**19MAT301**

**BELIEF NETWORKS**

Anuvarshini S.P Poduru Rama Sailaja

(CB.EN.U4AIE19011) (CB.EN.U4AIE19045)

CSE-AI CSE-AI

Amrita Vishwa Vidyapeetham

**ABSTRACT**

Bayesian network is a combination of probabilistic model and graph model. It is applied widely in machine learning, data mining, diagnosis, etc. because it has a solid evidence-based inference which is familiar to human intuition. However, Bayesian network may cause confusions because there are many complicated concepts, formulas and diagrams relating to it. Such concepts should be organized and presented in such a clear manner that understanding it is easy. This is the goal of this report. This report includes principles of Bayesian network. Main contents of this reported are extracted from the book David Barber, Bayesian Reasoning and Machine Learning, Cambridge University Press, 2012. This report focuses on discrete Bayesian network whose nodes are discrete random variables.

**ACKNOWLEDGMENT**

This project would not have been possible without the support and guidance Prof. K.P Soman who has provided us with sufficient knowledge and guidelines even during these unprecedented and trying times.

We would also like to thank our entire Computer Science and Engineering - Artificial Intelligence Department for giving us this opportunity to nurture and hone our skills.

Furthermore, we would like to thank the Amrita Vishwa Vidyapeetham management for providing ample resources to avail our project needs and a platform for online lectures during this lockdown.

**TABLE OF CONTENTS**

Contents

[Graphical Models 5](#_Toc89956799)

[Probabilistic Graphical Models 6](#_Toc89956800)

[Belief Network 7](#_Toc89956801)

[Math Behind Bayesian Networks 8](#_Toc89956802)

[**What Is Joint Probability?** 8](#_Toc89956803)

[**What Is Conditional Probability?** 8](#_Toc89956804)

[Example of a simple Bayesian network: 9](#_Toc89956805)

[Independence: 9](#_Toc89956806)

[An example of Conditional Independence: 9](#_Toc89956807)

[Role of conditional independence: 10](#_Toc89956808)

[Inference for Wet grass model 13](#_Toc89956809)

[Number of Probabilities in Bayesian Networks 14](#_Toc89956810)

[Prerequisites for D-Separation 15](#_Toc89956811)

[Conditional Independence: 15](#_Toc89956812)

[Local Structures and Independencies: 15](#_Toc89956813)

[**Common Parent** 16](#_Toc89956814)

[**Cascade** 16](#_Toc89956815)

[**V-structure** 16](#_Toc89956816)

[D-Separation 17](#_Toc89956817)

[Why D-Separation? 18](#_Toc89956818)

[What is D-Separation? 19](#_Toc89956819)

[Active/Inactive Paths 19](#_Toc89956820)

[D-Separation based on Moralization: 21](#_Toc89956821)

[Advantages and Drawbacks of Bayesian Networks 25](#_Toc89956822)

[Code 26](#_Toc89956823)

[Burglar Model 26](#_Toc89956824)

[26](#_Toc89956825)

[Wet Grass Model 28](#_Toc89956826)

[Conclusion 30](#_Toc89956827)

[References: 30](#_Toc89956828)

