Statistics For Data Science I

WEEK 5 - WEEK 8

Prof. Usha Mohan
Department of Management Studies
Indian Institute of Technology Madras

IIT Madras
ONLINE DEGREE

INDEX

Week 8

Statistics for Data Science - 1
Professor Usha Mohan
Department of Management Studies
Indian Institute of Technology, Madras
Permutations and Combinations - Basic Principles of counting

Now, we start the module 2 of this course. So, you recall that in the module 1 we spent our time understanding about descriptive statistics, in particular we looked at how we summarize a categorical variable and a numerical variable and we also spend time to understand how we describe association between variables when both the variables are categorical, when both of them are numerical and one of them is categorical and numerical.

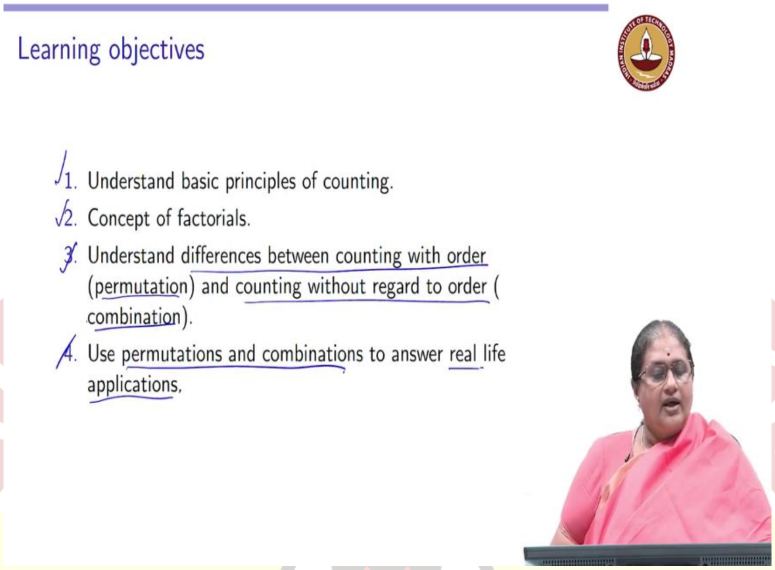
So, till this time we spent our energy and understanding to look at how we describe data that is given to us, we did not use the data to look into the future, there was everything was clear about the data, it was certain about the data, there was no element of uncertainty that was there in our data that we looked so far.

But however, we also know that we want to use the statistics as a tool to infer about the unknown. In other words, we are faced with uncertainty most of the times and whenever we are faced with uncertainty we want to see whether we can have a framework or a tool that would help us predict about the uncertain or understand the uncertainty with a certain degree of confidence.

Now, probability is a very powerful tool to help us understand this uncertainty. So, in this module we are going to understand about probability, we are going to introduce the notion of probability and we are going to look at how we can formalize certain intuitions which we already have in an uncertain world in the domain or in the framework of probability.

But before we go and understand probability what we are going to do is to try and understand about certain basic counting principles. All of us would have studied in high school about permutations, combinations and fundamental principles of counting, we just revisit that basic fundamental principle of counting along with a revision of concepts from permutation and combination this week. So, let us get started.

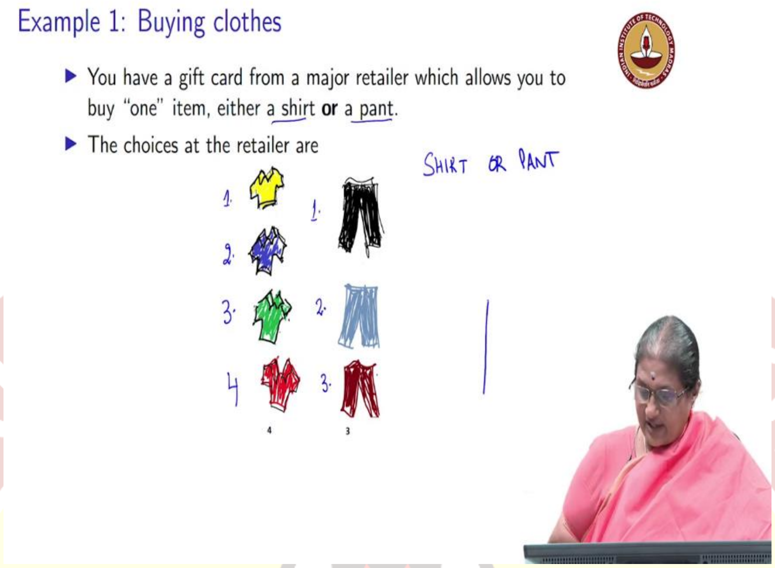
(Refer Slide Time: 3:18)



So, what are the learning objectives here? So, you can see that the learning objectives I am going to have here are the following. So, the basic objectives, the learning objectives of this week are we understand what are the basic principles of counting, then we introduce the concept of factorials and what are the simplified expressions that we can have using the concept of factorials.

Then this is the key thing will understand what is the counting with order and counting without order. In other words what is the difference between permutation and combination. And finally we will use the concepts of permutation and combinations to answer a few applications that arise in real life. So, this is the agenda for the week.

(Refer Slide Time: 4:07)

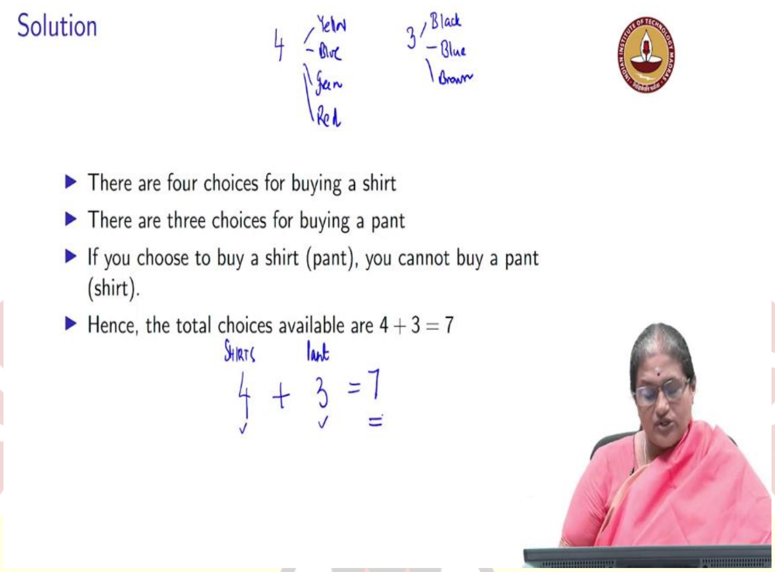


Now, let us start going, so the first thing is we start with a basic example of counting. Now, what you can see on the screen is suppose I have a gift card from a major retailer which specifies the following, you can buy 1 item and that item can either be a shirt or a pant, so the specification is you can buy a shirt or a pant, we assume now that the shirts and the pants are priced the same, we are not assuming it, so you have a gift card, a gift card for certain denomination and what is specified in the card is you can buy a shirt or a pant.

Now, when you go to the shop, you notice the following, you see that you have a choice the first choice which is a yellow shirt or a blue shirt or a green shirt or a red shirt but when it comes to pant I have 3 choices I can either buy a black pant or a blue pant or a brown pant. Now, if I exercise my gift card or I use my gift card to buy a yellow shirt I have exhausted my option. Similarly, if I use my gift card to buy a blue pant I have exhausted my option.

So, the question here we are asking is how many different ways can I use my card? So the answer is I can either buy a yellow shirt or a blue shirt or a green shirt or a red shirt. So, I have basically the number of options available to me, so the question is how many different ways can I use my card?

(Refer Slide Time: 6:15)

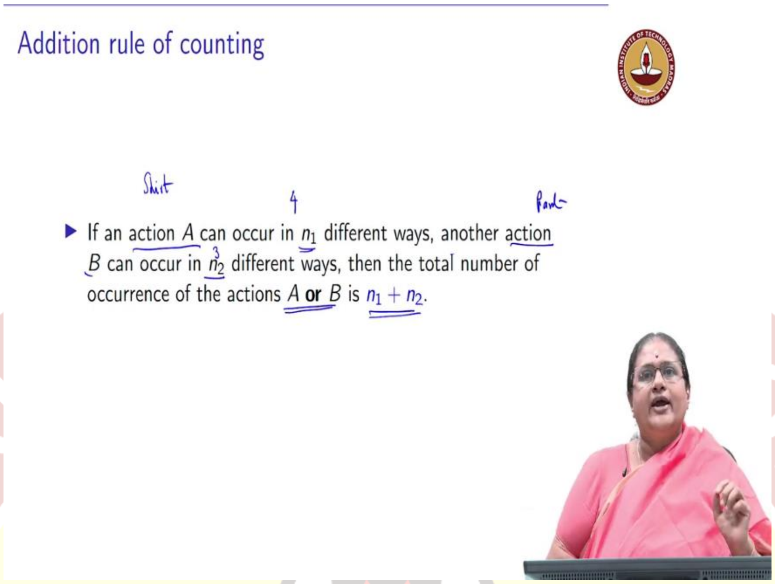


To answer this question I have how many choices to buy a shirt? I have 4 different choices to buy a shirt, I could have bought a yellow shirt or a blue shirt, those are the choices that are available to me or a green shirt or a red shirt these are the 4 different choices available to me, I have 3 different choices when it comes to a pant I can buy a black pant or a blue pant or a brown pant.

But what I need to remember is if I buy a shirt I cannot buy a pant, if I buy a pant I cannot buy a shirt, so the total number of choices that are available are the 4 choices which come from the shirts and the 3 choices that come from the pant and the total number of choices 4 plus 3 which is equal to 7.

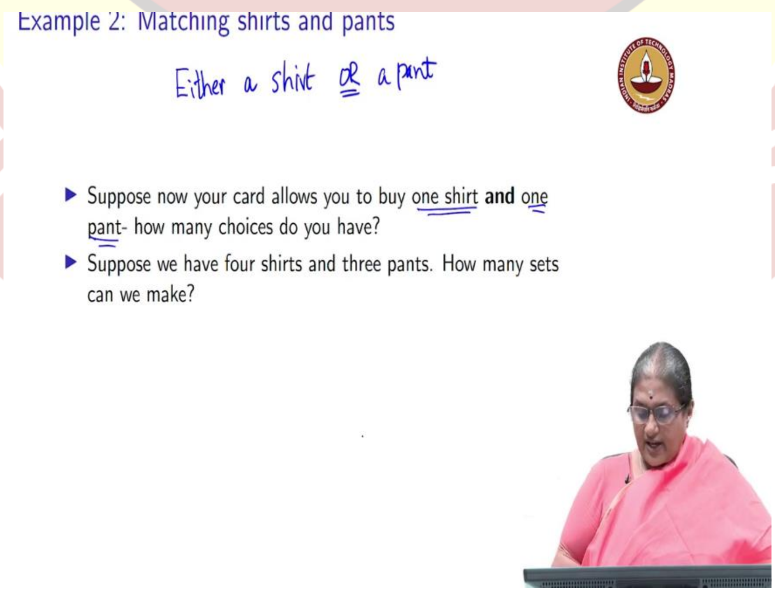
So, here when I am talking about a dependency that is what do I mean by dependency, here the actioner is you either buy a shirt or you buy a pant but I cannot, the actions are dependent what do I mean by dependent, I cannot buy a pant if I have bought a shirt and I cannot buy a shirt if I bought a pant, so in a sense that actions are dependent on each other I have 4 choices for my first action which is buying a shirt, I have 3 choices for the second event which is buying a pant and the total number of choices available here are the addition of the choices available for each of the action which is 4 plus a 3 which is equal to 7.

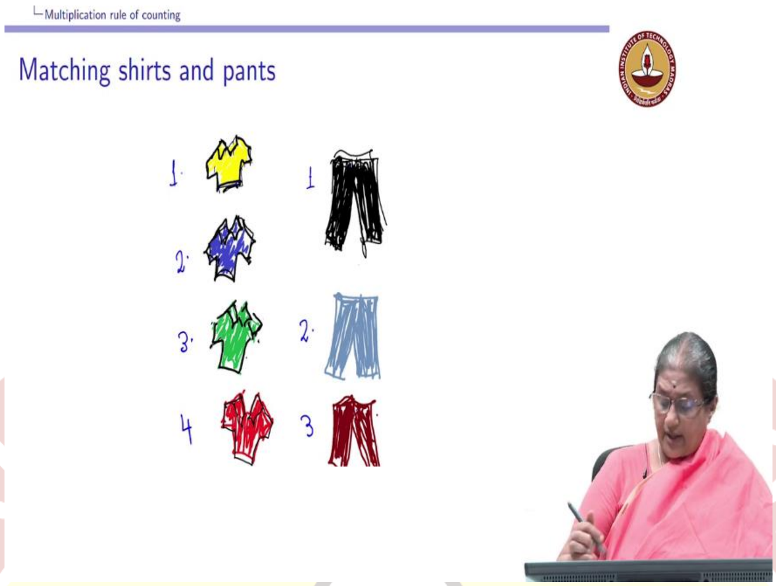
(Refer Slide Time: 8:15)



So, you can see that this is what I can formally state as the addition count rule of counting which I can state as if an action A, the action A in my example could be buying a shirt can occur in n1 different way, so n1 in my example was 4, another action B, action B could be buying a pant, this can occur in n2 different ways which is 3, then the total number of occurrence of actions A or B is given by n1 plus n2 which is given by n1 plus n2. So, now let us, this is what we refer to the addition rule of counting. Now, let us revisit the same example.

(Refer Slide Time: 9:05)

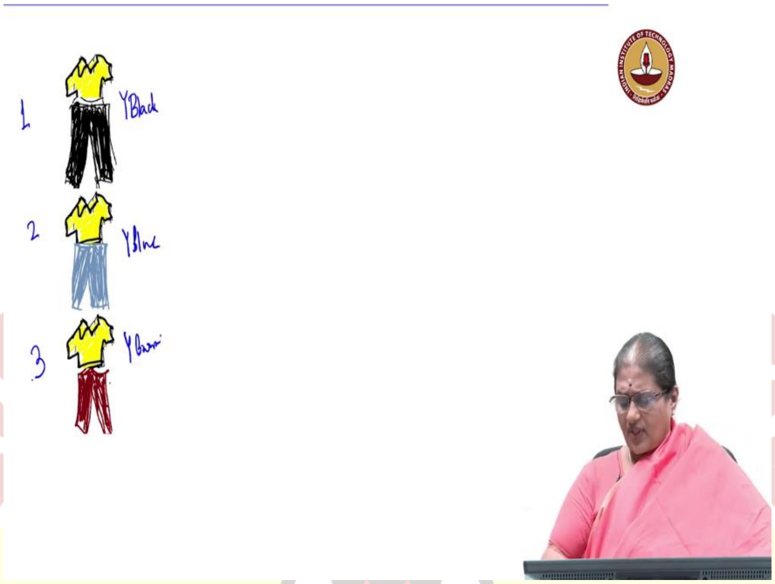


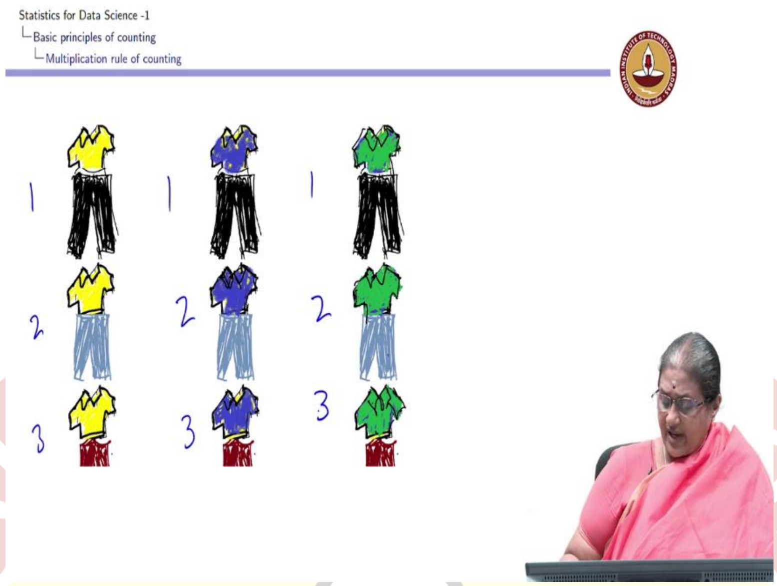


Now, suppose my card allows the following, initially my card said you choose either a shirt or a pant. So, it was very clear, the card said you choose either a shirt or a pant. But now, suppose the card allows you to buy 1 shirt and 1 pant, it is not allowing you to buy 2 shirts or 2 pants, it is allowing you to buy 1 shirt and 1 pant.

Now, how many choices do we have in this case? Again let us go back and we can see that I have 4 shirts, so this is my shirt 1 a yellow shirt, a blue shirt, a green shirt and a red shirt, I have a black pant, I have a blue pant and I have a brown pant. I can combine a yellow shirt with a black pant, a yellow shirt with a blue pant, a yellow shirt with a brown pant, this is 1, so how many ways can I buy it?

(Refer Slide Time: 10:22)



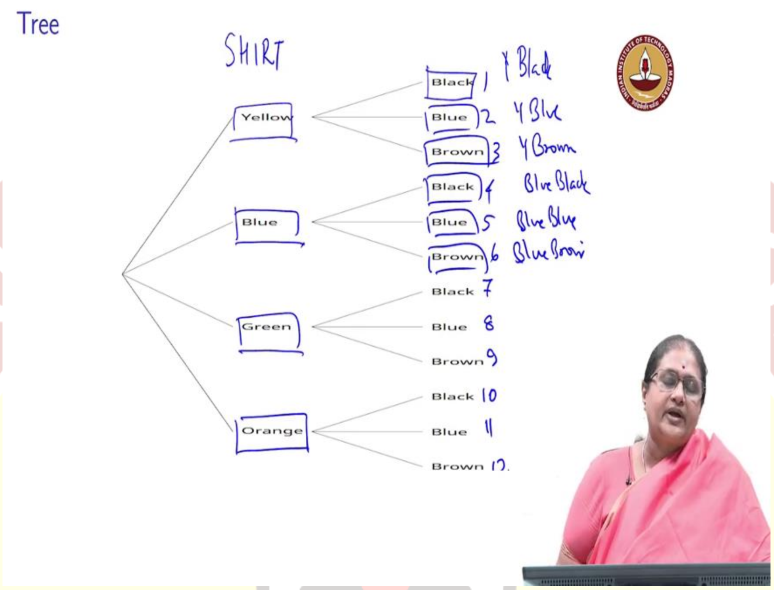


I can, when I am looking at the yellow shirt then I can see that a yellow black, a yellow blue and a yellow brown. So, that will give me 3 options, 1, 2 and 3. Similarly, what I can do is I can have a blue, so you can see a blue with black, a blue with blue and a blue with brown. So, I have 3 options from here 1, 2, 3. Blue again I have another 3 options I go to the next 1, a green with a black, green with a blue and a green with a brown, so I had 3 options here, I had 3 options with blue, I have 3 options with green and I finally have 3 options with orange also.

So, this is my 3, I have 3 options here, I have 3 options here, I have 3 options here, and I can see that what is the total number of things I have 3 here plus a 3 plus a 3 plus a 3 so I have a total of

12 options that are available for me to pick a shirt and a pant together and buy a pair of shirt and pants, these are 12 options that are available. Now, what do we mean by this?

(Refer Slide Time: 11:58)



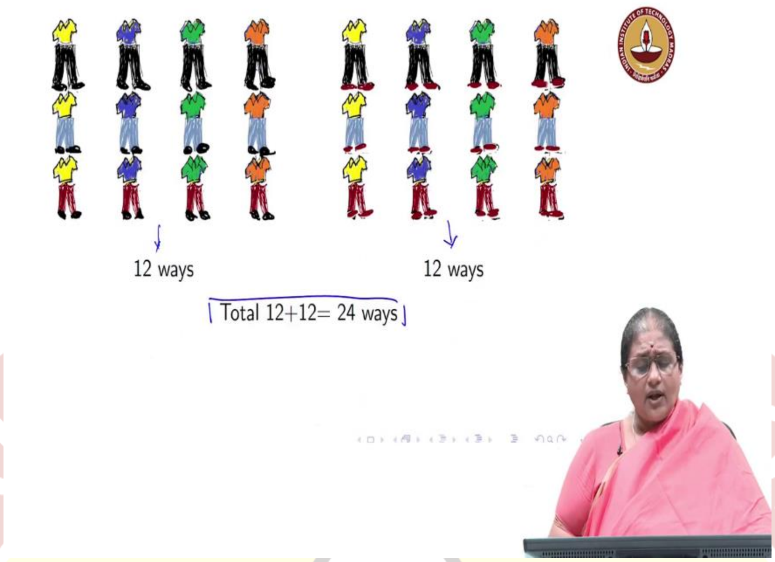
So, now let us look at this, I can represent it by what I call is a tree. So, if you look at the shirt options which is given at this level I have a yellow shirt, I have a blue shirt, I have a green shirt and I have a orange shirt. With every yellow shirt I could either have a black pant, so this is yellow with black, a blue, so I have a yellow with blue, a brown, yellow with brown. Similarly, I have a blue with black, blue with blue and blue with brown.

And you can see that the total number of ways I can do it is 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12. So, this sets the ground to understand what is the, what we refer to as the multiplicative rule of counting. What is the multiplicative rule of counting tell me? Before I go to multiplicative rule of counting, let us go to another example.

(Refer Slide Time: 13:18)

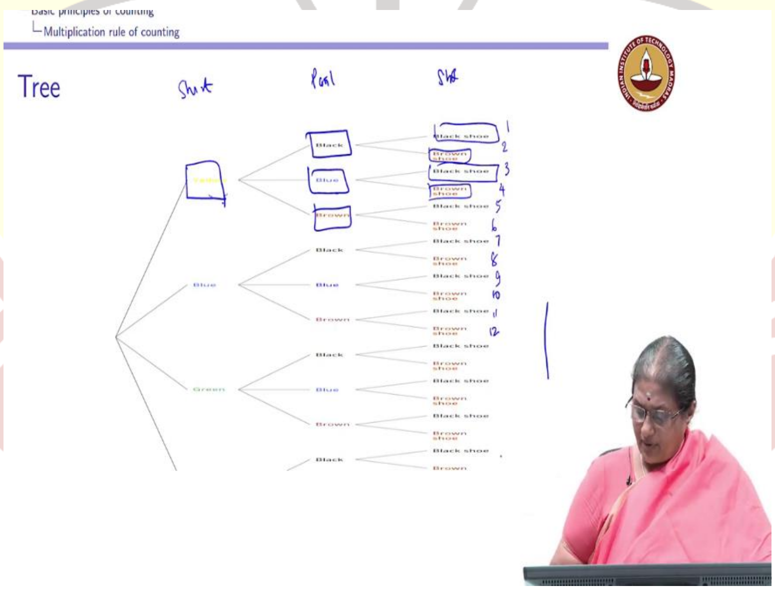
So, suppose I am having now I have 4 shirts, I have 3 pants and now I am allowing with the gift card to choose either black pair of shoes or brown pair of shoes, so I have 2 pairs of shoes I can choose. Now, I know that I can have 12 if I am pairing or matching shirts and pants, now with every pair of your pant and shirt you can either go with a black shoe or a brown shoe.

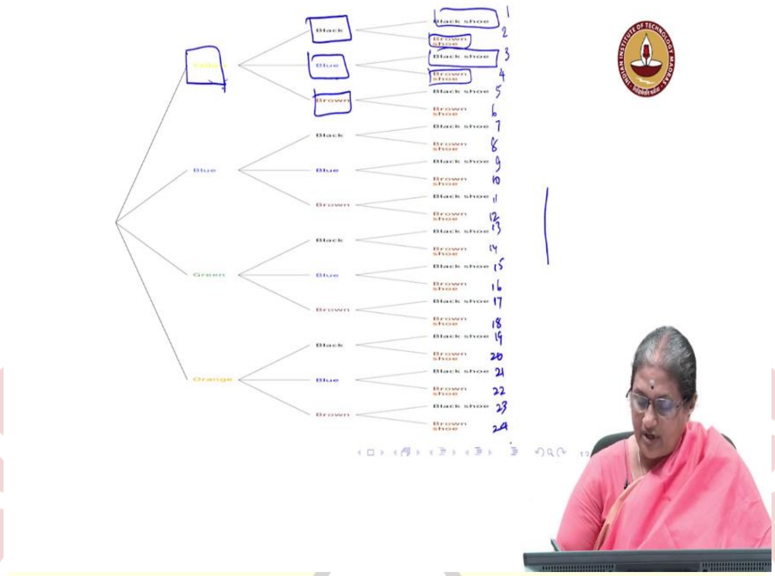
(Refer Slide Time: 13:56)

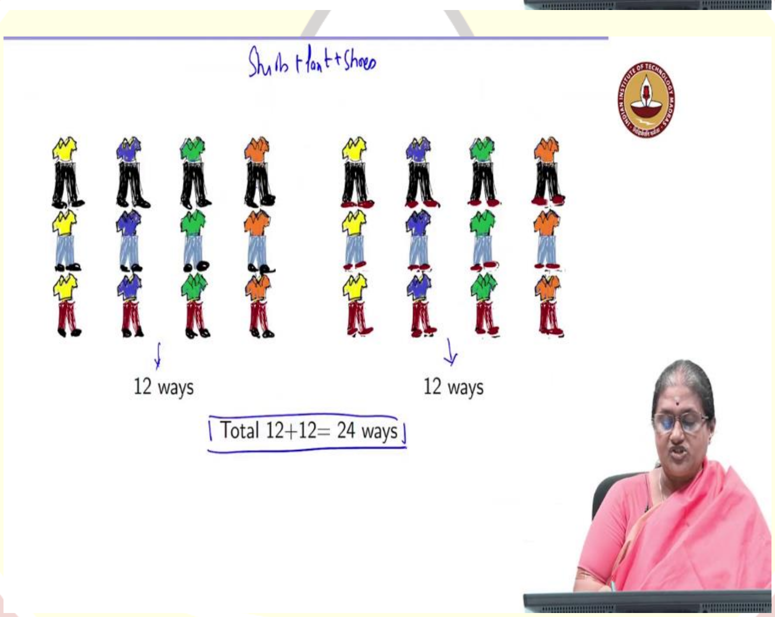


So, what I get is with each I had 12 pairs of my yellow and black, pant and shirt. And now you can see that I have for each of those 12 pairs I am having a black shoe, so this gives me 12 ways, the same 12 pairs with a brown shoe gives me 12 ways, so I have total 24 ways of matching the shirts with the pant and the shoes.

(Refer Slide Time: 14:33)



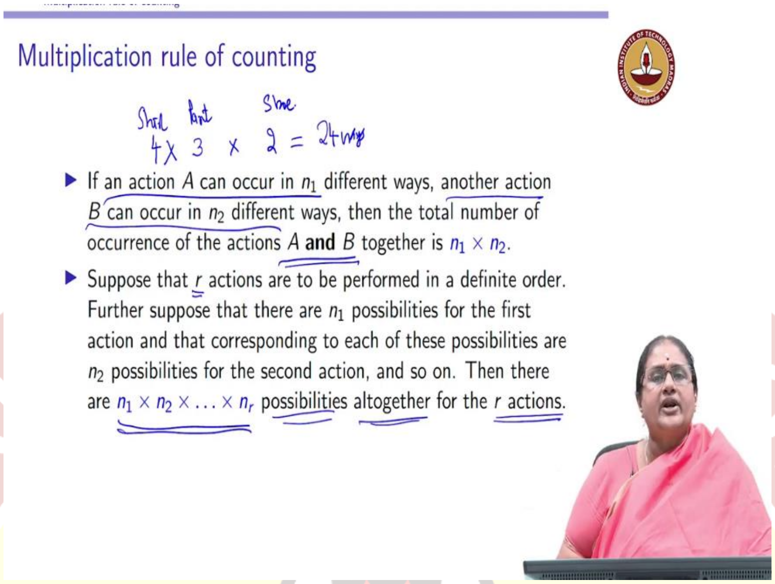




So, now if you go back and look at a decision or a tree, if you go back and look at a tree diagram, this is the huge tree. So, I have a yellow shirt, with a yellow shirt I could either have a black pant, a blue pant or a brown pant, with a yellow shirt and a black pant I could either have a black shoe or a brown shoe, with a yellow shirt a blue pant, so this is my pant, this is my shirt, this is my shoe. With the yellow shirt and a blue pant I could again have a black shoe or a brown shoe.

So, you can see if I number the total base of having this I have 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, I can keep going and you can see that the total number of ways is going to be so I have a 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, this is precisely what we had in our earlier slide that total twelve plus twelve is equal to 24 ways. So, what are we doing here?

(Refer Slide Time: 15:55)

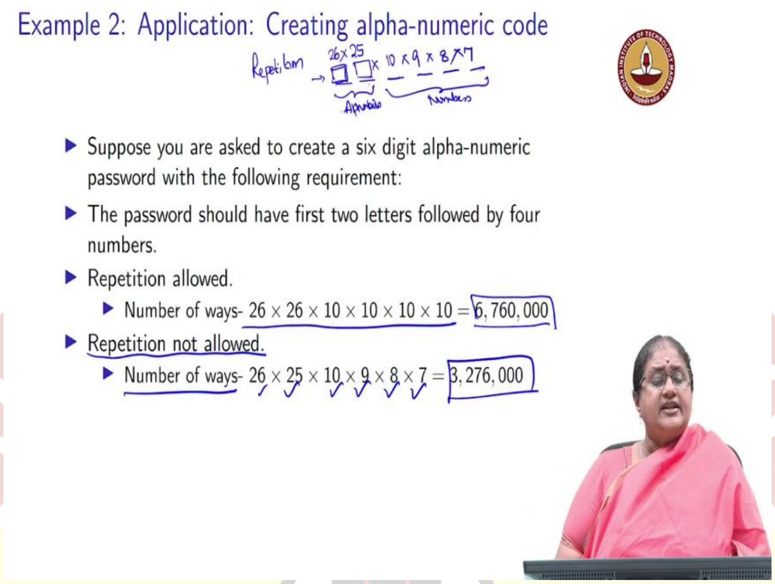


So, you can see that I can now I can state what is popularly known as the multiplication rule of counting. What does the multiplication rule of counting says? It says that if an action A can occur in n_1 different ways what is an action A, it was choosing a shirt. How many ways I could choose a shirt? I could have chosen a shirt in 4 different ways.

Another action B can occur in n_2 different ways, choosing a pant and that could have occurred in 3 different ways, then the total number of A and B, A and B earlier we looked at A or B, now we are looking at A and B is 4×3 is 12 ways. Suppose, I have r action, so in a third example we had 3 actions, what was it in addition to the shirt and pant?

I had to choose a shoe, so the shirt was occurring in n1 which is 4 ways, pant in 3 ways, shoe can be chosen in 2 ways, the total number of ways I can have, the all the 3 actions occurring that is choosing a shirt and a pant and a shoe is $4 \times 3 \times 2$ which is 24 ways. I can extend this to any sequence or order of r actions and $n_1 \times n_2 \times \ldots . \times n_r$ will give me the total number of possibilities all together for all these r actions to occur together. And this is what we refer to as the multiplication rule of counting.

(Refer Slide Time: 17:57)



So, now let us apply the multiplication rule of counting to create an alpha numeric code. Many at times whenever we login into some website has asked us to generate a password, suppose, I am logging in into a website which is asking me to generate a password and what it is telling me is you generate a 6 digit password, the first 2 digits should be alphabets and the next 4 digits should be numbers.

So, the first action is to fill in the first space. How many ways can this be done? I can choose any 1 of the 26 alphabets to be here. Again this second space is again an alphabet, I can choose any 1 of the 26 alphabets again here. So, in the first case if I allow repetition of my alphabets, the first can be chosen in 26 way, the second can be chosen in another 26 way, now numbers if I include 0 this has a 10 I can do this choice for this blank in 10 ways, this again and 10 ways, this again in 10 ways.

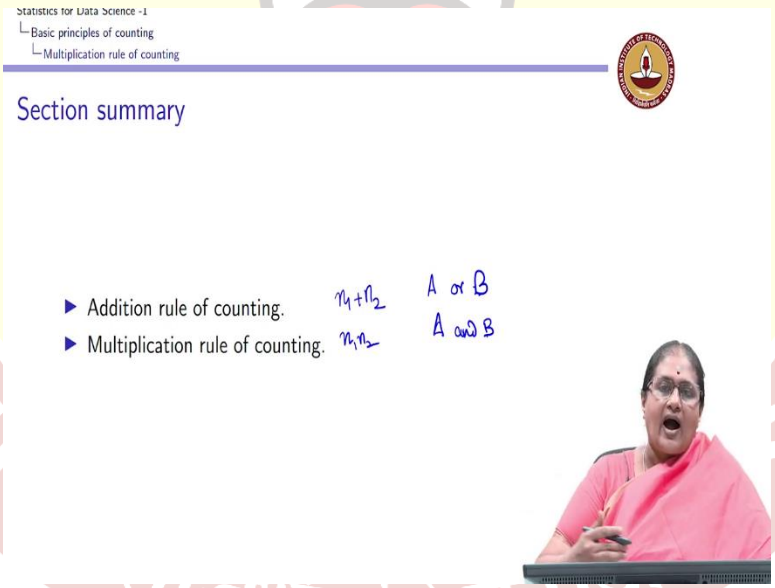
So, I have actually 6 actions, the first action is to choose an alphabet for the first blank, the second action is to choose an alphabet for the second blank, the third action is to choose a number for the third blank and so forth the 6th action is to choose a number for the 6th blank, so the total number of ways I can create or total number of codes I can create is $26 \times 26 \times 10 \times 10 \times 10$.

So, that is what I have here which is $26 \times 26 \times 10 \times 10 \times 10$ which is this number which is 6,760,000 ways to do it. Now, suppose I do not allow repetition. What do I mean? You can see that when I look at the alphabets here if I have chosen an alphabet to fill in the first blank I have only 25 alphabets available to fill in the second blank, so I have 25 here, so these 2 blanks or the first 2 blanks together can be filled up in 26×25 or 25×26 ways.

Similarly, the third blank which has to be a number for the first blank I have 10 choices, now once I have the 10 choices and I am not repeating that number, I only have a 9 choices for the second blank, I am again not repeating it, I have 8 choices that are left over for the third blank, and for the last blank I have only 7 choices that are left over.

Now, applying the same, so I have here r actions 26 choices for the first action, 25 choices for the second action, 10 for the third, 9 for the fourth, 8 for the fifth and 7 for the sixth action which gives me the total number of ways is $26 \times 25 \times 10 \times 9 \times 8 \times 7$ which is 3276, sorry which is this number 3,27,6000. So, you can see that how we can apply a basic principle of counting to something which we generate almost on a daily basis and every time you want to generate an alpha numeric code you know that you can choose from this number.

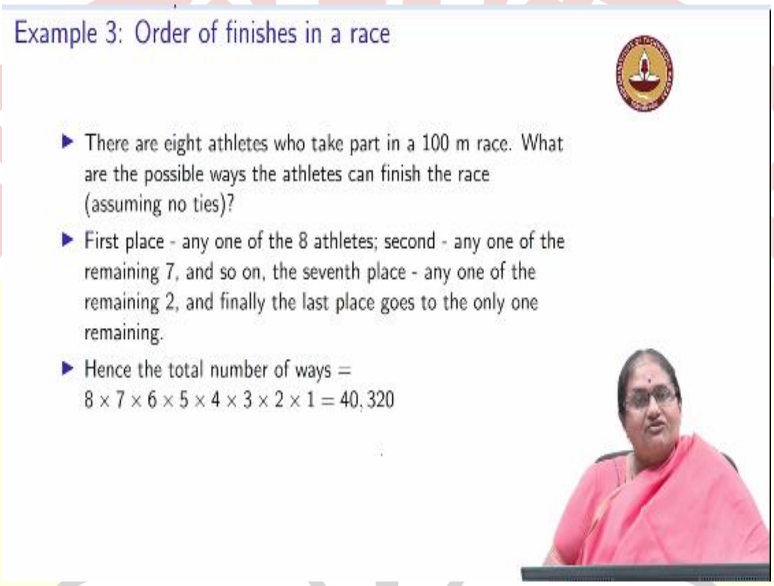
(Refer Slide Time: 22:12)



So, in summary what we have learnt so far is, first we have introduced what is the fundamental principle of counting, we say that the fundamental principle of counting is again we looked at the addition rule of counting here it is $n_1 + n_2$, here basically it is action A or action B, $n_1 \times n_2$where it is action A and action B and I can extend this logic to r actions.

Statistics for Data Science-I
Professor. Usha Mohan
Department of Management Studies
Indian Institute of Technology, Madras
Lecture 5.2
Permutations and Combinations - Factorials

(Refer Slide Time: 00:14)



So now let us introduce a very important concept which is referred to as a factorial notation or a factorial. Now let us look at a race it is a 100 meter race and I have 8 people who are running in this race. So they are distinct people so I have person 1, person 2, person 3, person 4, person 5, person 6, person 7 and person 8. These are all distinct people they are running this race.

Now what are the possible ways these athletes can finish the race. So I have position 1, 2, 3, 4, 5, 6, 7, 8 these are the positions that is what I mean by this is the first position of finish, the person who come second and the person who comes 8 because they are only 8 people these are the only finishing position. So what is it I am interested in I am interested in knowing how many orders of finish that is possible when 8 people are participating in a particular race.

And not assuming any ties now I am assuming that there is a clear ranking or a clear order of finish. So now let us look at this position and I am also assuming that all of the 8 of them are equally capable so there could be anybody can come first on any given day and the difference is not going to be too much. Now how many choices do we have for the first person? So if I put person 1 here just a hypothetical situation.

Then person 1 cannot appear in any other position because he has already come in the first position. So given that person 1 is in the first position, person 1 is not available for any other position so for person 2 to position 2 I have only a choice of 7 people because person 1 has already occupied the first position. So, suppose person 2 comes to the second person for the third position.

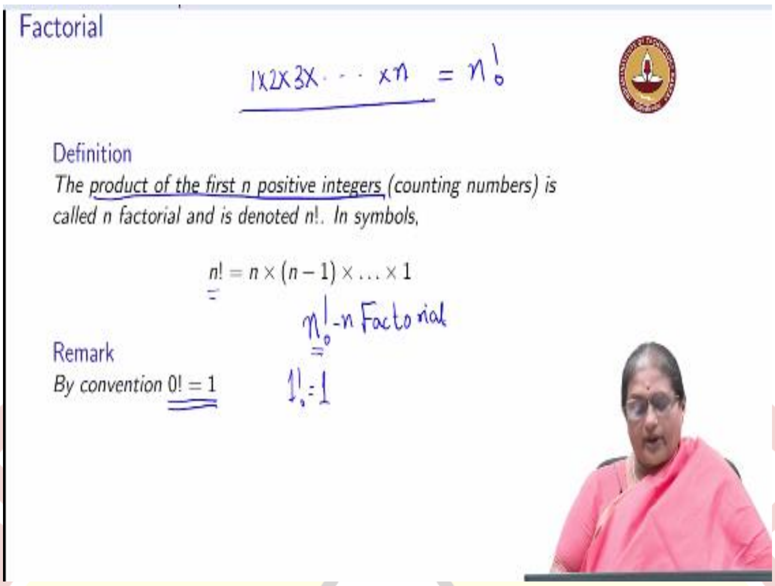
I do not have both person 1 and person 2 I have only a choice from the other 6 remaining and suppose person 4 is in the third position for the fourth position I have only 5 choices it could be person 7 and so forth I have person 8, I have person 6, I will have person 3 and I have person 5. So this is one order. So if I am looking at it the number of choices to fill in the first position I have 8 choices.

If I fill up the position 1 with one of these people for the second thing I have only 7 choices. I have 6 choices here, I have 5 choices here, 4, 3, 2 and 1 these are the choices I have to fill in these positions from the available number of people. These are distinct people. So I can in other words if I turn the action first action is to choose from these 8 people number of people who would fill in the first position.

Second action is to come up with the number of people who will fill in the second position and so forth 8 action is number of people who will fill in the 8 position then I can apply my multiplication rule of counting to know that the total number of possible ways in which all these 8 athletes can actually complete the race is going to be $8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1$ this is the large number.

So this is basically it helps me so instead of writing as $8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1$. Can I have a simpler way to express this number. So what is the simpler way to express this number which is $8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1$ this is where I define a notation which is referred to as a factorial notation.

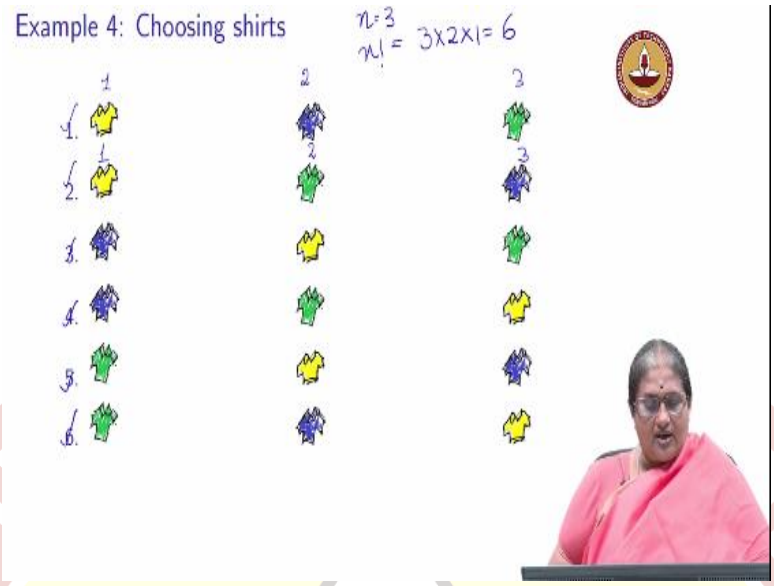
(Refer Slide Time: 05:20)



What is a factorial notation? Factorial notation is the product of the first n positive integers. What are the first n positive integer $1 \times 2 \times 3 \times \ldots \times n$. This product is what I referred to as n factorial $n!$. So this is the notation as n with a exclamation mark and I refer to it as a it is called n factorial and it is written as n factorial $n!$. by convention $0! = 1$, 1! is also equal to 1.

So now if I have this factorial notation my earlier problem becomes very simple I can write that the total number of ways I can express it as equal to 8 factorial which is far more elegant than writing it as $8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1$ I can just express it as 8 factorial, but this factorial notation is extremely useful.

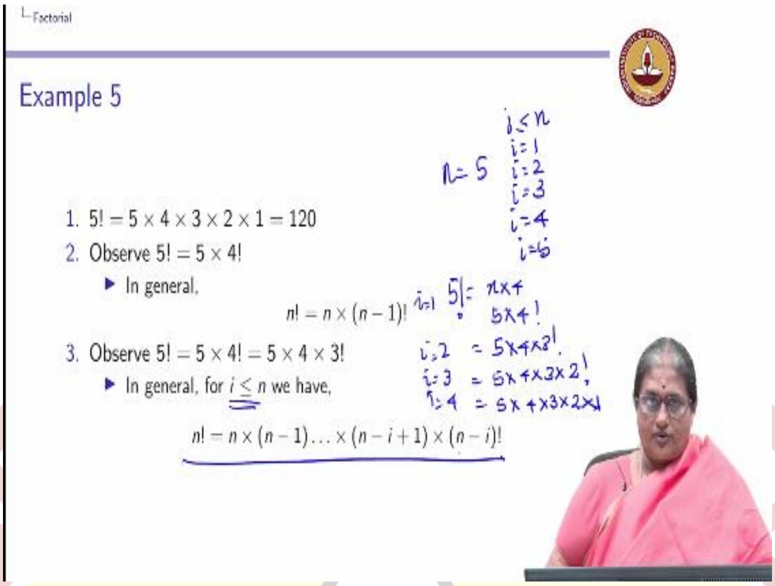
(Refer Slide Time: 08:43)



Again let us go back to looking at this example where I have three shirts. Now the first choice is again suppose in the same choosing this one I have only 3 people and this is again the order of finish of these 3 people or wearing a yellow t shirt, a blue t shirt and a green t shirt. The first order could be the yellow t shirt as 1 the blue t shirt is second, the green t shirt is third. The second is yellow t shirt is first, green is second, blue is third.

So you can see that the total number of ways is 1, 2, 3, 4, 5, 6. The number of object distinct object n is equal to 3 and you can see that n factorial the way I defined it was $3 \times 2 \times 1$ which is giving me a 6 here total number of choices which is $n!$ which is equal to 6.

(Refer Slide Time: 07:53)



So now let us go back and look at the factorial notation 5!. What is 5!? 5! is $5 \times 4 \times 3 \times 2 \times 1$ now if you look at this portion if you look at this portion it is $4 \times 3 \times 2 \times 1$. I know this is nothing, but 4!. Hence I have 5 factorial which can be written as $5 \times 4!$. In other words what we say is similarly if I look at 6! Is $6 \times 5 \times 4 \times 3 \times 2 \times 1$. Now this portion is nothing but 5!.

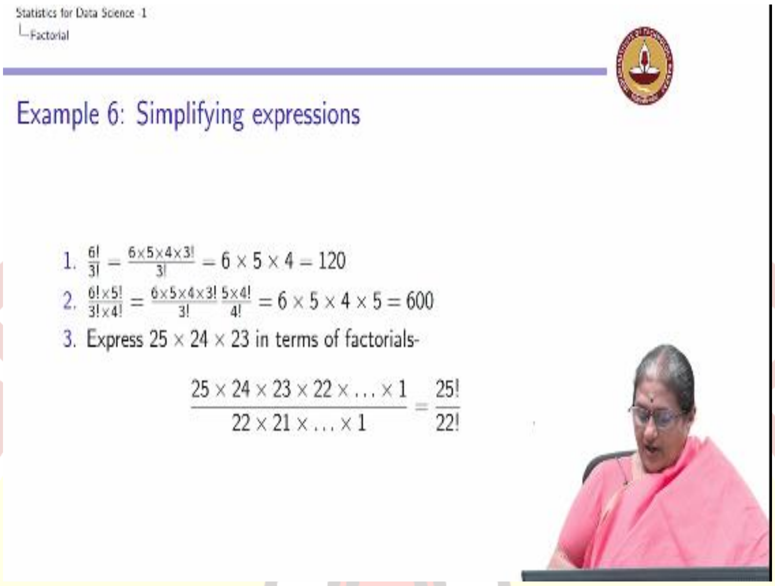
So I can write 6! as $6 \times 5!$. In other words for any integer $n! = n \times (n-1) \times (n-2) \ldots . \times 1$ this is $n \times (n-1)!$. So the first expression is $n! = n \times (n-1)!$. So I have $n! = n \times (n-1)!$. Now again look at the following I 5! is $5 \times 4 \times 3 \times 2 \times 1$. So I can write this also as $5 \times 4 \times 3 \times 2 \times 1$ where this is nothing but 3 factorial.

So I can write this as $5 \times 4 \times 3!$ Similarly if I have a 6! is $6 \times 5 \times 4 \times 3 \times 2 \times 1$ I see that this is equal to 4!. So I can express 6! as $6 \times 5 \times 4!$. So you can see that if I have 5! I can write it as $5 \times 4 \times 3!$. 6! or $6 \times 5 \times 4!$. So in general for any $i \leq n$ so my n if it is 5 what are the i's that are less than or equal to n.

I can have $i = 1$. I can have $i = 2$, $i = 3$, $i = 4$ or $i = 5$ I have $n!$ which is 5! which is n which is equal to 5×4 so my $i = 1$ $n - i$ is 4! it is $5 \times 4!$, $i = 2$ it is $5 \times 4 \times 3!$, $i = 3$ $n - i$ I have $5 \times 4 \times 3 \times 2!$, $i = 4$ is $5 \times 4 \times 3 \times 2 \times 1!$. So this is another expression which we can use.

And this will help us simplify a lot of expressions when we are encountering counting when we do our permutations and combinations and probability problems.

(Refer Slide Time: 11:56)



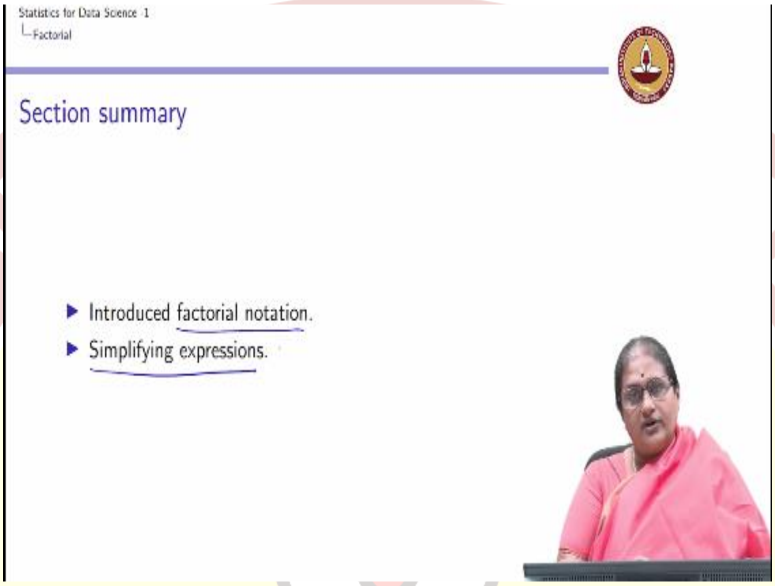
Now let us look at further simplification of factorial notation. Now what is the factorial notation we are looking at what is 6!. 6! I know is $6 \times 5 \times 4 \times 3 \times 2 \times 1$. From my first definition 3! I just retained that way. I recognize that this quantity is nothing, but 3! so I can cancel out this with this and I get $\frac{6!}{3!}$ is nothing but 6! is $6 \times 5 \times 4$ which is 120.

Now let us look at the next problem $6! \times 5!$ so I have a 6 into 5 into 4 we have already seen $\frac{6!}{3!}$ is $6 \times 5 \times 4$. Now $\frac{5!}{4!}$is going to be 5 factorial divided by 4 factorial into 4 factorial. So I can see that I can simplify this by cancelling out the 4! with 4! and what I have is 600 as the answer. Now sometimes you might want to express a product in terms of factorial.

So suppose the example I have taken here is to express $25 \times 24 \times 23$ in terms of factorials let me write it I am having $25 \times 24 \times 23$. I am going to multiply the numerator so I am going to write it as a fraction into 21. So forth with 1 I only want this portion so I need to divide the denominator with this portion which I have put with double line and what is that portion that portion is nothing, but $22 \times 21 \times 20 \times \ldots . \times 2 \times 1$.

Now what is this denominator? Denominator is nothing but 22!. My numerator is nothing but25!. So I can express this $25 \times 24 \times 23$ as $\frac{25!}{22!}$. So I can express this $25 \times 24 \times 23$ as $\frac{25!}{22!}$. So this notation of expressing the product in terms of factorial would come in news when we learn more about permutations and combinations.

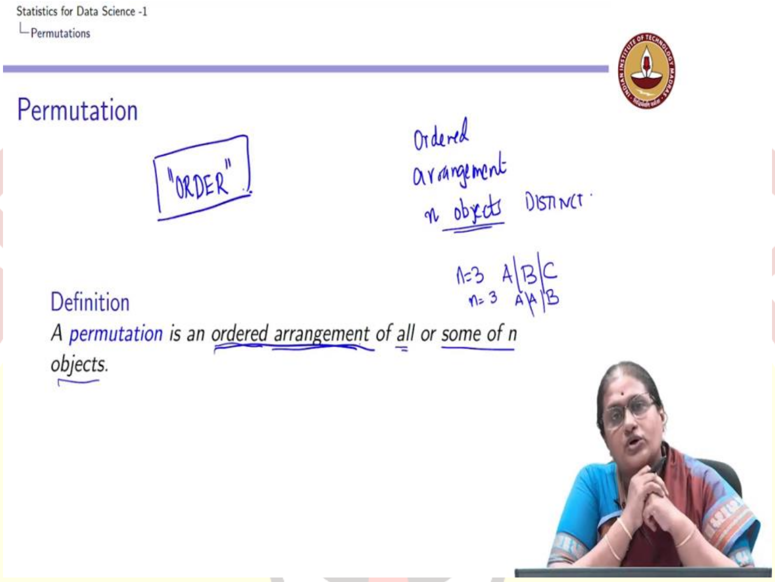
(Refer Slide Time: 15:14)



So, in summary what we have learned today is we introduced the factorial notation and how do we simplify expressions using the factorial notation, that is what we have learned.

Statistics for Data Science – 1
Professor. Usha Mohan
Department of Management Studies
Indian Institute of Technology, Madras
Lecture No. 5.3
Permutations and Combinations – Permutations: Distinct Objects

(Refer Slide Time: 00:13)

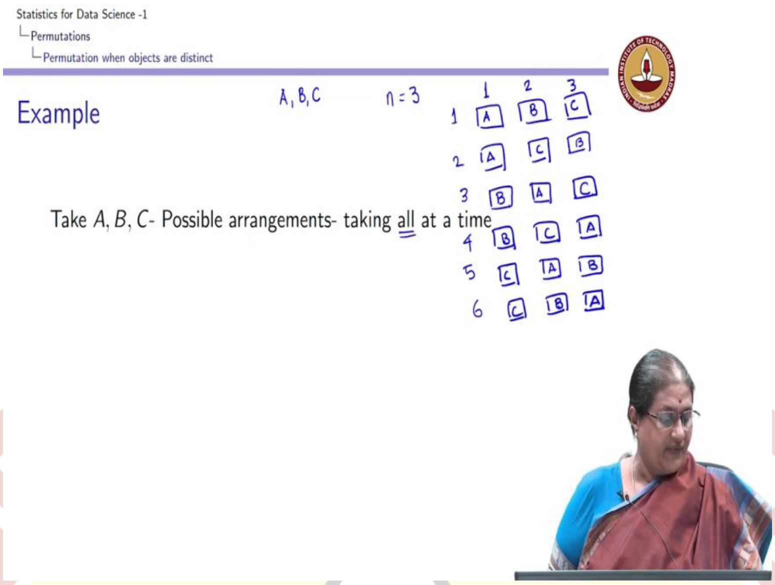


We now introduce what is a Permutation. Permutations and Combinations are extremely important when we are going to learn about probability. So, what is, we formally define what is a permutation? We can formally define what is a permutation as an ordered arrangement of all or some of n objects. I want to know you all to notice something very carefully. We have 3, qualified the statement with 3 important words. The first word is ordered, the second is arrangement, and the third is n objects.

Now, for, now, we are going that assume that these n objects are distinct objects. Now, what do I mean by distinct object? For example, if I take A, B, and C, these could be 3 people. When I am referring to 3 people, they are distinct people. So, first we are going to understand, and then we will start looking at what would happen if they are not distinct. For example, I can again, so n = 3 here, A, B, C are distinct. I can again take n = 3 objects or I can have an A, A and B.

It could be a case that I have, 2 red balls and 1 blue ball. So here, I do not have all the 3 of them distinct, but if I had a red, blue and a yellow ball, all of them were distinct. So, now, a permutation is an ordered arrangement. So, the key we need to notice here is this idea of order. And this is what defines what is a permutation.

(Refer Slide Time: 02:05)



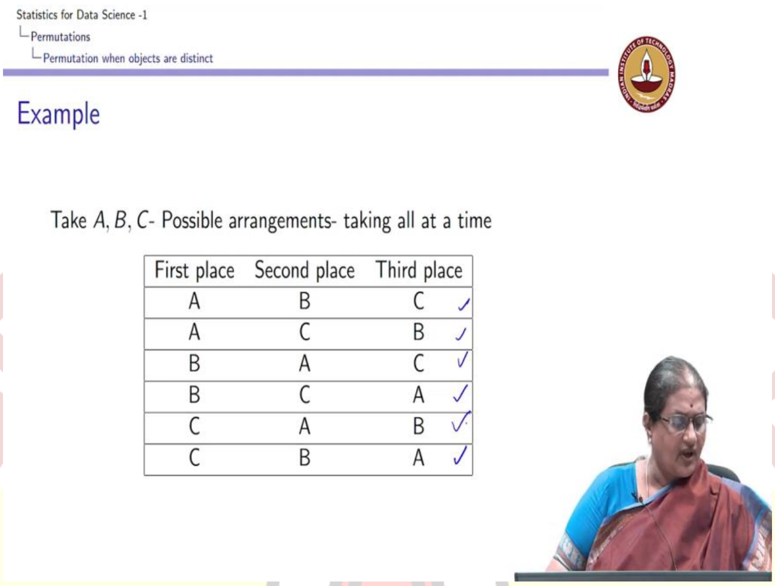
So, what is a permutation? Let us start with a very simple example, where we will extend what we learned about the fundamental principle of counting. So, what is this example? So, let us look at A, B and C. When we look at A, B and C, the number of objects or number of articles or number of things, whatever we want to refer to it, n equal to 3, because I have a A, I have a B, and I have a C. Now, let us look at all possible arrangements taking all at a time. In a sense, I want to see what are the all the possible arrangements of this A, B and C when I take all of them at a time.

So, the way I can look at it is, the another way to look at it is, I have 3 boxes, and I am looking at filling these 3 boxes with this A, B, C. And I want to know how I can do it, how many ways I can do this. So, you can start simply, intuitively. If I put a A here, let me fix the first box with an A. In the second box, I could have a B or a C. So, A, B, C, and A, C, B are possible ways of filling this box.

The another way I can look at it is, I can start with a B. Similarly, the second box and the third box. Second box could be a A, in which I can have the third box, only a C. Now, I am allow, I am not allowing the same repetition. When do I say I am allowing a repetition? For example, if I have, if I want to pick up some balls, I pick it up, I note the number and put it back. I am not allowing that. I am not allowing the repetitions in this case.

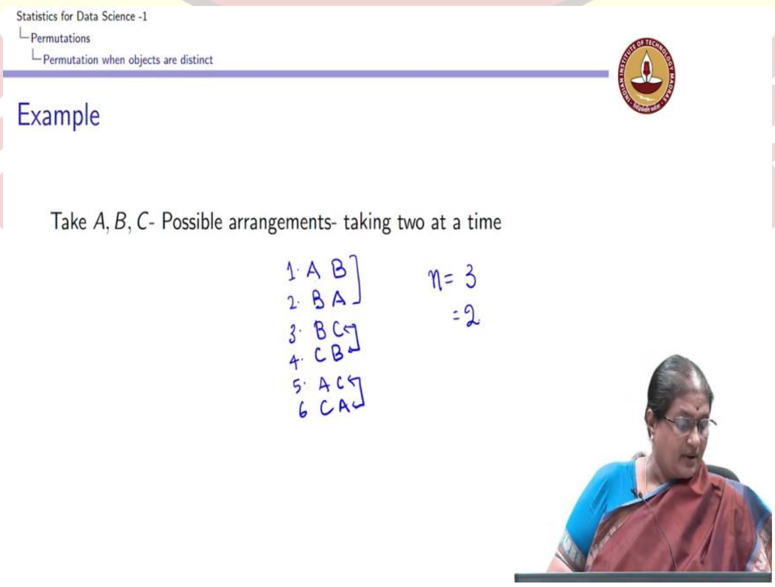
Similarly, if I have a B, I could have a C and a A. If I have the first, if my first blank is a C, first box is a C, the second could be a A, the third could be a B. The first as a C, the second could be a B and third could be a A. So, I can enumerate the number of possible arrangements when I am taking all the 3 alphabets here, A, B, C, which are distinct. If I count the number of possible arrangements, I have a 1, 2, 3, 4, 5 and 6.

(Refer Slide Time: 05:06)



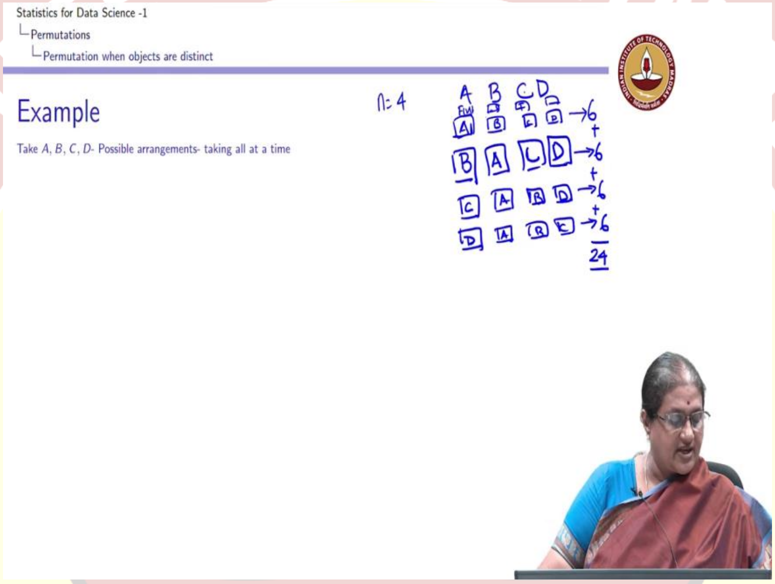
So, the total number of possible arrangements that are possible here are 6. So, ABC, ACB, BAC, BCA, CAB and CBA. Now, let us look at a slightly, so if you go back and see in the definition, I said it is an ordered arrangements of all or some. So, it is not necessary that I should take all the 3 of them together. So, I am interested, the second thing I am interested in knowing is, what are the possible arrangements of this ABC if I only take 2 of them?

(Refer Slide Time: 05:46)



That is I am choosing 2 out of A,B, and C. How many ways can I choose it, we will come to this in a quite a time and we are going to talk about combinations. But I can look at A, B and C, taking 2 at a time, I can choose A and B. And within A and B, I can arrange them as AB or BA. Similarly, I can choose B and C. Within B and C, CB, because this ordering is different from this ordering. And I can look at AC and CA. So, when I am taking out of n = 3, I am taking 2 at a time, I see, again the number of possible arrangements are 1, 2, 3, 4, 5 and 6. In this case, both the first answer and second answer are the same.

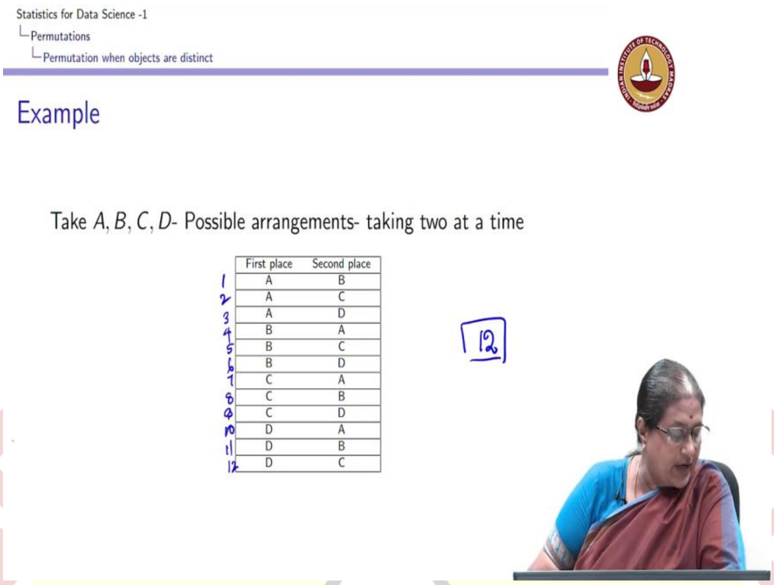
(Refer Slide Time: 06:51)



But then after this, we will, let us look at another example where I am looking at now, n = 4. And I have A, B, C, and D. Now, one way to look at it as you can fix A in my first place. And I know that B, C, D, if I look at B, C, D from a earlier example, I know if I have 3 places, the second, third and fourth, and within, between BCD, the way I can fill these 3 places with B, C and D are 6. So, if I fix A in my first place, I get 6 ways of getting hold of different possible arrangements. Is it clear?

So, the second thing, if I fix B, I can do the similar thing. I have a A, C and D to choose from for the 3 places. And I also know again that there are 6 ways of doing this. Similarly, if I fix a C, I can choose from A, B and D. And again, there are 6 ways of doing this. If I fix D as the first place, I have A, B and C. And there are again 6 ways of doing this, giving me 6 + 6 + 6 + 6, a total number of 24 arrangements, when n = 4. I can list all the arrangements in that following way.

(Refer Slide Time: 08:35)

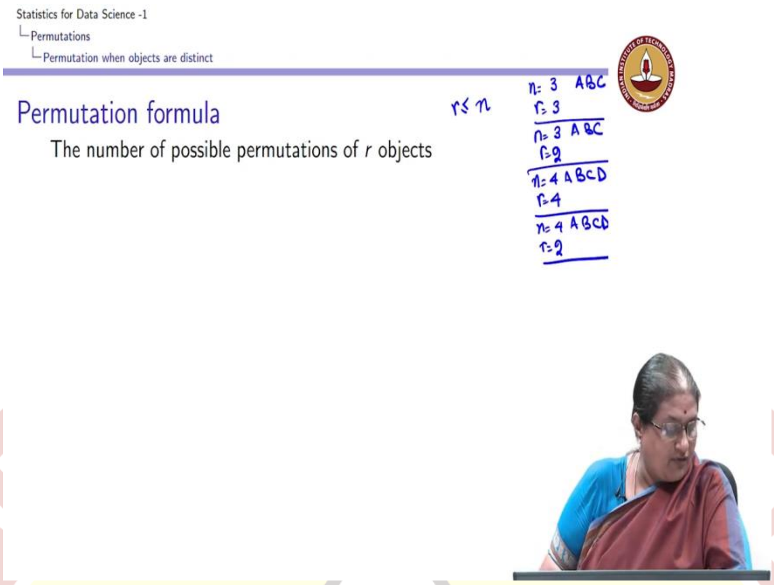


So, I have 6 from A, 6 from B. I can, again I can show you that I have. So, these are the 6 from A, I have 6 from B. I have 6 from C and 6 from D. And the total number of possible arrangements, the total number of possible arrangements in this case, you can see is 6 + 6 +6 + 6. I have 24 possible arrangements of A, B, C, D taking all at a time.

Now let us look at the same thing. Now if I am taking looking at the same possible arrangements A, B, C, D but I am now taking 2 at a time. Now again, if I have if I take A as my first object, the second could be a B, C or D. So, it could be AB, AC or AD. But that AB, I can have another arrangement BA, CA and DA. So again, with B, I could either have a A, but I already have listed that here. So, BC, BD, which will give me CB and DB. I have CD and DC.

So, the total number of arrangements I have, taking 2 at a time are AB, AC, AD, BA, BC, BD, CA, CB, CD, DA, DB, and DC, I have 12 arrangements of taking 2 at a time. Now, let us formalize this example to define what is a permutation?

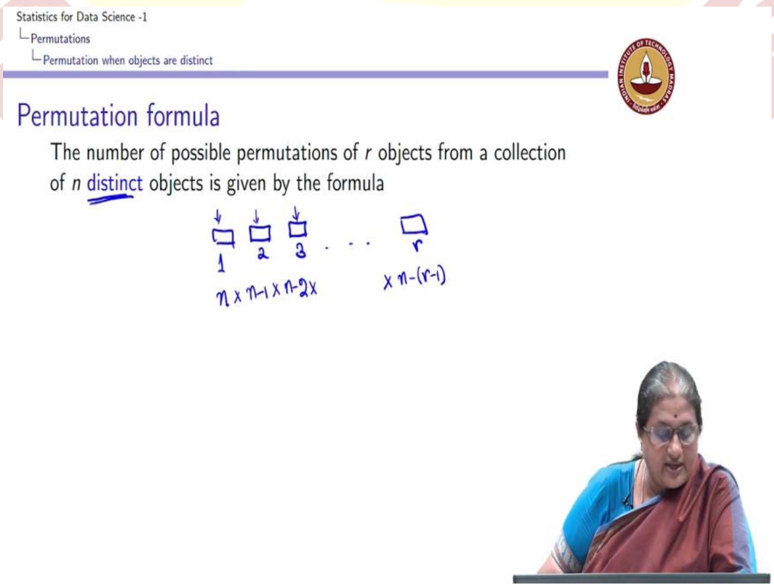
(Refer Slide Time: 10:35)



So, we are now ready to have the definition of a permutation. I have n objects of this, I am choosing r object. So, r ≤n. In my first example, when n was 3, I looked at ABC, I took at all at a time for my r was also 3. The second thing I took n = 3 again, ABC, but I was looking at 2 at a time, so my r was 2.

My third example, n was 4. I looked at ABCD, I again looked at all at a time, r was for 4. My fourth example, I had a ABCD, I again looked at 2 at a time, my r was 2. So, the question I am asking now is, how many possible permutations, these arrangements are also referred to as permutations. So, the question we are asking is, how many possible permutations are of r objects from a collection of n distinct?

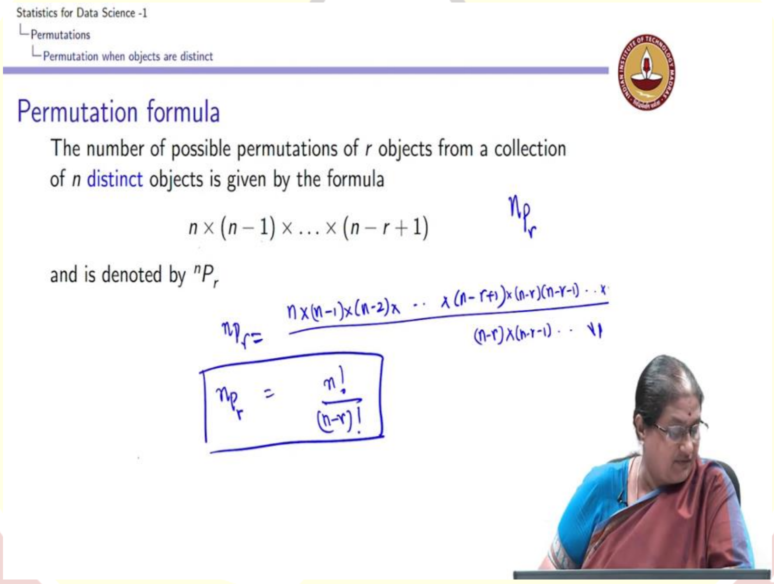
(Refer Slide Time: 11:42)



It is very important for us to understand that we are having it as distinct objects is given by the formula. So now, let us put in r blocks, 1, 2, 3. So, this is my first block, this is my second block, this is my third block, I can go so forth to r block. The number of ways, I can fill in this first block is n base. Now, once I have chosen from these n distinct objects, I have chosen 1 object to fill in this block. I do not have that available with me for the remaining. So, number of ways, I can fill in the second is n - 1.

So now, I have, the first and second are filled with 2 objects from these n objects. So, third, I can fill in n - 2 possible ways. So forth I can go, the rth, I can fill in n – (r – 1) possible ways. So, the first one can be filled in n way, second in n - 1 ways, third in n - 2 ways and the rth can be filled in n – (r – 1) ways.

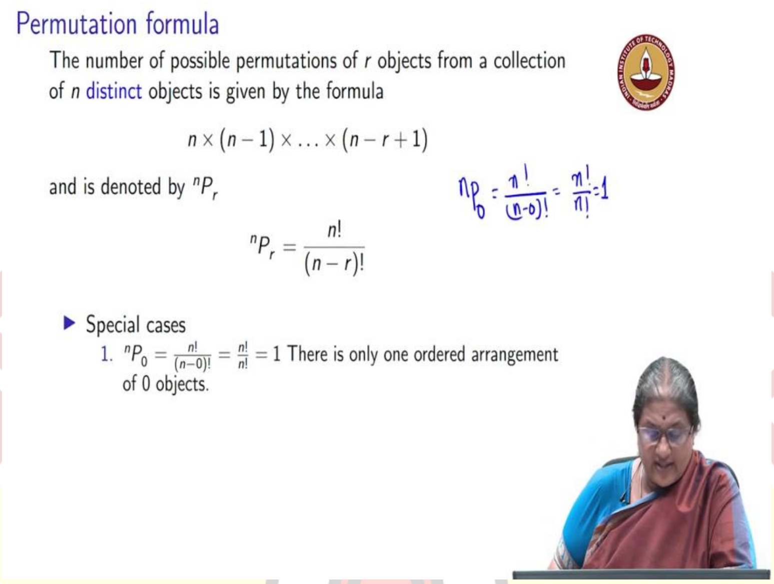
(Refer Slide Time: 13:20)



Again, using the fundamental multiplication rule of counting, I know the total number of ways I can fill, these r boxes n(n - 1) (n - 2) ..(n - r + 1), which can be written formally as n(n - 1) (n - 2) ..(n - r + 1). It is denoted by ${}^n_rP$. The way I can express this is, the collection from a collection of n objects, I am choosing r objects, this is given by ${}^n_rP$.

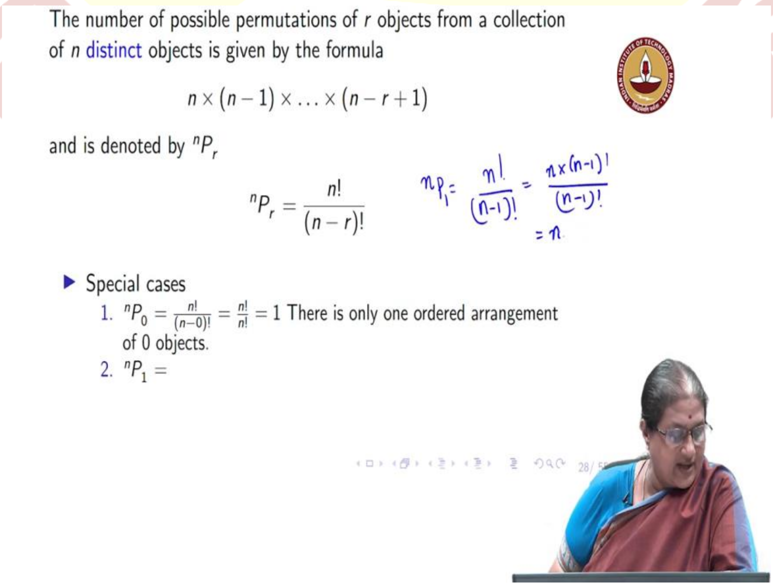
Now, let us look at this n into, now let us look at n into n minus 1 into n minus 2. So, I know nPr from the fundamental theorem of counting is given by this expression. Now, can I simplify this expression? If you look at it, if I can multiply my numerator with n minus r into n minus r minus 1 up to 1, I divide both numerator and denominator with the same quantity. The numerator now is nothing but n factorial. The denominator is nothing but n minus r factorial. So, I have a simplified expression for my nPr, which is n factorial by n minus r factorial.

(Refer Slide Time: 14:53)



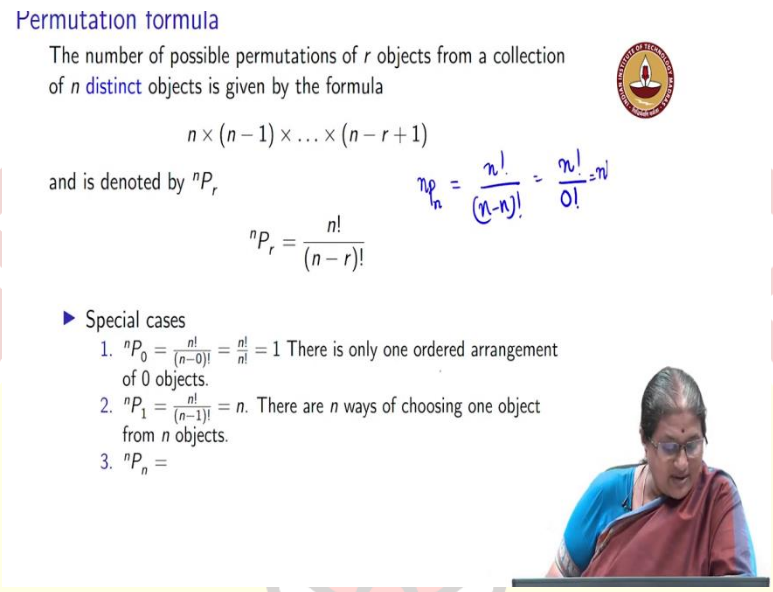
So, nPr is n factorial by n minus r factorial. Now, going back to the examples we have just seen, we can see that there are certain special cases of this formula. What are the special cases of this formula? What is nP0? The special cases of this formula, if you look at nP0, I apply the same formula, it is n factorial by n minus 0 factorial, n minus 0 is again n. So, I have n factorial by n factorial, which is 1. So, what does this mean? It means that when I am looking at no object in nP0, I have only 1 ordered arrangement of 0 objects. Now, if I am looking at the second special case, what is the second special case?

(Refer Slide Time: 16:06)



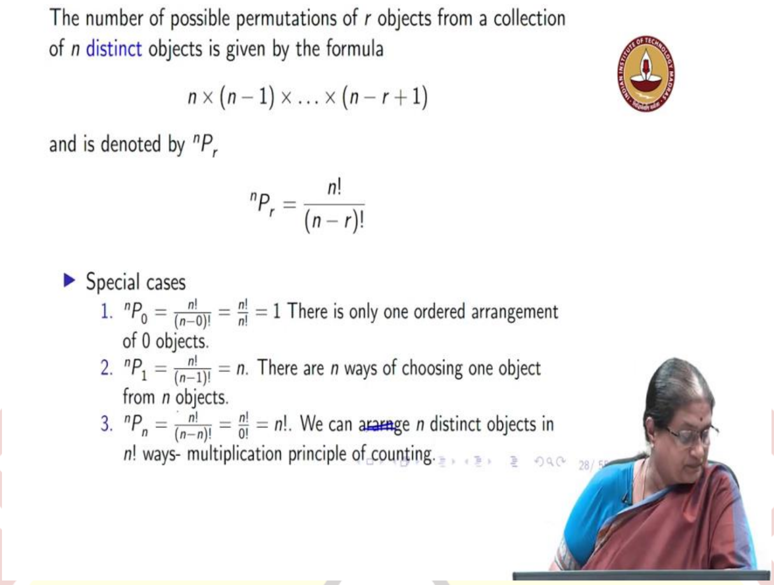
The second special case is, let me look at, nP1, what is nP1? Again, nP1 equal to n factorial by n minus 1 factorial. Now, I can write this numerator as n into n minus 1 factorial. i divided by n minus 1 factorial, so I have n as the answer. Now, nP1 is n, which is, this also can be viewed as the number of ways I can fill in 1 blank when I can choose from n objects.

(Refer Slide Time: 16:44)



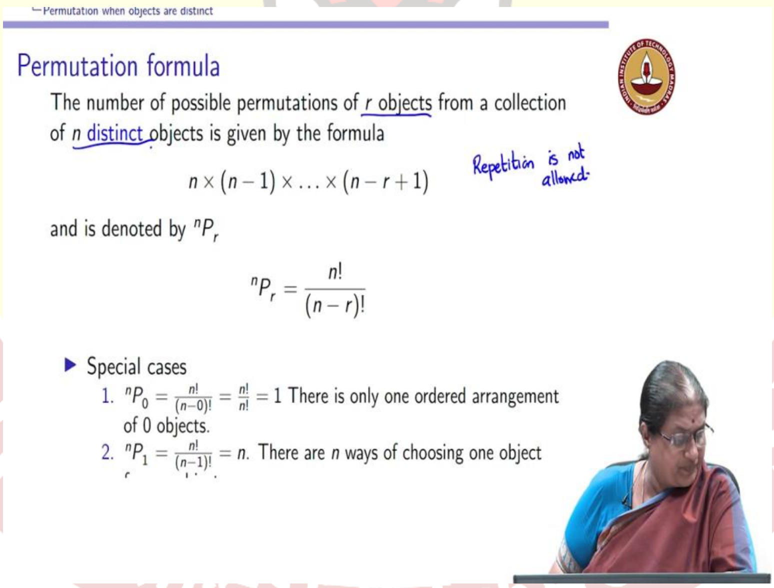
So, nP1 is n. There are n ways of choosing 1 objects from available and distinct objects. The third thing I want, we want to see is what is nPn. nPn is again number of ways of choosing n objects from an available n objects. So, nPn, if we apply the formula again, is n factorial by n minus n factorial, which is n factorial by 0 factorial. We already know, 0 factorial is 1, which is n factorial. So, nPn is the same as n factorial, which we have, we can see that this actually is what we got when we applied the fundamental theorem of multiplication, theorem of counting.

(Refer Slide Time: 17:43)



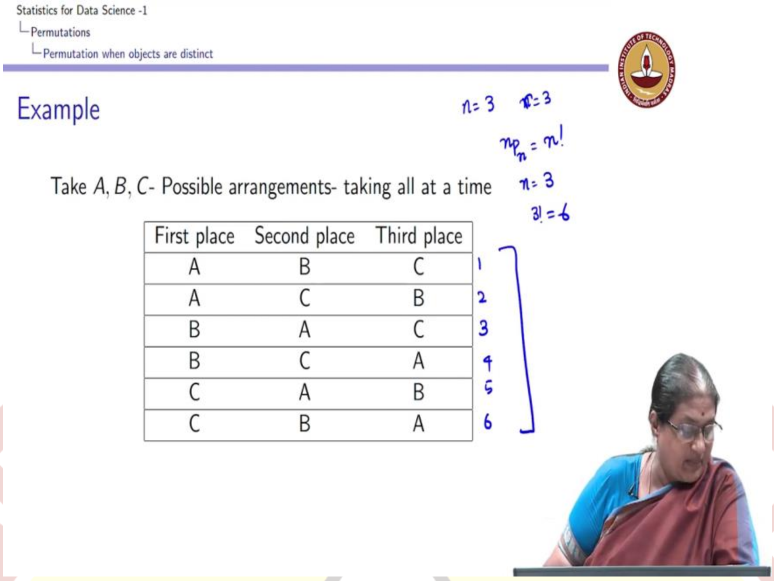
So, I have nPn equal to n factorial, where I am arranging the n distinct objects in n factorial ways.

(Refer Slide Time: 17:55)



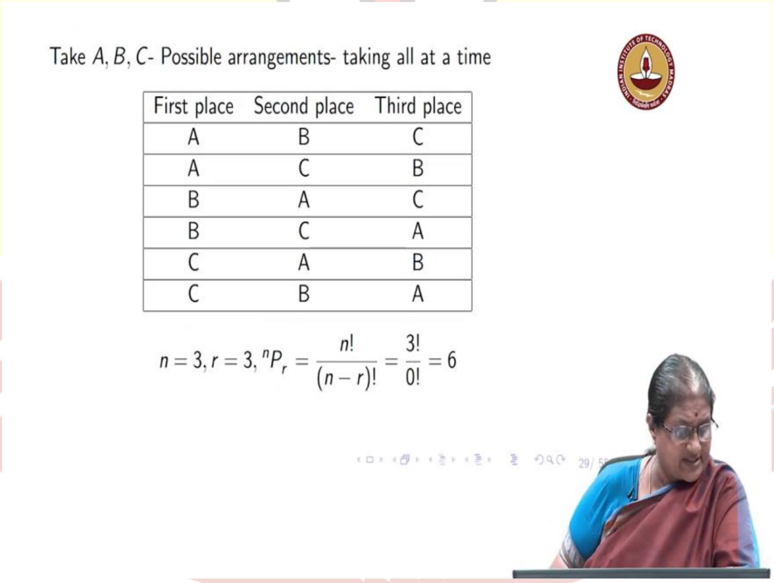
So, this is the main permutation formula, which tells me the possible permutations of r objects from n distinct objects. The key thing to remember here is repetition is not allowed. That is a key thing too. So, in that case, I have nPr is n factorial by n minus r factorial.

(Refer Slide Time: 18:32)



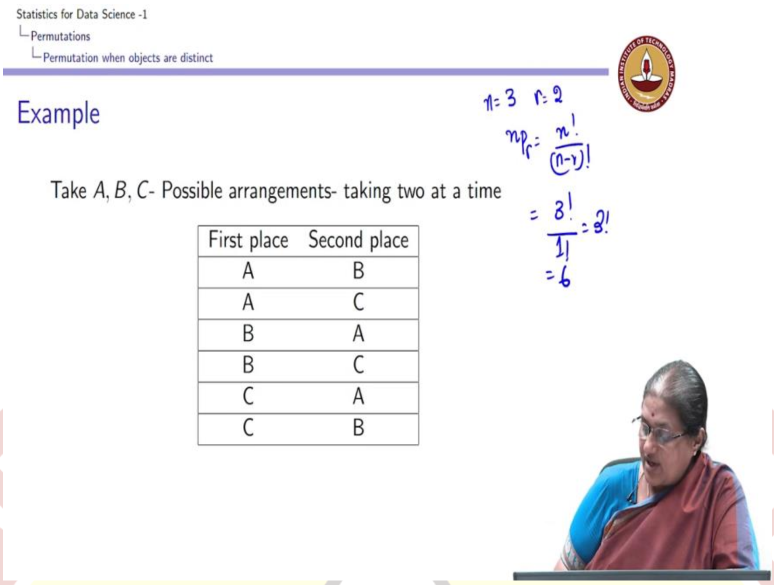
So, let us apply that to what we have seen so far. Now, when, the first example my n was 3, I am taking all at a time, so my r is also equal to 3. So, we can say that this would reduce the nPn, which is n factorial when n equal to 3. I know 3 factorial is 6.

(Refer Slide Time: 19:03)



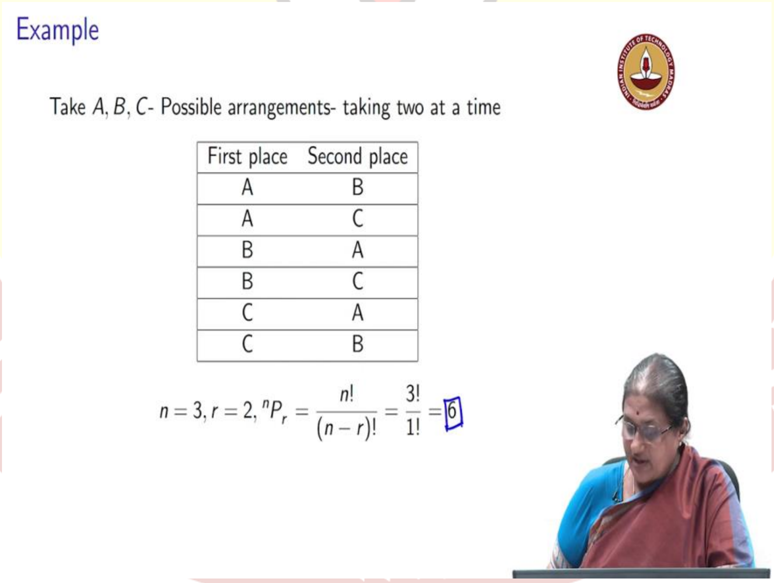
And I have 1, 2, 3, 4, 5, 6, which is the total number of ways in my first example, which we discussed. So, in the first example, I have nPr, which is equal to 6, which is nothing but n factorial. Now, let us look at the second case.

(Refer Slide Time: 19:17)



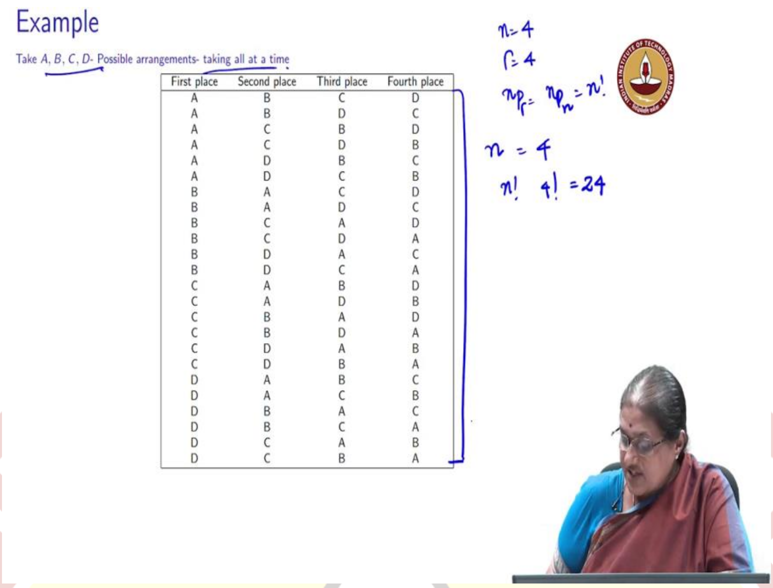
Now in the second case, I am looking at, again I have 3, n equal to 3, but my r equal to 2. So, nPr is n factorial by n minus r factorial, which is 3 factorial by 1 factorial, which is again 3 factorial, which is 6.

(Refer Slide Time: 19:40)



So here, in this case, my n equal to r 3, and r equal to 2, and I have the total number of possible arrangements is again 6. Now let us look at this third example, which we considered.

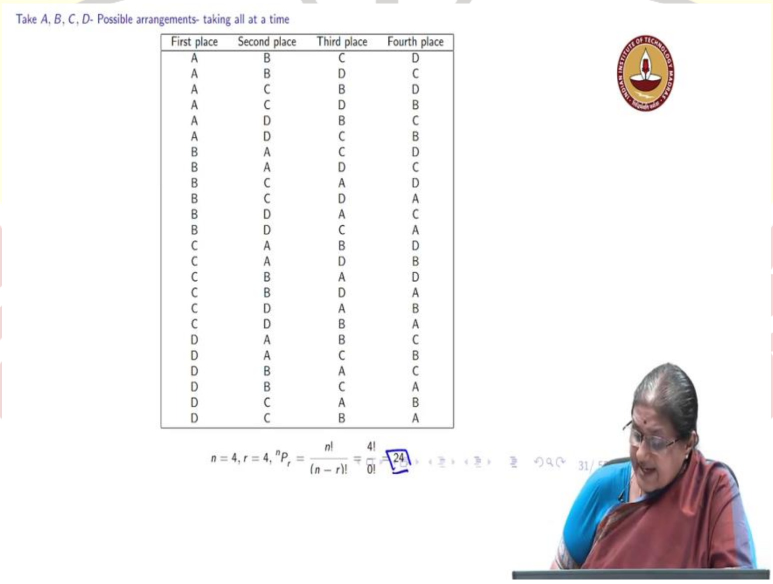
(Refer Slide Time: 19:55)



First place	Second place	Third place	Fourth place
A	B	C	D
A	B	D	C
A	C	B	D
A	C	D	B
A	D	B	C
A	D	C	B
B	A	C	D
B	A	D	C
B	C	A	D
B	C	D	A
B	D	A	C
B	D	C	A
C	A	B	D
C	A	D	B
C	B	A	D
C	B	D	A
C	D	A	B
C	D	B	A
D	A	B	C
D	A	C	B
D	B	A	C
D	B	C	A
D	C	A	B
D	C	B	A

In the third example, I had all possible arrangements of ABCD taking all at a time. So here, my n is equal to 4, my r equal to 4. It is again nPr, which is the same as nPn, which is equal to n factorial, which I know here n equal to 4, n factorial is 4 factorial, which is 24. And I have 24 possible arrangements here.

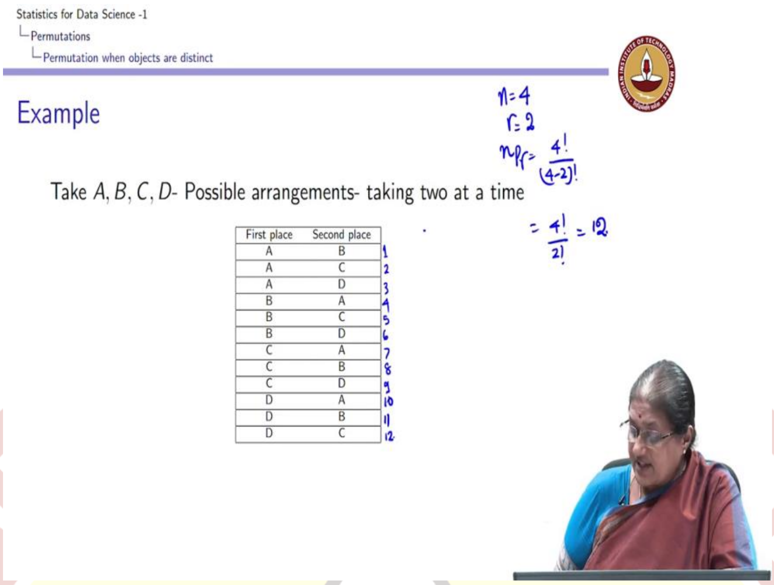
(Refer Slide Time: 20: 35)



First place	Second place	Third place	Fourth place
A	B	C	D
A	B	D	C
A	C	B	D
A	C	D	B
A	D	B	C
A	D	C	B
B	A	C	D
B	A	D	C
B	C	A	D
B	C	D	A
B	D	A	C
B	D	C	A
C	A	B	D
C	A	D	B
C	B	A	D
C	B	D	A
C	D	A	B
C	D	B	A
D	A	B	C
D	A	C	B
D	B	A	C
D	B	C	A
D	C	A	B
D	C	B	A

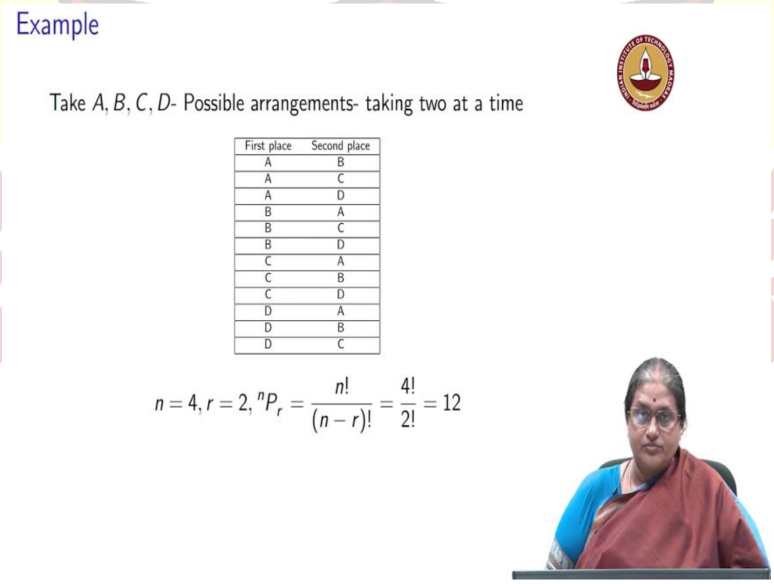
So, you can see that in this case, I have an, I am looking at all at a time, I have 4 factorial, and which is given by 24.

(Refer Slide Time: 20:43)



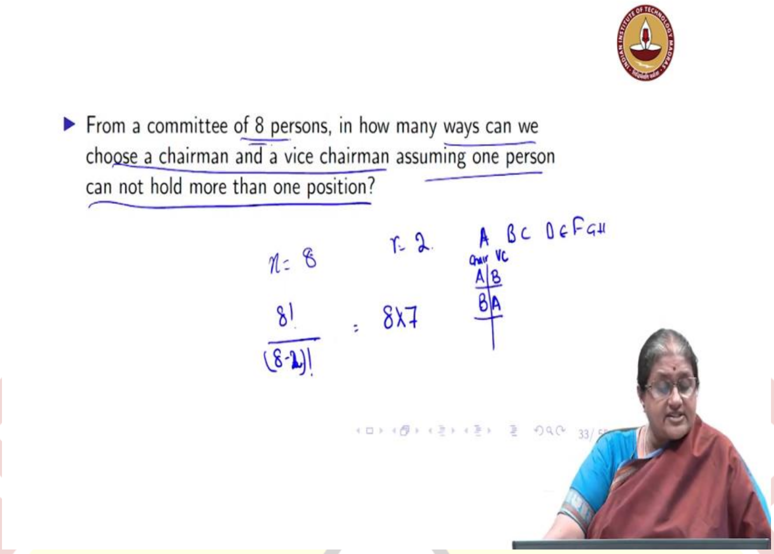
The last example, which we considered was, again ABCD taking 2 at a time again. Again my n is 4, my r is 2, my nPr equal to 4 factorial by 4 minus 2 factorial, which is 4 factorial by 2 factorial, I have 12 as the answer, and I can see that matches with what I have here 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12. I had 12 possible arrangements, which is nothing but what is my nPr in that case, 12.

(Refer Slide Time: 21:24)



So, this is the first permutation formula, which is obtained when I have n objects, n distinct objects, and I am taking r from these n distinct objects, I am not allowing repetition.

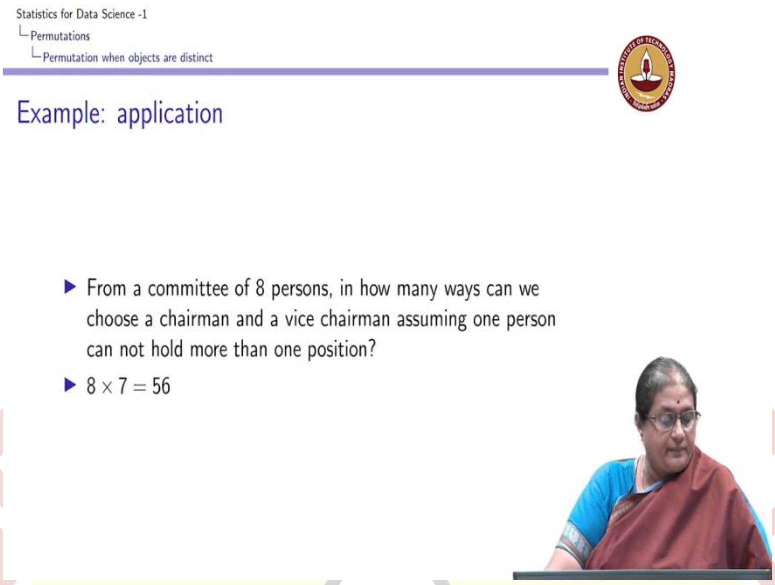
(Refer Slide Time: 21:43)



Let us look at a few applications of this formula. Suppose I want to form a committee of 8 people. The question is, how many ways can we choose a chairman and a vice chairman, assuming 1 person cannot hold more than 1 position? So, the first part, which we notice is 1 person cannot hold more than 1 position reflects the non-repetitive behaviour of our, whatever requirement which we have.

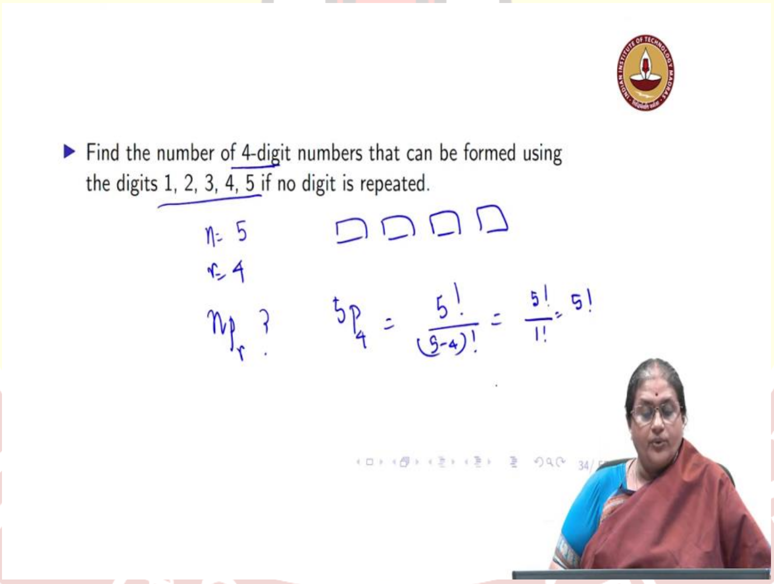
From a committee of n (persons), 8 persons, so this n equal to 8. I want to see how many ways can we choose a chairman and vice chairman. So, the r equal to 2, how many ways that is also something which I can, this one, so I need to know whether a chairman, whether if I have A, B, C, D, E, F, G, H, these are the 8 people. A could be a chairman, B could be vice chairman, B could be chairman, A could be vice chairman. So, these 2 orders are different. So, this is a chairperson, this is a vice chairperson. So, these 2 orders are different. So, applying the formula, I have 8 factorial by 8 minus 2 factorial, which is 8 into 7.

(Refer Slide Time: 23:27)



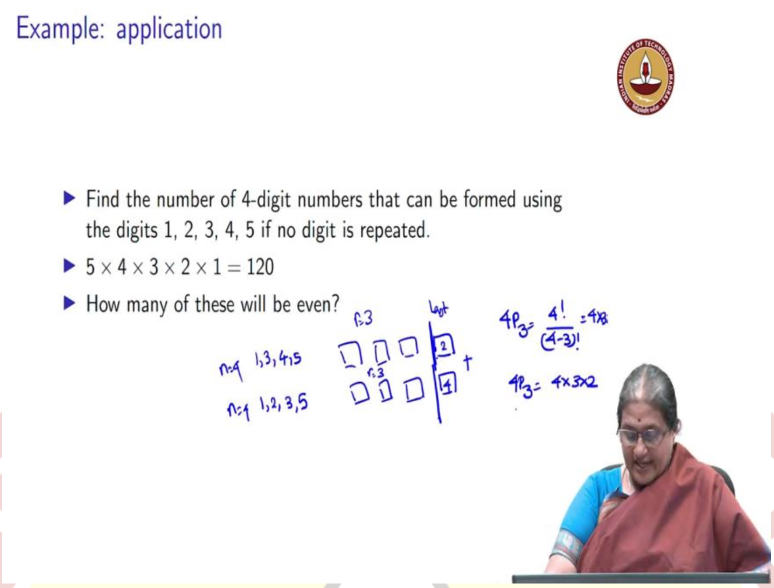
So, the total number of ways I can do this is 8 into 7, which is 56 ways.

(Refer Slide Time: 23:33)



Now, let us look at the next application. In the next application, I am looking at the number of 4-digit numbers that can be formed using digits 1, 2, 3, 4, 5. Again, no digit is repeated. So, my n in this case is 5, I am choosing from 5 distinct objects, I am looking at number of 4-digit numbers, so I have four 4 blanks. So, my r equal to 4. And I am asking the question, what is nPr again? It is 5p4, which is nothing but 5 factorial to 5 minus four factorial, which is same as 5 factorial by 1 factorial, which is 5 factorial.

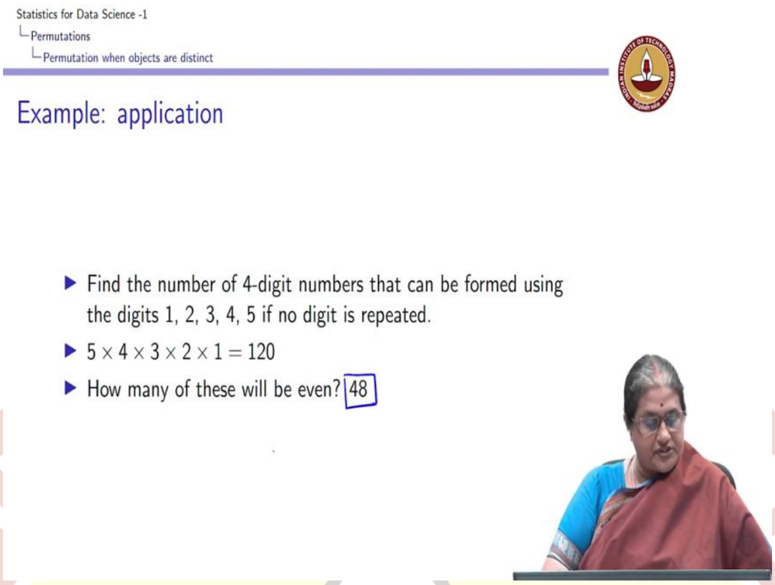
(Refer Slide Time: 24:29)



And I know that 5 factorial is equal to my 5 into 4 into 3 into 2 into 1, so I have a result, which is 120 ways of creating these 4-digit numbers or 4-digit codes, whatever we want. Let us look at an example. Let us slightly modify the application, which we looked. So earlier, I said find the number of 4-digit numbers, which can be formed using these digits. But now suppose I am saying that I want to know how many even numbers can be formed. I know if I want to count the number of even numbers, the last thing can either be a 2 or a 4. I am fixing the last digit here. So, if I fix the last digit, I have 3 blank spaces that are available.

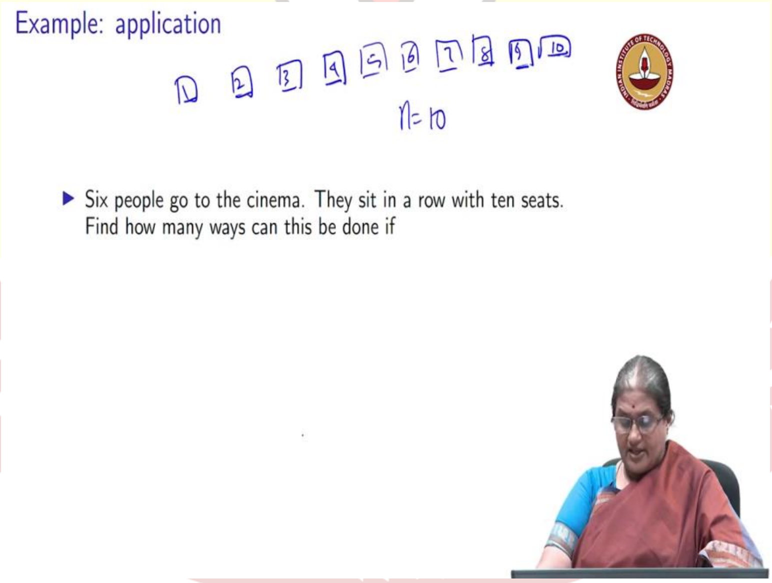
Now, once I fixed the last digit, these, now my r becomes 3, and I can choose from the remaining. So, if it is 2, and r is 3, I have only remaining 1, 3, 4, and 5, from which I can choose the remaining 3 numbers. And that can be done in 4p3 ways, which is 4 factorial by 4 minus 3 factorial, which is 4 into 3 into 2 ways. Similarly, if I fix a 4, I have again r equal to 3 blanks, and I have 1, 2, 3 and 5, my n equal to 4. Again, I can choose it in 4p3 ways, which is again 4 into 3 into 2 ways.

(Refer Slide Time: 26:28)



So, if I add these 2 up, I see that the total number of ways is 24 plus 24, which is 48. So, you can see that we can use this simple formula of permutation to count the number of arrangements that can arise in various situations.

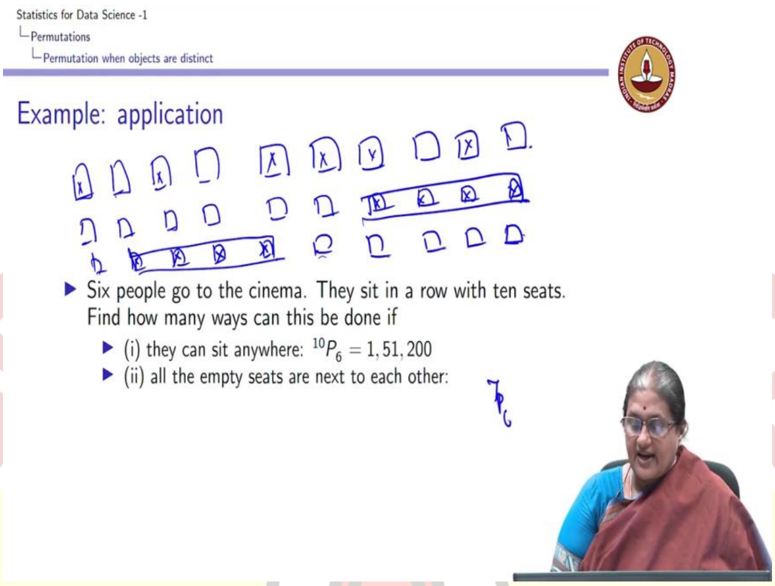
(Refer Slide Time: 26:54)

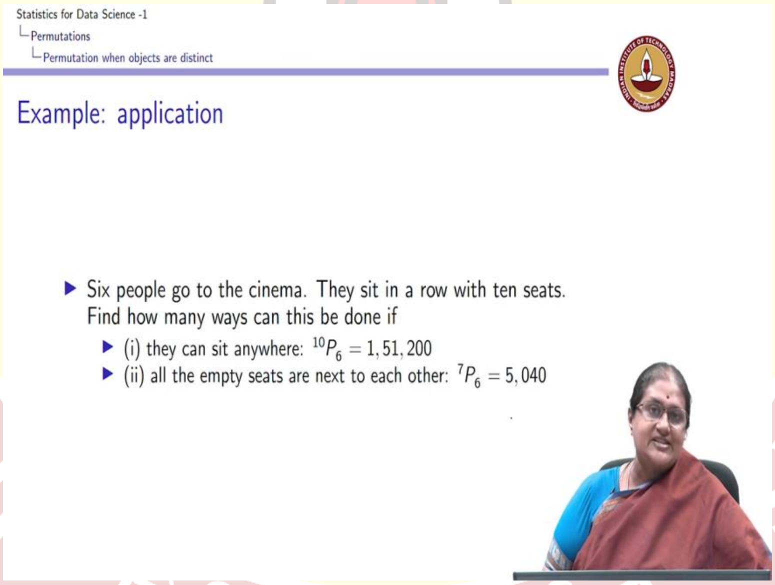


Moving forward, let us look at what would happen, another example, let us look at another application. Suppose I have 6 people who have gone to a cinema. In cinemas again, lets again remember that what we are talking about here is arranging people in a line, in a linear order. So again, if you look at cinema or stadium, typically people sit in a row. So, I go to a cinema, I have 10 seats. So, these are the 10 seats, I can number my seats as 1, 2, 3, 4, 5, 6, 7, 8, 9 and

10. These are the 10 seats that are available to me. So, 6 people go to a cinema. So, my n equal to 10. 6 people go to a cinema.

