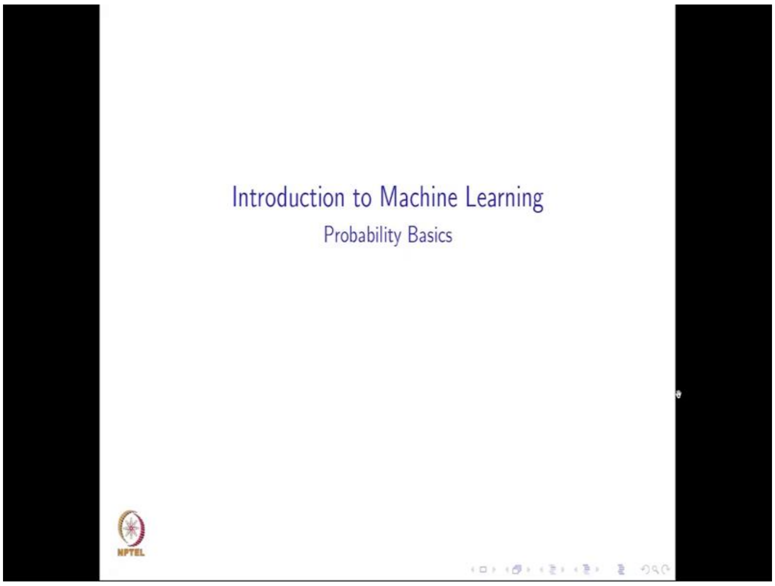
Introduction to Robotics
Professor Balaraman Ravindiran
Department of Computer Science
Indian Institute of Technology, Madras
Lecture - 31
Tutorial - 1: Probability Basics

(Refer Slide Time: 00:14)

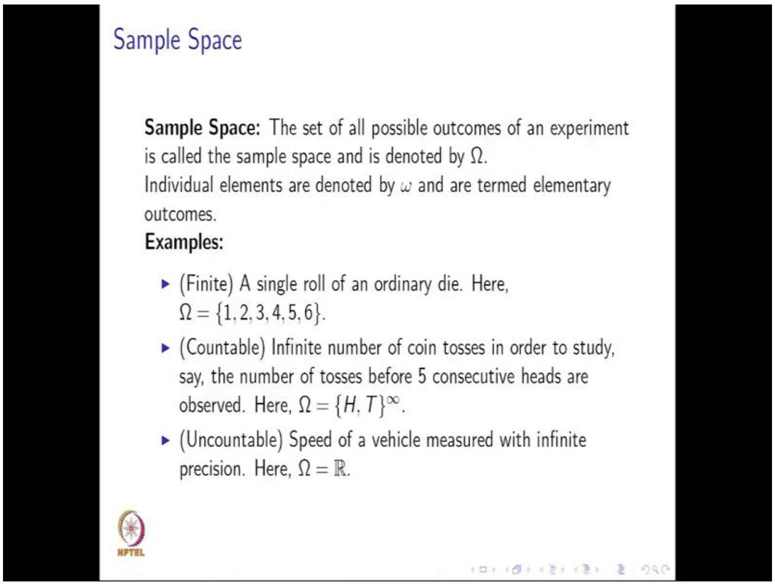


Hello and welcome to the first tutorial in the Introduction to Machine Learning course. My name is Priytosh, I am one of the teaching assistants for this course. In this tutorial, we will be looking at some of the basics of probability theory.

Before we start, let us discuss the objectives of this tutorial. The aim here is not to teach the concepts of probability theory in any great detail, instead, we will just be providing a high-level overview of the concepts that will be encountered later on in the course.

The idea here is that for those of you who have done a course in probability theory or are otherwise familiar with the content, this tutorial should act as a refresher. For others who may find some of the concepts unfamiliar, we recommend that you go back and prepare those concepts from, say, an introductory textbook or any other resource, so that when you encounter those concepts later on in the course, you should be comfortable with them.

(Refer Slide Time: 01:07)



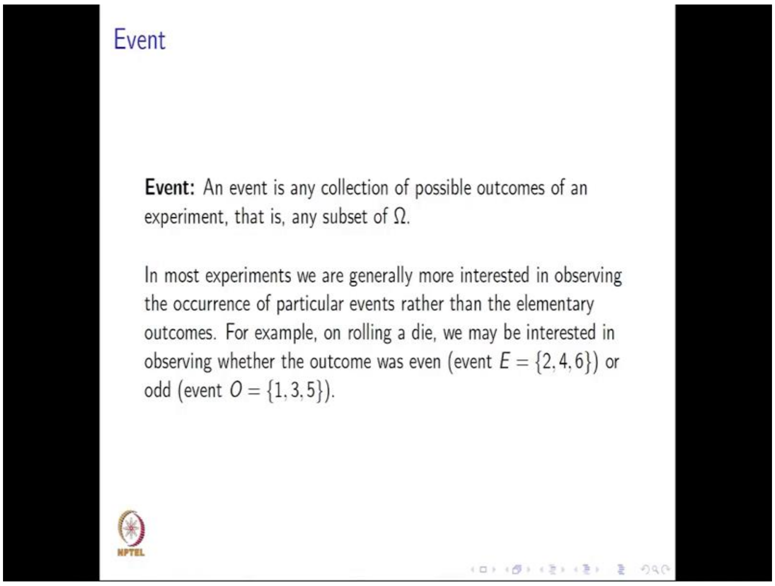
Okay. To start this tutorial, we will look at the definitions of some of the fundamental concepts. The first one to consider is that of the sample space. The set of all possible outcomes of an experiment is called the sample space and is denoted by capital Omega. Individual elements are denoted by small omega and our termed elementary outcomes.

