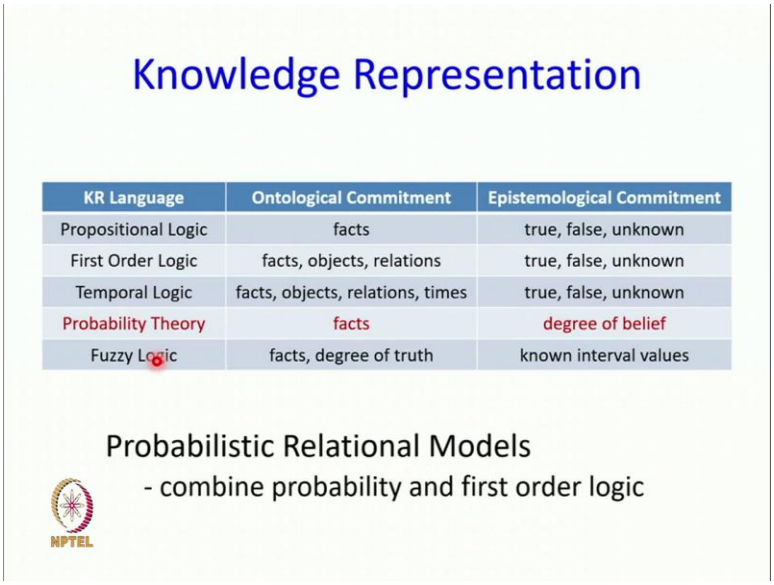
Artificial Intelligence
Prof. Mausam
Department of Computer Science and Engineering
Indian Institute of Technology - Delhi

Lecture-51
Uncertainty in AI: Basics of Probability, Part-1

Alright get started. So, today we have reached the point where we are going to move from the traditional AI to modern AI. I do completely different kind of AI before starting this class. And we will continue the language of knowledge representation right, the theme of knowledge representation that we have been working on.

(Refer Slide Time: 00:45)



So now, let go back to what AI have you been doing we have said that look state is very important in AI because state models the world, state models the problem, ah situation and then we can use that state as a fundamental unit to move in the search space and so an so. And then at some point we said look state cannot be indivisible, their properties in the state, we need to somehow those expose properties in a state. Moreover, in any problem there is a knowledge that the machine must have and that knowledge needs to be imported one way of the other.

And on the basis of that if we import such knowledge, the machine might be able to have some more inferences on knowledge. So then we started talking about some languages and we talk about constraint satisfaction is one of the, language where we have the variables in values and

that becomes part of the state. And then we talked about propositional logic where we are modelling facts and we always know what is is the fact true, is the fact false.

And if you do not know when the fact is true, or the fact is false, you say that I do not know the value of that particular fact and so. And then, there is something called first order logic where we have relations and objects and facts are formed on the basis of relation supply and object. And again, we know that whether particular fact is true, false or not. And then there may be many different kinds of logics as we have earlier discussed like the temporal logic, we add notion of time.

There may be Fuzzy Logic where we say that my fact itself may not be true or false. My fact itself may not represent something that is always true or false represent something that is the fuzzy that is too simple. And we may know it and you may not know it and then we said that there is another way of thinking about life which is probabilistic. And in a probabilistic world, we are still modeling facts just like propositional logic.

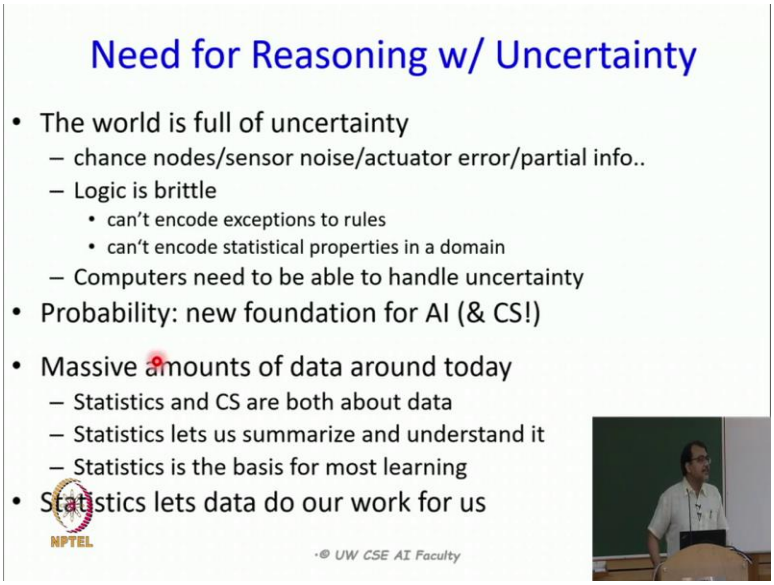
It is not as if the other the commitment on what we are capturing in terms of symbols and so and so and that has changed significantly. What is changed is what we know about it right? Fact with true facts may be false, but when we do not know a fact we do not say we just do not know it, if we do not say that we it is just say unknown, we additionally attribute some probability, some degree of belief.

