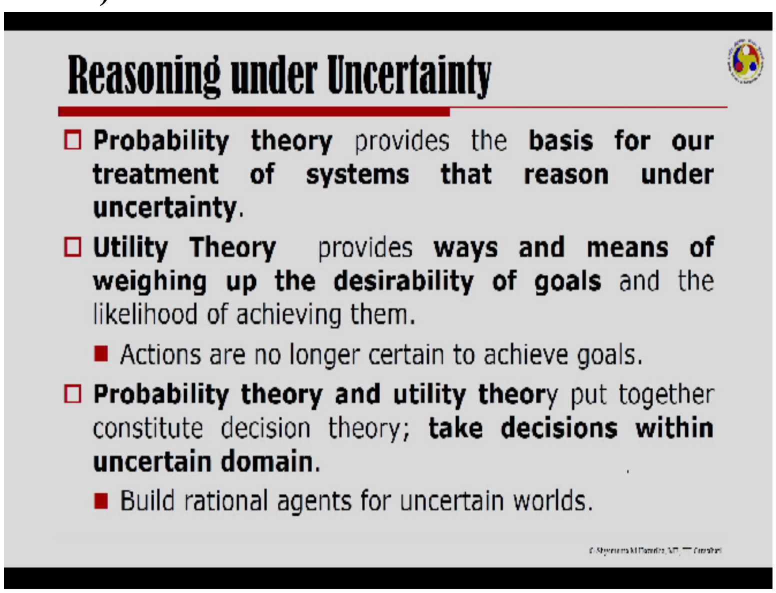
Fundamentals of Artificial Intelligence
Prof. Shyamanta M Hazarika
Department of Mechanical Engineering
Indian Institute of Technology- Guwahati

Lecture-18
Reasoning Under Uncertainty

Welcome to fundamentals of artificial intelligence. So far we have looked at the logical approach to AI, use first of the logic to represent facts and then use inference mechanisms to derive new facts from existing facts and rules. Under such an approach, the agent did not deal with any form of uncertainty in the domain of discourse. Today, we will reexamine the very logical approach and see how it could be adapted to work for domains that have uncertain knowledge, domains that involve some form of uncertainty. So, this module is on reasoning under uncertainty.

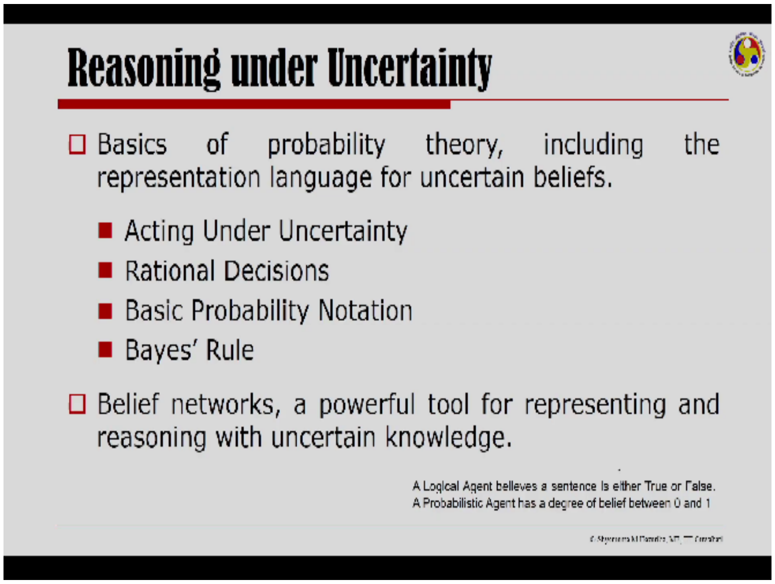
(Refer Slide Time: 01:32)



For reasoning under uncertainty probability theory provides the basis for our treatment of systems that reason under uncertainty, we would look at utility theory that would provide ways and means of weighing up the desirability of the goals. One needs to remember that in an uncertain domain actions are no longer certain to achieve goals. And therefore we need mechanisms that would weigh up the desirability of goals and the likelihood of achieving them.

And this we achieve through the utility theory. The probability theory and the utility theory come together to constitute what is called the decision theory, which involves ways and means of taking decisions with an uncertain domain. And finally, we will look at how to build rational agents for uncertain worlds.

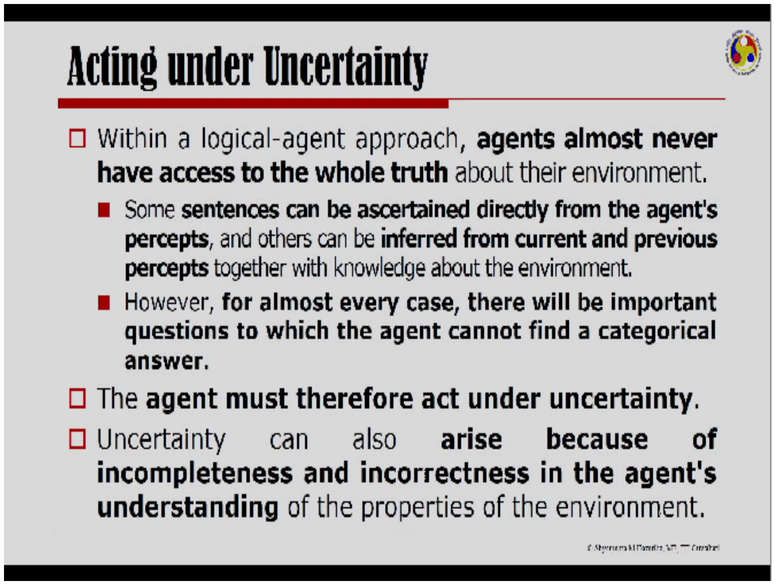
(Refer Slide Time: 02:51)



In this lecture today, we would look at the basics of probability theory, including some form of representation. For uncertain beliefs, it would cover the following 4 topics, we will look at acting under uncertainty, we will see what we mean by rational decisions. Thereafter, we will explore certain basic probability notations. And finally we look at the basis rule. Belief networks, which are a very powerful tool for representing and reasoning with uncertain knowledge will be covered in subsequent lectures.

One needs to really understand that until now any form of agent that we were dealing with was a logical agent. A logical agent believes a sentence to be either true or false. In these lectures that we start today, we would look at what are called probabilistic agents. Probabilistic agents are those that have a degree of belief about the validity of a given sentence. And the belief could be ranging from 0 to 1.

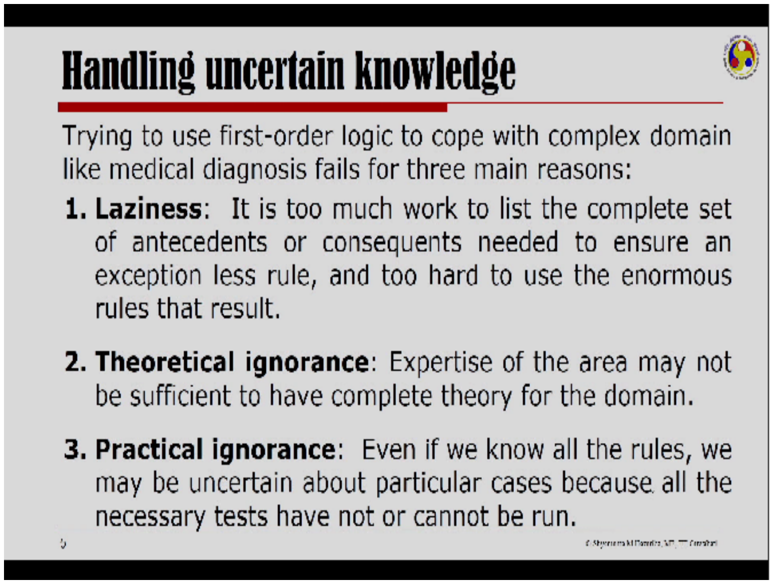
(Refer Slide Time: 04:22)



So let us first look at what do we understand by acting under uncertainty, within a logical agent approach, agents almost never have access to the whole truth about their environment. And sentences can be asserted directly from the agents precepts, or others can be inferred from the current and previous precepts. This is done together with the knowledge about the environment. However, for almost every case there would be certain important questions for which the agent cannot find a categorical answer.

This is where the agent must act under uncertainty. Uncertainty can arise because of incompleteness and incorrectness in the agents understanding of the properties of the environment.

