Uncertain knowledge and Reasoning

Unit 3

Syllabus Content

- Acting under Uncertainty, Basic Probability Notation, Inference Using Full Joint Distributions, Independence, Bayes' Rule and Its Use
- Representing Knowledge in an Uncertain Domain, The Semantics of Bayesian Networks, Efficient Representation of Conditional Distributions, Exact Inference in Bayesian Networks.
- Approximate inference in Bayesian Networks, Other Approaches to Uncertain reasoning-Fuzzy sets and Fuzzy logic.

Uncertainty

- In real life, it is not always possible to determine the state of the environment as it might not be clear. Due to partially observable or non-deterministic environments, agents may need to handle uncertainty
- When an agent knows enough facts about its environment. The logical

  plans and actions produces a guaranteed Work
- Unfortunately, agent never have access to the whole environment. Agents acts uncertainty.

Uncertainty

Uncertainty due to

- Uncertain data: Data that is missing, unreliable, inconsistent or noisy
- Uncertain knowledge: When the available knowledge has multiple causes leading to multiple effects or incomplete knowledge of causality in the domain
- Uncertain knowledge representation: The representations which provides a restricted model of the real system, or has limited expressiveness
- Inference: In case of incomplete or default reasoning methods, conclusions drawn might not be completely accurate.

CAR

Nature of Uncertain Knowledge

- The Diagnosis : medicine, automobile repair, or whatever-is a task that almost always involves uncertainty
- Let us try to write rules for dental diagnosis using first-order logic, so that we can see how the logical approach breaks down. Consider the following rule:
  - *∀p Symptom(p, Toothache) ⇒ Disease(p, Cavity) .*
- The problem is that this rule is wrong.
- Not all patients with toothaches have cavities; some of them have gum disease, swelling, or one of several other problems
  - *∀p Symptom(p, Toothache) ⇒ Disease(p, Cavity) V Disease(p, GumDisease) V Disease(p, Swelling) . . .*

Nature of Uncertain Knowledge

- to make the rule true, we have to add almost unlimited list of possible causes.
- We could try a causal rule:
  - $\forall p$ *Disease(p, Cavity)* $\Rightarrow$ *Symptom(p, Toothache).*
- But this rule is also not right either; not all cavities cause pain
- Toothache and a Cavity are unconnected, so the judgement may go wrong.

Nature of Uncertain Knowledge

- This is a type of the medical domain, as well as most other judgmental domains: law, business, design, automobile repair, gardening, dating, and so on.
- The agent take action, only a **degree of belief** in the relevant sentences.
- Our main tool for dealing with degrees of belief will be **probability theory**
- **The Probability** assigns to each sentence a numerical degree of belief between 0 and 1.

Sample Space

- In probability theory, the sample space is also called as a sample description space or possibility space.
- The sample space (Ω) is the set of possible outcomes of an experiment or random trial. Points ω in Ω are called sample outcomes, realizations, or elements. Subsets of Ω are called Events.

- **Example**. If we toss a coin twice then
  - Sample space (Ω) = {HH,HT, TH, TT}.
  - **The event that the first toss is heads is A** = {HH,HT}

- **Complement** - The event that A does not occur, denoted as **A´** , is called the complement of event A. **P(A´) = 1- P(A)**

- **Intersection** - The intersection of two events A and B, denoted by **A ∩ B**, is the event containing all elements that are common to A and B.

Sample Space

- **Union -** The union of the two events A and B, denoted by **A∪B,** is the event containing all the elements that belong to A or B or both.

- **Mutually Exclusive -** Events that have no outcomes in common are said to be ***disjoint or mutually exclusive.***

➢ Clearly, A and B are mutually exclusive or disjoint if and only if $A \cap B$ is a null set

➢ We say that events A1 and A2 are disjoint (mutually exclusive) if $Ai \cap Aj = \{\}$

  - Example: first flip being heads and first flip being tails

Sample Space

- **Collectively exhaustive -** A set of events is jointly or collectively exhaustive **if at least one of the events must occur**.

•The outcomes must be **collectively exhaustive**, i.e. on every experiment (or random trial) there will always take place some outcome.

- E1,E2 ,E3, …En are called **exhaustive events** if at least one of them necessarily occurs whenever the experiment is performed.

Sample Space

- **Sample space S = {1, 2, 3, 4, 5, 6}**
- Assume that A, B and C are the events associated with this experiment. Also, let us define these events as:

❑A be the event of getting a number greater than 3

❑B be the event of getting a number greater than 2 but less than 5

❑C be the event of getting a number less than 3

- We can write these events as:
- A = {4, 5, 6}
- **B = {3, 4}**
- **and C = {1, 2}**
- We observe that
- A U B U C = {4, 5, 6} U {3, 4} U {1, 2} = {1, 2, 3, 4, 5, 6} = S
- **Therefore, A, B, and C are called exhaustive events.**

Probability

- A **probability** is a number that reflects the chance or likelihood that a particular event will occur.
- **Probabilities** can be expressed as proportions that range from 0 to 1, and they can also be expressed as percentages ranging from 0% to 100%.
- A probability of 0 indicates that there is no chance that a particular event will occur, whereas a probability of 1 indicates that an event is certain to occur.
- A probability of 0.45 (45%) indicates that there are 45 chances out of 100 of the event occurring.

Axioms of Probability

- We will assign a real number P(A) to every event A, called the probability of A.
- To qualify as a probability, P must satisfy three axioms:

Axiom 1: For any event A, $P(A) \geq 0$.

Axiom 2: Probability of the sample space S is $P(S) = 1$.

Axiom 3: If $A_1, A_2, A_3, \cdots$ are disjoint events, then $P(A_1 \cup A_2 \cup A_3 \cdots) = P(A_1) + P(A_2) + P(A_3) + \cdots$

Axioms of Probability

Let us take a few moments and make sure we understand each axiom thoroughly. The first axiom states that probability cannot be negative. The smallest value for $P(A)$ is zero. If $P(A) = 0$, then the event A may be considered impossible, for practical purposes. The second axiom states that the probability of the whole sample space is equal to one, i.e., 100 percent. The reason for this is that the sample space S contains all possible outcomes of our random experiment. Thus, the outcome of each trial always belongs to S, i.e., the event S always occurs and $P(S) = 1$. In the example of rolling a die, $S = \{1, 2, 3, 4, 5, 6\}$, and since the outcome is always among the numbers 1 through 6, $P(S) = 1$.

Axioms of Probability

- The third axiom is probably the most interesting one. The basic idea is that if some events are disjoint (i.e., there is no overlap between them), then the probability of their union must be the summations of their probabilities.

Example 1.8

In a presidential election, there are four candidates. Call them A, B, C, and D. Based on our polling analysis, we estimate that A has a 20 percent chance of winning the election, while B has a 40 percent chance of winning. What is the probability that A **or** B win the election?

Axioms of Probability

In summary, if A_1 and A_2 are disjoint events, then $P(A_1 \cup A_2) = P(A_1) + P(A_2)$. The same argument is true when you have n disjoint events $A_1, A_2, \cdots, A_n$:

$$P(A_1 \cup A_2 \cup A_3 \cdots \cup A_n) = P(A_1) + P(A_2) + \cdots + P(A_n), \text{ if } A_1, A_2, \cdots, A_n \text{ are disjoint.}$$

In fact, the third axiom goes beyond that and states that the same is true even for a countably infinite number of disjoint events. We will see more examples of how we use the third axiom shortly.

Joint and Conditional Probabilities

- **Joint Probability:**
  - Joint probability is a statistical measure that calculates the likelihood of two events occurring together and at the same point in time.
  - Joint probability is the probability of event Y occurring at the same time that event X occurs.
- X, Y – Two different event that intersect
- P(X,Y)
- Probability of X and Y

The Formula for Joint Probability:

Notation for joint probability can take a few different forms. The following formula represents the probability of events intersection:

$$P\left(X \bigcap Y\right)$$

where:

$X, Y =$ Two different events that intersect

$P(X \text{ and } Y), P(XY) =$ The joint probability of X and Y

Conditional Probabilities

- Assuming that P(B) > 0, the **conditional** probability of A given B:

  $P(A|B)=P(AB)/P(B)$ (Conditioning restricts the sample space to those outcomes which are in the set being conditioned on)

- Two events A and B are independent if conditioning on one does not effect the probability of the other.

- Two events A and B are called independent if $P(A|B)=P(A)$

  Two events A and B are independent if and only if $P(A \cap B)=P(A)P(B)$. $=P(A)$.

  Thus, if two events A and B are independent and $P(B) \neq 0$, then $P(A|B)=P(A)$.