So when we do not know it, we say I do not know it, but I believe it to be true with the probability in a 0.9, 0.1 in a 0.5 and that allows us to do more fine grained reasoning more fine-grained in Science and as we shall see that became the basis of the modern AI. And its importance and but I want also point out that like you know we have in a first order logic which has facts object relations with true, false.

We also have first order probabilistic models also called probabilistic relational model also called statistical relational learning which combined of first order logic and the whole of late 1990's and early 2000 was the age in a region here. There was so much work that is going on in

the mining first order logic and probabilistic models. However, we would not go there. Actually, I would say that jury is still out that in the modern deep learning world that is still as important as it was considered it to be notorious and we do not know the answer to that, only time will tell, right.

(Refer Slide Time: 04:18)



Next thing I would say is that is reasoning with uncertainty is incredibly important as I was saying that this begin the basis of modern. The first thing you will realise the world actually is full of uncertainty not this is a Complicated question. Is the world really full of uncertainty when it toss the coin I said that is probability half I am going to get heads half I am going to get tails. Is it really what that is happening. Is it a problematic outcome in the real world?

How many believe that this is really a probabilistic outcome? Nobody will take outcome one probability people please raise your hands and how many people believe that is actually a deterministic process. Guys you can see many more people feel that my tossing the coin is actually deterministic process. Of course, they know that I do not know what the answer is going to be.

But if you could model all the Physics, if you could model exact coin, if you could model the exact impulse that I gave, the exact angle at which I give the impulse, the exact height from the ground when I in a toss the coin the exact molecules in the air et cetera et cetera, at some people

believe that we would exactly able to protect what the outcome is. So people may believe that the world is actually deterministic. But is the world really domestic? Why not?

Quantum right then, it may go to some atomic level then, you may start to believe it we do not know if the cat is alive or dead and the world is actually now what I have done for you is that I have started from a world where we said that it is probabilistic. Then said no, no the world is actually deterministic and then said, no the world is actually cowardice. Which of these is the truth is a very hard question, right.

And you can philosophize about that is up to you. The main point I want to raise here is that depending upon the abstraction at which we model the world we may or may not be able to capture the determinism in it even if it were present. Like for example let us believe that the coin tossing is a deterministic event, the amount of modeling I need to do the exact end molecules the exact exact this and that it is just not going to visible to practically model always.

What might be easy to say whether I am not going to model any of it or I will simply say you toss a Coin and already it is 0.5 probability to get heads in it. Why where you think it is 0.5 probabilities to get heads? The probability to get heads is we are assuming that the coin is fair and you know if a million trails, then you are very, very high chance that answer will have a relatively even distribution. There is no bias on one side of this heads.

Therefore, this leads to the very famous saying, very famous quotation that all models are wrong. I have said this before. Some are ok that is not the same. You can make your saying that all models say some are less wrong. That is not technically invalid as a saying. Now, it is a saying it is IIT class of Delhi is saying, but the famous thing is that all models are wrong, some are useful. Right answer you often ask the question, how much do I model?

Do I model all the complexity of the world? Sometimes, all the little complexity is very hard to take care of, because you do not have enough data to learn for it you, there so many corner cases and exceptions and you cannot enumerate all of them and so on and so forth. And therefore, I will come back to you again in class, therefore, we abstracted out we said that we will model the

world at some level abstraction and for all this corner cases in some randomization or some probability distribution. And this is exactly why we say -0.5 heads, 0.5 tails.

