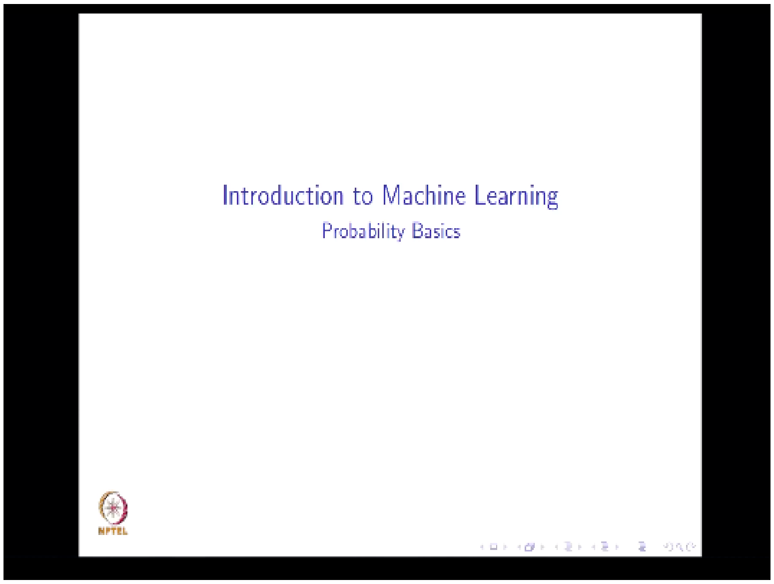
NPTEL

NPTEL ONLINE CERTIFICATION COURSE

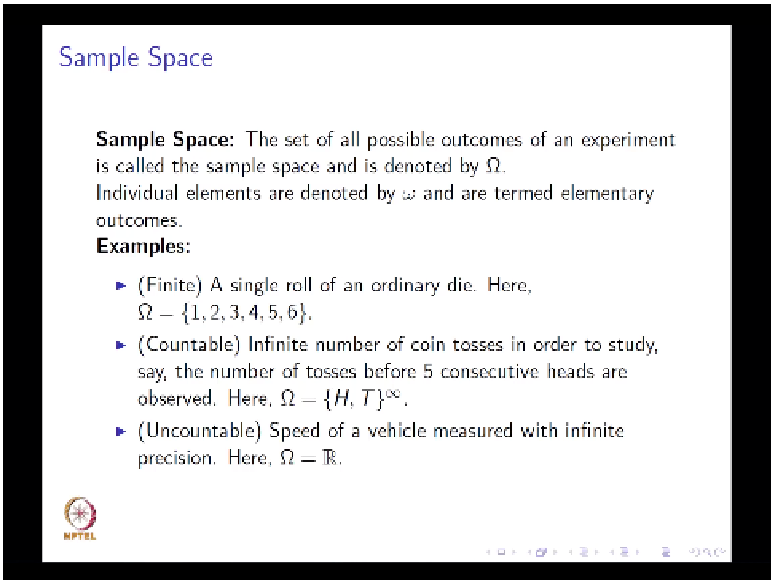
(Refer Slide Time: 00:18)



Hello and welcome to the first tutorial in the introduction of machine learning course my name is Priya Tosh 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 force 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.

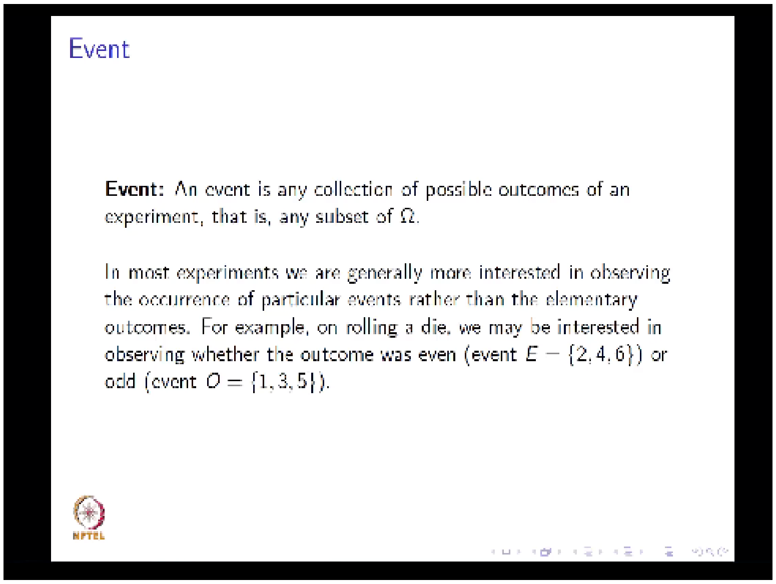
(Refer Slide Time: 01:13)



