Artificial Intelligence
Prof. Anupam Basu
Department of Computer Science and Engineering
Indian Institute of Technology, Kharagpur
Lecture # 29
Reasoning with Uncertainty - IV

In the last lecture we had introduced the concept of conditional probability and we had also introduced Bayes' rule. In this lecture we will first start with Bayes' rule and then we will continue to see how the probabilistic reasoning can actually be implemented using the notion of conditional probability. So, to start with we have a revisit of the Bayes' rule. As you can see here I want to know the probability of A given the probability of B's occurrence.

(Refer Slide Time: 02:18)

When I know the probability of occurrence of B how would I compute the probability of A occurring?
That is expressed in terms of A Priori probabilities of B and A and also the probability of B by A. So here we are trying to compute A by B in terms of the known probability of B by A and here is an example of how we compute that. The A Priori probability of the doorbell ringing is 0.001 and the A Priori probability of Mohan's waking up is 0.01 and A Priori probability of waking up given the doorbell ringing is 0.8 and from there we compute the probability of doorbell ringing given Mohan's waking up using Bayes' rule as 0.8 times P(D) by P(W) so we get 0.08. This is the basic notion of Bayes' rule. Such conditional probabilities are applied when the two events are not independent.

(Refer Slide Time: 03:55)


If two events are independent A and B two events are independent when do I call them independent?

In terms of conditional probability if A and B are independent then probability of A given B will be the same as the probability of A because A and B are actually independent. And probability of B by A will be the probability of B itself because they are independent. Therefore in case of independent events the joint probability P(A AND B) is P(A) times P(B). Here we can see a simple scenario that the P(A by B) is 0.2, P(A by NOTB) is 0.3, P(NOTA by B) is 0.1 and P(NOTA by NOTB) is 0.4. It is interesting to note that since they are joint probability the sum of all these are coming to one. Therefore here there are some nice formulations about computing the probabilities of events. We can compute the P(A) if we know the P(B) in terms of Bayes' rule and that is very much applicable in the case of rules where we deal with event e leading to hypothesis h so p(h by e) can be computed similarly. But there are of course some advantages and also there are some problems.
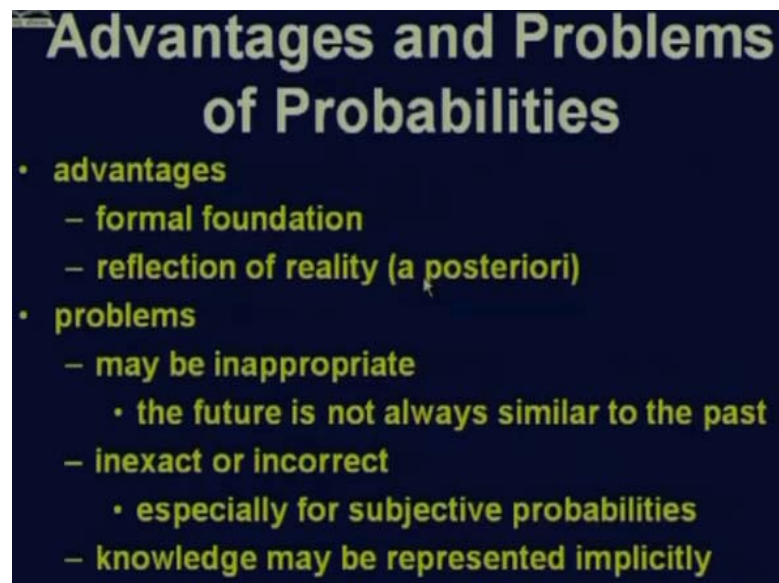
For example, the advantages are; Since it is based on the probability theory which has got years of formal mathematical foundation it is very rigorous and it is the reflection of a reality because we are computing a-posteriori probability. The word a-posteriori is just the opposite of A Priori. A Priori means something that is known beforehand and a-posteriori is something that we are trying to know and that is very much evident in the case of Bayes' rule as we had done that we want to the find the probability of this. That is, we want to the find the P(h by e) we want to know this.

We know the A Priori probability of e, we know the A Priori probability of h and we also know the A Priori P(e by h) that is same as h leads to e. These are known A Priori, these are known beforehand A Priori and using Bayes' rule using these three we are computing this one which is a-posteriori. That is a very nice thing to do. However, there are some inherent problems in the case of probabilistic approach. The problems are, it may be inappropriate because the real world scenario is changing. Therefore the future may not always be similar to the past. So we are taking A Priori probabilities which are based on the past information and we are going to predict the probability a-posteriori that is the futuristic probability. But that is valid as long as the world does not change. But in reality the world is changing. Therefore this may not be always appropriate.

The second thing is that it may be inexact or incorrect especially for subjective probabilities. If we compute the probabilities based on a large number of experiments that we do then there is no problem with that. But in many cases we cannot carry out so many experiments for all the rules. Therefore we often have to go by some subjective measures and subjective inputs and that is where the inaccuracies may creep in. So in order to apply the probabilistic reasoning rigorously we need to carry out a huge set of experiments to find out all the possible of conditional and joint probabilities and that is really an overhead because if there be n events then for all possible combinations we need to find 2^n number of probabilities.

The other thing is that knowledge may be represented implicitly. In that case we really cannot find out the probability corresponding to this. Just to avoid this overhead of finding out objective probabilities always and the huge amount of computation required we sometimes feel like bypassing this computational overhead. So we take recourse to some subjective measurements and have a quick but good enough solution using certainty factor approach. We have seen this is in MYCIN where it tries to avoid this overhead of the probabilistic approach or the Bayesian approach and take recourse to some subjective measurements or certainty factor algebra which assumed that all the probabilities which are dependent causally should be clubbed together in a rule. Therefore we will be dealing with only independent events then the computation becomes much simpler. However, Bayesian approach demands discussion in its own right.

In Bayesian approach the basic tenets or the steps of this we first derive the probability of a cause given a symptom. We get a symptom and we want to find out the probability of the cause. Although it required a lot of computational power now it is gaining more and more importance because we are having more and more powerful machines. So, more computational power is being available and better algorithmic methods are being available.