Moreover what I will show you later is a logic is brittle it is not easily possible to import exceptions to rules. It is definitely not possible to import statistical properties in a domain. And because of all of this, the probability became a new foundation for AI some people believe you know for computer science. By the way, the world itself is full of uncertainty, if I have a real robot in a risk, noise it is trying to say the obstacle is 4 feet away but there is some randomization some uncertainty is there et cetera et cetera.

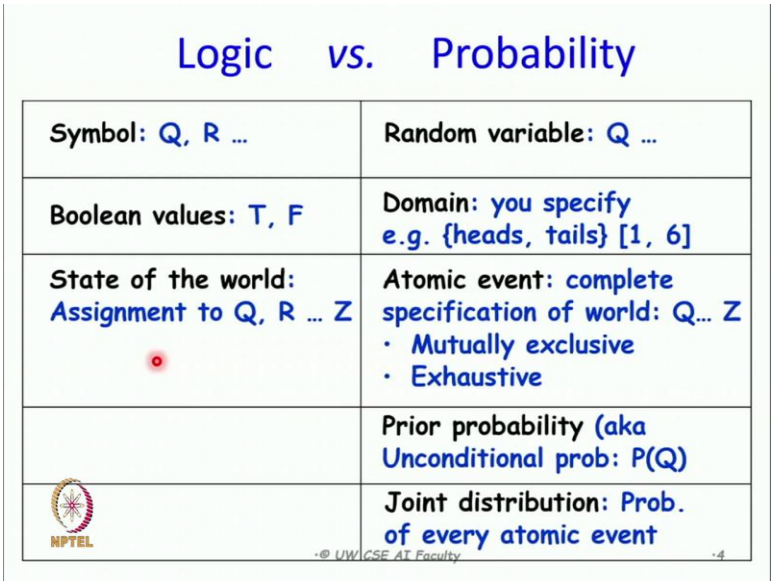
Many kinds of uncertaintie, we have talked about expecting minimax. That is the uncertainty in the chance node. What happens this happens that happens if there is a noise, which is reading uncertainty. It is actually I tried to remove my hand four feet but Only move 3.95 feet right it is possible. That will be in the motor rotation and I may not even know all the information. The uncertainty what I do not know like, I do not know where Kashika is today she is somewhere but she is not here today.

I do not know where she is right. The point is that is partial information. We do not exactly know everything in the world and survey them because of all of this partial information is reading and observation information on an uncertainty, action uncertainty et cetera et cetera we have to deal with it one way are the other. The other is an uncertainty becomes critical is that data has become very central in the field of AI.

When you relate to a machine learning course in the whole of the class will be about you know, how do I use data in a model in AI right. So, massive amount of data today, statistics and computer science are both about data. Cases of today let us summarise and understand it. It is the basis therefore for most learning, education learnt from data, we cannot set deterministic thing. We cannot be sure about everything we can only make probabilistic statements because it is possible that there are some data we have never seen before right. And so because of that, we need to somehow captcha problem.

There are many, many reasons why probabilistic AI has more real world applications than the pure logic based.

(Refer Slide Time: 11:17)



And in the class today, we will just quickly go over the very basics of probability, which you should already know, right? We will; may be introduce one important concept that it is going to be whole revision, most of it is. But if you have questions please stop, ok. So in logic we have symbols in Probability we have random variable. Random variables are what? Are the units of probability? Symbols can take values true and false, random variables may not it will take any value like from a domain, for example heads and Tails, a set of 1 to 6, True false also.

State the world was always the complete assignment to all the variables right. This is where in the logic came into picture is said you can go inside the State, there are many state variable standard state variables take values true and false in effect completely enumerate all the variables then that is the state. In the very same way probability folks used the word atomic event. The same thing as a state there is no difference.

