

Research article

A Survey Report of Parameter and Structure Learning in Bayesian Network Inference

Prof. (Dr.) V. N. Maurya

Professor & Ex Principal,
Shekhawati Engineering College, Rajasthan Technical University, India
E-mail: prof.drvmmaurya@gmail.com, prof_vnmaurya@yahoo.in

Diwinder Kaur Arora

Inspector of Police,
Group Centre, Central Reserve Police Force, Lucknow-226002, U.P.
Ministry of Home Affairs, Govt. of India
E-mail: hkdkarora@rediffmail.com, diwi.kaur1992@gmail.com

Er. Avadhesh Kumar Maurya

Assistant Professor, Department of Electronics & Communication Engineering
Lucknow Institute of Technology, U.P. Technical University, Lucknow-226002, India
E-mail: avadheshmaurya09@gmail.com

Abstract

This paper presents a survey report of parameter and structure learning in Bayesian network inference (BNI). The core objective of parameter learning is to compute the posterior density function. In this paper, parameter learning in case of data missing, expanding augmented Bayesian Network with multi-hypothesis nodes and updating posterior density function with multi-node Bayesian network inferences have been attempted successfully. Moreover, both score-based and constraint-based approaches have also been applied exhaustively for structure learning in Bayesian network inference. In each case of parameter and structure learning, numerical illustrations are presented. Finally, conclusions are drawn based on survey report.

Keywords: Bayesian network inference, parameter and structure learning, posterior density function, beta function, binomial sample, multi-node Bayesian network, score-based approach and constraint-based approach, EM algorithm etc.

1. Introduction

Bayesian network inference is applied widely in business statistics, graph theory, operations management, machine learning, data mining, diagnosis, etc. Bayesian network inference (BNI) has a solid evidence-based inference which is familiar to human intuition. The ultimate purpose of Bayesian network inference is to consolidate a hypothesis by collecting evidences. However Bayesian network causes a little confusion because there are many complicated concepts, formulas and diagrams relating to it. Such concepts should be organized and presented in clear manner so as to be easy to understand it. This is the goal of the present paper. This report includes two main significant domains of Bayesian network that cover parameter and structure learning in Bayesian network inference: Part 1: Parameter learning and Part 2: Structure learning. However, there is one more significant domain of Bayesian network called inference mechanism

which has recently dealt exhaustively by Maurya et al. [5]. By study of relevant literature, it is observed that very few researchers have contributed to analyse Bayesian network inference. However, scope of Bayesian network inference in our day-to-day life and particularly in industry and engineering is considerably too much. In this connection, [1,2,3...8] and references therein are some noteworthy researchers.

2. Essence of Parameter learning

2.1. Beta Function and Augmented Bayesian Network

There is a family of PDF which quantifies and updates the strength of conditional dependencies among nodes by natural way is called beta density function, denoted as $\beta(f; a, b)$ or $Beta(f; a, b)$ with parameters $a, b, N=a+b$ where a, b should be integer number > 0

$$\beta(f) = \frac{\Gamma(N)}{\Gamma(a)\Gamma(b)} f^{a-1} (1-f)^{b-1} \quad (2.1.1)$$

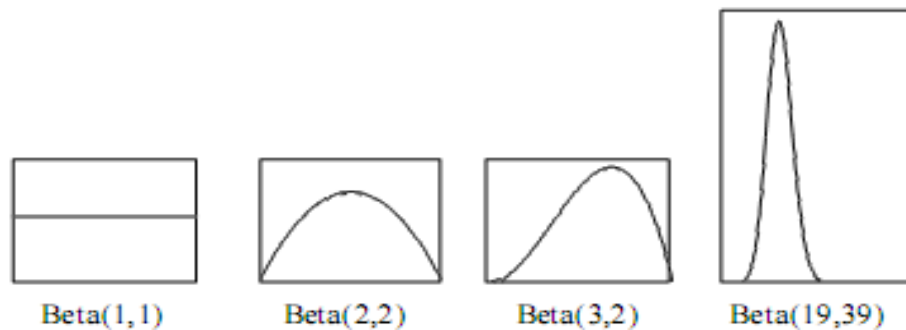


Figure 2.1.1: Beta functions

It means that, there are “ a ” successful outcomes (for example, $f=1$) in “ $a+b$ ” trials. Higher value of “ a ” is, higher ratio of success is, so, the graph leans forward right. Higher value of “ $a+b$ ” is, the more the mass is concentrate around $a/(a+b)$ and the more narrow the graph is. Definition of beta function is based on gamma function described below:

$$\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt \quad (2.1.2)$$

The integral will converges if $x>0$, at that time, $\Gamma(x) = (x-1)!$. Of course, we have

$$\frac{\Gamma(x+1)}{\Gamma(x)} = x \quad (2.1.3)$$

From formula 2.1.1, we can easily obtain

$$\int_0^1 f^a (1-f)^b df = \frac{\Gamma(a+1)\Gamma(b+1)}{\Gamma(a+b+2)} \quad (2.1.4)$$

Suppose there is one binary variable X in network and the probability distribution of X is considered as relative frequency having values in $[0, 1]$ which is the range of variable F . We add a dummy variable F (whose space consists of numbers in $[0, 1]$, of course) which acts as the parent of X and has a beta density function $\beta(f; a, b)$, so as to:

$$P(X=1|f) = f, \text{ where } f \text{ denotes values of } F$$

X and F constitute a simple network which is referred as augmented BN. So X is referred as real variable (hypothesis) opposite to dummy variable.

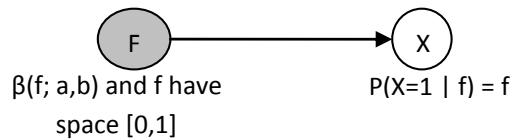


Figure 2.1.2: The simple augmented BN with only one hypothesis node X

Obviously, $P(X=1) = E(F)$ where $E(F)$ is the expectation of F

Proof: Owing to the law of total probability

$$P(X = 1) = \int_0^1 P(X = 1 | f) \beta(f) df = \int_0^1 f \beta(f) df = E(F)$$

Due to F is beta function, $E(F) = \frac{a}{N}$, so, $\Pr(X = 1) = \frac{a}{N}$ (2.1.5)

$$\begin{aligned} E(F) &= \int_0^1 f \beta(f) df = \int_0^1 f \frac{\Gamma(N)}{\Gamma(a)\Gamma(b)} f^{a-1} (1-f)^{b-1} df \\ &= \frac{\Gamma(N)}{\Gamma(a)\Gamma(b)} \int_0^1 f^a (1-f)^{b-1} = \frac{\Gamma(N)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(a+1)\Gamma(b+1)}{\Gamma(N+1)} \text{ (due to formula 2.1.2)} \\ &= \frac{a}{N} \text{ (applying formula 2.1.3)} \end{aligned}$$

The ultimate purpose of Bayesian inference is to consolidate a hypothesis (namely, variable) by collecting evidences. Suppose we perform M trials of a random process, the outcome of u^{th} trial is denoted $X^{(u)}$ considered as evidence variable whose probability $P(X^{(u)} = 1 | f) = f$. So, all $X^{(u)}$ are conditionally dependent on F . The probability of variable X , $P(X=1)$ is learned by these evidences.

We denote the vector of all evidences as $D = (X^{(1)}, X^{(2)}, \dots, X^{(M)})$ which is also called the sample of size M . Given this sample, $\beta(f)$ is called the prior density function, and $P(X^{(u)} = 1) = a/N$ (due to formula 3.1) is called prior probability of $X^{(u)}$. It is necessary to determine the posterior density function $\beta(f|D)$ and the posterior probability of X , namely $P(X|D)$. The nature of this process is the parameter learning. Note that $P(X|D)$ is referred as $P(X^{(M+1)} | D)$.