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 okay to start this tutorial we 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 Ω individual elements are denoted by small Ω and are 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 the set of numbers between one and six 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 the specified condition is observed here we are looking to observe five consecutive heads before terminating the experiment the sample space here is countable infinite we the individual elements are represented using a sequence of the right seventies where each entry is done for heads and tails respectively in the final example the experiment consists of measuring the speed of a vehicle with infinite rescission assuming that the vehicle speeds can be negative the sample space is clearly the set of real numbers here.

(Refer Slide Time: 02:37)



We observe that the samples case can be uncountable the next ones that 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.

(Refer Slide Time: 03:19)

Set Theory Notations

$$A \subset B \Leftrightarrow x \in A \Rightarrow x \in B$$

$$A = B \Leftrightarrow A \subset B \text{ and } B \subset A$$

$$A \cup B = \{x : x \in A \text{ or } x \in B\}$$

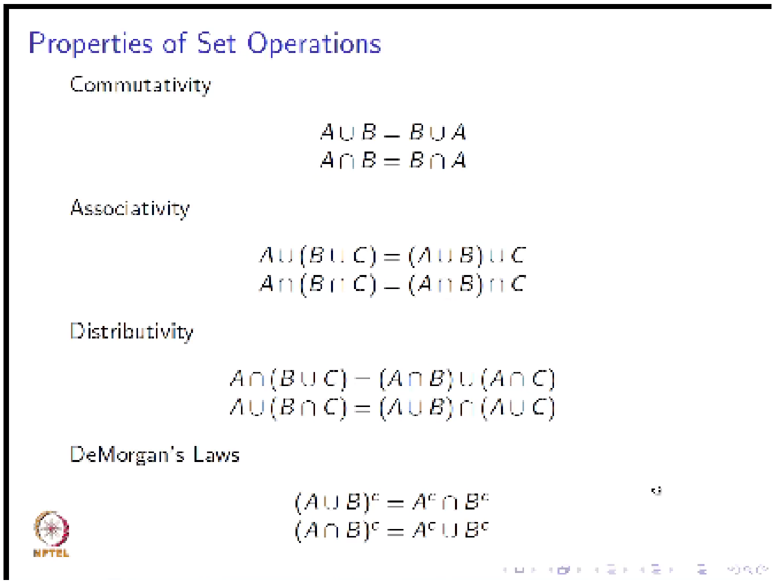
$$A \cap B = \{x : x \in A \text{ and } x \in B\}$$

$$A^c = \{x : x \notin A\}$$

We might be interested in observing whether the outcome was even or odd so for example on a specific role of a tie we have let us say we observe that the outcome was odd in this scenario whether the outcome was actually a one or a three or a five is not as important to us as the fact that it was odd since we are considering sets in terms of sample spaces and events we will quickly go to the basic set theory notations as usual capital letters indicate sets and small letters indicate set elements.

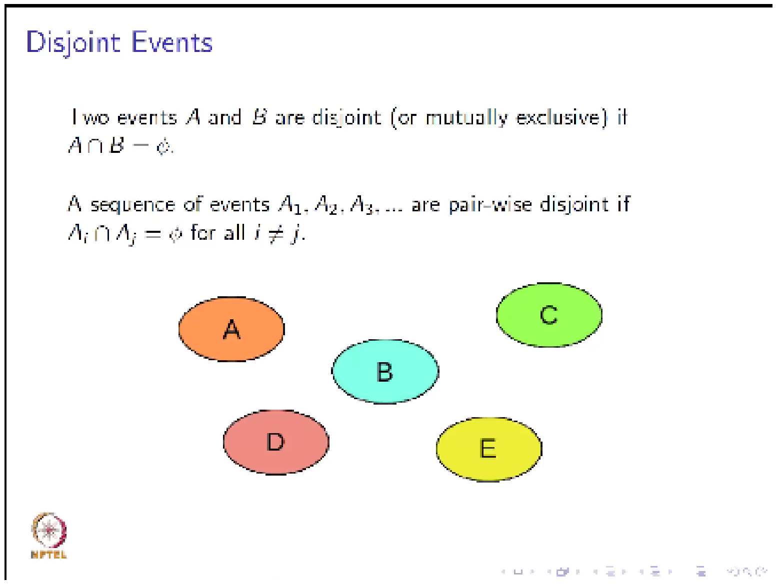
We first look at the subset relation for all legs if X element of a implies X element of B then we say that A is a subset of B or is contained in being two sets A and B are said to be equal if both A subset of B & B subset of a hold the union of two sets A and B gives rise to a new set which contains 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.

