INDIAN INSTITUTE OF TECHNOLOGY MADRAS

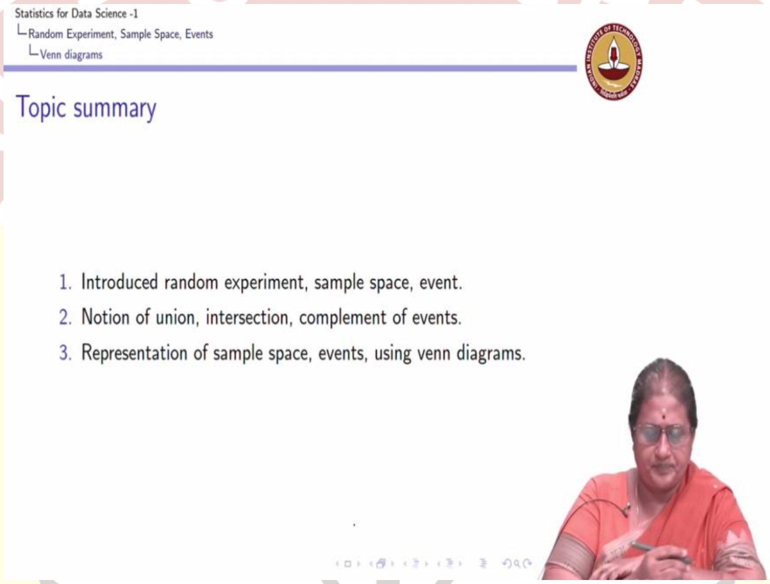
सिद्धिर्भवति कर्मजा

IIT Madras

ONLINE DEGREE

Statistics for Data Science - 1
Professor. Usha Mohan
Department of Management Studies
Indian Institute of Technology, Madras
Lecture No. 6.4
Probability – Properties of Probability

(Refer Slide Time: 00:13)

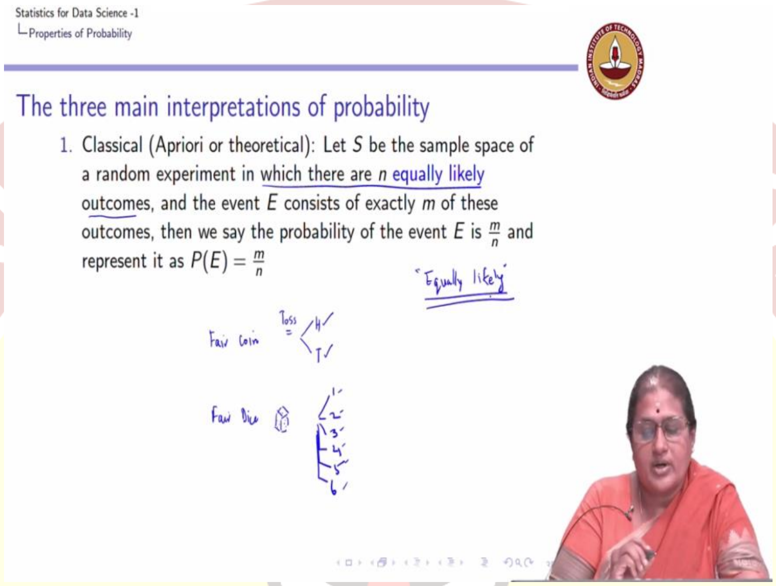


So, in the last lecture, we learned about what is a Random Experiment. Once we define what is a random experiment, we introduced the notion of a Sample Space. A sample space is the set of all basic outcomes of the random experiment. And we also introduced what was a Event. Now, we quickly realize that the sample space and events are sets. So, once I define them as sets, that is when I say define all the set possible collection of all basic outcomes of sets., I introduce the notion of basic set operations on these events.

So, we notion of union intersection and complement of events. A word of caution here is we do not view them as sets. That is a word of caution here is we are not interested only in the mathematical representation of these events as set but we need to understand why we are using them as sets and we need to actually articulate the situation we are facing in as these sets as union of sets or as intersection of sets.

Then finally, we looked at how we represent the sample spaces and events using Venn diagrams. The reason this representation is useful for us is when we are trying to find out the probability of events that, that is our main purpose. And we are able to express a particular event as union or intersection of events and how would we apply these fundamental concepts to actually work out or actually compute probability of events.

(Refer Slide Time: 02:04)



So, in today's lecture, we are going to look at mainly the development of what we call the axiomatic approach to probability or simply put, we would look at certain properties of probability. But even before I set up this properties of probability or the probability model, some books name it, I will just briefly touch upon what have been various interpretations of probability. But I am going to really focus on 3 main interpretations of probability.

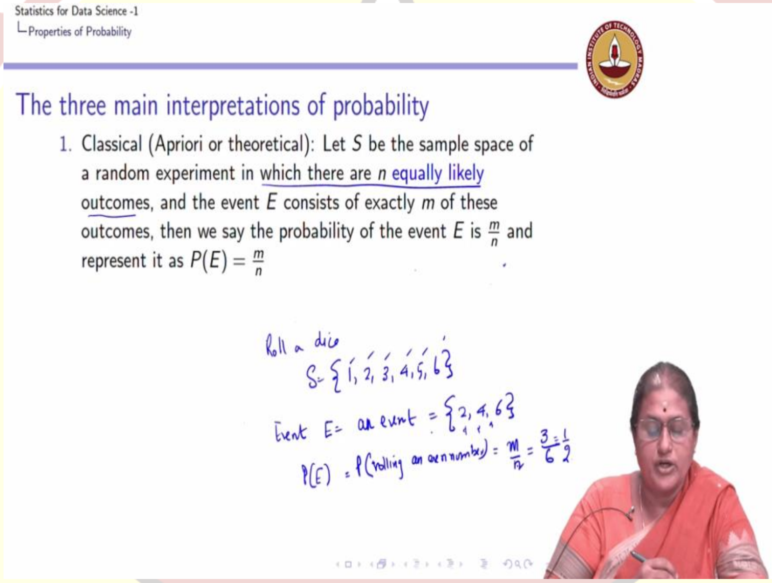
One of the main interpretations and most popular interpretation of probability is what we term as the classical approach. Now, this classical approach is also referred to as a priori or theoretical approach to probability. So, what this approach is, is let S be the sample space of a random experiment we have already introduced what is a random experiment and what is a sample space, but here we are trying to say in which there are n equally likely outcomes.

So, we are talking about what we mean by equally likely outcomes? Now certain examples natural examples of equally likely outcome is when we refer to what we call a fair coin, and we toss a coin, we expect the head and tail, which are the outcomes of the toss of a coin to be equally likely, that is, I expect head to occur as likely or equally likely as a tail and vice a versa.

Similarly, if a roll a fair dice, a 6 edged or 6 sided die, which we use to play. So, this is how a 6 sided die would look.

So, I win a roll a die, I would expect any one of these outcomes which are namely 1, 2, 3, 4, 5, and 6 to appear equally likely. I do not so, their chance of me getting any one of these outcomes are the same. That is what I mean by equally likely outcomes. So, even when I have an experiment when my outcomes are equally likely, and an event consists exactly m of these outcomes.