**INTRODUCTION**

# Graphical Models

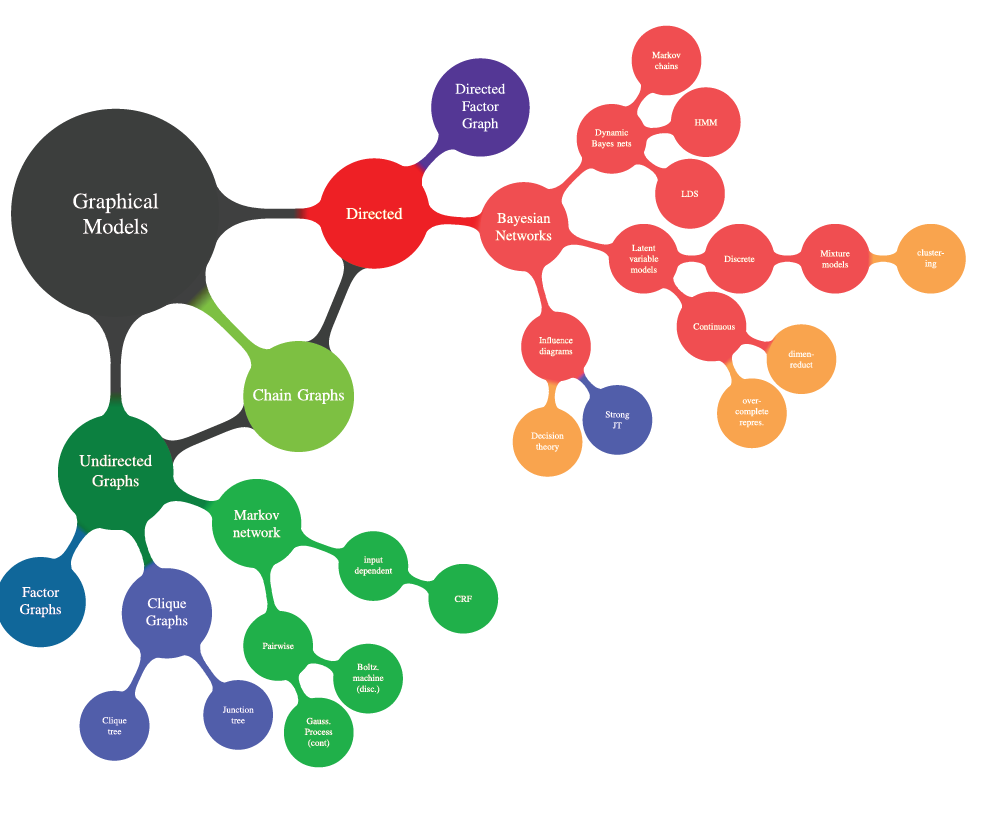
Graphical models have their origin in several areas of research. They are a union of graph theory and probability theory. They are a useful framework for representing, reasoning with and learning complex problems. The techniques are useful for multivariate (multiple variable) probabilistic systems and encompass many standard schemes, for example

• mixture models;

• factor analysis;

• hidden Markov models;

• Kalman filters



Probabilistic models can define relationships between variables and be used to calculate probabilities.

For example, fully conditional models may require an enormous amount of data to cover all possible cases, and probabilities may be intractable to calculate in practice. Simplifying assumptions such as the conditional independence of all random variables can be effective, such as in the case of Naive Bayes, although it is a drastically simplifying step.

An alternative is to develop a model that preserves known conditional dependence between random variables and conditional independence in all other cases. Bayesian networks are a probabilistic graphical model that explicitly capture the known conditional dependence with directed edges in a graph model. All missing connections define the conditional independencies in the model.

As such Bayesian Networks provide a useful tool to visualize the probabilistic model for a domain, review all of the relationships between the random variables, and reason about causal probabilities for scenarios given available evidence.

## **Probabilistic Graphical Models**

A [probabilistic graphical model](https://en.wikipedia.org/wiki/Graphical_model) (PGM), or simply “graphical model” for short, is a way of representing a probabilistic model with a graph structure.

The nodes in the graph represent random variables and the edges that connect the nodes represent the relationships between the random variables.

*A graph comprises nodes (also called vertices) connected by links (also known as edges or arcs). In a probabilistic graphical model, each node represents a random variable (or group of random variables), and the links express probabilistic relationships between these variables.*

* **Nodes**: Random variables in a graphical model.
* **Edges**: Relationships between random variables in a graphical model.

There are many different types of graphical models, although the two most commonly described are the Hidden Markov Model and the Bayesian Network.

The [Hidden Markov Model](https://en.wikipedia.org/wiki/Hidden_Markov_model) (HMM) is a graphical model where the edges of the graph are undirected, meaning the graph contains cycles. Bayesian Networks are more restrictive, where the edges of the graph are directed, meaning they can only be navigated in one direction. This means that cycles are not possible, and the structure can be more generally referred to as a directed acyclic graph (DAG).