(Refer Slide Time: 04:28)



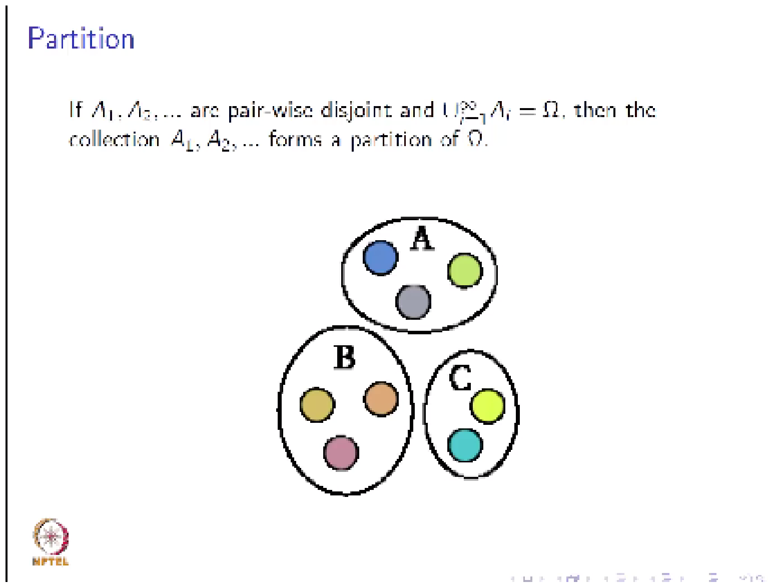
The sample space this slide lists out the different properties of set operations such as commutativity associativity and distributivity which you should all be familiar with it also this out the de Morgan's laws which can be very useful for include de Morgan's laws the complement of the set A union B is equal to a complement intersection B component similarly the complement of the set a intersection B is equals to a complement Union B complement the de Morgan's laws presented here are for two sets.

(Refer Slide Time: 05:05)



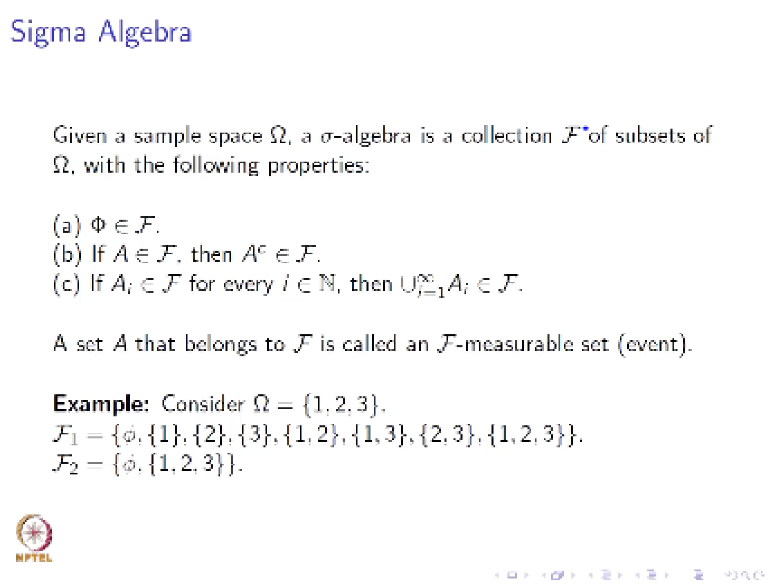
They can easily be extended for more than two sets coming back to events two events A and B are said to be disjoint on disjoint exclusive if the intersection of two sets is empty extending this concept to multiple sets we say that a sequence of events a1 a2 a3 and so on are pair wised is joint if AI intersection AJ is equals to null for all i not equal to j in the example below if each of the letters represents an event then the sequence of events a through E are pair wise disjoint.