(Refer Slide Time: 27:56)





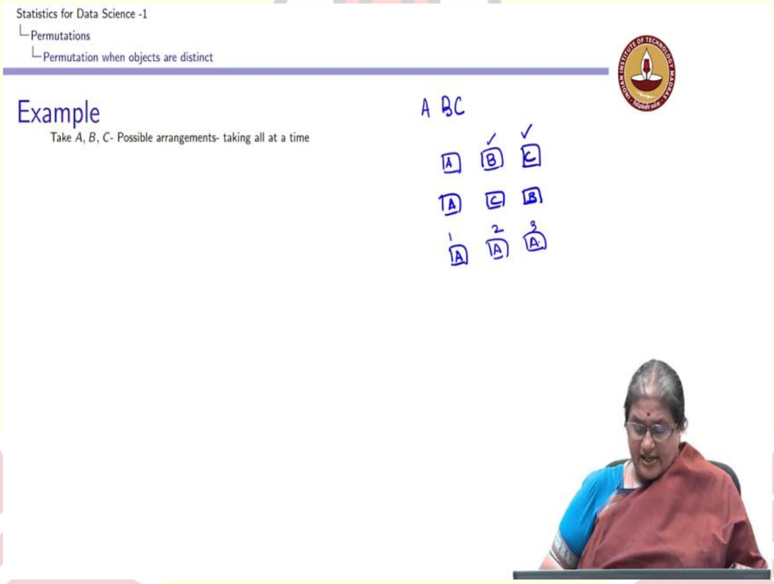
So, the question, the first question we are asking here is, how can I makes the 6 people sit in this row? I have 10 available seats, how many ways can they do if my people can sit anywhere? So, if these 6 people can sit anywhere, I know that n equal to 10, my r equal to 6 and number of arrangements applying my formula is 10p6.mSo, the first answer is, well if I can sit anywhere, the answer is 10p6, these are the number of ways 6 people can sit in available 10 seats.

But now, suppose I am giving another thing is all the empty seats are next to each other. How do we solve this problem? For example, I cannot have a situation of this kind. I again have the 10 seats, I cannot have a person occupying this seat, this seat 1, 2, 3, 4, 5, 6. So, these empty seats are not next to each other, whereas this is permitted. So, this is permitted for me. I can have all the 4 empty seats next to each other, or even this is permitted.

What we need to understand here is these four empty seats form a block. And I can consider these 4 empty seats as a distinct object in it, because I am not allowing them to move. Now, if these 4 empty seats are considered as 1 distinct object, then the total number of distinct objects are 1, 2, 3, 4, 5, 6 and 7. I need to see how many arrangements can be done, that is this. So now, I have 7 places available to me. I want to set these 6 people in these 7 places, and that can be done in 7p6 ways.

So, these are application of this simple, these are applications of the simple formula of arranging r objects out of n distinct objects when repetition is not allowed.

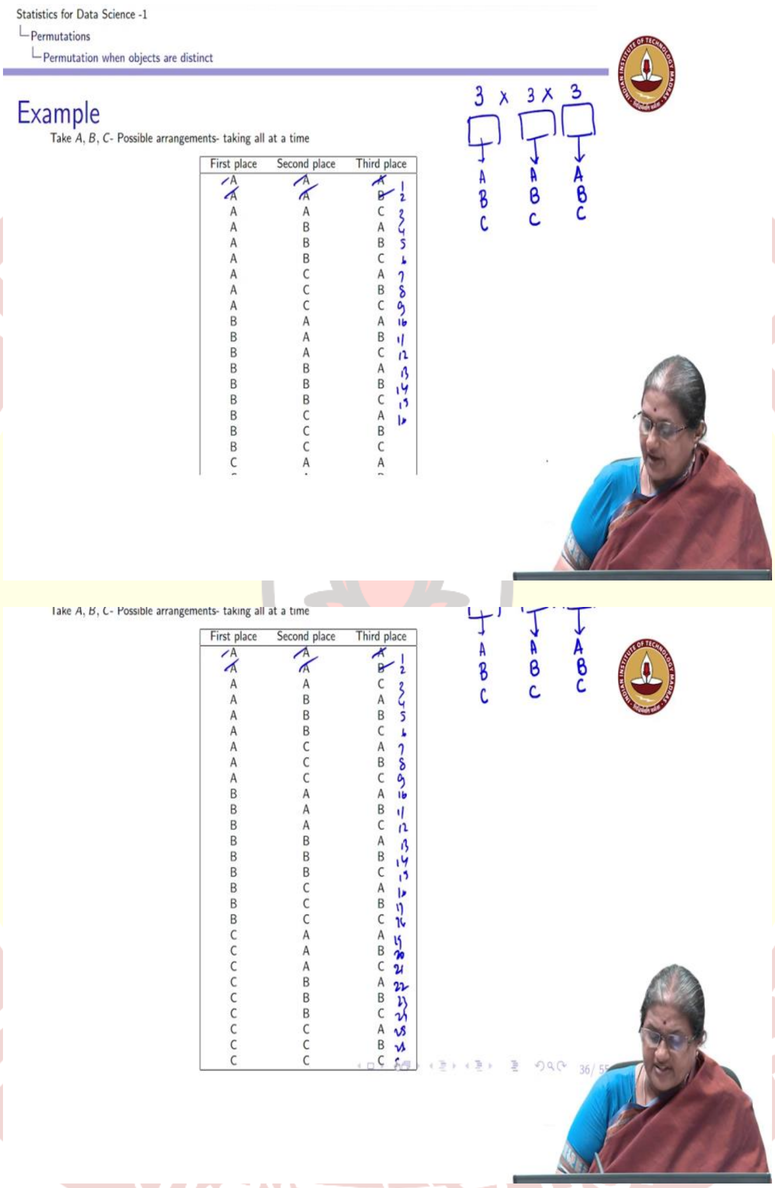
(Refer Slide Time: 30:57)



Now, suppose, let us go to the next case, where I want to know about what is that number of permutations or where a number of arrangements when the objects are not distinct, when I am looking at repetitions, so when I am looking at the same (example). So now, let us move forward to understand what is the number of possible arrangements again taking all at a time, but now I am allowing repetition. So, let us look at the first example. I looked at ABC.

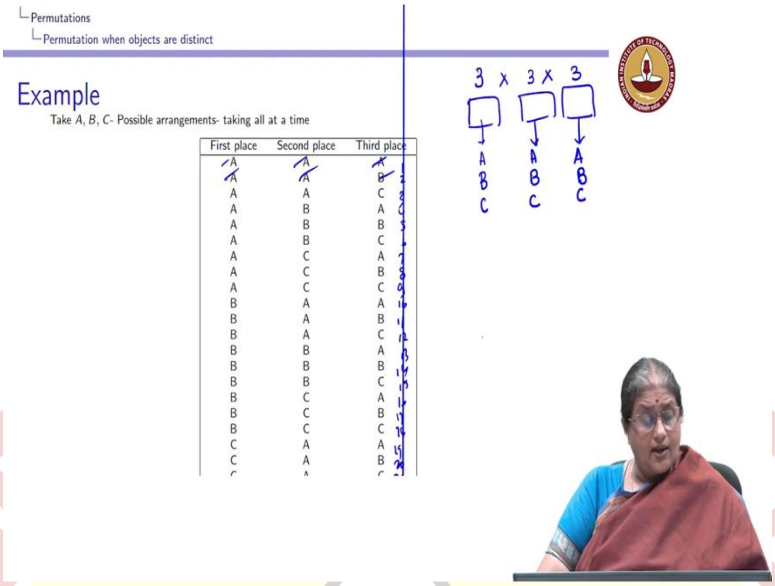
Now, I did not allow repetition earlier, so when I looked at the 3 boxes to be filled, once I fill in A in one of the boxes, I do not have A available to fill in B or C. So, only either B or C could get into the second box. So, I had a A, so I had a A with BC or CB. But suppose I am allowing repetitions, so if I fill A in my first box, A is available for the second box, A is available for the third box also. So, AAA is a valid arrangement, because I am allowing repetition. I am not saying that once I have put a particular thing in a box, that is not available for filling up the other 2 boxes.

(Refer Slide Time: 32:45)



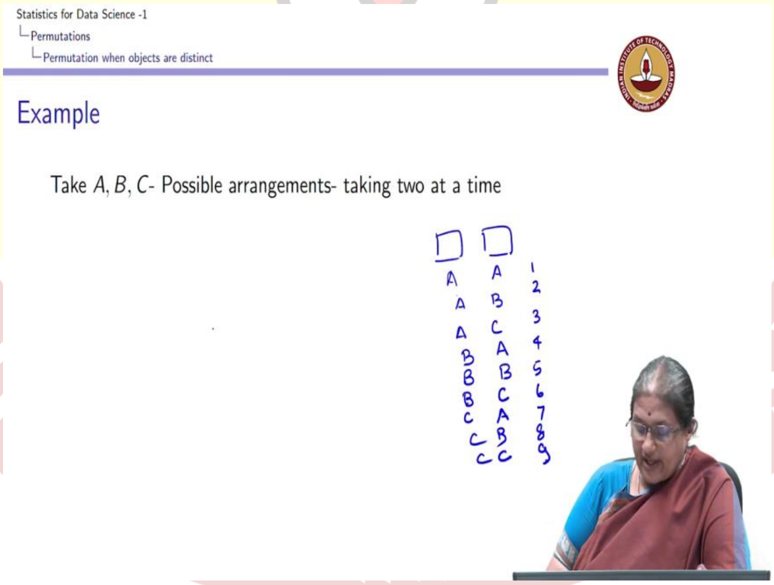
So, if you look at this case, so I am looking at the number of possible arrangements. So, AAA is a possible arrangement, AA with B is a possible arrangement. So, the number of possible arrangements, now if any, the first box can be filled with A or B or C, the second box can be filled with A or B or C, the third box can be filled with a A or B or C. So, there are 3 choices available to fill this box A, there are 3 choices available to fill the box, second box, there are 3 choices available to fill the third box, because all the choices are available for all the boxes.

(Refer Slide Time: 34:17)



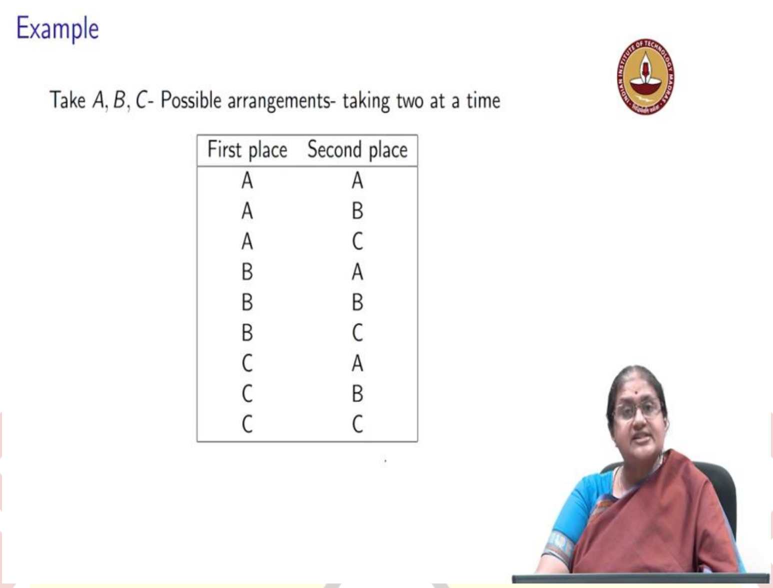
So, the total number of ways I can fill all the 3 boxes together are 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, and you can see that, that goes all the way up to, and you can see that, that 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26 and 27, I have up to 27.

(Refer Slide Time: 34:27)



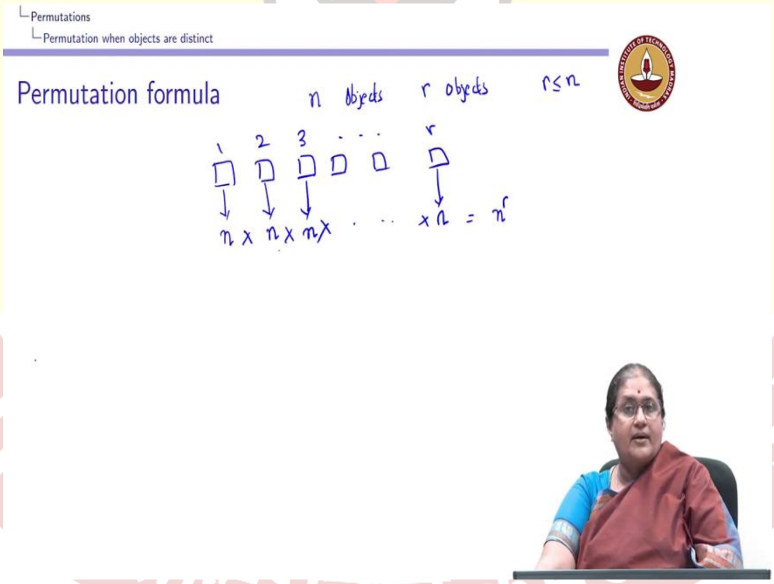
So, let us look at the second example. Now, suppose I am looking at the same possible arrangements but now I am taking 2 at a time. So, I have 2 boxes. The first box can be an A, the second box can be an A, so AA. First is an A, second can be a B. First is an A, second can be a C. Similarly, first is a B, second can be an A. First as a B, second can be a B. First is a B, I have a BC. C goes with an A, B or C, I am allowing repetition, so you can see the total number of possible arrangements possible in this case is 9.

(Refer Slide Time: 35:14)



So, we can actually come up with a formula to, so I have 9, if I take 3 objects, 2 at a time, I allow for repetitions, I can see that the total number of possible arrangements is 9 here.

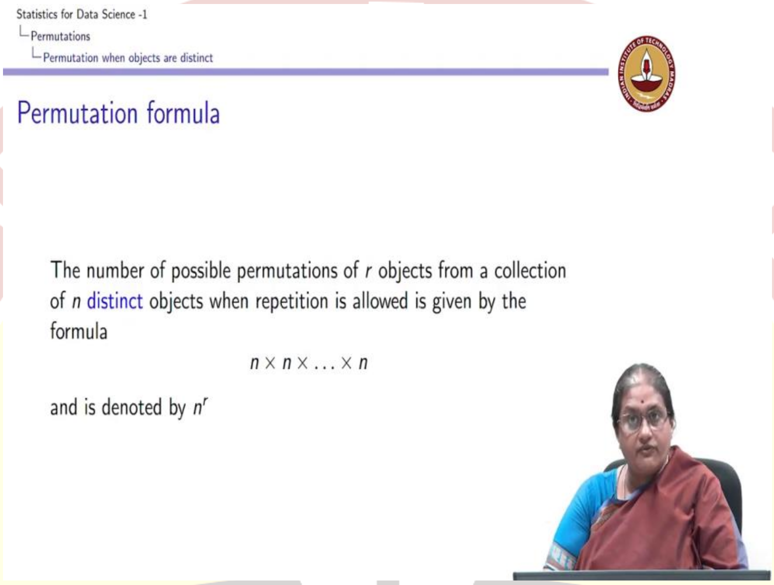
(Refer Slide Time: 35:29)



So, let us look at what is the permutation formula. So, I have in this case, n, again I have n distinct objects. I am taking again r objects out of this n distinct objects, r is less than or equal to n. So, how do I, I am allowing repetitions, so if I have r blocks or r blanks, this is my first, this is my second, this is my third and this is my rth block. The first block can be filled with any of the n objects. The second can also be filled with any of the n objects. Third can also be filled with any of the n objects. The rth block can also be filled with any of the n objects.

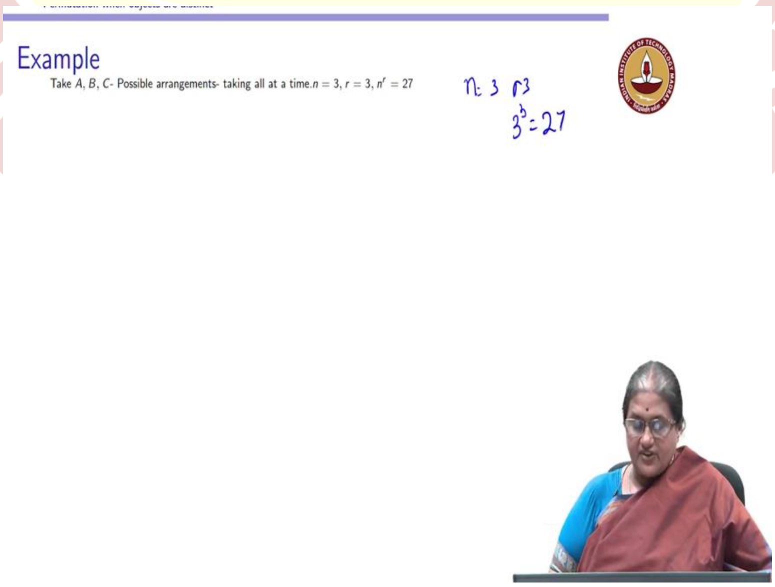
Again, I apply the multiplication rule of counting. If I apply the multiplication rule of counting, the total number of ways I can fill all these r boxes is nothing but n into n into n r times, which is n to the power of r. So, we can see that this is, this second rule of permutation, where I again have n distinct objects, I am choosing r distinct objects from this n distinct objects, but I am allowing repetition.

(Refer Slide Time: 36:59)



So, how do, what is the formula there? So, if I have r objects from n distinct objects, repetition is allowed. It is given by the formula, the formula is given here, it is n to the power of r, this is what we have seen.

(Refer Slide Time: 37:18)

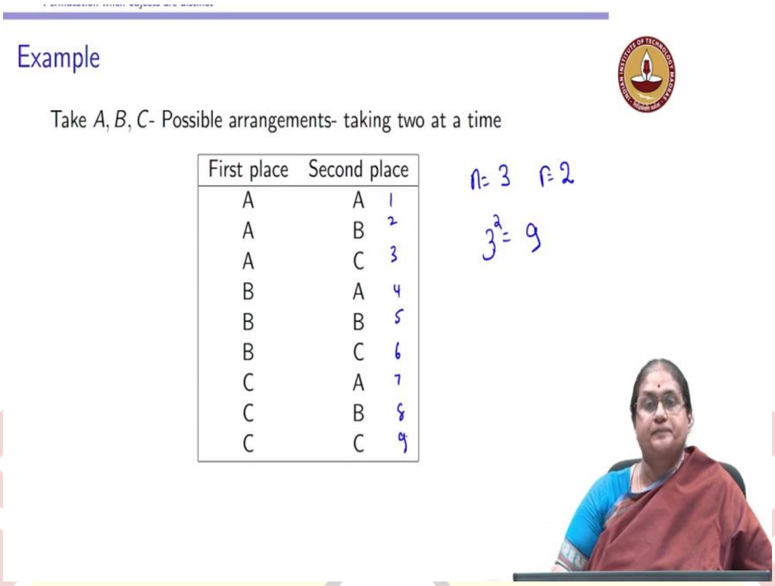


So, in the earlier example, when I had an A,B,C and I am taking all at a time, I had n equal to 3, r is also equal to 3. 3 to the power of 3 is 27.

(Refer Slide Time: 37:31)

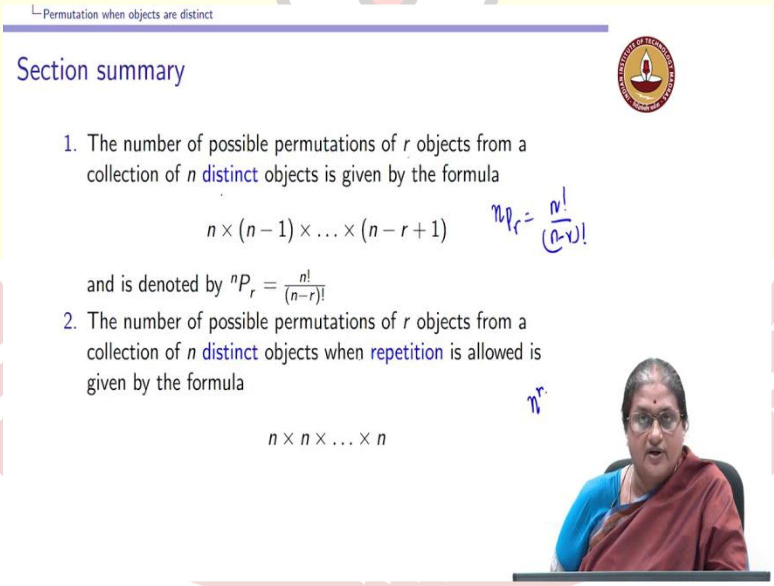
And we saw that we had the 27 arrangements, which were possible.

(Refer Slide Time: 37:37)



Similarly, when I looked at this case taking 2 at a time, my n equal to 3, r equal to 2, n to the power of r is 3 to the power of 2, which is 9. And we have seen that there are 9 distinct arrangements possible in this case.

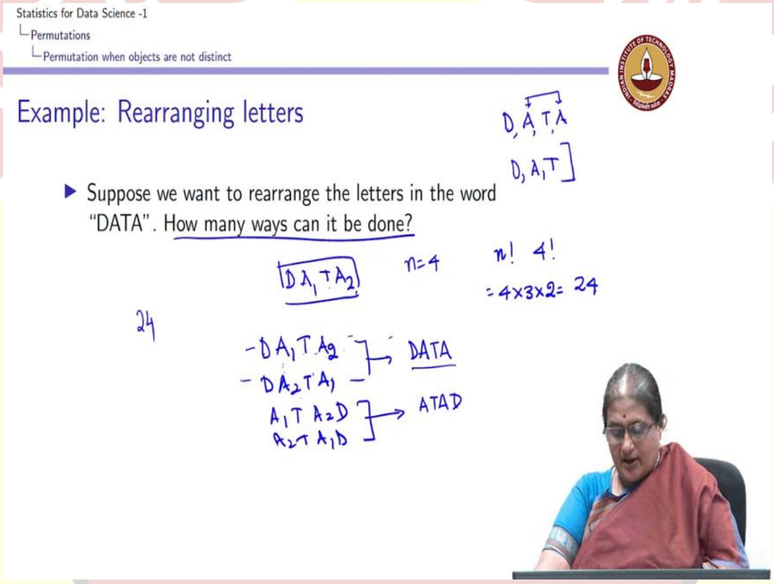
(Refer Slide Time: 38:01)



So, let us look at a summary of what we have learned so far in this topic. So, we started with permutations when objects are distinct. We again look at 2 cases in this case. One is when repetitions are not allowed. And the second when repetitions are allowed. When repetitions are not allowed, the number of possible permutations of r objects from a collection of n objects is given by and denoted by the formula nPr, which is n factorial by n minus r factorial. When repetitions are allowed, it is given by n to the power of r. This is what we have learned so far.

Statistics for Data Science – 1
Professor Usha Mohan
Department of Management Studies
Indian Institute of Technology, Madras
Lecture No. 5.4
Permutations and Combinations – Permutations: Objects not distinct and Circular permutations

(Refer Slide Time: 0:14)



Now, let us look at the case when the objects are not distinct. So, when we looked at the case when the objects were distinct, we looked at different ways of arranging A, B, C. Now, these A, B, C could be people, it could be objects, but they are distinct. But now suppose I have a case where they are not distinct. So, let us look at an example where I am looking at how do I rearrange the letters in the word "DATA". Now, if you look at the word DATA and just look at this as alphabets, I see that D, A and T are the 3 distinct objects or distinct alphabets, because A appears twice.

Now, I am asking how many ways can it be done. To understand this, let us do the following, I will write the DATA, I will write as A_1 and A_2. Basically, A_1 and A_2 are the same alphabet, but I know how to find out for this, I already know that if I have DA_1TA_2, my $n = 4$, I am taking all at a time so this is going to be $n!$, this is what we have already seen because it is taking all at a time is $n!$. So, I can have $n!$ which is 4! which is again $4 \times 3 \times 2$ which is 24 ways of arranging DA_1TA_2

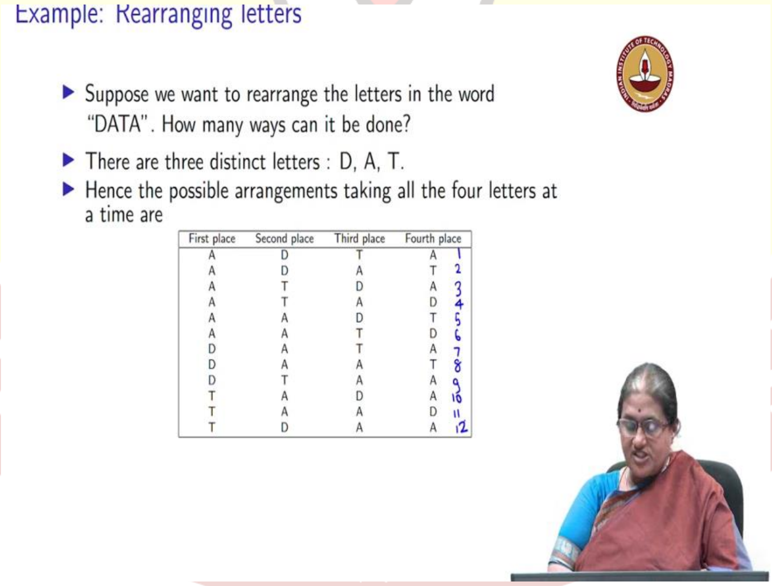
.

Let us look at these two arrangements DA_1TA_2 and DA_2TA_1

. Now these two arrangements are different when I consider A_1 and A_2as distinct objects. But we know that A_1 or $A_2$are essentially the same object. So, these two together will just give me one arrangement which is DATA, whether the first A comes before the second A or the second A comes before this first A is irrelevant. So, DA_1TA_2 and DA_2TA_1 give me the same arrangement which is DATA. Similarly, A_1TA_2D and A_2DA_1D these two together will again give me the same arrangement which is ATAD.

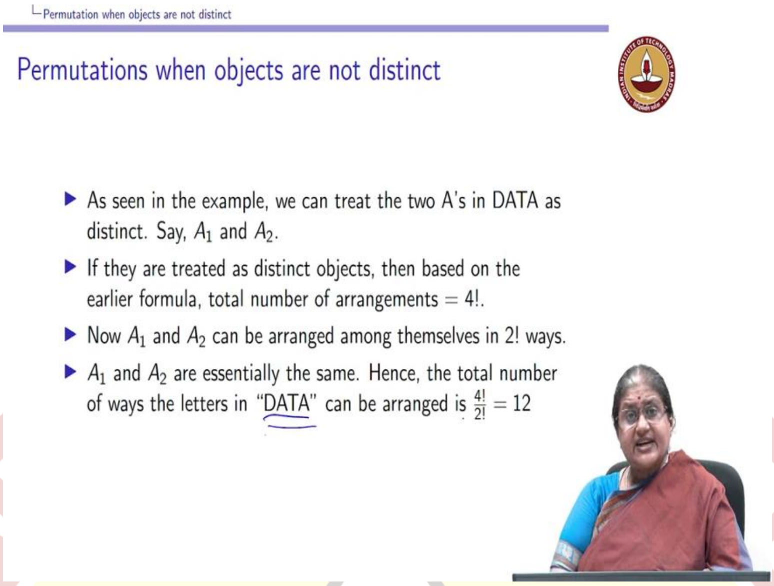
So, for every two arrangements which I had in my original, original which is these 24 arrangements, I can see that every 2 arrangements will give me 1 arrangement that is, whenever I have A_1 and A_2, A_2 and A_1, that which can I have 2! ways because when I have 2 elements, I have 2! ways. So, I can erase that for every 2 I have 1 arrangement distinct because these two are the same. In other words, out of these 24 arrangements, I can eliminate half of them so I get 12 arrangements when I have one alphabet which is appearing twice.

(Refer Slide Time: 4:01)



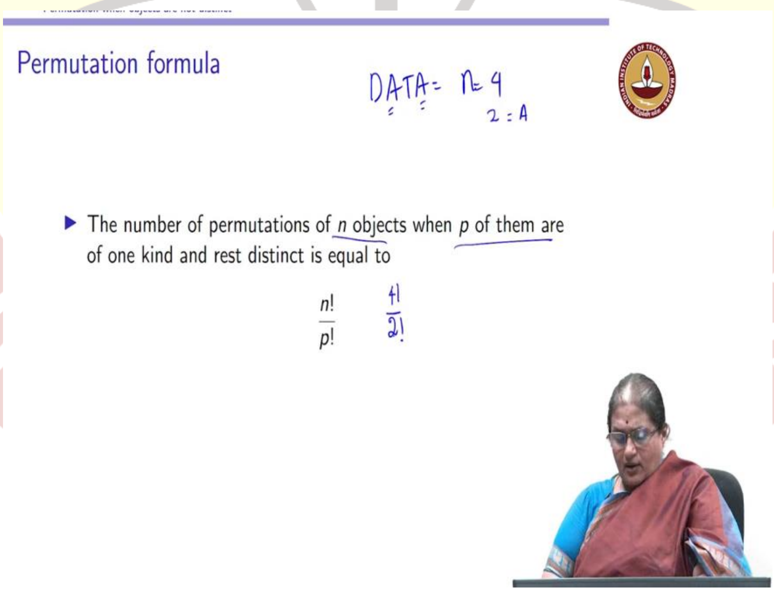
In other words, if I look at the 3 distinct alphabets, I have the possible arrangements that are possible are $1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11$ and 12 and how we have arrived at these 12 arrangements is what we had just discussed.

(Refer Slide Time: 4:26)



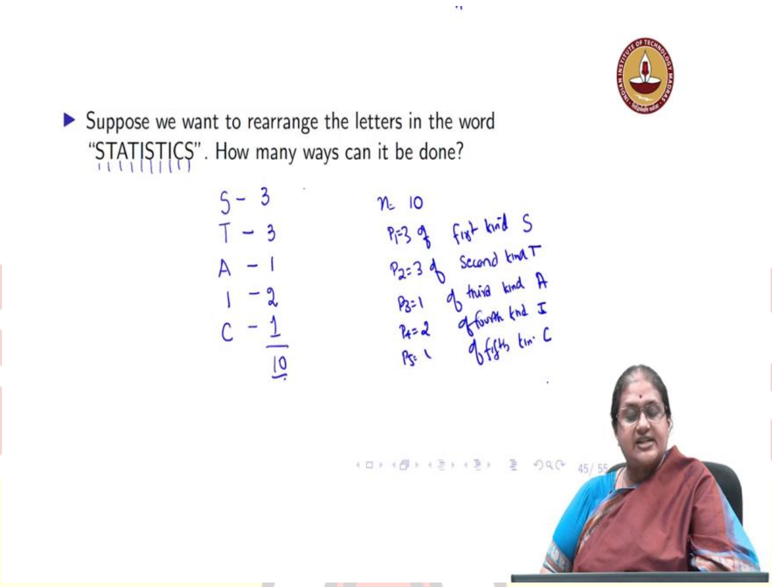
And the logic is I treat both the A's as A_1 and A_2. I looked them as distinct objects in which case I know I have 4! ways to do it, but I know effectively A_1 and A_2 are the same object hence I divide the total number of arrangements which is 4! by 2! and I get that the total number of ways the letters in DATA can be arranged is $\frac{4!}{2!}$ which is 12 ways.

(Refer Slide Time: 5:06)



I can generalise this formula to say that the number of permutations of n objects when p of them are of one kind, so in DATA, my n was 4 so I can generalise this into what I refer to as the permutation formula. The number of permutations of n objects when p of them are of one kind and rest distinct. So, what are the p of them here? $n = 4$, because I have 4 objects but I have p of them, I have 2 of them A which are of one kind and other at distinct equal to $\frac{4!}{2!}$ and I get the answer to be 12.

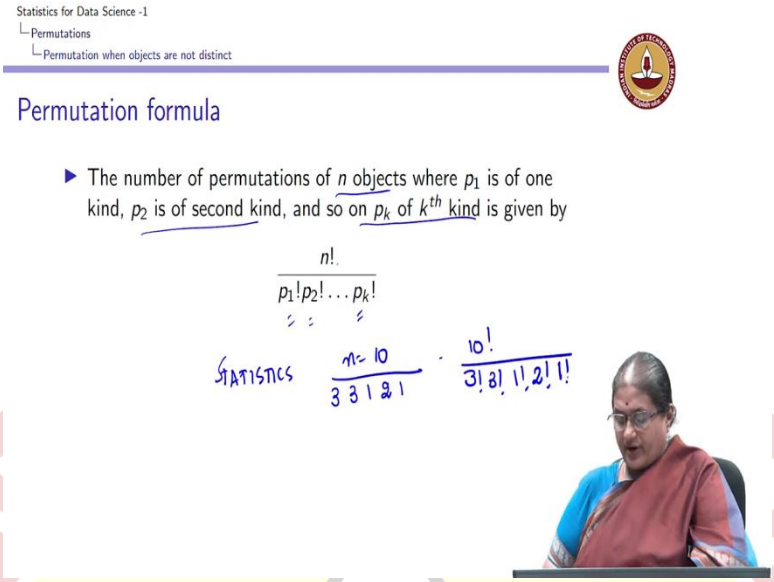
(Refer Slide Time: 5:58)

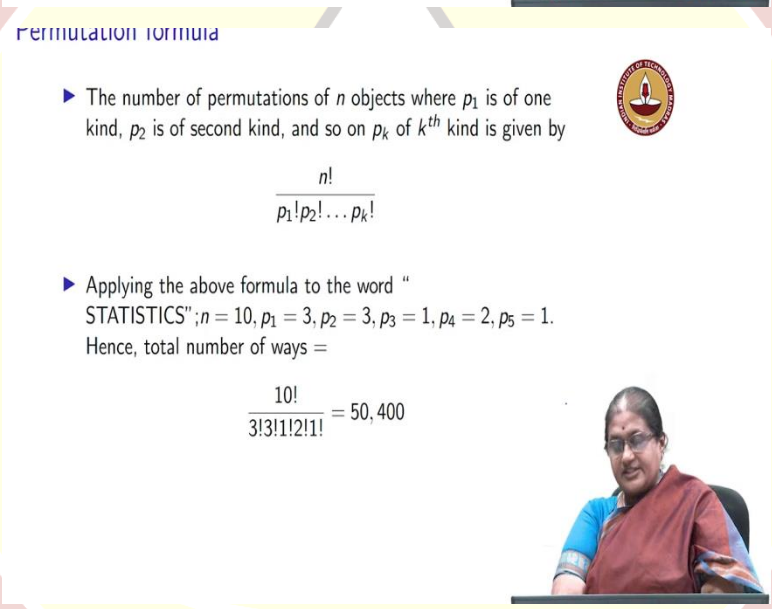


Now, let us extend this idea. Let us look at the same idea and say that now suppose I want to rearrange the letters in the word STATISTICS. How many ways can it be done? Now, if you look at this STATISTICS, I look at now the distinct alphabets in this word STATISTICS. Now if you look at the distinct alphabets, I have a S, I have a T, I have a A, I have an I and I have a C. The total number of alphabets in this word is $1, 2, 3, 4, 5, 6, 7, 8, 9, 10$. Now, of $n = 10$, I have 3 S, I have 3 T's. I have a 1 A, I have a 2 I and I have a 1 C. So, this adds up to 10.

In other words, of this $n = 10$ alphabets, I have say 3, I have 3 S, so I have p_1 of one kind, the first kind which is S. So, $p_1 = 3$ of the first kind, $p_2 = 3$ of second kind. What is my first kind? My first kind is S, my second kind if T. I have $p_3 = 1$ of third kind, my third kind is a 1, $p_4 = 2$ of fourth kind which is an I, this is A, this is I and $p_5 = 1$ of fifth kind which is C. So, how many possible arrangements can be done in this case?

(Refer Slide Time: 8:06)



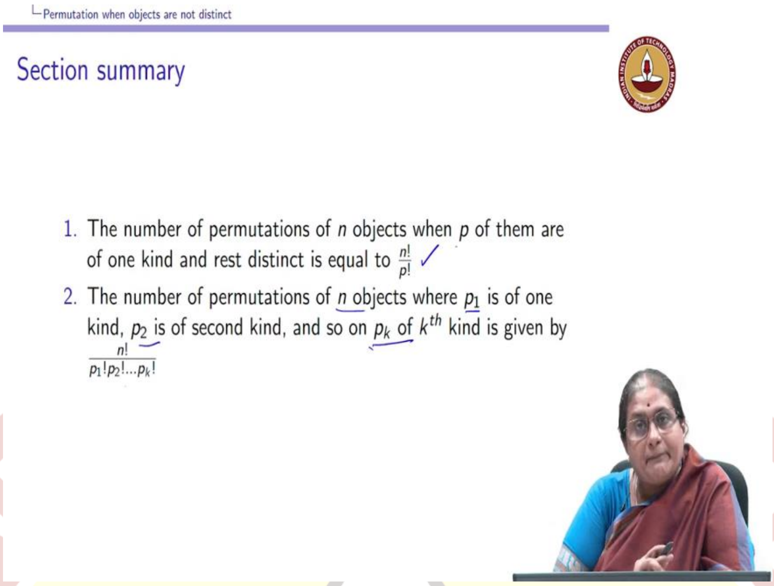


The formula that governs this of a thing which is just an extension of what we have seen before is that the number of permutations when of n objects, when p_1 is of one kind, p_2 of second kind and p_k of third kth kind is $\frac{n!}{p_1!p_2!\dots p_3!}$

In my STATISTICS example, we have already seen my $n = 10, p_1 = 3$, so I have, $p_2 = 3$, $p_3 = 1$, then I had a 2, $p_5 = 1$, so my total number of permutations is 3!, 3!, 1!, 2!and 1! is the total number of things in this STATISTICS thing and you can see that, that is equal to, that value is equal to 50400.

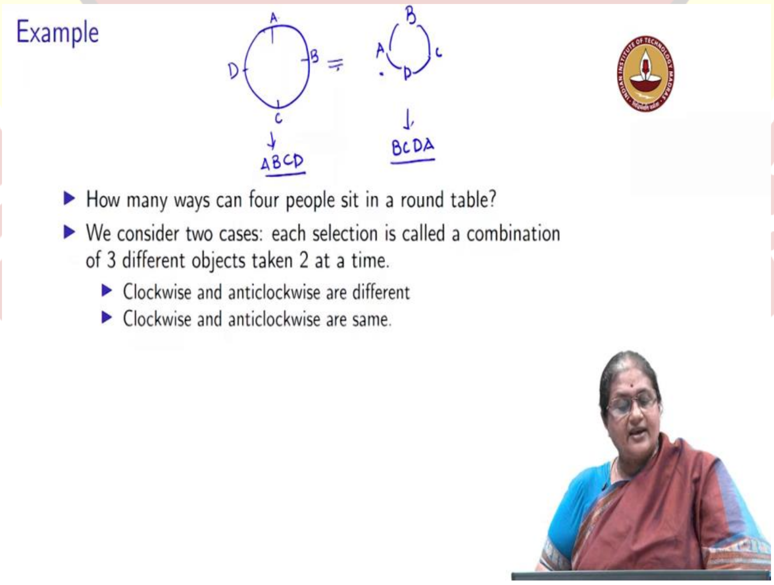
So, if you apply the formula, the total number of arrangements that are possible out of this word STATISTICS is 50400. So, we have applied this when objects are not distinct formula to the STATISTICS example.

(Refer Slide Time: 9:39)



So, what we have learned so far is when I am looking at linear ordering that is by linear ordering, I am arranging objects or things or people in a row, then the number of permutations of n objects when p of them are of one kind and the rest distinct = $\frac{n!}{p!}$ The number of objects when p_1 is of one kind, p_2 of second kind, p_k of kth kind is $\frac{n!}{p_1!p_2!\ldots p_k!}$

(Refer Slide Time: 10:17)



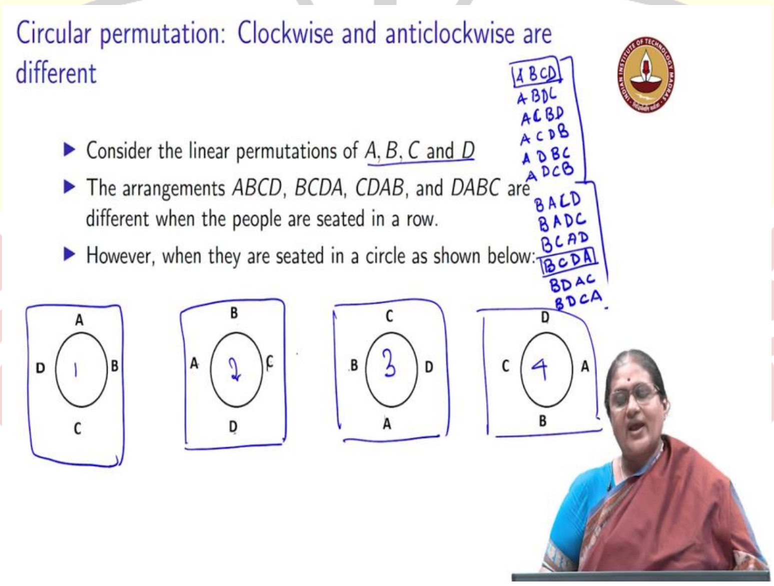
So, so far what we have seen is how do we count the number of arrangements when I am arranging n distinct or n objects, it could be distinct or not distinct objects in a linear row or in a linear fashion. But sometimes we might we might be interested in knowing about the number of ways we can arrange people in a nonlinear and the most common thing is if I have to arrange people in a circle. Then I am interested in knowing how many ways can I do it?

For example, suppose I have a round table, I am interested in knowing how can I seat 4 people in a round table? So, I have a round table, I want to know how can I seat 4 people. Again, let me assume the 4 people are A, B, C, and D, they are distinct. I want to know how can they sit in this table. Now suppose A occupies this position, I can have one arrangement where B occupies this position, and D occupies this position.

Now, if you look at this arrangement, you can see that B is next to A and next to C. So, B is between A and C. Now, this is same as if B is next to C and A and D, so this arrangement is same as this arrangement. The linear equivalence of this arrangement is A, B, C, D, whereas the linear equivalence of this is B, C, D, A.

A, B, C, D is different from B, C, D, A when we are talking about linear arrangements. But when you are looking at a circular arrangement, this arrangement is the same as this arrangement. So, you can see that it is circular arrangements need to be counted in a different way and that is what we are going to learn now. So, let us look at formulising this notion.

(Refer Slide Time: 12:44)

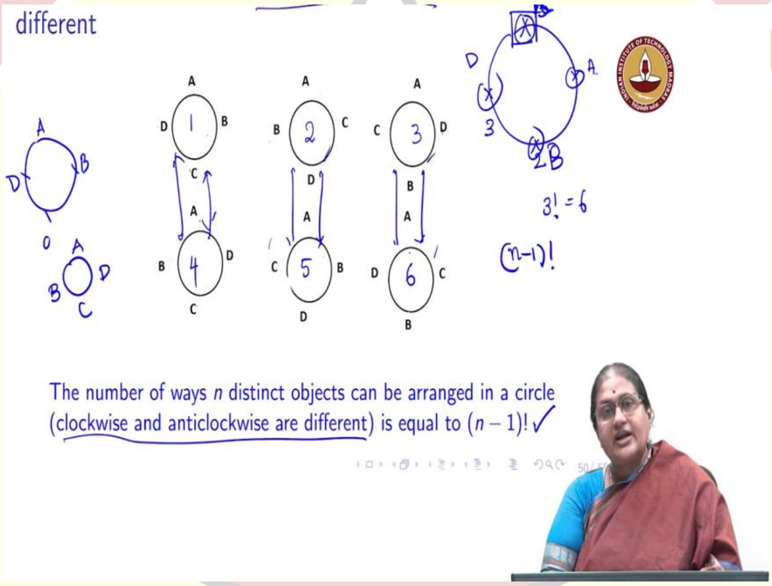


So, I have, first I will consider two cases again when clockwise and anticlockwise are different and when they are the same. So, I said that when you are looking at A, B, C, D, I know the possible suppose I fix A, I know it is a B, C, D. I know A, B, D, C; A, C, B, D; A, C, D, B; A, D, B, C and A, D, C, B. This is already something which we have seen. This has 6 possible linear arrangements if I fix A in one slot.

Now, if I fix B in the first slot, let me look at what are the possible arrangements here if I fix B, I can have a A, C, D again I am not allowing repetitions. B, I can have a B, A, D, C with B, C, A, D; B, C, D, A; B, D, A, C and a B, D, C, A. Now each one of these arrangements were distinct when I looked at a linear ordering. Now, let us look at these two A, B, C, D and let me look at this following B, C, D, A. Now, A, B, D, C corresponds to this circular arrangement. B, C, D, A corresponds to this arrangement.

Similarly, C, D, A, B corresponds to this arrangement and D, A, B, C correspond to this arrangement and we can actually see that these 4 circular arrangements 1, 2, 3 and 4 are the same arrangements.

(Refer Slide Time: 15:11)



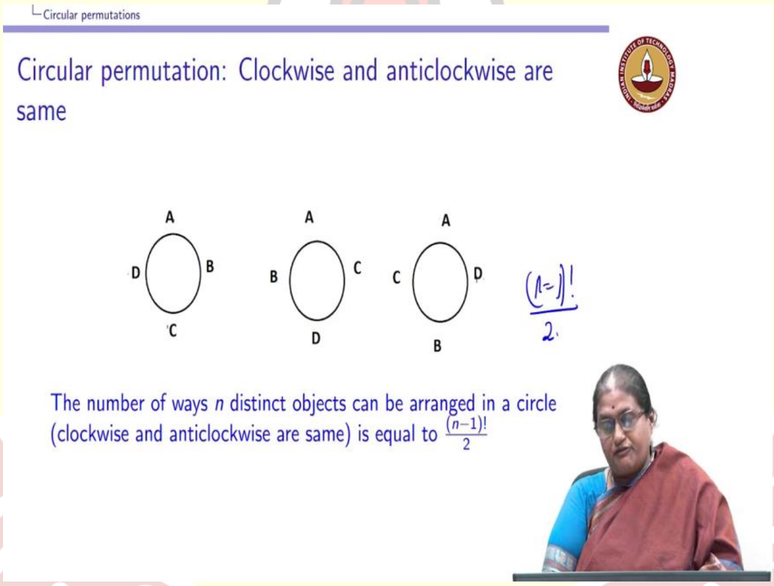
In other words, when we want to compute the circular arrangements, one way to look at it is A, B, D, C, these are the same so the way I can look at it is I fix, so I have a circle. In, within this circle if I fix the first or one position in the circle, the other 3 positions that are available if I fix this position in the circle, the other 3 positions that are available are position 1, 2 and 3 and this can be filled with the 3 available alphabets which are B, C, D and that can be done in 3! ways which is equal to 6.

So, if you look at these arrangements A, B, C, D is different from A, C, D, B is different from A, D, B, C which is different from A, D, C, B, which is different from A, B, D, C which is different from A, C, D, B. So, I have 6 distinct arrangements of people around a table. So, I fixed A. Now, if I fix B, I can do the same thing. So, B, A, C, D is one of the arrangements. But if I look at B, A, C, D, that is the same as A, C, D, B. It is the same arrangement.

So, if I fix C, I have C, A, B, D as one of the arrangements. If I fix C, and look at C, A, B, D, that is the same as A, B, D, C. Hence, whatever I am fixing here is irrelevant but these are the possible distinct permutations. But I am allowing clockwise and anticlockwise are different, so my A, B, C, D is different from A, B, C, D. These are different. So, I have a A, B, D, C which is not same as A, B, D, C, the clockwise. So, the mirror images are different. If you look at it, this is a mirror image of this and this is a mirror image of this. So, they are different.

So, the number of possible circular permutations of n objects if the clockwise and anticlockwise are different is $(n-1)!$ and the reason behind this is I fix one of the objects, the other $n-1$ objects can be arranged among themselves in $n-1$ ways. And this is the same for any of these objects. So, there are n $(n-1)!$ ways of arranging my circular permutations. Now, this is when clockwise and anticlockwise are different.

(Refer Slide Time: 18:42)



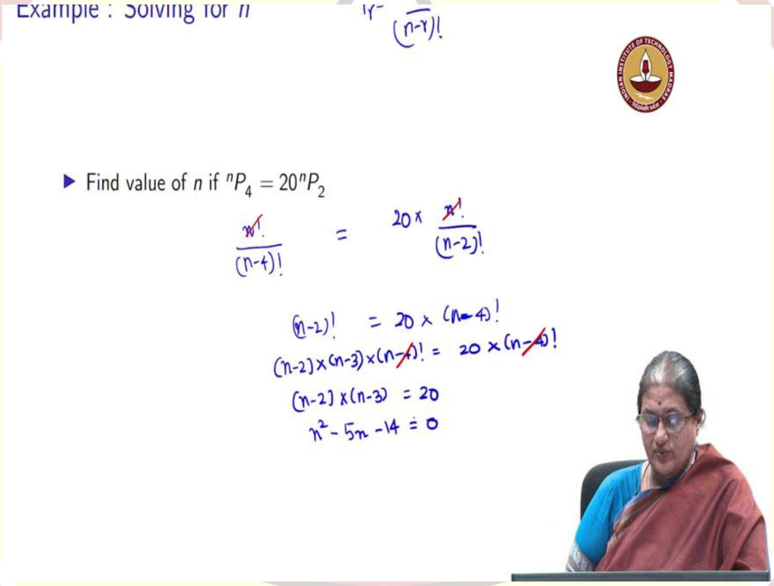
Now, if they are the same, these two are the same because the clockwise and anticlockwise are the same, these two are again the same and these two are again the same which gives me the total number of circular permutations when the clockwise and anticlockwise are same to be nothing but $\frac{(n-1)!}{2}$. So, this is how we compute the number of permutations when the order or when the objects are arranged in a circular fashion.

So, now we move forward. So, we had looked so far at linear transformations, when linear transformations, I mean, so we have looked so far at linear permutations, by linear permutations I mean that I am arranging n objects in a row, I am also looking at choosing r objects from these n objects and arranging them in a linear fashion or in a row.

First we looked at what would happen when the objects were distinct, where repetitions were allowed, repetitions are not allowed, we also looked when the objects were not distinct and then we looked at when I am placing the n objects, distinct objects in a circle, here whether clockwise and anticlockwise are same or different is what is considered.

Very often we would want to actually do some simplifications using the expressions we have gathered so far or we have described so far. Very often we would be looking at using the expressions that we have described so far. What are the expressions?

(Refer Slide Time: 20:30)

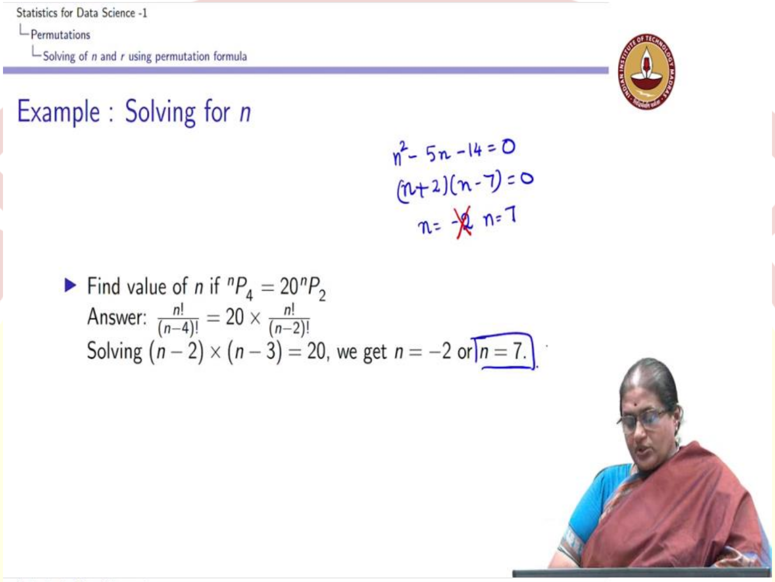


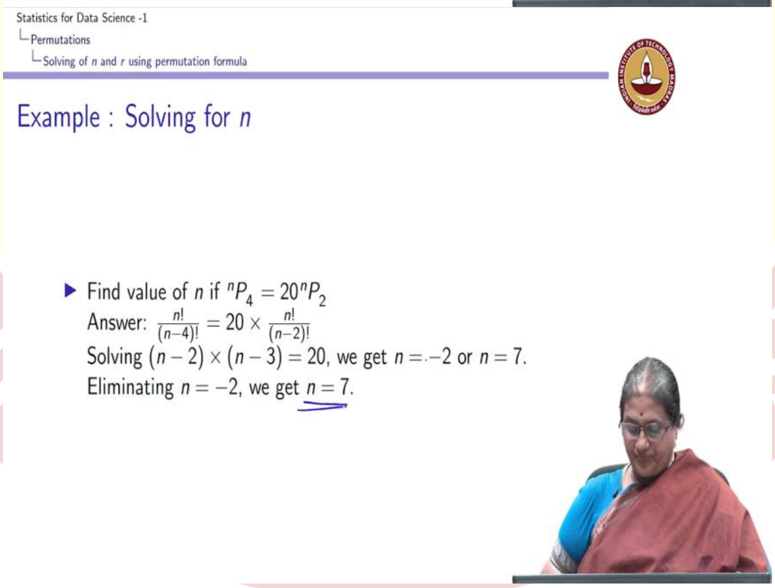
We looked at ${}^nP_r$. ${}^nP_r$ was $\frac{n!}{(n-r)!}$. So, when we are simplifying our notation we might have to solve for the n or r. Let us look at a few examples where we can apply the formula to obtain n or r given either of them that is, obtain n when we are given a expression concerning r or obtain r when we are given a expression concerning n. Let us look at a few examples.

For example, in the example we look at it, I want to find the value of n if ${}^nP_4$ is equal to $20 \times {}^nP_2$. Now the left hand side I have ${}^nP_4$ is $\frac{n!}{(n-4)!}$. This is from my definition. My right hand side I have $20 \times \frac{n!}{(n-2)!}$ and what is given to me is these two are equal and I have to solve for n. Now, if you are cross multiplying it, I can see that I can cancel off $n!$ here and this leaves me with $(n-2)! = 20 \times (n-4)!$.

Now, again applying what we learned from factorial, I know $(n-2)!$ is $(n-2)\times(n-3)\times(n-4)!$, again this comes from our discussion from factorials, this is equal to $20\times(n-4)!$. I can again cancel out these two $(n-4)!$ and now I have what I remain is I have $(n-2)\times(n-3)=20$. Now, this is what I have and I can quickly see that this is I can simplify this to get a quadratic which is $n^2-5n-14$ which is equal to 0.

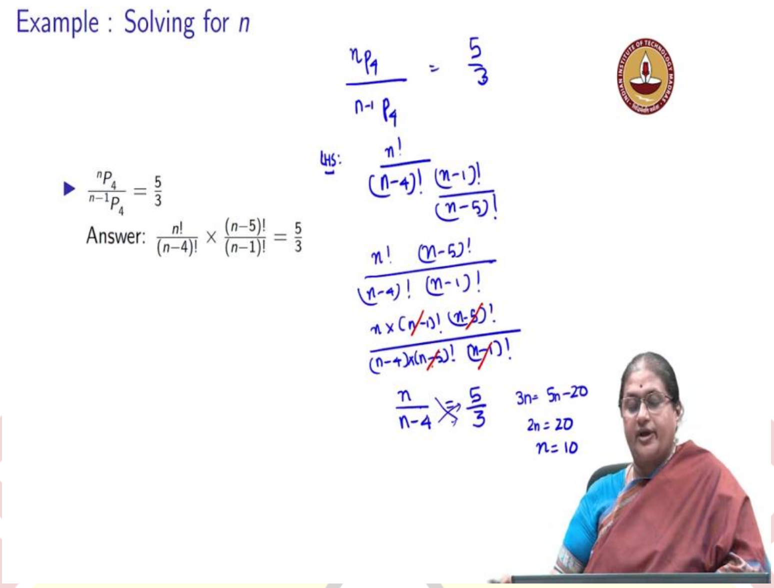
(Refer Slide Time: 23:11)

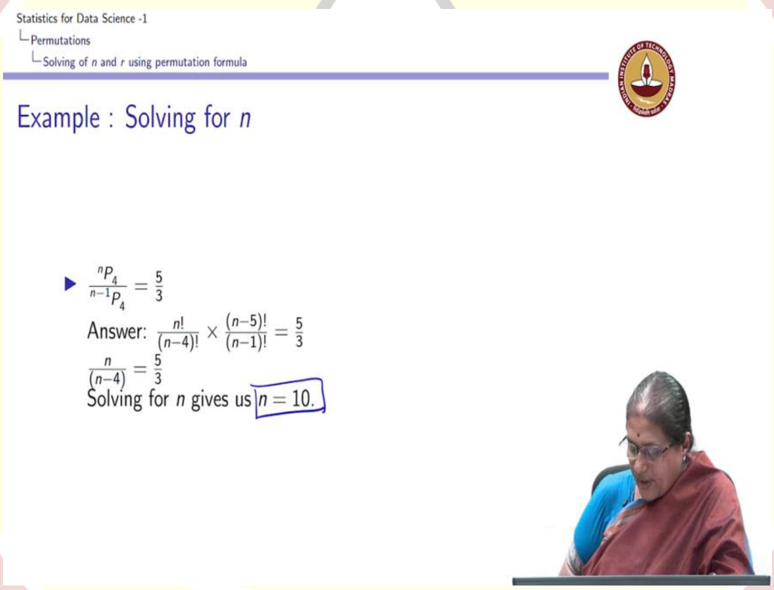




Now, if I solve this quadratic equation, I can so I have the following thing if I solve this quadratic equation, I get that n so I have $n^2-5n-14=0$, n square which I can write as $n-$, $n+2$ and $n-7=0$ factoring which gives me a $n=-2$ or $n=7$ and taking a negative value has to be cancelled out which leaves me with only the choice of $n=7$. So, the solution to this problem is $n=7$ because I am eliminating $n=-2$.

(Refer Slide Time: 24:09)



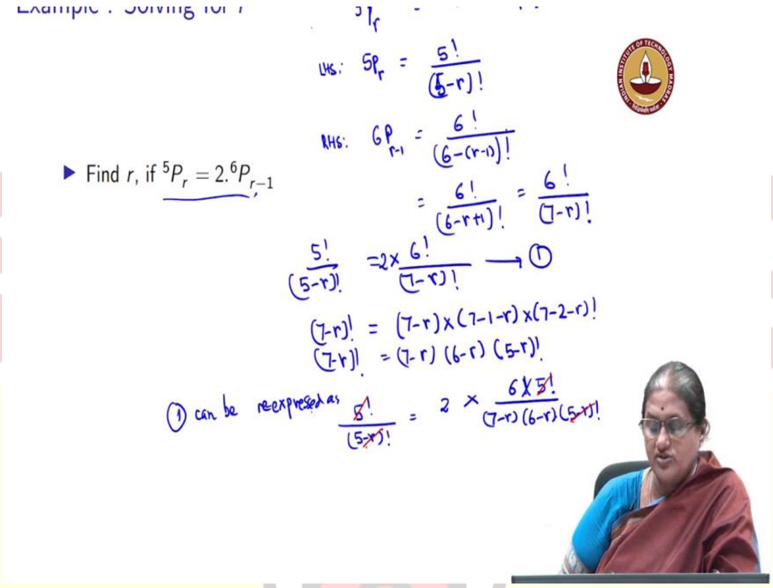


Now, let us look at another example. I have $n -$, $-= \frac{5}{3}$. Again I solve for n. Now, again my left hand side ${}^{n}P_4$, the numerator is $\frac{n!}{(n-4)!}$. The denominator is $\frac{(n-1)!}{(n-5)!}$, I can simplify this to write it as $\frac{n!}{(n-4)!} \times \frac{(n-1)!}{(n-5)!}$. Now, I know $n!$ is $n \times (n - 1)!$, this is what we know.

Similarly, $(n - 5)!$, $(n - 4)!$ is $(n - 4) \times (n - 5)!$. I retain $(n - 1)!$. Again, I can cancel out $(n - 1)!$ and $(n - 4)!$ and this leaves me with the following, this leaves me with in the numerator n, the denominator $(n - 4)$ equal to $\frac{5}{3}$. Cross multiplying I get, $3n = 5n - 20$ which further simplifies to $2n = 20$ and I can solve for $n = 10$.

So, you can see that I just solved from the first principles that $n = 10$ by expanding or applying whatever is the formula we have derived. Now, this kind of simplifying expressions would be very useful again when we get into probability and counting when we are doing the probability module. So, in the earlier case we solved for n given a particular expression.

(Refer Slide Time: 26:42)



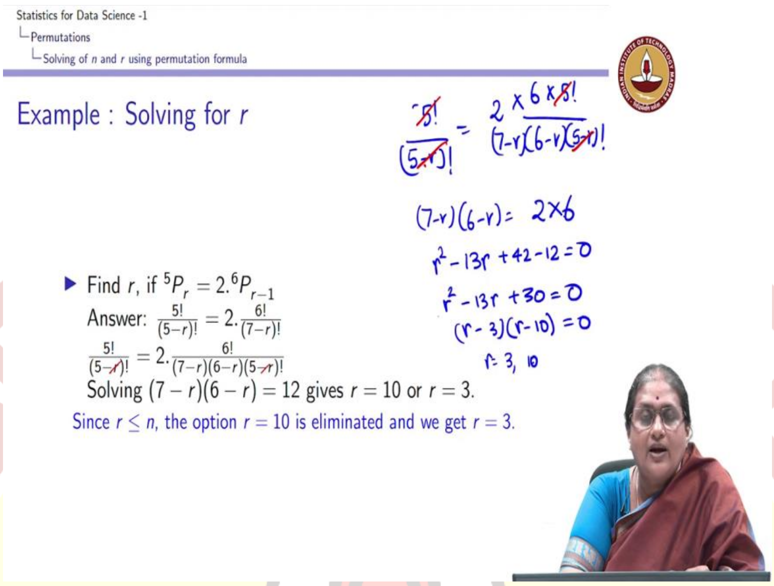
The next thing which we are interested in doing is how do we solve for r? So, in the earlier case I was solving for n, n was my unknown. Now, for example, I am given $^5P_r = 2 \times {}^6P_{r-1}$. Now, let us look at the left hand side. It is ^5P_r again applying the formula ^5P_r is $\frac{5!}{(5-r)!}$ this is what I have if I apply the formula. Let us look at the right hand side.

In the right hand side I have $^6P_{r-1}$. Again this is $\frac{6!}{(6-(r-1))!}$. We look at the denominator I can further simplify this denominator as $(6 - r + 1)!$. So, I have this as $\frac{6!}{(7-r)!}$. So, what is given to us is I $\frac{5!}{(5-r)!} = 2 \times \frac{6!}{(7-r)!}$ this that is what is given to us.

Further notice that $(7 - r)!$ can be re expressed as $(7 - r)(7 - 1 - r)(7 - 2 - r)!$ This is what we have already seen which is as $(7 - r)(6 - r)(5 - r)!$. So, I can re express as $(7 - r)!$ in this way. Now, once we do that, this expression which I have labelled as 1 can be re expressed as 1 can be re expressed as the following.

How can I re express 1? I can write it as 5!, but $(5 - r)!$ is $2 \times \frac{6!}{(7-r)(6-r)(5-r)!}$. Now, I can also write $6!$ as $6 \times 5!$, this is something which we have already seen. So, again we go back and we can see that I can cancel out $5!$ from that expression, $(5-r)!$ also from the expressions which gives us the following.

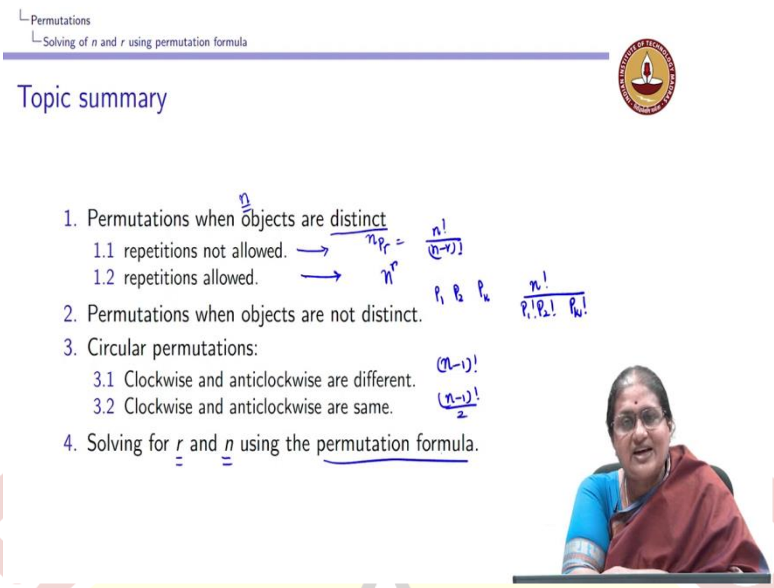
(Refer Slide Time: 29:43)



That when I cancel out both of them, it gives me this following. It gives me the expression that because I am cancelling out $(5-r)!$ and I can write $5, 6$. I have $7 -$ so what I have earlier was I had $5!$ so earlier what I had earlier was $\frac{5!}{(5-r)!} = 2 \times \frac{6\times 5!}{(7-r)(6-r)(5-r)!}$, I cancel out 5 factorial and $(5-r)!$i, and what I get is $(7-r)(6-r)$ is 2×6 which is 12.

I further I simplify this I get it as an r^2, this is an r^2, I get the I can write this as $r^2 - 13r + 42 - 12$ which is $r^2 - 13r + 30 = 0$ which I can see as $(r-3)(r-10) = 0$. I can factor this as $(r-3)(r-10) = 0$ which gives me r is either 3 or 10. So, this is how we can solve for r. Whenever you know what is an expression but you have an unknown you can simplify the and solve for either n or r by applying the formula.

(Refer Slide Time: 31:38)



So, when we look at the permutations what a summary of what we have seen in the permutation segment, first we started with n objects, these n objects were considered to be distinct. We looked at how many ways we can choose r objects from these n objects. When repetitions are not allowed we saw it was ^nP_r which was $\frac{n!}{(n-r)!}$. When repetitions are allowed, we saw it was n^r.

When objects are not distinct that is I had p_1 of one kind, p_2 of another kind and p_k of a kth kind, it was $\frac{n!}{p_1!p_2!...p_k!}$. Now, when we looked at circular permutation when anticlockwise and clockwise were different, we saw there are $(n-1)!$ ways to arrange these objects. When clockwise and anticlockwise were the same, it was $\frac{(n-1)!}{2}$. And finally, we saw how you can solve for r and n using the permutation formula when we are given an expression and we are asked to solve for r and n.

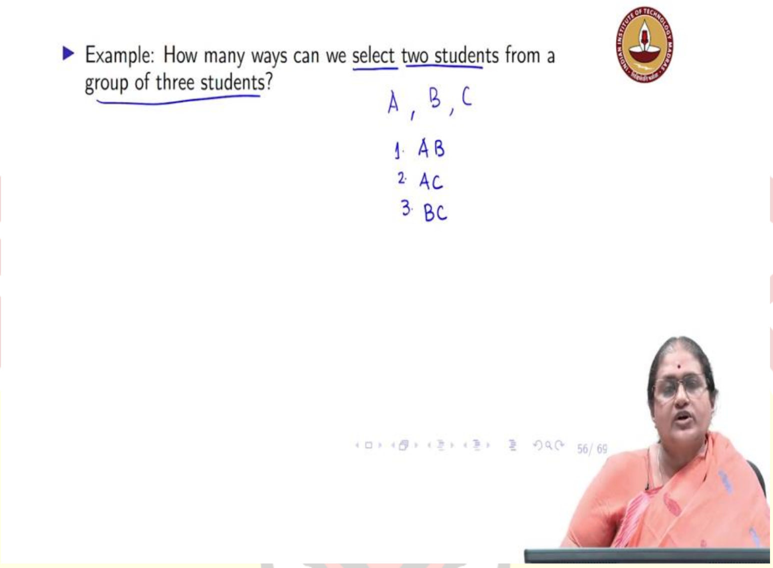
So, in permutation we actually looked that how do you arrange n objects or r objects out of n objects arrangement. A order was important. AB was different from BA. But many a time we might be just interested in how can we choose 2 objects out of 3 objects or how can we choose 3 people from a group of 10 people. In situations like this the order is not important. AB is same as BA. For example, I have 3 people and I just have to choose 2 people from 3 people, the order is insignificant.

Hence, we are going to look at what is called combinations, the number of ways you can choose 2 people or you can select 2 people or you can select r people in general from n

people gives us to what we referred to as combinations. That would be the next topic we are going to discuss.

Statistics for Data Science – 1
Professor Usha Mohan
Department of Management Studies
Indian Institute of Technology, Madras
Lecture 5.5
Permutations and Combinations – Combinations

(Refer Slide Time: 00:28)



So, the next topic we are going to learn about is the topic of combinations, we have learned about permutation, when we talked about permutation remember we were focusing on the word ordered arrangement. So, now let us go and understand what is a combination. So, let us start with an example, suppose I am asked to select I am clearly telling I want to only select two students from a group of three students, here we are not emphasizing on any arrangement or any order, we are very clearly specifying that we only want to select two students or choose two students, I am not bothered about whether the first student or second student, I do not have any order here.

So, in this case, let me start by telling that let me have A, B, C as the three students, just I am assuming that the first student is called A, B, and C, now from these three students if I have to make groups of two I can either have AB as a group, I could have AC as a group or I could have BC as a group, I can see that I have 1, 2, 3, so there are three ways of selecting two students from the group of three students.

(Refer Slide Time: 01:48)

Introduction

AB BA
1 2
A B
B A

- Example: How many ways can we select two students from a group of three students?
  - Let A, B, and C be the three students.
  - We can chose AB, AC, or BC.
  - Note, when we talked of permutations, the order was important, i.e., AB was different from BA.
  - In this case, they are the same- order is not important.

Statistics for Data Science -1
└ Combinations

Introduction

- Example: How many ways can we select two students from a group of three students?
  - Let A, B, and C be the three students.
  - We can chose AB, AC, or BC.
  - Note, when we talked of permutations, the order was important, i.e., AB was different from BA.
  - In this case, they are the same- order is not important.

Statistics for Data Science -1
└ Combinations

Introduction

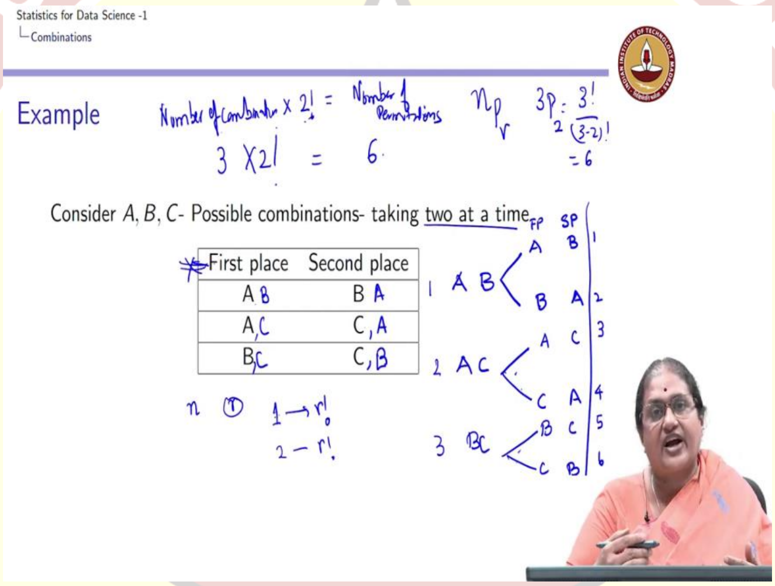
3 3 n AB COMBINATION
2 r AC
BC

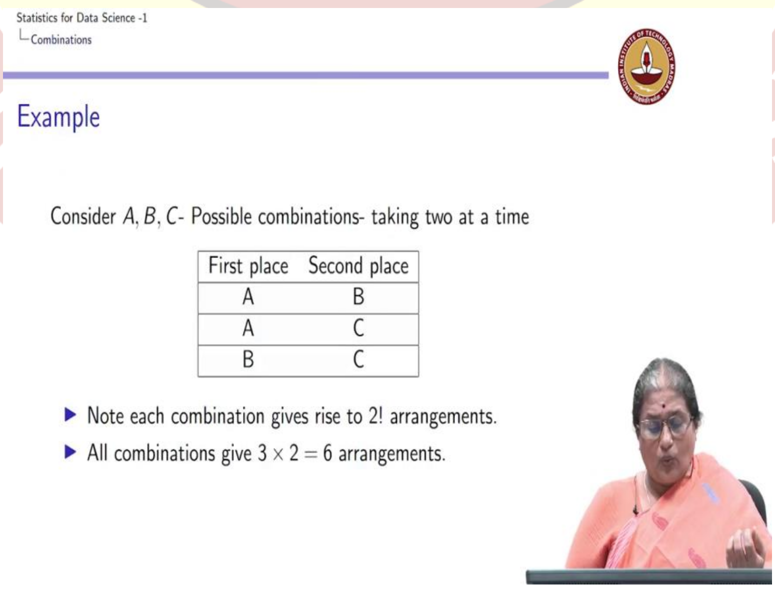
- Example: How many ways can we select two students from a group of three students?
  - Let A, B, and C be the three students.
  - We can chose AB, AC, or BC.
  - Note, when we talked of permutations, the order was important, i.e., AB was different from BA.
  - In this case, they are the same- order is not important.
- Each selection is called a combination of 3 different objects taken 2 at a time.

So, now this is AB, AC and BC. So, now here what you have to clearly understand is whether I write AB or BA is of no difference to me because the order in which I could choose A first and B first or B first and A first or both of them simultaneously there is no order in a combination. But when we talked about permutation whether A occupied the first block and B occupied the second or B occupied the first and A occupied the second there was a difference. So, you can see that when I am talking about combination the order is of no relevance to me, I do not have any importance to the order. Order is not important, that is something which we need to understand.

So, now each selection, so I have three selections, my selections are AB, AC and BC each selection is called a combination that is what we have written here. So, in general, here I had 3 objects and I selected 2 from this, so, I have from 3, I am selecting 2. In general, if I have n objects and I want to select r objects, I say I am choosing r from n or selecting r objects from n objects or I am choosing at a time, choosing r objects from n objects at the time.

(Refer Slide Time: 03:44)





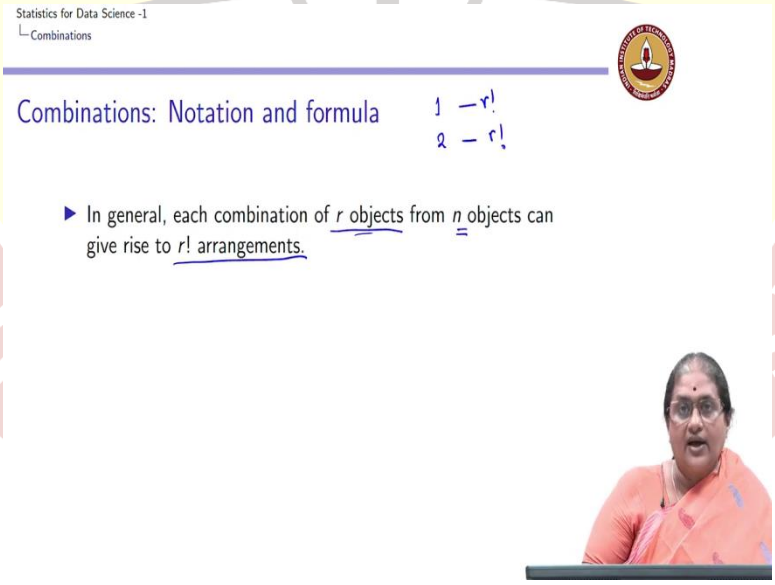
So, now going back to the example, I have suppose looking at the possible combinations taking two at a time, when we went back to our permutation, we had a first place and a second place here it really does not the order is of no importance here, so when I had the permutation, my first place could be an A or a B or it could be a B or A, the first place could be an A or second place could be a C or it could be the first place is a C and my second place is an A, first place is a B or a C or my second place is a C or a B.

These were the possible ways I could have had so you can see that if I have a AB combination, this has two arrangements one A in my first place, this is my second place and B in my second place or B in my first place and A in my second place, AC combination or AC selection I could have an A in my first place and C in my second place or C in my first place and A in my second place, BC combination could have a B in my first place, C in my second place or C in my first place and B in my second place. Recall, the number of permutations of 2 objects from 3, which was ${}^{3}P_{2}$ was $3! \times (3-2)!$ which was nothing but your 6, I have 1, 2, 3, 4, 5 and 6 possible permutations, the number of combination is 1, 2, 3.

So, from this example I have that the number of combinations which is equal to 3, each combination is giving rise to 2 permutation or 2! permutations is giving me the number of permutations, I repeat the number of combinations which is 3 each one is giving rise to 2! or 2 permutations, so finally I have the number of permutations which is 6, so number of combinations into 2! is giving me the number of permutations which was 6.

I can extend this logic to n thing, so if I have n objects and I am choosing r from these n objects, then I can say that if I have my first combination of n objects from r objects, each of these r objects can be so the one combination would give rise to $r!$ factorial arrangements, the second combination will also give rise to $r!$ arrangement, so every combination gives rise to $r!$ arrangements. Now, why is this important to us?

(Refer Slide Time: 07:26)



$1 \to r!$

Combinations: Notation and formula

$^nC_r \to \underbrace{^nC_r \cdot r!}_{\text{total no. of permut}}$

- In general, each combination of r objects from n objects can give rise to $r!$ arrangements.
- The number of possible combinations of r objects from a collection of n distinct objects is is denoted by ^nC_r and is given by

Statistics for Data Science -1
└Combinations

$1 \longrightarrow r!$

Combinations: Notation and formula

$^nC_r \times r! = {}^nP_r$

$^nC_r \times r! = \frac{n!}{(n-r)!}$

- In general, each combination of r objects from n objects can give rise to $r!$ arrangements.
- The number of possible combinations of r objects from a collection of n distinct objects is is denoted by ^nC_r and is given by

$$^nC_r = \frac{n!}{r!(n-r)!}$$

Statistics for Data Science -1
└Combinations

Combinations: Notation and formula

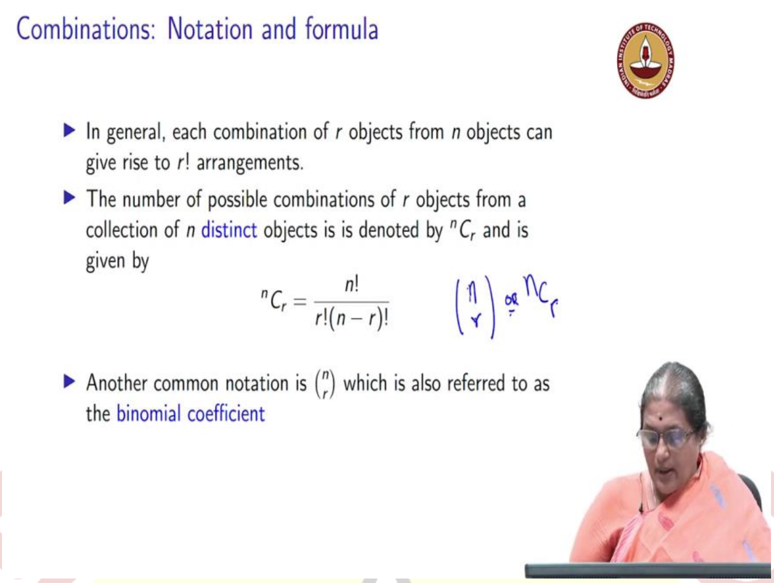
- In general, each combination of r objects from n objects can give rise to $r!$ arrangements.
- The number of possible combinations of r objects from a collection of n distinct objects is is denoted by ^nC_r and is given by

$$\boxed{^nC_r = \frac{n!}{r!(n-r)!}}$$

$^3C_2 = \frac{3!}{2!\,1!}$

$= 3.$

AB
AC
BC



From this, we can actually set up what we are referring to as a combination, notation, and formula. So, in general each combination of r objects from n objects give rise to $r!$ arrangements so that is what we have, if I have one combination then after it will give rise to $r!$, the second combination another $r!$ arrangements. So, if I am going to denote the number of possible combinations of r objects from a collection of n distinct object, I repeat the number of possible combinations of r objects from a number collection of n distinct object is denoted by ^nC_r, I am choosing r from n distinct objects.

So, this is the number of combinations, if I have from one combination I am going to have $r!$ permutations, so from ^nC_r, I am going to have $^nC_r \times r!$ arrangements. Now, this $^nC_r \times r!$ is nothing but the total number of permutations, total number of permutations this is what we saw in our example, so in general what do I have? I can write this ^nP_r .

Again, I repeat from one combination I get $r!$ permutations, so each combination because I have r objects these r objects among themselves can be arranged $r!$ ways if I have 2, I will have $2 \times r!$, ^nC_r is the number total number of combinations, so $^nC_r \times r!$ is going to give me the total number of permutations possible.

I already know ^nP_r is $\frac{n!}{(n-r)!}$, so I have $^nC_r \times r!$ is $\frac{n!}{(n-r)!}$ to give me the formula ^nC_r is $\frac{n!}{r!(n-r)!}$. So, this is the formula for ^nC_r and the way I express it is, the number of possible combinations of r objects from n distinct objects.

So, in earlier example I chose 2 objects from 3 distinct objects, so it is $^3C_2$ applying the formula it is 3! which is $n!$, $r!$ is $2!$, $(3-2)!$ is a $1!$, and I can see that this is equal to 3 which actually is the total number of combinations which I got namely AB, AC and BC, I had three combinations.

So, this formula is the number of possible combinations of r objects from n distinct objects. Another common notation which is used by many authors and book is $\binom{n}{r}$, this is also referred to as a binomial coefficient, some of the books use this either of these notations are fine $\binom{n}{r}$ or ^nC_r this is referred to also as a binomial coefficient.

(Refer Slide Time: 11:53)

Some useful results

$^nC_r = \frac{n!}{r!(n-r)!} = \frac{n!}{(n-r)!\,r!} = {}^nC_{n-r}$

1. $^nC_r =$

Statistics for Data Science -1
└Combinations

Some useful results

1 2 . . n

r $n-r$

1. $^nC_r = \frac{n!}{r!(n-r)!} = \frac{n!}{(n-r)!r!} = {}^nC_{(n-r)}$
   In other words, selecting r objects from n objects is the same as rejecting $n-r$ objects from n objects.

Statistics for Data Science -1
└Combinations

Some useful results

$^nC_n = \frac{n!}{n!(n-n)!} = \frac{n!}{n!\,0!} = 1$

$^nC_0 = {}^nC_{n-0} = {}^nC_n = 1$

$= \frac{n!}{0!(n-0)!} = \frac{n!}{n!} = 1$

1. $^nC_r = \frac{n!}{r!(n-r)!} = \frac{n!}{(n-r)!r!} = {}^nC_{(n-r)}$
   In other words, selecting r objects from n objects is the same as rejecting $n-r$ objects from n objects.
2. $^nC_n = 1$ and $^nC_0 = 1$ for all values of n

Some useful results

${}^{n}C_r = {}^{n-1}C_{r-1} + {}^{n-1}C_r$

1 2 3 4
A B C D E

A is in 3 chosen
A is not a part of 3

$n=5$
$r=3$

$$\boxed{\binom{5}{3} = \binom{4}{2} + \binom{4}{3}}$$

1. ${}^{n}C_r = \frac{n!}{r!(n-r)!} = \frac{n!}{(n-r)!r!} = {}^{n}C_{(n-r)}$
   In other words, selecting r objects from n objects is the same as rejecting $n-r$ objects from n objects.
2. ${}^{n}C_n = 1$ and ${}^{n}C_0 = 1$ for all values of n
3. ${}^{n}C_r = {}^{n-1}C_{r-1} + {}^{n-1}C_r; 1 \le r \le n$

Some useful results

${}^{n}C_r = {}^{n-1}C_{r-1} + {}^{n-1}C_r$

$n=5$
$r=3$

r has 'n'
r does not 'n'

${}^{n}C_r = \binom{n-1}{r-1} + \binom{n-1}{r}$

1. ${}^{n}C_r = \frac{n!}{r!(n-r)!} = \frac{n!}{(n-r)!r!} = {}^{n}C_{(n-r)}$
   In other words, selecting r objects from n objects is the same as rejecting $n-r$ objects from n objects.
2. ${}^{n}C_n = 1$ and ${}^{n}C_0 = 1$ for all values of n
3. ${}^{n}C_r = {}^{n-1}C_{r-1} + {}^{n-1}C_r; 1 \le r \le n$

Some useful results

${}^{n}C_r = {}^{n-1}C_{r-1} + {}^{n-1}C_r$

$n=5$
$r=3$

r has 'n'
r does not 'n'

${}^{n}C_r = {}^{n-1}C_{r-1} + {}^{n-1}C_r$

1. ${}^{n}C_r = \frac{n!}{r!(n-r)!} = \frac{n!}{(n-r)!r!} = {}^{n}C_{(n-r)}$
   In other words, selecting r objects from n objects is the same as rejecting $n-r$ objects from n objects.
2. ${}^{n}C_n = 1$ and ${}^{n}C_0 = 1$ for all values of n
3. ${}^{n}C_r = {}^{n-1}C_{r-1} + {}^{n-1}C_r; 1 \le r \le n$

Some useful identities combinatorial identities which will be useful for us to understand more about the combinatorial formula. The first thing is ^nC_r, so if I start with a ^nC_r by my formula I have this is $\frac{n!}{r!(n-r)!}$. Now, in the denominator I can write it as $\frac{n!}{(n-r)!r!}$, these two are equivalent, I am just changing the order in which I have written them.

So, what do I get now? Now, this if I am applying the same formula this is same as $^nC_{(n-r)}$. So, the first identity we have is $^nC_r = {}^nC_{(n-r)}$. In other words, what we mean is selecting ^nC_r, I can express as selecting r objects from n objects, $^nC_{(n-r)}$ is same as rejecting $(n-r)$ objects from r object. So, if I have n objects or n people, if I select r of them, I am basically rejecting $(n-r)$ of them, so if I have some ways of selecting these selecting these r objects is equivalent to rejecting these $(n-r)$ of them and that can be established by this identity.

The next important identity which we have is ^nC_n, again applying the formula I have ^nC_n is $\frac{n!}{n!(n-n)!}$, I have which is same as $\frac{n!}{n!0!}$, recall $0!$ is equal to 1, so I have ^nC_n equal to 1. Now, when we look at $^nC_0$ we can apply the earlier identity, I know $^nC_0$ is same as $^nC_{(n-0)}$ which is equal to ^nC_n, I have already established ^nC_n equal to 1, so $^nC_0$ is also equal to 1.

I can prove this from first principle, again $^nC_0$ is $\frac{n!}{0!(n-0)!}$ which is $\frac{n!}{n!}$, again $0!$ is 1, which is equal to 1. So, these are two important combinatorial identities that we need to understand. Now the third identity ^nC_r equal to $^{(n-1)}C_{(r-1)} + {}^{(n-1)}C_r$, this is a very important and a useful combinatorial identity. I am not going to formally prove this, this can be proved, you can give a proof for this also, but let me give you the intuition behind this identity by using an example.

For example, if n equal to 5 objects and I have to choose 3 objects out of them, so let me assume that I have A, B, C, D, and E who are my 5 students or 5 people or my 5 objects and I have to choose 3 from them, so now I choose one of them, for example if I choose A, I am choosing A, so there are 2 ways of looking at my total choices, 1 way is that the three group that is the group of 3 has A in it and the second is A is not a part of the 3 that is chosen, one way is A is a part of the 3 that is chosen and A is not a part of the 3 that is chosen.

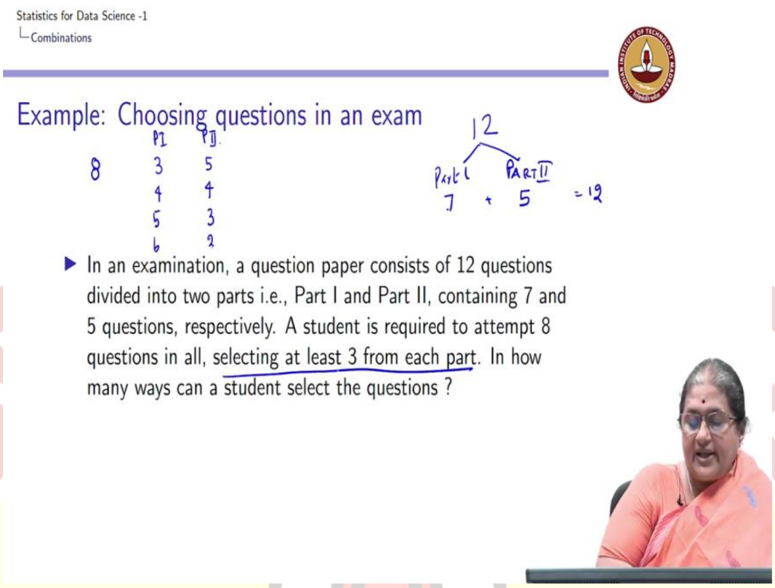
Now, if A is a part of the 3, so I am choosing 3 objects, if A is a part of the 3 that is chosen, I have to choose remaining 2, if A has already been chosen, I have to choose the remaining 2, for the remaining 2, if A is chosen I have 1, 2, 3, 4 available with me and I have to choose 2 and I can do that in $^4C_2$ ways, because A is already chosen I only have to choose 2 more objects, I have a remaining 4 objects to choose that from and that I can do in $^4C_2$ ways.

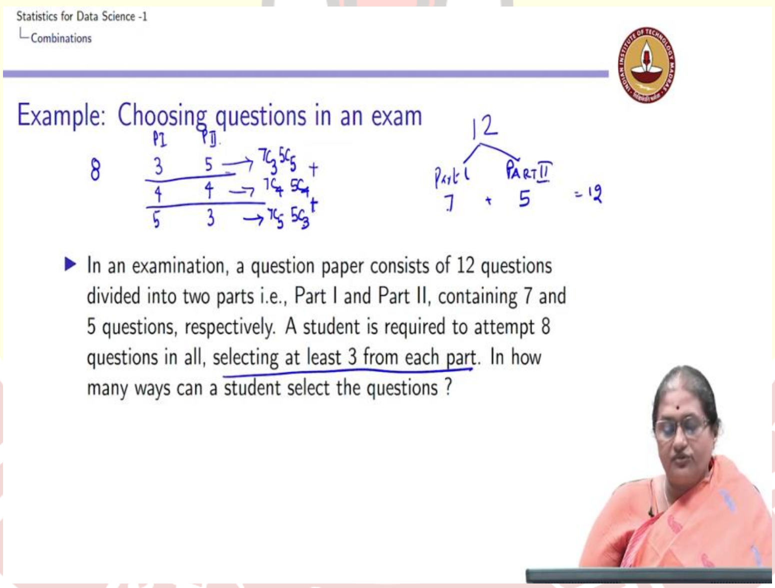
Now, suppose A is not a part of the 3 objects, then in that case I have 4 objects available with me, which is B, C, D, E and I have to choose all the 3 objects from this choice of 4 objects and that I can do in $^4C_3$ ways. So, the total number of ways I can choose 3 from this choice of A, B, C, D, E which is $^5C_3 = {}^4C_2 + {}^4C_3$, I can so you can see that n equal to 5, r equal to 3, (n-1) is 4, (r-1) is 2, plus (n-1) is 4, r is 3, so this is what we have got in this example. In general if I have n objects and I have to choose r from these n objects, I can fix, so the way I can understand this is I fix any one of the n objects, I fix any one of the n objects and I looked at the situation that the r objects has n and r does not have n.

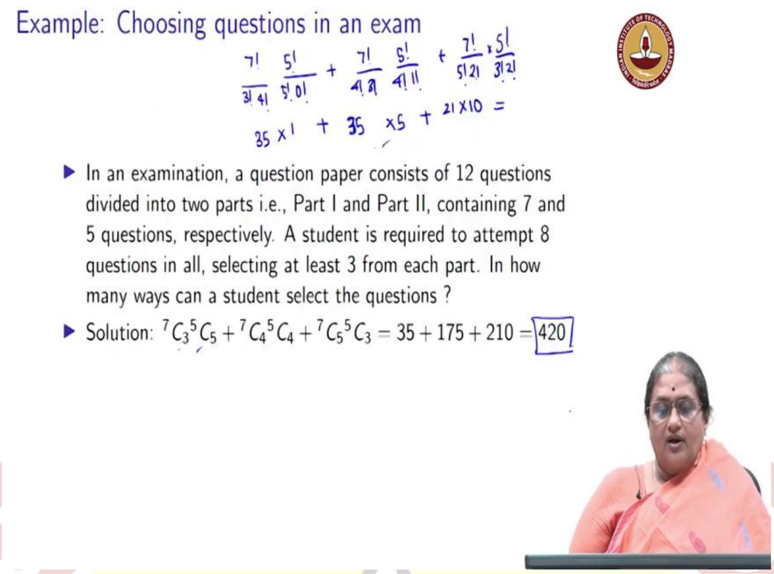
If r has n since I have already chosen one of the r objects, I only have to choose (r-1) of the objects from remaining (n-1) objects, here r does not have n, I have only remaining (n-1) because I am eliminating n, I choose r and these two put together give me ^nC_r. I use this

$^{(n-1)}C_{(r-1)}$, $^{(n-1)}C_r$ and this is ^nC_r. So, this is an important combinatorial identity which we can use or apply and we will be applying such identities whenever we go to the application aspect.

(Refer Slide Time: 19:49)







So, now let us look at a few examples to apply the concepts of combination we have learned so far, one is choosing a question in an exam. So, consider a situation where a paper consists of 12 questions, now these 12 questions are in 2 parts, part 1 and part 2, part 1 has 7 questions, part 2 has 5 questions, so 7 plus 5 is 12 questions. Now, a student has to attempt 8 questions in all, how many questions should a student attempt? A student has to attempt 8 questions.

So, the condition is he has to choose at least 3 from each part, so I have a part 1, I have a part 2, I can either choose 3 from here, if I choose 3 from part 1, I have to choose 5 from part 2, if I choose 4 from part 1, I have to choose 4 from part 2, if I choose 5 from part 1, I have to choose 3 from part 2, I cannot be choosing 6 and 2 because this would be violating the at least 3 from each part, so if it is at least 3 from each part, these are the possible ways I can choose these 8 questions assuming that the student is choosing exactly 8 questions out of the 12 questions.

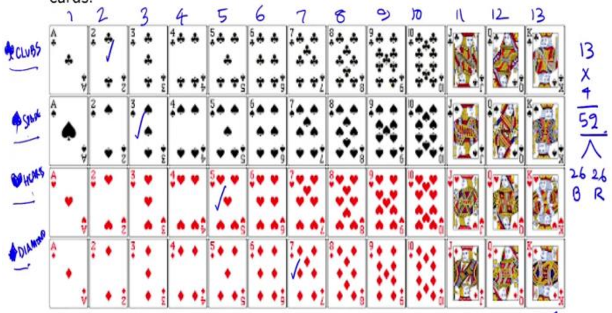
Now, how many ways can this be done? The number of ways this can be done is I have 7 questions in part 1, the number of ways I can choose 3 questions from 7 questions is $^7C_3$, the number of ways I can choose 5 questions from 5 questions is $^5C_5$, so I have a $^7C_3 \times {}^5C_5$ is a total number of ways I can choose 3 from part 1 and 5 from part 2. The number of ways I can do this is $^7C_4 \times {}^5C_4$, number of ways this can be done is $^7C_5 \times {}^5C_3$.

So, the total number of ways student can select question is $^7C_3 \times {}^5C_5 + {}^7C_4 \times {}^5C_4 + {}^7C_5 \times {}^5C_3$, which I can see is equal to 35, because $^7C_3 \times {}^5C_5$ is $\frac{7!}{3!\times4!}$, $^5C_5$ is $\frac{5!}{5!\times0!}$, so this is equal to 1, $\frac{7!}{4!\times3!} \times \frac{5!}{4!\times1!} + \frac{7!}{5!\times2!} \times \frac{5!}{3!\times2!}$, so $\frac{5!}{5!\times0!} = 1$, so $\frac{7!}{3!\times4!} = 35$, so 7×5, I have a 35.

Similarly, here I will get a 5, so from here I have a 7 into this is again a 35, here I have a 7×3 which is $21 \times 5!$, 5×2 which is a 10, so this in total gives me 1, this is 175, so I get a 35 + 175 + 210 which gives me a total answer of 420.

(Refer Slide Time: 24:02)

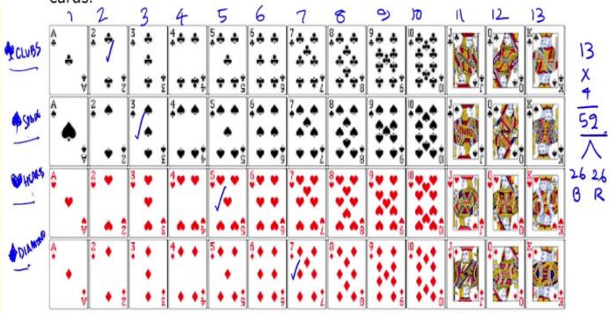
Lets consider the case of choosing four cards from a deck of 52 cards.



2♡ 3♡ 4♡ 5♡ X   X 2C 3S 5H 7D   2C 3C 6C 7C ✓
2D 4D 10D QD ✓

61 / 69

Lets consider the case of choosing four cards from a deck of 52 cards.



$^4C_1 \times {}^{13}C_4$

61 / 69

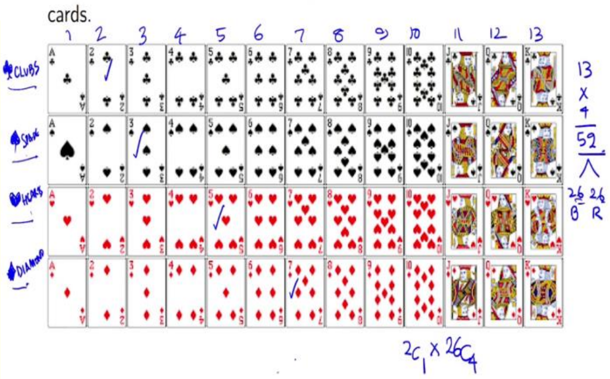
Example: Game of cards contd.

1. Total number of ways of choosing four cards from 52 cards = $^{52}C_4 = \frac{52!}{4!48!} = \boxed{2,70,725}$

Example: Game of cards contd.

1. Total number of ways of choosing four cards from 52 cards = ${}^{52}C_4 = \frac{52!}{4!48!} = 2,70,725$
2. All four cards are of the same suit ${}^{4}C_1 \times {}^{13}C_4 = 4 \times \frac{13!}{4!9!} = 2860$

Lets consider the case of choosing four cards from a deck of 52 cards.



Example: Game of cards contd.

1. Total number of ways of choosing four cards from 52 cards = ${}^{52}C_4 = \frac{52!}{4!48!} = 2,70,725$
2. All four cards are of the same suit ${}^{4}C_1 \times {}^{13}C_4 = 4 \times \frac{13!}{4!9!} = 2860$
3. Cards are of same colour ${}^{2}C_1 \times {}^{26}C_4 = 2 \times \frac{26!}{4!22!} = \boxed{2,99,00}$

So, now let us introduce you to a game of cards, the reason why I am introducing you to a game of cards is, when you look at problems and probability there are a lot of problems and probabilities which are based on a game of cards, so I will just introduce you to a game of cards, it is no gambling here, we are just is a nice way to understand the concept of probability that is why I have just introducing you to a formal deck of what we refer to as playing cards.

Now, when you look at playing cards these are called a suits, so I have clubs, a club is something of this kind this is what we refer to as a club, then I have what I refer to as a spade, then we have hearts and we have diamonds, a typical pack of cards has clubs, spade, hearts and, sorry this is a diamond.

So, now let me introduce you to a pack of cards, we are introducing you to a pack of cards because many problems in probability from many textbooks and whenever you learn probability, you will be introduced to questions which require you to understand something about a pack of cards.

So, when we look at a pack of cards, we have clubs, we have spade, we have hearts and we have diamond, now these are referred to as suits. So, typically I have 4 suits and what are the suits? They are clubs, they are spade, hearts and diamond, each of them have 13 denominations, that is I have A, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13. 13 denomination with 4 suits, which will give me a total of 52 cards and typically I will have of these 52, 26 are black and 26 are red, this is how a pack of cards is typically given to us.

Now, suppose I have to choose 4 cards from these 52 cards, I am not placing any other condition I am just telling choose 4 cards from this 52 cards, how many ways can we do it? The answer is pretty simple, it is just ${}^{52}C_4$ because I am not qualifying that what is it I require I am not saying choose 4 red cards, I am not saying choose for clubs, I am not saying choose 4 spades, I am just telling choose any 4 cards and those 4 cards could be a 2 clubs or a 3 spades or a 5 hearts or a 7 diamond, this is a perfect 4 choice where I have chosen these 4 cards. And how many ways can we choose 4 cards from 52 cards?

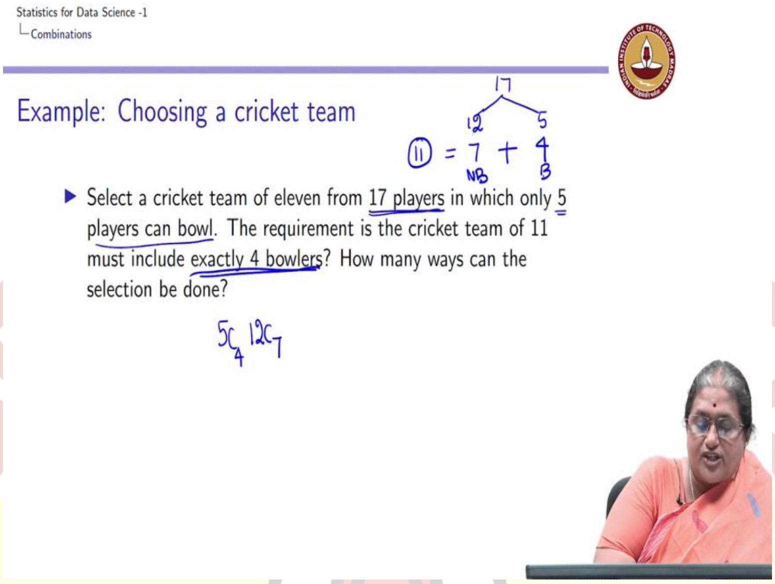
The ways I can choose 4 cards from 52 cards is just ${}^{52}C_4$ which is equal to this number which is a pretty large number. Now, suppose I am telling that all the 4 cards are of the same suit, so let us go back here, earlier I said that let us choose a 2 clubs, 3 spade, 5 hearts and 7 diamond, you can see that all these 4 cards or of different suits, I do not want this, this is not an acceptable selection for me, what is an acceptable selection? If I had a 2 clubs, 3 clubs, 6 clubs, and 7 clubs this is acceptable to me.

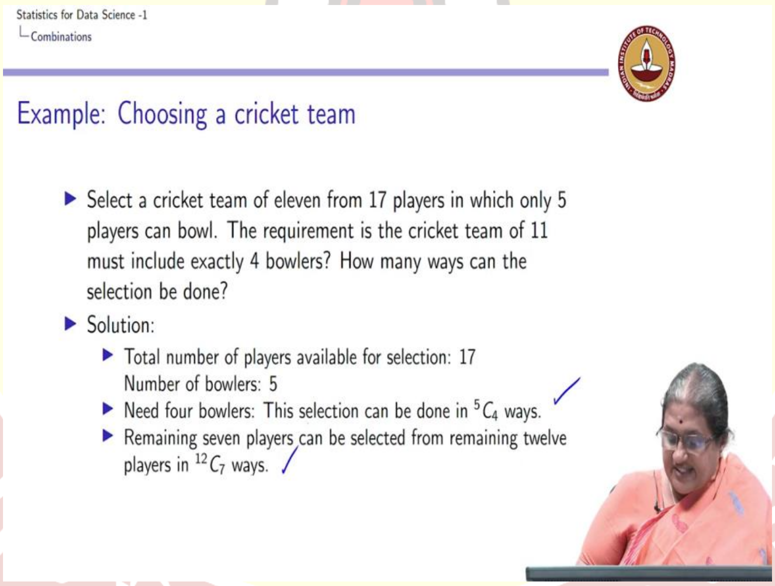
2 diamonds, 7 diamond, 10 diamond, and a queen diamond this is also acceptable to me, but a 2 hearts, a 3 hearts, a 4 diamond, and a 5 diamond it is not acceptable to me. So, how many ways can I choose this, again the requirement is choose 4 cards, but they should be of same suit. So, the way you can do it is, first let me choose a suit, how many ways can I do it? How many types of this? How many suits are there? There are 4, I need to choose 1 out of the 4 and how many ways can I do that? I can do that in ${}^{4}C_1$way.

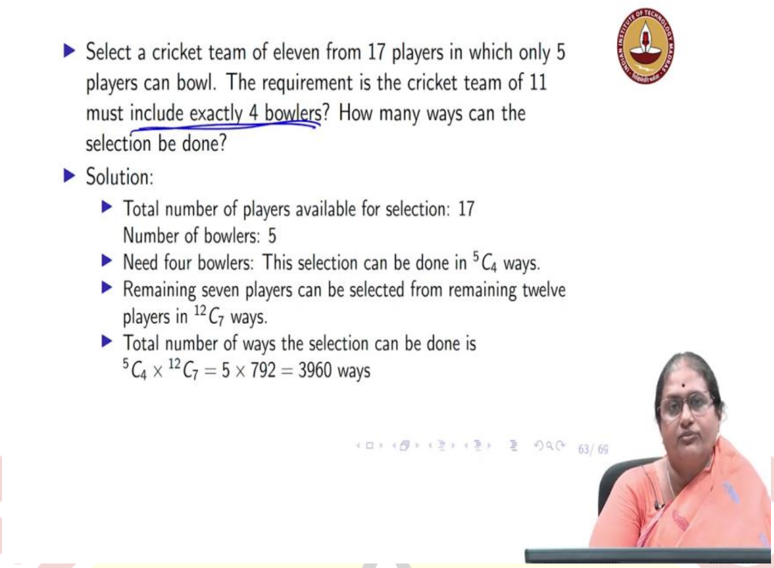
Now, within each suits I have 13 of each kind and I need to choose 4 from this 13 and that can be done in ${}^{13}C_4$ ways, so the total number of ways I can choose all the 4 cards from the same suits is ${}^{4}C_1 \times {}^{13}C_4$ which is equal to 2,860 ways of doing it. Now, let us go to another problem, now I am asking, how do you choose the cards of the same colour? Now, again go back, how many colours do we have? You can see that we have 2 colours. How many ways can I choose 1 colour from 2 colours? Again the number of ways of choosing 1 colour from 2 colour is ${}^{2}C_1$.

Now, within each colour I have 26 cards that is what we have seen and I need to choose 4 from each colour so I can do that in $^{26}C_4$ ways, so the total number of ways I can choose 4 cards of the same colour are $^2C_1 \times {}^{26}C_4$ which we can see is 29,900 ways of doing it.

(Refer Slide Time: 30:54)







So, let us move on to the next example. Now, suppose all of us know how we choose cricket team, a cricket team has 11 players, of these 11 players we need to have some players who are batsman, some who are all-rounders, we need a wicket keeper, we also need bowlers, suppose I have 17 players available with me, I have 17 players available with me, of these 17 players only 5 players can bowl, so I have total 17, so recall I can put it in the framework of my question paper, of these 17 question players I have 5 who can bowl, so $17 - 5$ is 12 who cannot bowl.

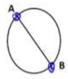
I need to choose 11 players, the requirement is, it should have exactly 4 bowlers, so if it should, the requirement is it should have exactly 4 players, I am not saying at least, I am saying exactly 4 players, if I have 4 bowlers, then I should have 7 non-bowlers in my team to make up to the 11 players, how many ways can the selection be done? The 4 bowlers can be chosen from the 5 bowlers in $^5C_4$ ways, which leads $^{12}C_7$ to be the number of ways I can choose the 7 remaining players from the 12 players to give me a total of $^5C_4$ and $^{12}C_7$ to give me a total number of ways as $^5C_4 \times {}^{12}C_7$, which I can simplify as 5×792 which is actually 3960 ways of selecting my team.

