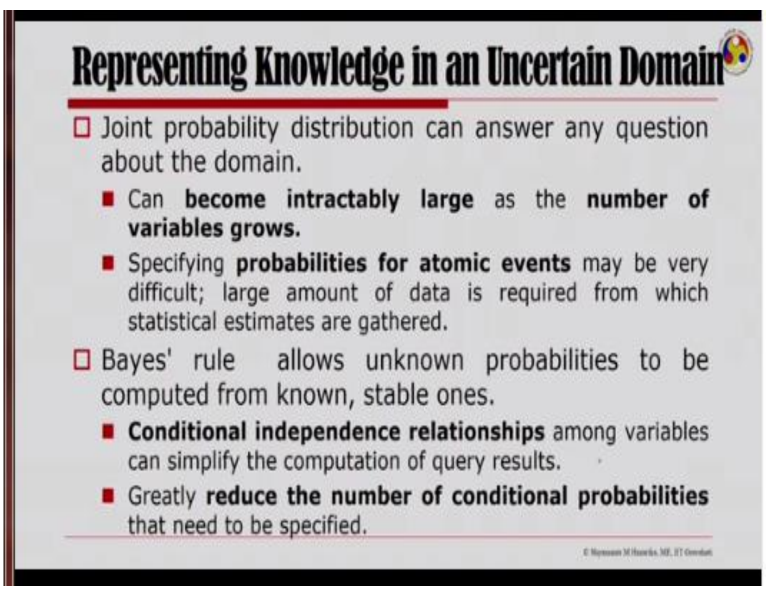
Fundamentals of Artificial Intelligence
Prof. Shyamanta M Hazarika
Department of Mechanical Engineering
Indian Institute of Technology- Guwahati

Lecture-19
Bayesian Network

Welcome to fundamentals of artificial intelligence, we continue our discussion on reasoning under uncertainty. In the last lecture, we have looked at the syntax and semantics of probability theory. Today we would introduce Bayesian networks, Bayesian networks or belief networks are the key technology to deal with probabilities with an artificial intelligence. We discuss how to capture uncertain knowledge in a natural and efficient way, how to arrive at the Bayesian networks, that is their syntax.

And look at how to interpret the encoded knowledge that is the semantics of the Bayesian networks. We start our discussion by quickly reviewing the characteristics of representing knowledge in an uncertain domain.

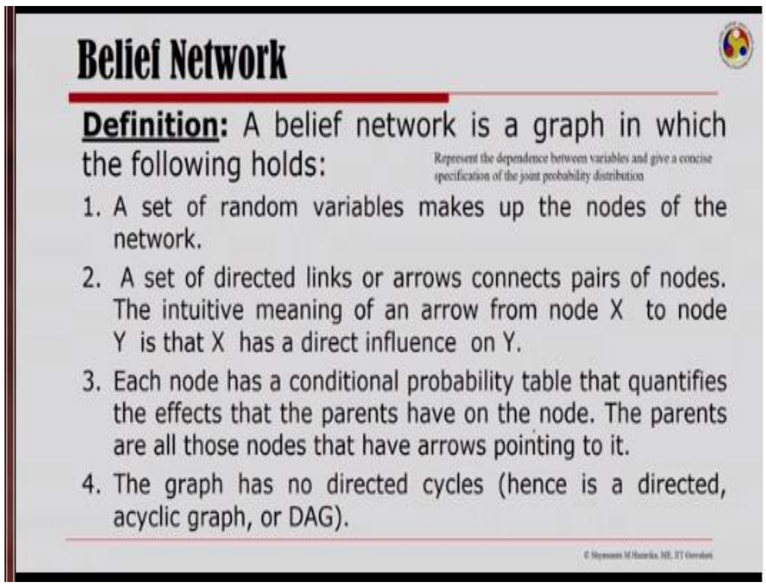
(Refer Slide Time: 01:48)



In the last lecture, we have seen that joint probability distributions can answer any question about the domain. However, it is important to realize that these joint probability distributions can become interactively large as the number of variables grow also specifying probabilities for atomic events may be very difficult. One would need large amount of data from which statistical estimates are to be gathered.

Bayesian’s rule allows unknown probabilities to be computed from known stable ones and we have seen in the last lecture, how conditional independence relations among variables, simplify the computation of query results.

(Refer Slide Time: 02:48)



Today we start our discussion on belief networks by giving a definition one needs to understand here that belief networks or Bayesian network represent the dependence between variables and gives a concise specification of the joint probability distribution. A belief network is a graph, in which the following 4 characteristics hold one a set of random variables make up the nodes of the network. Thereafter we have directed links or arrows that connects pairs of nodes. Now the intuitive meaning of an arrow from node x to node y is that x has a direct influence on y, the third characteristic is about each node having a conditional probability table that quantifies the effect that the parents have on the node. Now the parents are all those nodes that have arrows pointing to it, finally a belief network is a directed acyclic graph that is the graph has no cycles.

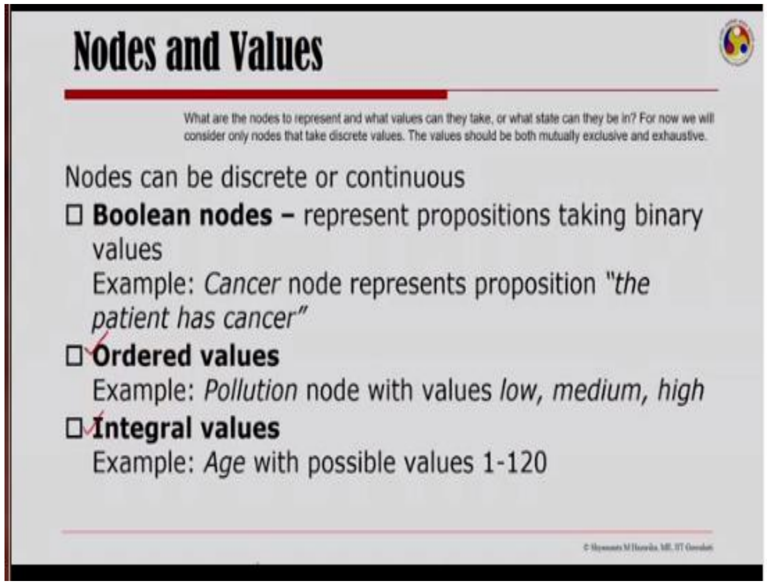
(Refer Slide Time: 04:15)

In order to understand a belief network and illustrate what it is nodes are and how we go about looking at the semantics, we would take help off an example here. This example is from the book by Corbin Nicholson Bayesian artificial intelligence from it is second chapter, so it is about lung cancer diagnosis. A patient has been suffering from shortness of breath, dyspnoea and visits the doctor worried that he has lung cancer.

The doctor knows that other diseases such as tuberculosis and bronchitis are possible causes as well as lung cancer for shortness of breath. She also knows that other relevant information includes whether or not the patient is a smoker, increasing the chances of cancer and bronchitis and what sort of air pollution the patient has been expose to. Thereafter, a positive x-ray would indicate either tuberculosis or lung cancer.

Now under this given scenario, if someone has to make a diagnosis, then there are a number of probabilities involve. When we are creating a Bayesian network out of this scenario we need to first understand what would be the different nodes here in this example shortness of breath or dyspnoea could be one node. The other could be that the patient has cancer now there are other informations which influence these two like the fact that he is a smoker or not a smoker. The air pollution that the person is expose to and finally the result of an x-ray, these could be the initial choices for the nodes.

