CMR INSTITUTE OF TECHNOLOGY

* UGC AUTONOMOUS *

* Approved by AICTE * Accredited by NAAC with 'A' Grade * All B.Tech programs Accredited by NBA*

(Kandlakoya (v),Medchal Road,Hyderabad - 501 401)

EXPLORE TO INVENT

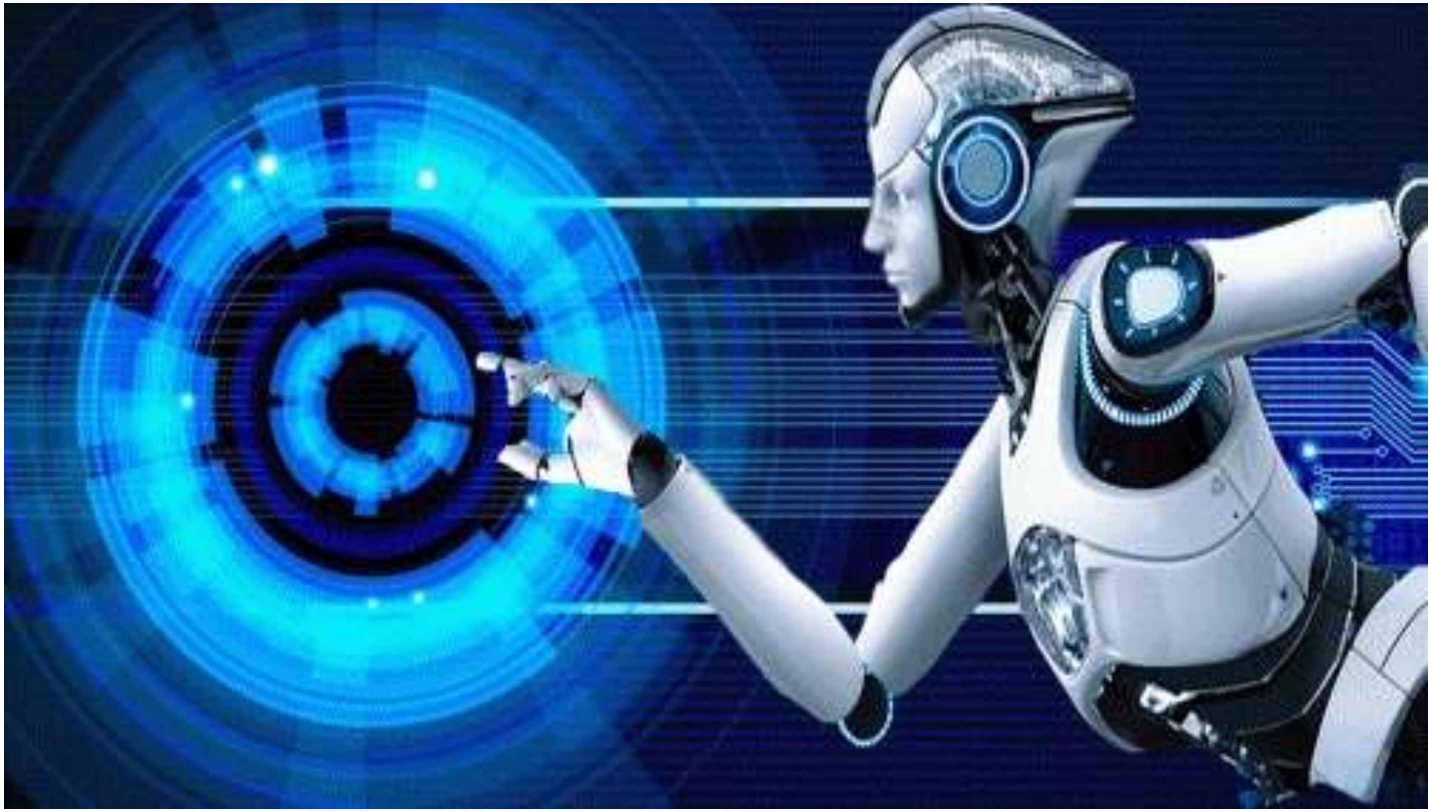
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

