The Accuracy of a Bayesian Network

Alexandros Pappas

November 2001

Supervisor: Dr. Duncan Gillies

Department of Computing
Imperial College of Science, Technology and Medicine
University of London

Abstract

A Bayesian network is a construct that represents a joint probability distribution, and can be used in order to model a given joint probability distribution.

A principal characteristic of a Bayesian network is the degree to which it models the given joint probability distribution accurately; the accuracy of a Bayesian network. Although the accuracy of a Bayesian network can be well defined in theory, it is rarely possible to determine the accuracy of a Bayesian network in practice for real-world applications. Instead, alternative characteristics of a Bayesian network, which relate to and reflect the accuracy, are used to model the accuracy of a Bayesian network, and appropriate measures are devised.

A popular formalism that adopts such methods to study the accuracy of a Bayesian network is the Minimum Description Length (MDL) formalism, which models the accuracy of a Bayesian network as the probability of the Bayesian network given the data set that describes the joint probability distribution the Bayesian network models. However, in the context of Bayesian Networks, the MDL formalism is flawed, exhibiting several shortcomings, and thus inappropriate for examining the accuracy of a Bayesian network.

An alternative framework for Bayesian Networks is proposed, which models the accuracy of a Bayesian network as the accuracy of the conditional independencies implied by the structure of the Bayesian network, and specifies an appropriate measure called the Network Conditional Independencies Mutual Information (NCIMI) measure. The proposed framework is inspired by the principles governing the field of Bayesian Networks, and is based on formal theoretical foundations.

Experiments have been conducted, using real-world problems, that evaluate both the MDL formalism and the proposed framework for Bayesian Networks. The experimental results support the theoretical claims, and confirm the significance of the proposed framework.

Contents

1. Introduction

1.1. Bayesian network

In essence, a Bayesian network is a construct that represents a joint probability distribution, and can be used in order to model a given joint probability distribution.

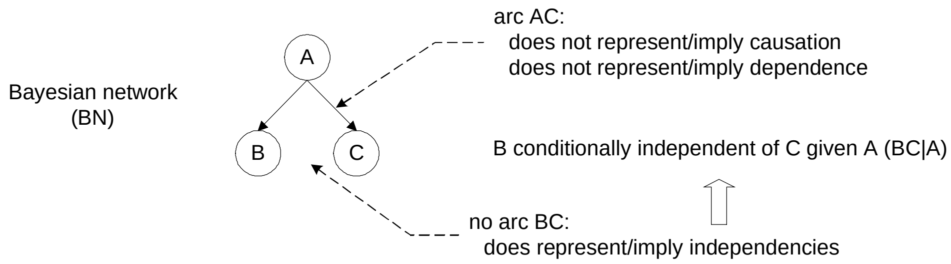


Figure 1

Definition 1.1 {1}

- V is a finite set of finite variables defined on the same probability space
- P is the joint probability distribution of V
- $G=(V,E)$ is a directed acyclic graph (DAG), where E are the edges
- $\forall v \in V : parents(v) \subset V$ is the set of all parents of v
- $\forall v \in V : descendants(v) \subset V$ is the set of all descendants of v
- $\forall v \in V : a(v) \subseteq V, a(v) = V - (descendants(v) \cup \{v\})$ is the set of variables in V excluding v and the descendants of v
- $\forall W : W \subseteq a(v)$, W and v are conditionally independent given the parents of v: if $P(parents(v)) > 0$, then $P(v \mid parents(v)) = 0$ or $P(W \mid parents(v)) = 0$ or $P(v \mid W \cup parents(v)) = P(v \mid parents(v))$

➢ $BN = (V,E,P)$ is a Bayesian network

If $parents(v) = \varnothing$, then W and v are unconditionally independent (conditionally independent given the empty set).

[1] [Neapolitan 1990] Definition 5.4.

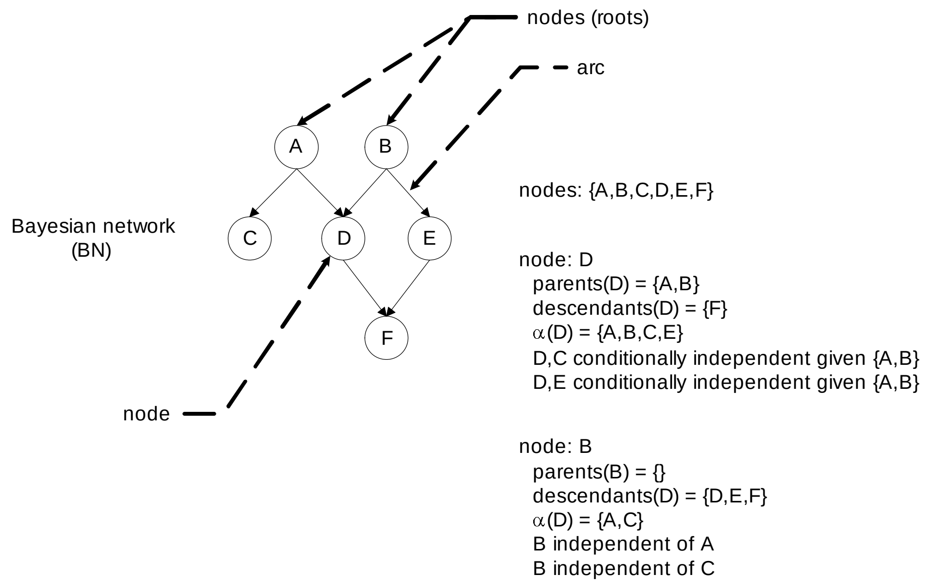


Figure 2

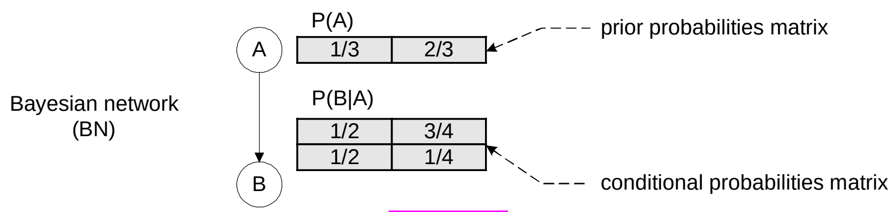


Figure 3

Theorem 1.1 [2]

For a Bayesian network ($BN = (V, E, P)$), the joint probability distribution the Bayesian network represents (P_{BN}) is given by:

$$P_{BN} = P(V) = \prod_{\substack{v \in V \\ P(parents(v))>0}} P(v \mid parents(v))$$

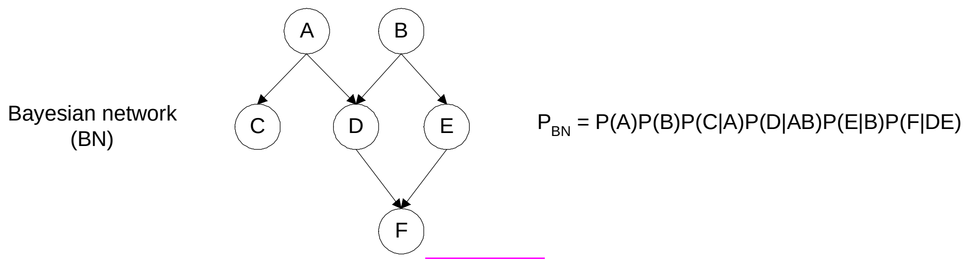


Figure 4

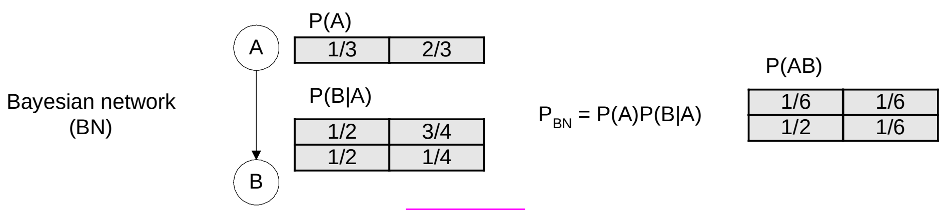


Figure 5

[2] [Neapolitan 1990] Theorem 5.1.

Theorem 1.2 [3]
For a joint probability distribution ($JPRD$), there exists a Bayesian network (BN) that models the joint probability distribution.
$\forall JPRD, \exists BN : P_{BN} = JPRD$

For a trivial proof of Theorem 1.2, consider any joint probability distribution and the corresponding fully connected Bayesian network*; a fully connected* Bayesian network *can model any joint probability distribution.*

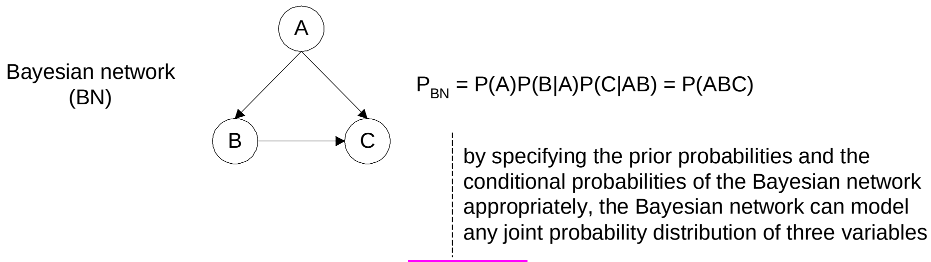


Figure 6

The significance of Theorem 1.2 is apparent, since it guarantees that for any joint probability distribution, there exists a Bayesian network that models the given joint probability distribution.
Theorem 1.2 constitutes a “proof of completeness” for the field of Bayesian Networks, indicating that it is always possible to use a Bayesian network instead of the joint probability distribution itself.

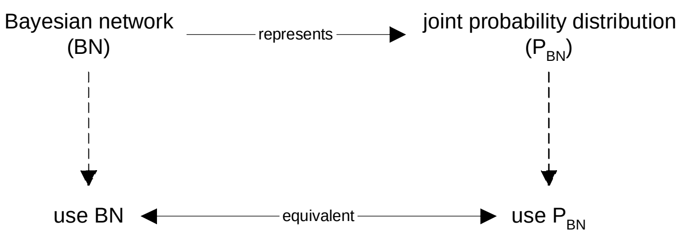


Figure 7

A Bayesian network represents a joint probability distribution, and so, in effect, a Bayesian network represents relationships between a set of variables.
A Bayesian network does not necessarily represent causation between a set of variables. For example, the parents of a node should not necessarily be considered causes of the node, although they can be interpreted as such in certain cases; instead, they should be viewed as shields against other influences.

A principal feature of a Bayesian network is the conditional independencies implied by the structure of the Bayesian network regarding the variables of the joint probability distribution it represents.
The absence of an arc (direct connection) between two nodes of a Bayesian network implies certain conditional independencies regarding these nodes.

3 [Neapolitan 1990] Theorem 5.4.

The structure of a Bayesian network does not imply dependencies regarding the variables of the joint probability distribution it represents.
The existence of an arc (direct connection) between two nodes of a Bayesian network does not imply dependence regarding these nodes.

All independencies implied by the structure of a Bayesian network are considered conditional; an unconditional independence can be viewed as a conditional independence given the empty set.

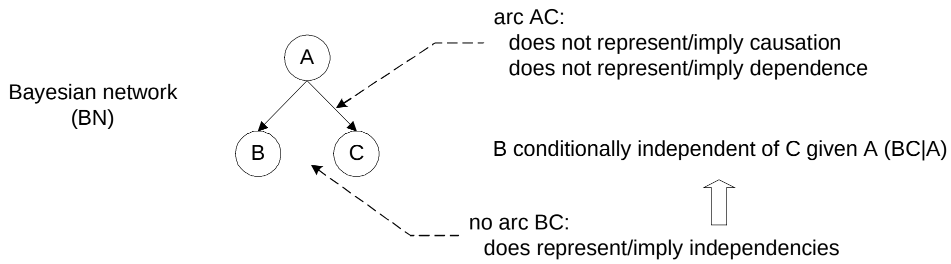


Figure 8

The reader may refer to [Neapolitan 1990] and [Pearl 1988].

1.2. Assumptions

The report makes assumptions regarding the Bayesian network, which limit the scope of the research.

The Bayesian network models a data set; the report considers only the case when a Bayesian network is used to model the joint probability distribution described (represented) by a given data set; the method used to build the Bayesian network is irrelevant.

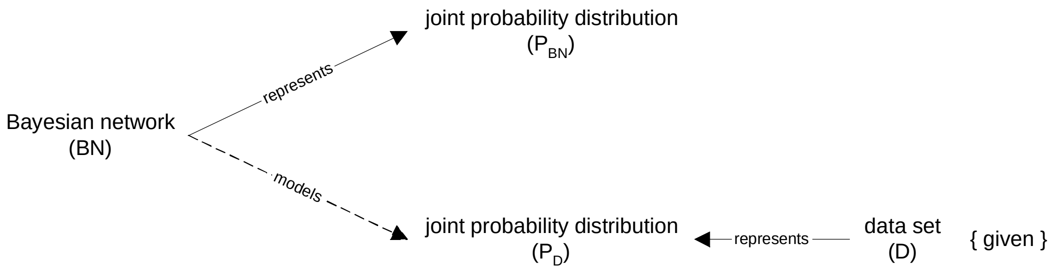


Figure 9

The data set is complete; the report assumes that there are no missing values for any of the variables of the data set.

The variables of the data set are finite; the report assumes that there are finite possible values for each of the variables of the data set.

1.3. Structure of the report

Chapter 2 introduces the subject of the report; the accuracy of a Bayesian network. A definition for the accuracy of a Bayesian network is presented, together with an appropriate measure of accuracy that can, unfortunately, be rarely used in practice for real-world applications.

Chapter 3 presents the Minimum Description Length (MDL) formalism, which examines the accuracy of a Bayesian network, by modelling it as the probability of the Bayesian network given the data set that describes the joint probability distribution the Bayesian network models. A lengthy discussion points out the shortcomings of the formalism, indicating that MDL is flawed, and thus inappropriate for examining the accuracy of a Bayesian network.

Chapter 4 presents a new framework for Bayesian Networks, which models the accuracy of a Bayesian network as the accuracy of the conditional independencies implied by the structure of the Bayesian network, and specifies an appropriate measure of accuracy called the Network Conditional Independencies Mutual Information (NCIMI) measure. The framework is formally established, with several definitions and theorems, and well-defined semantics. A lengthy discussion follows, which outlines the characteristics and properties of the proposed framework for Bayesian Networks, and indicates that the framework does not exhibit the shortcomings of MDL.