*Directed graphs are useful for expressing causal relationships between random variables, whereas undirected graphs are better suited to expressing soft constraints between random variables.*

# Belief Network

A [Belief Network](https://en.wikipedia.org/wiki/Bayesian_network), or simply “Bayesian Network,” provides a simple way of applying Bayes Theorem to complex problems.

The networks are not exactly Bayesian by definition, although given that both the probability distributions for the random variables (nodes) and the relationships between the random variables (edges) are specified subjectively, the model can be thought to capture the “belief” about a complex domain.

Bayesian probability is the study of subjective probabilities or belief in an outcome, compared to the frequentist approach where probabilities are based purely on the past occurrence of the event.

A Bayesian belief network describes the joint probability distribution for a set of variables.

A Belief Network is a distribution of the form



where  represent the parental variables of variable . Written as Directed Graph with an arrow pointing from a parent variable to child variable, a Belief Network is a Directed Acyclic

Graph (DAG) with the  vertex in the graph corresponding to the factor 

# Math Behind Bayesian Networks

As mentioned earlier, Bayesian models are based on the simple concept of probability.

### **What Is Joint Probability?**

Joint Probability is a statistical measure of two or more events happening at the same time, i.e., P(A, B, C), The probability of event A, B and C occurring. It can be represented as the probability of the intersection two or more events occurring.

### **What Is Conditional Probability?**

Conditional Probability of an event X is the probability that the event will occur given that an event Y has already occurred.

p(X| Y) is the probability of event X occurring, given that event, Y occurs.

* If X and Y are dependent events then the expression for conditional probability is given by:  
  P (X| Y) = P (X and Y) / P (Y)
* If A and B are independent events then the expression for conditional probability is given by:  
  P(X| Y) = P (X)

## ****Example of a simple Bayesian network:****

**p(A,B,C)=p(C|A,B)p(A)p(B)**

• Probability model has simple factored form

• Directed edges => direct dependence

• Absence of an edge => conditional independence

• **Also known as belief networks, graphical models, causal networks**

• Other formulations, e.g., undirected graphical models

# **Independence:**



Intuitively, two variables are independent when knowing about one variable adds nothing to what you know about the other. For example, hair colour and gender are independent. Knowing someone's hair colour adds nothing to your knowledge of their gender.

# An example of Conditional Independence:

Height and weight are dependent, however. Knowing someone's height does not determine their weight, but you know more about their weight after you have been told their height than before. For example, knowing nothing about them, you don't know anything about whether Clark is heavier than Peter. If I tell you that Clark is 6' 2" and Peter is 5'5", now you know something, and it’s a good guess that Clark weighs more, you have even more information.

Probabilistic or statistical independence is a way to formalize these intuitions. Two random variables X and Y are probabilistically independent if the probability distribution over X is the same as the probability distribution over X **conditional** on Y, for all values y that Y can take on. That is, for all y, P(X) = P(X|Y=y)

## **Role of conditional independence**:

**Constructing a simple Belief network: wet grass**

One morning Tracey leaves her house and realises that her grass is wet. Is it due to overnight rain or did she forget to turn off the sprinkler last night? Next she notices that the grass of her neighbour, Jack, is also wet.

If you know that the grass of Jack is also wet, then it’s very likely that the wet grass was caused by the Rain. In that case, the probability of having the sprinkler left on is vanishingly small.

• This is called the **“explaining away”** effect. The wet grass of Jack’s house “explains away” to some extent the possibility that her sprinkler was left on.

**Making a model**

We can model the above situation using probability by following a general modelling approach. First, we define the variables we wish to include in our model. In the above situation, the natural variables are

R {0,1} (R = 1 means that it has been raining, and 0 otherwise).

S {0,1} (S = 1 means that Tracey has forgotten to turn off the sprinkler, and 0 otherwise).

J {0,1} (J = 1 means that Jack's grass is wet, and 0 otherwise).

T {0,1} (T = 1 means that Tracey's Grass is wet, and 0 otherwise).

A model of Tracey's world then corresponds to a probability distribution on the joint set of the variables of interest p(T,J,R, S) (the order of the variables is irrelevant).