(Refer Slide Time: 11:18)

Bayesian approach is becoming more popular very much in the case of machine learning, information retrieval and language processing etc. The Bayesian approach is especially useful in diagnostic systems like medicine and computer diagnosis etc. Here the basic thing is, we are trying to find the inverse or a-posteriori probability. That is, inverse of the conditional probability of an earlier event given that a later one occurred. Now the Bayes' theorem needs a little bit of extension when there are a large number of hypothesis as we had presented earlier. So here let us see; let the evidence E be there with some probability.
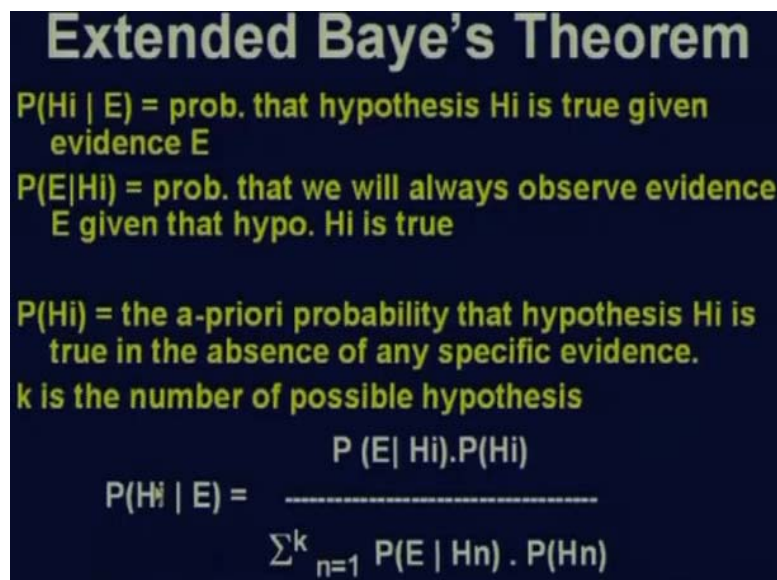
We want to find the probability of a hypothesis of a particular disease given a symptom. Given this evidence what is the probability of a particular hypothesis Hi?

So that is the probability that hypothesis Hi is true given the evidence E. And here we are having the A Priori probability known beforehand that probability will always observe the evidence E given that the hypothesis Hi is true. If the patient has malaria then always there will be fever.

What is the probability of that?
P(Hi) is the A Priori probability that the hypothesis Hi is true in the absence of any specific evidence. I go to a particular village and I straight away know that 60% of the people suffer from malnutrition. So, for that I did not really do a lot of statistics. If I have got that data I can start with that, that P(Hi) is probability that somebody in that village will suffer from malnutrition is 0.6. Now we have got an evidence E which can support a number of hypothesis such as n number of or k number of hypothesis. Then this evidence E and its contribution to a particular hypothesis is distributed over this entire k hypothesis.

(Refer Slide Time: 12:55)

Extended Baye's Theorem

$P(Hi \mid E)$ = prob. that hypothesis Hi is true given evidence E

$P(E \mid Hi)$ = prob. that we will always observe evidence E given that hypo. Hi is true

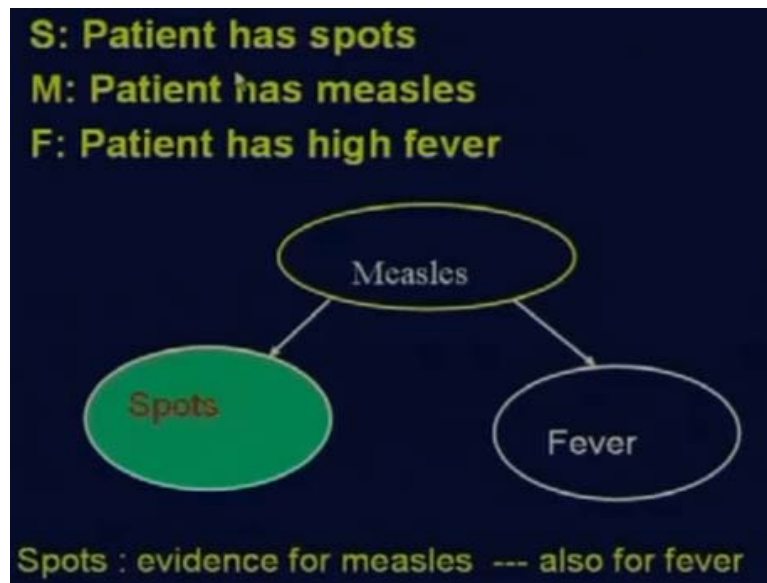$P(Hi)$ = the a-priori probability that hypothesis Hi is true in the absence of any specific evidence.

k is the number of possible hypothesis

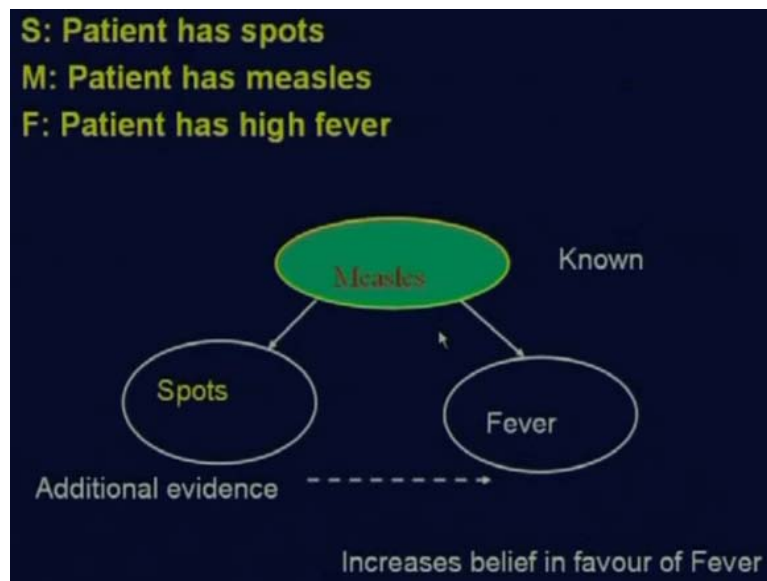$$P(Hi \mid E) = \frac{P(E \mid Hi).P(Hi)}{\sum^{k}_{n=1} P(E \mid Hn).P(Hn)}$$

So Bayes' rule or Bayes' theorem needs to be modified in this way P(Hi by E) is P(E by Hi) that was there already the A Priori probability of the evidence given the hypothesis times P(Hi) divided by, and here n varies from 1 to k because I am talking of k hypothesis then P(E by Hn) E by Hn times P(Hn) so earlier I was dividing at Bayes' theorem. I was just dividing by this evidence part or rather the B part.