Chapter 5 presents the experiments that have been conducted, and provides the experimental results that have been collected. Both the MDL formalism and the proposed framework for Bayesian Networks are evaluated, in view of the experimental results, which confirm what has been previously claimed in theory. Therefore, the experiments confirm that MDL is inappropriate for examining the accuracy of a Bayesian network, while the proposed framework for Bayesian Networks is not only appropriate but also performs remarkably well at determining the accuracy of a Bayesian network.

Chapter 6 concludes the report; it provides a review, it discusses the significance of the innovative research, and it indicates possible directions for future work.

The proofs for the theorems mentioned throughout the report are presented in Appendix A.

2. Accuracy of a Bayesian network

A Bayesian network can be used in order to model the joint probability distribution represented by a given data set.

A principal characteristic of a Bayesian network is the degree to which it models the joint probability distribution represented by the given data accurately; the accuracy of a Bayesian network with respect to the data set.

The degree of accuracy of a Bayesian network, with respect to a data set, depends on how well the Bayesian network models the given data set, and is determined by how well the joint probability distribution represented by the Bayesian network matches the joint probability distribution described by the given data set.

Definition 2.1
A Bayesian network (BN) is accurate with respect to a data set (D), if and only if, the joint probability distribution represented by the Bayesian network (P_{BN}) matches the joint probability distribution described by the data set (P_D).

$$\text{accurate } BN \text{ with respect to D} \Leftrightarrow P_{BN} = P_D$$

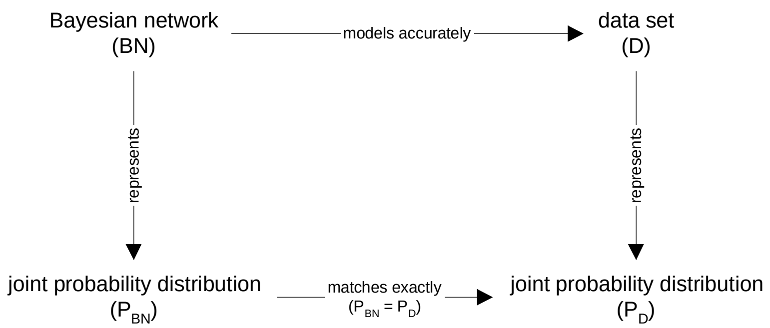


Figure 10

Theorem 2.1
For a data set (D), there exists a Bayesian network (BN) that models the joint probability distribution described by the data set (P_D) accurately.

$$\forall D, \exists BN : P_{BN} = P_D$$

Theorem 2.1 is derived directly from Theorem 1.2. The distinction between these theorems is the fact that Theorem 2.1 refers to a data set that describes a joint probability distribution, which can be represented by a Bayesian network, unlike Theorem 1.2 which refers to a joint probability distribution that can be represented by a Bayesian network

Theorem 2.1 guarantees that for any data set, there exists a Bayesian network that models the joint probability distribution described by the data set accurately.
Therefore, given any data set, an accurate Bayesian network can be constructed.

Both the joint probability distribution represented by the Bayesian network and the joint probability distribution described by the data set can be represented as n-dimensional matrices of $states_1 * ... * states_n$ elements, assuming the data set has n variables, and each variable v_i has $states_i$ states.
Since a matrix of n elements can be represented geometrically as a point in the $\Re^n$ space, then both the joint probability distribution represented by the Bayesian network and the joint probability distribution described by the data set can be represented geometrically as points in the $\Re^{states_1 * ... * states_n}$ space.

In view of the geometrical representation, the degree to which the joint probability distribution represented by the Bayesian network matches the joint probability distribution described by the data set is reflected by how close the corresponding points are.

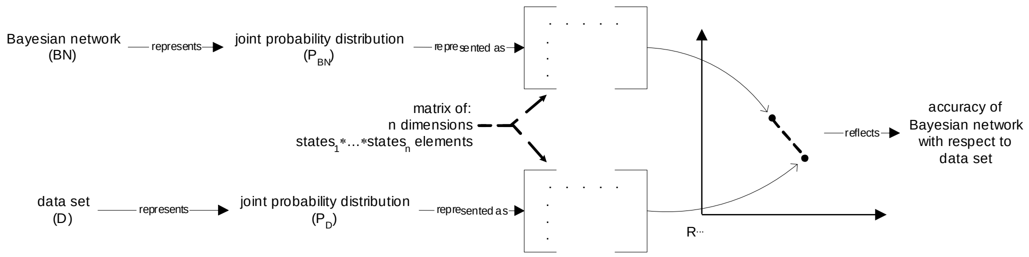


Figure 11

Definition 2.2
A Bayesian network is accurate with respect to a data set, if and only if, the point corresponding to the joint probability distribution represented by the Bayesian network matches the point corresponding to the joint probability distribution described by the data set.

Definition 2.3
The degree of accuracy of a Bayesian network, with respect to a data set, is determined by the geometrical distance between the point corresponding to the joint probability distribution represented by the Bayesian network and the point corresponding to the joint probability distribution described by the data set.

$$accuracy(BN) = accuracy(BN, D) = f_1(distance(P_{BN}, P_D))$$

The degree of accuracy of a Bayesian network is inversely related to the distance between the point corresponding to the joint probability distribution represented by the Bayesian network and the point corresponding to the joint probability distribution described by the data set; thus, it is more convenient to examine the inaccuracy of a Bayesian network, which is positively related to the distance.

Definition 2.4a
The degree of inaccuracy of a Bayesian network, with respect to a data set, is the geometrical distance between the point corresponding to the joint probability distribution represented by the Bayesian network and the point corresponding to the joint probability distribution described by the data set.

$$inaccuracy(BN) = inaccuracy(BN, D) = f_2(distance(P_{BN}, P_D)) = distance(P_{BN}, P_D)$$

It is assumed that $f_2(x) = x$ *for simplicity; another function could be used, such as* $f_2(x) = \frac{1}{\sqrt{2}} x$.

In the $\Re^n$ space, according to the Pythagorean Theorem principles, the distance between two points is given by the following formula.

$$distance(point_1, point_2) = \sqrt{\sum_{i=1}^{n} (p_i - q_i)^2}, \; point_1 = (p_1, ..., p_n), \; point_2 = (q_1, ..., q_n)$$

The Euclidean distance is used as a distance measure to derive one of the possible definitions for the inaccuracy of a Bayesian network; alternative measures of distance between joint probability distributions could be used, such as the Kullback-Liebler distance.

Definition 2.4b

The degree of inaccuracy of a Bayesian network, with respect to a data set, is the geometrical distance between the point corresponding to the joint probability distribution represented by the Bayesian network and the point corresponding to the joint probability distribution described by the data set.

$$inaccuracy(BN) = \sqrt{\sum_{v_1=1}^{states_1} \cdots \sum_{v_n=1}^{states_n} (P_{BN}(v_1, ..., v_n) - P_D(v_1, ..., v_n))^2}$$

The joint probability distribution described by the data set is determined directly from the data set, while the joint probability distribution represented by the Bayesian network is determined from the data set using Theorem 1.1.

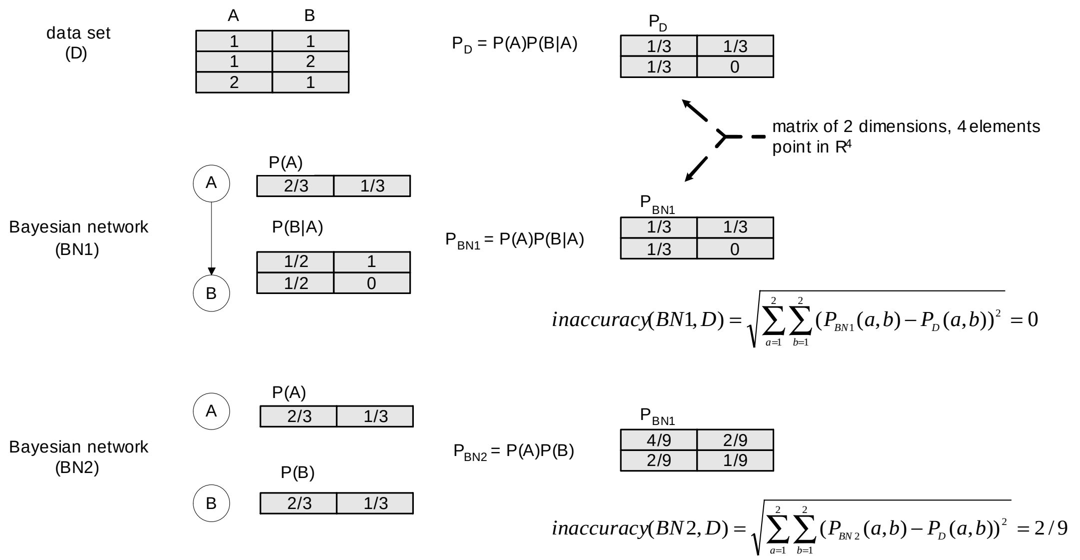


Figure 12

Theorem 2.2

The range of values for the actual degree of inaccuracy of a Bayesian network is $[0, \sqrt{2}]$.

$$0 \leq inaccuracy(BN) \leq \sqrt{2}$$

Unfortunately, although the accuracy of a Bayesian network is well defined in theory, and an appropriate measure of inaccuracy is specified, it is rarely possible to determine the degree of inaccuracy of a Bayesian network in practice for real-world applications.
This is due to the fact that in most cases it is computationally unfeasible to determine and use the joint probability distribution described by the data set, because of both processing and storage limitations; so it is not possible to determine the degree of inaccuracy of a Bayesian network, using Definition 2.4.
For example, for the Hepatitis C data set, which is used subsequently for experimentation, the matrix of the joint probability distribution described by the data set is 9 dimensional containing 14,696,640 elements. So, it is not only computationally unfeasible to determine the matrix, but also virtually impossible to use the matrix in practice.
This is principally the reason why a Bayesian network is used to model the joint probability distribution described by the data set, instead of the joint probability distribution itself.

3. Minimum Description Length (MDL)

Since it is rarely possible to determine the actual degree of accuracy of a Bayesian network in practice for real-world applications, alternative characteristics of a Bayesian network, which relate to and reflect the accuracy, are used to model and examine the accuracy of a Bayesian network.

A popular formalism that adopts such methods to examine the accuracy of a Bayesian network is the Minimum Description Length (MDL) formalism [Rissanen 1978] (see also [Friedman et al 1997]), which evaluates the accuracy indirectly, by examining alternative characteristics of a Bayesian network.

3.1. Formalism

The MDL formalism provides an evaluation scheme for a model that represents a data set, based on the length of the description of the data set; the sum of the length of the description of the model and the length of the description of the data set given the model.
The length of the description of the model reflects the model size and complexity, while the length of the description of the data set given the model is interpreted as the model accuracy.

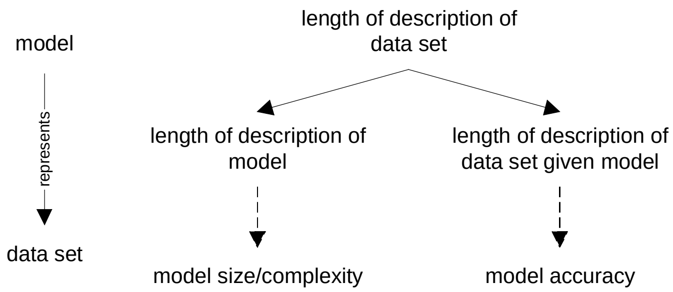


Figure 13

In the context of the Communication domain, the MDL formalism is employed in data compression, in order to identify the model that provides the shortest description of the data set.
In this case, the length of the description of the data set is the number of bits required to encode the data set; the sum of the number of bits required to describe the model and the number of bits required to encode the data set given the model.

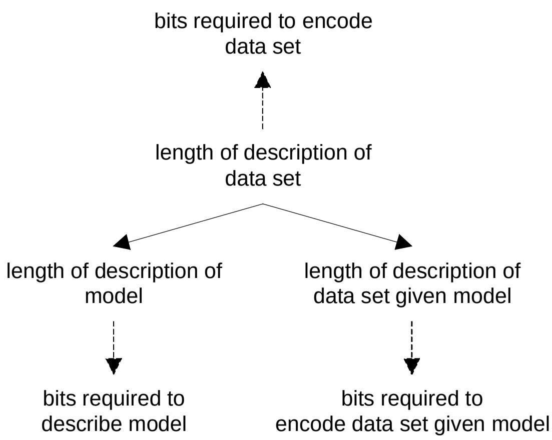


Figure 14

In the context of Bayesian Networks, the model that represents a data set is a Bayesian network that models the joint probability distribution described by a data set.
In this case, the length of the description of the data set given the Bayesian network is the negation of the log likelihood of the Bayesian network given the data set, which is interpreted as the degree of inaccuracy of the Bayesian network.

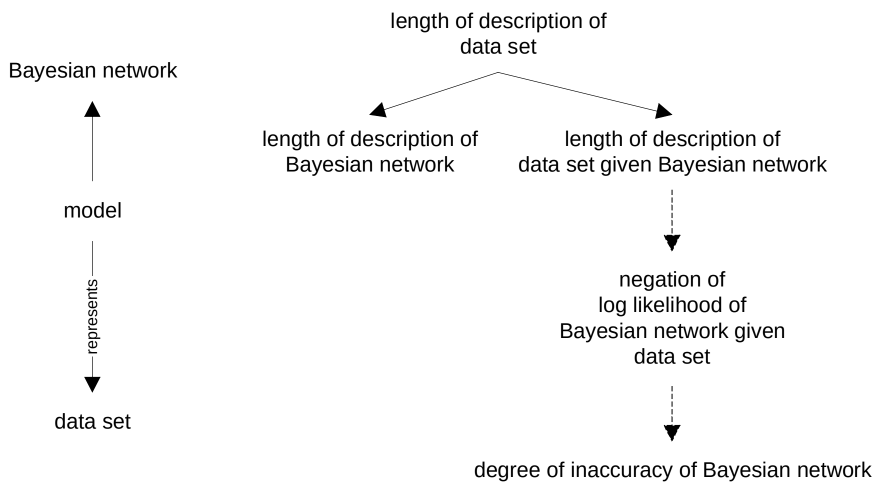


Figure 15

The MDL formalism models the accuracy of a Bayesian network with respect to a data set as the likelihood of the Bayesian network given the data set.
The complexity term of the MDL measure is missing, since the complexity of the structure of the Bayesian network is not taken into account when evaluating the accuracy of the network

Definition 3.1
The degree of accuracy of a Bayesian network, with respect to a data set, is the log likelihood of the Bayesian network given the data set.

$$accuracy(BN) = accuracy(BN, D) = \log_2(P(BN \mid D))$$

Definition 3.2a

The degree of inaccuracy of a Bayesian network, with respect to a data set, is the negation of the log likelihood of the Bayesian network given the data set.

$$inaccuracy(BN) = inaccuracy(BN, D) = -\log_2(P(BN \mid D))$$

It is apparent that the degree of inaccuracy of a Bayesian network is the negation of the degree of accuracy of the Bayesian network.