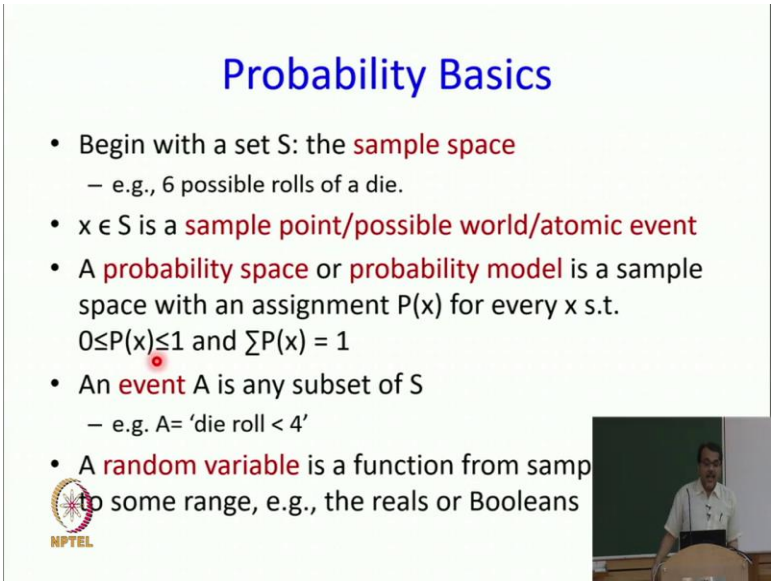
It is again a complete specification of the world, complete assignments to all the random variables. And this complete assignment is called an Atomic event. As atomic events if all these words are mutually exclusive that means two words cannot go a pair for an exhaustive, like if I can enumerate all the possible atomic events and states then I am not missing anything. In

addition we will need 2 new or 3 new definitions of probability which are very easy to understand.

One is prior probability distribution, one is going to be joint probability distribution and later when is posterior probability distribution. These three terms are very, very easy to learn, ok. These are very very easy terms. For example Prior probability of any random variable or any event may not be an atomic event may be any event is just its unconditional probability. I am not given information and I am saying what is the probability of a certain event?

On the other hand, joint distribution is probability of a state, probability of full specification of the atomic event. Full atomic, probability of atomic event is called the joint distribution and probability of any event without any additional information is called a prior Probability and again, if it does not get clearly will give you more examples.

(Refer Slide Time: 13:44)

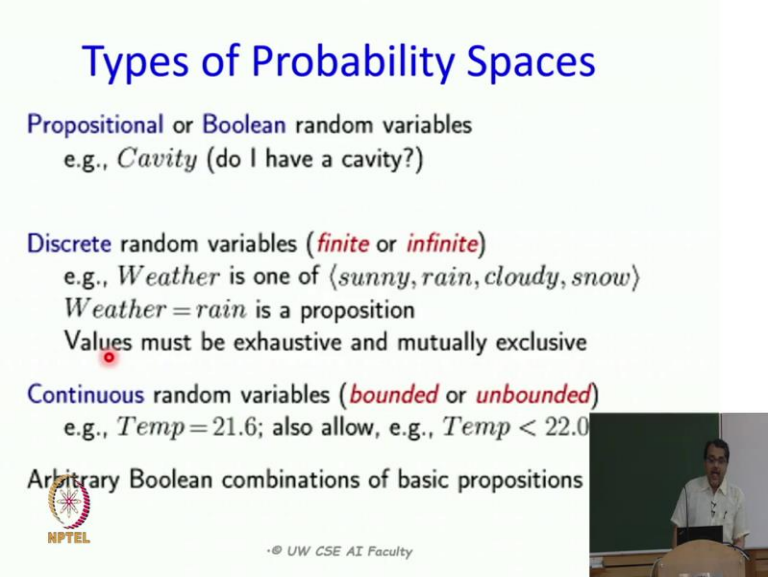


So to make it look at some examples that say I have all I always have a sample space. The sample space is all the possibilities of a random variable. For example, there are 6 rolls of a dice there the samples will be 1 2 3 4 5 6. And x would be a sample point, a point possible world atomic event. A probability space or a probability model is a sample space in assignment P of x for every x such that each of the P of x between 0 and 1 and some over all P of x is .So that an

event is any subset of S so, die roll less than 4 is an event is not an atomic event but it is an event.

And a Random variable is the functions of sample points like 1256 to some range, like range of real's or Booleans and sorry. If I suppose I have a sample space of red, blue, green, then, if we have a function that takes give red the value 1, blue the value 2 and green the value 3, then, that will be called a random variable.