III – B.Tech – I – Sem (A & C Sections)

ARTIFICIAL INTELLIGENGE (20-CS-PC-314)
(R20 Regulations)

By
Dr.A.Nirmal Kumar

ARTIFICIAL INTELLIGENCE
- The Future



SYLLABUS

Unit-I: Introduction **8 Hours**

Concept of AI, history, current status, scope, agents, environments, Problem Formulations, Review of tree and graph structures, State space representation, Search graph and Search tree.

Unit-II: Search Algorithms **10 Hours**

Random search, Search with closed and open list, Depth first and Breadth first search, Heuristic search, Best first search, A* algorithm, Game Search.

Unit-III: Probabilistic Reasoning **(6 + 4) 10 Hours**

Part-A: Probability, conditional probability, Bayes Rule, Bayesian Networks- representation, construction and inference.

Part-B: Temporal Model, Hidden Markov Model.

Unit-IV: Markov Decision Process **10 Hours**

MDP formulation, utility theory, utility functions, value iteration, policy iteration and partially observable MDPs.

Unit-V: Reinforcement Learning **10 Hours**

Passive reinforcement learning, direct utility estimation, adaptive dynamic programming, temporal difference learning, active reinforcement learning- Q learning.

Textbooks:

1. Elaine Rich & Kevin Knight, 'Artificial Intelligence', 3rd Edition, TMH, 2008.
2. Russel and Norvig, 'Artificial Intelligence', Pearson Education, PHI, 2003.

References:

1. Trivedi, M.C., "A Classical Approach to Artifical Intelligence", Khanna Publishing House, Delhi.
2. Saroj Kaushik, "Artificial Intelligence", Cengage Learning India, 2011.
3. David Poole and Alan Mackworth, "Artificial Intelligence: Foundations for Computational Agents", Cambridge University Press 2010.

COURSE OUTCOMES

Upon completion of the course, the student will be able

CO 1: To explain the concepts of artificial intelligence **(Unit – I)**

CO 2: To illustrate various search algorithms **(Unit – II)**

CO 3: To adapt various probabilistic reasoning approaches **(Unit – III)**

CO 4: To elaborate Markov decision process **(Unit – IV)**

CO 5: To perceive various reinforcement learning approaches **(Unit – V)**

UNIT – III
Probabilistic Reasoning

Part-A:

- Probability
- Conditional probability
- Bayes Rule
- Bayesian Networks- representation
- Construction and inference

Part-B:

- Temporal Model
- Hidden Markov Model

Probabilistic reasoning in Artificial Intelligence

Uncertainty:

- Till now, we have learned knowledge representation using first-order logic and propositional logic with certainty, which means we were sure about the predicates. With this knowledge representation, we might write $A \rightarrow B$, which means if A is true then B is true, but consider a situation where we are not sure about whether A is true or not then we cannot express this statement, this situation is called **uncertainty**.
- So to represent uncertain knowledge, where we are not sure about the predicates, we need uncertain reasoning or probabilistic reasoning.

Causes of Uncertainty:

Following are some leading causes of uncertainty to occur in the real world.

1. Information occurred from unreliable sources
2. Experimental Errors
3. Equipment fault
4. Temperature variation
5. Climate change

Probabilistic Reasoning

- Probabilistic reasoning is a way of knowledge representation where we apply the concept of probability to indicate the uncertainty in knowledge. In probabilistic reasoning, we combine probability theory with logic to handle the uncertainty.
- We use probability in probabilistic reasoning because it provides a way to handle the uncertainty that is the result of someone's laziness and ignorance.
- In the real world, there are lots of scenarios, where the certainty of something is not confirmed, such as "It will rain today," "behavior of someone for some situations," "A match between two teams or two players." These are probable sentences for which we can assume that it will happen but not sure about it, so here we use probabilistic reasoning.