Since the MDL formalism is searching for the most accurate Bayesian network given a data set, and assuming a uniform distribution over the possible Bayesian networks, the likelihood of a Bayesian network given the data set is equivalent to the likelihood of the data set given the Bayesian network.

$$\max_{BN \in B}\{\text{accuracy(BN, D)}\} \overset{\text{MDL}}{=} \max_{BN \in B}\{P(BN \mid D)\} =$$

$$\max_{BN \in B}\{\frac{P(D \mid BN)P(BN)}{P(D)}\} = \max_{BN \in B}\{P(D \mid BN)P(BN)\} = \max_{BN \in B}\{P(D \mid BN)\}$$

$$\text{since } \forall BN \in B, P(D) \text{ is constant, and } P(BN) \text{ is uniform}$$

The MDL formalism models the accuracy of a Bayesian network with respect to a data set as the likelihood of the data set given the Bayesian network.

Definition 3.3

The likelihood of a data set given a Bayesian network is given by:

$$P(D \mid BN) = \prod_{d \in D} P_{BN}(d)$$

Definition 3.2b

The degree of inaccuracy of a Bayesian network, with respect to a data set, is the negation of the log likelihood of the data set given the Bayesian network.

$$inaccuracy(BN) = -\log_2(P(D \mid BN)) = -\log_2(\prod_{d \in D} P_{BN}(d)) = -\sum_{d \in D} \log_2(P_{BN}(d))$$

The joint probability distribution represented by the Bayesian network is determined from the data set, using Theorem 1.1.

data set (D)

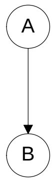
	A	B
d1	1	1
d2	1	2
d3	2	1
d4	2	2

$P_D = P(A)P(B|A)$

P_D

1/4	1/4
1/4	1/4

Bayesian network (BN)



P(A)

1/2	1/2

P(B|A)

1/2	1/2
1/2	1/2

$P_{BN} = P(A)P(B|A)$

P_{BN}

1/4	1/4
1/4	1/4

$$P(D \mid BN) = \prod_{d \in D} P_{BN}(d) = P_{BN}(d1) * P_{BN}(d2) * P_{BN}(d3) * P_{BN}(d4) =$$
$$(P_A(d1)P_{B|A}(d1)) * (P_A(d2)P_{B|A}(d2)) * (P_A(d3)P_{B|A}(d3)) * (P_A(d4)P_{B|A}(d4)) =$$
$$(1/2*1/2)*(1/2*1/2)*(1/2*1/2)*(1/2*1/2) = (1/2)^8$$
$$inaccuracy(BN) = -\log_2(P(D \mid BN)) = -\log_2((1/2)^8) = 8$$

Figure 16

Theorem 3.1a
The range of values for the MDL measure of the degree of accuracy of a Bayesian network, with respect to any data set, is $(-\infty, 0]$.
$$-\infty < accuracy(BN) \le 0$$

Theorem 3.1b
The range of values for the MDL measure of the degree of inaccuracy of a Bayesian network, with respect to any data set, is $[0, +\infty)$.
$$0 \le inaccuracy(BN) < +\infty$$

The range of values for the MDL measure of the degree of accuracy of a Bayesian network, with respect to a data set, depends on the size of the data set.

3.2. Discussion

3.2.1. Characteristics

The MDL formalism evaluates the likelihood of a Bayesian network given a particular data set, and so the accuracy of a Bayesian network with respect to a given data set depends on the nature of the data set.

The degree of accuracy of a Bayesian network, with respect to a data set, is affected directly by the size of the data set [Friedman & Yakhini 1996]. Apparently, the range of values for the degree of accuracy of a Bayesian network, with respect to a data set, depends on the size of the data set.

Therefore, it is not possible to determine whether a Bayesian network is accurate given a non-zero value for the degree of accuracy of the Bayesian network, or to determine the degree of accuracy of a Bayesian network given the Bayesian network is accurate, unless the size of the data set is considered.

Consider the data set D and the data set D' that consists of all the data entries of the data set D twice, so that $D' = 2D$. Given any Bayesian network, the degree of inaccuracy of the Bayesian network with respect to D' is twice the degree of inaccuracy of the Bayesian network with respect to D, according to Definition 3.2. However, both the data set D and the data set D' describe the same joint probability distribution, and so the degree of inaccuracy of the Bayesian network should be the same in both cases.

Evidently, the MDL formalism does not determine the "absolute" degree of accuracy of a Bayesian network, but instead, MDL determines the "relative" degree of accuracy of a Bayesian network for a particular data set.

Consequently, the MDL formalism cannot draw any conclusions about the "absolute" accuracy of a Bayesian network, but instead, MDL can only be used to compare "relative" degrees of accuracy for a set of Bayesian networks and for a particular data set.
Any results acquired and any conclusions drawn are valid for the specific set of Bayesian networks, and only in view of the particular data set.

3.2.2. Semantics

Although the MDL formalism is being used in the field of Bayesian Networks, it was initially developed for the Communication domain.

In the Communication domain, the focus is the transmission of a message; the reconstruction of a data set.
In the field of Bayesian Networks, the focus is the construction of a Bayesian network that models the joint probability distribution in a given data set; the reconstruction of the joint probability distribution described by a data set.

Evidently, the semantics of these two fields are different, and thus the MDL formalism, which was developed for the Communication domain, is being taken out of context when used in the field of Bayesian Networks.

The MDL formalism examines the accuracy of a Bayesian network with respect to the data entries of the data set, and not with respect to the joint probability distribution represented by the data set.

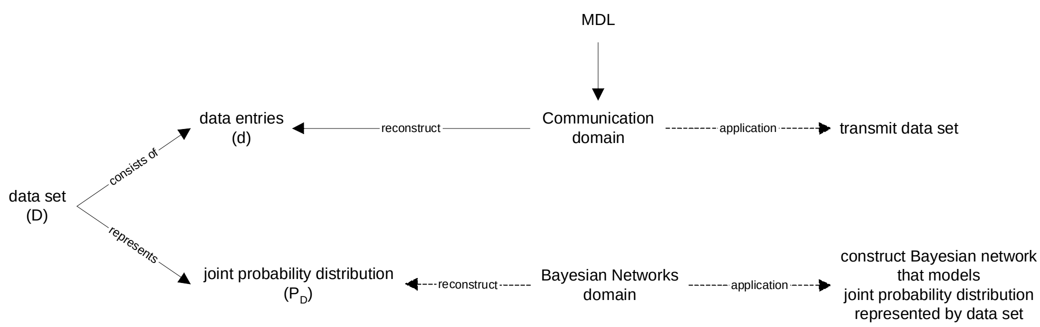


Figure 17

The above points are illustrated by the previous example, where two data sets are given, such that the second data set consists of all the data entries of the first data set twice.
In the context of the Communication domain, the length of the description of the second data set given a model is indeed twice the length of the description of the first data set given the same model.
However, in the context of Bayesian Networks, the degree of accuracy of a Bayesian network with respect to the first data set should be the same as the degree of accuracy of the Bayesian network with respect to the second data set, since both data sets describe the same joint probability distribution.

Consequently, it is evident that although the MDL formalism is appropriate for the Communication domain, its semantics are not completely appropriate for the field of Bayesian Networks, and thus MDL is not entirely appropriate for examining the accuracy of a Bayesian network.

3.3. Other approaches

The MDL formalism is one of the formalisms that provide an evaluation scheme for the characteristics, and in particular the accuracy, of a Bayesian network.

Other formalisms that have been developed to evaluate the characteristics of a Bayesian network are the Akaike Information Criterion (AIC) [Akaike 1974] and the Bayesian Information Criteria (BIC) [Schwarz 1978]. However, both of these formalisms are identical to the MDL formalism with regards to the evaluation of the accuracy of a Bayesian network, since they both model the accuracy as the log likelihood of the Bayesian network given the data set.

Even Bayesian network learning algorithms, such as the Maximum Weight Spanning Tree (MWST) algorithm [Chow & Liu 1968], provide some sort of an evaluation scheme for a Bayesian network. However, each algorithm is based on its own heuristic methods, which do not specify clearly what the characteristics of a Bayesian network are, and how these are evaluated.

4. Framework for Bayesian Networks

A framework for Bayesian Networks is presented, which is used to examine and evaluate the accuracy of a Bayesian network indirectly, by examining the conditional independencies implied by the structure of the Bayesian network.

4.1. Framework

The structure of a Bayesian network implies a set of conditional independencies regarding the joint probability distribution it models, which provides information about the accuracy of the Bayesian network.

Theorem 4.1 (NCI Theorem)
Given:

- A data set (D) representing a joint probability distribution (P) over variables (V)
  $P_D = P(V)$
- A Bayesian network (BN) of nodes (N), arcs (A), distribution (P')
  $BN = (N, A, P')$
  $P_{BN} = P'(N)$

Provided:

- The nodes of the Bayesian network are the variables of the joint probability distribution represented by the data set
  $N \equiv V$
  $\forall i, n_i \in N, v_i \in V : n_i \equiv v_i$
- The probabilities of the Bayesian network reflect the corresponding probabilities of the joint probability distribution represented by the data set
  $\forall i, n_i \in N : P'(n_i \mid parents(n_i)) = P(n_i \mid parents(n_i))$

➢ The Bayesian network is accurate with respect to the data set, **if and only if**, the conditional independencies implied by the structure of the Bayesian network are conditional independencies of the joint probability distribution represented by the data set
  $P_{BN} = P_D \Leftrightarrow$
  $\forall i, v_i \in V, \forall W, W \subseteq a(v_i) : P(v_i \mid W \cup parents(v_i)) = P(v_i \mid parents(v_i))$

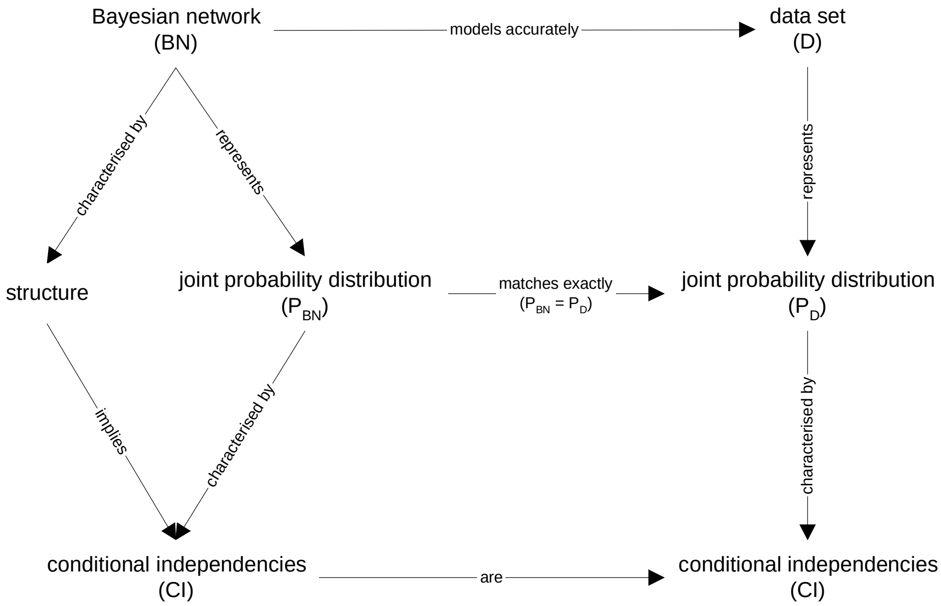


Figure 18

Since the report considers only the case when a Bayesian network is used to model the joint probability distribution represented by a data set, in order for the assumptions of the NCI Theorem to be satisfied, the nodes of the Bayesian network should model one-for-one the variables of the joint probability distribution represented by the data set, and the probabilities of the Bayesian network should be acquired directly from the data set.

Definition 4.1
A conditional independence implied by the structure of a Bayesian network is accurate with respect to a data set, if and only if, the conditional independence implied by the structure of the Bayesian network is a conditional independence of the joint probability distribution represented by the data set.

Theorem 4.2
A Bayesian network is accurate, if and only if, the conditional independencies implied by the structure of the Bayesian network are accurate.

The degree of accuracy of a Bayesian network is reflected by the degree of accuracy of the conditional independencies implied by the structure of the Bayesian network.

4.1.1. NCI & DCI

An individual conditional independence is denoted as CI .

Definition 4.2
The set of conditional independencies implied by the structure of a Bayesian network is the Network Conditional Independencies (NCI) set.

The NCI set contains only those conditional independencies implied by the structure of the network that are derived directly from the definition of a

Bayesian network (Definition 1.1). Additional conditional independencies implied by the structure of the network that are derived using d-separation are not included in the NCI set.

Definition 4.3
The set of conditional independencies of the joint probability distribution represented by a data set is the Distribution Conditional Independencies (DCI) set.

For a conditional independence implied by the structure of a Bayesian network to be accurate it is required that it is a conditional independence of the joint probability distribution represented by the data set (Definition 4.1), which can be formulated as $CI \in DCI$.

Definition 4.4
A conditional independence implied by the structure of a Bayesian network is accurate, if and only if, $CI \in DCI$.

For a Bayesian network to be accurate it is required that every conditional independence implied by the structure of the Bayesian network is a conditional independence of the joint probability distribution represented by the data set (NCI Theorem), which can be formulated as $NCI \subseteq DCI$.

Theorem 4.3
A Bayesian network is accurate, if and only if, $NCI \subseteq DCI$.

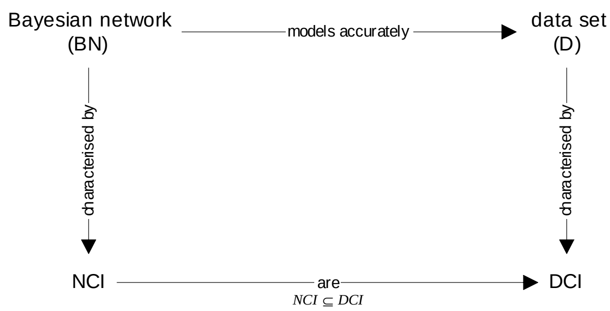


Figure 19

The condition $NCI \subseteq DCI$ reflects the soundness (accuracy) of the NCI; that is to say whether every conditional independence implied by the structure of the Bayesian network is a conditional independence of the joint probability distribution represented by the data set, and thus reflects the accuracy of the Bayesian network.