Now in this case this will be divided by the entire possible hypothesis that I know of so I am coming to this notion so this is the general structure of Bayes' theorem. That is why we are calling it as extended Bayes' theorem. Next let us look at an interesting example which will reveal to you the interaction of the causes and effects. You will see that a particular cause leads to an effect or a particular cause can lead to two possible effects and there is a very nice interaction among them.

S: Patient has spots
M: Patient has measles
F: Patient has high fever

Measles

Spots

Fever

Spots : evidence for measles --- also for fever

S: Patient has spots
M: Patient has measles
F: Patient has high fever

Measles

Known

Spots

Fever

Additional evidence

Increases belief in favour of Fever

S means the patient has spots or rashes; M means the patient has measles and F is that the patient has high fever. Now look at this causal diagram. If the cause is measles and measles leads to spots measles leads to fever. Both of these are effects. Now, suppose I observe the evidence spots, now these spots although it is an effect of measles obviously it will serve as an evidence for measles but it will also serve as an evidence for fever though there is no direct causality here but the causality is coming in an indirect way because if somebody has spots then he or she may have measles and if he or she has measles then they may have fever. That is the way in which this is also serving as an evidence also for fever.

In the earlier case we had assumed that my evidence was spots, spots were known. Next I see that measles is known the disease is known that somebody has Measles. Now obviously if somebody has measles and if it is known then obviously we can

infer this evidence that he will have spots. And once we infer spots then that is also supporting fever because then this is again serving as additional evidence and increases the belief in favor of fever. If I know this then of course I know this. So what happened is, if I start with spots I know I can infer measles as a causal or I also increase the belief in terms of fever.

(Refer Slide Time: 18:18)

- If both Spots and Fever are present they provide evidence in favour of Measles
- But they are not independent !!!
- Hence, we cannot smply sum their effects – but take conditional probabilities of their conjunction into account

If both spots and fever are present they could be present independently. But if both of them spots and fever is present they provide evidence in favor of measles. But unfortunately if they were independent I could have combined my belief as spot leading to measles and fever leading to measles independently as something like this. If someone has measles he will have spots, if someone has measles he will have fever. So using Bayesian I know spots independently and I know fever independently so I have got two evidences e_1 and e_2 and both of them are pointing to hypothesis h that would be a joint probability which we could have done but unfortunately if e_1 and e_2 were independent. But unfortunately as we have seen in this case spots and fevers are not independent because e1 is related to h and h is also related to e_2 so there is an unseen dependence here and these are not dependent therefore I cannot straight away combine them which I could if they were independent. So we cannot simply sum their effects but we have to take conditional probabilities of their conjunction into account.

- In general, given a prior body of evidence e and some new observation E, compute

$$P(H \mid E,e) = P(H \mid E) \cdot P(e \mid E,H) / P(e \mid E)$$

In a real world the size of joint probabilities that we need to compute grows as 2^n if there are n propositions to be considered.

So in general given a prior body of evidence e and some new observation E we have to compute P(H). This is the new observation and the earlier observation was this. I knew that he had spots so I have inferred something about measles. Now I also know about fever that is denoted by e here. That means now I have to compute this conditional probability as P(H by E) times P, this is the dependence that if somebody has spots and measles then what is the probability of his fever?

I have to multiply with this and then divide with the conditional probability of fever when somebody has spots. I have to find out this conditional probability as well as this joint probability and in a real world the size of joint probabilities that we need to compute grows as 2 power n if there are n propositions to be considered and that is the real problem in applying the Bayesian approach. However, in order to facilitate our computation some intelligent techniques have been devised. Now we will talk of one such technique which is known as belief network.
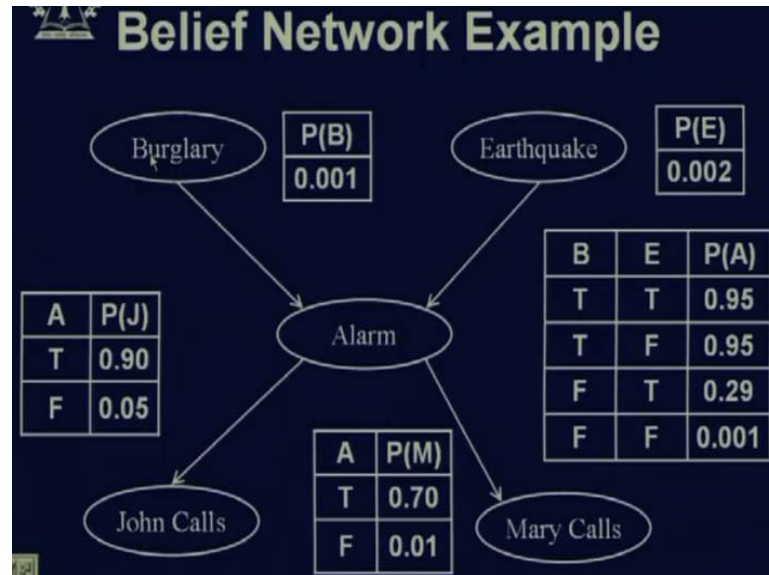
Belief Network

- Belief Network is a representation of causal connections between different events with associated conditional probabilities
- Nodes represent events
- Directed edge from node A to node B represents the fact that the event B *may* be an effect of event A.
- Conditional probabilities for a node is provided for each possible values of its parents.

What is a belief network?
A belief network is a representation of causal connections between different events with associated conditional probabilities. Here is a picture of a belief network.

(Refer Slide Time: 22:25)

All the nodes of a belief network are basically events and two nodes are connected by an edge if one node is the cause of the other. This is a typical example of a burglar's alarm case. The alarm this event has got a parent burglary that means burglary will cause the alarm, earthquake is another parent so earthquake will cause alarm and each of these edges show that there is a causal relationship. Therefore nodes are representing events. And that network is the representation of causal connections between different events.

