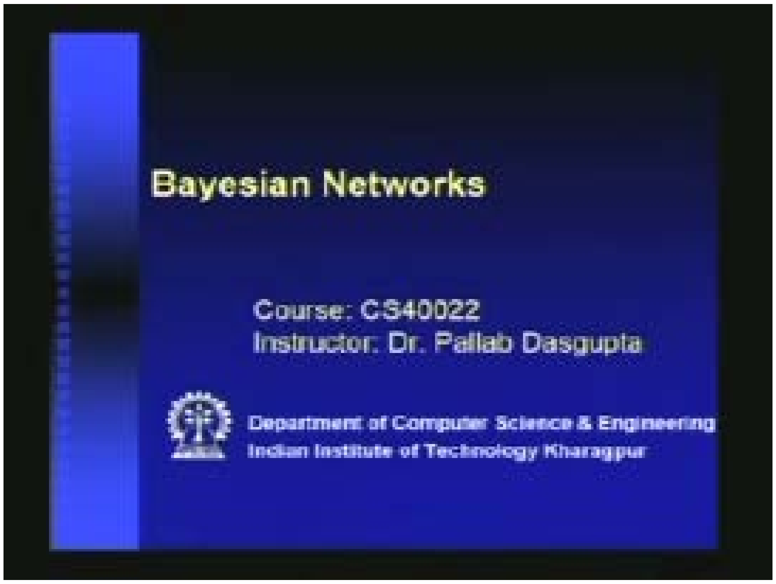
Artificial Intelligence

Prof. P. Dasgupta
Department of Computer Science & Engineering
Indian Institute of Technology, Kharagpur

Lecture No - 22
Bayesian Networks

In the last lecture, I had briefly outlined the meaning of what is a belief network or a Bayesian network. In this lecture, we look into the Bayesian network in details and also try to understand that why do we require such networks and what are the things that we can inferred in it.

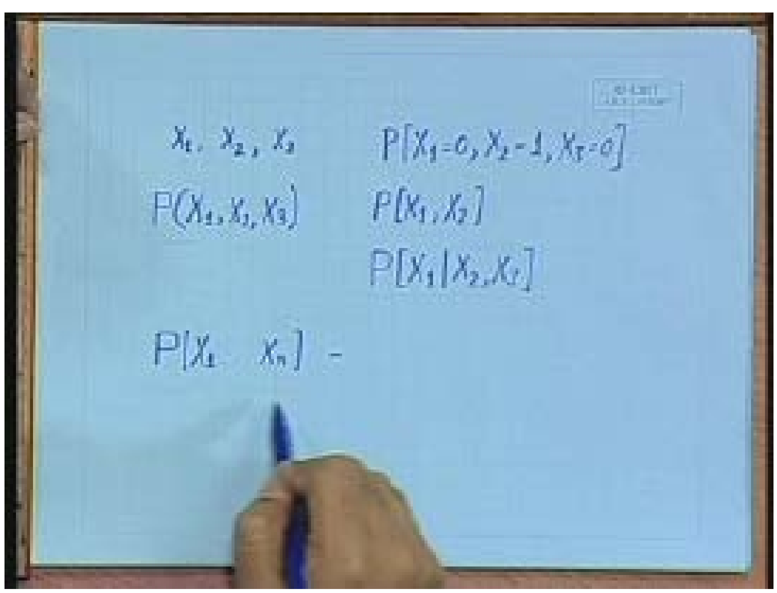
(Refer Slide Time: 01:07)



When we talk about reasoning under uncertainty, then, what we have is, we have a set of events and some events can be casually related to other events with certain probability, because there might be certain other factors which we do not know of; there can be certain other factors which we know of, but we do not know the exact degree of influence that it has. All these unknowns are partially known factors, can be modeled in the form of probability and we can reason under uncertainty once we have some idea about those probability. So, what we are interested in knowing about is the joint probability distribution. In other words, that suppose I have events X1, X2 and X3, then, I am interested in knowing the probabilities of different combinations of these X1, X2 and X3 in practice.

And we are also interested in finding out the probability of 1 or more of these variables, given that the other variables are true or false. For example, we might be interested in knowing in our burglar's alarm example, that what is the probability that John calls when there has been an earthquake, right? Given that the earthquake variable is true, we are interested in determining the value of probability of John calls, right? Effectively, let us- we should understand that what we have is a set of variables, let us say X1, X2, X3, and we are interested in the joint probability distribution. So, we are interested in things like pX1X2X3. Now, this distribution actually encompasses the probabilities of setting these different random variables to 0s and 1s.

(Refer Slide Time: 06:05)

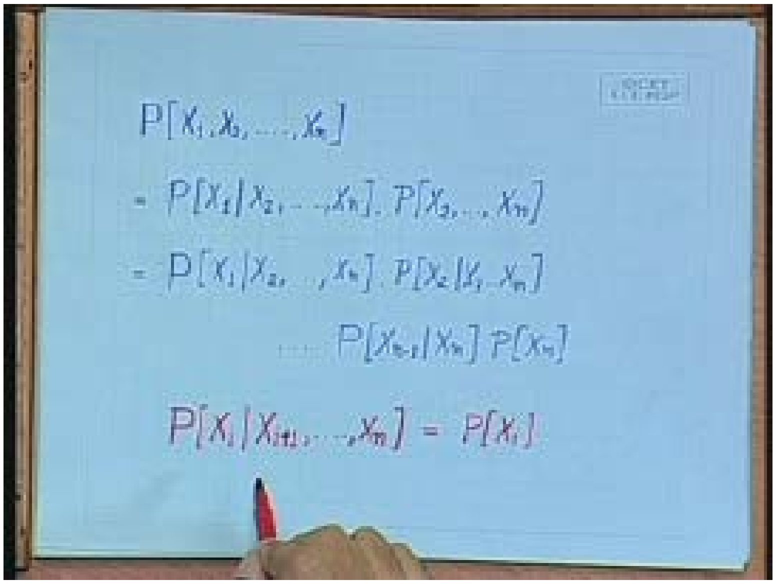


For example, this distribution will be able to give us the probability that X1 equal to 0, X2 equal to 1, X3 equal to 0, right? It will also be able to give us the probability of X1 equal to 1, X3 equal to 0, and so on, right? It is also going to give us the probability of X1 and X2 and what is this? This is the probability of X1, X2 and X3 plus probability of X1, X2, X3 bar, right? And from the same distribution, we should also be able to find out what is the probability of X1 given X2 and X3.