Definition 4.5
The NCI is sound (accurate), if and only if, $NCI \subseteq DCI$.

Theorem 4.4
A Bayesian network is accurate, if and only if, the NCI is sound (accurate).

The degree to which $NCI - DCI = \varnothing$ reflects the degree of soundness (accuracy) of the NCI, which, in turn, reflects the degree of accuracy of the Bayesian network.

Theorem 4.5 (NCI Soundness Theorem)
For a data set, there exists a Bayesian network so that the NCI is sound.
$\forall D, \exists BN : NCI \subseteq DCI$

The NCI Soundness Theorem indicates that for any data set, there exists a Bayesian network, so that every conditional independence implied by the structure of the Bayesian network is a conditional independence of the joint probability distribution represented by the data set.
Consequently, given any data set, it is possible to construct a Bayesian network that satisfies the condition $NCI \subseteq DCI$, and is thus accurate.

The condition $DCI \subseteq NCI$ reflects the completeness of the NCI; that is to say whether every conditional independence of the joint probability distribution represented by the data set is implied by the structure of the Bayesian network.

Definition 4.6
The NCI is complete, if and only if, $DCI \subseteq NCI$.

The degree to which $DCI - NCI = \varnothing$ reflects the degree of completeness of the NCI.

Theorem 4.6 (NCI Incompleteness Theorem)
There exists a data set, so that for any Bayesian network, the NCI is incomplete.
$\exists D, \forall BN : DCI \not\subseteq NCI$
$\exists D, \neg\exists BN : DCI \subseteq NCI$

The NCI Incompleteness Theorem indicates that there exists at least one data set, for which there exists no Bayesian network, such that every conditional independence of the joint probability distribution represented by the data set is implied by the structure of the Bayesian network.
Consequently, given certain data sets, it is not possible to construct a Bayesian network that satisfies the condition $DCI \subseteq NCI$, since the structure of the Bayesian network fails to imply all the conditional independencies of the joint probability distribution represented by the data set. Thus, given an arbitrary data set, it may not be possible to construct a Bayesian network whose NCI is complete.

The condition $NCI = DCI$ reflects the soundness (accuracy) and completeness of the NCI; whether every conditional independence implied by the structure of the Bayesian network is a conditional independence of the joint probability distribution represented by the data set, and whether every conditional independence of the joint probability distribution represented by the data set is implied by the structure of the Bayesian network.

Definition 4.7
The NCI is sound (accurate) and complete, if and only if, $NCI = DCI$.

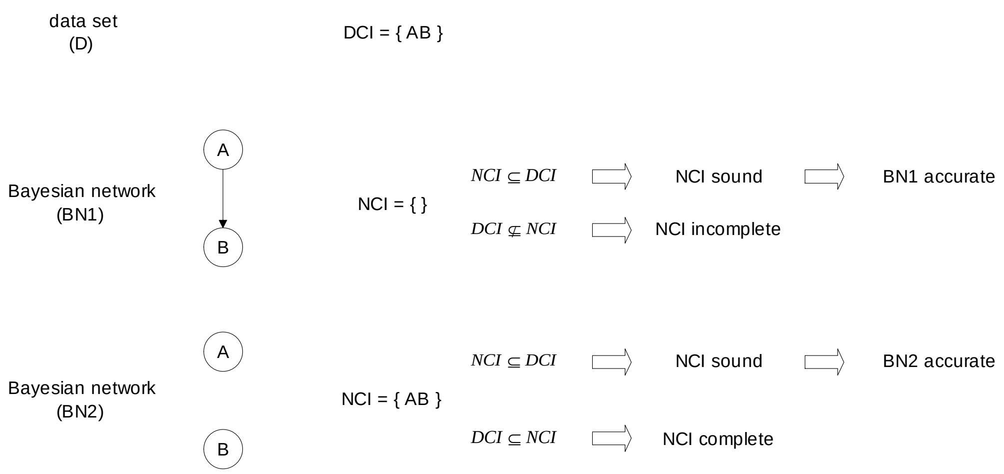


Figure 20

4.1.2. R/$\overline{R}$ & C/$\overline{C}$

Definition 4.8

The set of conditional independencies implied by the structure of the Bayesian network that are not conditional independencies of the joint probability distribution represented by the data set is the "unrealistic" set $\overline{R}$.

$\overline{R} = NCI - DCI$

Definition 4.9

The set of conditional independencies implied by the structure of the Bayesian network that are conditional independencies of the joint probability distribution represented by the data set is the "realistic" set R.

$R = NCI - \overline{R}$

The "unrealistic" set $\overline{R}$ represents the conditional independencies implied by the structure of the Bayesian network that are not accurate, while the "realistic" set R represents the conditional independencies implied by the structure of the Bayesian network that are accurate.

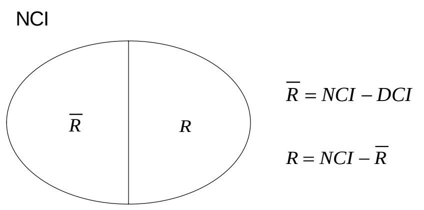


Figure 21

Definition 4.10

The set of conditional independencies of the joint probability distribution represented by the data set that are not implied by the structure of the Bayesian network is the "uncaptured" set $\overline{C}$.

$\overline{C} = DCI - NCI$

Definition 4.11

The set of conditional independencies of the joint probability distribution represented by the data set that are implied by the structure of the Bayesian network is the "captured" set C.

$C = DCI - \overline{C}$

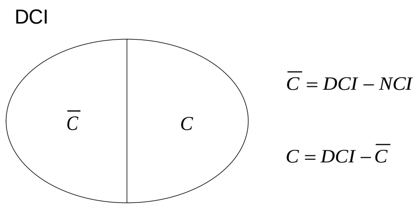


Figure 22

Theorem 4.7

The R set and the C set are identical.

$R = C$

Theorem 4.8

The R set is a subset of the DCI set.

$R \subseteq DCI$

Theorem 4.9

The C set is a subset of the NCI set.

$C \subseteq NCI$

Theorem 4.10

The $\overline{R}$ set and the $\overline{C}$ set are disjoint.

$\overline{R} \cap \overline{C} = \varnothing$

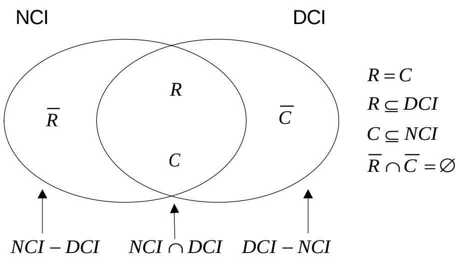


Figure 23

The $\overline{R}$ set reflects the condition $NCI \subseteq DCI$, and thus the soundness (accuracy) of the NCI, and, in turn, the accuracy of the Bayesian network

Definition 4.12

The NCI is sound, if and only if, $\overline{R} = \varnothing$.

Theorem 4.11

A Bayesian network is accurate, if and only if, $\overline{R} = \varnothing$.

The degree to which $\overline{R} = \varnothing$ reflects the degree to which $NCI - DCI = \varnothing$, and thus the degree of soundness (accuracy) of the NCI, and, in turn, the degree of accuracy of the Bayesian network

The $\overline{C}$ set reflects the condition $DCI \subseteq NCI$, and thus the completeness of the NCI.

Definition 4.13
The NCI is complete, if and only if, $\overline{C} = \varnothing$.

The degree to which $\overline{C} = \varnothing$ reflects the degree to which $DCI - NCI = \varnothing$, and thus the degree of completeness of the NCI.

Definition 4.14
The NCI is sound (accurate) and complete, if and only if, $\overline{R} = \varnothing \wedge \overline{C} = \varnothing$.

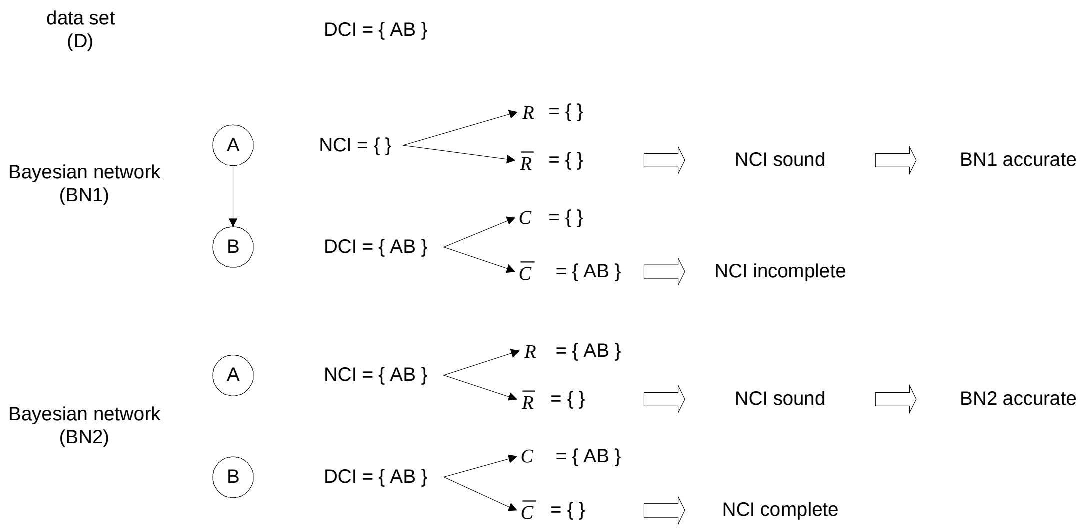


Figure 24

4.1.3. Inaccuracy of a conditional independence

Consider a dependency measure for variables that determines the degree of dependence of variables.
The reader may refer to [Johnston 2000].

The dependency measure can be used to determine either the degree of conditional dependence of variables given another variable or the degree of unconditional dependence of variables.

The report employs the Mutual Information as a dependency measure for variables.

Definition 4.15a
The unconditional dependency measure (DM_u) for variables A and B is the unconditional Mutual Information (MI) for variables A and B.

$$DM_u(A,B) = MI(A,B) = \sum_A \sum_B P(A,B) \log_2 \frac{P(A,B)}{P(A)P(B)}$$

Definition 4.15b

The conditional dependency measure (DM_c) for variables A and B given variable C is the conditional Mutual Information (MI) for variables A and B given variable C.

$$DM_c(A,B \mid C) = MI(A,B \mid C) = \sum_A \sum_B \sum_C P(A,B,C) \log_2 \frac{P(A,B \mid C)}{P(A \mid C)P(B \mid C)}$$

Theorem 4.12

The unconditional dependency measure (DM_u) for variables A and B is the conditional dependency measure (DM_c) for variables A and B given the empty set $\varnothing$.

$$DM_u(A,B) = DM_c(A,B \mid \varnothing)$$

A value of zero for the dependency measure indicates that the variables are independent, while a non-zero value indicates that the variables are dependent in some degree.
In particular, the higher the value for the dependency measure, the more dependent the variables are.

The dependency measure can identify either a conditional independence of variables given another variable or an unconditional independence of variables.

Theorem 4.13a

The unconditional dependency measure (DM_u) for variables A and B has a value of zero, if and only if, the variables A and B are unconditionally independent.

$$DM_u(A,B) = 0 \Leftrightarrow A,B \text{ unconditionally independent}$$

Theorem 4.13b

The conditional dependency measure (DM_c) for variables A and B given variable C has a value of zero, if and only if, the variables A and B are conditionally independent given C.

$$DM_c(A,B \mid C) = 0 \Leftrightarrow A,B \text{ conditionally independent given } C$$

The dependency measure for variables can be used as an inaccuracy measure for conditional independencies; the degree of inaccuracy of a conditional independence is reflected by the degree of dependence of the variables described by the conditional independence.
If the variables described by the conditional independence are dependent, then the conditional independence is inaccurate, whereas if the variables described by the conditional independence are independent, then the conditional independence is accurate.

Definition 4.16

The inaccuracy measure for a conditional independence is the dependency measure for the variables described by the conditional independence.

$$inaccuracy(CI) = DM(CI)$$

The inaccuracy measure determines whether a conditional independence implied by the structure of a Bayesian network is indeed a conditional independence of the joint probability distribution represented by the data set.

A value of zero for the inaccuracy measure indicates that the implied conditional independence is accurate and thus a conditional independence of the joint probability distribution represented by the data set, while a non-zero value indicates that the implied conditional independence is inaccurate in some degree, and thus not a conditional independence of the joint probability distribution represented by the data set.
In particular, the higher the value for the inaccuracy measure, the more inaccurate the conditional independence implied by the structure of the Bayesian network is.

Theorem 4.14a
The degree of inaccuracy of a conditional independence implied by the structure of a Bayesian network has a value of zero, if and only if, the conditional independence implied by the structure of the Bayesian network is accurate.

$$inaccuracy(CI) = 0 \Leftrightarrow \text{accurate}\, CI$$

Theorem 4.14b
The degree of inaccuracy of a conditional independence implied by the structure of a Bayesian network, with respect to a data set, has a value of zero, if and only if, the conditional independence implied by the structure of the Bayesian network is a conditional independence of the joint probability distribution represented by the data set.

$$inaccuracy(CI) = 0 \Leftrightarrow CI \in DCI$$

For a set of conditional independencies, the inaccuracy is reflected by the inaccuracy of the individual conditional independencies of the set of conditional independencies.

Definition 4.17
The degree of inaccuracy of a set of conditional independencies is the sum of the degrees of inaccuracy of the individual conditional independencies of the set of conditional independencies.

$$inaccuracy(CI) = \begin{cases} \sum_i inaccuracy(CI_i) & CI \neq \emptyset, CI_i \in CI \\ 0 & CI = \emptyset \end{cases}$$

Apparently, the more inaccurate conditional independencies the set of conditional independencies contains, the higher the degree of inaccuracy for the set of conditional independencies.

Theorem 4.15

The degree of inaccuracy of a union of sets of conditional independencies is the sum of the degrees of inaccuracy of each of the sets of conditional independencies.

$$inaccuracy(\bigcup_{i=1}^{n} CI_i) = \sum_{i=1}^{n} inaccuracy(CI_i)$$

4.1.4. Inaccuracy of the NCI

The "realistic" set R, the "captured" set C, and the "uncaptured" set $\overline{C}$ represent conditional independencies that are conditional independencies of the joint probability distribution represented by the data set, and so the sets R, C, and $\overline{C}$ represent accurate conditional independencies.