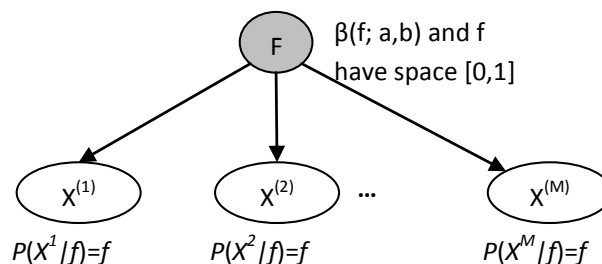


Figure 2.1.3: The sample $D=(X^{(1)}, X^{(2)}, \dots, X^{(M)})$ size of M

We only surveyed in the case of binomial sample, in other words, D having binomial distribution is called binomial sample and the network in figure 3 becomes a binomial augmented BN. Then, suppose s is the number of all evidences $X^{(i)}$ which have value 1 (success), otherwise, t is the number of all evidences $X^{(i)}$ which have value 0 (failed). Of course, $s + t = M$.

Owing the law of total probability, we have

$$\begin{aligned}
 E(f^s (1-f)^t) &= \int_0^1 f^s (1-f)^t \beta(f) df \\
 &= \int_0^1 f^s (1-f)^t \frac{\Gamma(N)}{\Gamma(a)\Gamma(b)} f^{a-1} (1-f)^{b-1} df \quad (\text{applying formula 2.1.1}) \\
 &= \frac{\Gamma(N)}{\Gamma(a)\Gamma(b)} \int_0^1 f^{a+s-1} (1-f)^{b+t-1} df \quad (2.1.6) \\
 &= \frac{\Gamma(N)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(a+s)\Gamma(b+t)}{\Gamma(a+b+s+t)} \quad (\text{due to formula 2.1.3}) \\
 &= \frac{\Gamma(N)}{\Gamma(N+M)} \frac{\Gamma(a+s)\Gamma(b+t)}{\Gamma(a)\Gamma(b)} \quad (\text{due to } s+t=M)
 \end{aligned}$$

And,

$$\begin{aligned}
 P(D) &= \int_0^1 \Pr(D|f) \beta(f) df = \int_0^1 \prod_{i=1}^M P(X^i | f) \beta(f) df \\
 &= \int_0^1 f^s (1-f)^t \beta(f) df = E(f^s (1-f)^t), \text{ due to } \prod_{i=1}^M P(X^i | f) = f^s (1-f)^t \quad (2.1.7)
 \end{aligned}$$

2.2 Parameter Learning

Parameter learning play a vital role to compute the posterior density function. Now, we need to compute the posterior density function $\beta(f|D)$ and the posterior probability $P(X=1|D)$. It is essential to determine the probability distribution of X .

$$\begin{aligned}
 \beta(f | D) &= \frac{P(D | f) \beta(f)}{P(D)} \quad (\text{Bayes' law}) \\
 &= \frac{f^s (1-f)^t \beta(f)}{E(f^s (1-f)^t)} \quad (\text{due to } P(D | f) = \prod_{i=1}^M P(X^i | f) = f^s (1-f)^t \text{ and apply formula 2.1.7}) \\
 &= \frac{f^s (1-f)^t \frac{\Gamma(N)}{\Gamma(a)\Gamma(b)} f^{a-1} (1-f)^{b-1}}{\frac{\Gamma(N)}{\Gamma(N+M)} \frac{\Gamma(a+s)\Gamma(b+t)}{\Gamma(a)\Gamma(b)}} \quad (\text{apply formula 2.1.1 \& 2.1.5}) \quad (2.2.1) \\
 &= \frac{\Gamma(N+M)}{\Gamma(a+s)\Gamma(b+t)} f^{a+s-1} (1-f)^{b+t-1} = \beta(f; a+s, b+t)
 \end{aligned}$$

Then the posterior density function is $\beta(f; a+s, b+t)$ where the prior density function is $\beta(f; a, b)$. According to formula 2.1.4, the posterior probability:

$$P(X=1|D) = E(\beta(f|D)) = \frac{a+s}{a+s+b+t} = \frac{a+s}{N+M} \tag{2.2.2}$$

In general, you should merely engrave the formula 2.1.1, 2.1.4, 2.2.1, 2.2.2 and the way to recognize prior density function, prior probability of X and posterior density function, posterior probability of X , respectively on your memory.

2.3 Expanding Augmented Bayesian Network with Multi-Hypothesis Nodes

Suppose we have a BN with two binary random variables and there is conditional dependence assertion between these nodes. See the network and CPT (s) in the figure below

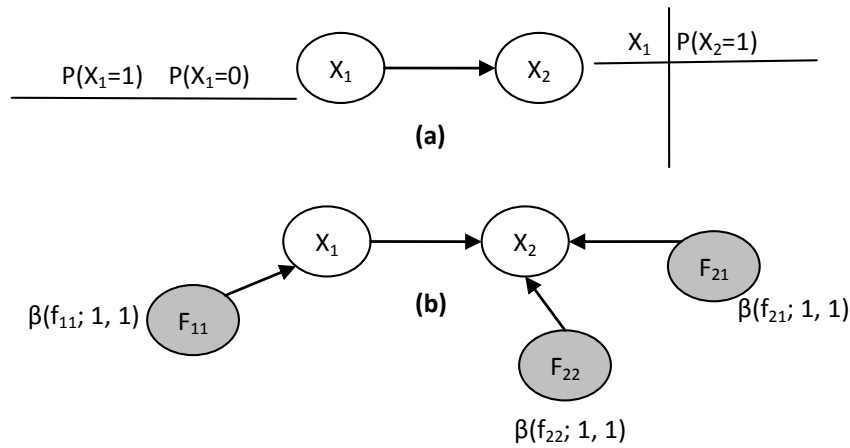


Figure 2.3.1: (a) Bayesian network and (b) expended augmented Bayesian network

For every node (variable) X_i , we add dummy parent nodes to X_i , obeying two ways below:

- If X_i has no parent (not conditionally dependent on any others), we add only one dummy variable denoted F_{i1} having the probability density function $\beta(f_{i1}; a_{i1}, b_{i1})$ so as to: $P(X_i=1|f_{i1})=f_{i1}$
- If X_i has a set of k_i parents and each parent pa_{il} ($l=1, k_i$) is binary, we add a set of $c_i=2k_i$ dummy variables $F_i = \{f_{i1}, f_{i2}, \dots, f_{i_{c_i}}\}$, in turn, instantiations of parents $PA_i = \{pa_{i1}, pa_{i2}, pa_{i3}, \dots, pa_{i_{c_i}}\}$. In other words, c_i denotes the number of instantiations of the parents PA_i . We have

$$P(X_i=1|pa_{i1}, f_{i1}, \dots, f_{i_{c_i}}) = f_{ij}$$

where

$$\beta(f_{ij}) = \frac{\Gamma(N_{ij})}{\Gamma(a_{ij})\Gamma(b_{ij})} f^{a_{ij}-1} (1-f_{ij})^{b_{ij}-1}$$

All f_{ij} have no parent and are mutually independent, so, $\beta(f_{i1}, f_{i2}, \dots, f_{i_{c_i}}) = \beta(f_{i1}) \beta(f_{i2}) \dots \beta(f_{i_{c_i}})$. Besides this local parameter independence, we have the global parameter independence if reviewing all variables X_i s, such below:

$$\beta(F_1, F_2, \dots, F_n) = \beta(f_{11}, f_{12}, \dots, f_{1_{c_1}}) \beta(f_{21}) \beta(f_{22}) \dots \beta(f_{n_{c_n}})$$

All variables X_i and their dummy variables form the expended augmented BN representing the trust BN in figure 4. In the trust BN, the conditional probability of variable X_i with the instantiation of its parent pa_{ij} , in other words, the ij^{th}

conditional distribution is given by $P(X_i=1 | pa_{ij}=1) = E(F_{ij}) = \frac{a_{ij}}{N_{ij}}$ (3.8), that's to say the expected value of F_{ij} .