So, we know mechanisms of the elementary probability to compute each of these, provided we are given the distribution of X1, X2 and X3. Now, is it clear how we can deduce each of these? Now, Bayes network is a succinct way of representing these distributions. Let us see what is the succinct way. How can we reduce the size of this representation? Now, if we did not know about any of this relationship between these X1, X2, X3, suppose they were all.

We do not know about their inter-causal relationships and we store this probability distribution explicitly. Then, what is the size of that? Of the distribution table, for a set of variables- X1 through Xn- it is going to be 2 to the power on n. Why the size of this is going to be 2 to the power of n? Because that for each- these X1 through Xn can take the values of true or false, these are proportional variables. So, we can have a true or false set to each of them, so, the number of entries in the distribution table is going to be 2 to the power of n. This is a n dimensional table and we left 2 to the power of n entries in it.

(Refer Slide Time: 09:21)

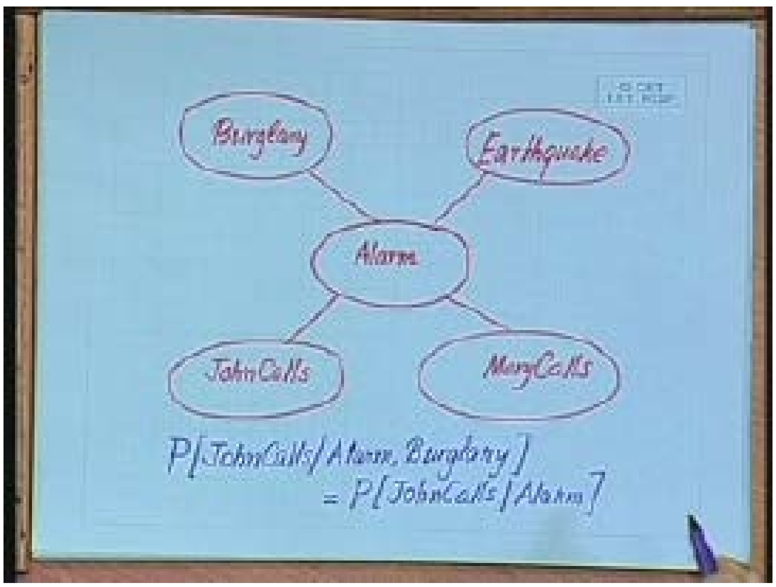


Now, storing that kind of a information is quite difficult, so, people try to see if we know about the inter-causal relationships or we are able to deduce the inter-causal relationships, then, can we get a more succinct representation, right? And to do that, let us first look at the way in which we can rewrite this joint probability distribution. So, what we have is probability of X1, X2 to Xn. We can write this as probability of X1, given X2, through Xn times probability of X2 Xn. And then, we can again rewrite this; we can unfold this term also in a similar way.

If we do that, then, what we are going to have is probability of X1 given times probability of X2 given- this is X3 to Xn- and so on, until we have probability of Xn minus 1 given Xn times probability of Xn. Now, there are several ways in which these individual terms can be simplified. For example, if we have a case where we have some p of Xi given Xi minus 1 through Xn, Xi plus 1 through Xn and if it turns out that Xi is independent of all these variables, so, this is going to reduce to simply p of Xi, provided that Xi is independent of all this.

Also, it could be the case that 1 of these- 1 or more of these variables- subsume the others, because let us understand in what way can they subsume. We will introduce a terminology called conditional independence. So, we will have this kind of scenario, when it is totally independent, but we will also look at a scenario where it is conditional independence. Now, for example, if you recall that burglar's alarm example, we had alarm here in the network, and then, we have 2 causes for the alarm possible. Causes- 1 is an actual burglary and the other is earthquake and the effect of the alarm is again- can be 2- 1 is John calls and the other was that Mary calls.

(Refer Slide Time: 11:35)

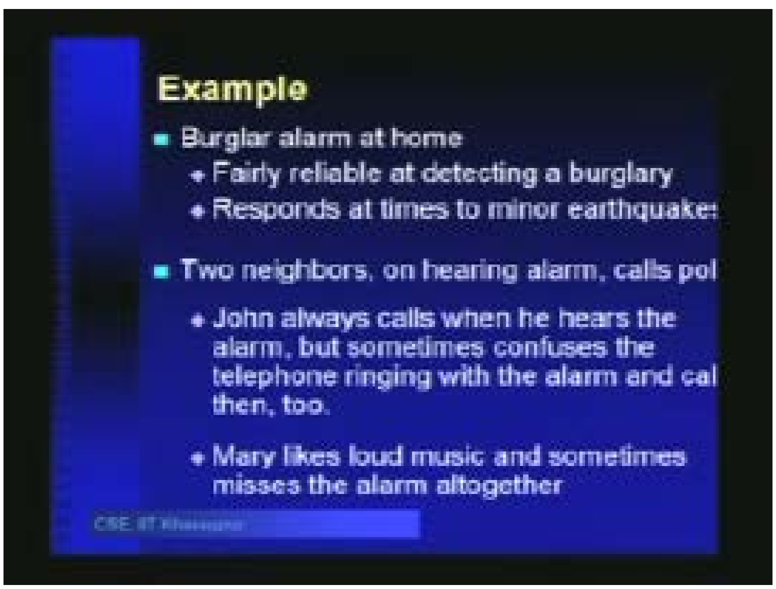


Now, let us look at the probability of John calls, given alarm and burglary. Now, see, my claim is that I can rewrite this; I can rewrite this as probability of John calls, given only alarm and my claim is that if you know that the alarm has gone off, then, it does not really matter whether the burglary has taken place or not, because the John calls is affected only by the fact whether the alarm has gone off or not. So, for John, all it matters is whether the alarm has gone off or not.

So, if you are given that the alarm has gone off, then, the probability of John's calling is not dependent on whether the burglary has actually taken place or not. Therefore, this is a scenario of conditional independence. It is this alarm that actually subsumes this burglary and so you do not require this burglary anymore in the conditional probability. It is this fact that we know, that John calls when he hears the alarm; it is this fact which has prompted us to say that okay, that if alarm is known, then John calls is independent of burglary.