Theorem 4.16
The degree of inaccuracy for each of the sets of conditional independencies R, C, and $\overline{C}$ has a value of zero.

$$inaccuracy(R) = 0$$

$$inaccuracy(C) = 0$$

$$inaccuracy(\overline{C}) = 0$$

The DCI set represents the conditional independencies of the joint probability distribution represented by the data set, and so the DCI set represents accurate conditional independencies.

Theorem 4.17
The degree of inaccuracy of the DCI has a value of zero.

$$inaccuracy(DCI) = 0$$

The "unrealistic" set $\overline{R}$ represents conditional independencies that are not conditional independencies of the joint probability distribution represented by the data set, and so the set $\overline{R}$ represents inaccurate conditional independencies.

Theorem 4.18
The degree of inaccuracy for the set $\overline{R}$ has a non-zero value, when the $\overline{R}$ set is not empty, while the degree of inaccuracy for the set $\overline{R}$ has a zero value, when the $\overline{R}$ set is empty.

$$inaccuracy(\overline{R}) \neq 0 \Leftrightarrow \overline{R} \neq \emptyset$$

$$inaccuracy(\overline{R}) = 0 \Leftrightarrow \overline{R} = \emptyset$$

The NCI set represents conditional independencies that are not all necessarily conditional independencies of the joint probability distribution represented by the data set, and the degree of inaccuracy of the NCI is determined by the degree of inaccuracy of the set $\overline{R}$.

Theorem 4.19
The degree of inaccuracy of the NCI is the degree of inaccuracy of the set $\overline{R}$.

$$inaccuracy(NCI) = inaccuracy(\overline{R})$$

Theorem 4.20
The degree of inaccuracy of the NCI has a non-zero value, when the $\overline{R}$ set is not empty, while the degree of inaccuracy of the NCI has a zero value, when the $\overline{R}$ set is empty.

$$inaccuracy(NCI) \neq 0 \Leftrightarrow \overline{R} \neq \varnothing$$

$$inaccuracy(NCI) = 0 \Leftrightarrow \overline{R} = \varnothing$$

Definition 4.18
The degree of inaccuracy of the NCI is the sum of the degrees of inaccuracy of the individual conditional independencies implied by the structure of the Bayesian network.

$$inaccuracy(NCI) = \begin{cases} \sum_i inaccuracy(NCI_i) & NCI \neq \varnothing, NCI_i \in NCI \\ 0 & NCI = \varnothing \end{cases}$$

4.1.5. Inaccuracy of a Bayesian network

The accuracy of a Bayesian network is reflected by the condition $\overline{R} = \varnothing$ (Theorem 4.11), and so the degree of inaccuracy of the set $\overline{R}$ reflects the accuracy of the Bayesian network.

Theorem 4.21
A Bayesian network is accurate, if and only if, $inaccuracy(\overline{R}) = 0$.

Since the degree of inaccuracy of the NCI is the degree of inaccuracy of the set $\overline{R}$, the degree of inaccuracy of the NCI reflects the accuracy of the Bayesian network.

Theorem 4.22
A Bayesian network is accurate, if and only if, $inaccuracy(NCI) = 0$.

The degree of accuracy of a Bayesian network is determined by the degree of accuracy of the NCI.

Definition 4.19
The degree of inaccuracy of a Bayesian network is the degree of inaccuracy of the NCI.

$$inaccuracy(BN) = f(inaccuracy(NCI)) = inaccuracy(NCI)$$

It is assumed that $f(x) = x$ *for simplicity.*

The accuracy of a Bayesian network is modelled as the soundness (accuracy) of the NCI; the accuracy of the set of conditional independencies implied by the structure of the Bayesian network.

Definition 4.20
The degree of inaccuracy of a Bayesian network is determined by the following formulas:

$$inaccuracy(BN) = inaccuracy(NCI)$$

$$inaccuracy(NCI) = \begin{cases} \sum_i inaccuracy(NCI_i) & NCI \neq \varnothing, NCI_i \in NCI \\ 0 & NCI = \varnothing \end{cases}$$

$$inaccuracy(NCI_i) = DM(NCI_i)$$

$$DM(NCI_i) = MI(NCI_i)$$

data set (D)

A	B
1	1
1	2
2	1

P(A,B)

1/3	1/3
1/3	0

P(A)

2/3
1/3

P(B)

2/3
1/3

$$DM(A,B) = MI(A,B) = \sum_{i=1}^{2}\sum_{j=1}^{2} P_{AB}(i,j)\log_2 \frac{P_{AB}(i,j)}{P_A(i)P_B(j)} = \log_2 3 - 4/3$$

DCI = { }

Bayesian network (BN)

NCI = { AB }

$$inaccuracy(BN) = inaccuracy(NCI) = inaccuracy(AB) = DM(AB) = \log_2 3 - 4/3$$

Figure 25

4.1.6. Semantics

A Bayesian network is characterised by a set of nodes and the corresponding prior probabilities, together with a set of arcs connecting the nodes and the corresponding conditional probabilities regarding the nodes each arc connects.
A joint probability distribution is characterised by a set of variables and the corresponding prior probabilities, together with a set of relationships between the variables and the corresponding conditional probabilities.

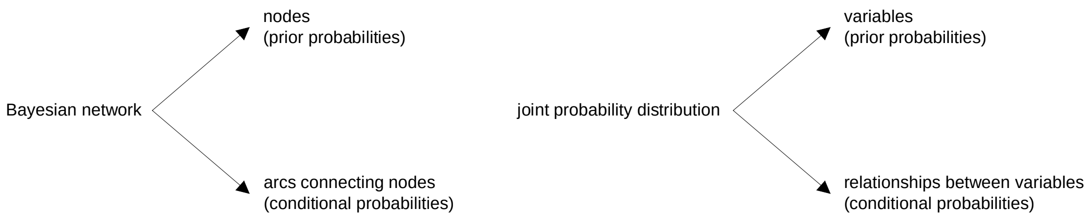


Figure 26

The nodes of the Bayesian network model the variables of the joint probability distribution, while the prior probabilities of the nodes of the Bayesian network model the prior probabilities of the variables of the joint probability distribution.
For every pair of nodes of the Bayesian network, there is either an arc connecting the nodes, or there is no such arc. When a pair of nodes is connected with an arc, the arc models the relationship between the variables of the joint probability distribution. In the case when the pair of nodes is not connected with an arc, the Bayesian network implies conditional independencies that model the relationship between the variables of the joint probability distribution.

The arcs of the Bayesian network and the conditional independencies implied by the structure of the Bayesian network due to the absence of arcs model the relationships between the variables of the joint probability distribution, while the conditional probabilities of the arcs of the Bayesian network model the conditional probabilities of the variables of the joint probability distribution.

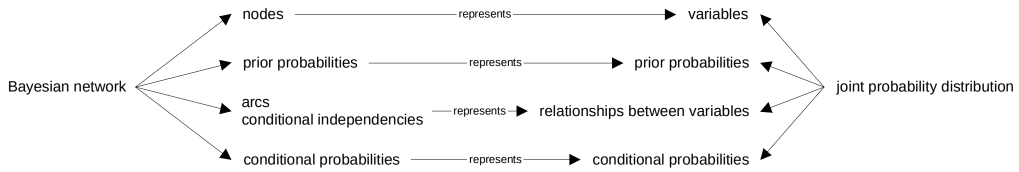


Figure 27

The Bayesian network is an accurate model of the joint probability distribution, if and only if, the Bayesian network is an accurate model of the variables of the joint probability distribution, and the Bayesian network is an accurate model of the relationships between the variables of the joint probability distribution.

The Bayesian network is an accurate model of the variables of the joint probability distribution, if:
1. the nodes of the Bayesian network model the variables of the joint probability distribution one-for-one,
2. the prior probabilities of the nodes of the Bayesian network model correctly (accurately) the prior probabilities of the variables of the joint probability distribution.

The Bayesian network is an accurate model of the relationships between the variables of the joint probability distribution, if:
1. the structure of the Bayesian network (the arcs and the conditional independencies implied due to the absence of arcs) models correctly (accurately) the relationships between the variables of the joint probability distribution,
2. the conditional probabilities of the arcs of the Bayesian network model correctly (accurately) the conditional probabilities of the variables of the joint probability distribution.

The structure of the Bayesian network models correctly (accurately) the relationships between the variables of the joint probability distribution, if:
1. the conditional independencies implied by the structure of the Bayesian network, due to the absence of arcs, are conditional independencies of the joint probability distribution represented by the data set.

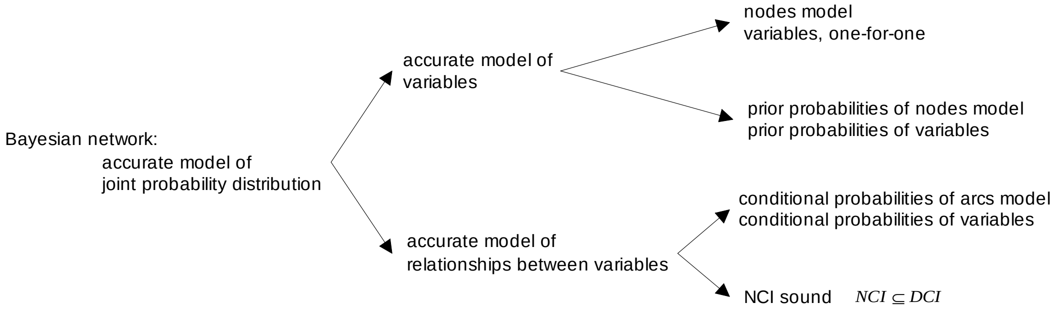


Figure 28

Provided the nodes of the Bayesian network model the variables of the joint probability distribution one-for-one, and given that the prior probabilities of the nodes of the Bayesian network and the conditional probabilities of the arcs of the Bayesian network are acquired directly from the data set, then the Bayesian network is an accurate model of the variables of the joint probability distribution.
In this case, the Bayesian network is an accurate model of the joint probability distribution, if and only if, the Bayesian network is an accurate model of the relationships between the variables of the joint probability distribution.
So, under these conditions, the Bayesian network is an accurate model of the joint probability distribution, if and only if, the NCI is sound ($NCI \subseteq DCI$).

4.2. Discussion

The framework for Bayesian Networks has been developed specifically for the field of Bayesian Networks; the framework is inspired by the fundamental principles governing the field of Bayesian Networks, employing concepts and methodologies associated with Bayesian networks.

The framework is based on formal theoretical foundations; the framework is formally established, with several definitions and theorems, and precise semantics.

The framework examines a Bayesian network with respect to a specific data set, and in particular with respect to the joint probability distribution represented by the specific data set.
A Bayesian network models the joint probability distribution represented by a given data set, it does not model the data entries of the data set.

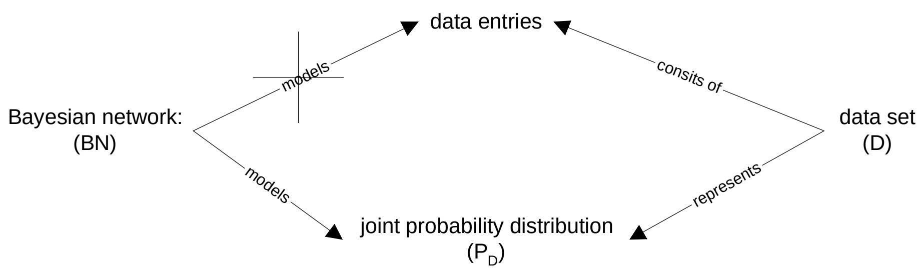


Figure 29

The framework identifies as a principal feature of a Bayesian network the conditional independencies implied by the structure of the Bayesian network; the NCI set.
The soundness (accuracy) of the NCI and the completeness of the NCI are principal characteristics of a Bayesian network, and reflect properties of a Bayesian network, under certain conditions.
In particular, when the nodes of the Bayesian network model the variables of the joint probability distribution represented by the data set one-for-one, and the probabilities of the Bayesian network are acquired directly from the data set, the soundness (accuracy) of the NCI reflects the accuracy of the

Bayesian network, while the completeness of the NCI affects the complexity and size of the Bayesian network.

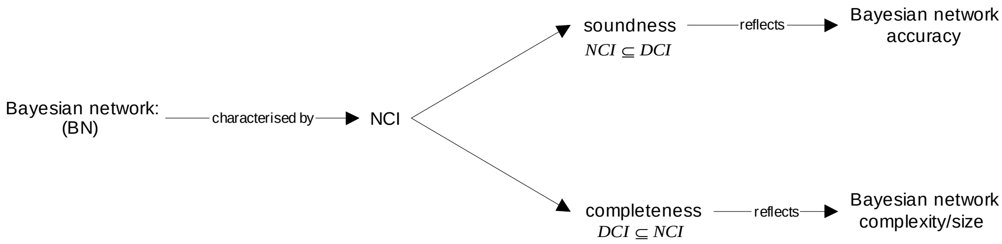


Figure 30

The framework models the accuracy of a Bayesian network as the soundness of the NCI; the accuracy of the conditional independencies implied by the structure of the Bayesian network.
The framework provides formal specifications for examining the soundness (accuracy) of the NCI, and supplies an evaluation scheme for the accuracy of a Bayesian network that allows the degree of inaccuracy of a particular Bayesian network to be determined precisely.

The framework presents the NCI Soundness Theorem and the NCI Incompleteness Theorem, which are remarkably significant for the field of Bayesian Networks.
The NCI Soundness Theorem indicates that given a data set there exists a Bayesian network whose NCI is sound, and so guarantees that given any data set it is possible to construct an accurate Bayesian network.
The NCI Incompleteness Theorem indicates that there exist data sets (at least one) for which there exist no Bayesian networks whose NCI is complete, and so points out that given any data set it might be impossible to construct a Bayesian network whose structure implies all the conditional independencies of the joint probability distribution represented by the data set.

The framework can be used in and applied to many areas of the field of Bayesian Networks.
The framework facilitates the comparison of individual Bayesian networks with regards to their characteristics, and provides the theoretical means to justify when and why a particular Bayesian network is more accurate than another Bayesian network.
The framework can clarify the procedure of the addition and deletion of arcs and nodes from the structure of a Bayesian network [Sucar 1991][Sucar et al 1993], and provide the means to illustrate and explain the effects of such actions on the characteristics of a Bayesian network.
The framework can supply a theoretical rationale to the process of the introduction of a hidden node within the structure of a Bayesian network [Kwoh 1995][Kwoh & Gillies 1996] and the effects of such an action.
The framework can be employed to develop Bayesian network construction (learning) algorithms [Tahseen 1998] as tree search algorithms that examine the accuracy of a Bayesian network and use the degree of inaccuracy as the evaluation function of the tree nodes.

