**Artificial Intelligence**
Prof. Anupam Basu
Department of Computer Science and Engineering
Indian Institute of Technology, Kharagpur
Lecture # 28
Reasoning with Uncertainty III

In the earlier lecture we discussed about reasoning with uncertainty and there we discussed about certainty factor algebra. And we had also shown how MYCIN uses certainty factors in order to deal with uncertain reasoning. A very powerful tool for reasoning with uncertainty is the probability theory and probabilistic techniques. In another term that is called statistical reasoning. Many of you might be familiar with probability and we will have a brief recapitulation of the basics of probability theory today and then from there we will proceed towards reasoning using probability of events. We are having some confidence associated with some rules. However, these confidences or the certainty factors or the measure of beliefs or disbeliefs were being attributed by subjective decisions taken by experts.

However, if we had actually conducted a number of experiments, for example, a particular rule a implies b, out of thousand times when a is true how many times did b become true?
If we could make a through experiment about this and based on the frequency of the times when a implies b really holds we could have really given a much more realistic value or a more realistic strength of belief to the rule. That is the basic idea of probabilistic reasoning. However, there are associated problems and complications as well. Before going towards the probabilistic reasoning we will re-look on the issue of uncertainty in decision making problem.

(Refer Slide Time: 03:54)
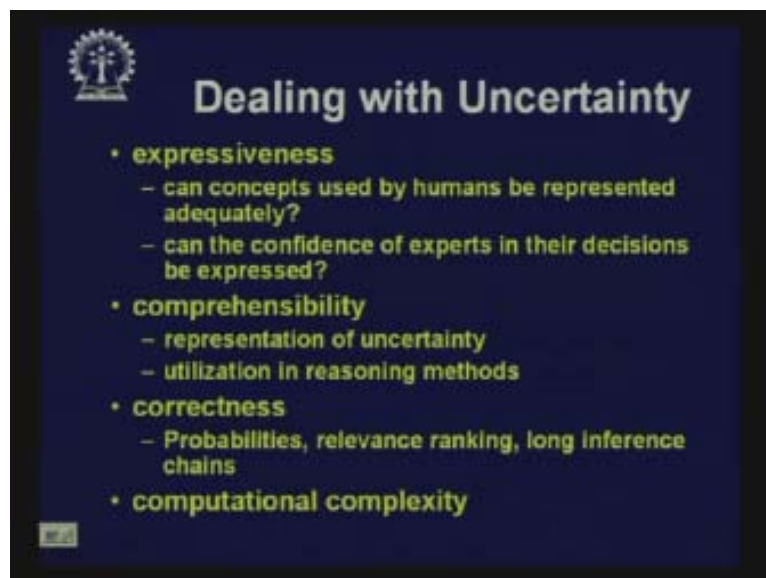
In the earlier lecture we had talked different sources of uncertainty, here we are looking at the same issues but they are being presented in a different way. One issue is expressiveness. We have to deal with uncertain situations because of our inherent

limitation of expressing very concretely some of the facts that we deal with. So the question is can concepts used by humans be represented adequately? Can the confidence of the experts in their decisions be expressed?

The experts take some actions. When they are faced with the situation they apply their intuition, their common sense, their experience and take some actions. But what really led them to take that action is often not very clearly expressed. If you recall while discussing about knowledge acquisition we mentioned the same thing that it is very difficult to extract expertise from the experts.

Now when the experts take a decision under an uncertain situation how do we ensure to extract the exact mechanism or confidence or the probability whatever the expert applies. The second issue is comprehensibility. How do we represent the uncertainty? First of all that uncertainty has to be represented and secondly that representation must be such that we will be able to utilize that in the standard reasoning method like Modus ponens or rule based inference or whatever. If you recall we will not call an organization of information to be knowledge unless that is usable to perform inferencing. The same is true here.

We have to represent the uncertainty in a way which can be reasoned with it. The third factor is correctness, how correct is it? When we take probabilities obviously anything that is having a probability of 0.7 is not known with certainty. So if I take this 0.7 probability as a measure or as a key factor to take any decision how correct would my decision be? A probability of 0.7 means, in 70% of the cases it is true but in 30% of the cases it is not. So obviously there is a chance that the decision that I take will not be correct if it falls in that 30%. Therefore the probability must be very reliably calculated and gathered from a huge set of data.

The other thing is relevance ranking and long inference chains. We have seen in our earlier discussion about certainty factors that as we propagate through rule chaining for example a implies b, b implies c. As we propagate or proceed through a longer chain and when our starting point a is known with some uncertainty which means not 100% certain and a implies b is also an uncertain rule then obviously b is inferred with a lesser certainty. So some error may have got in when I assigned that uncertainty in the rule or in the fact a, we will propagate b. Then when I use the rule b implies c and take a uncertain decision about c with some belief then if there were some error that error will be propagated. Therefore that is another issue.