Directed edge from one node to another node A to node B represents the fact that event B may be an effect of the event A. The usage of the word maybe has to be noted here. And we are storing conditional probabilities for a node and conditional probability of a node is provided for each possible value of its parents. That means, the conditional probability of A, think of this node John calls what is the probability of John calling? What is the parent of John calls? It is the alarm so corresponding to this node John calls in the belief network I am giving a least of probability table with a list of probabilities of John calling for every possible value of the alarm.

The alarm may ring that is true in that case John calls is 0.9 and if there is false it is 0.05. Here for this conditional probability it may not add to 1. With this definition now let us come to this example. There is a burglar's alarm fixed at home. Now this burglar's alarm is fairly reliable. That means whenever there is a burglary the alarm will ring. If a burglar breaks open your house it will start ringing. However, it is not 100% accurate. It also responds at times to minor earthquakes.

Whenever there is a minor earthquake and the door is shaking then also it starts ringing but that is a minor scenario. The other thing is that there are two neighbors John and Mary who on hearing alarm calls the police. Now John always calls when

the alarm rings. John is very particular about it. If the alarm rings John will call the police. But sometimes he confuses the telephone ring with the alarm and even if there is a telephone ring he thinks that it is an alarm and calls the police, that is his problem. And Mary has got the habit of listening to loud music and sometimes misses the alarm altogether, just misses the alarm. But if she gets the alarm then certainly she will call the police. That is the scenario. So here is the corresponding belief network burglary, earthquake, alarm, John calling, Mary calling which are the possible events and for each event we have made a node and for each of these nodes burglary we have put an A Priori probability of burglary as 0.001 and A Priori probability of an earthquake occurring is 0.002.

Now, alarm may be caused because of the burglary or because of the earthquake. So, if both of these things happen then the probability of alarm ringing is 0.95. If burglary occurs and there is no earthquake then also in 95% cases it will ring. If there is an earthquake then also sometimes it rings 0.29 and if none of them is there then also sometimes for spooky reasons it rings. This is the probability of alarm ringing with respect to these conditions burglary and earthquake. But here you see these two which have got no parents are A Priori probabilities.
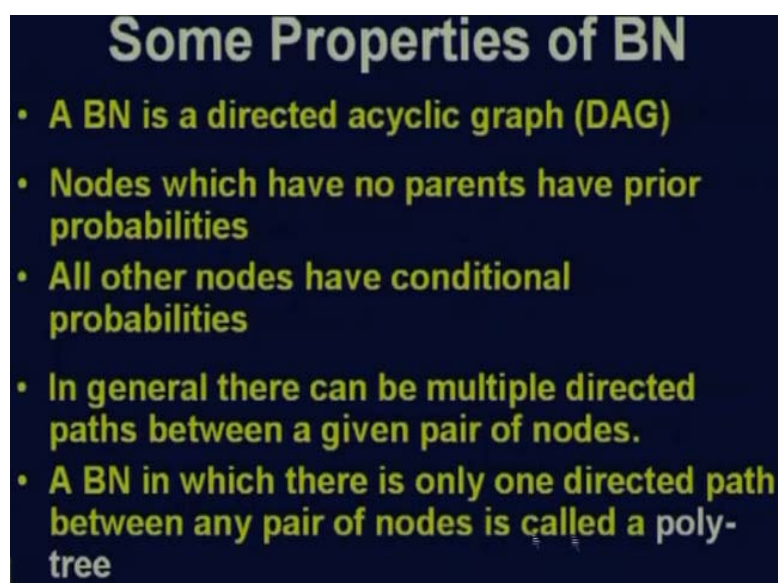
What is the probability of John calling?
If there is an alarm he calls but sometimes he also calls when there is a telephone ring so no alarm but he has called. And for Mary if there is an alarm and she really gets it she will call. But sometimes even if there is no alarm she calls. Therefore this is associated with this node and this is associated with this node and like that I have formed the belief network.

Properties of the belief network:
As we have seen a belief network is a directed acyclic graph or in short DAG there is no cycle in the graph as you have seen here because in that case there would be a problem.

(Refer Slide Time: 27:48)

Some Properties of BN
- A BN is a directed acyclic graph (DAG)
- Nodes which have no parents have prior probabilities
- All other nodes have conditional probabilities
- In general there can be multiple directed paths between a given pair of nodes.
- A BN in which there is only one directed path between any pair of nodes is called a poly-tree

If John calling again starts burglary or starts alarm then there would be a problem and it will be much more difficult to handle. Nodes which have no parents have prior probabilities as we have just now shown. The nodes which have got no parents are given A Priori probabilities. All other nodes have been given conditional probabilities. In general there can be multiple directed paths between a given pair of nodes. It is a possibility that there will be a multiple path from one node to another but that case will not consider. We will consider the case when there is only one directed path or only one path between any pair of nodes and that scenario is known as a poly-tree.

(Refer Slide Time: 29:44)
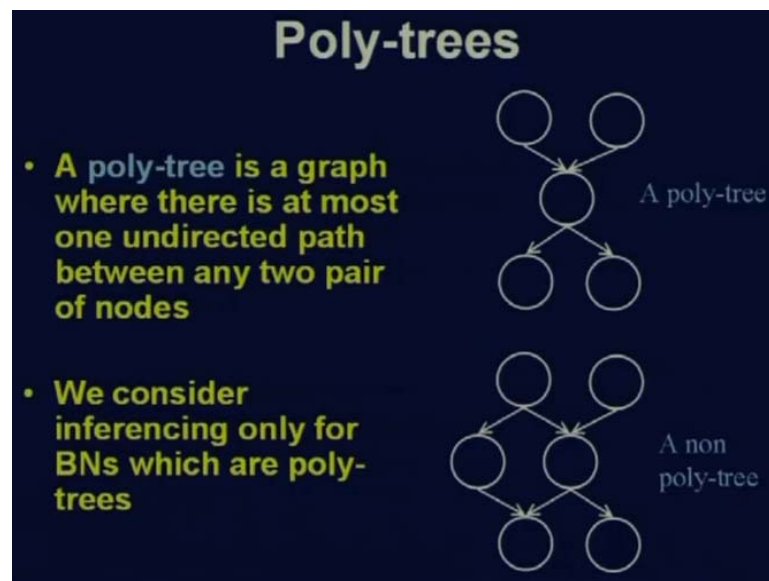
What is a poly-tree?
A poly-tree is a graph where there is at most one undirected path between any pair of nodes. Between these two nodes there is only one path either here or here and why is it undirected?
From this node to this node also there is only one path whereas this is not a poly-tree because from this node to this node I have got two paths. Here of course there is only one path but from this node to this node I have got two paths like this. So this is a poly-tree and this is not a poly-tree. We will restrict our cases only where it is a poly-tree.