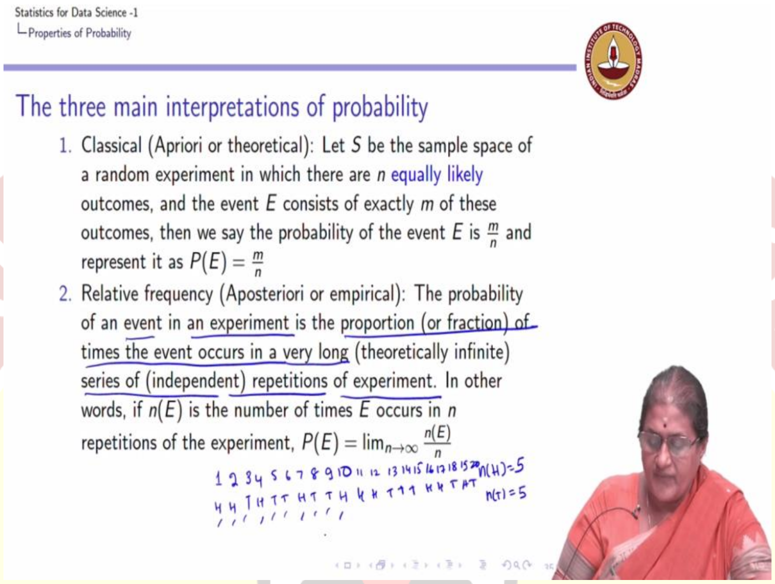
(Refer Slide Time: 04:52)



For example, again, let us roll a die or roll a dice. My sample space here is going to be 1, 2, 3, 4, 5 and 6, and each one of these outcomes are equally likely, let me define my event to be the event of getting an even number. So, the outcomes that satisfy this event are going to be 2, 4, and 6, the number of outcomes in this event are 3. So, the probability of getting a probability of E which is same as probability of getting or rolling an even number by the definition of classical probability equal to $\frac{m}{n}$, $n = 6$, which is the total number of outcomes in my sample space.

The number of outcomes which satisfy this event equal to 3, which is equal to half. So, my classical probability approach to compute the probability of an event is equal to $\frac{1}{2}$. I can extend this notion and I can apply it to various examples, but one thing we need to understand and be aware of, and always remember is this approach assumes that all of the outcomes are equally likely.

(Refer Slide Time: 06:35)

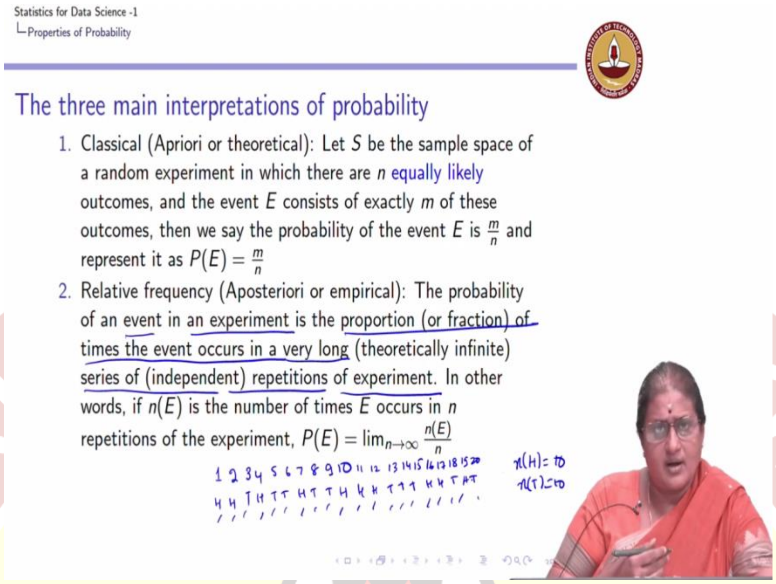


The next approach which is also a very popular approach is termed as a relative frequency or empirical or Aposteriari approach or interpretation to probability. What this approach tells us is the probability of an event in an experiment is the proportion of times the event occurs in a very long series of independent repetitions of the experiment. Let us understand what this statement tells us suppose my experiment is to toss a coin.

So, I am tossing a coin. So, my first experiment is toss a coin. So, I have I am noting down the experiment the observe the outcome. So, suppose I toss a coin once and I find a head I note down a head. I toss a coin twice this is also a head I note down the number of heads is equal to 2 number of tails equal to 0. I toss a coin thrice I note this now, what happens is this becomes 2 and a 1. The fourth trial I get ahead again, again this changes to 3 1. So, suppose I continue to do this experiment and I get the following results.

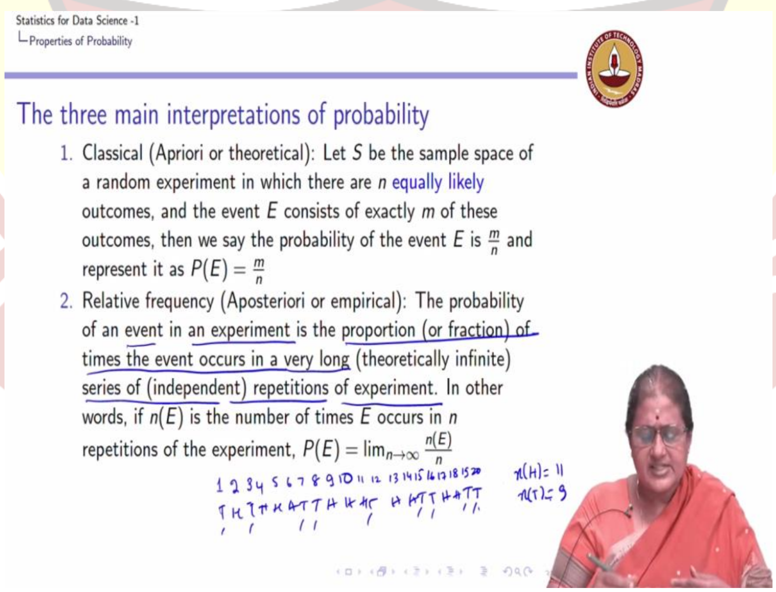
Now, this can easily be replicated by any one of us we can take a coin and keep tossing the coin. So here in 10 tosses I find 1, 2, 3, 4, 5 heads and 1, 2, 3, 4, 5 tails. I repeat this experiment, so I can just go to 11, 12, 13, 14, 15, 16, 17, 18, 19, 20. So, if I keep repeating this experiment and I get a head, head, tail, tail, tail, head, head, tail, head, tail.