Unconditional Probability (marginal probability)

- Refers to a probability that is unaffected by previous or future events
- It is the probability of a single event without consideration of any other event
- Called simple probability.
- If we select a child at random (by simple random sampling), then each child has the same probability (equal chance) of being selected, and the probability is 1/N, where N=the population size.
- Thus, the probability that any child is selected is 1/5,290 = 0.0002.
- In most sampling situations we are generally not concerned with sampling a specific individual but instead we concern ourselves with the probability of sampling certain types of individuals.
- For example, what is the probability of selecting a boy or a child 7 years of age? The following formula can be used to compute probabilities of selecting individuals with specific attributes or characteristics.

P(characteristic) = persons with characteristic / N

	Age (years)						
	5	6	7	8	9	10	Total
Boys	432	379	501	410	420	418	2,560
Girls	408	513	412	436	461	500	2,730
Totals	840	892	913	846	881	918	5,290

- **What is the probability of selecting a boy**

If we select a child at random, the probability that we select a boy is computed as follows

P(boy) = 2,560/5,290 = 0.484 or 48.4%.

It is an unconditional probability, because the denominator for each is the total population size (N=5,290) reflecting the fact that everyone in the entire population is eligible to be selected.

Probability

- Probability can be defined as a chance that an uncertain event will occur. It is the numerical measure of the likelihood that an event will occur. The value of probability always remains between 0 and 1 that represent ideal uncertainties.

$0 \leq P(A) \leq 1$, where $P(A)$ is the probability of an event A.

$P(A) = 0$, indicates total uncertainty in an event A.

$P(A) = 1$, indicates total certainty in an event A.

Probability (Cont.)

We can find the probability of an uncertain event by using the below formula.

$$\textbf{Probability of occurrence} = \frac{\text{Number of desired outcomes}}{\text{Total number of outcomes}}$$

- $P(\neg A)$ = probability of a not happening event.
- $P(\neg A) + P(A) = 1$.

Basic probability notation

1. Propositions

- Complex proposition can be formed using standard logical connectives.
- For example:
  1. [(cavity=true) ^ (toothache = false)]
  2. [(cavity ^ ~toothache)]
- **Random variables:**
  - Random variables are used to represent the events and objects in the real world.
  - Random variables are like symbols in propositional logic.
  - For example:
    - P(a)= 1- P(~a)

Basic probability notation
2. Atomic event

- An atomic event is a complete specification of the state of the world about which agent us uncertain.
- Example:
  - If the world consists of cavity and toothache the there are four distinct atomic events,

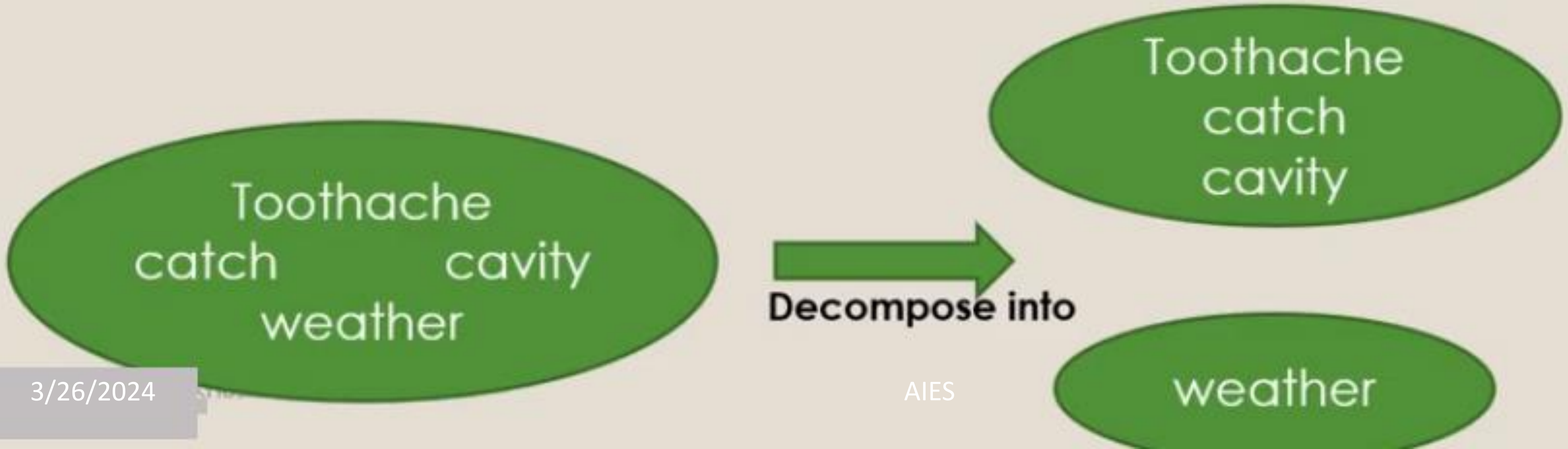
1. Cavity= false ^ toothache = True
2. Cavity= false ^ toothache = false
3. Cavity= true ^ toothache = false
4. Cavity= true ^ toothache = true

Basic probability notation
3. Unconditional probability

- It is the degree of belief accorded to a proposition in the absence of any other information.
- Written as a P(a)
- Example
  - Ram has cavity
  - P(cavity=true)=0.1 **OR** P(cavity)=0.1
  - When we want to express probabilities of all possible values of a random variable, then vector of value is used.
    - P(WEATHER)= <0.7,0.2, 0.08,0.02>
    - P(WEATHER=sunny)=0.7
    - P(WEATHER=rain)=0.2

Basic probability notation
4. Independence

- It Is relation between 2 different set of full joint distributions. It is also called as marginal or absolute independence of the variable.
- Independence indicates that whether the 2 full joint distributions affects probability each other.
- The weather is independent of once dental problem.
- **P**(toothache, catch, cavity, weather)= **P**(toothache, catch, cavity) **P**(weather)



Basic probability notation

5. Conditional Probability

- Conditional probability is a probability of occurring an event when another event has already happened.
- Let's suppose, we want to calculate the event A when event B has already occurred, "the probability of A under the conditions of B", it can be written as:

$$P(A|B) = \frac{P(A \wedge B)}{P(B)}$$

- **Where P(A ^ B)= Joint probability of a and B**
- **P(B)= Marginal probability of B.**

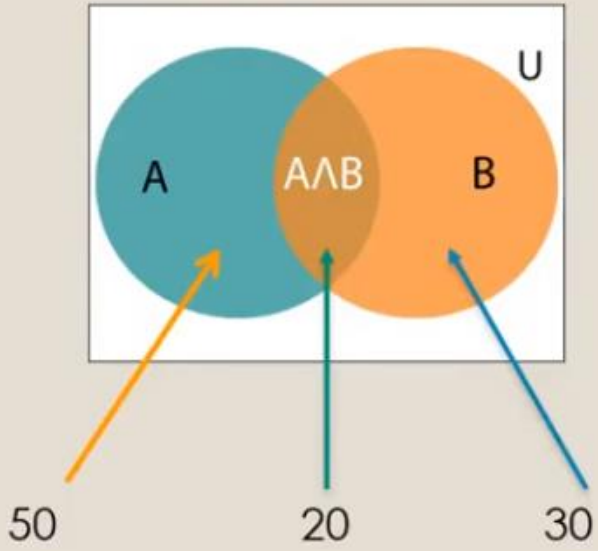
Basic probability notation

- If the probability of A is given and we need to find the probability of B, then it will be given as:

$$P(B|A) = \frac{P(A \wedge B)}{P(A)}$$

Basic probability notation

- $P(A|B) = {}^{P(A \wedge B)}/_{P(B)}$
- P(B) =30/100 = 0.3
- P(A ^ B) = 20/100 = 0.2
- P(A | B)=0.2/0.3 = 0.67



Inference using full joint distribution

- **Probabilistic inference:** The computation of posterior probabilities for query propositions given observed evidence.
- The **full joint probability distribution** specifies the probability of each complete assignment of values to random variables.
  - It is usually too large to create or use in its explicit form,
  - However, when it is available it can be used to answer queries simply by adding up entries for the possible worlds corresponding to the query propositions.

Inference using full joint distribution

- **Marginalization / summing out:** The process of extracting the distribution over some subset of variables or a single variable (to get the **marginal probability**), by summing up the probabilities for each possible value of the other variables, thereby taking them out of the equation.

Basic probability notation

6. Inference using Full joint Distribution

- Probability inference means, computation from observed evidence of posterior probabilities, for query propositions. The knowledge based answering the query is represented as full joint distribution.