Need of probabilistic reasoning in AI:

- When there are unpredictable outcomes.
- When specifications or possibilities of predicates becomes too large to handle.
- When an unknown error occurs during an experiment.

In probabilistic reasoning, there are two ways to solve problems with uncertain knowledge:

- **Bayes' rule**
- **Bayesian Statistics**

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.
- Probability implies 'likelihood' or 'chance'.
- When an event is certain to happen then the probability of occurrence of that event is 1 and when it is certain that the event cannot happen then the probability of that event is 0.
- Hence the value of probability ranges from 0 to 1. Probability has been defined in a varied manner by various schools of thought.

Classical Definition of Probability

- As the name suggests the classical approach to defining probability is the oldest approach. It states that if there are n exhaustive, mutually exclusive and equally likely cases out of which m cases are favorable to the happening of event A.

Example

- **Problem Statement:**

  A coin is tossed. What is the probability of getting a head?

- **Solution:**

  Total number of equally likely outcomes (n) = 2 (i.e. head or tail)

  Number of outcomes favorable to head (m) = 1

- $0 \leq P(A) \leq 1$, where P(A) is the probability of an eve nt A.
- P(A) = 0, indicates total uncertainty in an event A.
- P(A) =1, indicates total certainty in an event A.

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

- Probability of Occurrence = (Number of desired Outcomes / Total No of Outcomes)
- $P(\neg A)$ = probability of a not happening event.
- $P(\neg A) + P(A) = 1$.

Terminologies in Probability Theory

- **Event:** Each possible outcome of a variable is called an event.
- **Sample space:** The collection of all possible events is called sample space.
- **Random variables:** Random variables are used to represent the events and objects in the real world.
- **Prior probability:** The prior probability of an event is probability computed before observing new information.
- **Posterior Probability:** The probability that is calculated after all evidence or information has taken into account. It is a combination of prior probability and new information.

Mathematical probabilities use percentages. For example, when a meteorologist says "you can expect a 70% chance for thunderstorms," that is a probability. They are speaking in mathematical probabilities. This is done by more than just meteorologists, too. Explore other fun mathematical probability examples.

- There is a 20% chance of rain tomorrow.
- When flipping a coin, there is a 50% probability it will be heads.
- On a spinner that has four colors occupying equally sized spaces, there is a one in four probability it will land on any one color.
- In a drawer of ten socks where 8 of them are yellow, there is a 20% chance of choosing a sock that is not yellow.
- There is a 50% chance of snow tonight.
- There are 100 silver cars on the sales lot with 200 total cars. The probability of a customer choosing a silver car is 50%.
- There are 9 red candies in a bag and 1 blue candy in the same bag. The chance of picking a blue candy is 10%.
- In my closet are 5 pairs of shoes, 4 of which are black. The chances of picking a black pair of shoes are 4 out of 5.
- The probability of winning the lottery is 1 in many millions.

Basic Probability Rules

- Probability Rule One (For any event A, $0 \leq P(A) \leq 1$)
- Probability Rule Two (The sum of the probabilities of all possible outcomes is 1)
- Probability Rule Three (The Complement Rule)
- Probabilities Involving Multiple Events.
- Probability Rule Four (**Addition** Rule for Disjoint Events)
- Finding P(A and B) using Logic.

How is probability used in real life?

- You **use probability** in **daily life** to make decisions when you don't know for sure what the outcome will be. Most of the time, you won't perform actual **probability** problems, but you'll **use** subjective **probability** to make judgment calls and determine the best course of action.

What is a certain event?

- A **certain event** is an **event** that is sure to happen. E is a **certain event** if and only if P(E) = 1. Example. In flipping a coin once, a **certain event** would be getting a head or a tail.
- The probability formula provides the ratio of the number of favorable outcomes to the total number of possible outcomes. The probability of an Event = (Number of favorable outcomes) / (Total number of possible outcomes) $P(A) = n(E) / n(S)$

What is impossible event?