(Refer Slide Time: 08:47)



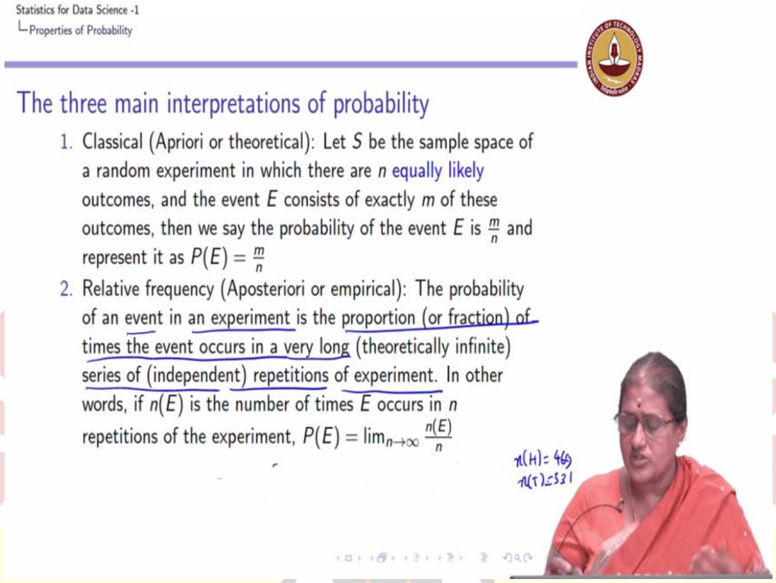
So now the number of heads I have is 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 and tail is also 10, but will this pattern always occur? need not be. For example, I could have a completely different suppose another person is tossing 20 tosses of the coin and they just randomly, this is the sequence they get.

(Refer Slide Time: 09:19)



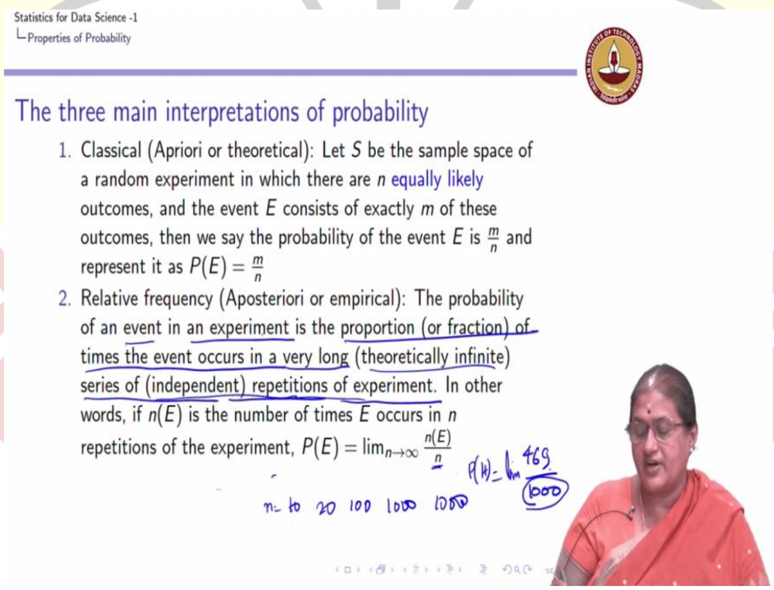
They get a tail, head, tail, tail, head, head, tail, tail, head, head, head, tail, head, head, tail, tail, head, head, tail, tail. So here, if I count the number of tails, I have 1, 2, 3, 4, 5, 6, 7, 8, 9. And number of heads is 11.

(Refer Slide Time: 09:55)



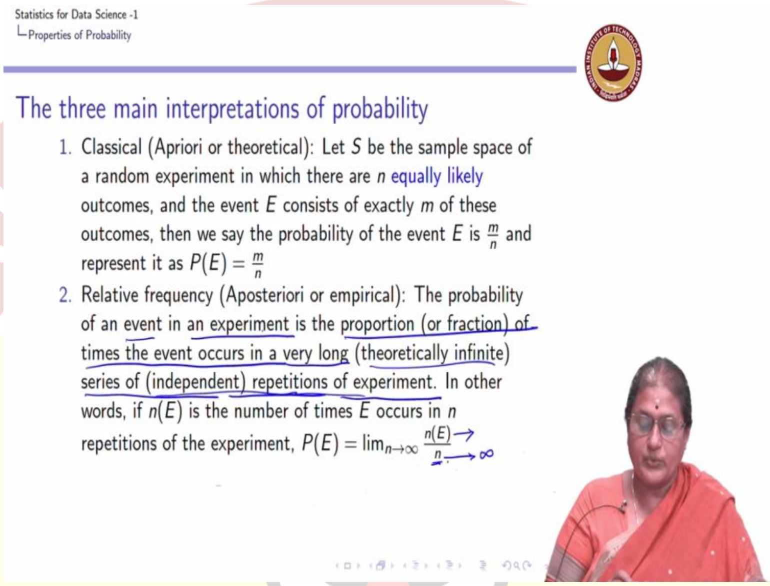
So, imagine that instead of doing this experiment just 10 times or 20 times I am doing it a 1000 times, then it could be very well possible that I am getting 469 heads. I am doing it a 1000 times I am getting 531 tails, I repeat it 10,000 times. So, we are counting the number of times.

(Refer Slide Time: 10:12)



So, what we mean is, we are repeating the setup independent repetitions. So, the first time I repeated 10 times, the next time I did it 20 times and I made it 100 times and I made it 1000 times, then 10,000 times. So, you can see that, what this refers is to this n, if I am repeating this experiment infinitely theoretically infinite number of times. And I am counting the number of outcomes that satisfy a particular event. So, that even could be coming either head or tail. So, if I have 469 in 1000 turns, so my probability of getting a head is going to be the limit as this tends to infinity.

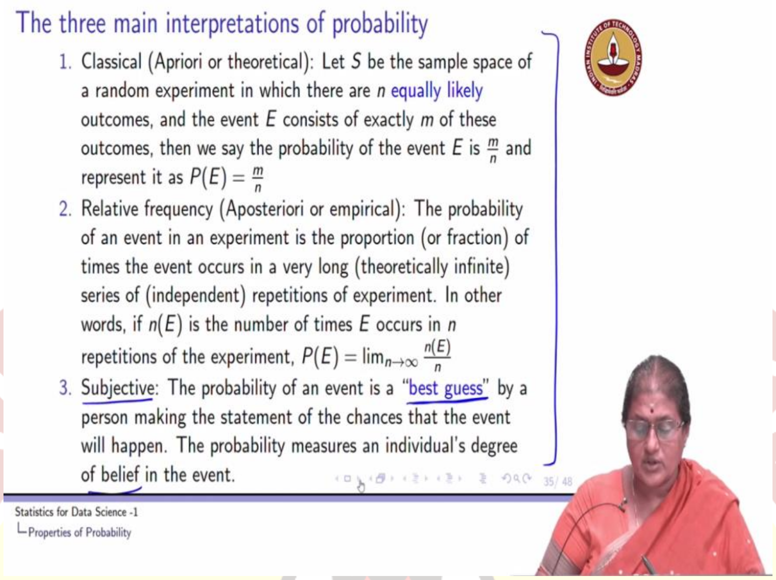
(Refer Slide Time: 11:12)



I keep repeating the same experiment and we can see that, as this number increases or tends to infinity, this tends to a particular value. So, this is so what is so this is what is popularly referred to as the relative frequency frequentist approach or the empirical approach to compute probabilities. The disadvantage of this approach to computing probabilities is we are expecting that the experiment can be repeated for a very long period, but not really all experiments can be repeated in this fashion.

However, this idea is, if I repeat a particular experiment independently, and n is the number of times E occurs in n repetitions, then probability of E is limit n tending to infinity. Number of times this particular event occurs to the total number of repetitions. So, this is the next approach, we started with a classical approach and what is the relative frequency approach.