	Toothache		~Toothache	
	Catch	~Catch	Catch	~Catch
Cavity	0.108	0.012	0.072	0.008
~Cavity	0.016	0.064	0.144	0.576

Basic probability notation

6. Inference using Full joint Distribution

- One particular common task in inferencing is to extract the distribution over some subset of variables or a single variable. This distribution over some variables or single variable is called as **marginal probability (Marginalization/ Summing).**

- **P(Cavity)** = 0.108+ 0.012 + 0.072 + 0.00 8

  = 0.2

Basic probability notation

6. Inference using Full joint Distribution

- Computing probability of a cavity, given evidence of a toothache is as follow:
- $$\mathbf{P(Cavity \mid Toothache)} = {}^{P(Cavity \wedge Toothache)}/_{P\ (Toothache)}$$

$$= {}^{0.108+\ 0.012}/_{0.108+\ 0.012+0.016\ +\ 0.064}$$

$$\boxed{= 0.6}$$

- Just to check also compute the probability that there is no cavity goven toothache is as follow:
- $$\mathbf{P(\sim Cavity \mid Toothache)} = {}^{P(\sim Cavity \wedge Toothache)}/_{P\ (Toothache)}$$

$$= {}^{0.016\ +\ 0.064}/_{0.108+\ 0.012+0.016\ +\ 0.064}$$

$$\boxed{= 0.4}$$

Bayes Theorem

- **Bayes' rule, Bayes' law**, or **Bayesian reasoning**, which determines the probability of an event with uncertain knowledge.
- it relates the **conditional probability** and **marginal probabilities** of two random events.
- Bayes' theorem named after the British mathematician **Thomas Bayes**.
- It is a way to calculate the value of **P(B|A)** with the knowledge of **P(A|B).**
- Allows **updating the probability prediction of an event** by observing new information of the real world.

Bayes Theorem

- Bayes' theorem can be derived using **product rule** and **conditional probability** of event A with known event B:

$P(A \wedge B) = P(A|B)\, P(B)$

or

$P(B \wedge A) = P(B|A)\, P(A)$

$$P(A|B) = \frac{P(B|A)\, P(A)}{P(B)} \quad \text{....(a)}$$

The above equation (a) is called as **Bayes' rule** or **Bayes' theorem**.

Bayes Theorem

- This equation is basic of most modern AI systems for **probabilistic inference**.
- It shows the simple relationship between **joint and conditional probabilities**.
- P(A|B) is known as **posterior,** Probability of hypothesis A when we have occurred an evidence B.
- P(B|A) is called the **likelihood**, in which we consider that hypothesis is true, then we calculate the probability of evidence.
- P(A) is called the **prior probability,** probability of hypothesis before considering the evidence
- P(B) is called **marginal probability,** pure probability of an evidence.

Bayes Theorem Applications

- It is used to calculate the next step of the robot when the already executed step is given.
- Bayes' theorem is helpful in weather forecasting.

Bayes Theorem

- A doctor is aware that disease meningitis causes a patient to have a stiff neck, and it occurs 80% of the time. He is also aware of some more facts, which are given as follows:

➢ The Known probability that a patient has meningitis disease is 1/30,000.

➢ The Known probability that a patient has a stiff neck is 2%.

Find the probability that a patient with the stiff neck has meningitis.

Bayes Theorem

Let a be the proposition that patient has stiff neck and b be the proposition that patient has meningitis.

- $P(a|b) = 0.8$
- $P(b) = 1/30000$
- $P(a) = .02$

$$P(b|a) = \frac{P(a|b)P(b)}{P(a)} = \frac{0.8 * (\frac{1}{30000})}{0.02} = 0.001333333.$$

Joint probability distribution

- The **full joint probability distribution** specifies the probability of values to random variables.
- It is usually too large to create or use in its explicit form.
- Joint probability distribution of two variables X and Y are

Joint probabilities	X	X’
Y	0.20	0.12
Y’	0.65	0.03

- Joint probability distribution for n variables require 2^n entries with all possible combination.

Drawbacks of joint Probability Distribution

- Large number of variables and grows rapidly
- Time and space complexity are huge
- Statistical estimation with probability is difficult
- Human tends signal out few propositions
- The alternative to this is Bayesian Networks.

Introduction to Bayesian Network:

- It is also called a Bayes network, belief network, decision network, or Bayesian model.
- Dealing with probabilistic events and to solve a problem which has uncertainty
- A Bayesian network is a probabilistic graphical model which represents a set of variables and their conditional dependencies using a directed acyclic graph
- This is another method of reducing the complexity of certainty factors. It is a directed graph in which each node is annotated with quantitative probability information.

Introduction to BN Contd..

- **The full specification of Bayesian Network is as follows:**

1. A set of random variables makes up the nodes of the network. Variable may be discrete or continuous.
2. A set of directed links or arrows connects pairs of nodes. If there is an arrow from node x to node y, x is said to be parent of y.
3. Each node X_i has a conditional probability distribution

   $P(X_i \mid \text{parents } (X))$ which quantifies the effect of parents on the node.

1. The graph has no directed cycles

Introduction to BN Contd..

- The **topology of the network**—the set of nodes and links, specifies the conditional independence relationships which hold in the domain.
- The intuitive meaning of an arrow in a properly constructed network is usually that X has a direct influence on Y.

Bayesian networks

- A simple, graphical notation for conditional independence assertions and hence for compact specification of full joint distributions

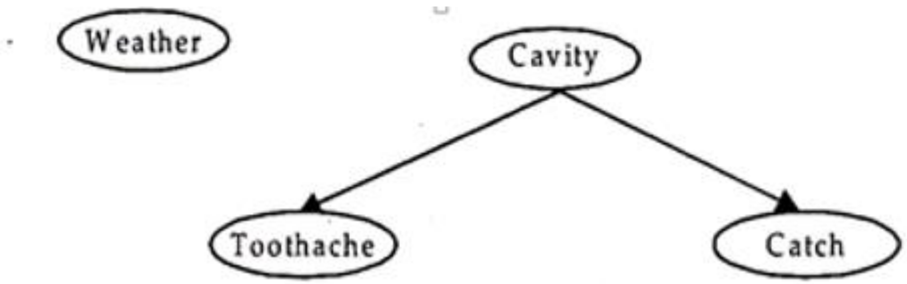
- Syntax:
  - a set of nodes, one per variable
  - a directed, acyclic graph (link ≈ "directly influences")
  - a conditional distribution for each node given its parents:

$$\mathbf{P}(X_i \mid \text{Parents}(X_i))$$

- In the simplest case, conditional distribution represented as a conditional probability table (CPT) giving the distribution over X_i for each combination of parent values

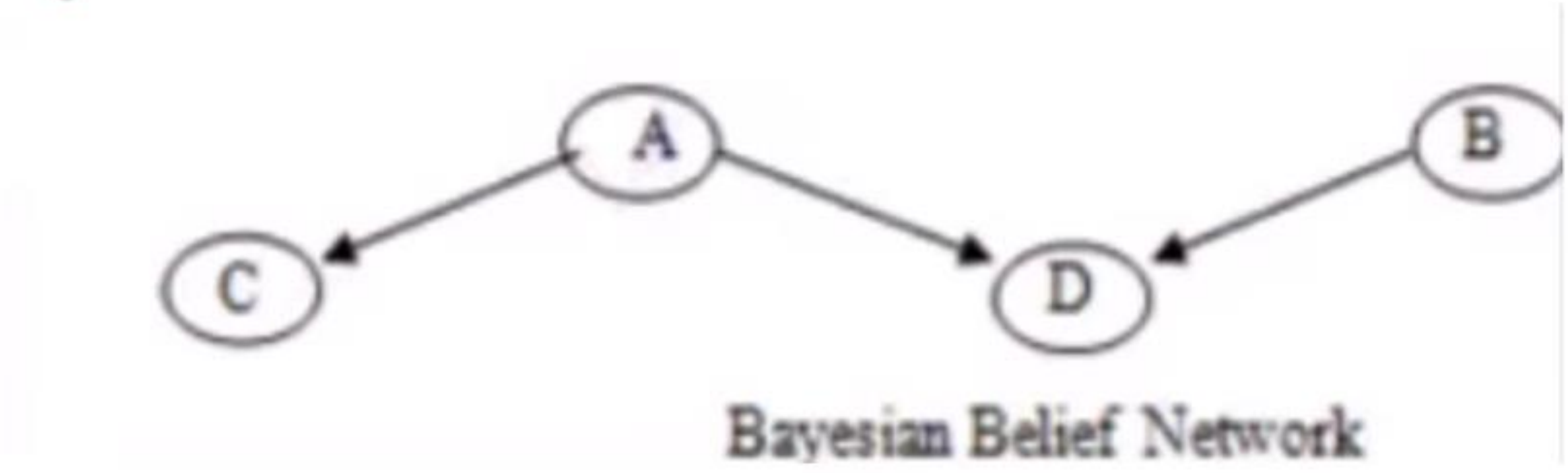
Example

- Take the world consisting of simple variables Toothache, Cavity, Catch and Weather. Assuming that weather has no influence on toothache, weather is independent of the other variables; toothache and catch are conditionally independent, given a cavity.
- Topology of network encodes conditional independence assertions:



Bayesian Network: Weather variable is independent from other three variables. Toothache and catch are conditional independent given cavity.