Now, these are the kinds of things that we attempt to discover from the domain and use them to deduce the size of the conditional of the joint probability distribution, and the net result of that is what we have in Bayes network. We try to order the variables in such a way that we can make maximum use of these conditional independence, so that in each of the tables that we stored in the- besides the entries of the Bayes network is place. Now, let us do 1 thing: we will return to this example and see what happens if we use a proper ordering and what happens if we use a improper ordering.

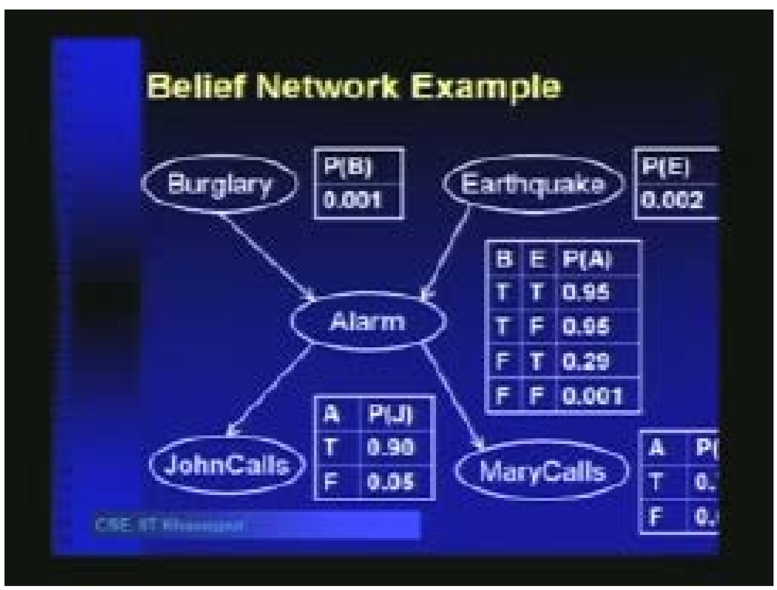
(Refer Slide Time: 14:20)



Let us recap: the example was that we have a burglar alarm at home, which responds to burglary or earthquakes. There are 2 neighbors, who on hearing the alarm, calls the police. John always calls when he hears the alarm but sometimes confuses the telephone ringing and then also calls, and Mary does not always call, even when the alarm goes off, because Mary is fond of loud music.

This was the belief network example that we were looking at, so, in this case, the ordering of the variables is John calls, Mary calls, alarm burglary, earthquake, or Mary calls, John calls, alarm burglary, earthquake. This is the ordering in which we are considering the joint probability distribution.

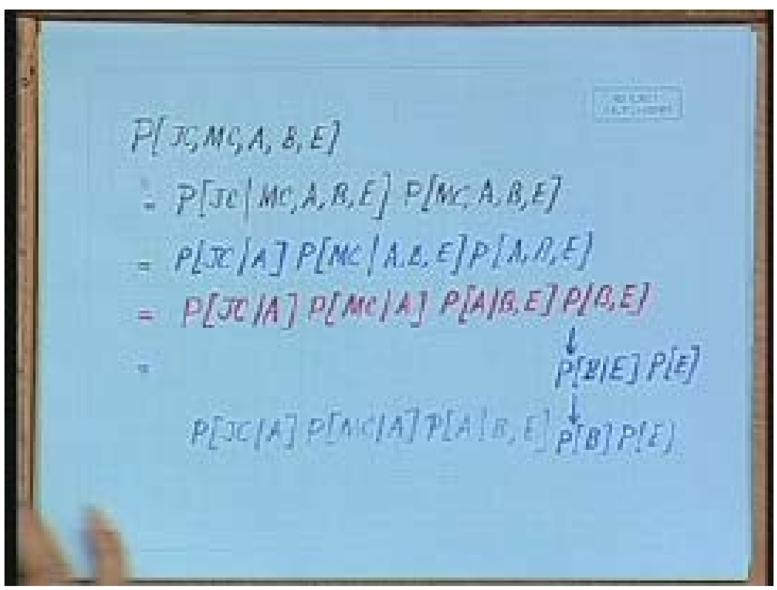
(Refer Slide Time: 14:39)



Effectively, what we are doing is, we are starting with probability of John calls, Mary calls, alarm burglary, earthquake, and then rewriting this as probability of John calls given Mary calls, alarm burglary, earthquake. Now, if you are given alarm, then, all these others are irrelevant, because John calls will depend only on the alarm. If the alarm was not given, if we do not have this variable here, if it is unknown, then, these may not be independent. Then, the fact that Mary has called is increased evidence that the alarm had went off.

Suppose alarm was not here; suppose we had only John calls, given Mary calls, a burglary and earthquake- then, the fact that Mary had called is increased evidence that the alarm had gone off. So then, John calls and Mary calls are not independent, but given alarm, they are independent. Are you getting my point? And then, we have probability of Mary calls alarm, so, the first term will reduce to probability of John calls, given alarm only. Now, let us look at what we can do with the second one. The second 1 will be Mary calls, given alarm burglary and the earthquake.

(Refer Slide Time: 19:57)



Again, if the alarm has gone off, if we know about the truth of the alarm variable, then, burglary and earthquake do not matter in terms of Mary’s calling. I mean, not in this case, of course it could have been the case that when Mary also detects the earthquake and hears the alarm, then, she might infer that it is a false alarm which is gone off. But we are talking about that, so, in this case, again, this alarm renders Mary calls to be conditionally independent of this. Again, we will have- okay, we had another term out there, which was- and that term can now be expanded and therefore, what is going to happen to this alarm, given burglary and earthquake?