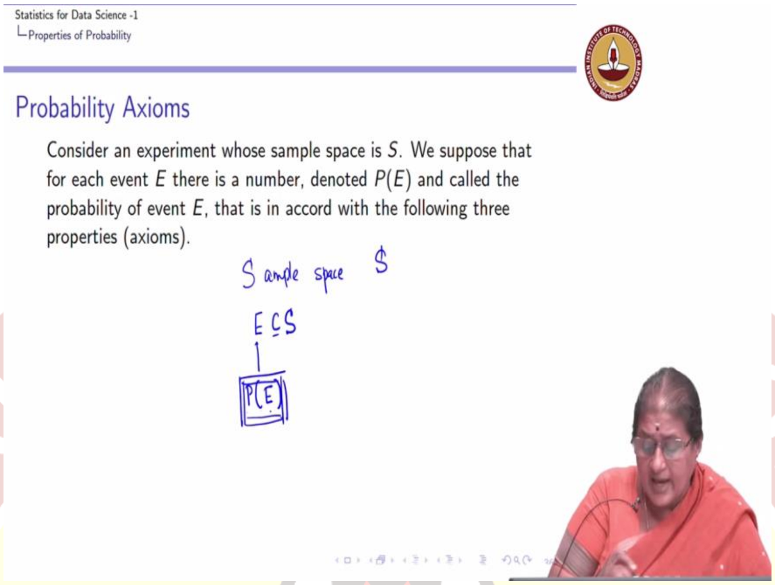
(Refer Slide Time: 12:23)



The third approach is what we refer to as this subjective approach. Now, this approach is basically it assigns a best guess, by a person making a statement. Very often we see people making statements like this a 30% chance that it would rain tomorrow, or the 75% chance that a person would win a particular election even though that person is standing for an election for the first time.

So, this statements of this kind actually measures an individual's degree of belief in a particular event. So this type of interpretation or this interpretation of probability is referred to as the subject to interpretation of probability.

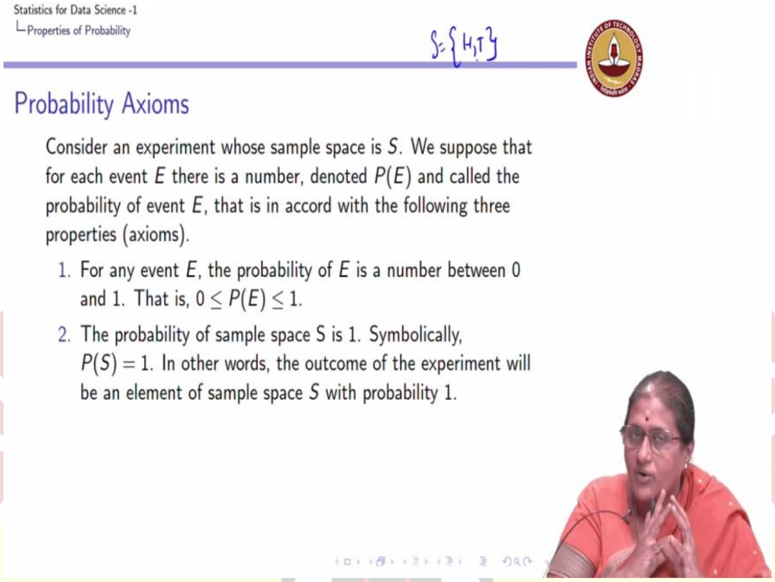
(Refer Slide Time: 13:30)



Now, these interpretations have been popular for a long time, but then afterwards, there has been a unified framework to interpret this entire probability theory and that is what is popularly referred to as the Axiomatic approach to probability. Going deeper into the mathematical framework is beyond the scope of this course. But however, we are trying to give you the foundations that are essential for you to help you understand this statistical inference course. So, what is the Axiomatic or what are what we refer to popularly as the probability axioms?

So, let us go back to our random experiment, we learned what was a random experiment. Let the sample space of a random experiment be S. So, sample space is given what is the sample space? The sample spaces S, and what we know is an event is a subset of a sample space. This is what we have already learned. Now associated with every event, I am going to associate a number and I denote that number with this notation, P(E) which is spelled out as probability of a event E. And what, what do what are the properties or axioms that this P(E) have?

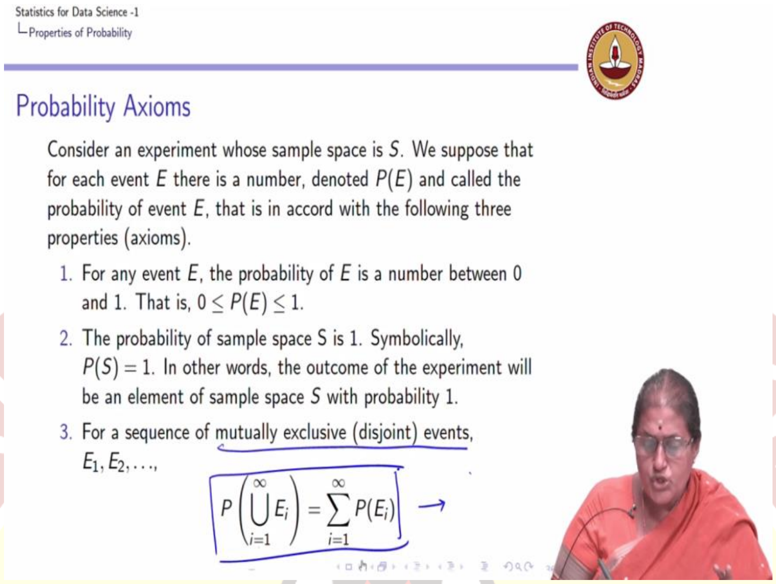
(Refer Slide Time: 14:56)



The first axiom is for any event. The probability of the event E is a number between 0 and 1, that is $0 \leq P(E) \leq 1$. This is what I refer to as my first axiom of probability. Now, the next axiom of probability is the probability of the sample space is equal to 1, I symbolically write that as $P(S)=1$. In other words, what this means or what this signifies is, the outcome of any random experiment will be an element of this sample space.

So, if I toss a coin, what are the outcomes head and tail. So, both these outcomes are elements of the sample space, which I am going to tell, so I cannot have an outcome of an experiment, which does not belong to my sample space, the sample space is the collection of all outcomes. So, that is why probability of the sample space is equal to 1.

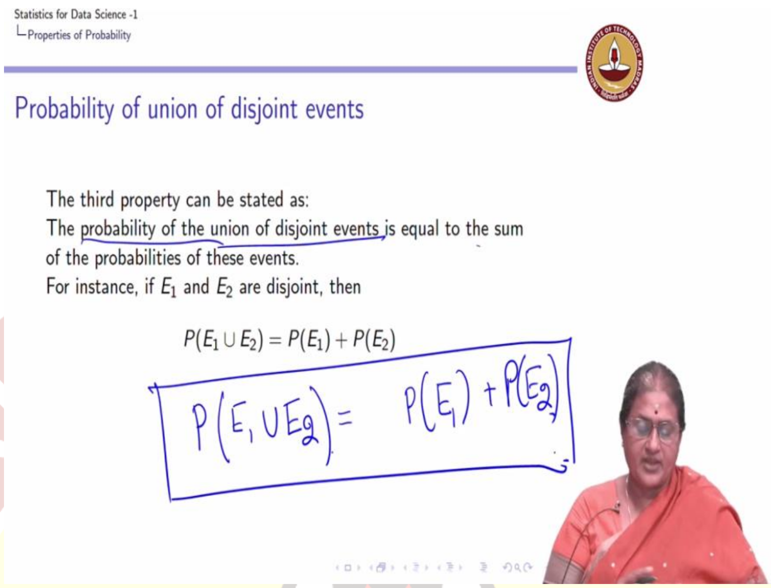
(Refer Slide Time: 16:22)



The third axiom states that if I have a sequence of mutually exclusive events, we have already defined what is an event. And we also know that I can construct events given Simple Events with simple events, I am just talking about Singleton sets, I can construct a lot of events from simple events. Then what the third axiom says is the probability of the union. Well, I am taking union 1 equal to infinity, I am assuming that there are many events countably finite events.