Here the qualifier was I should include exactly 4 bowlers and I can choose these 4 bowlers from 5 bowlers and the remaining 7 players from the remaining 12 non-bowlers. So, this is how I can have my team selection and apply the combinatorial or combination formula which we have stated earlier.

(Refer Slide Time: 33:39)

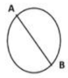
Example: Drawing lines in a circle

- Given n points on a circle, how many lines can be drawn connecting these points?
- $n = 2$ points, one line can be drawn connecting the points



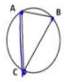
line segment: AB

- Given n points on a circle, how many lines can be drawn connecting these points?
- $n = 2$ points, one line can be drawn connecting the points



line segment: AB

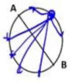
- $n = 3$ points, three line can be drawn connecting the points



AB
AC
BC

line segments: AB, AC, and BC

- Given n points on a circle, how many lines can be drawn connecting these points?
- $n = 2$ points, one line can be drawn connecting the points

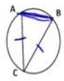


A, B → AB → 1

$^2C_2 = \frac{2!}{2!\,0!} = 1$

line segment: AB

- $n = 3$ points, three line can be drawn connecting the points



A
B
C

AB
AC
BC

3

$^3C_2 = \frac{3!}{2!\,1!} = 3$

line segments: AB, AC, and BC

- In general, given n points, number of line segments that can be drawn connecting the points is $\boxed{{}^nC_2}$ /

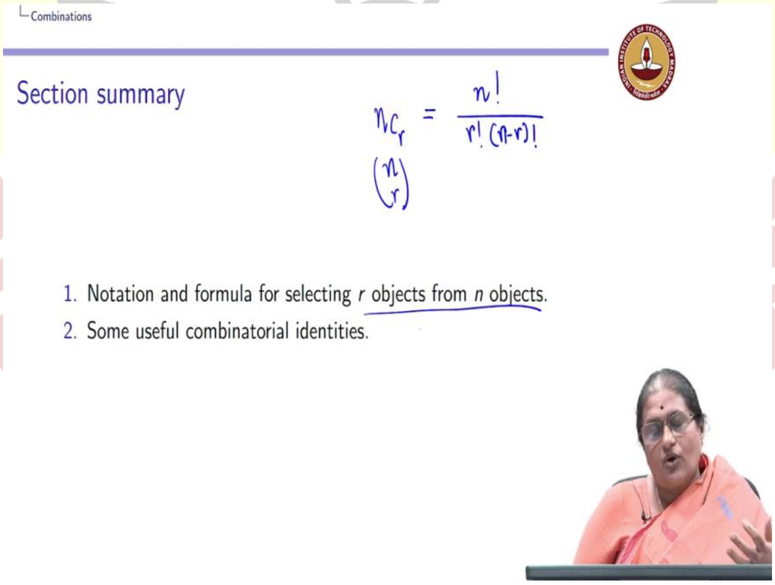
Now, let us look at another example, suppose I am given n points in a circle, the question is how many lines can be drawn connecting these points? So, now let us look at a very simple case of n equal to 2, let me turn the points A and B, so given A and B I know, I can draw only 1 line connecting A and B, so you can see that I am referring to that line segment as AB given 2 points on the circle, so n equal to 2, I can write I can connect it only with 2 points, sorry given 2 points I can connect it using 1 line.

Suppose I am given 3 points, so now you can see that I have A, B, and C are the 3 points I can connect AB with a line segment, I refer to that line segment as AB, I can connect AC with a line segment, I can also connect BC with a line segment. So, I can refer to my line segments as AB, AC, and BC.

So, I am given 3 points, what are the 3 points? My 3 points are so in the when n equal to 2, I had points A and B, my line segment was AB, I had 1 line, when I had points A, B, and C, I could form segments AB, AC, and BC, so I have 3 lines segments, so immediately you can notice that if I have to draw a line segment I need 2 points, I have 3, out of 3 points I can select 2 points at a time in ${}^{3}C_{2}$ ways and that ${}^{3}C_{2}$ we have already seen ${}^{3}C_{2} = \frac{3!}{2!1!}$ which is 3, so I have 3 lines 1, 2, and 3.

Here ${}^{2}C_{2}$ is $\frac{2!}{2!0!}$ which is 1 and hence I have that number which is equal to 1. So, in general if I am given n points, in general if I have n points on the circle, to draw lines I need to choose any 2 points, that is if I fix this point I can draw lines between any 2 of these points, given n points I can choose these points in ${}^{n}C_{2}$ways. So, the number of lines segments that can be drawn connecting the n points is ${}^{n}C_{2}$. Again, here there is no direction in this line I just have a line which is AB, the order is not important because there I refer to BA or AB, it is the same here, so I am looking at a combination.

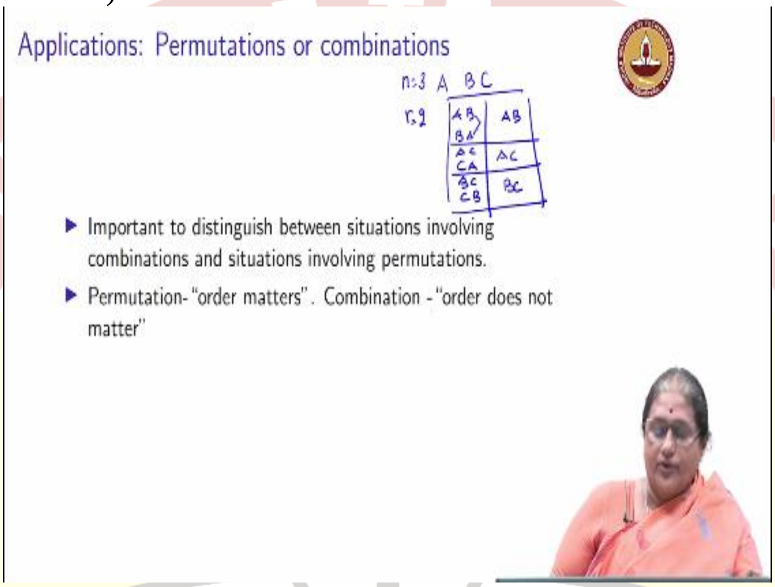
(Refer Slide Time: 37:11)



So, in summary we have started or we gave what is the notation and formula for selecting r objects from n objects, we represented that by ${}^{n}C_{r}$, some books represented by $\binom{n}{r}$, the binomial coefficient, we saw ${}^{n}C_{r}$ is given by $\frac{n!}{r!(n-r)!}$, this is the important combinatorial formula which we gave and then we gave some useful combinatorial identities, we describe them and qualified it and we also gave a few applications of combinations, that is what we have done in this section.

Statistics for Data Science-I
Professor Usha Mohan
Department of Management Studies
Indian Institute of Technology, Madras
Lecture – 5.6
Permutations and Combinations – Applications

(Refer Slide Time: 00:14)

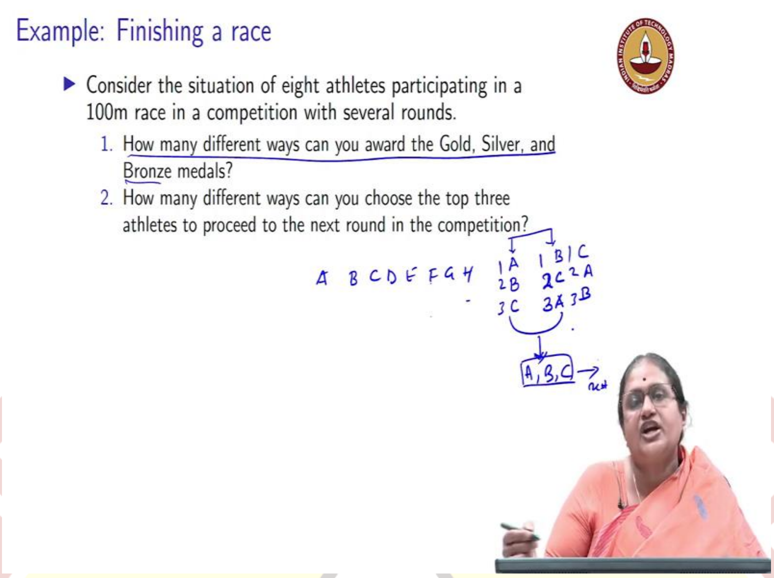


Now, we look at a very important application of whatever we have learned so far. So, given a situation it is very important for us to distinguish as to whether the situation under consideration requires us to answer in terms of permutation or in terms of combination. So, we need to distinguish the situations involving combinations or permutations that is the key understanding which we want to give now.

So, what is it? Remember when we talk about permutation the order matters, but when we talk about combination the order does not matter. Again recall that when I had three objects and I was choosing or arranging two objects or when my n was equal to 3 and I was arranging two objects I could do it in AB, BA, AC, CA, BC and CB, 6 ways.

Now, when I am talking about a combination these two are the same. So, I had an AB, AC and BC the order in which the objects are arranged did not matter. This was the key difference between a permutation and combination that is the absence of order in a combination and order mattering in a permutation.

(Refer Slide Time: 02:03)



Now, let us look at a few examples to articulate this difference. For example, suppose you consider the situation of 8 athletes participating in a 100 meter race. Now, if the situation says that okay there are 8 people participating in a race and I want to award three medals, namely the gold medal, the silver medal and the bronze medal to these 8 people then if I have these 8 people who are A B, C, D, E, F, G and H.

These are the 8 people then the order in which they are finishing. So, A could be first followed by D and F this could be one order. D could be first followed by A and third place F this could be another order so here A gets the gold, D the silver and F gets the bronze. In this case D gets the gold, A gets silver and F gets the bronze. So, the three people who are actually there are ADF in both this situation and in this situation.

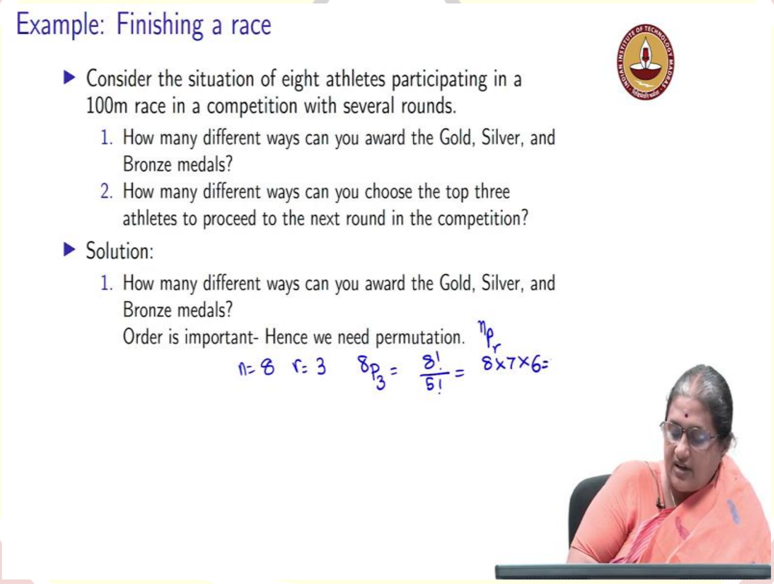
But clearly they are different because the order of finishing is different. So, when I want to know how many different ways I can award the gold, silver and bronze. I need to look at permutations because here the order matters. So, whenever you are asked a question the first thing we need to do is to understand whether order matters or not. Now, the second question says that suppose my question I have again the 8 people who are participating in the race.

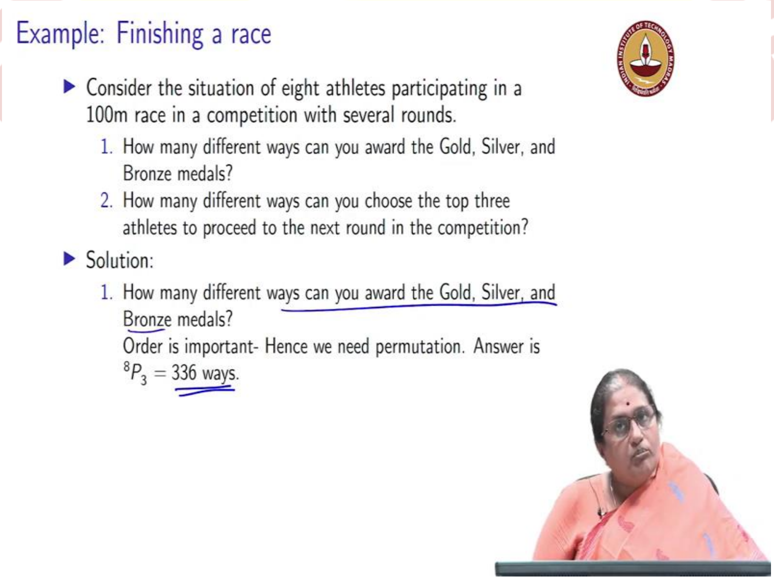
But here I am looking at a heat and I am just going to send the top three people to the next round. When I am going to send the top three people to the next round whether A came first, B came second and C came third or B came first, C came second and A came third in both these cases A, B, and C go to the next round. So, in this case I am not bothered or I am not concerned about the actual order in which A, B, C finish the race because I am just choosing the top three people to go into the next round.

So, in this second case A, B, C, B, C, A, C, A, B whatever has been their order of finishing are the same thing because they proceed to the next round. So, in the second case the order is not important to me when I choose the top three and I will be looking at a combination. So, this distinction is very important to us. So, when we encounter situations with the question we need to first see whether the order matters or the order does not matter.

When the order matters you will apply a permutation. When the order does not matter you will apply a combination.

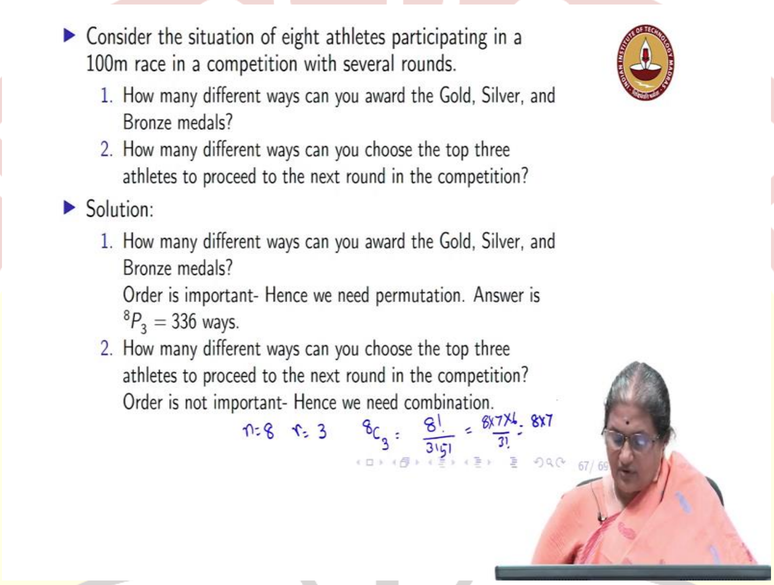
(Refer Slide Time: 6:14)

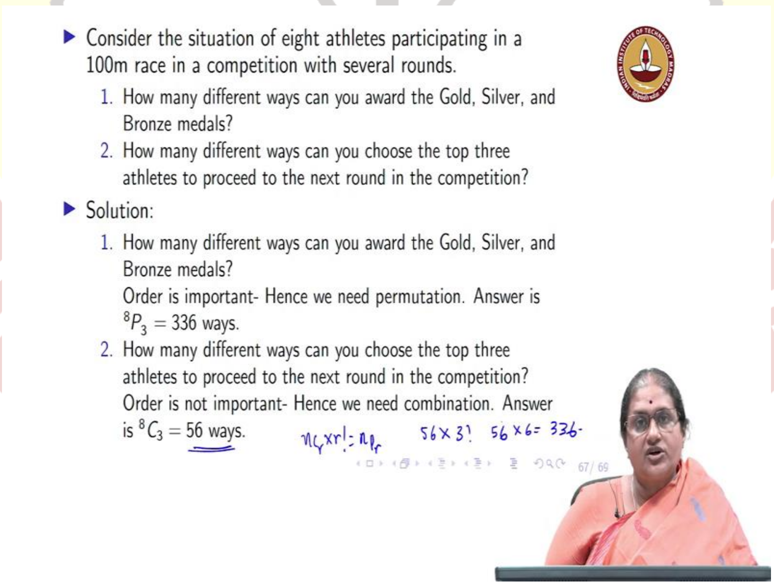




So to answer this question, in the first case we see that the total number of ways the order is important you go with the permutation. So, what is n here, so what is the permutation I am going to seek npr. In this case I have n equal to 8, r equal to 3 so I am looking at 8p3 which is $\frac{8!}{5!}$ which reduces to $8 \times 7 \times 6$. That is what I have which is 336 ways I can award the gold, silver and bronze medals.

(Refer Slide Time: 7:00)

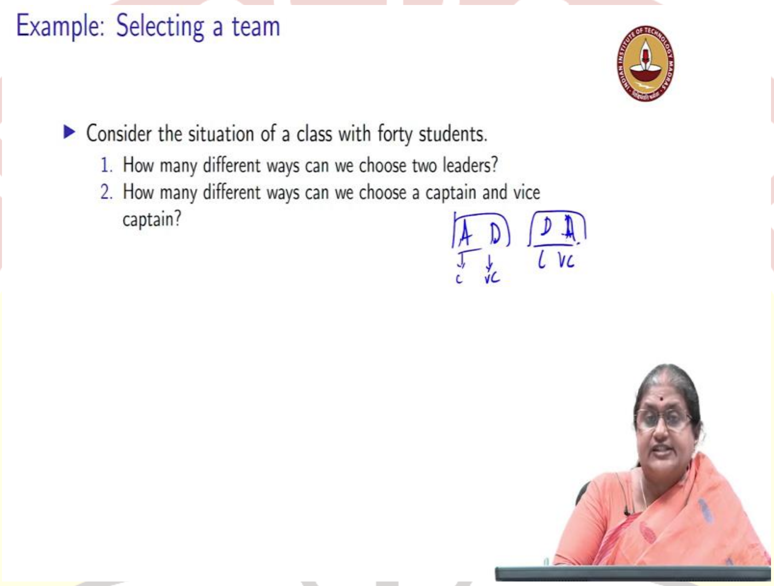




In the second way how many different ways can you choose the top 3? I again realize that the order is not important hence I proceed with combination again n equal to 8 r equal to 3 I look at 8 choose 3 which applying the formula is 8 factorial by 3 factorial which is $\frac{8!}{3! \times 5!}$.

So, which is $\frac{8!}{3!\times 5!}$ which I can see is $\frac{8\times 7\times 6}{3!}$ which is 8×7 is finally what I have the answer is just 56 ways of choosing the top 3 athletes to proceed into the next round of competition. You can see that $56\times 3!$ which is 56×6 is 336 which is my 8p3 this is the verification of your n choose r times r factorial is npr. We can check that n choose r into r factorial is your npr. Now, let us look at another few examples.

(Refer Slide Time: 08:24)



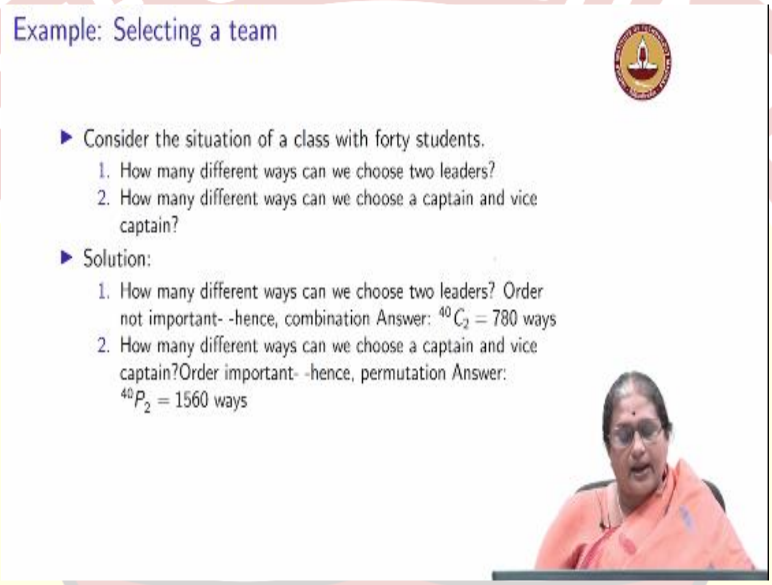
Now, looking at selecting a team. Consider a class of 40 students now the two questions I have here is. The first question is how many different ways can I choose two leaders or two representatives from this class. I can call them class representatives. I am not giving any order so I am not saying one of them is superior to the other or the other is superior.

So, I am just telling I just want to choose two class representatives from these 40 students. I want to know how many ways I can do it. The second question is different. I want to choose a captain and a vice captain from the 40 students. So, in the first case I am asking the question how many different ways can you choose two class representatives? In the second case I not only want to choose two students I am also asking the question I want to have one of them as a captain and another as a vice captain.

So, clearly when I am talking about the second question I am interested in having an order. For example, if I am choosing A and D in the first case choosing A and D as class representatives whether I choose A and D or D and A it is not going to make a difference, but in the second case if A is a captain and D is a vice captain and D is a captain and A is a vice captain these two are different in each other.

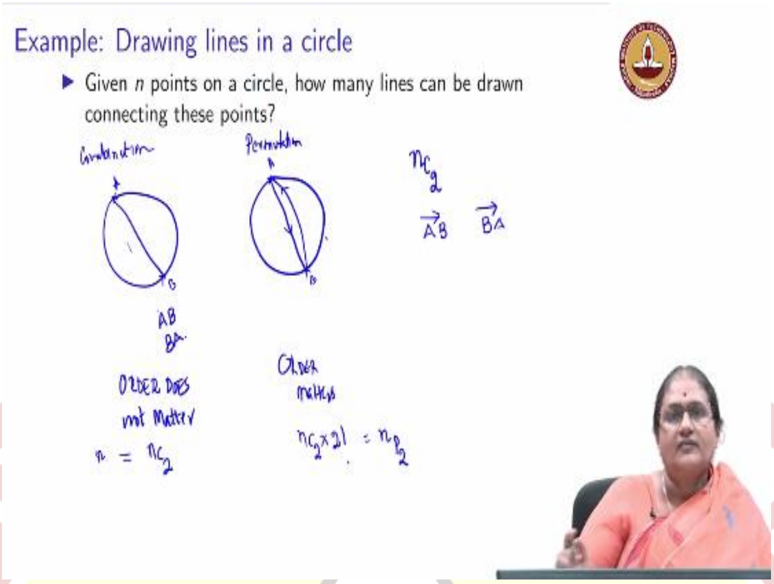
So, the second case needs a permutation to answer the question whereas the first case needs a combination to answer. Again it is extremely important for us to develop this idea or this skill of understanding which question for which question we need to apply a combination and for which question we need to apply a permutation.

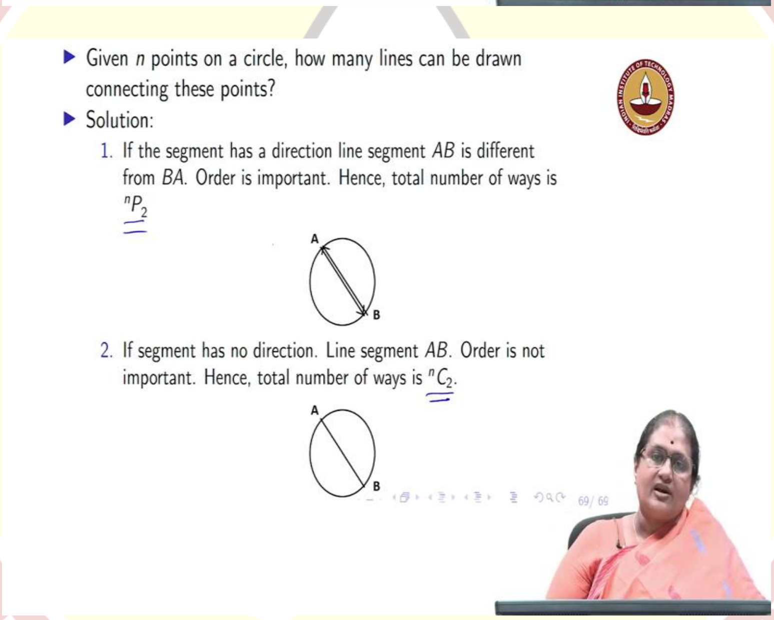
(Refer Slide Time: 10:59)



So, in this case I have how many different ways of choosing two leaders. I just have n equal to 40 when I have n equal to 40 I am going in for a combination because here the order is not important. My n equal to 40, r equal to 2, my n choose r is 40 choose 2 which gives me 780 ways of choosing a two class representatives or two leaders. Whereas in the second case, the different ways I should choose a captain and a vice captain from 40 students. I require a permutation because order is important again I will write that as 40P2 and 40P2 we know is 1560 ways of choosing a captain and a vice captain.

(Refer Slide Time: 11:49)





Now, let us go to another example. We have already seen how many ways we can draw line segments joining two points in a circle. So, if I am given a circle and I have two points say A and B, I know I can join it with a line drawn from A to B. So, suppose I am given a circle and I have two points A, B, I can join these two with a line. So, we have already seen that if we have given n points I can join them using n choose 2 is the total number of lines that can be drawn through n points which lie on a circle.

In other words I can have n choose 2 chords when if I have n points on the circle. Now, this chord is an undirected line. For example, if A and B were two locations on this point and I am interested in knowing from going from A to B to B to A. In that case A to B will be different from B to A because I have a direction from going from A to B to B to A which was different from the case when I just joined two points using a chord here the line segments AB was same as BA.

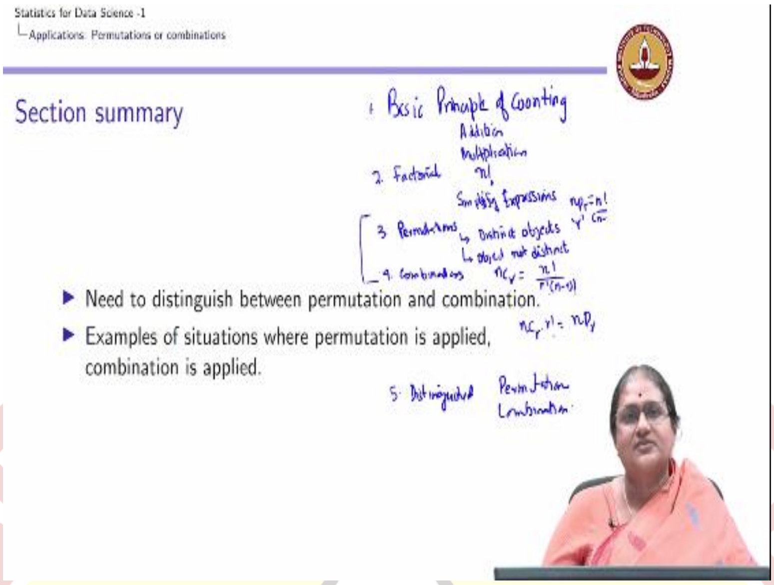
So, you can see that when there is a direction involved A to B and B to A are different and the order matters in the second case whereas in the first case A to B and B to A are the same order does not matter here. Here order matters. So, in this case if I am given n points then I have n choose 2 is the total number of lines I can draw between these points here I can again have n choose 2.

But I know that should be multiplied by 2! because every line is giving 2 directed lines and this is nothing, but np2. Again this is an example of in the same situation one of the answers needs a combination which is an answer and the other needs a permutation which gives me the answer. So, this is about drawing circle.

So you can see that when I have a directed line or a directed arc or a directed quad I have the order which is important. Whereas when I am talking about a line segment the order is not important and I just have total number of ways n choose 2 in the earlier case when the direction was there it was np2.

So, in this section we have just seen that it is very important for us to distinguish between permutation and combination depending on the situation and the question that is asked on the situation, not all situations require us to answer in terms of permutations or combinations. So, it is very important for us to distinguish these two. There could be situations where neither of them are necessary we have not looked into that situations in this module. We also looked at a few examples to illustrate the point as to how we distinguish between whether permutation has to be used or combination has to be used.

(Refer Slide Time: 16:33)



So, in summary what we have learned in this week is we started with the basic principle of counting. When we looked at the basic principle of counting we introduced both the addition principle of counting and the multiplication principle of counting.

The multiplication principle of counting helped us introduce what we call the factorial notation which is n! and this is a short hand for us to write the link the expression which we got from the multiplication rule for counting. Once we establish the factorial notation we looked at how do we simplify expressions using the factorial notation that is what we looked at.

Then we introduced permutations now when we looked at permutation we started with distinct objects and choose r objects from n distinct objects. We then extended it to objects are not distinct. So, first is I could have some of them which are of the same type. So, first we looked at distinct objects we looked at repetition not allowed, repetition allowed then we looked at objects were not distinct you could have p of one kind or p1 of one kind, p2 of the other kind and we saw how we can actually. So, objects are not distinct. We obtain formulae to come and understand how to do that.

The next thing we looked as combinations. We introduce and choose r, here we introduce what we was npr, npr was $\frac{n!}{(n-r)!}$. Here we introduce n choose r as $\frac{n!}{r!\times(n-r)!}$ by observing n choose r into r factorial is npr we get hold of this formula.

We looked at applications of combinations. The last thing we did was to actually look at examples where we distinguished whether we need a permutation or a combination we looked at a few examples wherein depending on the nature of the questioners you decide on whether to go with the permutation or with combination. Now, with this background we move forward to introduce you all to the probability module.

We start by again basics of probability namely we will introduce you to the concept of a random experiments, sample space, events, addition of events and all of that. So, that by that end you can compute probabilities of events. When you have to compute probabilities of events this notion of how you count becomes very useful. Hence this is a prerequisite for you to understand the probability module. Thank you.

Statistics for Data Science – 1
Prathyush P
Support Team
Indian Institute of Technology, Madras
Week - 5 Tutorial - 1

(Refer Slide Time: 00:16)

Hello, statistics students. In the tutorials for week five, we will be doing problems on counting principles. So in this problem, we are looking at the permutations of the word GRAPH. And they are arranged in dictionary order, which means they go alphabetically.

So if we were to put them down, the very first word would begin with A, which is the first letter. And then what is the next one? We have G coming up next. This is in the alphabetical order. And then we have H, then we have P, and then we have R. So this would be the very first word.

And then the second word would be what is next, you still have A, you would still have G, you would still have H, and these two are going to be exchanged. So this would be the second word. And then the third one and this order would be; A still there, G still there. But in the place of H, we put P, and then H R, and so on, and so forth, you will get so and so number of permutations.

How many permutations do we get on a five-letter word? We would get 5! because there are no repetitions. So 5! is 120. So you will have 120 permutations. And the last one is just going to be the reverse of the first one. So you will have RPHGA at the end. So what is the 73rd word is what we are looking at.

(Refer Slide Time: 02:05)

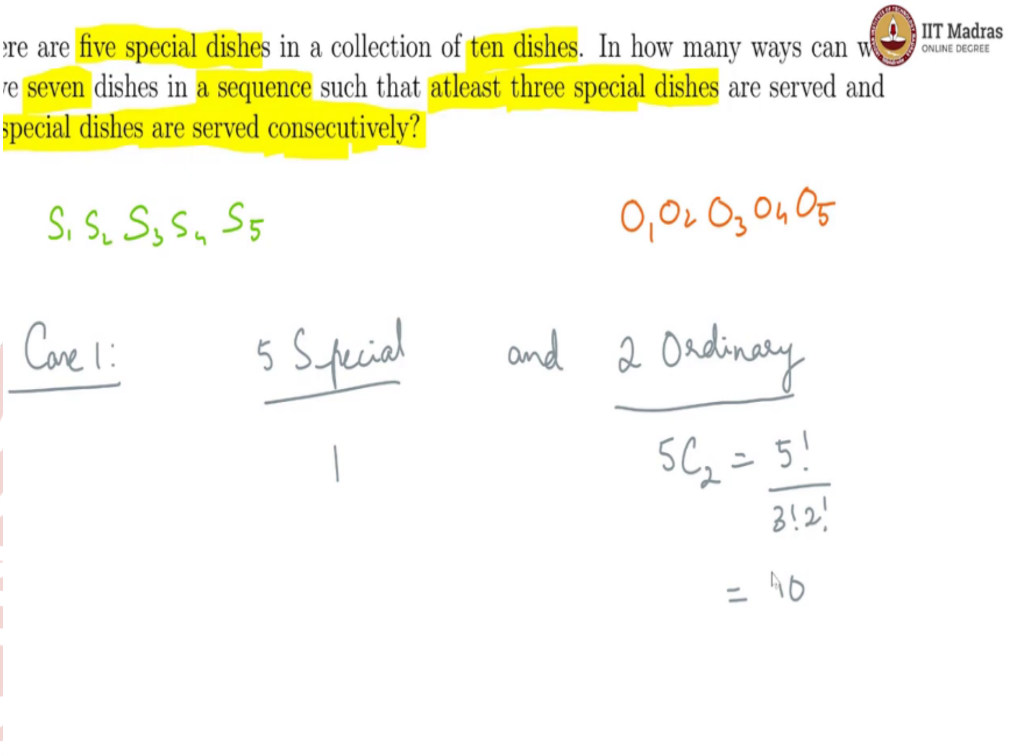
For that, now let us look at how many words we will get which begin with A. So if we fix A as the first letter, then we have four remaining boxes to be filled. And how many such permutations can we get? We will get 4! because there are four boxes to be filled and all of them are filled by different objects.

So 4! here is 24. So the number of words starting with A is 24. Then let us look at the number of words that start with G, which is the next in the alphabetical order. And again, we have 4. So again, 4! which gives us 24 words with G. Then we go next to H which again gives us 4!. So again 24.

So, so far, till this point, we have seen 72 words which means the next one in order is the 73rd word. And what is the next one in order? It has to start with P. So the first letter has to be P because after A and G and H comes P. And then the remaining we have to place in alphabetical order. So what are the remaining? Here, you will have A, this will be G, this is H, and this is R. So the permutation number 73 is PAGHR.

Statistics for Data Science - 1
Prathyush P
Support Team
Indian Institute of Technology, Madras
Week - 5 Tutorial - 2

(Refer Slide Time: 00:15)



In this question, they are saying there are 5 special dishes in a collection of 10 dishes. So there are 5 special dishes, let us call them S_1, S_2, S_3, S_4, and S_5 as special. And there are ordinary dishes which are O_1, O_2, O_3, O_4, and O_5.

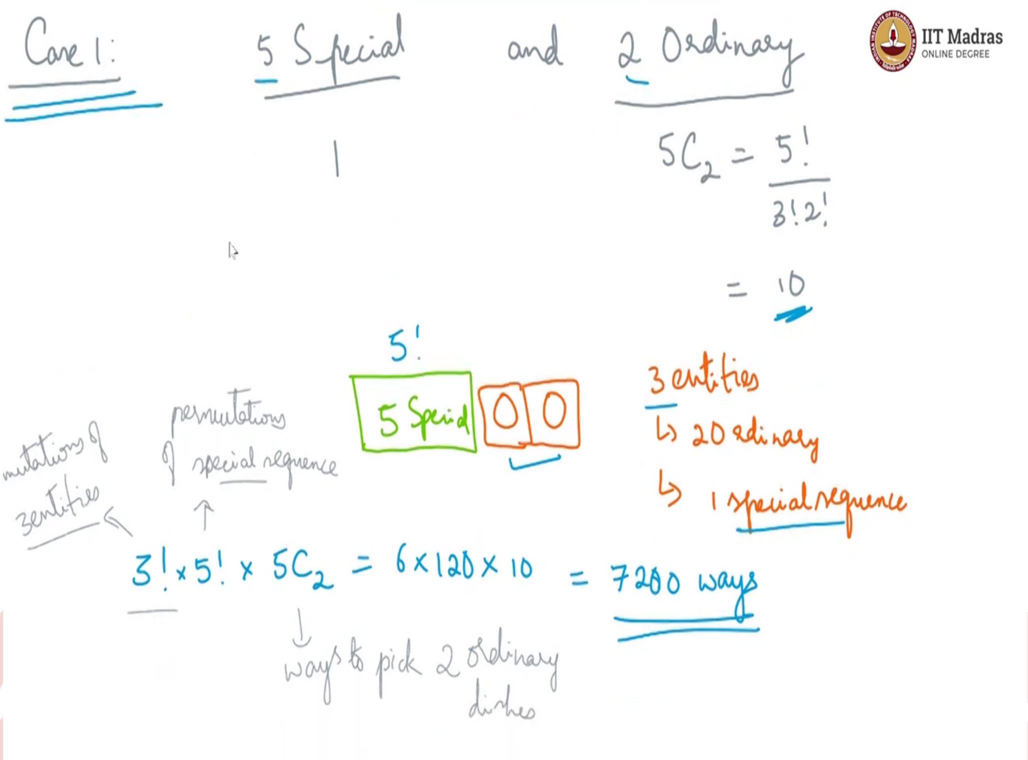
So these are the dishes and now, we have to serve these dishes; serve 7 dishes. Out of these 10, we have to pick 7 and serve them in a sequence. So we are looking at permutations, we are not looking at combinations. And they are saying at least 3; at least 3 special dishes should be served.

So of the 7, you can have 3 special dishes or 4 special dishes or all 5 special dishes. And all special dishes are served consecutively, that is, they are served together; there is no serving of an ordinary dish in the middle of the special dishes sequence.

So we consider the 3 cases which are case 1, let us consider the easiest one which is 5 special and 2 ordinary. Before we look at the permutations, the ordering, let us see how many ways we can pick these dishes in the first place without any sequencing. So there is only 1 way to pick the 5 special dishes because there are only 5 special dishes.

And here, you can do $^5C_2$ which is $\frac{5!}{2!\times 3!}$ which gives us 10. So in 10 ways, we pick 2 ordinary dishes and in 1 way, we pick 5 special dishes.

(Refer Slide Time: 02:14)



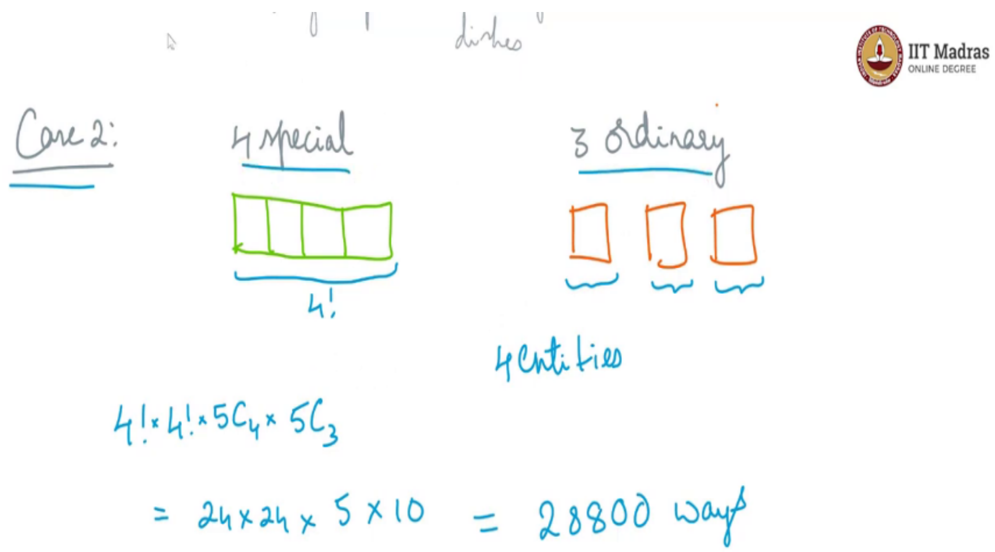
Now we have our 7 with us. However, these 5 put together, they make 1 block; 5 special. So other than that, we have 2 ordinaries. So in effect, it is like we are dealing with 3 entities which is basically 2 ordinary and the 1 sequence of 5 special dishes.

So now, these 3 entities can be permuted in 3! ways. And within them, our special sequence has 5! permutations within itself. So each of our 3 entity permutations gets 5! permutations. And then the picking of these 2 ordinary dishes happens in 10 ways. So we further multiply this by 10, which is basically ${}^5C_2$. And this should be our answer for the case 1.

So let us make sure we understand these terms. This 3! is to account for permutations of the 3 entities. And these 5 factorials are the permutations of the special sequence. So for every permutation of these 3 entities, you get 5! permutations of the special sequence. And for every such permutation of 3 entities, we have to pick, this is the number of ways to pick 2 ordinary dishes.

So for every such permutation, we have ${}^5C_2$ options of picking the ordinary dishes. So all put together, this is the answer we get for case 1, which gives us 3! is 6 into 5! is 120 into ${}^5C_2$ is 10. So we get 7200 ways. This is for case 1 where you have 5 special and 2 ordinary.

(Refer Slide Time: 05:03)



Now, let us look at case 2 which is 4 special. So these 4, we get in a sequence. And 3 ordinary, these we get as independent entities. So now, we have as, by our previous logic, we have 4 entities; these 4 special together is 1 entity, and then each of them is, each of the ordinaries is 1 entity.

So since you have 4 entities, we will have 4! permutations. So every permutation of our 4 entities also should account for the 4! permutations within the sequence. And now, we also have to account for how many ways we can pick 4 special items from 5. So we have ${}^5C_4$.

And how many ways we are picking 3 ordinary items from 5, which will give us ${}^5C_3$. So this is going to give us $24 \times 24 \times 5 \times 10$. So that is 28,800 ways for case 2, which is 4 special and 3 ordinary.

(Refer Slide Time: 06:51)

$= 24 \times 24 \times 5 \times 10 = 28800$ ways

Case 3:

3 special

4 ordinary

5 entities.

$5!\ 3!\ {}^5C_3\ {}^5C_4$

$= 120 \times 6 \times 10 \times 5 = 36000$ ways

That leaves us with case 3, which is 3 special and 4 ordinary. So these 3 special, they come together like this as a special sequence, whereas the 4 ordinary gives us 4 independent entities.

So this time, we have 5 entities leading to 5! permutations of these entities. And again, each of these permutations will have to consider the permutations within the special sequence, therefore, that is 3!.

And now, we look at how many ways we can pick these 3 from 5. So that is ${}^5C_3$. And again, picking these ordinary ones from 5 that is ${}^5C_4$. So this will give us $120 \times 6 \times 5 \times 10$ which is equal to 36,000 ways.

(Refer Slide Time: 08:13)

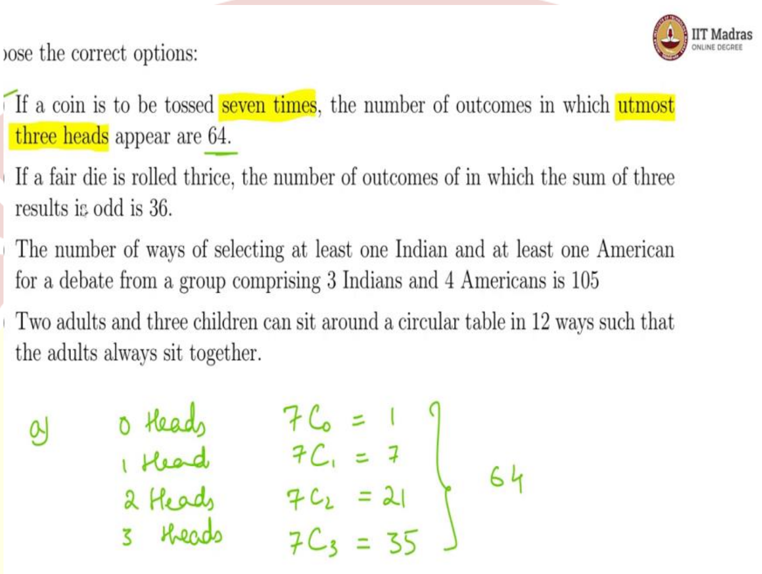
$= 120 \times 6 \times 10 \times 5$

$$
\begin{array}{r}
{\scriptstyle 2\ 1} \\
36000 \\
+\ 28800 \\
+\ 7200 \\
\hline
72000
\end{array}
\ \text{ways}
$$

So all put together, we are getting 36,000 for case 3, plus 28,800 for case 2, plus 7,200 for case 1. So that gives us 72,000 ways in all.

Statistics for Data Science - 1
Prathyush P
Support Team
Indian Institute of Technology, Madras
Week - 5 Tutorial - 3

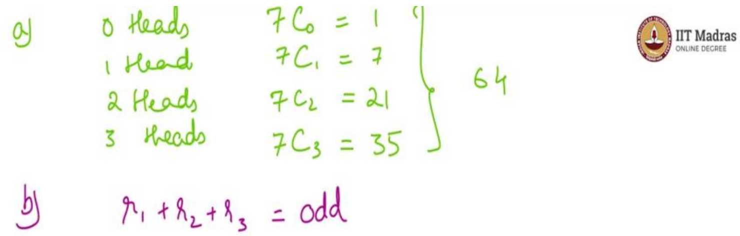
(Refer Slide Time: 00:14)

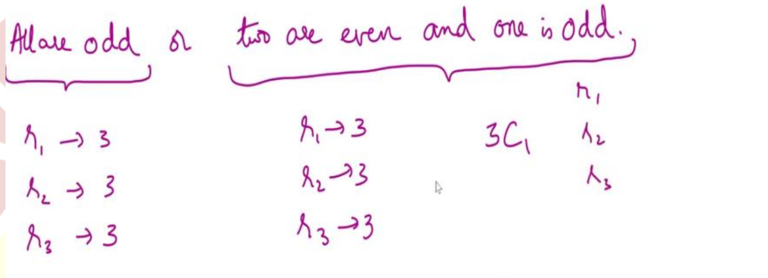


In this question, we are supposed to choose the correct options from these 4. So let us look at the first option. In the first option, a coin is to be tossed 7 times, the number of outcomes in which utmost 3 heads appear. So we need a maximum of 3 heads. So that gives us 4 cases which is 0 heads, 1 head, 2 heads, and 3 heads.

So 0 cases of the 7, we choose no toss at all. So that will become $^7C_0$ which is equal to 1. And for 1 head, we choose 1 toss of the 7 which is 7. And for 2 heads, we choose 2 tosses of the 7 which is 21. And finally, 3 heads is 3 choices from 7 which is 35. So all of these put together gives us 64 outcomes. So this is 64 and this is correct.

(Refer Slide Time: 01:41)





Now, for the second part, if a fair die is rolled thrice, the number of outcomes in which the sum of the three results is odd. So the sum of the results should be odd and that we are looking for the number of outcomes.

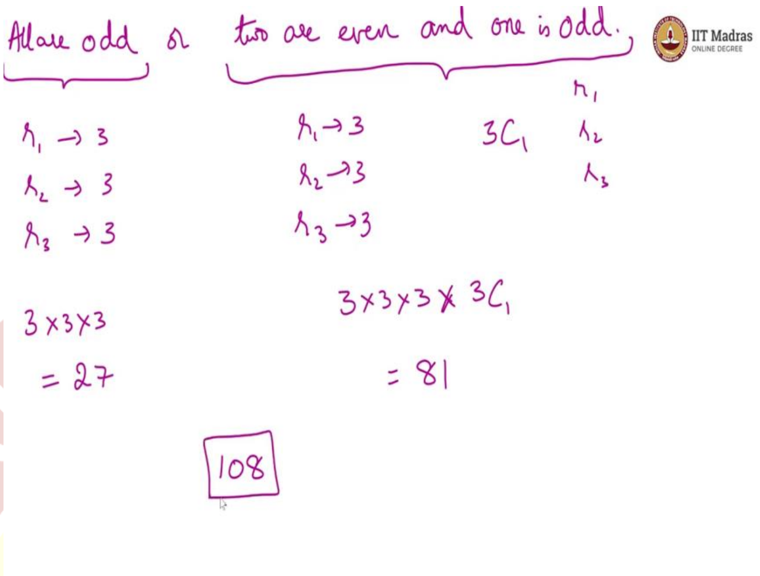
So the sum of the three results should be odd which means, let us call these results $r_1 + r_2 + r_3$, and this is odd. And this can only happen if all three of them are odd or two are even and one is odd. These are the only two cases which is, all are odd or two are even and one is odd.

So why are these only two cases? Suppose we consider the other cases where all or even. If all are even, you are going to get the sum as even, and if there is only one even and two odd, the sum of two odds will be even; so even plus even will give you even. So these are the only cases we have and now, let us look at them.

All are odd. So in the first result, we have 3 options; in the second result, we have 3 options. And in the third result, we have 3 options. Because you have 1 3 5, 1 3 5, 1 3 5 in all of these. Whereas, in the next case, again you will have, first result will have 3 options; second result will have 3 options, and third result will have 3 options.

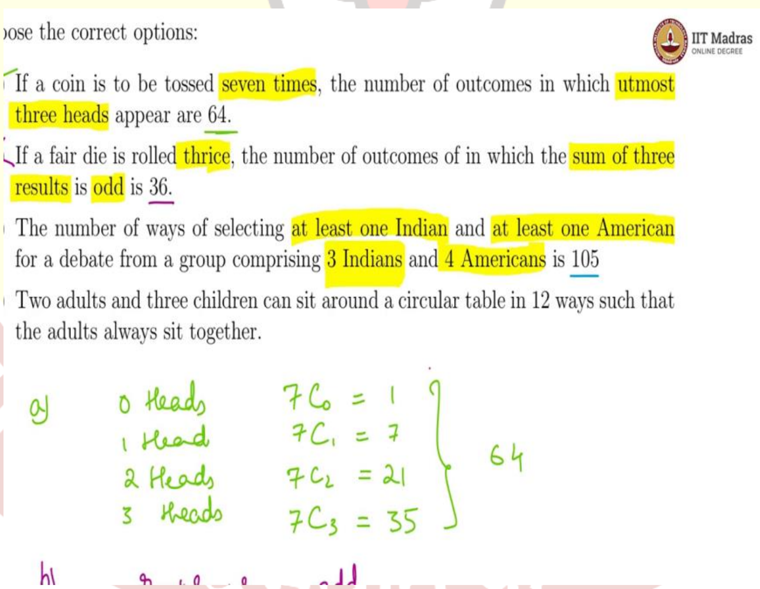
So these options are 3 3 3 because there are also 3 even numbers. So let us assume that r_1 is odd, then r_1 has the option of being 1 3 5, but r_2 and r_3 will then have the options of being 2 or 4 or 6. And now, there is a further concern of which one is the odd result. So that can be chosen in $^3C_1$ ways which is basically r_1 is odd or r_2 is odd or r_3 is odd.

(Refer Slide Time: 04:11)



Now, if we count the total number, will have $3 \times 3 \times 3 = 27$ for all are odd. Whereas, in the other case, will have $3 \times 3 \times 3 \times {}^3C_1$=81. The sum is then $27 + 81 = 108$.

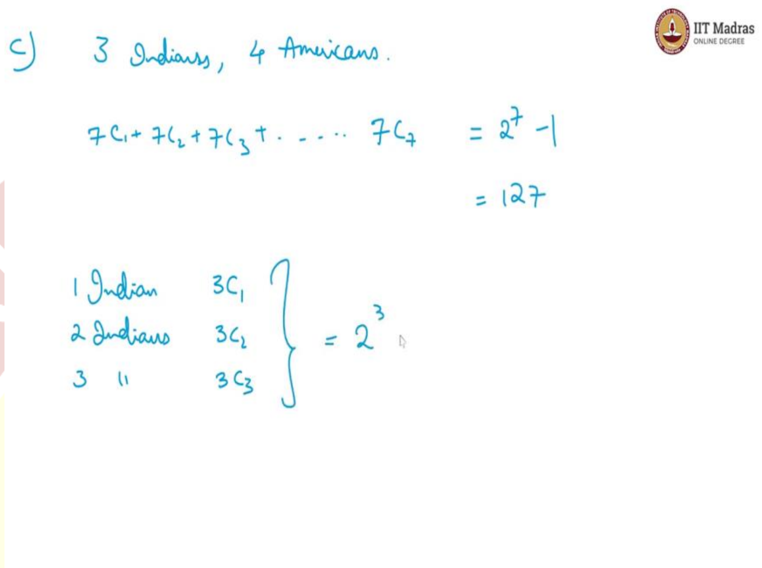
(Refer Slide Time: 04:29)

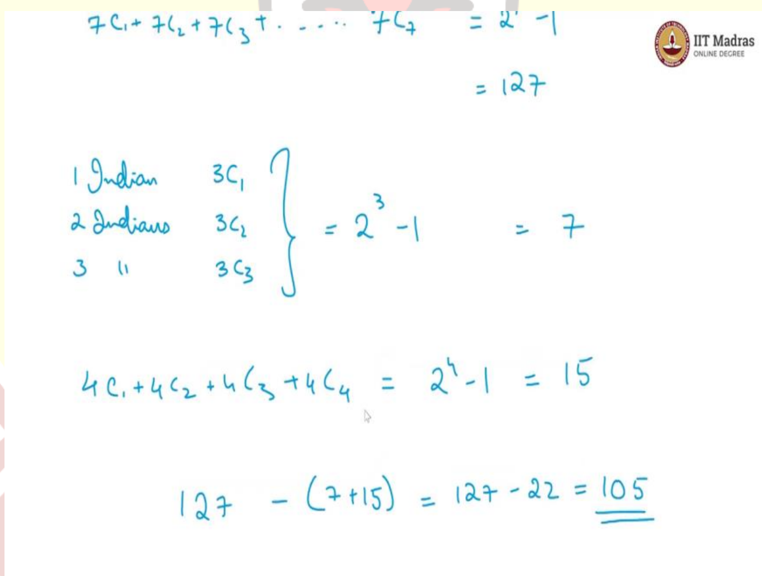


But here, they are saying it is 36 which is not true. So this is not true. Going further. The number of ways of selecting at least one Indian and one, at least one American for a debate from a group comprising of 3 Indians and 4 Americans is 105. So we have at least one Indian and at least one American.

Now, the problem does not say how many people need to be selected. So presumably, you can choose any number.

(Refer Slide Time: 05:13)



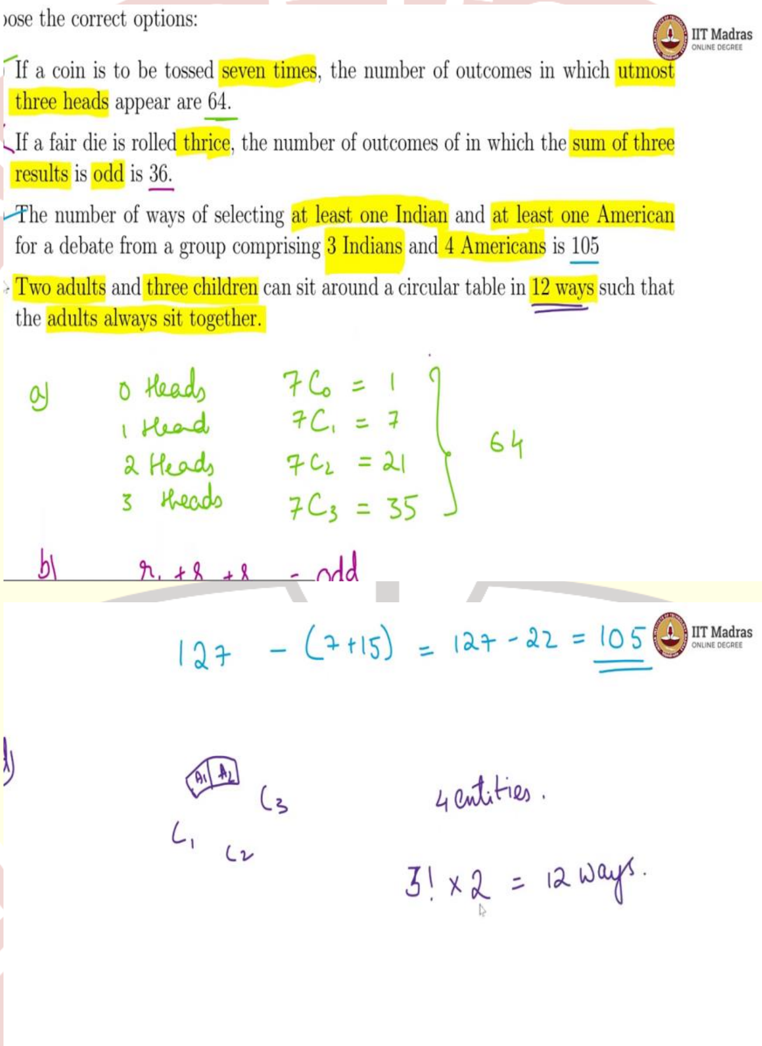


So if you can choose any number, there are 3 Indians and 4 Americans. So there are 7 in all. So if you can choose any number, what you are getting is ${}^7C_1$ plus ${}^7C_2$ plus ${}^7C_3$, so on till ${}^7C_7$. This is if there are no restrictions. And this is equal to 2^7-1, that is, 127.

Of these, we should remove the cases where there are no Indians or no Americans. So the cases where there are no Americans let us take, for example. Then you can pick 1 Indian or 2 Indians or 3 Indians. So that will be ${}^3C_1$ plus ${}^3C_2$ plus ${}^3C_3$ which again is equal to 2^3-1.

And similarly, if we choose 1 American or 2 Americans or 3 Americans or 4 Americans without any Indians, you will get ${}^4C_1$ plus ${}^4C_2$ plus ${}^4C_3$ plus ${}^4C_4$ which is equal to 2^4-1. This is 7 and this is 15. So of the total possibilities, we are subtracting 7 plus 15 which gives us 127 minus 22 which is equal to 105.

(Refer Slide Time: 07:07)



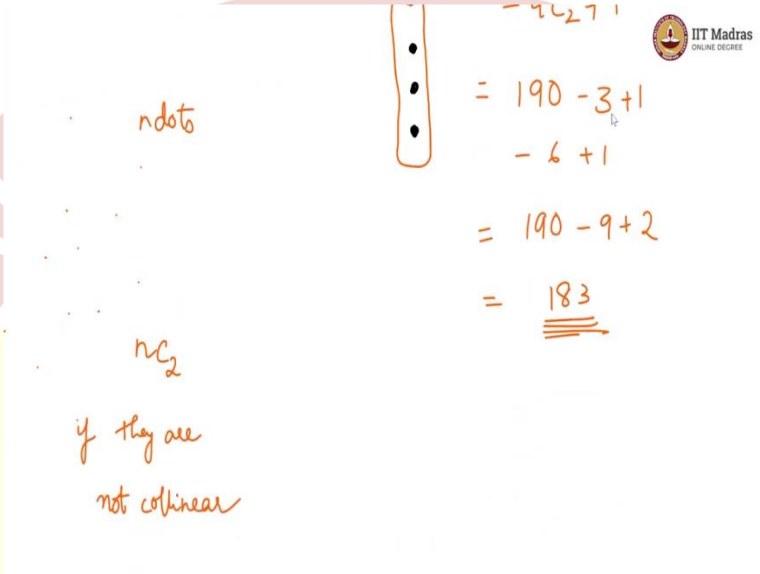
So this is a number of ways to pick a debate team in which there is at least one Indian or one American and 105 is correct. So C is also correct. Now, let us look at the last option which is, 2 adults and 3 children can sit around a circular table in twelve ways such that the adults are always sitting together.

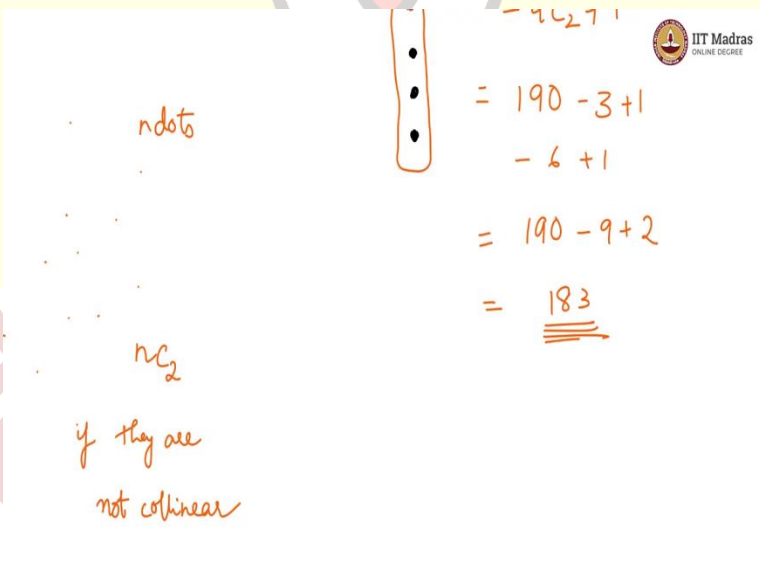
So then, we have a circular permutation here, there are 2 adults. Now, these 2 adults are treated as one entity, and then they have C_1, C_2, and C_3 who can sit around them, the three children. So technically, these are 4 entities totally and in a circular permutation for n entities, you will get n minus 1 factorial, that is, 3 factorial ways in this case.

But the A_1 and A_2, the adult 1 and adult 2 can be interchanged in each of these permutations so you get additionally 2. That gives us 6 into 2, 12 ways. So there is these 12 ways, which means D option is also correct.

Statistics for Data Science - 1
Prathyush P
Support Team
Indian Institute of Technology, Madras
Week - 5 Tutorial - 4

(Refer Slide Time: 00:14)





In this question, they are saying that there are 4 black dogs and 16 brown dogs and they are running on a ground such that no three dogs are in a straight line. So basically, no three of them are in a straight line. Any two of them will be on a straight line because a straight line is defined by two points. So I need two dogs will be on a straight line but they are saying no three are collinear together.

Now, suddenly, two of the brown dogs started following one black dog. So we have two brown dogs following one black dog. So these three have now become collinear. And then further, the remaining three black dogs started following one of the remaining brown dog.

So the other three black dogs are now following one brown dog. So these four are also on a straight line. And then they are asking how many straight lines can be found passing through the dogs.

So now, suppose there are n dots which are all non-collinear; suppose there are n dots. So in order to make a straight line, you just have to pick two of these n dots. And how many ways can you pick is the number of straight lines you get, that is basically ${}^{n}C_2$. And this is if they are not collinear.

Now, of course, any two points will be collinear. The funda is no three should be collinear. So if no three of them are collinear, you choose any two you will get a unique straight line. If the points are not collinear, you get ${}^{n}C_2$ straight lines.

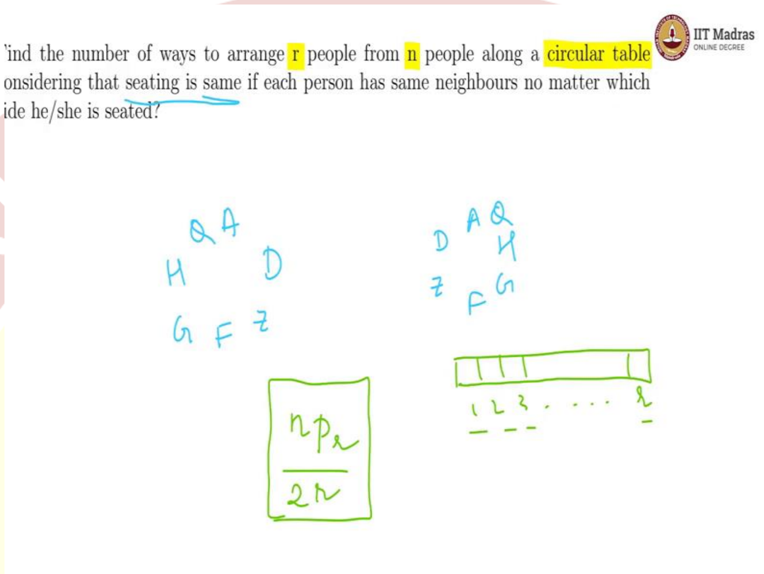
However, in our problem, there are 20 dots. And so, in this case, the each dot is a dog but there are some collinear. So what we do is we will calculate the number of straight lines for all 20 when they are not collinear, which is ${}^{20}C_2$ minus the number of lines which should have come from this, which is ${}^{3}C_2$ but we include the one line that is all coming from this particular three dogs being in a straight line. So we will add 1.

And then similarly, we will subtract the lines that were supposed to come from this which is ${}^{4}C_2$ but we will add the one individual line that they make anyway, which is 1. So we now have ${}^{20}C_2 - {}^{3}C_2 + 1 - {}^{4}C_2 + 1 = 190 – 3 + 1 – 6 + 1 = 190 – 9 + 2 =183$. So there are a 183 unique lines you can draw through these dogs.

Here, the questions is how many straight lines can be formed passing through the dogs. If you are only considering one dog per line, then you can get infinite. So what it actually is, is supposed to be how many straight lines can be found passing through at least two of the dogs; that is the actual question and that number is 183.

Statistics for Data Science - 1
Prathyush P
Support Team
Indian Institute of Technology, Madras
Week - 5 Tutorial - 5

(Refer Slide Time: 00:14)



In this question, we are expected to arrange r people from n people along a circular table. So it is a circular permutation. However, they are saying that the seating is same if each person has same neighbors no matter which side he is seated.

So let us consider this. We have, say, A, D, Z, F, G, H, and Q sitting together like this. According to what their saying, as long as the person has same neighbors, so if I flip this instead of clockwise to anti-clockwise; suppose I have A, and this side that is D, there is Z, and then F, and G, and H, and Q.

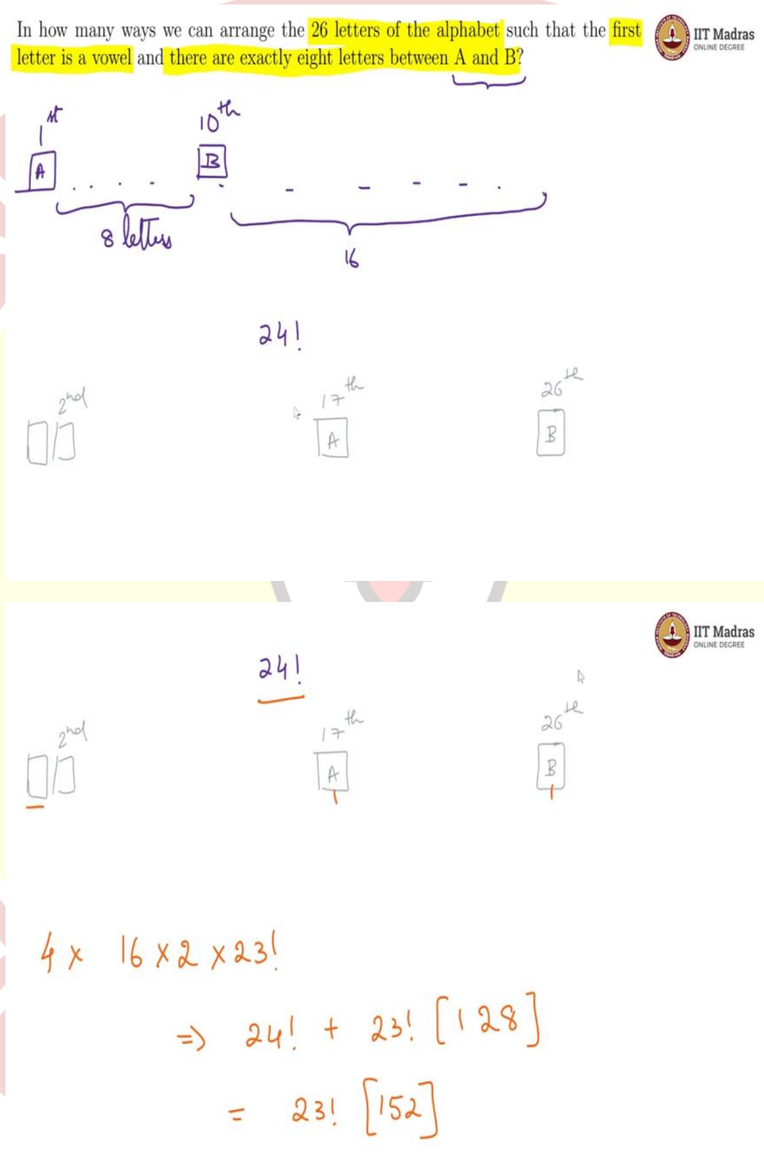
So now, A has the same neighbors D and Q, so here also Q and D; H has the same neighbors Q and G, Q and G; similarly, F has G and Z, and F has G and Z. So what has happened here is this is clockwise and this is anti-clockwise and they both are supposed to be considered the same. So the number of circular permutations we will get, first of all, if we look at it is taking r out of n people.

So we can first arrange them as nP_r, which will give us in a sequence like this. So this is 1, 2, 3, so on till r. Now, in a circular permutations, you get r repetitions because it does not matter which person you start from. So each of these can be the starting point and it will be the same circular permutation, so we will get r repetitions. So, therefore, divided by r.

And now, since clockwise and counter-clockwise does not matter, each of these is repeated once which means you have to divide by 2. So this will be the answer for our question. Thank you.

Statistics for Data Science - 1
Prathyush P
Support Team
Indian Institute of Technology, Madras
Week - 5 Tutorial - 6

(Refer Slide Time: 00:14)



In our last question, we are looking at the number of ways we can arrange the 26 letters of the alphabet, that is, A to Z such that the first letter is a vowel. So the first letter is A or E or I or O or U and there are exactly 8 letters between A and B.

So this is the first box and then we have the rest. And this can be filled in 5 ways but since we are looking at this condition here, let us start with filling it with A. Then B will be somewhere here, in the 10th position; this is the first position and there are 8 letters in between. So now, these are 10. So what is left is the remaining 16 letters and these 8+16, the 24 letters can be rearranged in 24! ways and all of them are valid by our conditions.

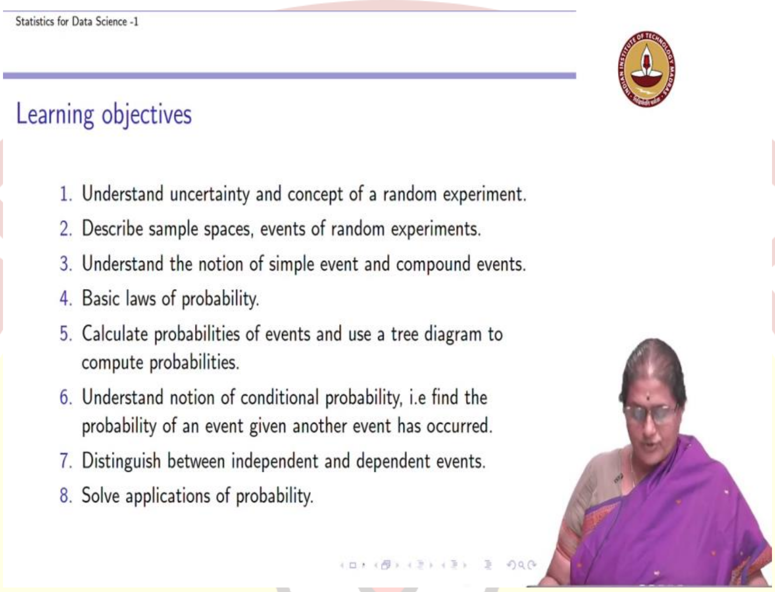
Now, suppose the first letter is not A. So then the first letter can be E or O or I or U. And we also know that the last letter if it is B, then A has to be the, if this is the 26th, then A has to be the 17th letter so as to accommodate 8 letters in between which mean A can go anywhere from the 2nd position to the 17th position.

So we have 4 choices for the first one which gives us 4. And now A can go from, A can be anywhere from 2nd to the 17th. So there are 16 choices for A and since you can shuffle A and B in this case, I mean B can come first and A can come afterwards, so we can multiply by 2 now. And then what are the remaining letters? First letter is fixed and we have A and B filled. So 3 letters are filled. So the remaining 23 can be rearranged in 23! ways.

So we earlier had 24 factorial and now, we have $4 \times 16 \times 2 \times 23!$, which gives us $24! \times 23!\,(128)$. If we take further 23 factorial common, we will get $152 \times 23!$.

Statistics for Data Science 1
Professor Usha Mohan
Department of Management Studies
Indian Institute of Technology, Madras
Lecture 6.1
Probability – Basic Definitions

(Refer Slide Time: 00:24)



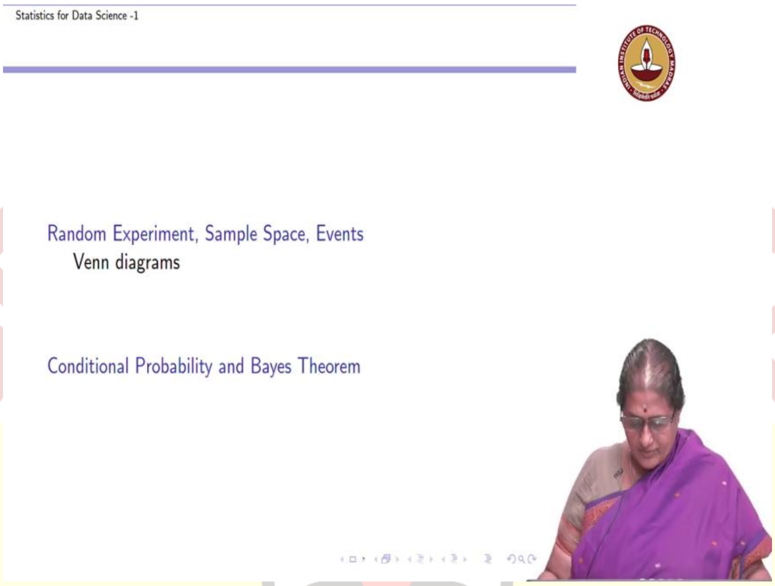
So having learned about permutations and combinations, we now move on to learn about the fundamentals of probability. So what are we expected to learn in the next two weeks is, we are going to first understand what we mean by uncertainty. And we will understand about a very important concept that is random experiment.

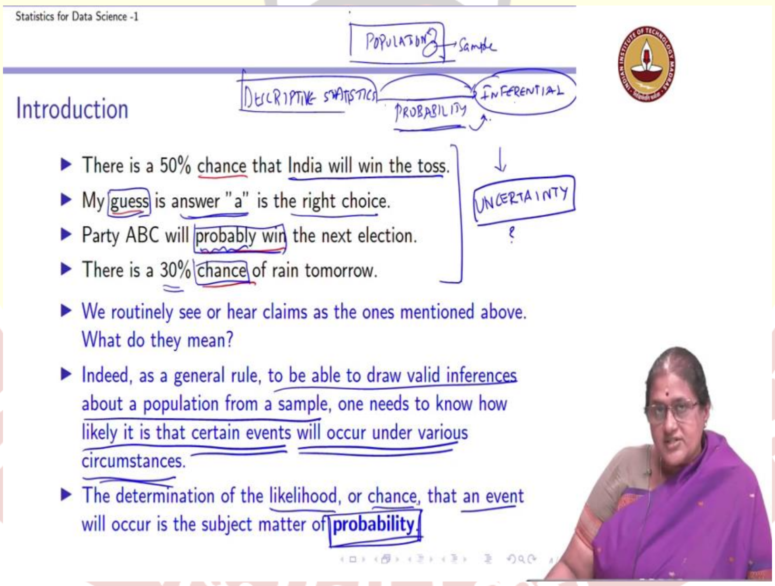
Now, the concept of random experiment and what are events, sample spaces, events these are extremely important in our understanding of probability. We then move forward to understand what is the notion of simple event and compound event, we will introduce the basic laws of probability.

We will spend some time to compute the probability of events through visual representations which we call tree diagrams. Then, a very important notion of conditional probability is introduced. Because we need the notion of conditional probability to define what we finally know, distinguish or to distinguish between what we refer to as independent events and dependent events.

We will also solve applications of probability throughout the module. So, these are, this is what you should be knowing at the end of these two weeks, which we have devoted to understand the concepts of probability.

(Refer Slide Time: 01:56)





Now, let us move forward and understand what, how to set up this formal definition of probability. Every time we hear things like this that there is a 50% chance that India will win a toss when you have captains who go for a toss at the beginning of a game, say the game of cricket. Usually, they toss the coin and the captains decide whether they want to field first or bat first.

So, we always say when there is a toss of a coin, there is a chance of the person winning the toss of a coin, typically we assign a 50% chance of a person will or a country would win a toss. In the, if we do not know the answer to a multiple choice question, most of the times we guess an answer. Suppose I guess an answer A, then afterwards I want to say that okay, this guess I do not know whether what I have guessed is the right answer or not.

Similarly, when we come to elections there is always we hypothesize or we say that a party would probably win the next election. Similarly, we have also heard statements of the kind that there is a 30% chance of rain tomorrow. Now, when you look at these statements, you always see that in the first statement, there was something called a chance that we are talking about. Again, my guess probably win chance of rain. So what we see immediately is that there is an element of uncertainty.

What do we mean by this element of uncertainty? We are not sure that are certain that India will win a toss. India could win, could not win. So the outcome is uncertain. I am not confident that answer A is the right choice. I am guessing, whenever I guess I am not certain. The uncertainty element is there. Similarly, we are not sure, we are not certain that party ABC will win the next election. We are probably it would win the next election.

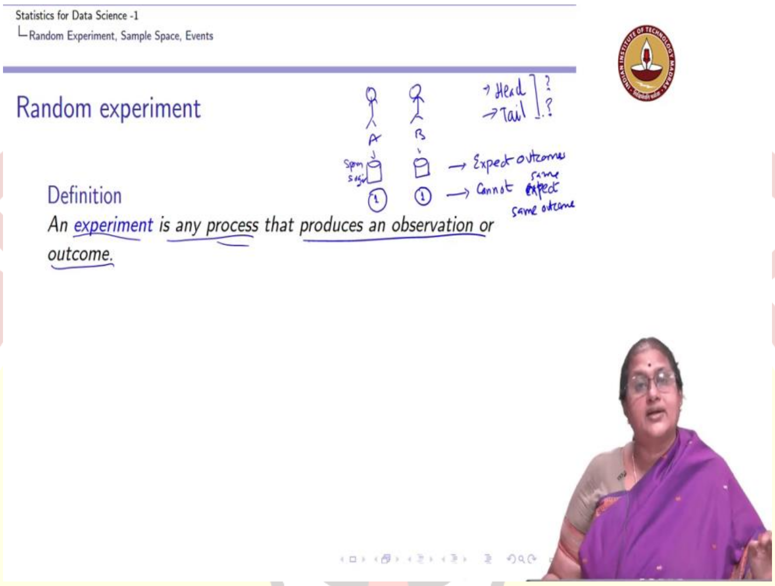
Similarly, I am just telling that there could be a chance of rain. Am I certain it would rain tomorrow? The answer is no. I am just saying there could be a chance of rain and I am quantifying that chance with a 30% chance. So, we routinely keep hearing claims of this kind, where we are addressing this element of uncertainty.

So, why do we need to learn about this here? Remember when we talked in our descriptive statistics module, you said that we have a larger subset or larger set which we refer to as a population. We refer to the smaller set, subset of a population as a sample. And, I said I want to draw valid inferences about this larger set using this smaller subset.

So, if I want to draw valid inferences about a population from a sample, I need to know how likely it is that certain events will occur under certain circumstances. So, if I am want to move from my descriptive statistics which we learned in our first module to inferential statistics which you will be learning in your second course.

I need certain tools to capture the uncertainty that would manifest here and that uncertainty that is the determination of the likelihood or chance that an event will occur is what is the subject matter or what we refer to as probability. Hence, if I have to go from this to this, I need to understand probability and that is what we are going to do in the next two weeks. We are going to set up the foundation because this is again remember a foundational level course. We are going to set up this foundation which will help you understand inferential statistics.

(Refer Slide Time: 07:18)



So, we start this foundation by setting up the building blocks of help in create this framework. Now what is an experiment? An experiment is any process that produces an observation or outcome at monitoring the time taken to, for example, if I am just monitoring the time taken to measure, Okay, certain output from us certain machines, Okay, or if I am measuring the diameter of nuts or diameter of bolts which are produced by a certain machine.

This is my experiment because it is producing an observation. What is the observation? I am measuring a particular diameter, Okay. Now, I can define an experiment as a process that produces any observation or outcome. Now, if I conduct an experiment in a completely controlled setup.

For example, I am just having two people; let me call this as person A and this is person B and I give each one of them a glass of water, same identical glass of water and I asked them to add a spoon of sugar to the glass of water. Same amount of water, identical, and asking them to add a spoon of sugar and ask them to stir it, we would expect both the outcomes to be the same because we are having exactly the controlled experiment.

Both of them are given the same type of glass, everything is similar. I am asking them to just add a spoon of sugar. Sugar is also same. Everything is absolutely identical, same. They are just mixing it. I would expect the both the outcomes to be the same. Now let us repeat another experiment. This is an experiment because it is a process, it is producing an observation. What is observation it is producing? I am mixing water and salt.

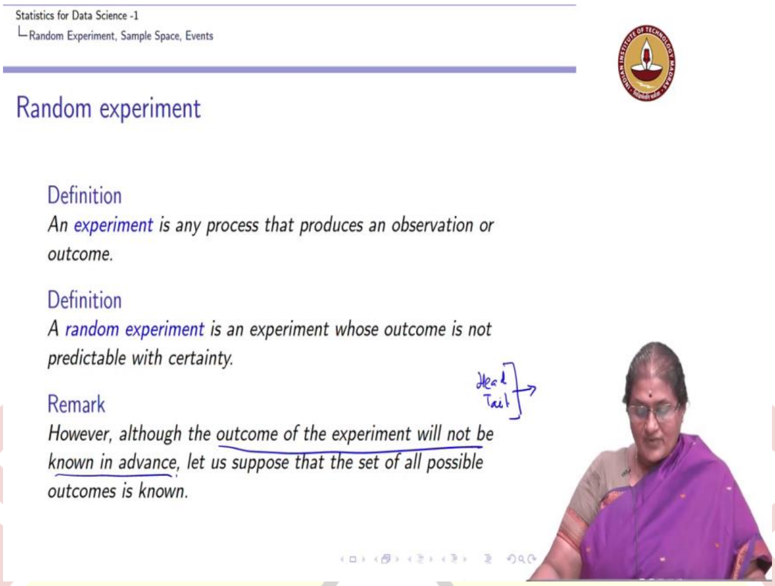
So I am seeing what is the outcome, it is going to be a mixture of sugar and water this is what I am mixing. Now suppose I give both of them 1 rupee coin, okay I am giving both of them a 1 rupee coin and I am asking them to toss the coin, that is each one of them gets identical 1 rupee coin and both of them are asked to do the experiment of tossing a coin.

So, when you toss a coin, you have an outcome of the experiment. Now when I toss a coin, now, in this case, I cannot expect the same outcome from both. So cannot expect same outcome, even though they are in the same setup, identical setup, identical coin, identical force, thrust everything given to them, we cannot expect the same outcome.

So, in the first case, it is a some sort of a deterministic experiment, whereas in the second case, because my outcome is uncertain when I toss a coin, I do not know which of the outcomes would come. So, when it toss a coin, I know, I could either get a head or a tail in a coin. I know the outcomes, but I do not know whether it would be a head or a tail.

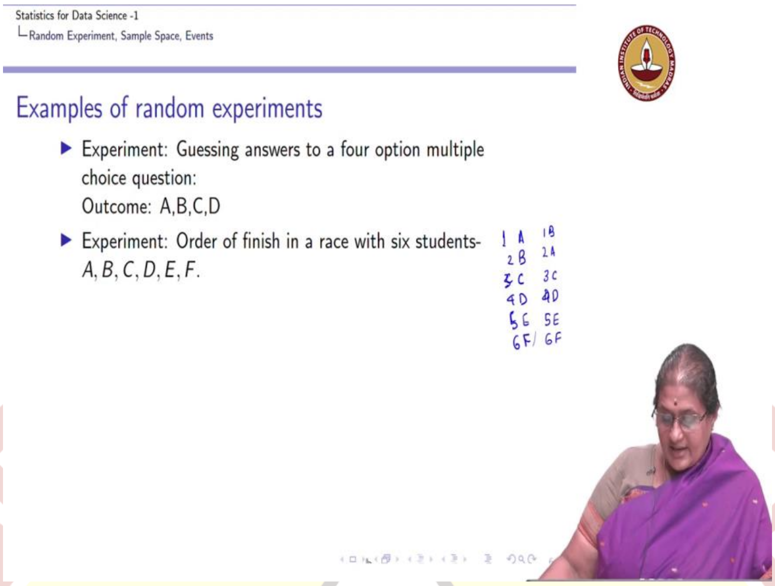
So the uncertainty is which of the outcomes is actually going to happen. It is not a random experiment. It is not an experiment with unknown outcomes. Here, I know what are my outcomes. In this experiment, I know my outcome is either a head or a tail but the uncertainty is whether it is a head or tail.

(Refer Slide Time: 12:05)



So, we formally define what is a random experiment as an experiment whose outcome is not predictable with uncertainty or the outcome is uncertain. So, this is an extremely important understanding. So, here I want to qualify it. The outcome of the experiment is known in advance. So, when I am tossing a coin, I know it would be a head or a tail. But I do not know whether it would actually be a head or tail, though two possible outcomes I know what are the possible outcomes, but I do not know whether it would be actually a head or a tail.

(Refer Slide Time: 12:53)

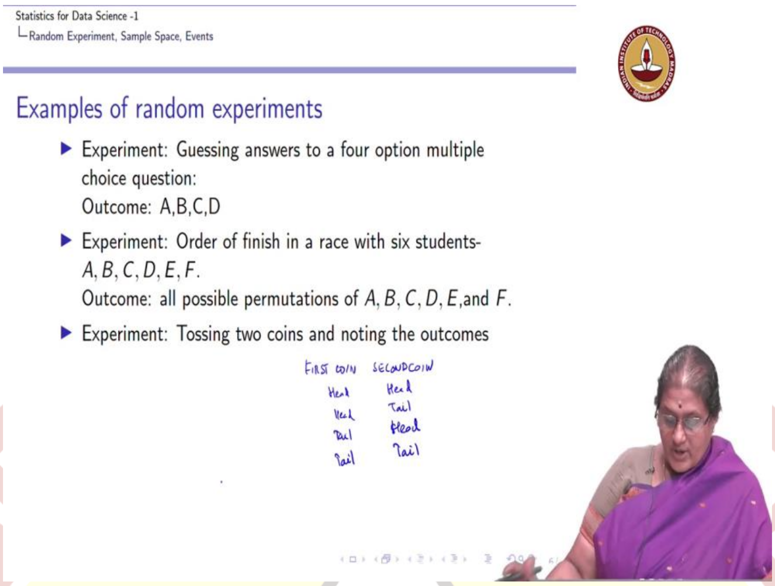


So, now let us look at examples of a few random experiments. Suppose, my experiment is guessing answer to a four option multiple choice question. Let my options be A, B, C, D. The possible outcomes of this experiment are I have A, B, C, and D. Those are the possible outcomes because the experiment is guessing answer. I just have four possible answers because it is a four option multiple choice question. So, my outcomes are A B, C, D, where these are my options.

Order of finishing a race with 6 students, I have A, B, C, D, E, F. By order of finish I am, I mean that for example, one order of finishes, A comes first, B comes second, C comes third, D comes forth, E comes fifth and F comes sixth. This is one order of finish. Another order of finish is, B comes first, A comes second, C comes third, D comes forth, E comes fifth and F comes sixth. This is another order of finish.

So, the experiment is to actually note down what are the order of finish is when I have 6 students? What are the possible ways the order of finish can happen? So, we can see that this is one particular order, this is another particular order. We quickly recognize that this is nothing but the number of ways you can arrange the six people.

(Refer Slide Time: 14:52)

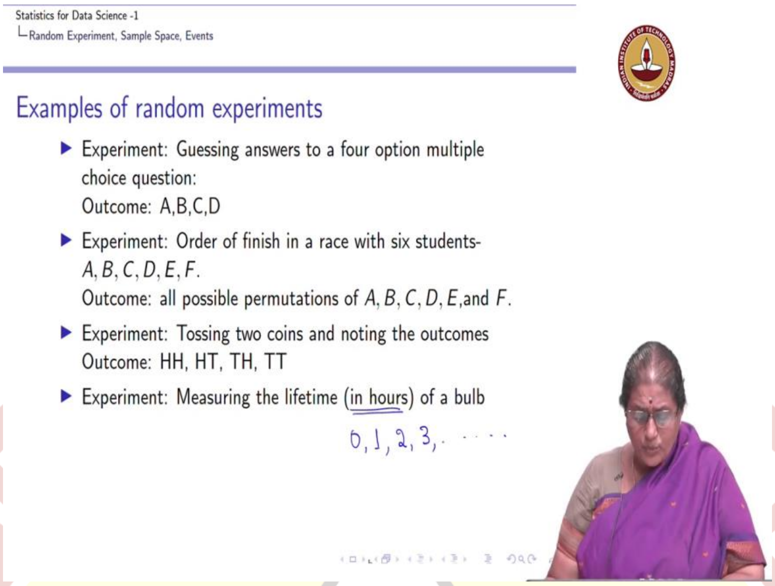


So my total number of outcomes is nothing but I can list all possible permutations of A, B, C, D, E, and F. Those are my possible outcomes of the experiment where I need the order of finish. Now, why is this a random experiment? Because when six people are competing in a race, I do not know who is going to come first or second. There is an element of uncertainty. We do not know who is going to come first, second, third, fourth, fifth, and sixth, there is an element of uncertainty.

Any one of them could come first or second or any of the positions. So, the outcome is all possible permutations of the 6 students who are named A, B, C, D, E, F. Now, suppose I toss two coins, I am tossing two coins. So, what are the possible outcome? I have my first coin. I have my second coin.

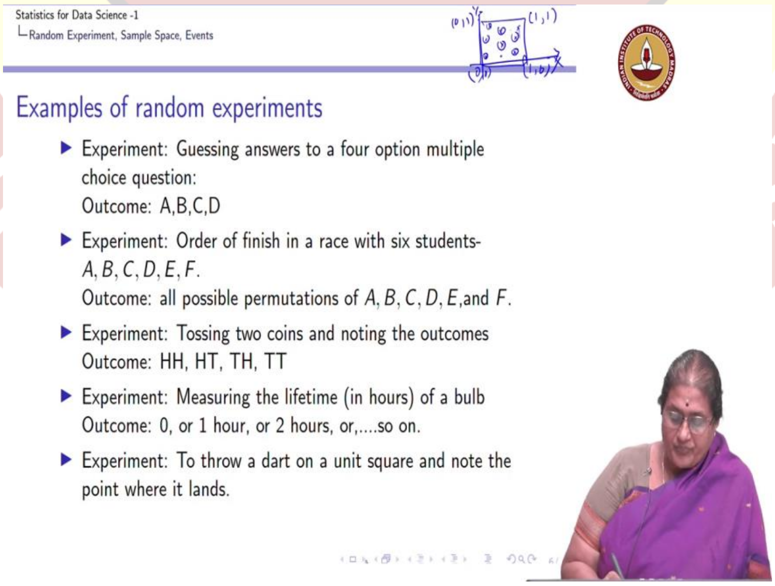
So, the possible outcome is I have a head on my first coin a head on my second coin. I have a head on the first coin, a tail on my second coin. A tail in my first coin, a head in my second coin. A tail in my first coin and a tail in the second coin. These are the possible outcomes of the experiment of tossing both the coins together.

(Refer Slide Time: 16:33)



So, I can represent that as HH, HT, TH and TT. Now, suppose the experiment is to measure the lifetime in hours. So, if, if a bulb is lasting even for 30 minutes, I will say that it is lasting for zero hours, if it is lasting for 1 hour, 30 minutes, I says that it has lasted for 1 hour. So that is how we are measuring the lifetime of a bulb. So, what are the possible outcomes? It could last for 0 hours, or 1 hour or 2 hour or 3 hours, I can keep going. So it could be 0 hours, 1 hour, 2 hours or 3 hours.

(Refer Slide Time: 17:19)



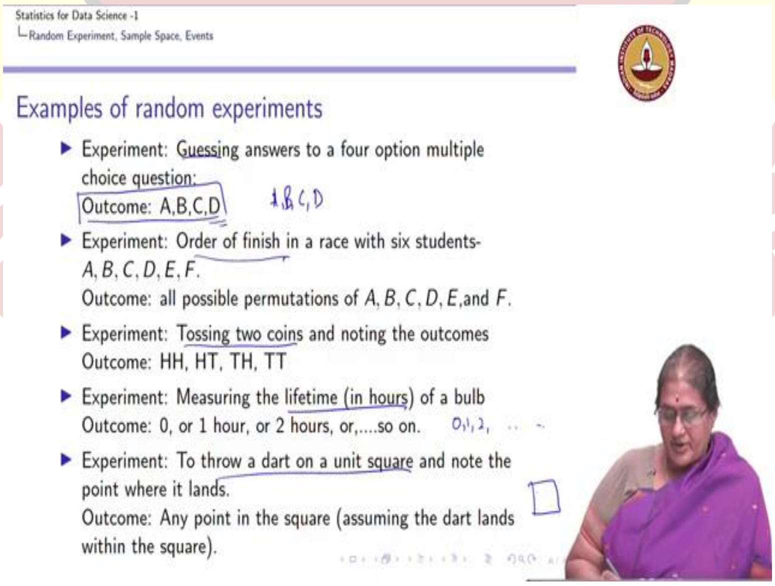
It could be any amount of time. So, it is 0 or 1 hour or 2 hours or so on. These are the possible outcomes. I do not know how long a bulb is going to last. It could last for 30 minutes, 45 minutes, in which case my outcome is, I am going to record it as a 0 hour. If it lasts for or it depends, if it, if I decide that anything less than 1 hour, I will record as 1 hour.

Anything less than 2 hours, between 1 hour and 2 hours, I would record as 2 hours. It would be 30 minutes is 1 hour. It depends on how you are defining this convention. So, you can see that the outcome again is uncertain. It is a random experiment because I do not know how long would any light bulb last. It could be anything between 1 hour, 2 hours, 40 hours, 45 hours, we do not know. So, it was a random experiment with whatever is the outcome.

Similarly, let us go to the next experiment. The next experiment we are going to look at is the experiment where I am throwing a dart. So, suppose I have a unit square here. For example, if I am having an unit square, this is the point 0, 0; this is the point 1, 0; this is a point 0, 1; this is my X axis, this is my Y axis and this is a point 1, 1. I am throwing a dart and I assume that this dart is going to actually land within the square.

Now, the dart could land at this position, it could land at this position, it could land anywhere. So, there is an uncertainty as to where this dart actually lands. So, if I am if my experiment is to throw a dart and my outcome is to note down, the observations I am going to note down is where it is actually landing in this unit square. So, that is my observation so where it lands.

(Refer Slide Time: 19:22)

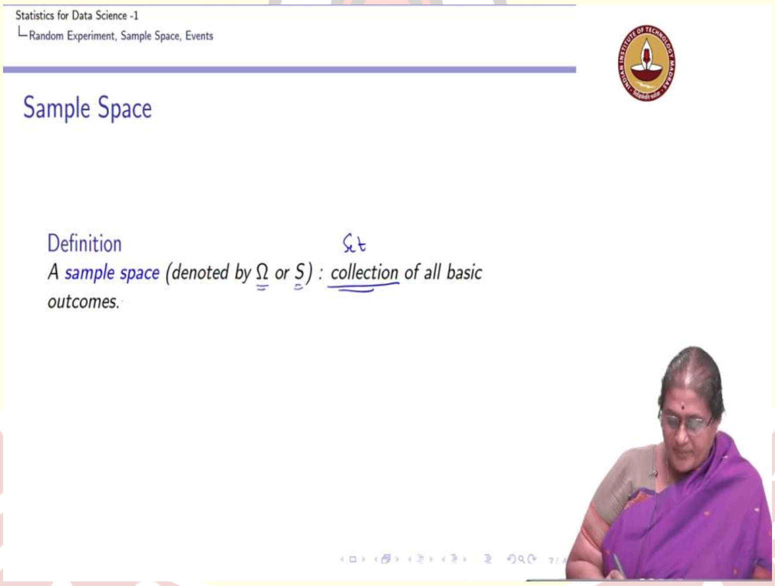


So, any point in this square is a possible outcome to the random experiment. Why is it random? I know it would be any point in the square but I really do not know which of these points in this square would be the actual outcome. So, these are examples of random experiments. You can see that they are random because I am guessing order of finish is not known. When I am tossing a 2 coin, it could be either head or tail. Measuring lifetime, it could be anything. There is an element of uncertainty.

Similarly, when I am throwing a dart in a unit square, it could be any point in the unit square. So, these are some examples of random experiments. Now, when you look at these examples, here I had a finite, so A, B, C, D only four choices, they were discrete finite. Whereas here, I can writing it as 0, 1, 2. You can see that again, when I am recording it as 0, 1, 2 and so on again discrete choices, and I am recording it as hours, but it is infinite.

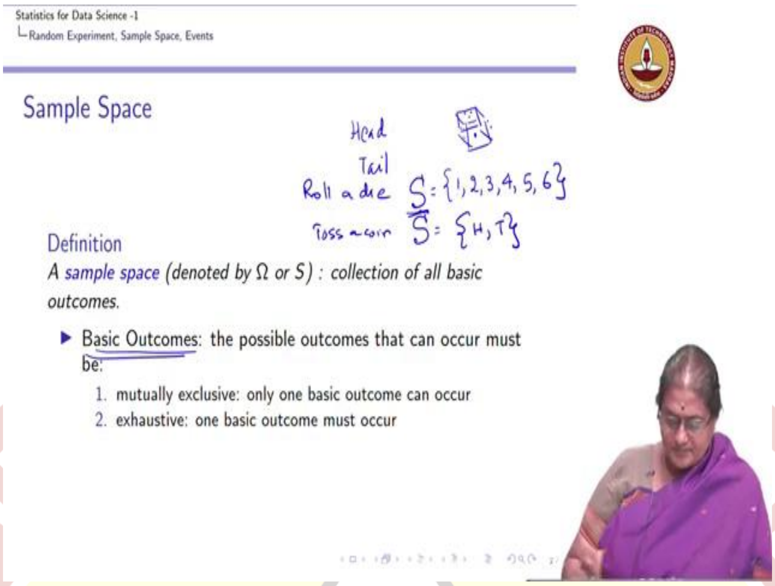
Now again, when I look at this point in the square, I am recording it as a coordinate. Again, there are infi... countable, but there are many points in this square. So immediately, you see that I need to have a way to represent all the outcomes of this random experiment.

(Refer Slide Time: 21:07)



And, the formal way to do it is through what we define as a sample space. So, some books use the notation omega to represent a sample space, I am just going to use the notation S, because it is a collection or set of all basic outcomes of a random experiment.

(Refer Slide Time: 21:33)

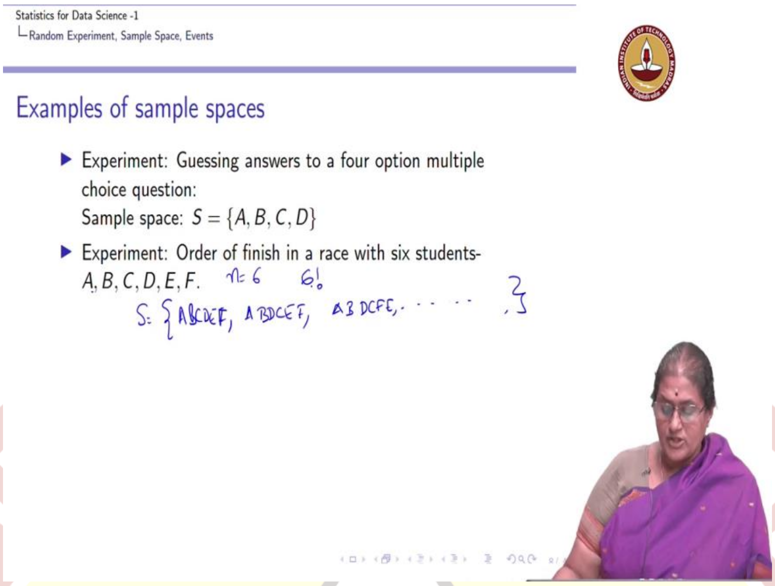


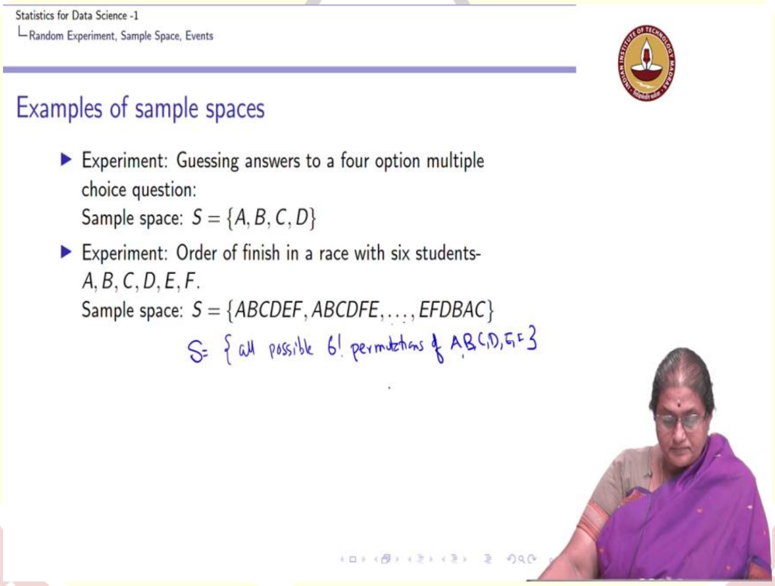
Now, what do we mean by basic outcomes? The basic outcomes are all possible outcomes of a random experiment which should be mutually exclusive. By mutually exclusive I mean that only one not the outcome can occur. So when I am tossing a coin, the outcomes are head or a tail, these are the 2 possible outcomes. If a head occurs, a tail cannot occur; if a tail occurs, a head does not occur. So, this is what we refer to as a basic outcome.

For example, if I roll a die, die you all know what is a die. So, this has one dots here, it might have 3 dots here, it has 4 dots here, it might have 2 dots here and this is a die. So, the possible, when I roll a die whatever turns up on the top is what I record as my observation. It could be either a 1 or a 2 or a 3 or a 4 or a 5 or a 6. So, one of the basic outcomes will occur, I represent this by my sample size.

So when a toss a coin, my S is represented by head, tail, when a roll a die, I have a sample space. This is for tossing a coin. When I roll a die, I have sample space which can take the values 1, 2, 3, 4, 5 and 6. And it is exhaustive in the sense that when I roll a die, one of these outcomes have to occur. When I toss a coin one, it should either be a head or tail. So, this collection or the set of all basic outcomes of a random experiment is defined to be what we refer to as a sample space.

(Refer Slide Time: 23:34)

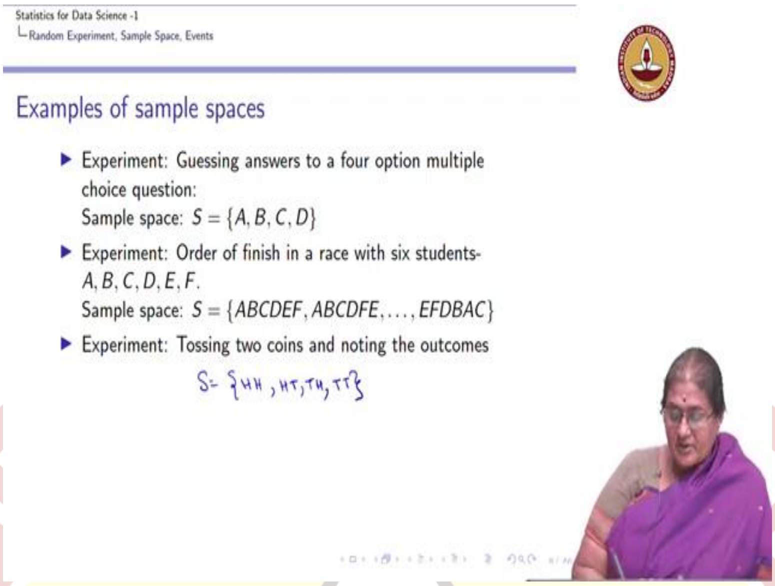


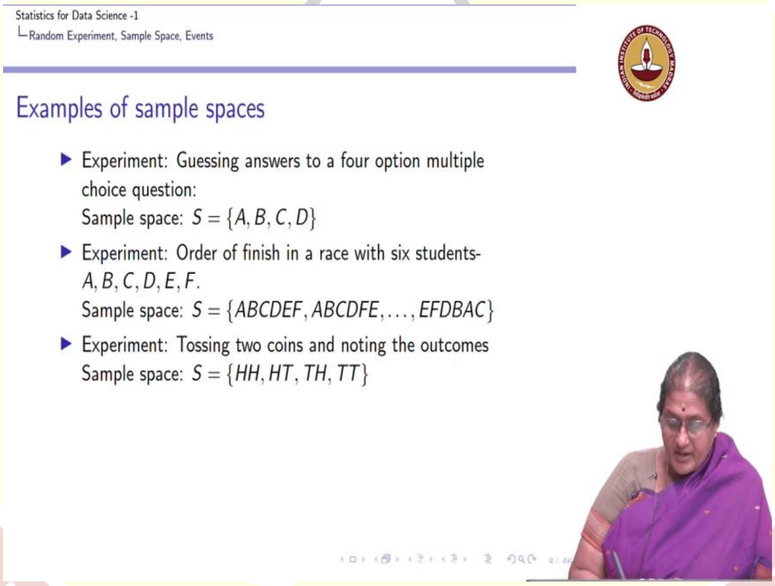


So, now let us go back to the experiments which we have seen earlier and represent them now as sample space. Recall sample space is a set of all outcomes of my experiment. So guessing answers, the sample space is going to be I have 4 outcomes so A, B, C, D I represent S equal to A, B, C, D. Order of finish I have, so now, my sample space is going to be ABCDEF which is one order of finish. ABDCEF, this is another order of finish. ABDCFE, so forth all possible and I know my n=6 here. So, I have 6! permutations which are possible.

So, my sample space is going to have a total of 6! elements which I can represent as ABCDEF. The dots represent all the other. I am not listing all the 6! elements here. Or I can write it as all possible permutations, 6! permutations of my A, B, C, D, E, and F form my sample space in this case.

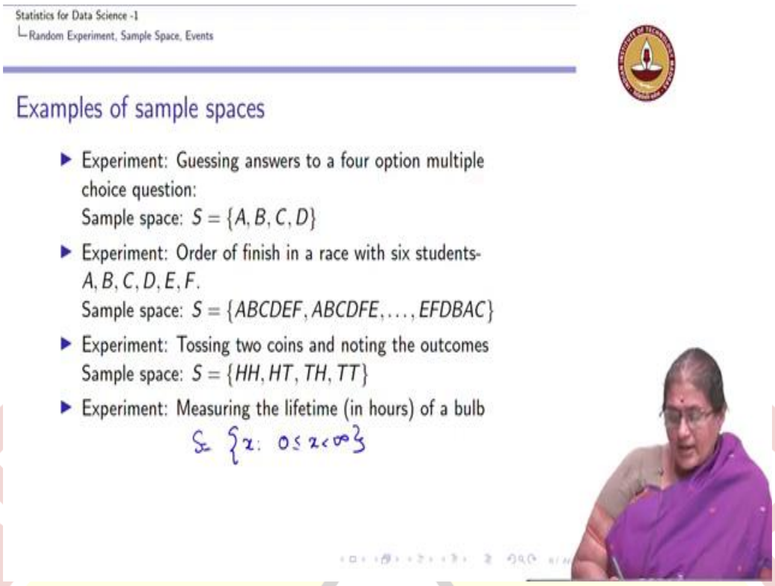
(Refer Slide Time: 25:12)

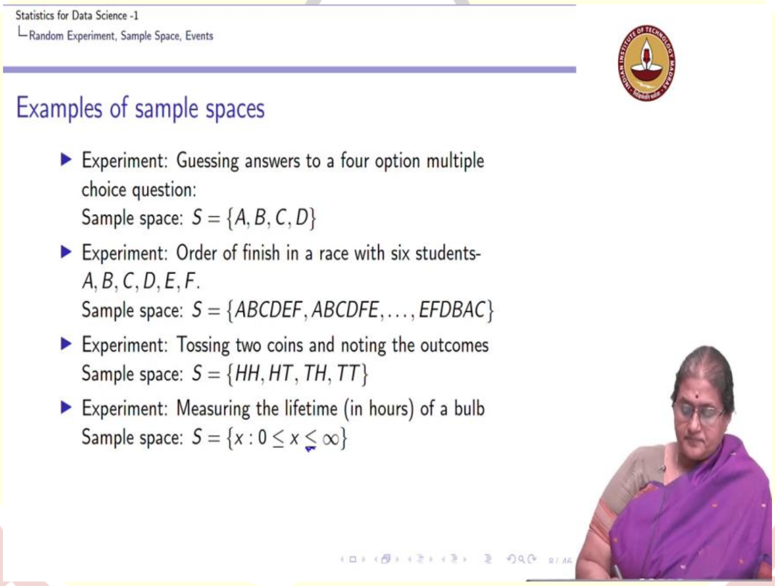




Now, let us go on to the next example. So, I again have 2 coins. So, my sample space is, I can either have a head-head; head in my first toss, head in the second toss or head in the first coin, head in the second coin, head in the first coin, tail in the second coin, tail in the first coin, head in the second coin and tail in the first and tail in the second. So, I can represent my sample space of this experiment in this way, HH, HT, TH and TT.