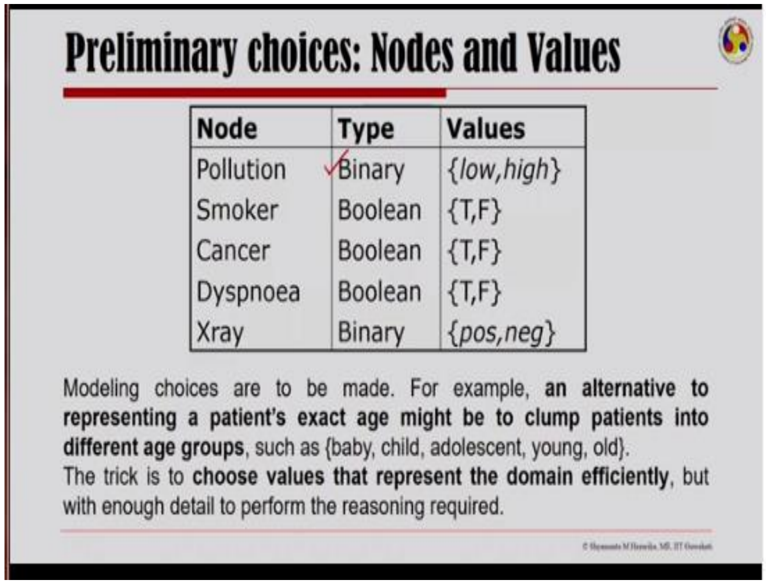
(Refer Slide Time: 06:29)



So these nodes represent and what values they take or what states they can be in is what is our concern here. For this example we will only consider nodes that have discrete values now the value should be both mutually exclusive and exhaustive. Nodes can be discreet or continuous we could have Boolean nodes like Boolean nodes represent propositions taking binary values. For example in the illustrative lung cancer diagnosis example we are talking off.

The cancer node could represent the proposition the patient has cancer and it could be either true or false. The ordered valued node is one which allows a range of values for the particular node, for example here the pollution node could have values low, medium and high and could be an ordered value node. I could have nodes that have integral values like if I was somehow also capturing the age of the person who is involved in this example. Then the age could be a with possible values between some 1-120 and I could think of that as an integral value node.

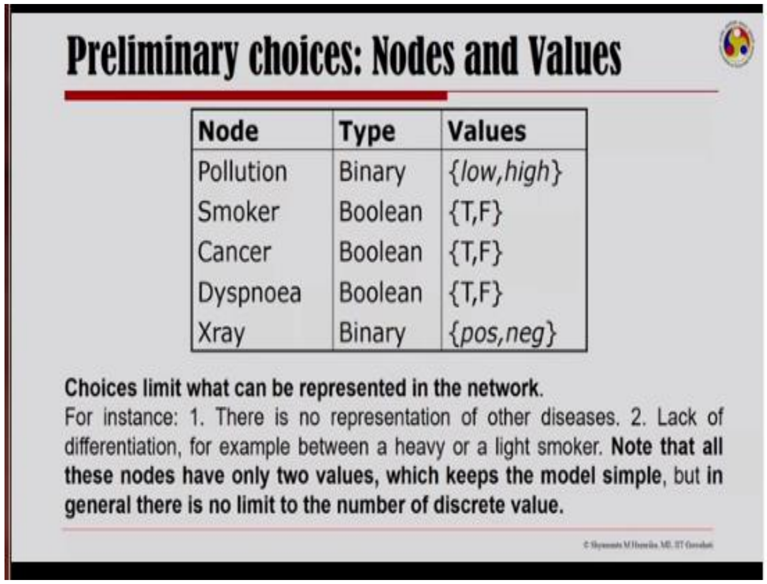
(Refer Slide Time: 08:13)



For our example, the preliminary choice of node and it is values could be that I could think of pollution as a node that has 2 values low and high. Therefore it is a binary node, I could think of smoker as a Boolean node somebody could be a smoker or somebody is not a smoker. Thereafter cancer as already discuss is a Boolean node, dyspnoea is a Boolean node which is about true and false and x-ray again is our binary node which have positive or negative values.

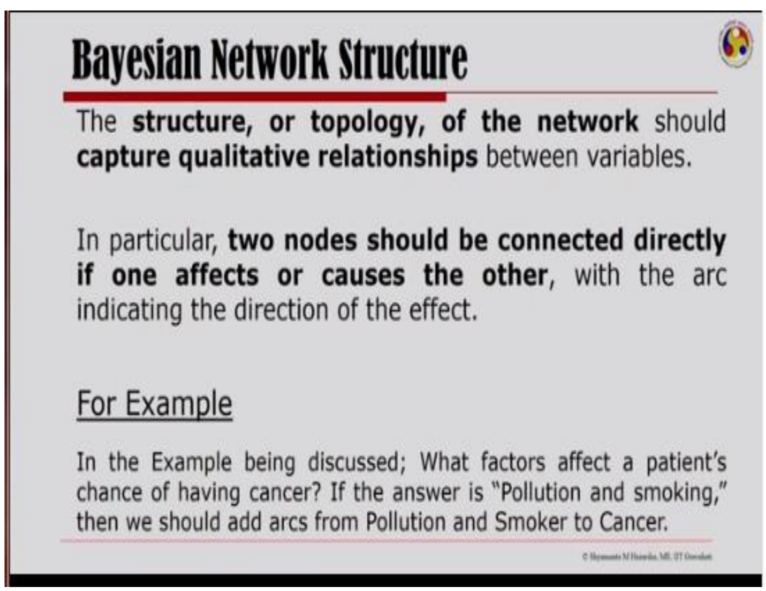
Now, one needs to understand this that the modeling choices are to be made before we start creating the Bayesian network. We could have alternate representations and alternative for example to representing the exact age might be to club patients into different age groups such as I could talk off baby, child, adolescent, young and old. Now whenever I am talking of choices for nodes and it is values, the trick is to choose those values that represent the domain efficiently but with enough detail to perform the reasoning required.

(Refer Slide Time: 09:48)



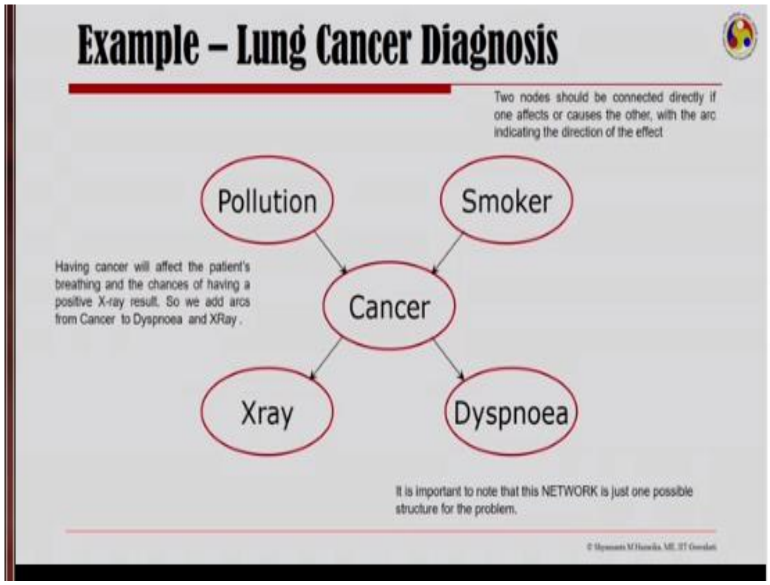
Now, one should also realize that such choices that we make or the choices that we pick up, limit what can be represented in the network. For instance in this choice that we have made for our self taking pollution, smoker, cancer, dyspnoea and x-ray as the nodes, there is no representation of other diseases. Further we have not made any differentiation between, for example a heavy smoker and a light smoker all these nodes that we have picked up have only 2 values which keeps the model simple. But one should realize that there is no limit to the number of discrete values that I can pick up.

(Refer Slide Time: 10:39)



So let us now try to understand the structure or the topology of the network. The structure or the topology of the Bayesian network should capture the qualitative relationship between the variables, in particular 2 nodes should be connected directly if one affects or causes the other with the arc indicating the direction of the effect. For example, what factor affects a patient's chance of having cancer in this lung cancer diagnosis example. If the answer is pollution and smoking then we should have arcs from pollution and the smoker node to cancer.

