NPTEL

NPTEL ONLINE CERTIFICATION COURSE

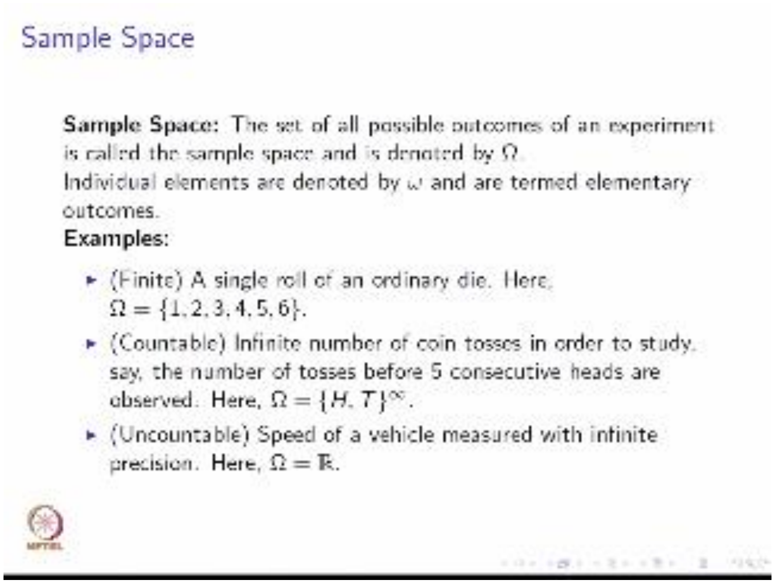
Introduction to Machine Learning

Probability Basic-1

Hello and welcome to the first tutorial in the introduction to machine learning course my name is Preyathosh I am one of the teaching assistance for this course and this tutorial will looking at some of the basics of probability theory before we start let us discuss the objective of this tutorial the aim here is not to teach the concept of probability theory in any great detail instead we will just be eroding high level over view of the concepts start will we uncounted later on in the course the idea here is that for those of you have done course and probability theory or are otherwise from wilier with the content this tutorial should act as a refresher.

For others who may find some of the concepts and some wilier we recommend that you go back and prepare those concepts from say introductive text book or any other resource so that when you encounter those concepts later on in that course you should be comfortable with that okay.

(Refer Slide Time: 01:12)

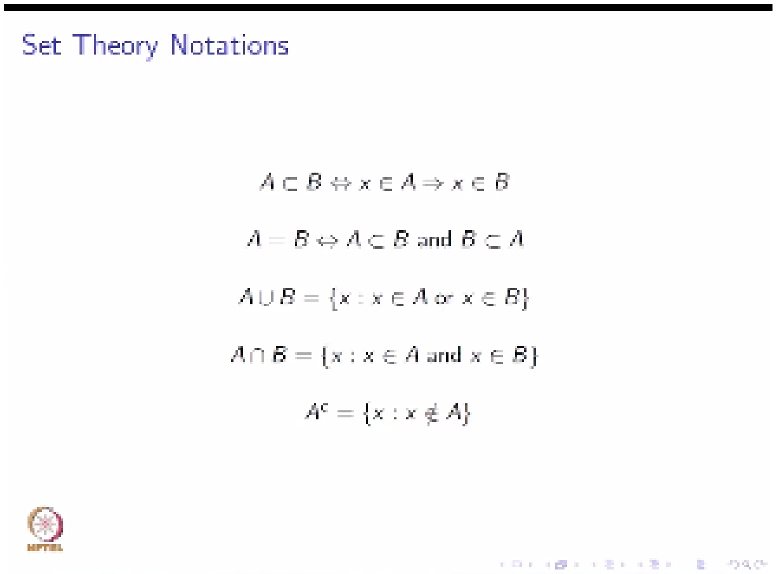


To start this tutorial will look at the technicians or some other fundamental concepts the first one to consider is that are the sample space the set of all possible are comes even experiment is call the sample space and it is noted by **Ω**. Individual elements are denoted by ω and are term elementary out comes let us consider some examples in the first example the experiment consist of rolling and ordinary die the sample space here is a set of number between one and six each in which will element here represents one of the six possible outcomes rolling a die.

Note that in this example the samples space is finite in the second example the experiment consist of tossing a coin repeatedly and till a specify condition is observed, here we are looking to observe five consecutive heads before terminating the experiment, a sample space here is uncountable infinite we the individual elements are represented using a sequence of the action T and T is where action T is turn for H T respectively and the final example experiment consist on measuring the speed of the vehicle with infinite precision.

Assuming that the vehicle speeds can be negative the sample space is clearly the set of real numbers here we are serve the sample space can be uncountable.

(Refer Slide Time: 02:34)



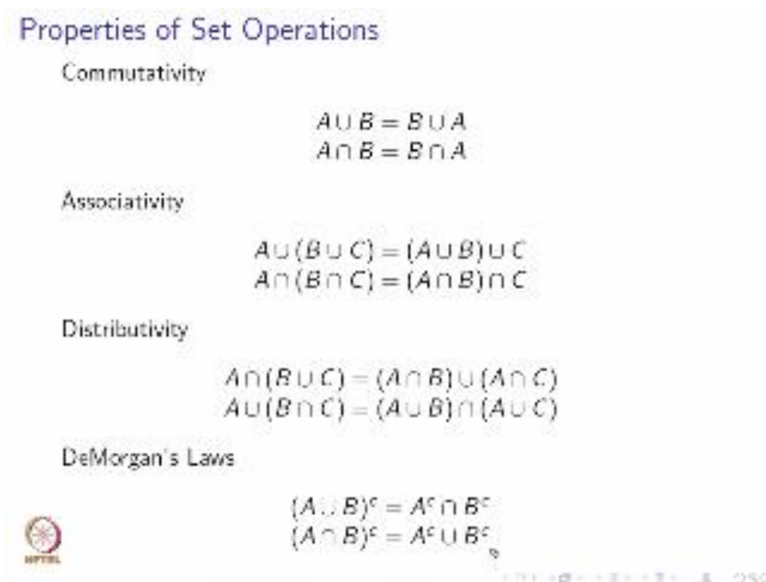
The next concept we look at is that is of an event an event is any collection of possible outcomes of an experiment that is any subset with the sample space. The reason why events are important to us is because in general when we conduct an experiment we really that not interested in the elementary outcomes rather we are more interested in some sub sets of the elementary outcomes. For example on rolling a die we might be interested in observing whether the outcome was even or hard.