4.3. Comparison

The framework for Bayesian Networks examines a Bayesian network with respect to the joint probability distribution represented by a particular data set, unlike the Minimum Description Length (MDL) formalism [Rissanen 1978] that examines a Bayesian network with respect to the data entries of a particular data set.
Consequently, the semantics of the framework are correct with regards to the field of Bayesian Networks, and therefore the framework does not exhibit the shortcomings of the MDL formalism.

The degree of inaccuracy of a Bayesian network is not affected by the size of the data set. When different data sets are used that represent, however, the same joint probability distribution, the degree of inaccuracy of a particular Bayesian network remains the same, regardless of the data set used to examine the accuracy of the Bayesian network.

The framework determines the "absolute" degree of inaccuracy of a Bayesian network, and so it can be used both for drawing conclusions about the accuracy of an individual Bayesian network and for comparing the accuracy of a set of Bayesian networks that model either different or the same data sets.

Of course, the framework has been developed specifically for the field of Bayesian Networks, unlike the MDL formalism, and it is entirely appropriate for examining the accuracy of a Bayesian network.

5. Experiments

So far, the report has examined the field of Bayesian Networks from a theoretical point of view; the definition of the accuracy of a Bayesian network, the presentation of the Minimum Description Length (MDL) formalism [Rissanen 1978], and the development of a framework for Bayesian Networks.

Now, the report examines the field of Bayesian Networks from an experimental point of view; several experiments are conducted to examine the MDL formalism and the proposed framework for Bayesian Networks.

The experiments constitute a significant part of the report, aiming to provide insight about the MDL formalism and the proposed framework for Bayesian Networks, and to support what has been previously stated in theory.

5.1. Present

The experiments examine two distinct real-world problems, each with different characteristics, and so two distinct real-world data sets are used.
The first data set represents information concerning Hepatitis C patients [Bang 1999], while the second data set provides information about the morphological development of neurons [Kim 2001][Kim & Gillies 1998].
The Hepatitis C data set has 1672 data entries; each being an individual Hepatitis C patient, characterised by 9 distinct variables.
The Neurons data set has 44 data entries; each being an individual neuron, characterised by 6 distinct variables.

Data set	**Hepatitis C (HC)**	**Neurons (N)**
Entries	1672	44
Variables	9 (A, B, C, D, E, F, G, H, I)	6 (A, B, C, D, E, F)
States for variables	A:2, B:10, C:9, D:7, E:9, F:12, G:6, H:9, I:2	A:4, B:8, C:8, D:6, E:6, F:6

The experiments are conducted using part of the real-world data sets; at any time, only three variables of the original real-world data set are used to carry out the experiments and collect experimental results.
Such a limited approach is adopted, so that it is computationally feasible to calculate the joint probability distribution represented by the data set. This is essential for conducting the experiments, and determining the actual degree of accuracy of a Bayesian network.
Therefore, a data set used for the experiments is uniquely identified by the original data set (that it is derived from), and the set of variables (that are being actually used).
For example, "$HC:ABC$" denotes that the data set used for the experiments is the data set obtained from the Hepatitis C data set using only the variables A, B, and C, while "$N:BCD$" denotes that the data set used for the experiments is the data set obtained from the Neurons data set using only the variables B, C, and D. Of course, the "$HC:ABC$" data set is different from the "$HC:BCD$" data set, which is, in turn, different from the "$N:BCD$" data set.

Data set	HC:ABC
Entries	1672
Variables	3 (A, B, C)
States for variables	A:2, B:10, C:9

The experiments are conducted for a collection of different data sets, which are acquired from the original Hepatitis C data set and the original Neurons data set, by randomly selecting the set of variables.

$$D_{HC} = \{HC:ABC, HC:ABD, HC:ABE, HC:ABF, HC:BCD\}$$
$$D_{N} = \{N:ABC, N:ABD, N:ABE, N:ABF, N:BCD\}$$
$$D = D_{HC} \cup D_{N}$$

The experiments are conducted for a collection of different Bayesian networks. In particular, for a given data set, all possible tree structured Bayesian networks that can model the given data set are used to carry out the experiments and collect experimental results.

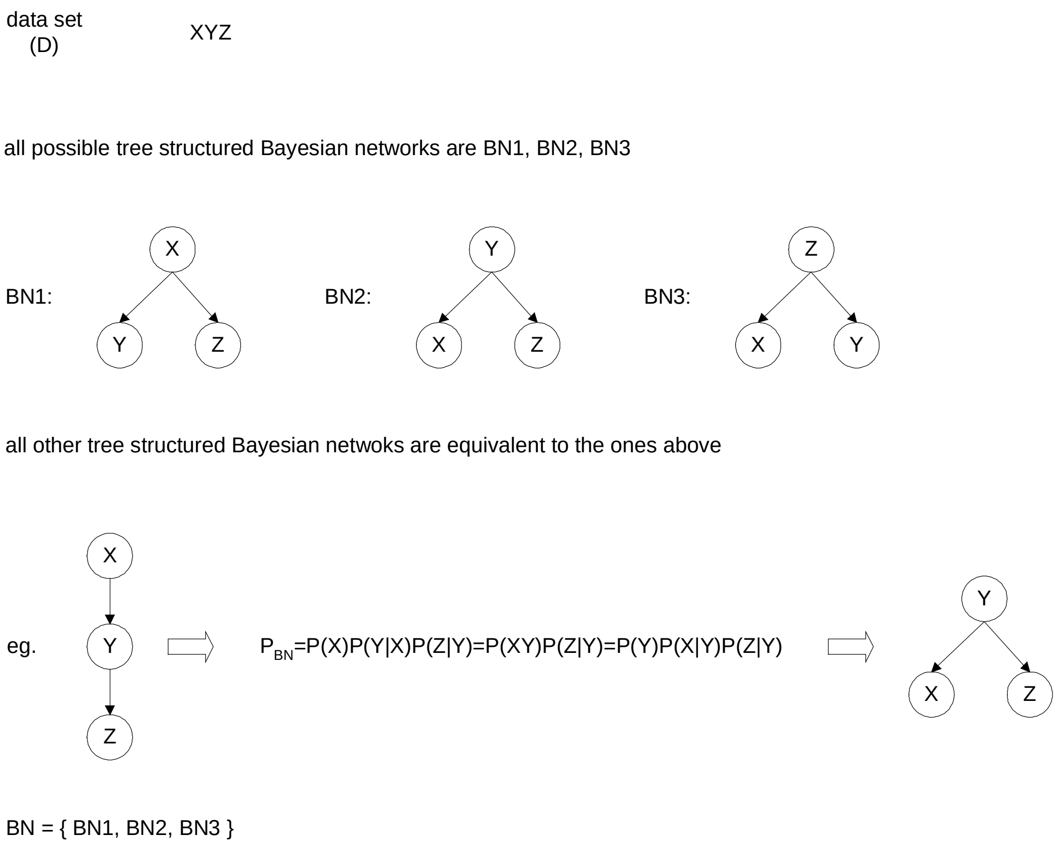


Figure 31

Therefore, the experiments are conducted for a collection of different data sets, and for the set of all possible tree structured Bayesian networks for each particular data set.
For a given data set and a given Bayesian network, the experiments determine the accuracy of the Bayesian network with respect to the data set.
In particular, the experiments determine the actual degree of inaccuracy of the Bayesian network (Definition 2.4), the degree of inaccuracy of the

Bayesian network according to the Minimum Description Length formalism (Definition 3.2), and the degree of inaccuracy of the Bayesian network according to the proposed framework for Bayesian Networks (Definition 4.20).

$\forall D, \forall BN$
determine:

1. $inaccuracy(BN)$, according to the actual measure (Definition 2.4)
2. $inaccuracy(BN)$, according to MDL (Definition 3.2)
3. $inaccuracy(BN)$, according to NCIMI (Definition 4.20)

5.2. Experimental results

A software prototype has been implemented, which conducts the experiments for the given collection of data sets, and the corresponding sets of Bayesian networks.

	Data Set	Bayesian Network	inaccuracy(BN) actual	inaccuracy(BN) MDL	inaccuracy(BN) NCIMI
Hepatitis C (1672)	ABC	B ← A → C	0.159	8496.75	0.567
		A ← B → C	0.018	7607.67	0.036
		A ← C → B	0.025	7615.20	0.040
	ABD	B ← A → D	0.121	7537.36	0.287
		A ← B → D	0.012	7080.98	0.014
		A ← D → B	0.025	7097.85	0.024
	ABE	B ← A → E	0.021	6992.09	0.063
		A ← B → E	0.032	6954.81	0.041
		A ← E → B	0.029	6931.39	0.027
	ABF	B ← A → F	0.018	8861.48	0.053
		A ← B → F	0.018	8820.34	0.028
		A ← F → B	0.021	8828.10	0.033
	BCD	C ← B → D	0.127	8572.08	0.457
		B ← C → D	0.023	8018.15	0.125
		B ← D → C	0.099	8460.20	0.390

	Data Set	Bayesian Network	inaccuracy(BN) actual	inaccuracy(BN) MDL	inaccuracy(BN) NCIMI
Neurons (44)	ABC	B ← A → C	0.108	245.20	0.710
		A ← B → C	0.109	256.41	0.965
		A ← C → B	0.085	233.99	0.455
	ABD	B ← A → D	0.114	253.70	0.897
		A ← B → D	0.093	243.56	0.667
		A ← D → B	0.100	244.31	0.684
	ABE	B ← A → E	0.090	217.81	0.671
		A ← B → E	0.132	232.76	1.010
		A ← E → B	0.103	209.28	0.477
	ABF	B ← A → F	0.103	234.85	0.748
		A ← B → F	0.101	234.41	0.738
		A ← F → B	0.148	250.83	1.111
	BCD	C ← B → D	0.098	273.21	0.932
		B ← C → D	0.106	273.35	0.935
		B ← D → C	0.105	275.17	0.976

The experimental results are also presented in a graphical form; the graphs are used to illustrate the properties of the Minimum Description Length formalism and the proposed framework for Bayesian Networks.

The graphs are scatterplots [Johnston 2000] of the actual degree of inaccuracy of the Bayesian network, against either the degree of inaccuracy of the Bayesian network according to MDL, or the degree of inaccuracy of the Bayesian network according to the proposed framework.
Hence, different graphs are used for the experimental results collected for MDL, and the experimental results collected for the proposed framework.

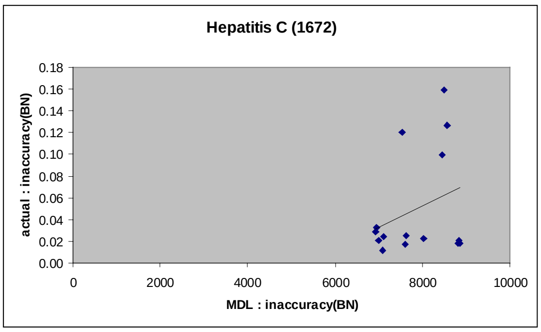
There are two types of graphs; the first type of graphs presents the experimental results as different sets of points, grouped according to the data set used; the second type of graphs presents the experimental results as one set of points, regardless of the data set used.

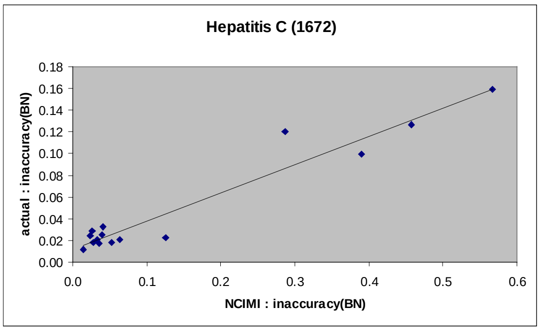
In addition, different graphs are used for the experimental results collected for the data sets of the Hepatitis C domain, and the experimental results collected for the data sets of the Neurons domain.

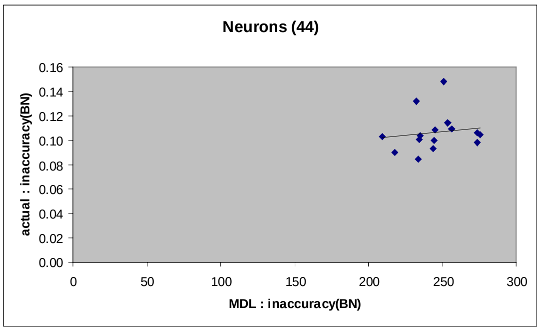
The graphs also display the best trendline for the points of the graph [Johnston 2000]; the line that best fits the points, so that the overall distance of the points from the trendline is minimum.

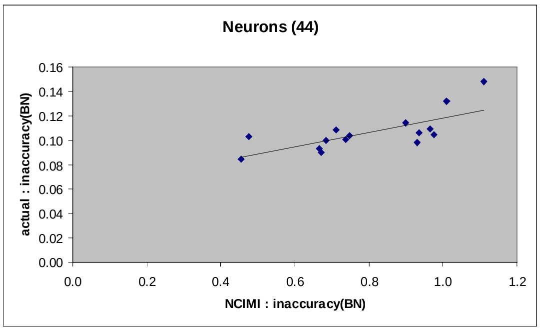
A trendline is characterised by the R^2 value, which reflects the reliability of the trendline. The closer the R^2 value of the trendline is to 1, the more reliable the trendline is, and the better it fits the points of the graph. However, in the case when the points of the graph are either too few or too close to each other, the best trendline is least reliable, and fits the points poorly.

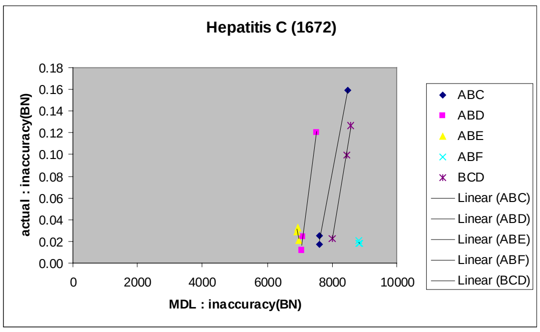
The best trendline for the points of a graph is associated with the correlation of the variables represented in the graph. A low correlation coefficient indicates that the variables are unrelated and so the best trendline fits the corresponding points of the graph very poorly, while a high correlation coefficient indicates that the variables are correlated and so the best trendline fits the corresponding points of the graph well.