Now, these notion of countability and finite events and everything is something you would be learning in your math courses in due course. I will just tell what is essential with regard to this course. So, what this axiom says is the probability of the union is the sum of the probabilities when I look at a sequence of mutually exclusive events. This is crucial for us to understand that what this tells us

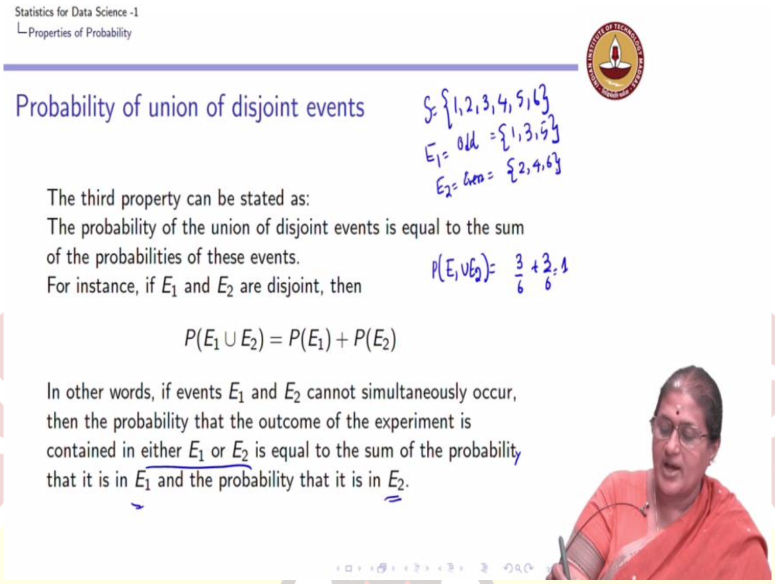
(Refer Slide Time: 17:37)

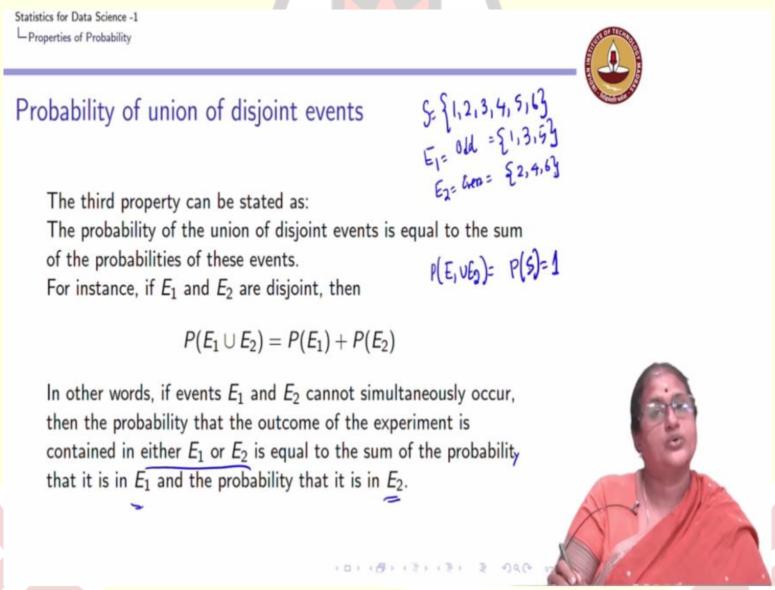


If I have mutually disjoint events, so I for the purpose of this course, I am just going to look at the finite union even though this applies for larger or countable union, we are not going into the mathematical foundations of understanding that at this point of time, but for this course, I can state the third property as the probability of union of disjoint sets is equal to the sum of probabilities.

Let us look at only finite union. What does that mean? Suppose I have E_1 and E_2 as disjoint sets, then I can represent the union of these sets by what does this mean suppose I have E_1 and E_2 which are my disjoint set, I know the union is represented by $E_1 \cup E_2$. So, the union of these disjoint sets is $E_1 \cup E_2$. The probability of the union is $P(E_1 \cup E_2)$ it is equal to the sum of probabilities. So, this is equal to $P(E_1 + E_2)$. This is what my statement or third axiom tells me.

(Refer Slide Time: 19:00)





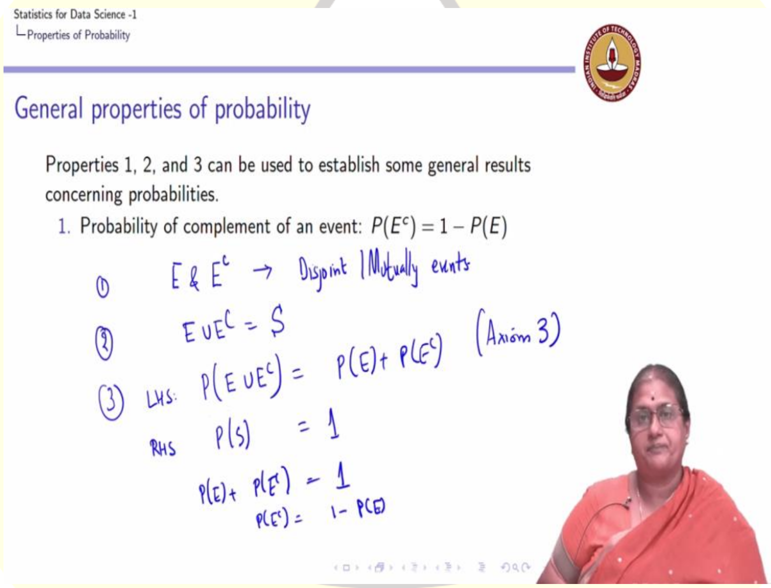
So, what is the understanding and why is this property important? It states that if E_1 and E_2 are 2 events that are disjoint. In other words, they cannot occur simultaneously, then the probability of the outcome that the experiment is contained in either E_1 or E_2 is equal to the sum of probabilities that is an E_1, and the probability that it is in E_2. So, this is what is the probability of union of disjoint events.

For example, let us again take the case of rolling a die. I know this is my sample space. Suppose I define $E_1$to be the event of me getting a odd number, then I know the outcomes are going to be 1, 3 and 5. If $E_2$is the event of getting an even number, the outcome are 2, 4 and 6. I know $E_1$and

E_2 are disjoint that is I cannot have an event where I can get an odd and even number in the same throw of a die.