Let us consider some examples. In the first example, the experiment consists of rolling an ordinary die. The sample space here is a set of numbers between 1 and 6. Each individual element here represents one of the six possible outcomes of rolling a die. Note that in this example, the sample space is finite.

In the second example, the experiment consists of tossing a coin repeatedly until a specified condition is observed. Here, we are looking to observe five consecutive heads before terminating the experiment. The sample space here is countably infinite.

We, the individual elements are represented using the sequences of their Hs and Ts, where H and T stand for heads and tails respectively. In the final example, the experiment consists of measuring the speed of a vehicle within finite precision. Assuming that the vehicle speeds can be negative, the sample space is clearly the set of real numbers. Here, we observe that the sample space can be uncountable.

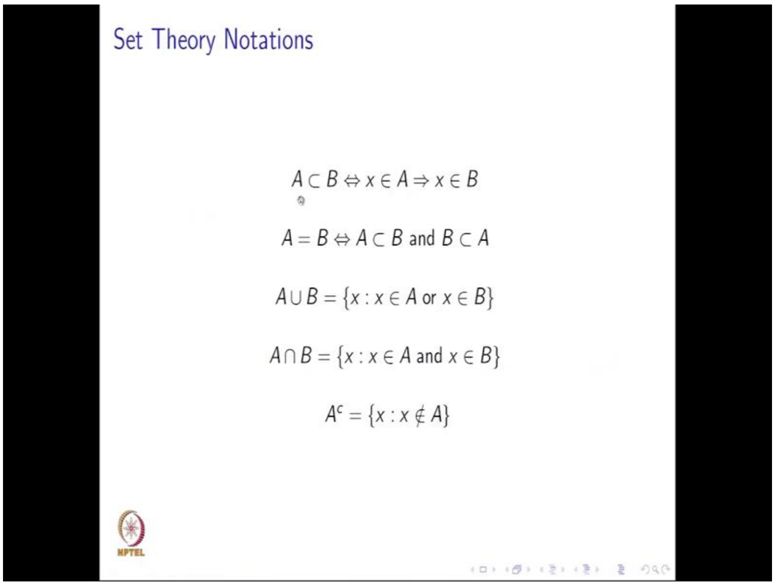
(Refer Slide Time: 02:28)



The next concept we look at is that of an event. An event is any collection of possible outcomes of an experiment; that is, any subset of the sample space. The reason why events are important to us is because, in general, when we conduct an experiment, we are not really that interested in the elementary outcomes. Rather, we are more interested in some subsets of the elementary outcomes.

For example, on rolling a die, we might be interested in observing whether the outcome was even or odd. So, for example, on a specific roll of a die, we, let us say we observed that the outcome was odd. In the scenario, whether the outcome was actually a 1 or a 3 or a 5 is not as important to us as the fact that it was odd.

(Refer Slide Time: 03:14)

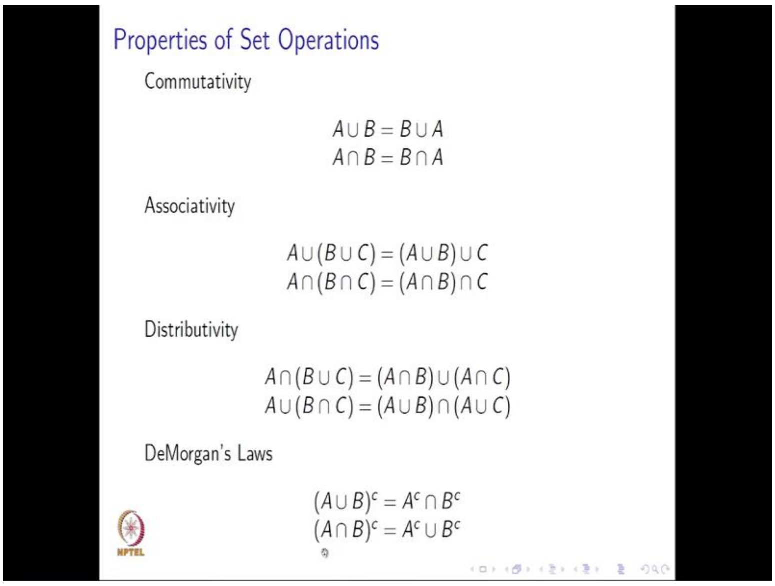


Since we are considering sets in terms of sample spaces and events, we will quickly go through the basic set theory notations. As usual, capital letters indicate sets, and small letters indicate set elements.

We first look at the subset relation. For all x, if x element of A, implies x element of B, then we say that A is a subset of B, or A is contained in B. Two sets A and B are set to be equal if both, A subset of B and B subset of A holds. The union of two sets, A and B, gives rise to a new set, which contains elements of both A and B. Similarly, the intersection of two sets gives rise to a new set, which contains of only those elements, which are common to both A and B.

Finally, the complement of a set A is essentially the set, which contains all elements in the universal set, except for the elements present in A. In our case, the universal set is essentially the sample space.