- An **impossible event** is an **event** that cannot happen. E is an **impossible event** if and only if P(E) = 0. Example. In flipping a coin once, an **impossible event** would be getting BOTH a head AND a tail.
- The **probability line** is a line that shows probabilities and how these probabilities relate to each other. Since the probability of an event is a number from 0 to 1, we can use the probability line above to show the possible ranges of probability values.

Four perspectives on probability are commonly used:

- Classical (sometimes called "A priori" or "Theoretical") ...
- Empirical (sometimes called "A posteriori" or "Frequentist") ...
- Subjective. ...
- Axiomatic.

What are the different types of probability distributions?

- There are many different classifications of probability distributions. Some of them include the normal distribution, chi **square** distribution, binomial distribution, and Poisson distribution.

What are probability models?

- A **probability model** is a mathematical representation of a random phenomenon. It is defined by its sample space, events within the sample space, and **probabilities** associated with each event. The sample space S for a **probability model** is the set of all possible outcomes.

Who is known as father of probability?

- A gambler's dispute in 1654 led to the creation of a mathematical theory of probability by two famous French mathematicians, Blaise Pascal and Pierre de Fermat.

What are the 3 axioms of probability?

- For any event A, $P(A) \geq 0$. In English, that's “For any event A, the **probability** of A is greater or equal to 0”.
- When S is the sample space of an experiment; i.e., the set of all possible outcomes, $P(S) = 1$.
- If A and B are mutually exclusive outcomes, $P(A \cup B) = P(A) + P(B)$.

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:

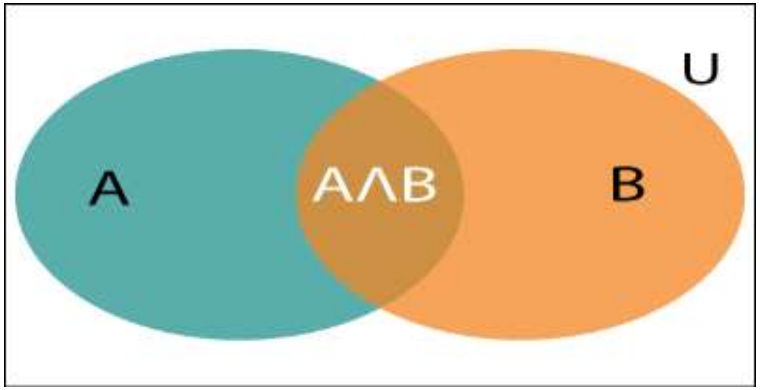
$$P(A|B) = \frac{P(A \wedge B)}{P(B)}$$

Where $P(A \wedge B)$= Joint probability of A and B
P(B)= Marginal probability of B.

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

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

It can be explained by using the below Venn diagram, where B is occurred event, so sample space will be reduced to set B, and now we can only calculate event A when event B is already occurred by dividing the probability of P(A⋀B) by P(B).



- **Conditional probability** is the **probability** of one event occurring with some relationship to one or more other events. **For example:** Event A is that it is raining outside, and it has a 0.3 (30%) chance of raining today. Event B is that you will need to go outside, and that has a probability of 0.5 (50%).

What is the difference between probability and conditional probability?

- Their only difference is that the conditional probability assumes that we already know something -- that B is true.

How do you calculate conditional proportions?

- The analog of **conditional proportion** is **conditional** probability: P(A|B) means “probability that A happens, if we know that B happens”. The **formula** is P(A|B) = P(A and B)/P(B).

Independent Events

- Events can be "**Independent**", meaning each event is **not affected** by any other events.

Example: Tossing a coin.

- Each toss of a coin is a perfect isolated thing.
- What it did in the past will not affect the current toss.
- The chance is simply 1-in-2, or 50%, just like ANY toss of the coin.
- So each toss is an **Independent Event**.