So for example on a specific roll of a die we of let us say we observed at the outcome was hard in the scenario whether the outcome was actually a one or three or a five is not as important to us as the fact there it was hard since we are considering sets in terms of sample spaces and event we will quickly go through basic set theory notations. As we show capital letters indicates sets and small letters indicate set elements.

We first look at the sub set relation for all x if x element of a in flies x element of P can we say that a is a subset of B or A is contained ion B, two sets same be are set a b equal if both a sub set of b and b sub set of a holds. The union of two sets same b is raise to a new set which contains elements of both a and b, similarly the intersection of two sets gets rise to a new set which contains of only those elements which are common to both a and b.

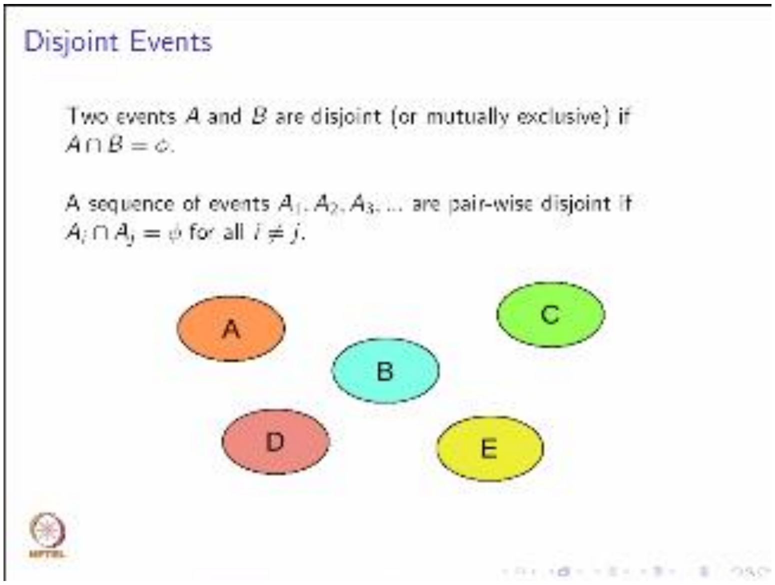
Finally the complement of the set a is essentially the set which contains all elements in the universal set except for the element present in A in our case the universal set is essentially the sample space.

(Refer Slide Time: 04:25)



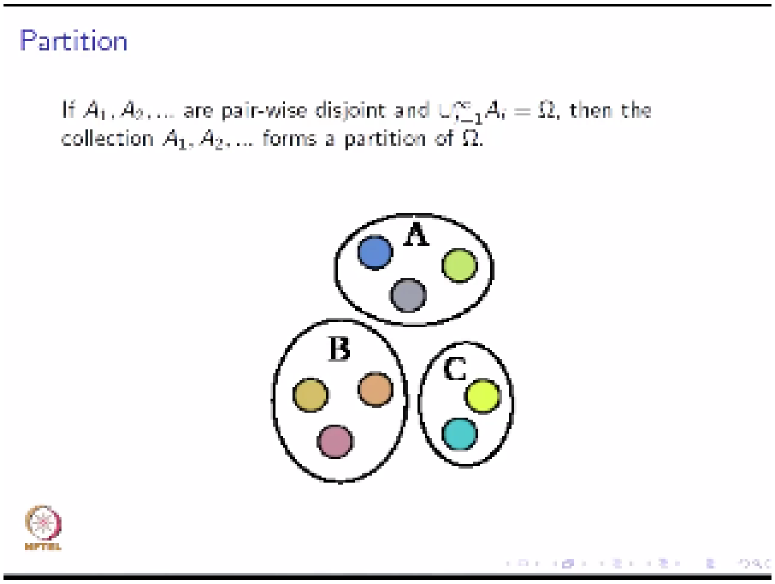
This slide list out the different properties of set operation such as commutativity assocoativity and distributivity which is should all come here with it is also list out the DeMorgan’s Laws which can be very useful, according to DeMorgan’s Law the complement of the set a union b is equal to a complement inter section b complement similarly the complement of the section a in to section b is equals to a complement union b complement that DeMoragan law is presented here or for two sets they can easily extended for more than two sets.

(Refer Slide Time: 04:59)



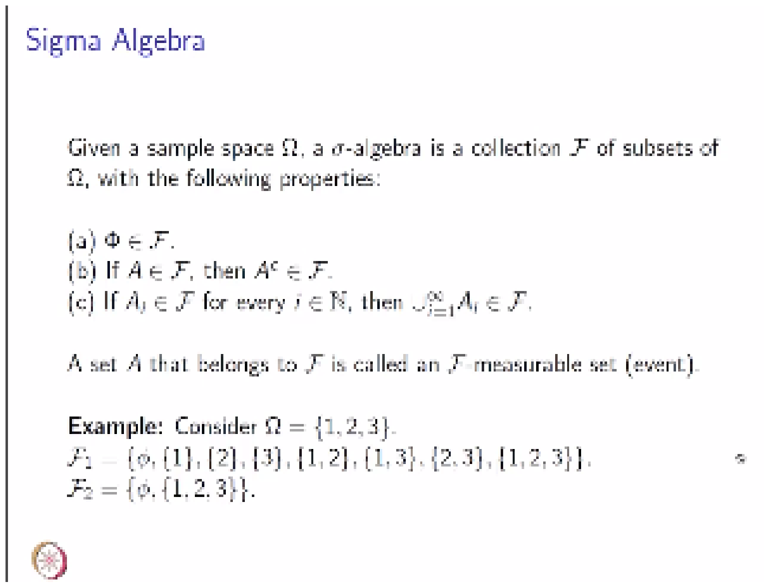
Coming back to events two events a and b are set to be disjoint or mutually exclusive in the inter section of two set is empty. Extending this concept to multiple sets we say that our sequence of events A1A2A3 and so on are pair wise disjoint if Ai inter section Aj is request to null for all I not equals to j the example below if each of the letters represents an event then the sequence of the events A3 are there wise rejoined since the introduction of any pair is empty.