Proof: We have following equation

$$P(X_i = 1 | pa_{ij} = 1) = \int_0^1 \dots \int_0^1 P(X_i = 1 | pa_{ij} = 1, f_{i1}, \dots, f_{ic_i}) \beta(f_{i1}) \dots \beta(f_{ic_i}) df_{i1} \dots df_{ic_i}$$

$$= \int_0^1 \dots \int_0^1 f_{ij} \beta(f_{i1}) \dots \beta(f_{ic_i}) df_{i1} \dots df_{ic_i} = E(F_{ij})$$

(due to F_{ij} (s) are mutually independent, $P(X_i = 1 | pa_{ih_i} = 1, f_{i1}, \dots, f_{ic_i}) = P(X_i = 1 | pa_{ij} = 1, f_{ij}) = f_{ij}$)

Suppose we perform M trials of random process, the outcome of i^{th} trial which is BN like figure 2.3.1 is represented as a

random vector $X^{(u)} = \begin{pmatrix} X_1^{(u)} \\ \dots \\ X_n^{(u)} \end{pmatrix}$ containing all hypothesis variables in network. $X^{(u)}$ is also called evidence vector (or

evidence, briefly). M trials constitute the sample of size M which is the set of random vectors denoted as $D = \{X^{(1)}, X^{(2)}, \dots, X^{(M)}\}$. D is also called evidence matrix. We review only in case of binomial sample; it means that D is the binomial BN sample of size M . For example, this sample corresponding to the network in figure 4 is showed below:

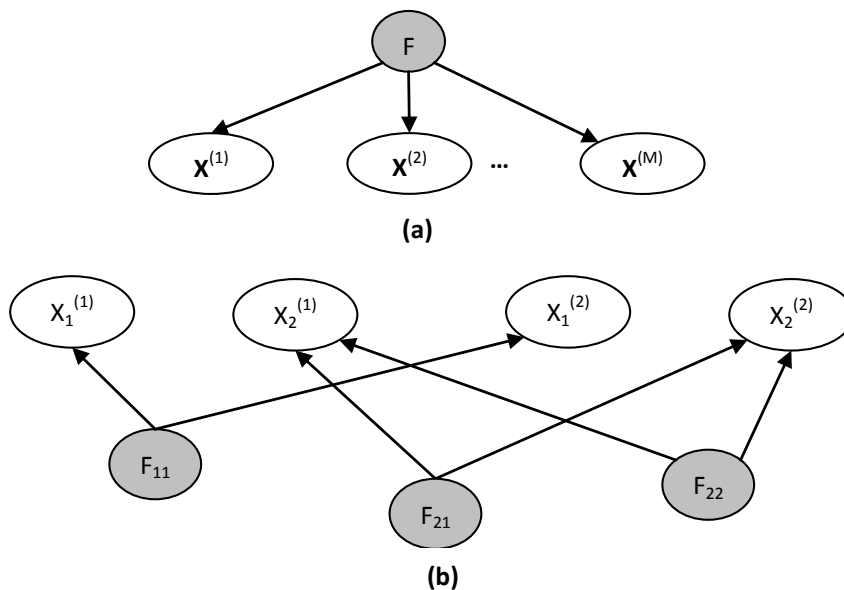


Figure 2.3.2: Expanded binomial BN sample of size M

After occurring M trial, the augmented BN was updated and dummy variables' density functions and hypothesis variables' conditional probabilities changed. We need to compute the posterior density function $\beta(f_{ij}|D)$ of each dummy

variable F_{ij} and the posterior condition probability $P(X_i=1/ pa_{ij} = 1, D)$ of each variable X_i . Note that the samples $X^{(u)}$ are mutually independent with all given F_{ij} . We have,

$$\prod_{u=1}^M P(X_i^{(u)} | pa_i, F_i) = \prod_{j=1}^{c_i} (f_{ij}^{s_{ij}})(1 - f_{ij})^{t_{ij}}$$

where

- c_i is the number of instances of $X_i^{(u)}$'s parents. In binary case, each $X_i^{(u)}$ (s) parent has two instances/values, namely, 0 and 1.
- s_{ij} , respective to f_{ij} , is the number of all evidences that variable $X_i = 1$ and $pa_{ij} = 1$
- t_{ij} , respective to f_{ij} , is the number of all evidences that variable $X_i = 1$ and $pa_{ij} = 0$.

Proof: We have,

$$P(D | F_1, \dots, F_n) = \prod_{i=1}^n \prod_{u=1}^M \Pr(X_i^{(u)} | pa_i, F_i) = \prod_{i=1}^n \prod_{j=1}^{c_i} (f_{ij}^{s_{ij}})(1 - f_{ij})^{t_{ij}} \quad (2.3.1)$$

$$P(D) = \prod_{i=1}^n \prod_{j=1}^{c_i} E(f_{ij}^{s_{ij}}(1 - f_{ij})^{t_{ij}}) \quad (2.3.2)$$

$$P(D) = \prod_{i=1}^n \left(\int_{F_i} \prod_{u=1}^M P(X_i^{(u)} | pa_i, F_i) \beta(F_i) dF_i \right)$$

(due to the law of total probability and the joint probability distribution)

$$= \prod_{i=1}^n \left(\int_{F_i} \prod_{j=1}^{c_i} (f_{ij}^{s_{ij}})(1 - f_{ij})^{t_{ij}} \beta(F_i) dF_i \right)$$

$$\left(\text{applying formula } \prod_{u=1}^M P(X_i^{(u)} | pa_i, F_i) = \prod_{j=1}^{c_i} (f_{ij}^{s_{ij}})(1 - f_{ij})^{t_{ij}} \right)$$

$$= \prod_{i=1}^n \prod_{j=1}^{c_i} \int_0^1 (f_{ij}^{s_{ij}})(1 - f_{ij})^{t_{ij}} \beta(f_{ij}) df_{ij}$$

$$= \prod_{i=1}^n \prod_{j=1}^{c_i} E(f_{ij}^{s_{ij}}(1 - f_{ij})^{t_{ij}})$$

There is the question “how to determine $E(f_{ij}^{s_{ij}}(1 - f_{ij})^{t_{ij}})$ ”. Applying formula 2.1.5, we have:

$$E(f_{ij}^{s_{ij}}(1 - f_{ij})^{t_{ij}}) = \frac{\Gamma(N_{ij})}{\Gamma(N_{ij} + M_{ij})} \frac{\Gamma(a_{ij} + s_{ij})\Gamma(b_{ij} + t_{ij})}{\Gamma(a_{ij})\Gamma(b_{ij})} \quad (2.3.3)$$

where $N_{ij}=a_{ij}+b_{ij}$ and $M_{ij}=s_{ij}+t_{ij}$

2.4 Updating Posterior Density Function with Multi-Node Bayesian Network

$$\beta(f_{ij} | D) = \frac{(f_{ij}^{s_{ij}})(1 - f_{ij})^{t_{ij}} \beta(f_{ij})}{E(f_{ij}^{s_{ij}}(1 - f_{ij})^{t_{ij}})} = \text{beta}(f_{ij}; a_{ij} + s_{ij}, b_{ij} + t_{ij}) \quad (2.4.1)$$

Proof: In the light of Baye’s law we have

$$\begin{aligned} \beta(f_{mn} | D) &= \frac{P(D | f_{mn})\beta(f_{mn})}{P(E)} \\ &= \frac{\left(\int_0^1 \dots \int_0^1 P(D | F_1, F_2, \dots, F_n) \prod_{ij \neq mn} \beta(f_{ij}) df_{ij} \right) \beta(f_{mn})}{P(E)} \quad (\text{law of total probability}) \\ &= \frac{(f_{mn})^{s_{mn}} (1 - f_{mn})^{t_{mn}} \left(\prod_{ij \neq mn} \int_0^1 (f_{ij})^{s_{ij}} (1 - f_{ij})^{t_{ij}} \beta(f_{ij}) df_{ij} \right) \beta(f_{mn})}{\prod_{i=1}^n \prod_{j=1}^{c_i} E(f_{ij}^{s_{ij}} (1 - f_{ij})^{t_{ij}})} \\ & \quad (\text{apply formula 2.3.1, 2.3.2}) \\ &= \frac{(f_{mn})^{s_{mn}} (1 - f_{mn})^{t_{mn}} \beta(f_{mn})}{E(f_{mn}^{s_{mn}} (1 - f_{mn})^{t_{mn}})} \\ &= \frac{(f_{mn})^{s_{mn}} (1 - f_{mn})^{t_{mn}} \frac{\Gamma(N_{mn})}{\Gamma(a_{mn})\Gamma(b_{mn})} (f_{mn})^{a_{mn}-1} (1 - f_{mn})^{b_{mn}-1}}{\frac{\Gamma(N_{mn})}{\Gamma(N_{mn} + M_{mn})} \frac{\Gamma(a_{mn} + s_{mn})\Gamma(b_{mn} + t_{mn})}{\Gamma(a_{mn})\Gamma(b_{mn})}} \\ & \quad (\text{expansion of } \beta(f_{mn}) \text{ and applying formula 2.3.3 to } E(f_{mn}^{s_{mn}} (1 - f_{mn})^{t_{mn}})) \\ &= \frac{\Gamma(N_{mn} + M_{mn})}{\Gamma(a_{mn} + s_{mn})\Gamma(b_{mn} + t_{mn})} (f_{mn})^{a_{mn} + s_{mn} - 1} (1 - f_{mn})^{b_{mn} + t_{mn} - 1} \\ &= \text{beta}(f_{mn}; a_{mn} + s_{mn}, b_{mn} + t_{mn}) \end{aligned}$$