(Refer Slide Time: 14:50)

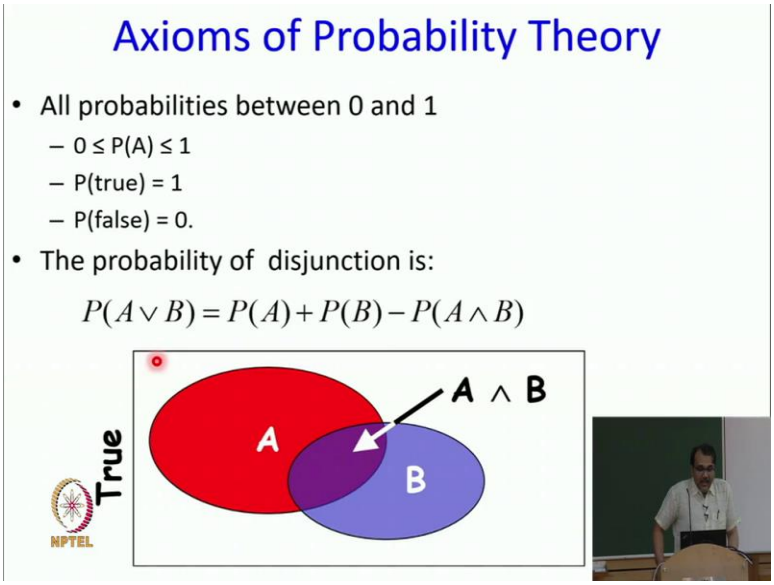


And my probability spaces maybe propositional like Boolean for example, do I have a cavity in my teeth? We will be using Mr. Bean's dentist appointment as an example today. Demonstrate it in a few minutes. The random variables may be discrete like is it sunny or rainy or cloudy or snowy? The random variable may be continuous is the temperature 21.6, is it 22.0, or India is the temperature 31.2, 32.0 and these are continuous.

They may be bounded, they may be unbounded right as the global warming is going on its sort of slowly looking like unbounded but we shall see. May it will not be unbounded because we will all die before the bound gets, that is also possible. It's a Joke look like you are not following the class today so did it to laugh. Too serious of a joke for you to laugh it is not clear whether you will live your whole life to see you know the earth at that level.

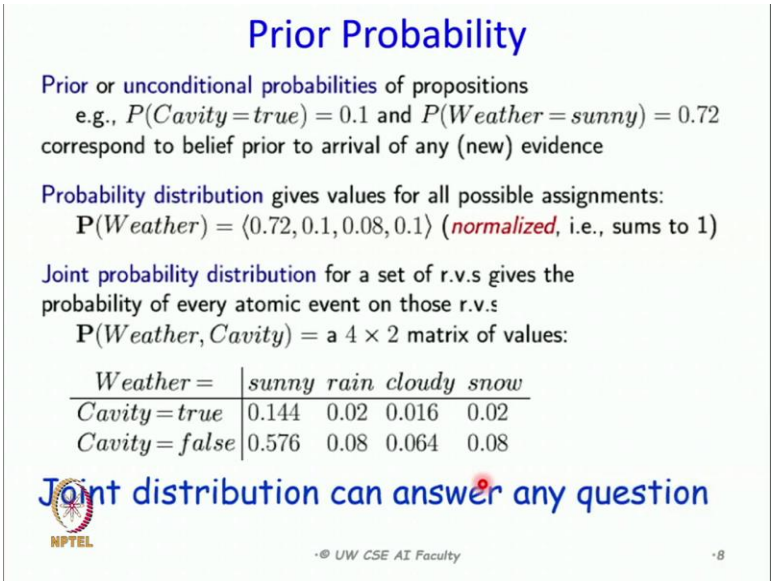
I hopefully will live my life. Ah you know you must think about how to use AI for global warming at some point. But you cannot do that in the U.S. because U.S. does not believe in global warming. So you have to do it in India or somewhere else Germany ok, good. Atleast you are here.

(Refer Slide Time: 16:21)



So, and then we know this basic probability of two events A or B, is probability of A + probability of B – Probability of A intersection B. I am pretty sure everybody knows this probability of A capturing this red space, probability of B capturing this blue space. One common space that has been captured twice you have to subtract it. Ok.

(Refer Slide Time: 16:44)



So, now this is what is prior probability. Prior probability is also called the unconditional probability of any event, right? So what is the probability that Mr. Bean had cavity? What is the probability that the weather is Sunny in the month of September in Delhi? Reasonable probability 0.726, reasonable probability, Prior probability corresponds to our belief prior to arrival of any new evidence. For example, in the case of Mr Bean suddenly he may observe that in suddenly one of his tooth may sort of moving or he may observe that he has tooth ache.