(Refer Slide Time: 05:17)

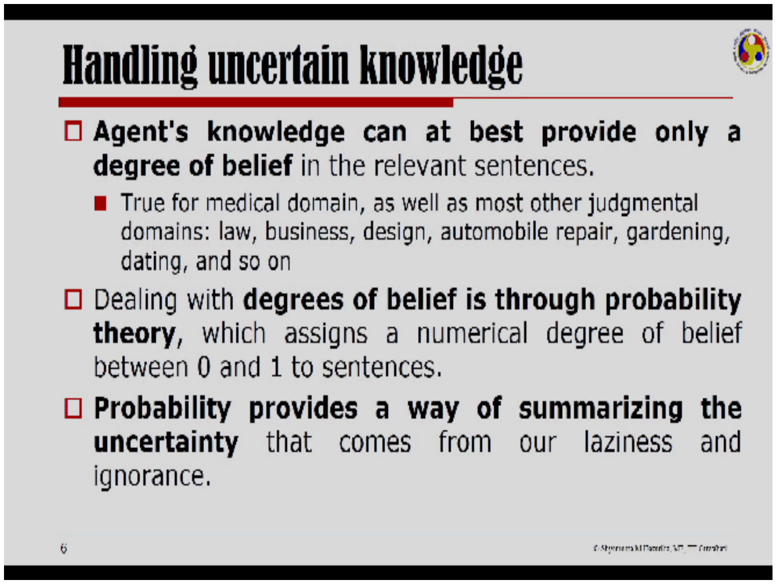


Now, when we are talking of handling uncertainty, one needs to recall that if I am using first order logic to cope with complex domains, like medical diagnosis or criminal investigation or some form of methods and techniques to figure out false in a system, there would be 3 main reasons why such first order logic systems would fail. One, we call it laziness, this is about too much work involved to create the complete set of antecedents or consequence, needed to ensure an exception less rule.

And it is too hard to use the enormous rules that result out of this. So, if you are looking for completely covering one of these domains, then it would be too much of task either to list the complete set of incidents or consequence for a given rule or it would be even hard to really get all the rules in the system. Number 2 is about certain ignorance, which is referred to as the theoretical ignorance, it is about the expertise of the area, which may not be sufficient to have complete theory for the domain that is being worked with.

And finally we have what is called the practical ignorance. Suppose we know all the rules yet we may be uncertain about particular cases, because all the necessary tests or evaluations may have not been or possibly cannot be done for that particular case.

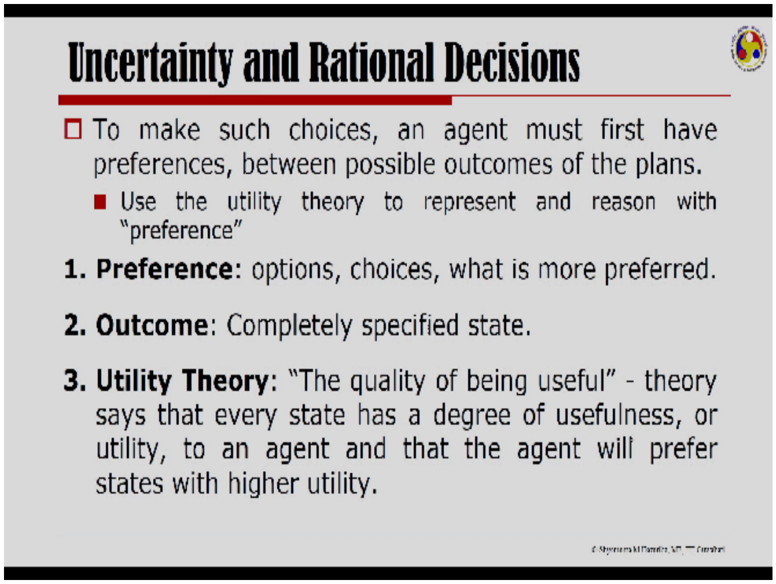
(Refer Slide Time: 07:20)



So agent's knowledge under such a situation can at best provide only a degree of belief in the relevant sentences. And that is why every sentence in under such a scenario cannot evaluate to being just true or false. We have to associate to each sentence that the knowledge of the agent comprises of to a degree of belief. This is not only true for the medical domain that I have been emphasizing. It is also true for most other judgmental domains like law, business, design, automobile repair, gardening, so on and so forth.

So dealing with degrees of belief is what is done through the probability theory, which assigns a numerical degree of belief between 0 and 1 to the sentences. Probability provides a way of summarizing the uncertainty. And one needs to realize that this uncertainty comes from our laziness and ignorance. Laziness here refers to our inability to completely quantify the domain and ignorance their reference to either our theoretical ignorance of the domain or the practical ignorance when I am working with certain cases.

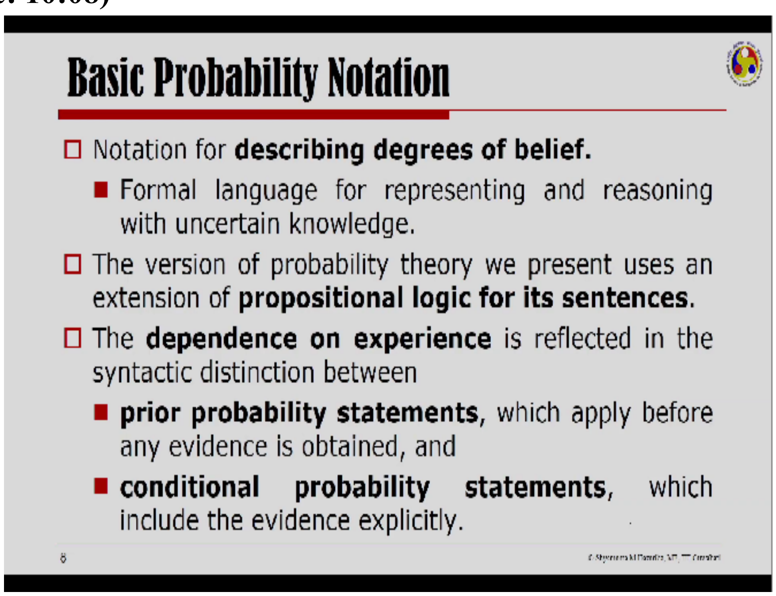
(Refer Slide Time: 09:03)



So to make choices that is to make rational decisions, an agent must first have preferences between the possible outcomes of the plans and this is where we use what is called the utility theory. The utility theory is used to represent and reason with preferences. Here preference refers to options choices, and other alternatives of what is more preferred, outcomes are completely specified state and the utility theory is about figuring out which is more useful, or in terms we say the quality of being useful.

So, the theory says that every state that I get has a degree of usefulness or utility to an agent and the agent will prefer states with higher utility.

(Refer Slide Time: 10:08)

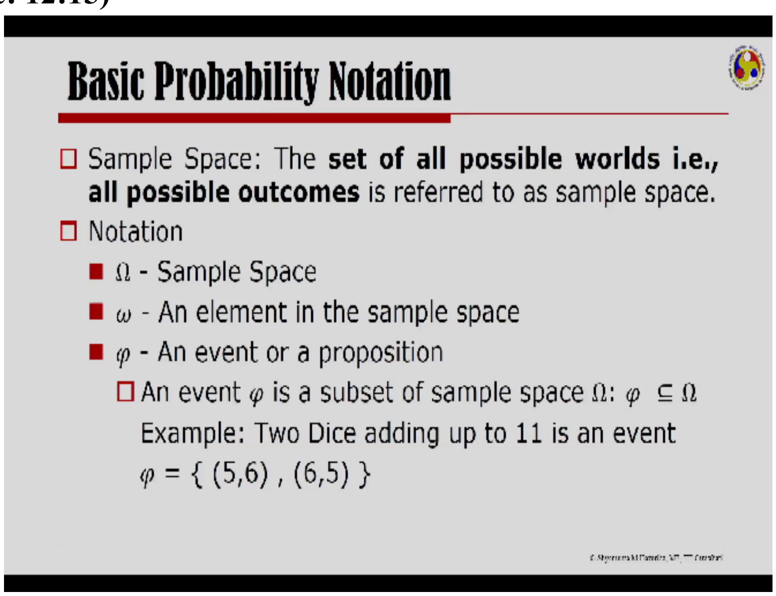


So let us now focus on the basic probability notations that we will use during the course of our discussion in this portion of artificial intelligence. So notation for describing the degree of belief is what is important, we need a formal language for representing and reasoning with uncertain knowledge. We use a version of probability theory. And we use propositional logic for its sentences.