According to formula 2.2.2 and 2.4.1,

$$P(X_i=1 | pa_{ij}=1, D) = E(F_{ij}) = E(\beta(f_{ij}|D)) = \frac{a_{ij} + s_{ij}}{a_{ij} + s_{ij} + b_{ij} + t_{ij}} = \frac{a_{ij} + s_{ij}}{N_{ij} + M_{ij}} \quad (2.4.2)$$

In short, in case of binomial distribution, if we have the real/trust BN embedded in the expanded augmented network such as figure 3.4 and each dummy node F_{ij} has a prior beta distribution $\beta(f_{ij}; a_{ij}, b_{ij})$ and each hypothesis node X_i has the prior conditional probability $P(X_i=1 | pa_{ij}=1) = E(\beta(f_{ij})) = \frac{a_{ij}}{N_{ij}}$, the parameter learning process based on a set of evidences is to update the posterior density function $\beta(f_{ij}|D)$ and the posterior conditional probability $P(X_i=1 | pa_{ij}=1, D)$. Indeed, $\beta(f_{ij} | D) = \text{beta}(f_{ij}; a_{ij} + s_{ij}, b_{ij} + t_{ij})$ and $P(X_i=1 | pa_{ij}=1, D) = E(\beta(f_{ij}|D)) = \frac{a_{ij} + s_{ij}}{N_{ij} + M_{ij}}$

Example 2.4.1: Suppose we have the set of 5 evidences $D = \{X^{(1)}, X^{(2)}, X^{(3)}, X^{(4)}, X^{(5)}\}$ owing to network in figure 2.3.1

	X_1	X_2
$X^{(1)}$	$X_1^{(1)} = 1$	$X_2^{(1)} = 1$
$X^{(2)}$	$X_1^{(2)} = 1$	$X_2^{(2)} = 1$
$X^{(3)}$	$X_1^{(3)} = 1$	$X_2^{(3)} = 1$
$X^{(4)}$	$X_1^{(4)} = 1$	$X_2^{(4)} = 0$
$X^{(5)}$	$X_1^{(5)} = 0$	$X_2^{(5)} = 0$

Table 2.4.1: Set of evidences D corresponding to 5 trials (sample of size 5)

Note that the first evidence $X^{(1)} = \begin{pmatrix} X_1^{(1)} = 1 \\ X_2^{(1)} = 1 \end{pmatrix}$ implies that variable $X_2=1$ given $X_1=1$ occurs in the first trial. We need to

compute all posterior density functions $\beta(f_{11}/D)$, $\beta(f_{21}/D)$, $\beta(f_{22}/D)$ and all conditional probabilities $P(X_1=1)$, $P(X_2=1/X_1=1)$, $P(X_2=1/X_1=0)$ from prior density functions $\beta(f_{11}; 1, 1)$, $\beta(f_{21}; 1, 1)$, $\beta(f_{22}; 1, 1)$. In fact,

$$\begin{aligned} s_{11} &= 1+1+1+1+0=4 & t_{11} &= 0+0+0+0+1=1 \\ s_{21} &= 1+1+1+0+0=3 & t_{21} &= 0+0+0+0+1=1 \\ s_{22} &= 0+0+0+0+0=0 & t_{22} &= 0+0+0+0+1=1 \end{aligned}$$

$$\beta(f_{11}/D) = \beta(f_{11}; a_{11}+s_{11}, b_{11}+t_{11}) = \beta(f_{11}; 1+4, 1+1) = \beta(f_{11}; 5, 2)$$

$$\beta(f_{21}/D) = \beta(f_{21}; a_{21}+s_{21}, b_{21}+t_{21}) = \beta(f_{21}; 1+3, 1+1) = \beta(f_{21}; 4, 2)$$

$$\beta(f_{22}/D) = \beta(f_{22}; a_{22}+s_{22}, b_{22}+t_{22}) = \beta(f_{22}; 1+0, 1+1) = \beta(f_{22}; 1, 2)$$

and $P(X_1=1)$, $P(X_2=1/X_1=1)$, $P(X_2=1/X_1=0)$ are expectations of $\beta(f_{11}/D)$, $\beta(f_{21}/D)$, $\beta(f_{22}/D)$. Then,

$$P(X_1=1) = \frac{5}{5+2} = \frac{5}{7} \quad P(X_2=1/X_1=1) = \frac{4}{4+2} = \frac{2}{3} \quad P(X_2=1/X_1=0) = \frac{1}{1+2} = \frac{1}{3}$$

Network in figure 2.3.1 changed as follows:

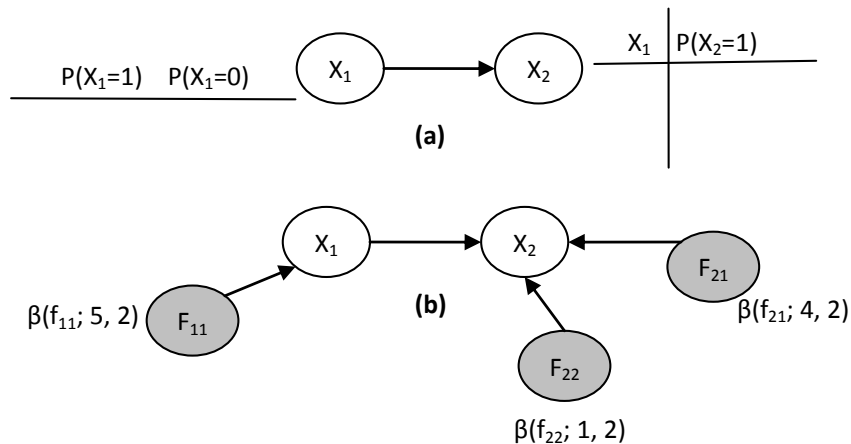


Figure 2.4.2: Updated version of BN (a) and augmented BN (b) in figure 2.3.1

2.5 Parameter Learning in Case of Data Missing

In practice there are some evidences in D such as $X^{(i)}$ (s) which lack information and thus, it stimulates the question “How to update network from data missing”. We must address this problem by artificial intelligence techniques, namely, expectation maximization (EM) algorithm – a famous technique solving estimation of data missing.

Example 2.5.1: Like above example, we have the set of 5 evidences $D = \{X^{(1)}, X^{(2)}, X^{(3)}, X^{(4)}, X^{(5)}\}$ along with network in figure 4 but the evidences $X^{(2)}$ and $X^{(5)}$ have not data yet.