Dependent Events

- But events can also be "dependent" ... which means they **can be affected by previous events.**

Example: Marbles in a Bag

- 2 blue and 3 red marbles are in a bag.
- What are the chances of getting a blue marble?
- The chance is **2 in 5**
- **But after taking one out** the chances change!
- So the next time:
- if we got a **red** marble before, then the chance of a blue marble next is **2 in 4**
- if we got a **blue** marble before, then the chance of a blue marble next is **1 in 4**

Bayes Rule

- Bayes' theorem is also known as **Bayes' rule, Bayes' law**, or **Bayesian reasoning**, which determines the probability of an event with uncertain knowledge.
- In probability theory, it relates the conditional probability and marginal probabilities of two random events.
- Bayes' theorem was named after the British mathematician **Thomas Bayes**. The **Bayesian inference** is an application of Bayes' theorem, which is fundamental to Bayesian statistics.
- It is a way to calculate the value of P(B|A) with the knowledge of P(A|B).

- Bayes' theorem allows updating the probability prediction of an event by observing new information of the real world.
- **Example**: If cancer corresponds to one's age then by using Bayes' theorem, we can determine the probability of cancer more accurately with the help of age.
- Bayes' theorem can be derived using product rule and conditional probability of event A with known event B:
- As from product rule we can write:

$$P(A \land B) = P(A|B)\,P(B) \text{ or}$$

Similarly, the probability of event B with known event A:

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

Equating right hand side of both the equations, we will get:

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

- The above equation (a) is called as Bayes' rule or Bayes' theorem. This equation is basic of most modern AI systems for probabilistic inference.
- It shows the simple relationship between joint and conditional probabilities. Here,
- P(A|B) is known as posterior, which we need to calculate, and it will be read as 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.
- In the equation (a), in general, we can write P (B) = P(A)*P(B|Ai), hence the Bayes' rule can be written as:

$$P(A_i|B) = \frac{P(A_i)*P(B|A_i)}{\sum_{i=1}^{k} P(A_i)*P(B|A_i)}$$

Where A1, A2, A3,........, An is a set of mutually exclusive and exhaustive events.

Applying Bayes' Rule:

Bayes' rule allows us to compute the single term P(B|A) in terms of P(A|B), P(*B*), and P(A). This is very useful in cases where we have a good probability of these three terms and want to determine the fourth one. Suppose we want to perceive the effect of some unknown cause, and want to compute that cause, then the Bayes' rule becomes:

$$P(\text{cause} \mid \text{effect}) = \frac{P(\text{effect} \mid \text{cause})\, P(\text{cause})}{P(\text{effect})}$$

- **Example-1:**

  Question: what is the probability that a patient has diseases meningitis with a stiff neck?

  Given Data:

- 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%.
- Let a be the proposition that patient has stiff neck and b be the proposition that patient has meningitis. , so we can calculate the following as:
- $P(a|b) = 0.8$
- $P(b) = 1/30000$
- $P(a)= 0.02$

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

Hence, we can assume that 1 patient out of 750 patients has meningitis disease with a stiff neck.

Example-2:

Question: From a standard deck of playing cards, a single card is drawn. The probability that the card is king is 4/52, then calculate posterior probability P(King|Face), which means the drawn face card is a king card.

Solution:

$$P(king|face) = \frac{P(Face|king) * P(King)}{P(Face)} \quad ....(i)$$

P(king): probability that the card is King= 4/52= 1/13
P(face): probability that a card is a face card= 3/13
P(Face|King): probability of face card when we assume it is a king = 1

Putting all values in equation (i) we will get:

$$P(king|face) = \frac{1 * (\frac{1}{13})}{(\frac{3}{13})} = 1/3,$$ **it is a probability that a face card is a king card.**

Application of Bayes' theorem in Artificial Intelligence:

- 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.
- It can solve the Monty Hall problem.