**CASE 1:(without assuming conditional independence)**

Since each of the variables in this example can take one of two states, it would appear that we naively have to specify the values for each of the = 16 states

However, since there are normalisation conditions for probabilities, we do not need to specify all the state probabilities. To see how many states need to be specified, consider the following decomposition.

Wlog:The term is used to indicate the assumption that follows is chosen arbitrarily, narrowing the premise to a particular case, but does not affect the validity of the proof in general.

Without loss of generality (wlog) and repeatedly using Bayes' rule, we may write



That is, we may write the joint distribution as a product of conditional distributions. The first term p(T |J,R, S) requires us to specify = 8 values -> we need p(T = 1|J,R, S) for the 8 joint states of J,R, S. The other value p(T = 0|J,R, S) is given by normalisation : p(T = 0| J,R,S) = 1 - p(T = 1|J,R, S).

Similarly, we need 4 + 2 + 1 values for the other factors, making a total of 15 values in all. In general, for a distribution on n binary variables, we need to specify values in the range [0,1].

**CASE 2:(with assuming conditional independence)**

The modeler often knows constraints on the system. For example, in the scenario above, we may assume that Tracey's grass is wet depends only directly on whether or not is has been raining and whether or not her sprinkler was on. That is, we make a **conditional independence** assumption

p(T |J,R, S) = p(T|R, S)

Similarly, since whether or not Jack's grass is wet is influenced only directly by whether or not it has been raining, we write

p(J|R, S) = p(J|R)

and since the rain is not directly influenced by the sprinkler,

p(R|S) = p(R)

which means that our model now becomes :

p(T,J,R, S) = p(T|R, S)p(J|R)p(R)p(S)

We can represent these conditional independencies graphically, as in figure given below. This reduces the number of values that we need to specify to 4 + 2 + 1 + 1 = 8, a saving over the previous 15 values in the case where no conditional independencies had been assumed.

**Wet Grass Model:**

|  |  |
| --- | --- |
|  | P(S) |
| S=1 | 0.1 |
| S=2 | 0.9 |

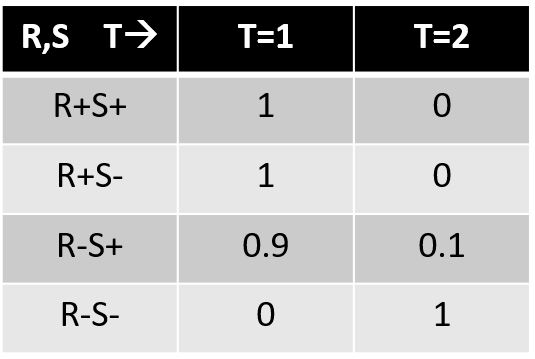
|  |  |
| --- | --- |
|  | P(R) |
| R=1 | 0.2 |
| R=2 | 0.8 |

**Rain**

**Sprinkler**

**Jack**

**Tracey**



↓

|  |  |  |
| --- | --- | --- |
| **R** ↓**J🡪** | **J=1** | **J=2** |
| R=1 | 1 | 0 |
| R=2 | 0.2 | 0.8 |

## Inference for Wet grass model

The essence of Bayesian reference is to compute the posterior probabilities of nodes given evidences.

Now that we've made a model of an environment, we can perform inference.

**🡪Let's calculate the probability that the sprinkler was on overnight, given that Tracey's grass is wet: p(S = 1|T = 1).**

To do this, we use Bayes' rule



The belief that the sprinkler is on increases above the prior probability 0.1, due to the fact that the grass is wet.

**🡪What is the probability that Tracey's sprinkler was on overnight, given that her grass is wet and that Jack's grass is also wet?**



The probability that the sprinkler is on, given the extra evidence that Jack's grass is wet, is lower than the probability The probability that the sprinkler is on, given the extra evidence that Jack's grass is wet, is lower than the probability that the sprinkler is on given only that Tracey's grass is wet.

That is, the probability of grass is wet due to the sprinkler is (partly) **explained away** (vanishingly small) due to the fact that Jack's grass is also wet -this increases the chance that the rain has played a role in making Tracey's grass wet.