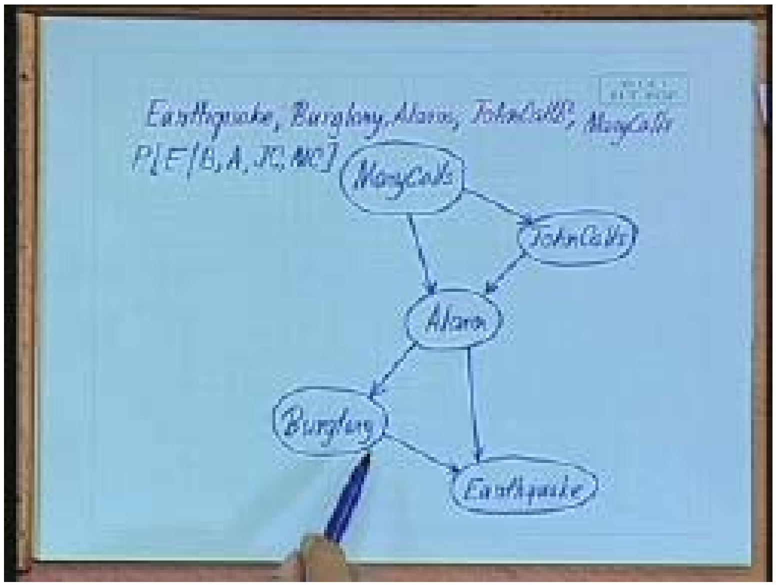
Here, none of them can subsume the other. Even if we know that there is no burglary, the alarm can go off because of the earthquake and vice versa. So, this does not reduce any further, so, we will again expand up this to probability of B given E and probability of E. Now, this term- we know that the probability of burglary has got nothing to do with whether there has been an earthquake or not. I mean, it is true that burglars will not be willing to do a burglary when there is an earthquake going on.

But well, for the example that we are having here, we will consider this to be independent, so, this is going to reduce further to probability of burglary. So, the total thing that we have- let us see- will be probability of John calls, given alarm, probability of Mary calls, given alarm, then, probability of alarm, given BE and then, these 2. This is exactly the Bayes network that we have, so, we have the parent of John calls is the alarm, the parent of Mary calls is the alarm, parent of alarm is burglary and earthquake and burglary and earthquake are the top level events which are not casually dependent on any things.

Let us see what happens if we try some other ordering. Suppose we look at the other ordering, where the ordering that we are going to look at is earthquake followed by burglary, then, alarm. So, this is exactly the opposite what we took last time, followed by Mary calls. Let us start constructing the belief network with this in mind. We will have earthquake here and we will be looking at probability of earthquake, given burglary alarm, John calls, Mary calls, right? Now, let us see.

Does earthquake depend on burglary? Can we make it independent of burglary? We are trying to find out the probability, we are trying to reason about the probability of earthquake, given burglary alarm, John calls and Mary calls. (Students speaking). Yes, is it not? Therefore, this is not going to be independent, so, we are going to have will alarm affect earthquake? Yes, because whether the alarm has gone off or not is increased evidence. If the alarm is gone off, then, is increased evidence, however minor of the earthquake having taken place.

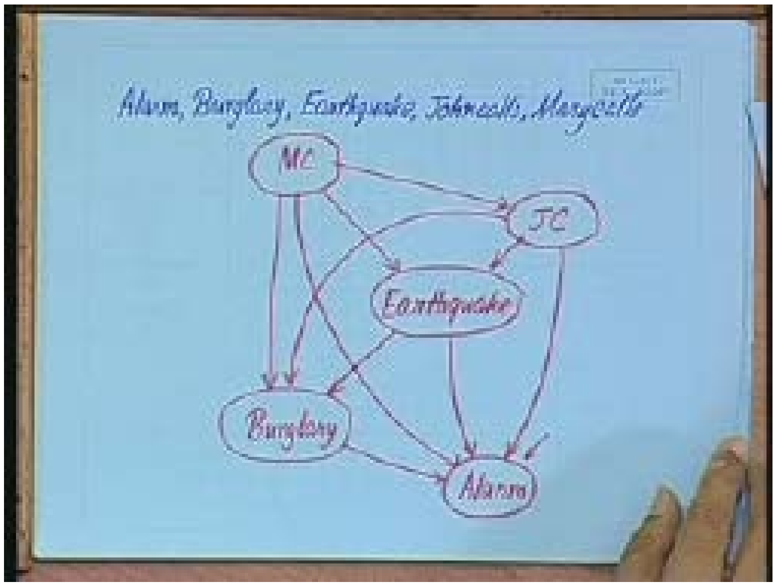
(Refer Slide Time: 24:43)



Then, let us look at burglary. Given an alarm, I think the earthquake will be independent of John calls and Mary calls. So, then, let us look at burglary, given alarm, John calls and Mary calls. The probability of burglary for the same reason- it will depend on alarm, but given alarm, it is not going to depend on John calls and Mary calls. Then, let us look at alarm, given John calls and Mary calls. Now, this is going to depend on this: definitely, this is going to depend on this, but what about probability of John calls, given Mary calls? It depends, so, this is going to be the belief network.

Now, this is not much different in size as compared to the original belief network, but it requires us to produce probability values which are very confusing. For example, we have to produce the probability value of earthquake given burglary. How are we going to produce that kind of probability? If we construct it in incorrect way, then, problem number 1 that comes up is that you are required to specify the conditional probability of some very confusing combination of events, which is very difficult in practice. There is another problem, so, let us look at a slightly different ordering, where we have alarm burglary.

(Refer Slide Time: 29:38)

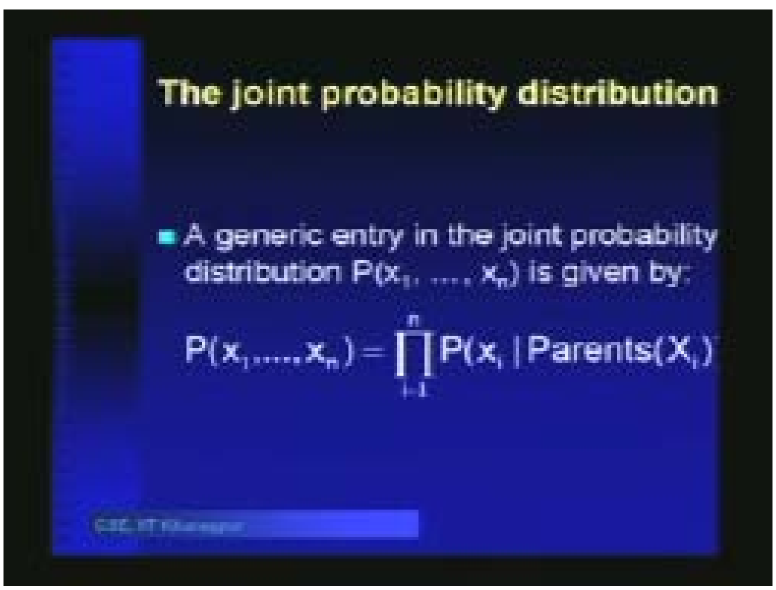


Let us start with this one. We will have, first of all, alarm. It will depend on burglary of course. Is it clear to you why the alarm becomes conditionally dependent on burglary? Then, we have earthquake, John calls and Mary calls. Let us look at the different kinds of things that we will have, when we look at alarm, given burglary, earthquake, Mary calls, John calls. This is going to be dependent on all of this; none is going to subsume the other. We will have earthquake to burglary and then, we will have the ones that we had previously, namely these 3. We will also have from Mary calls to burglary and Mary calls to earthquake and John calls to burglary and John calls to earthquake. So, this is going to be the Bayes network for this ordering.