(Refer Slide Time: 05:35)



If events A1 A2A3 so on are pair wise disjoined and the union of the sequence of events describes to the samples face then the collection A1 A2 and so on is set to form of patrician of the sample space Ω, this is illustrated in the figure below.

(Refer Slide Time: 05:57)



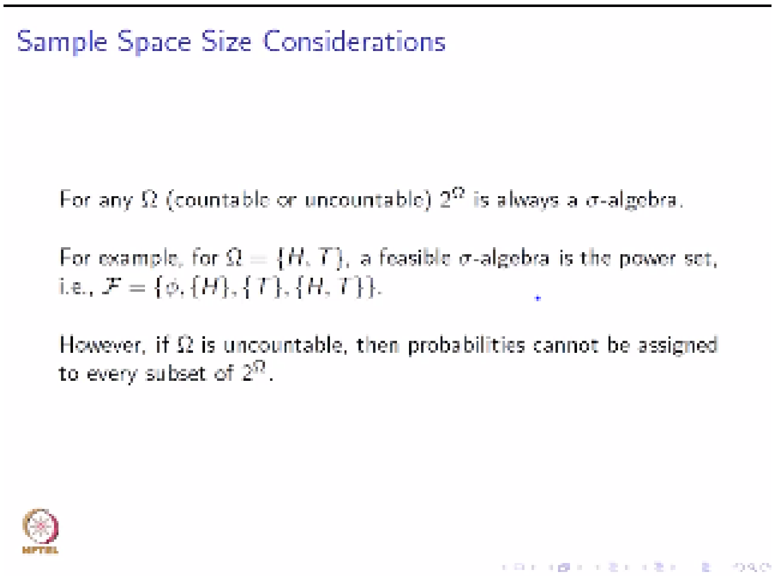
Next we come to the concept over σ algebra given a sample space **Ω** a σ algebra is a collection as of sub sets of the sample space with for following properties the null set is an element of F, if a is an element of F and a complement is also measurement of F if Ai is an measurement of f for every I belong to the natural numbers then union I = 1 to infinity Ai is also an element of F as set a that belongs to F is called as a F measure of full set this is what we naturally understands as an event.

So going back to the third property what this essentially says is that if there are number of events which belong in the set σ algebra then the countable union of this event also belongs in the σ algebra, let us considering an example considering a **Ω** = 123 this is our sample space with this sample space we can constrict a number of different σ algebra here the first σ algebra F1 is essentially the power set of the sample space.

All possible events are present in the first σ 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 out satisfied. Now that we know what a σ algebra is let us try and understand how this concepts is useful.

(Refer Slide Time: 07:34)

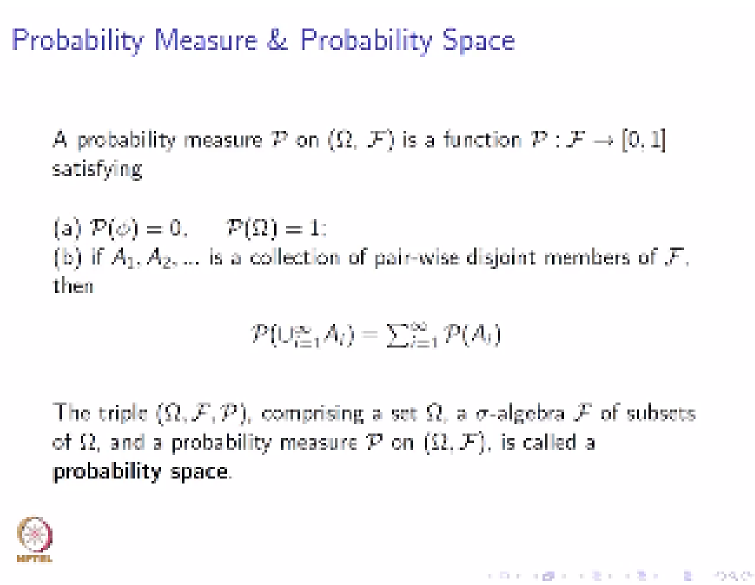


First of all for any Ω countable or uncountable the power set is always a σ algebra for example form the sample space comprising of two elements H, T a feasible σ algebra is the power set this is not the only feasible σ algebra as we have seen in the previous example but always the power set will give you a feasible σ algebra, however if Ω is uncountable the probability is cannot be assign to every subset of the power set.

This is the crucial point which is why we need the concept was σ algebra so just to recap if the sample space is finite or countable then we can kind away ignore the concept of σ algebra because in such a scenario we can consider all possible events, that is the power set of the sample space and meaningfully apply probability to each of these events however this cannot be done when the seems sample space is uncountable that is AΩ is uncountable and probabilities cannot be assign to every sub set of due to the Ω.

This is where the concept of σ algebra shows its 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 important to us and use this along with the 3 properties listed in the previous slide to construct a σ algebra and probabilities will then assign to the collection of sets in the σ algebra.