(Refer Slide Time: 05:39)



Since the intersection of any pair is empty if events a1 a2 a3 soon are pair wise disjoint and the union of the sequence of events gives rise to the sample space then the collection a 1 a 2and soon we set to form a partition of the sample space Ω this is illustrated in the figure below.

(Refer Slide Time: 06:00)

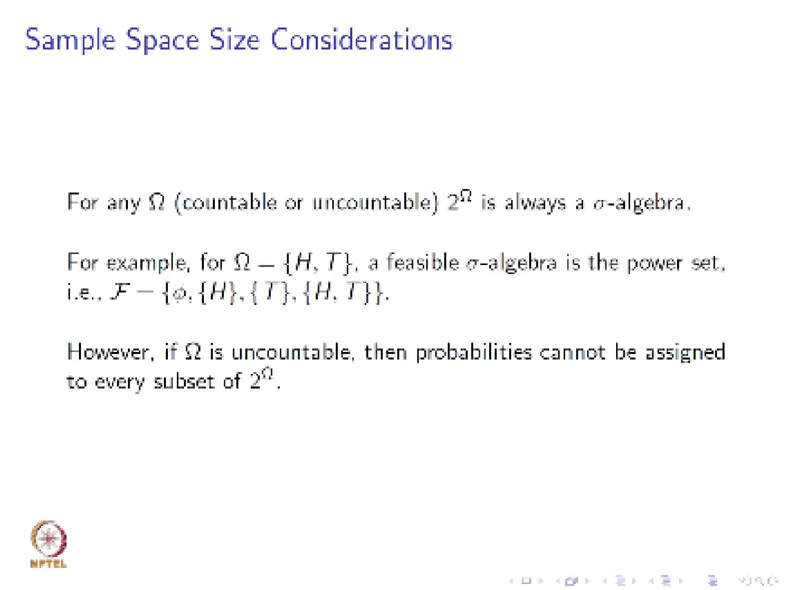


Next we come to the concept of a ∑ algebra given a sample space Ω a ∑ algebra is a collection F of subsets of the sample space with the following properties the null set is an element of F if A is an element of F then a compliment is also an element of F if AI is an element of F for every I long the natural numbers then Union I 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 ∑ algebra then the countable union of these events also belongs in the ∑ algebra let us consider an example consider at Ω equals to 1 2 3 this is our sample space with this sample space we can construct a number of different ∑ algebra here the first thing we are in algebra f1 is essentially the power set of the sample space all possible events are present in the first segment algebra.

However if we look at f2in this case there are only two events the null set or the sample space itself you should verify that for both f1 andf2 all three properties listed above are satisfied now that we know what a ∑ algebra is let us try and understand how this concept is useful first of all for any Ω countable or uncountable the power set is always a ∑ algebra for example for the sample space comprising of two elements H comma T a feasible ∑ algebra is the power set this is not the only feasible ∑ algebra.

(Refer Slide Time: 07:39)