And the dependence on experience is reflected in the syntactic distinction between what we call the prior probability statements, which are statements which applies before any evidences obtained and conditional probability statements, which include the evidence explicitly. So what we have here is some form of representation and that we have to create to capture the uncertainty. And that is captured in terms of the probabilities.

And the statement itself is some form of proportional logic statement. But we associate to each of such statement, some degrees of belief. And for experience, whether we have dependence on experience, when we are trying to talk off the probability associated with a particular statement is reflected in the way the statement is written. So for statements that have no experience of any evidence being obtained prior to the statement is being made. So those are called prior probability statements. And when we include evidence explicitly, we have what are called conditional probability statements.

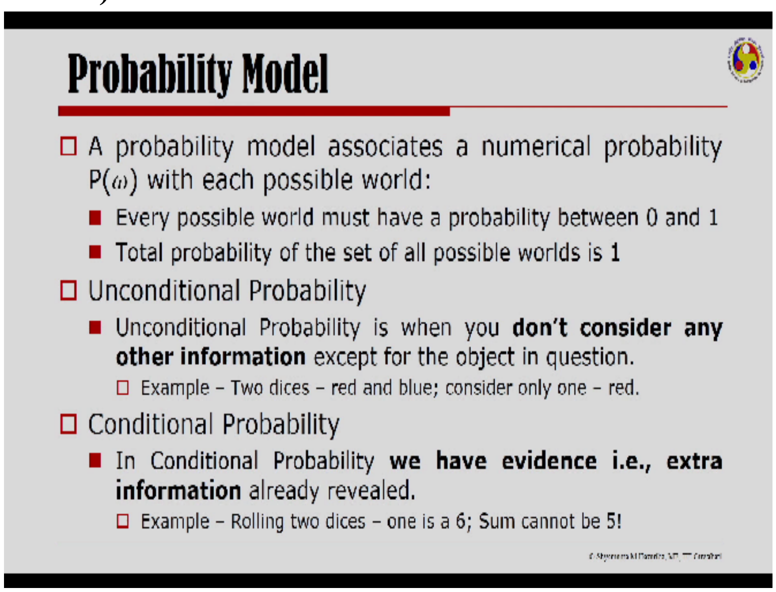
(Refer Slide Time: 12:15)



Now, when we are talking of such a representation, we have to talk about what is called the sample space, the sample space refers to the set of all possible worlds that is all possible outcomes. And we use the upper case omega to represent the sample space. Omega is an element in the sample space, phi is an event. Now, this is interesting to realize that phi the event, or a proposition is a subset of the sample space.

Like let us say, I am rolling 2 dice. And I want to figure out, if every time I get the 11, between the pair of dices. So the event that I have is about the 2 dices, adding up to 11. And that I could represent as pairs 5, 6 and 6, 5 because if 1 die has a 5, the other dice shows a 6, I have 11, or the first one is a 6, the second one would need to be 5 to have an 11. So that would be an even, which would be from the sample space of all pair of dice that I am rolling.

(Refer Slide Time: 13:49)

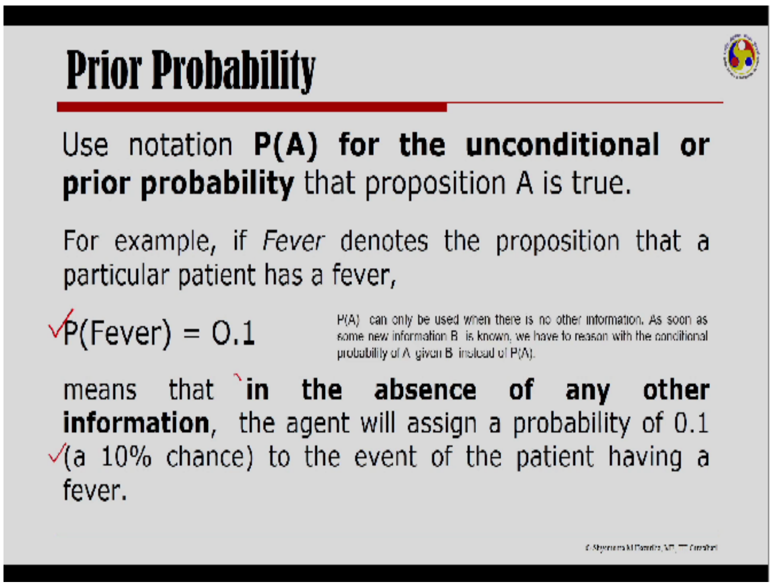


A probability model associates a numerical probability with each possible world. So every possible world must have a probably between 0 and 1. And then the total probability of the set of all possible worlds must be 1. Unconditional probability is when you do not consider any other information except for the object in question. Like, let us go back to our example of rolling a pair of dice I have 2 dice, let us say one, a red one, the other a blue one.

Now in unconditional probability, I may only consider one the red. And I would be not concerned about what is the result on the blue. Whereas in conditional probability, we have evidence that is we have information already revealed to us. So information that has been already revealed to us puts a lot of constraints on what could be the other possibilities, like coming back to the rolling of the 2 dices that I was talking about.

Let us say, I am talking of a sum for the 2 dices. Now, if I see my red dice has a 6. Immediately I know that the sum of the 2 dice cannot be less than 6. That is it cannot be 5. And those type of scenarios is what we have under conditional probability. Where we have evidence or extra information, already revealing to us what could be the possibilities remaining with us.

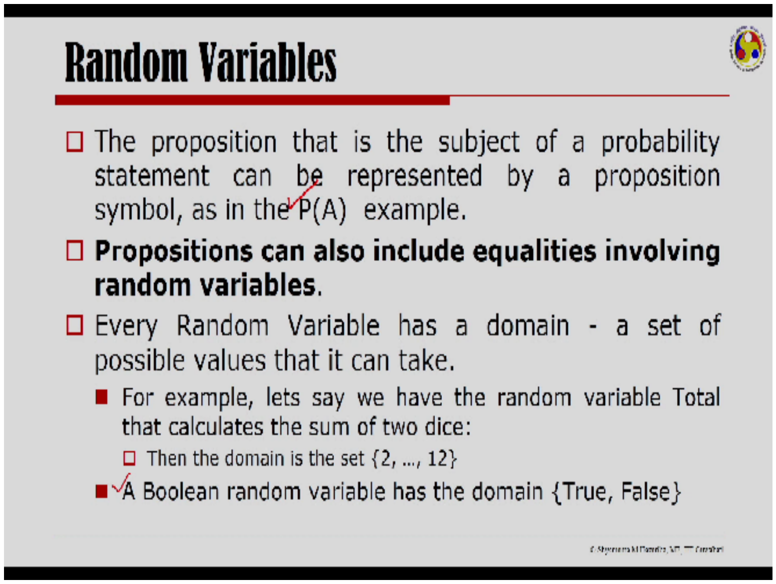
(Refer Slide Time: 15:54)



Now coming back to the 2 probabilities, the unconditional probability is also referred to as prior probability. So the notation used for it is P of A for the unconditional or prior probability that proposition A is true. For example, if fever denotes the preposition that a particular patient has a fever, then I can write probability of fever is equal to 0.1 means so when we are talking of unconditional or prior probabilities, we need to understand that it is in the absence of any other information.

So under such a scenario, the agent will assign a probability of 0.1, which is a 10% chance to the event of the patient having a fever. On the other hand, we need to understand that conditional probabilities will have certain information. So unconditional probabilities, can only be used when there is no other information. As soon as some new information B is known to us, we have to reason with the conditional probability of A given me, instead of the prior probability.