	X_1	X_2
$X^{(1)}$	$X_1^{(1)} = 1$	$X_2^{(1)} = 1$
$X^{(2)}$	$X_1^{(2)} = 1$	$X_2^{(2)} = v_1?$
$X^{(3)}$	$X_1^{(3)} = 1$	$X_2^{(3)} = 1$
$X^{(4)}$	$X_1^{(4)} = 1$	$X_2^{(4)} = 0$
$X^{(5)}$	$X_1^{(5)} = 0$	$X_2^{(5)} = v_2?$

Table 2.5.1: Set of evidences D (for network in figure 4) with data missing

As known, s_{21} , t_{21} and s_{22} , t_{22} can't be computed directly, it means that it is not able to compute directly the posterior density functions $\beta(f_{21}/D)$ and $\beta(f_{22}/D)$. In evidence $X^{(2)}$, v_1 must be determined. Obviously, v_1 obtains one of two values which are according to two situations respectively:

- $X_1^{(2)} = 1$ and $X_2^{(2)} = 1$, it is easy to infer that $v_1 = P(X_2^{(2)} = 1 | X_1^{(2)} = 1) = E(\beta_{21}) = \frac{a_{21}}{a_{21} + b_{21}} = 1/2$
- $X_1^{(2)} = 1$ and $X_2^{(2)} = 0$, it is easy to infer that $v_1 = P(X_2^{(2)} = 0 | X_1^{(2)} = 0) = E(\beta_{22}) = \frac{a_{22}}{a_{22} + b_{22}} = 1/2$

We split $X^{(2)}$ into two $X^{(2)}$ (s) corresponding to two above situations in which the probability of occurrence of $X_2 = 1$ given $X_1 = 1$ is estimated as $1/2$ and the probability of occurrence of $X_2 = 0$ given $X_1 = 1$ is also considered as $1/2$. We perform similarly this task for $X^{(5)}$.

	X_1	X_2
$X^{(1)}$	$X_1^{(1)} = 1$	$X_2^{(1)} = 1$
$X^{(2)}$	$X_1^{(2)} = 1$	$X_2^{(2)} = 1/2$
$X^{(2)}$	$X_1^{(2)} = 1$	$X_2^{(2)} = 1/2$
$X^{(3)}$	$X_1^{(3)} = 1$	$X_2^{(3)} = 1$
$X^{(4)}$	$X_1^{(4)} = 1$	$X_2^{(4)} = 0$
$X^{(5)}$	$X_1^{(5)} = 0$	$X_2^{(5)} = 1/2$
$X^{(5)}$	$X_1^{(5)} = 0$	$X_2^{(5)} = 1/2$

Table 2.5.2: New split evidences D' for network in figure 2.4.2

So, we have $\left(\begin{matrix} s'_{21} = 1 + \frac{1}{2} + 1 = \frac{5}{2} \\ t'_{21} = \frac{1}{2} + 1 = \frac{3}{2} \end{matrix} \right)$ and $\left(\begin{matrix} s'_{22} = \frac{1}{2} \\ t'_{22} = \frac{1}{2} \end{matrix} \right)$ where s'_{21} , t'_{21} , s'_{22} , t'_{22} are the counts in D' . Then

$$\beta(f_{21}/D) = \beta(f_{21}; a_{21} + s'_{21}, b_{21} + t'_{21}) = \beta(f_{21}; 1 + 5/2, 1 + 3/2) = \beta(f_{21}; 7/2, 5/2)$$

$$\beta(f_{22}/D) = \beta(f_{22}; a_{22} + s'_{22}, b_{22} + t'_{22}) = \beta(f_{22}; 1 + 1/2, 1 + 1/2) = \beta(f_{22}; 3/2, 3/2)$$

$$P(X_2 = 1 | X_1 = 1) = E(\beta(f_{21}/D)) = \frac{7/2}{7/2 + 5/2} = \frac{7}{12} \quad P(X_2 = 0 | X_1 = 1) = E(\beta(f_{22}/D)) = \frac{3/2}{3/2 + 3/2} = \frac{1}{2}$$

If there are more evidences, this task repeated more and more brings out the EM algorithm having two steps.

1. **Step 1.** We compute s'_{ij} and t'_{ij} based on the expected value of given $\beta(f_{ij})$, $s'_{ij} = E(\beta(f_{ij}))$ and $t'_{ij} = 1 - E(\beta(f_{ij}))$. Next, replacing missing data by s'_{ij} and t'_{ij} . This step is called Expectation step.
2. **Step 2.** We determine the posterior density function f_{ij} by computing its parameters $a_{ij} = a_{ij} + s'_{ij}$ and $b_{ij} = b_{ij} + t'_{ij}$. Note that s'_{ij} and t'_{ij} are recomputed absolutely together on occurrence of s'_{ij} and t'_{ij} . Terminating algorithm if the

stop condition (for example, the number of iterations approaches k times) becomes true, otherwise, reiterating step 1. This step is called the Maximization step.

After k^{th} iteration, we have $\lim_{k \rightarrow \infty} Expectation_{ij} = \lim_{k \rightarrow \infty} \frac{a_{ij} + s_{ij}^{(k)}}{a_{ij} + s_{ij}^{(k)} + b_{ij} + t_{ij}^{(k)}}$ which will approach a certain limit. Don't worry about the case of infinite iterations; we will obtain approximate s'_{ij} , t'_{ij} , posterior f_{ij} if k is large enough due to certain value of $\lim_{k \rightarrow \infty} Expectation_{ij}$

3. Structure Learning

As discussed in Maurya et al. [5], directed acyclic graph (DAG) that contain the same given nodes V are Markov equivalent if they satisfy Markov condition and have the same d-separations. In other words, they entail the same conditional independences and their joint conditional probabilities are identical. Let the pattern gp represent this Markov equivalent DAG (s). Such pattern gp is called Markov *equivalent class*. Of course given a set of nodes V , there are a lot of equivalent classes. Let GP be random variable whose values are pattern gp . The basic idea of structure learning approaches is to find out the pattern gp that satisfy some condition best. Instead of searching many individual DAG According to given condition, there are two main learning approaches:

- Score-based approach: For each pattern $gp \in GP$, the gp which gains the maximal scoring criterion $score(D, gp)$ given training data set D is the best gp . Because the essence of score-based approach is find out the most likely structure, it is also called *model selection* approach.
- Constraint-based approach: Given a set of conditional independences (a set of d-separations), the best gp is the DAG which satisfy Markov condition overall and only these conditional independences. Such independences play the role of the “door latch” for learning algorithm.

Note that in structure learning context, Bayesian network or pattern gp is mentioned as a DAG.

3.1 Score-Based Approach

Given a set of random variables (nodes) $V = \{X_1, X_2, \dots, X_n\}$, let (G, P) be possible Bayesian network where P is joint conditional probability density and $G=(V, E)$ is the DAG. Let $(G, F^{(G)}, \beta^{(G)})$ be the augmented BN with equivalent sample size N where $F(G)$ is augmented variables (nodes) attached to every nodes in V and $\beta(G)$ represents beta distributions for augmented (see section about parameter learning). Pattern gp also represents Markov equivalent augmented BN. Score-based approach has three following steps:

1. **Step 1.** Suppose all augmented BN (s) has the same equivalent sample size N .
2. **Step 2.** Let r_i be the number of possible values of variable X_i . If X_i is binary then $r_i = 2$. Let q_i be the number of distinct instantiations of parents of X_i . For example, if X_i and its parents are binary and X_i have l parents then $q_i = 2^l$. All augmented variables F_{ij} representing the conditional probability of X_i given instantiation j of its parent are assigned to uniform distribution according to equivalent sample size N :

$$a_{ijk} = \frac{N}{r_i q_i}$$

3. **Step 3.** Given $D=\{X^{(1)}, X^{(2)}, \dots, X^{(M)}\}$ is the training data set size M , where $X^{(h)}$ is a trial. Note that $X^{(h)}=(X^{(h)}_1, X^{(h)}_2, \dots, X^{(h)}_n)$ is a n -dimension vector which is a outcome (instantiation) of variable X_i . $X^{(h)}_i$ has the same space to X_i . Each DAG gp which is connected by variables in V is assigned a value so-called scoring criterion $score(D, gp)$. This score is the posterior probability of gp given training data set D .

$$score(D, gp) = P(gp | D) = \frac{P(gp)P(D | gp)}{P(D)}$$

where

$$P(D | gp) = \prod_{i=1}^n \prod_{j=1}^{q_i} \left(\frac{\Gamma(N_{ij})}{\Gamma(N_{ij} + M_{ij})} \prod_{k=1}^{r_i} \frac{\Gamma(a_{ijk} + s_{ijk})}{\Gamma(a_{ijk})} \right)$$

$P(gp)$ is the prior probability of gp . $P(D)$ is constant.