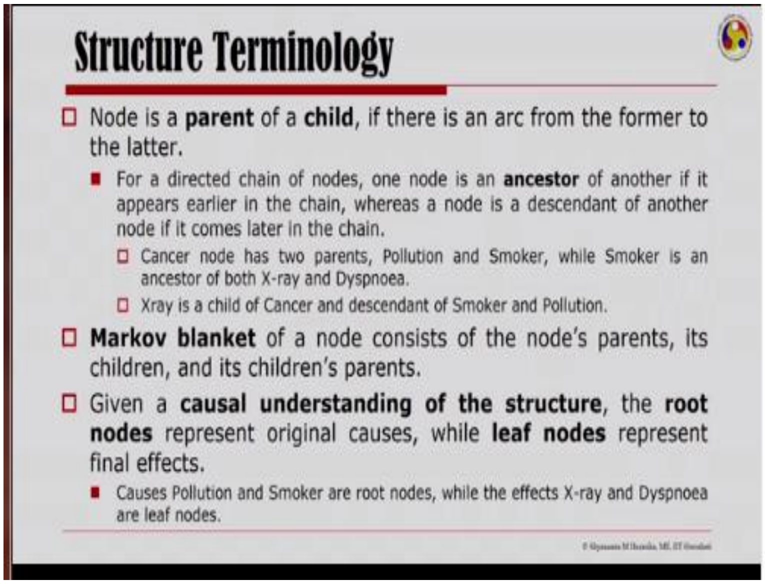
(Refer Slide Time: 11:26)



So here is a network that is representing the lung cancer diagnosis now it is important to note that this network is just one possible structure for the problem. Now here as discussed we have 2 nodes the smoker node and the cancer node now they are connected directly. Because one affects or causes the other and the arc indicates the direction of the effect having cancer will affect the patient’s breathing, he will have shortness of breath and you also have chances of having a positive x-ray.

So we add arcs from cancer to dyspnoea and to x-ray, so this is one Bayesian network for the lung cancer diagnosis example that we have been introduced.

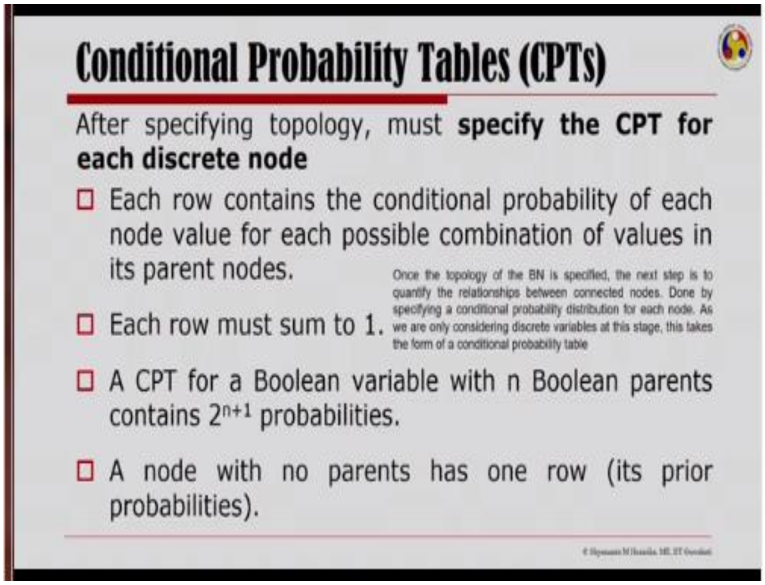
(Refer Slide Time: 12:29)



Now let us get to certain terminologies of the structure, the first of them is a parent, node is a parent of a child if there is an arc from the former to the later. Now for a directed chain of nodes one node is an ancestor of another if it appears earlier in the chain and we have a concept of a descendant, where a node is a descendant of another node if it comes from later in the chain. So if you recall the Bayesian network that we have introduce.

The cancer node has 2 parents pollution and smoker, while smoker is an ancestor of both x-ray and shortness of breath, x-ray is a child of cancer and descendant of smoker and pollution we have a term called the markov blanket. Now the markov blanket of a node consist of the nodes parents it is children and it is children's parents, so given a causal understanding of the structure the root nodes represent the original causes, while the leaf node represents the final effects. Now in this example, pollution and smoker are the route nodes while x-ray and dyspnoea are our leaf nodes.

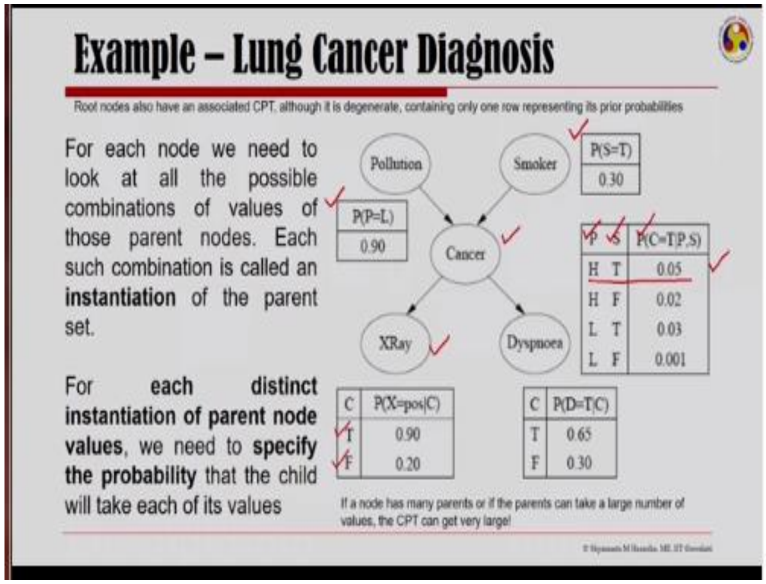
(Refer Slide Time: 14:03)



After specifying the topology one must specify the conditional probability tables for each discrete node. Once the topology of the Bayesian network is specified you need to quantify the relationship between connected nodes. This is done by the conditional probability distribution now we are only considering in this example, discrete variables. So this takes the form of the conditional probability table.

Each row within the conditional probability table contains the conditional probability of each node value, for each possible combination of values in it is parent nodes and each row must some to one. A conditional probability table for a Boolean variable with n Boolean parents contains 2 to the power n + 1 probabilities. A node which do not have any parents has just 1 row its prior probabilities, for without parents we do not have any conditional probabilities possible for that particular node.

(Refer Slide Time: 15:24)



So here is the lung cancer diagnosis a Bayesian network with the conditional probability tables. If you look at the root nodes here, the smoker node and the pollution node all they contain is 1 row representing its prior probabilities. So in a sense, the conditional probability table is degenerate here it is saying that the probability that the person is a smoker is 30%, pollution level being low is 90%, for each node, we need to look at all possible combinations of values for the parent nodes.

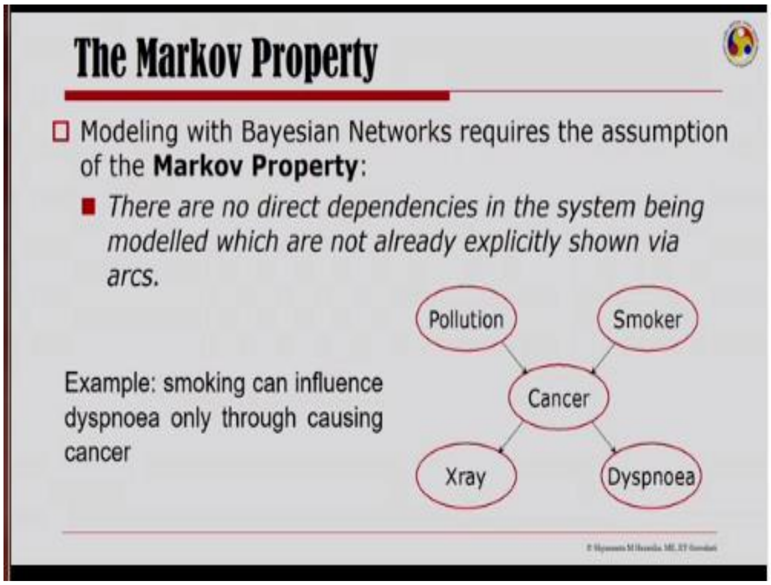
So such combinations is called an instantiation of the parents set, so in this example, for the cancer node we have the pollution and the smoker, as it is parents set and this combination is called an instantiation. So we have pollution, we have smoker and we have these possibilities. Now, one needs to remember that pollution we had possibility of being high and low and smoker was a Boolean node, so we had possibility of being true and false.

