) + 2^{n-1}(M - D)\Big)$
- $P(D) = \frac{1}{2^n S}\Big((2D - M)s(\Omega) + 2^n(M - D)\Big)$

# 4. Multi-hypothesis diagnostic relationship

Conditional probability, posterior probability, and transformation coefficient of X-D network according to formula 4.4

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

# 4. Multi-hypothesis diagnostic relationship

**Diagnostic theorem**

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

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

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

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

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

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.

# 4. Multi-hypothesis diagnostic relationship

## Proof of diagnostic theorem

The transformation coefficient is rewritten as follows: $k = \frac{2^{n-1}S}{2D(s(\Omega)-2^{n-1})+M\left(2^n-s(\Omega)\right)}$

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

Where $C$ is constant. We have: $\sum_D \left(2D(s(\Omega) - 2^{n-1}) + M\left(2^n - s(\Omega)\right)\right) = \sum_D C \Rightarrow 2S(s(\Omega) - 2^{n-1}) + NM\left(2^n - s(\Omega)\right) = NC \Rightarrow 2^n S = NC$

It is implied that $k = \frac{2^{n-1}S}{2D(s(\Omega)-2^{n-1})+M\left(2^n-s(\Omega)\right)} = \frac{NC}{2C} = \frac{N}{2}$

This holds $2^n S = N\left(2D(s(\Omega) - 2^{n-1}) + M\left(2^n - s(\Omega)\right)\right) = 2ND(s(\Omega) - 2^{n-1}) + 2S\left(2^n - s(\Omega)\right)$

$$\Rightarrow 2ND(s(\Omega) - 2^{n-1}) - 2S(s(\Omega) - 2^{n-1}) = 0$$

$$\Rightarrow (ND - S)(s(\Omega) - 2^{n-1}) = 0$$

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

In general, the event that $k$ is constant with regard to $D$ is equivalent to the event $s(\Omega) = 2^{n-1}$.

# 4. Multi-hypothesis diagnostic relationship

Probabilities and transformation coefficient according to X-D network with AND-gate reference called *AND-D network* according to formula 4.5.

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

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

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

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

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

For convenience, we validate diagnostic condition with a case of two sources $\Omega = \{X_1, X_2\}$, $p_1 = p_2 = w_1 = w_2 = 0.5$, $D \in \{0,1,2,3\}$. By applying diagnostic theorem stated for AND-D network, because $s(\Omega) = 0.25$, AND-D network does not satisfy diagnostic condition.

# 4. Multi-hypothesis diagnostic relationship

- 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.
- Formula 4.6 specifies probabilities of **SIGMA-D network** in order to validate it, as follows:

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

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

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

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

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

- By applying diagnostic theorem stated for SIGMA-D network, we have $s(\Omega) = 2^{n-1} \sum_i (w_i +$

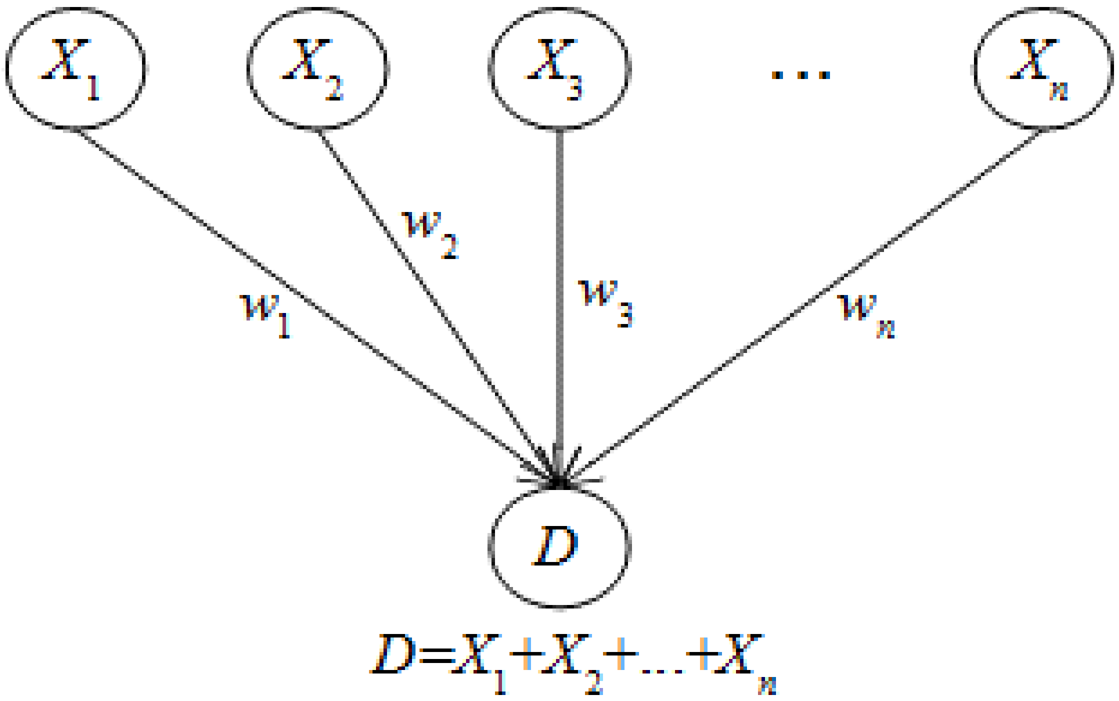
# 4. Multi-hypothesis diagnostic relationship