Given this scenario you can see that it is a poly-tree. What is the probability that the alarm has sounded? But neither a burglary nor earthquake has occurred and both Mary and John has called. In this case I want to find the probability that John has called, Mary has called but burglary and earthquake none of these things occurred. Now obviously John has called because is it just when there was no alarm or John really heard that there was an alarm, what is it? Was there an alarm or it was a false thing? None of these things occurred so how would I reason out this?

I am trying to model this scenario. Probability that there was no earthquake and there was no burglary and there was alarm, the alarm sounded, John called and Mary called.

This is the probability that I need to compute and that is equal to probability of John calling on alarm multiplied by probability of Mary calling on alarm.
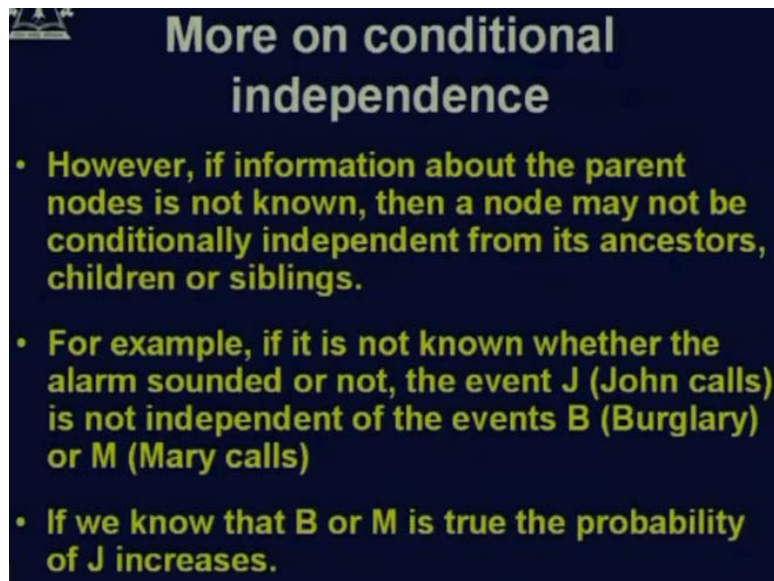
(Refer Slide Time: 32:44)

Probability of alarm ringing was when there was no earthquake and there is no burglary and A Priori probability that there is no earthquake and the A Priori probability that there is no burglary. Now in this graph poly-tree we have got all these values so we will compute this from the same formula. We can see that probability of John calling on alarm was 0.9, probability of Mary calling on alarm was 0.7, and probability that the alarm has occurred when there was no burglary and there is no earthquake is this 0.001. This figure I have to take and times this, when I take this joint I have to take 1 minus 0.002. There is a probability of NOTE because the probability of E was 0.002 and probability of burglary was 0.001 so the probability of NOT burglary is 1 minus 0.001 that is the basic of probability. So we find the value as 0.00062 that is how we compute in the case of using the belief network.

(Refer Slide Time: 34:11)

Now if the information about the parent node is not known then a node may not be conditionally independent from its ancestors. For example, in general in a belief network I will consider the conditional probability for the nodes which has got some causal relationships and those which are not related I will not consider them.

For example, if I am in a scenario where I know that burglary has got some relationship with alarm and earthquake has got some relationship with siren or some other thing which is not related then in this graph I will just restrict myself only to the nodes where there is a conditional dependency. For example, John calls etc and I will not consider these nodes here. But if I do not know anything that burglary occurred or John called if for this event I do not have any information about its parent then any of these can be participating in the decision making.

Now our problem was that in the case of Bayesian approach we have to compute a huge number of probabilities. So, in a belief network there are so many nodes and edges. Which of the parts we have to really look into and take into our computation?
In order to reduce the size of computation we can very intelligently restrict ourselves to a subset of the belief network. In order to appreciate that, first let us look very carefully into these definitions. A set of nodes E, some set of nodes E d-separates two set of nodes X and Y if every undirected path from a node on X to a node in Y is blocked given E.

Example: d-separation
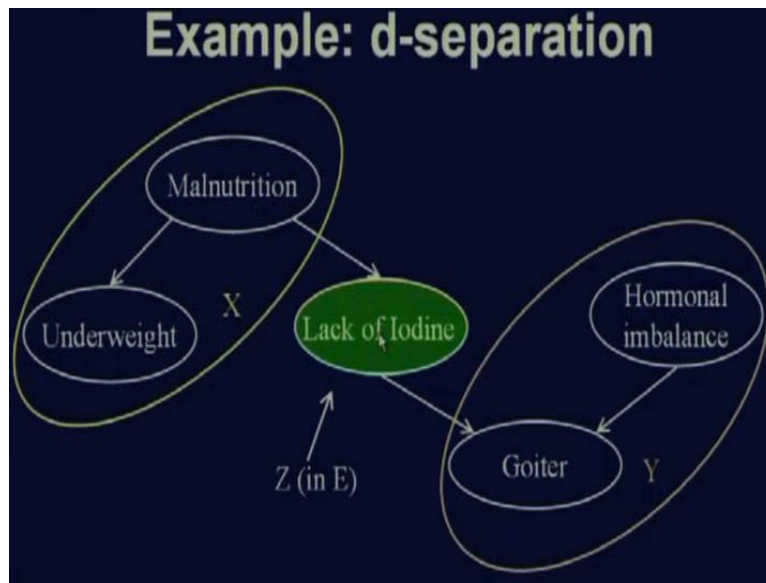
Conditional Independence Relations

- If every undirected path from the nodes in X and nodes in Y is d-separated by a given set of evidence nodes E, then X and Y are conditionally independent given E
- A set of nodes E d-separates two set of nodes X and Y if every undirected path from a node in X to a node in Y is blocked given E.