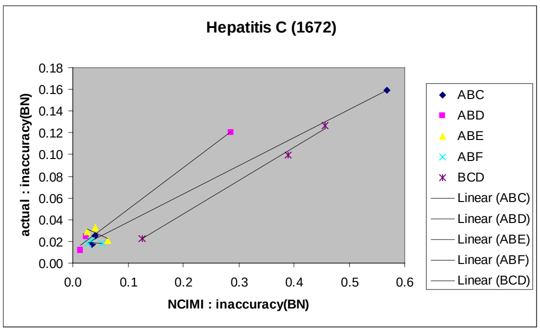


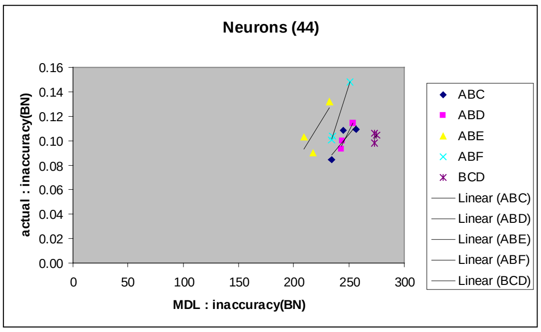


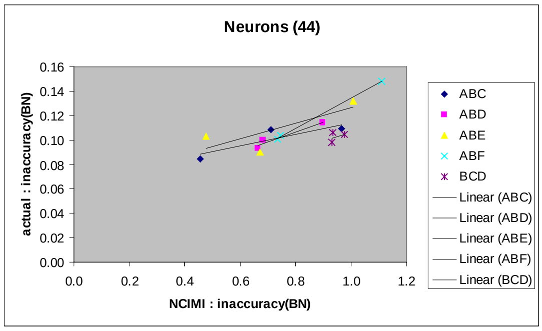












5.3. Discussion

The experiments are conducted to examine the MDL formalism and the proposed framework for Bayesian Networks, and determine how well each of these methodologies models the inaccuracy of a Bayesian network.
Such an examination requires the identification of the characteristics and properties of the MDL formalism and the proposed framework for Bayesian Networks, and the classification of the advantages and disadvantages of each methodology in view of the experimental results.

5.3.1. Minimum Description Length (MDL) formalism

According to the MDL formalism, the degree of inaccuracy of a Bayesian network, with respect to a data set, is affected directly by the size of the data set [Friedman & Yakhini 1996].
For the data sets of the Hepatitis C domain, the average degree of inaccuracy of the Bayesian networks examined is 7858.30; for the data sets of the Neurons domain, the average degree of inaccuracy of the Bayesian networks examined is 245.26. Thus, the average degree of inaccuracy of the Bayesian networks for the data sets of the Hepatitis C domain is 32.04 times greater than the average degree of inaccuracy of the Bayesian networks for the data sets of the Neurons domain. This result is not only of the same magnitude but also very close to the size ratio of the Hepatitis C data set and the Neurons data set, which is 38.
The experimental results point to a linear correlation between the degree of inaccuracy of a Bayesian network, with respect to a data set, and the size of the data set, which agrees with what is predicted in theory.

The correlation coefficient between the actual degree of inaccuracy of a Bayesian network and the degree of inaccuracy of a Bayesian network according to MDL is significantly low, while the best trendlines of the corresponding scatterplots are significantly unreliable and fit the points of the scatterplots poorly.

	Hepatitis C	**Neurons**
Correlation coefficient	0.31	0.14
R^2 **value**	0.09	0.02

So, the experimental results point to a significantly weak correlation, indicating that the actual degree of inaccuracy of a Bayesian network and

the degree of inaccuracy of a Bayesian network according to MDL are not correlated.
Therefore, the MDL formalism does not actually determine the "absolute" degree of inaccuracy of a Bayesian network.

The correlation coefficient between the actual degree of inaccuracy of a Bayesian network and the degree of inaccuracy of a Bayesian network according to MDL, for a particular data set, is reasonably high, while the best trendlines of the corresponding scatterplots are reasonably reliable and fit the points of the scatterplots well.

	HC:ABC	**HC:ABD**	**HC:ABE**	**HC:ABF**	**HC:BCD**
Correlation coefficient	0.99	0.99	-0.76	-0.26	0.99
R^2 **value**	0.99	0.99	0.58	0.07	0.99

	N:ABC	**N:ABD**	**N:ABE**	**N:ABF**	**N:BCD**
Correlation coefficient	0.88	0.97	0.79	0.99	0.42
R^2 **value**	0.78	0.95	0.62	0.99	0.18

So, the experimental results, for a particular data set, point to a reasonably strong correlation, indicating that the actual degree of inaccuracy of a Bayesian network and the degree of inaccuracy of a Bayesian network according to MDL, for a particular data set, are correlated.
Therefore, although the MDL formalism does not actually determine the "absolute" degree of inaccuracy of a Bayesian network, it does determine, to some extent, the "relative" degree of inaccuracy of a Bayesian network for a particular data set.

Since the MDL formalism does not determine the "absolute" degree of inaccuracy, it is not possible to examine the "absolute" inaccuracy of a Bayesian network.
For example, the degree of inaccuracy according to MDL of the Bayesian network $B \leftarrow A \rightarrow C$ with respect to the data set $HC:ABC$, which has a value of 8496.75, cannot be used to draw conclusions about the inaccuracy of the Bayesian network.

The MDL formalism should only be used to compare the "relative" inaccuracy of Bayesian networks for a particular data set. Any results acquired and any conclusions drawn are valid for the specific set of Bayesian networks, and only in view of the particular data set. A comparison between the degrees of inaccuracy of Bayesian networks for different data sets is not valid.
For example, the degree of inaccuracy according to MDL of a Bayesian network with respect to a data set of the Hepatitis C domain should never be compared to the degree of inaccuracy according to MDL of a Bayesian network with respect to a data set of the Neurons domain. Besides, such a comparison would point to the bizarre conclusion that any of the Bayesian networks used for the data sets of the Neurons domain is less inaccurate than any of the Bayesian networks used for the data sets of the Hepatitis C domain.
However, the degree of inaccuracy according to MDL of the Bayesian network $B \leftarrow A \rightarrow C$ with respect to the data set $HC:ABC$, which has a value of 8496.75, can be compared to the degree of inaccuracy according to MDL of

the Bayesian network $A \leftarrow B \rightarrow C$ with respect to the data set $HC:ABC$, which has a value of 7607.67, since the Bayesian networks model the same data set.

5.3.2. Framework for Bayesian Networks

The correlation coefficient between the actual degree of inaccuracy of a Bayesian network and the degree of inaccuracy of a Bayesian network according to the proposed framework is significantly high, while the best trendlines of the corresponding scatterplots are significantly reliable and fit the points of the scatterplots well.

	Hepatitis C	**Neurons**
Correlation coefficient	0.97	0.73
R^2 **value**	0.94	0.53

So, the experimental results point to a significantly strong correlation, indicating that the actual degree of inaccuracy of a Bayesian network and the degree of inaccuracy of a Bayesian network according to the proposed framework are correlated.
Therefore, the proposed framework for Bayesian Networks does determine the “absolute” degree of inaccuracy of a Bayesian network.

Since the proposed framework for Bayesian Networks determines the “absolute” degree of inaccuracy, it is possible to examine the “absolute” inaccuracy of a Bayesian network.
For example, the degree of inaccuracy according to the proposed framework of the Bayesian network $B \leftarrow A \rightarrow C$ with respect to the data set $HC:ABC$, which has a value of 0.567, can be used to draw conclusions about the inaccuracy of the Bayesian network.

Furthermore, the proposed framework for Bayesian Networks can also be used to compare Bayesian networks with regards to their inaccuracy. The comparison of the degrees of inaccuracy of the Bayesian networks can be carried out, regardless of whether the Bayesian networks model the same data set or different data sets.
For example, the degree of inaccuracy according to the proposed framework of a Bayesian network with respect to a data set of the Hepatitis C domain could easily be compared to the degree of inaccuracy according to the proposed framework of a Bayesian network with respect to a data set of the Neurons domain.
Of course, comparisons can be carried out for Bayesian networks that model the same data set as well, and so the degree of inaccuracy according to the proposed framework of the Bayesian network $B \leftarrow A \rightarrow C$ with respect to the data set $HC:ABC$, which has a value of 0.567, can be compared to the degree of inaccuracy according to the proposed framework of the Bayesian network $A \leftarrow B \rightarrow C$ with respect to the data set $HC:ABC$, which has a value of 0.036.

5.4. Conclusion

The experiments offer substantial support to the theory, and constitute, in some degree, a proof of soundness of the theory.
The experimental results support the theoretical statements, and agree with what is predicted by the theory.

Despite the use of distinctly different data sets from two unrelated domains, the experimental results point to the same conclusions regardless of the data set used.

The proposed framework for Bayesian Networks does not exhibit the shortcomings of the Minimum Description Length (MDL) formalism, but instead, it overcomes the limitations of MDL, and manages to model the accuracy of a Bayesian network correctly.

The MDL formalism fails to model the accuracy of a Bayesian network precisely, so that MDL can only be used in a limited context, examining only the “relative” accuracy of a Bayesian network for a particular data set.

On the other hand, the proposed framework for Bayesian Networks models the accuracy of a Bayesian network correctly, so that the proposed framework can be used in a more general context than MDL, examining the “absolute” accuracy of a Bayesian network.

	Correlation (Hepatitis C)	**Correlation (Neurons)**	**R^2 value (Hepatitis C)**	**R^2 value (Neurons)**
MDL	0.31	0.14	0.09	0.02
NCIMI	0.97	0.73	0.94	0.53

Both methodologies can be employed in the field of Bayesian Networks; however, the proposed framework for Bayesian Networks is evidently more appropriate than the MDL formalism.

6. Conclusion

The aim of the research is the study of a Bayesian network; the identification and the examination of the features and the characteristics of a Bayesian network.

The report examines the accuracy of a Bayesian network with respect to a given data set; the degree to which the Bayesian network models the joint probability distribution represented by the given data set accurately.

6.1. Review

Initially, the report provides a formal definition for a Bayesian network, based on solid theoretical foundations. The features of a Bayesian network are presented, and the conditional independencies implied by the structure of the network are identified as a principal feature.

Then, the report identifies the accuracy as a principal characteristic of a Bayesian network. A formal definition of the accuracy of a Bayesian network with respect to a given data set is provided, together with a method for determining the degree of accuracy of a Bayesian network. However, the methodology presented is computationally unfeasible in practice for real-world applications.

The report examines the Minimum Description Length (MDL) formalism as an alternative methodology for evaluating the accuracy of a Bayesian network in practice for real-world applications. However, the semantics of the MDL formalism are not entirely appropriate for the field of Bayesian Networks, and thus, MDL exhibits several shortcomings. Other approaches, such as the Akaike Information Criterion (AIC) [Akaike 1974] and the Bayesian Information Criteria (BIC) [Schwarz 1978], are similar to the MDL formalism, and so, exhibit several weaknesses.

Subsequently, the report proposes a framework for Bayesian Networks, which attempts to overcome the shortcomings and weaknesses of previous formalisms. The proposed framework examines the accuracy of a Bayesian network, by examining the conditional independencies implied by the structure of the Bayesian network. The framework is formally established, with several definitions and theorems, and well-defined semantics.

Finally, the report presents experiments that have been conducted in order to examine the MDL formalism and the proposed framework for Bayesian Networks. The experimental results provide significant insight, and confirm what has been previously stated in theory.

6.2. Discussion

The main contribution of the research is the development of a framework for Bayesian Networks.

The framework has been specifically developed for the field of Bayesian Networks, inspired by fundamental principles governing this field, and employing concepts and methodologies associated with Bayesian networks.

The framework is based on formal theoretical foundations, employing several definitions and theorems, and characterised by precise semantics.

The framework provides formal specifications in order to examine the accuracy of a Bayesian network with respect to a given data set, and specifies an evaluation scheme in order for the degree of inaccuracy of a Bayesian network to be determined precisely.

Although the report is focused on the accuracy of a Bayesian network, the framework examines a Bayesian network from a general point of view, identifying several other features and characteristics of a Bayesian network.
The framework can be used in a wide context, and examine all the features and characteristics of a Bayesian network.

The framework can be used in and applied to many areas of the field of Bayesian Networks; it can facilitate the comparison of individual Bayesian networks with regards to their characteristics, clarify the procedure of the addition and deletion of arcs and nodes from the structure of a Bayesian network, supply a theoretical rationale for the process of the introduction of a hidden node within the structure of a Bayesian network, develop construction (learning) algorithms for Bayesian networks, and so on.

The framework has a substantial impact on the field of Bayesian Networks. It is a rigid and precise, yet simple and intuitive, framework that accurately reflects the essence of a Bayesian network.

6.3. Future work

So far, the research into Bayesian Networks has been considerably systematic; however, it is not yet complete. Several issues need to be investigated in order for the research to be thoroughly comprehensive.

The framework for Bayesian Networks has to be refined and improved, by enriching the theoretical foundations, and identifying subtle and distinctive features of the theory. In addition, several assumptions have to be relaxed or even lifted in order to augment the framework.
Although this research is a uniquely systematic examination of the features and characteristics of a Bayesian network, there is past relevant research. In particular, the framework has subtle similarities with research by Pearl [Pearl 1988], Neapolitan [Neapolitan 1990], and Chow & Liu [Chow & Liu 1968]. Further investigation is required to identify these similarities, and to amalgamate these seemingly different methodologies into a unified framework.
The theoretical foundations of the framework focus on the structure of a Bayesian network and the implied conditional independencies. However, the framework can be refined to examine not only the conditional independencies implied by the structure of a Bayesian network but also the

conditional independencies implied by the conditional probabilities matrices of the arcs of a Bayesian network.

Although the report is focused on the accuracy of a Bayesian network, the framework can be used in an extended context, and facilitates the identification and study of additional characteristics of a Bayesian network.
In particular, the framework can be extended to examine the complexity, the classification performance, and the classification accuracy of a Bayesian network.

The experiments conducted to examine the framework for Bayesian Networks are relatively simple. Further experimentation is essential, and supplementary experiments need to be carried out.
The Bayesian networks employed for the experiments are singly-connected networks whose nodes have single parents. Supplementary experiments would employ singly-connected Bayesian networks whose nodes have multiple parents, and multiply-connected Bayesian networks.
The data sets employed for the experiments are data sets of just three variables. Supplementary experiments would employ larger data sets of more than three variables. However, the data sets have to remain relatively small so that the experiments are computationally feasible.
In addition to real-world data sets, experiments can also be carried out employing artificial data sets.

The refinement and the extension of the framework will result in an even greater understanding of Bayesian Networks, while the supplementary experiments will offer further insight, and additional evidence in support of the proposed framework.

The research is to be completed within the following nine months.