Another important thing is the computational complexity. We may like to represent uncertainty in different ways but ultimately the computer will have to take the decisions. So the computational complexity maybe of the order n or n square whether it is polynomial or whether it is exponential all these issues will rise because we have to take the decision in sometime. In some cases we have to take it in real time but in some other cases also there must be a finite time we cannot spend a lot of time to just make a computation for one uncertainty handling and there will be so many other uncertain cases and for every time we will not be able to devote that much computational time. Therefore computational complexity becomes an issue because we want to have fast decision making.

(Refer Slide Time: 10:00)

The other source of uncertainty is data. Missing data: Some data may not at all be available. That is, suppose we have got sensors put in different places but there are some very difficult places where it is very difficult to approach and put in a sensor then we will not get data from there. There may be that we are collecting data from hundred sensors but some of the sensors fail and we are not getting data. But in the absence of that data also we have to carry out the reasoning. This is known as missing data.

The other thing is unreliable data. The data may be unreliable because when we receive the data the data is propagated through some channel and that channel may be noisy. For example, you are carrying out a blood test and whatever machine you are using for the blood test is not very reliable so you get a gross idea of the blood report but that is unreliable. In some cases it is ambiguous and imprecise. We do not very clearly specify those.

We have already seen some expressions like very tall, quite big, sometimes, frequently, seldom etc are some of the statements we make which are not precise, seldom, often, how often? How many times in a day? How many times in a week? Sometimes the data may be inconsistent. You may get the data or the information from multiple sources and the data may be inconsistent. Sometimes the data may be subjective and sometimes the data may be derived from defaults. Now this requires a special mention.

What do we mean by default?
Default means assuming something when nothing is known to the contrary. If you recall in the passing I mentioned about abductive reasoning, and deductive reasoning has been discussed. We have also mentioned about abductive reasoning. There is another very interesting type of reasoning called default reasoning.

(Refer Slide Time: 14:02)

What does this default mean?

Default means assume something like x to be true when nothing is known to the contrary. That means I do not know, I do not have any evidence, no evidence is right now available is right now available that contradicts x that says not x, there is nothing such. So when nothing is known to the contrary then assume something. That is called default reasoning.

Here is an example; Fido is a dog and you know dogs have four legs. Therefore I can infer that Fido has four legs. Now if I know that Fido met with an accident and had one leg amputated then obviously this reasoning will fail. Now I am feeling tempted to ask you what is this sort of reasoning?

Fido is a dog so I can say dog Fido and dog x implies four legs x from there I infer four legs Fido.

What is this form of reasoning?

This example relates to Modus ponens which looks like P implies Q and P is true so we assume Q is true.

(Refer Slide Time: 16:02)
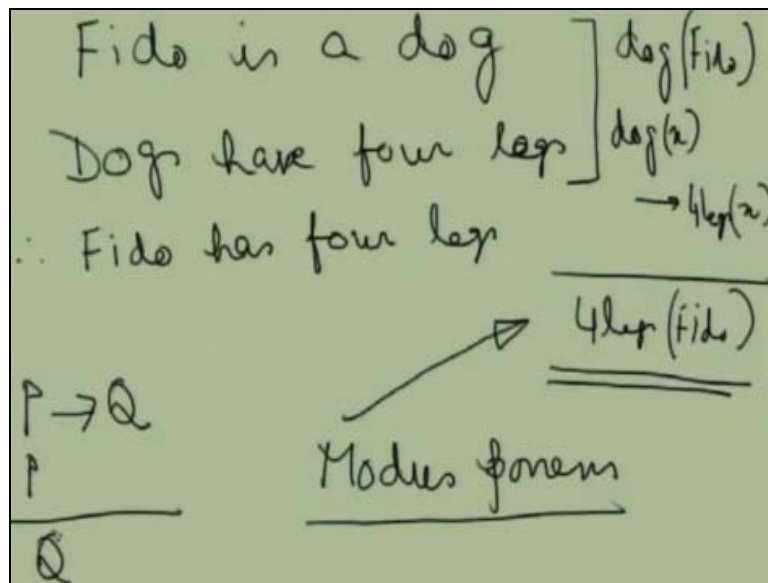
Now if I had known that Fido met with an accident and had one leg amputated then obviously this would have failed. But in the absence of that knowledge I will certainly apply this rule and assume that Fido has four legs. Now say I have got some facts from the domain of cricket matches, so I want to know about batsman x and I have got some details about batman x, I know average runs scored by him, number of centuries he made etc.

Suppose in my knowledge base I do not know whether he is a right handed batsman or a left handed batsman, it is not known nothing is there. Now on a sudden a query comes as which hand does x uses while batting?
Normally given these facts I will not be able to answer these questions because I do not have that information.