See, the problem here is that if you look at the conditional probability table here, because I have 4 parents, so, I will have to store 2 to the power of 4 entries here. This fellow will have 3 parents, so, 2 to the power of 3 entries here. These conditional probability tables are going to be quite large, because you will have many parents of the same node. So, if you use an improper ordering, this is another side effect that takes place, but we can always say that this particular 1 and this 1 and the 1 that we started off with, they are all valid Bayes networks, in the sense that we are able to recover if we are able to produce the probability tables for these- the conditional probability tables for each of these- then, we can recover the joint probability distribution from all of them.

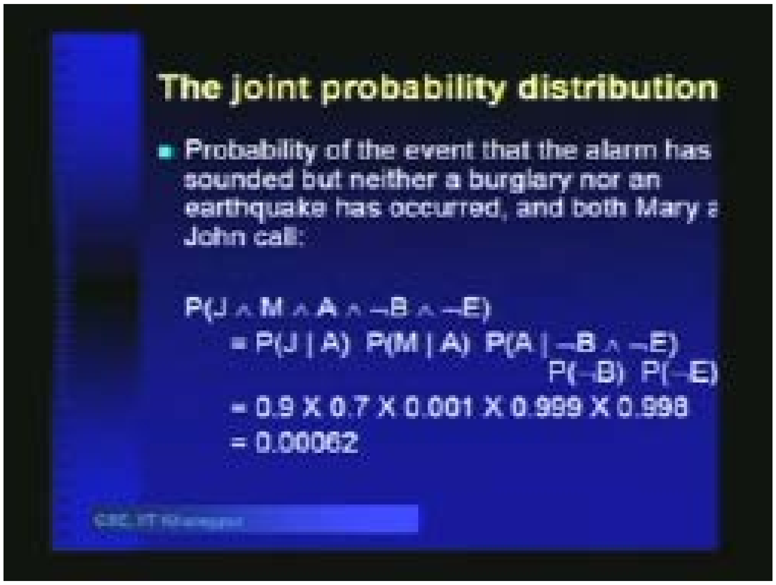
We can recover the joint probability distribution from all of them, so, all of them are valid representations of the joint probability distribution. But in this case, we were having to produce conditional probability values which are not natural and which are difficult to obtain experimentally and in this case, the size of the representation is quite large as compared to the 1 that we started off with.

(Refer Slide Time: 29:51)

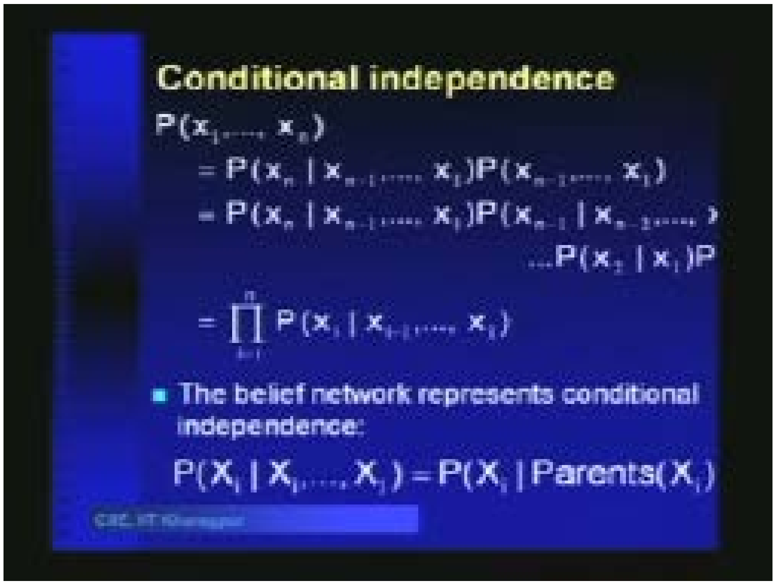


From here, from what we have seen so far, let us quickly see the way in which we can construct a Bayes network. The way in which we have talked about right now-

(Refer Slide Time: 29:53)

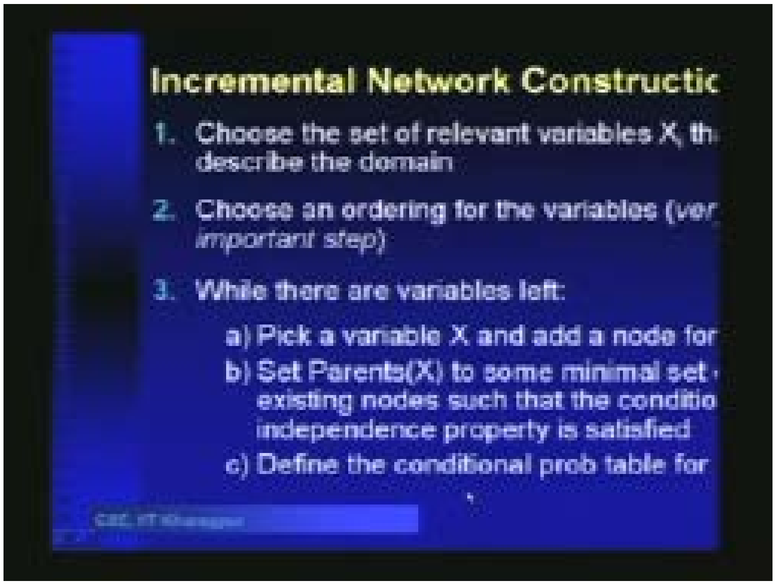


(Refer Slide Time: 29:54)



We first choose the set of relevant variables that describe the domain, then choose an ordering for the variables, which is, as we now know, a very important step. While there are variables left, pick a variable X and add a node for it. The way in which we did was, we started from X1 and then picked up X2, then picked X3, and so on. Set the parents of X to some minimal set of existing nodes, such that the conditional independence property is satisfied. Define the conditional probability table for X. So, this is the way in which we can construct the Bayes network.