So we look at each combination and which is pollution high, smoker true, pollution high, smoker being false, pollution low smoker being true and pollution low smoker being false. So we look at this probabilities of cancer being true given the 2 parents pollution and smoker, for each distinct instantiation that we have created of the parent node values we are specifying the probability that the child will take for each of it is values.

So this is what the conditional probability table gives us. So here in this case, for the case that this is H and T, I have specified a probability which is 0.05 of cancer being true and I have done this for all instantiations of the parents set. Similarly if we look at one of these here, x-ray I could see would depend on cancer being true or false and I could then write the probabilities of x-ray being positive given cancer.

So there is a 90% probability that given cancer, the x-ray is positive and 20% even if the cancer the x-ray could be positive. So one needs to realize that if a node has many parents or if the parents can take a large number of values then the conditional probability tables that I have for each node can get very large.

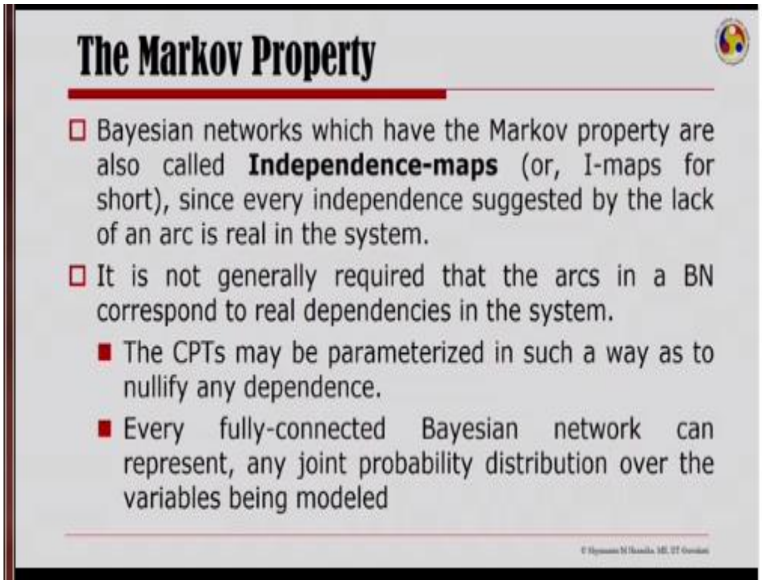
(Refer Slide Time: 19:05)



So the modeling with Bayesian networks requires the assumption of the markov property, the markov property states that there are no direct dependencies in the system being model which are not already explicitly shown via the arcs. So whatever is the influence of one node on the other I have to explicitly show it using the arcs in the Bayesian network. So for example in our case we saw that smoking can influence shortness of breath only through causing cancer.

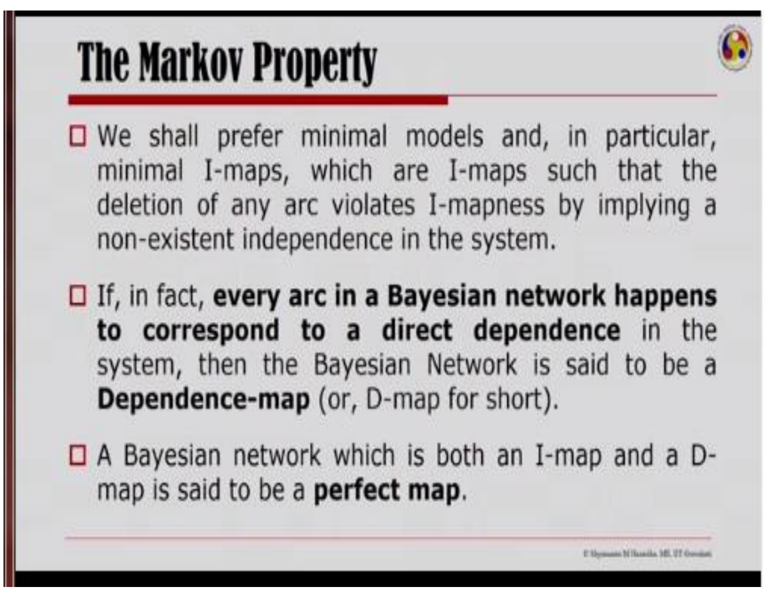
And therefore in our example, we have a link from smoking to cancer and then to shortness of breath but there is no direct link from here because this direct link does not exist. So there are no direct dependencies in the system being model which are not already explicitly shown via this arcs. So here we could say that smoking can influence shortness of breath only through causing cancer and that is there in the Bayesian network.

(Refer Slide Time: 20:22)



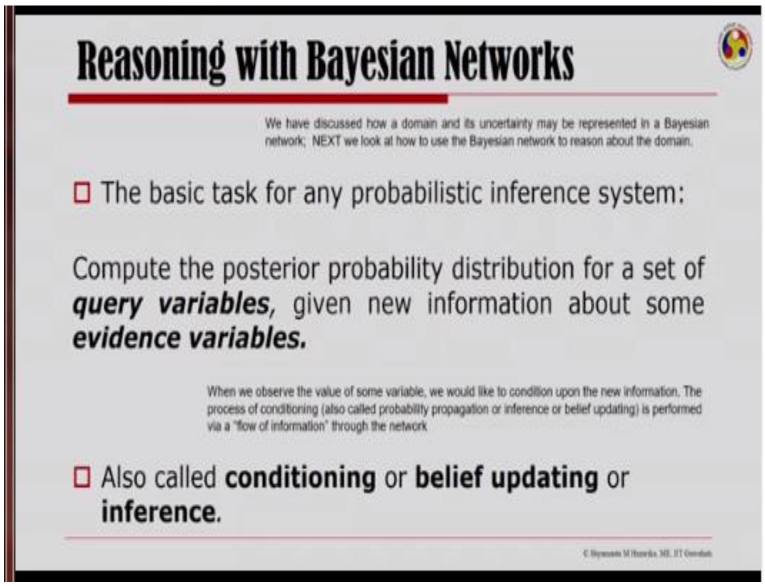
So Bayesian networks which have the Markov property are also called independence maps or I-maps since every independence suggested by the lack of an arc is real in the system. It is not generally required that the arcs in our Bayesian network correspond to real dependencies in the system. The conditional probability tables maybe parameterized in such a way as to nullify any dependency. So every fully connected Bayesian network can represent joint probability distribution over the values, that is being model.

(Refer Slide Time: 21:07)



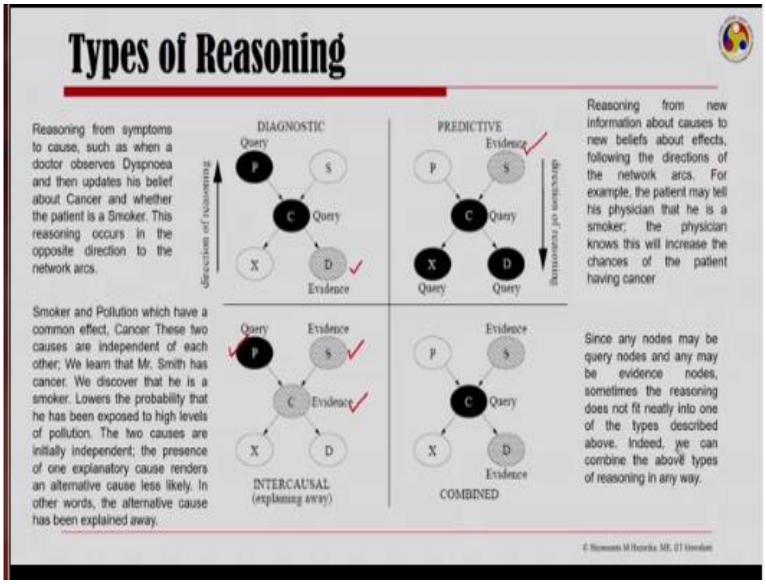
We shall prefer minimal models or what are called minimal I-maps such that if you delete any arc that violates the I-mapness by implying a non-existent independence in the system. If in fact every arc in a Bayesian network happens to correspond to a direct dependence in the system then such a Bayesian network is said to be dependence-map or D-map. Now a Bayesian network is a perfect map if it is both an I-map and a D-map that is a Bayesian network which is both an independence map and a dependence map is said to be a perfect map.