- 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, which composes so-called **direct SIGMA-D network**.
- CPT of direct SIGMA-D network is determined by formula 4.7.

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

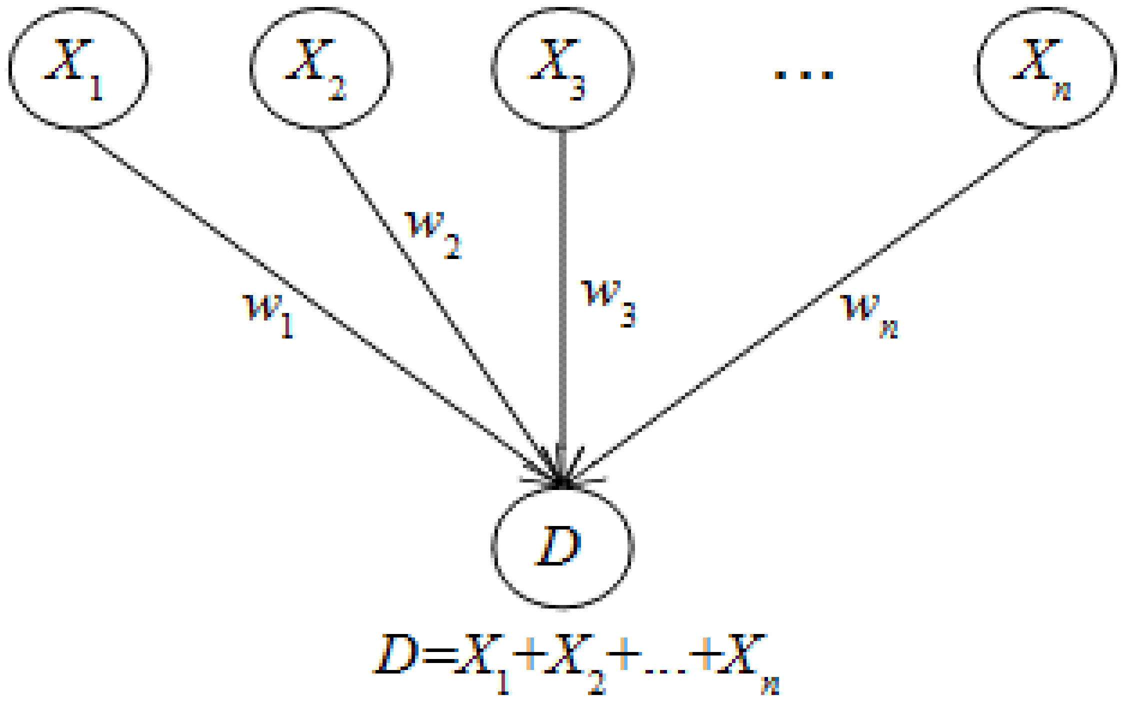
Where the set of $X_i$ (s) is complete and mutually exclusive.

$$\sum_{i=1}^{n} w_i = 1 \text{ and } X_i \cap X_j = \emptyset, \forall i \neq j$$



# 4. Multi-hypothesis diagnostic relationship

- Direct **SIGMA-D network** shares the same conditional probabilities $P(X_i|D)$ and $P(D|X_i)$ with SIGMA-D network, as seen in formula 4.6.
- Direct SIGMA-D network also satisfies diagnostic condition when its $s(\Omega) = 2^{n-1}$.