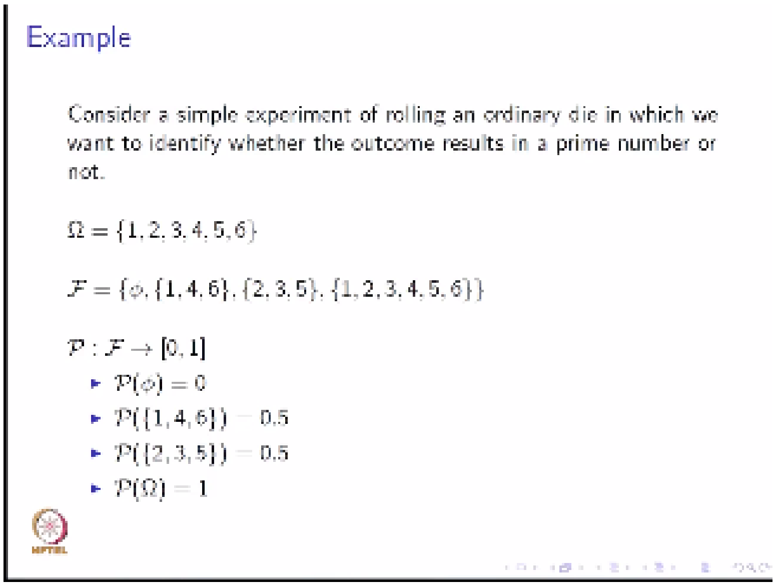
(Refer Slide Time: 09:31)



Next we look at the important concepts of probability measure in probability space a probability measure P on a specific sample space Ω and σ algebra F is a function from F to the close intervals 0, 1 satisfies the following properties, probabilities and null set it goes to 0 probability of Ω = 1and if A1 A2 and so on is a collection of pair wise disjoint number of F and probability of the union of also is remember is equals to the sum of their individual probabilities.

Now that the scolds because the sequence A1 A2 is pair wise disjoint the triple Ω FP comprising as sample space Ω or σ algebra F which are sub sets of Ω and a probability measure P defined on Ω, F this is called a probability space, for every probability problem that we come across the resist of probability space comprising of the triple Ω FP, even though you may not always especially taking two concentration this probability space is when you solve the problem it should always remain the back of our heads.

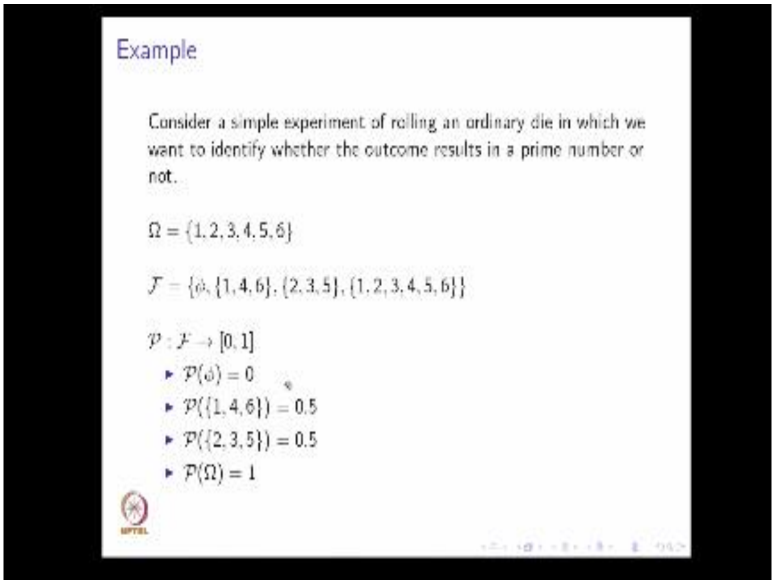
(Refer Slide Time: 10:45)



Let us now look at an example well we two consider the probability spaces involved in the problem consider a simple experiment of rolling an ordinary die in which we want to identify whether the outcome results in a prime number or not, the first in to consider as a sample space is there only six possible outcomes in our experiment a sample space here is consist of a numbers between 1 to 6 next we look at the σ algebra now that is the sample space is finite you might is well consider all possible events that is the power set of the sample space.

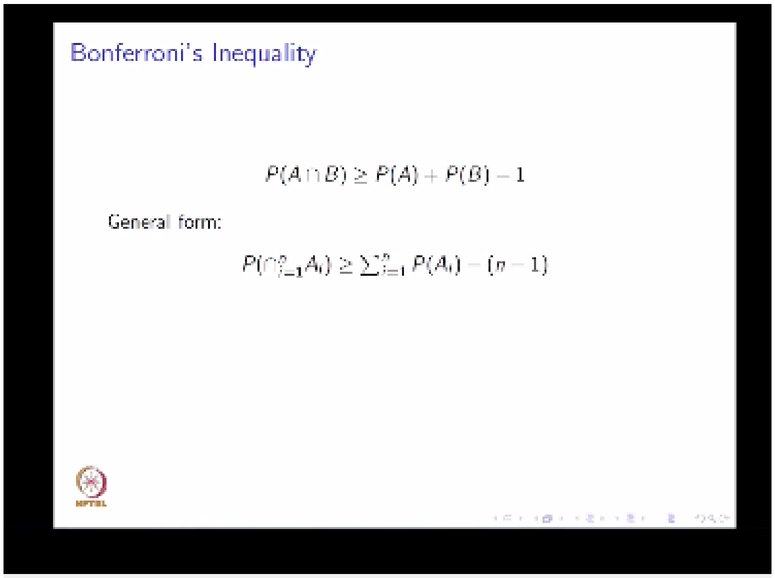
However not that the problem declares there we are only interested in two possible events that is whether a number whereas the outcomes is prime or not. This restating ourselves to these two events we can consider to simple σ algebra here we have two events which corresponds to the adventure interested in and the remaining two events follow from the properties which is σ algebra has to follow, please verify that the σ algebra listed here does actually satisfied the 3 property that we have discussed above the final component is the probability measure the probability measure assigns value it is 0 and 1 that is the probability value to each of the components of the σ algebra.