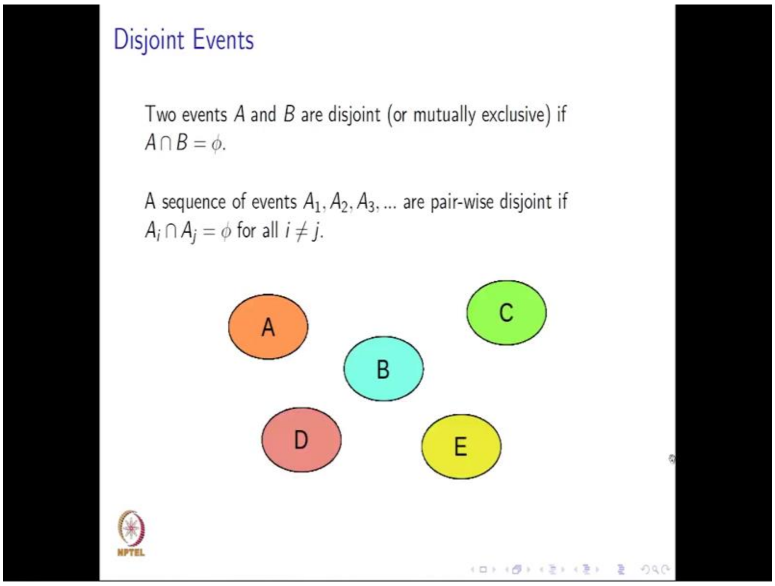
(Refer Slide Time: 04:23)



This slide list out the different properties of set operations, such as commutativity, associativity, and distributivity, which you should all be familiar with. It also list out the DeMorgan's Laws, which can be very useful.

According to the DeMorgan's Laws, the complement of the set A union B is equals to A complement intersection B complement. Similarly, the complement of the set A intersection B is equals to A complement union B complement. The DeMorgan's Laws presented here are for two sets, they can easily be extended for more than two sets.

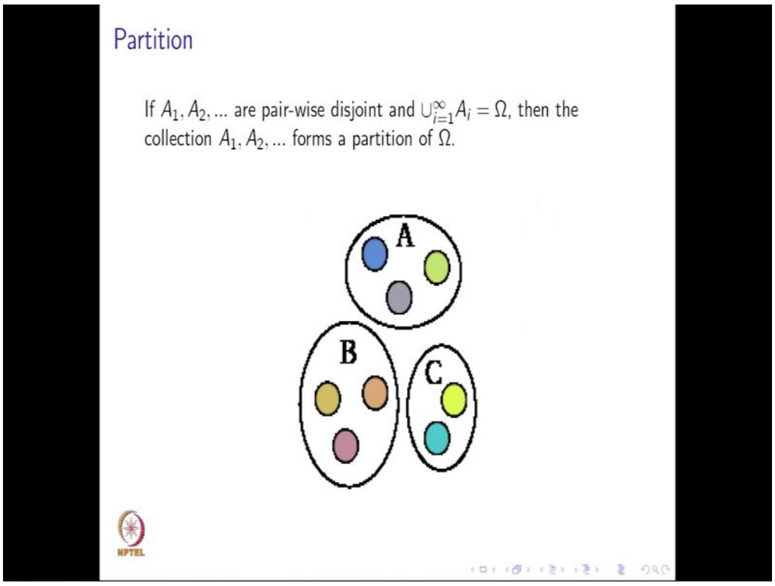
(Refer Slide Time: 04:57)



Coming back to the events, two events A and B are set to be disjoint or mutually exclusive if the intersection of the two sets is empty. Extending this concept to multiple sets, we say that a sequence of events, A1, A2, A3, and so on are pairwise disjoint. If Ai intersection Aj is equals to null, for all i not equals to j.

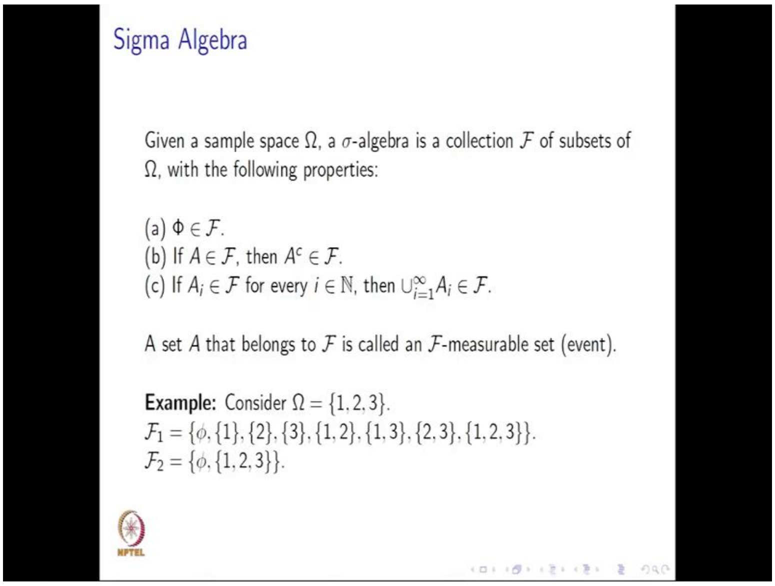
In the example below, if each of the letters represents an event, then the sequence of events A through E are pairwise disjoint since the intersection of any pair is empty.

(Refer Slide Time: 05:33)



If events A1, A2, A3, so on are pairwise disjoint and the union of the sequence of events gives rise to the sample space, then the collection A1, A2, and so on is set to form a partition of the sample space Omega. This is illustrated in the figure below.