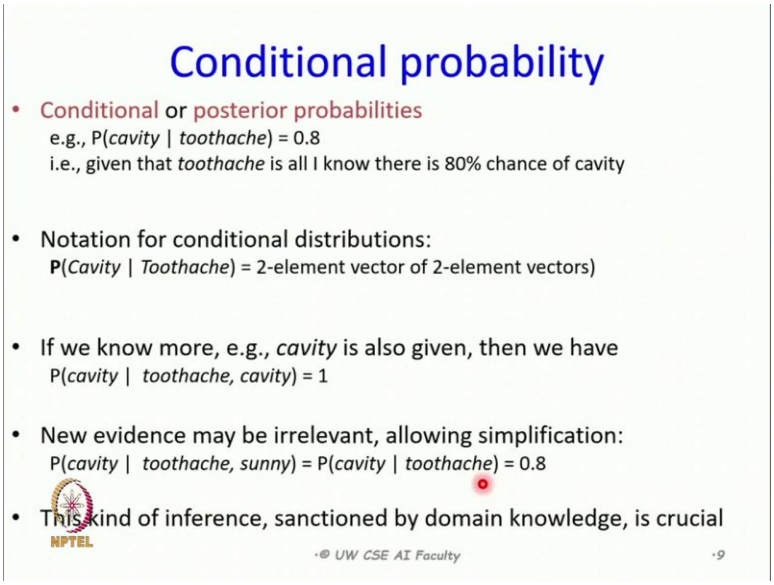
And that new observation, that new evidence might change his distribution of whether he has cavity or not. Make sense. But suppose I do not give you this new evidence of any new evidence then what is the initial belief before you get any information of whether Mr. Bean is cavity or not that is called the prior probability. Prior probability is the probability prior to seeing any evidence, ok.

Now, probability distribution is the probability of the whole distribution of all the possible events, atomic events and their values like for weather. So you may have a snowy weather, cloudy weather, rainy weather, sunny weather. And let us say the probability distribution would sum to one obviously and that would be all the values may be called a distribution. Now, suppose, I want to model 2 random variables in my domain, one is what is the weather, one is whether Mr. Bean has cavity or not.

Then I can define a joint probability distribution. A joint probability distribution would be a set of random variable given a set of probability for every atomic event on those random variable, right. For example for probability weather, cavity I will have 4,2 matrix of values for each possible weather condition and each possible cavity condition. For example, this might look like the distribution. You may say that whether Sunny and cavity true is 0.144, weather cloudy and cavity falls is probability 0.064.

Now, what is very interesting is that we will very soon show that joint distribution can answer every question that you may have about this probability distribution. And if it is not clear to you it will be clear to you in a few points. Joint distribution is what we want. If we have the joint distribution, we have everything.

(Refer Slide Time: 19:43)



But before that we have to define a conditional Probability again you know, what is conditional probability. It is also called the posterior probability. Prior probability was before seeing any evidence; posterior probability is posterior seeing some evidence. It is the conditional probability conditioned on the evidence that you have seen. What is the probability of the cavity if I am given 2 things? And that is the; and that may be very high because your tooth ache may have a good chance.

Question answer when in the previous probability we had the probability of whether it is Sunny and and so on so, can we call it the prior probability? Yes, any prior probability, which is without giving any new evidence. So, when you are not giving any evidence we will call it a prior probability. However, a prior probability which is giving a value for all possible random variable, the full state is also called the joint probability.

And notice that the probability of cavity given to tooth ache is a two element vector of two element vectors because tooth ache maybe true or false and for that cavity may be true or false. For toothache true, I have two values; for toothache false, I have two values. Of course, those values independently given tooth ache is true, the probability of cavity and probability of not cavity sums to 1, sums to 1. I can get away by only giving one value I can use this to compute.