That means here is a belief network, malnutrition leads to underweight, malnutrition leads to lack of iodine, lack of iodine leads to goiter and hormonal imbalance also leads to goiter. Now, this node is very critical because this node d-separates this node and this node. Suppose if in this node I know that there is lack of iodine then in order to reason whether he has got goiter or not it is sufficient for me to look into whether hormonal imbalance is there and I need not look into malnutrition. But unless I know this I could have tried with goiter whether he has got malnutrition and from there I would have started to reason. So I would have needed more conditional probabilities to compute. But in this case I can simply separate out between two sets of nodes X and Y which are basically d-separated by this node lack of iodine.
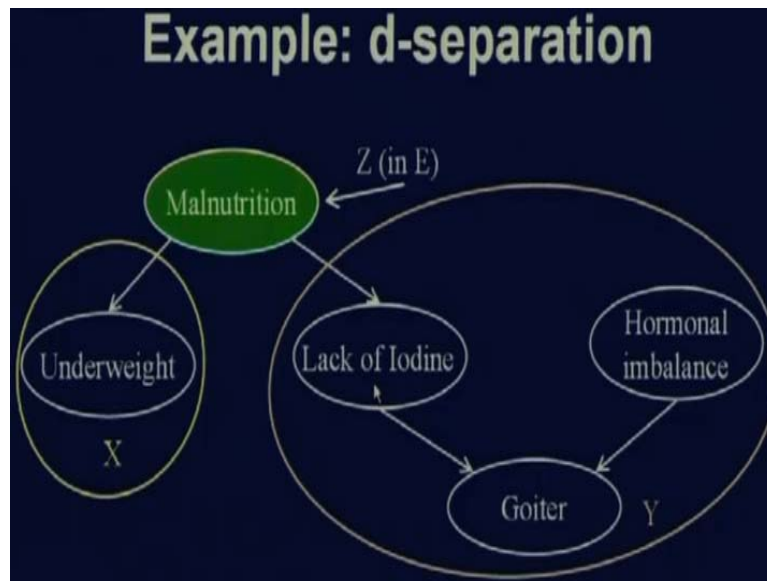
Another example is:
If I know that somebody has malnutrition then for reasoning about underweight I need not look into the probabilities of lack of iodine, hormonal imbalance or goiter. Similarly, for finding goiter I need not take underweight into consideration. So this node d-separating is making two clusters this one and this one. Therefore lack of iodine is again clustering these.

Definition:
A set nodes d-separates, in this case lack of iodine was the case or malnutrition was the case which d-separates two sets of nodes X and Y if every undirected path from X to Y is blocked or goes through E as was the case here. Every node from this set was going through this. There was no node which was by-passing this.

When do we say it is blocked?

A path is blocked given a set of nodes E if there exists a node Z in the path for which one of the following holds: Z is in E and Z has one arrow on the path leading in and one is leading out is the case. This is node which is in E that I know and that is a blocking node. So there is one edge coming in and one edge going out which is the first condition. Another case is Z so what is that?

Z is a node that is given a set of nodes E Z is in E and Z has both path arrows leading out, this is the example. Z is in E, in my evidence set Z is there, a particular node is there and all of them are going out from there. So obviously the descendants of this are separate. The third case is neither Z nor its descendants are in E and both path arrows lead into Z. This is the evidence.

Although this is the evidence but the blocking node is this Z which is separating out Y both the path arrows from these clusters it is coming into the goiter. So goiter is blocking out these two so there is no dependence between these. This is a nice idea of d-separation and therefore I can restrict myself to the computation of the probability with respect to the d-separated clusters only.
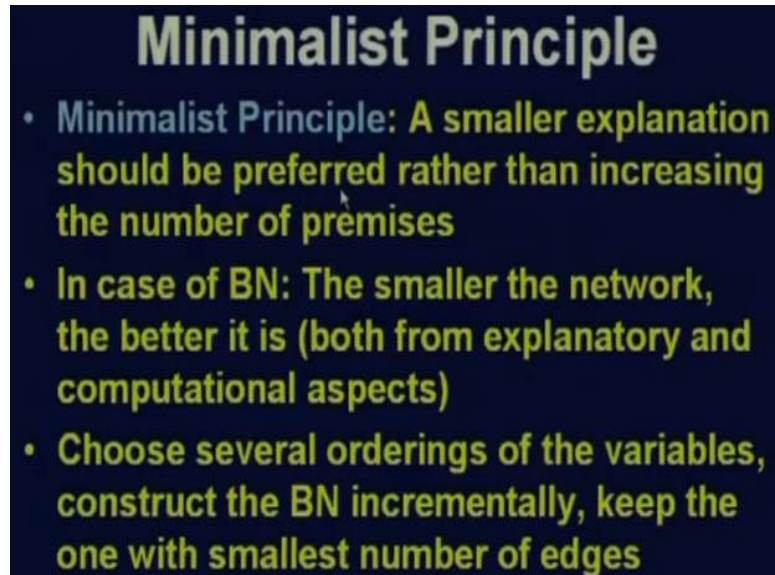
(Refer Slide Time: 41:55)

Now it is very important to learn how we build in the belief network?

When we build the belief network we have got a number of variables. We have to select the variables in such a way that we will get a node or it will be possible to create d-separated clusters. For our reasoning the computation of conditional probability becomes less so we choose the set of variables that describe the domain like all those basic nodes such as malnutrition, goiter, lack of iodine or for burglars alarm you can think of burglary or earthquake, alarm, John calls, Mary calls are the variables I have to define.

Another very important step is to choose an ordering for the variables because I will first pick up a variable X and a node for it and set the parent of X to some minimal set of existing nodes such that conditional independence is satisfied. And then I will go on doing this and in this way I will define the probability table for X. So, given this I

have to select from the possibilities. I will first select one node namely lack of iodine then I have the select parent in such a way so that I can create a conditional independence set and then we define the probability of this. Therefore ordering is much important and this approach is based on the minimalist principle.