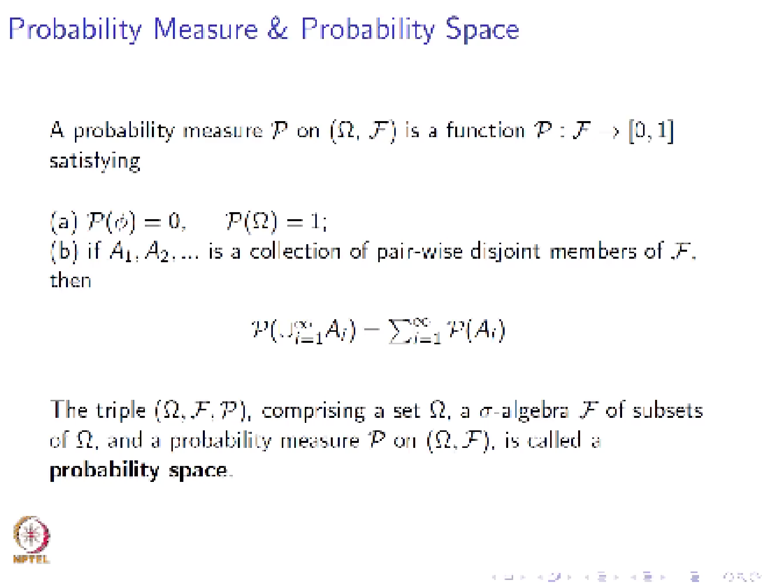


As we have seen in the previous example but always the power set will give you up feasible ∑ algebra however if Ω is uncountable then probabilities cannot be assigned to every subset of the power set this is the crucial point which is why we need the concept of ∑ algebra so just to recap if the sample space is finite or countable then we can kind of ignore the concept of ∑ 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 Ω is uncountable then probabilities cannot be assigned to every subset of true to the Ω this is where the concept of ∑ algebra shows use when we have an experiment in which the sample space is uncomfortable.

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 ∑ algebra and probabilities will then be assigned to the collection of sets in the single algebra.

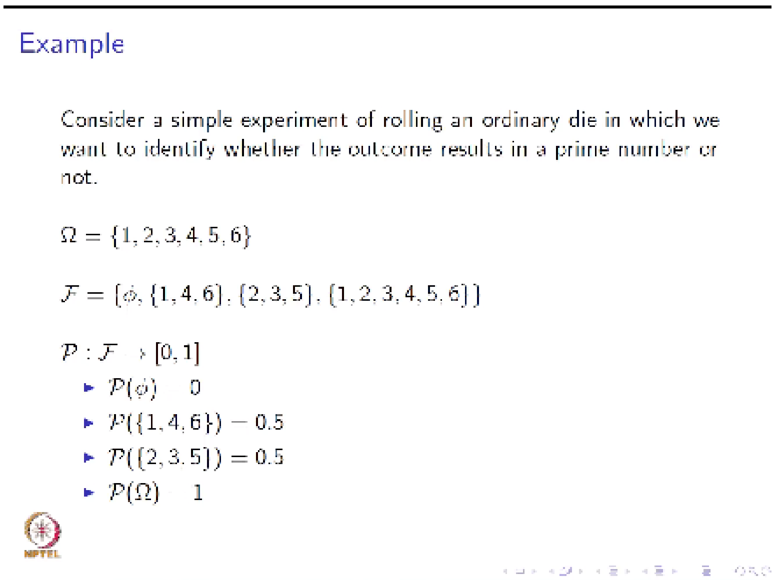
(Refer Slide Time: 09:32)



Next we look at the important concepts of probability measure in probability space a probability measure P on a specific sample space Ω and ∑ algebra F is a function from F to the closed interval 0 comma 1 it satisfies the following properties Robert here the null set equal to zero probability of Ω equals to 1 and if a1 a2 and so on is a collection of pair wise disjoint members of F then probability the union of all such members is equal to the sum of their individual probabilities note that.

This holes because the sequence a1 a2 is pair wise disjoint the triple Ω FP comprising us sample space Ω a ∑ algebra F which are subsets 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 there exists a probability space comprising of the Triple Ω FB even though we may not always explicitly take into consideration.

(Refer Slide Time: 10:47)



This probability space with me solve the problem it should always remain in the back of our heads let us now look at an example where we do consider the probability space involved in the problem consider a simple experiment of rolling and ordinary in which we want to identify whether the outcomes ends in a prime number on the first thing to consider is the sample space is there are only six possible outcomes in our experiment.

The sample space here is consists of the numbers between one to six next we look at the ∑ algebra note that since the sample space is finite you might as we will 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 ∑ 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 is ∑ algebra has to follow please verify that the ∑ algebra listed here does actually satisfy the three properties that we had discussed about the final component is the probability measure the probability measure assigns a value between zero and one that is the probability value to each of the components of the ∑ algebra here the values Lister assumes that the die is fair in the sense that the probability of each phase is equals to 1 by 6 having covered some of the very basics of probability in the next few slides.

(Refer Slide Time: 12:24)

Bonferroni's Inequality

$$P(A \cap B) \geq P(A) - P(B) - 1$$

General form:

$$P(\cap_{i=1}^{n} A_i) \geq \sum_{i=1}^{n} P(A_i) - (n-1)$$

Gives a lower bound on the intersection probability which is useful when this probability is hard to calculate.

Only useful if the probabilities of individual events are sufficiently large.