However, if my reasoning mechanism says that for any such query if nothing is known if not known assume right handed. Then with this power I will be able to answer this question and I will say right handed x. However, this may not be true but this is the default reasoning that we do because might be now if here it is said left handed x then obviously this information which is present will overwrite this default and this information will not be holding any longer. This is what is known as default reasoning. So whenever we take some conclusion based on default reasoning obviously some amount of uncertainty mixed in that.

The second point we are coming to is expert knowledge. As we are mentioning time and again there may be an inconsistency between different experts. Often the experts talk about plausibility that is a best guess of experts, best guess so again it is not a certainty.
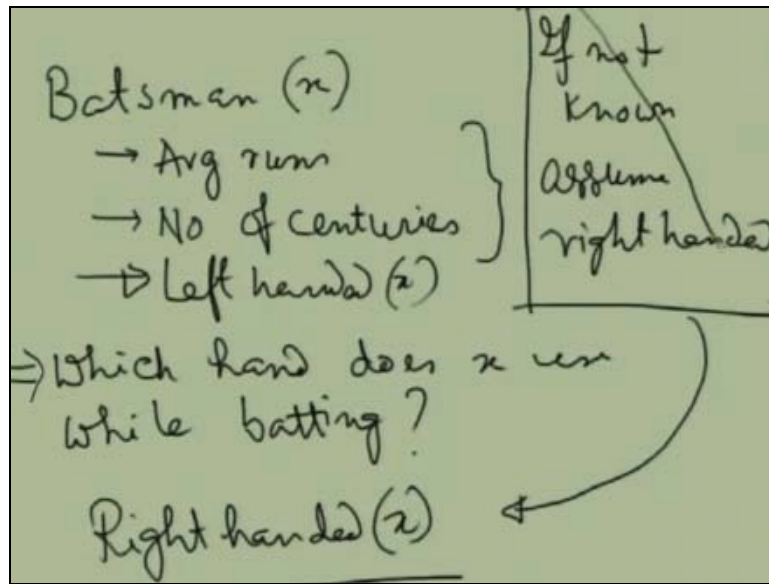
The other important factor is the quality the causal knowledge. If we know the causality the cause effect relationship, for example the resistance of an electric circuit affects the current that is flowing and also the voltage causes the current to flow. Now this is the effect and if we really know for example this causality is known by the well known law of Physics the Ohms law. We have got a deep understanding of the causality the relationship between the cause and the effect. But in many cases that is not true. We do not know the area to such depths.

For example, psychology we are still trying to model the human brain. We do not have any clear grasp on how the mind or how the psychology works so we cannot make a very strong causal commitment over there. Instead in many cases where the causal knowledge of the deep understanding is not available we go through statistical associations which are observations. We look at several cases and based on that we carry on our decisions.

Knowledge representation: This is another issue. How do we represent knowledge? When we discussed about knowledge base systems we said that we cannot really make a complete knowledge base. Therefore in order to make it really good we have to narrow down the domain so that we can go deep into it. Hence we actually work with a restricted model of the real system. The actual system may be very diverse but we are working with a restricted version a narrow version of that and limited expressiveness of the representation mechanism.
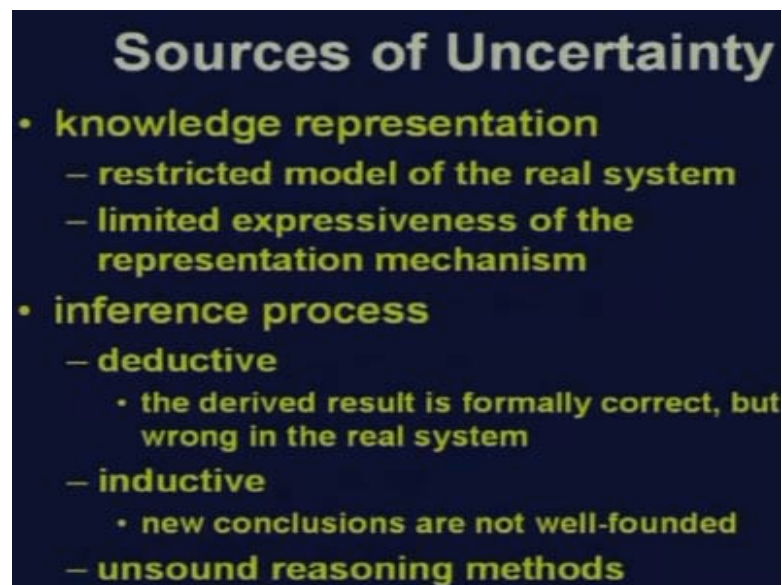
Inference process is also another important issue. We know what deductive reasoning or deduction is. We know that there is sound so the derived result is formally correct but we may find that it is wrong in the real system because the rule may itself be wrong. There is a subtle difference, my reasoning is right but my rule is wrong. So again coming back to our Modus ponens P implies Q and P I infer Q this inference process is fine, this inference rule is correct but the correctness of this result depends on the correctness of this rule. If this rule is not correct then obviously this inference also will not be correct. So there is a difference between the soundness of the inference process and the correctness of the result. Even if the inference process is sound the result may not be correct unless the axiom or the theorem or the rule that I have written over here is not correct.