Example



- A & B are unconditional, independent, evidence and parent nodes
- C & D are conditional, dependent, hypothesis and child nodes.

Conditional Probability Table

P(A)	=	0.3
P(B)	=	0.6
P(C|A)	=	0.4
P(C|~A)	=	0.2
P(D|A, B)	=	0.7
P(D|A, ~B)	=	0.4
P(D|~A, B)	=	0.2
P(D|~A, ~B)	=	0.01

Conditional Probability Tables

P(A)	P(B)
0.3	0.6

A	P(C)
T	0.4
F	0.2

A	B	P(D)
T	T	0.7
T	F	0.4
F	T	0.2
F	F	0.01

- $P(A,B,C,D) = P(D|A,B)*P(C|A)*P(B)*P(A)$

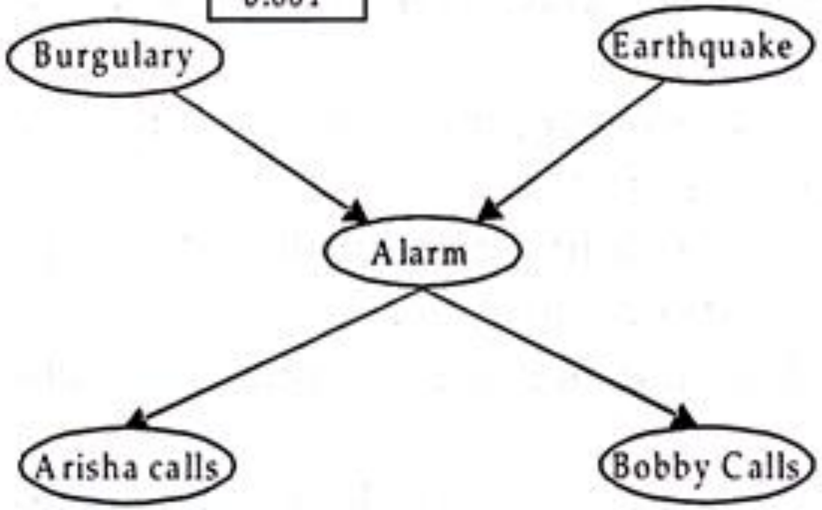
Bayesian Network - Burglar Alarm

- You have installed a new burglar alarm at home.
- It is fairly reliable at detecting a burglary, but also responds on occasion to minor earthquakes.
- You also have two neighbors, John and Mary, who have promised to call you at work when they hear the alarm.
- John always calls when he hears the alarm, but sometimes confuses the telephone ringing with the alarm and calls then, too.
- Mary, on the other hand, likes loud music and sometimes misses the alarm altogether.
- Given the evidence of who has or has not called, we would like to estimate the probability of a burglary.

Example contd.

P (B)
0.001

P (E)	
0.002	t
0.998	f



B	E	P (A)
t	t	0.95
t	f	0.94
t	t	0.29
f	f	0.001

A	P (M)
t	0.70
f	0.01

A	P(J)
t	0.90
f	0.05

- In burglary network, the topology shows that
  - burglary and earthquakes directly affect the probability of the alarm,
  - but John and Mary call depends on the alarm.
- Our assumptions from the network,
  - they do not perceive any burglaries directly,
  - they do not notice the minor earthquakes, and
  - they do not discuss before calling.

- Notice that the network does not have nodes corresponding to
  - Mary is currently listening to loud music or
  - The is telephone ringing and confusing John.
- These factors are summarized in the uncertainty, associated with the links from *Alarm* to *JohnCalls* and *MaryCalls.*
- This shows both laziness and ignorance in operation.

Conditional Probability Tables - CPTs

- The *conditional probability tables* in the network give the probabilities for the values of the random variable depending on the combination of values for the parent nodes.
- Each row must sum to 1
- All variables are Boolean, and therefore, the probability of a true value is p, the probability of false must be 1–p
- A table for a Boolean variable with k parents contains 2^k independently specifiable probabilities
- A variable with no parents has only one row, representing the prior probabilities of each possible value of the variable

Example

- We can calculate the probability that the alarm has sounded, but neither a burglary nor an earthquake has occurred, and both John and Mary call.

Example

- We can calculate the probability that the alarm has sounded, but neither a burglary nor an earthquake has occurred, and both John and Mary call.

$$P(j \wedge m \wedge a \wedge \neg b \wedge \neg e)$$
$$= P(j|a)P(m|a)P(a|\neg b \, A \neg e)P(\neg b)P(\neg e)$$
$$= 0.90 \ x \ 0.70 \ \text{x} \ 0.001 \ \text{x} \ 0.999 \ \text{x} \ 0.998 = 0.0006'2.$$

Semantics of Bayesian Networks:

- **There are two ways in which we can understand Semantics of Bayesian networks:**

1. See the network as representation of the joint probability distribution. This is useful in understanding how to construct networks.

2. See the networks as an encoding of a collection of conditional independence statements. This is useful in designing inference procedures. However, the two ways are equivalent.

Semantics of Bayesian Networks:

Representing the Full Joint Distribution:

- A belief network provides a complete description of the domain.
- Every entry in the joint probability distribution can be calculated from the information in the network
- Generic entry in the joint is the probability of a conjunction of particular assignments to each variable
- Each entry in the joint is represented by the product of the appropriate elements of the conditional probability tables (CPTs) in the belief network.
- Joint distribution can be used to answer any query about the domain.

Semantics of Bayesian Networks

Representing the Full Joint Distribution:

- We explain it, by calculating the probability that the alarm has sounded, but neither the burglary nor an earthquake has occurred, and both Jonny and Merry telephone you.

$$
\begin{aligned}
P(j \wedge m \wedge a \wedge \neg b \neg e) &= P(j \mid a)\, P(m \mid a)\, P(a \mid \neg b \wedge \neg e)\, P(\neg b)\, P(\neg e) \\
&= 0.90 \times 0.70 \times 0.001 \times 0.999 \times 0.998 \\
&= 0.00062
\end{aligned}
$$

- We can generalize the example by writing: $P(x_1, \ldots, x_n) = \prod_{j=1}^{n} P(x_i \mid parents\,(x_i))$

Constructing Bayesian networks

- 1. Choose an ordering of variables $X_1, \ldots, X_n$
- 2. For $i = 1$ to n
  - add X_i to the network
  - select parents from $X_1, \ldots, X_{i-1}$ such that

$$\mathbf{P}(X_i \mid Parents(X_i)) = \mathbf{P}(X_i \mid X_1, \ldots X_{i-1})$$

This choice of parents guarantees:

$$\mathbf{P}(X_1, \ldots, X_n) = \pi_{i=1}^{n} \mathbf{P}(X_i \mid X_1, \ldots, X_{i-1}) \quad \text{(chain rule)}$$

$$= \pi_{i=1}^{n} \mathbf{P}(X_i \mid Parents(X_i)) \quad \text{(by construction)}$$

$$\mathbf{P}(X_i|X_{i-1}, \ldots, X_1) = \mathbf{P}(X_i|Parents(X_i)) \, ,$$

$$\mathbf{P}(MaryCalls|JohnCalls, Alarm, Earthquake, Burglary) = \mathbf{P}(MaryCalls|Alarm) \, .$$

Inferences in Bayesian Networks

Purpose:

- Inference system is to compute the posterior probability distribution for a set of **query variables,** given some observed **event.**
- That is, some assignment of values to a set of **evidence variables.**

Inferences in Bayesian Networks

Notation:

- **X** → denotes the query variable
- **E** → set of evidence variables $E1, \ldots, Em$.
- **e** → particular observed event
- **Y** → The non-evidence, non-query variables $Y1, \ldots, Yn$ (called the **hidden variables**)
- The complete set of variables is $\mathbf{X} = \{X\} \cup \mathbf{E} \cup \mathbf{Y}$.
- A typical query asks for the posterior probability distribution $\mathbf{P}(X \mid \mathbf{e})$

Inferences in Bayesian Networks

- Inference by Enumeration
- Inference by Variable Elimination

Inferences in Bayesian Networks

- Any conditional probability can be computed by summing terms from the full joint distribution. More specifically, a query **P**(X | **e**) can be answered using Equation:

$$\boxed{\mathbf{P}(X \mid \mathbf{e}) = \alpha\, \mathbf{P}(X, \mathbf{e}) = \alpha \sum_{\mathbf{y}} \mathbf{P}(X, \mathbf{e}, \mathbf{y})\,.}$$