(Refer Slide Time: 21:56)



So reasoning with Bayesian networks is about computing certain probabilities for a set of query variables given new information about some evidence variables. We have discussed up till now how a domain and it is uncertainty may be represented in a Bayesian network. So the next best thing we do is to use that to reason and when we say reason using Bayesian network the basic task is to compute posterior probability distribution for a set of query variables given new information about some evidence variables that is when we observed value of some variable we would like the condition upon the new information or the process of conditioning called probability propagation or it is also called inference or belief updating to have in essence flow of information through the network. So that I can finally come up with what are the probabilities for the query variable.

(Refer Slide Time: 23:06)



Now there are different types of reasoning using Bayesian network let us look at each of them. The first of the reasoning using Bayesian networks is called diagnostic reasoning, now diagnostic reasoning is reasoning from symptoms to cause such as when a doctor observes dyspnoea and then updates his belief about cancer and whether the patient is a smoker. Now this reasoning occurs in the opposite direction to the network arc.

So actually it is about the causes of a smoker coming out as cancer leading to having shortness of breath but when you are doing diagnostic reasoning the evidence is about shortness of breath. And given this evidence reasoning occurs in the opposite direction to the network arc to find out or to be sure about his belief about cancer, whereas we have another type of reasoning which is called predictive reasoning, where reasoning from new information about causes to new beliefs about effects follow the direction of the network.

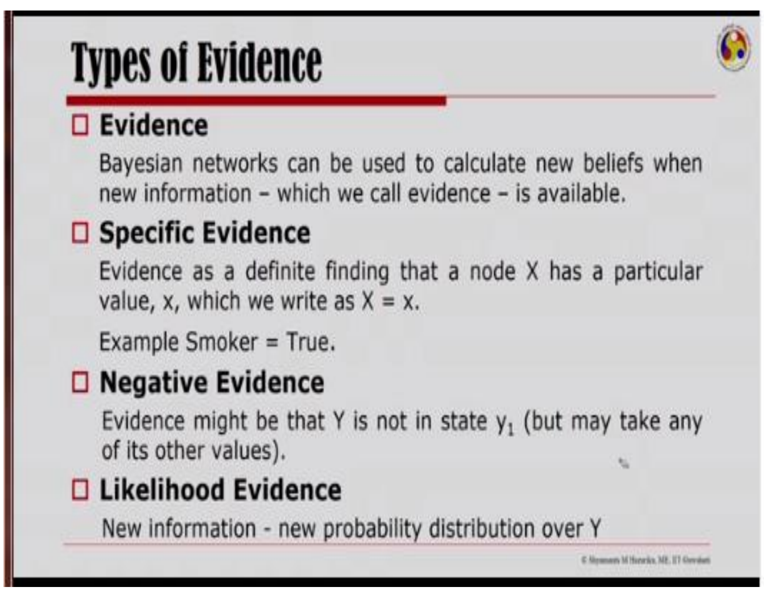
So you have evidence here that the person is a smoker and the doctor knows that this will increase the chances of the patient having cancer or having a positive x-ray. So here the direction of reasoning is along the direction of the network, we have a third type of reasoning using Bayesian network which is called inter causal or explaining away type of reasoning, now here you have an evidence that the person has cancer.

We also learn that the person is a smoker, so this lowers the probability that he has been expose to lot of pollution. Initially if you see the 2 causes of cancer, like pollution and smoking are actually independent however the presence of one in this case smoking renders the alternative cause less likely. In other words the alternative class has been explained away, this type of reasoning is called inter causal or explaining away reasoning.

Then we have certain scenarios in which query nodes and maybe evident nodes does not fit quite neatly into any of the 3 above reasoning types. We need to combine the above types of reasoning in some way and those type of reasonings are called combined reasoning. So to just review what we have discuss in types of reasoning we have A diagnostic reasoning which is reasoning from symptoms to causes that is what most of the physicians do.

Then we have reasoning from causes to new beliefs about effects that is called predictive reasoning that is along the direction of the network. We have third type of reasoning called explaining away which is about presence of one explanatory cause rendering an alternative cause less likely. And then there may not be so neat separation between queries and evidence nodes and we may need to combine these 3 types of reasoning to get to an answer and that would be a combined reasoning.

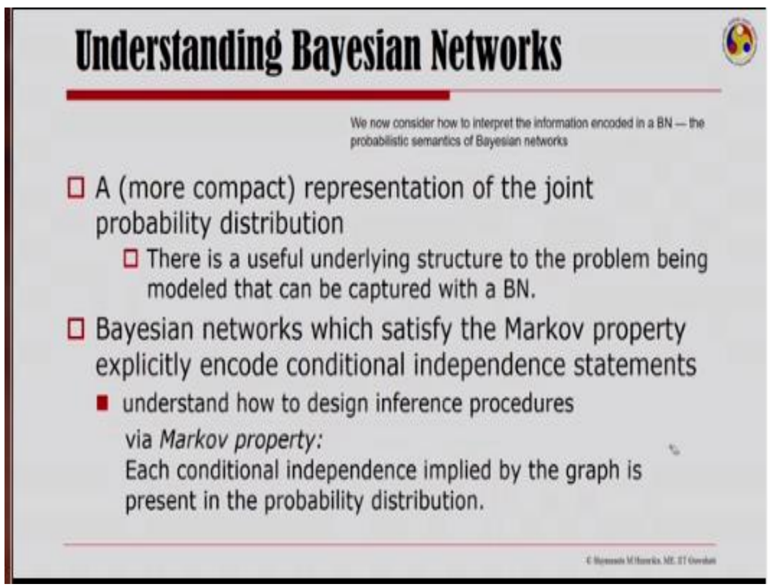
(Refer Slide Time: 27:23)



So when we are talking of evidence that is new information, what are the different types of evidence for a Bayesian network. So Bayesian networks calculate new beliefs when new information which we call evidence is available, evidence could be very specific. So evidence as a definite finding that a particular node x has a particular value like we could realize when talking to a person who has shortness of breath and we suspect cancer that he is a smoker.

So that is a specific evidence smoker = true, evidence might also be negative like it might be that y is not in state y1 but make take any other of it is values. Then there is something called the likelihood evidence when we have new information available that would need that we create a new probability distribution for that particular node. And then it would be called likelihood evidence.

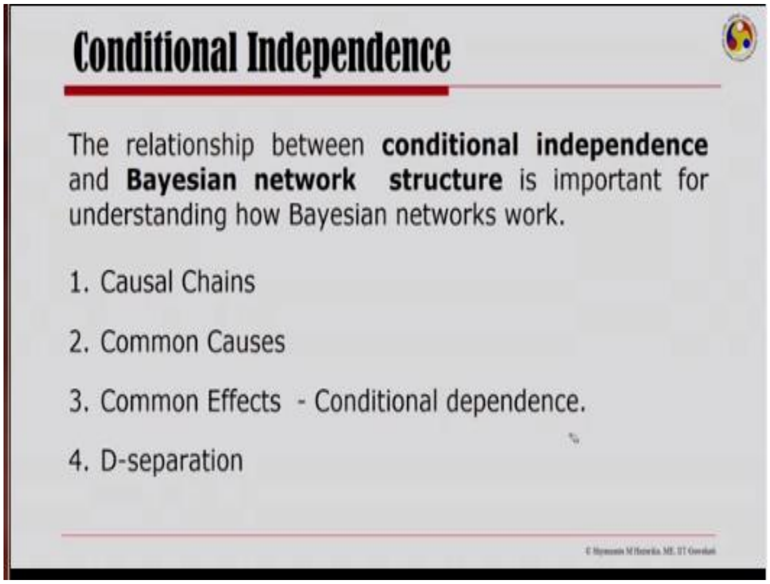
(Refer Slide Time: 28:37)



So we now consider how to interpret the information encoded in a Bayesian network this is the probabilistic semantics of Bayesian networks. So as already stated Bayesian network is a compact representation of the joint probability distribution there is a useful underlying structure to the problem being model that can be captured with the Bayesian network. Bayesian networks which satisfy the Markov property explicitly encode conditional independent statements.