Inductive reasoning: The new conclusions that we get are always not well founded. Again the different inference mechanisms with which we are by now familiar is deductive the other one is abductive.

Now I introduce another one that is inductive reasoning. And actually it is not new because you will see that day-to-day we often use this. For example, to a child I show one ball a particular ball that is a round shaped object. So, a_1 is some round shaped object and I tell the child that it is a ball so the child associates the round shaped object with ball. Next I show him another ball which is another round shaped object of a different color and I say this is a ball. I show him another might be a football but the earlier two were smaller balls like cricket ball, hockey ball etc, now I show him a football which is also a round shaped object but only the size is different and I say this is a ball. In this way I go on.

Gradually the child learns that for any round object that is shown to him is a ball. Therefore all these are instances and based on that the child learns a rule that is for all a that is round for all round objects a_i implies ball. Now I show him an apple or an orange which is round a well shaped orange and the child will infer it as a ball because he has inferred that all the round shaped objects is a ball because I have

shown him large number of instances n number of instances and from there I have made a generalization. Now if these instances are proper then for most practical purposes this will be a very useful way of reasoning.

How do we infer? How do we learn?
Actually most of the things we do is through induction. This is the process that goes on. So this is known as inductive reasoning.

(Refer Slide Time: 28:18)
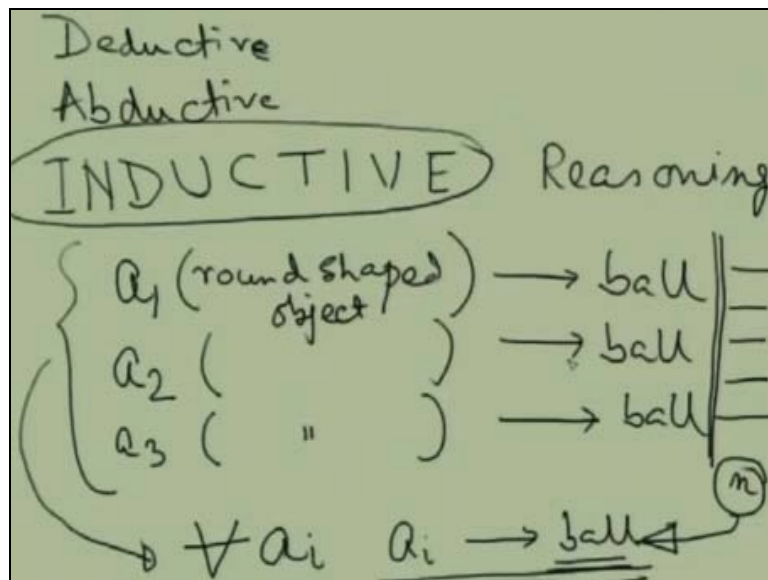
However, as we have shown here this is not sound because there may be counter examples. The same unsoundness was there in the case of abductive reasoning also. So inductive reasoning in that the new conclusions that we arrived at are not always well-founded.

The other thing is individual rules. There can be errors in the individual rules, there can be representation errors, inappropriate application of the rules etc. Now we have already seen this case that the likelihood association of the evidence we have seen just as we did in the case of certainty factors we had put in several certainty factors with the premise for the conclusion and the combination of evidence from multiple premises. All these examples we have already seen while discussing about certainty factors.

In Rule Based System there can be another issue like uncertainty can come up when we have got multiple parts and there are multiple rules to use. Therefore here we bring in conflict resolution which means from the set of multiple rules we select one or two and apply them. Now, the one which is selected might give rise to some sort of uncertainty. The compatibility, the contradiction between rules, the subsumption where one rule may be more general than the other one, there may be missing rules. When we collect the data from multiple sources of different reliability we will again get a varied result, the reliability of the ultimate fused data set will not be the same. Now, in order to handle such complicated scenario a probabilistic approach is required.

Basics of probability theory:

Probability theory is a mathematical approach for processing uncertain information. And we talk about a sample space. What is a sample space?

A sample space is the set of all possible events. For example, when we deal with a dice when you toss a die which has got six faces the possible outcomes are 1 2 3 4 5 6 so that is a set. It can be discrete or continuous. For example, when I think of the possible temperatures that can be there tomorrow the typical weather report what will be the maximum temperature?

It can vary in a continuous range. So this sample space can be considered as a set.

What is a probability?