But I know that if I define an event that the outcome is either an even or an odd number, then the probability of getting either an even or an odd number is equal to the $P(E_1) + P(E_2)$, which is$\frac{3}{6} + \frac{3}{6}$, which is 1, which is nothing but the probability of your sample space which is equal to 1. So, this is what we refer to as the probability axioms. And these 3 axioms help us from these 3 axioms, we get hold of many other properties of probability. And that is what we are going to see now.

(Refer Slide Time: 21:10)

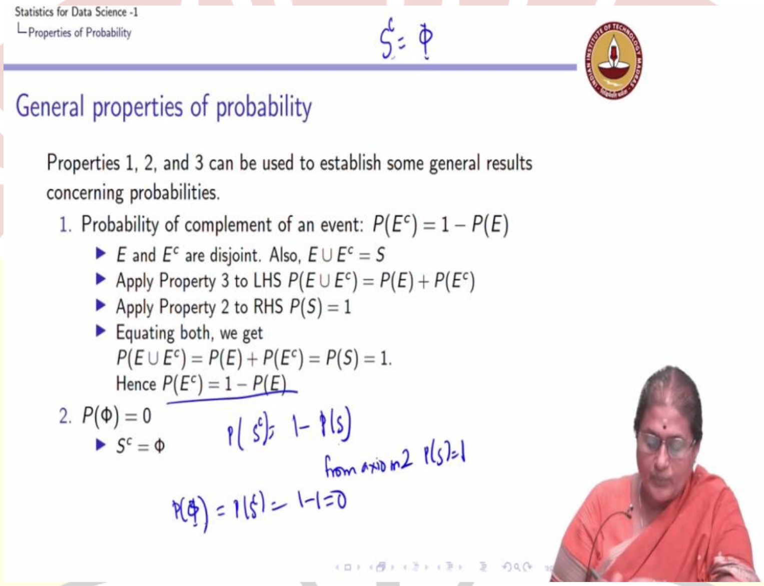


The first problem property of probability is, remember when we talked about sets, we talked about union intersection and complement of sets? So can we work out what is the probability of a complement of a set? Now when we look at a compliment, so if E is a event, E^C a sets event, we know the first thing we need to understand is E and E^C are disjoint. Why is this true? By the way we define E^C is the set of outcomes that are not in E. Hence, E and E^C will not have anything in common. These 2 sets or these 2 events are disjoint, or mutually exclusive events.

Now, another thing which we notice is $E \cup E^C$ is your entire sample space. So given an event, E, I know first E and E^C are disjoint. And the second thing I observe as $E \cup E^C$ is my entire sample space. Now, if E and E^C are disjoint, then I know that from my third axiom, I know the $P(E \cup$

$E^C) = P(E) + P(E^C)$. This is through axiom 3. By axiom 3 states that when I have disjoint, or mutually exclusive set, the probability of the union is the sum of the probabilities.

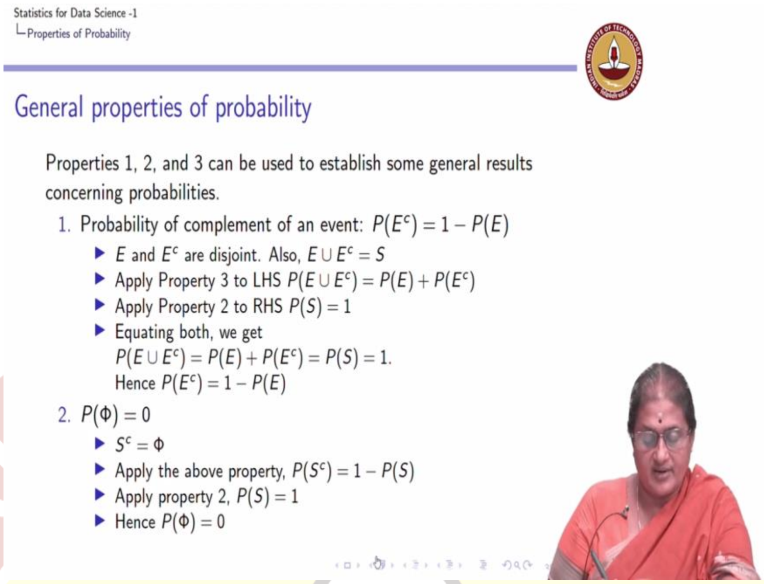
Now, axiom 2 states probability of the sample spaces is 1. So, my left hand side says the $P(E \cup E^C) = P(E) + P(E^C)$. My right hand side states $P(S) = 1$. I equate these 2 to get $P(E) + P(E^C)$ =1, which gives me $P(E^C) = 1 - P(E)$ (Refer Slide Time: 23:27)



So, that is what we have. And we can, so hence I have probability of E^C, which is equal to$1 - P(E)$. Now the next important property is we also define what was an Null event. An Null event is an event which has no outcomes. So, I can define what is a null event, this is an event which has no outcomes. So, when I am defining an null event, then I am interested in knowing what is the probability of the null event? That is the next thing which we are interested in knowing.

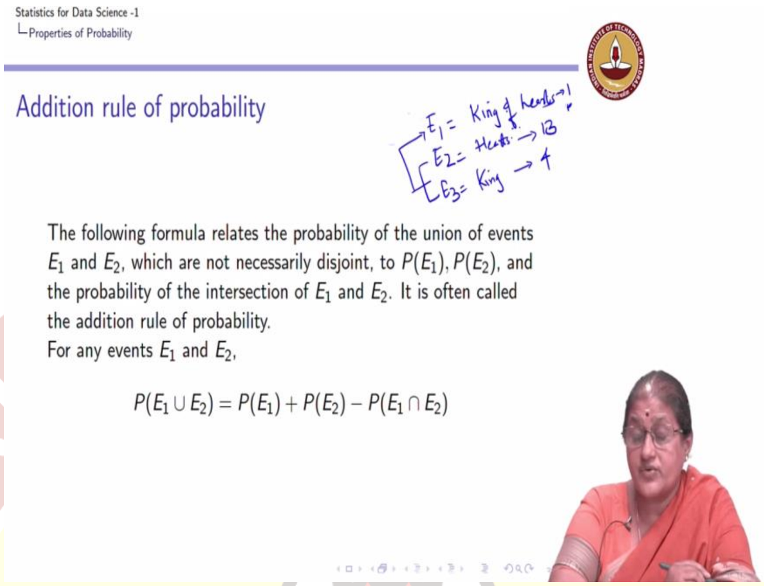
So what, what is the probability of an Null event? Now what is the complement? I know the complement of my sample spaces in our living sample space is the collection of all the outcomes null event does not have any outcome. Hence, the complement of the sample space is my null event. And I know from my earlier property, probability of a sample space, the complement equal to 1 minus probability of that event, sample space is also the entire event it is a set. And my axiom 2 from axiom 2, I know probability of my sample space is 1, which gives me $P(S^C) = 1 - 1$ which is a 0, but I recognize S^C is my null set, hence probability of my null set is 0.

(Refer Slide Time: 25:05)



So, these 2 properties have been derived from the axioms or the 3 prime properties which we stated earlier.