We look at some rules which allow us to estimate probability values the first thing we look at is known as the bonferroni's inequality according to this inequality probability of A intersection p is greater than equals to probability of a plus travel T 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 it 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 values are very small when this -1 term dominates and the result has been make much sense.

(Refer Slide Time: 13:17)

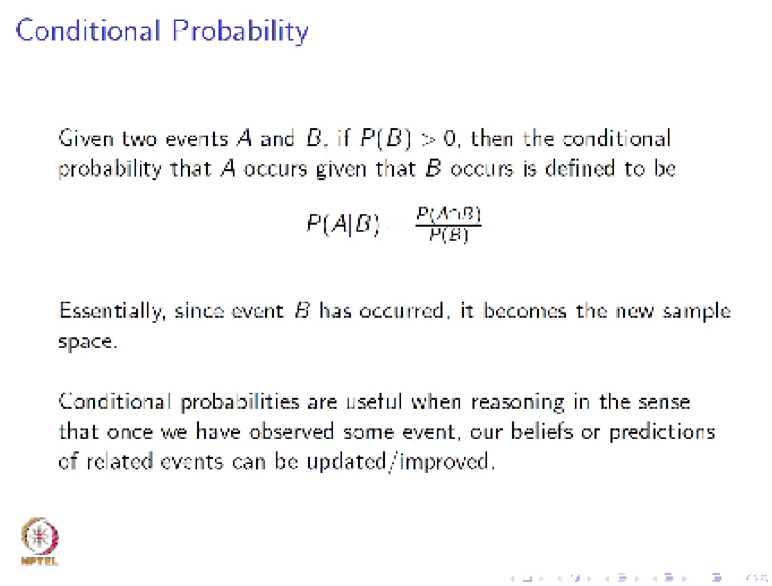
Boole's Inequality

$$P(\cup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} P(A_i), \text{ for **any** sets } A_1, A_2, \ldots$$

Gives a useful upper bound for the probability of the union of events.

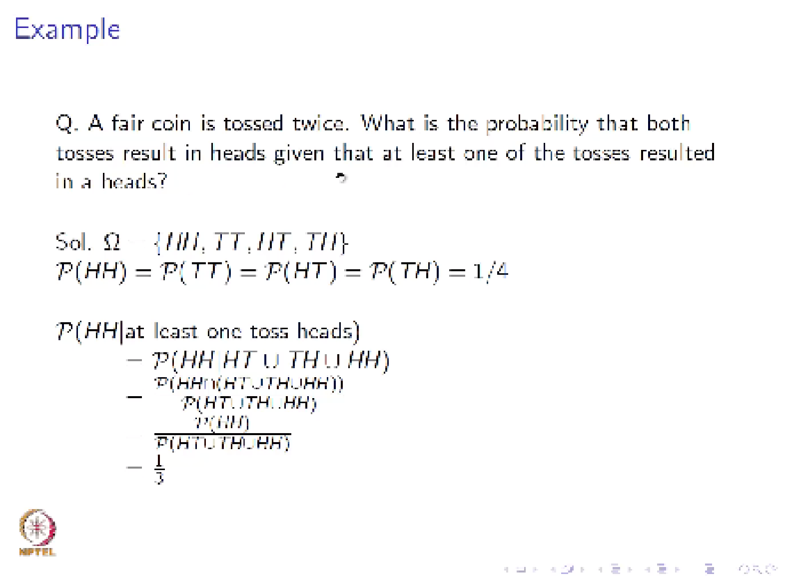
According to the Boole's inequality for any sets a1 a2 and so on the probability of the union of these sets is always less than equal to the sum of their individual probabilities clearly this gives us a useful upper bound for the probability of the Union events notice that this equality will all only hold when these sets are pair wise disjoint.

(Refer Slide Time: 13:43)



Next we look at conditional property 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 equal 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 on or improved.

(Refer Slide Time: 14:13)



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 process resultant heads given that at least one of the causes resulted in the 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 game of tails.

And the other way around since we are assuming that the coin is fair each of the elementary outcomes are the same probability of occurrence equals to1 by 4 now we are interested in the probability that both tosses come up heads even there at least one resulted in the head 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 one by three note that in the denominator each of these events is mutually exclusive thus the probability of the union of these R events is equal to the summation of the individual probabilities as an exercise trans all the same problem with the modification that we observe only the first toss coming up heads that is we want the probability that both tosses her to enhance given that the first toss resulted in a heads does this change the problem.