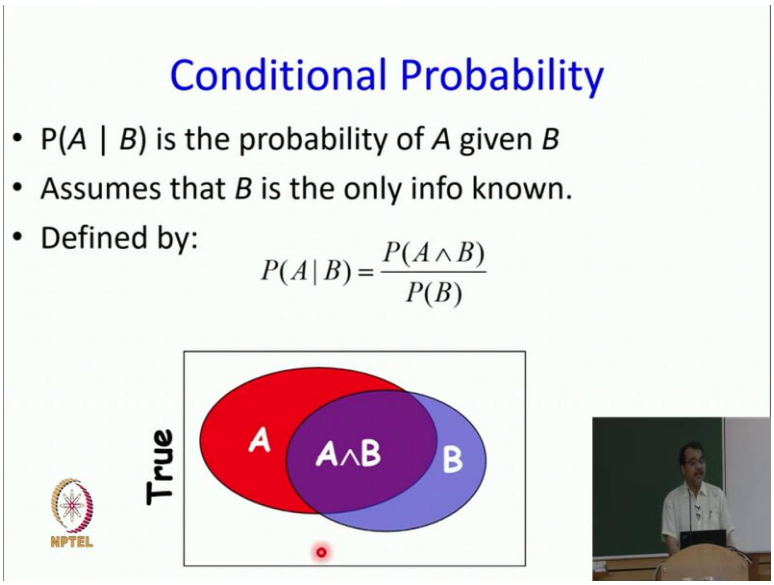
And now one of the most interesting things that is going to happen in our world and we will talk more about this very soon is that new evidence maybe irrelevant. By knowing the new evidence, I do not get any new information about the probability that I am interested in. For example probability of cavity given tooth ache and that weather is Sunny will be equal to probability of cavity given tooth ache because Sunny is irrelevant.

Unless it remain Sunny so much and it is so Sunny that because of that the microbes increase and because the microbes increases cavity and because the cavity this is possible. But again remember, it is its level do we model the problem, right? So, you can say that probability that part has cavity given as that does not have cavity. Now you put that these are the two independent things. By knowing whether cavity or not, I do not get any new information about path.

And so I will say that probability path cavity has given as cavity is equal to probability path on the cavity. On the other hand, you could say oh, one person has cavity, maybe because they live in the same world in a more people have cavities by knowing one person has cavity, we increase the overall probability distributions that people here have cavities, therefore, we increase the probability of that path. There is no; nothing wrong in this logic, per se.

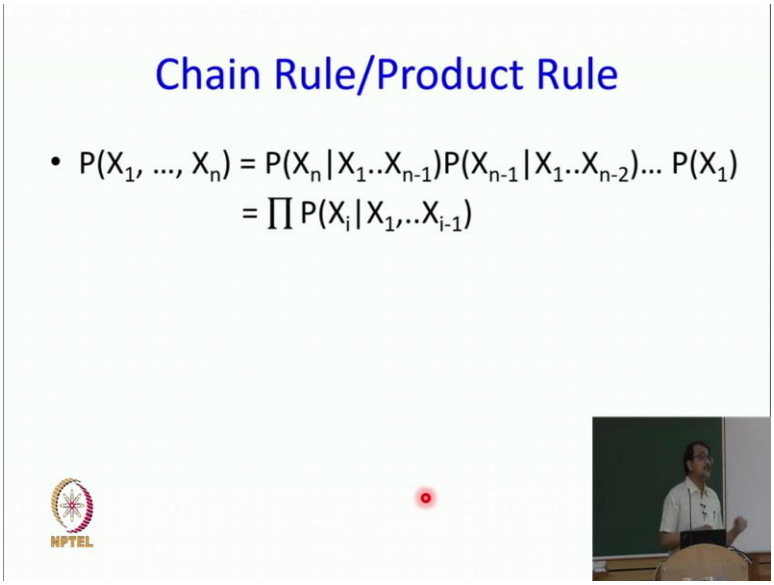
But is it something we want to model or is it not something we want to model. If you are not modeling this kind of behaviour and phenomenon because he said that the influence is very weak. I just said that sort of new evidence does not matter and the technical term we use for this is that the two variables are Independent. And we will see more about this in a minute. And by the way, this kind of inference is the most important influence for probability modeling and we shall see that too, ok.

(Refer Slide Time: 23:58)



So let us get into it. Oh no, we have to define conditional probability. Factorial is A given B, A Intersection B which is the common area divided by the ability of B which is is the Blue area. Because in a given B we are only interested in the blue part, nothing else matter.

(Refer Slide Time: 24:14)



And the last but not the least this is Chain rule and everybody knows the Chain rule and chain rule is most important. The probability of a joint distribution is in any order start with prior probability X n, time X1 given X2 times, probability of X3, times X1 to Xn-2, probability of X1 times X1 to X2 so on times probability of given all the other is always true is does not require any additional knowledge about the variable in any variable (())(24:35).

This is called the Chain rule of the product. So, at this point we have quickly covered everything that you already know about probability distributions.