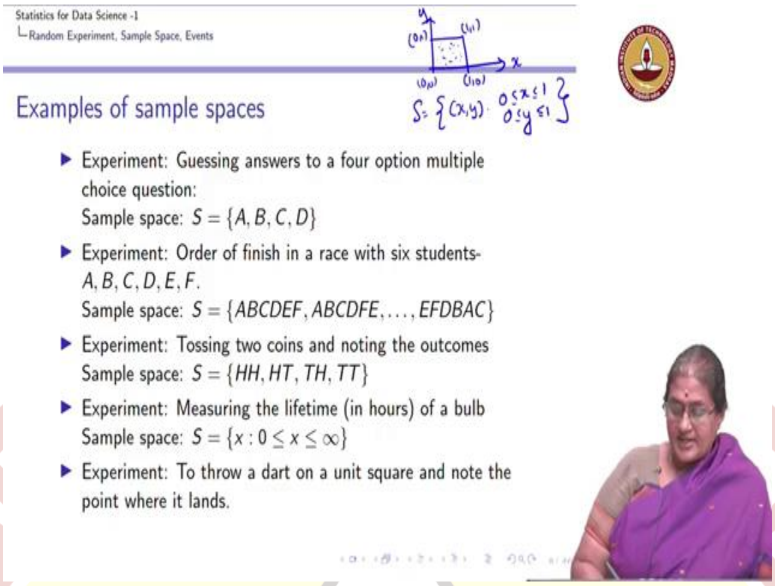
(Refer Slide Time: 25:44)

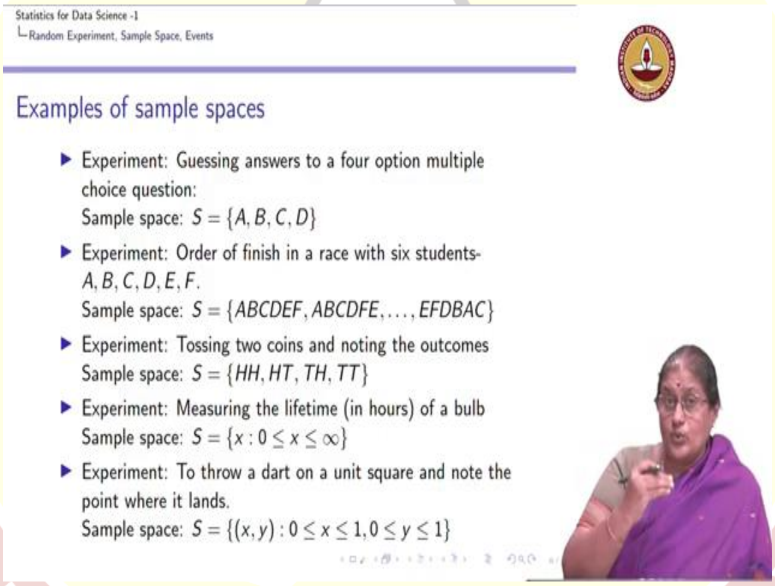




Now measuring the lifetime, I said it could be a 0 hour, it could be a 1 hour. So, I can express this as x where $0 \leq x < infinity$. And, this would actually I can, I know that this is either a 0 hour or a 1 hour or a 2 hour, it is less than infinity, sorry, okay.

(Refer Slide Time: 26:11)

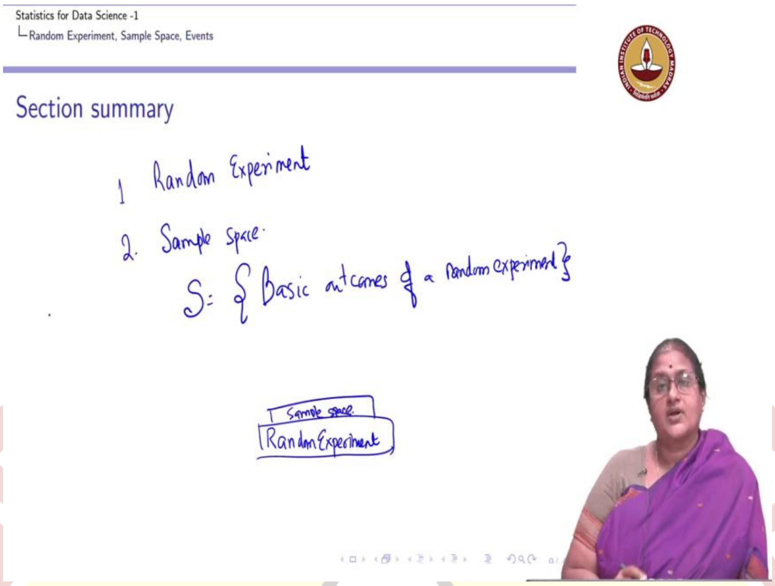




The next thing is when I throw a dart, I have already told you that this is going to be my X axis Y axis, this is going to be the square. If I represent my cord… points in the square using my coordinates of the X, Y could include my boundary points also. So, my S is going to be the set of all ordered coordinate points such that $0 \leq x \leq 1$, $0 \leq y \leq 1$.

That is what is my sample space where the sample space is nothing but all the points which are enclosed within the square. So, given, if we are, so if we encounter any situation with uncertainty, we need to understand what is the experiment, what are we observing, we need to know how to list down the sample space and the how to rep…we need to know how to represent the sample space.

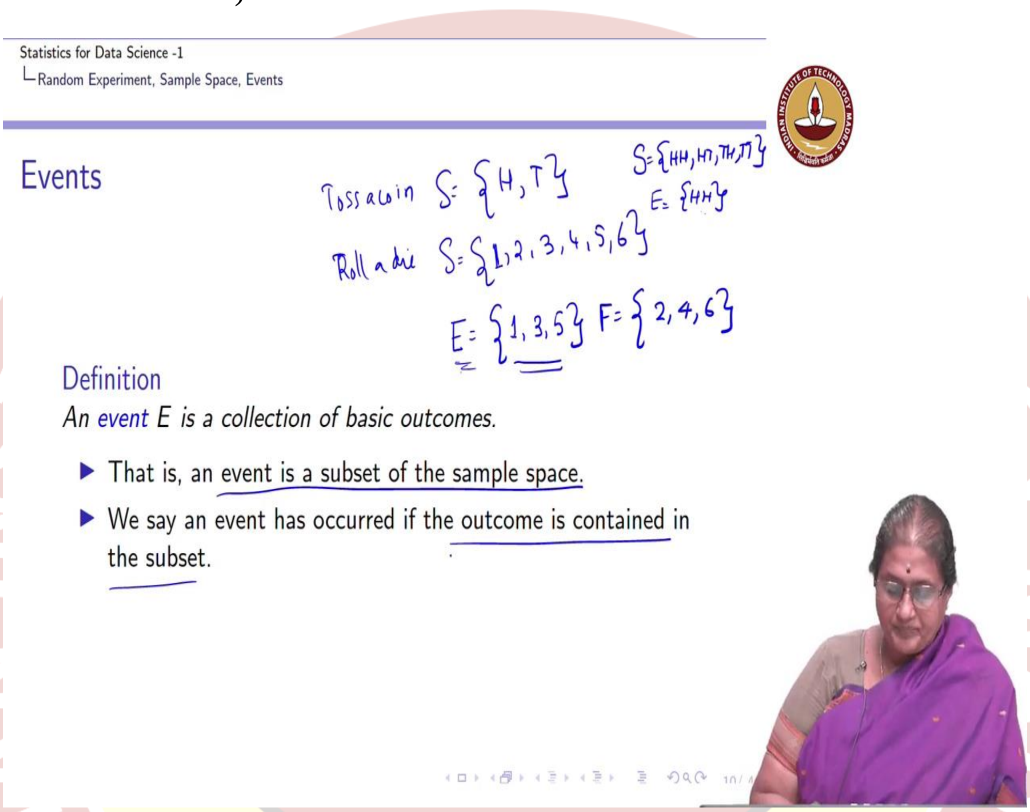
(Refer Slide Time: 27:23)



So, what have we learned so far? The main thing is we are learned what is a random experiment. The second thing we have learnt is what is a sample space and, we can list a sample space is nothing but the set of all basic outcomes of a random experiment, okay. So that is what we have learned. Now what is the next building block? So the first building block is the notion of a random experiment. And once we have the notion of a random experiment we understood what was the sample space.

Statistics for Data Science 1
Professor Usha Mohan
Department of Management Studies
Indian Institute of Technology, Madras
Lecture 6.2
Probability- Events and Basic Operations on Events

(Refer Slide Time: 00:14)



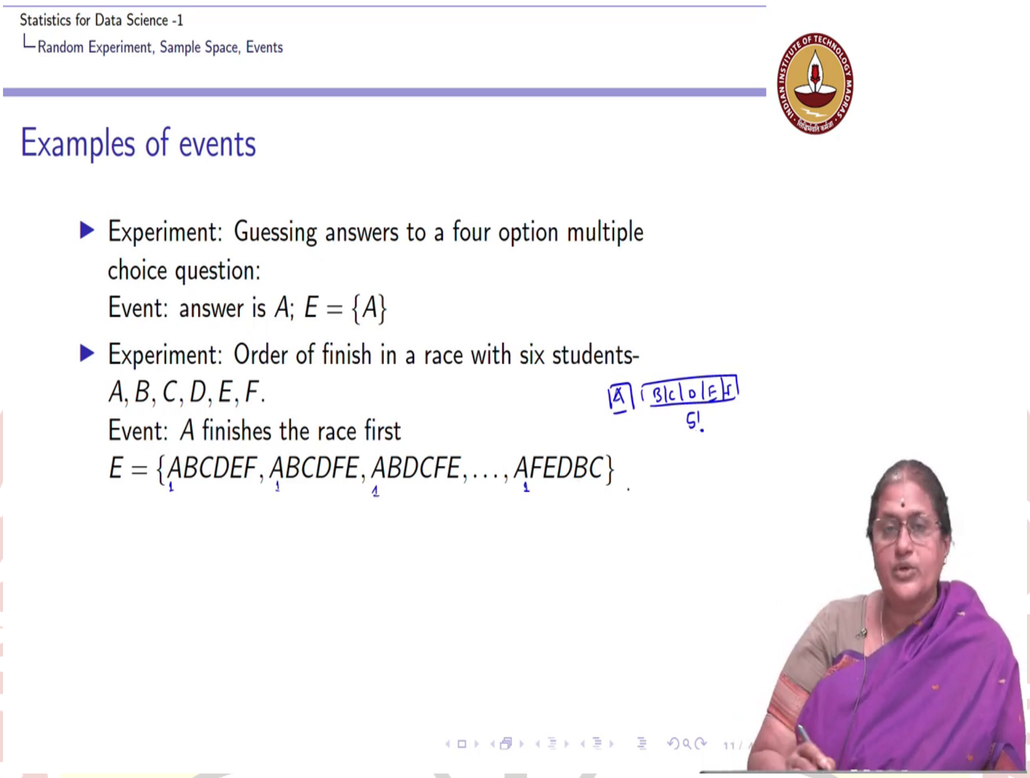
So, the next building block, we need to understand, is what is a event. So, event is a collection of basic outcomes. For example, when I toss a coin once, I know my basic outcome is a head or a tail, when I roll a die. So, this is a, when I toss a coin, when I roll a die; my collection is {1, 2, 3, 4, 5, 6}. A basic outcome is what I have, when I actually roll a die. That is what, is a basic outcome.

Now, I can define an event which is a collection of these basic events. So, for example, if I define an event {1, 3, 5}; I know the outcome of rolling a die is an odd number. So, it is again a collection of basic outcomes, but it is a subset of my total sample space. I could similarly define another event which is {2, 4, 6} which is again a subset of my total sample space. It is again a collection of the basic outcomes.

Now when I toss a coin twice, I know my sample space is $\{HH, HT, TH, TT\}$. Now, I can define an event which is $\{HH\}$, which I can see, it is again a. It is a set of basic outcome here it is a singleton. It is only 1 element, it is only 1 basic outcome, but nevertheless it is a event. So, we need to understand that, given a random experiment, we need to understand, how we define these events. We need to also understand that, these events are subsets of my sample space.

So, event is a subset of the sample space, and we articulate it as saying that, an event has occurred if the outcome is contained in the subset. So, I will say a event, that is a event of an odd number has occurred, if the outcome, what is a outcome, 1 is contained in my subset.

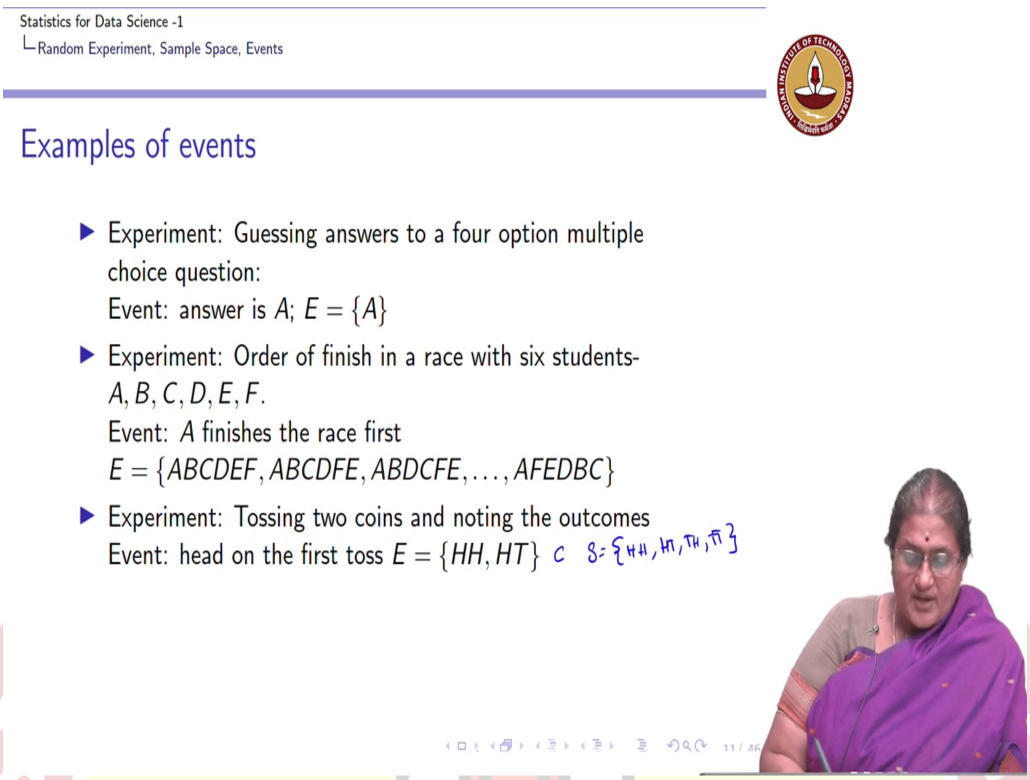
(Refer Slide Time: 02:45)

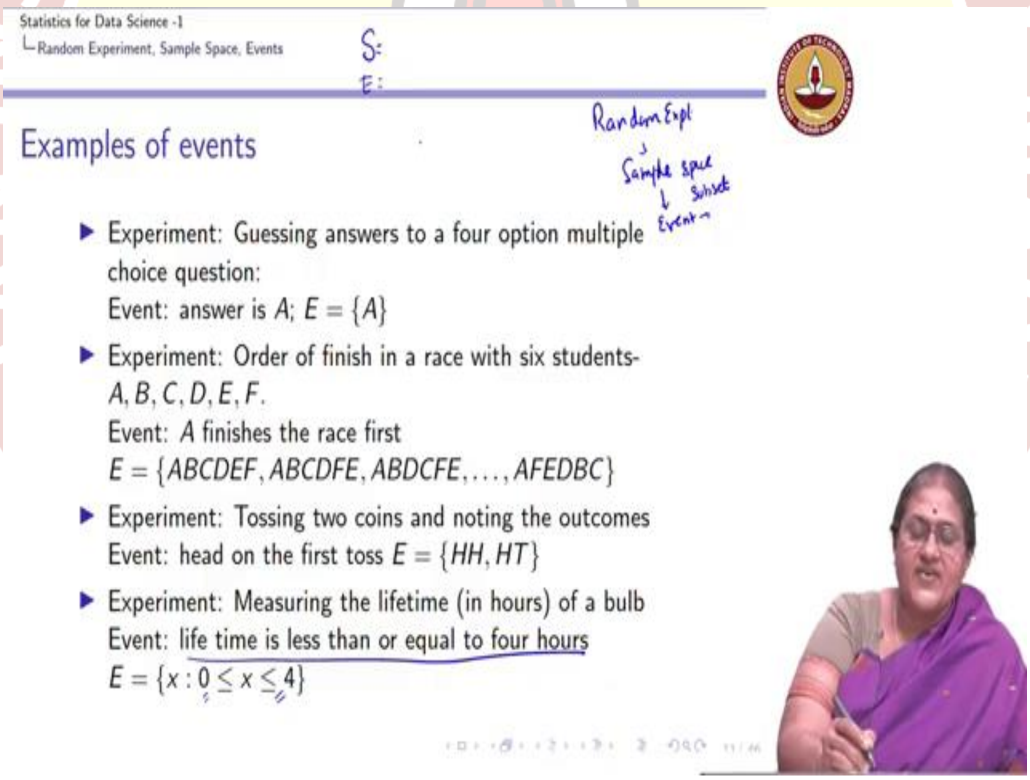


So now, let us go back to our examples. So, when I am guessing an answer to a 4 option multiple choice question. I can just define an event, that the answer is A. My event, I can represent it as a singleton set $\{A\}$. Again order of finish in a race with 6 students, suppose, I define my event to be A is the finisher, or the first person who has come first in the race.

So, you can see that, the event would be all possible permutations where A has appeared first. So, there are, we know that, if I fix A in the first position; the other B, C, D, E and F can be arranged in 5! ways. So, the event E, in itself will have about 5! outcomes. We know that, this is a subset of my original sample space.

(Refer Slide Time: 03:54)

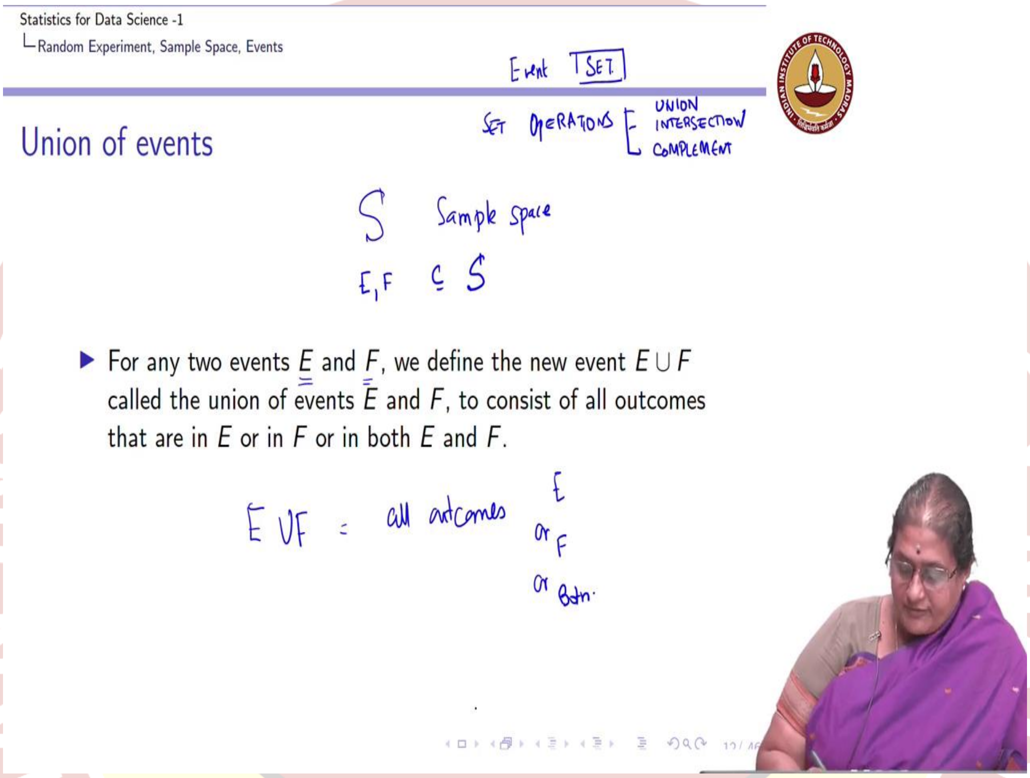


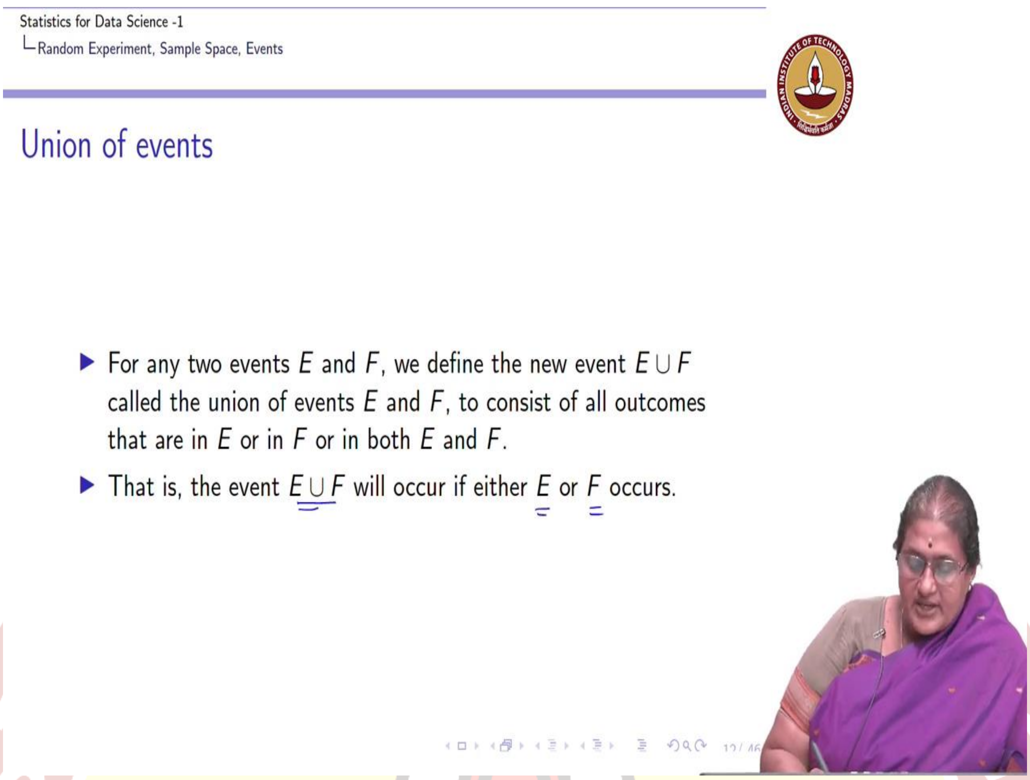


Now when I toss a coin twice, and note the outcomes; the event head on the first toss. So, I know that my sample space here is $\{HH, HT, TH, TT\}$. The event that, head appears in the first toss, again is a subset of my sample space. Measuring the life time, I can define an event, that my bulb has lasted for 4 hours, or has, is actually my bulb fails after 4 hours. In which case my event is, the life time is less than or equal to 4 hours, or my bulb has failed within 4 hours; my event can be represented by $\{x : 0 \leq x \leq 4\}$.

So, we can see that, coming from a random experiment and a sample space; sample space is the set of all possible basic outcomes. I define what is an event which is a subset of my sample space but this is also a set of basic outcomes. So, when we have a sample space and E is the set of basic outcomes we know that we can define all possible subsets of this sample space as events. Now, do all the possible subsets makes sense that is something which we need to see. But we know that, theoretically we can define all subsets as events. In a theoretical way, whether they make sense or not, we need to see, depending on the context of the experiment.

(Refer Slide Time: 05:47)



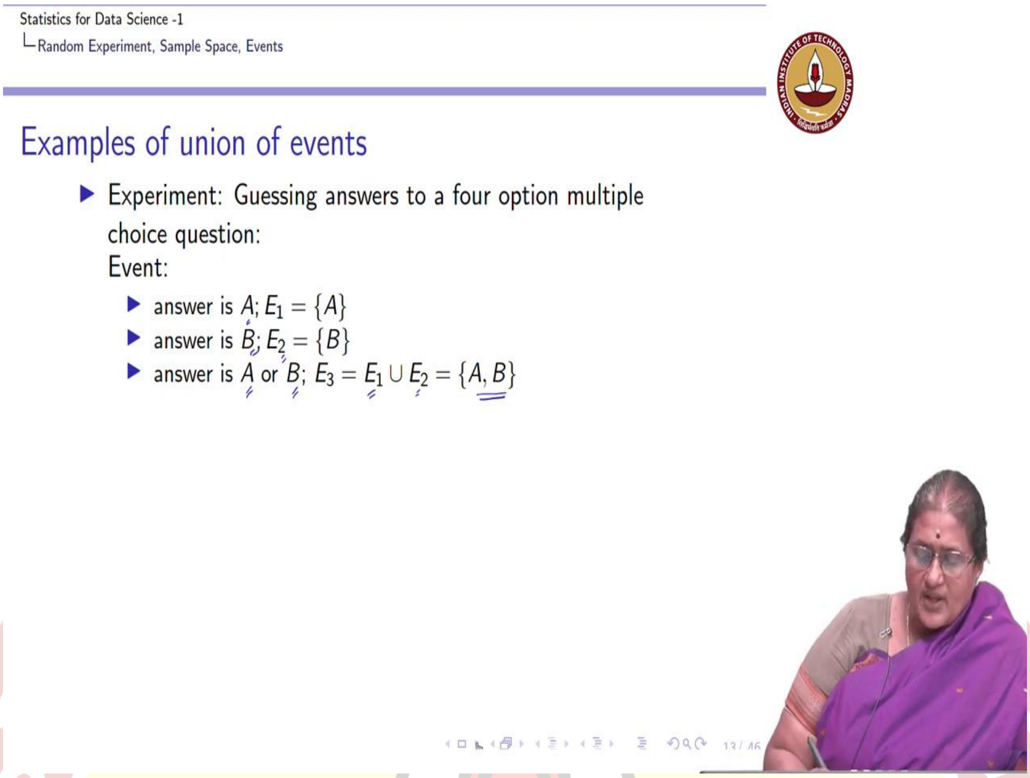


Now, because events are again subsets. So, I have an event. Event is also a set. Now I know, the minute I say that, E is an event; then I can talk about all possible set operations. So, the 3 key set operations we are going to see, are. So, I have what I can refer to as set operations on my events because I realize an event is a set. What are the possible set operations? The 3 set operations, basic set operations, we are going to talk over.

What do we mean by union of two events? What do we mean by intersection of two events? And what do we mean by complement of a event? So, these are the basic set operations, which we are interested in knowing. So, suppose I am given two events; E and F. I know that S is my sample space. Given S is my sample space I know both E and F are subsets of my sample space because they are events.

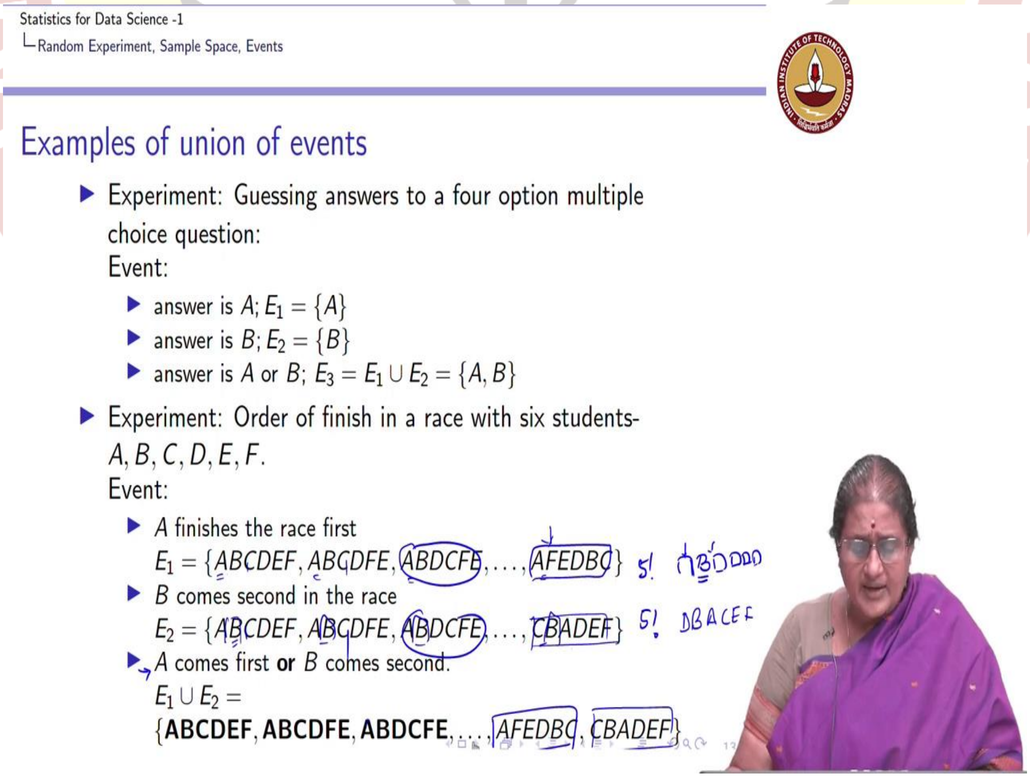
Now I can be define a new event $(E \cup F)$ to consist of all outcomes that are in E or F or both; that is what I refer or I mean by $(E \cup F)$. So, the event $(E \cup F)$ will occur, if either E or F occurs.

(Refer Slide Time: 07:35)



So now, let us look at a few examples. Suppose my event E is E_1, suppose my event E_1 is answer A, event E is answer is B, answer is either A or B can be defined by the event $(E_1 \cup E_2)$ which can be represented by $\{A, B\}$.

(Refer Slide Time: 08:01)



Let us go to the next example. Order of finish in a race, suppose my E_1 event is A finishes the race first. So, in this I have all the outcomes where A is finishing first. Event B is, B is coming second in the race. So, you can see that B is occupying the second position here. Again, I fix B in the second position. Total number of outcomes here is again 5! because I fixed B in the second position.

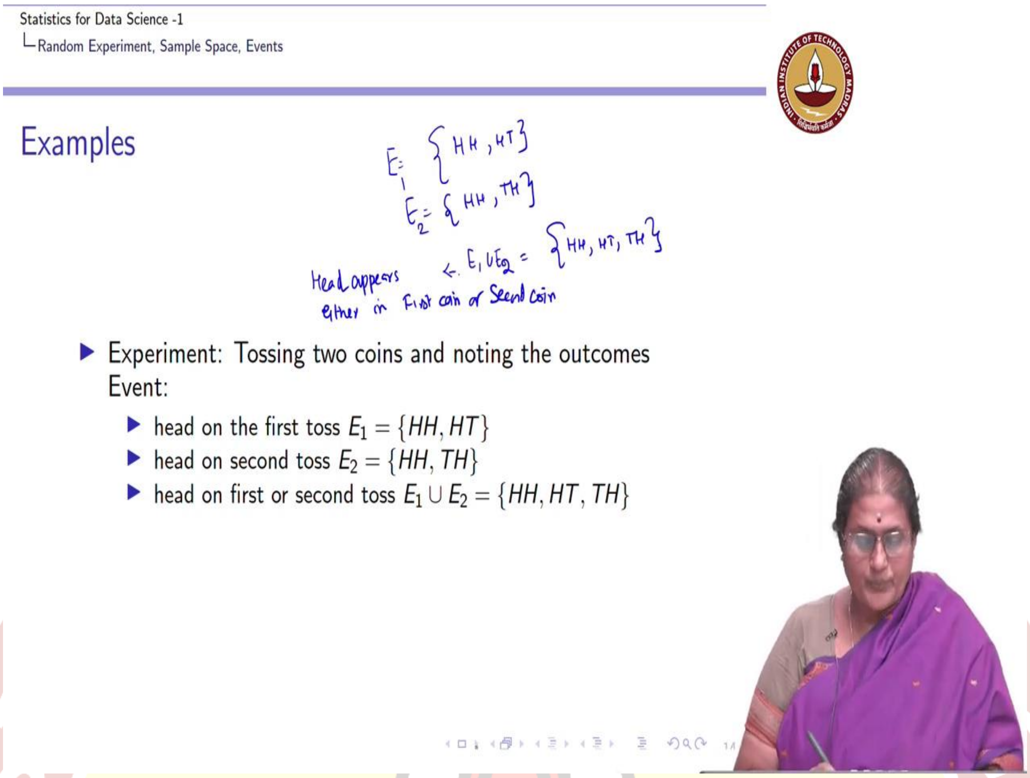
The other things are, I have 5 places which have, can be occupied by the 5 available alphabets, and that can be done in 5! way. So, the event A union, or $(E_1 \cup E_2)$ can be described as either A comes first or B comes second and that is. So, you can see that, this includes ABCDEF which is common to these two; ABCDFE is common to these two, this event ABDCFE is also common to both of them. But this event is not in B, but I am including it here because it is occurring in A.

So, this event AFEDBC does not have B in the second position, but I am including it in $(E_1 \cup E_2)$ because it is occurring in A. And this outcome CBADEF is not in A because A is not the first finisher here. In this event CBAD, in this outcome CBADEF, A is not the finisher. So, this outcome is not a part of A, but we are including it. It is not, this outcome is not a part of your E_1 but we are including it in $(E_1 \cup E_2)$ because it is an outcome in E_2.

So, the set of outcomes where either A comes first, or B comes second constitutes this $(E_1 \cup E_2)$ event. And I can describe it as either A comes first or B comes second. So, it may, it is a meaningful event to have. So, would this include the event where B comes first? The answer is would this include the event where B comes first? B comes second, so, B comes first is not included here. But would it include C coming first? Answer is yes because we have seen here.

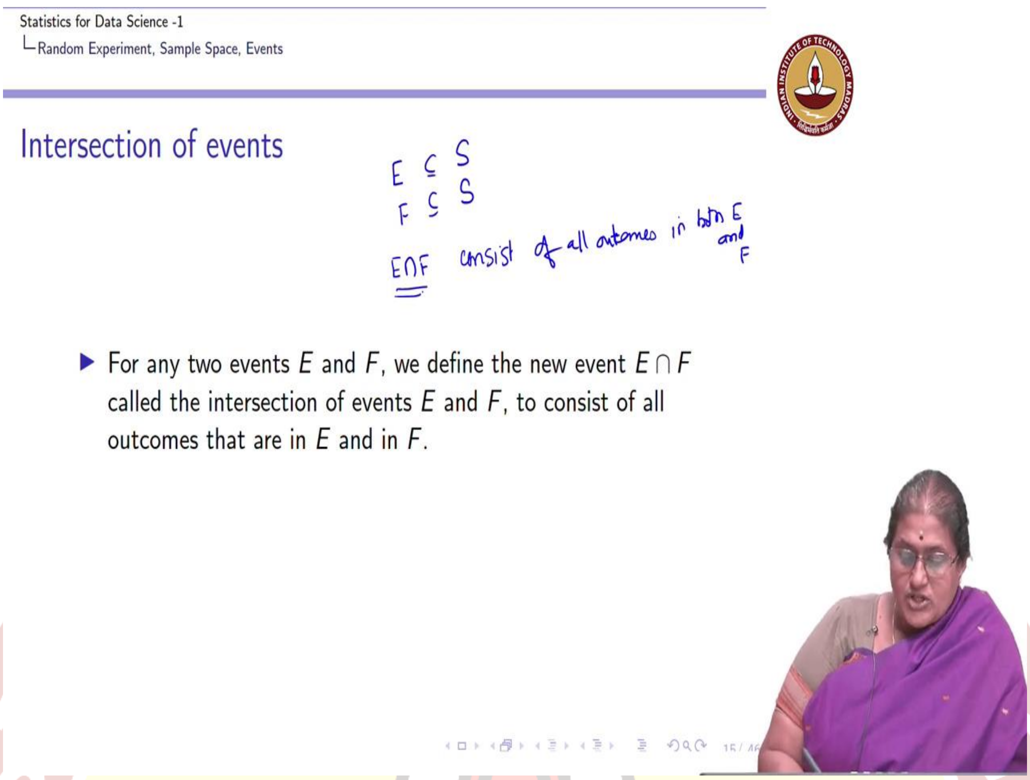
Would it include D coming first? Yes, if the outcome is a DBACEF, yes because that is DBACEF is a outcome in my E_2 set. So, you can see that, it includes all the possible outcomes where either A is first or B is second.

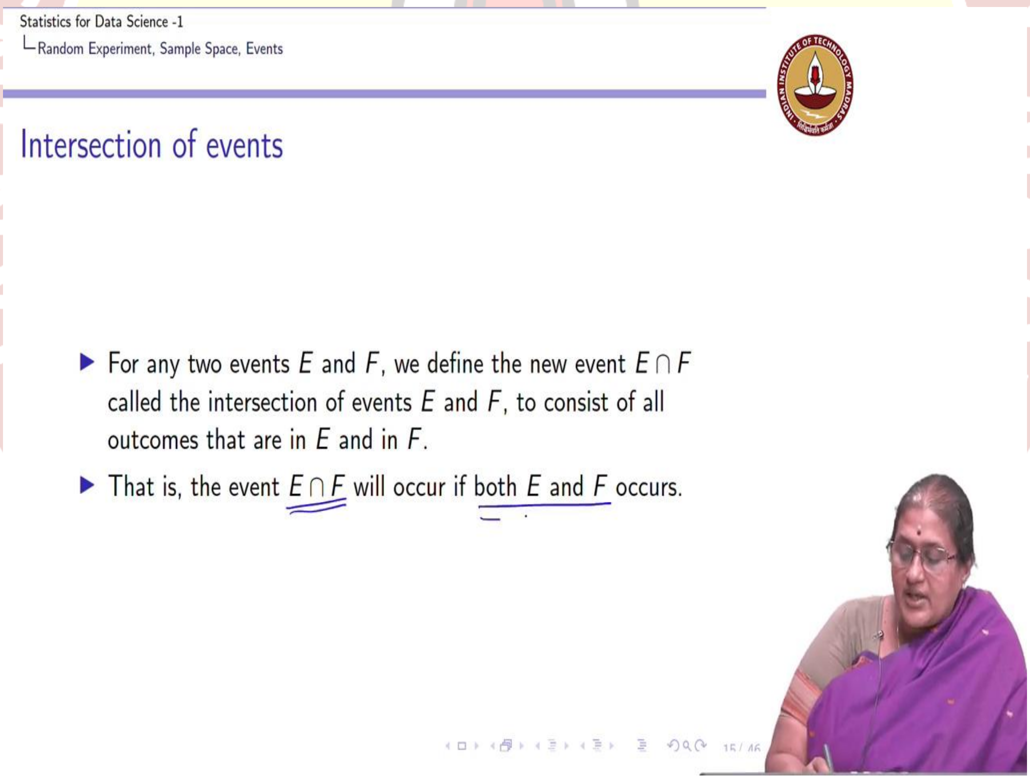
(Refer Slide Time: 11:13)



Now let us look at an example of tossing 2 coins, and noting the outcome. Let me define my first event as head on the first toss. I know if it is a head on the first toss, E_1 is {HH, HT}. Head on the second toss is {HH, TH}. This is the head on the second toss. So, if I am defining an event where I say that head occurs either in the first toss or in the second toss. So, my outcomes $(E_1 \cup E_2)$ is {HH, HT, TH}. So, this tells that head appears either in my first toss, or head appears either in first coin or second coin. That is how we can express this event $(E_1 \cup E_2)$.

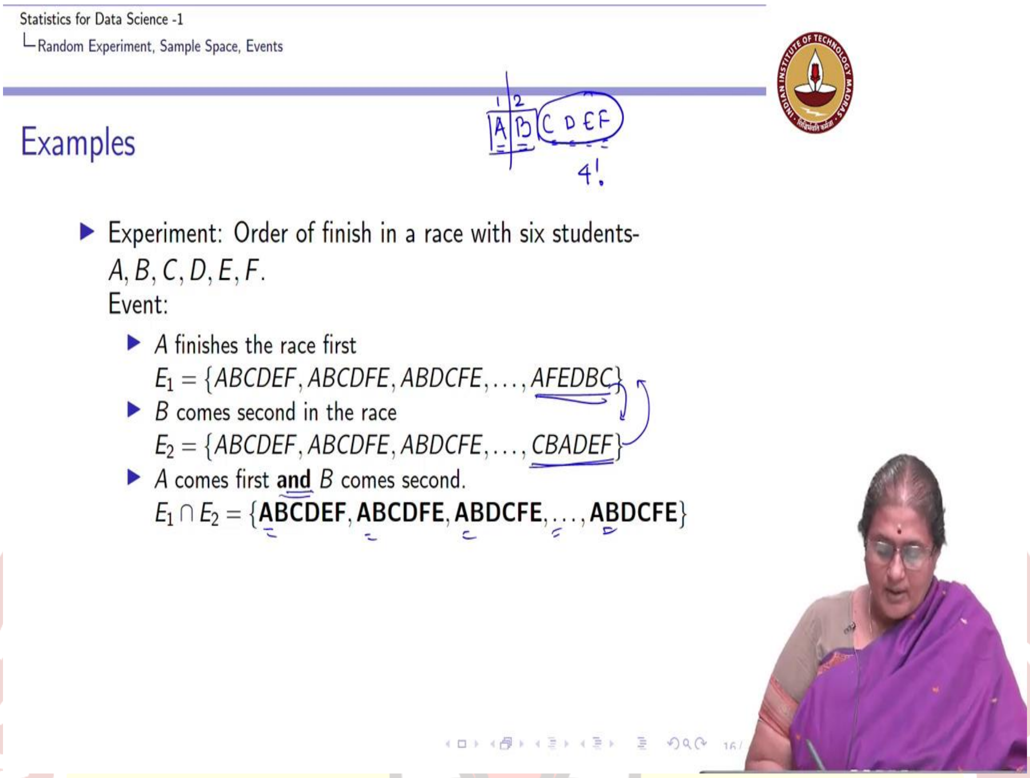
(Refer Slide Time: 12:19)





Now, the next set operation is what we refer to as a intersection of events. So, now given 2 events E and F. Again, I know these are subsets of my sample space. I can define a new event, and I represent it by $(E \cap F)$. Now, this event $(E \cap F)$ consists of all outcomes, that are in outcomes that are in both E and F. Earlier it was either E or F. Here it is both E and F. So, that is, so E, the event $(E \cap F)$will occur if both E and F occur. So that is how we define an event $(E \cap F)$.

(Refer Slide Time: 13:20)

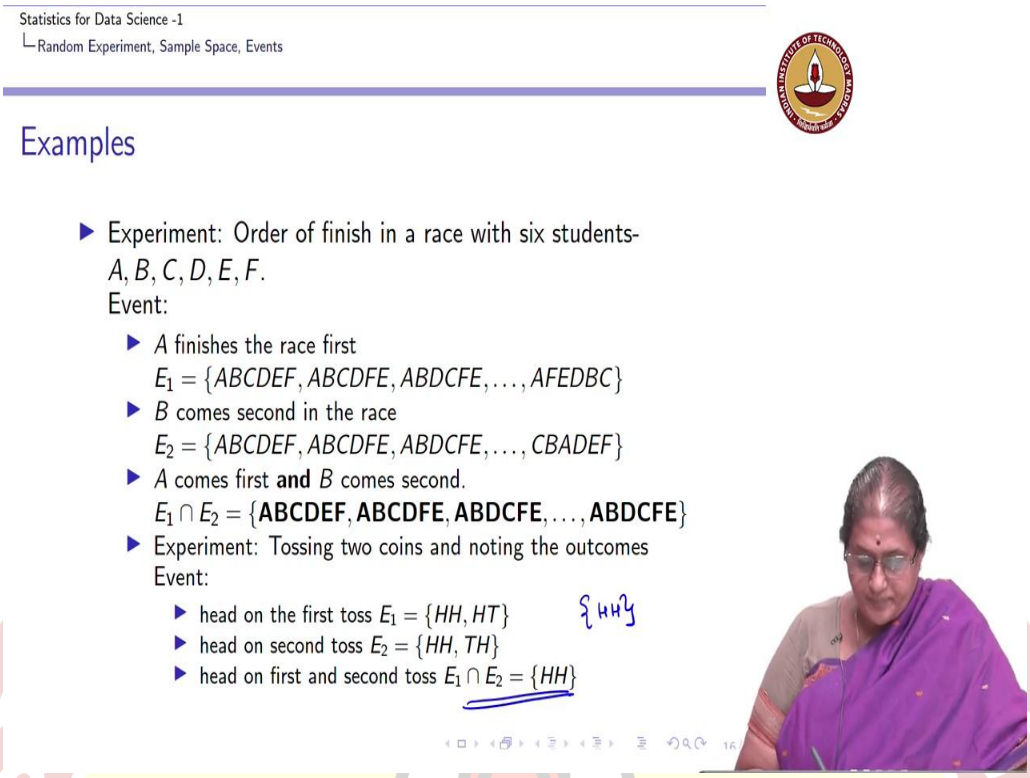


Let us look at examples. Again, when I have an order of finish in a race with 6 students. Let me go back, and define the events in a similar way. I have A finishes the race, which is ABCDEF. B comes second in the race is ABDCEF. Now if I am looking at events where A comes first, and B comes second.

Then I am looking at all possible arrangements where A is in my first place, B is in my second place, and the rest of the elements are arrangements of C, D, E, and F which can happen in 4! ways, because I am fixing A and B to be in my first and second space. And I am looking at all other possible arrangements of C, D, E, F among themselves. So, this is an event where A comes first, and B comes second.

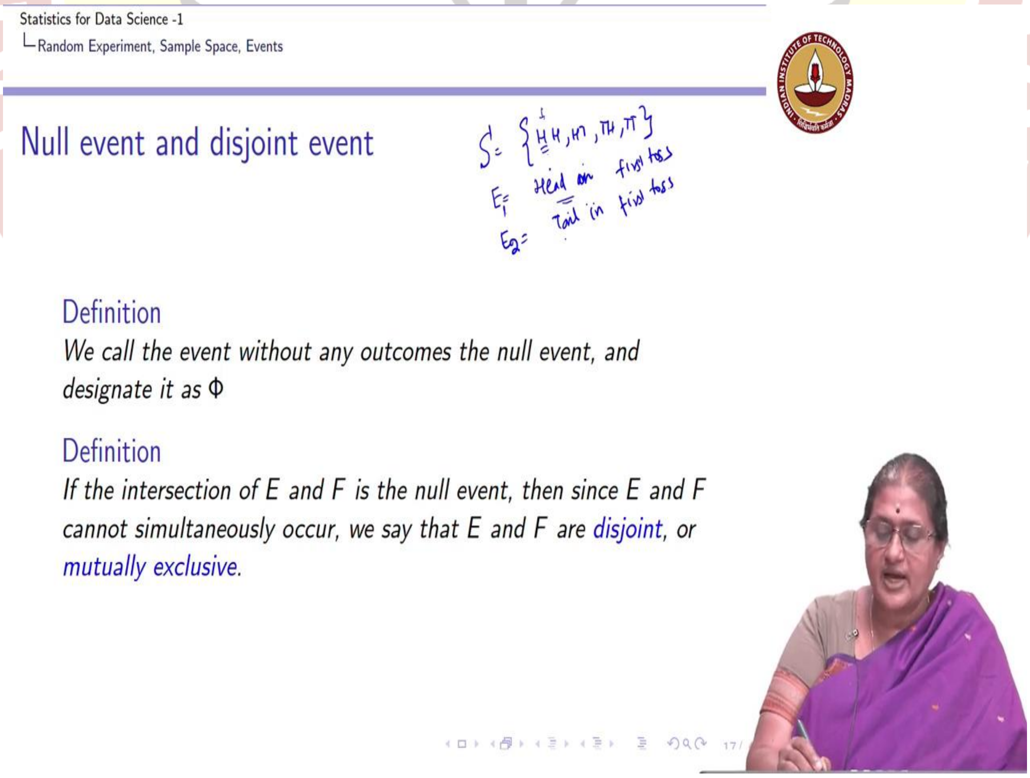
Notice that this element CBADEF will not be an outcome in the set. Similarly, AFEDBC which was an event will not be an element of the $(E_1 \cap E_2)$, because this is not an element of E_2, and this is not an element of E_1.

(Refer Slide Time: 14:47)



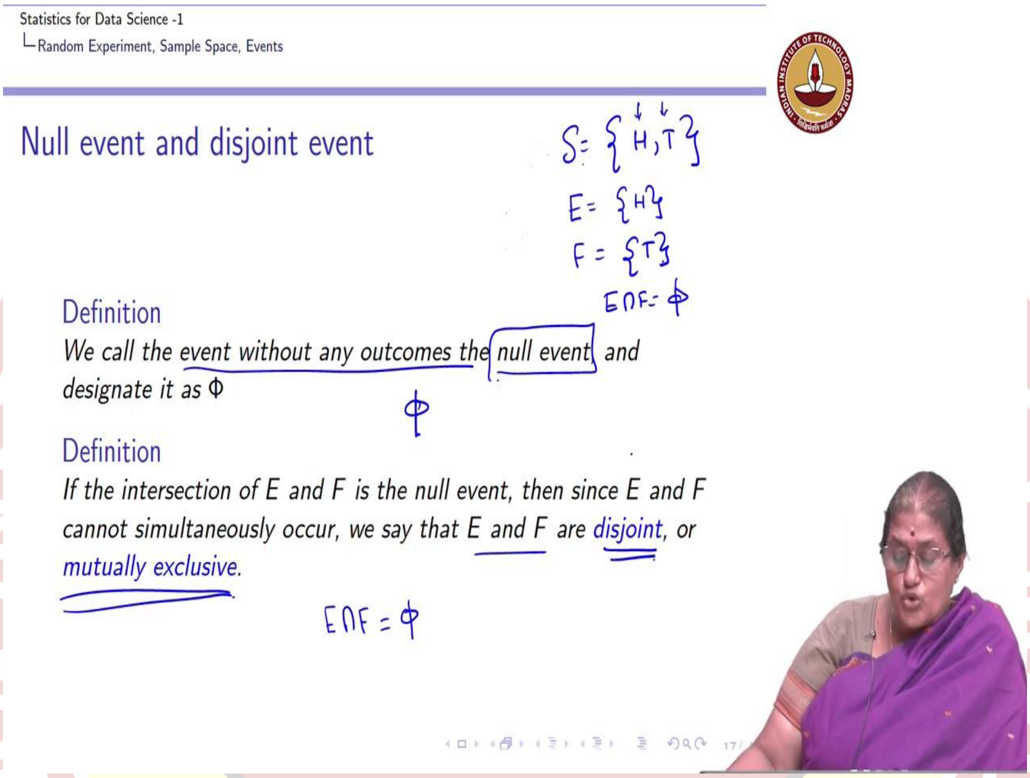
Now, let us toss two coins and note the outcomes. Again, let head on the first toss, head on the second toss, head on both the tosses would just be the event heads {HH}, which you can see is $(E_1 \cap E_2)$, would represent the event that head appears in both the tosses.

(Refer Slide Time: 15:12)



So, now suppose I go back here, and I define here that let S be my sample space again. And let me define an event E which is head in my first toss and E_2 to be head in second toss, or I can define my E_2 to be tail in my first toss. Now, I know that, this event $E1$ an E_2 cannot happen at the same time because if I have a head in my first toss, I cannot have a tail. If I have a head in my first toss, I know that I cannot have a tail also in my first toss. So, there are events where which cannot occur together.

(Refer Slide Time: 16:30)

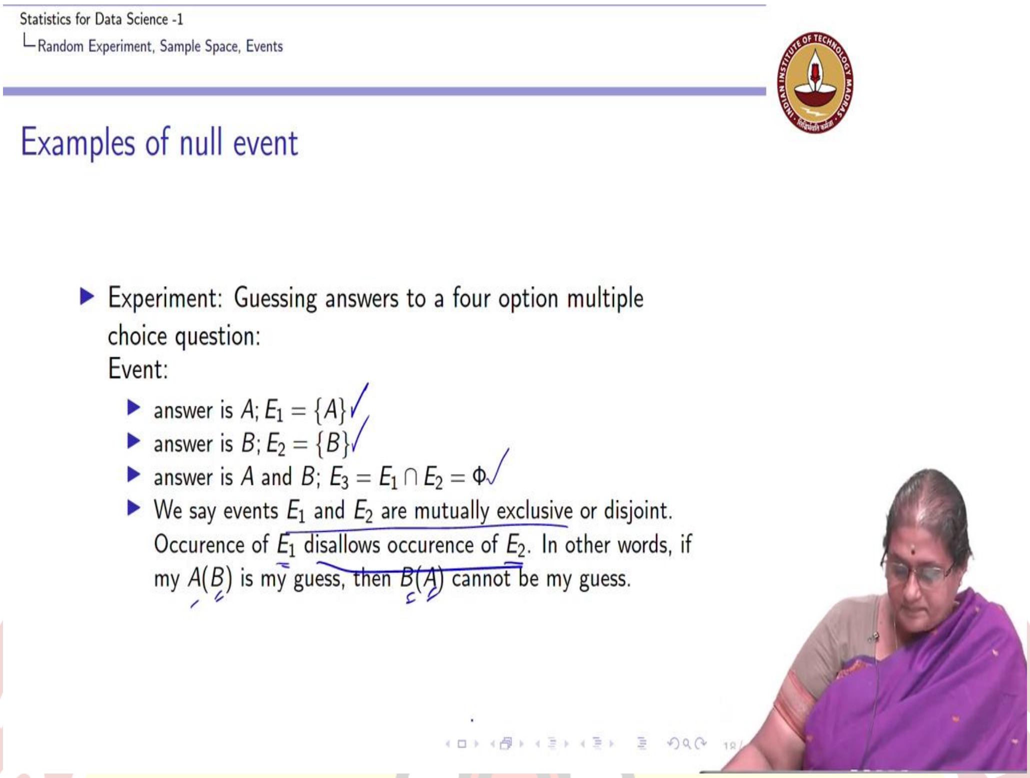


Similarly, when I have an answer to a multiple-choice question, and my sample space is one of these answers, or one of these guesses. I cannot have an event which would say A and B are correct answers, where only exactly one of the answers are correct to my multiple-choice questions with 4 options. So, an event without any outcome is called a null event, and you can determine or designate it by this alphabet or by this symbol $\emptyset$.

So, if the intersection of $(E \cap F)$is a null event, then we say E and F are disjoint events, or mutually exclusive events. For example, I have just toss a coin once, and I have head or tail. I can define my event E to be the outcome of a head, I can define my event F to be outcome of a tail, I can see that $(E \cap F)$is my disjoint set.

In other words, I can say the outcome head and tail are mutually exclusive that if I get a head, I cannot get a tail, if I get a tail, I do not get a head. So, these are mutually exclusive events.

(Refer Slide Time: 18:02)



So, suppose I am guessing an answer to a four option multiple choice question. If E_1 is $\{A\}$, E_2 is $\{B\}$, E_3 which is $(E_1 \cap E_2)$, is a null set, $(E_1 \cap E_2)$, are mutually exclusive. If A that is occurrence of an event disallows the occurrence of E_2. So, if my guess is A or B. If my guess is A then B cannot be my guess. If my guess is B then A cannot be my guess. So, these are mutually exclusive events. Again, a concept that is very-very important for us to understand about probability of events.

(Refer Slide Time: 18:52)

Complement of an event

Toss a coin $S = \{H, T\}$

$E = \{H\}$ $S \setminus E = \{T\}$

Definition

The complement of E, denoted by E^c, consists of all outcomes in the sample space S that are not in E.

Statistics for Data Science -1
└ Random Experiment, Sample Space, Events

Complement of an event

Toss a coin $S = \{H, T\}$

$E = \{H\}$ $E^c = \{T\}$

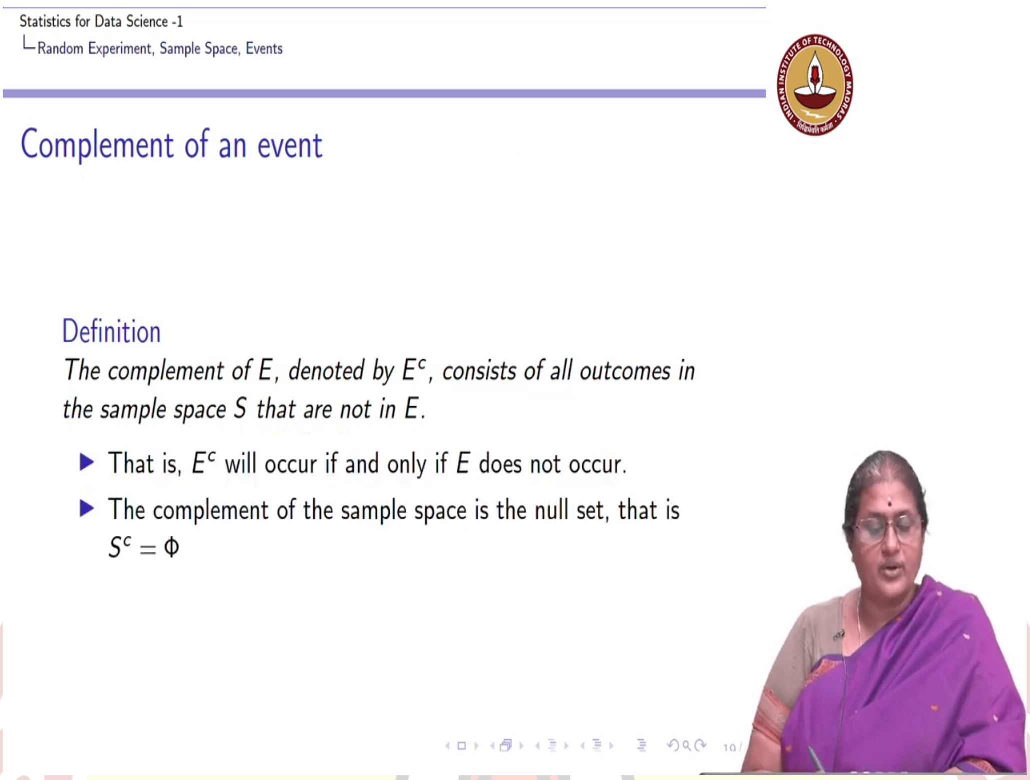
Definition

The complement of E, denoted by E^c, consists of all outcomes in the sample space S that are not in E.

Toss a coin twice $S = \{HH, HT, TH, TT\}$

$E = \{HH, HT, TH\}$ → At least coin is a head

$E^c = \{TT\}$ → Both the coin are Tail

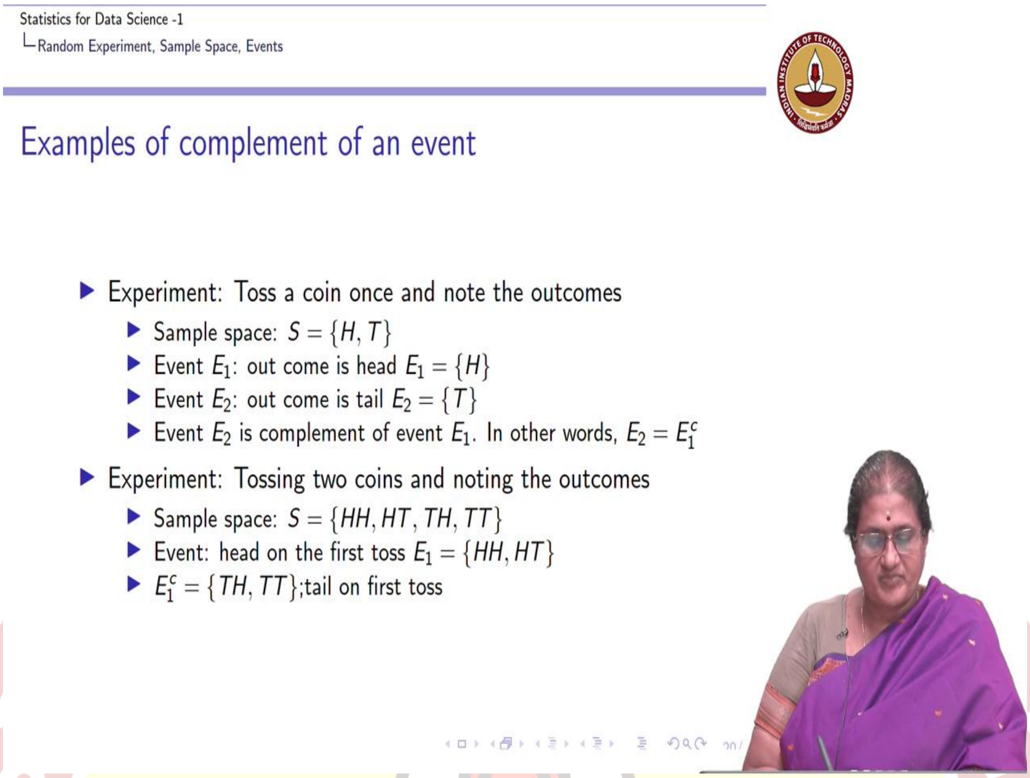


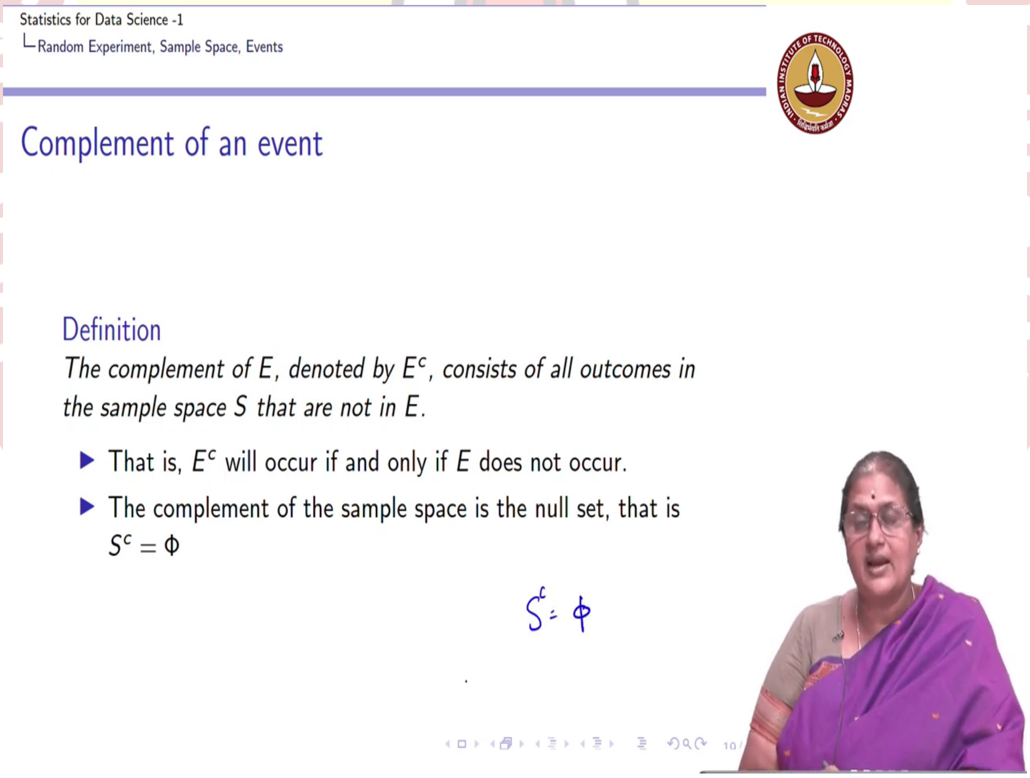
The next set operation which we are going to talk about is, what we refer to as a compliment of a set. So, I have my sample space. Again, I have $\{H, T\}$. I toss a coin once. I know this is my sample space. If I define E to be the outcome of a head, I know the complement of E is $(S \setminus E)$ set, which is $\{T\}$. I can refer to this as E^c, complement of a set.

Similarly, what is the E^c? It consists of all those outcomes that are not in E. So again, I toss a coin twice, and it define, I know S is $\{HH, HT,\ TH,\ TT\}$. If I define an event that at least one of the coin is a head, then I know these events are $\{HH, HT, TH\}$ are the outcomes in this event E. So, the E^c is $\{TT\}$, that is this element which is not in E, and I can see that this corresponds to both the coins are tail.

Now, this is at least one coin is a head. So, the complement of E is both the coins are tail. So, you can see that this is a very important operation again, and how we have represented this operation also. So now let us go back, and look at our example.

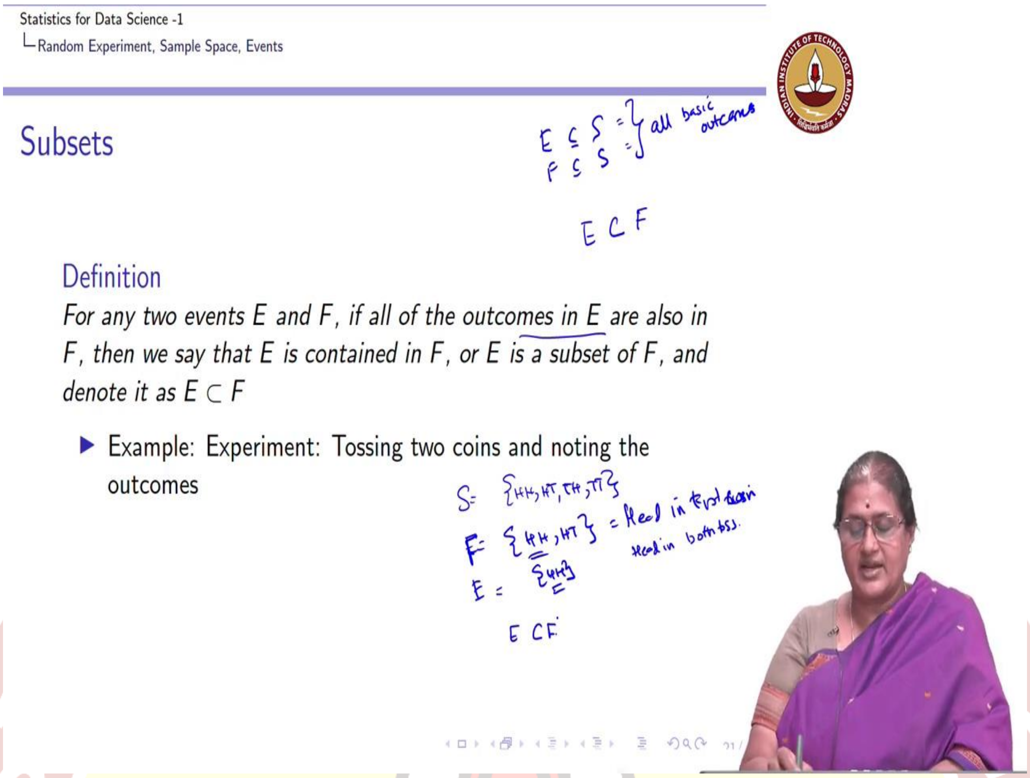
(Refer Slide Time: 20:46)

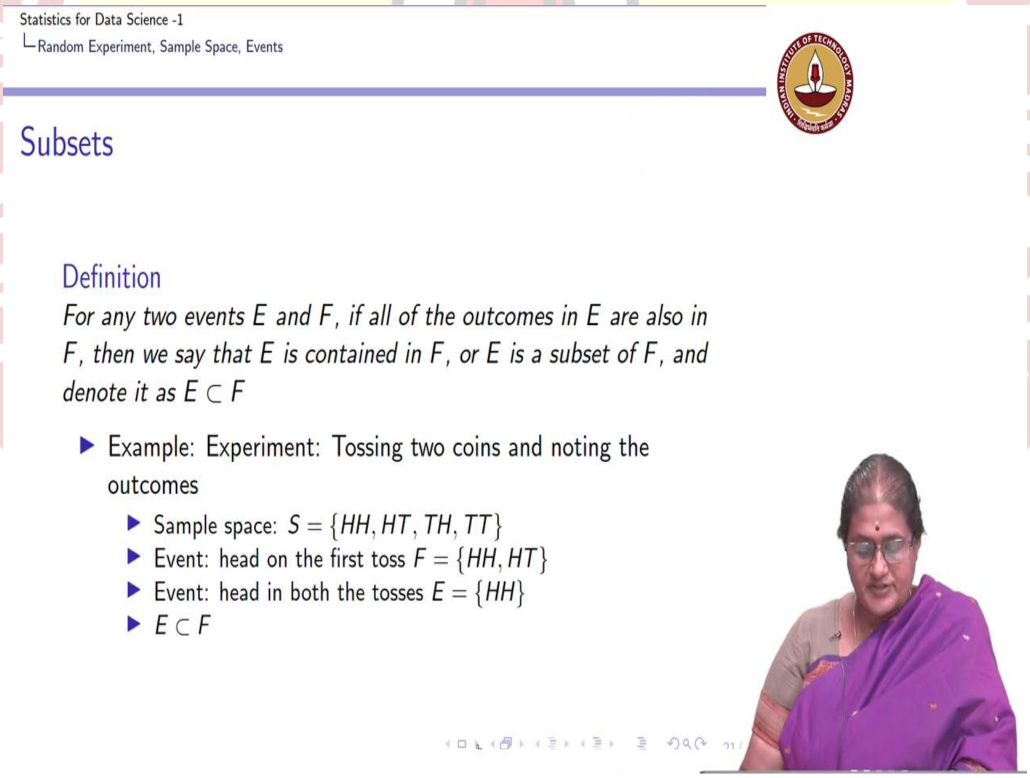




So, I toss a coin once, and note the outcomes. I know that $E_2^c$is E_1^c. Toss a coin twice, event is head on the first toss, event 2 is tail on the first toss and E_1^c is $\{TH,\ TT\}$. One thing we need to observe is the complement of the sample space is the null set. This is again, or the null event.

(Refer Slide Time: 21:22)





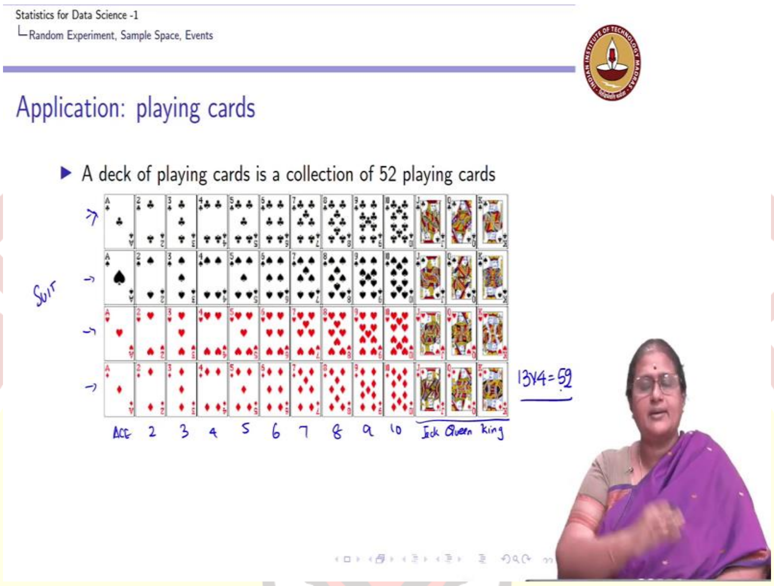
So, now suppose we are given 2 events. Again, remember events are subsets of my sample space, where sample space in itself is a set of all basic outcomes; all possible basic outcomes of my random experiment. So, if all the outcomes in E are also in my event F, then I say E is contained in F, and I denote it by $E \subset F$.

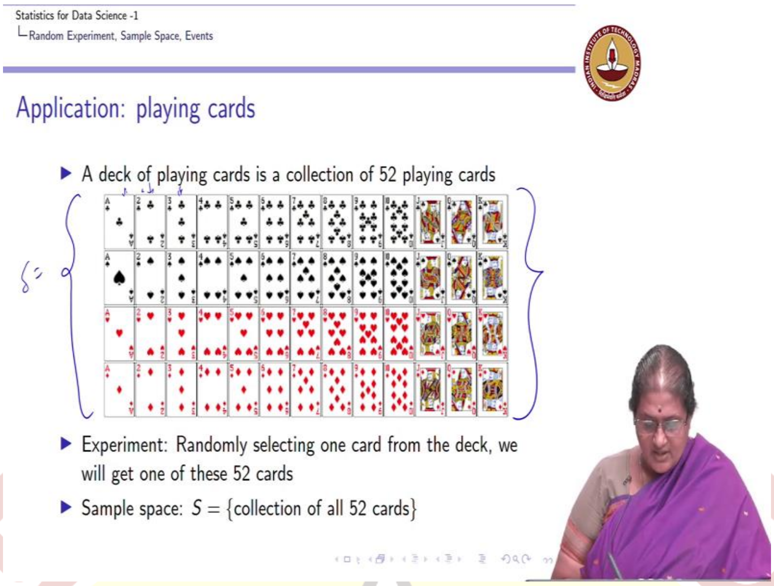
Suppose, I again go back, I toss a coin twice. I know $\{HH, HT, TH, TT\}$ are my outcomes of the sample space. Let me define an event which is head in my first toss. So, the event of head in my first toss is $\{HH,\ HT\}$. This is my event of my head in my first toss. Let me define or event F is my head in the first toss. Let me define E to be the event, head in the both coins, or head in both tosses. I am tossing 2 coins, so head in both toss, head in first toss of, first coin, head in both toss of the coin. Then I know E is $\{HH\}$. So, the outcomes in E, what is the outcome, I have only one outcome $\{HH\}$. It is also in F.

Then I know that, I can say, $E \subset F$. In terms of event, I know that head occurring in both the coins is a subset event of the event, where I am saying head is occurring in my first coin. So, head occurring in the first coin can happen in the outcome $\{HH, HT\}$, and head in both toss is a subset of this event. Now, these notions would become very important when we want to actually derive probabilities of the events.

Statistics for Data Science 1
Professor Usha Mohan
Department of Management Studies
Indian Institute of Technology, Madras
Lecture 6.3
Probability- Random experiment, Sample space, events

(Refer Slide Time: 00:13)





So, let us learn the concepts you have learned so far to one example or application of decks of cards or playing of cards. Recall in the permutation and combination module, we introduced you all to what we mean by a deck or a playing card deck.

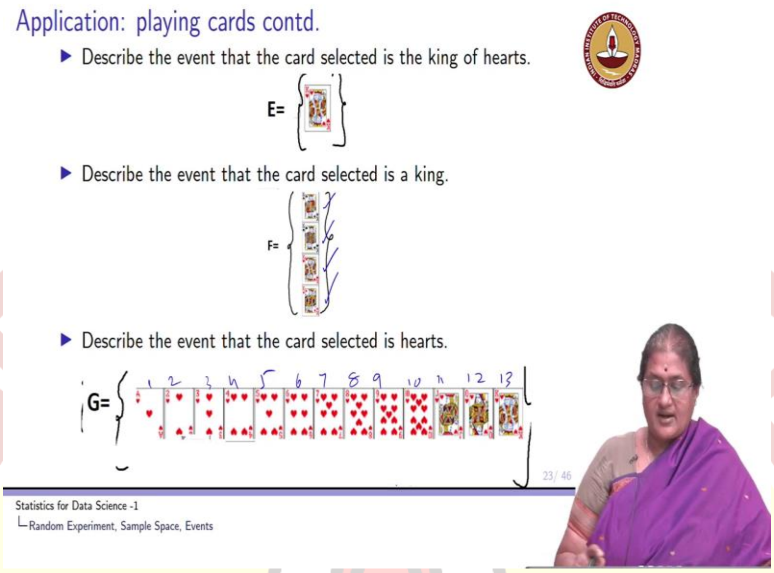
So, a deck of playing cards is actually this collection of these 52 playing cards. Now these clubs, the clubs, the spade, the heart and diamond is what we referred to as a suit of the card, it goes from Ace, this is an ace, this is a 2 3 4 5 6 7 8 9 10. Now these are referred to as face cards, you have this is a jack, this is a queen and this is a king.

So, they are totally 13 types. So, I have four shoots and I have 13 types. So, making it 13×4, which is 52 playing cards. Now, if I am laying down a deck of playing cards and I asked a person to choose one of the cards from this deck, I am laying, all these playing cards are laid face down. So, this is a face of a card. So, each one of them is laid facedown and I ask a person to choose one of the card.

Now, the random experiment in this case is to choose or randomly you are selecting one card from this deck of 52 cards. Why is it a random experiment? Because I am choosing one card. I know this card can be any one of these 52 cards, I know these are the possible outcomes of the experiment. But I really do not know what is the outcome, hence randomly selecting one card from these 52 cards will form my random experiment.

So, what are the outcomes? The outcomes are any one of these 52 cards, so the collection of all these 52 cards. So this, this was one of the outcome, these are my basic outcomes. So, a clubs is one of my basic outcome, this is the second basic outcome. So, I have these 52 outcomes. So, I can say my sample set is these 52 outcomes of the random experiment. So, that is my sample space.

(Refer Slide Time: 02:57)



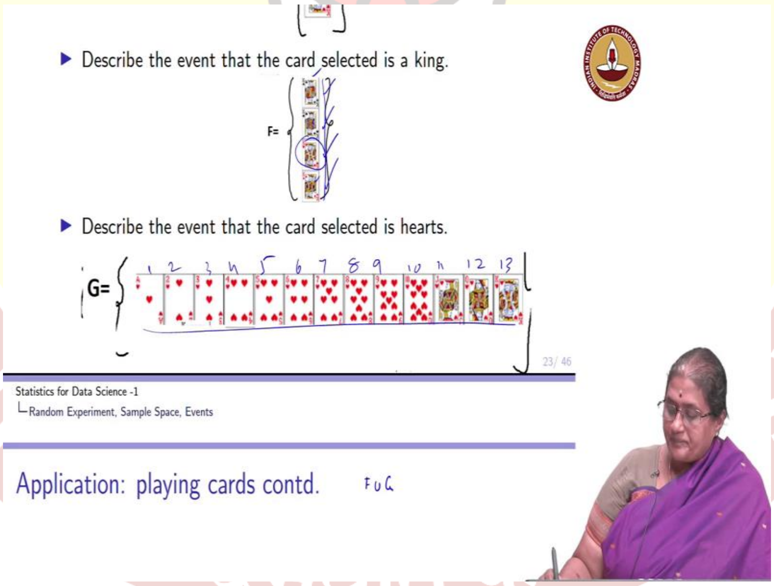
Now, let us go and look at describing different events. So, suppose my event is that the card selected is a king of hearts. So, if that is my event, I can go back and I can look at what is that event. So, if I go back here, I have 52 cards, describing an event which says that the card selected is a king of hearts, I can see that there is only one king of hearts in this entire 52 cards. So this event E which I am going to represent as the King of cards, this event has only one basic outcome and that is nothing but the king of hearts.

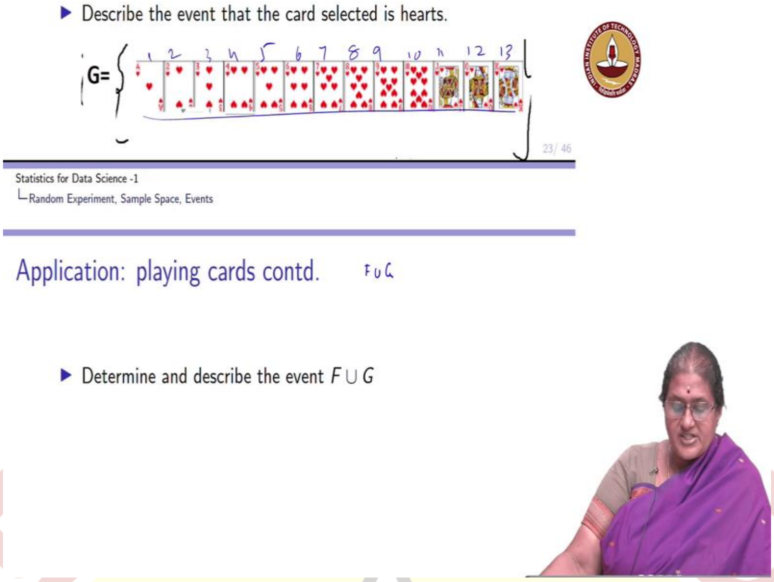
The card selected is a king again, let us go back and see that when we want to see what is the card selected is a king. Again, I go back to my earlier slide and I can see that I have king of clubs, it could be a king of spades, it could be a king of hearts or it could be a king of diamonds. So, the set of possible outcomes which are actually favouring the statement that the card selected is a king are any one of these four outcomes.

So, my event that the card selected is a king will have these four events which I have listed as king of clubs, king of spade, king of hearts and king of diamonds. So now let us look at the next event. Now the next event I am going to define is the event of the selected card selected as a hearts.

Again, let us go back to our deck of cards. You can see that the outcome which satisfies this even that the card selected is a heart can be any if I choose any one of these cards this is just ace hearts or 2 hearts, or a 3 or a 4 or 5, 6, 7, 8, 9, 10, jack hearts, queen hearts and a king hearts I have 13 outcomes, which satisfy this outcome that the card selected is a heart and I can define that or describe that event as G, which has ace, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 11, 12, 13, these 13 outcomes are my event G. So, you can see that we can describe these events using whatever we have defined earlier. So, now let us look at set operations on this card sets.

(Refer Slide Time: 05:52)

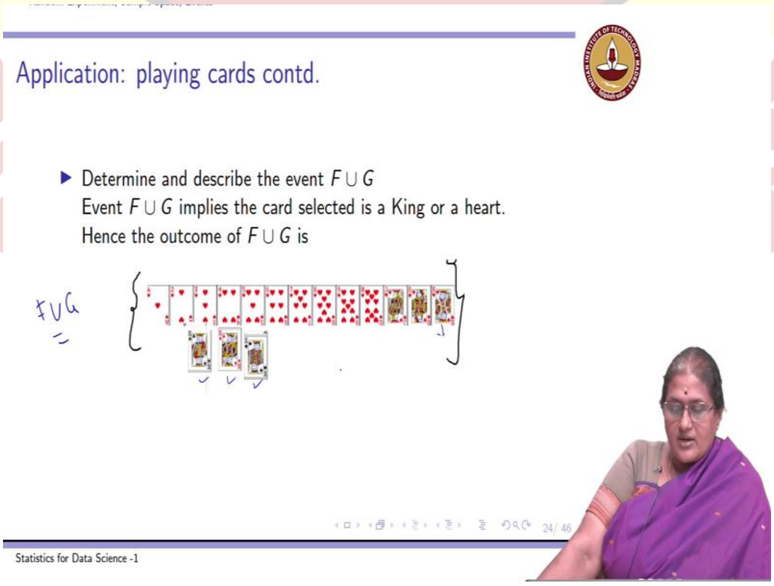




Now, suppose I want to define union of $F \cup G$, suppose I want to describe the event $F \cup G$. Now, F is the event that the card selected is a king, G is a event that the card selected is a heart. So the, by a definition the event $F \cup G$ is the event that card selected is either a heart or a king.

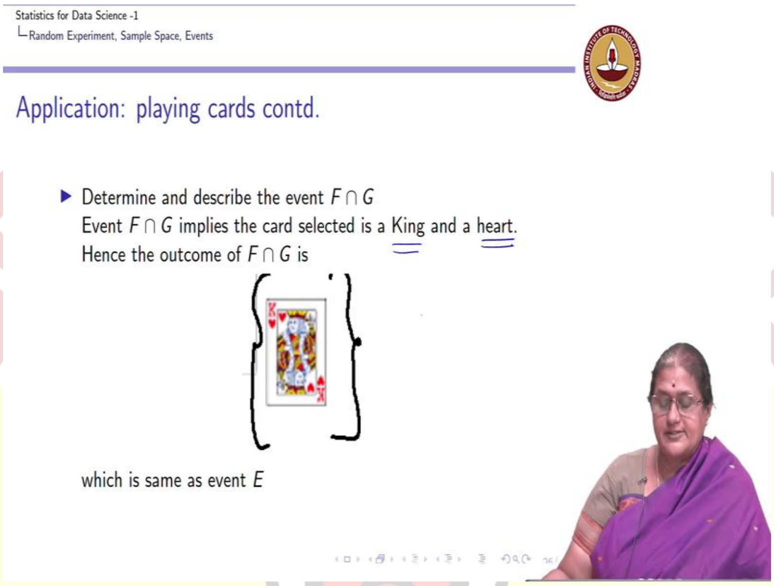
So, if you look at whatever is the, what are the cards that would satisfy it, these cards are hearts, these cards are king and I can see that this card that is king of hearts is actually common to both of them. So in addition to these 13 cards, I will have a king of spades, king of clubs and a king of diamonds, which would make 13 + 3, 16 possible outcomes in my set.

(Refer Slide Time: 07:07)



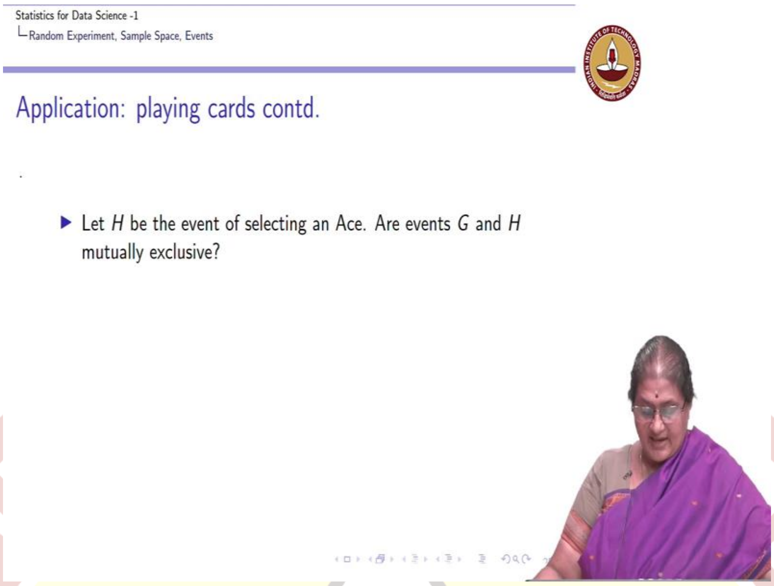
So, I can see that I can describe this event, $F \cup G$ to be this event here. Which has 16 possible outcomes, 13 which are from the hearts and three of the king, in inclusion with the king of hearts, I have these are my $F \cup G$. So, again to the deck of cards, I have applied the set operation $F \cup G$ and I have defined a new event which is my card is either a king or a heart. Now suppose I define an event that my card which I have taken is a king, and heart if that is the event I am going to define.

(Refer Slide Time: 07:57)



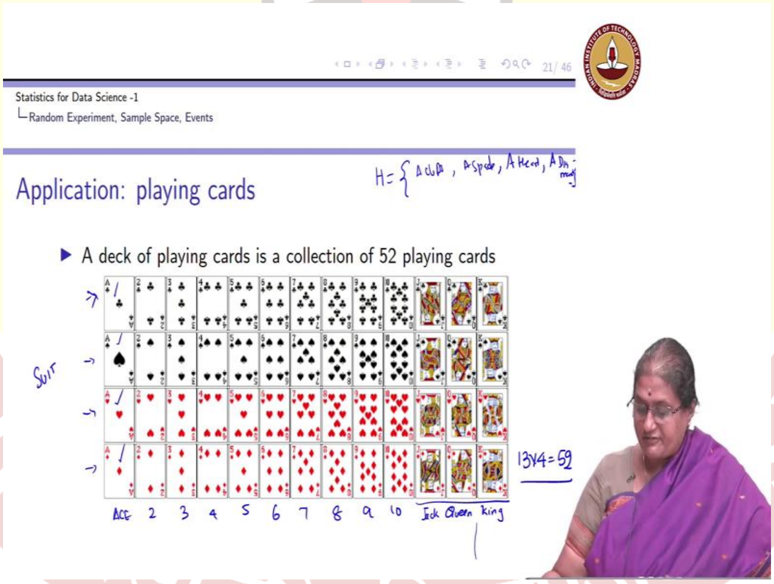
Now, suppose I have the event and I want to describe the event that my card selected as a king and a heart, recall F is a king and G is a this one, king and a heart, you can see that there is only one card that satisfies that it is both a king and a heart and you can actually notice that this is exactly the event E, where the event E was choosing a king of hearts.

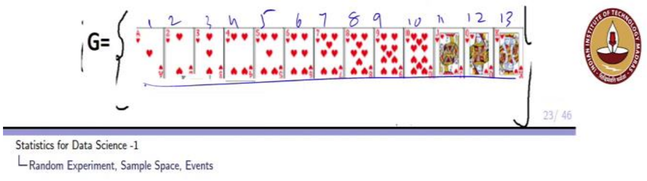
(Refer Slide Time: 08:30)

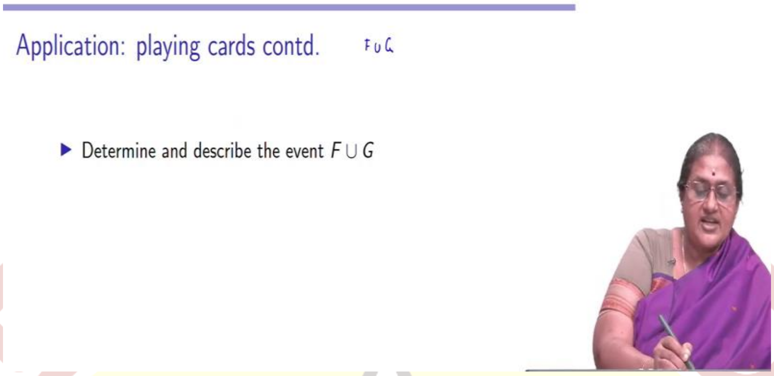


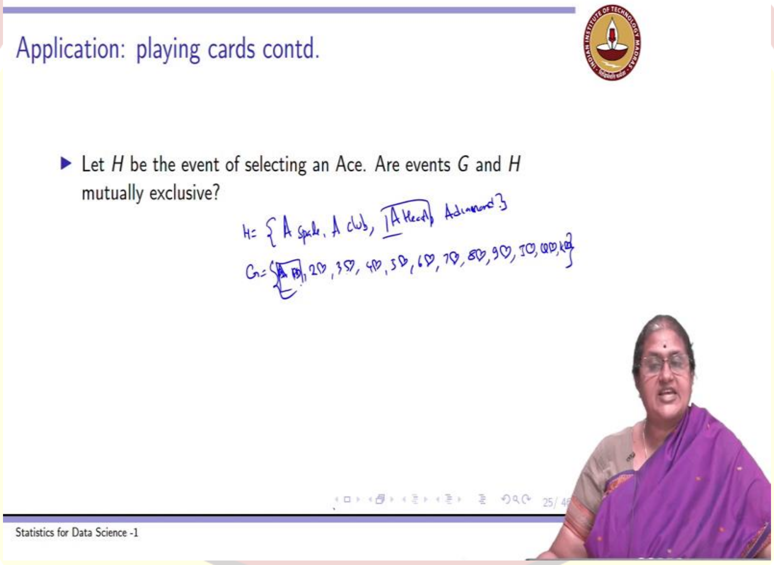
So, now let us continue with our application on playing cards. Suppose H is the event of selecting an ace. Again, let us go back to our deck of cards. So this is our deck of cards here.

(Refer Slide Time: 08:51)









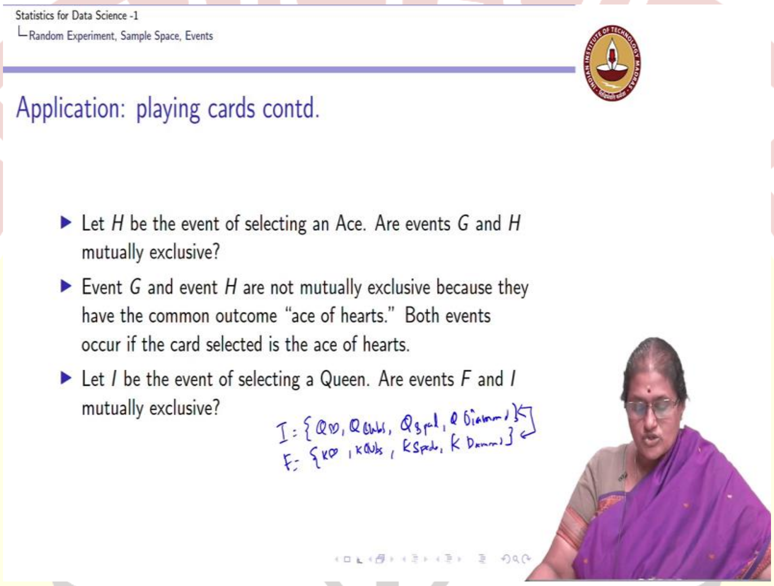
So, I have 52 cards and my event is selecting a ace. So, you can see that I have four aces, ace club ace, spade ace, heart ace and diamond ace. So, if I am selecting an ace and I am defining H to be the event of selecting an ace, it could be ace clubs or ace spade or a heart or a diamond.

Any one of these four outcomes that is my H. So if I have H is defined in that way and I want to see whether I am defining H as this is a collection of playing cards. So H, so my event H is A spade, A club, A heart and A diamond. Recall what is G? G is the event of having a heart that is how we define the event G.

If you recall G, go back to our event G, this is our event G, these are the outcomes in my event G we can see that the outcomes in my event G are a hearts, two hearts, we can see that a hearts is one of the outcomes in my event G and you can see that this a hearts is also an outcome in my, so a all heart. So, I have a heart, I have 2 hearts, I have 3 hearts, I have 4 hearts, 5 heart, 6 heart, 7 heart, 8 heart, 9 heart. So, this is a 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, jack heart, queen heart and king heart.

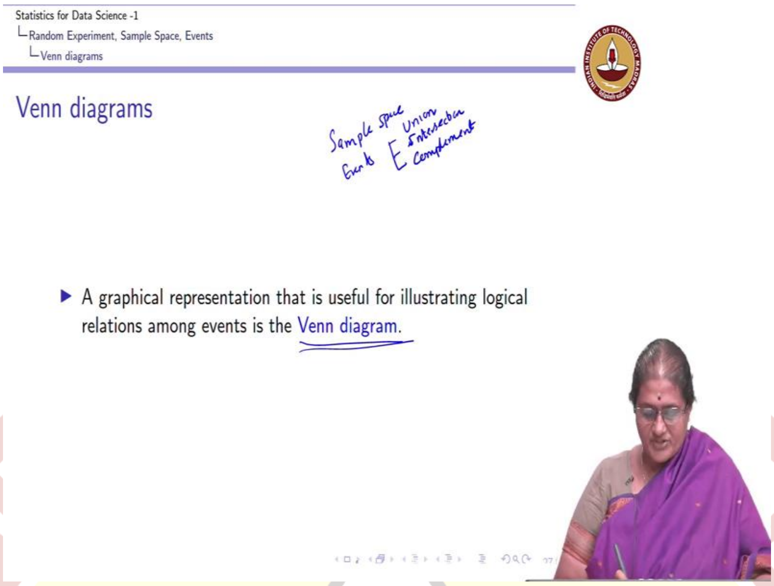
So, this a heart appears in both G and H. So, events G and H are not mutually exclusive. So, if H is the event of selecting an ace and G is an event of selecting an heart, we see that these events have H and G are not mutually exclusive.

(Refer Slide Time: 11:32)



However, G and H, if I is the event of selecting a queen again, Q is going to be queen of hearts, queen of clubs, queen of spades and queen of diamonds. My F is the event, this is event I my F is the event of king of hearts, king of clubs, king of spade and king of diamond, I see that there is no common outcome in both these events. So, the events F and I are obviously mutually exclusive that I can either select a queen card or a king card I cannot select both, a card which is both king and queen. So, the events F and I are mutually exclusive.

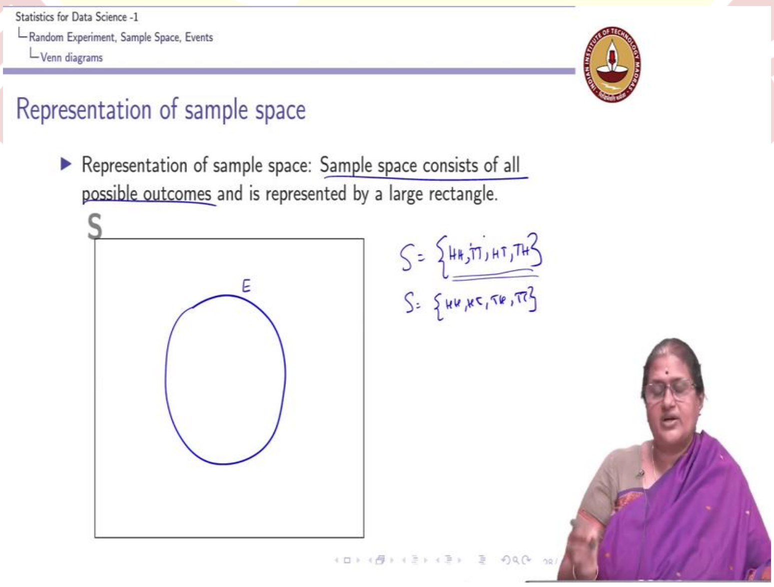
(Refer Slide Time: 12:42)

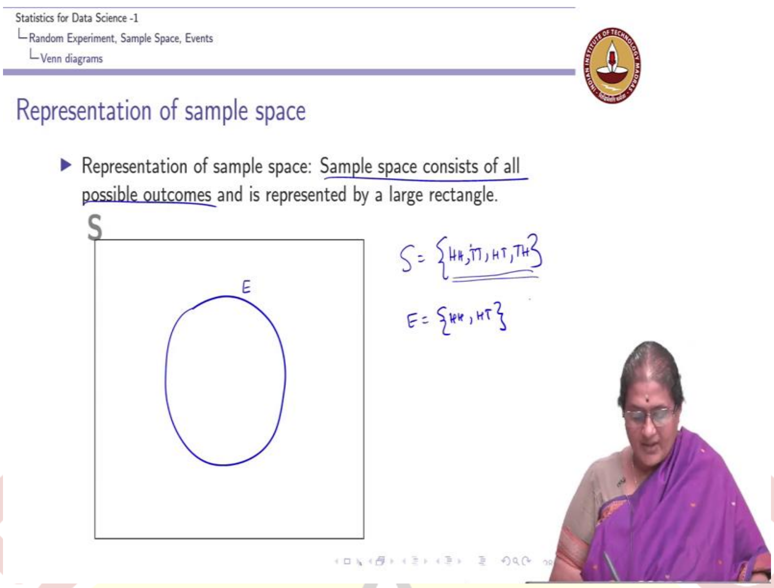


So, now we move forward to look at how to represent the events the sample space or how do we represent this in a useful way. A graphical representation that is very useful in illustrating logical relations, because we have seen what we have seen so far is, we have a sample space and we have events which are subsets of the sample space and we could define all set operations namely the union, intersection and complement on events.

So, if I can have a graphical representation, then logical relations between these events can be easily represented. Recall when we have sets, you have learned about what is called a Venn diagram. So similarly, in our probabilistic framework, we are going to represent these events and the sample spaces using our Venn diagrams.

(Refer Slide Time: 13:59)



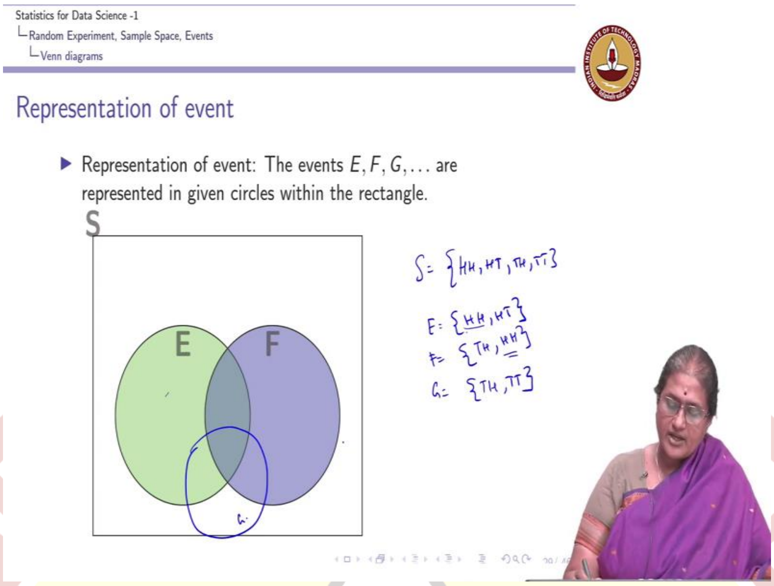


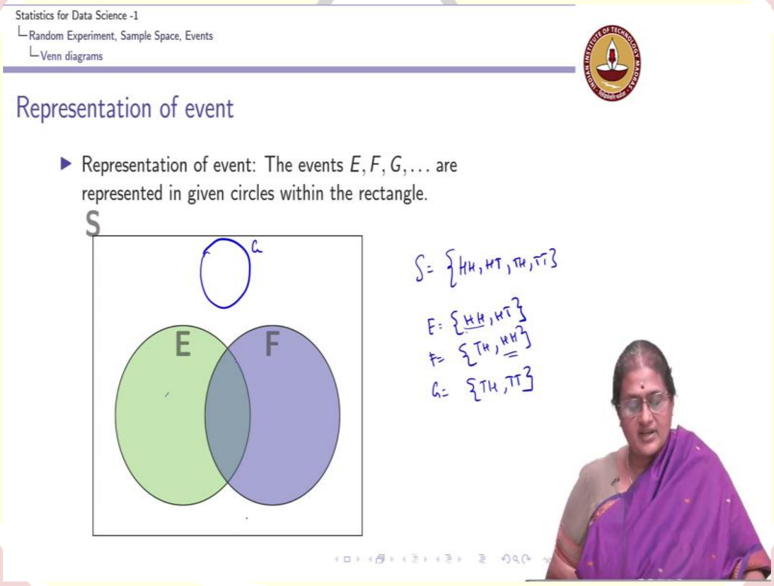
So, what is the convention? How do we represent a sample space, because sample space consists of all possible outcomes, it is just represented by a large rectangle. I am not listing all the outcomes, I am just representing the sample space by a large rectangle. Now, a event is a subset of this sample space. So, we represent a event using a circle.

For example, this is my event. So if I have a sample space, which is tossing a coin twice and the outcomes are HH, TT, HT, TH again notice the order in which I list the outcomes is not of any relevance. So, I could have, it is the same thing that if I listed as HH, HT, TH and TT this equivalent to this set.

So, the order in which I list the outcomes is not of any great importance, but they represent the sample space with the rectangle and suppose E is my event of head in my first toss. So, HH, HT are the outcomes which are in my event E, E is a subset which I represent using a circle.

(Refer Slide Time: 15:25)

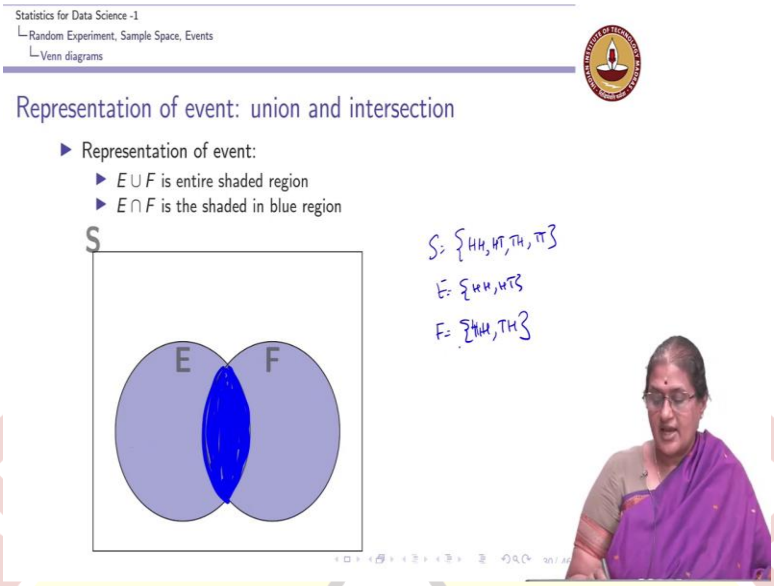


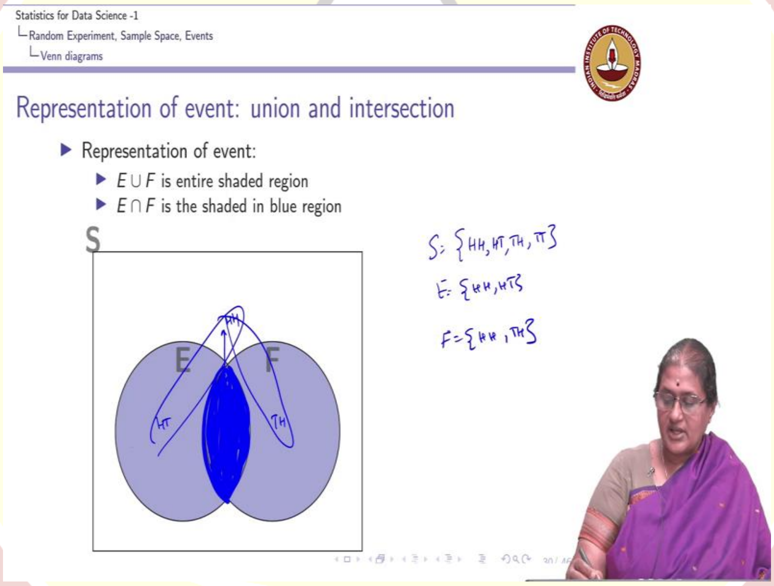


So, the first thing is we represent an event using a circle. So, I could have many events, so I have an event E, I have an event F, so I can represent, so for example, if S is my HH, HT, TH, TT that is represented by my rectangle. Event E could be head in my first toss which is HH, HT, event F could be head in the second toss which is TH, HH, what is common between these two is this outcome HH is common, I could have another event G which is tail in my first toss, which is TH and TT.

So, we can have all possible events and I could have another event G which is again listed here and you can see that there is nothing common to all the three events. So, G could be an event here, which is not common to this event E and F.

(Refer Slide Time: 16:37)



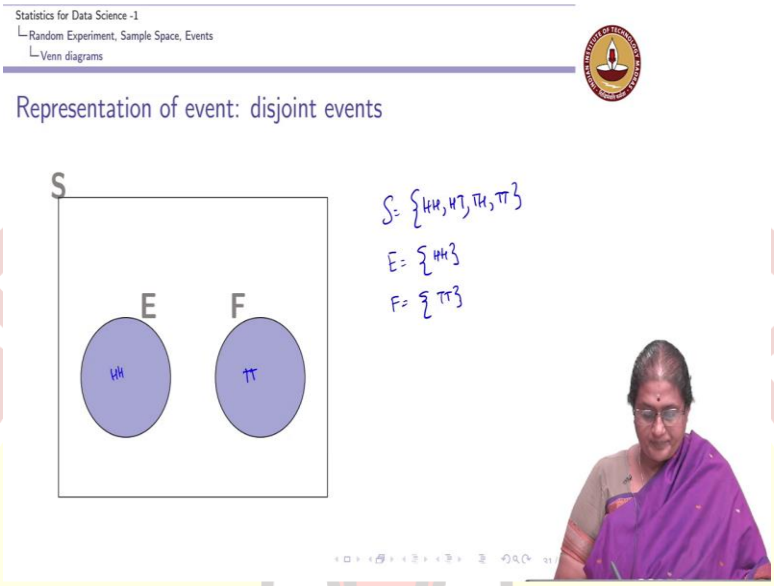


So, that is comes to us how we represent a union and intersection. So, this E ∪ F the entire shaded region is my E ∪ F, whereas this region which is shaded in blue, it does not look like blue, but let me put a different colours here. So, if this is, so this region, which I am going to highlight now.

So, this region, which is in blue, fluorescent blue, is my E ∩ F event. So, if we are going back to our example. If my sample space is H tossing of two coins, the outcomes of tossing of two coins and my F is head in my first toss. So, this are my outcomes HH and HT, HT. F is head in my second toss.