# 4. Multi-hypothesis diagnostic relationship

- The most general nonlinear X-D network is U-D network whereas SIGMA-D network is linear one. Aforementioned nonlinear X-D network such as AND, OR, NAND, NOR, XOR, and NXOR are specific cases of X-D network. **Now we validate if U-D network satisfies diagnostic condition**.
- The U-gate inference given arbitrary condition on $U$ is $P(X_1 \uplus X_2 \uplus \ldots \uplus X_n) = \sum_{U \in \mathcal{U}} \left( \prod_{i \in U \cap K} p_i \prod_{i \in U \cap L} (1 - \right.$

# 4. Multi-hypothesis diagnostic relationship

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

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

In any case of degradation, there always exist expression $Expr$ (s) having at least 2 sub-products ($\Pi$), 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 $\rho_i$ 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 $n \geq 5$. Without loss of generality, each $p_i$ or $\rho_i$ is sum of variable $x$ and a variable $a_i$ or $b_i$, respectively. Note that all $p_i$, $\rho_i$, $a_i$ 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 $n \geq 5$.

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

Where coefficients $C_i$ (s) are functions of $a_i$ and $b_i$ (s). According to Abel-Ruffini theorem (Wikipedia, Abel-Ruffini theorem, 2016), equation $g(x) = 0$ has no algebraic solution when $m \geq 5$. Thus, abstract variables $p_i$ and $\rho_i$ cannot be eliminated entirely from $g(x)=0$, which causes that **there is no specification of U-gate inference $P(X_1 \mathrm{x} X_2 \mathrm{x} \ldots \mathrm{x} X_n)$ so that diagnostic condition is satisfied**.

# 4. Multi-hypothesis diagnostic relationship

- 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.

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

- The GL-gate inference is singular if $\alpha_i$ and $\beta_i$ are functions of only $X_i$ as follows:

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

# 4. Multi-hypothesis diagnostic relationship

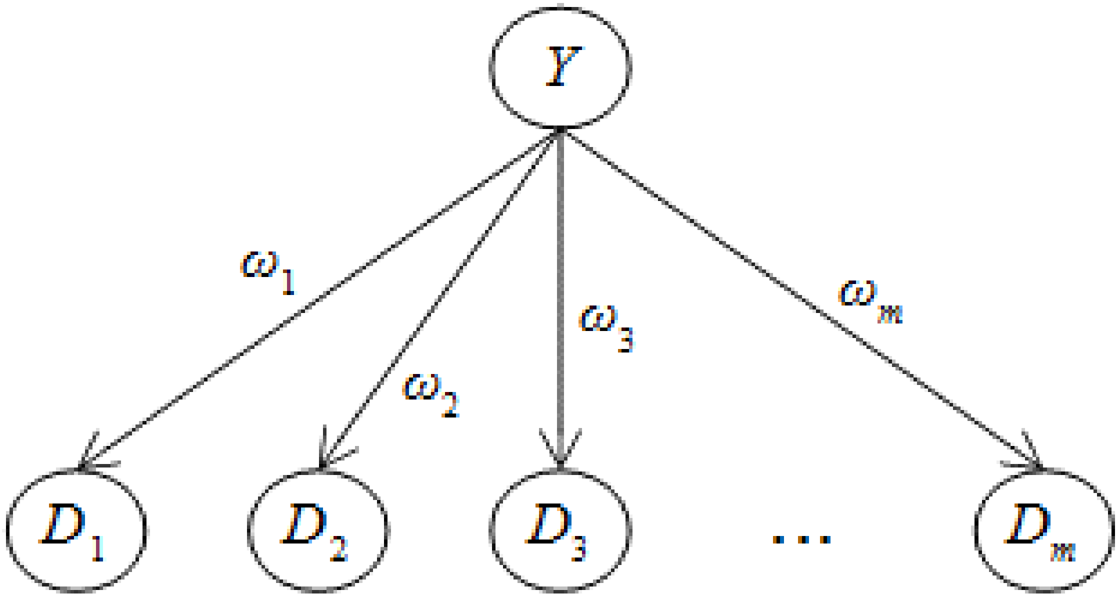
- Suppose $h_i$ and $g_i$ are probability mass functions with regard to $X_i$. For all $i$, we have: $0 \leq h_i(X_i) \leq 1$, $0 \leq g_i(X_i) \leq 1$, $h_i(X_i = 1) + h_i(X_i = 0) = 1$, $g_i(X_i = 1) + g_i(X_i = 0) = 1$
- **GL-D network** satisfies diagnostic condition if $s(\Omega) = 2^n C + 2^{n-1} \sum_{i=1}^{n}(w_i + d_i) = 2^{n-1} \Rightarrow 2C + \sum_{i=1}^{n}(w_i + d_i) = 1$
- Suppose the set of $X_i$ (s) is complete, we have $\sum_{i=1}^{n}(w_i + d_i) = 1$. This implies $C$=0.
- Shortly, formula 4.10 specifies the singular GL-gate inference so that GL-D network satisfies diagnostic condition.