(Refer Slide Time: 12:02)



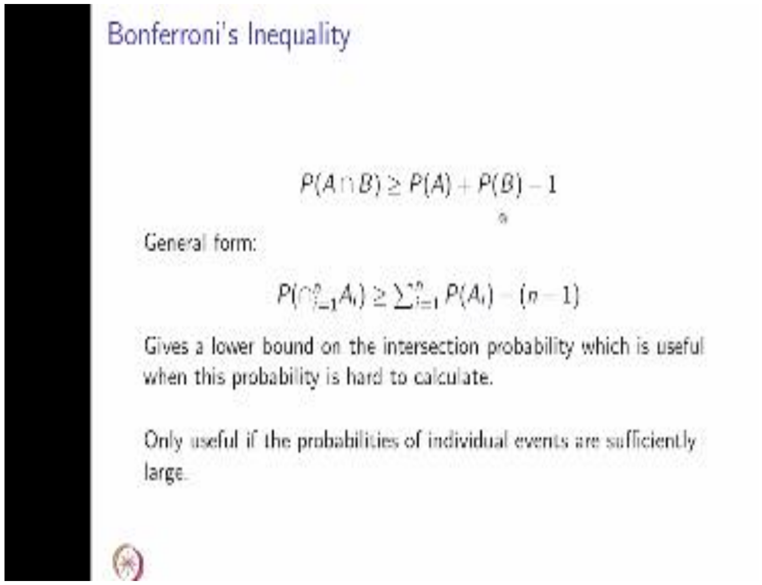
Here the values listed assumes that, the die is here in these that, the probability of each face = 1/6. Having covered some of the very basics of probability, in the next slides we look at home rules which allows estimating probability values.

(Refer Slide Time: 12:22)



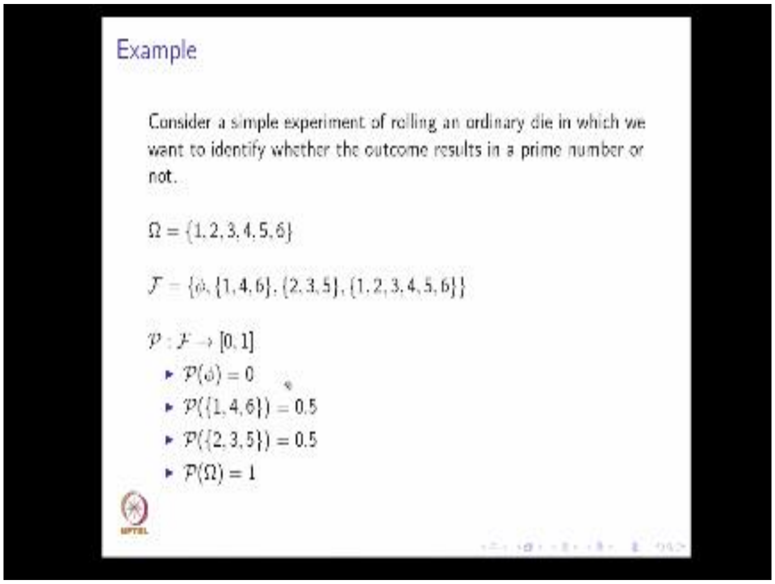
The first thing we look at is known as the Bonferroni's Inequality. According to this inequality, the probability of A intersection B I greater than or equal to probability of A + probability of B - 1, the general form of this inequality is also listed. What this equality allows u to do is to get the lower on in the interaction probability; this is useful when the intersection probabilities are to calculate.

(Refer Slide Time: 12:56)



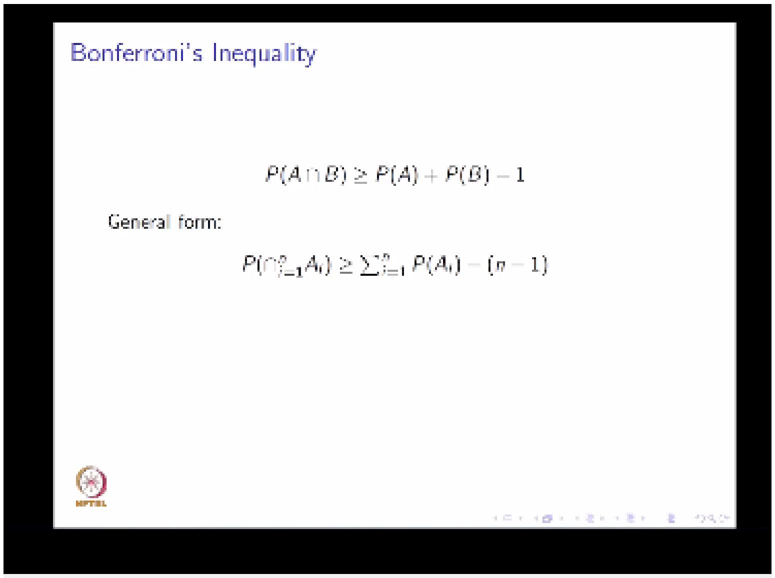
However if you notice the right hand side inequality, you should observe that this result is only meaningful when the individual probability are sufficiently large or example, if the probability of A and probability of B, both these values are very small and when this -1 from down leaves and the result assume, make much sense. According to the Boole' inequality, any sets, A1,A2 and so on, the probability of the union of the set is always less than or equals to the sum of their individual probabilities.