Inferences in Bayesian Networks

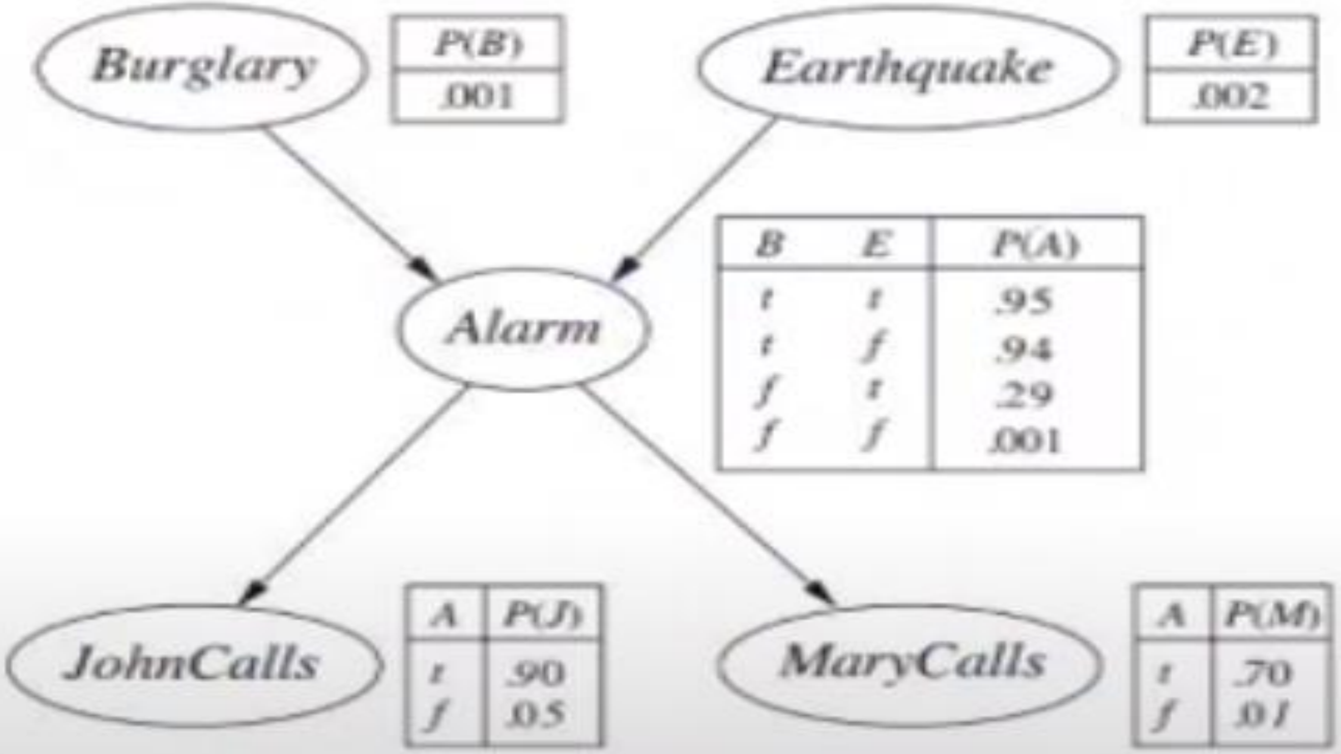


Figure 14.2 A typical Bayesian network, showing both the topology and the conditional probability tables (CPTs). In the CPTs, the letters B, E, A, J, and M stand for *Burglary*, *Earthquake*, *Alarm*, *JohnCalls*, and *MaryCalls*, respectively.

Inferences in Bayesian Networks

- Consider **P(Burglary | John Calls = true, Mary Calls = true)**
- **Burglary → Query variable (X)**
- **John calls → Evidence variable 1 (E1)**
- **Mary calls → Evidence variable 2 (E2)**
- **The hidden variables for this query are Earthquake and Alarm.**

Inferences in Bayesian Networks

- Using initial letters for the variables to shorten the expressions, we have

$$\boxed{\mathbf{P}(B \mid j, m) = \alpha\,\mathbf{P}(B, j, m) = \alpha \sum_{e} \sum_{a} \mathbf{P}(B, j, m, e, a,)\,.}$$

Inferences in Bayesian Networks

- The semantics of Bayesian networks, gives us an expression in terms of CPT entries. For simplicity, we do this just for **Burglary = true** as follows:

$$\boxed{P(b \mid j, m) = \alpha \sum_{e} \sum_{a} P(b) P(e) P(a \mid b, e) P(j \mid a) P(m \mid a)}$$

Inferences in Bayesian Networks

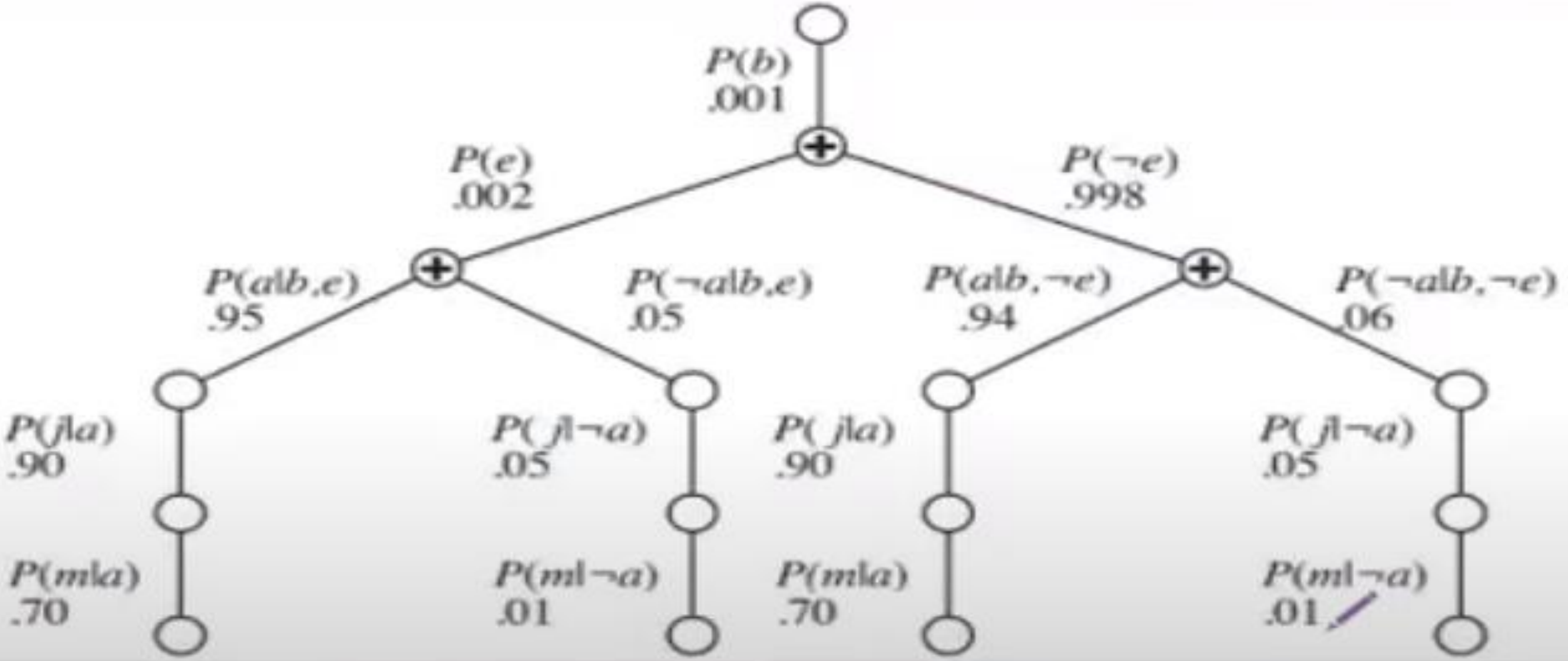


Figure 14.8 The structure of the expression shown in Equation (14.4). The evaluation proceeds top down, multiplying values along each path and summing at the "+" nodes. Notice the repetition of the paths for j and m.

Inferences in Bayesian Networks

- Inference by Enumeration
- Inference by Variable Elimination

Inferences in Bayesian Networks

Inference by Variable Elimination :

- The enumeration algorithm can be improved substantially by eliminating repeated calculations.
- The idea is simple: do the calculation once and save the results for later use. This is a form of dynamic programming.