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

Where $h_i$ and $g_i$ are probability mass functions and the set of $X_i$ (s) is complete $\sum_{i=1}^{n}(w_i + d_i) = 1$

# 4. Multi-hypothesis diagnostic relationship

- According to authors (Millán & Pérez-de-la-Cruz, 2002), a hypothesis can have multiple evidences as seen in the next figure. This is *multi-evidence diagnostic relationship* opposite to aforementioned multi-hypothesis diagnostic relationship, which is called shortly **M-E-D network**.
- The joint probability of M-E-D network is $P(Y, D_1, D_2, \dots, D_m) = P(Y) \prod_{j=1}^{m} P\left(D_j \middle| Y\right) = P(Y)P(D_1, D_2, \dots, D_m | Y)$
- The possible transformation coefficient is $\frac{1}{k} = \prod_{j=1}^{m} P\left(D_j \middle| Y = 1\right) + \prod_{j=1}^{m} P\left(D_j \middle| Y = 0\right)$



# 4. Multi-hypothesis diagnostic relationship

- 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 following formula 4.11 has 2*m* real roots $P(D_j|Y)$ for all $m \geq 2$.

$$\prod_{j=1}^{m} P\left(D_j \middle| Y = 1\right) + \prod_{j=1}^{m} P\left(D_j \middle| Y = 0\right) = 1 \quad (4.11)$$

- Suppose equation 4.11 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 equation 4.11, it holds $\begin{cases} a_1 = a_2 = 0 \\ b_1 = b_2 \\ a_1^2 + b_1^2 = 1 \\ b_1 = 2 \end{cases}$ or $\begin{cases} a_1 = a_2 = 0.5 \\ b_1 = b_2 \\ a_1^2 + b_1^2 = 1 \\ b_1 = 1.5 \end{cases}$ which leads a contradiction ($b_1$=2 or $b_1$=1.5) and so **it is impossible to apply the diagnostic condition into M-E-D network**.