(Refer Slide Time: 12:02)



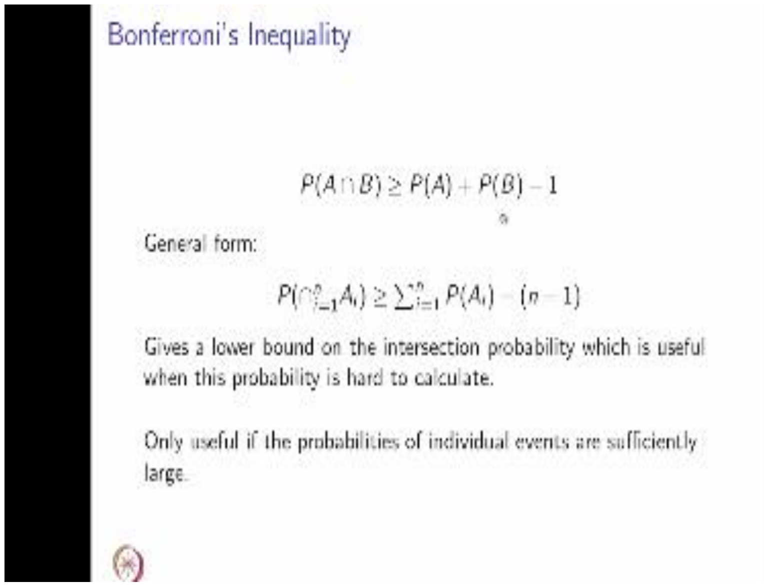
The final component is the probability measure, the probability measure assigns, value between 0 and 1, that is the probability value, to each other component of the σ algebra, here the values listed assumes that, the die is here in these that, the probability of each face = 1/6. Having covered some of the very basics of probability, in the next slides we look at home rules which allows estimating probability values.

(Refer Slide Time: 12:22)



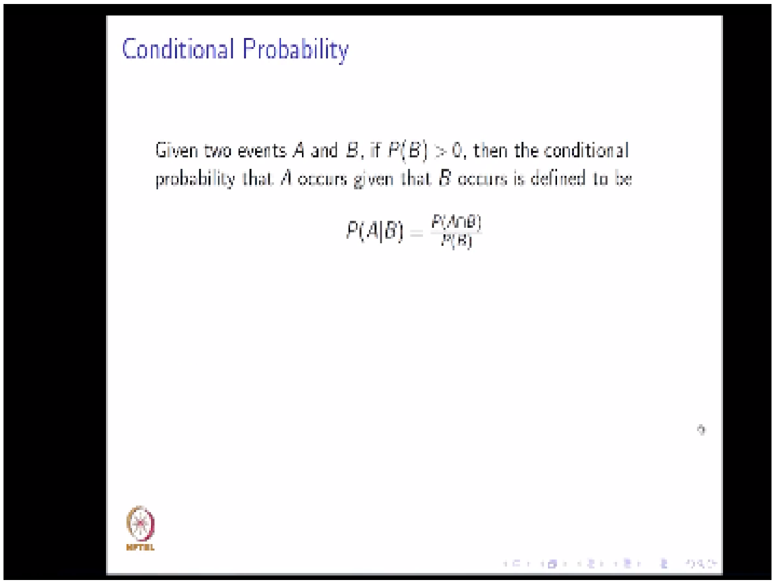
The first thing we look at is known as the Bonferroni's Inequality. According to this inequality, the probability of A intersection B I greater than or equal to probability of A + probability of B - 1, the general form of this inequality is also listed. What this equality allows u to do is to get the lower on in the interaction probability; this is useful when the intersection probabilities are to calculate.

(Refer Slide Time: 12:56)



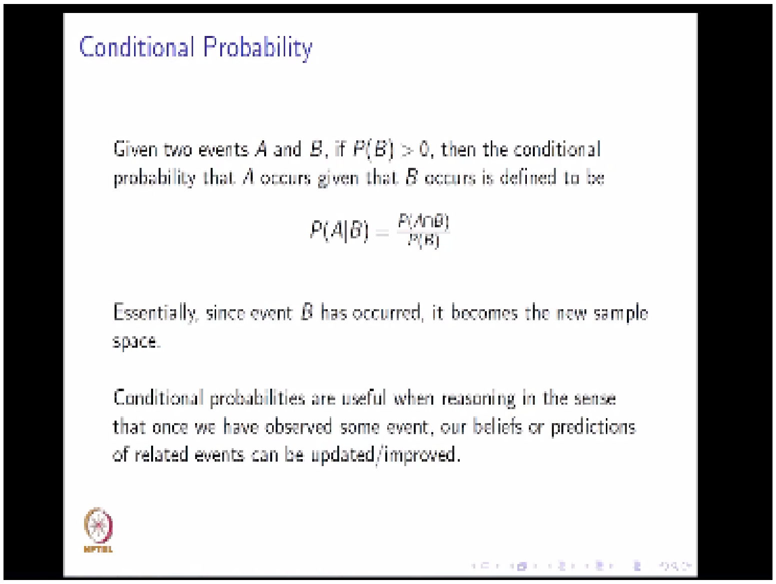
However if you notice the right hand side inequality, you should observe that this result is only meaningful when the individual probability are sufficiently large or example, if the probability of A and probability of B, both these values are very small and when this -1 from down leaves and the result assume, make much sense. According to the Boole' inequality, any sets, A1,A2 and so on, the probability of the union of the set is always less than or equals to the sum of their individual probabilities.

(Refer Slide Time: 13:43)



Clearly it gives a useful upper bound for the probability of the union of events. Notice that this equality only will hold, when these set are pair we just joined. Next we look at conditional probability, giving two events A and B, if probability of B I greater than 0, than the conditional probability A occur, given that B occurs is defined to the probability of A given B, due to probability of A intersection B, the probability of B.