(Refer Slide Time: 16:07)

Bayes' Rule

We have:

$$\mathcal{P}(A|B) = \frac{\mathcal{P}(A \cap B)}{\mathcal{P}(B)}$$

$$\mathcal{P}(A \cap B) = \mathcal{P}(A|B)\mathcal{P}(B)$$

$$\mathcal{P}(A \cap B) = \mathcal{P}(B|A)\mathcal{P}(A)$$

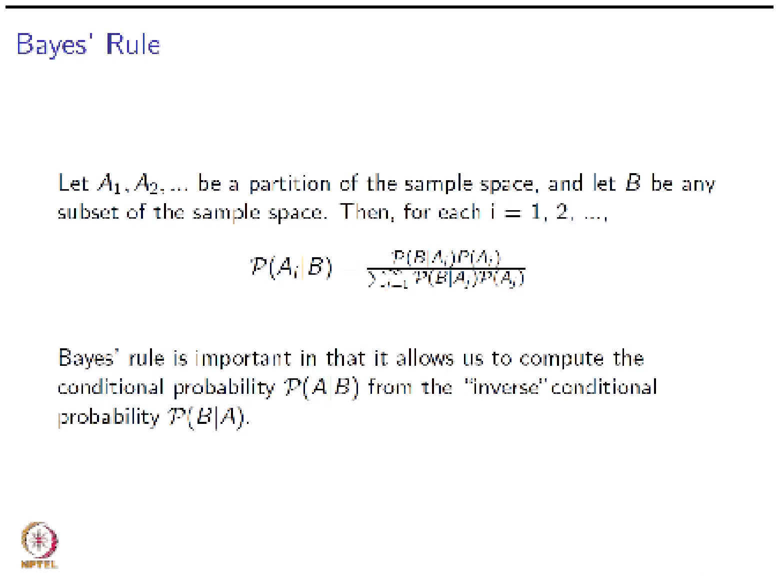
$$\mathcal{P}(A|B)\mathcal{P}(B) = \mathcal{P}(B|A)\mathcal{P}(A)$$

$$\mathcal{P}(A|B) = \frac{\mathcal{P}(B|A)\mathcal{P}(A)}{\mathcal{P}(B)} \text{ (Bayes' Rule)}$$

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 going to probably a intersection B by probability of B rearranging we have probability of A intersection B it was too proud to a given B into probability of B now instead of starting from product with probability of a given B if I started with probability of B given a you would have got probably way intersection B equal to going to be 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 going to fold your be given a into probability of A.

Now taking this probability of B to the right hand side we get probability of A given B is equal to probability of B given A into 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 happen I can use the information of probability of B given a along with the knowledge of all to end probably to get this value.

(Refer Slide Time: 17:29)



As you can see this is a very important formula here we are can present the Bayes' rule in an expanded form where a 1 a2 and so on for partition of the sample space as mentioned Bayes' rule is important in that it allows us to compute the conditional probability per round here a given B from the inverse conditional broad view role to be given A.

(Refer Slide Time: 17:50)

Example

Q. 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?

Sol. Let K be the event that the student knows the question, and C be the event that the student answers the question correctly.
We have $\mathcal{P}(K) = p$, $\mathcal{P}(\neg K) = 1 - p$, $\mathcal{P}(C|K) = 1$, $\mathcal{P}(C|\neg K) = q$

$$\mathcal{P}(K|C) = \frac{\mathcal{P}(C|K)\mathcal{P}(K)}{\mathcal{P}(C)} = \frac{\mathcal{P}(C|K)\mathcal{P}(K)}{\mathcal{P}(K)\mathcal{P}(C|K) + \mathcal{P}(\neg K)\mathcal{P}(C|\neg K)}$$