(Refer Slide Time: 44:18)

There is a smaller explanation. This is a very general philosophical term and is also applicable to AI. A smaller explanation should be preferred rather than increasing the number of premises. For example, X has happened because A has happened, B has happened, C has happened, D has happened etc. Instead, if we could say K and P happened and therefore A happened. So, that is the minimalist approach and that is exactly what we are trying to do that smaller the network the better it is both from explanatory and as well as from computational aspects. So, in order to really find a good belief network we have to choose several orderings of the variables and construct the belief network incrementally and ultimately keep the one with smallest number of edges. This is a very important step.

(Refer Slide Time: 45:45)

How do we infer in belief network?

The inferences in belief network can be divided into four categories. One is diagnostic inferences. That means from effects we want to go to the causes. Suppose given that John calls what is the probability of burglary?

So the evidence was John has called and what is the probability of burglary that is the diagnostics so I have seen the effect. My car is not starting. What is the probability? There is some problem in the spark plug. This is what we are trying to do so it is diagnostic.

What is the probability of burglary given John Calling?

The other thing is causal inferences from causes to effects. Given burglary what is the probability that John will call or Mary will call, there is a probability of John calling given B. For example, if the spark plug of my car gets dirty then what is the probability that it will not start? That is causal. The third type is intercausal inference. One common event can be given by multiple causes.

(Refer Slide Time: 46:55)

For example, alarm could be caused by burglary as well as earthquake. Given alarm what is the probability of burglary? But alarm can be caused by both earthquake and burglary from there if one is known then what is the probability of the other one?

Given alarm what is the probability of burglary? It is that I know the probability of burglary given alarm.

Now given earthquake what is the probability of burglary? Alarm has been there so probability of A and E, what is the probability of B? The other thing is mixed inference. Some things are causes and some of the effects are known. For example, I know John has called and I know that is an effect and I also know that no earthquake has occurred that is the status of some cause I know so what is the probability of alarm? There is a probability of alarm when John has not called and there is no earthquake. So, figuratively we can see that there are four patterns.

(Refer Slide Time: 48:18)
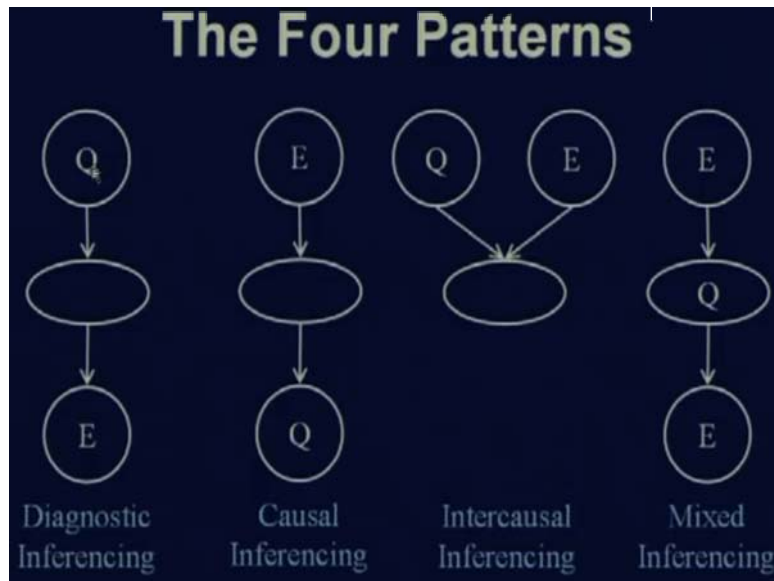
This is the diagnostic and this evidence is known. What is the probability of a particular cause to have occurred?

Causal is, I know that the voltage has increased, what is the probability that the current flow will increase?

That is probability one but I know the cause that with some probabilities burglary has occurred and what is the probability that John will call?

The third is intercausal.

Something has happened and I know there are multiple causes I know some probability of earthquake occurring and I also know the probability of alarm being sounded on earthquake.

What is the probability that burglary has also occurred?

So there are multiple causes and I want to find this one. The other one is mixed inference. That is, I know some of the causes and some of the effects.

I know that John has called, I know that there has been no earthquake, what is the probability of burglary?
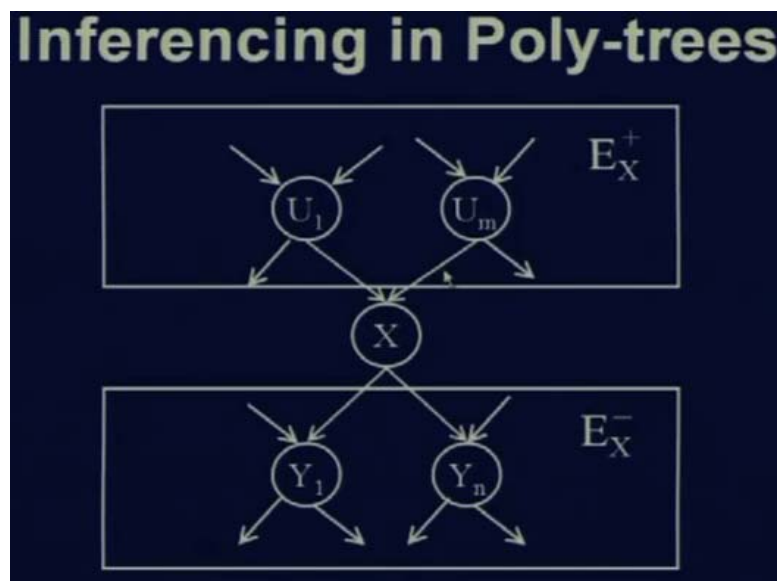
This is mixed inference. So there are four patterns. Now poly-tree we have already explained.

(Refer Slide Time: 49:58)

(Refer Slide Time: 50:18)

Let us consider X to be a query variable, we want to know the probability of X given an evidence E. So I know the probability of E and I want to know the probability of X by E. And X is my query so I want to know the probability of X by E and suppose there are a set of nodes which are parents of X and a set of nodes which are children of X so I want to know the probability of this. I can separate out, there can be some parts which are neither parent nor children of X, X is the query variable whose conditional probability I want to know so we separate out these two so I call one to be E_X plus and another to be E_X minus these two sets. These are the effect set and you can say this is the cause set.

So E_X plus is the set of causal support for X that the variables which are connected through its parents which are known whose status and probabilities are known and E_X minus is the set of evidential support for X that is variables below X are connected through its children.