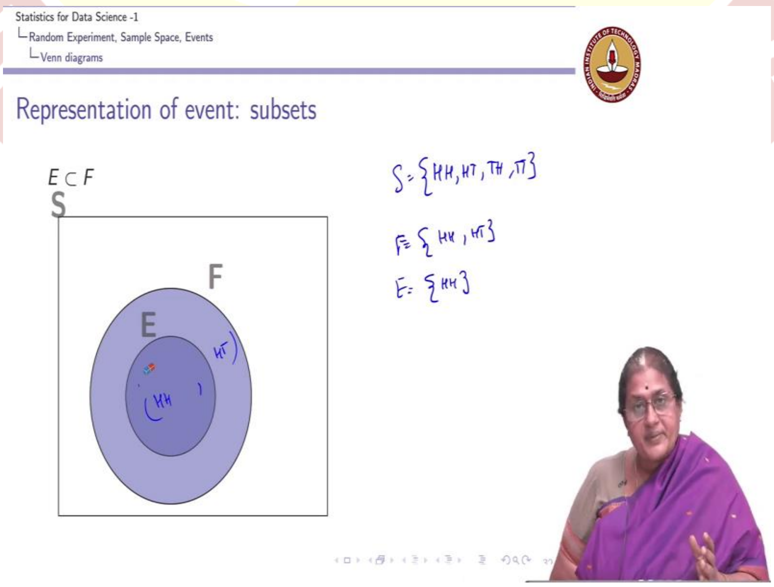
So, this has, suppose my E is head in my first toss, F is. Suppose, E is head in my first toss, let F be head in the second toss. So, you can see that the outcome HH is in my shaded region, HT would be here and TH which would be here. So, these two outcomes are in my F these two outcomes are in my T and this outcome HH is common to both E and F which will appear in the shaded region

(Refer Slide Time: 18:48)



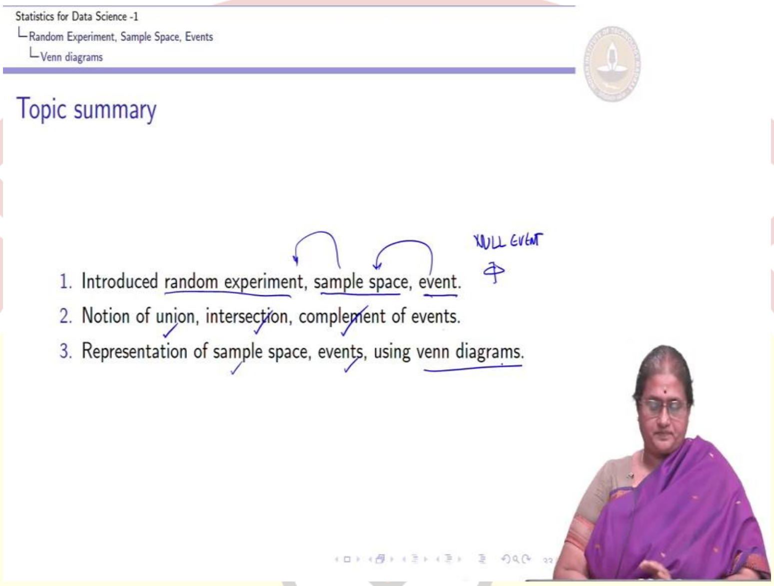
Now, if two events are disjoint, for example, again go back to our tossing a coin twice, I have HH, HT, TH, TT, I define an event E which is head in my first, head in both the tosses together it is HH. F is an event where I define tail in both the tosses. So, you can see that HH is an event here, TT is an event here, there is nothing common in these two or they are mutually exclusive that is head in both the tosses and tail in both the tosses are mutually exclusive of each other and they are disjoint events and I can represent that event in this way.

(Refer Slide Time: 19:32)



Now, suppose I again go back I have a sample space HH, HT, TH, and TT head in the first toss is my let F be in my event, head in my first toss, so that would be HH and HT. Let E be my event head in my first and second toss, so that is going to be HH, we can see that E which is only HH is a subset of F which has both HH and HT and this is a way I can visually represent subset events where E is a subset of F.

(Refer Slide Time: 20:13)

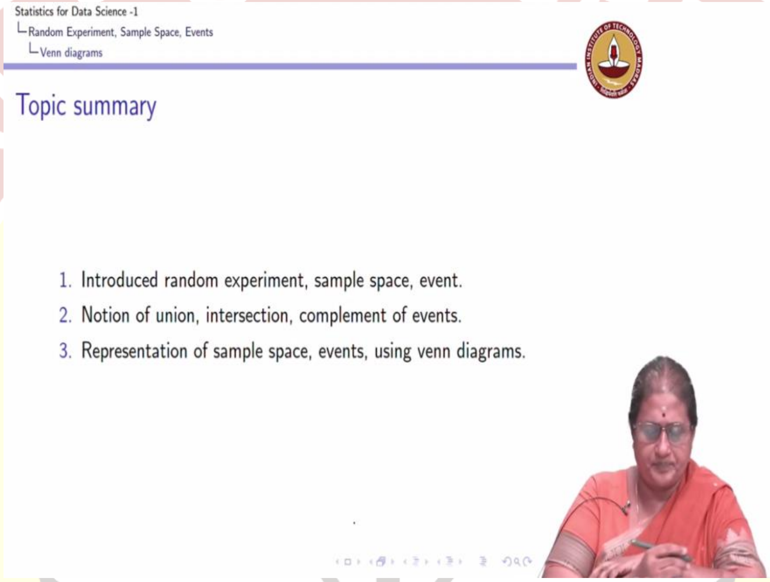


So, what we have learned in this lecture is, we have introduced the notion of a random experiment, sample space and event. Sample space is a set of all possible basic outcomes of the random experiment, event as a subset of a sample space. We also introduced a notion of what we call a null event.

Since these are sets we can introduce a notion of union intersection and complement of events. We have seen how to define these events, given a basic event and finally, we looked at how do we represent a sample space and event using Venn diagrams, because Venn diagrams is a powerful graphical representation of sets, since my sample space and events are sets, we have looked at how to represent these events using Venn diagrams. The next thing we are going to learn is how we assign probabilities to these events.

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)

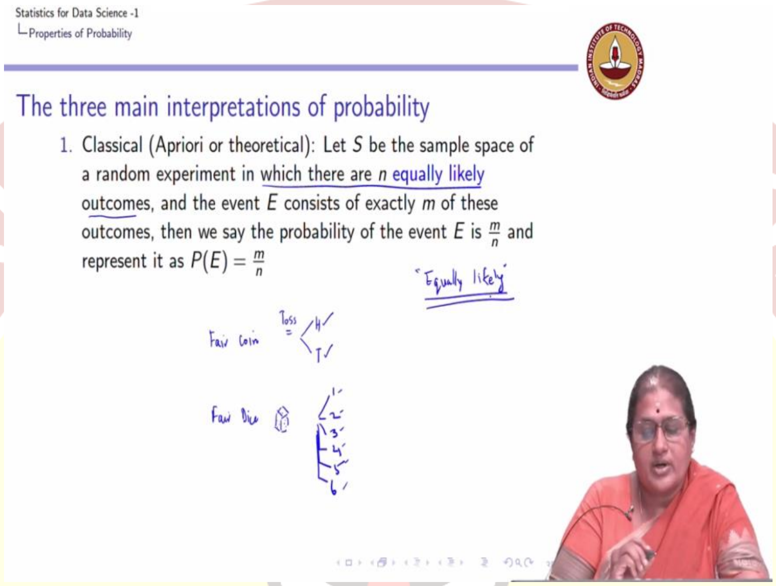


So, in the last lecture, we learned about what is a Random Experiment. Once we define what is a random experiment, we introduced the notion of a Sample Space. A sample space is the set of all basic outcomes of the random experiment. And we also introduced what was a Event. Now, we quickly realize that 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.

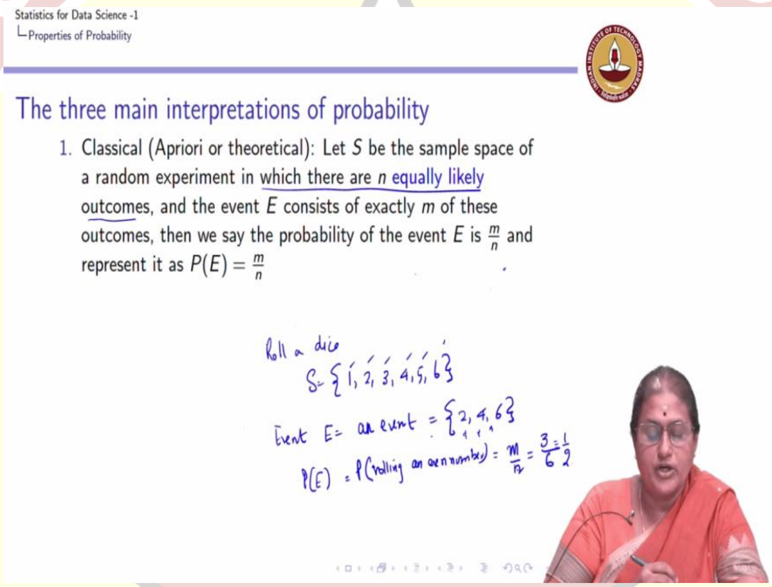
One of the main interpretations and most popular interpretation of probability is what we term as the classical approach. Now, this classical approach is also referred to as a priori or theoretical approach to probability. So, what this approach is, is let S be the sample space of a random experiment 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.

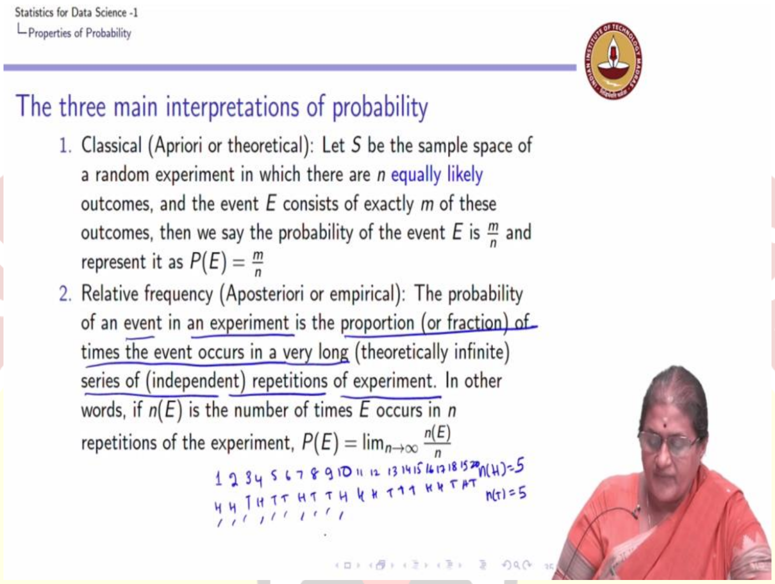
(Refer Slide Time: 04:52)



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)

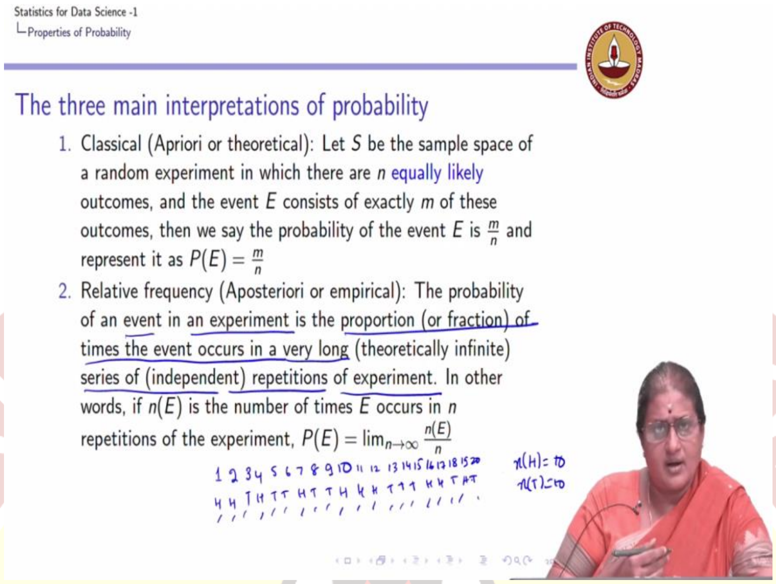


The next approach which is also a very popular approach is termed as a relative frequency or empirical or Aposteriari approach or interpretation to probability. What this approach tells us is the probability of an event in an experiment is the proportion of times the event occurs in a very long series 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.

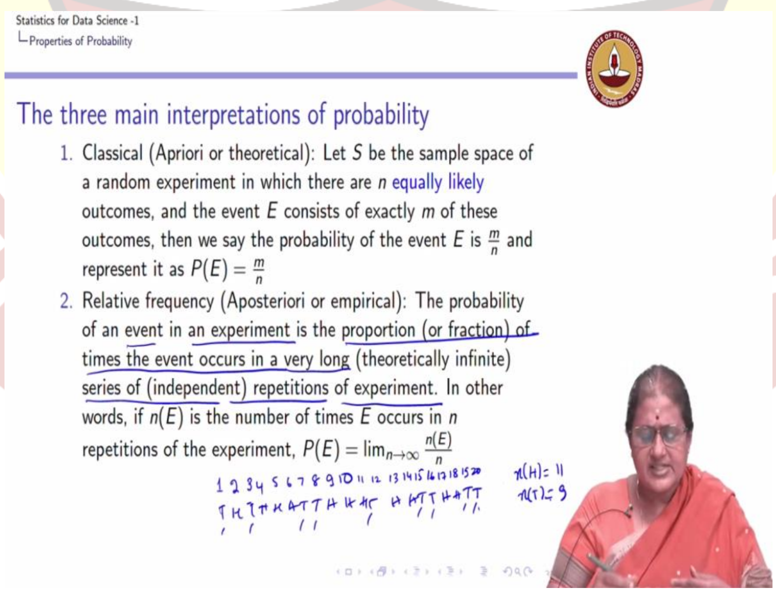
Now, this can easily be replicated by any one of us we can take a coin and keep tossing the coin. So here in 10 tosses I find 1, 2, 3, 4, 5 heads and 1, 2, 3, 4, 5 tails. I repeat this experiment, so I can just go to 11, 12, 13, 14, 15, 16, 17, 18, 19, 20. So, if I keep repeating this experiment and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

(Refer Slide Time: 08:47)



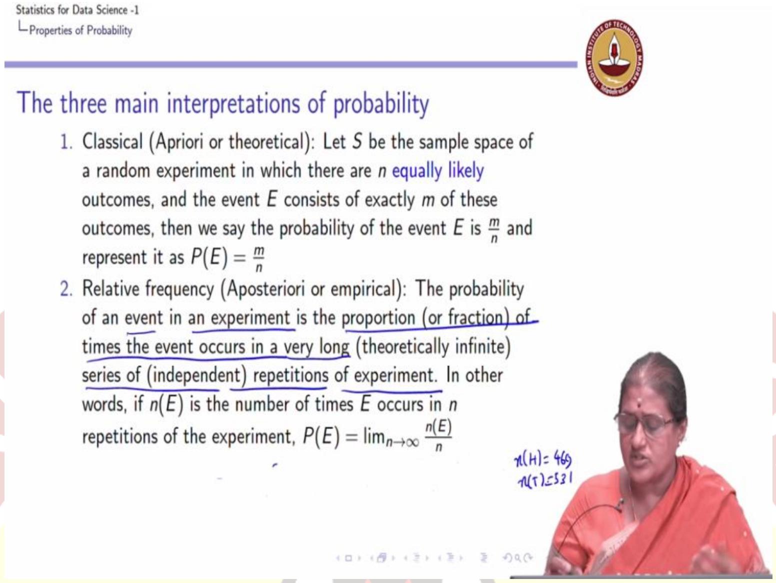
So now the number of heads I have is 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 and tail is also 10, but will this pattern always occur? need not be. For example, I could have a completely different suppose another person is tossing 20 tosses of the coin and they just randomly, this is the sequence they get.

(Refer Slide Time: 09:19)



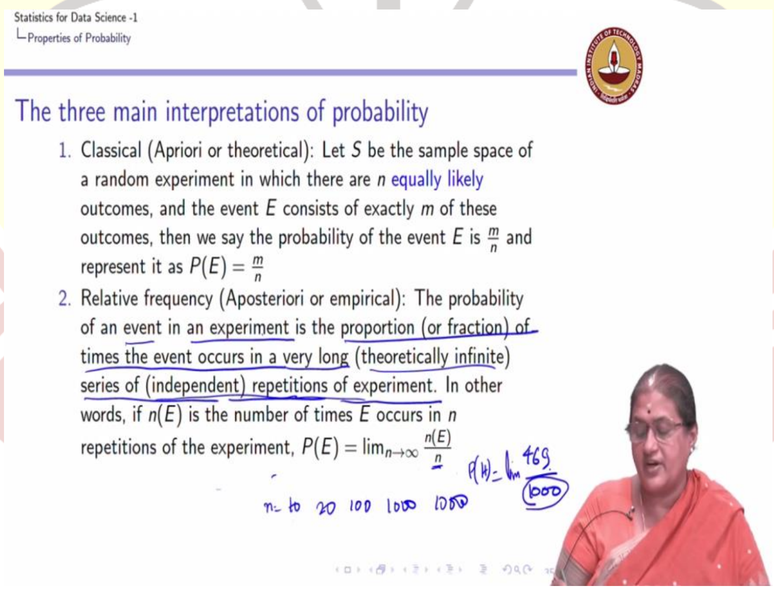
They get a tail, head, tail, tail, head, head, tail, tail, head, head, head, tail, head, head, tail, tail, head, head, tail, tail. So here, if I count the number of tails, I have 1, 2, 3, 4, 5, 6, 7, 8, 9. And number of heads is 11.

(Refer Slide Time: 09:55)



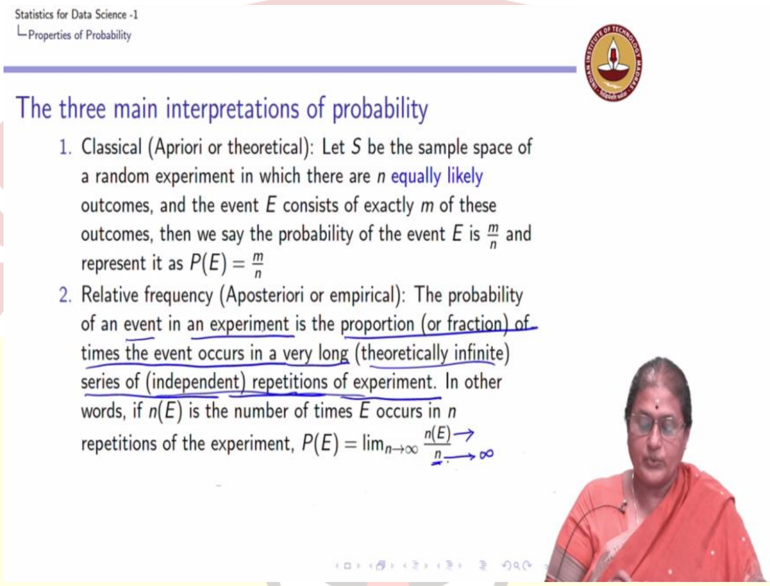
So, imagine that instead of doing this experiment just 10 times or 20 times I am doing it a 1000 times, then it could be very well possible that I am getting 469 heads. I am doing it a 1000 times I am getting 531 tails, I repeat it 10,000 times. So, we are counting the number of times.

(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.

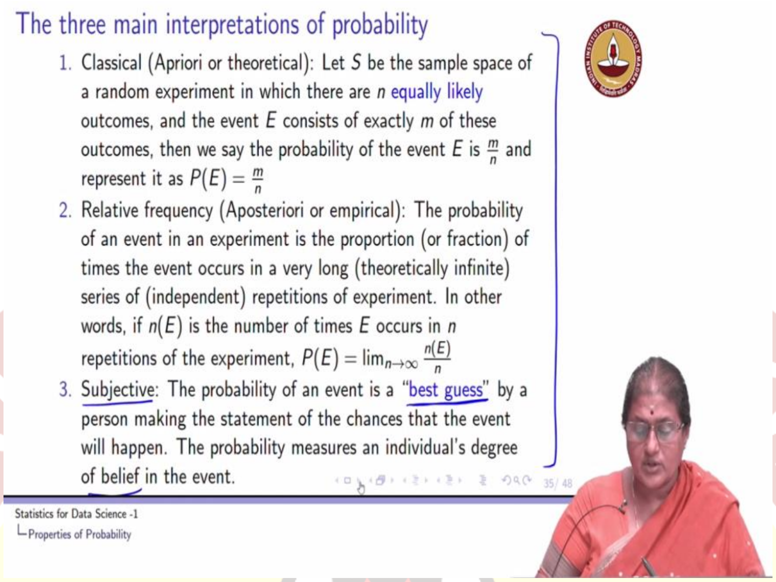
(Refer Slide Time: 11:12)



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.

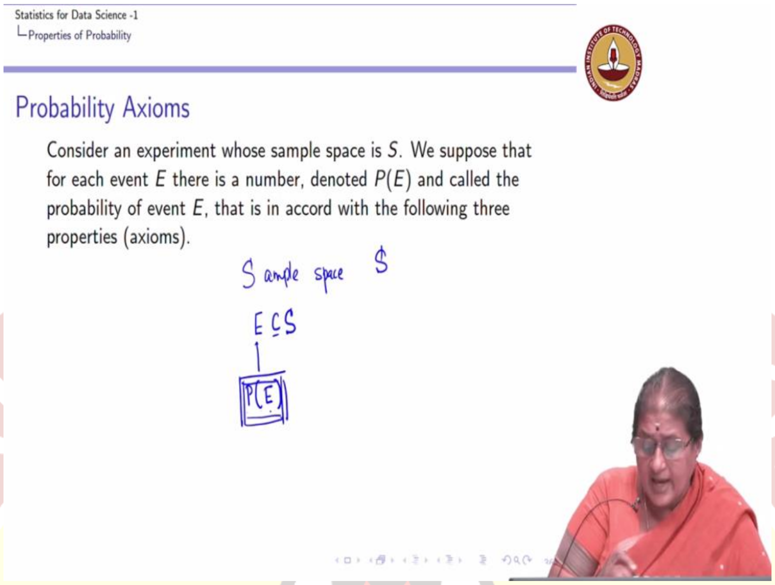
(Refer Slide Time: 12:23)



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.

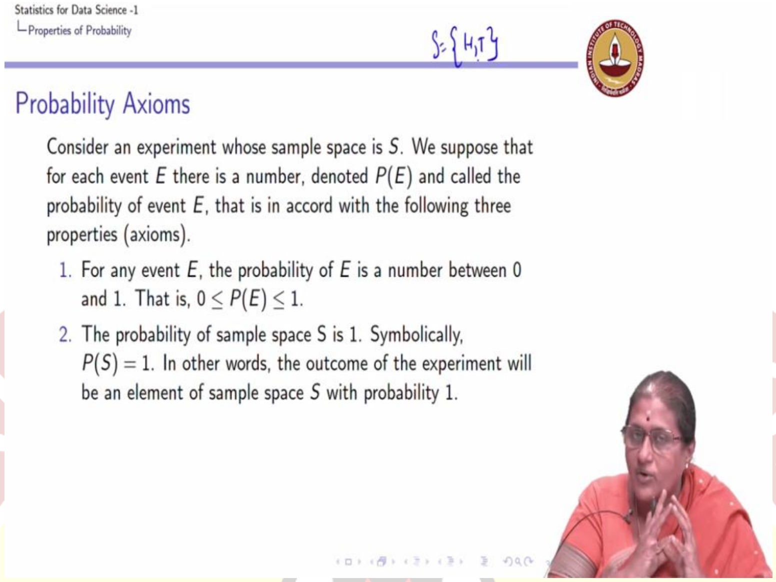
(Refer Slide Time: 13:30)



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?

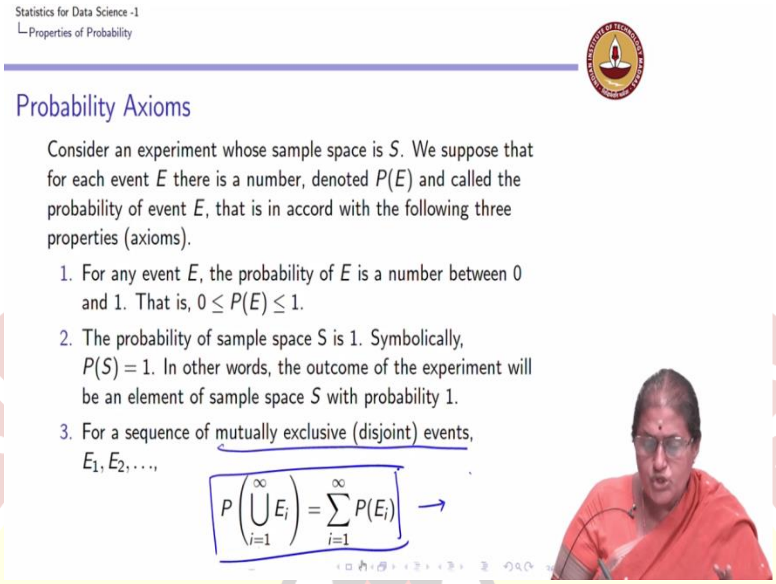
(Refer Slide Time: 14:56)



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.

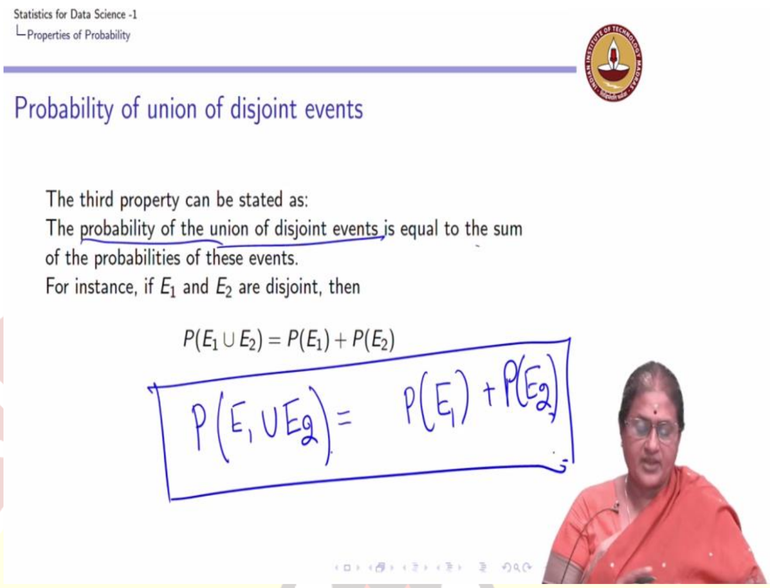
(Refer Slide Time: 16:22)



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

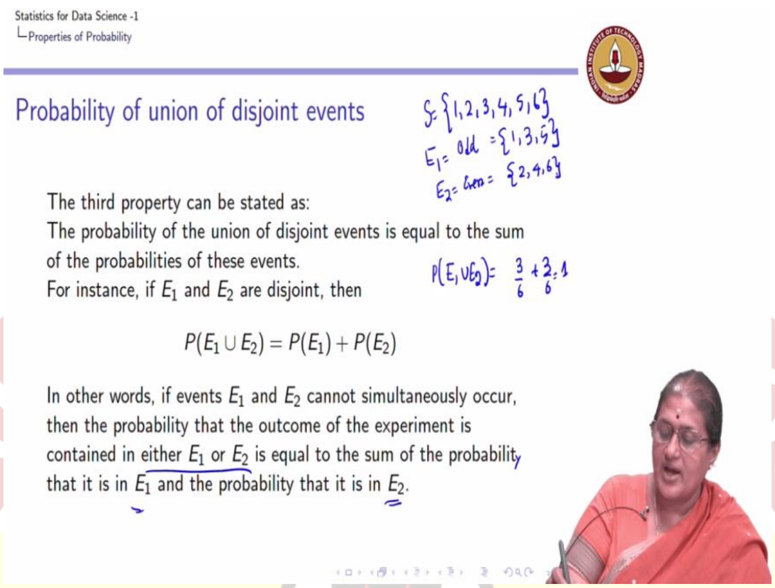
(Refer Slide Time: 17:37)

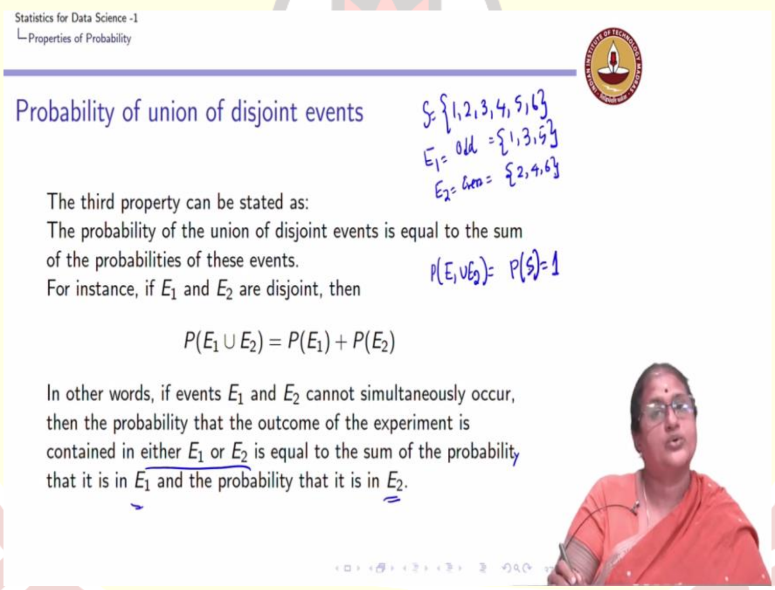


If I have mutually disjoint events, so I for the purpose of this course, I am just going to look at the finite union even though this applies for larger or countable union, we are not going into the mathematical foundations of understanding that at this point of time, but for this course, I can state 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)





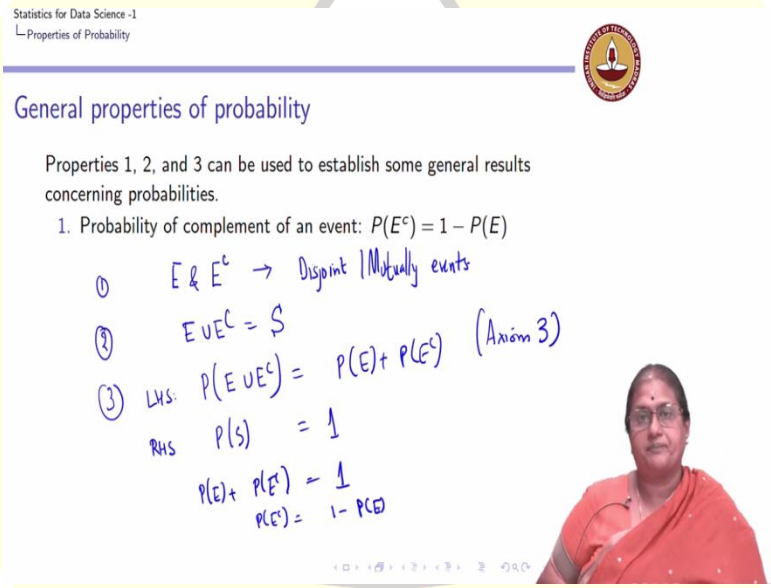
So, what is the understanding and why is this property important? It states that if E_1 and E_2 are 2 events that are disjoint. In other words, they cannot occur simultaneously, then the probability of the outcome that the experiment is contained in either E_1 or E_2 is equal to the sum of probabilities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

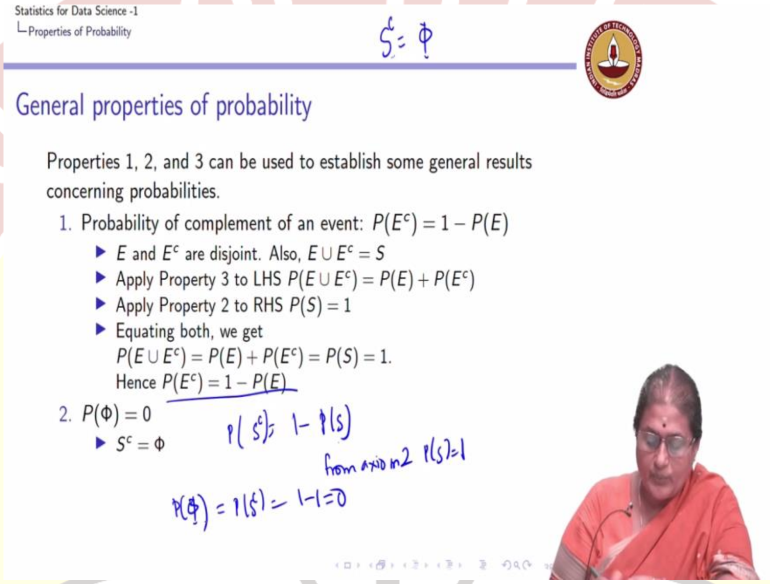


The first problem property of probability is, remember when we talked about sets, we talked about union intersection and complement of sets? So can we work out what is the probability of a complement of a set? Now when we look at a compliment, so if E is a event, E^C a sets event, we know the first thing 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.

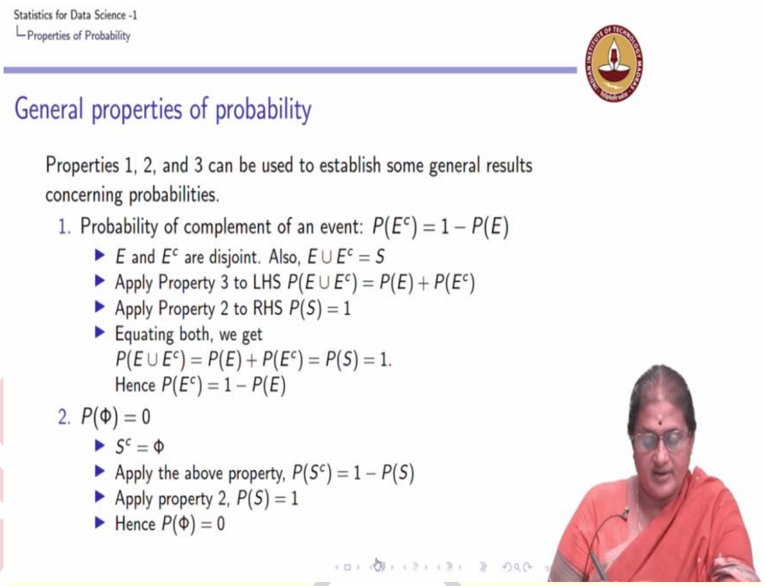
Now, axiom 2 states probability of the sample spaces is 1. So, my left hand side says the $P(E \cup E^C) = P(E) + P(E^C)$. My right hand side states $P(S) = 1$. I equate these 2 to get $P(E) + P(E^C)$ =1, which gives me $P(E^C) = 1 - P(E)$ (Refer Slide Time: 23:27)



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.

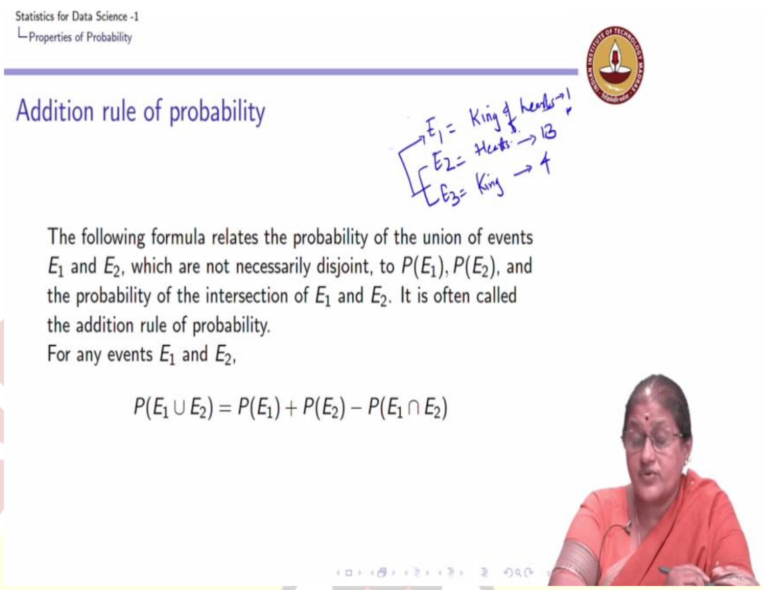
So what, what is the probability of an Null event? Now what is the complement? I know the complement of my sample spaces in our living sample space is the collection of all the outcomes null event does not have any outcome. Hence, the complement of the sample space is my null event. And I know from my 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.

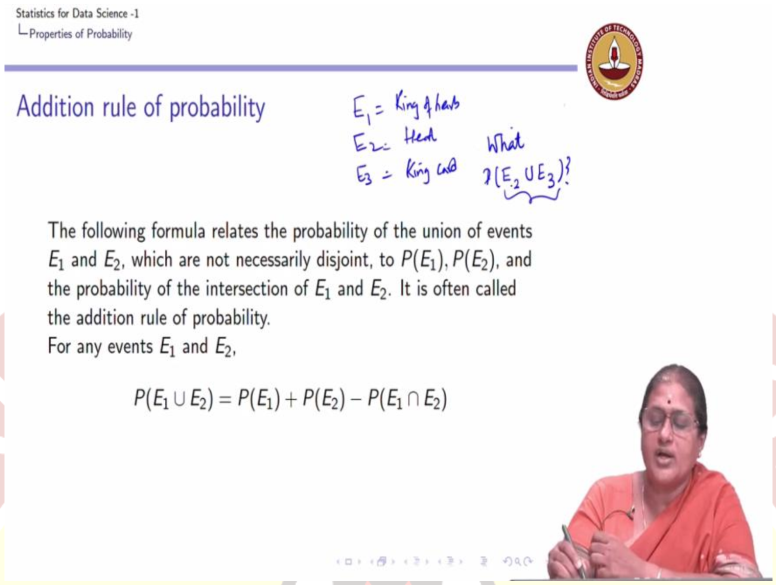
(Refer Slide Time: 25:15)



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.

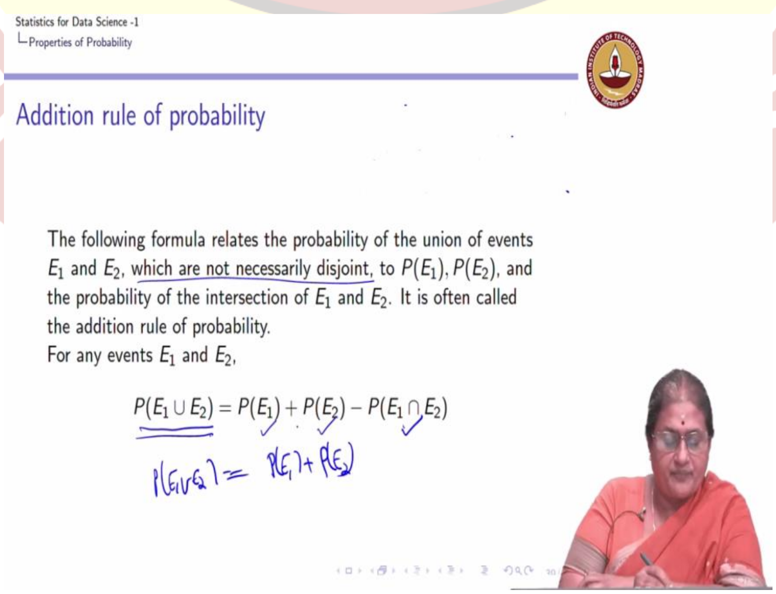
The second event we defined was hearts. And then after that, you can see that these 2 events, this had 13 cards, this had only 1 card, the third event we defined was a king. So, this had 4 cards, $E_2$and E_3had 1 card which is in common, and that was a king of hearts. So, you can see that many a time 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)



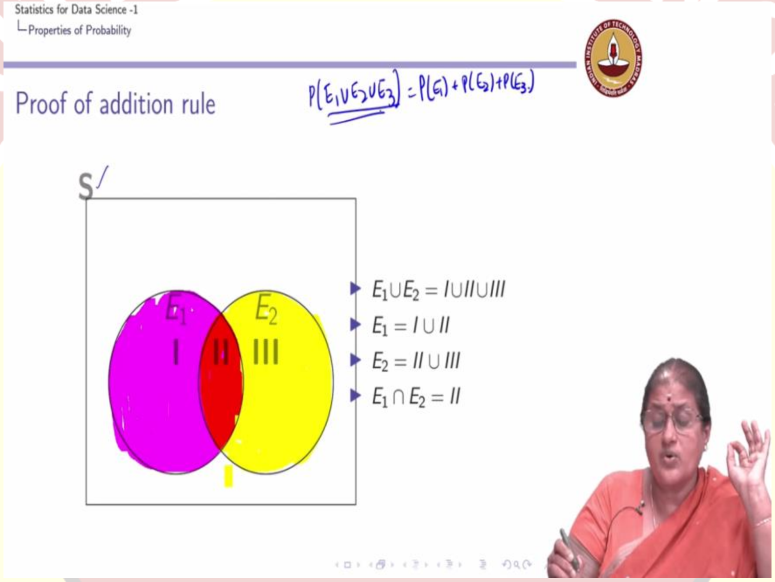
So, if you go back my E_1 was king of hearts. My E_2 was a heart. My E_3 was a king card. So, if my question is what is the probability that I have the draw card? What is the probability that I either have a heart or a king? This is a probability I am interested in finding out but now, E_2 and E_3 are 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.

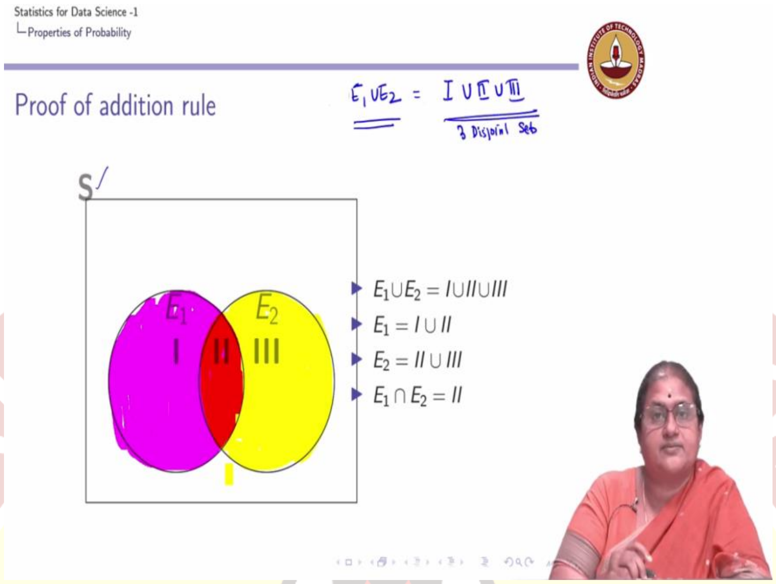
(Refer Slide Time: 28:57)



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 $E_1 \cap E_2$.

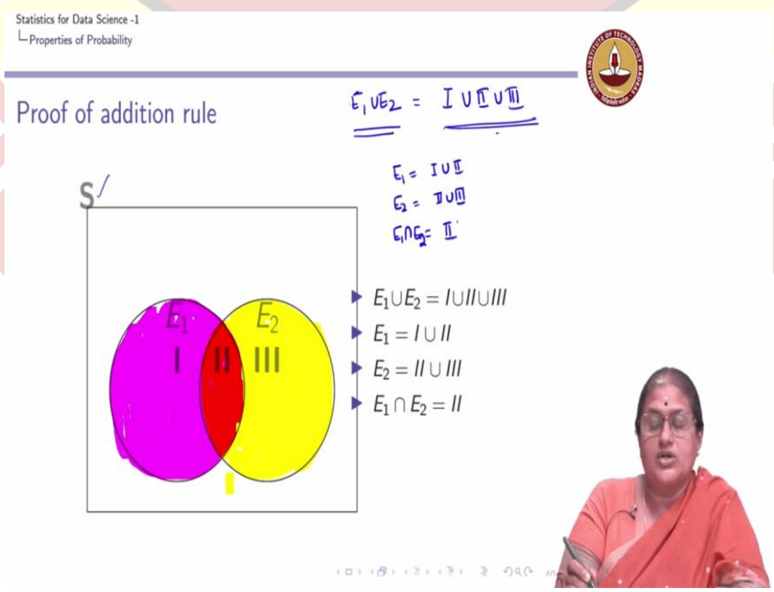
So we see that there, event E_1 and E_2 are not disjoint and they are not mutually exclusive. But I have a theorem or I have I but, but I have an axiom, which says $P(E_1 \cup E_2 \cup E_3) = P(E_1) + P(E_2) + P(E_3$, provided they are mutually exclusive or disjoint.

(Refer Slide Time: 30:59)



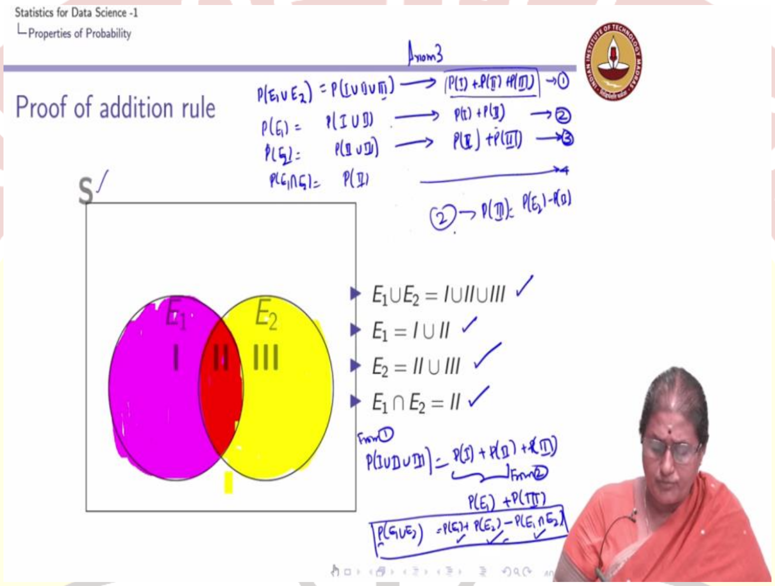
So, what I am planning to do, we recognize that this $E_1 \cup E_2$ is this entire shaded region of the purple, orange and yellow shaded region, I write this entire shaded region as the following. I write it as 1 ∪ 2 ∪3 were 1, 1 and 2. And 3, as you can see from the Venn diagram, are disjoint, I do 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.

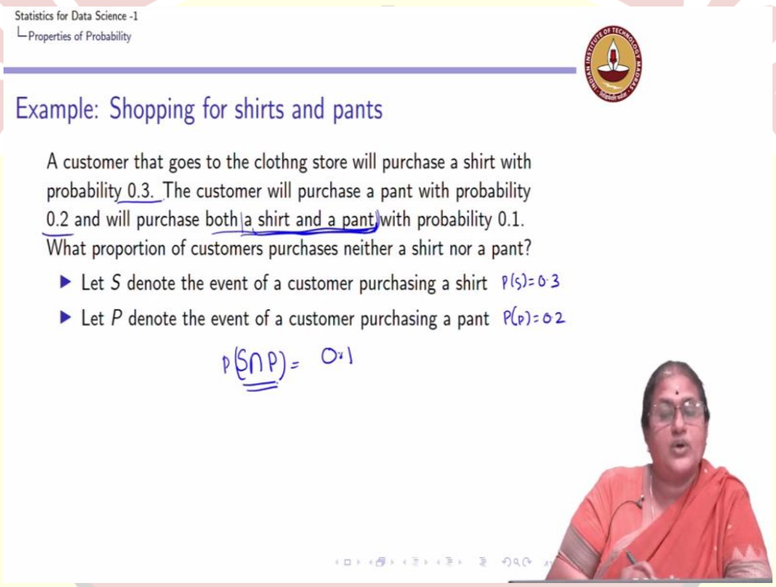
Now, let us actually express this probability in terms of E_1, E_2, and $(E_1 \cap E_2)$. Probability of 1 plus probability, so if I look at P(1 ∪ 2 ∪ 3), I have this is P(1)+P(2)+P(3). This comes from 1. Now, 2 from 2 I have this is equal to $P(E_1) + P(3)$ but $P(3)$is that invert $P(E_2) - P(2)$,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

Statistics for Data Science 1
Professor Usha Mohan
Department of Management Studies
Indian Institute of Technology, Madras
Lecture 6.5
Probability - Applications

So, let us look at a few applications of this addition rule.

(Refer Slide Time: 36:50)



So, suppose we have a customer who goes to a clothing store, and we know that somebody gives us the following information that the person can purchase a shirt with the probability of 0.3, this is a clothing store, the customer will purchase a pant with a probability of 0.2 and will purchase both a shirt and a pant with probability 0.1. So, the question asked is what proportion of customers' purchases neither a shirt nor a pant? That is a question that is being asked.

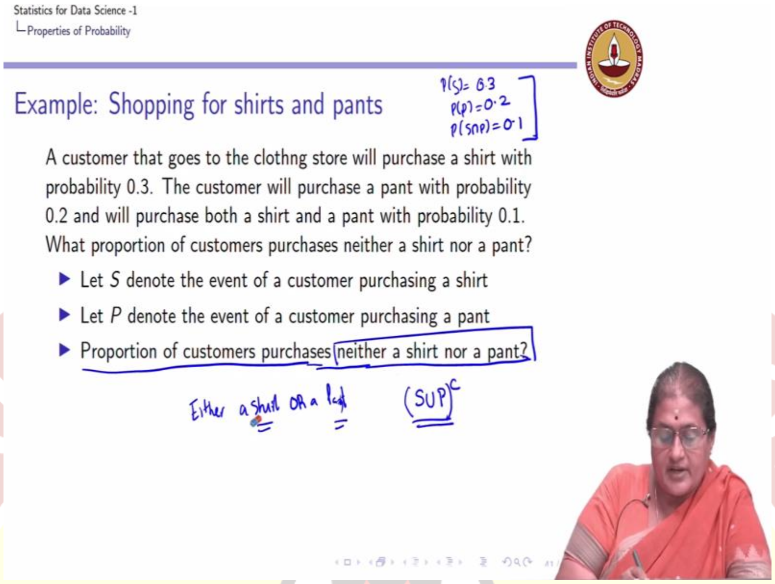
So, if I am a shopkeeper, I would be interested in knowing the proportion of customers who visit me who will make no purchase. So, how do I translate this problem to what we have learned so far? So, first step is to identify the information as events. So let us identify the events. So, the first event is let me denote that event by S. So what is S? S is the event of a customer purchasing a shirt.

Now, if the shopkeeper is certain that every person who enters the shop will make a purchase there is no uncertainty here. But we all know that when a first customer enters a shop, they might buy a shirt. It is not will buy a shirt so there is that element of uncertainty. And this I am denoting by S is the event of a customer purchasing a shirt and what is given to us this the probability with which your customer buys a shirt is 0.3.

Similarly, the next event is the event of a customer purchasing a pant I know probability of S is 0.3. Similarly what is given to me is the probability a customer purchases a pant is given to me as 0.2. And some other information that is given to me is the probability that a customer would purchase both the shirt and a pant. Now this information a shirt and a pant can be captured by $S \cap P$.

Now, this is what is most important that is we are translating an information from a situation in real life, we are abstracting it to a mathematical expression. What is the mathematical expression $S \cap P$ is the event that a shirt and pant is being purchased and that probability is given to be 0.1. So, this is what is given to us.

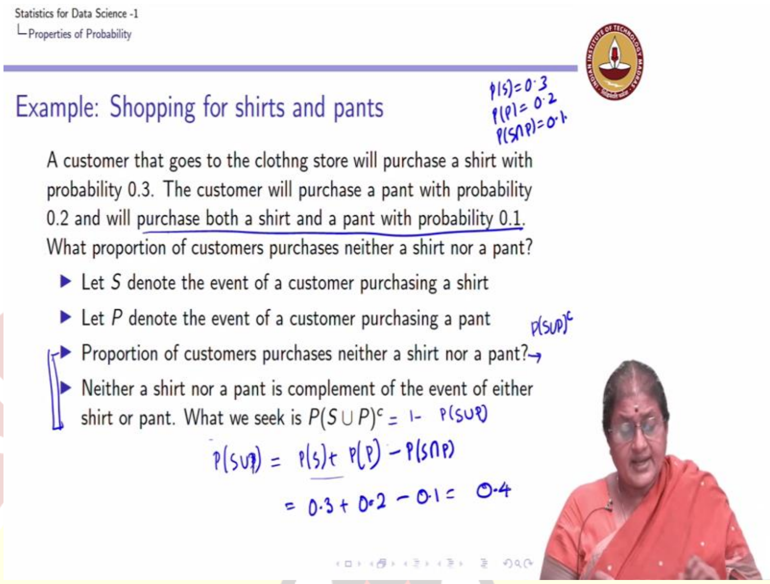
(Refer Slide Time: 40:34)



What is needed now, so, I am given $P(S) = 0.3$ $P, (P) = 0.2$ and $P(S \cap P) = 0.1$ this is the information that is given to us. We are interested in knowing the proportion of customers or probability that a customer does not make a purchase in other words neither a shirt or a pant. Now, what is that expression that can be used to capture this.

Now, recall either a shirt or a pant, either A or B is captured by $S \cup P$, either shirt or pant is captured by a operation union. So, neither a shirt nor a pant is the complement of this event.

(Refer Slide Time: 41:54)



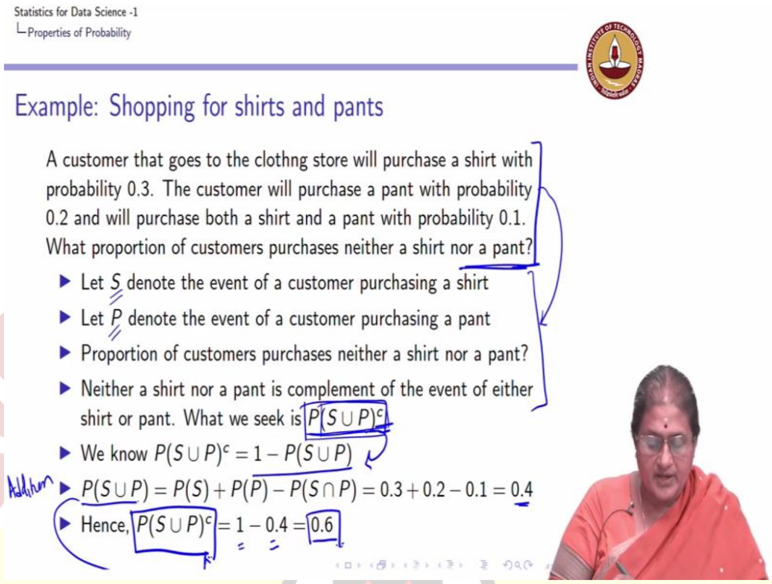
So, either a shirt or a pant is the union of this event, neither a shirt or a pant is the complement of the event either shirt or pant. So, what we seek to answer this is probability of $(S \cup P)^c$. Now, these are crucial observations and these are the crucial things that you need to know. Nobody is going to give us find out $P(S \cup P)^c$ directly. We need to understand that given the problem, this is what is being asked or this is what is being sought.

Once we pose a problem in this framework, then we apply the laws of probability or the axioms of probability to come up with the answer. So, what is that that we seek we seek $P(S \cup P)^c$. Now, I know the probability of A complement is 1 minus probability of S union P this is what we had just derived a few minutes before. So, what is $P(S \cup P)$? So, we first asked the question are S and P disjoint?

What do we mean by that? Is does a person who buy a shirt does not buy a pant on vice a versa? The answer is no because I have a probability of a person who could buy both a shirt and pants. So, the event of purchase of shirt and pant together can happen. Hence, they are not disjoint events. So, the next thing is if they are not just disjoint I know from the additive rule of probability, this is $P(S) + P(P) - P(S \cap P)$.

Now, I have all these 3 probabilities I know $P(S) = 0.3$, I know $P(P) = 0.2$, I know $P(S \cap P) = 0.1$. Hence, I can just plug in those values to get $0.3 + 0.2 - 0.1 = 0.4$.

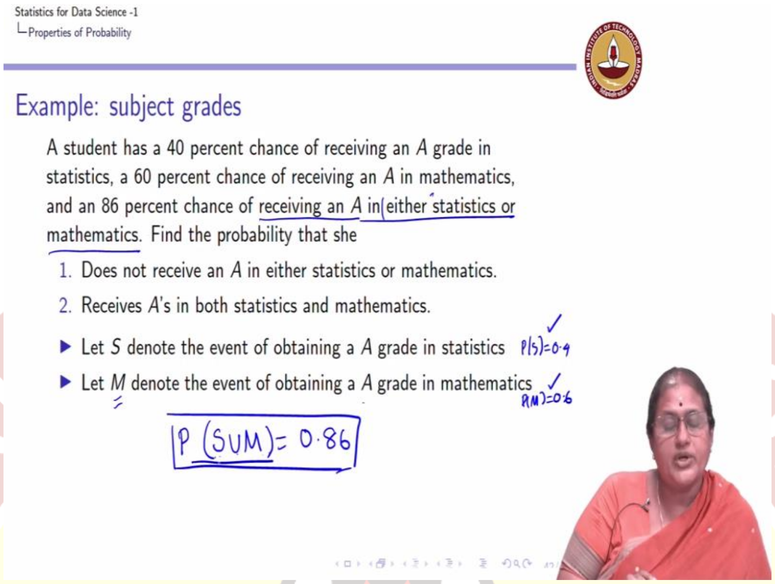
(Refer Slide Time: 44:09)



Hence, I can say that the probability of a customer or the proportion of customers who purchase neither a shirt nor a pant is, $1 - 0.4 = 0.6$. Because this is the probability I was seeking here. Hence this is the answer is 0.6. What I want you all to understand here is typically in real life situations, this is the situation that is given to us. From this situation, we abstracted and put it in a mathematical format or a framework by identifying what were the events and what is set in terms of the probability of which set are we seeking.

Once we post it in this format, we realize that this probability of complement is $1 - P(S \cup P)$, we applied the addition rule to find out the probability here and then we got the $P(S \cup P)^c$ through the complement rule. So, this is what we need to train ourselves to first identify what is given in a problem and post it in the mathematical framework and then tell it in the language that is sought after find out what is being asked. That is extremely important.

(Refer Slide Time: 45:51)



Now, let us look at another example. This is an example which students would relate to. So, I have a student who has a 40% chance of receiving an A grade in statistics, a 60% chance of receiving an A grade in mathematics and an 86% chance of receiving an A in either statistics or mathematics. I repeat, the student has a 40% chance of receiving an A grade in statistics, a 60% chance of receiving an A grade in mathematics, but an 86% chance of receiving an A in either statistics or mathematics.

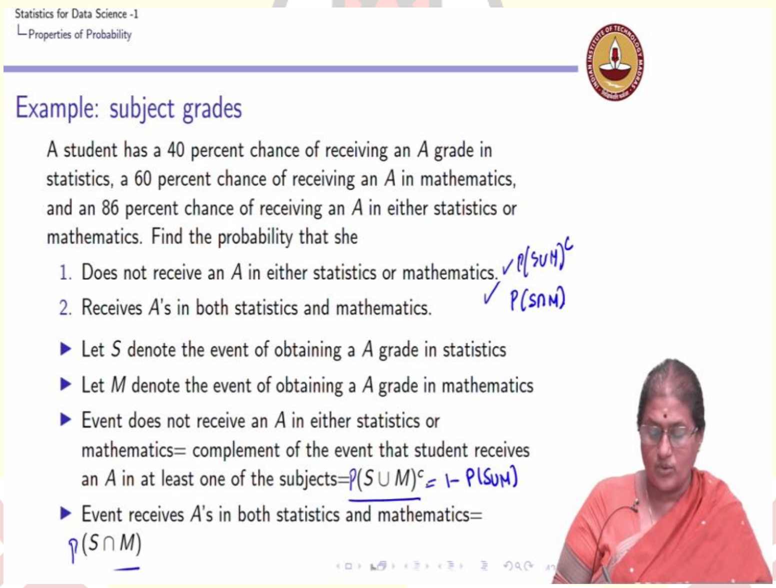
What is being asked in this question? What is the probability that the student does not receive an A grade in either statistics or mathematics? The second question is receives an A grade in both statistics and mathematics. This is the question that is being asked. So, again, we start what is the sample, what is the experiment here? I have a student who has written an exam. So, the outcome is the grades the student obtains in statistics and in mathematics.

So, we can see that this I have a class and in this class, I can find out what is the grades obtained by different students in subjects or I given. So, first, what we do here is first we identify the events, how do we identify the events? So, again, I go back, I identify my event, let S denote the event of a student obtaining of the student obtaining A grade in statistics. So, the first thing we identify is what is the event. Let S denote the event of obtaining an A grade and statistics this probability is given to me and that is given to me by this 40%. I write that as 0.4.

Similarly, the next thing which is given to me, let me denote M to be the event of obtaining a, A grade in mathematics. So again, I have a probability of S is 0.4. I have probability of M is 0.6. Now, what does this mean receiving an A in either statistics or mathematics. Recall either statistics or mathematics is represented by the union of these two events and it is given probability of statistics or mathematics is 0.86.

So, the information given to us is probability of receiving a A grade in statistics is 0.4. Probability of receiving an A grade in mathematics is 0.6. And the probability of getting in either of the subjects is 0.86, which we recognize is the probability of the union of these two events. So, the first question that is being asked this, what is the probability that the student does not receive an A in either statistics or mathematics?

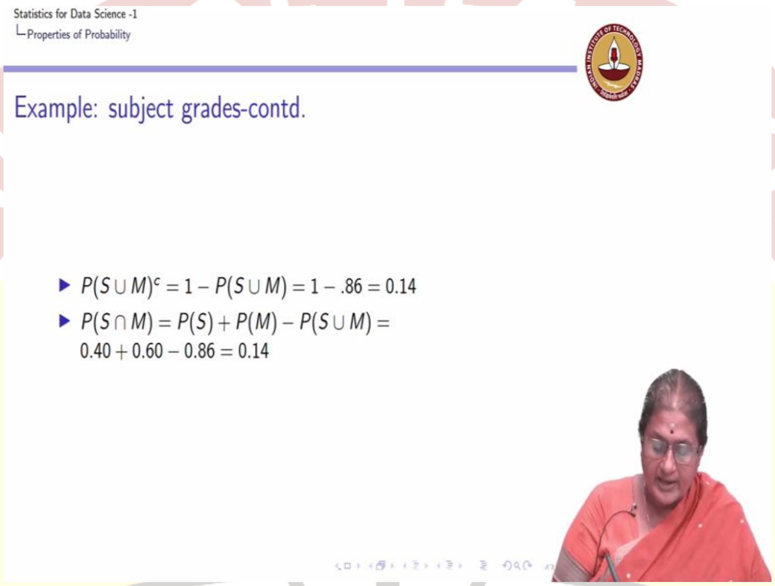
(Refer Slide Time: 49:26)



Again, like we have done in the earlier problem, the event that she does not receive an A is the complement of the event that the student receives A in at least one of the subjects she does not receive in either of the subjects is same as the compliment that she receives in at least one of the subjects, which is again S union M complement. And again, we apply the probability loss of probability to get that.

So, the first thing we understand that this is S union M complement, and the second is receives an A in both statistics and mathematics is we know, it is very simple, it is nothing but the intersection of these two events. So, let us apply our probability properties or axioms to find out the probability. So, I have to find out what is the probability of S union M complement and probability of S intersection M that is what we need. So, this is probability S union M complement this is probability of S intersection M. Now, $P(SUM)^c = 1 - P(S \cup M)$. This is what we know.

(Refer Slide Time: 50:51)



So, we need to understand what is $P(S \cup M)^c$ this I know is equal to $1 - P(S \cup M)$. But what is $P(S \cup M)$ that is given to be 0.86. Hence, I have the probability of S union M complement is nothing but probability 1 minus so that I have an 86% chance of either statistics or mathematics. So, apply whatever I have to find out that probability of S union M complement is 0.14.

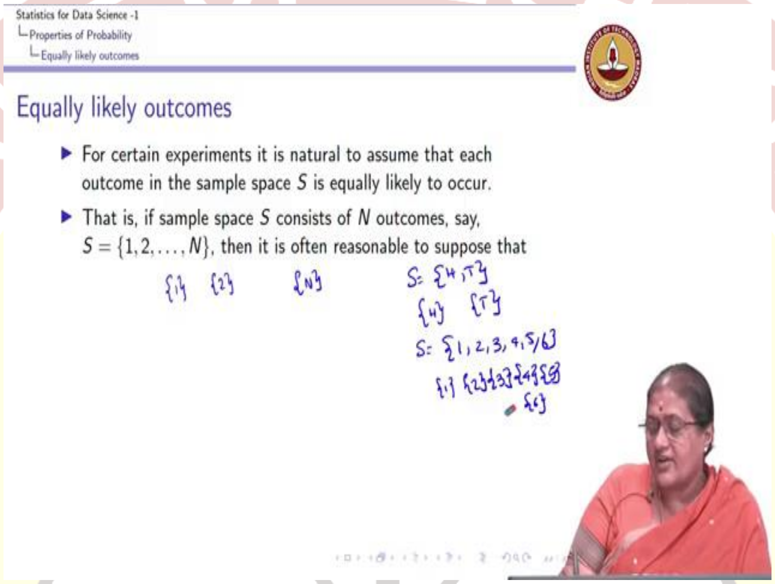
Now, what is the next probability, the next probability is $P(S \cap M)$. Now, again recall S and M are not disjoint. Whatever addition rule says is $P(S \cup M) = P(S) + P(M) - P(S \cap M)$. And this is what my addition rule says, I have this probability I have this probability I have this probability I need to find out what this is.

So, what I do is I can rewrite this addition rule as $P(S \cap M) = P(S) + P(M) - P(S \cup M)$ and I have just rewritten my addition rule. I know probability of S is 0.4, I know probability of M is equal 0.6, I know probability of S union M is 0.86 to give me a 0.14 probability of the candidate receiving A grade in both these subjects together, there is a 0.4 or a 14% chance that the candidate, there is a 0.4 or a 14% chance that the candidate will get A grade in both the subjects.

So, these were examples of how you apply the addition rule. What we need to understand is, we need to first identify the events given the scenario and then apply the addition rule.

Statistics for Data Science 1
Professor Usha Mohan
Department of Management Studies
Indian Institute of Technology, Madras
Lecture 6.6
Probability - Equally likely outcomes

(Refer Slide Time: 00:15)

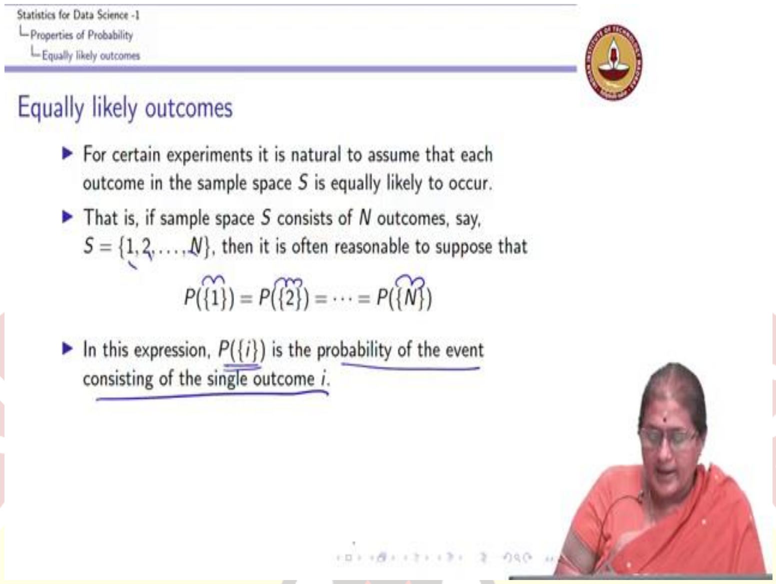


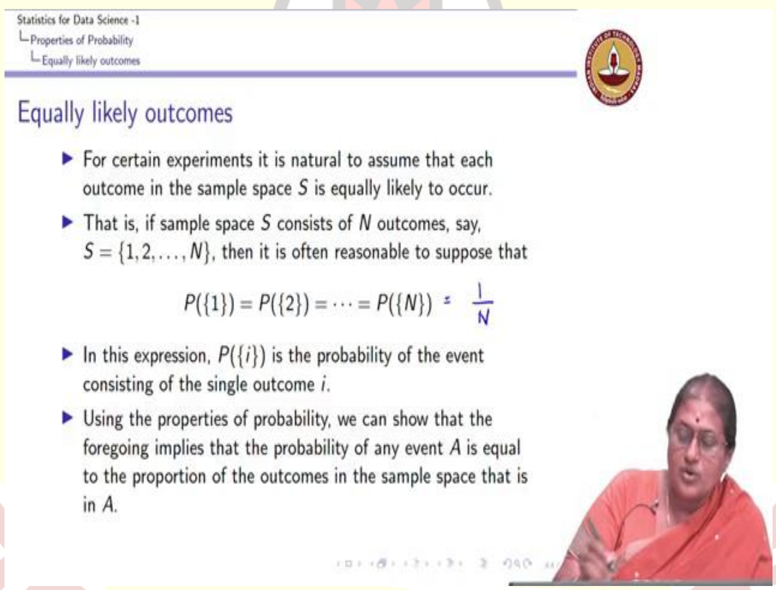
So, now we will focus on equally likely outcomes. Recall when we discussed about the classical interpretation of probability we said that we are assuming in an experiment the outcomes are equally likely. So how do we compute the probabilities of events when I have equally likely outcomes using the properties of probability.

So, suppose I have a random experiment and I am assuming that each outcome in the sample spaces is equally likely to occur then suppose my sample space has n outcomes, finitely many n outcomes and I am assuming that each one of these outcomes are equally likely to occur. What are the outcomes? I have 1 as an outcome, 2 as an outcome, n as an outcome, each one of them is an outcome.

For example, if I am tossing a coin I have head and tail; head is an outcome, tail is an outcome. When I am rolling a dice 1, 2, 3, 4, 5, 6; now 1, 2, 3, 4, 5 and 6 are the outcomes and we are assuming in each of these cases that each of the outcomes are equally likely to happen.

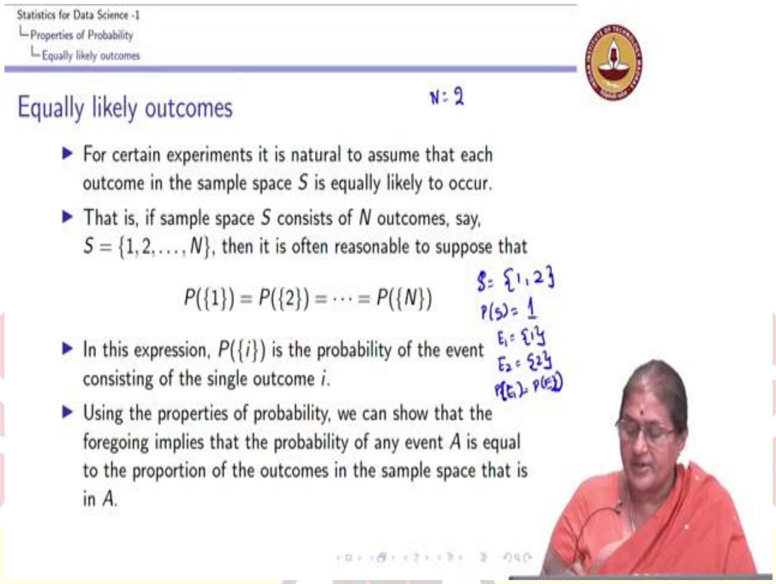
(Refer Slide Time: 01:56)

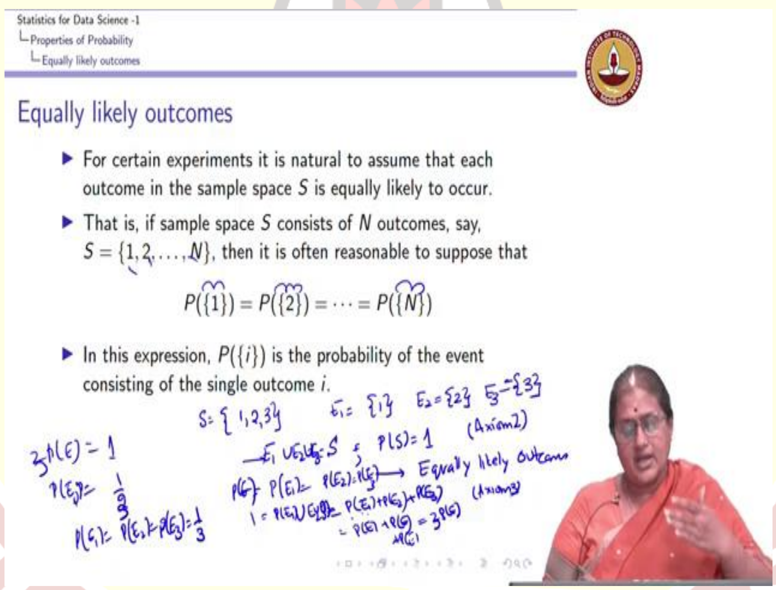




In other words what we are trying to say is a probability with which this outcome happens is the probability with which this outcomes happens with the probability with which this outcome happens.

(Refer Slide Time: 02:35)





There are n outcomes. so let us find out what would be the probability of the event consisting of a single outcome.

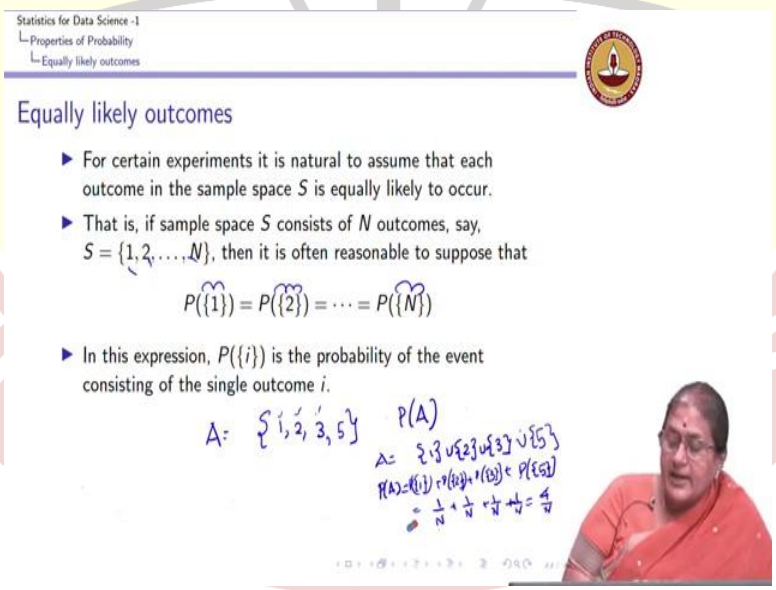
So, let us look at a simple example where I have s equal to 1 and 2 just n equal to 2, I have two outcomes. Now I am assuming let me define E1 to be the outcome 1, E2 to be the outcome 2. Now, I know $E1 \cup E2$ is my sample space and I also know my probability of my sample space is equal to 1, this is from my axiom 2, this is what I have from my axiom 2.

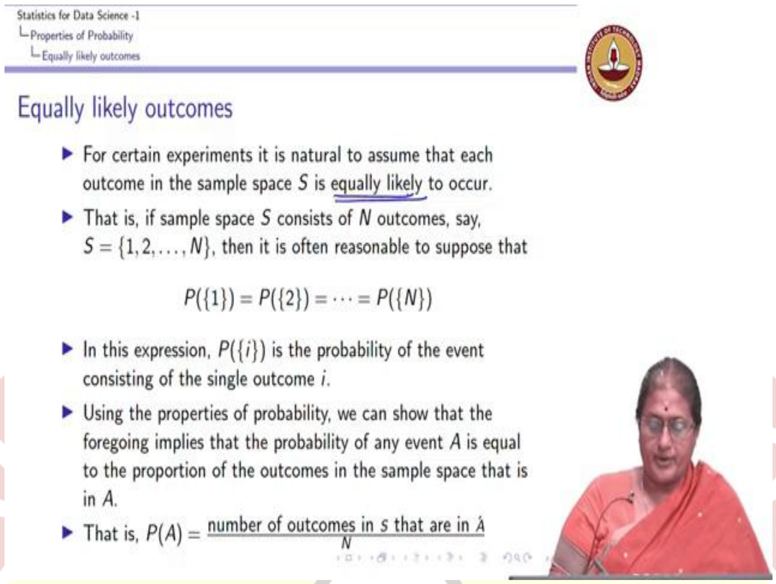
Now I also know $P(E1) = P(E2)$ this is from my assumption of equally likely outcomes. Now from my axiom 3 I know $P(E1 \cup E2) = P(E1) + P(E2)$. Now this $P(E1 \cup E2) = 1$ and I know probability of E1 plus probability of E2 and E1 equal to E2 let me call it some probability of E so this equal to probability of E plus probability of E which is equal to 2 times probability of E.

So, I have $P(E) = 1$ which gives me $P(E) = \frac{1}{2}$. So $P(E1) = P(E2) = \frac{1}{2}$.. I can extend this logic to 1 2 3 I have E1 equal to E2 equal to E3 equal to 3, so $(E1 \cup E2 \cup E3) = S$, so I have now this is equal to probability of E I extend the logic I have a E3 here, so this would be probability of E3, so I have probability of E plus probability of E plus probability of E which is 3 times probability of E.

I get $P(E) = \frac{1}{3}$, so I get $P(E1) = P(E2) = P(E3) = \frac{1}{3}$. So we can extend the same logic to a sample space having n outcomes and we can show that the probability of each of these outcomes is equal to 1 by N, the probability of each of these outcomes if I have n outcomes we can show is equal to 1 by n.

(Refer Slide Time: 06:33)



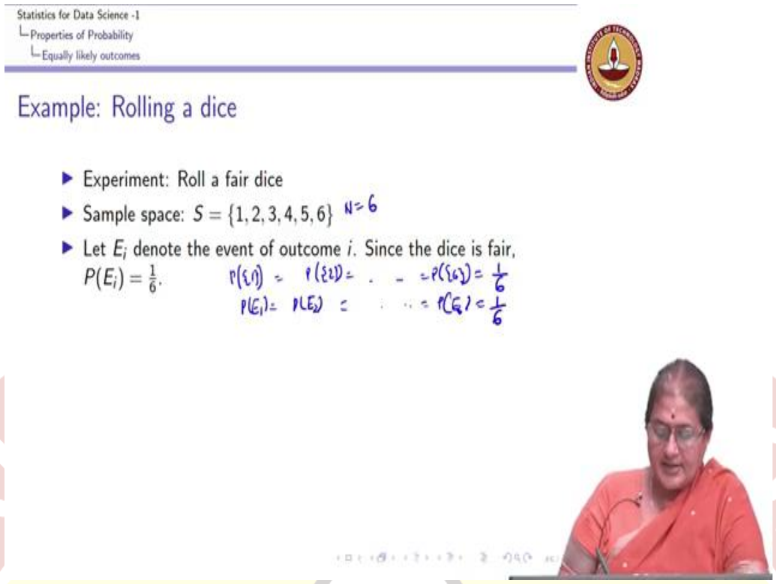


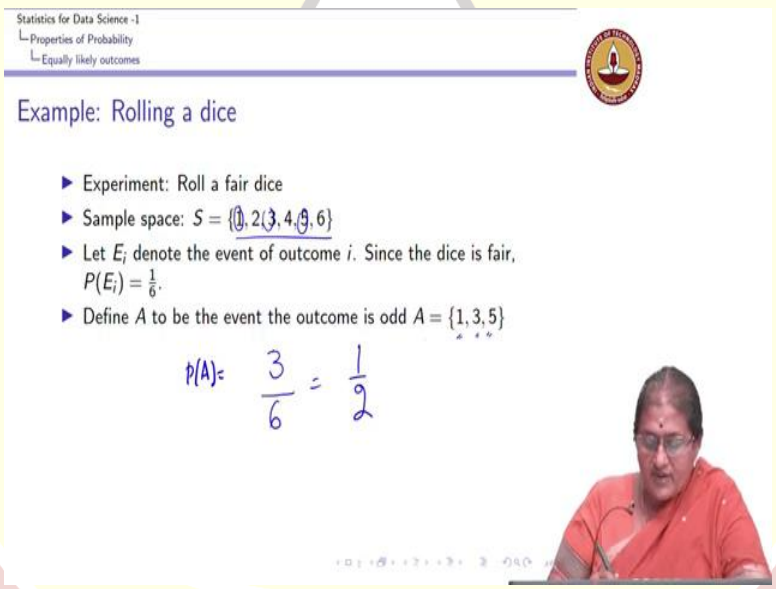
So, now the question is suppose I have an event A which is a subset of this, for example if I have an event A which is 1, 2 and 3. Now what is the probability of this A? I know outcome 1, outcome 2 and outcome 3 are equally likely to happen. Now I can express this A as 1 union 2 union 3 I can express A as 1 union 2 union 3 and by applying my earlier logic of mutually disjoint events I can get probability of A is probability of 1 plus probability of 2 plus probability of 3, plus probability of 3 and we have just worked out these probabilities to 1 by n plus 1 by n plus 1 by n which is equal to 3 by n.

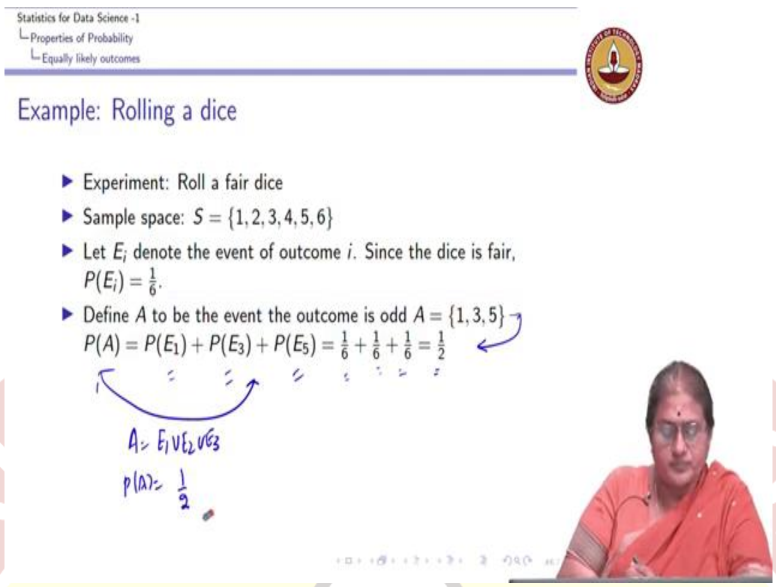
Suppose I have a 1 2 3 5 I will have a union 5 here, I will have a plus probability 5 here I will have a plus probability 5 here and I will have a 1 plus 1 by n here which will give me 4 by n. So if I have equally likely events I can find out the probability of any event by applying the loss of probability and what we can see is I can use the probabilities of probability to show that the probability of any event A is equal to the number of outcomes from S that are in A to the total number of outcomes which I give by n.

This actually is what we refer to in the classical interpretation of probability and hence if you recall I said the assumption of equally likely events or equally likely outcomes is extremely important.

(Refer Slide Time: 09:15)


Statistics for Data Science -1
Properties of Probability
Equally likely outcomes
Example: Rolling a dice
Experiment: Roll a fair dice
Sample space: $S = \{1, 2, 3, 4, 5, 6\}$ $N=6$
Let E_i denote the event of outcome i. Since the dice is fair, $P(E_i) = \frac{1}{6}$.
$P(\{1\}) = P(\{2\}) = \ldots = P(\{6\}) = \frac{1}{6}$
$P(E_1) = P(E_2) = \ldots = P(E_6) = \frac{1}{6}$


Statistics for Data Science -1
Properties of Probability
Equally likely outcomes
Example: Rolling a dice
Experiment: Roll a fair dice
Sample space: $S = \{1, 2, 3, 4, 5, 6\}$
Let E_i denote the event of outcome i. Since the dice is fair, $P(E_i) = \frac{1}{6}$.
Define A to be the event the outcome is odd $A = \{1, 3, 5\}$
$P(A) = \frac{3}{6} = \frac{1}{2}$



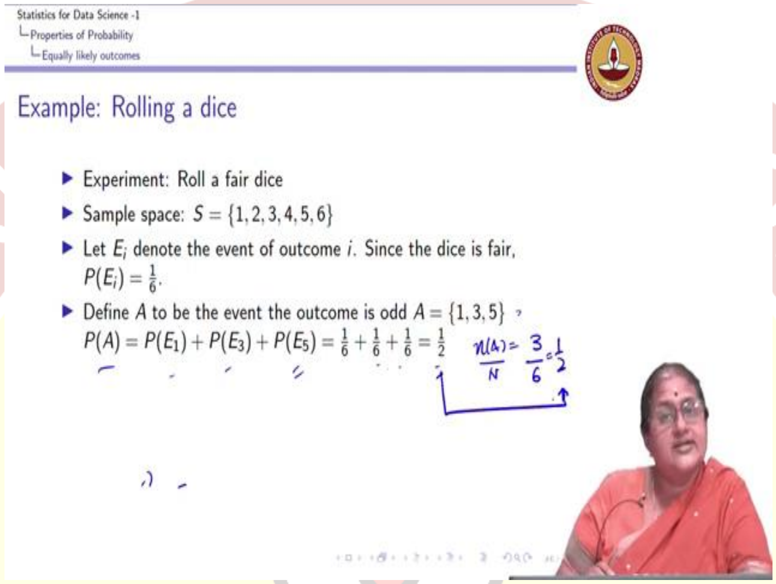
Now, let us look at a few examples where my outcomes are equally likely in a day to day life. So the first is I roll a fair dice, my sample space is 1 2 3 4 5, the minute I say fair I assume it is unbiased, it is not loaded or it is not weighted dice.

The chance I get any one of these outcomes is equal to same, so any one of these outcomes are equally likely to happen. So I have an experiment wherein the outcomes are equally likely to happen I can show that from our earlier, so my n equal to 6 in this case, so I can show that from my earlier discussion probability of 1 occurring equal to probability of 2 occurring which is equal to the probability of 6 occurring and all of them are same and equal to 1 by 6.

So, if Ei is the probability of an outcome i probability of E1 equal to probability of E2 which is equal to the probability of E6 which is equal to 1 by 6. So that is what we can see. Now suppose I define the event that the outcome is odd, so I know that the outcomes are 1, 3 and 5. Now how do I find out the probability of this event A? From my expression I find out how many so if you go back by my definition if I have equally numb likely outcomes probability of A is number of outcomes from S that are in A to the total number of outcomes.

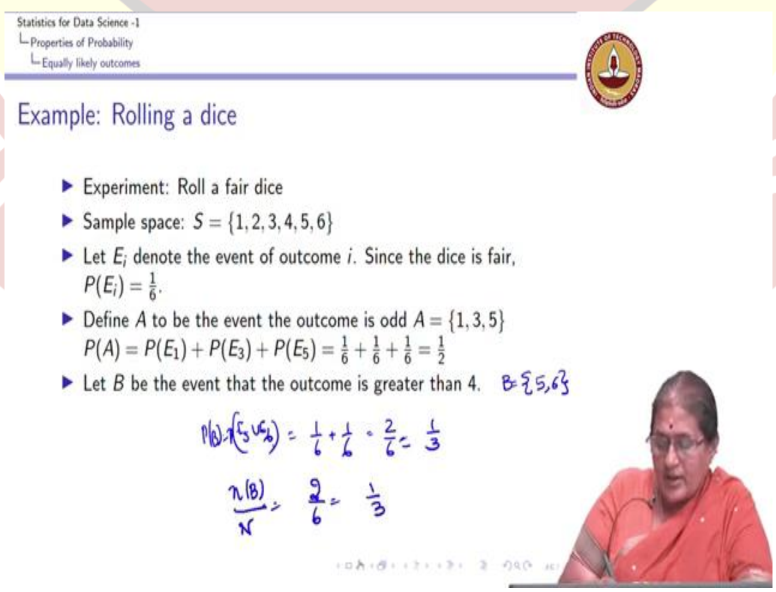
I just apply that logic, so the number of outcomes from S, the number of outcomes in S that are in A are 1, 3 and 5 which is equal to 3. And the total number of outcomes is 6 giving me a probability of A is 1 by 2 which is also very evident that when I toss a die the chance that I get an odd number is 50%, it makes a lot of sense we have just got it from the properties which we have just defined which is probability of A, I can get it from the definition and I can verify that this is the same as probability of E1 plus probability of E3 plus probability of E5 which is 1 by 6, 1 by 6, 1 by 6 which is 1 by 2. This is from my axiom I have expressed A to be $E1 \cup E2 \cup E3$ and I work out probability of A is $\frac{1}{2}$.

(Refer Slide Time: 12:13)



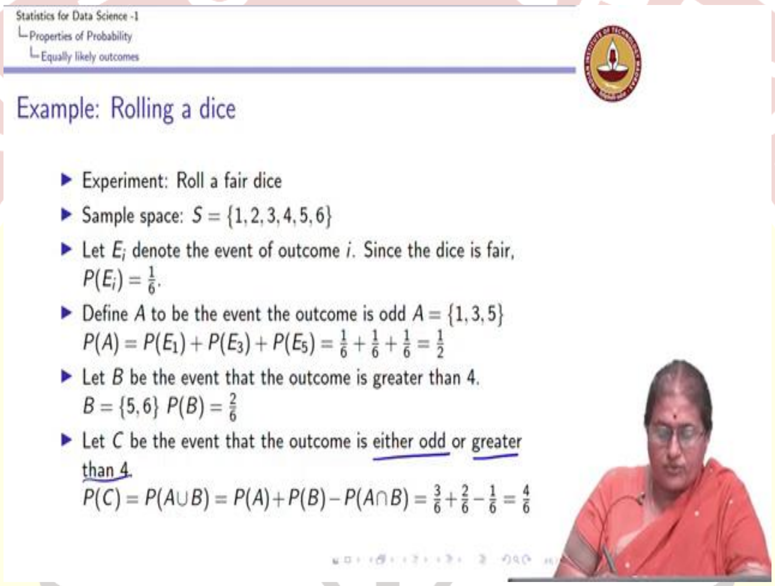
And this we have verified is the same as number of elements of A which is 3 to total number of elements which is 6 which is 1 by 2 and we can see that these two are the same.

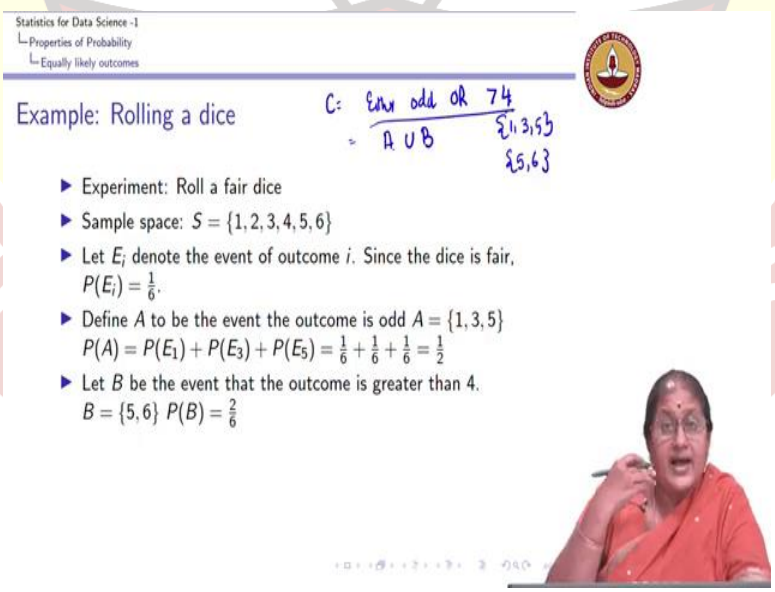
(Refer Slide Time: 12:32)



So, this, let us look at another event which says that the outcome is greater than 4. So I know the elements in B are greater than 4, so this is 5 and 6. Again I can express this as E5 union E6 probability of B equal to $P(E5 \cup E6)$where E5 and E6 are again the outcome is 5 and 6 so I get a 1 by 6 plus 1 by 6 which is a probability so I get this is 2 by 6 which is 1 by 3 I can also find out that as number of elements in B to total number which is 2 by 6 which is again 1 by 3.

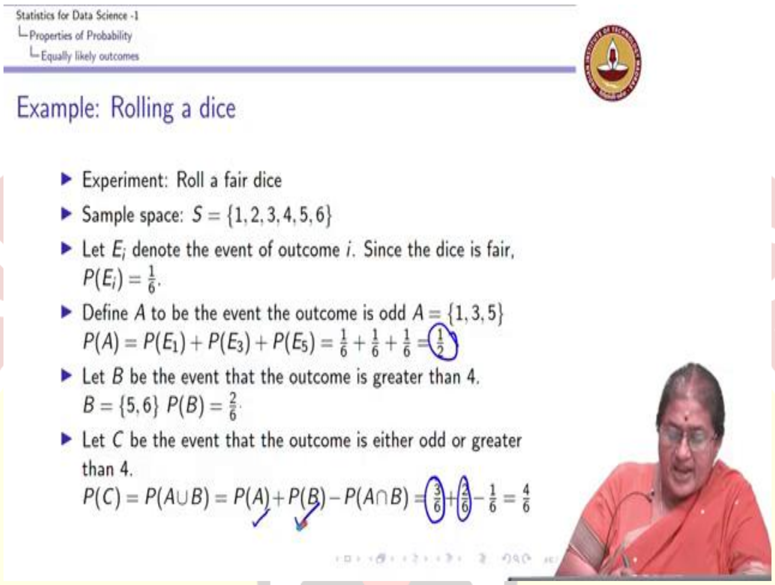
(Refer Slide Time: 13:20)

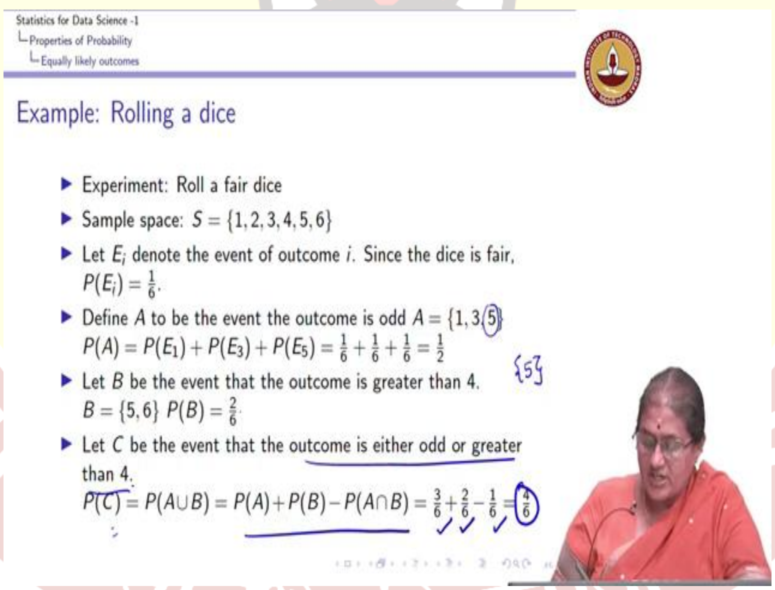




So, now let us look at the third event which says that it is either odd or greater than 4. So now we have defined an event C which is either a odd number or it is greater than 4. So I know C can be expressed as A union B. A was the outcome 1, 3, 5. B was the outcome 5 and 6. So you can see that A union B are not disjoints, so either odd and greater than 4, A union B are not mutually exclusive.

(Refer Slide Time: 14:05)





I apply the additive rule of probability to get probability of C is probability of A union B which is same as probability of A which I have obtained here which is 1 by 2 or 3 by 6. Probability of B which is 2 by 6, what is $P(A \cap B)$? So what do I mean by A intersection B? I mean that the event is both it is just this 5 which is greater than 4 and odd which is just the event 5 and that probability is 1 by 6 because $P(E5) = \frac{1}{6}$ and I see that the probability of C which is either odd or greater than 4 by application of the additive rule is $\frac{3}{6}+\frac{2}{6}-\frac{1}{6}$ which is equal to 4 by 6 which is equal to further equal to 2 by 3.

(Refer Slide Time: 15:04)

Now, let us look at another example which we have been looking at right from the beginning which is about taking a card from a deck of 52 cards. So again when you draw a card from a standard deck of 52 cards, the question is what is the probability that the card is either a red or a queen? Now, first when I am having a card or a deck of 52 cards and I am drawing a card from these 52 cards the chance that I draw any one of these 52 cards is the same.

So, if I define my sample space as drawing a card so I know my sample space is going to be these 52 outcomes and each one of these outcome is equally likely to happen. So the chance that I am going to draw an A club is the same as the chance I am going to draw 9 hearts which is the same as the chance I am going to get a 10 clubs which is the same as I could get a queen diamond and so forth.

So, each one of these outcomes are equally likely to happen. With what probability; I have 52 cards, so each one of them can happen with the probability of 1 by 52. So now the question is what is the chance that the card is either a red or a queen? So I need to define the events.

(Refer Slide Time: 16:42)

So, the first event I define is the event that the card drawn is a red card. So I can see that I have 26 of these cards which are red cards. So the number of cards which are red cards are 26, the total number of cards are 52 which gives me the probability of drawing a red card is 26 by 52 which is half. Another way to look at it is I have exactly 26 of cards of black and 26 are red and so hence the chance I draw a red card is going to be 50% which is given by $\frac{1}{2}$.

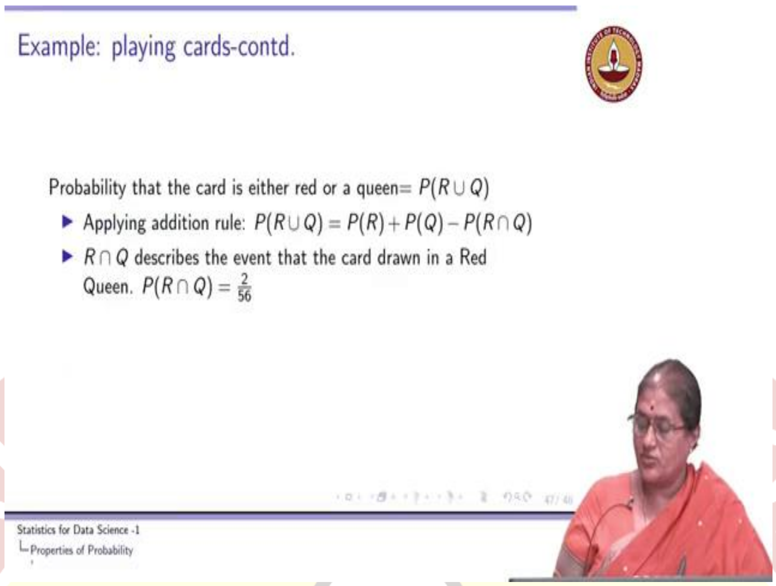
(Refer Slide Time: 17:28)

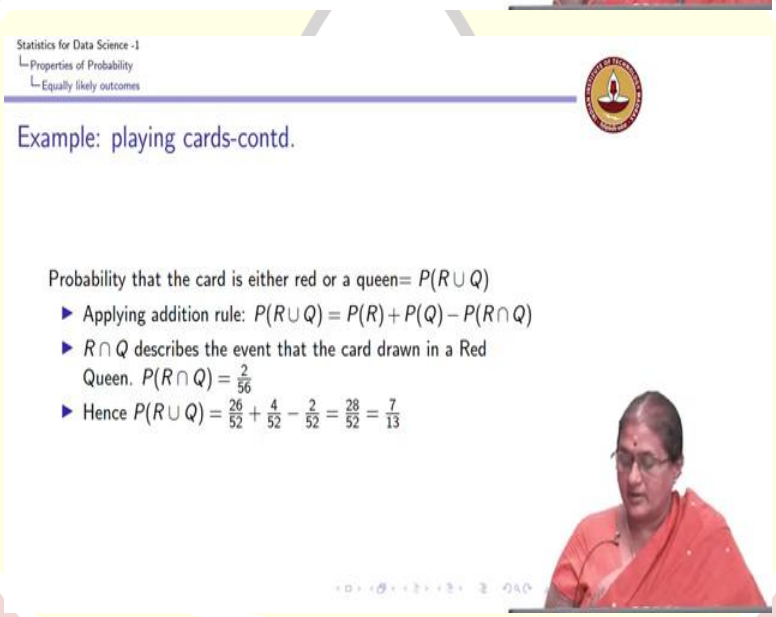
Now, what is the chance of me drawing a queen? So I can see that I have 4 queens and drawing each queen so the chance of drawing a queen is I either get this card or this card or this card or a this card and I can again apply my laws of probability to know that the probability of drawing a queen is nothing but $\frac{4}{52}$ again applying the equally likely outcomes and number of outcomes that satisfy this event that it is a queen which is 4 I get which is $\frac{1}{13}$.

But that is not what we want, we want either a red or a queen. So I can express this as $R \cup Q$, now are they disjoint events? In other words, can I have a card which is both red and a queen so let us look at this, the answer is yes. I have a card, I have in fact two cards which are both red and queen hence red and queen are not mutually exclusive or not disjoint.

So, how do we find out what is probability of red union Q? The way I find out the probability of red union Q is I find out probability of red plus probability of Q minus probability of red intersection Q I have which is $\frac{26}{52} + \frac{4}{52} - \frac{2}{52}$ which gives me $\frac{28}{52}$ which I can further see that this is $\frac{7}{13}$. So hence I have the probability of red or a queen to be $\frac{7}{13}$. So what have we done so far?

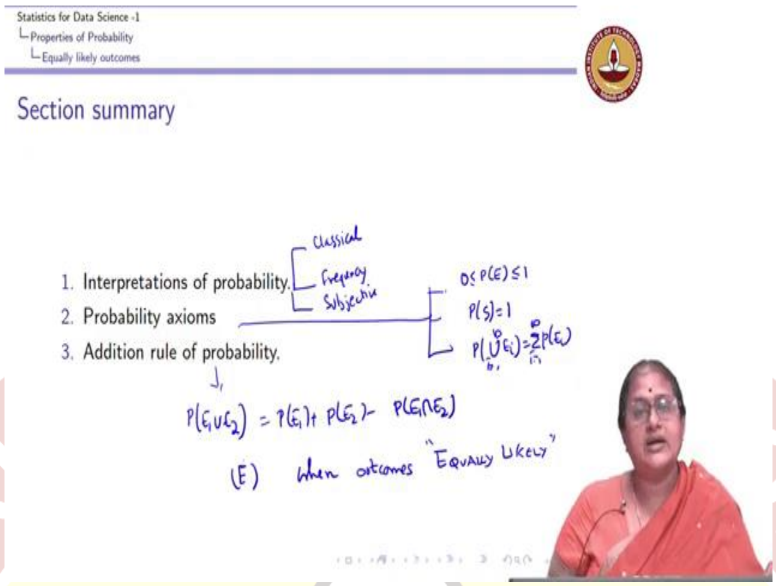
(Refer Slide Time: 19:45)





So, I apply the addition rule and after applying the addition rule I get that $P(R \cup Q) = \frac{7}{13}$.

(Refer Slide Time: 19:58)



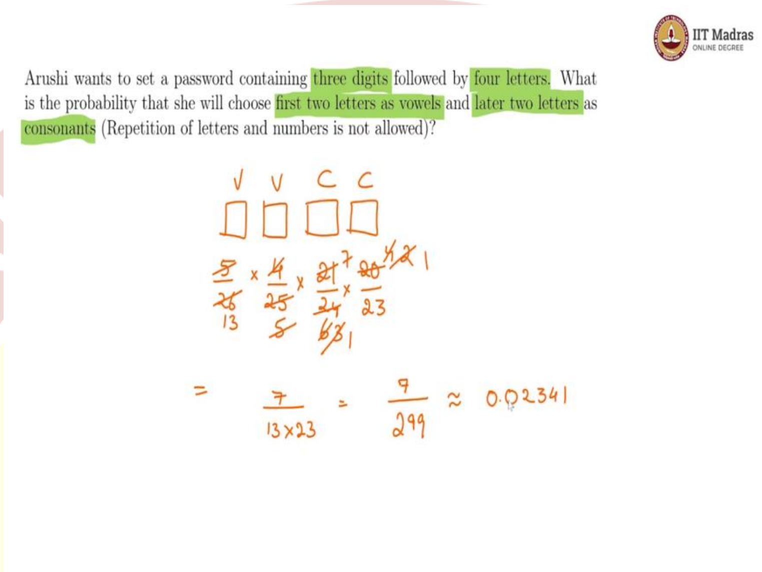
So, quick summary of what we have done in this session; we first introduce a 3 interpretations of probability, we started with a classical, so here we said the three popular interpretations are classical. The relative frequency or the empirical and the subjective interpretations of probability then we introduce the probability axioms by probability axioms we have three main axioms; one is given an event from a sample space probability of the event would be a number between 0 and 1.

Probability of the sample space equal to 1 and probability of the union of a sequence of events is equal to the sum of the probabilities of the events these are the three axioms. Then we extended this axiom and or derived the probability of a union of any two events not necessarily disjoint to be probability of E1 plus probability of E2 minus probability of E1 intersection E2 and looked at applications of this addition rule of probability.

We finally ended by how to compute the probability of an event when the outcomes of an experiment are equally likely when the outcomes of an experiment are equally likely to happen. So the next thing which we are going to address is the notion of conditional probability.

Statistics for Data Science – 1
Professor Usha Mohan
Department of Management Studies
Indian Institute of Technology, Madras
Week 6 - Tutorial 1

(Refer Slide Time: 0:19)



Hello statistics students. In this week's tutorials we look at problems related to probability. In this problem they are saying Arushi wants to set a password containing three digits followed by four letters and then they are asking what is the probability that she will choose the first two letters as vowels and the later two letters as consonants. Repetition of letters and numbers is not allowed.

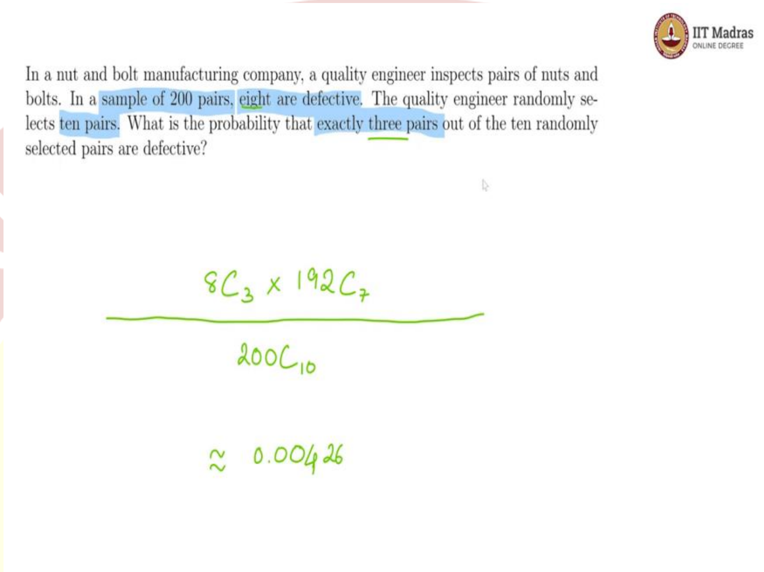
So, the first thing we realize here is there is no condition which is set on the digits, so the digits are as good as not being there. So, we are only interested in the four letters. So, we will look at the probability regarding these four letters and here you want a vowel, this is also a vowel, this is a consonant, this is a consonant and there is no repetition.

So, let us see for the first box to get a vowel the probability would be —. Now, in the next box this vowel, the one in the first box should not come up, so we get — because one letter is taken away and then we have 24 letters left but these are consonants now, so you have 21 consonants so — here and lastly this would be 20 consonants remaining divided by 23.

So, this calculation is what we are supposed to do, if we can cancel out some things let us try 4×6 is 24 and 5×5 is 25 so 5×4 is 20 again, so two 2 two 3 two 1 and there is 13 so you will now get, actually we can further cancel three 1 and this is 7. So, you now get $\frac{7}{13 \times 23}$which is equal to $\frac{7}{299}$ which is roughly equal to 0.02341 so this is the probability.