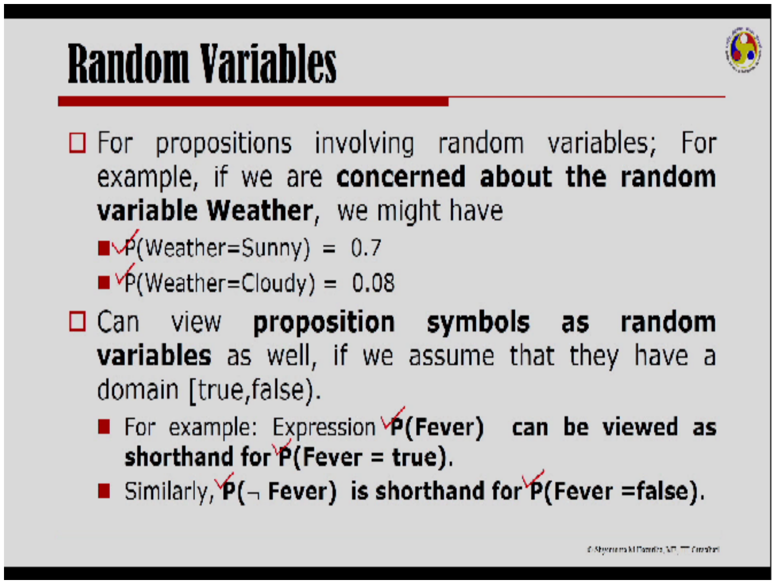
(Refer Slide Time: 17:23)



We will come back to conditional probability in a minute. Before that, let us look at what we understand by random variables. The proposition that is the subject of a probability statement can be represented by a propositional symbol. As in the previous case, I wrote probability of A, A being a propositional symbol. Propositions can also include equalities. And when they include the qualities they involve random variables.

So every random variable has a domain itself a possible values that it can take. For example, let us say we have the random variable total, that calculates the sum of the 2 dices, then the domain is the set 2 up to 12, because 2 dices can maximum add up to 6 + 6 = 12. So on the other hand, a Boolean random variable has the domain of either true or false.

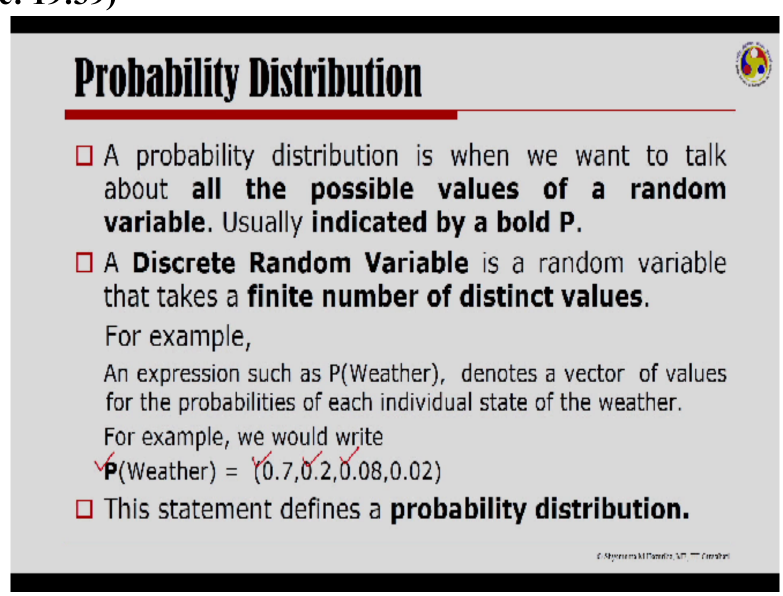
(Refer Slide Time: 18:29)



For propositions involving random variables. For example, if we are concerned about the random variable weather, we might write statements like the probability of weather being sunny is 0.7, or the probability of the weather being cloudy is 0.08. We can also view proposition symbols as random variables, if we assume that that they have a domain that is true or false. For example, recall the expression that we have used probability of fever.

Now probability a fever can actually be viewed as a shorthand for probability of fever equal to true. Similarly, probability, not fever, could be actually representing probability, fever equal to false. So we can see that even propositional symbols can be viewed as random variables.

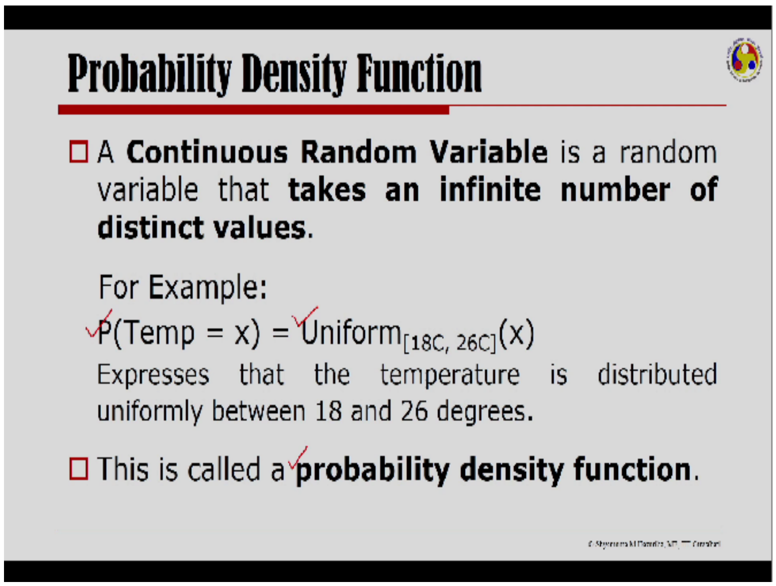
(Refer Slide Time: 19:39)



A probability distribution is when we want to talk about all the possible values of a random variable. And it is usually indicated by bold phase P, we could have what is called a discrete random variable, which is a random variable that takes a finite number of distinct values. So an expression such as probability of whether I could denote a vector of values for the probability for each individual state of the weather.

So it could be like probability of weather being sunny, cloudy, rainy, or some other options. Now, this statement defines the probability distribution, because it gives us all the possible values of a random variable.

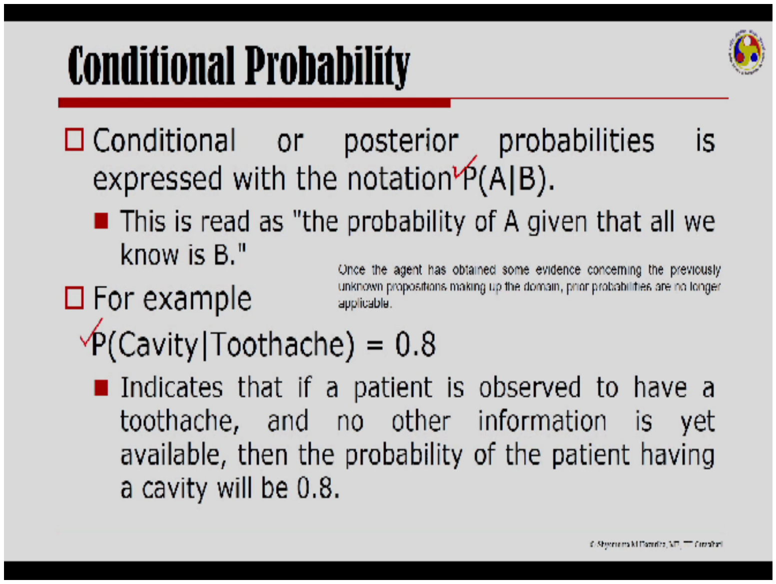
(Refer Slide Time: 01:22)



On the other hand, if I have variable which is called continuous random variable, a continuous random variable is one that takes an infinite number of distinct values, for example, I could be talking of the temperature. So the probability of temperature being uniformly distributed between 18 degree and 23 degree is called a probability density function. So a continuous random variable when I am talking off, and trying to express what its value would be over the complete distribution, it is a probability density function.

But when I am talking of a discrete random variable, and talking of all the values of the variable, then it is a probability distribution.

(Refer Slide Time: 21:25)

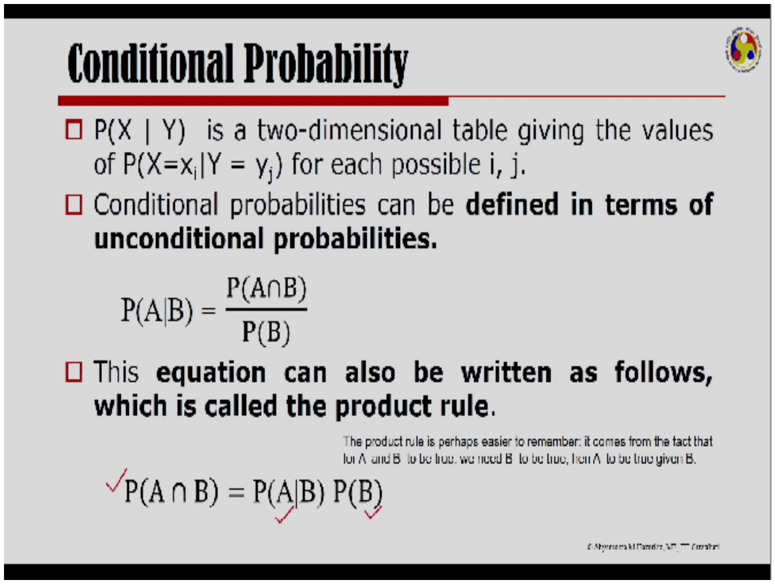


Let us now come back to conditional probability. Conditional probability is also called posterior probability. And is expressed by the notation probability of A given that all we know is B or read shortly as probability of A given B, it is about having some evidence concerning the previously unknown proposition making up the domain. So, when I have such information or evidence, then prior probabilities or what we call the unconditional probabilities are no longer applicable.