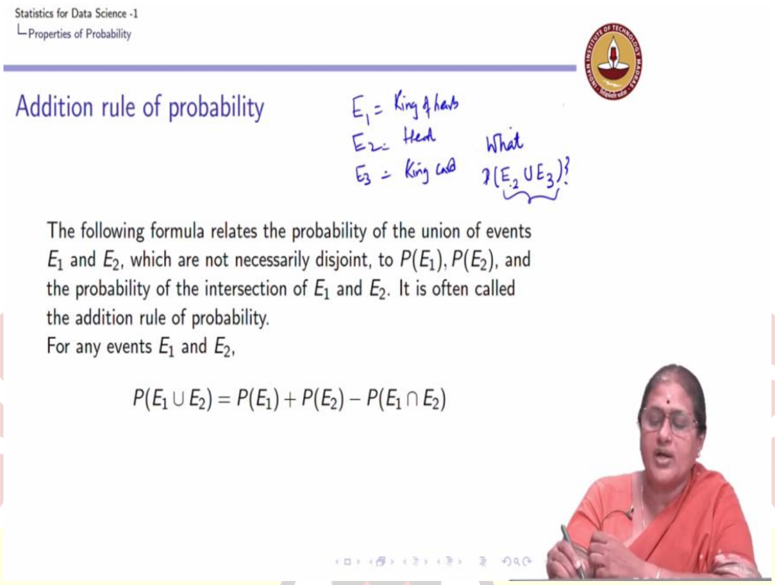
(Refer Slide Time: 25:15)



Now, the next thing is what we refer to as an addition rule in probability. The third axiom says very clearly it needs both E_1 and E_2 , if I am talking about more events it one qualifier in the third axiom of probability was these events were mutually exclusive, or disjoint. That was the important qualifier we had here. But sometimes we are having events. For example, if you recall our card example, one of the events I defined was king of hearts.

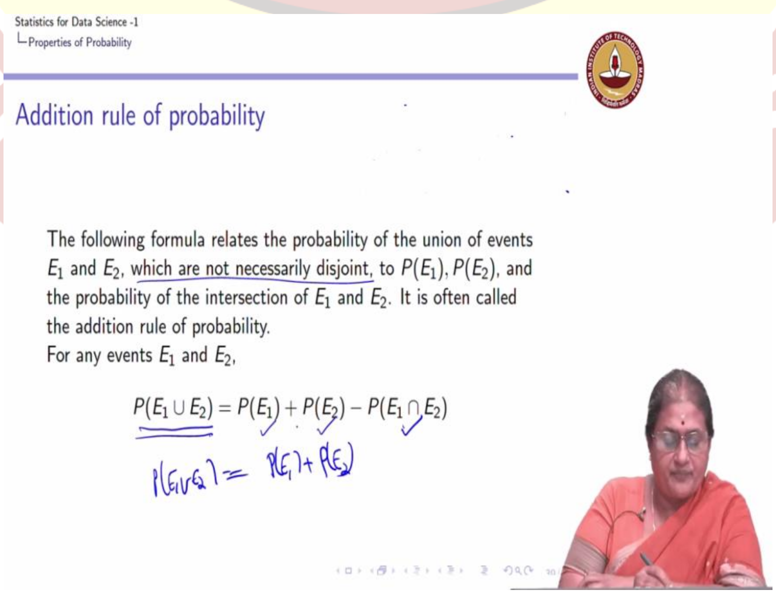
The second event we defined was hearts. And then after that, you can see that these 2 events, this had 13 cards, this had only 1 card, the third event we defined was a king. So, this had 4 cards, $E_2$and E_3had 1 card which is in common, and that was a king of hearts. So, you can see that many a time we have events which are not disjoint or which are not mutually exclusive. So, the natural question to ask is, how do I compute the probability of the union? So, why would I be interested in this?

(Refer Slide Time: 26:52)



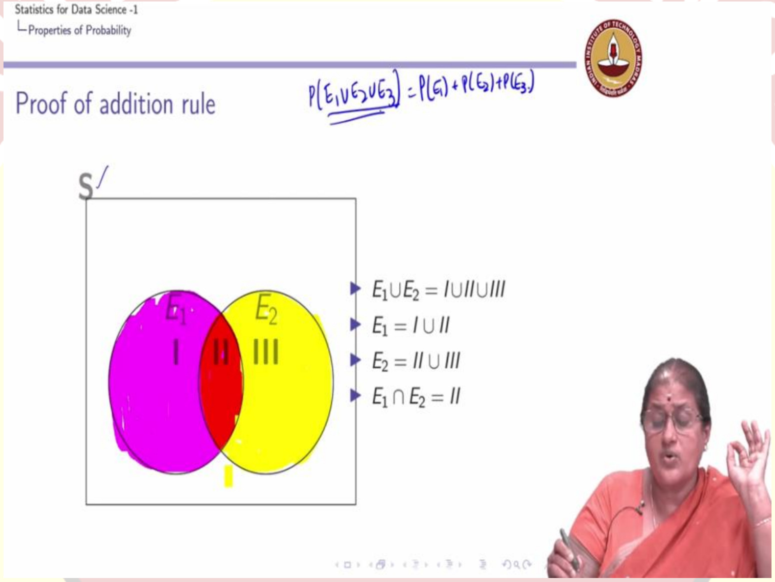
So, if you go back my E_1 was king of hearts. My E_2 was a heart. My E_3 was a king card. So, if my question is what is the probability that I have the draw card? What is the probability that I either have a heart or a king? This is a probability I am interested in finding out but now, E_2 and E_3 are not mutually exclusive. I know that they are not disjoint. How do we find the probability of such an event? This event is a legitimate event. This event is what is the chance that the card I have picked is either a heart or king. How do we compute these probabilities using the properties which we have just studied? That is a question.

(Refer Slide Time: 27:55)



So the addition rule helps us to find out these probabilities, what does the addition rule state. If I have 2 events, which are not necessarily disjoint, then the probability of the union of these 2 events is $P(E_1) + P(E_2) - P(E_1 \cap E_2)$. So, if they were disjoint, this would have been 0. And this reduces to $P(E_1 \cup E_2) = P(E_1) + P(E_2)$, which is what we started with and this was the third axiom. So, the addition rule is a more generalized rule for finding out probability of union of events. So, let us try and prove this rule using our basic properties of probability.

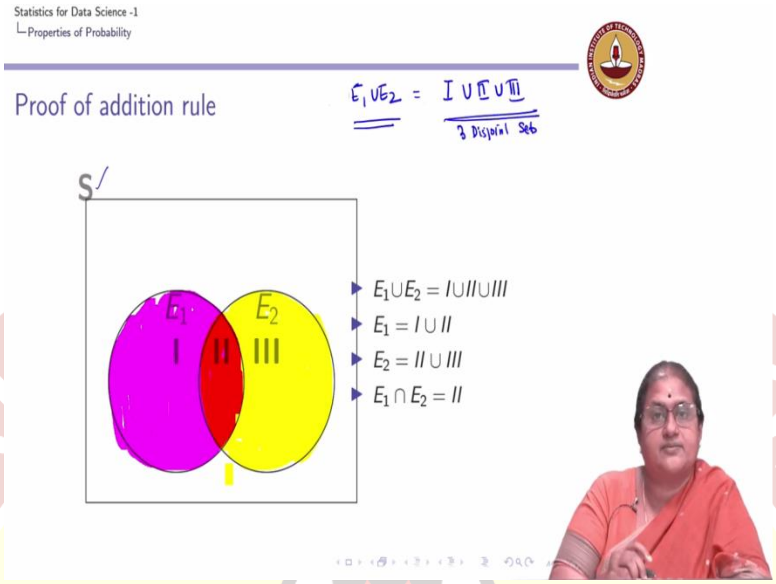
(Refer Slide Time: 28:57)



We represent this So, I have an experiment, I have a sample space, we saw that we can represent the events and the sample space using a Venn diagram. So, this block so, if you if I can, so the red circle represents my so, if you look at it the purple circle, I will shade the area. The purple circle is my E_1. The yellow is what E_2. So, the yellow let me shade it again. So the yellow shaded region is my E_2 the purple shaded region is my E_1 and the orange shaded region is by $E_1 \cap E_2$.