(Refer Slide Time: 30:34)

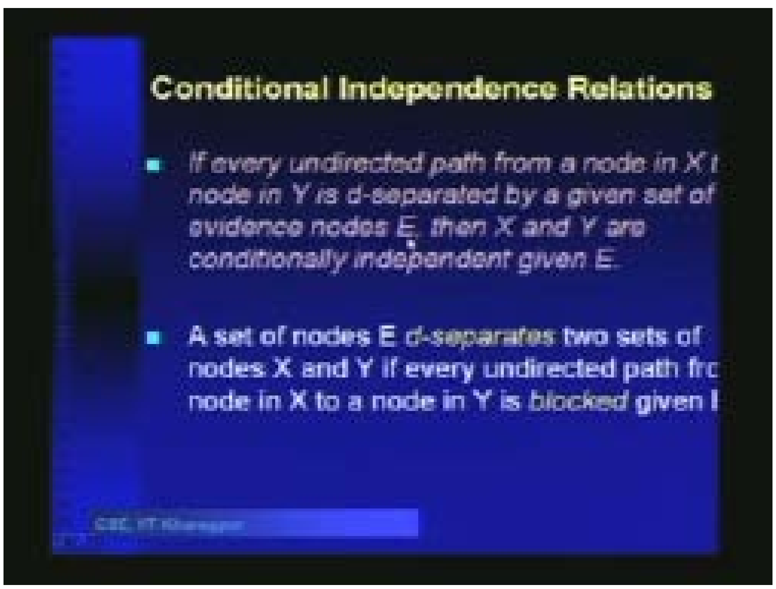


As we have seen, the conditional independence relations play a very important role in the Bayes network structure. We will now formally make some definitions about conditional independence, which will be very useful when we start querying the Bayes network. The first step is to construct the Bayes network. Then, why do we construct the Bayes network? We construct the Bayes network because if we have any queries related to the joint probability distribution which includes queries like okay, if I have cavity and if there is a past history of extraction, then, what is the probability that I will have to go for extraction again?

These are the kinds of diagnostic inferences whose probabilities we want to evaluate and that is what we are trying to do using the joint probability distribution. So far, we have only studied how to construct the Bayes network. Next, we are moving into how to use the Bayes network to infer about other combinations of events. If every undirected path from a node in X to a node in Y- X and Y are sets of nodes in the Bayes network is d-separated by a given set of evidence nodes E. Evidence nodes are the nodes whose variable values are given, like when we say the probability of John calls given alarm, so, alarm then becomes an evidence node, because we are given that alarm has gone off or probability that many calls given not alarm, then also, alarm is an evidence node.

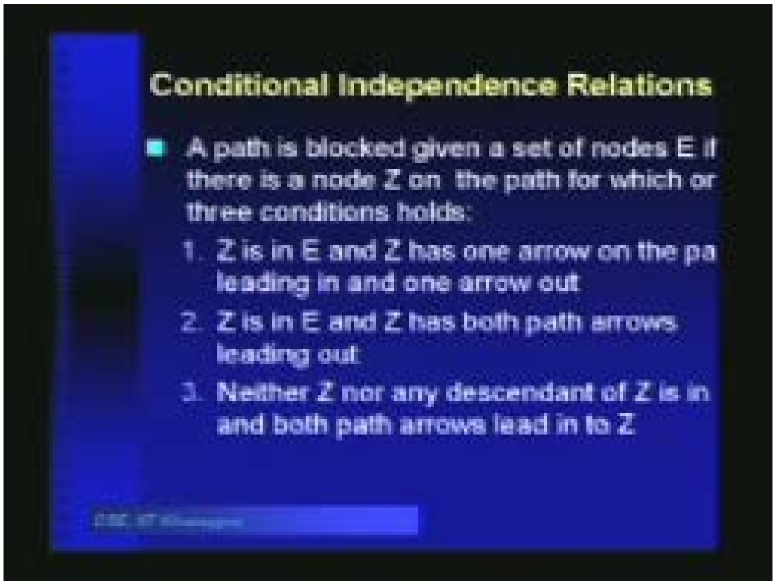
So, if we have a set of nodes X and the set of node Y and every undirected path from X to Y is d-separated- we are going to define what is meant by d- separated by a given set of evidence nodes E, then, X and Y are conditionally independent given E. Now, the definition of d-separation is like this: a set of nodes E d-separates 2 sets of nodes X and Y. If every undirected path from a node in X to a node in Y is blocked, given E:

(Refer Slide Time: 32:36)



Now, we have to define what is blocked. A path is blocked given a set of nodes E. If there is a node Z on the path for which 1 of the 3 condition holds, Z is in E and Z has 1 arrow on the path leading in and 1 arrow out.

(Refer Slide Time: 36:11)

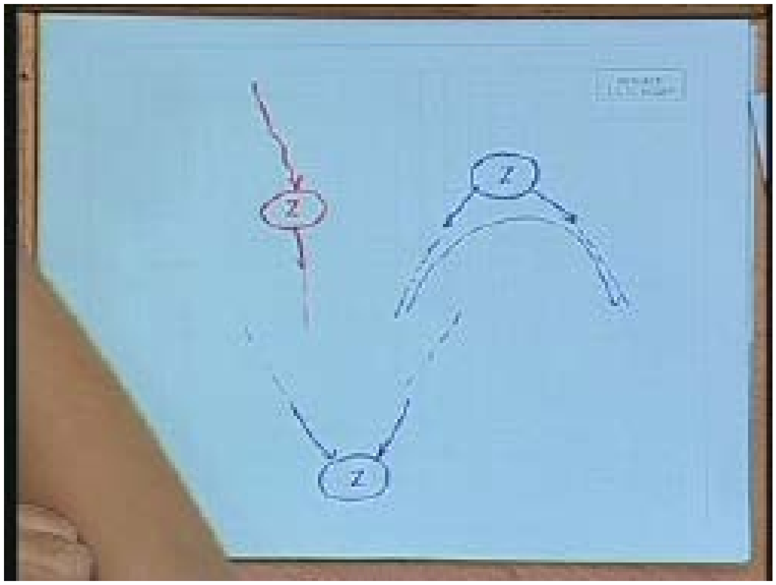


We are going to come back to this. I will first give some examples, then, it will be easier to understand this. Just keep this in mind, that a path is blocked given a set of nodes E if there is a node Z on the path, for which 1 of the 3 conditions hold. That is that the node Z is in E and Z has 1 arrow on the path leading in and 1 arrow out. So it is like this, that we have in the path a, node Z and 1 arrow is leading into Z and 1 is leading out of Z, then, this path is blocked by Z or Z is in E and Z has both path arrows leading out. So, it could be that in a Bayes network, you can have paths like this, also this is also a path, say, if we take this path.