References

[Akaike 1974]
Akaike H.
"A new look at the statistical identification model"
IEEE Transactions on Automatic Control, 19:716-723, 1974

[Bang 1999]
Bang J.W.
"Medical prognostic expert system using Bayesian network in Hepatitis C"
PhD report (2nd year), Imperial College, 1999

[Chow & Liu 1968]
Chow C.K., Liu C.N.
"Approximating discrete probability distributions with dependence trees"
IEEE Transactions on Information Theory, 14:462-467, 1968

[Friedman & Yakhini 1996]
Friedman N., Yakhini Z.
"On the sample complexity of learning Bayesian networks"
Proceedings of the 12th Annual Conference on Uncertainty in Artificial Intelligence, 1996

[Friedman et al 1997]
Friedman N., Geiger D., Goldszmidt M.
"Bayesian network classifiers"
Machine Learning, 1997

[Gillies & Guo 2000]
Gillies D., Guo Y.
"Bayesian networks"
Lecture notes, Intelligent data analysis and probabilistic inference, M.Sc. Advanced Computing, Imperial College, 2000

[Guo 1997]
Guo Y.
"Bayesian networks"
Lectures notes, Artificial Intelligence, M.Sc. Advanced Computing, Imperial College, 1997

[Jensen 1996]
Jensen F.V.
"An introduction to Bayesian networks"
UCL Press, 1996

[Johnston 2000]
Johnston I.
"I'll give you a definite maybe: an introductory handbook on probability, statistics, and Excel"
Lecture notes, Liberal Studies, Malaspina University-College, 2000

[Kim 2001]
Kim J.
"Automatic morphometric analysis of neural cells"
Ph.D. Thesis, Imperial College, 2001

[Kim & Gillies 1998]
Kim J., Gillies D.F.
"Automatic Morphometric Analysis of Neural Cells"
Machine Graphics and Vision 7(4), 1998

[Kwoh 1995]
Kwoh C.K.
"Probabilistic reasoning from correlated objective data"
Ph.D. Thesis, Imperial College, 1995

[Kwoh & Gillies 1996]
Kwoh C.K., Gillies D.F.
"Using Hidden Nodes in Bayesian Networks"
Artificial Intelligence 88:1-38, 1996

[Pappas & Gillies 2002]
Pappas A., Gillies D.
"A New Measure for the Accuracy of a Bayesian Network"
Proceedings of the Mexican International Conference on Artificial Intelligence (MICAI), 2002 (to appear)

[Neapolitan 1990]
Neapolitan R.E.
"Probabilistic reasoning in expert systems: theory and algorithms"
Wiley-Interscience, 1990

[Pearl 1988]
Pearl J.
"Probabilistic reasoning in intelligent systems: networks of plausible inference"
Morgan Kaufmann, 1988 (4th printing, 1997)

[Rissanen 1978]
Rissanen J.
"Modelling by shortest data description"
Automatica, 14:465-471, 1978

[Russell & Norvig 1995]
Russell S., Norvig P.
"Artificial Intelligence: a modern approach"
Prentice Hall International, 1995

[Sucar 1991]
Sucar L.E.
"Probabilistic reasoning in knowledge-based vision systems"
Ph.D. Thesis, Imperial College, 1991

[Sucar et al 1993]

Sucar L.E., Gillies D.F., Gillies D.A.
"Uncertainty Management in Expert Systems"
Artificial Intelligence 61:187-208, 1993

[Schwarz 1978]
Schwarz G.
"Estimate the dimension of a model"
The Annals of Statistics, 6(2):461-464, 1978

[Tahseen 1998]
Tahseen T.
"A new approach to learning Bayesian network classifiers"
Ph.D. Thesis, Imperial College, 1998

Appendix A – Proofs

Theorem 2.1

For a data set (D), there exists a Bayesian network (BN) that models the joint probability distribution described by the data set (P_D) accurately.
$\forall D, \exists BN : P_{BN} = P_D$

Proof

This is a trivial proof, since in this case, the data set (D) uniquely specifies a joint probability distribution (P_D), and so it is apparent according to Theorem 1.2 that for the joint probability distribution described by the data set (P_D), there exists a Bayesian network (BN) that models the joint probability distribution accurately.

Theorem 4.1 (NCI Theorem)

Given:

- A data set (D) representing a joint probability distribution (P) over variables (V)
  $P_D = P(V)$
- A Bayesian network (BN) of nodes (N), arcs (A), distribution (P')
  $BN = (N, A, P')$
  $P_{BN} = P'(N)$

Provided:

- The nodes of the Bayesian network are the variables of the joint probability distribution represented by the data set
  $N \equiv V$
  $\forall i, n_i \in N, v_i \in V : n_i \equiv v_i$
- The probabilities of the Bayesian network reflect the corresponding probabilities of the joint probability distribution represented by the data set
  $\forall i, n_i \in N : P'(n_i \mid parents(n_i)) = P(n_i \mid parents(n_i))$

➢ The Bayesian network is accurate with respect to the data set, **if and only if**, the conditional independencies implied by the structure of the Bayesian network are conditional independencies of the joint probability distribution represented by the data set
$P_{BN} = P_D \Leftrightarrow$
$\forall i, v_i \in V, \forall W, W \subseteq a(v_i) : P(v_i \mid W \cup parents(v_i)) = P(v_i \mid parents(v_i))$

Proof

1. $\forall i, v_i \in V, \forall W, W \subseteq a(v_i) : P(v_i \mid W \cup parents(v_i)) = P(v_i \mid parents(v_i)) \Rightarrow$
$P_{BN} = P_D$

Since the Bayesian network is a Directed Acyclic Graph (DAG), then there is an ancestral ordering of the nodes of the Bayesian network. [4]
Let $[v_1,...,v_n]$ be the ancestral ordering, considering that $\forall i : n_i \equiv v_i$
It follows from the ancestral ordering that:
$\forall i : parents(v_i) \subseteq \{v_1,...,v_{i-1}\}$
$\forall i : descendants(v_i) \subseteq \{v_{i+1},...,v_n\}$
So, it holds that:
$\forall i : parents(v_i) \cup \{v_1,...,v_{i-1}\} = \{v_1,...,v_{i-1}\}$
The set W is any subset of the set $a(v_i)$
$a(v_i) = V - (descendants(v_i) \cup \{v_i\})$
Consider $W = \{v_1,...,v_{i-1}\}$, then $W \subseteq a(v_i)$
It holds that:
$W \cup parents(v_i) = \{v_1,...,v_{i-1}\} \cup parents(v_i) = \{v_1,...,v_{i-1}\}$

For the Bayesian network, it holds that:
$$P_{BN} = P'(N) \overset{1}{\Leftrightarrow}$$
$$P_{BN} = \prod_{i=1}^{n} P'(n_i \mid parents(n_i)) \overset{2}{\Leftrightarrow}$$
$$P_{BN} = \prod_{i=1}^{n} P(n_i \mid parents(n_i)) \overset{3}{\Leftrightarrow}$$
$$P_{BN} = \prod_{i=1}^{n} P(v_i \mid parents(v_i)) \overset{4}{\Leftrightarrow}$$
$$P_{BN} = \prod_{i=1}^{n} P(v_i \mid W \cup parents(v_i)) \overset{5}{\Leftrightarrow}$$
$$P_{BN} = \prod_{i=1}^{n} P(v_i \mid \{v_1,...,v_{i-1}\}) \overset{6}{\Leftrightarrow}$$
$$P_{BN} = P(V) \Leftrightarrow$$
$$P_{BN} = P_D$$
q.e.d.

$\overset{1}{\Leftrightarrow}$: Because of Theorem 1.1
$\overset{2}{\Leftrightarrow}$: Because the probabilities of the Bayesian network reflect the corresponding probabilities of the joint probability distribution represented by the data set (provided)
$\overset{3}{\Leftrightarrow}$: Because the nodes of the Bayesian network are the variables of the joint probability distribution represented by the data set (provided)

[4] [Neapolitan 1990] Theorem 3.7

$\overset{4}{\Leftrightarrow}$: Because the conditional independencies implied by the structure of the Bayesian network are conditional independencies of the joint probability distribution represented by the data set (condition)

$\overset{5}{\Leftrightarrow}$: Because of $W \cup parents(v_i) = \{v_1,...,v_{i-1}\}$ (proven)

$\overset{6}{\Leftrightarrow}$: Because of the probabilities chain rule

2. $P_{BN} = P_D \Rightarrow$
$\forall i, v_i \in V, \forall W, W \subseteq a(v_i) : P(v_i \mid W \cup parents(v_i)) = P(v_i \mid parents(v_i))$

This is a trivial proof, since in this case, the joint probability distribution represented by the Bayesian network is identical to the joint probability distribution described by the data set, and so it is apparent that the conditional independencies implied by the structure of the Bayesian network are indeed conditional independencies of the joint probability distribution described by the data set.

It holds that:

$$P_{BN} = P_D \Leftrightarrow$$
$$P'(N) = P(V) \overset{1}{\Leftrightarrow}$$
$$P'(N) = P(N) \Leftrightarrow$$
$$P' = P$$

$\overset{1}{\Leftrightarrow}$: Because the nodes of the Bayesian network are the variables of the joint probability distribution represented by the data set (provided)

For the Bayesian network, it holds that:

$$\forall i, v_i \in V, \forall W, W \subseteq a(v_i) : P'(n_i \mid W \cup parents(n_i)) = P'(n_i \mid parents(n_i))$$

So, it holds that:

$$\forall i, v_i \in V, \forall W, W \subseteq a(v_i) :$$
$$P'(n_i \mid W \cup parents(n_i)) = P'(n_i \mid parents(n_i)) \overset{1}{\Leftrightarrow}$$
$$P(n_i \mid W \cup parents(n_i)) = P(n_i \mid parents(n_i)) \overset{2}{\Leftrightarrow}$$
$$P(v_i \mid W \cup parents(v_i)) = P(v_i \mid parents(v_i))$$
$$q.e.d.$$

$\overset{1}{\Leftrightarrow}$: Because of $P' = P$ (proven)

$\overset{2}{\Leftrightarrow}$: Because the nodes of the Bayesian network are the variables of the joint probability distribution represented by the data set (provided)

Theorem 4.5 (NCI Soundness Theorem)

For a data set (D), there exists a Bayesian network (BN) so that the NCI is sound.
$\forall D, \exists BN : NCI \subseteq DCI$

Proof

According to Theorem 2.1:
For a data set (D), there exists a Bayesian network (BN) that models the joint probability distribution described by the data set (P_D) accurately.
$\forall D, \exists BN : P_{BN} = P_D$

According to Definition 2.1:
A Bayesian network (BN) is accurate with respect to a data set (D), if and only if, the joint probability distribution represented by the Bayesian network (P_{BN}) matches the joint probability distribution described by the data set (P_D).
$\text{accurate } BN \text{ with respect to D} \Leftrightarrow P_{BN} = P_D$

So, it holds that:
For a data set (D), there exists a Bayesian network (BN) that is accurate with respect to the data set (D).
$\forall D, \exists BN : \text{accurate } BN \text{ with respect to D}$

According to Theorem 4.3:
A Bayesian network is accurate, if and only if, $NCI \subseteq DCI$.

So, it holds that:
For a data set (D), there exists a Bayesian network (BN) so that $NCI \subseteq DCI$ (i.e. the NCI is sound).
$\forall D, \exists BN : NCI \subseteq DCI$

Theorem 4.6 (NCI Incompleteness Theorem)

There exists a data set (D), so that for any Bayesian network (BN), the NCI is incomplete.
$\exists D, \forall BN : DCI \not\subseteq NCI$
$\exists D, \neg\exists BN : DCI \subseteq NCI$

Proof

According to Pearl [Pearl 1988], there are models of dependencies (i.e. joint probability distributions) that cannot be represented by Bayesian networks (i.e. $NCI = DCI$). [5]
$\exists JPRD, \neg\exists BN : NCI = DCI$

---

[5] weak transitivity and chordality are some of the dependencies that are not representable by Bayesian networks

So, there are models of dependencies (i.e. joint probability distributions) and data sets that describe such models of dependencies, for which it holds that:

$\exists D : \neg \exists BN : NCI = DCI \Leftrightarrow$

$\exists D : \neg \exists BN : (NCI \subseteq DCI) \wedge (DCI \subseteq NCI) \Leftrightarrow$

$\exists D : \forall BN : \neg((NCI \subseteq DCI) \wedge (DCI \subseteq NCI)) \Leftrightarrow$

$\exists D : \forall BN : (\neg(NCI \subseteq DCI)) \vee (DCI \not\subseteq NCI) \Leftrightarrow$

$\exists D : \forall BN : (NCI \subseteq DCI) \rightarrow (DCI \not\subseteq NCI)$

Similarly, it holds that:

$\exists D : \forall BN : (DCI \subseteq NCI) \rightarrow (NCI \not\subseteq DCI)$

So, for certain data sets, if there exist Bayesian networks whose NCI is sound, then for those Bayesian networks the NCI is incomplete.
This indicates that, for certain data sets, there do not exist Bayesian networks whose NCI is both sound and complete.

Of course, this allows for the case when there exist Bayesian networks whose NCI is not sound but may be complete.
This agrees with Pearl's research, which indicates that there exist Bayesian networks that employ auxiliary nodes and whose NCI is complete (and thus not sound).

However, if the argument is limited to the case of Bayesian networks that do not use auxiliary nodes, then it can be proven that there exist data sets so that there exist no Bayesian networks whose NCI is complete.

$\exists D, \neg \exists BN : DCI \subseteq NCI$

This is a complex proof, which employs results of research by Pearl.
For the theoretical background, and an in-depth discussion of the fact that there are models of dependencies (i.e. joint probability distributions) that cannot be represented by Bayesian networks, the reader is advised to refer to Pearl [Pearl 1988], and in particular sections 3.1.4 and 3.3.3.

Below, we provide a "proof by example" of the fact that there exist data sets so that there exist no Bayesian networks (that do not use auxiliary nodes) and whose NCI is complete.

Consider a joint probability distribution so that $DCI = \{BC, BC \mid A\}$, and a Bayesian network employing only three nodes A, B, C to model the given joint probability distribution.
For the conditional independency BC to be implied by the structure of the Bayesian network, these are the possible structures:

$A\,B\,C$

$A \rightarrow C\;B$

$A \leftarrow C\;B$

$A \rightarrow B\;C$

$A \leftarrow B\;C$

$B \rightarrow A \leftarrow C$

For the conditional independency $BC \mid A$ to be implied by the structure of the Bayesian network, these are the possible structures:

$B \leftarrow A \rightarrow C$

$B \rightarrow A \rightarrow C$

$B \leftarrow A \leftarrow C$

It is evident that there is no common structure that represents both conditional independencies.

Therefore, there is a joint probability distribution, so that there is no Bayesian network, which does not use auxiliary variables, and whose NCI is complete.