Probability is a real number $P(x_i)$ where x_i represents any of these events. So we call these to be the events and any of these events is associated with a real number which denotes the likelihood of the event to occur. And as you know it is a non negative value between 0 and 1 and total probability of the sample space is 1. These are the basics and here is an example. We can always think about it as a set. It is a set of possible events in the case of a dice throwing. The possible events are 1 2 3 4 5 6. So when I throw a die any of these discrete events will occur. Either 4 will occur or 5 will occur and each of them have got a probability associated with it. Here this is x_i and if the dice is fair then you know that the probability of an event x_i is 1 by 6 in this case because all of them are having equal probability. And the total probability of the sample space is 1. That is, when all these are added together that will be 1.

Basics of Probability
- for mutually exclusive events, the probability for at least one of them is the sum of their individual probabilities
- *experimental probability*
  - based on the frequency of events
- *subjective probability*
  - based on expert assessment

Now for mutually exclusive events, what is meant by mutually exclusive events?
Two events cannot take place together. When I throw a coin it cannot be head and tail, it will be either head or tail. The probability of at least one of them is the sum of their individual probabilities. The probability can be of two types namely experimental probability or subjective probability.

Experimental probability is based on the frequency of events. If I carry out a fair experiment of throwing the die or if I take a biased coin and throw it ten times may be it will not be five times head and five times tail. If it is biased then may be I will get seven times head and three times tail. But with that experiment being done quite a number of times we can come to a fair degree of association of the frequency of the events. The other thing is the subjective probability that is based on the expert's assessment of the whole thing. The next thing is compound probabilities.
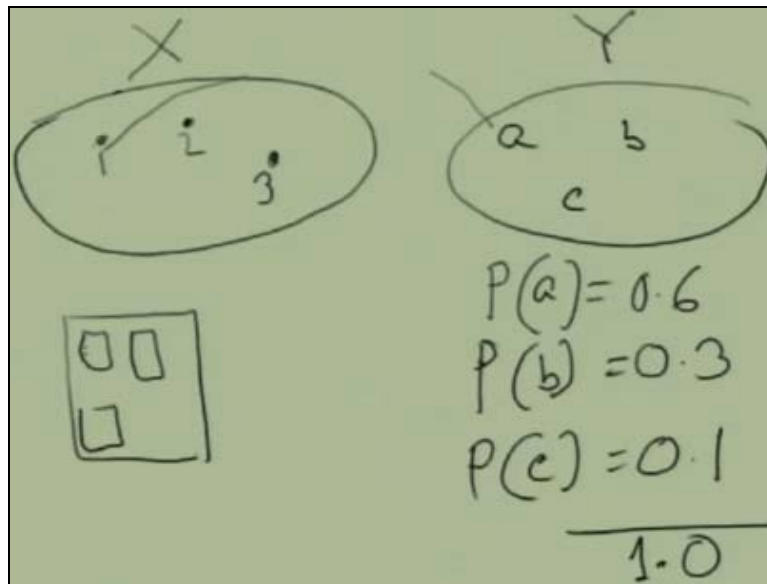
Compound Probabilities
- describes *independent* events
  - do not affect each other in any way
- *joint* probability of two independent events
- A and B
  $$P(A \cap B) = n(A \cap B) / n(s) = P(A) * P(B)$$
  where n(S) is the number of elements in s
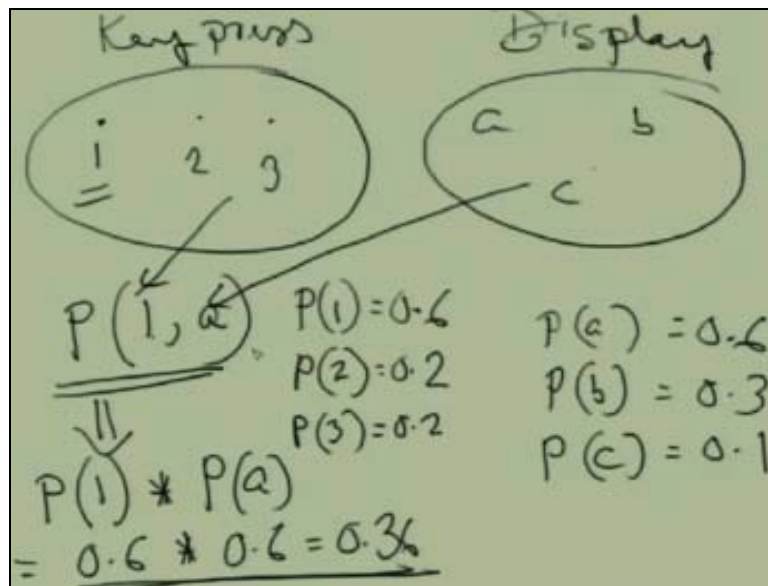
(Refer Slide Time: 37:00)

In the last lecture we tried to explain the idea of independent events. Suppose there are two sets here the possible outcomes are 1 2 3 that is my space X and there is another space Y, when I press the bell 1 as this is a sort of a button or I can press any one of them, and here there are some other events say a b c that will appear as I press a button. Now if there be a direct association, that if 1 is pressed then (a) will be displayed, when 2 is pressed (b) will be displayed then there is a direct causal relationship between these two and they are not independent.