(Refer Slide Time: 35:19)

Now, in this path this Mary calls is 1 which is a d-separating or blocking this path, because it is in the evidence. Set E and I have 2 edges coming out of the node Z. This is 1 scenario. The other scenario is that the path has Z like this and the path is like this, so, the path is not necessarily a directed path.

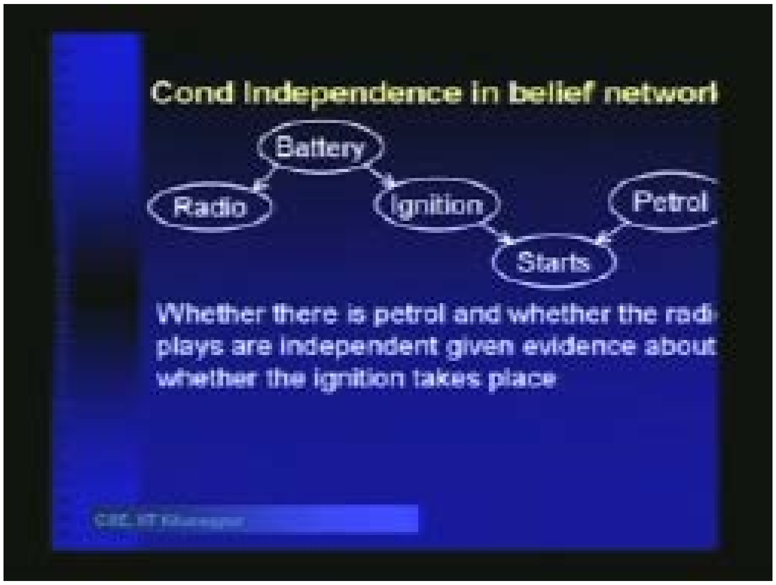
(Refer Slide Time: 36:42)



Then, neither Z nor any descendent of Z is in E and both path arrows lead into Z, so, we have Z and both path arrows are leading into Z and neither Z or any of its descendents are in E, so, the path is like this, but this is not in E and neither of its successors are in E. Now, these are the cases where Z blocks the path and we say that if Z blocks the path, then, the set of nodes E d-separates 2 sets of nodes X and Y, if every undirected path from a node X to a node in Y is blocked, given E.

Let us take an example. We have the following scenario: we have a car radio, we have the car battery, we have the ignition, we have the petrol gauge and we have the event- whether the car starts or not. For example, if the car does not start, then, it could be that either it is out of petrol or it could be that the ignition is not working properly or it could be that the batteries down, but if the batteries down, then, the radio will not work either. So, if the radio is working, then, it is evidence- increased evidence- that the battery is fine.

(Refer Slide Time: 38:22)



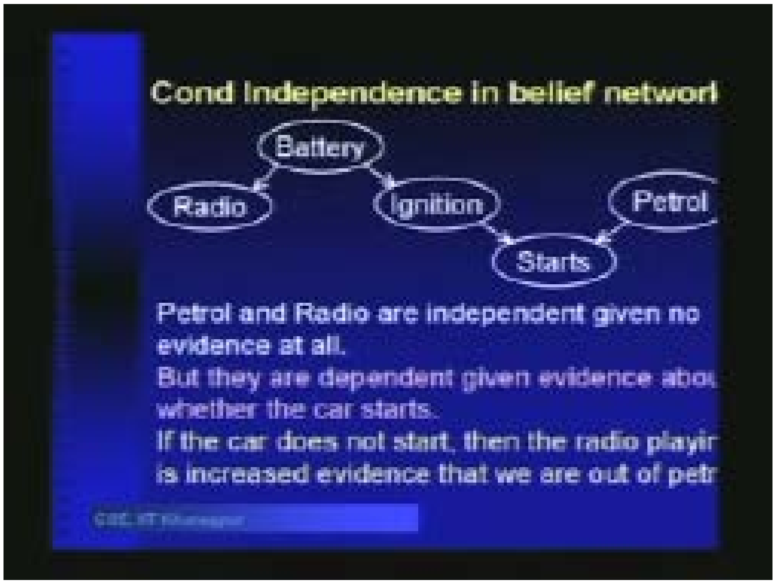
This is the set of event that we have and the Bayes network structure is given here. Now, whether there is petrol and whether the radio plays are independent, given evidence about whether the ignition takes place. Suppose we are given that the ignition takes place. Then, whether there is petrol and whether the radio plays will become independent. Is it clear? Why is it- (Students speaking). So, formally, the node ignition which is part of the evidence will d-separate the radio and the petrol, so, the path from radio to petrol is d-separated by ignition.

This is which of these cases? Out of these cases, it is this 1, right? Because the node Z is in E and this is the node ignition, we have 1 arrow coming into it, 1 arrow going out of it. But is it intuitively understood that why radio and petrol are independent? Once we know whether ignition has taken place or not, see, because if ignition has taken place, given that ignition has taken place or not taken place. Then, the probability whether it starts or not is not going to depend any more on the battery. If you know whether the battery is good or not does not matter anymore, because we are given whether the ignition has taken place or not.

The probability of starts becomes independent of battery, similarly, the probability of whether there is petrol or not becomes independent of battery and of radio. Petrol and radio are independent if it is known whether the battery works again for the same reason.
If you are given whether the battery works or not, then, petrol and radio will be independent. Petrol and radio are independent given no evidence at all. (Students speaking). Yes. (Students speaking). No, radio is directly from the battery, it does not go through the ignition.