Bayesian Belief Network

- Bayesian belief network is key computer technology for dealing with probabilistic events and to solve a problem which has uncertainty. We can define a Bayesian network as:

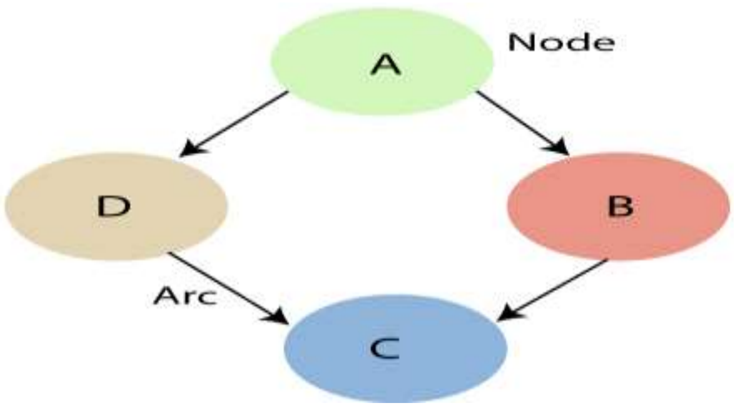
  "A Bayesian network is a probabilistic graphical model which represents a set of variables and their conditional dependencies using a directed acyclic graph."
- It is also called a **Bayes network, belief network, decision network**, or **Bayesian model**.
- Bayesian networks are probabilistic, because these networks are built from a probability distribution, and also use probability theory for prediction and anomaly detection.

- Real world applications are probabilistic in nature, and to represent the relationship between multiple events, we need a Bayesian network. It can also be used in various tasks including **prediction, anomaly detection, diagnostics, automated insight, reasoning, time series prediction**, and **decision making under uncertainty**.
- Bayesian Network can be used for building models from data and experts opinions, and it consists of two parts:

  Directed Acyclic Graph

  Table of conditional probabilities.
- The generalized form of Bayesian network that represents and solve decision problems under uncertain knowledge is known as an **Influence diagram**.

A Bayesian network graph is made up of nodes and Arcs (directed links), where:



•Each **node** corresponds to the random variables, and a variable can be **continuous** or **discrete**.

•**Arc or directed arrows** represent the causal relationship or conditional probabilities between random variables. These directed links or arrows connect the pair of nodes in the graph. These links represent that one node directly influence the other node, and if there is no directed link that means that nodes are independent with each other

- In the above diagram, A, B, C, and D are random variables represented by the nodes of the network graph.
- If we are considering node B, which is connected with node A by a directed arrow, then node A is called the parent of Node B.
- Node C is independent of node A.
- The Bayesian network graph does not contain any cyclic graph. Hence, it is known as a **directed acyclic graph or DAG.**
- The Bayesian network has mainly two components:
    1. Causal Component
    2. Actual numbers

    Each node in the Bayesian network has condition probability distribution P(Xi |Parent(Xi)), which determines the effect of the parent on that node.

- Bayesian network is based on Joint probability distribution and conditional probability. So let's first understand the joint probability distribution:

Joint probability distribution:

- If we have variables x1, x2, x3,....., xn, then the probabilities of a different combination of x1, x2, x3.. xn, are known as Joint probability distribution.

  P[x1, x2, x3,....., xn], it can be written as the following way in terms of the joint probability distribution.

  = P[x1| x2, x3,....., xn]P[x2, x3,....., xn]

  = P[x1| x2, x3,....., xn]P[x2|x3,....., xn]....P[xn-1|xn]P[xn].

Explanation of Bayesian Network:

Let's understand the Bayesian network through an example by creating a directed acyclic graph:

Example: Harry installed a new burglar alarm at his home to detect burglary. The alarm reliably responds at detecting a burglary but also responds for minor earthquakes. Harry has two neighbors David and Sophia, who have taken a responsibility to inform Harry at work when they hear the alarm. David always calls Harry when he hears the alarm, but sometimes he got confused with the phone ringing and calls at that time too. On the other hand, Sophia likes to listen to high music, so sometimes she misses to hear the alarm. Here we would like to compute the probability of Burglary Alarm.