On the other hand, if it be the case that both the appearances of 1 2 or 3 or a b c are absolutely random based on their own probabilities, there are some probabilities of P(A) may be 0.6, most of the time A occurs P(B) is 0.3 and P(C) is 0.1 so all these are adding up to 1 for this space. Most of the times probably I will get A irrespective of whether I press 1, 2 or 3. So in that case the appearance of the pressing of the button 1, 2 or 3 and the appearance of the symbols a b c are absolutely independent, there is no causality among them. In such cases the probabilities do not affect each other in any way.

The joint probability of two independent events A and B is given by P
probability of A and B that is the number of intersections A and B divided by the total number of elements in the set and that can be computed as the probability of A times probability of B. For example, again thinking of this same example 1 2 3 and a b c this is the display that will be displayed and this is the key that will be pressed. Now I can compute the number of times that 1 as been pressed and a has appeared. Now, if this be random it has got its own probability. here we said that P(a) was 0.6, P(b) was 0.3 and P(c) was 0.1 and similarly here we have said P(1) is 0.6, P(2) is 0.2 and P(3) where 3 is being pressed is also 0.2 because that should also add up to 1. Then the joint probability of that 1 will be pressed and a will appear will be probability of 1 multiplied by probability of a occurring that will be equal to .6 times 0.6 is equal to 0.36.

(Refer Slide Time: 39:44)

(Refer Slide Time: 41:21)
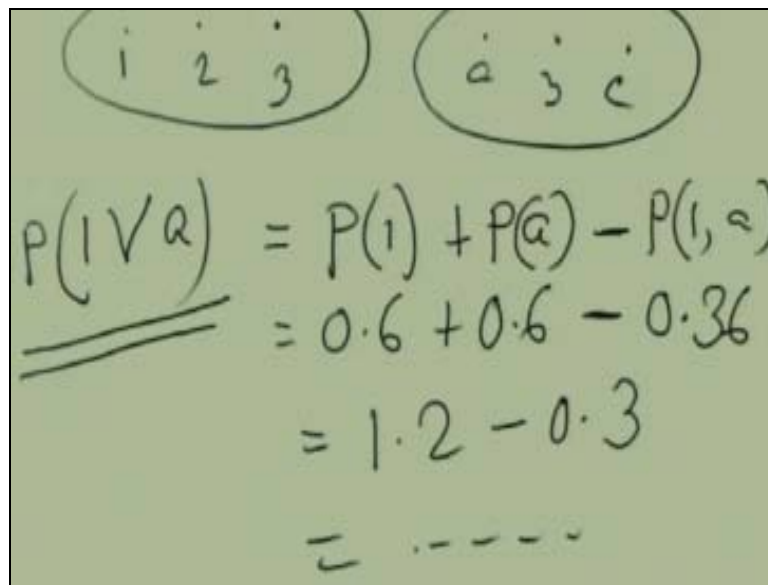
Now this will hold only if this and this are independent. They are absolutely independent, this is coming from one set and this is coming from an independent set. That is known as the joint probability or compound probability. Similarly, A or B, that is the union probability of two independent events will be P(A) plus P(B) the probability of A plus probability of B minus the joint probability when they are occurring together. In the earlier example they have got their probability 1, 2 and 3.

What is the probability that 1 will be pressed or a will appear?
Obviously 1 will be pressed has got the probability to be 0.6 as we have done earlier. The fact that a will appear has also got a probability 0.6. But there is a joint probability in that I am already considering the cases where 1 and a have jointly occurred so that was 0.36 so minus 0.36. Therefore it will come to 1.2 minus 0.36 and that will be the union of this. Basic axioms of probability:

Probability is a value between 0 and 1. P(E) denotes the probability of the event E. If the event is always true then P true is 1 that it is a certain event. There is no uncertainty involved in it.