And this is where I have to take help of the posterior probabilities. For example, here is a probability of cavity given toothache. So indicates that if a patient is observed to have a toothache, and no other information is yet available, then the probability of the patient having a cavity will be almost 80%. So probability of cavity given toothache is 0.8. Now, this probability that I have written here has some evidence concerning the previously unknown proposition making up the domain.

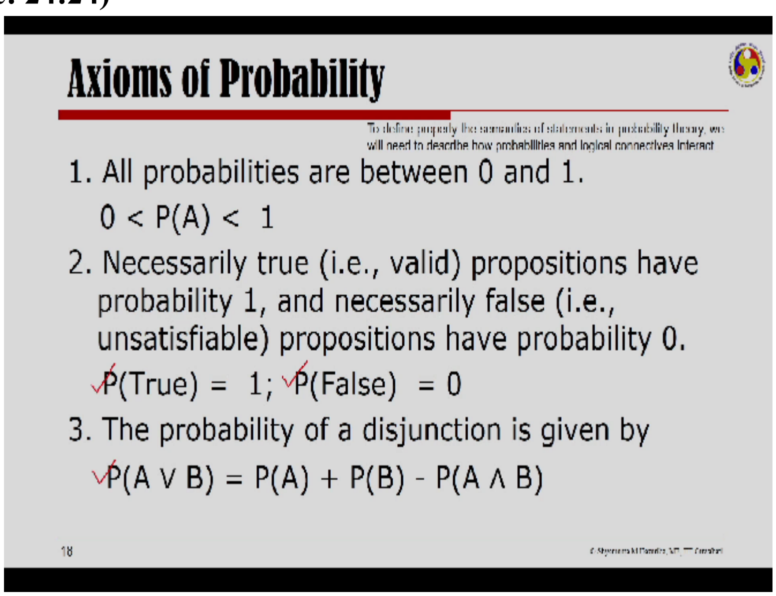
And therefore, under the observation that the patient is supposed to have toothache the probability of the patient having cavity is a conditional probability.

(Refer Slide Time: 23:05)



So conditional probability, probability of X given Y is a 2 dimensional table, which gives the value of Xis and Yjs for each possible IJ pairs. Conditional probabilities can be also defined in terms of unconditional probabilities, like here I have the probability of A given B written as the probability of A intersection B by the probability of B. This equation can also be written as probability of A intersection B is probability of A given B into probability of B. This is called the product rule. And the welcome back to this during our course on discussion of reasoning, using uncertainty. So the product rule is perhaps easier to remember, if you think in the following lines, it comes from the fact that for A and B to be true, we need B to be true, then A to be true given B. So that is the product rule.

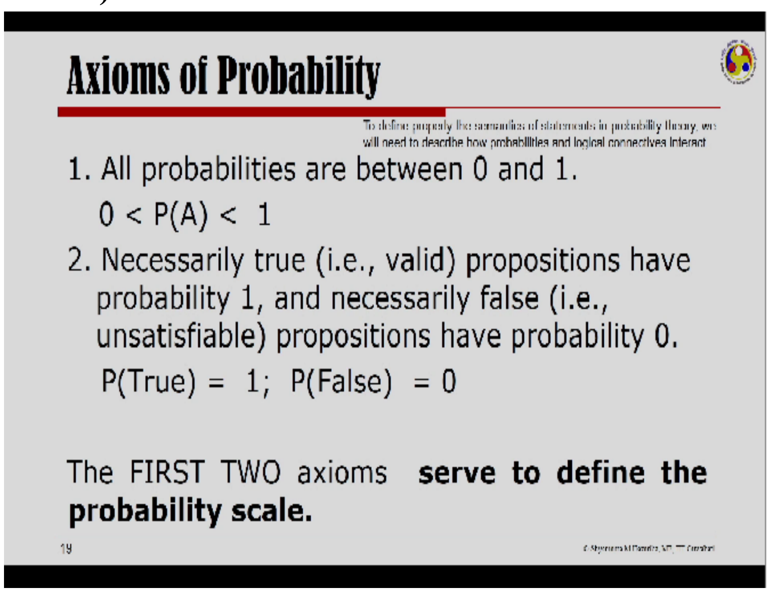
(Refer Slide Time: 24:24)



To define properly. the semantics of statements In probability theory, we would need to describe how probabilities and the logical collectives interact. We have looked at the semantics of the logical collectives during our discussion on knowledge representation and reasoning. Here, we will look at how does the probabilities interact with the logical collectives, the first axioms of probability states all probabilities are between 0 and 1.

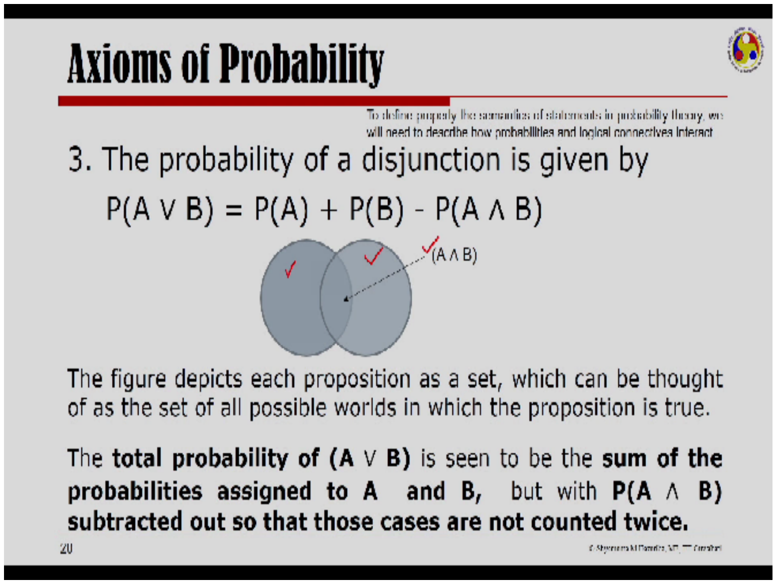
Now, necessarily true. That is valid propositions have probability 1, and necessarily false that is unsatisfiable propositions have probability 0, in the sense that the probability of true is 1 and the probability of false is 0. The third axioms of probability is the following. It gives us the probability of a disjunction probability of A or B is the probability of A plus the probability of B minus the probability of A and B.

(Refer Slide Time: 25:55)



So let us try to understand this with a small figure. But prior to this, we need to take note that the first 2 axioms of probability serve to define the probability scale.

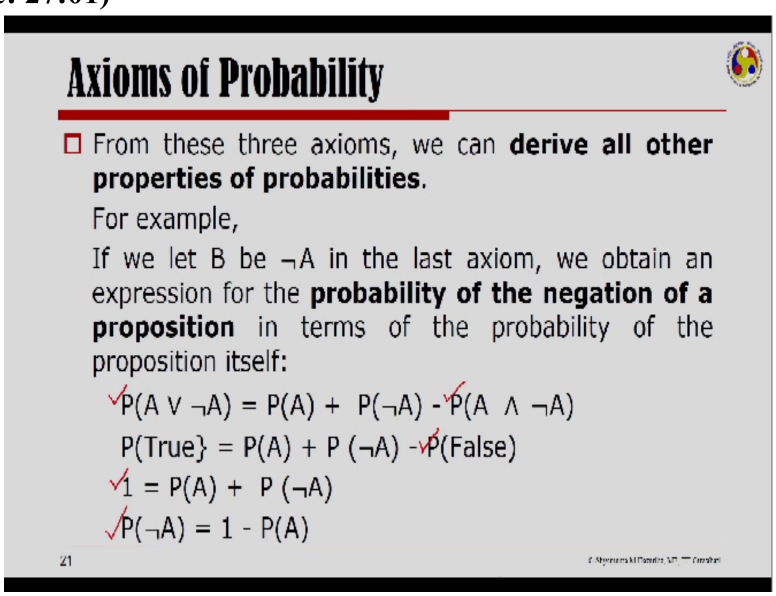
(Refer Slide Time: 26:07)



The probability of disjunction can best be understood by looking at this figure where each proposition is shown as a set. And this can be thought of as the set of all possible worlds in which the proposition is true, the total probability of A or B is seemed to be the sum of the probabilities assigned to A and that assigned to B, but we need to subtract from it, the probability of those cases that lies in A and B.