Inferences in Bayesian Networks

<u>Inference by Variable Elimination :</u>

- Variable elimination works by evaluating expressions such as Equation :

$$\boxed{P(b \mid j, m) = \alpha P(b) \sum_{e} P(e) \sum_{a} P(a \mid b, e) P(j \mid a) P(m \mid a) .}$$

Inferences in Bayesian Networks

- Intermediate results are stored, and summations over each variable are done only for those portions of the expression that depend on the variable.
- Let us illustrate this process for the burglary network. We evaluate the expression:

$$\boxed{\mathbf{P}(B \mid j,m) = \alpha \underbrace{\mathbf{P}(B)}_{\mathbf{f}_1(B)} \sum_e \underbrace{P(e)}_{\mathbf{f}_2(E)} \sum_a \underbrace{\mathbf{P}(a \mid B,e)}_{\mathbf{f}_3(A,B,E)} \underbrace{P(j \mid a)}_{\mathbf{f}_4(A)} \underbrace{P(m \mid a)}_{\mathbf{f}_5(A)}}$$

Inferences in Bayesian Networks

- For example, the factors **f**4(A) and **f**5(A) corresponding to P(j | a) and P(m| a) depend just on **A** because **J** and **M** are fixed by the query. They are therefore two-element vectors:

$$\mathbf{f}_4(A) = \begin{pmatrix} P(j \mid a) \\ P(j \mid \neg a) \end{pmatrix} = \begin{pmatrix} 0.90 \\ 0.05 \end{pmatrix} \qquad \mathbf{f}_5(A) = \begin{pmatrix} P(m \mid a) \\ P(m \mid \neg a) \end{pmatrix} = \begin{pmatrix} 0.70 \\ 0.01 \end{pmatrix}$$

Inferences in Bayesian Networks

For example, given two factors f1(A,B) and f2(B,C), the pointwise product f1 ×f2 =f3(A,B,C) has 2^{1+1+1} =8 entries, as illustrated in below Figure :

A	B	$\mathbf{f}_1(A,B)$	B	C	$\mathbf{f}_2(B,C)$	A	B	C	$\mathbf{f}_3(A,B,C)$
T	T	.3	T	T	.2	T	T	T	$.3 \times .2 = .06$
T	F	.7	T	F	.8	T	T	F	$.3 \times .8 = .24$
F	T	.9	F	T	.6	T	F	T	$.7 \times .6 = .42$
F	F	.1	F	F	.4	T	F	F	$.7 \times .4 = .28$
						F	T	T	$.9 \times .2 = .18$
						F	T	F	$.9 \times .8 = .72$
						F	F	T	$.1 \times .6 = .06$
						F	F	F	$.1 \times .4 = .04$

Figure 14.10 Illustrating pointwise multiplication: $\mathbf{f}_1(A,B) \times \mathbf{f}_2(B,C) = \mathbf{f}_3(A,B,C)$.

Approximate inference in Bayesian Networks

- It is a method of estimating probabilities in Bayesian networks also called 'Monte Carlo' algorithms.
- We will discuss two types of algorithms: ***Direct sampling*** and ***Markov chain sampling.***
- ***Why use approximate inference?***
- Exact inference becomes intractable for large multiply-connected networks
- Variable elimination can have exponential time and space complexity
- Exact inference is strictly HARDER than NP-complete problems (#P-hard)

Direct Sampling

- In direct sampling we take samples of events. We expect the frequency of the samples to converge on the probability of the event.
- **Rejection Sampling** —
- Used to compute conditional probabilities $P(X|e)$
- Generate samples as before
- Reject samples that do not match evidence
- Estimate by counting the how often event X is in the resulting samples

Markov Chain Sampling

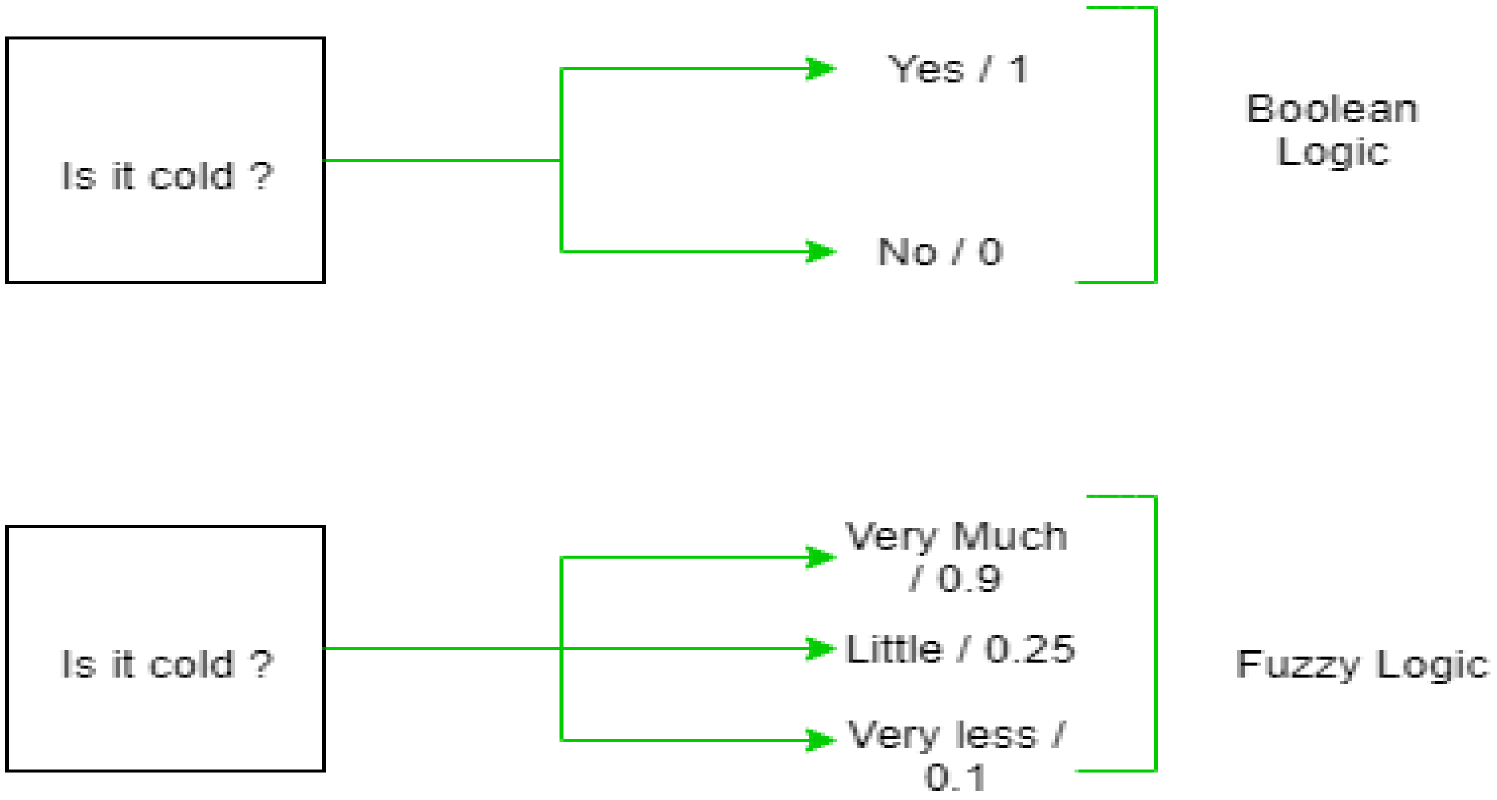
- Generate events by making a random change to the preceding event
- This change is made using the Markov Blanket of the variable to be changed
- Markov Blanket = parents, children, children's parents
- Tally and normalize results

Some Applications of BN

- Medical diagnosis
- Troubleshooting of hardware/software systems
- Fraud/uncollectible debt detection
- Data mining
- Analysis of genetic sequences
- Data interpretation, computer vision, image understanding

Other Approaches to Uncertain reasoning- Fuzzy sets and Fuzzy logic.

What is Fuzzy Logic?



Fuzzy logic

- Fuzzy means uncertain, indefinite, vague, or unclear.
- It is a computing technique that is based on the degree of truth.
- A fuzzy logic system uses the input's degree of truth and linguistic variables to produce a certain output.
- The state of this input determines the nature of the output.
- This technique is different from *boolean logic*
- In *boolean logic*, two categories (0 and 1) are used to describe objects
- **Example – the temperature in water served in glass may be**

BL - *High* (1) or *Low* (0)

fuzzy logic- *very cold*, *very warm*, or *warm*.

Fuzzy logic

- **Example -** Suppose we have a question that we need to answer

BL – *Yes or No*

Fuzzt Logic - *possibly yes*, *possibly no*, or *certainly no*.

What is Fuzzy Logic?

- The inventor/father of fuzzy logic is Lotfi Zadeh
- Fuzzy Logic (FL) is a method of reasoning that resembles human reasoning.
- The approach of FL imitates the way of decision making in humans that involves all intermediate possibilities between digital values YES and NO.
- The conventional logic block that a computer can understand takes precise input and produces a definite output as TRUE or FALSE, which is equivalent to human's YES or NO.
- Fuzzy Logic in AI provides valuable flexibility for reasoning

What is Fuzzy Logic?