(Refer Slide Time: 41:48)

(Refer Slide Time: 43:29)
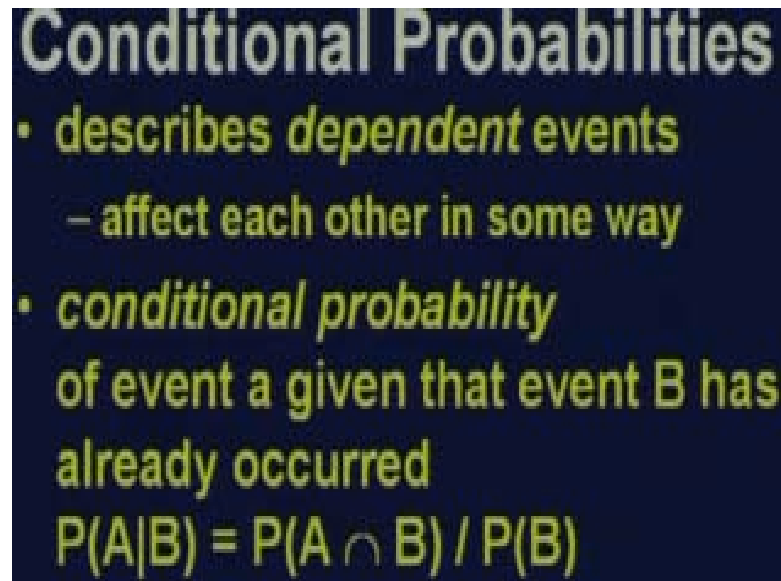
Similarly, P(false) is 0 that is it is an impossible event. P(A) there is a probability of A occurring is equal to 1 minus probability of 0A occurring. Since the total space is 1 then obviously the probability of A occurring must be 1 minus this. Similarly, again this example will tell you a b c and P(a) is 0.6, P(b) is 0.2, P(c) is 0.2 so what is the probability of 0a?
Probability of 0a is either probability of b or probability of c so that will be the union of this so it is 1 minus the probability of a. That is, basically if you look at the probability of b in this case the probability of c taken together so it will be 1 minus 0.6 is equal to 0.4. The basics are P, A or B. These are the basic axioms of probability

which will lead to use. Till now we were discussing about independent events. Now let us look into dependent events.

(Refer Slide Time: 43:48)

Dependent events mean they affect in some way. One of the ways they can affect is the causality. If the voltage across a circuit is increased the current flow will increase. There are other ways also in which our reasoning can affect. For example, if there is a murder then there will be some blood stains, so, if we see blood stains from there we can also reason back towards murder. So evidences also sometimes help in reasoning and that is very useful in the case of diagnosis. When we look at some fault we try to reason out the cause of the fault.

Now, if two events are dependent then that is the fact that one event has occurred will certainly influence the occurrence of another event. This is known as conditional probability. That is, conditional probability of an event A given that event B has already occurred that is given by P(A by B). Now here you will again find some similarity with our certainty factor case where we are time and again writing MB(h by e). Similarly, here we are saying that the probability of an event A occurring if B has occurred.

What is the probability of this?
Now B has occurred or B has got a particular probability of occurrence. In that case what is the probability of A's occurring?
That is known as conditional probability. And that is given by P(A) AND B. That is A has occurred and B has occurred divided by probability of B. It is P(A by B) is equal to P(A AND B) by P(B). Now there are two types of probability. Here you can see that the probability of A and B occurring is being divided by the probability of B. Now the probability of B is known A-Priori.

Here we have to talk about two types of probabilities. One is A-Priori probability. That is the probability of A-Priori and we also write as apriorie that is known before hand. P(B) we have computed irrespective of A. Irrespective of A I have computed

the P(B) through experiment. And suppose I find that in general B occurs with a frequency of 0.6. You go to a village and carry on some experiments just on the occurrence of malaria in that village. So P(malaria) you find that out of every ten persons six persons are inflicted with malaria. So the probability of malaria is 0.6.

Now we would like to see probability of fever P(fever by malaria). This may not be a very good example when they had malaria. When they had given they have malaria what is the probability of fever? How many of these people had fever? Assume that all malaria cases are not giving rise to fever. So in that case what we are trying to say it is the probability that somebody will have fever, he has fever and malaria and I find that the probability is 0.9 in those cases where he had fever and malaria. Now out of these malaria has got an independent probability A-Priori probability that every six people will have malaria. So they got fever because of malaria not be right. People can have fever for some other reason also.   So we compute this by dividing with the A-Priori probability of malaria. That is because they are not independent I divide it with 0.6 and whatever value comes up that is my conditional probability of fever given malaria.

(Refer Slide Time: 49:44)

(Refer Slide Time: 49:55)

This is known as conditional probability. A-Priori probability:
The prior probabilities of events A and B suppose we know P(A) and P(B) we know the prior probabilities. We know the prior probabilities of P(A) and P(B). But if we know A for sure then probability of B might change. If we know A for sure then this is represented as the conditional probability of B by A.

Now let us go back to the doorbell problem we had seen earlier. Doorbell rings at midnight so it is D. W is Mohan wakes up at midnight. And suppose the A-priori probability of the doorbell ringing at midnight is 0.001 as the probability is really less and Mohan is waking up at midnight for different reasons such as for the doorbell or for some other reason maybe feeling very hot or uncomfortable etc is 0.01. And we say that the probability of Mohan's waking up with doorbell when the doorbell rings is 0.8. Now this is a conditional probability. These are A-Priori probabilities that doorbell or no doorbell the probability of Mohan's waking up is 0.01.