(Refer Slide Time: 05:56)



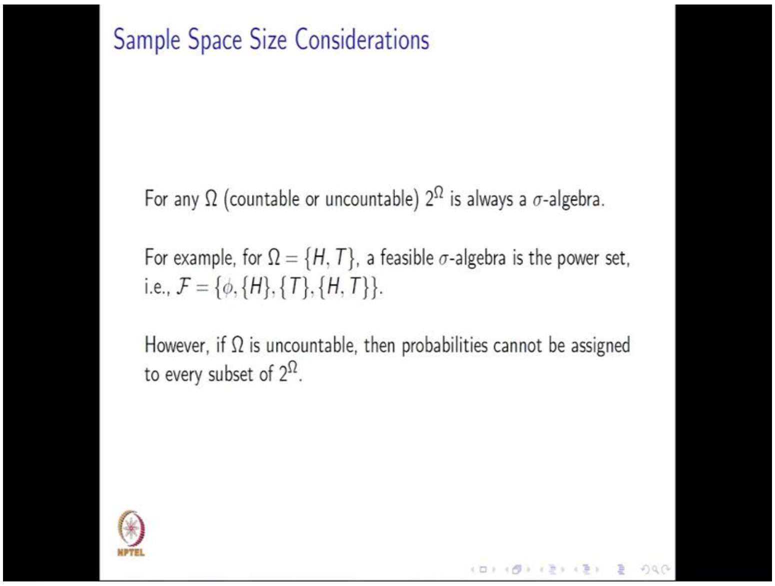
Next, we come to the concept of a sigma-algebra. Given a sample space, Omega, a sigma-algebra is a collection F of subsets of the sample space with the following properties. The null set is an element of F. If A is an element of F then A complement is also an element of F.

If Ai is an element of F for every i belonging to the natural numbers, then union i equals to 1 to infinity Ai is also an element of F. A set A that belongs to F is called an F measurable set. This is what we naturally understand as an event. So going back to the third property, what this essentially says is that, if there are a number of events which belong in the sigma-algebra, then the countable union of these events also belongs in sigma-algebra.

Let is considered an example. Consider a Omega equals 1, 2, 3. This is our sample space. With this sample space, we can construct a number of different sigma-algebras. Here, the first sigma-algebra F1 is essentially the power set of the sample space. All possible events are present in the first sigma-algebra.

However, if we look at F2, in this case, there are only two events; the null set or the sample space itself. You should verify that for both F1 and F2, all three properties listed above are satisfied. Now that we know what a sigma-algebra is, let us try and understand how this concept is useful.

(Refer Slide Time: 07:32)



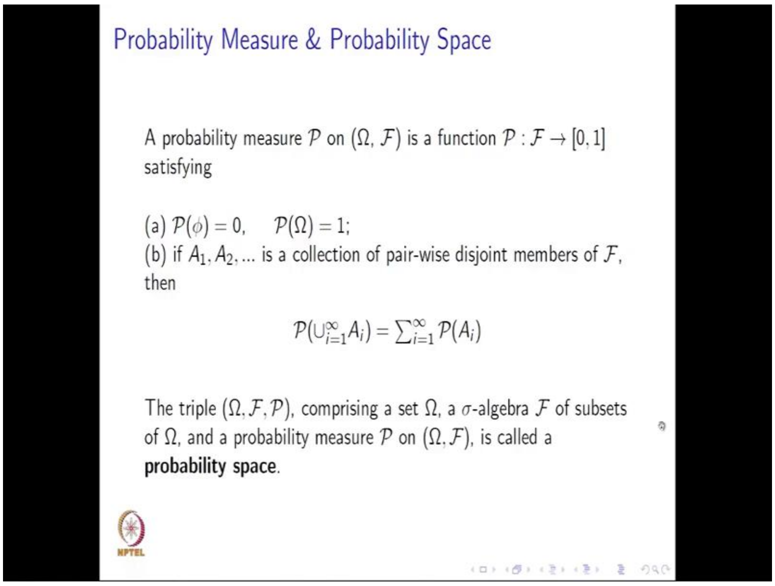
First of all, for any Omega, countable, or uncountable, the power set is always a sigma-algebra. For example, for the sample space comprising of two elements H comma T, a feasible sigma-algebra is the power set. This is not the only feasible sigma-algebra as we have seen in the previous example, but always the power set will give you a feasible sigma-algebra.

However, if Omega is uncountable, then probabilities cannot be assigned to every subset of the power set. This is the crucial point, which is why we need the concept of sigma-algebras. So just to recap, if the sample space is finite or countable, then we can kind of ignore the concept of sigma-algebra because, in such a scenario, we can consider all possible events, that is the power set of the sample space, and meaningfully apply probabilities to each of these events.

However, this cannot be done when the sample space is uncountable. That is if Omega is uncountable, then probabilities cannot be assigned to every subset of 2 to the Omega. This is where the concept of sigma-algebra shows it use.

When we have an experiment in which the sample space is uncountable, for example, let us say the sample space is the set of real numbers. In such a scenario, we have to identify the events which are of importance to us and use this along with the three properties listed in the previous slide to construct a sigma-algebra and probabilities will then be assigned to the collection of sets in the sigma-algebra.