So as those cases are not counted twice, so we have P(A) + P(B), but these cases which are A and B I have to subtract them to ensure that do not double count them.

(Refer Slide Time: 27:01)

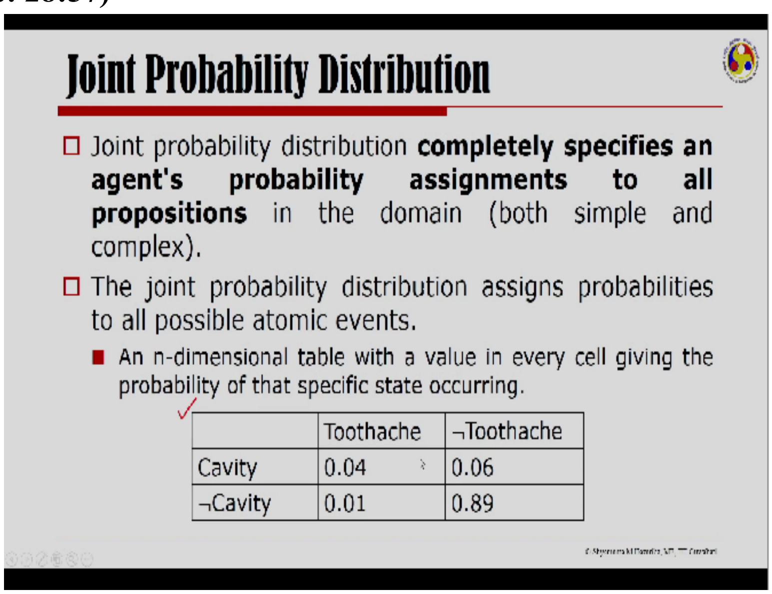


So from these 3 axioms, we can derive all other properties of probabilities. For example, if we let B be not A in the last axioms, we obtain an expression for the probability of negation of a preposition, in terms of the probability of the proposition itself. Let us look at the derivation here. So I have the probability of A or B, but now I have replaced B with not of A, so I have probability of A plus probability of not of A minus probability of A and nothing.

Now this is where we need to understand the logical connective, probability of A or not A is going to be 1, because that is a tautology which is always true. And therefore I have 1, the probability of A and not A, so this was a disjunction A or not A is going to be true always, therefore giving A1. Whereas on this side, if you see probability of A and not A is a contradiction, going to be false always, and therefore is 0.

So what I have finally is 1 equal to probability of A plus probability of not A . And from that, I can write probability of not of A is equal to 1 minus probability of A. So the probability of the negation of a proposition in terms of the probability of the proposition itself comes out as a consequence of the 3 axioms that we have written. From these 3 axioms, we can derive other properties of probabilities as well.

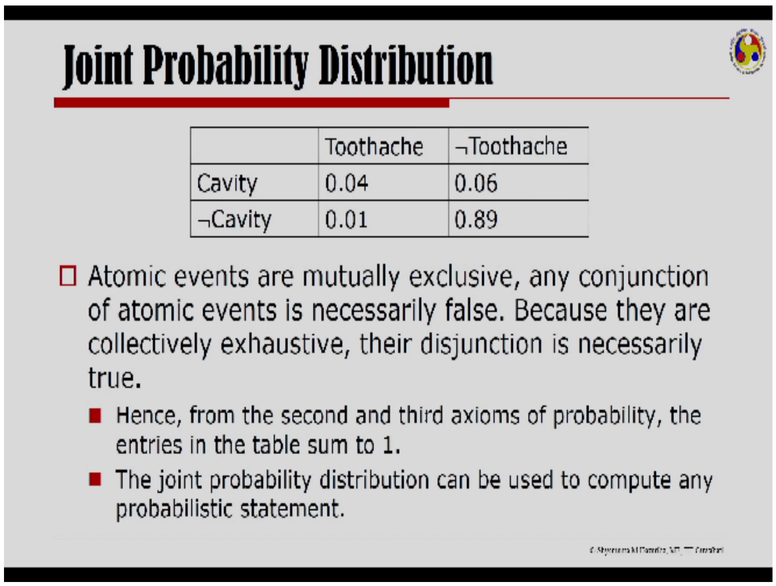
(Refer Slide Time: 28:57)



So now, let us try to define something called the joint probability distribution. A joint probability distribution completely specifies an agent's probability assignment to all propositions in the domain. So, it assigns probabilities to all possible atomic events, here I have an m dimensional table with a value in every cell giving the probability of that specific state occurring. Now, coming back to this toothache and cavity problem, I could think of cavity and not cavity toothache and not toothache.

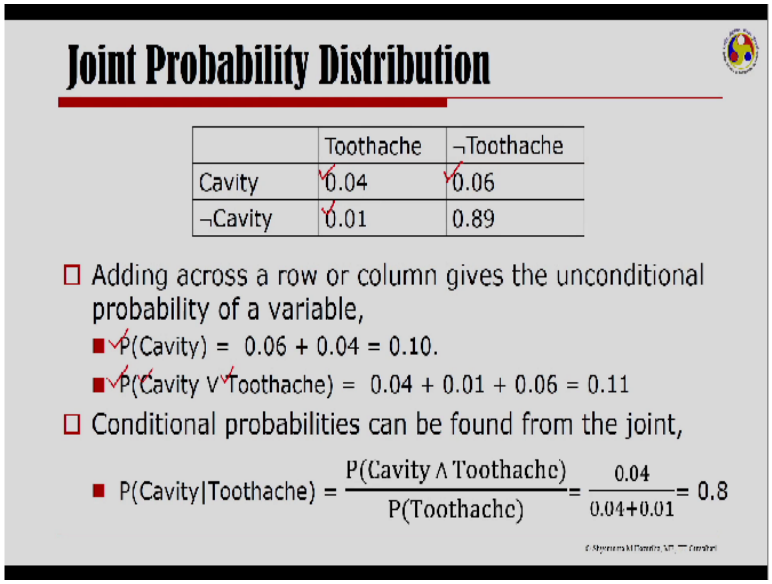
And completely specify the probability assignments and this would be a joint probability distribution.

(Refer Slide Time: 29:50)



Now, there are certain things that needs to be understood in a joint probability distribution. The first one is that we are talking of atomic events. Now, atomic events are mutually exclusive. So any conjunction of atomic events is necessarily false. And because they are collectively exhaustive, their disjunction is necessarily true. Hence, from the second and third axioms of probability, the entries in the table sum to 1

(Refer Slide Time: 30:36)



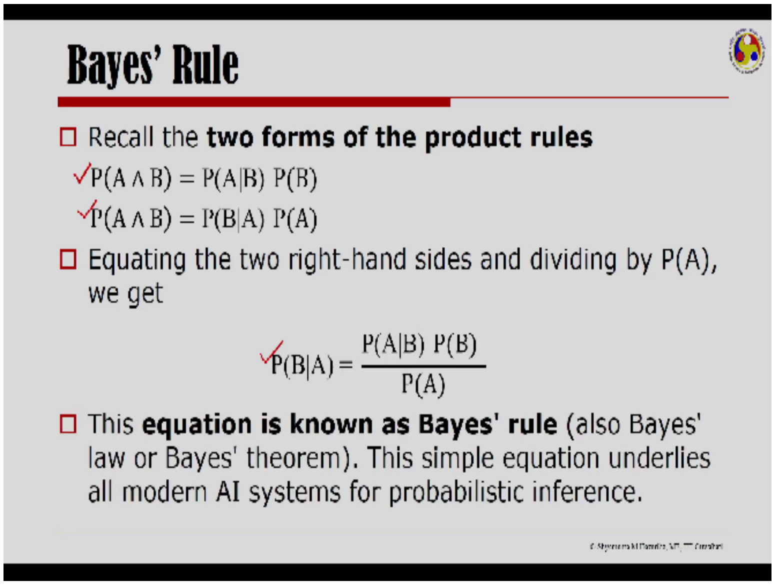
And the joint probability distribution can be used to compute any probabilistic statement, like adding across a row or column gives the unconditional probability of a variable, like for the probability of cavity, all I need to do is add up whether it is toothache or whether it is not toothache because I am talking of the unconditional probability of cavity, so that I add up the row. For that matter, if I am looking for are only the toothache, I would add up a column.