For any reason the A Priori probability of the doorbell ringing is 0.001. And this is a conditional probability that if the doorbell rings then Mohan will wake up that is 0.8. So there is a difference between the types of probabilities. Now, in order to handle this sort of cases there is a very interesting and famous rule that is known as Bayes' rule that is used for computing the conditional probabilities. Probability of an event A by B that is if B is known then the probability of A can be computed by this formula as, here we are trying to find out the conditional probability.

Conditional of probability of A occurring when B is known. Now, if we know the A-Priori probability the reverse relationship I know the A-Priori P(A) and I know the A-Priori P(B) and I also know the A-Priori P(B) occurring when A occurs. What we are trying to do is a cause effect relationship. Then this is the formula that gives me P(A by B) that can be derived with P(B by A) multiplied by the A Priori P(A) by A Priori P(B).

Now this can be easily derived like this:
P(A AND B) is, what is this? It is P(A by B) times P(B) and again P(A AND B) is P(B A) (P A) if I just change it. Now from this I can say this one is P(A AND B) by P(B) and P(A B) is P(B A) by times P(A).

What is the implication of this?
Before going to the implication of this here is an example. A Priori probability of doorbell ringing is 0.001, A Priori probability of Mohan's waking is 0.01, A Priori probability we know that if the doorbell rings Mohan will wake up is 0.8 then what is the probability that when Mohan woke up?
He woke up because the doorbell rang. I want to find the reverse thing. Then by Bayes' rule we will simply use this P(W D) multiplied by the A Priori probability of this denominator. This is the event and this is the hypothesis. The event is 001 and that is divided by the A Priori probability of waking up because I have to divide by this because anyway that would have happened so I get a normalized probability of P(D W) which is 0.08. This is a very important technique and it has got far reaching implication. We will see what it really implied. We are given some event and from there we try to come to a hypothesis. And I really want to find out the probability that this hypothesis is true given this event.

(Refer Slide Time: 55:33)
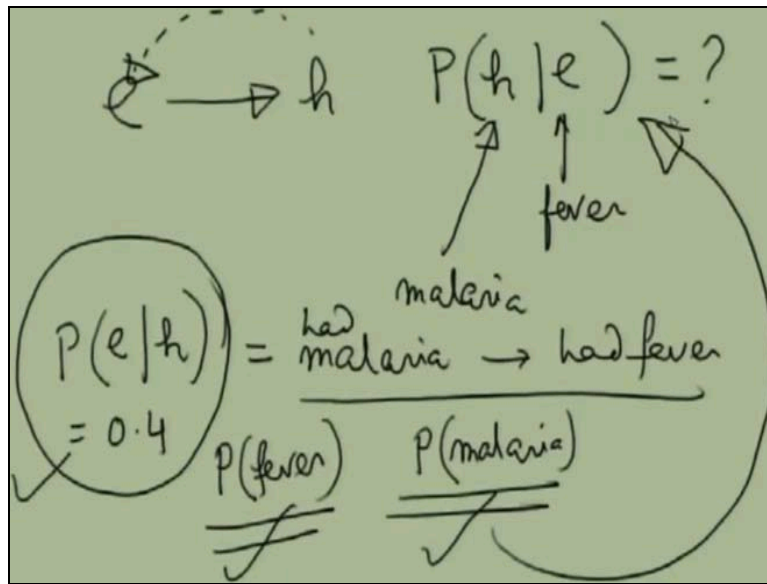
For example, if this event is fever and if he has fever he had malaria. What is the probability that if he has fever he will have malaria?

I want to find that out that fever is implying malaria. I want to find out the probability of this. Now I know P(e by h). What does it mean?

This means that somebody who had malaria had fever also, how many times it happened?

I find that this probability is something like 0.4 so I know this thing A Priori and I also know the probability of fever in a particular locality, I also know the probability of malaria in a particular locality and these are the A Priori probabilities. So using this one, this one and this one I am trying to find out this. So I know these A Priori probabilities and I also know that if somebody has malaria then he has fever this part is known. So, from the knowledge of this I am trying to derive this and that is being done using exactly what is being done in the case of Bayes' theorem. And this conditional probability is very important for all purpose for the simple reason as we will realize that it is a way of dealing with these rules that e implies h. This is nothing but a rule and I will have to reason through these implications and if I know the probability of this then it will be a much stronger measure than what we had in the case of certain defectors which was rather a subjective measure. So, Bayes' theorem forms a basis of probabilistic reasoning.

In the next lecture we will start with the Bayesian reasoning techniques using this Bayes' theorem. And one very popular powerful tool that has been designed for reasoning in the probabilistic domain is the belief network. But this forms the basis of the basic probability theory requirements. The main thing is the idea about the conditional probability and the Bayes' theory.