# 4. Multi-hypothesis diagnostic relationship

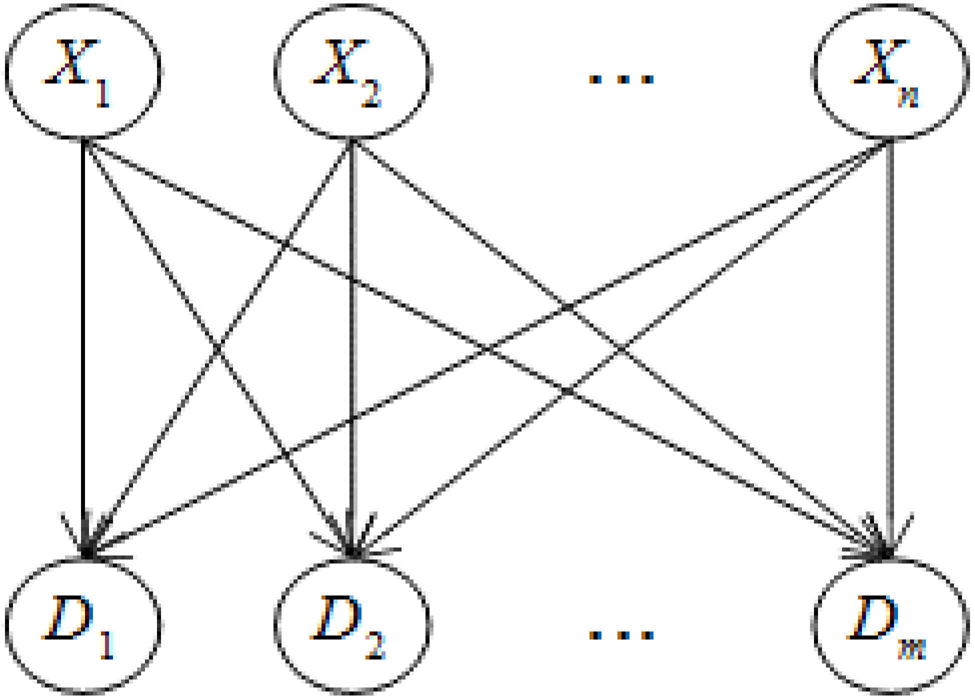
- It is impossible to **model M-E-D network by X-gates**. The potential solution for this problem is to group many evidences $D_1$, $D_2$,…, $D_m$ into one representative evidence $D$ which in turn is dependent on hypothesis $Y$ but this solution will be inaccurate in specifying conditional probabilities because directions of dependencies become inconsistent (relationships from $D_j$ to $D$ and from $Y$ to $D$) except that all $D_j$ (s) 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.

# 4. Multi-hypothesis diagnostic relationship

- Another solution to model **M-E-D network** by X-gates 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:
- The joint probability of M-E-D is: $P(Y, D_1, D_2, \ldots, D_m) = P(Y)\prod_{j=1}^{m} P\left(D_j \middle| Y\right)$
- M-E-D network satisfies partial diagnostic condition because $P\left(Y \middle| D_j\right) = \frac{1}{2} P\left(D_j \middle| Y\right)$
- 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.

# 4. Multi-hypothesis diagnostic relationship

- 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 $X_1$, $X_2$,…, $X_n$ and $m$ evidences $D_1$, $D_2$,…, $D_m$. Such extended network represents *multi-hypothesis multi-evidence diagnostic relationship*, called **M-HE-D network**.
- The M-HE-D network is the most general case of diagnostic network, which was mentioned in (Millán & Pérez-de-la-Cruz, 2002, 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 is adhered to semantics of relationships. The weak point of logic gates is to require that all variables must be binary.
- In order to lessen the impact of such weak point, I use numeric evidence for extending capacity of simple BN. However, combination of binary hypothesis and numeric evidence leads to errors or biases in inference. Therefore, I propose the *diagnostic condition* so as to confirm that numeric evidence is adequate to make complicated inference tasks in BN.
- A large BN can be constituted of many simple BN (s). Inference in large BN is hazardous problem. In future, I will research effective inference methods for the special BN that is constituted of X-gate BN (s).
- Moreover, I try my best to research deeply M-E-D network and M-HE-D network whose problems I cannot solve absolutely now.

# 5. Conclusion

- Two main documents I referred to do this research are the book “Learning Bayesian Networks” by the author (Neapolitan, 2003) and the article “A Bayesian Diagnostic Algorithm for Student Modeling and its Evaluation” by authors (Millán & Pérez-de-la-Cruz, 2002).
- Especially, the SIGMA-gate inference is based on and derived from the work of the authors Eva Millán and José Luis Pérez-de-la-Cruz.
- This research is originated from my PhD research “A User Modeling System for Adaptive Learning” (Nguyen, 2014).
- Other references relevant to user modeling, overlay model, and Bayesian network are (Fröschl, 2005), (De Bra, Smits, & Stash, 2006), (Murphy, 1998), and (Heckerman, 1995).