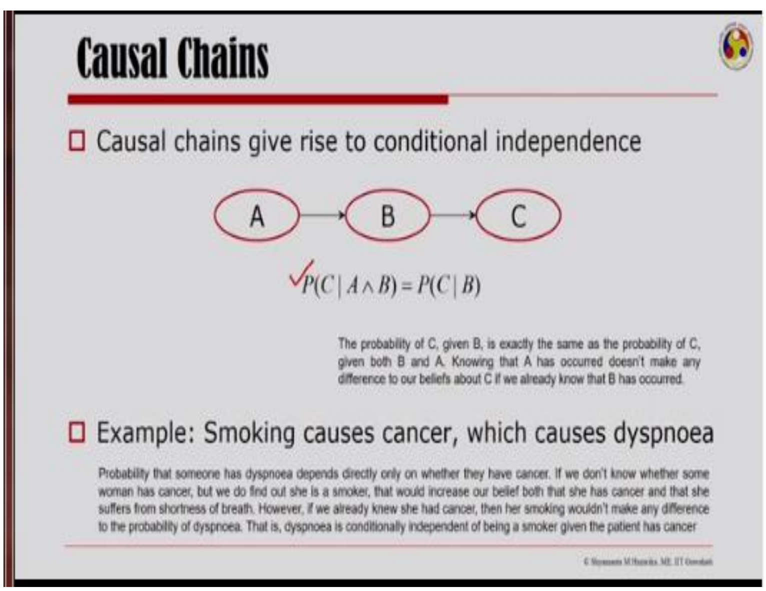
So understanding how to design inference procedures via the Markov property would mean that each conditional independence implied by the graph is present in the probability distribution.

(Refer Slide Time: 29:31)



So let us try to understand what we mean by conditional independence the relationship between conditional independence and Bayesian network structure is important for understanding how Bayesian networks work, we will look at the following 4 ways of trying to understand conditional independence, causal chain, common causes, common effects and D-separation. Now common effects is about conditional dependence we will see it when we discuss common effects in a minute.

(Refer Slide Time: 30:12)

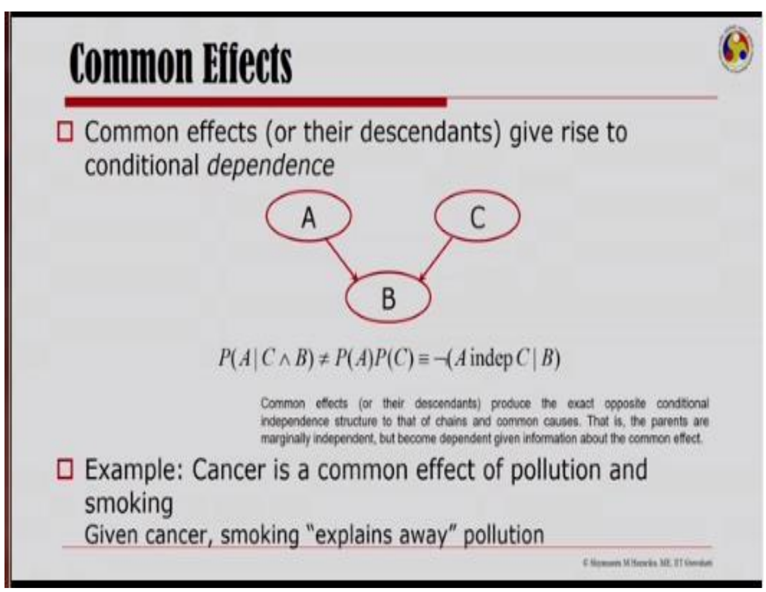


So let us first look at causal chains giving rise to conditional independence, so here is a causal chain A affects B and B affects C. So if you are looking for the probability of C given A and B now actually this is same as probability of C given B knowing that A has occurred does not make any difference to our belief about C if we already know that B has occurred. Now if you take our own example of the lung cancer diagnosis problem we know smoking causes cancer which causes shortness of breath.

Now probability that someone has dyspnoea depends directly only on whether they have cancer. If we do not know whether some woman has cancer but we do find out she is a smoker that would increase our belief both that she has cancer and she is suffers from shortness of breath. However if we know that she has cancer then her smoker would not make any difference to the probability of shortness of breath.

So dyspnoea is conditionally independent of being a smoker given that the patient is with cancer this is what we are showing here in terms of the causal chain. So causal chain gives rise to conditional independence this is a very important relation when we are analyzing a Bayesian network. So the probability of C given A and B, where A B and C forms a causal chain then is same as the probability of C given B.

(Refer Slide Time: 32:09)



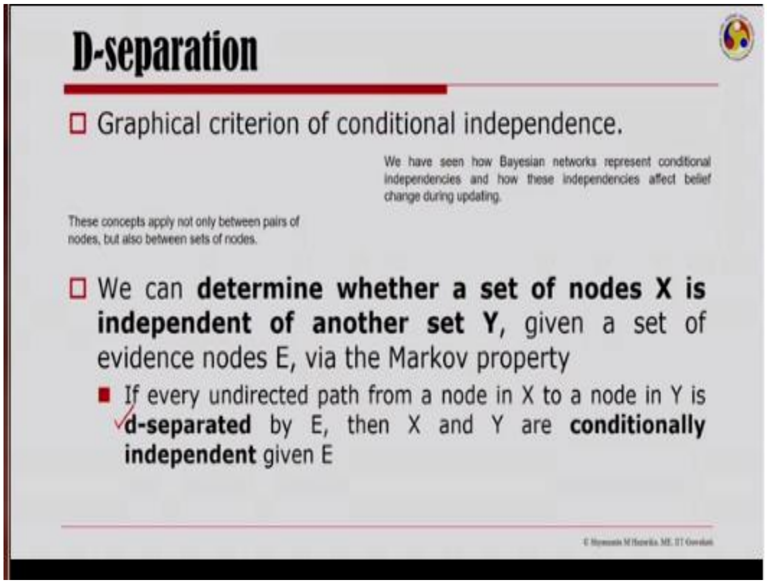
The next conditional independence rises because of common cause or common ancestors. Now here in this case, I am looking for the probability of C given A and B, if there is no evidence or information about a common cause like there is no evidence or information about cancer. Then learning that one symptom is present will increase the chances of cancer and which will increase the chances of the other symptom. However, if we know that the person has cancer then an additional positive x-ray would not tell us anything about the chances of dyspnoea.

So basically if I am looking for the probability of C given A and B and is same as the probability of C given B, I could say that A is independent of C given B. So common causes also give rise to conditional independence. Common effects, on the other hand gives rise to conditional dependence common effects or their descendants produce the exact opposite of what we have just discuss about conditional independent structure to that of chains and common causes.

Parents are marginally independent but they become dependent given information about the common effect. So if I am looking for the probability of A given C and B, I cannot say that it is equal to probability of A and probability of C, A is not independent of C given B. So here is from our example cancer is a common effect of pollution and smoking. And as I was discussing the types of reasoning given cancer, smoking could explain away pollution.

So you could see that when I have a common effect which is cancer of 2 nodes which is pollution and smoking having cancer one no longer remains independent of the other. Because smoking would explain away pollution and then that is about giving rise to conditional dependence.