Therefore now I can compute P(X by E), P(X) I want to compute given E_X plus and E_X minus because other parts of the poly-tree are independent. How you could I find out which are the independent parts because I have found out some nodes by which I could d-separate the entire belief network. Now I have restricted myself only to a cluster after d-separation. The other part is not important. So I have reduced the number of variables. Therefore I have also reduced the number of parents and number of children corresponding to a particular node X. So I have got a smaller set of E_X plus and a smaller set of E_X minus. Now we can simply apply Bayes' theorem that

P(X by E) can be computed as P(E_X minus by X) and E_X plus times P(X by E_X) plus by P(E_X minus by E_X plus) that is standard Bayes' theorem.

(Refer Slide Time: 52:52)
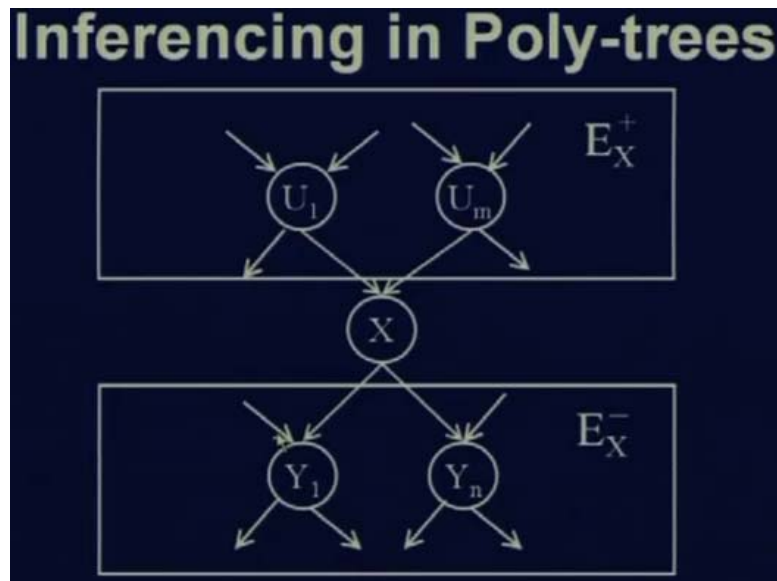
Since X is d-separating of E_X minus from E_X plus here you see X is also d-separating these two clusters because all of them are coming through X. Therefore I can further simplify it from this Bayes' theorem part. We can say P(X, E) will be, we just take the numerator P(E_X minus by X and P(X by E_X plus) and we multiply it with alpha where 1 minus alpha is a constant representing the denominator because they are d-separated. E_X plus and E_X minus are d-separated so I can find out the independent A Priori probabilities useful for this. So basically if they are independent then probability of E_X minus by E_X plus is basically the P(E_X minus) if they are independent and then only this part is there and this independence is coming from this d-separation.

(Refer Slide Time: 53:55)

Now both the terms P(X by E_X minus) and P(E_X plus by X) can be computed recursively using the conditional independence relations. That is a simple way of dealing with the thing. We have seen four types of inferencing. Now we will see an example of casual inferencing.

What does it mean?
I know that there has been a burglary, I know the cause. I want to find the probability of the effect. So that is my query. So this green part, burglary I know has occurred. So that is my evidence and I want to know the probability, with which John called, how would I compute this?
Obviously I will have a d-separated graph and I will compute it here out of this.

(Refer Slide Time: 55: 20)
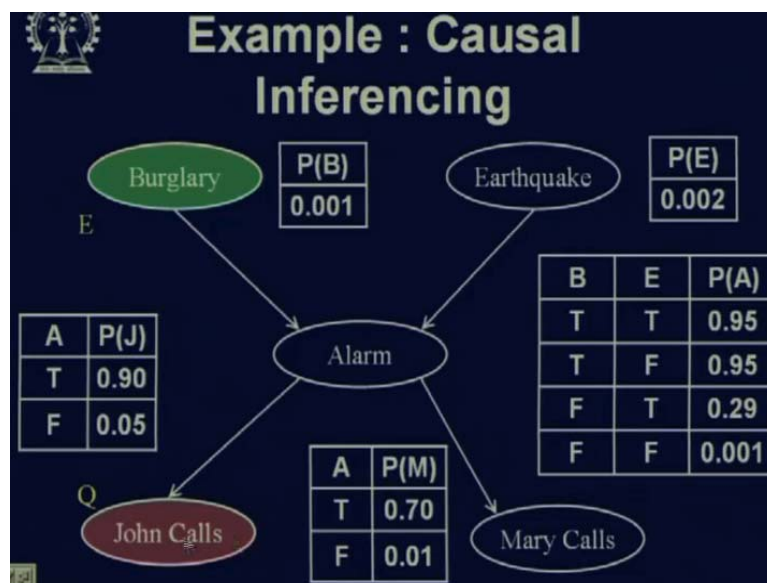
What is E_X minus and what is E_X plus?
E_X minus is, for this query node there is no children. This is my query node so this is my node X so there is no children. So E_X minus is null. And what is E_X plus? J is in E_X plus. John called is in E_X plus. So we need to compute probability of John calling given burglary. Now this alarm is d-separating B and J. So, using the Bayes' theorem we can write A can be true or false.

(Refer Slide Time: 55:55)

So probability of John calling given burglary is what I want to know. Is probability of John calling NOTA given no alarm?
Probability of no alarm given B given there is a burglary plus the probability of John calling given the alarm and the probability of alarm given burglary. So, I have restricted it to three nodes of the graph and I can compute it as 0.858. Even after this reduction there may be considerable amount of computation.

The fundamental thing is apply Bayes' theorem, the Bayesian approach. Try to find out that d-separated part. And only restrict yourself to this E_X plus and E_X minus sets and compute the Bayesian probabilistic value based on these nodes. That is the basic approach of belief networks or Bayesian networks. Here is an example of causal reasoning, causal inferencing. Here the scenario is John has called so it is diagnostic.

What is the probability of burglary?
That is the thing we want to find out. Here you get a confidence about applying Bayesian networks for managing uncertainty. These are some of the approaches. So we have seen certainty factors and we have seen in quite detail the probabilistic approach and the Bayesian approach.