In practice, $score(D, gp)$ is only dependent on $P(D/gp)$ when $P(D)$ is ignored and $P(gp)$ is initialized subjectively.

$$score(D, gp) \approx \prod_{i=1}^n \prod_{j=1}^{q_i} \left(\frac{\Gamma(N_{ij})}{\Gamma(N_{ij} + M_{ij})} \prod_{k=1}^{r_i} \frac{\Gamma(a_{ijk} + s_{ijk})}{\Gamma(a_{ijk})} \right) \quad (3.1.1)$$

Which gp gaining maximal $score(D, gp)$ is chosen.

Example 3.1.1: Suppose there are two variables X_1, X_2 , we don't know exactly their relationship but the training data D is observed as below:

X_1	1	1	1	1	0	0
X_2	0	0	0	1	1	0

Let gp_1 be the DAG in which X_1 is parent of X_2 ; otherwise let gp_2 be the DAG in which X_1 and X_2 are mutually independent. Given the sample size is $N = 4$

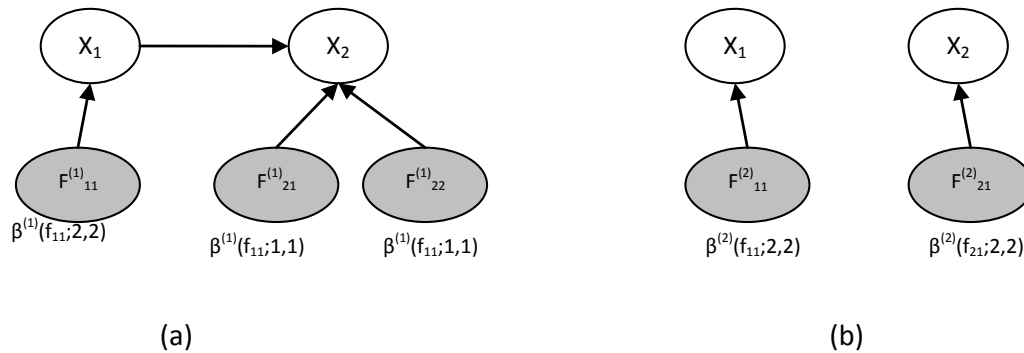


Figure 3.1.1: Augmented Bayesian networks of gp_1 (a) and gp_2 (b)

We have:

$$\begin{aligned} score(D, gp_1) &\approx P(D | gp_1) = \left(\frac{\Gamma(N_{11}^{(1)})}{\Gamma(N_{11}^{(1)} + M_{11}^{(1)})} \left(\frac{\Gamma(a_{11}^{(1)} + s_{11}^{(1)})}{\Gamma(a_{11}^{(1)})} \frac{\Gamma(b_{11}^{(1)} + t_{11}^{(1)})}{\Gamma(b_{11}^{(1)})} \right) \right) \\ &\times \left(\frac{\Gamma(N_{21}^{(1)})}{\Gamma(N_{21}^{(1)} + M_{21}^{(1)})} \left(\frac{\Gamma(a_{21}^{(1)} + s_{21}^{(1)})}{\Gamma(a_{21}^{(1)})} \frac{\Gamma(b_{21}^{(1)} + t_{21}^{(1)})}{\Gamma(b_{21}^{(1)})} \right) \right) \left(\frac{\Gamma(N_{22}^{(1)})}{\Gamma(N_{22}^{(1)} + M_{22}^{(1)})} \left(\frac{\Gamma(a_{22}^{(1)} + s_{22}^{(1)})}{\Gamma(a_{22}^{(1)})} \frac{\Gamma(b_{22}^{(1)} + t_{22}^{(1)})}{\Gamma(b_{22}^{(1)})} \right) \right) \\ &= \left(\frac{\Gamma(4)}{\Gamma(4+6)} \frac{\Gamma(2+4)}{\Gamma(2)} \frac{\Gamma(2+2)}{\Gamma(2)} \right) \times \left(\frac{\Gamma(2)}{\Gamma(2+4)} \frac{\Gamma(1+1)}{\Gamma(1)} \frac{\Gamma(1+3)}{\Gamma(1)} \right) \times \left(\frac{\Gamma(2)}{\Gamma(2+2)} \frac{\Gamma(1+1)}{\Gamma(1)} \frac{\Gamma(1+1)}{\Gamma(1)} \right) \\ &= \frac{\Gamma(4)\Gamma(4)}{\Gamma(10)} = 9.9 \times 10^{-5} \end{aligned}$$

$$\begin{aligned}
 score(D, gp_1) &\approx P(D | gp_2) = \left(\frac{\Gamma(N_{11}^{(1)})}{\Gamma(N_{11}^{(1)} + M_{11}^{(1)})} \left(\frac{\Gamma(a_{11}^{(1)} + s_{11}^{(1)})}{\Gamma(a_{11}^{(1)})} \frac{\Gamma(b_{11}^{(1)} + t_{11}^{(1)})}{\Gamma(b_{11}^{(1)})} \right) \right) \\
 &\times \left(\frac{\Gamma(N_{21}^{(1)})}{\Gamma(N_{21}^{(1)} + M_{21}^{(1)})} \left(\frac{\Gamma(a_{21}^{(1)} + s_{21}^{(1)})}{\Gamma(a_{21}^{(1)})} \frac{\Gamma(b_{21}^{(1)} + t_{21}^{(1)})}{\Gamma(b_{21}^{(1)})} \right) \right) \\
 &= \left(\frac{\Gamma(4)}{\Gamma(4+6)} \frac{\Gamma(2+4)}{\Gamma(2)} \frac{\Gamma(2+2)}{\Gamma(2)} \right) \times \left(\frac{\Gamma(4)}{\Gamma(4+6)} \frac{\Gamma(2+2)}{\Gamma(2)} \frac{\Gamma(2+4)}{\Gamma(2)} \right) \\
 &= \left(\frac{\Gamma(6)\Gamma(4)\Gamma(4)}{\Gamma(10)} \right)^2 = 1190 \times 10^{-5}
 \end{aligned}$$

Because $score(D, gp_1)$ is larger than $score(D, gp_2)$, the equivalent pattern gp_1 is chosen as Bayesian network appropriate to training data set.

In above example we recognize that it is difficult to determine all DAG (s). So the score-based approach becomes ineffective in case of many variables. The number of DAG (s) which is surveyed to compute scoring criterion gets huge. It is impossible to do brute-force searching over DAG (s) space. There are some heuristic algorithms to reduce whole DAG (s) space to smaller space so-called candidate set of DAG (s) obeying some restriction, for example, the prior ordering of variables. Such heuristic algorithms are classified into approximate learning. The global score can be defined as a product of local scores:

$$score(D, gp) = \prod_{i=1}^n score(D, X_i, PA_i)$$

Where $score(D, X_i, PA_i)$ is the local score of X_i given its parents PA_i .

$$score(D, X_i, PA_i) = P(X_i | PA_i, D) = \prod_{j=1}^{q^{(PA_i)}} \left(\frac{\Gamma(\sum_{k=1}^{r_i} a_{ijk})}{\Gamma(\sum_{k=1}^{r_i} a_{ijk} + \sum_{k=1}^{r_i} s_{ijk})} \prod_{k=1}^{r_i} \frac{\Gamma(a_{ijk} + s_{ijk})}{\Gamma(a_{ijk})} \right) \quad (3.1.2)$$

Let $q^{(PA_i)}$ be the number of distinct instantiations of parents of X_i

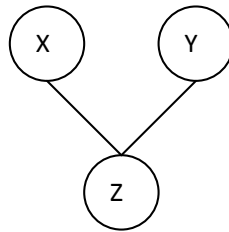
The K_2 algorithm tries to find out the pattern DAG gp whose each variable X_i maximizes local score $score(D, X_i, PA_i)$ instead of discovering all DAG (s). It means that K_2 algorithm finds out optimal parents PA_i of each X_i . Note that it expects that the global score will be approached by maximizing each partial local score. K_2 algorithm has following steps:

- **Step 1.** Suppose there is an ordering (X_1, X_2, \dots, X_n) . There is no backward edge, for example, the edge $X_i \leftarrow X_j$ (if exist) where $i < j$ is invalid. Let $Pre(X_i)$ be the set of previous nodes of X_i in ordering. Let PA_i is parents of X_i . K_2 's mission is to find out PA_i for every X_i . Firstly, each PA_i (s) is set to be empty and each local $score(D, X_i, PA_i)$ is initialized with such empty PA_i .
- **Step 2.** Each X_i is visited according to the ordering. When X_i is visited, which node in $Pre(X_i)$ that maximizes the local $score(D, X_i, PA_i)$ is added to PA_i .
- **Step 3.** Algorithm terminates when no node is added to PA_i .