Statistics for Data Science – 1
Professor Usha Mohan
Department of Management Studies
Indian Institute of Technology, Madras
Week 6 - Tutorial 2

(Refer Slide Time: 0:16)



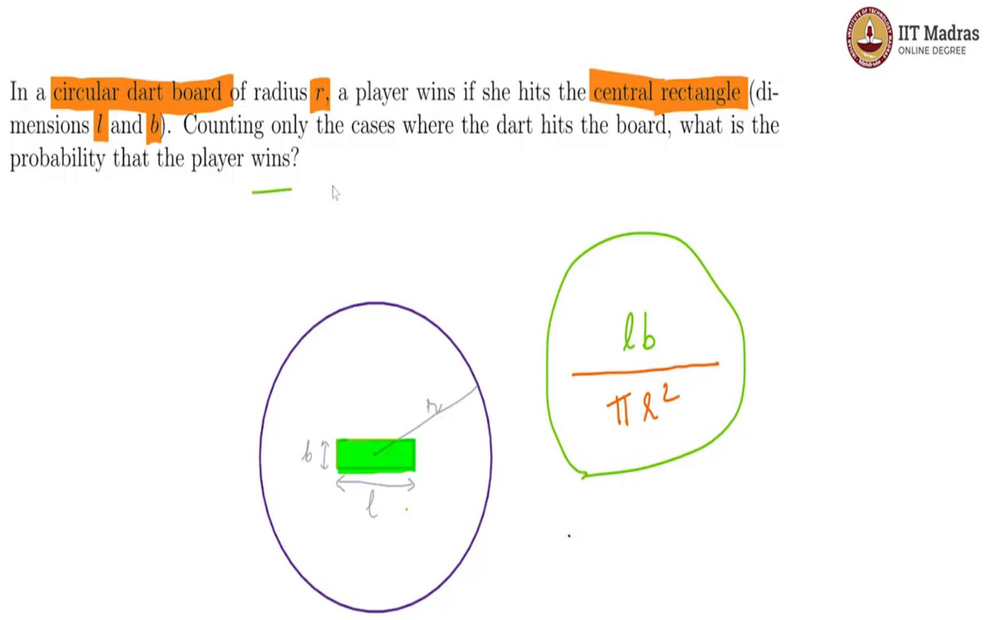
In a nut and bolt manufacturing company, a quality engineer inspects pairs of nuts and bolts. In a sample of 200 pairs, 8 are defective, so out of the sample is of 200 and 8 are defective. And the quality engineer randomly selects 10 pairs, so of these 200, the quality engineer is selecting 10 pairs. What is the probability that exactly three pairs out of the 10 are defective?

So, if we look at this as number of possibilities where our condition is being satisfied divided by the total number of possibilities. We can look at it this way the denominator which is the total number of possibilities will be ${}^{200}C_{10}$. So, you are picking 10 pairs out of 200 randomly and the number of ways you can do it is ${}^{200}C_{10}$ these are the total number of possibilities.

And now for these exactly three pairs to be defective those three should come from these 8, so that would be ${}^{8}C_{3}$ multiplied by the remaining 7 should come from the remaining 192 pairs. So, that would be ${}^{192}C_{7}$. So, this should be our answer, if you calculate it, it is going to come to about 0.00426. So, that is quite small, anyway this is the probability.

Statistics for Data Science – 1
Professor Usha Mohan
Department of Management Studies
Indian Institute of Technology, Madras
Week 6 - Tutorial 3

(Refer Slide Time: 0:16)

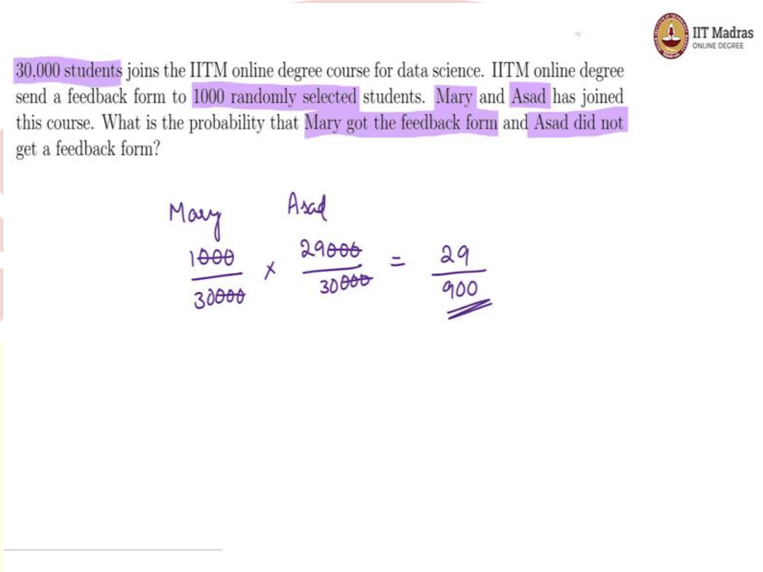


In this question there is a circular dart board radius is r, and a player wins if she hits the central rectangle, so there is a rectangle at the center and the dimensions are l and b. Counting only the cases where the dart hits the board, what is the probability that the player wins? So, a dart board is something like this and the radius is given to be r and then they are saying there is a central rectangle which if the dart hits the player wins.

And this rectangle's dimensions, this is b, the breath is b and the length is l. So, the sample space here is the total circle. So, our denominator will be the area of the total circle which is πr^2 and the specific condition we are looking for is hitting the center which is this area, just the rectangle's area and that comes out to be lb, so this should be the probability of winning.

Statistics for Data Science – 1
Professor Usha Mohan
Department of Management Studies
Indian Institute of Technology, Madras
Week 6 - Tutorial 4

(Refer Slide Time: 0:15)

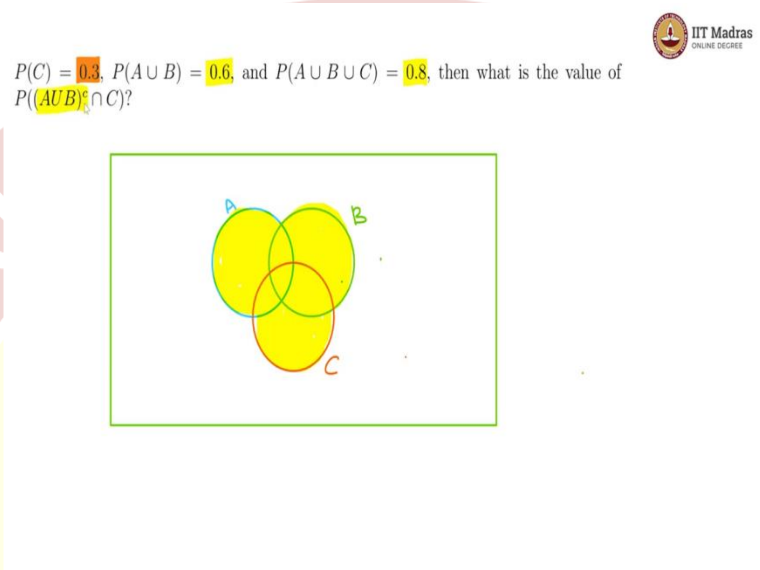


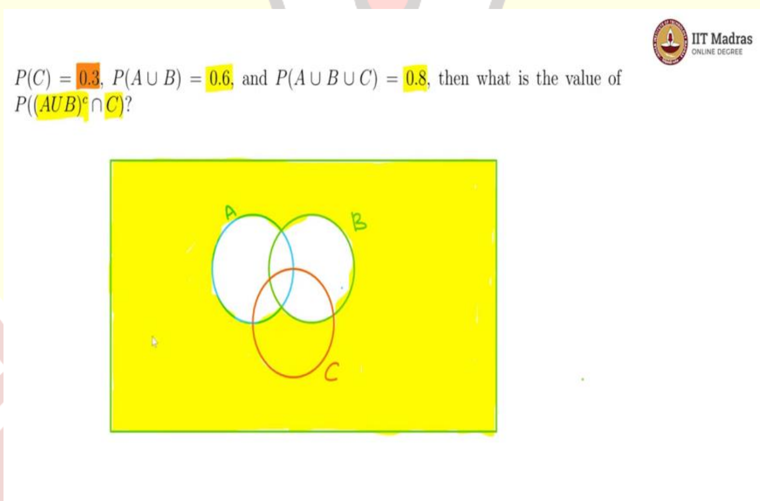
In this question we have 30000 students who have joined the IITM online degree course in Data Science and the degree program sent a feedback form to 1000 randomly selected students. Now, Mary and Asad have joined this program. What is the probability that Mary got the feedback form and Asad did not?

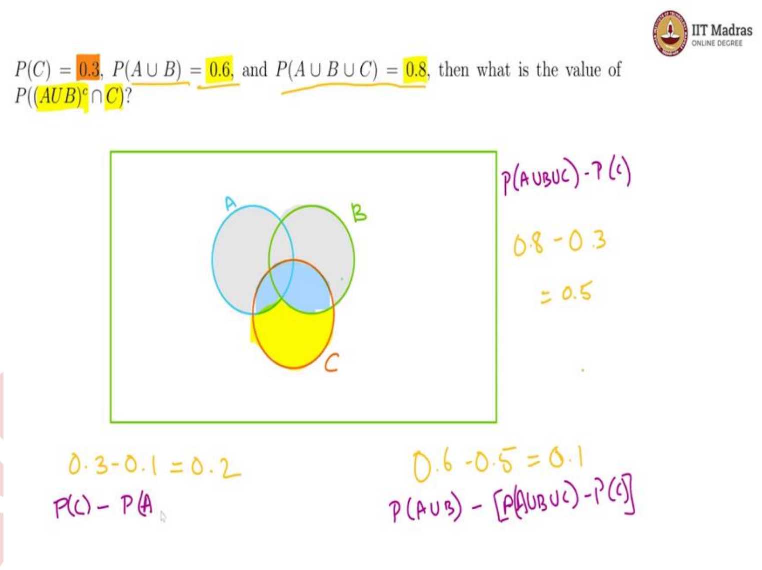
So, here we have the two students Mary and Asad. Probability that Mary got the feedback from would be if she is in the 1000, so that would be $\frac{1000}{}$ and the probability that Asad did not get the feedback form would be the remaining, he Asad should be in the remaining ____. And for both of these to happen together we multiply the probabilities. So, these two cancel off, these two cancel off, so we get $\frac{29}{}$, this is the probability.

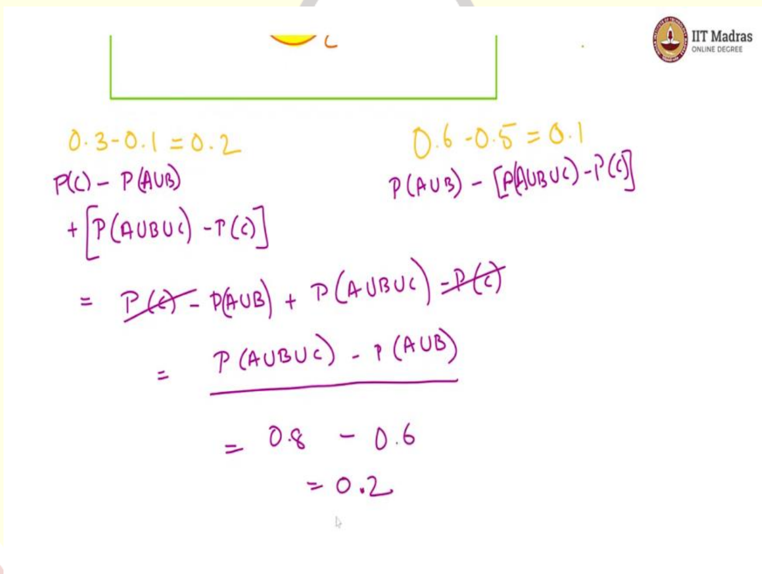
Statistics for Data Science – 1
Professor Usha Mohan
Department of Management Studies
Indian Institute of Technology, Madras
Week 6 - Tutorial 5

(Refer Slide Time: 0:16)









In this problem we are looking at some sets, the probability of set C is 0.3, so for this let us try to draw the Venn diagram. So, let this be the universal set, so this would be the sample space and within this we have three sets C and A and B, let this be set A and this is set B and this is set C. So, this 0.3 indicates that this area which is the set C's area is 0.3 of the total area that is 30% of the total area.

And then this 0.6 is the union of A and B which is roughly this, so that is 60% of the total sample space and lastly they are saying that 0.8 is for A∪B∪C, so the union of all of this is 0.8. So, now they are asking what is the value of $(A\cup B)^C \cap C$, so $(A\cup B)^C$ is everything outside of A∪B which is all of this.

And you want the intersection with C, so we only look at the part where C is also involved which is essentially this portion. So, we want this one's area we know that totally C is 0.3 and we know that totally $A \cup B \cup C$ is 0.8. So, if we remove C from $A \cup B \cup C$, we will be looking at this portion which is then $0.8-0.3$. So, this gray portion is 0.5.

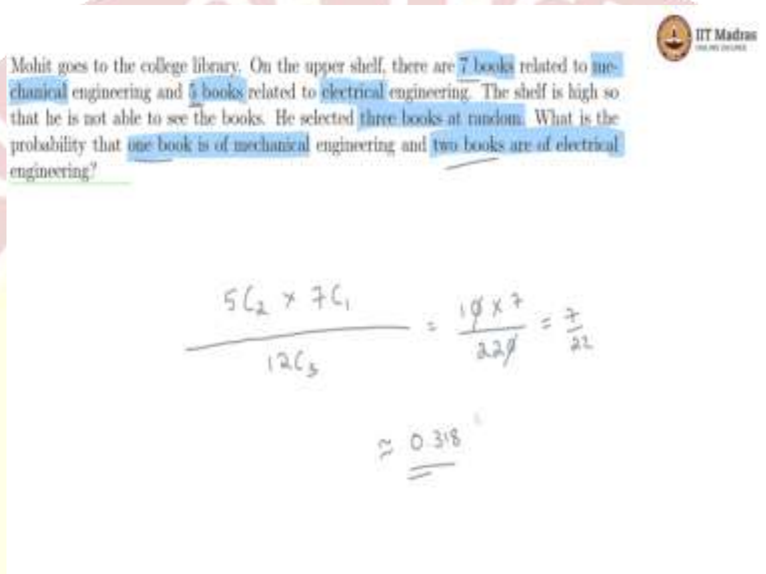
And further we know that $A \cup B$ is 0.6 so this blue region here that is going to be $A \cup B$ minus the grey region which is $0.6-0.5$ is equal to 0.1 and we know that this yellow region is essentially all of C minus the blue region so that is $0.3-0.1$ which is equal to 0.2. So, here we are saying $P(A \cup B \cup C)-P(C)$ is 0.5 and here we are saying $P(A \cup B)$ minus this 0.5 which is $P(A \cup B \cup C)-P(C)$.

And lastly here we are saying it is P(C) minus whatever this whole thing is. So, if we applied that we will get $P(C)-P(A \cup B)+P(A \cup B \cup C)-P(C)$. So, this gives us $P(C)-P(A \cup B)+ P(A \cup B \cup C)-P(C)$, so P(C) is getting cancelled here.

So, we are effectively left with $P(A \cup B \cup C)-P(A \cup B)$ which also makes full sense because this yellow region is essentially the union of the 3 minus the union of A and B. So, yeah if we substituted that directly here also we would get $P(A \cup B \cup C)$ is 0.8 minus $P(A \cup B)$ is 0.6. So, we get 0.2.

Statistics for Data Science – 1
Professor Usha Mohan
Department of Management Studies
Indian Institute of Technology, Madras
Week 6 - Tutorial 6

(Refer Slide Time: 0:16)

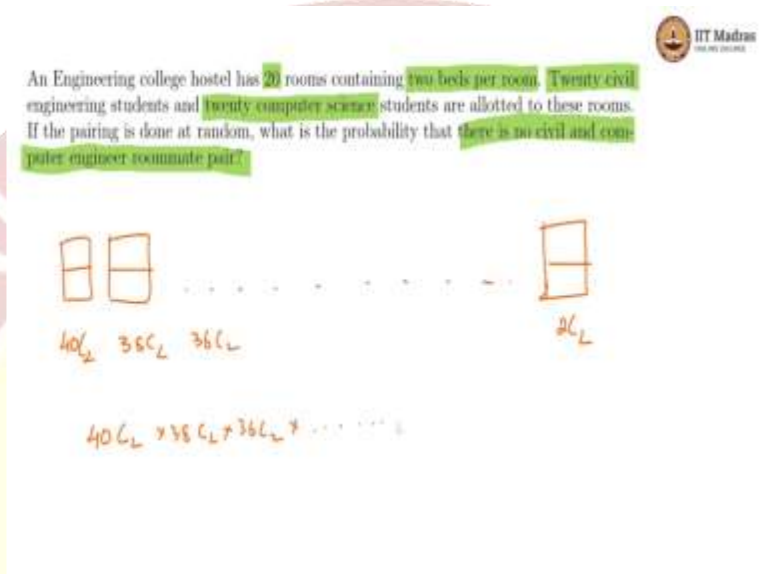


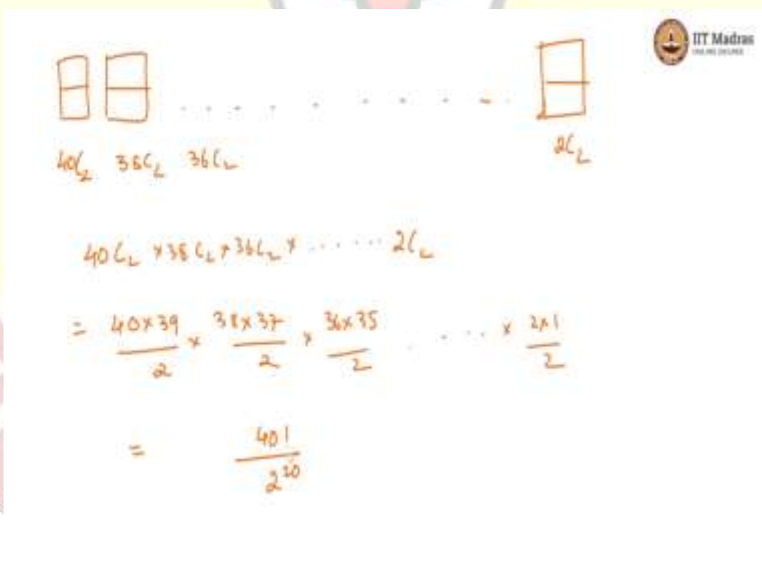
Mohit goes to the college library. On the upper shelf, there are 7 books related to mechanical engineering and five books related to electrical engineering. The shelf is high so he is not able to see the books. He selected three books at random. What is the probability that one is mechanical and the other two are electrical?

So, again we are overall the sample space is number of ways of picking three books out of twelve, so that will be ${}^{20}C_3$ and over here in the numerator we have ${}^{5}C_2$ which is two books from electrical, five books of electrical into ${}^{7}C_1$ which is one book of mechanical out of the 7. So, that gives us —— = — which is roughly 0.318. So, this is the probability.

Statistics for Data Science – 1
Professor Usha Mohan
Department of Management Studies
Indian Institute of Technology, Madras
Week 6 - Tutorial 7

(Refer Slide Time: 0:16)



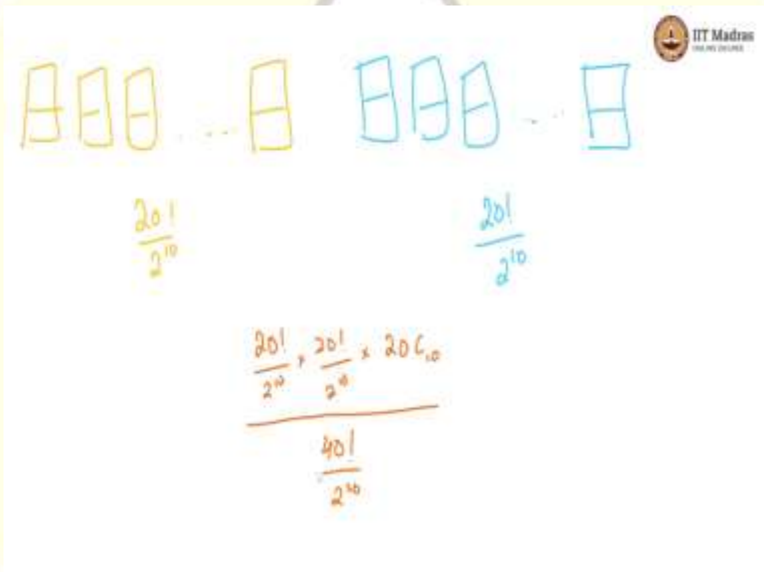


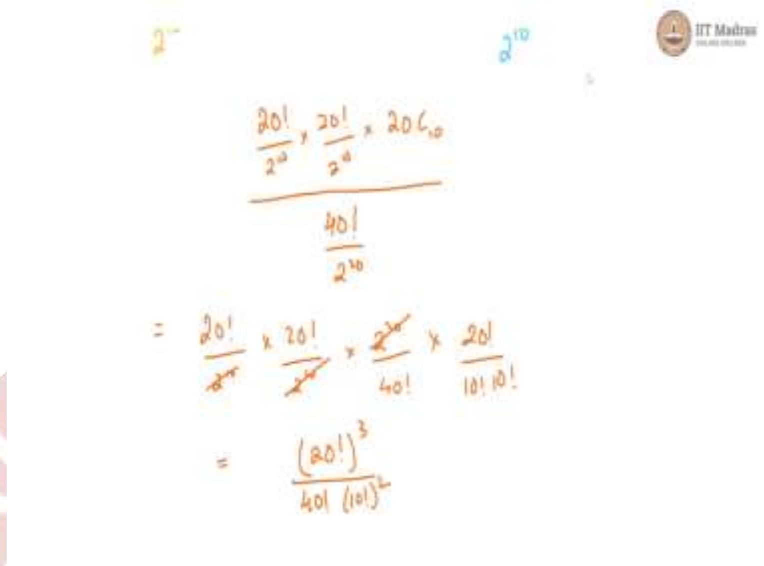
An engineering college hostel has 20 rooms and two beds per room. There are 20 civil engineering students, 20 computer science students who are allotted to these rooms. Pairing is done at random, what is the probability that there is no civil and computer engineer roommate pair?

So, all the civil engineering students are in 10 separate rooms and all the computer science students are in 10 separate rooms. So, for solving this problem let us look at the rooms which are like this, each room has two beds so on and there are 20 rooms all together and there are 40 students, there 20 in computer science and 20 in civil and we are looking at the total possibilities.

So, the first room gets ${}^{40}C_2$ ways of being filled. Then the next one has ${}^{38}C_2$ and the next one will have ${}^{36}C_2$ so on till we are left with ${}^{2}C_2$, so we have ${}^{4}C_2 \times {}^{38}C_2 \times {}^{36}C_2 \times \ldots.. {}^{2}C_2$ which is basically $\frac{40\times39}{2} \times \frac{38\times37}{2} \times \frac{36\times35}{2} \times \ldots \ldots \ldots \times \frac{2\times1}{2}$. So, this will give us $\frac{40!}{2^{20}}$. So, this is the total ways of being able to fill these rooms.

(Refer Slide Time: 2:11)





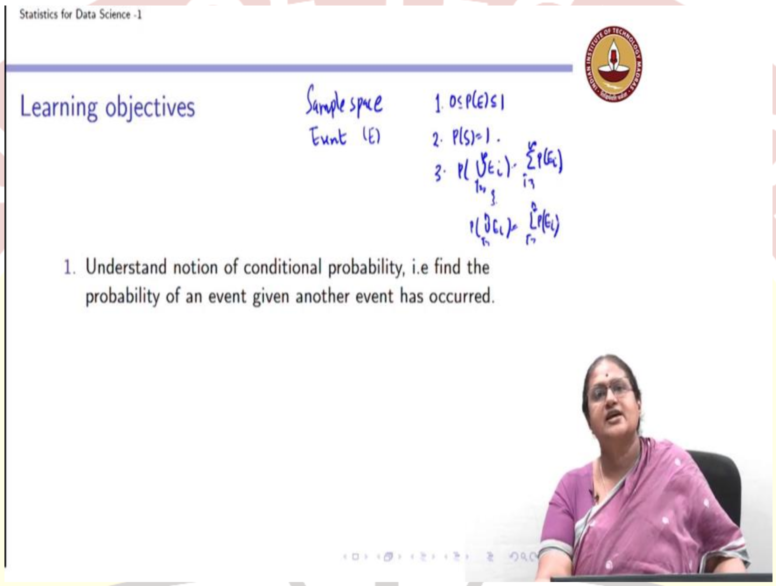
Now, imagine if we filled separately 10 rooms with just the civil engineering students and again separately 10 rooms with just the computer science students. Then we will get a very similar logic for both of these. You will get $\frac{}{10}$ for these and $\frac{}{10}$ for these as well.

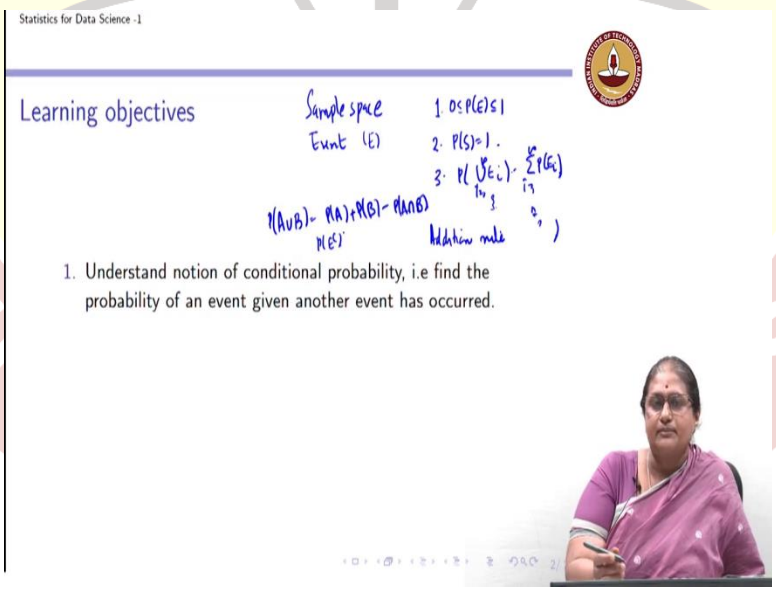
However, how we choose the rooms also should matter. So, we will be getting $\frac{}{10} \times \frac{}{10} \times {}^{20}C_{10}$. After these rooms are chosen, these are the ways that they respectively get filled. And then this divided by the $\frac{}{20}$. This gives us the simplification would be $\frac{}{10} \times \frac{}{10} \times \frac{}{} \times \frac{}{} = \frac{}{2}$.

Thank you.

Statistics for Data Science – 1
Professor. Usha Mohan
Department of Management Studies
Indian Institute of Technology, Madras
Lecture No. 7.1
Conditional Probability – Contingency Tables

(Refer Slide Time: 00:21)





So, what we have learned so far is we set up the axioms of probability, but before we set up the axioms of probability we define what was a sample space, we define what was a event and we if we have a event E, the axioms of probability said that $0 \leq P(E) \leq 1$, probability of a sample space is equal to 1 and probability of a countable union is equal to the sum of the probability when I have disjoint or mutually exclusive events.

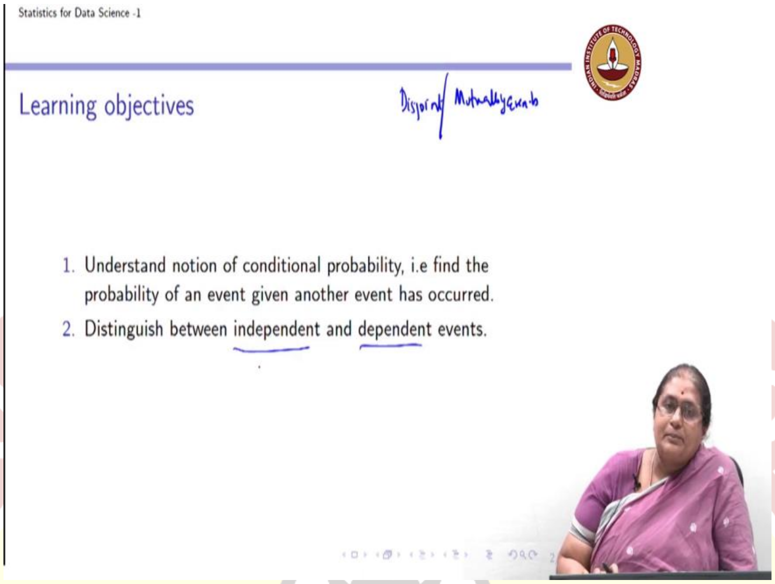
From here, we looked at a finite union and we said that the probability of a union of a disjoint set is the same as the sum of the probabilities. We also looked at what we refer to as the addition rule and we looked at when I do not necessarily have events that are mutually disjoint I come up with a probability of the union and I refer to that rule as the additional rule which says that the probability of a union of any two events is $P(A) + P(B) - P(A \cap B)$.

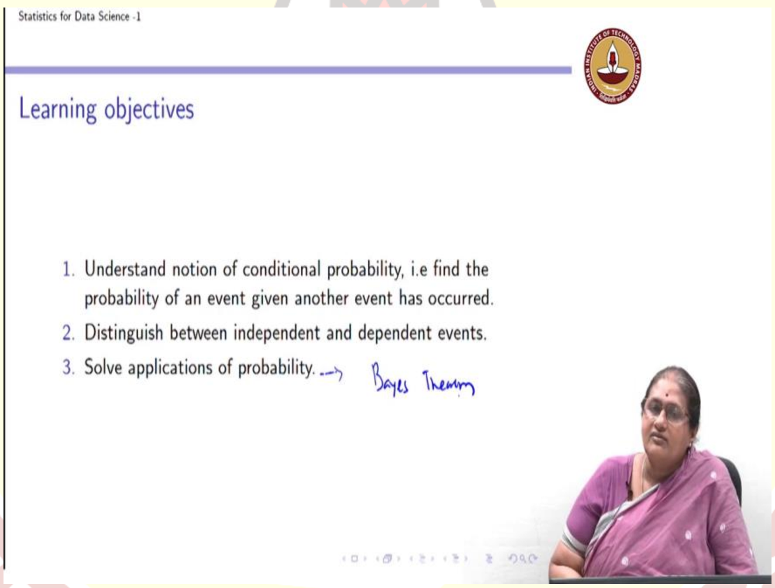
We also define what is the probability of a complement of a event and then afterwards we worked out certain examples of this theory which we develop. So, the next important thing which we are going to understand today is how do I compute probabilities when an event is conditioned on another event. Now, why is this important? Sometime I but have information or partial information of something that has happened.

Now, if I want to know what is the probability of a event happening conditioned on that event, for example I could condition I am tossing a coin twice I know that my first toss is a head, I would want to know that what is the chance of me getting ahead in the second toss conditioned on the fact that the first toss was a head.

So, I am conditioning, so I am using the information I have from the outcome of the first toss. So, this is where we introduce the notion of conditional probability, which in other words is the probability of an event occurring conditioned on or given that another event has occurred. So, we will understand what is conditional probability.

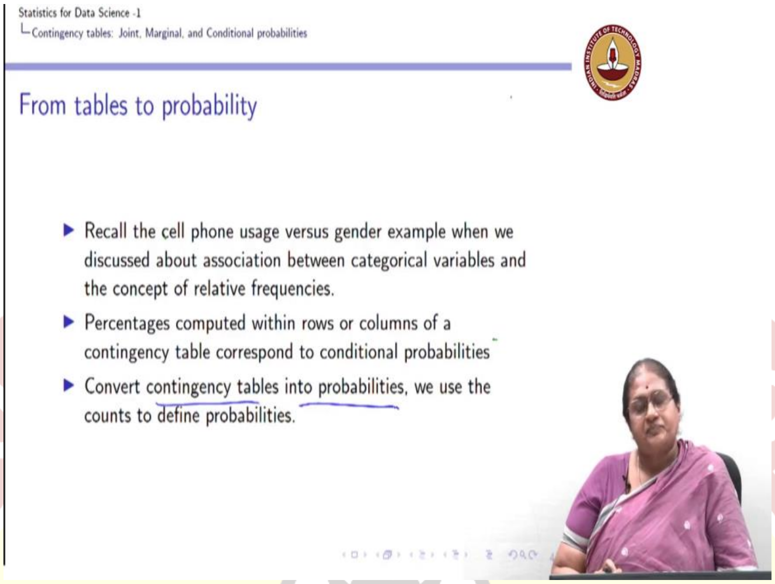
(Refer Slide Time: 03:09)





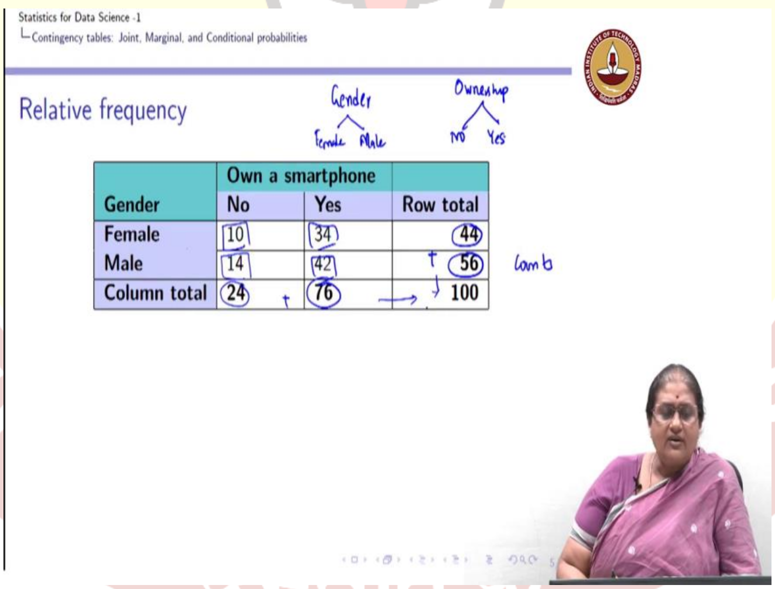
The next important concept is once we understand what is conditional probability we will introduce the notion of what are independent events and hence we will understand what are dependent events. We have already seen what are disjoint or mutually exclusive events, we are going to introduce what are independent and dependent events. And finally, we will apply all the concepts which we have learnt in for we will also introduce what is called Bayes theorem and we will apply certain concepts of whatever ever we have learnt in real-world applications. So, how do we motivate our need for conditional probability?

(Refer Slide Time: 04:01)



Recall that we looked at association between variables in particular when we looked at association between categorical variables we introduced what we refer to as contingency tables.

(Refer Slide Time: 4:16)

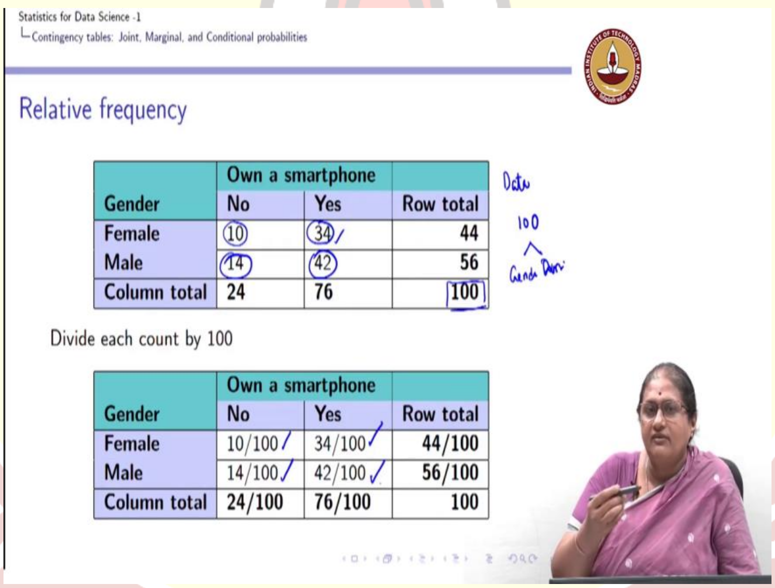


So, if you recall the cell phone versus gender usage this was what we had as a contingency table, I had two variables the first variable was gender and there are two categories of this variable which is female and male, I had the second variable which was ownership of a smartphone, which was off again two categories, no or yes and I was interested in knowing about an association between these two variables.

So, the questions we post was what is the relative frequency of a female if a person is a female what is the relative frequency of a female owning a smartphone or of a male owning a smartphone or what is the frequency of a person who owns a smartphone being a male or a female. These were the equations which we asked. So, we are going to revisit that contingency table and we are going to actually convert these contingency tables into probabilities.

So, recall what are the this 10 and 34 and 42 where the counts, by counts what I mean is of 100 people, 10 people are both female and do not own a smartphone, 34 people are female and own a smartphone, 14 are male and do not own a smartphone and 42 are male and own a smartphone. So, you can see that people who do not own a smartphone are 24 people who own a smartphone or 76, number of female candidates are 44, number of male are 56, this adds up to 100 and this also adds up to 100 people.

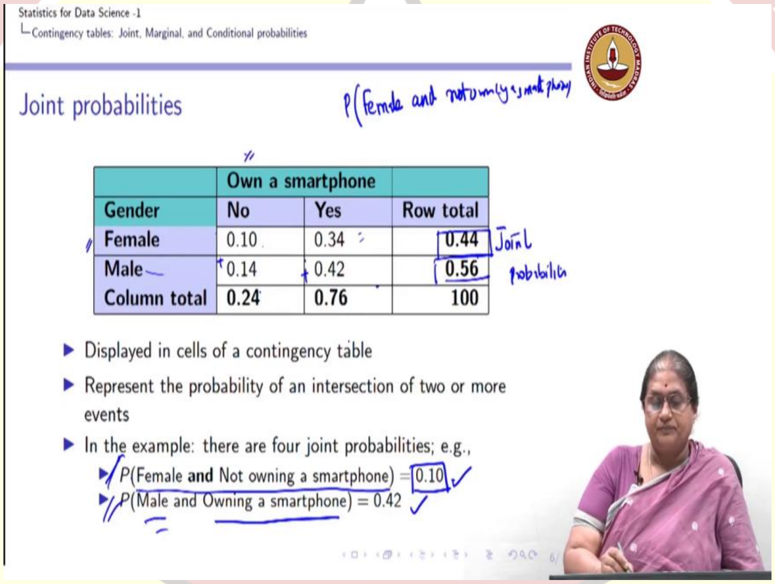
(Refer Slide Time: 06:27)



So, now let us construct a relative frequency table, we introduce what we are row relative frequencies and column relative frequencies when we studied about contingency tables. Now, we are going to introduce what is a relative frequency table. Now, what do you how do I get a relative frequency table? The total number of participants are 100, I divide each one of the entries in the table by this total number of participants, so I get a $\frac{10}{100}$, I get a $\frac{34}{100}$, I get a $\frac{14}{100}$ and I get a $\frac{42}{100}$, so these are my relative frequencies.

Now, if I have to visualize or if I have to interpret this as probabilities, so first remember this is from a data I collected from 100 people where I come I actually recorded what was their gender and whether they owned a smartphone or not, ownership. Now, suppose I want to convert this into a chance, suppose I have the hundred and first person, so the questions I might ask is what is the chance that the hundred and first person is a female, what is the chance that the hundred and first person owns of smartphone or what is the chance that the hundred and first person is both female and owns a smartphone? So, these are what are the probable questions I can ask. So, I want to use these numbers which I have here to convert them into probabilities.

(Refer Slide Time: 08:17)

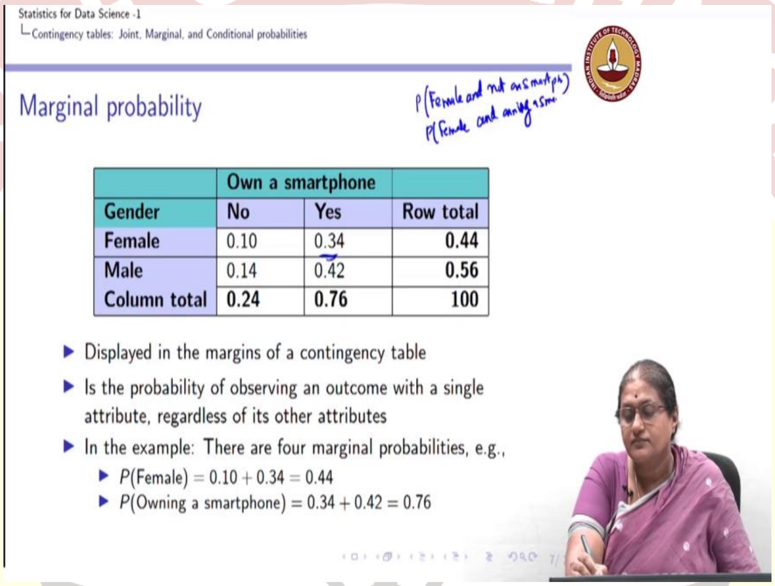


The first probability will introduce is what is referred to as the notion of joint probabilities. So, again recall this is my $\frac{10}{100}$, this is my $\frac{34}{100}$, so let us see what this entry means, this entry is equivalent to the probability of a person being a female and not owning a smartphone, that is what I have listed here, female and not owning a smartphone and that probability is given to us by 0.10, probability of a male and owning a smartphone, so male and owning a smartphone is 0.42.

So, what are the entries here, these entries are referred to as joint probabilities, the reason why they are referred to as joint probability it is the probability of two events happening together and what are these events that are happening together either female with owning a smartphone or female with either female with not owning a smartphone or female with owning a smartphone, male not owning a smartphone and male owning a smartphone.

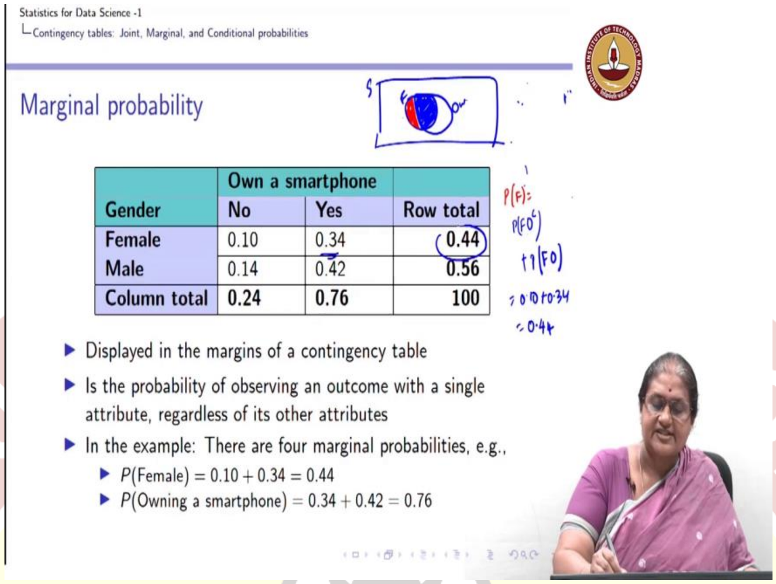
So, the entries inside this contingency table are referred to as joint probabilities. Now, we can see that this 0.44 is 0.10+ 0.34 and 0.56 is 0.14+0.42, similarly you can see that 0.10+ 0.14 is a 0.24, 0.34+0.42 is a 0.76. Now, let us understand what is this 0.10.

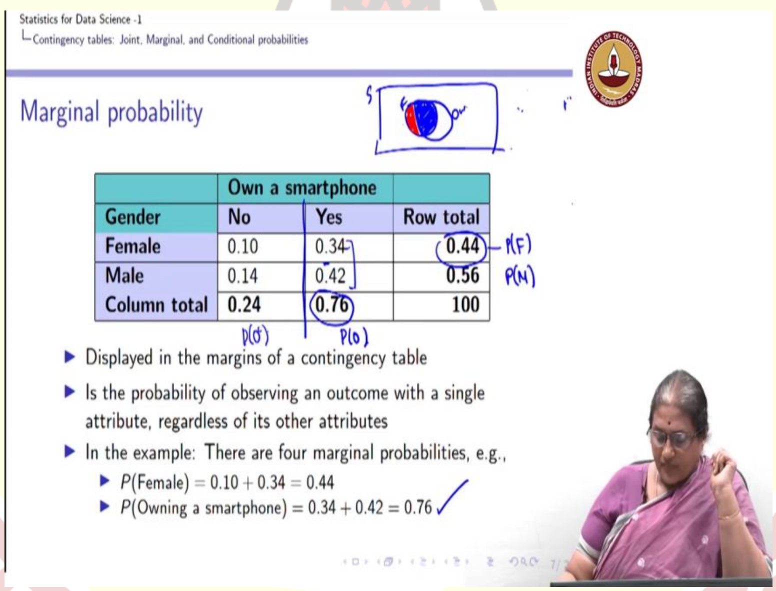
(Refer Slide Time: 10:29)



0.10 is nothing but the probability of a female being a female and not owning a smartphone, whereas 0.34 is the probability of being a female and owning a smartphone.

(Refer Slide Time: 11:05)





So, if I want to know what is the chance that you had a female let us understand this using a Venn diagram, so this is my sample space, this is my event of a person being a female, let me have this as event of owning a smartphone, so this region which is being shaded in blue is the region of being a female and owning a smartphone. The blue region is the region of being a female and owning a smartphone.

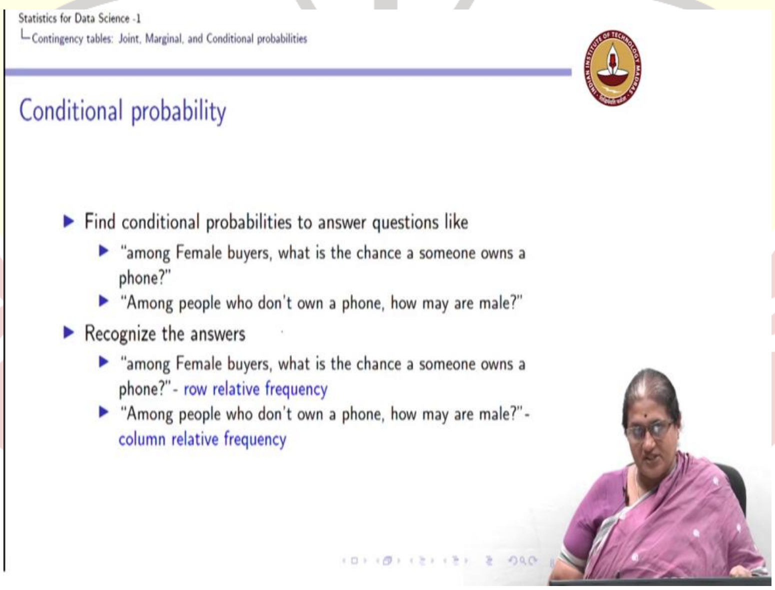
The region of red is being a female and not owning a smartphone, so you can see that if I want to know what is probability of F, this probability of F is nothing but the union of the red region, the red region is being a female and not owning a smartphone that is why I am putting o compliment union, probability of being a female and owning a smartphone.

Now, this is given to us to be 0.10, this is given to us to be 0.34, so you can see that this works out to be 0.44, which is precisely what I have here and these are what we refer to as marginal probabilities and this is what how I have converted the entries of a contingency table to represent A joint probabilities and B now marginal probability.

Similarly, if you look at the variable of owning a smartphone so you can look at this column which is a 0.34, so I can own a smartphone either, so the probability of owning a smartphone is probability of being a male and owning a smartphone and probability of being a female and owning a smartphone. So, if I add up these two entries I get 0.76 and that is again my marginal probability.

So, I have 4 marginal probabilities, this is probability of being a female, this is probability of being a male, this is probability of owning a smartphone and this is probability of not owning a smartphone which is probability of O^C.

(Refer Slide Time: 13:38)

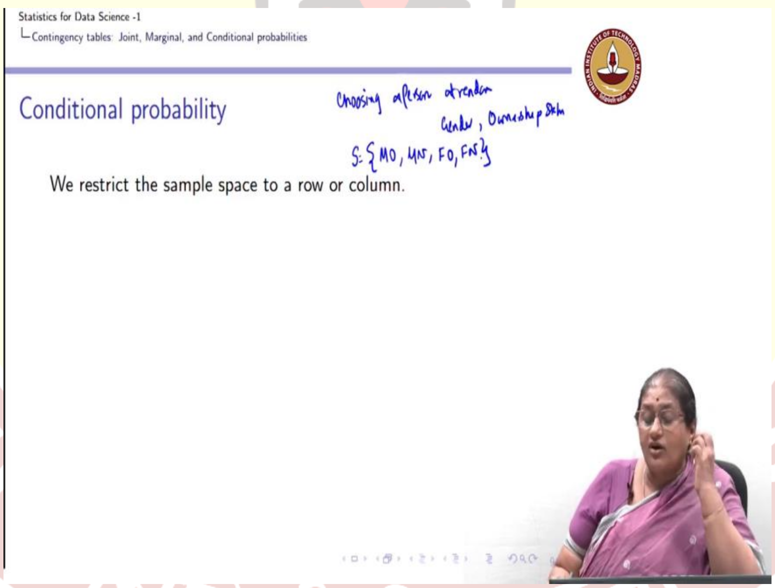


So, using these three let us introduce now the notion of conditional probability. So, when we talk about conditional probability, we are seeking answers to questions like among female buyers, what is the chance that someone owns a smartphone? So, what is the partial information that is given to us? The partial information that is given to us here is among or I am conditioning it on female buyers or among female buyers, I know that the buyer set I am looking at now as the female buyer set. So, among these buyers, so how many females do we have? We have 44 females I am interested I know how many among them or smartphone? Or do not own a smartphone.

The second question is among the people who do not own the first smartphone, so how many of them are there? 24 people. So, these are the total number of people who do not own a smartphone, I know there are 24 people, I am interested in knowing what are the chance that I have a male or how many are male. Remember we again introduced the concept of a row relative and column relative frequency when we discussed about the contingency table.

So, to answer the first question basically I am looking at a row relative frequency whereas to answer the second question I am looking at a column relative frequency. In other words to put it in the framework of a conditional probability I am restricting my sample space.

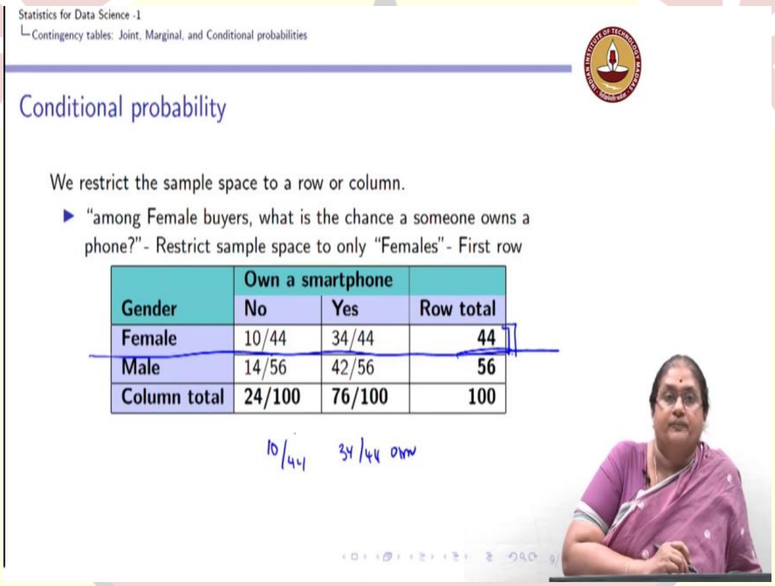
(Refer Slide Time: 15:34)

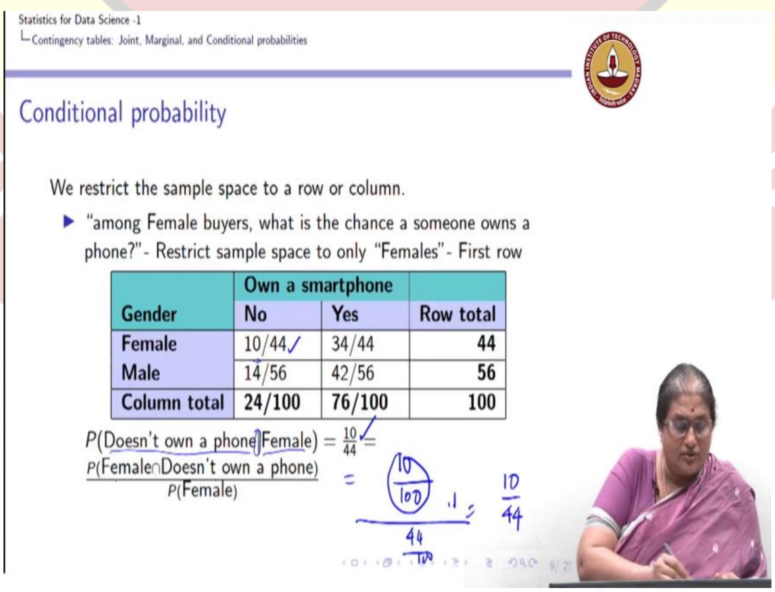


So, if my sample space earlier was having so if I am one way I can so first I need to understand what is my experiment here. So, in the experiment I am choosing a person at random, I am recording their gender and their ownership status, by ownership status I mean yes or no. So, the outcomes possible outcomes are I could have a male and I could have a person who owns I could have a male and who does not own M with does not own a female who owns and female who does not own.

But here remember there are not equally likely. So, I need to find out a way of how to compute these probabilities. So, the way to compute this probabilities I first I work on what I call a restricted sample space. So, by restricted sample space if my original sample space is male owner male does not own female owns and female does not own and I know some information about the person, suppose the information is given that the person is female, then I restrict that sample space to only female owns and female does not own, I restrict the sample space. So, how do we compute it you restrict a sample space.

(Refer Slide Time: 17:12)



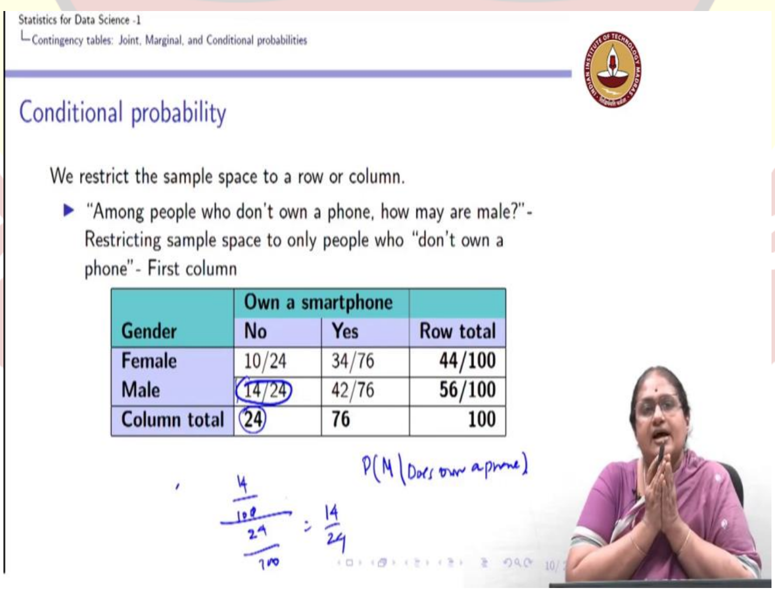


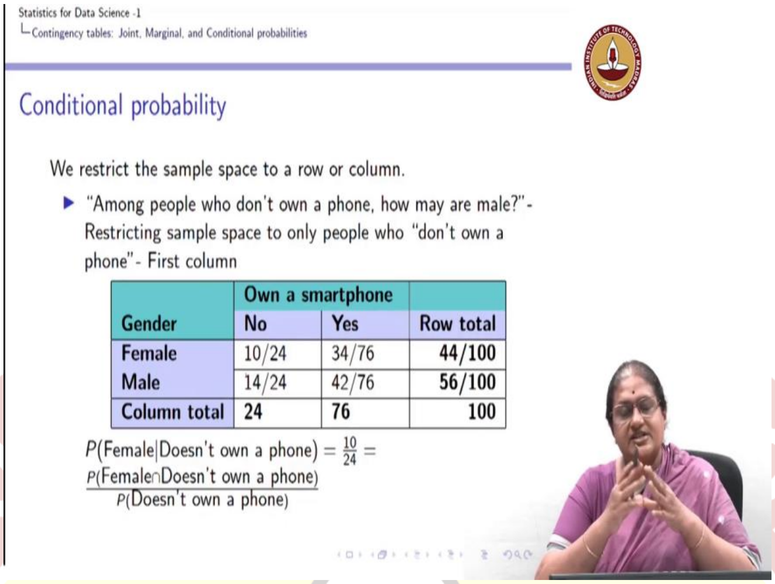
So, now if I restrict the sample space to only females, so I am not bothered about the sample space of perhaps the person is a male among these females, so earlier I computed $\frac{10}{100}$because I had the expanded sample space, now I am only finding what is the relative frequency, related to what? Related to the person being a female, I know 10 out of 44 do not own and 34 out of 44 own a smartphone.

Similarly, 10 out of 56 again 56 is the number of males, 14 out of 56 do not own a phone and 42 out of 56 own a phone. So, I am restricting the sample to females only or males only which is my first row. And I know that the chance that someone owns a phone is $\frac{34}{44}$. And the chance that a person, so if I am looking at probability of does not own a phone given a female this is we will explain this notation in a while, but if I am looking at a chance that I know a person is a female what is the probability that they do not own a phone, I see that it is $\frac{10}{44}$.

And we can see that this is nothing but 10 was $\frac{10}{100}$ was a probability that I had a female and does not own a phone, probability of a female is $\frac{10}{44}$ and I can see that this $\frac{10}{44}$ can be given by $\frac{10}{100}$ divided by $\frac{44}{100}$which is $\frac{10}{44}$, this is something which I observe.

(Refer Slide Time: 19:21)



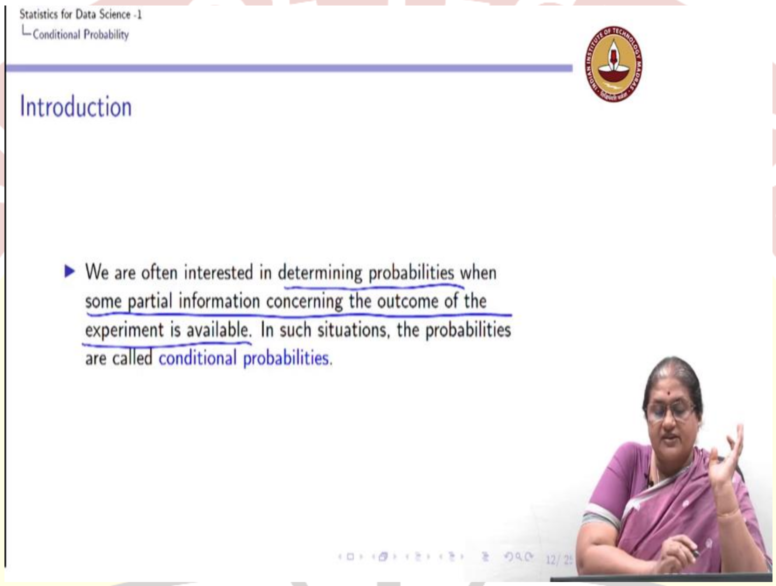


Similarly, if I restrict among the people who do not own a phone, who are the people who do not own a phone, I have 24 people who do not own a phone, how many are males? So, I know 14 people in this, so jointly I know 14 people out of a 100 people do not 14 males out of 100 males do not own a phone. But given so if I am want restricted to the sample space of not owning a phone I know only 24 do not own a phone.

So, if I want to find out the conditional probability, so if I want to find out the conditional probability of what is the chance of being male given a person does not own a phone that would reduce to $\frac{14}{24}$ divided by my $\frac{14}{100}$ divided by $\frac{24}{100}$ which is $\frac{14}{24}$ and that is what I have here. In other words when we are talking about conditional probability we are actually restricting our sample space to include all those events conditioned on the event that has occurred.

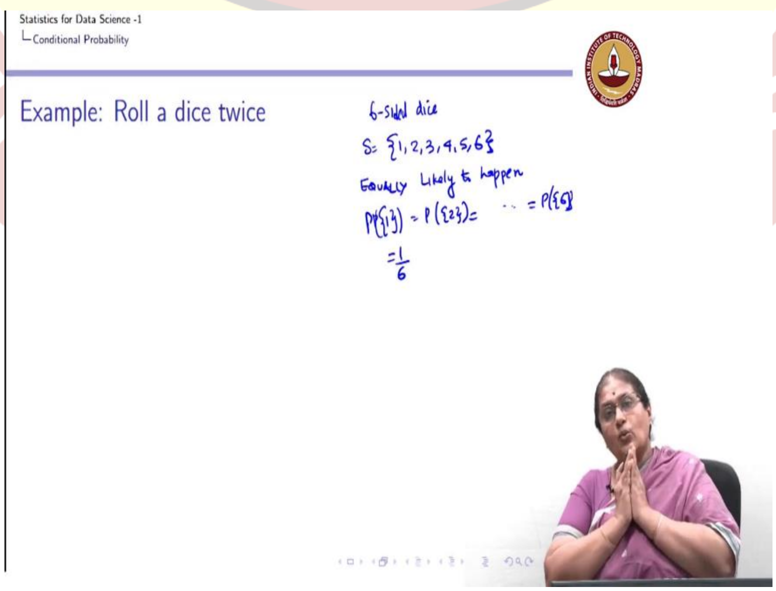
Statistics for Data Science - 1
Professor. Usha Mohan
Department of Management Studies
Indian Institute of Technology, Madras
Lecture No. 7.2
Conditional Probability – Conditional Probability Formula

(Refer Slide Time: 00:14)



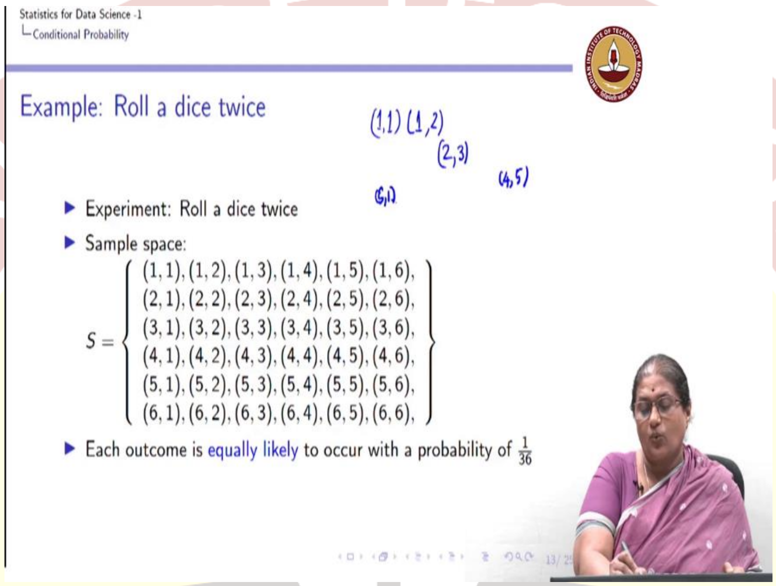
So, now let us actually introduce the notion of a conditional probability. Why do we have to learn about conditional probability? Because most of the times are very often we are interested in determining probabilities when some partial information concerning the experiment is available. So, let us understand it through an example.

(Refer Slide Time: 00:43)



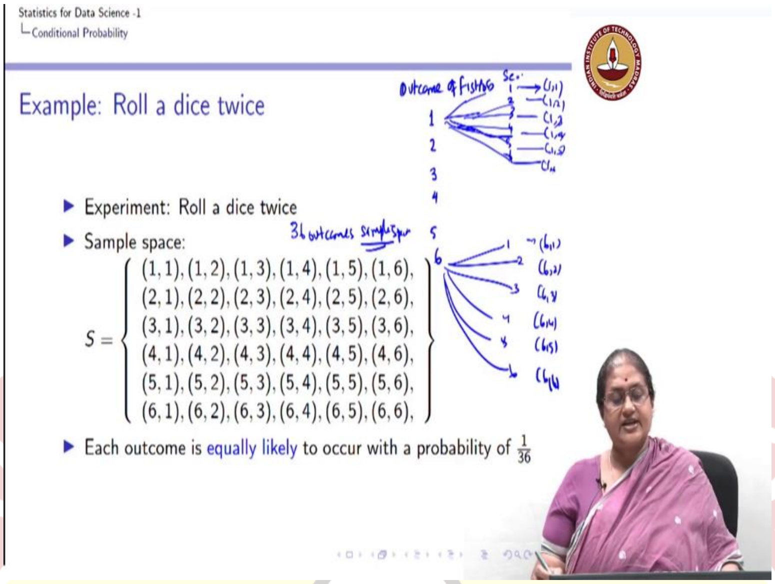
Let me have the experiment of rolling a dice twice I am assuming this dice is fair in the sense that we know that it is a 6 sided dice, so if I throw it once the sample space is I can have any one of the outcomes 1, 2, 3, 4, 5, 6, I again assume all the outcomes are equally likely to happen, we introduce what we mean by equally likelihood and this means that the probability of this 1 happening is the same as probability of 2 happening, which is the same as probability of each of the outcomes happening equally likely which is going to be $1/6$.

(Refer Slide Time: 1:51)



So, now suppose I toss a coin twice or so now suppose I am tossing or rolling this dice twice, so what are the outcomes so the again the experiment is to roll this fair dice twice, my sample space in this case is going to be, so if I my first toss is a 1, my second roll is also a 1, first is a 1, my second could be a 2, my second could be a my first toss could be a 1, my second could be a 3, my first toss could be a 4, my second could be a 5, my first toss would be a 5, my second could be a 1, so you can see that the set of possible outcomes is with every first toss.

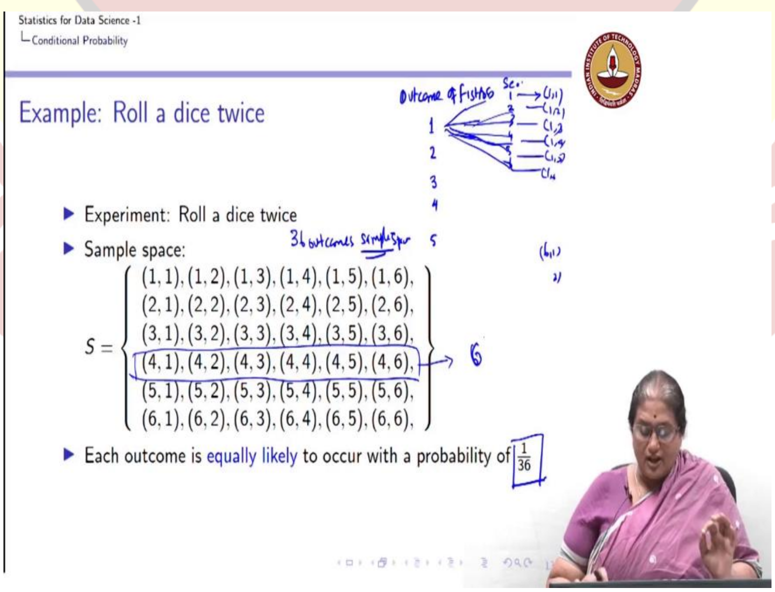
(Refer Slide Time: 2:25)



So, I have the first so I can depict this in this way the outcome of first toss the outcome of a first toss could be any one of this 1, 2, 3, 4, 5 and 6, now if my first toss is a 1 the second toss could again be so if I have second toss it could be a 1, 2, 3, 4, 5, 6, with 2 also I could have a 1 so this final outcome of the experiment is (1,1), this is (1,2), this is (1,3), this is (1,4),this is a (1,5), this is a (1,6), similarly with each one of these outcomes with 6 ,I could have a 1, 2, 3, 4, 5 and 6.

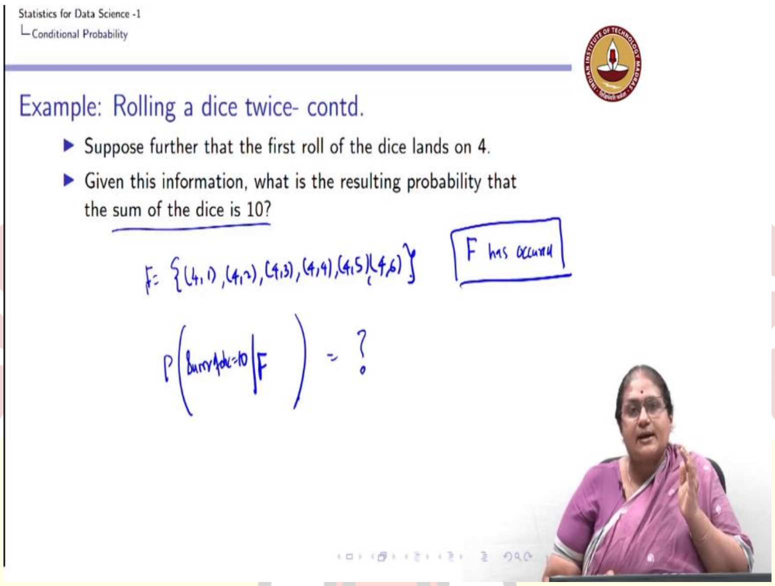
So, this outcome corresponds to (6 ,1), (6 2), (6, 3), (6, 4), (6,5) and (6, 6), so you can see with each one, I have 6 more outcomes and that gives me a total of 36 outcomes in my sample space, each outcome is equally likely to happen I have 36 outcomes, so each of these outcomes are equally likely to happen with the probability of $1/36$.

(Refer Slide Time: 04:07)



Now, suppose further I know that the first roll of the dice lands in a 4, so what are the chances? So, what are the outcomes? So, this you can see that these outcomes correspond to the first roll being a 4. How many outcomes do I have here? I have 6 outcomes and these 6 outcomes and these 36 outcomes actually correspond to the outcome that the first roll is a 4.

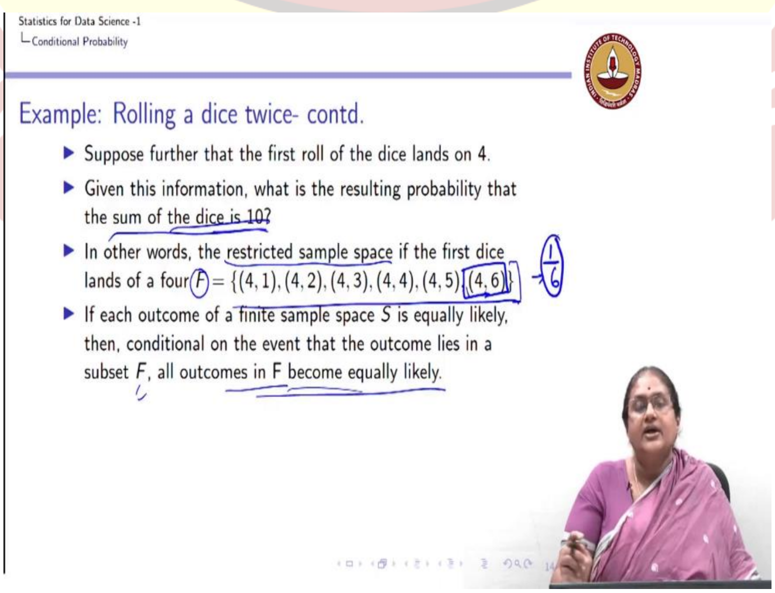
(Refer Slide Time: 04:42)



So, I want to know that given that the first outcome is a first role is a 4, what is the chance of the resulting probability that the sum or what is the chance that the sum of the dice is a 10? So, if I know that the first outcome is 4, let me define the event F to be $\{(4,1),(4,2),(4,3),(4,4),(4,5),(4,6)\}$, so what is given to us?

The information that is given to us is F has occurred, this is the partial information that has been given to us is F has occurred. So, given this information, so given this information I want to know what is the chance that some of dice is equal to 10? That is the question which we are asking. Given that the first roll is a 4, what is the chance that the sum is equal to a 10?

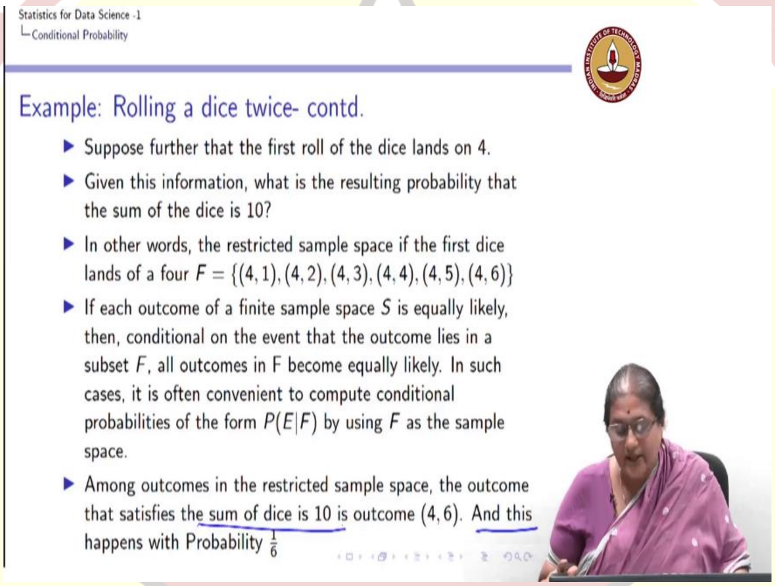
(Refer Slide Time: 06:03)



So, if an outcome of a finite sample space, what is the finite sample space we consider here I have 36 outcome, so if an outcome of a finally finite sample space is equally likely then conditional on the event that the lies in a subset F the outcomes in F are also equally likely. So, if I am conditioning at on this event the outcomes of this event or this subset are also equally likely.

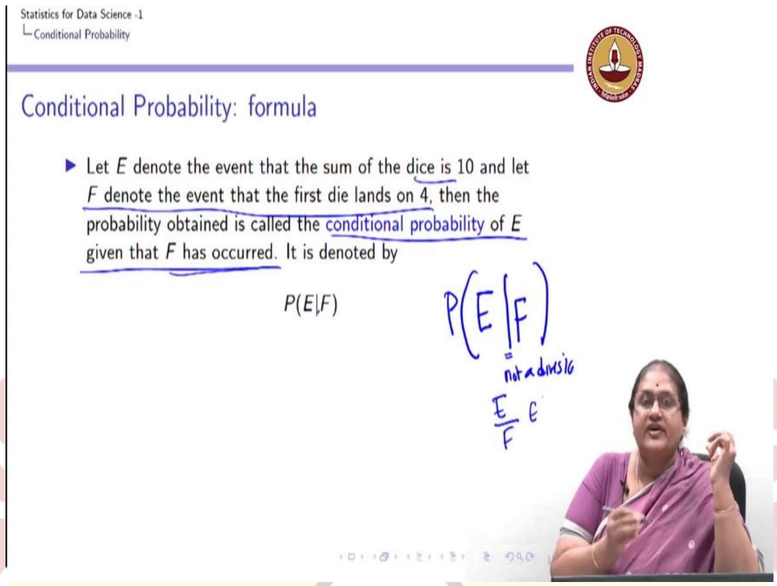
So, conditioned on what I refer to as a restricted sample space, what is the restricted sample space? I have this full sample space, but I am not interested in all these outcomes here, I am only interested in those outcomes that are favourable to this event that the first roll is a 4 and I am looking at these events also being equally likely, now within this restricted sample space you can see that the outcome that satisfies that the sum of dice is 10 is this outcome. And what is the chance of that assuming all the outcomes here are equally likely in this space I have 6 outcomes of which one of the outcome satisfies this event giving me a probability of $1/6$.

(Refer Slide Time: 07:40)



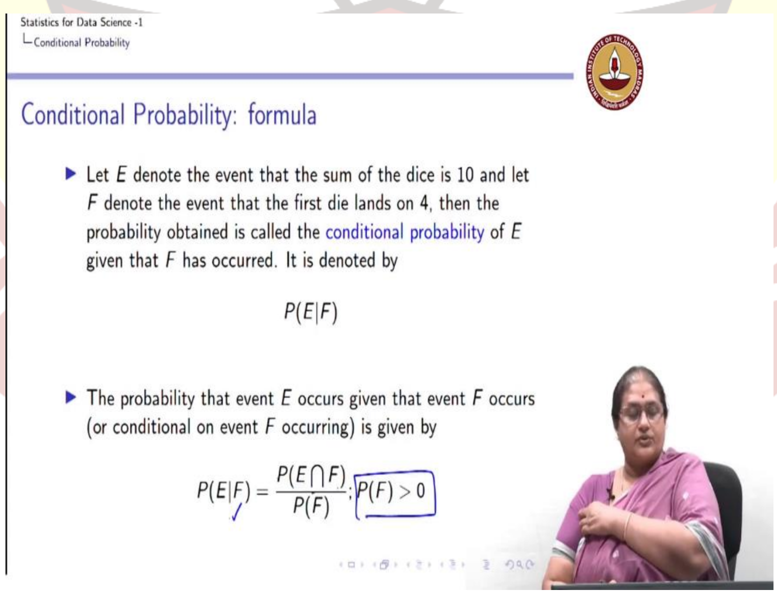
So, you can see that the $P(E|F)$ can be obtained by this logic as $1/6$. So, among the outcomes in the restricted sample space, since the outcomes in the restricted sample space are also equally likely, I have 6 outcomes in my restricted sample space, I can say that the probability of sum of dice is equal to 10 is as $1/6$.

(Refer Slide Time: 08:13)



So, now let us come to a formal formula for the conditional probability. Now, what do we have here? If E denotes the event that the sum of dice is 10, so this is the event on which I am conditioning or this is the conditioning event and let F denote the event that the first die lands on a 4, so I am interested in knowing what is the conditional probability of E given that F has occurred. I denote that by $P(E|F)$. I repeat I write that or denote it as probability of E conditioned on F, please remember that this is not a division symbol, it is not E divided by F or it should not be told E by F, always practice yourself to articulate it as probability of E conditioned on F or probability E given F.

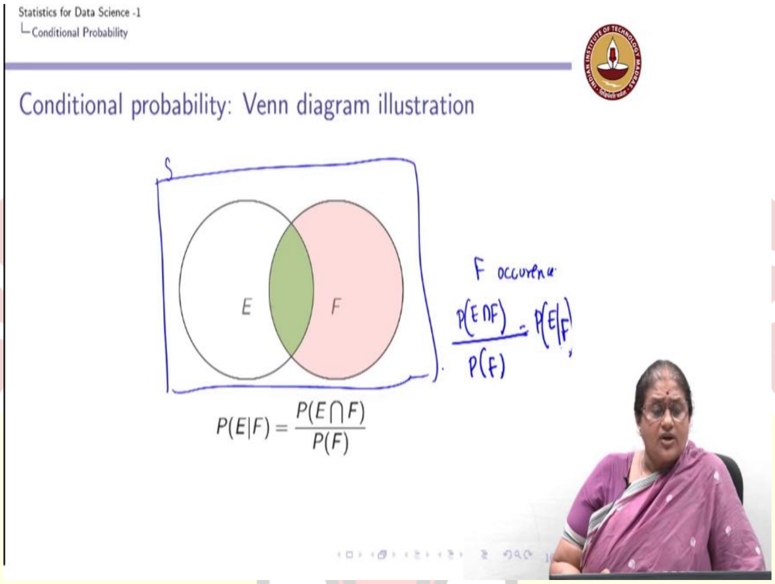
(Refer Slide Time: 9:34)



So, what is this $P(E|F)$? What is the formula for this probability? The probability of an event E occurring given that an F occurring is given by $\frac{P(E \cap F)}{P(F)}$and we know that this is defined if I have

$P(F) > 0$ or in other words I always condition on a non null event. So, the conditional probability formula is $P(E|F) = \frac{P(E \cap F)}{P(F)}, P(F) > 0$

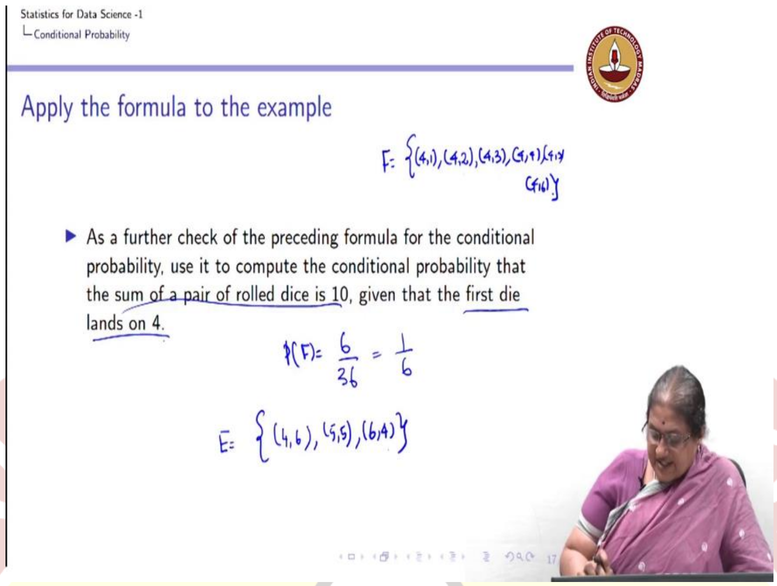
(Refer Slide Time: 10:16)

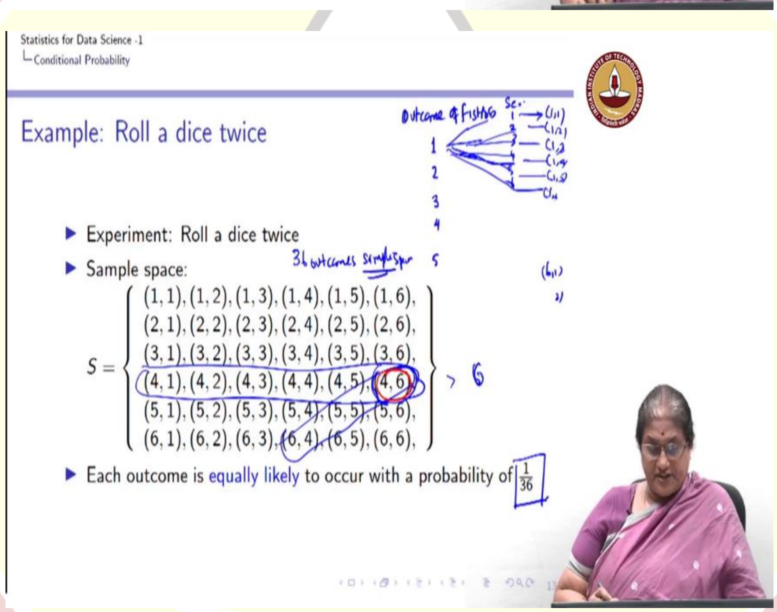


So, now let us understand it through a Venn diagram, we also saw that Venn diagrams illustrate the concept of probability well, so suppose I have a sample space and I have two events E and F which are subsets of my sample space, now I am interested in knowing so this is my E event and this is my F event, so the green portion is my $E \cap F$. Now, the way I can view this is suppose I want to know what is the chance of E happening given F has happened, so I can look at this.

So, this green portion is the portion of your entire pink portion entire pink portion is F occurrence and this E portion this actually this shaded green portion is the chance of E occurring because given F has occurred, pink portion is F occurring, the green portion is the chance of E occurring given F has occurred and we can see that the green portion is nothing but $E \cap F$, so $\frac{P(E \cap F)}{P(F)}$ this green portion to the entire pink portion is my $P(E|F)$, that is one way of visualizing the conditional probability formula.

(Refer Slide Time: 12:05)

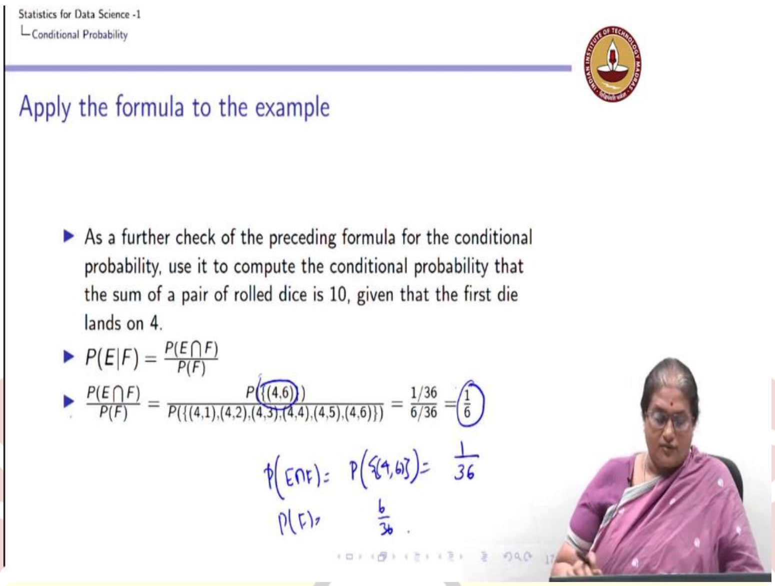




So, now let us apply the formula to the example that we have stated. The example was again here we check the preceding formula, so what is it we are interested in knowing? I have this event F which is the event that the first dice lands on a 4 and I know the outcomes of this experiment are 4 of this event is $\{(4,1),(4,2),(4,3),(4,4),(4,5),(4,6)\}$. Now, the probability of this F is 6 outcomes all of them are equally likely again recall this is going to be 6 /36 which is my 1 / 6.

Now, let E be the event that the sum is 10 and I know the events or the outcomes that satisfy this are $\{(4,6),(5,5),(6,4)\}$, these are the 3 events that satisfy that the sum is equal to 10, I can see that those events here are though outcomes here are this outcome, so I can write down that these 3 outcomes are the outcomes that satisfy my event E and this are my outcomes that satisfy my event that the first roll is a and we can see that the intersection of these 2 events is this outcome which is $\{(4,6)\}$.

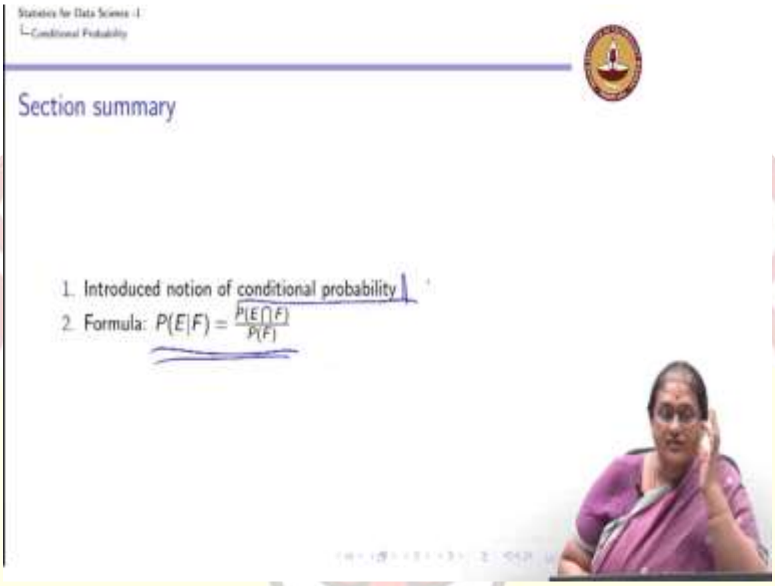
(Refer Slide Time: 14:06)



So, if we define E to be the event which has the so you can see that $E \cap F$ is my outcome which is $\{(4, 6)\}$ and the $P(E \cap F)$ is the same as the probability of this outcome (4, 6) happening, which is $1/36$ again I am assuming all my outcomes are equally likely, probability of F we have already computed it to be $6/36$, so by applying the conditional probability formula this is going to be $\frac{1/36}{6/36}$ which is $1/6$, which is what we obtained when we restricted the sample space and assume that the outcomes of the restricted sample space are also equally likely and computed the probability.

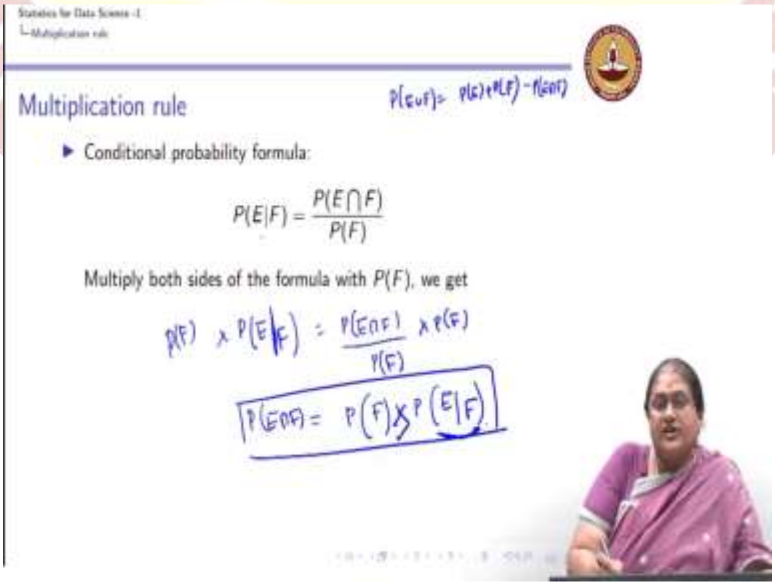
Statistics for Data Science – 1
Professor Usha Mohan
Department of Management Studies
Indian Institute of Technology, Madras
Lecture No. 7.3
Conditional Probability - Multiplication rule

(Refer Slide Time: 0:14)



So, what we have introduced is the important notion of conditional probability which takes care or discusses about how to compute probability of event’s condition or given the partial information that some other event has occurred and this is the formula for conditional probability, $P(E|F) = \frac{P(E\ \ F)}{P(F)}$ with $P(F)$

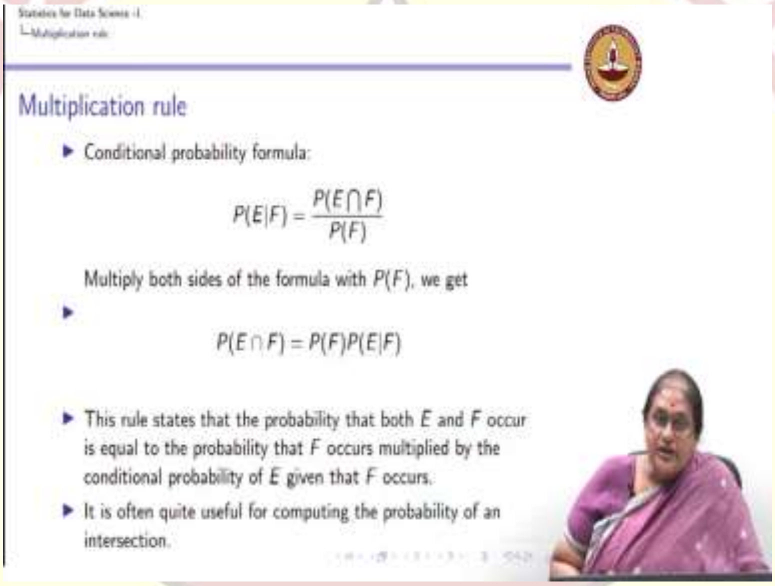
(Refer Slide Time: 1:00)



Now, from this formula which we have derived or we have given for conditional probability, we derive an important rule which we refer to as the multiplication rule. Recall that the probability, conditional probability formula tells, $P(E|F)$, again I repeat that this is not E divide by F, it is E given F or E conditioned on F is $\frac{P(E \cap F)}{P(F)}$.

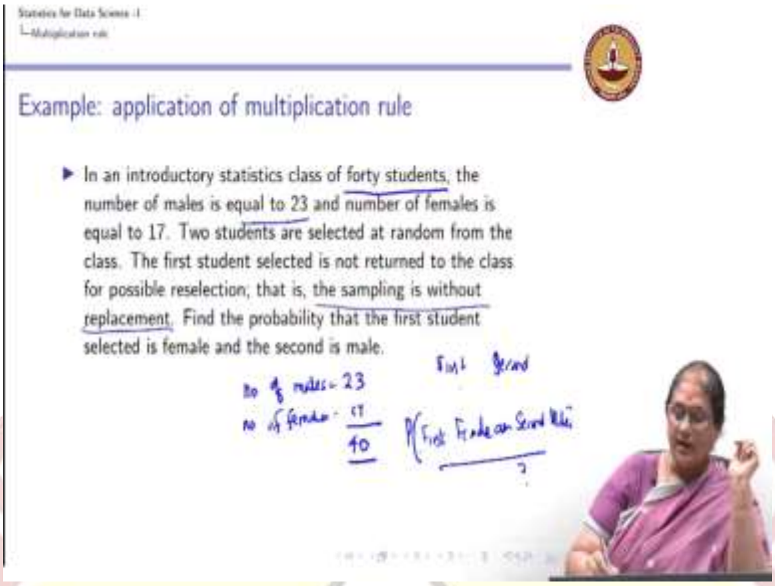
I can multiply both the sides with $P(F)$, so I get $P(E \cap F)$. So, if I multiply both the sides with $P(F)$ because $P(F) > 0$, I will get $P(E \cap F) = P(F)\,P(E|F)$. So, recall the additional rule gave us $P(A \cup B)$ or $P(E \cup F) = P(E) + P(F) - P(E \cap F)$. This is what our addition rule stated. Now, the multiplication rule states that $P(E \cap F) = P(F)\,P(E|F)$.

(Refer Slide Time: 2:26)



So, this multiplication rule actually states that the probability of both E and F occurring together, that is the probability of intersection is equal to that the probability that F occurs that is this multiplied by the conditional probability of E given F occurs which is this. So, this multiplication rule is useful to compute the probability of intersection of events.

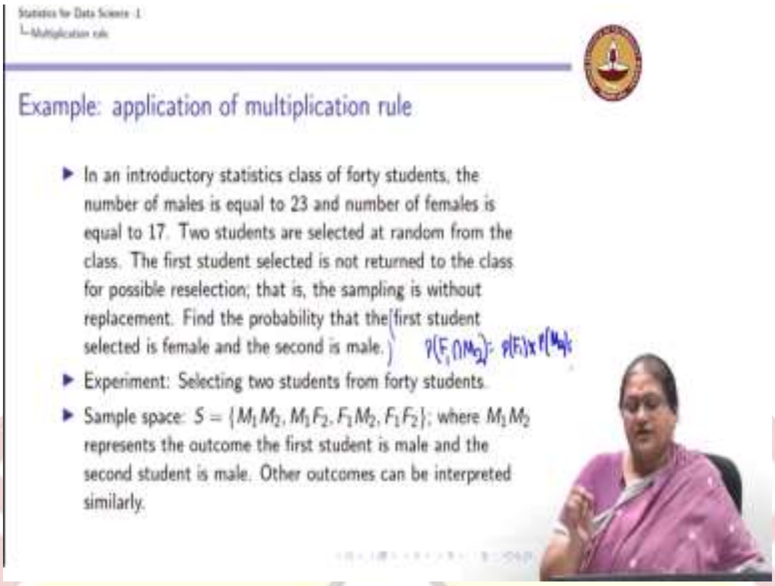
(Refer Slide Time: 2:58)



So, now let us look at an example of application of the multiplication rule. So, in an introductory statistics class of 40 students, so total number of students in the class is 40, the number of males = 23 and number of females = 17. And total number is 40 students. I am selecting 2 students at random and the way I am selecting these 2 students is I have a first student and a second student, if I am selecting a first student that student is not available for selection. In other words it is called a sampling without replacement.

So, if I select the first student, this student is not available for selection for the second student, so this is called a sampling without replacement, I am not having that student available. So, the first student and the second student. So, what are we asking? We are asking for probability that the first student who is selected was a female and second student is a male. This is the question we are asking.

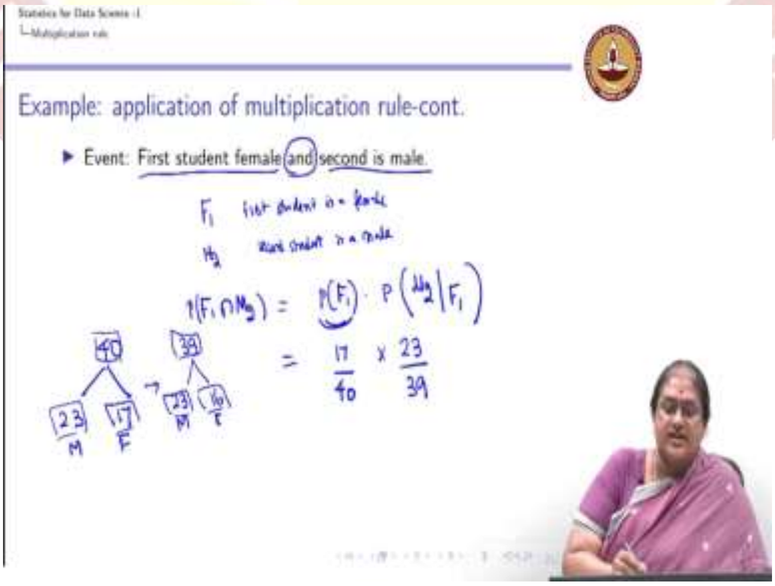
(Refer Slide Time: 4:35)

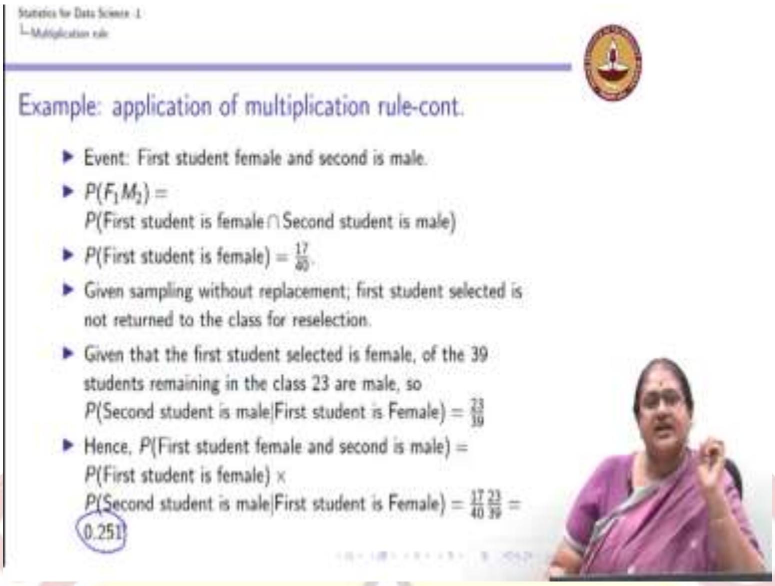


So, now if I put this in the framework which we know the random experiment here is to select 2 students and how are we selecting these 2 students, we are picking up one student, keeping him aside, then we have the remaining 39 picking up another student. So, the random experiment is selecting 2 students from the 40 students. I can list by sample space in the following way M_1M_2, M_1F_2, F_1M_2 and F_1F_2, where M_1M_2 refers to the first student being a male and the second student also being the male.

F_1M_2 is first student is a female and the second student is a male. M_1F_2 is a first student is a male and the second student is a female and F_1F_2 is the chance that both the first and the second student selected are female.

(Refer Slide Time: 5:37)





Now, from this event, I am defining the event that the first student is a female and second student is a male. So, F_1 represents the outcome that the first student is a female, M_2 is the outcome second student is a male. Now I want to know $P(F_1 \cap M_2)$ because here it is asking what is the chance that the first student is a female, and second student is a male.

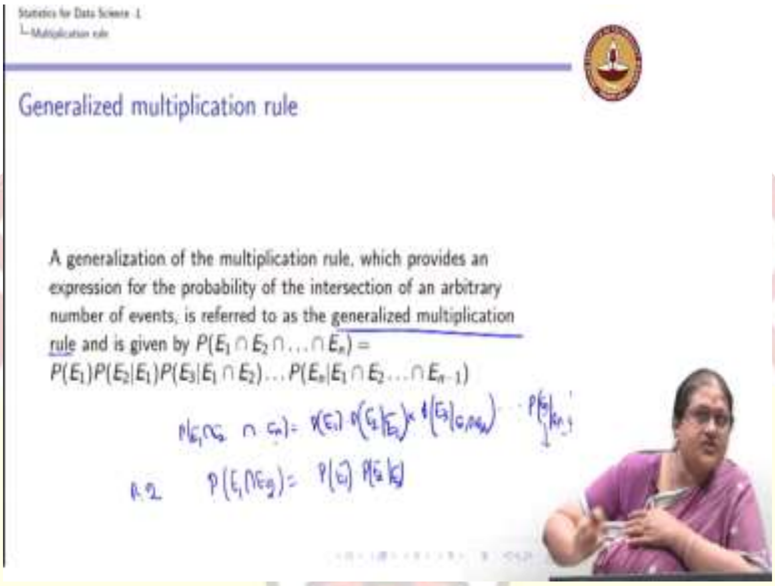
Now, if I apply the multiplication rule, know that this is probability of first student being a female into probability of the second student being chosen is a male conditioned on the fact that the first student was a female. Now if you look at what is the probability of the first student being a female, I have 40 students of which I have 23 of them who are male and 17 of them who are female.

So, the chance of my first student being a female is 17/40, because I have 17 females and 40 total student, so the chance of me selecting a female out of 40 students is 17/40. Now once I have selected this female student, the total number of students become 39 and this I have already selected a female so I have only 16 females remaining, but I have 23 males remaining.

So, given that there are 20, the first student was a female, now the chance of me finding a male reduces to 23/39 because the total number of students is now 39 with 23, this distribution or this separation or this classification that 23 are male and 16 are female. So, the chance of the first student being female and second student being male is $\frac{17}{40} \times \frac{23}{39} \approx 0.251$.

So, this is what we referred to as the multiplication rule and we have seen an example where the multiplication rule has been used to compute the probability of a intersection. So, what we need to understand is translate the problem in terms of the events so that we understand what is the event that we are seeking. For example, here the event that we are seeking is female 1 and the second is a male, it is expressed as an intersection of the events and once we have expressed it as an intersection of events, we applied the multiplication rule to get hold of the probability of the event.

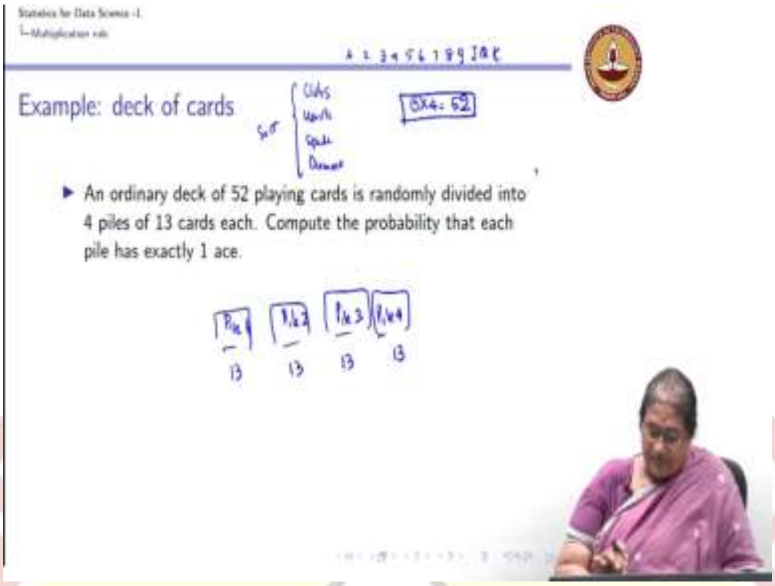
(Refer Slide Time: 9:06)



So, this multiplication rule can be generalised to arbitrary number of events. Suppose I have n events, then I can refer to that as a, and I refer to this as a generalized multiplication rule. What is the generalized multiplication rule? If I have n events, $P(E_1 \cap E_2 \cap \ldots \ldots \cap E_n) = P(E_1) \times P(E_2|E_1) \times P(E_3|E_1 \cap E_2) \ldots \ldots . . \times P(E_n|E_1 \cap E_2 \cap \ldots \ldots \cap E_{n-1})$.

So, when my $n = 2$, it reduces to what we already have is $P(E_1 \cap E_2) = P(E_1)\, P(E_2|E_1)$, because if $n = 2$, it reduces to this term just because $P(E_2|E_1)$ because E_{n-1} is E_1 which is what I already have seen. So, which is, $E_2|E_1$ which is something which we have already computed.

(Refer Slide Time: 10:26)

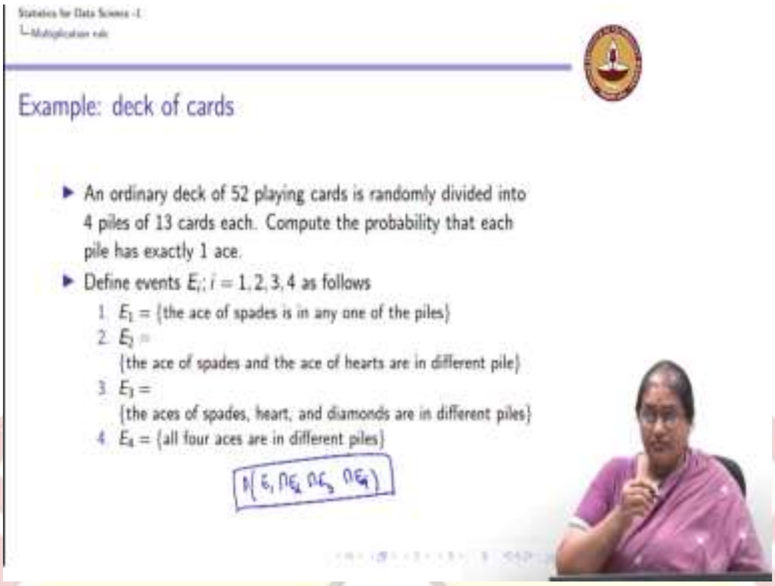


So, let us look at an example of the application of the generalized multiplication rule. Again consider a deck of 52 playing cards. Now all recall when I talk about a playing cards, I have clubs, I have hearts, I have spade and I have, so I have 13 faces with these are what I referred to as a suit of a card. So, 13 × 4, I have 52 cards. We have already seen a card example when we discussed about probability of events.

So, what we are doing from these 52 cards is I have picked up, I have randomly dividing this into 4 piles of 13 cards each. So, we are randomly choosing cards we are preparing pile 1, pile 2, pile 3, so I have 4 piles. So, what are we doing? We are picking up cards at random and putting them into 4 piles or 4 buckets whatever way we would like to choose, pile 3 and pile 4. So, I pick up 1 card, I put. So, I start putting, I shuffle these cards and I start putting them into these 4 buckets or 4 piles randomly. So each pile now has 13 cards.

So, now we want to compute the probability that each pile as exactly 1 ace. So, I could have a ace of hearts here, I could have a ace of diamond here, I could have a ace of spade here and a ace of club here, but I cannot have a situation where pile 1 has an ace of heart and a ace of club, ace of spade here, ace of diamond here. This cannot, this is not permitted. But I want each of these piles to have exactly one ace. How do I apply the general multiplication rule to solve this problem?

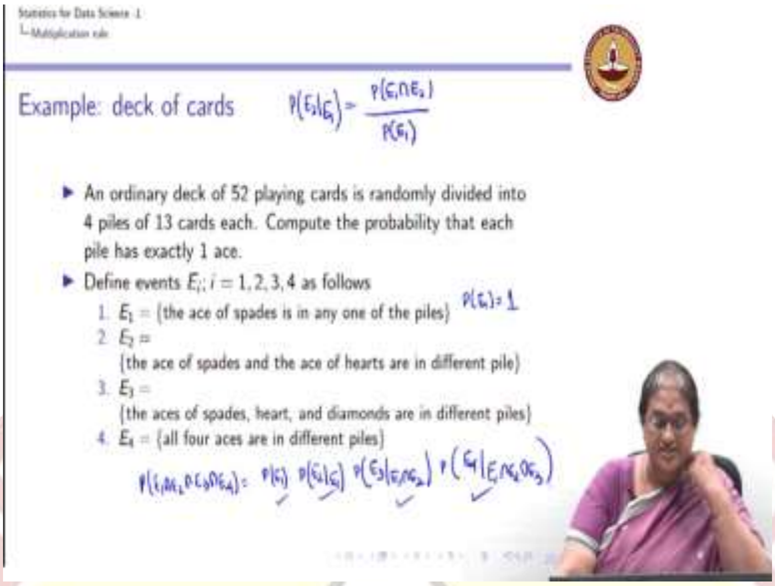
(Refer Slide Time: 12:52)



So, now let us define a few events. Now let E_1 be the event that the ace of spades is in any one of the pile. So, remember I have 4 piles. I want to define an event that ace of spades is in any one of these piles. The second event is ace of spades and ace of hearts are in different piles. There is no sanctity about starting with ace of spades, I could have started with ace of hearts and defined ace E_2 to be the event ace of hearts and ace of spades are in different piles.

The third event is ace of spades, heart and diamonds are in different piles and the fourth event is all 4 aces are in different piles. So, the question that compute that the probability that each pile has exactly 1 ace is equivalent to finding out the probability that $P(E_1 \cap E_2 \cap E_3 \cap E_4)$, this is what I want to find out that each pile has exactly 1 ace is equivalent to finding out this probability.