(Refer Slide Time: 14:17)

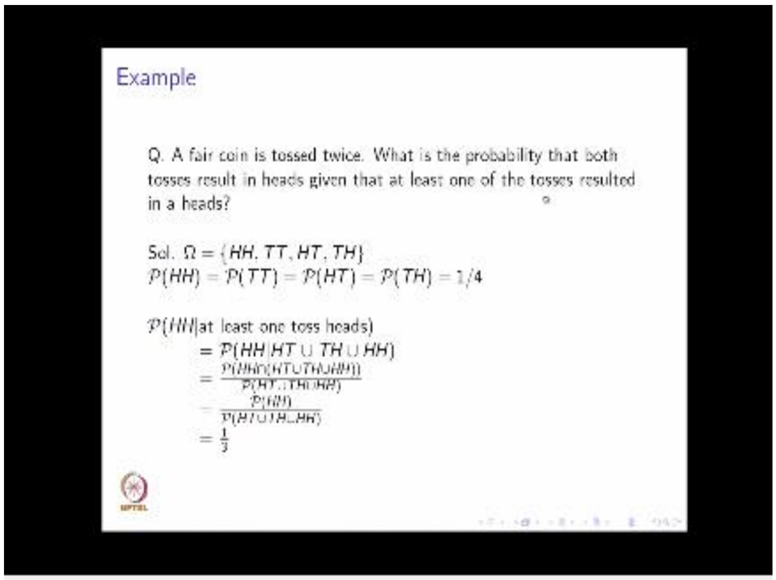


Essentially , since event B has occurred, it becomes the now sample space and now the probability of A is accordingly modified, Conditional probability that very useful when listening, in the sense that once we have observed a event or beliefs are prediction of relative difference Can be updated or improved. Let us try working pout a problem in which conditional probabilities are used, A fair coin is tossed twice, what I the probability that the both tosses result in heads given that at least one of the toes resulted in heads?

Go ahead and pause the video here and try working out the problem yourself. From the question it is clear that, that there are four relative outcomes, both tosses resulted in heads, both came up tails, the first came up heads while the second toss came up tails and the other way round. Sin ewe are assuming that the coin is fair, which of the elementary outcomes, probability of occurrence equal to 4. Now we are interested in the probability that both tosses come up heads, given that at least one resulted in heads.

Applying the conditional probability formula, here probability of A, given B, because the probability of A intersection B is 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 1/3, now that in the denominator, each of the events is mutually associative, thus the probability of union of third events, request to the summation of the individual probabilities.

(Refer Slide Time: 15:50)



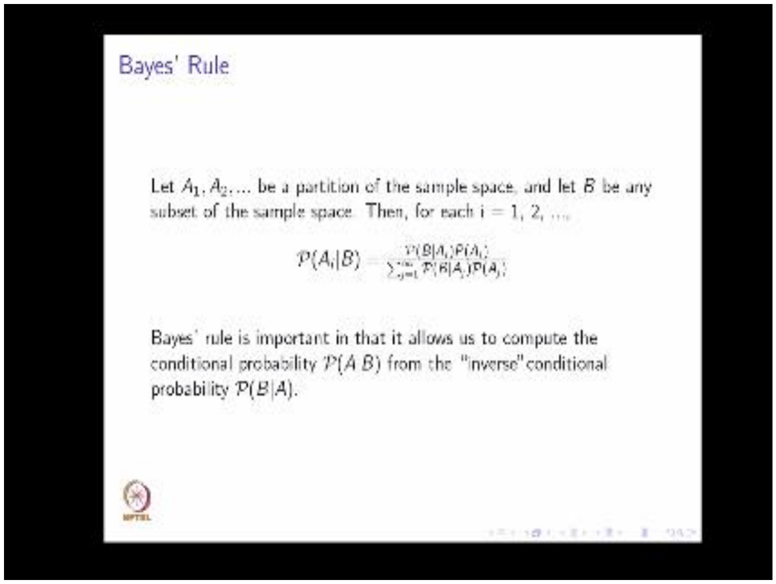
As an exercise try some of the problem with modification that w observed only the first toss, coming up heads that is we want the probability that four tosses is relate to heads giving that toughest tosses in heads, does this change the problem? Next we have come to the very important theorem called the Bay's theorem or the Bay's Rule; we start with the equation for the conditional probability.

Probability of A union B = Probability of A intersection B/ Probability of B, rearranging we have probability of A intersection B =Probability of A given B * Probability of B, now instead of starting from probability of A given the equation started with B given A, you would have got probability of A intersection B = P(B/A) P(A), these two right hand sides can be equated, P(A/B)P(B) =P(B/A)P(A), now taking this P(B) to the right hand side you get, P(A/B)=P(B/A)P(A)/P(B), this is what is known as the Bay's rule.

(Refer Slide Time: 17:20)

Note that what it essentially says is, if I want to find the probability of A, given that B happen, I can use information, P (B), given along with the knowledge of P(A)P(B), you get this value. As you can see this is the very important formula, here again we present the Baye’s rule in expanded form and so on from the partition of the samples. As mentioned Baye’s rule is important in order to compute the probability of A = B, for the reverse condition of Probability, P(B/A).

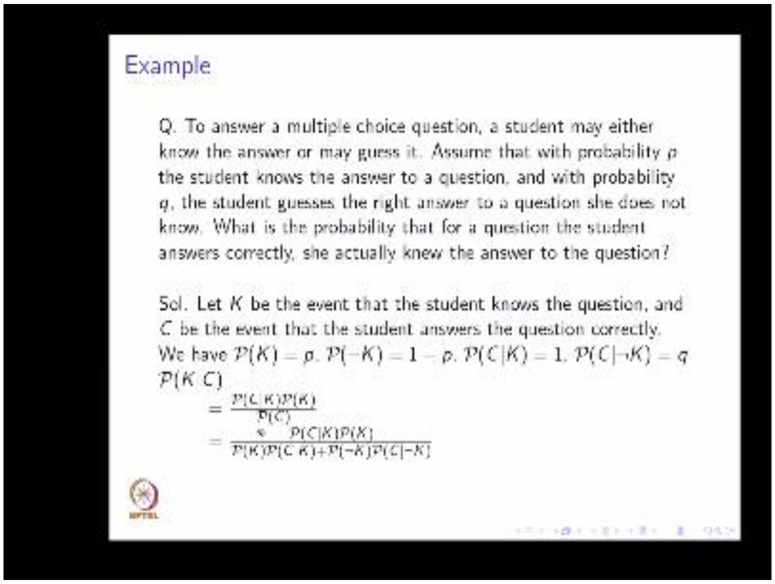
(Refer Slide Time: 17:45)



Let us look at a problem where, Bay's rule is applicable, to answer a multiple choice question a student may know the answer or any guess it, assume that the probability P that the student knows the answer to the question and the probability Q, guesses the right answer to question she does not know, what is the probability that, for the question that student answers correctly, she actually knew the answer to the question? Okay how's the bit you are hearing, try solving the problem yourself.