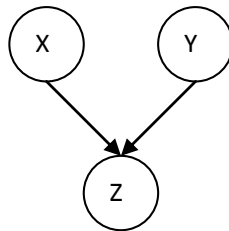
3.2 Constraint-Based Approach

Given (G, P) let IND_p be a set of conditional independences. IND_p is considered as the set of constraints. Constraint-based approach tries to find out the DAG that satisfies IND_p based on theory of d -separation. In other words the set of d -separations of the best DAG pattern are the same as IND_p .

Example 3.2.2: Suppose we have $V = \{X, Y, Z\}$ and $IND_p = \{I(X, Y)\}$. Because X and Z is n't d -separated from any set, there must be a link between X and Z . In similar way, there must be a link between Y and Z . We have:



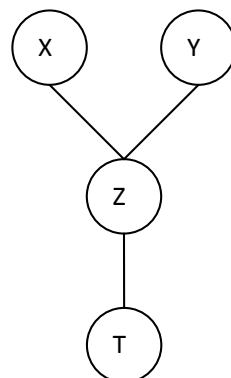
Because $X-Z-Y$ is uncoupled chain and there is a d -separation $I(X,Y)$, the chain $X-Z-Y$ should be converged.



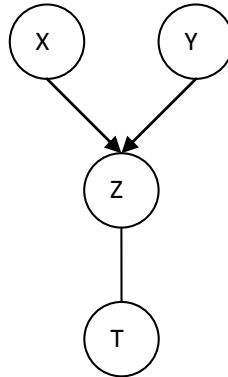
If the number of variables is large we need effective algorithms. The simple algorithm includes two steps:

1. Firstly, the structure of DAG is drafted as “skeleton”. If there is no conditional independence relating to X_i and X_j then the link between them is created. So skeleton is the undirected graph which contains variables (nodes) and links.
2. The second step is to determine direction of links by applying four following rules in sequence rule 1, rule 2, rule 3, rule 4:
 - **Rule 1:** If the uncoupled chain $X-Z-Y$ exists and Z isn't in any set that d -separate X from Y then this chain is assumed convergent: $X \rightarrow Z \leftarrow Y$
 - **Rule 2:** If the uncoupled chain $X \rightarrow Z \rightarrow Y$ exists (having an edge $X \rightarrow Z$) then this chain is assumed serial path: $X \rightarrow Z \rightarrow Y$.
 - **Rule 3:** If the edge $X \rightarrow Y$ caused a directed cycle at a position in network then it is reversed: $X \leftarrow Y$. This rule is applied to remove directed cycles so that the expected BN is a DAG.
 - **Rule 4:** If all rules 1, 2, 3 are consumed the all remaining links have arbitrary direction.

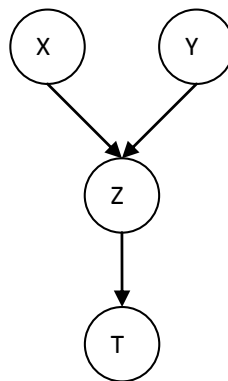
Example 3.2.3: Suppose we have $V = \{X, Y, Z, T\}$ and $IND_p = \{I(X,Y), I(X,T), I(Y,T)\}$. Because there is no conditional independence between X and Y , between Z and T , the “skeleton” is drafted as below:



Applying *rule 1*: Because the uncoupled chain $X-Z-Y$ exists and Z isn't in any set that d-separate X from Y , this chain is assumed convergent: $X \rightarrow Z \leftarrow Y$



Applying *rule 2*: Because the uncoupled chain $X \rightarrow Z \rightarrow T$ exists, we have the assumed serial path: $X \rightarrow Z \rightarrow T$.



4. Conclusions

In this paper, parameter learning and structure learning in Bayesian network inference have been discussed exhaustively. The essence of parameter learning and structure learning as two significant domains of Bayesian network Bayesian network are presented successfully with numerical illustrations. The parameter learning and structure learning domains indicate how to build up Bayesian Network. The ideology of Bayesian Network is to apply a mathematical inference tool (namely Bayesian rule) into a graph with expectation of extending and enhancing the ability of such tool so as to solve realistic problems, especially diagnosis domain. After reviewing and examining the article on survey report of parameter and structure learning in Bayesian network inference, we draw mainly following conclusions:

- The parameter and structure learning become difficult when training data is missing (not complete). Missing data problem is introduced in section 2.5 but its detail goes beyond this report. We expect that we may have a chance to discuss about it in connection to some relevant work.
- The first-order Markov condition has important role in Bayesian network study when there is an assumption “nodes are dependent on only their direct parents”. If the first-order Markov condition is not satisfied, many inference and learning algorithms go wrong. We think that BN will get more potential and enjoyable if first-order (Markov) condition is replaced by n-order condition.
- In the process of developing BN, there are many problems involving in real number (continuous case) and nodes dependency, refer Maurya [5]. This report focuses on discrete case when the probability of each node is discrete CPT, not continuous PDF.

- The Bayesian network discussed here is “static” BN because the temporal relationships among nodes aren’t concerned. The “static” BN is represented at only one time point. Otherwise dynamic Bayesian network (DBN) aims to model the temporal relationships among nodes.
- The process of inference is concerned in time series; in some realistic case this is necessary. However the cost of inference and learning in DBN is much higher than BN because the size of DBN gets huge for long-time process. Because of the limitation of this report, the algorithm that keeps the size of DBN intact (not changed) isn’t introduced here.
- Finally, the essence of such algorithm is to take advantage of both Markov condition and knowledge (inference) accumulation. Due to the complexity of DBN, we should consider to choose which one (BN or DBN) to apply into concrete domain. It depends on domain and purpose chosen by the decision makers.

Acknowledgements



Dr. V. N. Maurya; principal author of the present paper and former founder Director at Vision Institute of Technology, Aligarh (Uttar Pradesh Technical University, Lucknow (India), former Principal/Director at Shekhawati Engineering College (Rajasthan Technical University, Kota) and former Professor & Dean Academics, Institute of Engineering & Technology, Sitapur, UP, India; is now the Professor & Head of Shekhawati Engineering College (Rajasthan Technical University, Kota). He is the Chief Editor of Editorial Board of American Journal of Modeling and Optimization; Science and Education Publishing, New York, USA and Statistics, Optimization and Information Computing; International Academic Press, Hong Kong and Advisory Editor of World Research Journal of Numerical Analysis and Mathematical Modeling; Bioinfo Publications, Pune, India and Member of Editorial and Reviewer Board of over 50 Indian and Foreign International journals published by leading publishers of USA, Italy, Hong Kong, Austria, U.K., Algeria, Nigeria and other European and African countries.. He has been associated with leading Indian Universities-U. P. Technical University, Lucknow during 2005-06 and Chhatrapati Shahu Ji Maharaj University, Kanpur for three terms during 2000-2004 for significant contribution of his supervision as Head Examiner of Central Evaluation for Theory Examinations of UG (B.Tech./B.Pharm.) and PG (MA/M.Sc.) programmes.