When I am looking for a probability of cavity or toothache I would add up values of cavity 0.04 and 0.06, and 0.01, which is a probability of toothache. Now recall that this is unconditional probability, so here toothache does not matter whether I have a cavity or do not have a cavity, both of them adds needs to be added up on the cavity, it does not matter whether I have toothache or no toothache, both of them needs to be added up.

So when I am looking for conditional probabilities, conditional probabilities can be found from the joint as well. So this is something that I need to mention here, that the joint probability distribution is also referred to as joint simply. So here is a conditional probability, probability of cavity given toothache so that too I can find out from the joint probability distribution, which is probability of cavity and the toothache and probability of toothache.

If you remember, the product rule that we were discussing little while ago, that could lead me to conditional probabilities from the joint probability distribution.

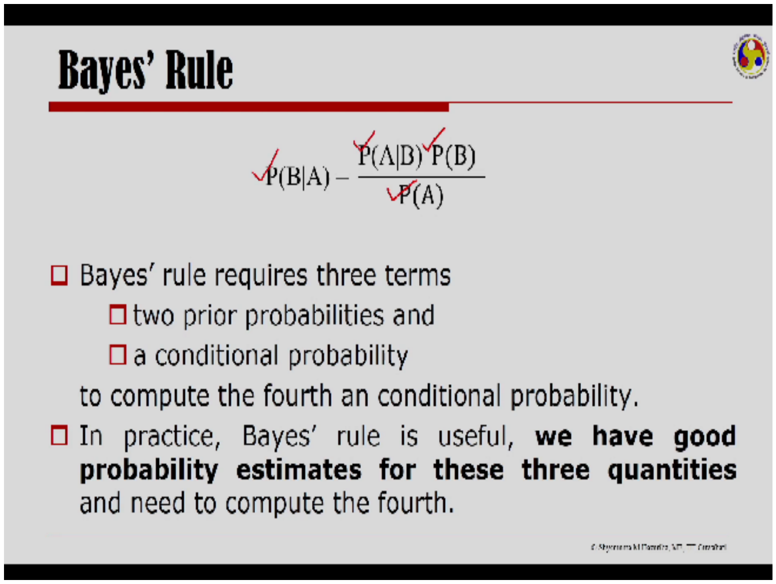
(Refer Slide Time: 32:31)



Now, let us recall the 2 forms of the product rules, probability of A and B is probability of A given B probability of B and probability of A and B can also be written in terms of probability of B given A and probability of A. So equating the 2 right hand sides and then dividing by P(A), we could get the problem quality of B given A is actually probability of A given B multiplied by probability of B whole divided by probability of A.

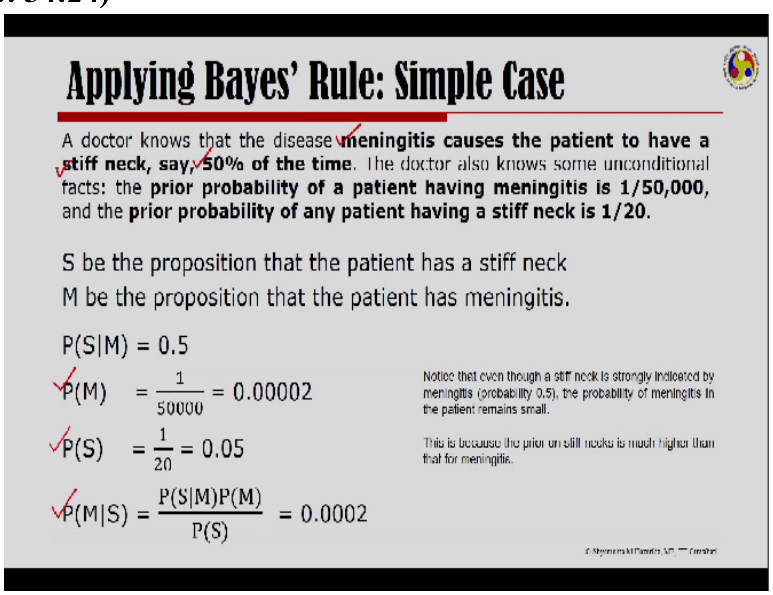
This is a equation which is known as the Bayes rule, also called the Bayes law, or the Bayes theorem. This simple equation surprisingly, underlines all modern artificial intelligence systems for probabilistic inference.

(Refer Slide Time: 33:34)



Now, just said it is simple. First rule, one fails to understand the power of the Bayes' rule, let us try to understand why it is so important for probabilistic reasoning with an AI, the Bayes' rule, request 3 terms. If you have noticed properly, 2 of them are prior probabilities and one of them is a conditional probability, these 3 computes the 4, which is a conditional probability. In practice, the Bayes' rule is very useful because we have good probability estimates for the 3 quantities that I have listed above.

(Refer Slide Time: 34:24)

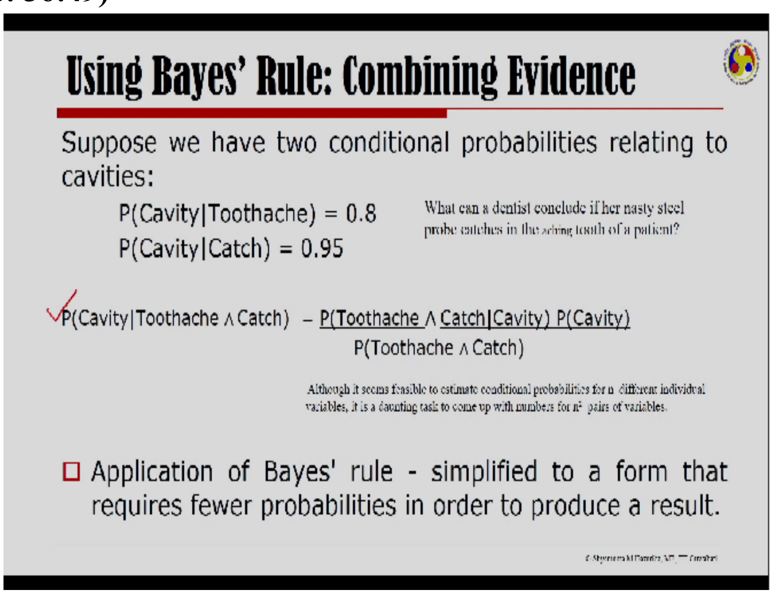


Now let us take an example. And try to understand the application of Bayes' rule to a very simple case. This is an example from Russell and Novig's book. So here we have a statement which says, a doctor knows that the disease, meningitis causes the patient to have a stiff neck, say 50% of the time. The doctor also knows some unconditional facts. One, the prior probability of a patient having meningitis is 1 over 50,000.

And the prior probability of any patient having a stiff neck is 1 over 20. Let S be the proposition that the patient has a stiff neck, and M be the proposition that the patient has meningitis. Now, probability of stiff necks given M is 0.5. It is 50%. That is what it says stiff neck. Because of the disease, probability of the disease itself is 1 over 50,000. And that is a prior probability, unconditional, and then I have a probability of stiff neck, which is 1 over 20 = 0.05.

Now, if one wants to know what is the probability of the disease that he has a stiff neck, then the probability is 0.00002, which is very low. Notice that even though a stiff neck is strongly indicated by meningitis, which is a probability of 0.5, the probability of the disease in the patient with stiff neck remains very small. This is because the prior on stiff necks is much higher than that of the disease itself. Now, this was a very simple application of the Bayes' rule to understand what would be the probability of the disease given stiff neck.