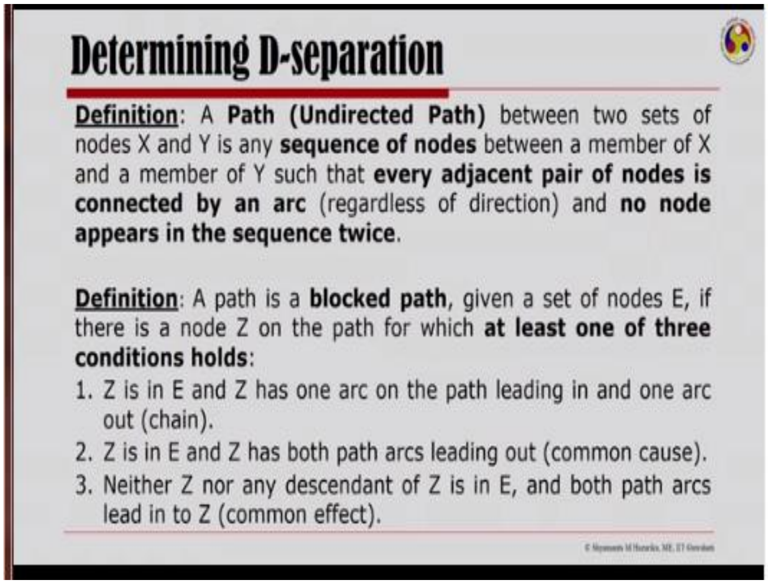
(Refer Slide Time: 34:46)



We have seen now how Bayesian networks represent conditional independences and how this independences affect belief chain, let us look at a graphical criteria of conditional independence up till now, the common causes, common effects and causal chains that we have discuss apply between pair of nodes. But one needs to realize that these concepts apply not only between pair of nodes but could also apply between sets of nodes.

So here we have d-separation which determine whether a set of node X is independent of another set Y given a set of evidence nodes E via the mark of property. Now it says that if every path from node in X to node in Y is D-separated by the evidence node E then X and Y are conditionally independent but let us look at what we mean by being d-separated.

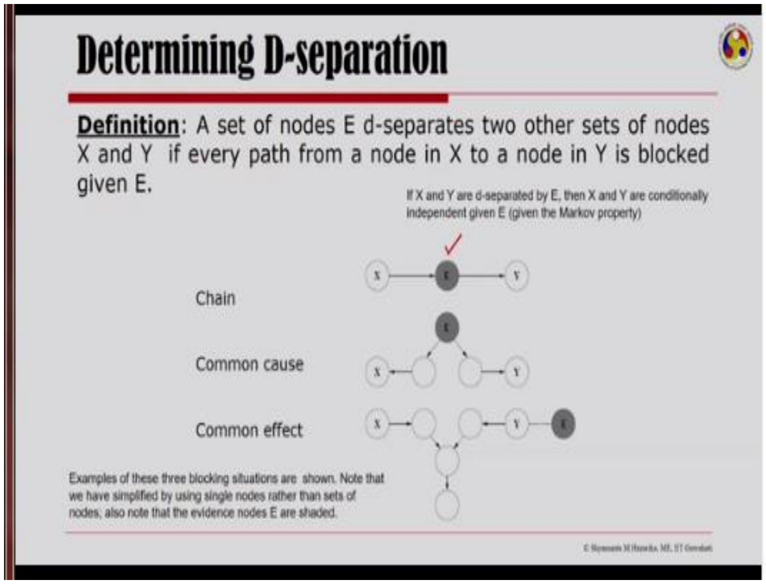
(Refer Slide Time: 35:56)



So in order to define D-separation let us introduce 2 more definitions one about the path, a path between 2 sets of nodes X and Y is any sequence of nodes between a member of X and a member of Y, such that every adjacent pair of nodes is connected by an arc and no nodes appears in the sequence twice. So we have a path and then we define something called a block path, now a path becomes a block path if given a set of nodes E.

If there is a no Z on the path for which at least one of the 3 condition holds, one Z is in E and Z has 1 arc on the path leading in and 1 arc out that means it is in a chain or Z could be in E and Z has both path arcs leading out. So it is about the common cause now neither Z nor any descendant of Z is in E and both path arcs lead into E. So it is about being a common effect, so when I have a chain I have a common cause or a common effect involving Z then Z on the path for which at least it has 1 of the 3, it creates what is called the block path.

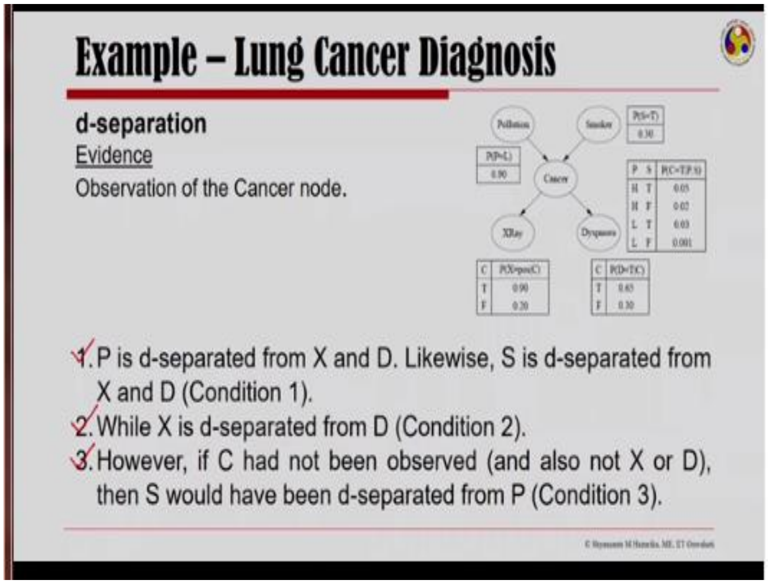
(Refer Slide Time: 37:31)



So now we are in a position to define what is d-separation. A set of nodes E is said to D-separate 2 other set of nodes X and Y if every path from a node in X to a node in Y is blocked given E and when I say block we could have one of the 3 either a common cause a common effect are in a chain. So if X and Y are d-separated then X and Y are conditionally independent given E, so here is the example now these examples are shown just using single nodes rather than set of nodes.

And this need not misleading to thinking that what we have in X and Y are single nodes, X and Y are sets of nodes and the evidence node E is what D-separates the 2 sets of nodes. In the first case E creates a block path using a chain the second is that of a common cause and the third is that of a common effect. And if X and Y are D-separated then we know that X and Y are conditionally independent given E.

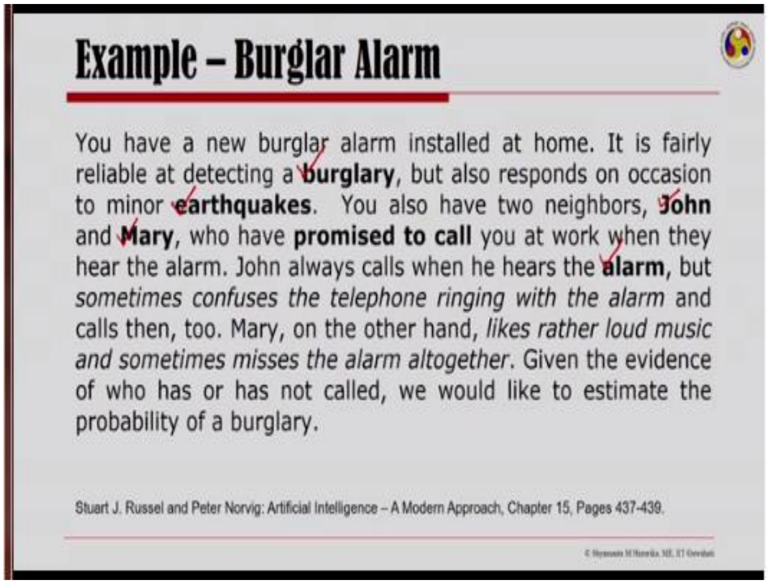
(Refer Slide Time: 38:56)



Now, let us look at the d-separation for the example that we are discussing for this lecture the lung cancer diagnosis. So let us say we have observation of the cancer node now given this on your right hand top corner, the Bayesian network for the lung cancer diagnosis example. We could see that here P is d-separated from X and D ok and similarly S is d-separated from X and D, S is also d-separated from X and D.