(Refer Slide Time: 09:29)



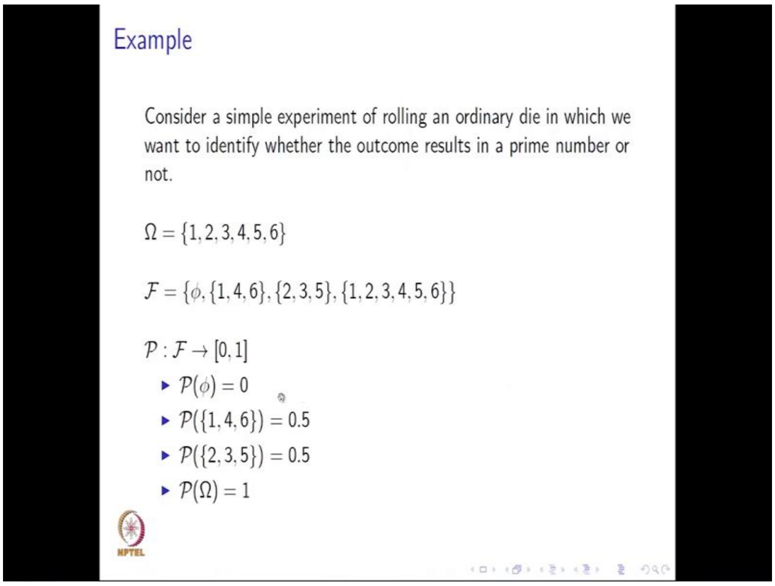
Next, we look at the important concepts of probability measure and probability space. A probability measure P on a specific samples space Omega and sigma-algebra F is a function from F to the close interval 0 comma 1 and satisfies the following properties.

Probability of the null set equals to 0, probability of Omega equals to 1, and if A1, A2, and so on is a collection of pairwise disjoint members of F, then probability of the union of all such members is equal to the sum of their individual probabilities. Note that this holds because the sequence A1, A2 is pairwise disjoint.

The triple, Omega F P, comprising a samples space Omega, a sigma-algebra F, which are subsets of Omega and a probability measure P defined on Omega comma F. This is called a probability space.

For every probability problem that we come across, there exist a probability space comprising of the triple, Omega F P. Even though we may not always explicitly take into consideration this probability space when new solve the problem, it should always remain in the back of our heads.

(Refer Slide Time: 10:42)



Let us now look at an example where we do consider the probability space involved in the problem. Consider a simple experiment of rolling an ordinary die in which we want to identify whether the outcome results a prime number or not.

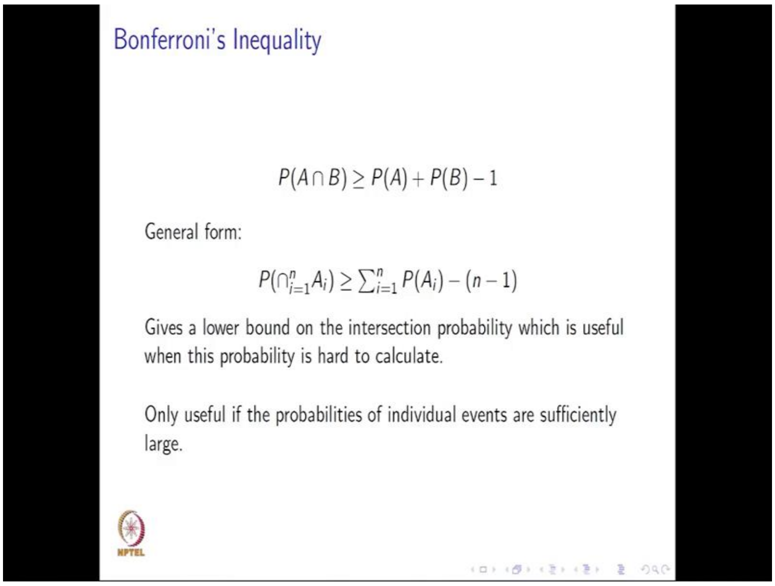
The first thing to consider is the sample space. As there are only six possible outcomes in our experiment, the sample space here is consists of the numbers between 1 to 6. Next, we look at the sigma and algebra. Note that, since the sample space is finite, you might as well consider all possible events. That is the power set of the sample space.

However, note that the problem dictates that we are only interested in two possible events, that is whether a number, whether the outcome is prime or not. Thus restricting ourselves to these two events we can construct a simpler sigma-algebra.

Here, we have two events which correspond to the events we are interested in, and the remaining two events follow from the properties, which the sigma-algebra has to follow. Please verify that the sigma-algebra listed here does actually satisfy the three properties that we had discussed above.

The final component is the probability measure. The probability measure assigns a value between 0 and 1 that is the probability value to each of the components of the sigma-algebra. Here, the values listed assumes that the die is fair in the sense that the probability of each face is equals to 1 by 6.

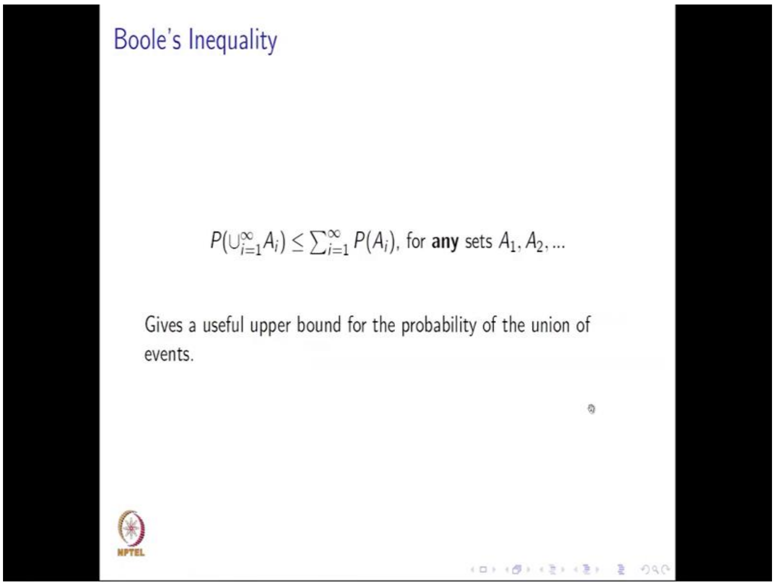
(Refer Slide Time: 12:14)



Having covered some of the very basics of probability, in the next few slides, we will look at some rules which allow us to estimate probability values. The first thing we will look at is known as the Bonferroni's Inequality. According to this inequality, probability of A intersection B is greater than equals to probability of A plus probability of B minus 1, the general form of this inequality is also listed.