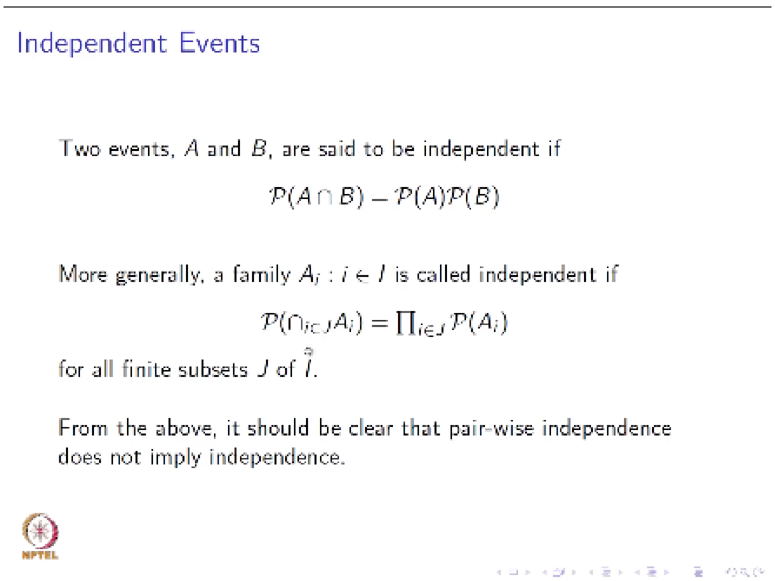
Let us look at a problem in which the Bayes’ rule is applicable to answer a multiple choice question or student neither know the answer or may guess it assume that with probability P the student knows the answer to a question and the probability Q the student guesses the right answer to a question she 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 at C be the event of 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 probability that the student does not know the question is goes 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 of C given K into probability of K by probability of C.

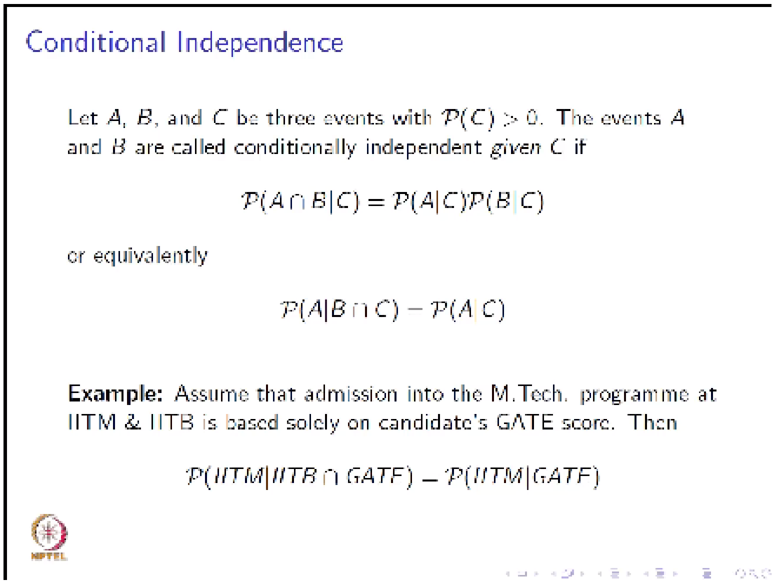
There are probability of answering the question correctly can be expanded in the denominator to consider the two situations probability I want see the question correctly given that the student knows the question and probably run through the question directly under the student does not know question now using the values which mean have gathered from the question we can arrive at the answer of P by P plus Q into 1minus 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 probably of K given C.

(Refer Slide Time: 20:13)



The two events A and B are set 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 probabilities of those events essentially what, 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.

(Refer Slide Time: 21:03)



This condition holds from this should be clear that pair wise independence does not imply independence that is pair wise independence is a weaker condition explaining 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 zero the events a and B are called conditionally independent given see if probability of a intersection B given C equals to probability of a given C into probability of B given see this condition is very similar in form to the previous condition.

For independence of events equivalently the events A and Bare conditionally independent given see if probability of A given B intersection C because a probability of a given see this latter condition is quite an informative what it says is that the probability of A calculated after often 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 conditions independent given C let us consider an example assume that admission into the M.Tech program at IIT Madras in IIT Bombay is based solely on candidates grade school then probability of admission into IIT madras given knowledge of the candidates admission status in IIT Bombay as well as the candidate score is equivalent to the probability calculated simply knowing the candidates grade school just knowing the status of the candidates admission into IIT Bombay does not provide an 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 candidates grade school.

IIT Madras Production

Funded by
Department of Higher Education
Ministry of Human Resource Development
Government of India

www.nptel.ac.in