And this is precisely by condition 1 now we could also see that X is d-separated from D by condition 2 and if C had not been observed and also not X or D, then S would have been d-separated from P through condition 3, so these are the d-separation examples leading us to conditional independence.

(Refer Slide Time: 40:24)

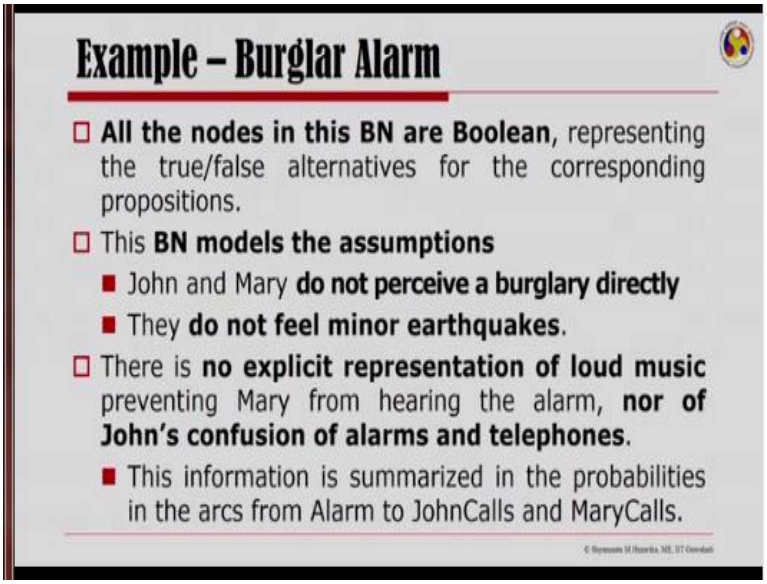


Now let us quickly introduce one example from literature we will not look at a great depth in to this example but we will just give you a flavor of how to identify the nodes and get to the Bayesian network. This example is the burglar alarm example from Russell and Norvig's book artificial intelligence a modern approach from it is 15th chapter. So you have a new burglar alarm installed at home it is fairly reliable at detecting a burglary.

But also responds on occasion to minor earthquakes now there are 2 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 them too. Mary on the other hand likes loud music and sometimes misses the alarm altogether now given the evidence of who has or has not called.

We would like to estimate the probability of a burglary, how do you do this with then reasoning under uncertainty you create a Bayesian network and then as I had been discussing you try to get to the probability of the query variable.

(Refer Slide Time: 41:54)

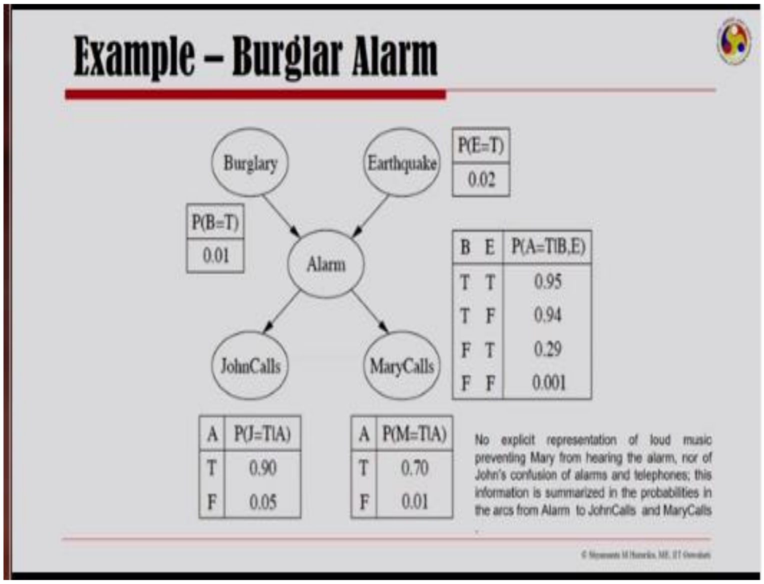


So all the nodes in this Bayesian network are Boolean they could represent true or false alternatives for the corresponding propositions. But then what are these nodes in the first place, so we could see that we are talking of the possibility of detecting a burglary. So that could be 1 node, now these alarm also talks of earthquakes, so that could be the other node, John and Mary has promised to call and depending on who is calling you want to estimate the probability of burglary.

So it would be nice to have 2 nodes, one for John and one for Mary calling you, so it could be like something like John calls and Mary calls. And then John or Mary would respond only when they hear the alarm, so alarm would be another node in your Bayesian network, all of these nodes are Boolean. The Bayesian network models the assumptions that John and Mary do not perceive a burglary directly they do not feel minor earthquakes all they do is respond to the burglar alarm system.

And there is no explicit representation of loud music preventing Mary from hearing the alarm nor of John's confusion of alarms and telephones. These type of information is to be summarize in the probabilities that are to be there in the arcs, from alarm to John calls and Mary calls.

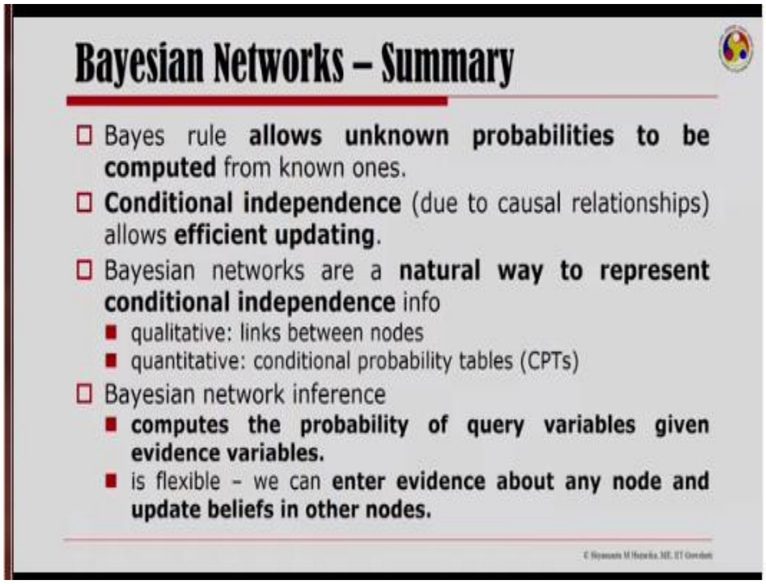
(Refer Slide Time: 43:40)



So here is the Bayesian network for the burglar alarm we have no explicit representation of loud music all that we have is the alarm system and earthquake or burglary giving rise to the alarm. The alarm in term giving rise to John calling or Mary calling but neither John's confusion of alarms and telephones nor Mary's hearing loud music and missing the alarm is represented. These need to be factored in, in the probabilities from alarm to John calls and Mary calls and the conditional probability tables.

If you see have instantiation for the parents if burglary is true an earthquake is true then the alarm sounding is almost 95% probable burglary true, earthquake false 0.94 falls burglary earthquake true of minor earthquakes it would respond. So here it is 0.29 and then if both of them false there is a very small probability that even then the alarm would be true. And here for Mary calling or for John calling the only thing that decides is the alarm. So the alarm ringing there is a probability that john calls, John also calls at times by confusing with the telephone, so there is a 5% probability that he calls but actually the alarm is false. And then we have calls by Mary even under alarm being true she misses because of loud music and that is why it is 70%. So such confusions such missing of information is summarize in probabilities in the conditional probability tables and this is how you would represent a problem using the Bayesian networks.

(Refer Slide Time: 46:19)



So, let us quickly review what we have done in the lecture today and the previous one we have looked at Bayesian's rule that allows unknown probabilities to be computed from known ones. Then we had looked at conditional independence or due to causal relationship that allows efficient updating. Today we had looked at Bayesian networks as a natural way to represent conditional independence information qualitatively it is about the link between the nodes.

Quantitatively it is about the conditional probability tables that we have for each node, the Bayesian network inference is about computing the probability of the query variable given the evidence variable. And we have seen that Bayesian network inference is flexible we can enter evidence about any node and update beliefs in other nodes, so this is all about by Bayesian networks, thank you.