# **Number of Probabilities in Bayesian Networks**

• Consider n binary variables

• Unconstrained joint distribution requires O() probabilities

• If we have a Bayesian network, with a maximum of k parents for any node, then we need O(n\*) probabilities • Example – Full unconstrained joint distribution

For example, if,

• n = 30: need probabilities for full joint distribution – Bayesian network

• n = 30, k = 4: need 480 probabilities.

# **Prerequisites for D-Separation**

## Conditional Independence:

⫫ *y* | Ƶ

denotes that the two sets of variables and *y* are independent of each other provided we know the set of variables Ƶ. For full conditional independence, and y must be independent given all states of Ƶ.

This means that | Ƶ =  Ƶ   Ƶ for all states of , *y,*  Ƶ. In case the conditioning set is empty we may also write ⫫ *y* for ⫫ *y| ∅*, in which caseis (unconditionally) independent of *y .*If and *y* are not conditionally independent, they are conditionally dependent. This is

written as ⫫ *y* | Ƶ.

## Local Structures and Independencies:

There are three types of local structures in graphical models.

In general, the variables (nodes) may be split into two groups:

**• Observed variables** are the ones we have knowledge about.

**• Unobserved variables** are ones we don’t know about and therefore have to infer the probability.

### **Common Parent**



In this structure, two nodes A and C share the same parents. Fixing B decouples A and C, that is, A and C are independent given B.

### **Cascade**

In this structure, node A has an edge to node B, which has an edge to node C. Fixing B again decouples

A and C, that is, A and C are independent given B.



### **V-structure**

In this structure, node C has two parents A and B. Knowing C would couple A and B, meaning that A and B are originally independent if we don't know C.

This can be justified by thinking of a real-world example.

Let A denote the fact that the clock in the classroom is 5 minutes fast, B represent the statement that there is a traffic jam on Highland Park Bridge, and C denote the observation that Eric is late for class. Apparently having a traffic jam is independent

of any problem with the clock. However, if we know that Eric comes to class late, knowing that there is no traffic jam means a higher probability of the clock being fast. Therefore, the two events now become dependent.

For the v-structure, the joint probability can be factorised as



If Y is observed:



but we can see there is no way there is no way to get this to

 it implies that X and Z are dependent.

But when Y is unobserved, we can show that , X and Z are independent.



# **D-Separation**

In the early 1930s, a biologist named Sewall Wright figured out a way to statistically model the causal structure of biological systems. He did so by combining directed graphs, which naturally represent causal hypotheses, and linear statistical models, which are systems of linear regression equations and statistical constraints, into a unified representation he called path analysis.

Others after him, realized that the causal structure of his models (the directed graph) determined statistical predictions we could test without doing experiments.

For example, consider a model in which blood sugar causes hunger, but only indirectly.

blood sugar → stomach acidity → hunger

The model asserts that blood sugar causes stomach acidity directly, and that stomach acidity causes hunger directly. It turns out that no matter what the strength (as long as it’s not zero) of these causal connections, which are called "parameters," the model implies that blood sugar and hunger are correlated, but that the partial correlation of blood sugar and hunger controlling for stomach acidity does vanish.

**Can we specify an algorithm that will compute, for any directed graph interpreted as a linear statistical model, all and only those independence and conditional independence relations that hold for all values of the parameters (causal strengths)?**

Judea Pearl and his colleagues in the mid-1980s realized that uncertain information could be stored much more efficiently by taking advantage of conditional independence, and they used directed acyclic graphs (graphs with no loops from a variable back to itself) to encode probabilities and the conditional independence relations among them. **D-separation was the algorithm they invented to compute all the conditional independence relations entailed by their graphs**.

Fortunately, there is a relatively simple algorithm for determining whether two variables in a Bayesian network are conditionally independent: D-separation.

D-separation can be computed in linear time using a depth-first-search-like algorithm.

## Why D-Separation?

Dependencies and Independencies

* Crucial for understanding network behaviour
* Independence properties are important for answering queries
* Exploited to reduce computation of inference
* A distribution P that factorizes over G satisfies I(G)

## What is D-Separation?

The motivation for d-separation. The "d" in d-separation and d-connection stands for dependence. Thus, if two variables are d-separated relative to a set of variables Z in a directed graph, then they are independent conditional on Z in all probability distributions such a graph can represent. Roughly, two variables X and Y are independent conditional on Z if knowledge about X gives you no extra information about Y once you have knowledge of Z. In other words, once you know Z, X adds nothing to what you know about Y.