(Refer Slide Time: 14:22)

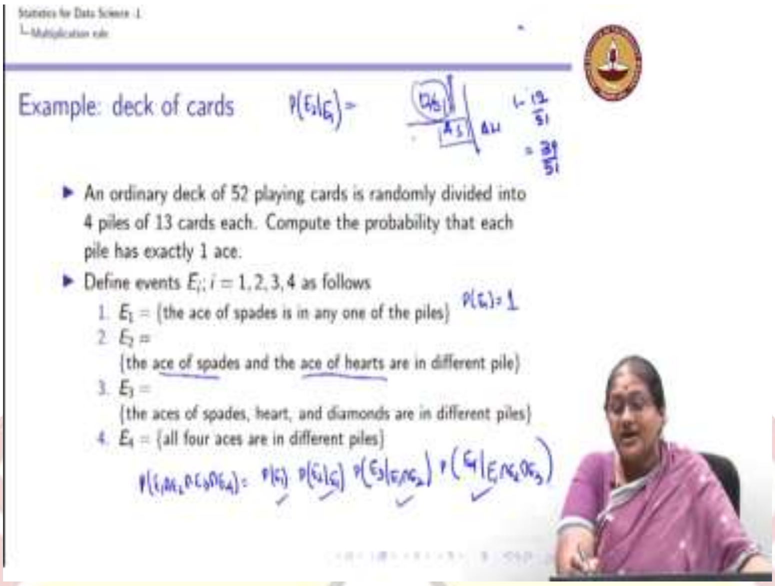


And by general rule of probability, I have that $P(E_1 \cap E_2 \cap E_3 \cap E_4) = P(E_1) \times P(E_2|E_1) \times P(E_3|E_1 \cap E_2) \times P(E_4|E_1 \cap E_2 \cap E_3)$. So, once I find out these 4 probabilities, I can answer the probability on the left hand side which is the probability we want to compute. Now let us look at each event's probability. What is event E_1?

Event E_1 is that the ace of spades is in any one of the piles. So, what does this mean that, what is the chance that I can find ace of spades in any one of the piles, I know that this is equivalent to the probability of the sample space that is ace of spade, so I have these 52 cards, I have put it in 4 piles. So, ace of spades is in this or this or this or this so the chance of finding ace of spades in any one of them is equal to 1, so I can write down $P(E_1) = 1$ or it is a sure event that is I can find the ace of spades being in any one of the piles is a sure event.

Now, let us look at $E_2|E_1$. From my definition I know $P(E_2|E_1) = \frac{P(E_1 \cap E_2)}{P(E_1)}$. Now what is $E_1 \cap E_2$? $E_1 \cap E_2$ is the ace of spades and ace of hearts are in different piles. So, now let me look at the conditioning part.

(Refer Slide Time: 16:34)

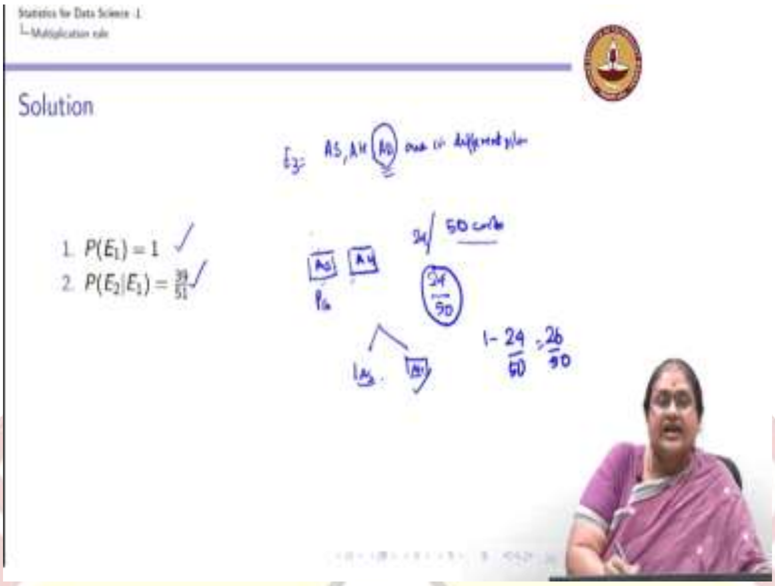


I look at E_1 and it is a ace of spade, I put that in one of the piles. Let me put it in say pile 1. Now, each pile has 13 cards, each of the piles have 13 cards, so if I put ace of spades as a one of the cards in this pile 1, the remaining 12 cards can be chosen from, so I am fixing this ace of spades, so I have 51 cards that are remaining so I can choose the remaining 12 cards in this pile of which ace of spades is there. See, I repeat, I take a pile on which I already have ace of spades, this could be any of the piles, but if ace of spades is one of the cards in a pile, I have to choose 12 remaining cards in the pile.

Now, these 12 remaining cards can be chosen from 51 cards. Now, how can they be chose? Why are they chosen from 51 cards? Because the ace of spades is not available for my choice now. And that can be chosen from 51 cards in 12/51 ways. Now I do not want ace of hearts to be one of these 12/51 choices. I do not want the ace of hearts because I do not want ace of spades and ace of hearts to be in the same pile.

The remaining 12 cards can be chosen by 12/51 ways. But I do not want the ace of hearts to be in this 12/51 ways. So the way that ace of spade and ace of hearts will be in different piles is going to be $1 - \frac{12}{51} = \frac{39}{51}$ ways. I am just using the conditional probability logic by if I am conditioning it on the fact that ace of spades is already occupying a law, the remaining 12 cards can be got in 12/51. I do not want ace of hearts to be in this choice, so the way ace of hearts cannot be in this choice is $1 - \frac{12}{51} = \frac{39}{51}$.

(Refer Slide Time: 19:12)

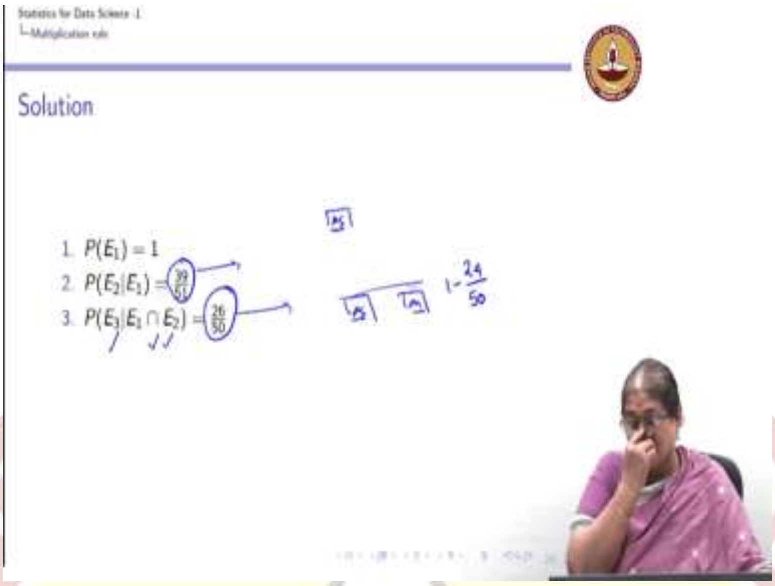


And hence, $P(E_2|E_1) = \frac{39}{51}$. Now what is E_3? E_3 is ace of spade, ace of hearts and ace of diamond are in different piles. So, I already have ace of spade in one pile, ace of hearts in one pile, so if I am fixing ace of spades in one pile and ace of hearts in one pile, I have 50 cards because these 2 cards are already fixed. Now to fill up this pile 1 say, I need a 12 cards, to fill up this I need another 12 cards. So, I need a total of 24 cards to be chosen out of 50 cards.

I repeat. I have ace of spades in one pile, ace of hearts in one pile, so if I fix ace of spades in one pile and ace of hearts in one pile, I have only 50 cards to be chosen from total 50 cards. In pile 1, I can, I need to choose 12 cards, pile 2 I need to choose 12 cards, so in total I need to have 24 out of 50 that has to be chosen. I am writing 24 out of 50 because we can interchange these choices and now 24/50 ways I have AS and A hearts are in different piles.

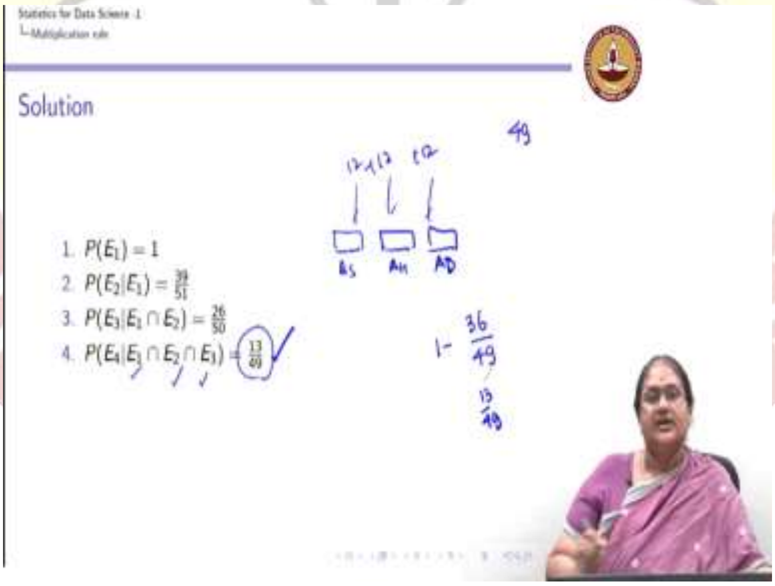
Now, I do not want A diamond to be in either of these piles. So, the number of ways ace of diamond is not in these of any of these 2 piles, I know there are 24/50 ways it can be in these 2 piles, so $1 - \frac{24}{50} = \frac{26}{50}$ ways I will have ace of diamond not in the piles in which already ace of spades and ace of hearts is.

(Refer Slide Time: 21:37)



So, $P(E_3|E_1 \cap E_2) = \frac{26}{50}$. Now similarly using the same logic so we got this 39/51 by fixing ace of spades, we got this 26/50 by fixing ace of spades and ace of hearts and we said that the remaining 24 cards that need to go into these 2 piles can be got by 24/50 ways, 24/50 ways and the way I can ensure that ace of diamond is not a outcome of this, so I get $1 - \frac{24}{50} = \frac{26}{50}$.

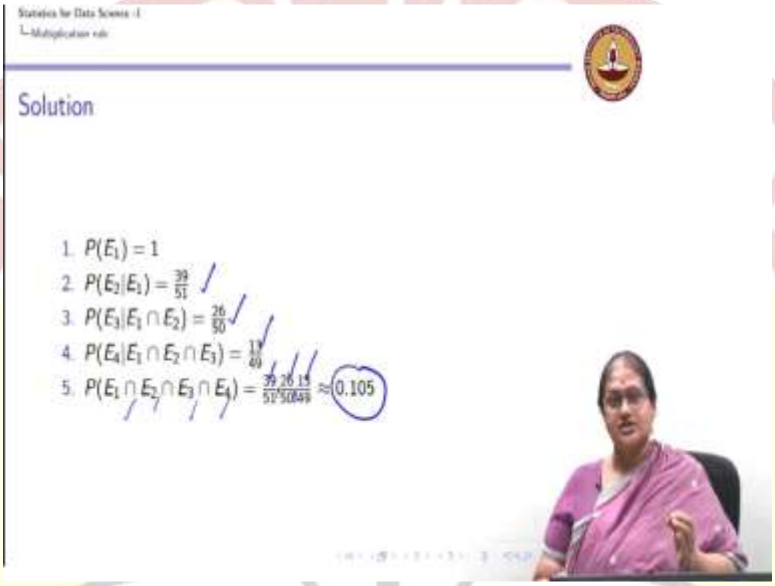
(Refer Slide Time: 22:27)



Similarly, I can argue that $P(E_4|E_1 \cap E_2 \cap E_3) = \frac{13}{49}$. Now I have 49. So, I already I am fixing ace of spades, ace of heart, ace of diamond, so I have 49 cards that are available. Again 12×3 which is 36 ways I can actually choose these cards, the remaining, so for each of these piles I have 12 + 12 + 12 which is 36 cards which can be chosen from the 49 cards in 36/49 ways.

I do not want my cards to be any one of these 36/49 choices. So, I look at $1-\frac{36}{49}$ which is given by the number here. So, we have computed all the, whatever we needed to compute $P(E_1 \cap E_2 \cap E_3 \cap E_4)$.

(Refer Slide Time: 23:41)



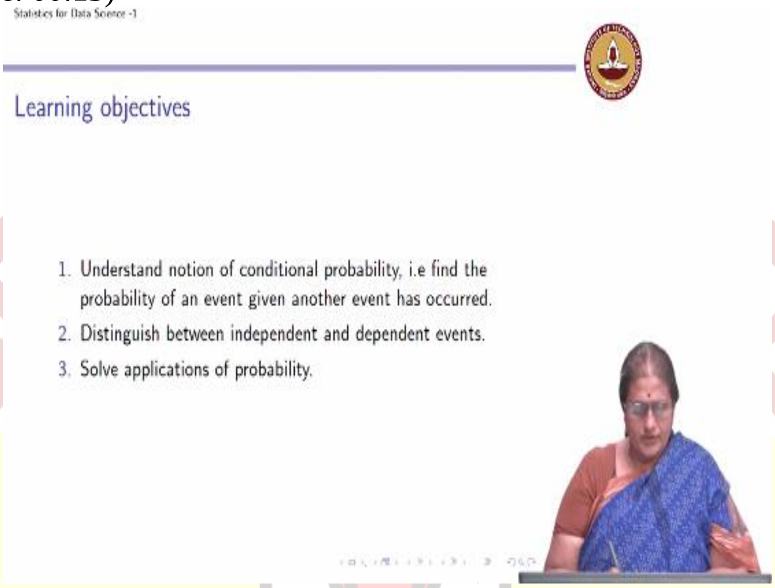
We just plug in these values to get $P(E_1 \cap E_2 \cap E_3 \cap E_4) = \frac{39}{51} \times \frac{26}{50} \times \frac{13}{49} \approx 0.105$

(Refer Slide Time: 24:04)

So, what we have learned so far in this section was, what is a multiplication rule, when I have two events I know $P(E_1 \cap E_2) = P(E_1) \times P(E_2|E_1)$. We extend this to give the generalized multiplication rule which is for any arbitrary n events, it is $P(E_1 \cap E_2 \cap \dots \dots \cap E_n) = P(E_1) \times P(E_2|E_1) \times P(E_3|E_1 \cap E_2) \dots \dots . . \times P(E_n|E_1 \cap E_2 \cap \dots \dots \cap E_{n-1})$. And we further use this generalized multiplication rule to compute probability of intersection of events.

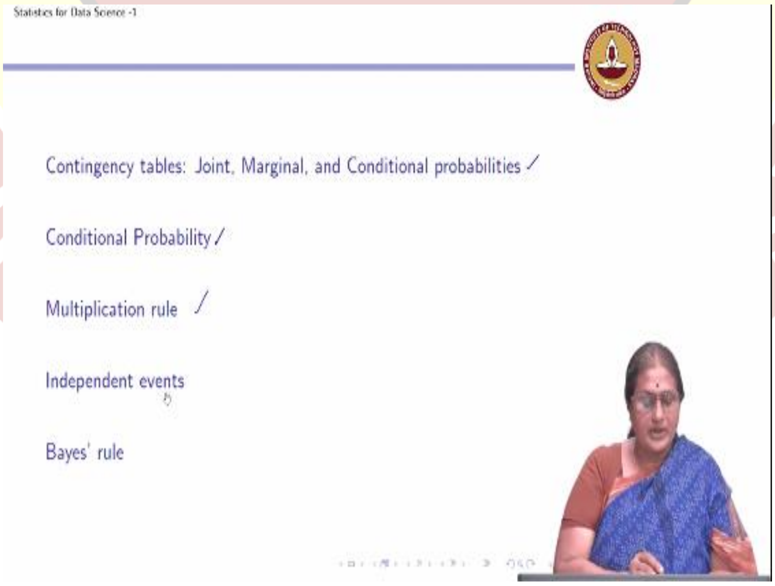
Statistics for Data Science-1
Professor Usha Mohan
Department of Management Studies
Indian Institute of Technology, Madras
Lecture – 7.4
Conditional Probability - Independent events

(Refer Slide Time: 00:13)



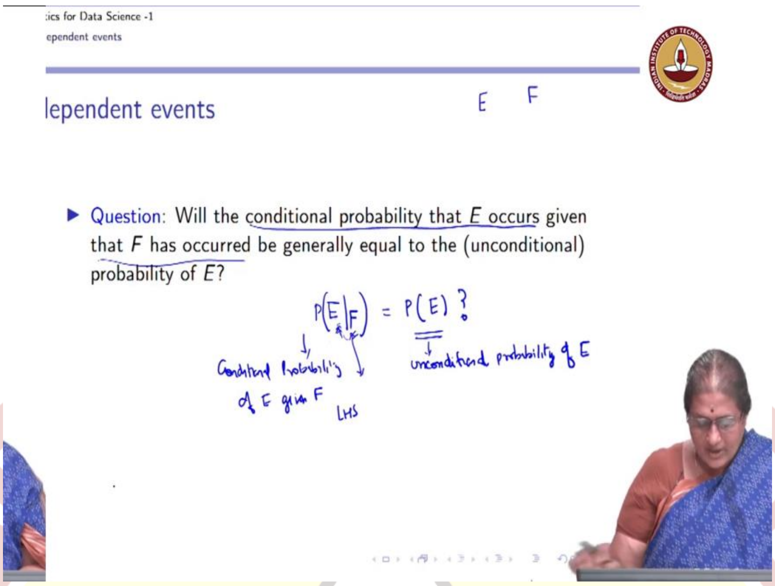
In this lecture, we are going to look at an extremely important concept which is called independence of events.

(Refer Slide Time: 00:23)



So, if you recall that we have already seen what are joint, marginal and conditional probabilities. We introduced the notion of conditional probabilities and the multiplication rules so far. So, today we are going to look at what do we mean by independent events?

(Refer Slide Time: 00:40)



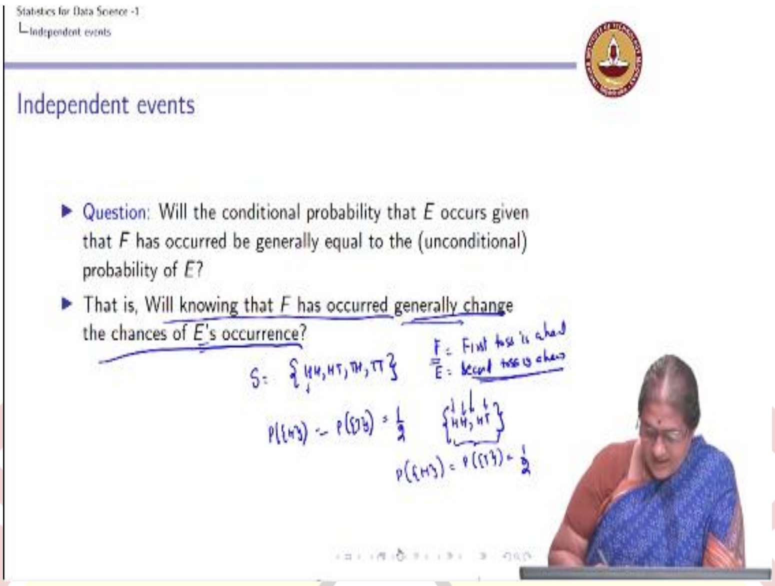
So, let us begin by asking a question, will the conditional probability of an event. So suppose I have two events let me call the events E and F. These are the two events defined on the same sample space. So the question we are asking is will the conditional probability that E occurs given that F has occurred and we have already seen that I am writing this as E occurs given F has occurred.

So, the question that is being asked is will the conditional probability of E given F has occurred be equal to the unconditional probability of E. So this is what we refer to as the unconditional probability of E and this is the conditional probability of E occurring given F occurring. So the question is will this situation happen generally or will it not happen so that is the question we are asking.

Now, let us understand what the left hand side means? It says that will the conditionality probability of E given F occurring be equal to the unconditional probability. So what we are asking is when will they if the conditional probability equals to the unconditional probability then it means that E does not depend on F. In other words, no matter what is this occurrence or non occurrence of this event F is unaffected because of it.

So, we are introducing a notion of whether there is any dependence of this event occurring given the information F has occurred or not. So, this is the question that we are going to answer and that is being answered by the following.

(Refer Slide Time: 03:09)

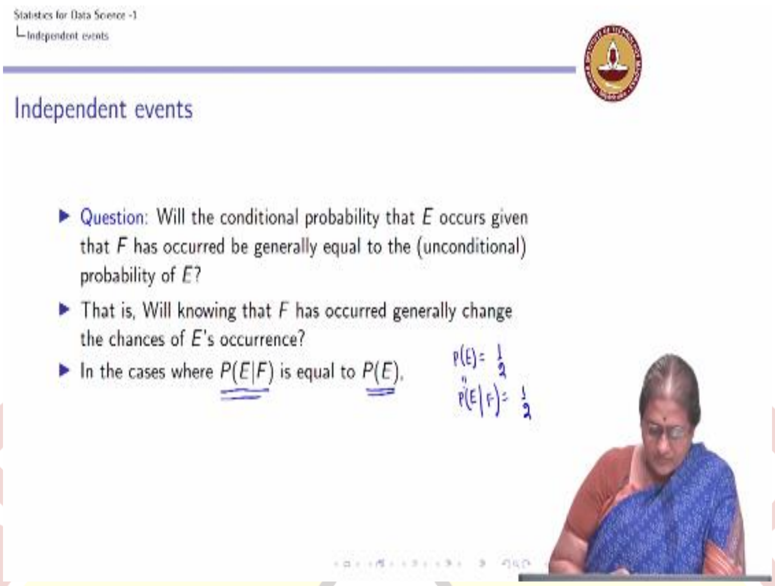


This is the question we are asking again is will knowing that F has occurred generally changed the chances of $E's$ occurrence. So, let us look at a simple example again I toss a coin twice I know my experiment is tossing a coin twice and this is my sample space. Now suppose I do not talk about probability or sample spaces I am just tossing a coin twice. Now the first toss is a head.

Now, the question we are asking is suppose I define an event E to be first toss is a head or let me define event F to be first toss is a head and event E to be second toss is a head. So the question we are asking is will the knowledge of the fact that the first toss is a head generally changed the chances of $E's$ occurrence in the sense that if I know that the first toss was a head will the chance of me getting a head or a tail in the second toss change.

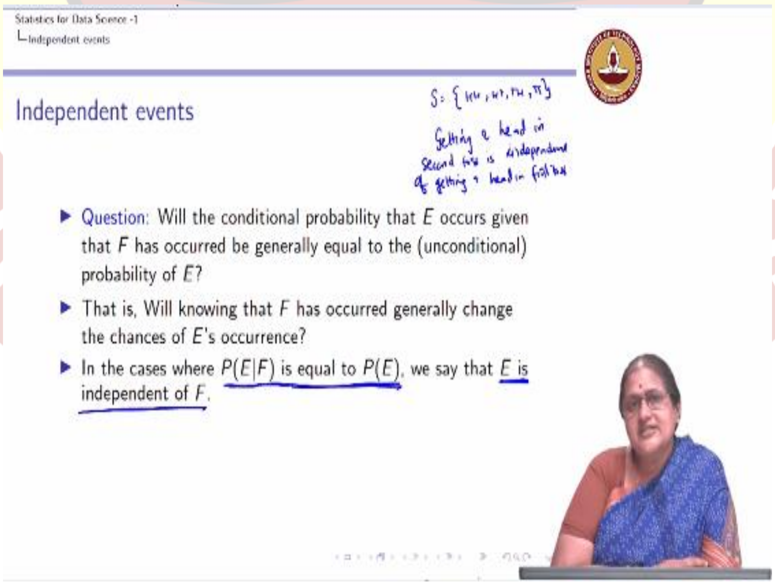
So, we know that independently the probability of you getting a head or a probability of getting a tail in a fair coin is equal to half. Now if you look at this, if my toss is a head I know a head has happened. So this is my reduce sample space. Now in this sample space you again see the probability of getting a head or a tail is the same. In other words, you can see that the chance or the information knowing that I had a head in my first toss did not change the chance of a head in the second toss.

(Refer Slide Time: 05:21)



So, the question that is being asked is in general do we think that where are cases, are there cases where the conditional probability equals the unconditional probability. So we have just seen a case where probability of getting a tail or a head is equal to half and probability of getting head in the second toss given you have got a head in the first toss is also equal to half and these two are equal.

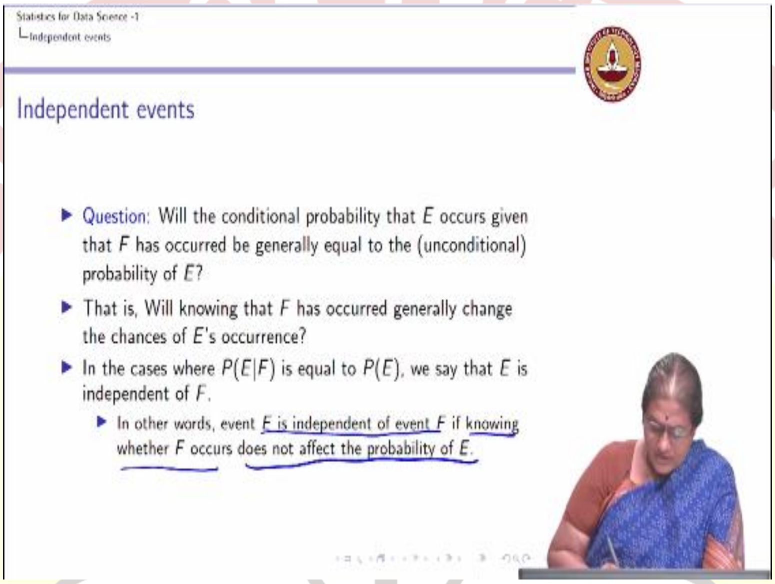
(Refer Slide Time: 05:58)



So, in cases where the conditional probability is equal to the unconditional probability we say that E is independent of F. I repeat in cases where the conditional probability is equal to the unconditional probability we say that E is independent of F. So in our coin tossing example you can see that getting a head in the second toss, getting a head in second toss is independent of getting a head in first toss.

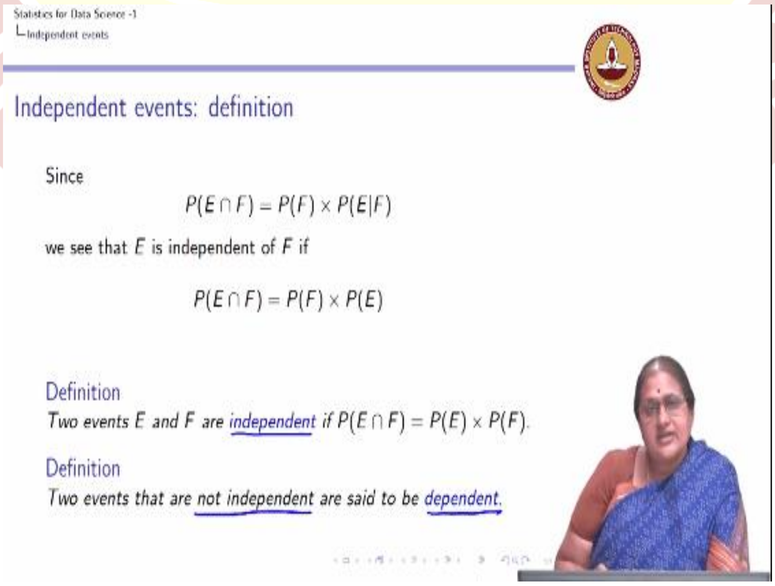
In other words, whether I got a head or tail in the first toss would not affect my chances of getting a head or tail in the second toss. So in this case we say the events are independent events.

(Refer Slide Time: 07:08)



So, in other words we say that E is independent of an event if knowing whether F occurs does not affect the probability of E. So this is the notion of independence of events.

(Refer Slide Time: 07:28)



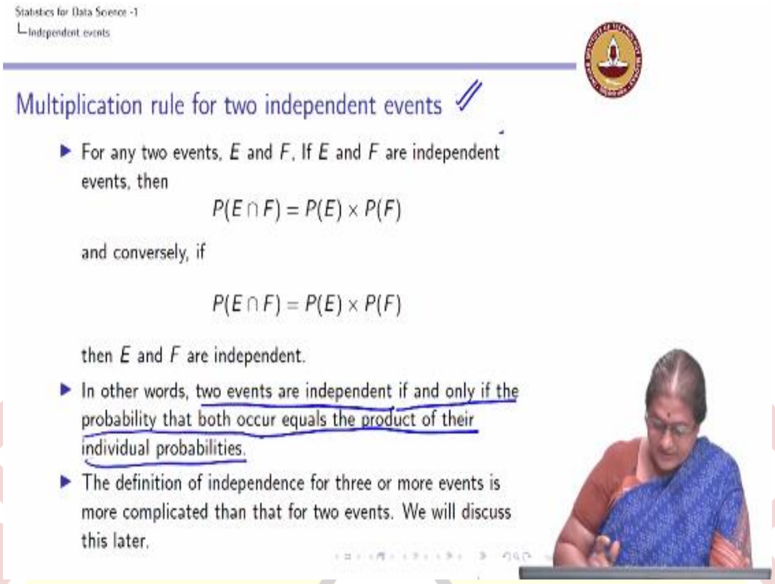
So, let us look at a formal definition to it. So since again from my multiplication rule I know probability given two events E and F I know the multiplication rule says that the probability of the intersection is $P(E) \times P(F)$. This is from my definition of conditional probability which says the $P(E|F) = \frac{P(E \cap F)}{P(F)}$ where $P(F) > 0$. I get $P(E \cap F) = P(F).P(E|F)$.

Now, E is independent of F, I said that the conditional probability is equal to the unconditional probability. Hence, we get $P(E \cap F) = P(F)P(E)$ which is nothing, but the product of the probabilities. Recall in the addition rule when I had disjoint a mutually exclusive events, recall in the addition rule when I had disjoint a mutually exclusive events the probability of the union was the sum of the probabilities.

Similarly, when I have independent events, probability of the intersection is product of the probabilities of the events. Hence I can pose my formal definition which says that two events E and F are independent if the probability of intersection is equal to the product of probabilities.

I repeat two events E and F are independent if the probability of the intersection of these two events is equal to the product of the probabilities. So once we know what is an independent event we are now in a position to define what is a dependent event. So two events that are not independent are said to be dependent.

(Refer Slide Time: 09:47)



So, this notion of independence is extremely important. So recall the multiplication rule for two events are $P(E \cap F) = P(E|F)P(F)$ where $P(F)$>0. So now for two independent events I will have $P(E)P(F)$ so that is my multiplication rule says that for any two events if E and F are independent than probability of the intersection is equal to the product of the probabilities.

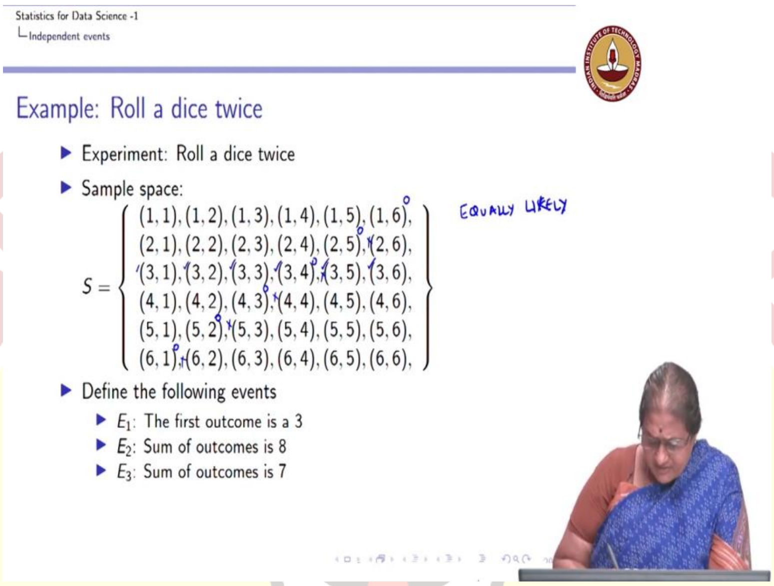
So, conversely it make sense for us to answer that if $P(E \cap F) = P(F)P(E)$ then can I say that E and F are independent that is the converse of the statement the answer is yes. Conversely, so this I have if E and F are independent then the probability of the intersection is product of probability. The converse is if probability of E intersection is product of the probabilities of events then E and F are independent.

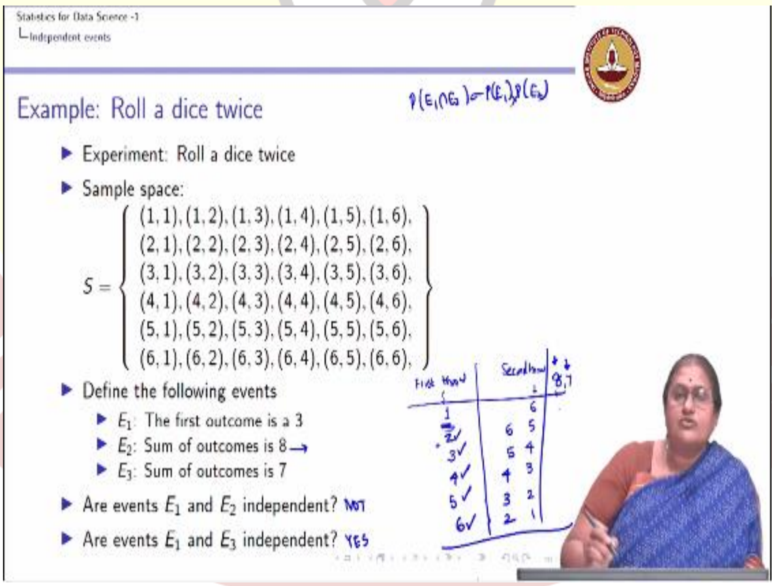
So, hence I have a basic rule which is referred to as a multiplication rule for two independent events which states that two events are independent if and only if and the if and only if comes from this if E and F are independent then this happens. If probability of the intersection is equal to the product of probability then E and F are independent. Hence this is nothing but a necessary and sufficient condition, but the word of caution is I am looking at two events.

So, if the probability of both of them occurring together equals the product of the individual probabilities then we state that the two events are independent. To look at more than two events the definition of independence where more than two events that is three or more events is slightly more complicated we will look at it later, but for now I want you all to understand that if two events are independent if and only if the probability of the intersection. In other words, the probability that both occur together equals the product of their individual probabilities. So this is what we refer to as multiplication rule for independent events.

Statistics for Data Science-1
Professor. Usha Mohan
Department of Management Studies
Indian Institute of Technology, Madras
Lecture 7.5
Conditional Probability - Independent Events Examples

(Refer Slide Time: 00:14)





Now let us look at an example we have already established the sample space of the experiment of rolling a dice twice. We know that the sample space has 36 outcomes and these are the 36 outcomes that are listed in the sample space. Each one of them are equally likely that is another assumption we have and we also know it is a good assumption it is not a bad assumption to have.

So now let us define three events. E_1 is the event that the first outcome is the 3 so we can see that this outcome, this outcome, this outcome, this outcome, this outcome and this outcome namely (3,1), (3, 2), (3,3), (3, 4), (3, 5), (3, 6) are the outcomes in my event E_1. E_2 is the outcome, sum of outcomes is a 8. So now I can see that let me denote it by cross 5+3, 4+4, 3+5, 2+ 6 and 6+ 2.

These are the outcomes that satisfy this and they are in this event E_2. Now when I look at sum of outcomes as a 7 I see the outcomes I put a circle here (1,0), (2,5), (3,4), (4,3), (5,2) and (6,1) these are the outcomes in my event E_3, but now having defined this events the question we are asking is are E_1 and E_2 independent and are E_1 and E_3 independent. So before we go to verify whether E_1 and E_2 and E_1 and E_3 are independent events.

Let us look at the intuition behind independence. Now E_2 is an event where the sum of outcome is 8. Now let us write down the following the first throw of a dice and the second throw of a dice. Now to get a sum of 8 if the first row was a 1 I cannot get a sum of 8 because the maximum that I can get in a second throw is 6 and 1 plus 6 is 8. So if the first throw is a 1 I cannot get a sum of 8.

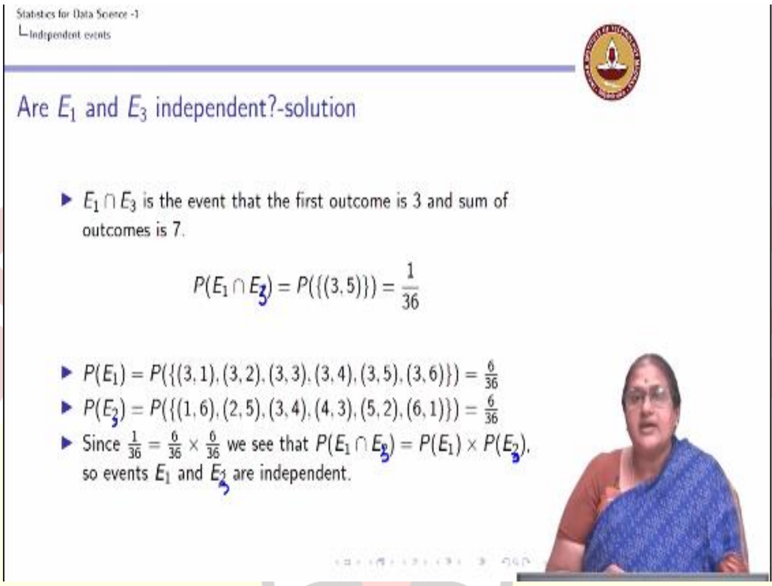
The first row is a 2 I need to have a 6, 3, 5, 4, 4, 5, 3 and 6, 2. In a sense to get a sum of 8 I need the first throw to be either 2, 3, 4, 5 or 6 to get a sum of 8. Now let us repeat the experiment with 7. So if I have sum 7 written here the first throw is a 1 then if the second throw is a 6 I get a sum of 7 5 I get a sum of 7 3 plus 4 is a 7, 4 plus 3 is a 7, 5 + 2 is a 7 and 6 + 1 is a 7.

So to get a sum of 7 I see that the first throw can be of these 6 values and I get a sum of 7 which is possible with this second throw whereas to get a sum of 8 we saw that if the first throw was a 1 it is impossible to get a sum of 8. In other words to get a sum of 8, the sum 8 is dependent on what was your first throw whereas the sum 7 was independent of what was your first throw so this is the intuition.

So sum to get an 8 I could not have got a sum of 8 if my first throw was a 1 whereas a sum 7 is independent even if it was a 1 and I throw a 6 I get a sum of 7 if it was a 2 and I throw a 5 so it is independent of whatever was my first throw there is a chance of me a sum of 7 whereas to get a sum of 8 I needed the first throw to be either a 2 or a 3 or a 4 or a 5 or a 6. So intuitively we can say that E_1 and E_2 are not independent.

Whereas E_1 and E_3 are independent events. Now let us apply the formula to actually verify it. Recall the formula sates that if E_1 and E_2 are independent then the probability of the intersection has to be the product of the probability so this is what we have to check.

(Refer Slide Time: 05:26)

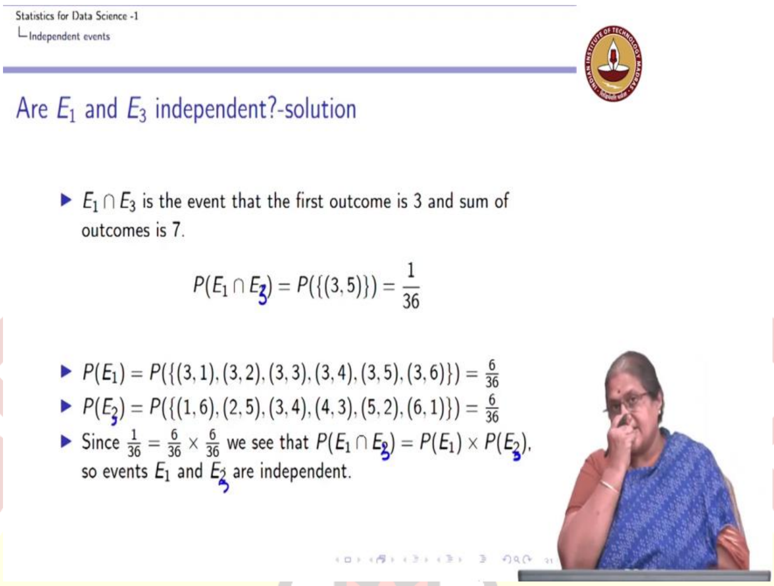


So let us go to the first thing E_1 and E_2 are independent. So I know $E_1 \cap E_2$ is a event that the first outcome is a 3 and the sum of outcomes as a 8 that is how I can articulate the event $E_1 \cap E_2$. I know that only this outcome and since all the outcomes are equally likely I know the probability of this outcome alone is equal to 1 by 36 and I know that is what I have as the probability of $E_1 \cap E_2$ which is $\frac{1}{36}$.

Now probability of E_1, E_1 is the event that the first outcome is a 3 again we saw that the outcomes the first outcomes is a 3 has these 6 outcomes where the outcome is a 3. So my E_1 event has the outcomes the outcomes of the E_1 event is going to be (3,1), (3,2), (3,3), (3,4), (3,5) and (3,6) again using that all of them are equally likely and the probability of events I get probability of E_1 is $\frac{6}{36}$.

Probability of E_2 where the sum of outcomes is a 8 we also saw that there are 5 outcomes which give an outcome of 8 hence probability of E_2 equal to is $\frac{5}{36}$. Now you can see that is $\frac{1}{36}$ is not equal to is $\frac{6}{36} \times \frac{5}{36}$. Hence, again using the if and only if condition we can conclude that probability of E_1 intersection E_2 is not equal to the product of probabilities hence E_1 and E_2 are not independent events.

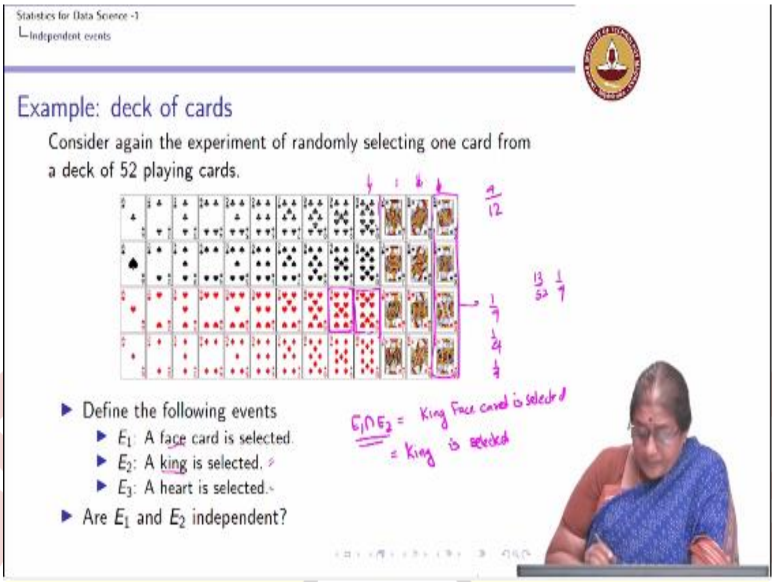
(Refer Slide Time: 7:34)



So the next question we are asking is are E_1 and E_3 independent? So E_1 intersection E_3 is the events that the first outcome is a 3 and sum of outcomes is a 4 so this is probability of 3, 4 which is is —. So I have probability of E_1 intersection E_3 is probability of 3,4 which is is —. So I have probability of E_1 as before which is equal to is — I can see that this is equal to is –.

I have probability of E_3 is probability of (1,6), (2,5), (3,4), (4,3), (5,2) and (6,1) which is equal to is — which is also equal to is –. Hence I have $P(E_1 \cap E_3)$ =$P(E_1) \times P(E_3)$ in this case which tells me that probability of, hence $E_1 \cap E_3$ is probability of E_1 into probability of E_3. Hence the events E_1 and E_3 are independent events.

(Refer Slide Time: 08:45)



Now let us go and apply this notion of independence to our cards or deck of cards example. So again consider the experiment of randomly selecting one card from this deck of 52 playing cards. Let me define the following events E_1 is a face card is selected. What is a face card? Face card is a card which has a face on it so I have 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 of these cards which are referred to as face cards.

These cards which has a face namely jack, queen or king are the cards that are referred to as a face card. So, a face card is selected you can see that I have 12 cards which gives me probability of E_1 is 12/52. Now probability of E_2 is the event that a king is selected again you can see that the event a king is selected I have 4 kings and the chance of choosing a king from 52 card is going to be 4/52.

A heart so you can see an event an heart is selected I have here 13 hearts so the probability or the chance that I am selecting a heart probability of E_3 is is —. So these are the actual probabilities which I can have. Now the question that is being asked is are E_1 and E_2 independent events. So now let us understand what is $E_1 \cap E_2$. Now E_1 is a face card E_2 is a king card.

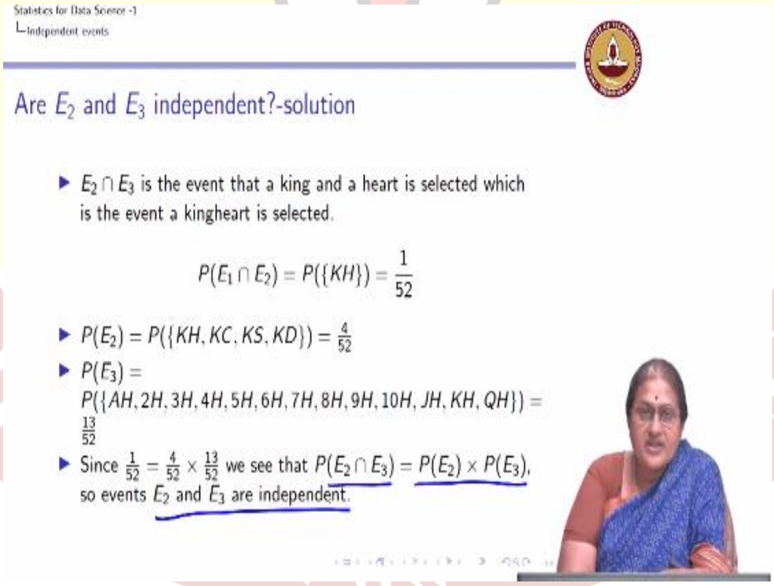
So $E_1 \cap E_2$ is the event that a king face card is selected which is equivalent to the probability a king is selected. Similarly, E_2 and E_3 what does the event $E_2 \cap E_3$? E_2 is a king E_3 is a heart so $E_2 \cap E_3$ is the event a king and heart. In other words, a king of hearts is selected. So now the question we are asking first question are E_1 and E_2 independent events?

So now let us look at the first question which is asking about $E_1 \cap E_2$ that is knowing that a face card is selected will that affect my chance of having a king. So you can see that there are 12 face cards. Now if I know that a face card is selected then I know that the chance of getting a king from these 12 cards is is — intuitive explanation, but whereas if I know that a king is selected the chance of you having a heart is just is –.

Now the chance of you having a heart from these 52 cards is also is — which is is –. Now this whether you choose a king or a queen again a chance of you getting a heart if you had selected a queen is also is – whether jack is also is –, 10 also the chance of getting a heart is is –, 9 is –. So you see that the chance of you getting a heart is not dependent on whatever has been the card.

Whereas the chance of you getting a king changes if you know the information that what you had taken was a face card.

(Refer Slide Time: 13:35)



So now us move forward and verify this. So $E_1 \cap E_2$ is the event that a face card and a king is selected. So we know that this probability is nothing, but the probability of having a king. We know there are 4 king card which is king of hearts, king of clubs, king of spades and king of diamonds. Hence, the probability of a face card and the king is the same as the probability of a king which is 4/52. So what is the probability of a face card?

Again I know there are 12 face cards, jack heart, jack clubs, jack spades, jack diamond, king hearts, king clubs, king space, king diamonds, queen hearts, queen clubs, queen spades and queen diamonds. Hence, the probability of this event E_1 that is getting a face card is 12/52, probability of getting a king is 4/52 you can verify $— \neq — \times —$.

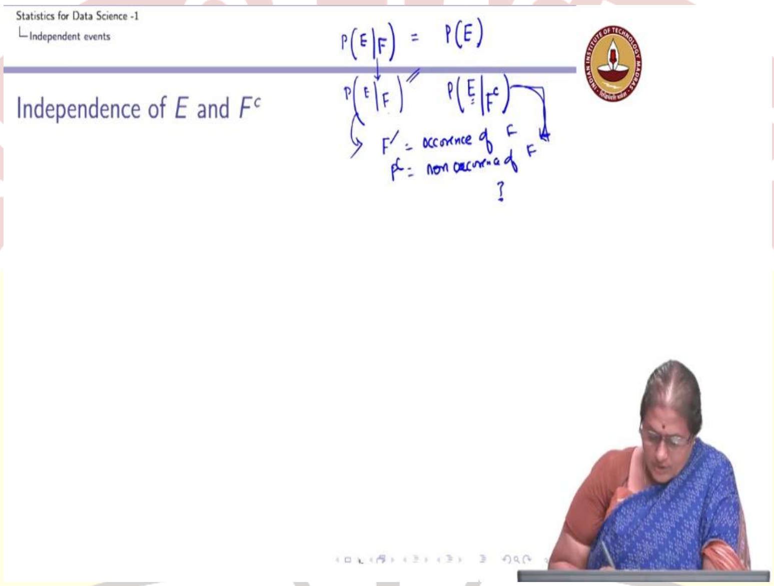
Hence, my $P(E_1 \cap E_2) \neq P(E_1) \times P(E_2)$. So, E_1 and E_2 are not independent of each other something which we could explain intuitively. Now let us look at the case of E_2 and E_3. Now what is event E_2? E_2 is the event that there is a king, E_3 is the event of a heart so $E_2 \cap E_3$ is the event of a king and a heart. In other words you are selecting a king of hearts card.

So probability of $E_1 \cap E_2$ is just the probability of choosing the king of heart cards which is 1/52 considering all of them are equally likely events. Now probability of choosing a heart. So probability of choosing a king is 4/52, probability of choosing a heart is I can have a 2, 3,4, 5, 6, 7, 8, 9, 10 jack king or queen so 13 possible choices it is 13/52. We observe $— \quad — \quad —$.

So this is 13 so I get this is 1/52 hence I can say that since). E_2 and E_3 are indeed independent events. This is something which we even checked intuitively.

Statistics for Data Science – 1
Professor. Usha Mohan
Department of Management Studies
Indian Institute of Technology, Madras
Lecture No. 7.6
Conditional Probability - Independent events: properties

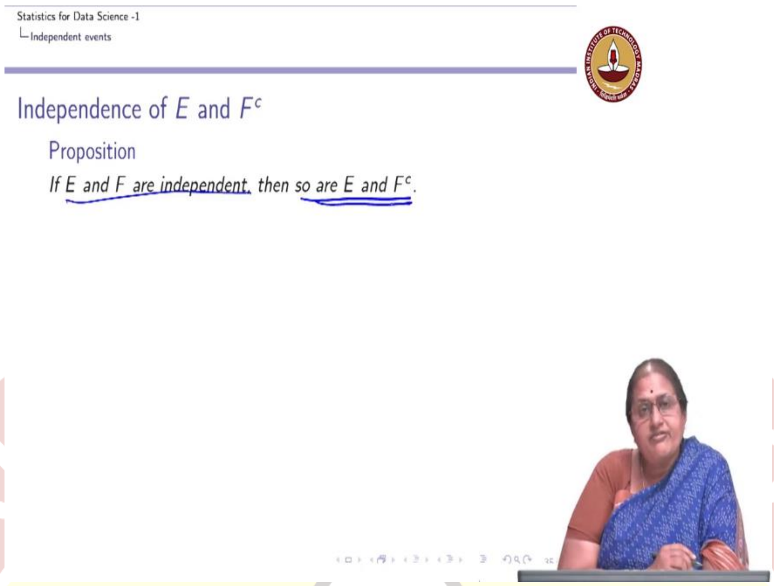
(Refer Slide Time: 0:14)

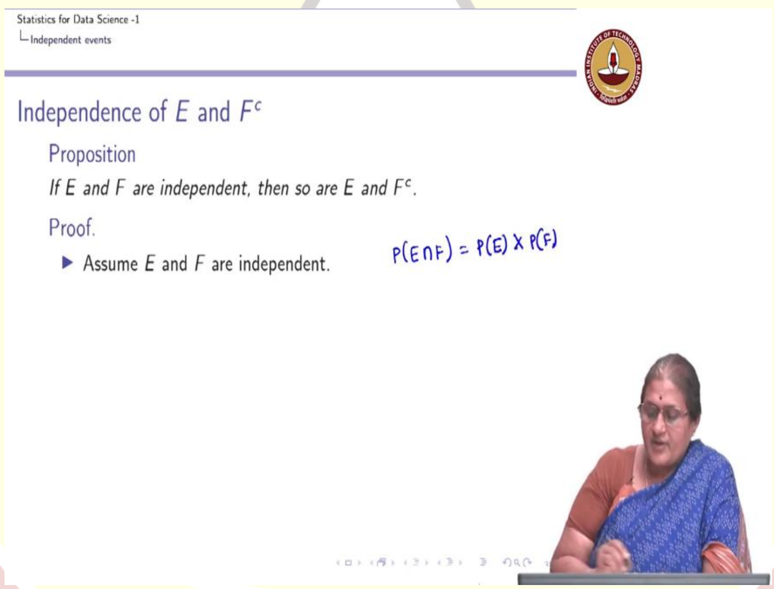


So, the question now we want to ask is, now independence means that the conditional probability of a event E happening conditioned on the occurrence of F is equal to the unconditional probability of an event happening. Recall this says that E conditioned on the occurrence of F, so the natural question to ask is what would happen with if E is conditioned on occurrence of F is does not affect the unconditional probability of E, then what can you say about the probability of this event happening on the non occurrence of this event F.

Recall F is occurrence of event F, so F^c is non occurrence of event F, F is the occurrence of event F andF^c is the non occurrence of event F, so the question we are asking is if a event is conditioned on the occurrence of the event F and this is the event E conditioned on the nonoccurrence of event F. So, the question that is being asked is if E and F are independent can I say that E and F^c are also independent? That is the question, in other words if the occurrence of E is independent of the occurrence of F can I say that the occurrence of E is independent of the non occurrence of F also? So, that is the question that is being asked.

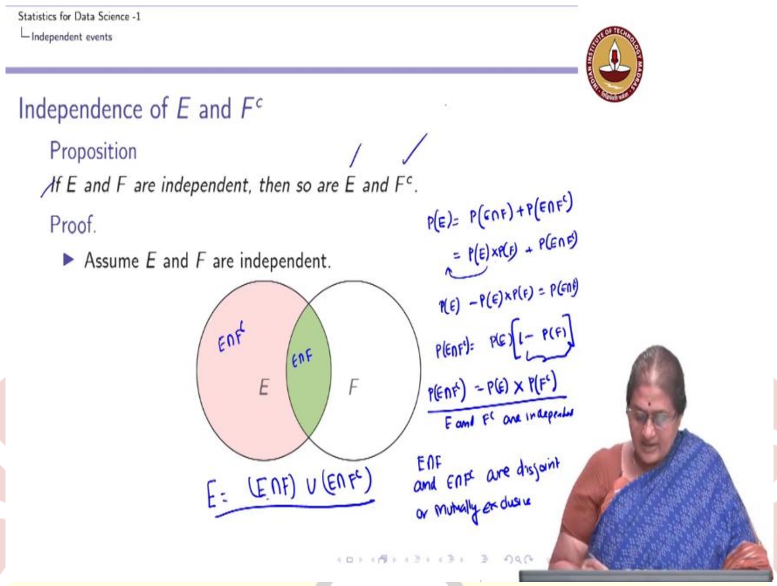
(Refer Slide Time: 2:04)





And the proposition is if E and F are independent then so are E and F^c. So, this requires a very simple proof. So, we can see that E and F are, let us assume E and F are independent, so if E and F are independent we have seen that $P(E \cap F) = P(E)P(F)$, so that is what we can see.

(Refer Slide Time: 2:35)

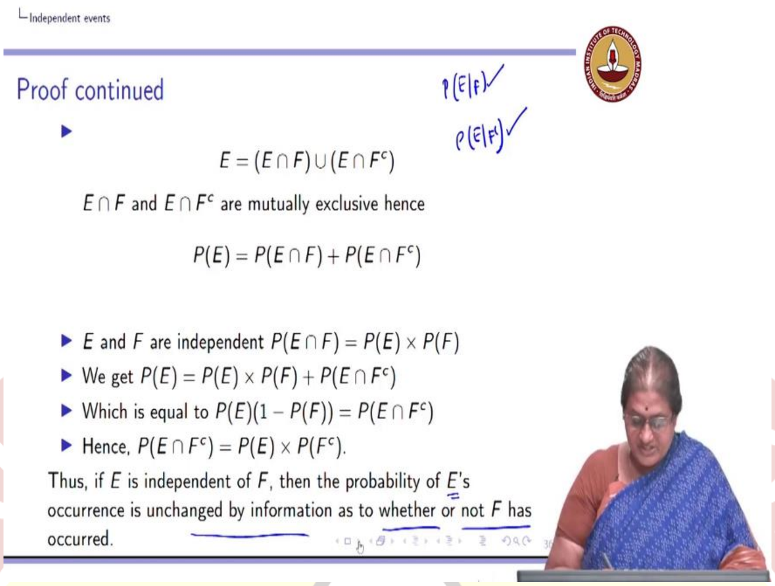


So, now you can look at this Venn diagram this green area is my$(E \cap F)$, I can write the pink area as $(E \cap F^c)$, so that I can write my E as $[(E \cap F) \cup (E \cap F^c)]$. Notice that $(E \cap F)$ and $(E \cap F^c)$ are disjoint or mutually exclusive.

Now, once I write E as $[(E \cap F) \cup (E \cap F^c)]$ I can apply my addition law to the disjoint sets to get $P(E) = [P(E \cap F) + P(E \cap F^c)]$. Now, I apply $[(E \cap F)]$, I apply my multiplication rule for independent events, $P(E \cap F) = P(E)P(F)$, I want to, I retain probability of $P(E \cap F^c) =$ the same way.

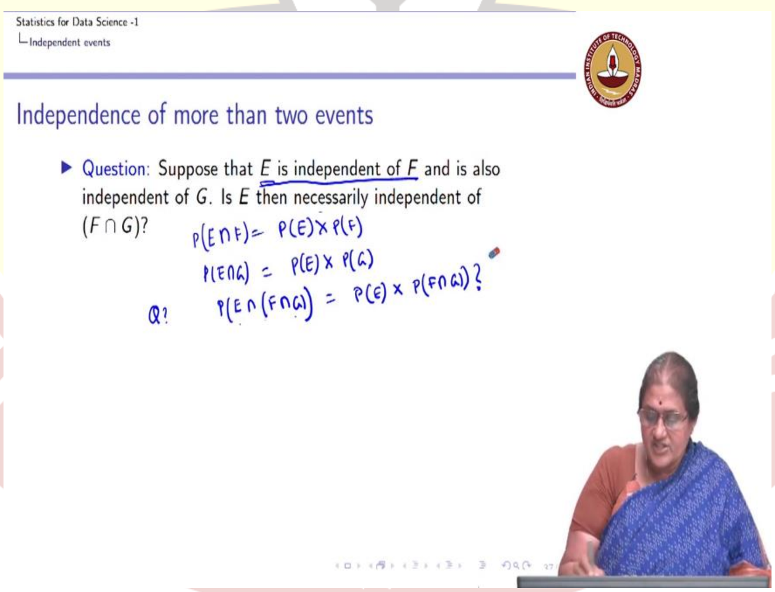
Now, I take this term to my left hand side, I have $P(E) - P(E)P(F) = P(E \cap F^c)$. Hence, I can write $P(E \cap F^c)$ is $P(E)[1 - P(F)]$. We know $1 - P(F)$ is $P(F^c)$ and hence I get $P(E \cap F^c) = P(E)P(F^c)$ and this tells me that E and F^c are independent. So, we can see that if E and F are independent so are E and F^c.

(Refer Slide Time: 5:01)



So, this is the proof which we have, we have just discussed, we have $P(E \cap F^c) = P(E)P(F^c)$. So, what this entails is the probability of E's occurrence is unchanged by the information as to whether F has occurred or F has not occurred. So, $P(E|F)$ or $P(E|F^c)$ whether F has occurred or F has not occurred if E and F are independent the probability of E's occurrence is unchanged, that is what this means.

(Refer Slide Time: 5:46)

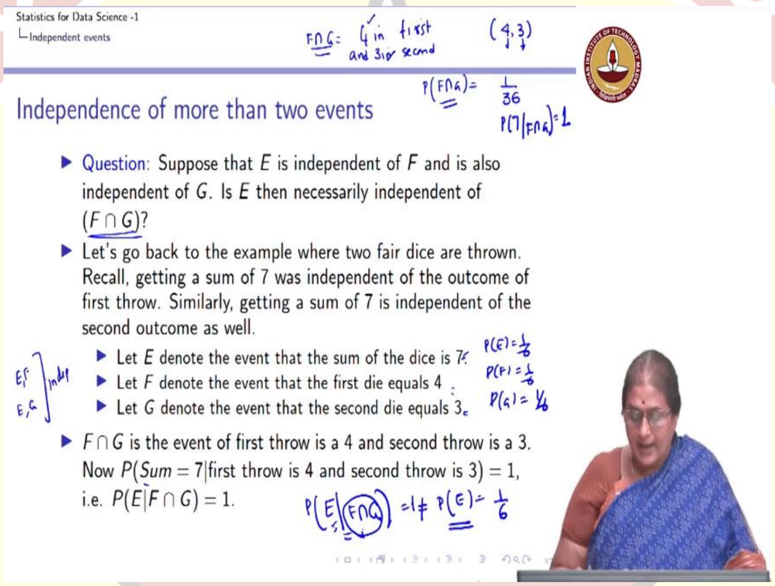


So, now we said that independence of 3 or more events is slightly more complicated than just discussing independence of 2 events. When we talked of independence of 2 events we said that , if I have 2 events probability of E and F, if probability of E is, if the probability of the intersection is equal to the product of probabilities then I can say E and F are independent and the converse is also true that if E and F are independent then the product; probability of the intersection as the product of the probabilities, this is what we have for 2 events.

Now, suppose I have, I want to extend this notion of multiplication or intersection of events to more than 2 events. So, let us, the question that is being asked is suppose E is independent of F, given that E is independent of F, so I know the $P(E \cap F) = P(E)P(F)$ this is what is given to us, E is also independent of G, so I have $P(E \cap G) = P(E)P(G)$, then the question we are asking is, is E necessarily independent of $F \cap G$.

So, I am asking if $P(E \cap (F \cap G)) = P(E)P(F \cap G)$, this is the question we are asking. So, let us look at it. So, this is an intersection E intersection F intersection G, this is the question we are asking. And we answer this question through a example, we are not going into the detailed mathematical implication here but we would like to establish the multiplication rule for more than 3 events through an example.

(Refer Slide Time: 8:01)



So, now let us recall this example. I rolled 2 fair dice, you also recall that we defined an event E, now let me define the event E is sum of 7 in the independent throw, so I am rolling a die twice and we saw that the sum of 7 probability of E was $1/6$ this is something which we saw, I had 6 out of 36 which is equal to $1/6$.

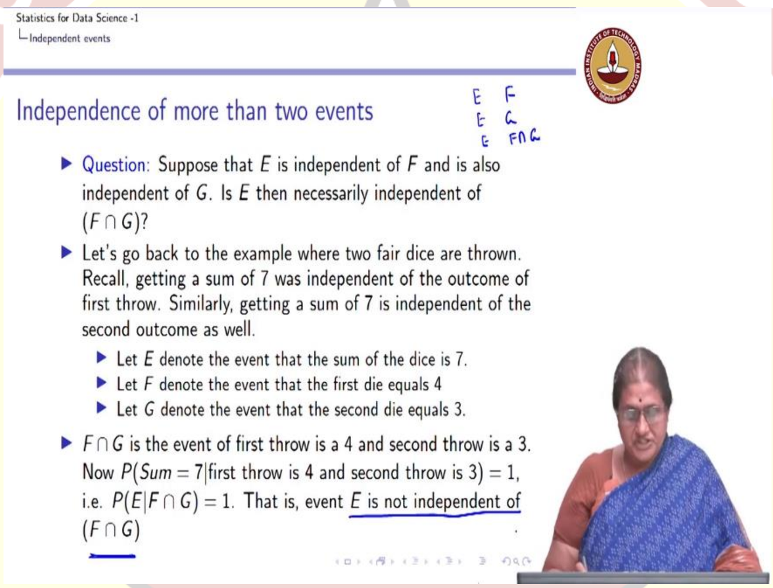
In other terms we saw that getting a sum of 7 is actually independent of whether your first throw was a 1 or a 2 or a 3 or a 4 or a 5 or a 6, the sum of 7 is independent of what your first throw was, I can use a similar logic to say that a sum of 7 is independent of what my second throw is also. So, if I define my events in the following way E is the event that the sum of dice is equal to 7 we know $P(E)$ is $1/6$, let me define F to be the event that the first dice equals a 4 again $P(E)$ is $1/6$, $P(F)$ is again 1 first, first dice is equal to 4 or the first outcome is a 4, again this is a $1/6$.

This G is the event that the second dice equals a 3, so $P(E)$is $1/6$, $P(F)$ is $1/6$, $P(G)$ is $1/6$, I know that E and F are independent because the sum is equal to 7 is independent of what was your first choice, the sum yet is equal to 7 is similarly independent of what is your second choice, so E and F are independent. Similarly E and G are also independent. Now, let us look at what is the event $(F \cap G)$.

So, now if I have defined all these events, the event$(F \cap G)$ is the event of getting a 4 in your first throw and a 3 in the second throw. So, the event $(F \cap G)$ G corresponds to the outcome 4 and 3 and I know that the $P(F \cap G)$is again $1/36$. This is the probability of F intersection G.

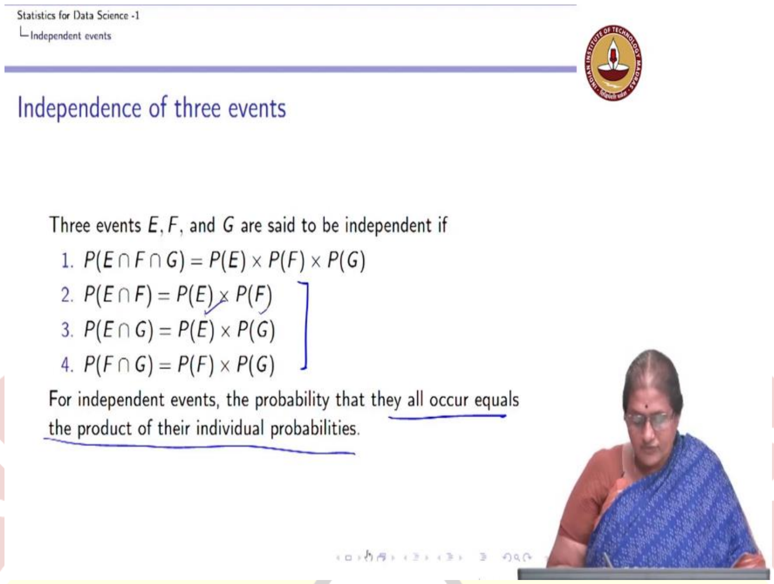
Now, the question that is being asked is what is the chance of E happening given the event $(F \cap G)$ has happened, I know that, if I know that this event $(F \cap G)$ has happened that is the first throw is a 4 and second throw is a 7 is 3, I know that the chance of me getting a 7 is given $(F \cap G)$ has happened is 1, given 4 and 3 has happened the sum equal to 7 is equal to 1. In other words, $P\big(E\big|(F \cap G)\big) = 1$. But I also know $P(E)$ is $1/\ 6$. So, we can see that the unconditional probability of E conditioned on this event F intersection G is not equal to the unconditional probability of E happening.

(Refer Slide Time: 12:07)



Hence, we can say that event E is not independent of event $(F \cap G)$. So, we have that, even though I have E is independent of F and E is independent of G, we have a condition where E is not independent of $(F \cap G)$.

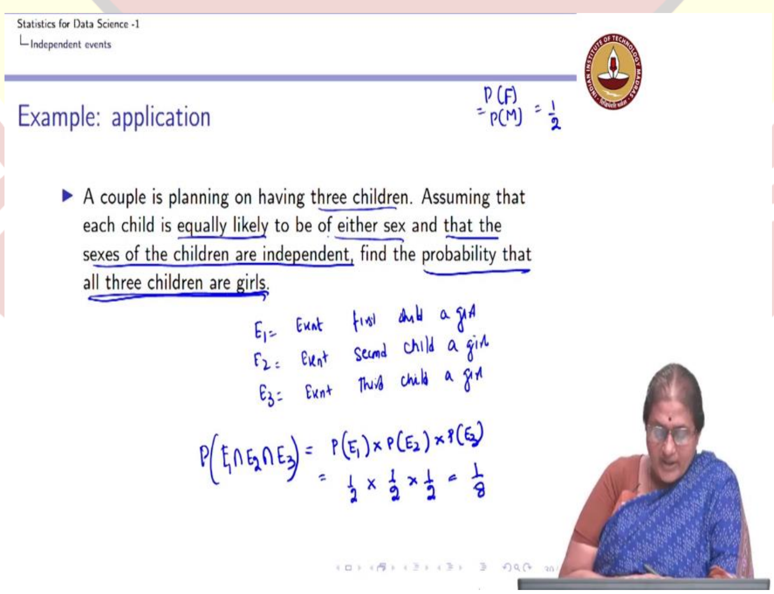
(Refer Slide Time: 12:33)



So, this helps us actually come up with the rule for 3 events to be independent, we are only stating the rule here, explaining and proving it is beyond the scope of this course but I prove, I state the independence of 3 events rule the following way. 3 events are said to be independent if probability of the intersection equal to the product of the probabilities, not only that I need to check the pairwise probabilities $P(E \cap F)$ is $P(E)P(F)$, $P(E \cap G)$ is $P(E)P(G)$, $P(F \cap G)$ is $P(F)P(G)$. If these 4 happen then I say the events E, F and G are independent events.

So, for independent events the probability that they all occur equals the probability of their individual probabilities and these probabilities which are looking at pairwise intersections that probability of pairwise intersection should be the product of the individual probabilities.

(Refer Slide Time: 13:52)

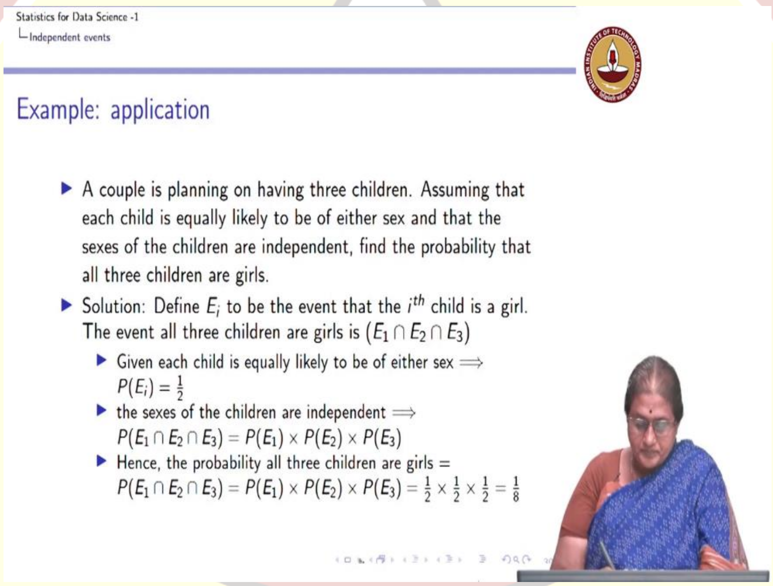


So, let us look at an application of the multiplication rule for more than 3 events. So, a couple is planning on having 3 children. Assuming that each child is likely to be of either sex, I am assuming female and male and the probability of female is equal to the probability of male equal to half that is what we mean by equally likely to be of either sex and that the sexes of the children are independent, then find the probability that all the 3 children are girls.

Now, let us define the events, the events are let us define E_1 to be the event that the first child is a girl, E_2 is the event second child is a girl, similarly, E_3 is the event third child is a girl. Now, the probability, now the event that all 3 children are girls are is the intersection, the event that all 3 children are girls are the intersection of these 3 events and what we require to find is the probability of this intersection.

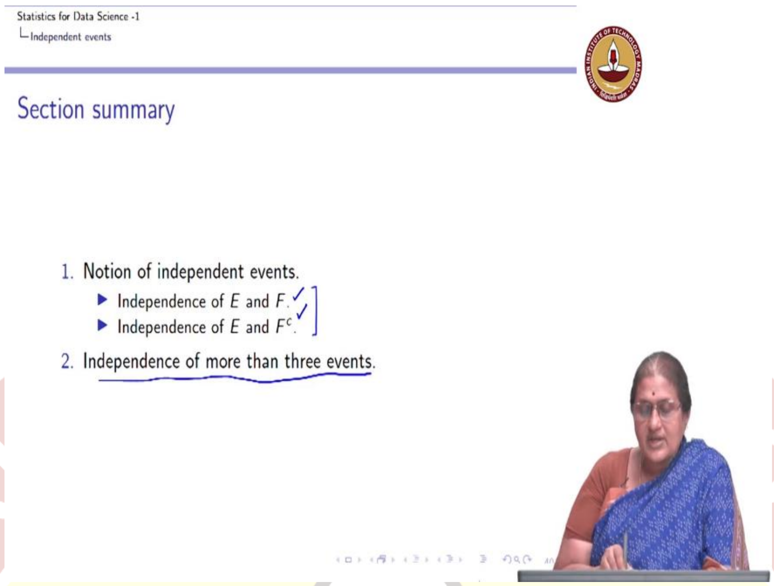
So, applying our multiplication rule I know that probability if they are independent I know this is equal to $P(E_1)P(E_2)P(E_3)$. What is $P(E_1)$? I know $P(E_1)$=1/2, $P(E_2)$=1/2, $P(E_3)$=1/2 giving me a probability of 1/8.

(Refer Slide Time: 16:02)



So, I have that the event the probability of all girls are independent is 1/2 into 1/2 into 1/2 which is equal to 1/8.

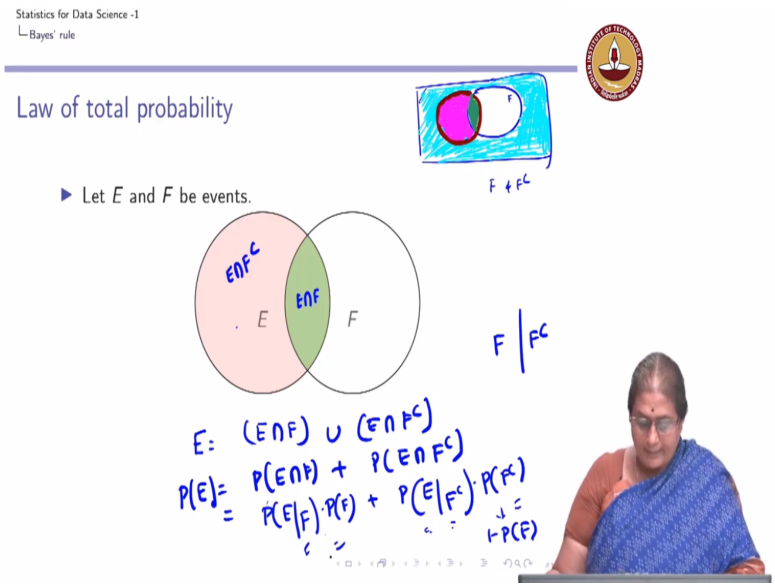
(Refer Slide Time: 16:19)



So, what we have seen so far is we introduce the important notion of independent events, we looked at the independence of E and F and E and F^c namely the occurrence of event E, given occurrence of event F and namely the independence of event E conditioned on occurrence of F, independence of event E conditioned on non occurrence of event F and we also extended the notion of independence to more than 3 events.

Statistics for Data Science - 1
Professor Usha Mohan
Department of Management Studies
Indian Institute of Technology, Madras
Lecture 7.7
Conditional Probability – Bayes' Rule

(Refer Slide Time: 00:14)



Now, we move on to discuss a very, very important concept, which is used in many applications of probability, which is referred to as Bayes' rule. But before we discuss about Bayes' rule, we introduced what we refer to as a Law of total probability. Now, let E and F be any events. In other words, if I have a recall, I have a sample space, suppose I have a event F, which is given here, suppose this is my event F, then this shaded region is my F complement.

In other words that I can mutually exclusive events. So, F and F complement are always disjoint, and mutually exclusive. So, let us share this event. So, let me use a different color to shade this F and F complement. So, I am just looking at F. So, I have this event F. So, let me look use a lighter color. So, I have F here, F is the white color here, and let me use a lighter color say this let me use a lighter color here, blue and that blue is going to be my F complement.

This light color blue is my F complement set. Now, let us look at an event E another event E that is happening, but the way this event E is happening is I am having information of this event F that has occurred this when I talk of this event F I know that there are mutually exclusive in the sense that F can occur or F complement can occur. Now, let me look at an event E that is happening. So, when I talk about an event E that is happening here.

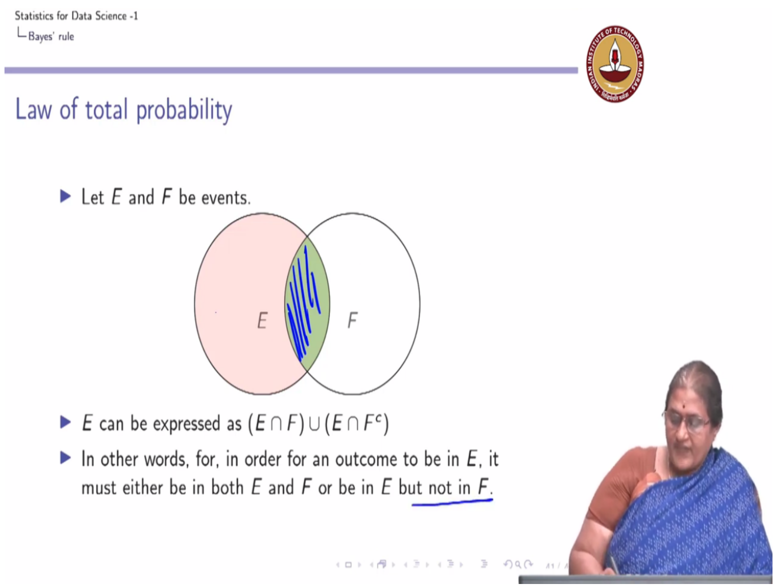
This so, this event E is happening here, so you can see the event E that is happening. Now, if you look at event E now you can look at this event E has an intersection with F which is given by this green portion. And the intersection of this event E with F complement is this pink portion, that is what we are trying to see.

So, I have a mutually exclusive F and F complement. When I looked at event E event E has an intersection with F which is referred to as this was this area which is E intersection F and E this portion is the intersection of E with F complement. So, I can actually write any event E as E intersection F union E intersection F complement.

So, what does the Law of total probability say is when I have mutually exclusive F and F complement from my sample space, then I can express my probability of E as probability of E intersection F plus probability of E intersection F complement this is from the addition rule. Now from the multiplication rule, I can state this as probability of E conditioned on F into probability of F plus probability of E conditioned on F complement into probability of F complement, recall probability of F complement is 1 minus probability of F.

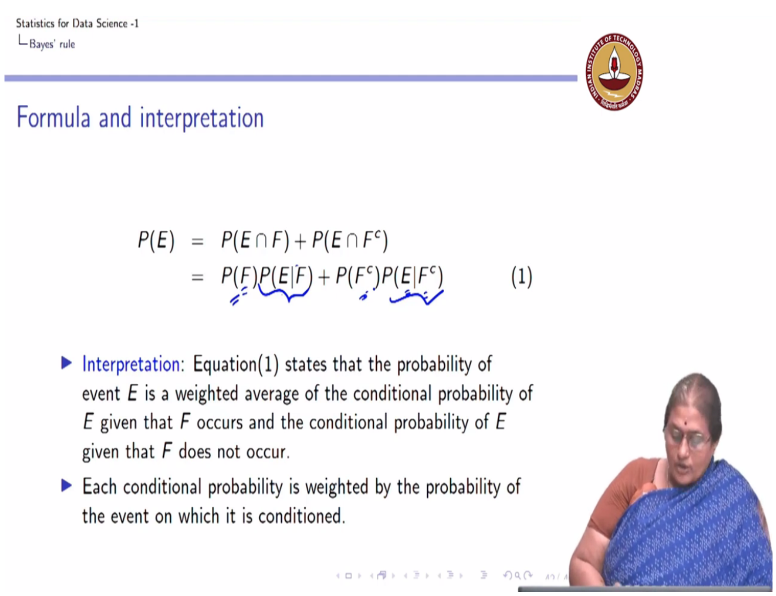
In other words, the way we can articulate it is the chance of this E happening is a weighted average of E happening, given F has happened. And the chance of E happening given F has not happened multiplied by the respective probability of F happening and F not happening.

(Refer Slide Time: 05:18)



And this is what is referred to as the Law of total probability, which says that in order for an outcome to be in E, it should be either in E and F, which is this green region, or it should be E and not an F.

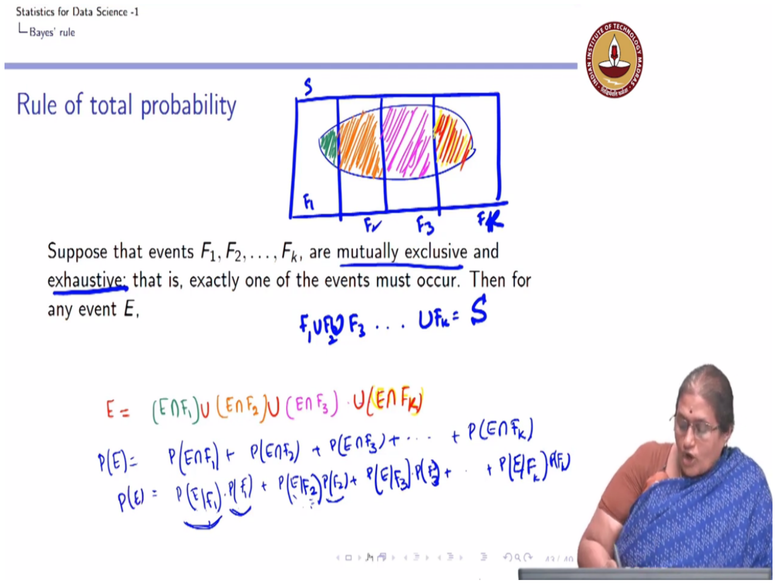
(Refer Slide Time: 05:37)



And I can tell that this is what we have got. Now, this expression probability of E equal to probability of F into probability of E given F plus probability of F complement into probability of

E given F complement is what is referred to as the Law of total probability. Where each conditional probability I have 2 conditional probabilities, it is weighted by the probability of the event on which it is conditioned, that is probability of E given F is conditioned on probability of F happening, probability of E conditioned on F complement is conditioned on probability of F not happening.

(Refer Slide Time: 06:22)

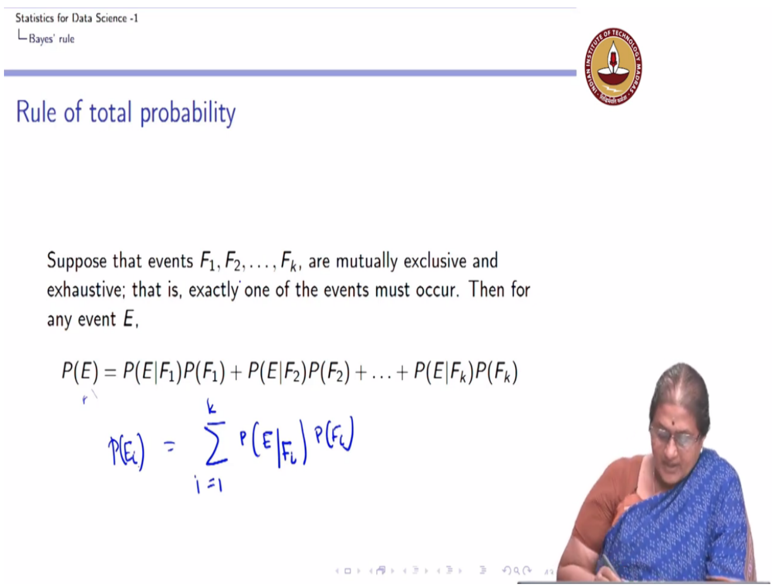


Now, we can actually extend this to many events. For example, again, let us go back to my sample space, I can partition my sample space into events F1, F2, F3 and Fk, such that they are mutually exclusive. So, you can see F1, F2, F3, Fk are disjoint, and they are exhaustive in the sense F1 union F2 union F3 union Fk should be my sample space, that is what I mean by exhaustive. Then for any event E so, let me draw an event E here, let this be my event E, and let me color it using a lighter color. So, this event E.

So, this green area is now E intersection F1. Similarly, the orange area is E intersection F2. I have the purple area is E intersection F3. Similarly, I have this area, the yellow area is E intersection Fk, yellow is not very visible, so, let me make it red. So, this is E intersection Fk, I can write my event E as the union of these disjoint sets. And now I can apply my addition rule which says probability of E is probability of E intersection F1 plus probability of E intersection F2 plus probability of E intersection F3 plus probability of E intersection Fk.

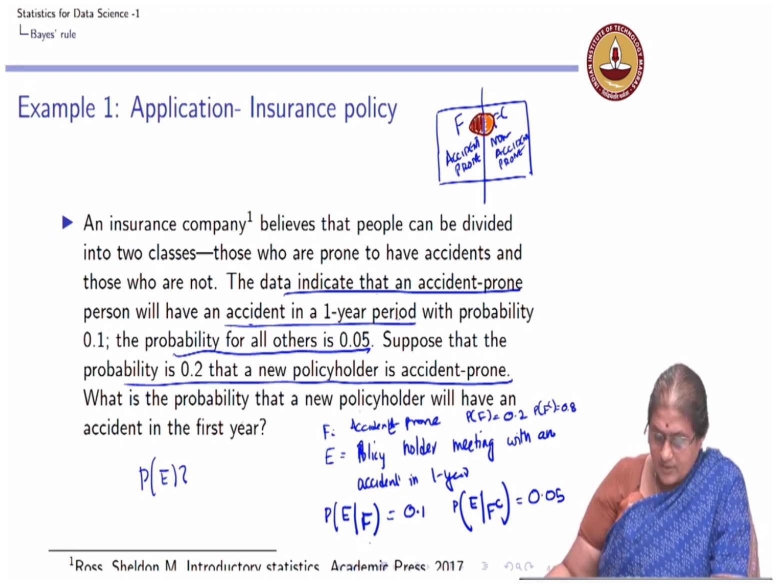
Applying my multiplication rule I get this is equal to probability E conditioned on F1 into probability of F1 plus probability E conditioned on F2 into probability of F2 plus probability E conditioned on F3 into probability of F3 plus probability of E conditioned on Fk into probability of Fk. So, we can see that each of these conditional probabilities of E happening weighted on the probability of the event it is conditioned on. Here again I have E given F2 into the weight of the probability of the event it is conditioned on which is F2. And this is what is referred to as the Law of total probability.

(Refer Slide Time: 09:40)



Which I can rewrite as the following I can rewrite this as summation i going from 1 to k probability E conditioned on Fi into probability of Fi. So, this is what is referred to as the rule of total probability or law of total probability, which states that suppose we have F1, F2, Fk, which are mutually exclusive and exhaustive, then for any event E, the probability of E happening is expressed as the weighted average of the conditional probabilities of E happening conditioned on these events.

(Refer Slide Time: 10:32)



So, now let us look at an Application of the Law of total probability. So, now, the application, an insurance company believes that people can be divided into two classes, those who are prone to have accidents and those who are not prone to have accidents. So, if my sample space is the entire population of people, the insurance company believes that I can divide this entire space into two classes, one class let me call it F is the accident prone and the other one is the non-accident prone. And we know that these two are disjoint, so I cannot have a person who is both accident prone and non accident prone.

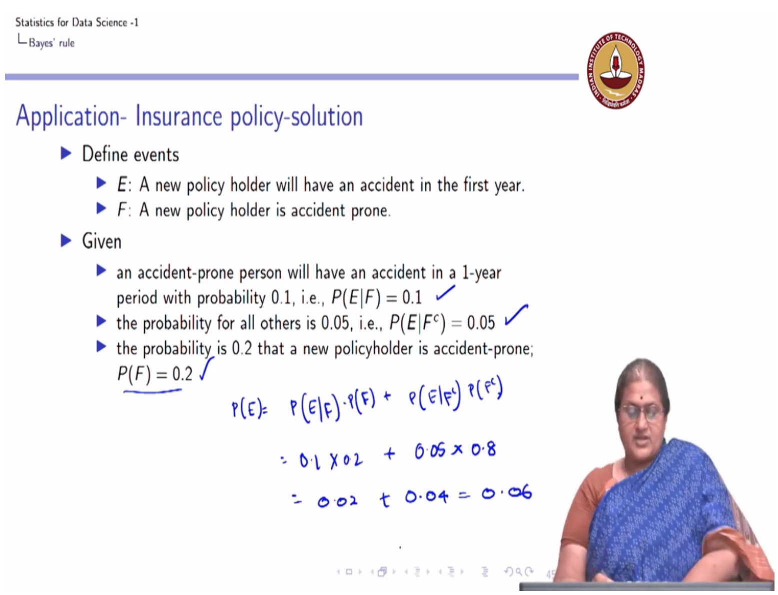
Now, the data indicates that the accident prone person will have an accident in a 1 year period with probability 0.1. So, I have another event that is being defined. And let me call that event to be E, which is the problem with the event is a person or a policy holder meeting with an accident in a 1-year period, or within a year. Now, what does this indicate? An accident prone person will have an accident with probability 0.1, there is a probability for all other is 0.05.

So what do we understand by this, the chance of a person having an accident within a year, given he belongs to the accident prone class, that is what this means that an accident prone person will have an accident within a 1 year period and this probability is given to be 0.1. Whereas the policyholder would meet with an accident so, the policyholder who is meeting with an accident also will belong to these two classes.

So, now my event E, which is the chance that a person meets with an accident within a year could happen in two ways. The person is meeting with an accident within a year, and is accident prone, or the person is meeting with an accident within a year, but as the person who comes from the non accident prone class. So, these are the probabilities that are given to us is this conditional probability is 0.1. And this conditional probability is 0.5.

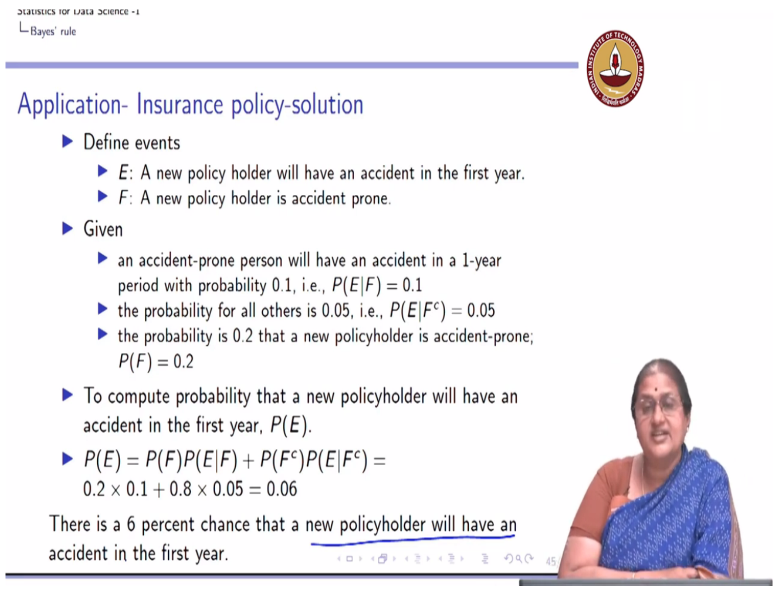
Suppose that the probability is 0.2 that a new policyholder is accident prone. So, we have defined F to be the probability of a person to be accident prone. And that probability we have is 0.2. Hence, probability of F complement is going to be 0.8. So, the question that is being asked is, what is the probability that a new policyholder will have an accident in the first year. So, that is for this probability of E. That is what is being asked. So, let us look at this solution. So, I have the following which is given to us. So, let us look at the solution. So, let us look at this solution in a more structured way. So, I have the following.

(Refer Slide Time: 14:30)



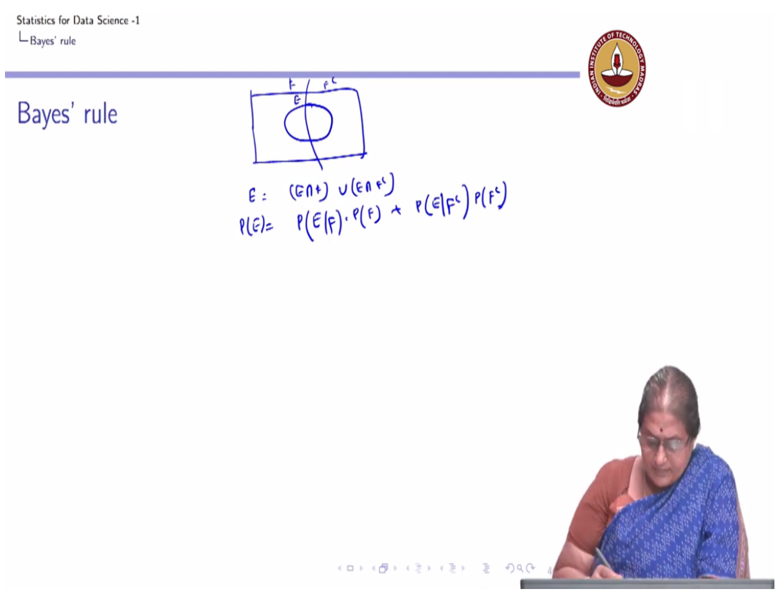
So, let us look at the solution. So again, we have defined as before, E is the event that a new policy holder will have an accident in the first year and F is the policy, a new policy holder is accident prone. So these are the probabilities given to us as before, probability E given us is 0.1 whereas probability E given F complement is 0.05. And probability of F is 0.2. So, my law of total Probability tells the probability of E is equal to probability of E given F into probability of F plus probability of E given F complement into probability of F complement, which is equal to 0.1 into 0.2 plus 0.05 into 1 minus 0.2 which is 0.8. So, I get a 0.02 plus a 0.04, which is equal to 0.06.

(Refer Slide Time: 15:35)



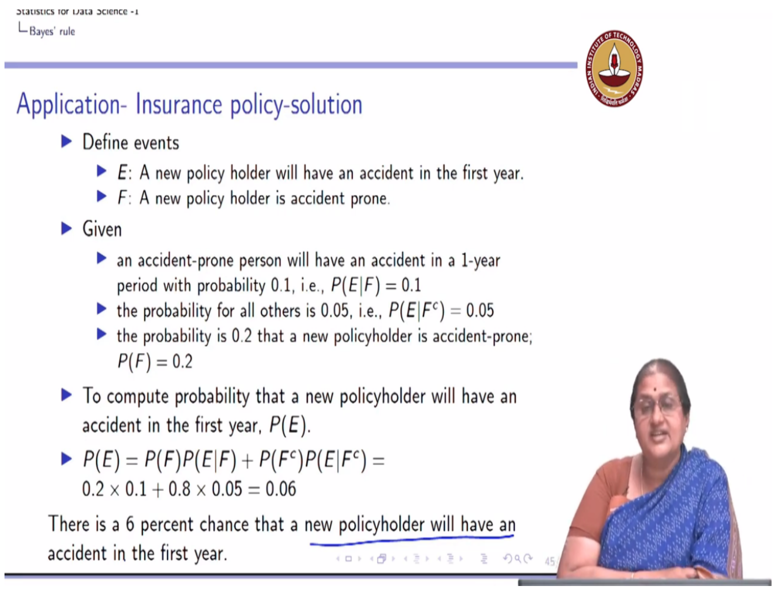
So, that is the answer which we have which is equal to 0.06. In other words, we can state that this is 6 percent chance that a new policy holder will have an accident in the first year, this is a 6 percent chance.

(Refer Slide Time: 15:57)



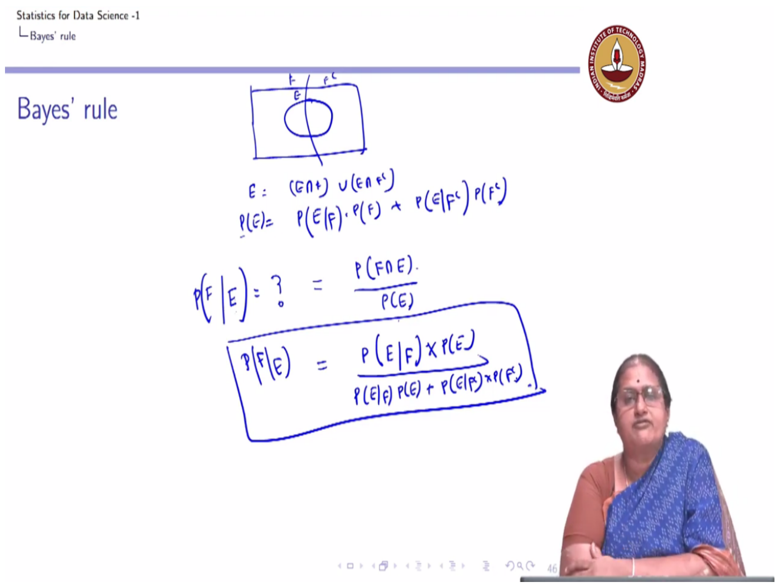
So, now, we come to have state what is very important and what is very popularly referred to as a Bayes' Rule. Recall what we had in the Law of total probability is suppose I have two disjoint events F, F complement and I have an event E here, I can write E as E intersection F union E intersection F complement to give the law of total probability is for each E given F into probability of F plus probability of E conditioned on F complement into probability of F complement. Now, suppose some additional information is given to us, for example,

(Refer Slide Time: 16:46)



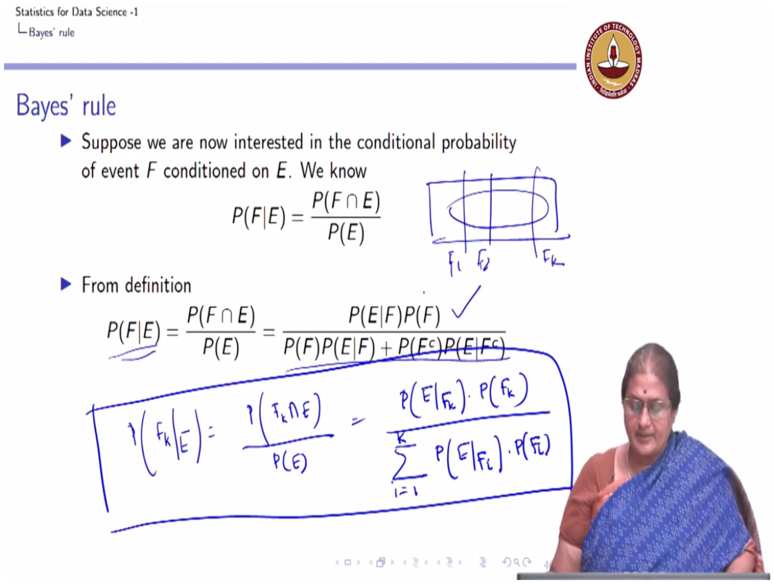
Here going back to this example, I am interested in knowing what is the chance that a new policy holder will have an accident in the first year. Now, I go back, and I see that a policy holder has had an accident in the first year. Now, this is the new information we have. Now, I will I might be interested in knowing that, given this information that a policy holder has had an accident in the first year, did he come originally from the accident prone class or the non accident prone class.

(Refer Slide Time: 17:25)



In other words, given E I am interested in knowing what is the chance he came from an accident prone class. Hence, probability of F given E is given by this expression. And this expression is what is popularly referred to as the Bayes' Rule.

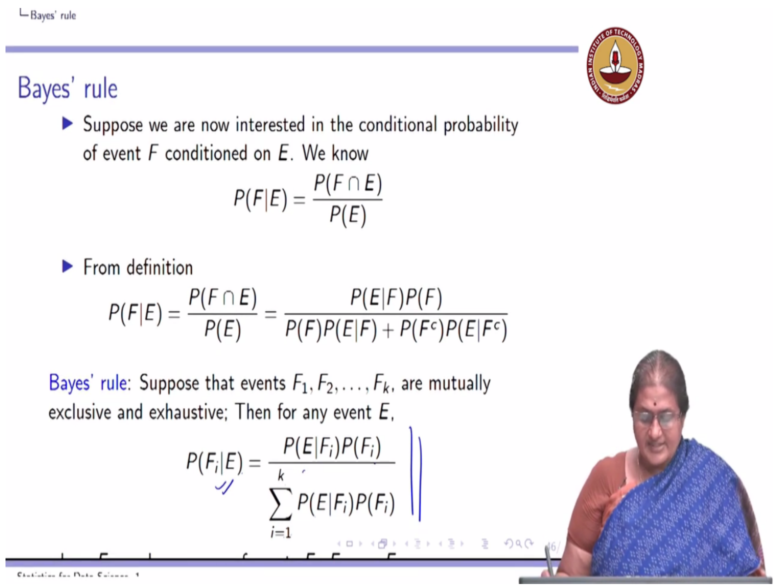
(Refer Slide Time: 18:30)



F conditioned on E is probability of E conditioned on F by the probability of E which is probability of F into probability E conditioned on of a into probability of F complement into probability of E conditioned on F complement. Now, I can extend this Bayes' Rule, I can extend this to my F1, F2, just as a extension extended it to my law of total probability to have k events and I had this event E.

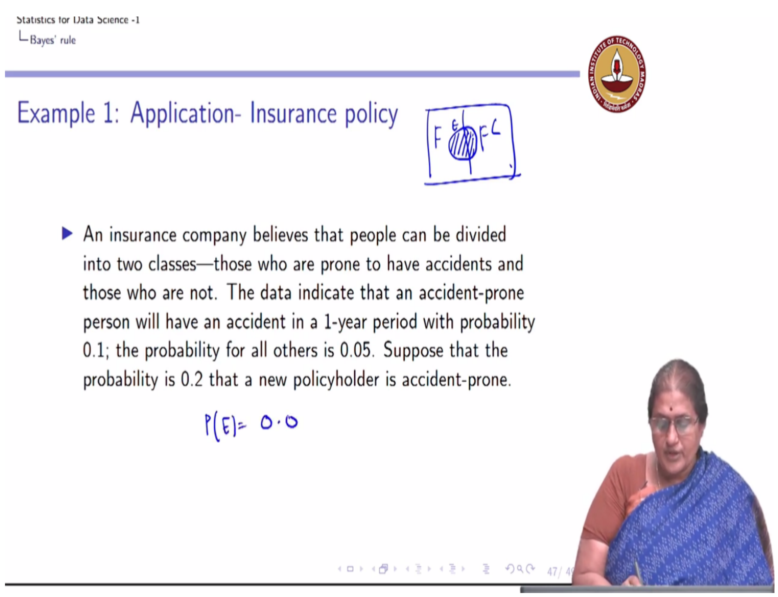
Now, for any event Fk conditioned on E, I know is equal to probability of Fk intersection E divided by probability of E. This is probability of E conditioned on Fk into probability of Fk, that is my numerator, whereas my denominator I saw was probability of E i going from 1 to k probability of E conditioned on Fi into probability of Fi.

(Refer Slide Time: 19:43)



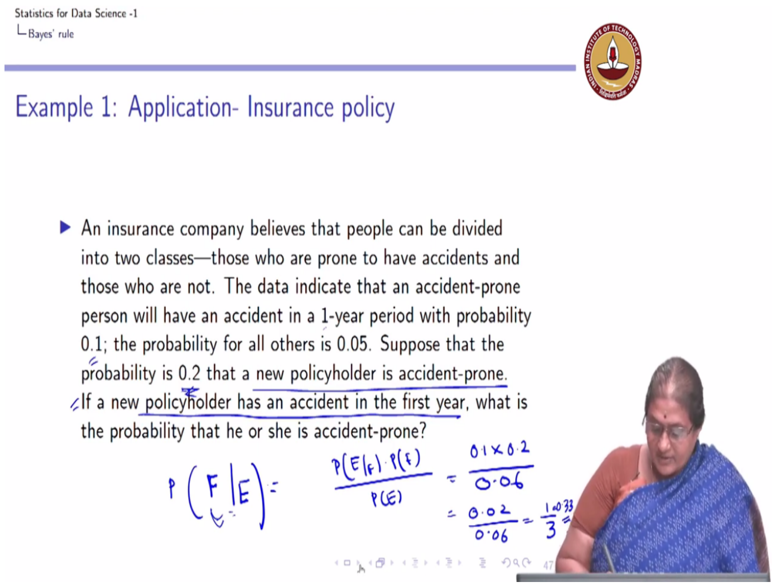
Now, this is what is popularly referred to as my Bayes theorem. Which states the following that suppose I have F1, F2, Fk, which are mutually exclusive and exhaustive. Then, given any event E, I can write probability of Fi conditioned on E is, probability of E given Fi into probability of Fi to summation i going from 1 to k E conditioned on Fi divided by probability of Fi. So, this is what we refer to very popularly as the Bayes' Rule.

(Refer Slide Time: 20:22)



Now, let us go back to our Insurance Policy problem and apply the result there. So, we know that again, I have two classes accident prone and non accident prone. We were interested in knowing what is the chance of a person E where E is the event of, E is the event of a person having an accident within a year. So, we found out probability of E was equal to 0.06.

(Refer Slide Time: 20:52)

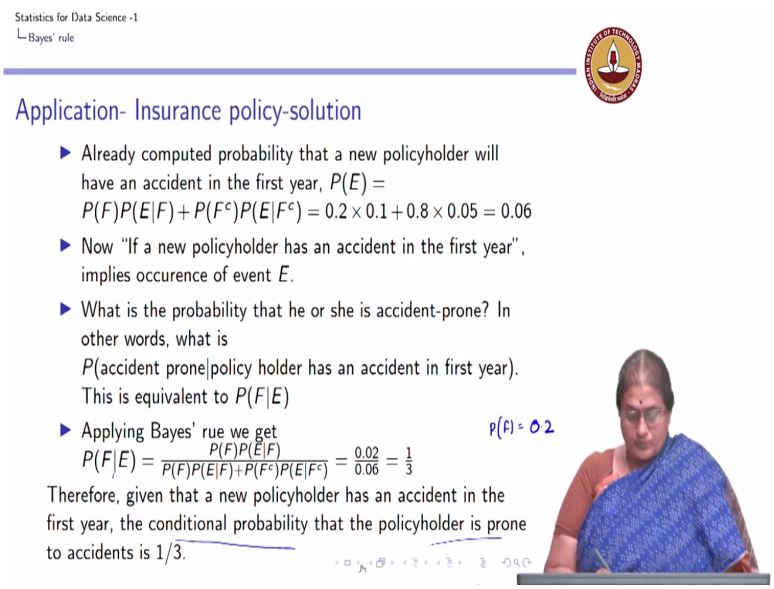


Now, the question is, if a new policy holder has had an accident within the first year, so I am conditioning the problem on E, the accident has happened within the first year, what is the chance that the person was accident prone? So, applying Bayes theorem, I know this is going to be E given F into probability of F divided by probability of E we already computed probability of E in our earlier problem, which was 0.06.

Probability of E given F I know is 0.1, probability of F is 0.2. So, I have this is 0.02 by 0.06, which is 1 third. In other words, if a new policy holder has had an accident in the first year, the chance that he or she is or accident prone is 0.33, do you see that independently this chance that a person will have was accident prone was just 0.1. But given the information that a new policy holder has had an accident, you can see, you can notice that the chance of a policy holder being accident prone was 0.2.

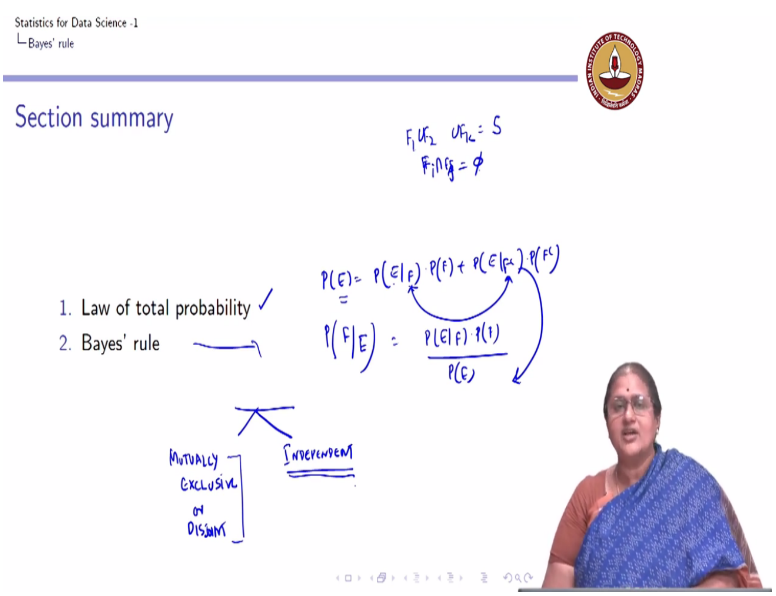
But given the extra information that a person has had an accident within a 1 year period, you see changes the probability of being accident prone to 0.33, or increases this probability. So, in the within the subset of people who have had an accident within a year, you can see that the chance of a new policy holder being accident prone is increased from 0.2 to 0.33.