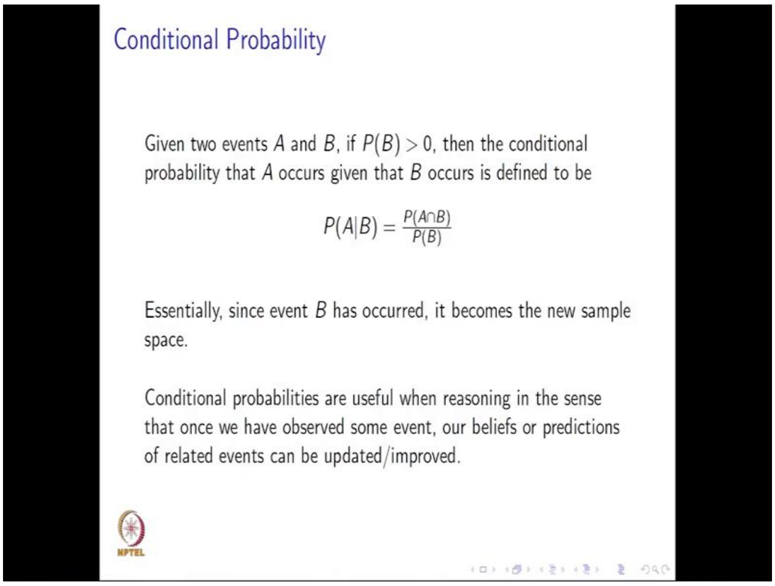
What this inequality allows us to do is to give a lower bound on the intersection probability. This is useful when the intersection probability is hard to calculate. However, if you notice the right-hand side of the inequality, you should observe that this result is only meaningful when the individual probabilities are sufficiently large. For example, if the probability of A and the probability of B, both these values are very small, then this minus 1 term dominates and the result does not make much sense.

(Refer Slide Time: 13:12)



According to the Boole's Inequality, for any sets, A1, A2, and so on the probability of the union of these sets is always less than equals to the sum of their individual probabilities. Clearly, this gives us a useful upper bound for the probability of the union of events. Notice that this equality will only hold when these sets are pairwise disjoint.

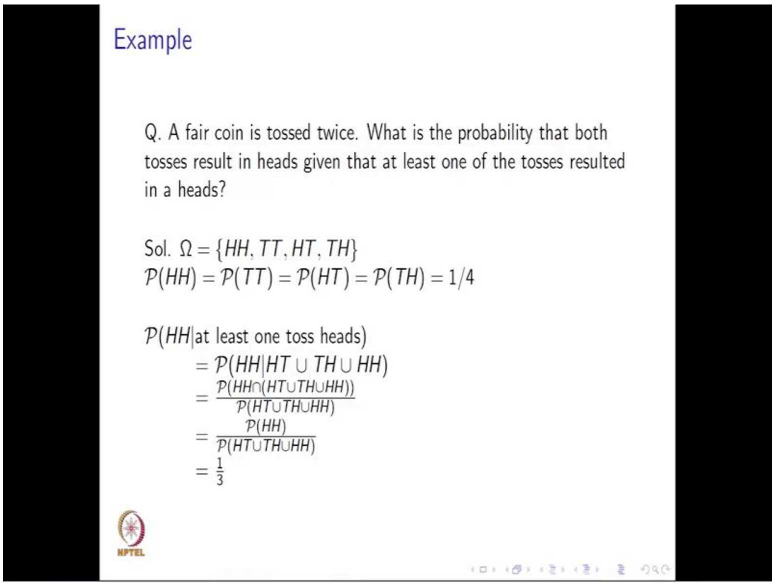
(Refer Slide Time: 13:39)



Next, we look at conditional probability. Given two events A and B, if probability of B is greater than 0, then the conditional probability that A occurs, given that B occurs is defined to be probability of A given B is equals to probability of A intersection B by probability of B.

Essentially, since event B has occurred, it becomes a new sample space. And now, the probability of A is accordingly modified. Conditional probabilities are very useful when reasoning in the sense that once we have observed some event, our beliefs or predictions of related events can be updated or improved.

(Refer Slide Time: 14:20)



Let us try working out a problem in which conditional probabilities are used. A fair coin is tossed twice, what is the probability that both tosses result in heads, given that at least one of the tosses resulted in a heads. Go ahead and pause the video here and try working out the problem yourself.

From the question, it is clear that there are four elementary outcomes. Both tosses resulted in heads, both came up tails, the first came up heads while the second toss came of tails, and the other way round. So means, we are assuming that the coin is fair, each of the elementary outcomes have the same probability of occurrence, equals 1 by 4.