- **Problem:**

  Calculate the probability that alarm has sounded, but there is neither a burglary, nor an earthquake occurred, and David and Sophia both called the Harry.

- **Solution:**
- The Bayesian network for the above problem is given below. The network structure is showing that burglary and earthquake is the parent node of the alarm and directly affecting the probability of alarm's going off, but David and Sophia's calls depend on alarm probability.
- The network is representing that our assumptions do not directly perceive the burglary and also do not notice the minor earthquake, and they also not confer before calling.

- The conditional distributions for each node are given as conditional probabilities table or CPT.
- Each row in the CPT must be sum to 1 because all the entries in the table represent an exhaustive set of cases for the variable.
- In CPT, a boolean variable with k boolean parents contains 2^K probabilities. Hence, if there are two parents, then CPT will contain 4 probability values.
- **List of all events occurring in this network:**
    - **Burglary (B)**
    - **Earthquake(E)**
    - **Alarm(A)**
    - **David Calls(D)**
    - **Sophia calls(S)**

- We can write the events of problem statement in the form of probability: **P[D, S, A, B, E]**, can rewrite the above probability statement using joint probability distribution:

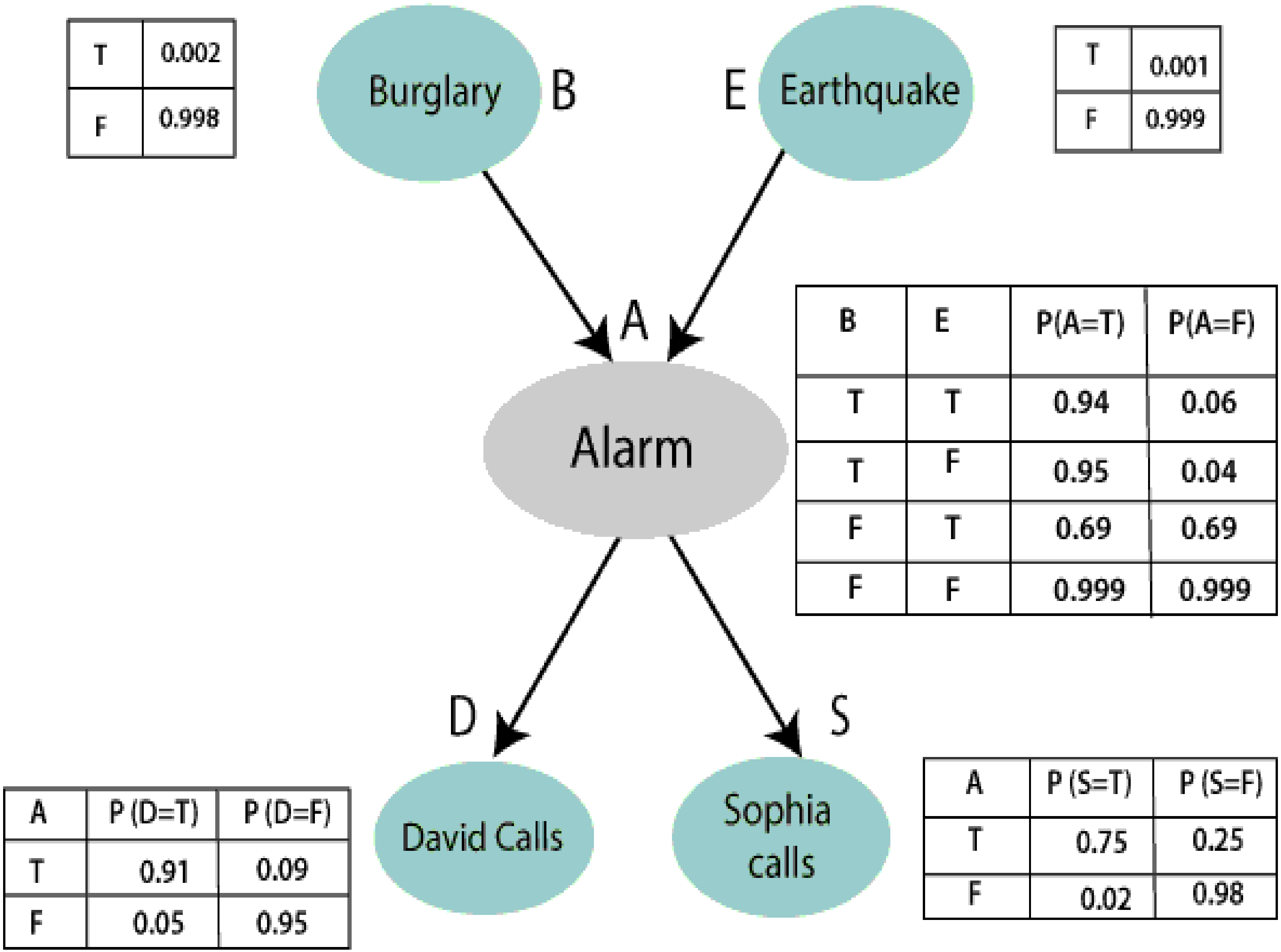
P[D, S, A, B, E]= P[D | S, A, B, E]. P[S, A, B, E]