Intuitively, a path is *active* if it carries information, or dependence. Two variables X and Y might be connected by lots of paths in a graph, where all, some, or none of the paths are active. X and Y are d-connected, however, if there is *any* active path between them. So, X and Y are d-separated if *all* the paths that connect them are *inactive*, or, equivalently, if no path between them is active.

**OUTLINE:**

* Study independence properties for subgraphs (connected triples)
* Analyse complex cases in terms of triples along paths between variables.

## Active/Inactive Paths

Question: Are X and Y conditionally independent given evidence variables {Z}?

* Yes, if X and Y “d-separated” by Z
* Consider all (undirected) paths from X to Y
* Otherwise (i.e. if all paths are inactive), then “D-separated” = independence is guaranteed
* If one or more paths is active, then independence not guaranteed

🡪A path is active if every triple in path is active:

* Causal chain A 🡪 B 🡪 C where B is unobserved (either direction)
* Common cause A🡨B 🡪 C where B is unobserved
* Common effect (aka v-structure) A 🡪 B 🡨 C where B or one of its descendants is observed
* If one of the paths is active, then not guaranteed to be independent

🡪 A path is inactive if just one of the triples is inactive:

* Causal chain A 🡪 B 🡪 C where B is observed (either direction)
* Common cause A🡨B 🡪 C where B is observed
* Common effect (aka v-structure) A 🡪 B 🡨 C where B or one of its descendants is unobserved
* if all paths are inactive, then guaranteed to be independent.

🡪 For conditional independence to be guaranteed, all paths connecting the two variables need to be inactive.

Active Triples Inactive Triples

Example: Check if U and V are conditionally independent given X.

**U V| X**

There are two paths possible:

U W V: It is a V structure and the middle node is unobserved and its descendant is also unobserved, hence the path is inactive. Once, it is inactive, we will check for other possible paths.

U W Y X V: Here, there are several triples, first one is [U,W,Y] which takes a causal chain structure with the middle node unobserved, leaving the triple active. If one of the triple in the path is active, we have to check for other triples too. Next triple is [W,Y,X] which takes a V structure with the middle node unobserved leaving the triple inactive. If one of the triple in a path is inactive, then the whole path becomes inactive.

Now, both the paths are inactive. Hence, we can say U and V are conditionally independent given X.

# **D-Separation based on Moralization:**

In [graph theory](https://en.wikipedia.org/wiki/Graph_theory), a **moral graph** is used to find the equivalent undirected form of a [directed acyclic graph](https://en.wikipedia.org/wiki/Directed_acyclic_graph). It is a key step of the [junction tree algorithm](https://en.wikipedia.org/wiki/Junction_tree_algorithm), used in [belief propagation](https://en.wikipedia.org/wiki/Belief_propagation) on [graphical models](https://en.wikipedia.org/wiki/Graphical_models).

The Bayes net assumption says:

**“Each variable is conditionally independent of its non-descendants, given its parents.”**

It’s certainly possible to reason about independence using this statement, but we can use

D-separation as a more formal procedure for determining independence. We start with an independence question in one of these forms:

“Are X and Y conditionally independent, given Z”

“Are X and Y marginally independent?”

For instance, if we’re asked to figure out “Is P(A|BDF) = P(A|DF)?”, we can convert it into an independence question like this: “Are A and B independent, given D and F?”

Then we follow this procedure:

1. Draw the ancestral graph. Construct the “ancestral graph” of all variables mentioned in the probability expression. This is a reduced version of the original net, consisting only of the variables mentioned and all of their ancestors (parents, parents’ parents, etc.)

2. “Moralize” the ancestral graph by “marrying” the parents. For each pair of variables with a common child, draw an undirected edge (line) between them. (If a variable has more than two parents, draw lines between every pair of parents.)

3. "Disorient" the graph by replacing the directed edges (arrows) with undirected edges (lines).