Now, we are interested in the probability that both tosses come up heads, given that at least one resulted in a heads. Applying the conditional probability formula, we have probability of A given B equals to probability of A intersection B divided by probability of B. Simplifying the intersection in the numerator, we get the next step.

Now, we can apply the probability values of the elementary outcomes to get the result of 1 by 3. Note that in the denominator, each of these events is mutually exclusive. Thus the probability of the union of these three events is equals to the summation of the individual probabilities.

As an exercise try solve the same problem with the modification that we observed only the first toss coming up heads. That is, we want the probability that both tosses result in heads given that the first toss resulted in a heads. Does this change the problem?

(Refer Slide Time: 16:04)

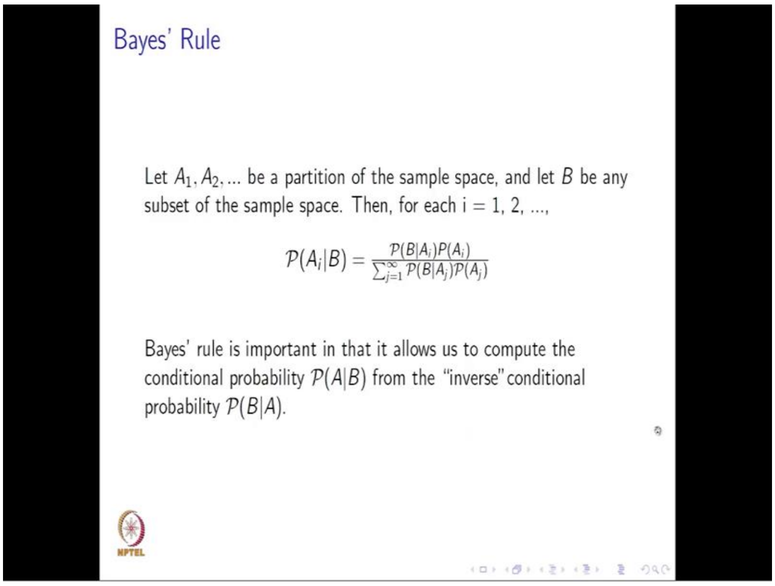
Next, we come to a very important theorem called the Bayes' Theorem or the Bayes' Rule. We start with the equation for the conditional probability. Probability of A given B is equal to probability of A intersection B by probability of B.

Rearranging, we have probability of A intersection B equals to probability of A given B into probability of B. Now, instead of starting from, with probability of A given B, if we have started with probability of B given A, we would have got probability of A intersection B equals to probability of B given A into probability of A.

These two right-hand sides can be equated to get probability of A given B is equal to, into probability of B is equals to probability of B given A into probability of A. Now, taking this probability of B to the right-hand side, we get probability of given B is equal to probability of B given A into probability of A by probability of B. This is what is known as the Bayes' Rule.

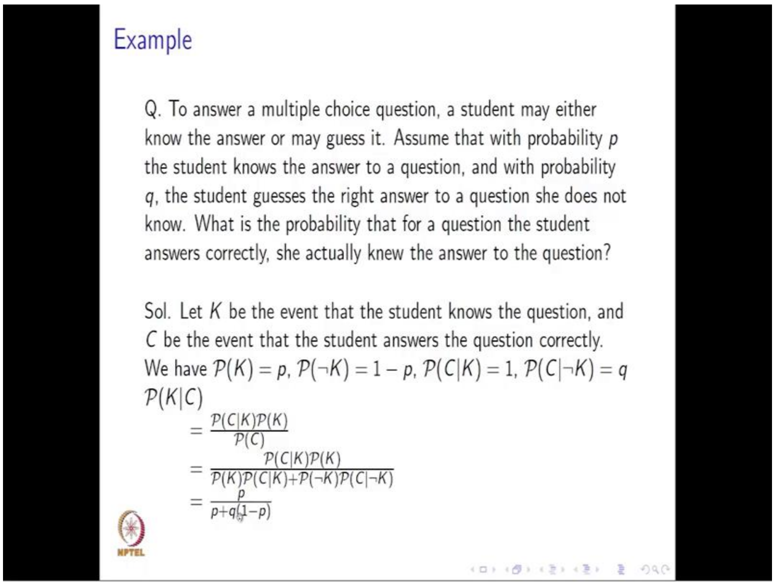
Note that, what it essentially says is, if I want to find the probability of A, given that B happened, I can use the information of probability of B given A along with the knowledge of probability of A and probability of B to get this value. As we will see, this is a very important formula.

(Refer Slide Time: 17:25)



Here, we again present the Bayes' Rule in an expanded form, where A1, A2, and so on form a partition of the sample space. As mentioned, Bayes' Rule is important in that it allows us to compute the condition of probability of A given B from the inverse conditional probability of B given A.

(Refer Slide Time: 17:45)



Let us look at a problem in which the Bayes' Rule is applicable. To answer a multiple-choice, question, a student may either know the answer or may guess it. Assume that with probability p, the student knows the answer to a question and with probability q, the student guesses the right answer to a question she does not know. What is the probability that for a question the student answers correctly, she actually knew the answer to the question? Again, pause the video here and try solving the problem yourself.