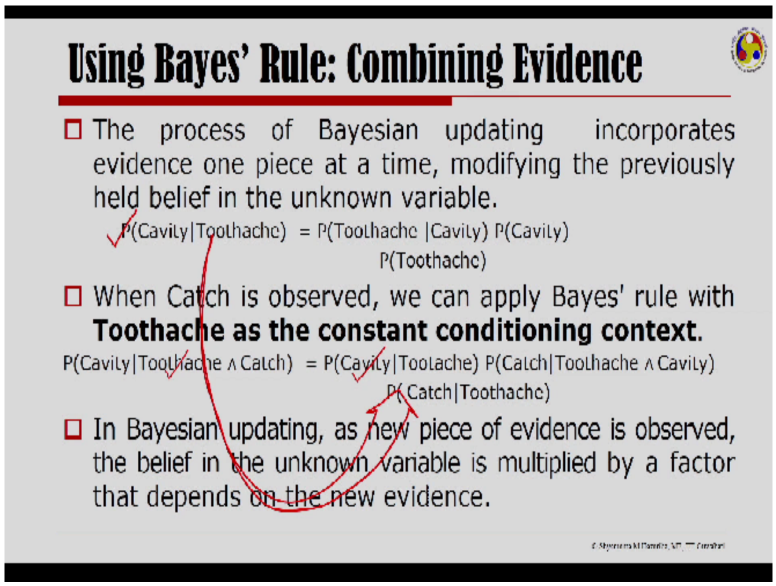
(Refer Slide Time: 36:49)



But Bayes' rule can also be use for combining evidences, like suppose we have 2 conditional probabilities relating to cavities. One saying the probability of a cavity given a toothache is 80%. And the probability there is a cavity because the dentist could catch it with his probe is 95%. Now what can a dentist conclude if it catches aching tooth of a patient that is, what is the probability of cavity if I have toothache and a catch given toothache and a catch.

Now, this is what the probability would look like. Probability of toothache and catch given cavity probability of cavity probability of toothache and catch. One needs to realize that although it seems feasible to estimate conditional probabilities, for n different individual variables, it is actually a daunting task to come up with numbers for n square pair of variables. And therefore, application of Bayes' rule needs to be simplified to a form that requires fewer probabilities in order to produce a result when combining evidences.

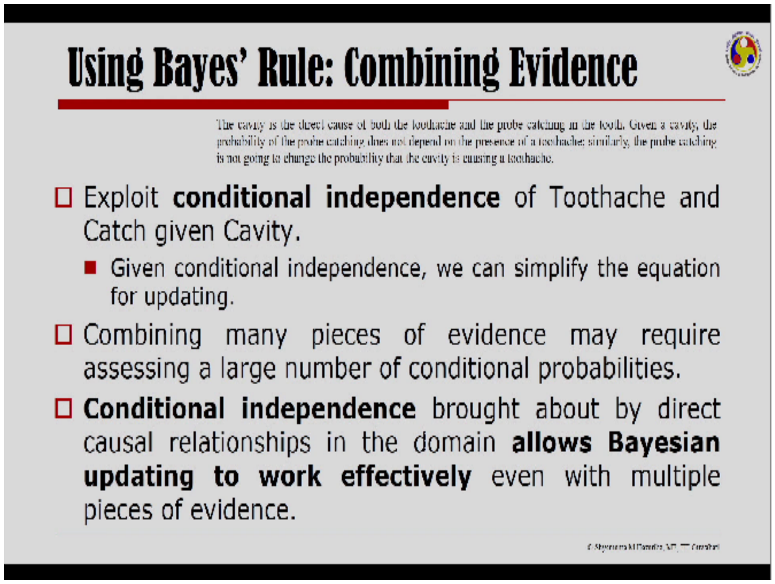
(Refer Slide Time: 38:15)



So the process of Bayesian updating, which incorporates evidence is done one piece at a time modifying the previously held belief in the unknown variable, so if I was talking of cavity toothache and a catch together, I would first look at cavity and toothache and then look at the other one. So, when catch is observed, we can apply the Bayes' rule at that point toothache is the constant conditioning context.

What that means is that I would be looking for definitely the toothache and catch, but when I am doing it, I would be looking for the cavity and toothache and the probability of cavity and toothache will be brought from here into this equation of the Bayes' rule here. So in Bayesian updating as new piece of evidence is observed, the believing the unknown variable is multiplied by a factor that depends on the new evidence.

(Refer Slide Time: 39:40)

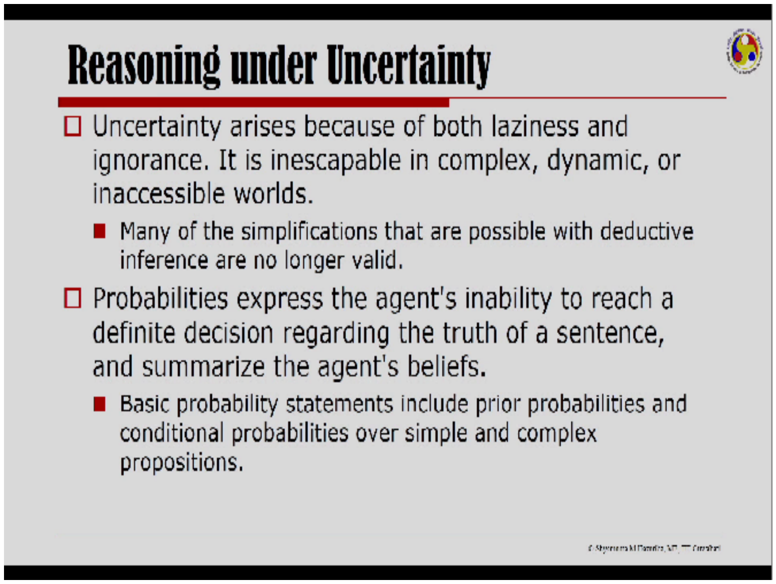


So the cavity, if you think of this in this way, that the cavity is the direct cause of both the toothache and the pro catching in the tooth. Given a cavity, the probability of the probe catching does not depend on the presence of toothache. Similarly, the probe catching it is not going to change the probability that the cavities causing a toothache. So all of this is about exploiting causal relationship in the domain between these variables.

And that is called conditional independence. So you exploit what is called conditional independence of toothache and catch given cavity. And given the conditional independence, we can simplify the equations for updating. Combining many pieces of evidence may require assessing a large number of conditional probabilities. But nevertheless, conditional independence, which as I mentioned before, is brought about by the direct causal relationship in the domain allows Bayesian updating to work very effectively, even with multiple pieces of evidence.

This is what we have for today. So let us quickly recall what we have done before we move on to something called the Bayesian network in the next lecture.

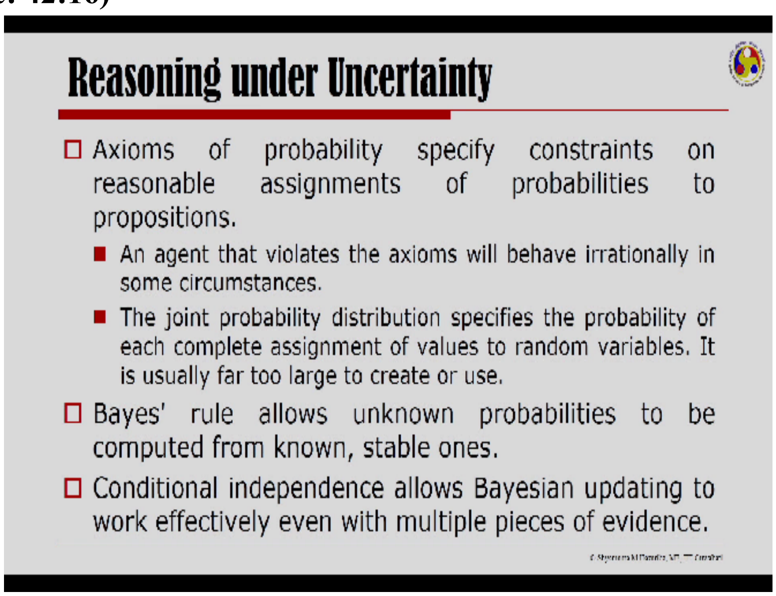
(Refer Slide Time: 41:27)



So, today, we looked at uncertainty and realize that uncertainty arises because of both laziness and ignorance. It is inescapable in complex dynamic or inaccessible worlds. Now, many of the simplifications that are possible with deductive inferences are no longer valid when we are talking of domains with uncertainty. Probabilities express the agents inability reach a definite decision regarding the truth of a sentence.

And summarizes the agents believe, a basic probability statements we have seen include prior probabilities and conditional probabilities.

(Refer Slide Time: 42:16)



We have looked at the axioms of probability, which specify constraints on reasonable assignment of probabilities to proposition. An agent that violates the axioms actually will behave irrationally. And the joint probability distribution that we have seen specifies the probability of each complete assignment of values to random variables.

It is usually far too large to create or use. And we have looked at the biases rule, which allows unknown probabilities to be computed from known stable ones. Finally, we have looked at conditional independence, somehow allowing Bayesian updating to work effectively, even in multiple pieces of evidence. Thank you very much.