# Thank you for attention

# References


1. De Bra, P., Smits, D., & Stash, N. (2006). The Design of AHA! In U. K. Wiil, P. J. Nürnberg, & J. Rubart (Ed.), *Proceedings of the seventeenth ACM Hypertext Conference on Hypertext and hypermedia (Hypertext '06)* (pp. 171-195). Odense, Denmark: ACM Press.
2. Díez, F. J., & Druzdzel, M. J. (2007). *Canonical Probabilistic Models*. National University for Distance Education, Department of Inteligencia Artificial. Madrid: Research Centre on Intelligent Decision-Support Systems. Retrieved May 9, 2016, from http://www.cisiad.uned.es/techreports/canonical.pdf
3. Fröschl, C. (2005). *User Modeling and User Profiling in Adaptive E-learning Systems*. Master Thesis, Graz University of Technology, Austria.
4. Heckerman, D. (1995). *A Tutorial on Learning With Bayesian Networks*. Microsoft Corporation, Microsoft Research. Redmond: Microsoft Research. Retrieved from ftp://ftp.research.microsoft.com/pub/dtg/david/tutorial.ps
5. Kschischang, F. R., Frey, B. J., & Loeliger, H.-A. (2001, February). Factor Graphs and the Sum-Product Algorithm. *IEEE Transactions on Information Theory, 47*(2), 498-519. doi:10.1109/18.910572
6. Millán, E., & Pérez-de-la-Cruz, J. L. (2002, June). A Bayesian Diagnostic Algorithm for Student Modeling and its Evaluation. (A. Kobsa, Ed.) *User Modeling and User-Adapted Interaction, 12*(2-3), 281-330. doi:10.1023/A:1015027822614
7. Millán, E., Loboda, T., & Pérez-de-la-Cruz, J. L. (2010, July 29). Bayesian networks for student model engineering. (R. S. Heller, J. D. Underwood, & C.-C. Tsai, Eds.) *Computers & Education, 55*(4), 1663-1683. doi:10.1016/j.compedu.2010.07.010
8. Murphy, K. P. (1998). *A Brief Introduction to Graphical Models and Bayesian Networks*. Retrieved 2008, from Kevin P. Murphy's home page: http://www.cs.ubc.ca/~murphyk/Bayes/bnintro.html
9. Neapolitan, R. E. (2003). *Learning Bayesian Networks*. Upper Saddle River, New Jersey, USA: Prentice Hall.
10. Nguyen, L. (2014, April). *A User Modeling System for Adaptive Learning*. University of Science, Ho Chi Minh city, Vietnam. Abuja, Nigeria: Standard Research Journals. Retrieved from http://standresjournals.org/journals/SSRE/Abstract/2014/april/Loc.html
11. Nguyen, L. (2016, March 28). Theorem of SIGMA-gate Inference in Bayesian Network. (V. S. Franz, Ed.) *Wulfenia Journal, 23*(3), 280-289.
12. Pearl, J. (1986, September). Fusion, propagation, and structuring in belief networks. *Artificial Intelligence, 29*(3), 241-288. doi:10.1016/0004-3702(86)90072-X
13. Wikipedia. (2014, October 10). *Set (mathematics)*. (A. Rubin, Editor, & Wikimedia Foundation) Retrieved October 11, 2014, from Wikipedia website: http://en.wikipedia.org/wiki/Set_(mathematics)
14. Wikipedia. (2015, November 22). *Factor graph*. (Wikimedia Foundation) Retrieved February 8, 2017, from Wikipedia website: https://en.wikipedia.org/wiki/Factor_graph
15. Wikipedia. (2016, June 10). *Abel-Ruffini theorem*. (Wikimedia Foundation) Retrieved June 26, 2016, from Wikipedia website: https://en.wikipedia.org/wiki/Abel%E2%80%93Ruffini_theorem
16. Wikipedia. (2016, June 2). *Logic gate*. (Wikimedia Foundation) Retrieved June 4, 2016, from Wikipedia website: https://en.wikipedia.org/wiki/Logic_gate