You can run the radio even when you have not turned on the ignition. (Students speaking). Yes, but if you know about the battery status, then, radio and ignition become independent, so, the dependence is through the battery and that is how we have depicted it here. In the slide, if you look into the slide- slides please- if you look at the slide, then, we have slight guess that the dependence between radio and ignition is through the battery. If you do not know about the status of the battery, then, obviously, radio and ignition are co-dependent.

(Refer Slide Time: 44:23)

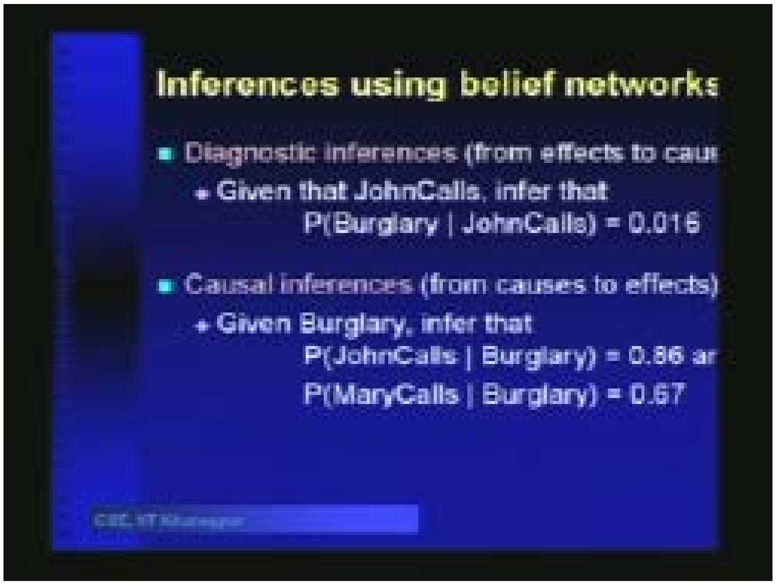


But if you know about the battery status, then, they are independent, then, the radio has got nothing to do with the ignition. But petrol and radio are independent, given no evidence at all. If you do not know of any evidence, then, they are independent. Why? Because starts is going to d-separate them. So, you do not know. If you do not know, do not have any evidence about whether the car starts or not, then, there is no dependence between petrol and radio, right? However, if you are given evidence about the car starting, then, petrol and radio are not independent. Why? Because if the car starts, then, the radio not running is- (Students speaking).

Will have less probability and also, why are petrol and radio going to be dependent once we know whether the car starts? (Students speaking). If the car does not start but the radio is on, then, that is increased evidence that we are out of petrol. If the car does not start but the radio is working fine, then, that is increased evidence that we are out of petrol, so, petrol and radio are not independent anymore, right? Now, we have seen the example of all 3 of these d-separating scenarios, right? Please recap this thing at home and try to think of some events where you have these kinds of separations. When we do the tutorial, we will do 1 example on Bayes network.

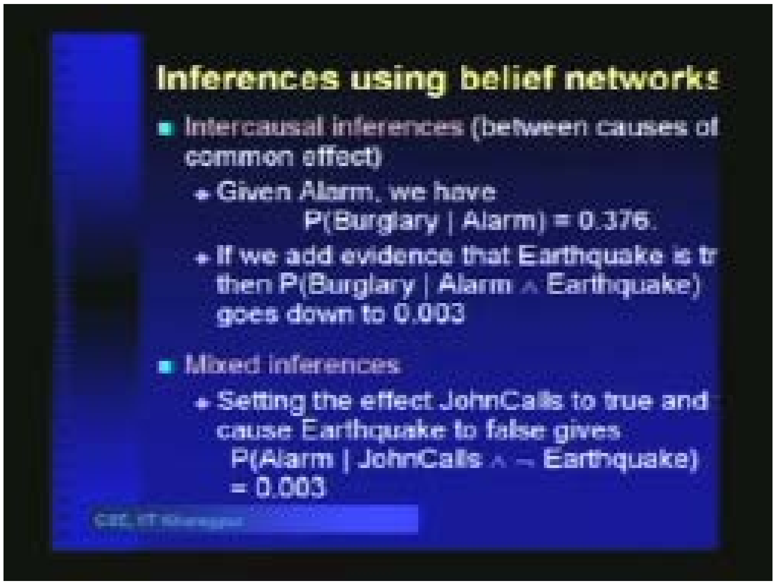
Let us see what we have next. If the car does not start, then, the radio playing is increased evidence that we are out of this, is what I just now said. Now, we are interested in doing different kinds of inferences using belief networks, right? I am just introducing them today and we will pick them up from here in the next lecture. So, 1 is diagnostic inferences, where we try to find out- go from effects to causes. Like, for example, given John calls, we can infer that probability of burglary, given John calls, is 0.016. This is from effects to causes. We can also have causal inferences which are from causes to effects, like probability of John calls, given burglary or probability of Mary calls, given burglary.

(Refer Slide Time: 45:13)



For example, when we are talking about medical diagnostics, then, a diagnostic inference will be that if you are given evidence about the causes, then, what is the probability that this is going to be the effect? And when you are talking about causal inferences, then, given the symptoms, what is the probability that this is the disease? Then, inter-causal interference inferences between causes of a common effect, so, given alarm, what is the probability of burglary? It is 0.376. If we add evidence that earthquake is true, then, probability of burglary, given alarm and earthquake goes down to 0.003. Why is that so? (Students speaking). Because earthquake itself is a very rare event.

(Refer Slide Time: 47:43)



And the probability that you are having a burglary at the same time as the earthquake takes place, is even less probable. This is the- between causes of a common effect, if you want to reason about, then, mixed inferences like setting the effect John calls to true and the cause earthquake to false gives probability of alarm, given John calls and not earthquake is this. In terms of Bayes networks these are- how are these dependent? In the first one- diagnostic inferences- we are inferencing the probability of a parent, given the child.

The second 1 is inferencing the probability of the child given the parent, and the third 1 in inter-causal, we are examining the probability of a parent given a child and a sibling. And in mixed inferences, we are examining the probability of an event, given its parent and its child. These are the 4 main kinds of inferences that we will talk about and we will examine some algorithms, such that given a Bayes network, how do we determine these probability values? With that, we come to the end of this lecture.