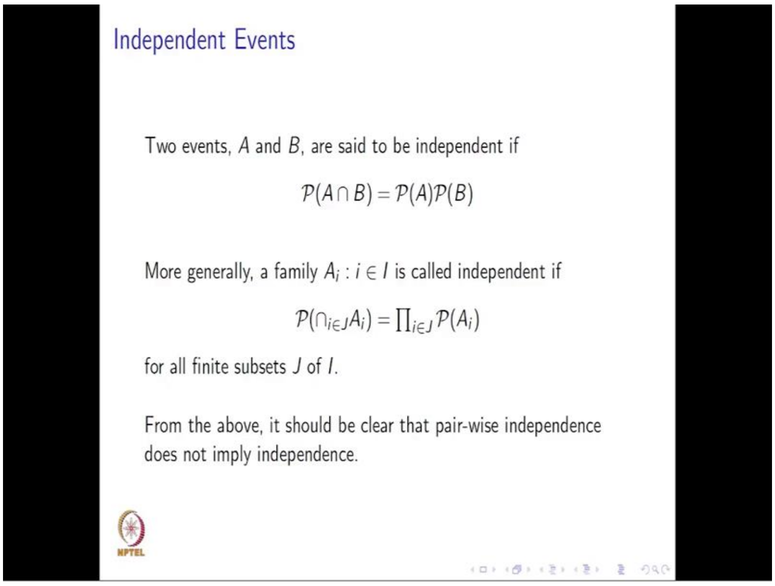
Okay. Let us first assume that K is the event that the student knows the question and that C be the event that the student answers the question correctly. Now, from the question, we can gather the following information.

The probability that the student knows the question is p, hence the probably that the student does not know the question is equals to 1 minus p. The probability that the student answers the question correctly, given that she knows the question is equals to 1 because if she knows the question, she will definitely answer it correctly. Finally, the probability that the student answers the question correctly, given that she makes a guess, that is, she does not know the question is q.

We are interested in the probability of the student knowing the question, given that she has answered it correctly. Applying Bayes' Rule, we have probability of K given C is equal to probability C given K into probability of K by probability of C. The probability of answering the question correctly can be expanded in the denominator to consider the two situations, probability of answer the question correctly, given that the student knows the question, and probability of answering the question correctly given that the student does not know the question.

Now, using the values which we have gathered from the question, we can arrive at the answer p by p plus q into1 minus p. Note here that the Bayes' Rule is essential to solve this problem because while, from the question itself, we have a handle on this value probability of C given K, there is no direct way to arrive at the value of probability of K given C.

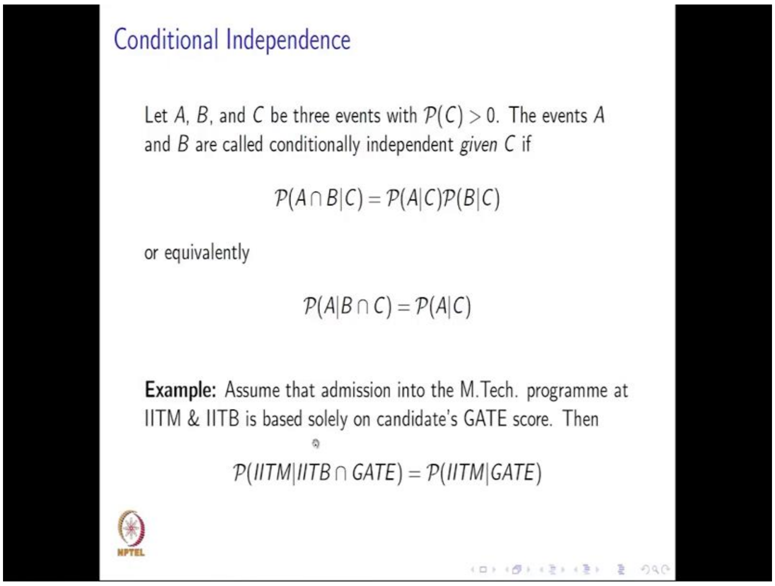
(Refer Slide Time: 20:04)



Two events A and B are said to be independent if probability of A intersection B is equals to probability of A into probability of B. More generally, a family of events Ai, where i is an element of the integers is called independent if probability of some subset of the events Ai is equal to the product of the probability of those events.

Essentially, what we are trying to say here is that, if you have a family of events Ai, then the independence condition holds only if, for any subset of those events, this condition holds. From this, it should be clear that pairwise independence does not imply independence, that is, pairwise independence is a weaker condition.

(Refer Slide Time: 20:57)



Extending the notion of independence of events, we can also consider conditional independence. Let A, B, and C be three events with probability of C strictly greater than 0. The events A and B are called conditionally independent given C if probability of A intersection B given C equals to probability of A given C into probability of B given C. This condition is very similar in form to the previous condition or independence of events.

Equivalently, the events A and B are conditionally independent given C if probability of A given B intersection C equals to probability of A given C. This latter condition is quite informative.

What it says is that the probability of A calculated after knowing the occurrence of event C is same as the probability of A calculated after having knowledge of occurrence of both events B and C, thus observing the occurrence or non-occurrence of B does not provide any extra information. And thus, we can conclude that the events A and B are conditionally independent given C. Let us considered an example.

Assume that admission into the M.Tech program at IIT Madras and IIT Bombay is based solely on candidate’s GATE score, then probability of admission into IIT Madras given knowledge of the candidate’s admission status in IIT Bombay, as well as the candidate’s GATE score is equal equivalent to the probability calculated simply knowing the candidate’s GATE score.

Thus, knowing the status of the candidate’s admission into IIT Bombay does not provide any extra information. Hence, since the condition is satisfied, we can say that admission into the program at IIT Madras and admission into the program at IIT Bombay are independent events, given knowledge of the candidate’s GATE score.