Dr. Maurya was born on 15th July 1974 and he is having an outstanding academic record. He earned his M.Sc. and Ph.D. Degree in Mathematics & Statistics with specialization in Operations Research with First Division from Dr. Ram Manohar Lohia Avadh University, Faizabad, UP, India in the year 1996 and 2000 respectively and thereafter he accomplished another two years Master’s Professional Degree-MBA with First Division (B+ Grade) with specialization in Computer Science from NU, California, USA in 2003. His Ph.D. Thesis titled as “A study of use of stochastic processes in some queueing models” submitted to Department of Mathematics & Statistics, Dr. R.M.L. Avadh University, Faizabad under supervision of Prof. (Dr.) S.N. Singh, Ph.D. (BHU); was offered to publish in Scholar’s Press Publishing Co., Saarbrucken, Germany in view of his excellent research work. Since his primary education to higher education, he has been a meritorious scholar and recipient of meritorious scholarship. He started his teaching career as Lecturer in 1996 to teach post-graduate courses MBA, MCA and M.Sc. and later he was appointed as Professor & Head, Department of Applied Sciences and Engineering at Singhania University, Rajasthan in the year 2004. Since then, Prof. V. N. Maurya has rendered his services as Professor & Head/Dean as well as keen Researcher for Post-Doctoral research and he has devoted his entire scientific and professional career in teaching at various premier technical institutions of the country such as at Haryana College of Technology & Management, Kaithal (Kuruchhetra University, Kuruchhetra); Institute of Engineering & Technology, Sitapur and United College of Engineering & Research, Allahabad. On the basis of significant research work carried out by him in the last 17 years of his professional career, Prof. V. N. Maurya has authored three textbooks and published more than 55 scientific and academic research papers including 25 research papers as Principal Author based on his Post-Doctoral work and D.Sc. Thesis in Indian and Foreign leading International Journals in the field of Mathematical and Management Sciences, Industrial Engineering & Technology. Some of his published research papers in India, USA, Algeria, Malaysia and other European and African countries are recognized as innovative contributions in the field of Mathematical and Physical Sciences, Engineering & Technology. Prof. V. N. Maurya is an approved Supervisor of UGC recognized various Indian Universities for Research Programs leading to M. Phil. & Ph.D. such as Shridhar University, Pilani (Rajasthan), Singhania University, Rajasthan and CMJ University, Sillong, Meghalaya and JJT University Jhunjhunu, Rajasthan and U.P. Technical University

Lucknow etc. and since last 7 years, he is actively engaged as Research Supervisor of M. Phil. & Ph.D. Scholars in wide fields of Operations Research, Optimization Techniques, Statistical Inference, Applied Mathematics, Operations Management and Computer Science. He has guided as Principal Supervisor and Co-Supervisor to several Research Scholars of M. Phil. and Ph.D.

During his tenure as the Director, Vision Institute of Technology, Aligarh (Uttar Pradesh Technical University, Lucknow) and as the Principal, Shekhawati Engineering College (Rajasthan Technical University, Kota); massive expansion of infrastructure, research facilities, laboratories upgradation/augmentation and other relevant facilities and services for B.Tech./M.Tech./MBA academic programmes in different branches had taken place to accommodate and facilitate the campus students. His major contribution was to enhance the result of weaker students of their University Examination. He planned strategically and developed some tools and methods and then finally implemented for getting successfully considerable better result of campus students particularly in numerical papers.

Prof. Maurya is also on active role of Fellow/Senior/Life Member of various reputed National and International professional bodies of India and abroad including Operations Research Society of India, Kolkata; Indian Society for Technical Education, New Delhi; Indian Association for Productivity, Quality & Reliability, Kolkata; Indian Society for Congress Association, Kolkata; International Indian Statistical Association, Kolkata; All India Management Association, New Delhi; Rajasthan Ganita Parishad, Ajmer and International Association of Computer Science & Information Technology, Singapore etc.



Diwinder Kaur Arora; co-author of the present paper accomplished MBA Degree with specialization in Human Resources from Pondicherry Central University, Pondicherry and she graduated with B.Sc. (Medical/ZBC Group) Degree in 1987 from Kanpur University, Kanpur, India and did Diploma also from Government Polytechnic College, Amethi, U.P. throughout in First Division. She has vast experience of more than 22 years of general administration and management as Police Officer of Central Reserve Police Force, Ministry of Home Affairs, Govt. of India. She was selected as Assistant Sub-Inspector (Non-Gazetted Officer) in 1991 and after successful completion of her services she was promoted as Sub-Inspector in 2004 and since 2012 she is working in the grade of Inspector of Police at Group Centre, Central Reserve Police Force, Lucknow, U.P. Apart from this, she has published more than 12 research papers in Indian and Foreign International journals of repute in the field of Management, Information Technology and Physical Sciences such as in World of

Sciences Journal, Engineers Press Publishing Group, Vienna, Austria; International Journal of Engineering Research and Technology, Engineering Science & Research Support Academy (ESRSA), Vadodara, India; International Journal of Electronics Communication and Electrical Engineering, Algeria; International Journal of Information Technology & Operations Management, Academic and Scientific Publisher, New York, USA.



Er. Avadhesh Kumar Maurya; co-author of the paper is having an outstanding academic record and accomplished his M.Tech. Degree with specialization in Digital Communication from Uttarakhand Technical University, Dehradun, UK and he was graduated with B.Tech. Degree in Electronics and Communication Engineering from Rajasthan Technical University, Kota (Rajasthan). He is recipient of four First Divisions in his Student Career with flying colours. Since last one year, Er. A. K. Maurya is serving as Assistant Professor in Department of Electronics and Communication Engineering at Lucknow Institute of Technology, U.P. Technical University, Lucknow. Prior to assuming the post of Assistant Professor at Lucknow Institute of Technology, U.P., he served as a Network Engineer for two years in National Informatics Centre, Department of Information Technology, Govt. of India with collaboration of HCL Co. He has worked on some projects such as Movable Target Shooter using Ultrasonic Radar and Hartley Oscillator. Apart from this, he has got

industrial training in Door Darshan Kendra, Lucknow, U.P. in the field of TV Program Generation and Broadcasting of different channels for partial fulfilment of his Degree and published also over 15 research papers in various Indian and Foreign International journals of repute in the field of Electronics & Communication Engineering, Computer Science & Information Technology and Physical Sciences such as in International Journal of Electronics Communication and Electrical Engineering, Algeria; World of Sciences Journal, Engineers Press Publishing Group, Vienna, Austria; International Journal of Information Technology & Operations Management, Academic and Scientific Publisher, New York, USA; International Journal of Engineering Research and Technology, Engineering Science & Research Support

Academy (ESRSA), Vadodara, India; International Journal of Software Engineering & Computing, Serials Publications, New Delhi, India and many more.

References

- [1] David Heckerman, A tutorial on learning with Bayesian networks, Technical Report MSR-TR-95-06, Microsoft Research Advanced Technology Division, Microsoft Corporation
- [2] Jian Ni, Performance analysis and inference of communication networks, Ph.D. Thesis, Yale University, 2008
- [3] Maurya V. N., Arora Diwinder Kaur, Maurya A. K. & Gautam R.A., Exact modelling of annual maximum rainfall with Gumbel and Frechet distributions using parameter estimation techniques, World of Sciences Journal, Engineers Press Publishing Group, Vienna, Austria, Vol. 1, No. 2, 2013, pp.11-26
- [4] Maurya V. N., Arora Diwinder Kaur & Maurya A.K., A survey report on nonparametric hypothesis testing including ANOVA and goodness-fit-test, Communicated for publication, 2013
- [5] Maurya V. N., Arora Diwinder Kaur & Maurya A.K., Significant role and special aspects of inference mechanism in Bayesian network, International Open Journal of Operations Research, Vol.1, No.2, 2013, pp. 1-24
- [6] Maurya V. N., Inferences on operating characteristics of the queue in power supply problems, Journal of Decision and Mathematical Sciences, New Delhi, India, Vol. 11, No.1-3, 2006, pp. 105-112
- [7] Maurya V. N., Inferences on operating characteristics of the system size distribution in an $M/G/\infty:(\infty;GD)$ queueing system in equilibrium state, Acta Ciencia Indica Mathematics, Vol. XXXII M, No.3, 2006, pp. 1093-1100, ISSN: 0970-0455 (Citation No. 015879, Indian Science Abstract, Vol. 43, No. 16, 2007)
- [8] Richard E. Neapolitan, Learning Bayesian networks, North-eastern Illinois University Chicago, Illinois, 2003