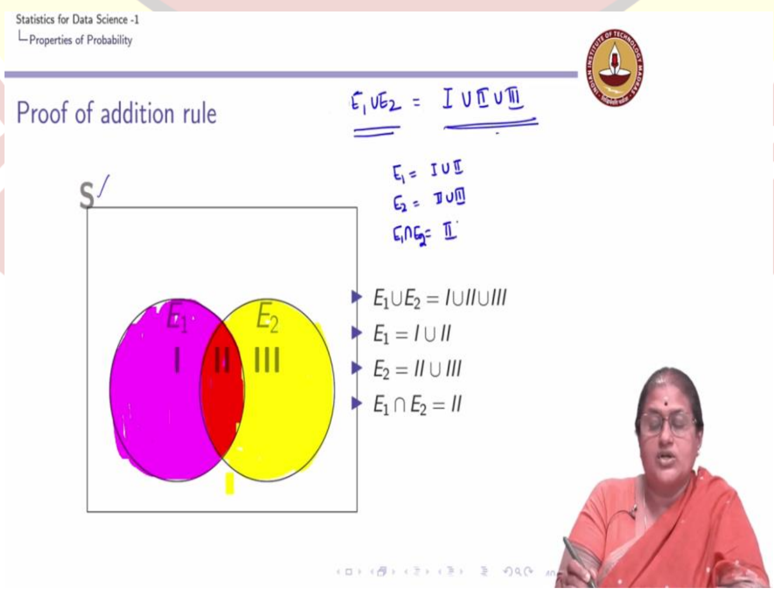
So we see that there, event E_1 and E_2 are not disjoint and they are not mutually exclusive. But I have a theorem or I have I but, but I have an axiom, which says $P(E_1 \cup E_2 \cup E_3) = P(E_1) + P(E_2) + P(E_3)$, provided they are mutually exclusive or disjoint.

(Refer Slide Time: 30:59)



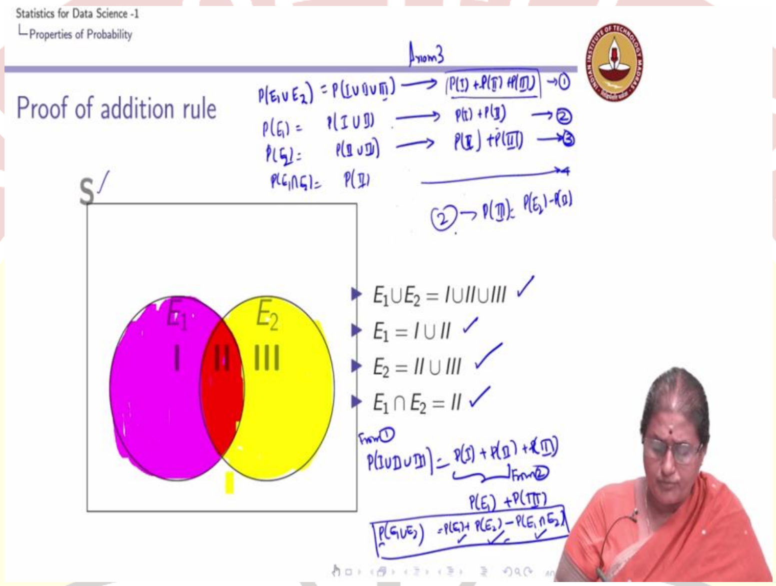
So, what I am planning to do, we recognize that this $E_1 \cup E_2$ is this entire shaded region of the purple, orange and yellow shaded region, I write this entire shaded region as the following. I write it as 1 ∪ 2 ∪3 were 1, 1 and 2. And 3, as you can see from the Venn diagram, are disjoint, I do not have anything in common between 1 and 2, I do not think has anything in common between 2 and 3, but I can see that E union $E_1 \cup E_2$ can be represented as 1 ∪ 2 ∪3. In other words, I am representing the union of $E_1 \cup E_2$ as a union of 3 disjoint regions or disjoint sets.

(Refer Slide Time: 32:01)



Now, similarly, if I look at the same Venn diagram, I see that E_1 can be represented as 1 ∪ 2, E_2 can be represented as 2 ∪ 3, and I also see that the orange region which is even intersection E_2 is 2. So, these are the things which we can see from the given diagram. And that is what we have typed.

(Refer Slide Time: 32:31)



We have here that is $E_1 \cup E_2$ need to is 1 ∪ 2 ∪ 3. E_1 is 1 ∪ 2, and E_2 is what 2 ∪ 3, $E_1 \cap E_2$ is 2. Now what are we interested in finding out we are interested in finding out $P(E_1 \cup E_2)$, which is the same as P(1 ∪ 2 ∪ 3). And I can apply axiom 3 to this. And I know, this P(1)+P(2)+P(3). Let me label this as my first equation.

Similarly, $P(E_1)$ is $P(1 \cup 2)$ again through an application of axiom 3, I get this as $P(1) + P(2)$, $P(E_2)$ is $P(2 \cup 3)$, which is P(2)+P(3). Let me label this as 3 and this is 2. And I have $P(E_1 \cap E_2)$.is P(2) I am labeling this as 4. In other words, I have just expressed $P(E_1 \cup E_2)$, $P(E_1), P(E_2)$ and $P(E_1 \cap E_2)$ in terms of probability of 1, 2, and 3 by applying the axiom 3.

Now, let us actually express this probability in terms of E_1, E_2, and $(E_1 \cap E_2)$. Probability of 1 plus probability, so if I look at P(1 ∪ 2 ∪ 3), I have this is P(1)+P(2)+P(3). This comes from 1. Now, 2 from 2 I have this is equal to $P(E_1) + P(3)$ but $P(3)$is that invert $P(E_2) - P(2)$, this comes from 2 so I can write this as $P(E_1) + P(E_2) - P(2)$ but I know from 4, $P(2)$ is equal to $P(E_1 \cap E_2)$. And hence, I can replace this with $P(E_1 \cap E_2)$.

And this actually tells me that $P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$.. This is what is referred to as the addition rule of probability. That is, I repeat, if I have 2 events, E_1 and E_2, not necessarily disjoint, then the probability of the union is $P(E_1) + P(E_2) - P(E_1 \cap E_2)$.. So the key is to write the union as a union of 3 disjoint sets and apply the axioms of probability to show that the probability of union is the $P(E_1) + P(E_2) - P(E_1 \cap E_2)$. which we can actually prove from whatever we have seen and then apply the addition rule.