- Inventor of Fuzzy logic, observed that unlike computers, the human decision making includes a range of possibilities between YES and NO, such as −

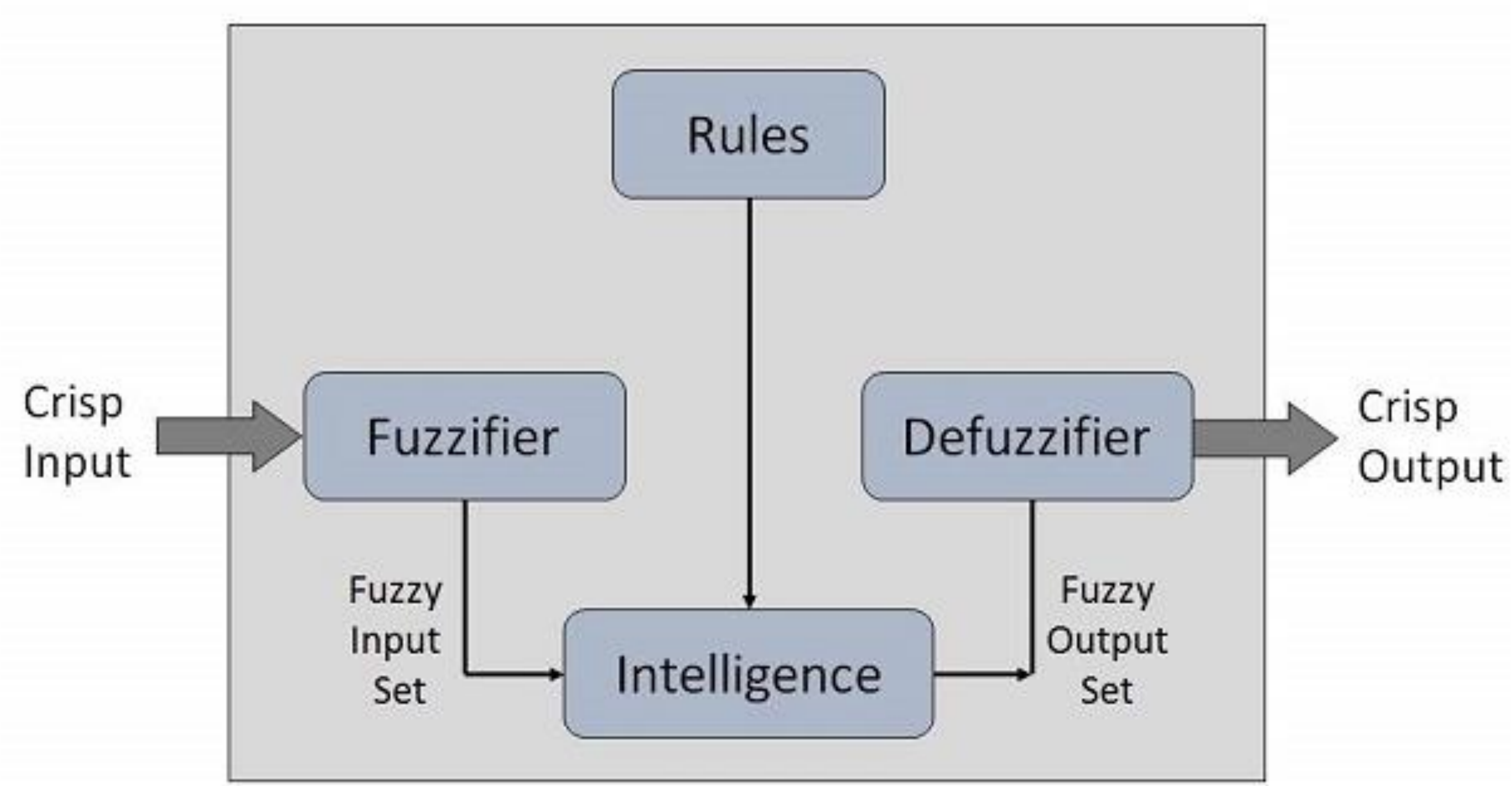
CERTAINLY YES
POSSIBLY YES
CANNOT SAY
POSSIBLY NO
CERTAINLY NO

- The fuzzy logic works on the levels of possibilities of input to achieve the definite output.

Why Fuzzy Logic?

- It solves the problem of uncertainty in the engineering field.
- When accurate reasoning is not available, it provides an accurate level of reasoning.
- Fuzzy logic has a simple structure that is easy to understand.
- It is an effective way of controlling machines.
- It provides solutions to various industrial problems (especially decision making).
- It requires little data to be executed.

Fuzzy Logic Systems Architecture



Fuzzy Logic Systems Architecture

- Fuzzy Logic Systems (FLS) produce acceptable but definite output in response to incomplete, ambiguous, distorted, or inaccurate (fuzzy) input.
- FSL has four main parts as shown −
  1. **Fuzzification Module** − It transforms the system inputs, which are crisp numbers, into fuzzy sets. It splits the input signal into five steps such as −

LP	x is Large Positive
MP	x is Medium Positive
S	x is Small
MN	x is Medium Negative
LN	x is Large Negative

Fuzzy Logic Systems Architecture

2. **Knowledge Base** – It stores IF-THEN rules provided by experts.
3. **Inference Engine** – It simulates the human reasoning process by making fuzzy inference on the inputs and IF-THEN rules.
4. **Defuzzification Module** – It transforms the fuzzy set obtained by the inference engine into a crisp value.

The **membership functions work on** fuzzy sets of variables

Membership Function

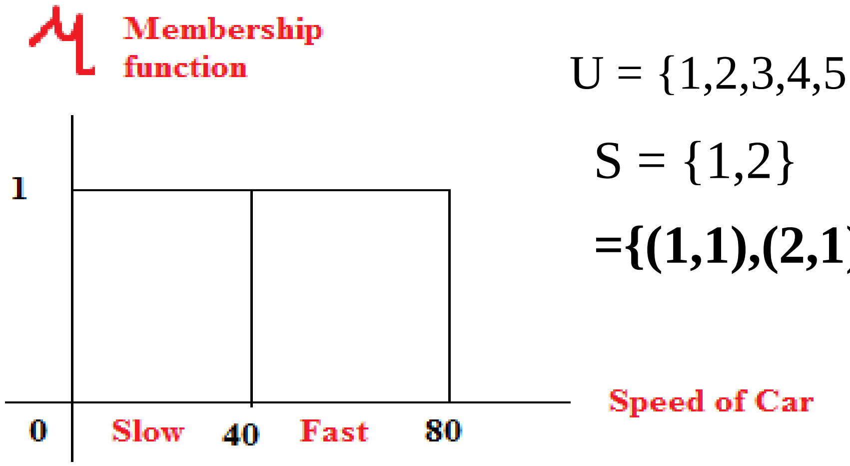
- It is a graph that defines how each point in the input space is mapped to membership value between 0 and 1.

- Allow you to quantify linguistic term and represent a fuzzy set graphically.

- A **membership function** for a fuzzy *set A* on the universe of discourse X is defined as $\mu_A : X \to [0,1]$.

- Here, each element of *X* is mapped to a value between 0 and 1. It is called **membership value** or **degree of membership**. It quantifies the degree of membership of the element in *X* to the fuzzy set *A*.
  - x axis represents the universe of discourse.
  - y axis represents the degrees of membership in the [0, 1] interval.

Membership Function



U = {1,2,3,4,5}

S = {1,2}

={(1,1),(2,1),(3,0),(4,0),(5,0) }

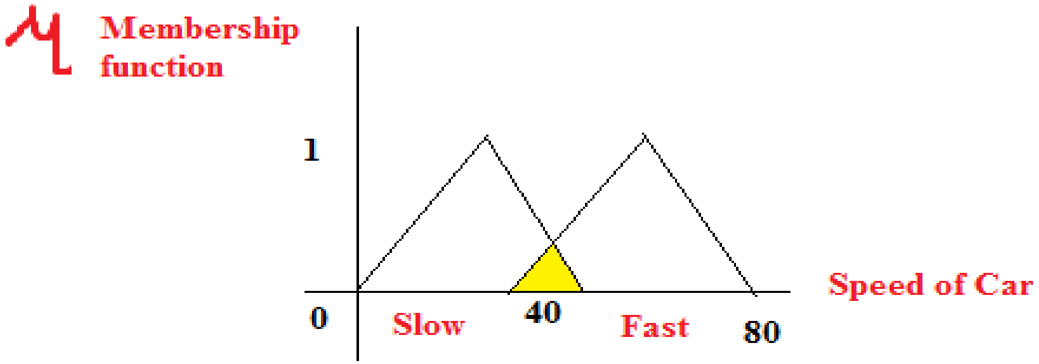
Degree of fastness =

0, if speed(x) <= 40

speed(x)-40/10, if 40 < speed (x) < 50

1, if speed(x) >= 50

Example –



x = 30 (30,0)