4. Delete the givens and their edges. If the independence question had any given variables, erase those variables from the graph and erase all of their connections, too. Note that “given variables” as used here refers to the question “Are A and B conditionally independent, given D and F?”, not the equation “P(A|BDF) =? P(A|DF)”, and thus does not include B.

5. Read the answer off the graph.

• If the variables are disconnected in this graph, they are guaranteed to be independent.

• If the variables are connected in this graph, they are not guaranteed to be independent. Note that “are connected” means “have a path between them”.

• If one or both of the variables are missing (because they were givens, and were therefore deleted), they are independent.

Practicing with the d-separation algorithm will eventually let us determine independence relations more intuitively. For example, we can tell at a glance that two variables with no common ancestors are marginally independent, but that they become dependent when given their common child node.

Example:

**Are A and B conditionally independent, given D and F?**

Draw Ancestral Graph Moralize Disorient

Delete Givens

We can infer from the last step that A and B are connected, so they are not required to be conditionally independent given D and F.

**Are D and E conditionally independent, given C?**

Draw Ancestral Graph Moralize Disorient

Delete Givens

We can infer from the last step that D and E are not connected, so they are conditionally independent given C.

# **Advantages and Drawbacks of Bayesian Networks**

**Advantages**

* Bayesian Networks offer a graphical representation that is reasonably interpretable and easily explainable;
* Relationships captured between variables in a Bayesian Network are more complex yet hopefully more informative than a conventional model;
* Models can reflect both statistically significant information (learned from the data) and domain expertise simultaneously;
* Multiple metrics can used to measure the significance of relationships and help identify the effect of specific actions;
* Offer a mechanism of suggesting counterfactual actions and combine

actions without aggressive independence assumptions.

**Drawbacks**

* Granularity of modelling may have to be lower. However, this may either not be necessary, or can be run in tangent to other techniques that provide accuracy but are less interpretable;
* Computational complexity is higher. However, this can be offset with careful feature selection and a less granular discretisation policy, but at the expense of predictive power;
* This is (unfortunately) not a way of fully automating Causal Inference.

# **Code**

## Burglar Model

## 

BURGLARY

EARTHQUAKE

MARY CALLS

JOHN CALLS

ALARM





clc;

clear all;

close all;

burglary=[0.001 0.999];

earthquake=[0.002 0.998];

alarm=[0.95 0.05;0.94 0.06;0.29 0.71;0.001 0.999];

john=[0.9 0.1;0.05 0.95];

mary=[0.7 0.3;0.01 0.99];

[P,M]=FPD(burglary,earthquake,alarm,john,mary)

sum(P)

function [P,M] = FPD(burglary,earthquake,alarm,john,mary)

% a = size(burglary);

% b = size(earthquake);

% c = size(alarm);

% d = size(john);

% e = size(mary);

P = zeros(32,1);

A=fullfact([2 2 2 2 2]);

M=A(:,(end:-1:1));

size(M)

id=[1 2;3 4];

for k=1:length(P)

burg=M(k,1);

earq=M(k,2);

ala=M(k,3);

jc=M(k,4);

mc=M(k,5);

%P(b,e,a,j,m)=P(b)P(e)P(a|b,e)P(j|a)P(m|a)

tempprob=burglary(burg)\*earthquake(earq);

idind=id(burg,earq);

tempprob=tempprob\*alarm(idind,ala); %comb of burg and eq in row,alarm in col

tempprob=tempprob\*john(ala,jc); %alarm in row, john in col

tempprob=tempprob\*mary(ala,mc); %alarm in row, mary in col

P(k)=tempprob;

end

end

OUTPUT:

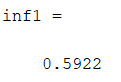
Inference 1:Probability that John and Mary called given Burglary occurred.

% P(J=1,M=1/B=1)= P(J=1,M=1,B=1)/P(B=1)

kk = (M(:,1)==1);

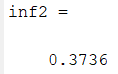
den = sum(P(kk));

kk = and(and((M(:,4)==1),(M(:,5)==1)),(M(:,1)==1));

****num = sum(P(kk));

inf1 = num/den

Inference 2:Probability that Earthquake occurred given Alarm is on.