=P[D | S, A, B, E]. P[S | A, B, E]. P[A, B, E]

= P [D| A]. P [S| A, B, E]. P[A, B, E]

= P[D | A]. P[S | A]. P[A| B, E]. P[B, E]

= P[D | A]. P[S | A]. P[A| B, E]. P[B |E]. P[E]



T	0.002
F	0.998

T	0.001
F	0.999

B	E	P(A=T)	P(A=F)
T	T	0.94	0.06
T	F	0.95	0.04
F	T	0.69	0.69
F	F	0.999	0.999

A	P (D=T)	P (D=F)
T	0.91	0.09
F	0.05	0.95

A	P (S=T)	P (S=F)
T	0.75	0.25
F	0.02	0.98

- Let's take the observed probability for the Burglary and earthquake component:
- P(B= True) = 0.002, which is the probability of burglary.
- P(B= False)= 0.998, which is the probability of no burglary.
- P(E= True)= 0.001, which is the probability of a minor earthquake
- P(E= False)= 0.999, Which is the probability that an earthquake not occurred.

- We can provide the conditional probabilities as per the below tables:
- **Conditional probability table for Alarm A:**

The Conditional probability of Alarm A depends on Burglar and earthquake:

B	E	P(A= True)	P(A= False)
True	True	0.94	0.06
True	False	0.95	0.04
False	True	0.31	0.69
False	False	0.001	0.999

Conditional probability table for David Calls:

The Conditional probability of David that he will call depends on the probability of Alarm.

A	P(D= True)	P(D= False)
True	0.91	0.09
False	0.05	0.95

Conditional probability table for Sophia Calls:

The Conditional probability of Sophia that she calls is depending on its Parent Node "Alarm”.

A	P(S= True)	P(S= False)
True	0.75	0.25
False	0.02	0.98

From the formula of joint distribution, we can write the problem statement in the form of probability distribution:

P(S, D, A, ¬B, ¬E) = P (S|A) *P (D|A)*P (A|¬B ^ ¬E) *P (¬B) *P (¬E).

= 0.75* 0.91* 0.001* 0.998*0.999

= 0.00068045.

Hence, a Bayesian network can answer any query about the domain by using Joint distribution.

The semantics of Bayesian Network

- There are two ways to understand the semantics of the Bayesian network, which is given below:

1. To understand the network as the representation of the Joint probability distribution.

It is helpful to understand how to construct the network.

2. To understand the network as an encoding of a collection of conditional independence statements.

It is helpful in designing inference procedure.

Types Probability Models

- Bayes' Net
- Temporal Probability Model
- Dynamic Bayes' Net (DBN)

  Special classes of DBN

  – Hidden Markov Model (HMM)
  – Kalman Filter