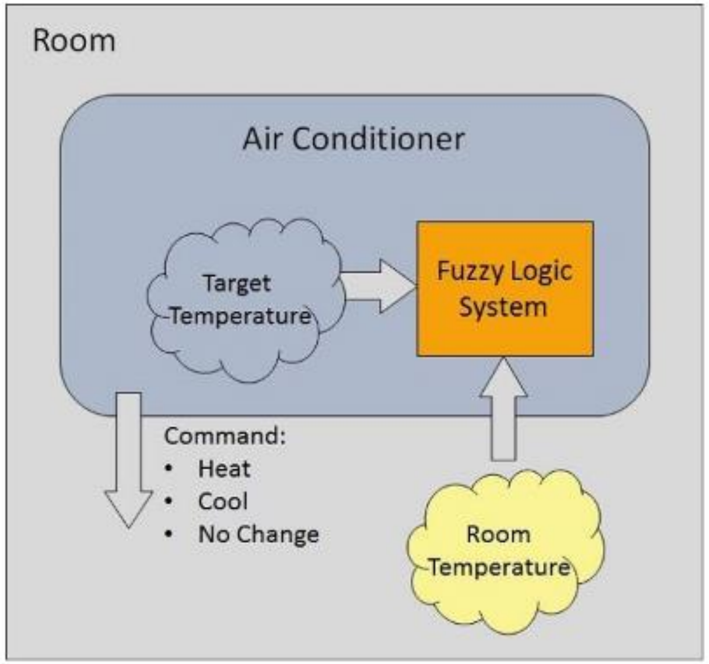
x = 60 (60,1)

x = 42 = 42 – 40/10 = 2/10 =0.2

x = 45 = 45-40/10 = 5/2 = 0. 5

Example of a Fuzzy Logic System

- Let us consider an air conditioning system. This system adjusts the temperature of air conditioner by comparing the room temperature and the target temperature value

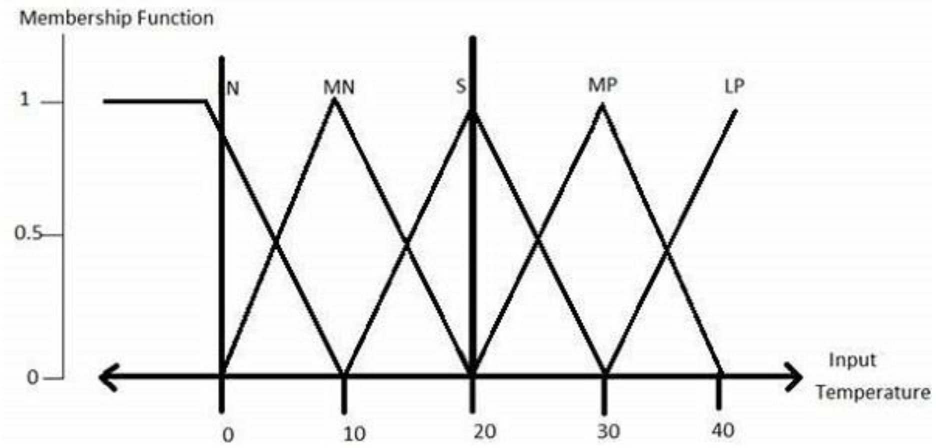


Algorithm

- Define linguistic Variables and terms (start)
- Construct membership functions for them. (start)
- Construct knowledge base of rules (start)
- Convert crisp data into fuzzy data sets using membership functions. (fuzzification)
- Evaluate rules in the rule base. (Inference Engine)
- Combine results from each rule. (Inference Engine)
- Convert output data into non-fuzzy values. (defuzzification)

Development of Fuzzy logic system

- **Step 1 − Define linguistic variables and terms**
  - Linguistic variables are input and output variables in the form of simple words or sentences. For room temperature, cold, warm, hot, etc., are linguistic terms.
  - Temperature (t) = {very-cold, cold, warm, very-warm, hot}
  - Every member of this set is a linguistic term and it can cover some portion of overall temperature values.
- **Step 2 − Construct membership functions for them**
  - Membership functions are represented by graphical forms.
  - The membership functions of temperature variable are as shown −



- **Step3 − Construct knowledge base rules**
- Create a matrix of room temperature values versus target temperature values that an air conditioning system is expected to provide.

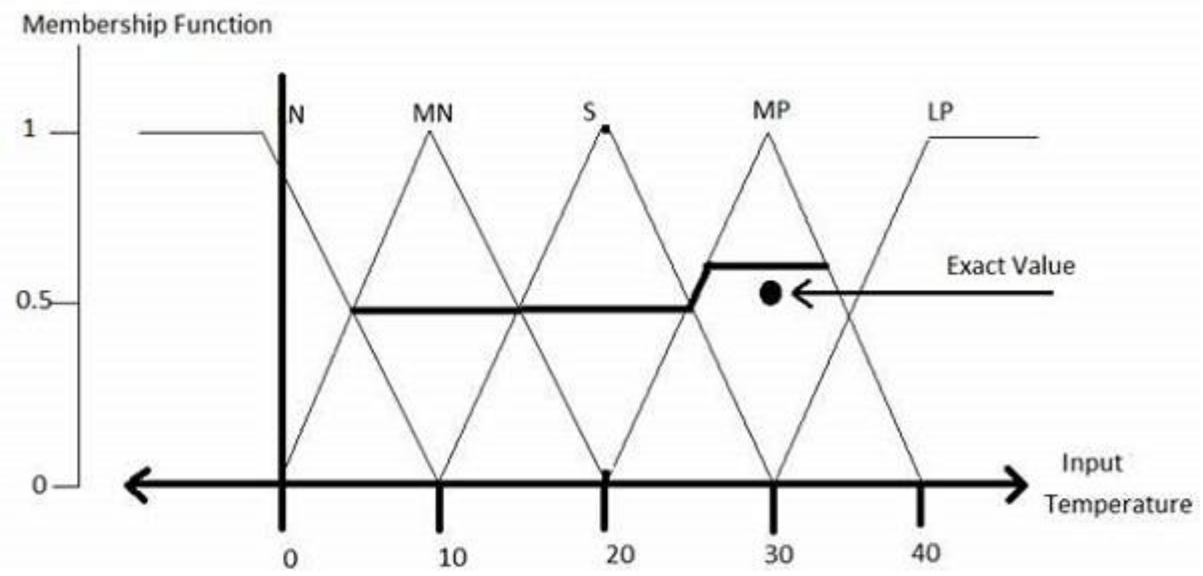
RoomTemp. /Target	Very_Cold	Cold	Warm	Hot	Very_Hot
Very_Cold	No_Change	Heat	Heat	Heat	Heat
Cold	Cool	No_Change	Heat	Heat	Heat
Warm	Cool	Cool	No_Change	Heat	Heat
Hot	Cool	Cool	Cool	No_Change	Heat
Very_Hot	Cool	Cool	Cool	Cool	No_Change

- Build a set of rules into the knowledge base in the form of IF-THEN-ELSE structures.

Sr. No.	Condition	Action
1	IF temperature=(Cold OR Very_Cold) AND target=Warm THEN	Heat
2	IF temperature=(Hot OR Very_Hot) AND target=Warm THEN	Cool
3	IF (temperature=Warm) AND (target=Warm) THEN	No_Change

- **Step 4 – Obtain fuzzy value**
  - Fuzzy set operations perform evaluation of rules. The operations used for OR and AND are Max and Min respectively. Combine all results of evaluation to form a final result. This result is a fuzzy value.

- **Step 5 − Perform defuzzification:**
  - Defuzzification is then performed according to membership function for output variable.



Application Areas of Fuzzy Logic

- The key application areas of fuzzy logic are as given −
- **Automotive Systems**
  - Automatic Gearboxes, Four-Wheel Steering, Vehicle environment control,etc
- **Consumer Electronic Goods**
  - Photocopiers, Still and Video Cameras, Television,etc
- **Domestic Goods**
  - Microwave Ovens, Refrigerators, Toasters, Vacuum Cleaners, Washing Machines, etc
- **Environment Control**
  - Air Conditioners/Dryers/Heaters, Humidifiers, etc

Advantages and Disadvantages of FLSs

- **Advantages of FLSs**
  - Mathematical concepts within fuzzy reasoning are very simple.
  - You can modify a FLS by just adding or deleting rules due to flexibility of fuzzy logic.
  - Fuzzy logic Systems can take imprecise, distorted, noisy input information.
  - FLSs are easy to construct and understand.
  - Fuzzy logic is a solution to complex problems in all fields of life, including medicine, as it resembles human reasoning and decision making.
- **Disadvantages of FLSs**
  - There is no systematic approach to fuzzy system designing.
  - They are understandable only when simple.
  - They are suitable for the problems which do not need high accuracy.