%P(E=1|A=1)

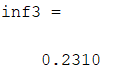
kk = (M(:,3)==1);

den = sum(P(kk));

kk = and((M(:,2)==1),(M(:,3)==1));

num = sum(P(kk));

inf2 = num/den %prob of earthquake given alarm= 0.2310

****Inference 3:Probability that Burglary occurred given Alarm went on.

%P(B=1|A=1)=P(B=1,A=1)/P(A=1)

kk = (M(:,3)==1);

den = sum(P(kk));

kk = and((M(:,1)==1),(M(:,3)==1));

num = sum(P(kk));

inf3 = num/den %prob of burglar given alarm = 0.3736

%We can conclude from the inference that the possibility for the alarm went off is due to the burglary

## Wet Grass Model

clc;

clear all;

close all;

rain=[0.2;0.8];

sprinkler=[0.1;0.9];

jack=[1 0;0.2 0.8];

tracey=[1 0;1 0;0.9 0.1;0 1];

[P,M]=FPD(rain,sprinkler,jack,tracey)

sum(P)

function [P,M] = FPD(rain,sprinkler,jack,tracey)

P = zeros(16,1);

A=fullfact([2 2 2 2]);

M=A(:,(end:-1:1));

size(M)

id=[1 2;3 4];

for k=1:length(P)

r=M(k,1);

s=M(k,2);

j=M(k,3);

t=M(k,4);

%P(r,s,j,t)=P(r)P(s)P(j|r)P(t|r,s)

tempprob=rain(r)\*sprinkler(s);

tempprob=tempprob\*jack(r,j);

idind=id(r,s); %mapping

tempprob=tempprob\*tracey(idind,t);

P(k)=tempprob;

end

end

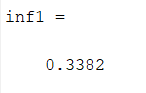
OUTPUT:

Inference 1: Probability that Sprinkler is on given Tracey’s grass is wet.

% P(S=1/T=1)= P(S=1,T=1)/P(T=1) Probability of Sprinkler on given that

% Tracey grass is wet

kk = (M(:,4)==1);

den = sum(P(kk));

kk = and((M(:,2)==1),(M(:,4)==1));

num = sum(P(kk));

inf1 = num/den %ans=0.3382

Inference 2: Probability that Sprinkler is on given both Tracey’s and Jack’s grass is wet.

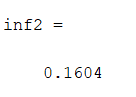
% P(S=1/T=1,J=1)= P(S=1,T=1,J=1)/P(T=1,J=1)

Probability of Sprinkler on given that

% both Tracey and Jack grass is wet

kk = and((M(:,4)==1),(M(:,3)==1));

den = sum(P(kk));

kk = and((M(:,2)==1),and((M(:,4)==1),(M(:,3)==1)));

num = sum(P(kk));

inf2 = num/den

%ans=0.1604 the fact that the sprinkler was on is explained away by including Jack grass is wet,

%which means the prob that the grass is wet due to rain is high.

# **Conclusion**

Bayesian belief networks is a class of highly data efficient and interpretable models for domains with causal relationships between variables. The trade-off is a dependency on good prior knowledge and often problem-specific adaptions and simplifications. For the problems where their strengths shine however, belief networks are well worth their trouble.

This can be applied in a wide range of areas in health services research (health economic evaluation, health quality measurement, health outcomes monitoring, cost-effectiveness analysis), but also in epidemiology, clinical research, medical decision making or public health**.**

# **References**:

[1] David Barber, Bayesian Reasoning and Machine Learning, Cambridge University Press, 2012.

[2] <http://web.mit.edu/jmn/www/6.034/d-separation.pdf>

[3] <https://cgi.csc.liv.ac.uk/~xiaowei/ai_materials/26-PGM-D-separation.pdf>

[4] <https://cs.uwaterloo.ca/~zsheikhb/slides/cs480-Lecture8-Winter2020.pdf>

[5]<https://www.researchgate.net/publication/282685628_Overview_of_Bayesian_Network>

[6] <https://cs.uwaterloo.ca/~zsheikhb/slides/cs480-Lecture8-Winter2020.pdf>