Let us first assume that K I given that the student knows the question and C be given question to the student answered correctly. Now from the question w can gather the following information, probability of student knows the question is B , hence the probability of student does not know the question is 1-P, The probability that the student answer the question correctly, given that she knows the question = 1, because if she knows the question, she answers the question correctly, finally the probability the student answer the question correctly, that she makes the guess, that he does not know the question is Q.

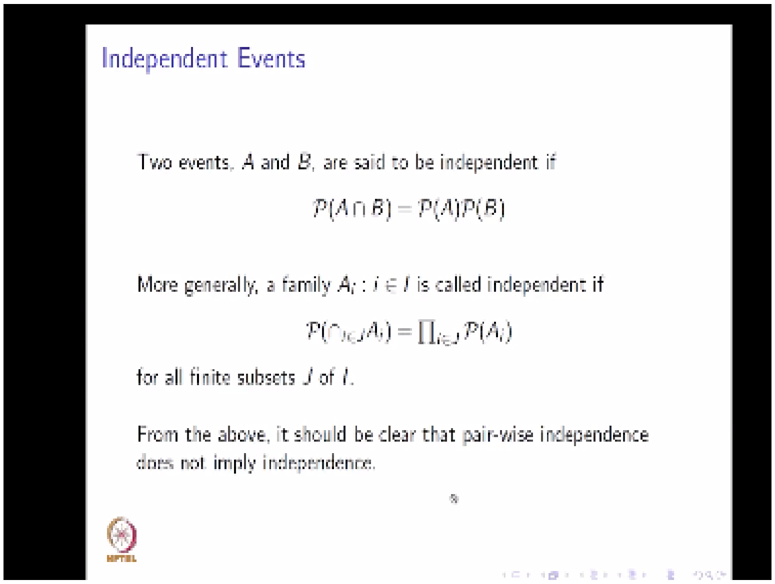
(Refer Slide Time: 18:47)



We are interested in the probability of the student knowing the question given that she answered it correctly, applying this rule we have (k/C)P(k)/P(c), there are probability of answering the question correctly can be expanded in the denominator to consider the two situations, probability of answering the questions correctly given that, the student knows the question and probability of answering the question correctly, may be the student does not know the question. Now using the values which we have gathered from the question, we can arrive the answer of P/P+Q(1-P).

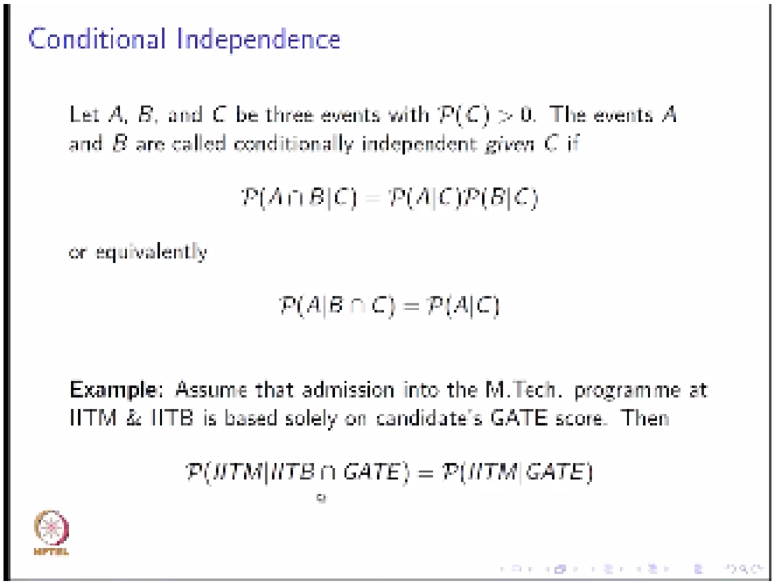
Note here that the babe's rule is essential to solve the problem, because while from the question itself, we have a handle on this P(C/K). There is no way to arrive t the value of (K/C). Two events A and B are aid to be dependent, P $(A \cap B) = P(A)P(B)$, more generally a family of Ai, where I is an element of integers, is called independent, if probability of some subset of the event Ai = product of the probability of those events, essentially what we are trying to tell hers is, if you have family of events Ai, then the independent condition holds only if for any subset of those events, this condition holds.

(Refer Slide Time: 20:55)



From this it is clear that that pair wise independence deserve weaker condition, extending the independence of events, we can also consider conditional independence. Let A, B and C be three events with p(C) strictly A and B are 0, the event A and B are called conditional independent given C if $P(A\cap D/C)=P(C) *P(A/C)$, this condition is very similar to the previous condition for independence of events, equivalently the events A and B are conditionally event dependent, $P(A\cap C)=P(A/C)$, this latter condition is quite informative, what it says is that the P(a) calculated after knowing the reference of event C is same as the P(A), calculated after having knowledge of occurrence of both events B and C.

(Refer Slide Time: 22:04)



Thus observing occurrences or non occurrence of P does not provide any extra information and thus we can conclude that the event C are conditional independent, event c, let us consider an example, assume that admission into the M tech programmer at IIT Madras and IIT Bombay, is based solely on candidate GATE score, then probability of admission into IIT Madras, given knowledge of the candidates admission status in IIT Bombay.

As well as the candidate GATE score is equivalent to the probability calculated simply knowing the candidates GATE score plus knowing the status of the candidate admission to IIT Bombay does not provide any extra information, hence, since the condition is satisfied, we can say that admission into program me at IIT Madras, or admission into program me at IIT Bombay are independent events given knowledge at the candidates GATE score

IIT Madras Production

Funded by

Department of Higher Education

Ministry of Human Resource Development