(Refer Slide Time: 22:47)



So, this is a solution, and I have the probability of F given E is 1 by third. Hence, given that the new policy holder is an accident, has had an accident in the first year the conditional probability is prone to accident is 1 by 3. So that is what we have the conditional probability is equal to 1 by 3, the unconditional probability was 0.2.

(Refer Slide Time: 23:18)

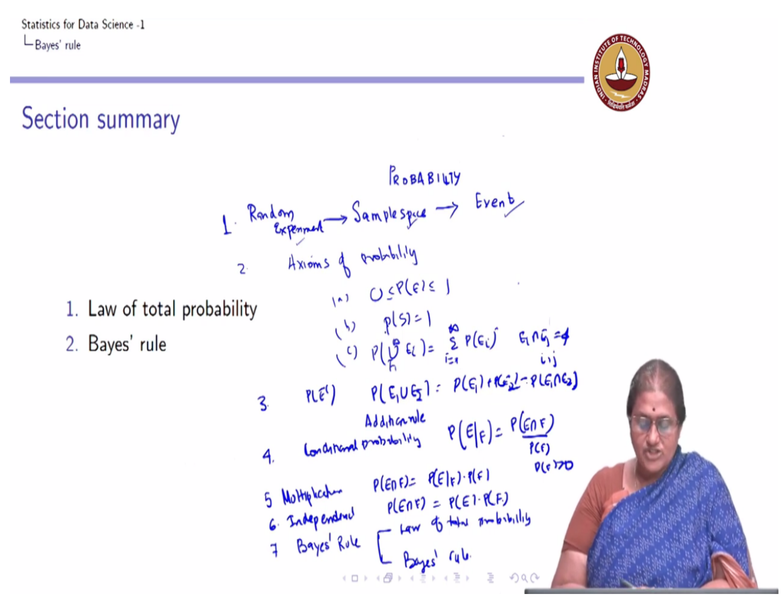


So, what we have learned in this section was first we learned what was the Law of total probability, which was probability of E is probability of E given F into probability of F plus probability of E given F complement into probability of F complement. In other words, the probability of any event happening is the weighted average of that event conditioned on the occurrence of an event and the non occurrence of an event these two the mutual exclusivity of these two events is very important, not only the mutual exclusivity the exhaustiveness So, when I have more than two events, I said that these two are mutually exclusive and they are exhausted. So F1 Fi intersection Fj equal to null set. This is the Law of total probability.

And the Bayes' Rules says states that given this information, then probability of E given F into probability of F divided by probability of E I can substitute this here to get the Bayes' Rule. A word of caution. So, here we have two important things which have learned one is what are mutually Exclusive or Disjoint Events and the other is what are Independent Events. These two should not be confused, they are not the same mutually exclusive events is the occurrence of an event prohibits the occurrence of the other.

For example, when I am tossing a coin, occurrence of a head does not allow a tail to be an outcome. That is when I call them to be mutually exclusive. Whereas independent event says that the unconditional probability of an event happening is the same as the conditional probability of an event happening conditioned on the occurrence or non occurrence of another event.

(Refer Slide Time: 25:39)



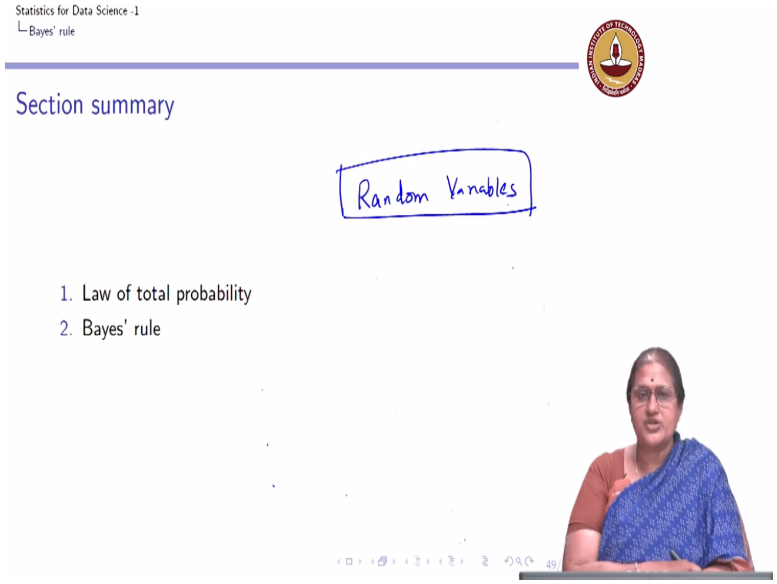
So, with this, we come to the end of the Introduction module of Probability where we first started with the building blocks of probability, we introduced through the concept of a Sample Space. Once we introduce the concept of a sample space, we spoke of what was an event even before the concept of a sample space, we discussed what was a random experiment. So, whenever we we are talking about a probabilistic scenario, we first identify what is the experiment the sample space and the events.

Then we introduce the Axioms of Probability, namely, probability of any event to lie between 0 and 1. Probability of the sample space is 1 and the probability of countable union of disjoint sets is probability of the sum of the probabilities. Ei intersection Ej is i phi all ij. Once we introduce this, we went on to talk about probability of the complement then we had the general addition rule which said probability of E1 union E2 equal to probability of E1 plus probability of E2 minus probability of E1 intersection E2.

This is what we refer to as the addition rule of probability. Then we introduce the notion of Conditional Probability, where we spoke about the probability of E conditioned on F to be probability of E intersection F by probability of F, provided probability of F is defined. Then we spend time and understanding what is a Multiplication rule, which is just another way of restating my conditional probability definition.

Then we introduced the notion of Independence of Events, we stated the multiplication rule for independent events to be the product of the probabilities. And finally, we looked at what was the Bayes' Rule or the Bayes theorem. Here we introduce the Law of total probability and the Bayes' rule. So, this is a summary of what we have done so far. Now, the next topic, which we are going to discuss is the following.

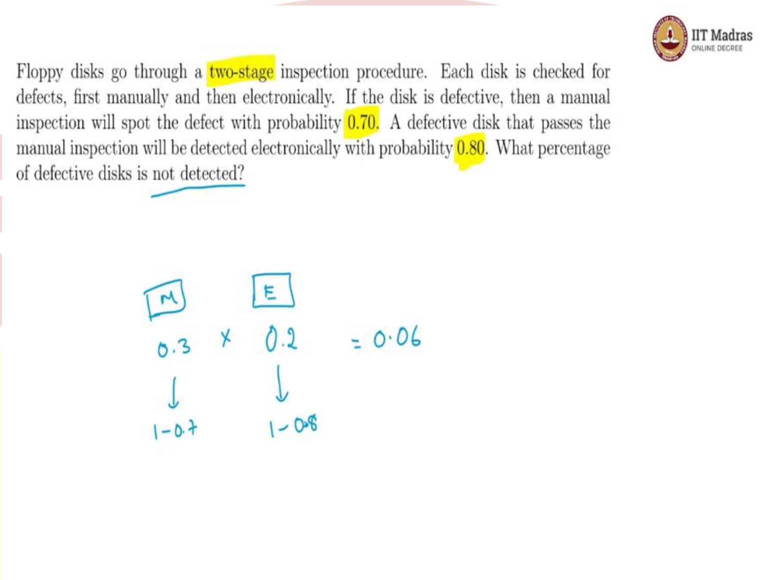
(Refer Slide Time: 28:44)



When we go to the next now, given this knowledge, we are going to go and understand what we mean by Random Variables. So, the next topic is to understand what we mean by Random Variables. When we talk about Random Variables, we will be introducing the notion of a distribution of a random variable, expectation of a random variable, standard deviation, the variance of a random variable, and other concepts. In particular, we will be focusing on two very well known distributions, namely the Binomial Distribution and the Normal Distribution.

Statistics for Data Science-1
Professor. Usha Mohan
Prathyush P (Support Team)
Department of Management Studies
Indian Institute of Technology, Madras
Week 7 Tutorial 1

(Refer Slide Time: 0:14)

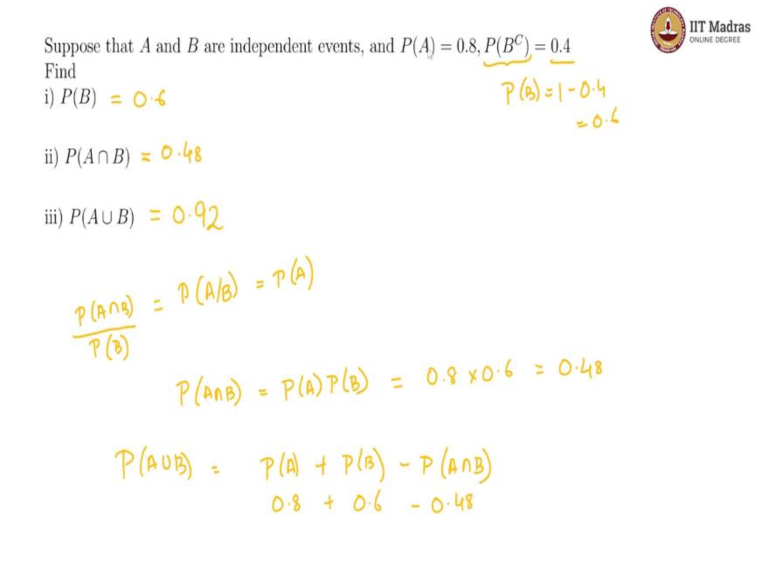


Hello statistics students. In this week's tutorial we will do problems related to conditional probability. In this problem, we have floppy disks which go through a two-stage inspection procedure, each disk is checked for defects first manually then electronically. If the disk is defective, then a manual inspection will spot the defect with probability 0.7. a defective disk that passes through the manual inspection will be detected electronically with a probability of 0.8. What percentage of defective disks is not detected?

So, there there is a two-stage process, one is manual and the other is electronic and in the first case so we are saying that the disk is not detected by these two stages, so the probability for that in the manual is 0.3 and the probability for that in the electronic stage is 0.2. So, now when we look at the combined probability we have 0.2 × 0.3 which is 0.06. So, this 0.3 is 1-0.7 because 0.7 is the probability of detection and the probability of not being detected is 0.3 likewise this is 1- 0.8.

Statistics for Data Science-1
Professor. Usha Mohan
Prathyush P (Support Team)
Department of Management Studies
Indian Institute of Technology, Madras
Week 7 Tutorial 2

(Refer Slide Time: 0:14)



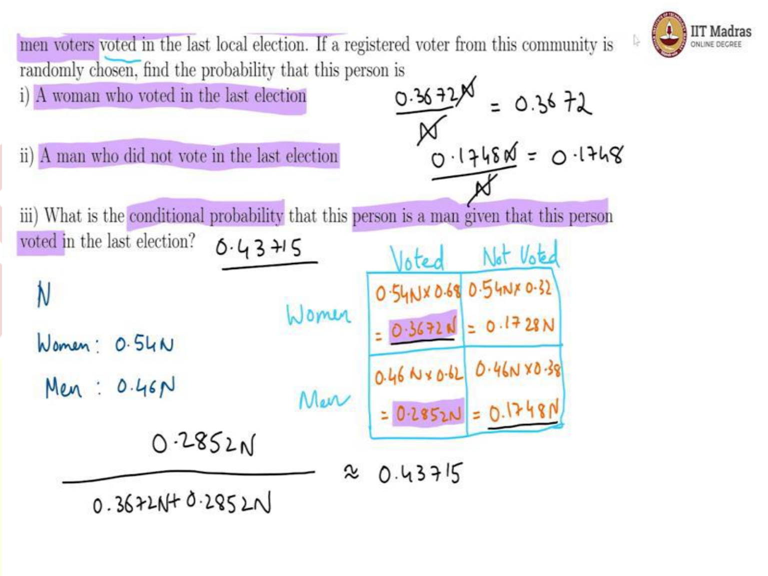
Suppose that A and B are independent events and $P(A)=0.8$ and $P(B^c)=0.4$, then find these particular probabilities. Now what are independent events? Independent events are those whose probability is not affected by the other event happening, so $P(A|\ B)$ should be the same thing as $P(A)$.

Now this is basically $P(A \cap B)$ that is the probability of A occurring with B within the space of B. So, divided by $P(B)$ this is what happens for independent events alone. Now we know from $P(B^c)=0.4$ that $P(B)=1 - 0.4$ which is equal to 0.6.

So first we have found out that $P(B) = 0.6$ then $P(A \cap B)$ for independent events would be just the product of these two and that is 0.8×0.6 which is equal to 0.48. So, this is equal to 0.48 and now the in probability of the union would be the sum of the independent probabilities minus the sum of the intersection that is $P(A) + P(B) – P(A \cap B)$ this is because the intersection is added twice once in A and once in B so we have to subtract it once, so here we will get it as $0.8+ 0.6-0.48 = 0.92$.

Statistics for Data Science-1
Professor. Usha Mohan
Prathyush P (Support Team)
Department of Management Studies
Indian Institute of Technology, Madras
Week 7 Tutorial 3

(Refer Slide Time: 0:15)



Of the registered voters in a certain community 54% are women and 46% are men. So, let us say there are N registered voters totally. So, 54% are women which means the women are 0.54N and the remaining 46% are men. So, men are 0.46N so these are the total number of women and men voters in this N. Now among them 68% of registered women voters, so of the 5.54 and 68% and 62 percent of the registered men voters which is 62% of the 0.46 N.

So, we have a situation like this let us draw a box so let this row be women and this row be men and what is the other criterion they voted in the last election, so the 68% of women and 62% of men voted in the election so this would be they voted and not voted. So, now we try to fill these boxes so of these 0.54N, 68% voted which means this box would be 0.54×0.68 and this is equal to 0.3672 N. So, this is a number of women who voted and of that 0.54 and the remaining 32 percent have not voted, so that would be 0.54 N×0.32 which will give us 0.1728N.

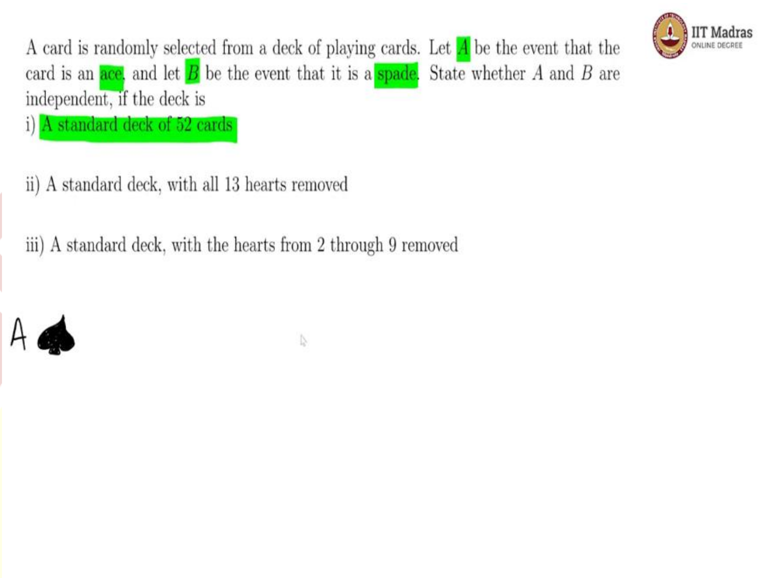
And for men also we have similar logic so of these 0.46 N 62% that is 0.62 have voted so this will give us 0.2852 N and the remaining 38% would be 0.46N×0.38 which is equal to 0.1748 N. So, we now have the total number of people who voted and did not vote and also their gender wise split. So, now we can try to answer these questions a woman who voted in the last election so if a registered voter from this community is randomly chosen find the probability that this person is a woman who actually voted and for that the denominator is clearly the total number of registered voters which is N and the numerator would be the number of women who voted which is 0.367 N.

And these two cancel so the probability is 0.3672 and the probability that this person is a man who did not vote and for that we are looking at men who did not vote which is this value and thus this would be 0.1748N/N, N and N cancels so we have this probability 0.1748 and then lastly we are looking for the conditional probability this person is a man given that this person voted.

So for that we have to look for only the voted cases which would be this plus this this would be the sum of these two would make the denominator so we would have 0.3672+0.2852 in the denominator N and N and the numerator would be the number of such men which would be 0.2852N this is coming out to be roughly 0.43715, so this value is 0.43715 this is a conditional probability.

Statistics for Data Science 1
Professor. Usha Mohan
Department of Management Studies
Indian Institute of Technology, Madras
Week 7 Tutorial 4

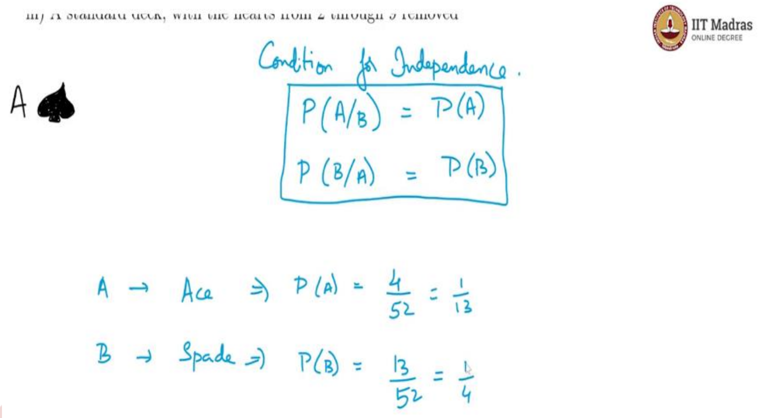
(Refer Slide Time: 00:14)



A card is randomly selected from a deck of playing cards, let A be the event that this card is an ace. And let B be the event that this card is a spade. So, if the card is a spade, then B is true. And if the card is an ace, then A is true. So, both are true when if it is the ace of spades. I do not know if you are familiar with the nomenclature of cards. So, spade is this particular symbol, which looks something like this filled with black.

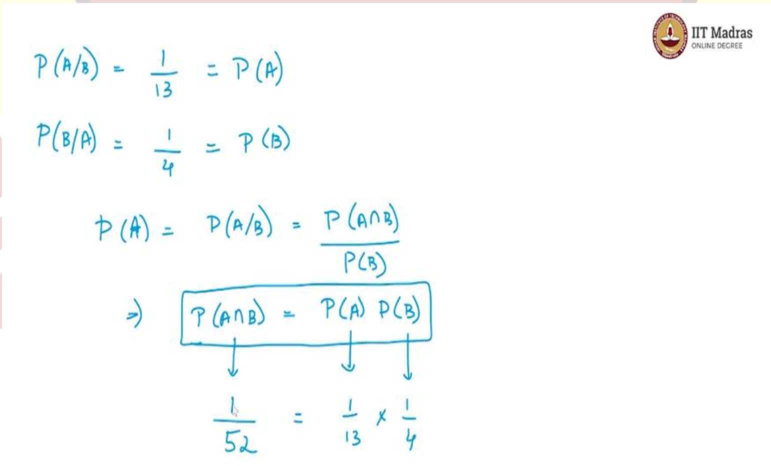
So, the ace of spades would be both A and B being true, any other card of spades B would be true, but A would not be true. And any other suit, an ace A would be true, but not B. Now, state whether A and B are independent if the deck is a standard deck of 52 cards. Now, the condition for independence is that the probability of one event should not change whether or not the other event has taken place, which means P(A) given that B happened should be the same thing as P(A). And likewise, P(B) given that A happened should be the same thing as P(B).

(Refer Slide Time: 01:45)



Now, our events are such that A is an ace the card being an ace, whereas B is the card being a spade. So, this would give us this is the condition for independence. This would indicate that P(A) that is the number of the probability of the card being an ace would be 4 by 52 in a standard deck, there are 52 cards in a standard deck and of them for races. So, that is 1/13. And P(B) is 13 by 52. There are 13 spades out of 52 cards, so you have 1 by 4 as a probability for a spade coming up.

(Refer Slide Time: 02:49)

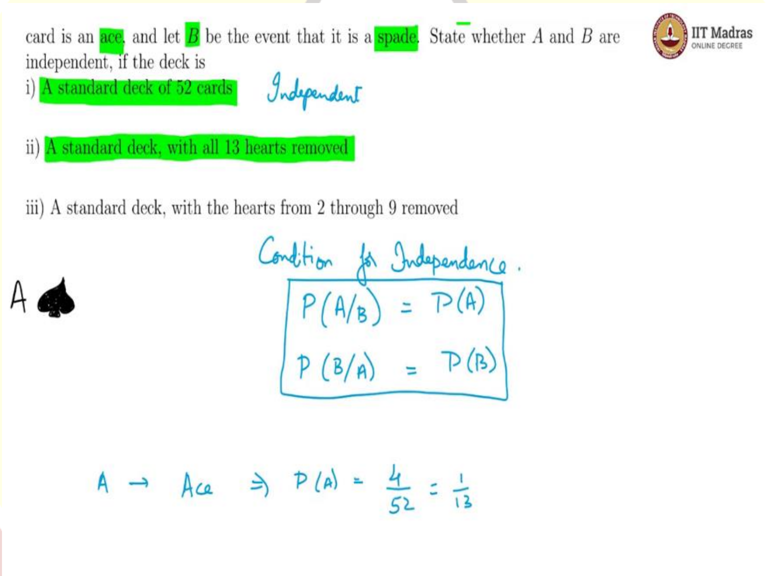


Now what is P(A) given B so in this case there is we know that the card is already a spade, so there are 13 spades, so your denominator is 13. And how many aces are there among the spades there is only 1 this is 1/13, which is equal to P(A). Likewise, if you did P(B) given A if you know your card is already an ace, there are 4 aces, how many spades would be there among these aces only 1 and that is equal to P(B).

So, that condition for independence is satisfied and we can further simplify this by saying that since P(A) is equal to P(A) given B, which is incidentally equal to P(A) intersection B divided by P(B), this condition can be simply written as P(A) intersection B of both of them happening together is simply the product of these two probabilities.

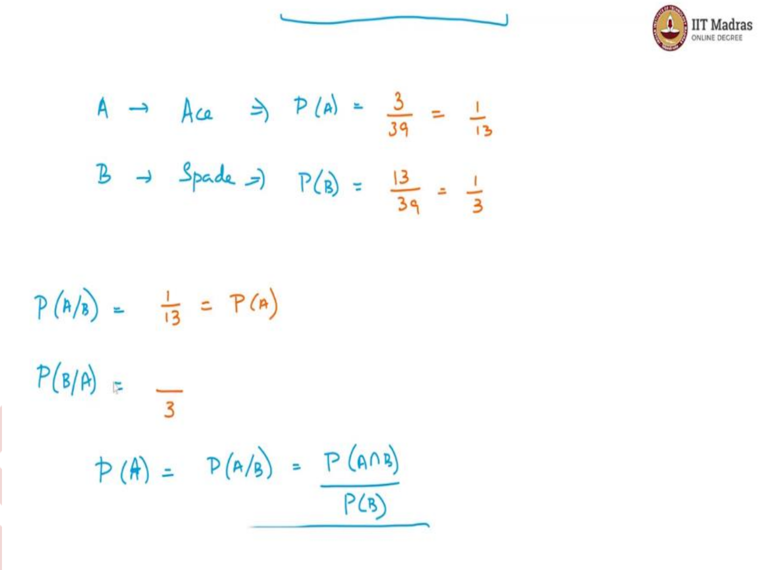
So, if this condition is satisfied, we can say that the events are independent and what is the condition here what is to be satisfied P(A) given A intersection B is both of them happening together. There is only one ace of spades overall. So that will give you 1 by 52 and P(A) is 1/13 P(B) is 1 by 4. So, their product is 1 by 52. So therefore, independence is satisfied.

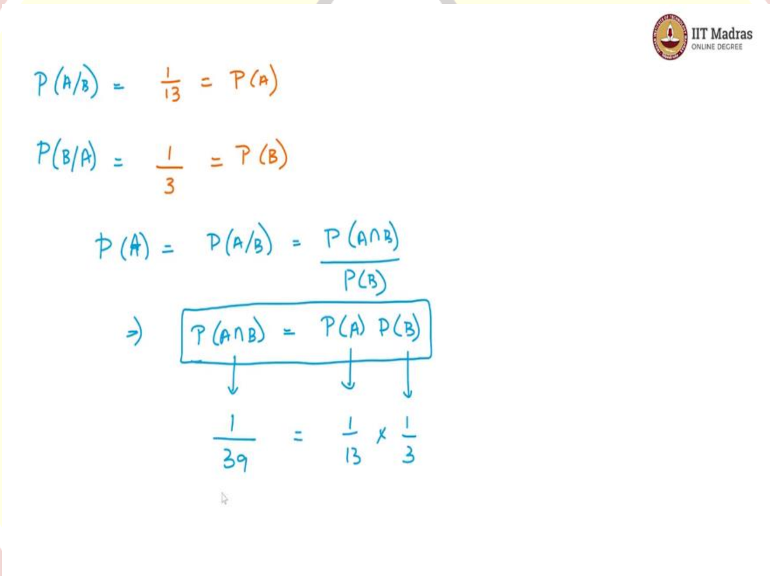
(Refer Slide Time: 04:32)



So, here standard deck when this happens, the two events are independent. Now, we look at a standard deck with all the 13 hearts removed. With all the 13 hearts removed, we will have to modify these let see if the conditions are still satisfied.

(Refer Slide Time: 05:05)

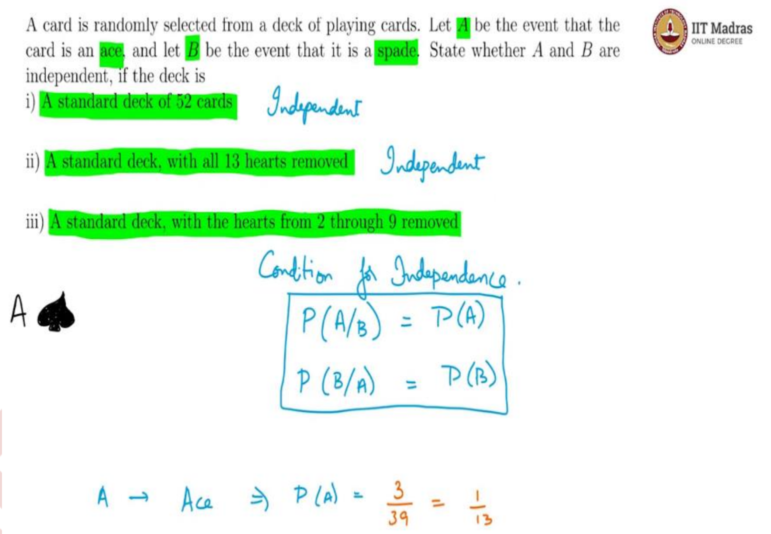




All the hearts are removed, which means probability of an ace coming up will be 3, but there are only 39 cards left. So, this is equal to 1/13 again. And what is the probability of spades there are now 13 spades out of 39. So that will give us 1/ 3 is P(B), then P(A) given B is basically given that it is a spade how many aces.

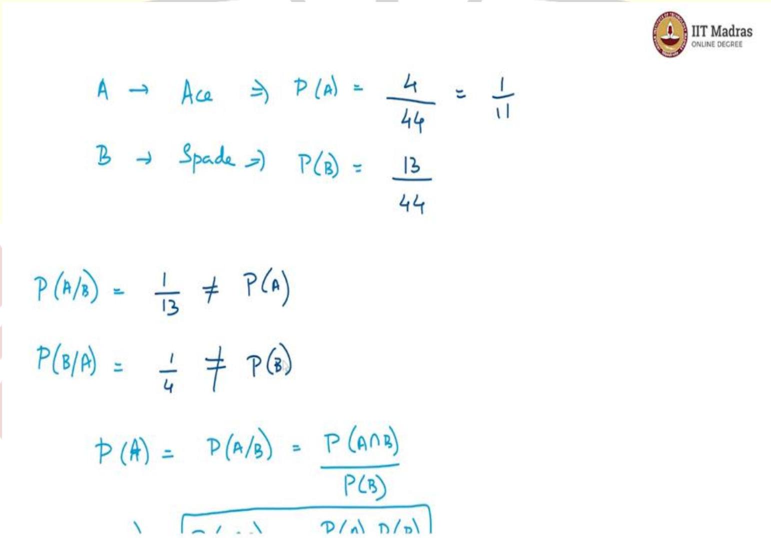
So, that would again be 1/13 which is equal to P(A). And what is P(B) given A this would be there are only 3 aces now, and how many of them would be spared only 1. So, this is equal to P(B), and again our condition is satisfied, except in this case, P(A) intersection B is going to be 1/39 there is one ace of spades in the 39 cards and P(A) is 1/13 and P(B) is 1/3 and again 1/13 and 1/3, 1/39. So, the condition is satisfied.

(Refer Slide Time: 06:20)



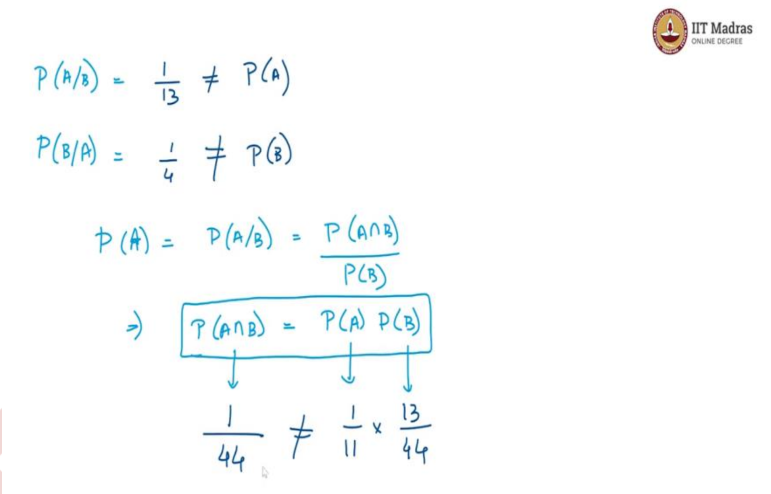
So, we can again claim that the second case also these two events will be independent. Now, let us look at the third case a standard deck with hearts from 2 through 9 removed that is heart 2, heart 3, heart 4, hearth 5, heart 6, heart 7, heart 8 and heart 9 are remote, which means 8 cards are removed.

(Refer Slide Time: 06:52)



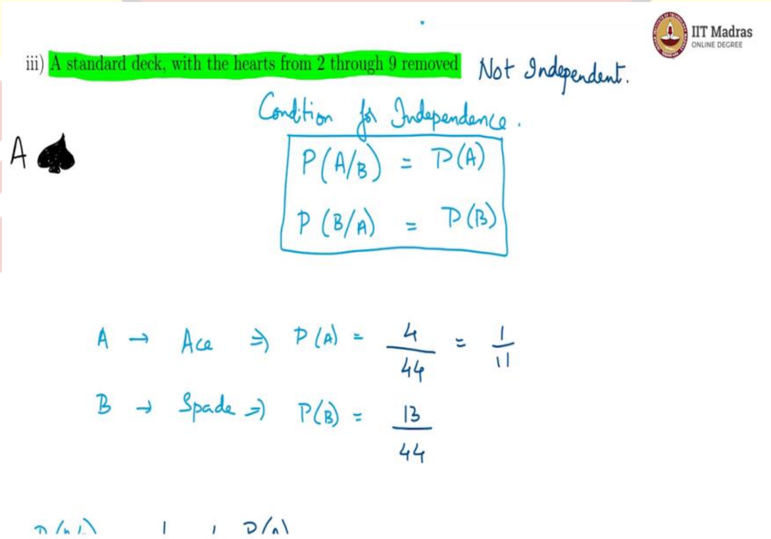
Now, let us see what happens to these probabilities. 8 cards are removed which means there are now 44 cards in the deck and among them how many aces are there are 4 aces, because the ace of hearts is not removed. So, you will get 1 by 11 is P(A). Now, what is P(B) of these 44 cards how many spades are there 13. So, 13 by 44 is P(B).

(Refer Slide Time: 07:28)



Now let us see what happens with P(A) given. So given that the card is a spade, how many of them is an ace. So, you have 1/13 here directly, but 1/13 is not equal to P(A) likewise P(B) given A. So, if it is an ace what is the possibility probability that it is a spade, that is again 1 by 4 because there are 4 aces and only 1 spade among them. And this is again not equal to P(B). If you check for this condition now P(A) intersection B is basically one card out of that is spade… aces spades is one card out of your 44 whereas, P(A) is basically 1/11 and P(B) is 13/44 and this product is not equal to 1/44.

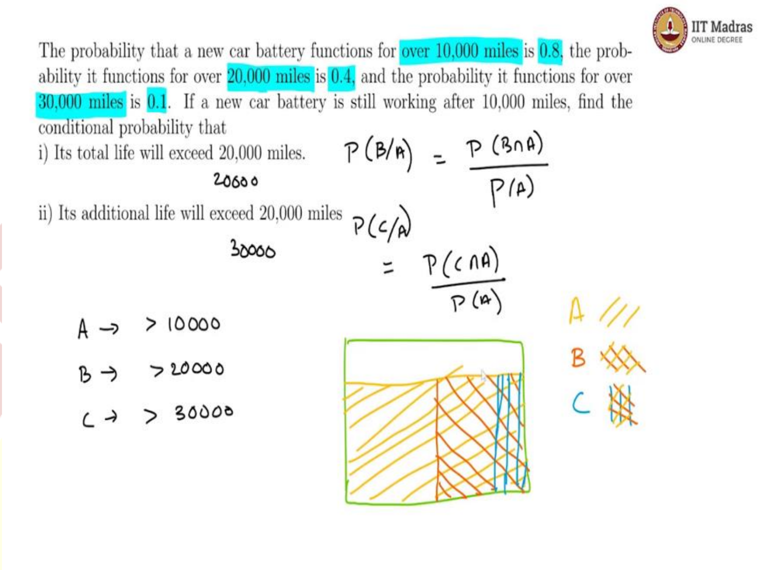
(Refer Slide Time: 8:32)



Therefore, in this case the two events are not independent.

Statistics for Data Science 1
Professor. Usha Mohan
Department of Management Studies
Indian Institute of Technology, Madras
Week 7 Tutorial 5

(Refer Slide Time: 00:15)



The probability that a new car battery functions for over 10,000 miles is pointed. And the probability that it functions for over 20,000 miles is 0.4. And lastly, the probability that it functions for over 30,000 miles is 0.1. And if a new car battery is still working after 10,000 miles, find the conditional probability that its total life will exceed 20,000 miles and additional life will exceed 20,000 miles that means, case 1 is for the 20,000 scenario, and case 2 is for the 30,000 scenario.

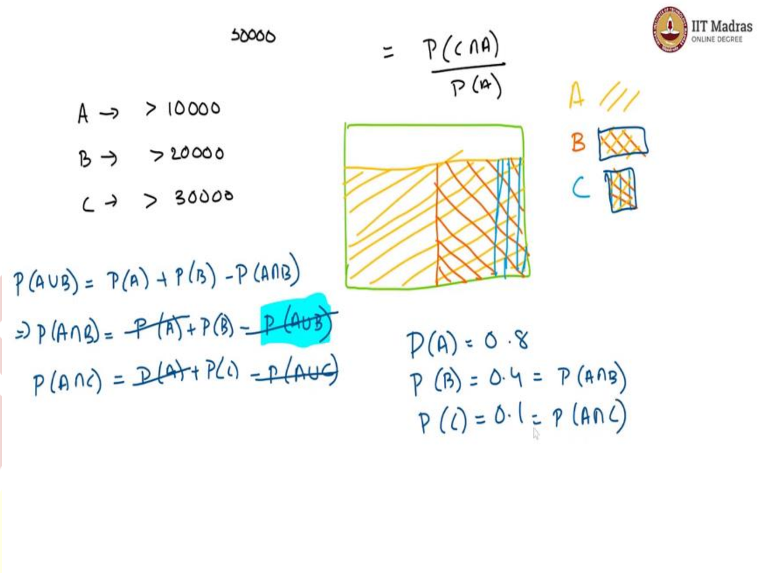
So, if we consider these as 3 events, where A is exceeding 10,000, B is exceeding 20,000 and C is exceeding 30,000 what is being asked of us here is P(B|A) and this is P(C|A). And we know this is equal to ——— and likewise, this would be equal to ———. And for this, if we draw Venn diagrams, we will observe the following.

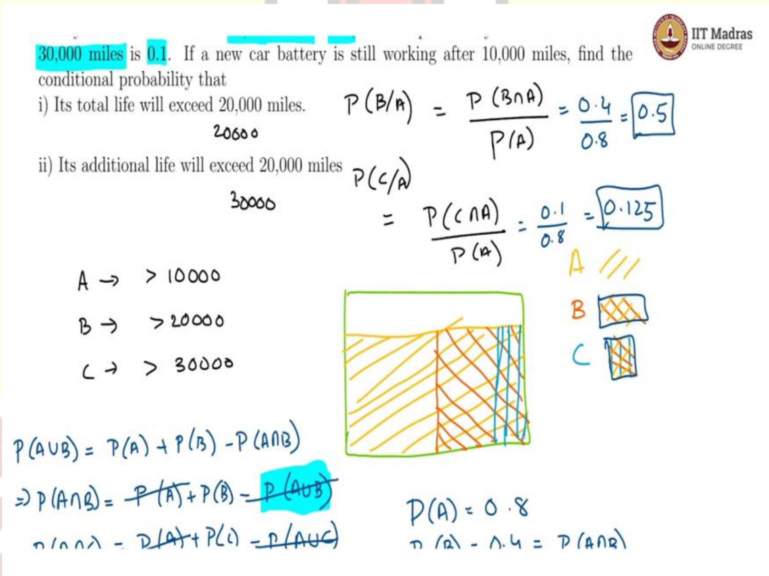
So, let this be the entire universe. And if there is 80% probability of this happening, so yeah, let us say this is A, this is A. And B is a case which is entirely within this and it is only 40% of the total universe. So, B would be something like this, because B has to happen only when A happens. So, B is completely within A and lastly, the 30,000 miles C happens within B. So, within A is B and within B is C. So, C is also within A and that would be something like this.

So, we have a situation where A is these lines and B is these lines and C is these lines. However, C will also have these lines. And these lines, because C is within embedded within

B and within A it and likewise B is going to have these lines because B is embedded within A.

(Refer Slide Time: 03:40)





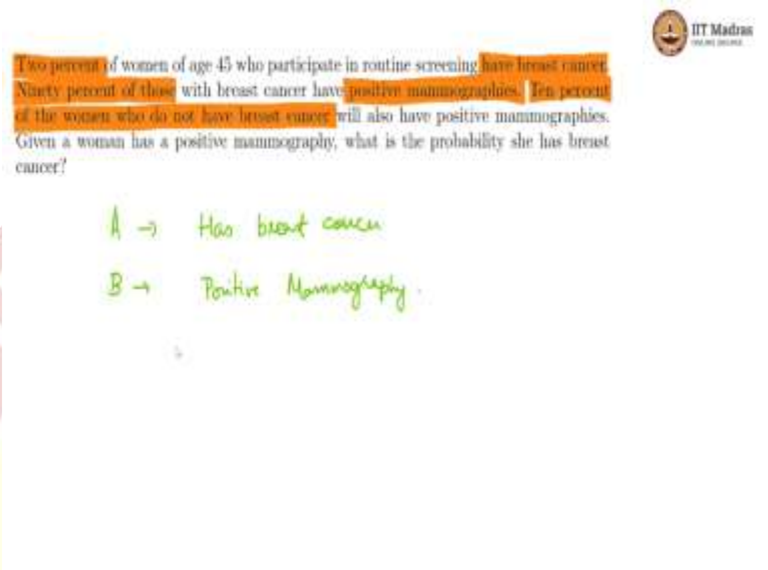
Anyway, now looking at this case, you know, P(A) is 0.8 and P(B)=0.4 and P(C) is equal to point how much was that 0.1 this is what is given to us. Now, for finding the intersection probabilities, what we should use is this probability of union is equal to P(A) + P(B) – P(A∩B). And that would give us P(A∩B) is equal to P(A) + P(B) – P(A∪B).

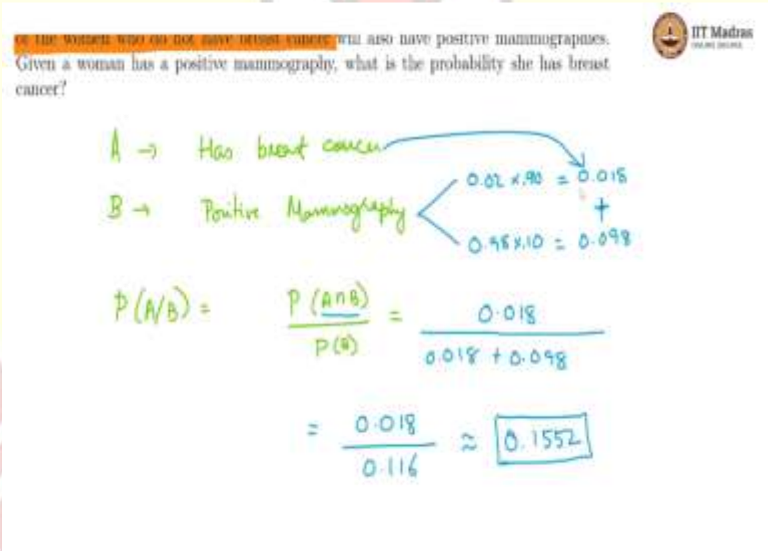
Likewise, P(A∩C) also will be P(A)+P(C) – P(A∪C). So, once we have this, what we have left to observe is this P(A∪B), the probability of the union is going to be just P(A) because B is fully embedded within A so the union is just A. So, therefore, P(A) and P(A∪B) cancelled here and likewise P(A∪C) and P(A) are the same thing because C is fully embedded within A. As you can see, B is completely within A and C is completely within B and A.

So, therefore P(B)=0.4 is also P(A∩B) and P(C) which is 0.1 is also P(A∩C) = 0.4/0.8 which is 0.5. And this would give us 0.1/0.8 which is 0.125. So, these are the probabilities we are looking for.

Statistics for Data Science 1
Professor. Usha Mohan
Department of Management Studies
Indian Institute of Technology, Madras
Week 7 Tutorial 6

(Refer Slide Time: 00:15)





2 percent of women of age 45 who participate in routine screening have breast cancer, so 2 percent of the women have breast cancer and off them 90 percent of those with breast cancer have positive mammographic. And 10 percent of the woman who do not have breast cancer will show positive mammographic. Given that a woman has a positive mammography, what is the probability that she has breast cancer?

So, we are doing conditional probability here where one event is that the woman has breast cancer. And the other event is that the mammography is positive. So, positive mammography.

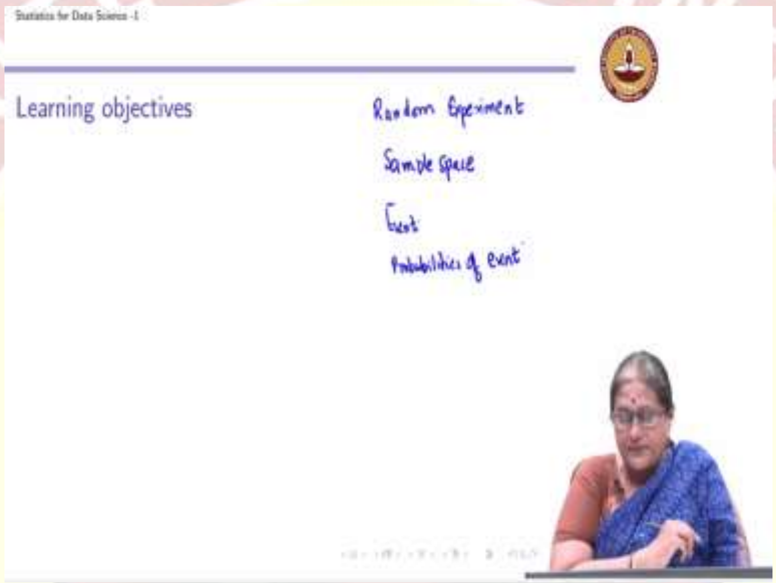
And what we are looking for is basically $P(A|B)$, which is essentially $\frac{P(A\cap B)}{P(B)}$. Now, we know $P(B)$ is affected by both people having breast cancer and not having breast cancer. So, people having breast cancer is 2 percent so 0.02. And of them 90 percent that is 0.9 show up was 2 mammographies and that gives us 0.018 and in the remaining cases of the 98 percent who do not have breast cancer, 10 percent show positive mammographies.

So, that gives us so that is point one, that gives us 0.098 This is a total probability of a mammography showing positive so that would be $P(B)$ = 0.018 + 0.098. And of them the $A \cap B$ case is this having breast cancer and having a positive mammography is this case which is $P(A \cap B)$ = 0.018. So, we will have $P(A|B)$ = 0.018/0.116. And that is roughly equal to 0.1552. So, this is the conditional probability of a woman who was given positive mammography having breast cancer and this pretty low. Thank you.

Statistics for Data Science - 1
Professor. Usha Mohan
Department of Management Studies
Indian Institute of Technology, Madras
Lecture No. 8.1
Discrete Random Variable

In this week, we are going to learn about the important concept of a Random Variable.

(Refer Slide Time: 00:22)

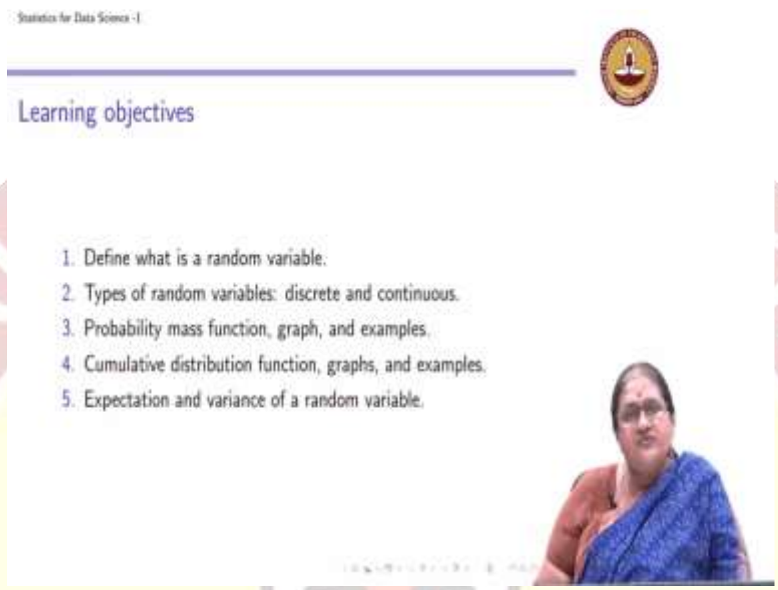


So, what we have learned so far is we have heard learned about what we call a Random Experiment. After a random experiment, we defined what was a sample space. A sample space is what we defined as a set of all outcomes of this random experiment. Then we define the notion of an event, which is a subset of a sample space. And then we define probabilities of the event.

When we define probabilities of an event, we approach that through the axiomatic approach to define probabilities. Now, what are we going to do after this then we introduce the notion of conditional probability we introduced what was based here. But basically what we have done is we have started introducing this entire framework of probability through the notion of random experiment and sample space.

So, by now, all of you should be really comfortable in computing probabilities of events. To help you compute the probabilities of events. We also introduced you to permutations and combinations. So today, and the next 2 weeks, we are going to focus on the following. So, we are going to learn about random variables.

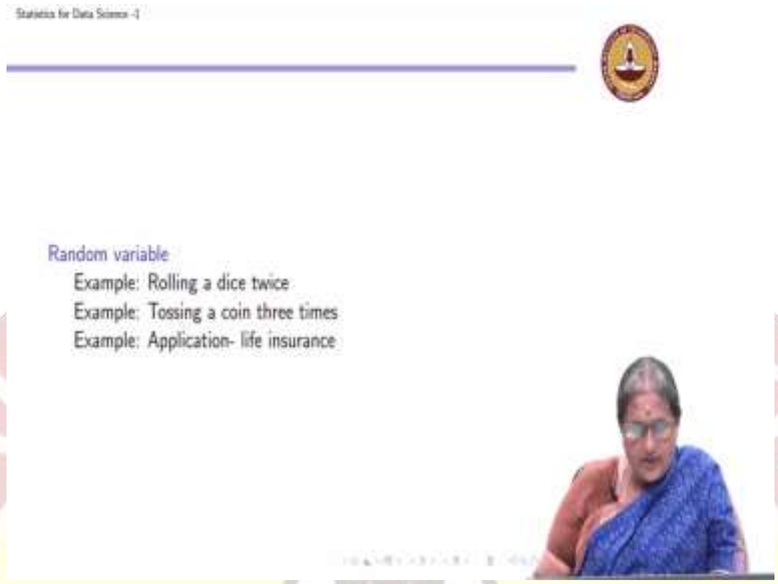
(Refer Slide Time: 02:00)



In particular, after the end of these 2 weeks, you should be able to know what is a random variable. And then you are going you need to understand what is a discrete random variable, what is a continuous random variable. When you are talking about discrete random variable, which is going to be the focus of the first 2 weeks, we will be introducing notions of a probability mass function, how the graph of a probability mass function would look and then afterwards, we'll also introduce the notion of a cumulative distribution function.

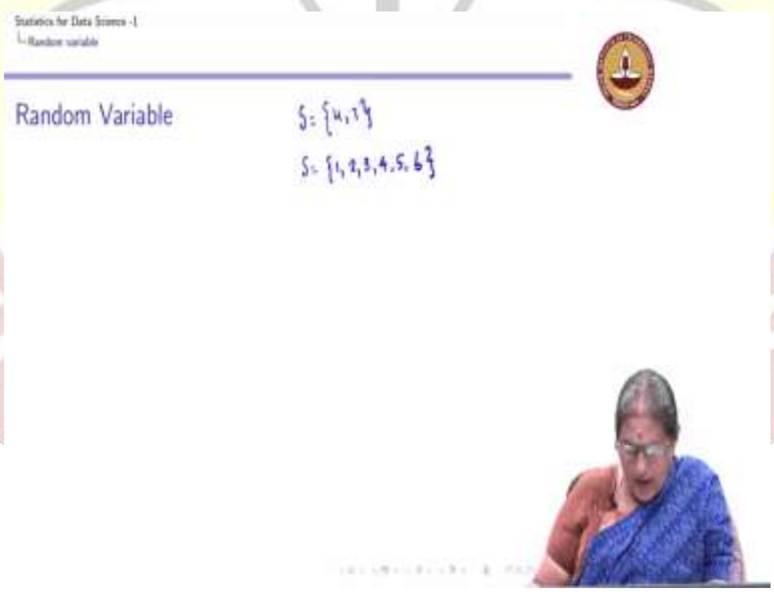
Now, when we talk about both the probability mass functions and accumulator distribution functions, we are going to give a lot of examples, which might be very useful. Finally, we are going to introduce important concepts, namely the expectation and variance of random variables.

(Refer Slide Time: 03:01)



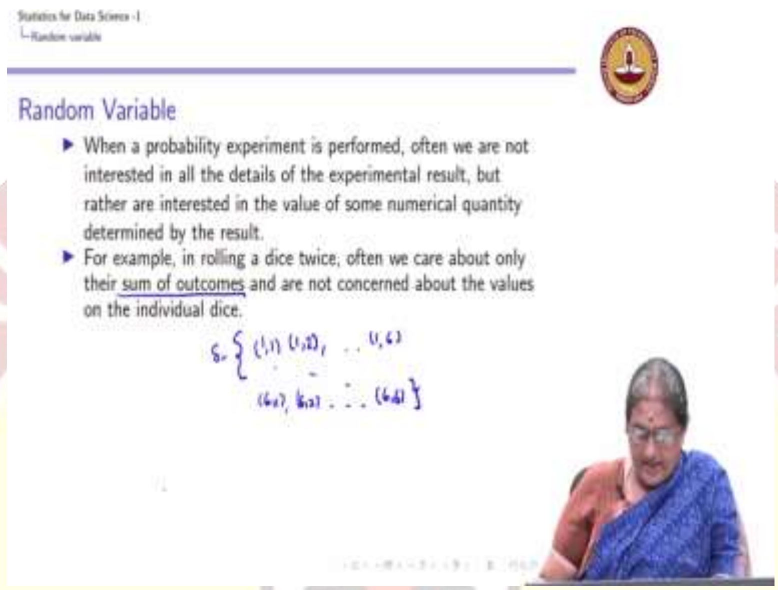
And we are going to see how we can apply this in our day to day lives. So, we begin with the notion of a random variable. So, what is the notion of a random variable? Let us start by revisiting a random experiment.

(Refer Slide Time: 03:19)



So, recall when we talk about a random experiment, for example, when a toss a coin, I have a random experiment, the sample space is head and tail. When I roll a die once I can have any 1 of these outcomes, 1, 2, 3, 4, 5, and 6, any 1 of these 6 outcomes can happen. So, we say that every time I can use the notion of a sample space to describe the outcomes of the experiment, but many time I might not be just interested in knowing about the outcome of a experiment.

(Refer Slide Time: 03:59)

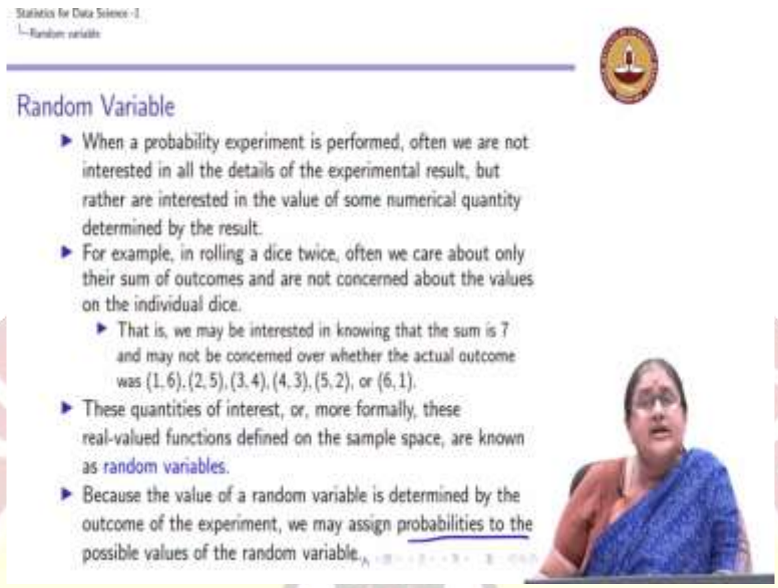


When a random experiment or a probability experiment is performed. We are not interested in the actual outcome. But we might be interested in the value of some numerical quantity determined by the result, I repeat, we might not be interested in the outcome itself, but we might be interested in the value of some numerical quantity that is determined by the result. What do we mean by this for example, when I roll a dice twice, I know there are 36 possible outcomes.

For example, I could have a 1 in my first toss, I could have 1 in my second toss. I could have 1 in my first toss a 2 in the second toss. I could have a 1 in a first toss and a 6 in the second toss. I notice that my sample space would have 36 such outcomes. This is something which we have already seen in our earlier lessons. But suppose I do not care about the individual outcomes, but I am interested only in the sum of the outcomes.

So, what is the numerical value or quantity I am associating and associating, and I am only bothered about the some of the outcomes.

(Refer Slide Time: 05:33)

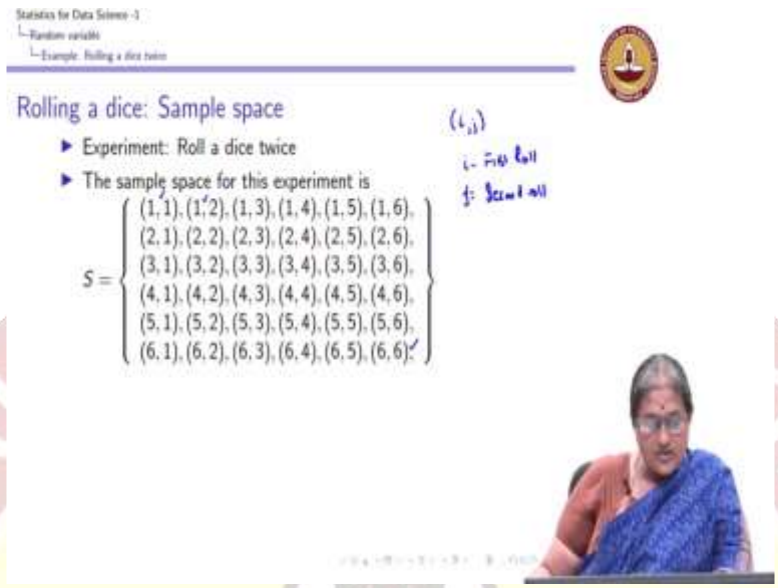


In this case, for example, I am not concerned, I am only bothered about whether the sum of outcome is 7. But I am really not concerned about whether that 7 has arisen because of outcome 1, 6, or 2, 5, or 3, 4, or 4, 3, or 5, 2, or 6, 1. I am not concerned about it, but I am interested in knowing that the sum is 7. Now, whenever I am talking about such a numerical quantity, these numerical quantities of interest, more formally, they are actually functions defined on the sample space, I am not going into the mathematical rigor of defining such functions.

In advanced courses, you will be subjected to the mathematical rigor. But what I want you to understand this with every outcome of the sample space, I am associating a quantity, a numerical quantity. So, this quantity of interest is what I refer to as a random variable. Since these random variables are, again, the values of the random variables are determined by the outcome of a random experiment.

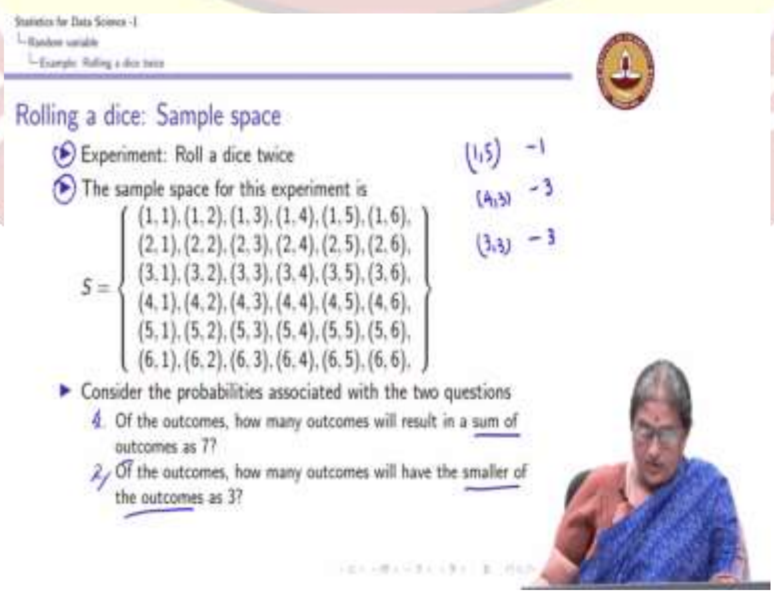
And I can actually assign values to these outcomes of the random experiment, I can assign probabilities to the possible values of the random variable. So, let us look at an example to understand what we mean by numerical quantities.

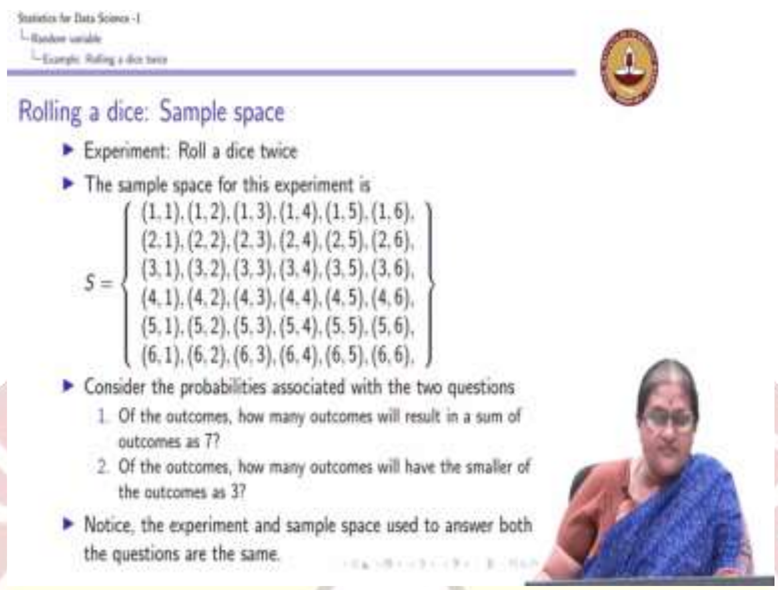
(Refer Slide Time: 07:19)



Let us revisit the experiment of rolling dice twice. So, when I roll a dice, twice, that is I have a 6 sided die. And I am rolling it twice. You already seen, I could have 36 possible outcomes. And those 36 possible outcomes are listed in the form of my sample space, which is given here I have s, I have 36 possible outcomes, it could be a (1, 1), (1, 2), up to (6, 6), by (i, j) I mean, i in the first toss, or the first roll of the dice, and j is outcome of the second roll of the dice. So, I am throwing a dice twice on my rolling a dice twice. So, this is something which you have already seen, we have defined events on this.

(Refer Slide Time: 08:21)



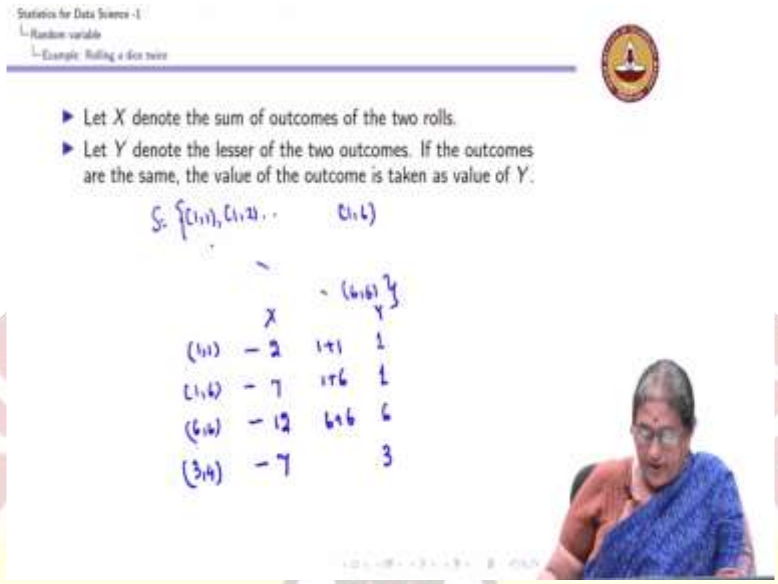


Now suppose I am asking two questions. So, I am asking the questions out of out of these outcomes, how many outcomes do we have we have 36 outcomes? How many outcomes will result in a sum of 7? Notice I am just interested in the final value of 7. I am not interested in the individual outcomes as such. The other question I am interested in knowing is how many outcomes will have the smaller of the outcomes as 3.

What do I mean by smaller of the outcome, if I have an outcome (1, 5), the smaller of these outcomes is 1, whereas and (4, 3), the smaller of the outcome is 3. And for convenience, if it is (3, 3), I am going to take it as 3, both the outcomes are the same, we say the smaller of the outcome is itself. So, these two are the questions I am interested in answering. So, how do we answer these 2 questions?

So, these questions, I notice the experiment is the same, the sample space is the same. The questions are actually based on the outcomes and associating some numerical quantity. 1 is the sum of outcomes. And 1 is what is the smaller of outcomes. So, it is important for us to notice that the random experiment and the sample space that I am going to use to answer both these questions are the same. So, what is the random variable?

(Refer Slide Time: 10:06)

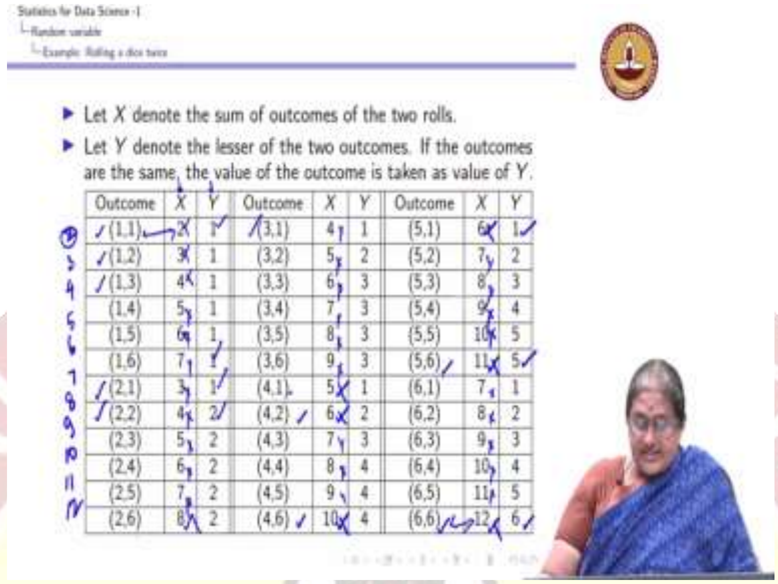


So, now let us start with the same example, let me denote the sum of the 2 roles by X. Let me denote Y to denote the lesser of the 2 outcomes. So for example, I know this is my sample space, I have (1, 1), (1, 2), up to (1, 6), and I have up to (6, 6), this is what we have listed earlier. This was a sample space given here.

So, I am just writing the sample space again here. And I am telling X is denoting the sum of outcomes. So if I have (1, 1) is my outcome, the value X will take is 2, which is 1 + 1. If I have (1, 6) is my outcome, the value X would take a 7, which is 1 + 6, the value is (6, 6) is my outcome, the value X would take is 12, which is 6 + 6. So, X is denoting the sum of outcomes. Similarly, Y is denoting the lesser of 2 outcomes.

So, what is the lesser of these 2 outcomes. I said, if the outcomes on both the roles are the same, I am going to take it as itself, the lesser of (1 , 6) is again a 1, the lesser of (6, 6) is a 6, suppose I had (3 , 4), the value of X would be a 7 and value of Y would be 3. So, this is how we are defining our X and Y here.

(Refer Slide Time: 11:52)



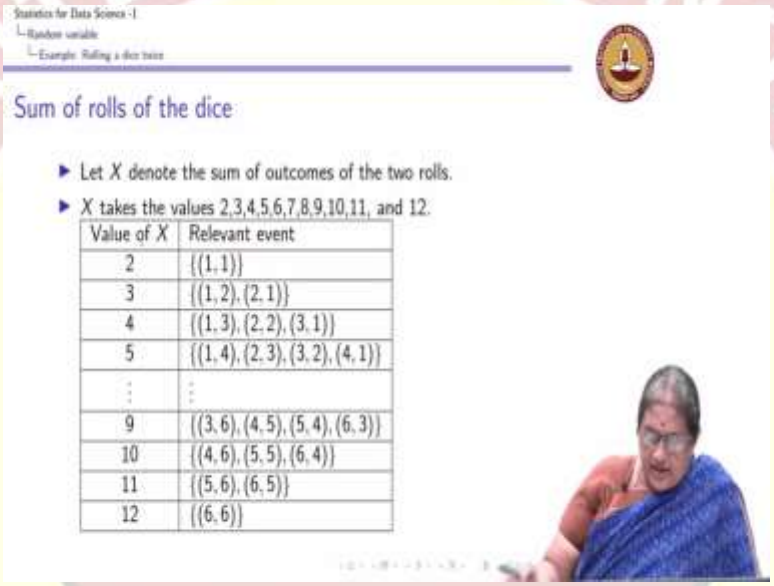
So, for the entire problem, we can see that these are I have 36 outcomes. For (1 , 1), the value is 2, (1, 2) is 3, 1 + 3 is 4, so 4, 2 + 6 is 8, the minimum of (1, 1) is 1, (1 , 6) is 1, (2, 1) is again, 1, (2, 2) is 2. So, you can see that 4 + 1 is a 5, 4 + 2 is a 6, 4 + 6 is a 10. And here you can see the outcome, 5 + 1 is a 6, minimum of (5, 1) is a 1, so minimum of (5, 6) is a 5, 5 + 6 is 11. 6 + 6 is at 12. And the minimum of (6 , 6) is a 6.

So, you can see that for each 1 of these 36 outcomes of my sample space, I have a value of X, and I have a value of Y, which is associated with it. Now why what, what, what next. Now, these outcomes are outcomes of a random experiment. And associated with each one of these outcomes is the value of X and the value of Y. So, when I have the outcomes, and these values, I can talk about associating, or I can talk about the following is, can I talk about probability of the values X taking the values?

Now let us go back here, X is taking a set of values, Y is taking set of values. On closer inspection, I can see X takes the value 2 takes the value again. Does it take the value 1, it does not take the value 1, it takes the value 2, it takes the value 3, I have an X taking a value here I take X again taking a value here, then I have yes, 4 yes, it takes a value 4 here it takes a value 4 here. 5 yes 5, 1. So, you can see that X is taking the value 5 yes 6, 7, 8, 9, 10, 11.

So, it is taking 7, 8, 9, 10, 11 and 12. So, if you look at the values this X is taking, it is taking the values 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, and 12. These are the values this X is taking. I can see that X is taking the value to when the outcome is (1, 1), it is taking the value 3 when the outcome is (1, 2), and it is taking the value when it is (2, 1), it is taking the value 4 when the outcomes are (1, 3), then it is also (2, 2), and it is taking when it is (3, 1). So, this way you can see that it is taking a particular value, where I have more than I could have only 1 outcome corresponding to the value as in this case, and in this case, or I could have more than 1 outcomes corresponding to the value. So, what are the values X is taking?

(Refer Slide Time: 15:32)



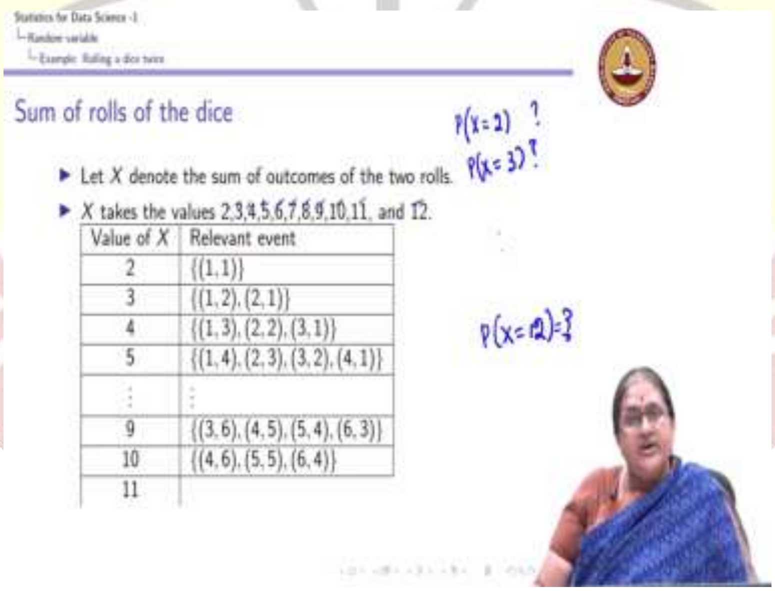
X is taking the values, which are which is denoted as the sum, it is taking the value 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 and 12. These are the values X is taking. And however defined X, I have defined X to be the sum of the outcomes of both the throws or both the roles of the dice. Now given this, now, let us do the following. I see what is the value of X. Now, let me take the value of X to be 2. Now, what is the relevant event, I know X will take the value too, when the outcome is (1, 1). So, the relevant event that is corresponding to 2 is just the event (1, 1), there is no other way X can take the value 2.

Now X can take the value 3, if my first outcome is a 1 and second outcome is a 2, first outcome is a 2 and second outcome is a 1. So, these is again the relevant event, which will give me the value of X equal to 3. So, if I am going to look at each one of them, I get 2 with the event (1, 1), 3 I have the event which is (1, 2) and (2, 1), 4 I have the event (1, 3), (2, 2) and (3, 1), this the sum of these outcomes, which I can define as a relevant event will give me the sum of 4, 5, it is going to be (1,4), (2 , 3), (3 , 2) and (4, 1), and so forth, they keep going.

And then you can see 9, it is going to be (3, 6), (4, 5), (5, 4) and (6, 3), all of them add up to 9. For 10, I see the events are going to be (4,6), (5 , 5) and (6 ,4), for 11, it is going to be (5, 6) and (6, 5), and 12 I see the relevant event is again, just (6, 6). So, what we have done is first we have associated a value with each of the outcomes. And then I have just nabbed I have seen that the X takes the value 2 through 12, X takes these values, I have mapped the relevant event for each value X takes.

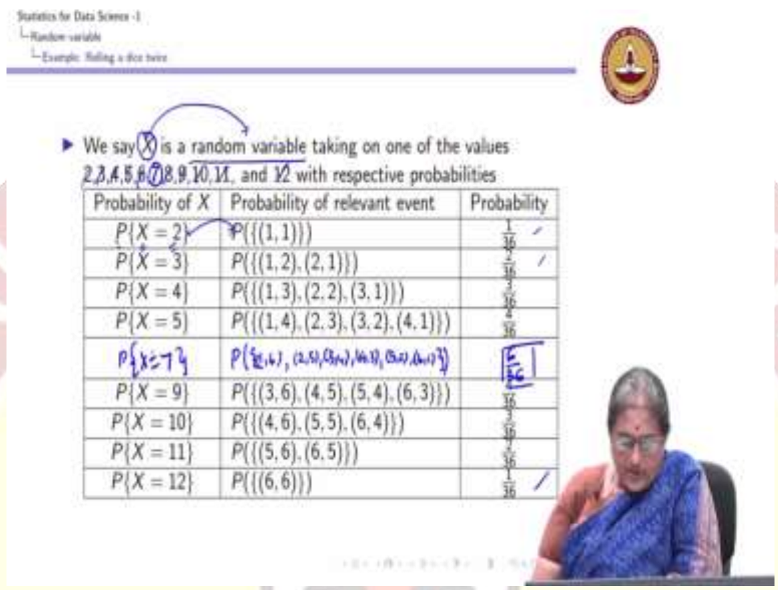
Now what is the point of mapping this relevant event? We recall what we said is because these random variables are defined on my probability sample spaces, what I can actually see and what we have defined earlier is the following is because the value of a random variable is determined by the outcome of an experiment; we may assign probabilities to the possible values. Now come back to the example what are the possible values of the random variable? I see that the possible values of the random variable are 2 through 12.

(Refer Slide Time: 18:58)



These are the values it can take the values 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12. So, the question is, can I assign a probability of X taking the value 2. Can I assign the probability of X taking a value 3? Can I do this for all the values X is taking? That is the question we are asking. To answer this question we are finding, and we are mapping to each value that X is staking what is the relevant event? We have already seen how to come up with probabilities of events.

(Refer Slide Time: 19:42)



Once we do this mapping, we can see that the probability of X is the same as a probability of the relevant event. So, what is the X value of X? I am looking at value of X, X takes the value 2. I am interested in knowing what is probability X = 2, I know X = 2, is the relevant even test, just this set, I need to know what is the probability of this set happening. And we know the probability of this happening is 1/36. This is something which we have already seen in an earlier discussions.

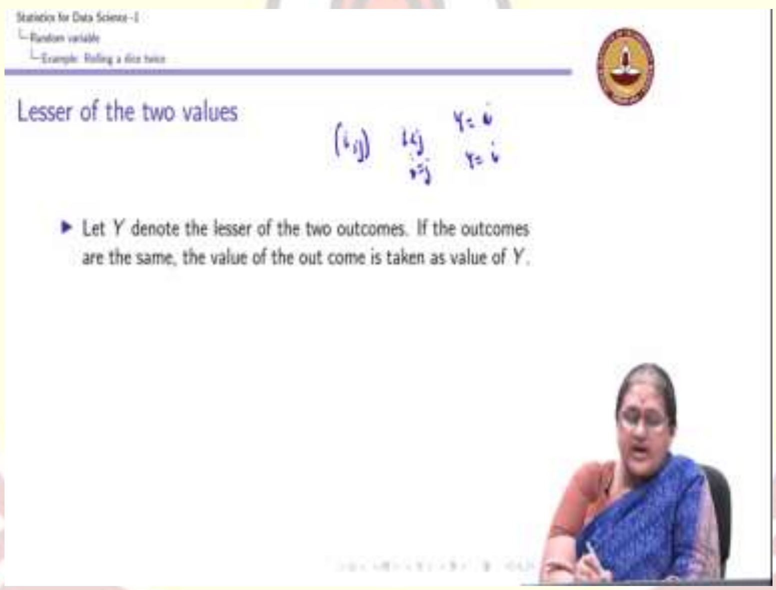
Similarly, if X takes the value 3, I am interested in knowing what is the probability of X = 3? So, I know what are the outcomes that give me X = 3. So again, I go back, and I see it is (1, 2), and (2, 1), these are the, this is the irrelevant event. And I know the probability of this event is 2/36. So, if I continue in this way, I get probability X = 2 is 1/36. Probability of $(X = 3) = 2/36$, X = 4 is 3/36, probability X = 5 is 4/36.

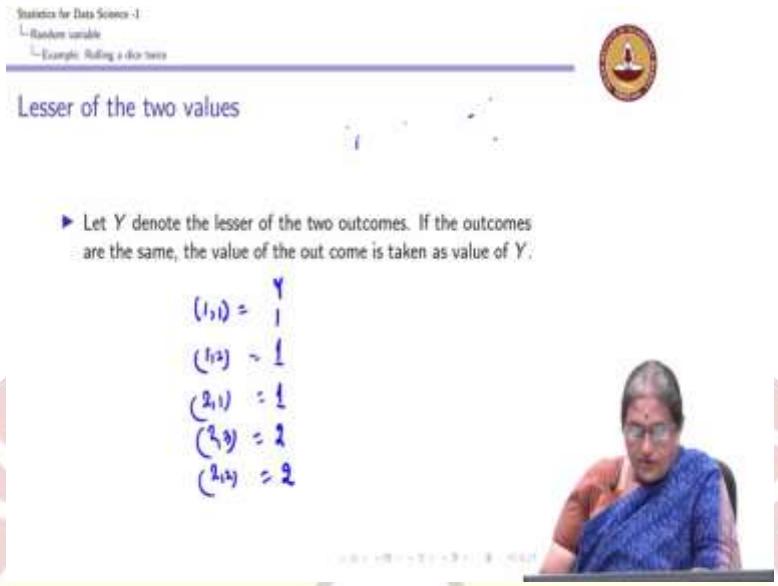
I can keep continuing it this way. And I can verify that probability X = 9 is 4/36, X = 10 is 3/36. Probability of X = 11 is 2/36. And probability X = 12 is 1/36. So, what we have established is, I know that this X, I say X is a random variable, which takes the values 2 3 4 5 6 7 8 9 10 11, and 12. And I can also assign a probability to the value X, taking a particular value, which I get from recognizing the probability X taking a particular value, I will find out what is the probability of the relevant event and I have assigned probabilities to each of the values X takes.

Now let us look at the other. So, I asked two questions. The first question I asked is, what is the probability of the sum equal to 7? That was a question. So, I am not interested in anything else, I just need to check whether X takes a value 7, I see that X takes a value 7, I need to know what is the probability X would take the value 7, I know probability X would take a value 7 is same as probability (1 6), (2, 5), (3, 4), (4, 3), (5, 2) and (6, 1), which is same, which would be 6/36.

So, I am not interested in the individual outcomes, but the probability that the sum is equal to 7. And I know that that happens with a probability of 6/36. So, what we have done is we have defined what is a random variable, and we have assigned probabilities to that random variable.

(Refer Slide Time: 23:17)



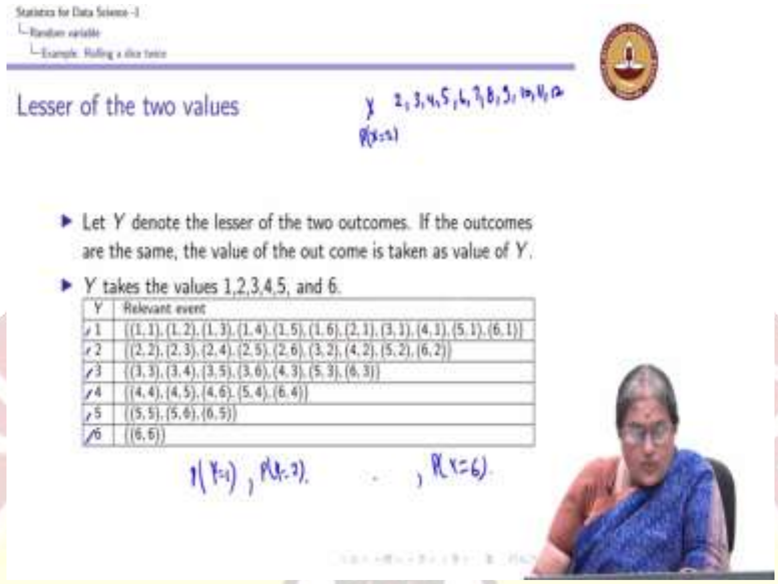


Now on the same sample space, what do we mean by same sample space, I have again, roll 2 die, I have the same sample space. Now I am going to define Y to be the lesser of the 2 outcomes? What do we mean by it? If (i, j) is an outcome, if $i < j$, then Y will take the value i. If $i = j$ Y will again take the value i, I saw, I can define that Y takes the value i if $i \leq j$. So, what are the values this Y would take?

So for example, if I have (1, 1), the value is 1, Y is equal to 1, (1, 2) the values again equal to 1. (2, 1), the values again equal to 1, (2, 3) the value would be equal to 2. (2, 2), the value is equal to 2. So, these are the values that I am associating with each of the outcomes. So, I know that for each of these 36 outcomes, I can see that again, I can go back here.

So if you go, so we can go back here, you can see here, the value is 1. So Y takes the value 1, the second value Y takes is 2. So this is where it takes the value 1, 2, it takes the value 3 takes the value 4, takes the value, 5, and 6. So you can see that on the same sample space, now I am interested in the value Y takes. So Y takes the values 1, 2, 3, 4, 5, and 6, these are the values this variable, or this Y takes.

(Refer Slide Time: 25:21)



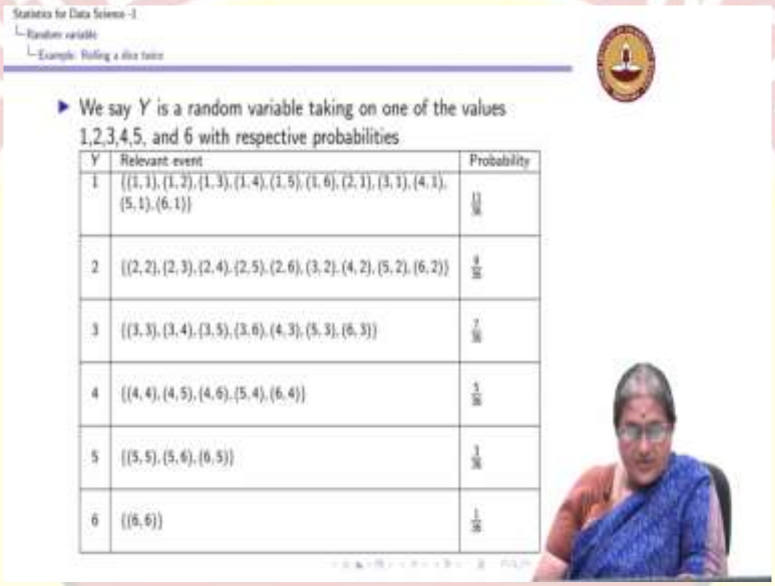
So, you can see why it takes the values 1 2 3 4 5, and 6. So, I can again repeat what I did for the earlier case. So, I see Y, with Y takes the value 1, the relevant event is when we would Y take the value 1, it takes the value 1 when my outcome is either a (1, 2), (1, 3), (1, 4) or all outcomes in which 1 appears either in the first toss, or in the second toss. So, what are the outcomes in which 1 appears either in the first toss or second toss. So, I have (2, 1), (3, 1), (4, 1), (5, 1), and (6, 1). So, these are the outcomes in which 1 appears.

And for these 2 outcomes, Y takes the value 1. So, I can say that Y is taking the value 1, the relevant event is this following, which is listed all the possible outcomes I have just discussed. Now, when we take the value 2, Y would take the value 2, when 2 is equal to (2, 2), (3, 2), (4, 2), (5, 2), (6, 2), (3, 2), (4, 2), (5, 2), and (6, 2). So, you can see that the outcomes were the lesser of the 2 outcomes is 2, and that you can see is (2, 2), (2, 3), (2, 4), (2, 5), (2, 6), (3, 2), (4, 2), (5, 2) and (6, 2).

The third thing, when would Y take the value 3, again, I have a (3, 3), I will have a (3, 4), I will have a (4, 3), I will have a (3, 5), (5, 3), (3, 6), and (6, 3), which I can list in the same way, I have all these outcomes, where Y takes the value 3. Similarly, for 4, Y takes the value when it is (4, 4), (4, 5), (4, 6), (5, 4) and (6, 4). 5 and it is (5, 5), (5, 6) and (6, 5), and Y takes the value 6, only when the outcome is (6, 6).

So, you can see immediately what you notice is the value Y is taking whereas X was the sum which took the value 2 3 4 5 6 7 8 9 10 11 and 12. And I obtained what was the probability X = 2, X = 3 up to X = 12. So similarly, here, I have Y, which is taking value 1, 2, 3, 4, 5, 6. And I can find out based on what is the relevant event, I can also tell what is probability Y = 1, Y = 2 up to probability Y = 6. I can tell this similarly, like the way we did earlier. So, let us see what are the probabilities.

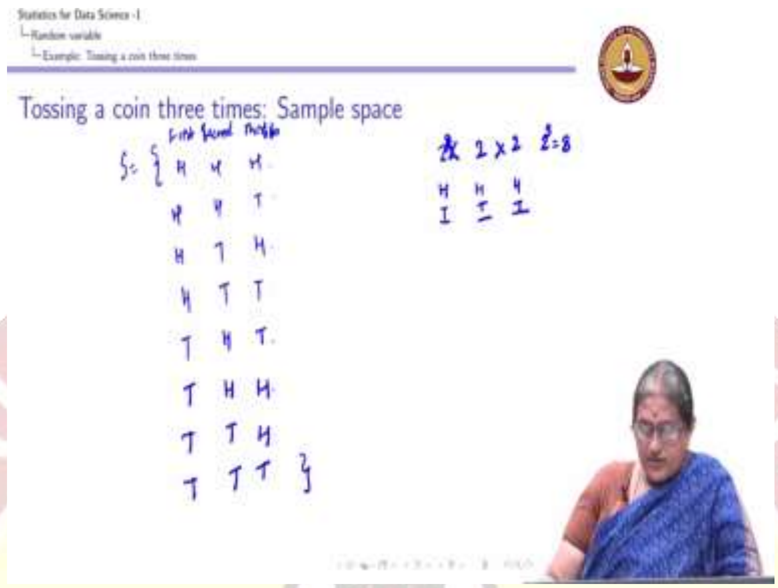
(Refer Slide Time: 28:26)



So, I have Y = 1, Y equal to 1, I have this is my relevant event. And I can see there are 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 in the event. And the probability of this event happening is 11/36 so probability Y = 1 is l1/36. Similarly, the probability with which Y takes the value 2, I have 1, 2, 3, 4, 5, 6, 7, 8, 9. And all of them are equally likely. So, it is 9/36, probability Y = 3, is 1, 2, 3, 4, 5, 6, 7, which will give me a 7/36. Probability Y = 4 is 5/36, Y = 5 is 3/36. And finally, probability Y = 6 is 1/36.

So, this was an example where we have defined 2 random variables, which are capturing different numerical quantities on the same sample space, which comes from the same experiment.

(Refer Slide Time: 29:45)

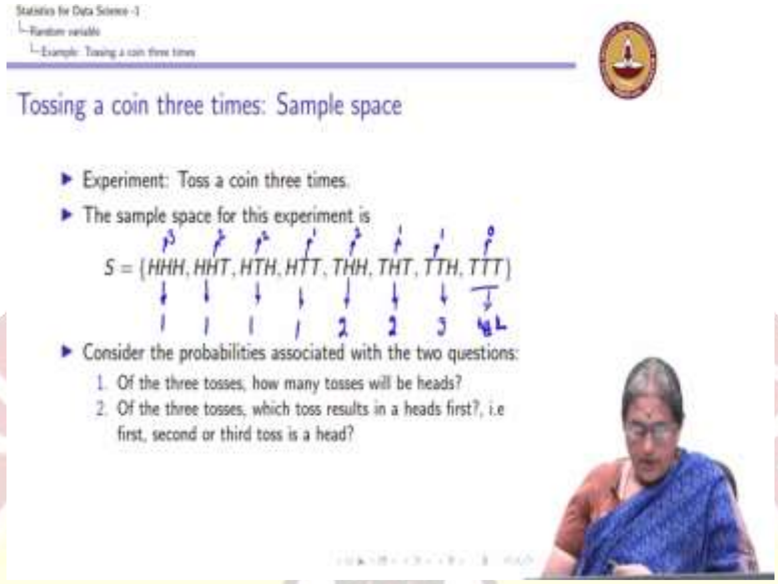


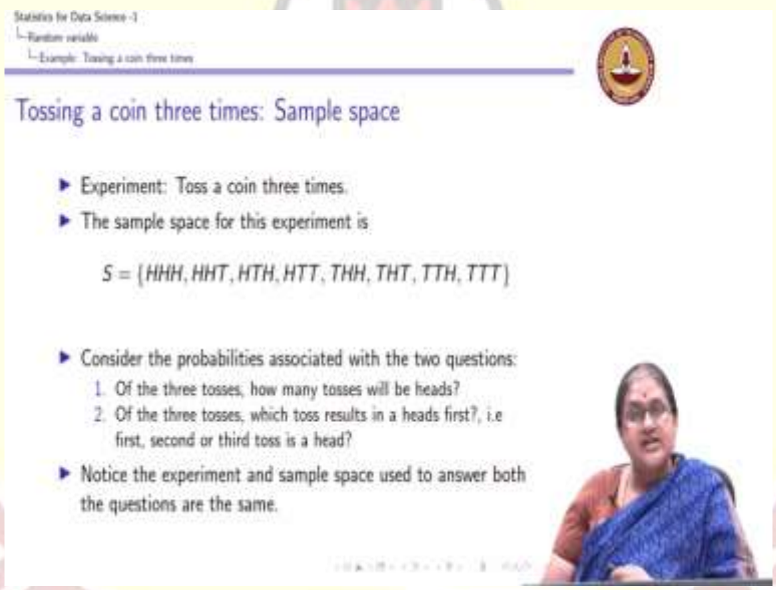
Now let us look at another problem. Let us look at another example. Again, this is an example which we have seen earlier, and I am tossing a coin 3 times. So, when I toss a coin 3 times I am observing what is the outcome of each of the tosses, I know the sample space in this is my first toss I record, what is my first toss, my second toss and my third toss, this is what I am recording.

So, if I look at the outcome, I could have a head, head, head, I could have a head in my first toss, head in the second toss or tail in my third toss, head in my first toss, tail in my second toss, head in my third toss, I could also have a head in my first toss, a tail and my second toss, a tail in the third toss, I could have a tail in my first toss, head in a second toss, tail in my third toss, tail, head, head; tail, tail, head and tail, tail, tail.

So, you can see that this 1, 2, 3, 4, 5, 6, 7, 8 are the set of possible outcomes that is I can have a head in my first head in the second. So, it is basically head appearing in there. So, I have 3 tosses. The first toss could be head or tail. So, there is 2 ways of happening it. Second toss also could be head or tail. Another 2 ways of happening third toss could also be head and tail, another 2, so $2 \times 2 \times 2$, which is 2^3 , which is 8 possible outcomes for this experiment.

(Refer Slide Time: 31:28)





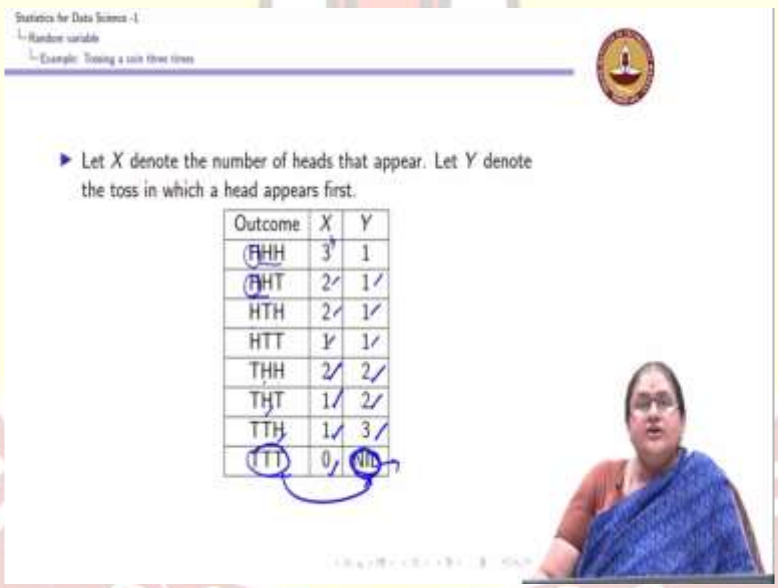
And the sample space for this experiment is hence given by these 8 possible outcomes. Now, suppose for this experiment, again, I am asking two questions. The first question I am asking is, I want to know how all of this 3 tosses how many tosses will be heads? In other words, I am counting the number of heads in each of the tosses. The second question I am asking is, of the 3 tosses, which toss results in a head first, what do I mean by it, is my first toss a head, or the second toss a head, or the third toss a head.

For example, in this outcome, my first toss is a head. In this outcome, my first toss is a head. In this outcome, again, the first toss is a head. In this, my first toss is a head. In this, my second, the head appears for the first time in the second toss, here also the head appears in the for the first time in the second toss, here, it appears for the first time in the third toss, here, it does not appear at all.

Now, when I am counting the number of heads, I know I have 3 heads here, I have 2 heads here, I have 2 heads here, I have 1 head here, I have 2 heads here, I have 1 head here, I have 1 head here, I have no head here. With a caution and not put a number here because here, I am not telling it is a zeroth toss because I do not know whether what is a zeroth toss. So I just left for now I am calling this nil.

So, you can see that on the same sample space and with the same outcomes we have defined 2 random variables or 2 numerical quantities which we seek to answer. Again, observe that the experiment and the sample space are the same.

(Refer Slide Time: 33:33)

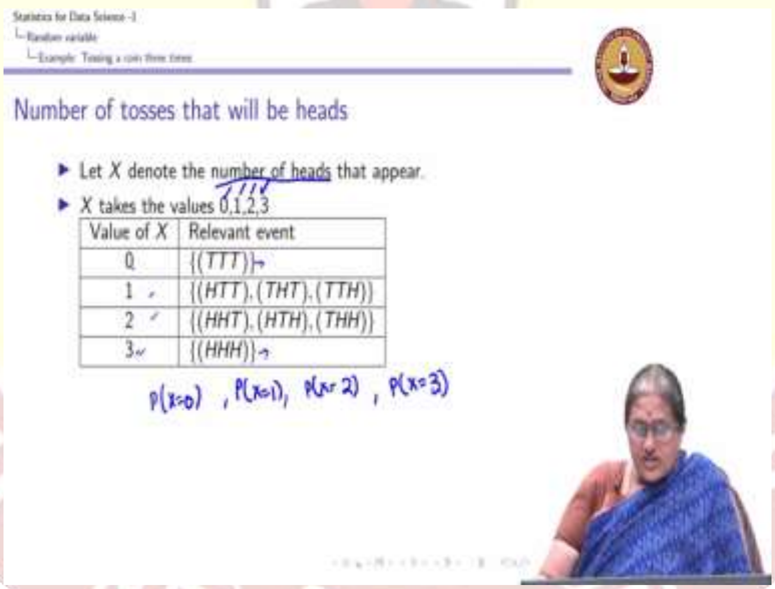


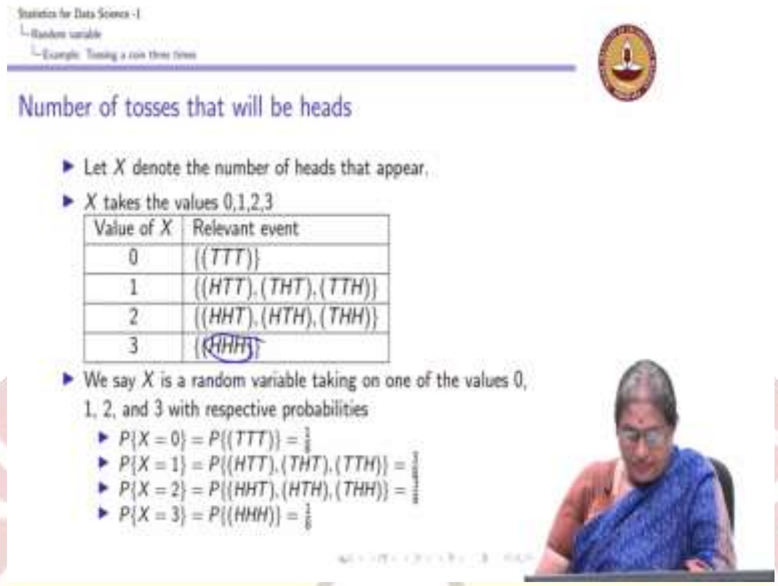
So, now let us define X to be the number of heads that appear. And Y to be that toss in which a head appears first that is the order of the toss, whether is the first toss, or the second, or the third toss. Again, as in the previous example, for each outcome, let us find out what is the value of X and what is the value of Y. So, what do I have here? What are the outcomes I have these 8 outcomes.

In this outcome I know the value of head there are 3 heads, so value of X is 3, and the head is appearing in the first store. So, the value of Y is 1. Here I have 2 heads, head appears in the first toss so value of Y is 1, X is 2. Again access to a head appears in the first toss to and 1 head appears in the first toss and there is only 1 head so both X and Y take the value 1, here head appears in the second toss, Y takes the value 2 and there are 2 heads.

Here again head appears in the second toss Y value is 2 but there are only 1 head, head appears in the third toss and there is only 1 head here there are no heads so X takes the value 0, whereas here, I am writing nil or none because Y does not appear does not appear in any of the tosses so I am just assigning a value nil. Now this value could be a very high value or not. But what I want you to see is the Y takes a value corresponding to this, what is the value I need to give to this nil is something it could be any real value. For now, I want you to understand that these are the values that Y takes. So, Y and X are defined on the same sample space.

(Refer Slide Time: 35:33)





And for each outcome, I have a value. So, now let us look at the X that is number of heads in the outcome. So what are the values this X is taking? So again, go back here, you can see that X takes the value 0 1 2, and 3, so X takes 4 values, and what are the values X is taking? X is taking the value 0 1 2 and 3. So, now let us look at the relevant events where X takes the value 0, I know X takes the value 0 means that all the 3 tosses result in a tail. So, the relevant event is TTT. X takes the value 1 when 2 tosses, so 1 of the tosses is a head. So, I have the first toss, second toss, third toss. So, the head in the first toss or head in the second toss or head in the third toss.

So, the possible ways it can happen are the following. So, what are the possible events, the relevant event of X taking the value 1 is HTT, THT or TTH. Similarly, X is taking the value 2. So I can have my first second and third toss, H can be the first 2 tosses, or the first and the third toss or the second and the third toss. So, the relevant event is all these 3 outcomes put together, which is HHT, HTH, and THH.

Similarly, X takes the value 3, if all the 3 tosses result in head, and that can happen only in 1 way. So, the relevant event is again, HHH. So what we have done now is I know x, which is the number of heads in an outcome that appear is taking the value 0, 1, 2, 3 X takes the value 0, the relevant event is all my 3 tosses are tails. And for X = 3, the relevant event is all the 3 tosses are heads, and for 1 and 2, I have listed what are my relevant events?

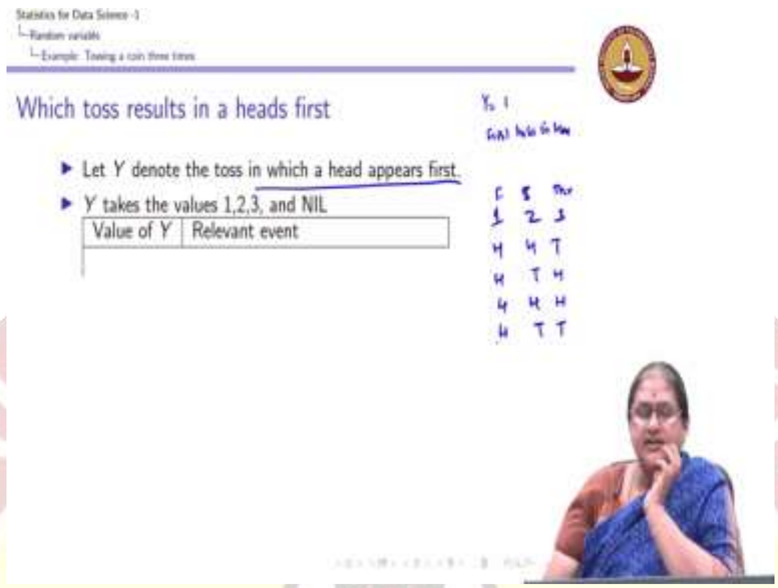
So, as in the earlier case, I can find out what is the probability of X = 0, X = 1, probability of X = 2 and probability of X = 3, which are nothing but the probability of the relevant events. So, what is probability of X = 0? Probability of X = 0 is same as probability of this event TTT happening and I know that that is equal to 1/8, because my sample space has 8 equally likely outcomes. Similarly, probability of X = 1 is probability of this happening, which is equal to 3/8.

Again, in the assumption of equal likelihood probability of X = 1 is 3/8. What is probability of X = 2? Again, there are 3 outcomes probability of X = 2 is also 3/8, whereas probability of X =3 is just 1/8 because there is only 1 outcome that satisfies the condition that X = 3.

So, what we have done here to see that from a sample space of tossing a coin, or an experiment of tossing a coin 3 times we list down the sample space define the random variable to be the number of heads we have got, what are the values his head can take, and what is the probability with which X takes those values. Now for the same sample space, we are interested in knowing which toss results in a head first. To understand this, let us go back to our earlier table.

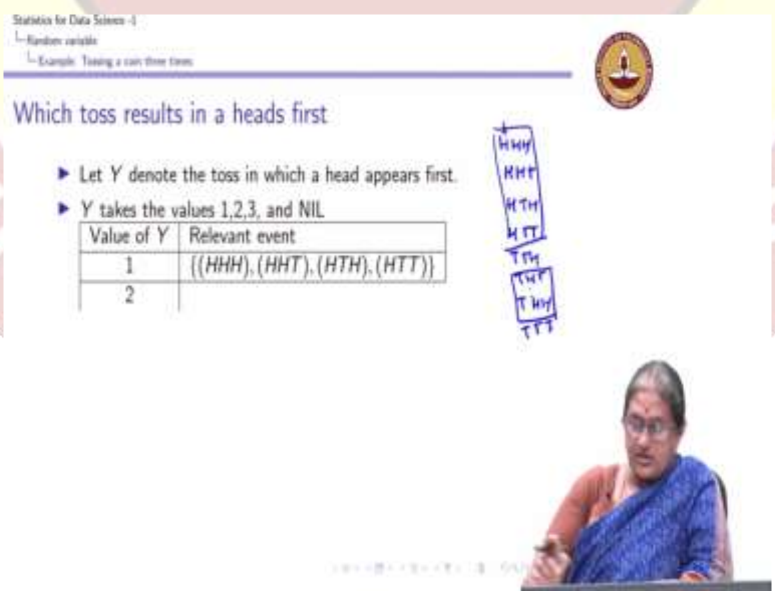
So, if you look at this table, you can see that Y takes the values what are the Y values, Y takes the value 1 2 3 and nil these are the values Y is taking. As again now I am not assigning a numerical value here, but I could assign a numerical value, I could assign a value 0, but for now I am just writing nil because Y equal to 0, what physically I can interpret it as the zeroth toss, this does not make meaning. So, I am just writing that there is no toss corresponding to this outcome where the first (toss) head appears for the first time. So, you can see that Y takes these 4 values that is what I can see that Y takes these values,

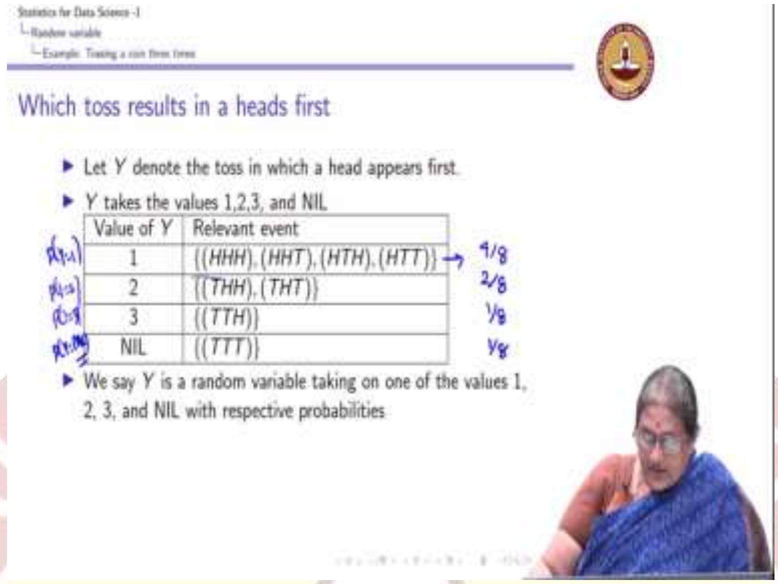
(Refer Slide Time: 40:56)



And now, again, let us see what when Y takes the value 1, when we take the value 1 again, Y would take the value 1 is same as first toss is head that is in which toss the head appears first So, again, I have my first toss, my second toss, my third toss that is 1 2 3 my first toss is a head, then my second toss could be a head or a tail, tail or a head, head or a head or tail or tail. Is there any other possibility that can happen other than this?

(Refer Slide Time: 41:41)



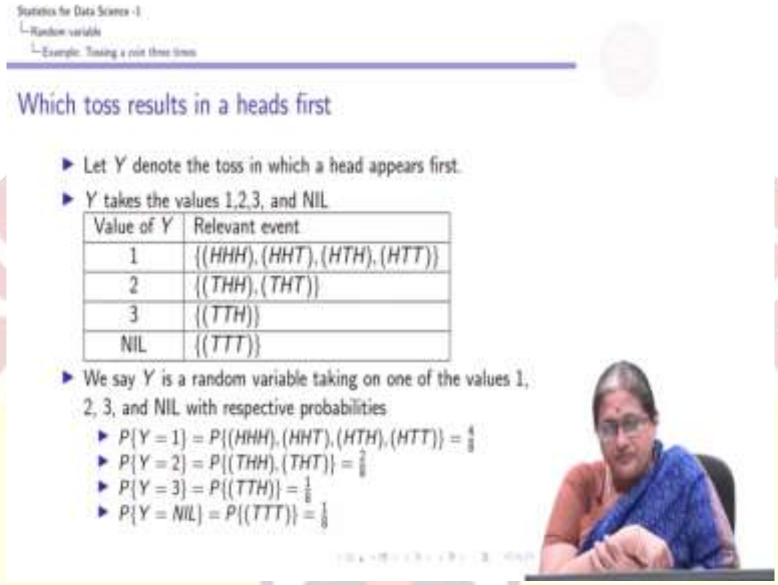


So, you can see that if my first toss is a head, the relevant events that could happen as head with both the tosses falling as a head or head tail head or head tail, tail head or head TT. So, these are the relevant events. So, now, let us look at Y taking the value 2. So, head Y is denoting the toss in which head appears first. So, again, I look at this HHH, HHT, HTH, HTT, TTH, THT, THT, THH and TTT. These are my possible outcomes. So, here I have head appearing for the first time in the first 4 tosses. Now, in this among these 2 tosses, I have head appearing first in the second toss for these 2,

So, the outcomes that satisfy that head appearing first for the first time and the second toss is THH and THT. For the third time again, we can see for the third time, it would be only this outcome, which is TTH, and that outcome, X head appearing for the first time in the third toss is given by TTH. And then of course, in the nil, which corresponds to this outcome, because head does not appear at all, and the way head does not appear. So, head appears first does not happen in this outcome that the head appears first.

So, these are the values that Y takes Y takes values 1, 2, 3 and nil. As earlier, I can also associate probabilities with the values Y takes, and what would be these probabilities. Again, you can see that the probability of Y taking the value 1 is equivalent to the probability of this event happening. Remember, there were 8 outcomes in my sample space, I have a 4/8. Probability of Y = 2 items, or 2/8, Y = 3 is just 1/8, probability Y equal to nil again, will come back to what is this nil is again, 1/8. I can represent this nil with any real valued number, which makes sense but for now, I am just telling that Y takes the value nil.

(Refer Slide Time: 44:25)

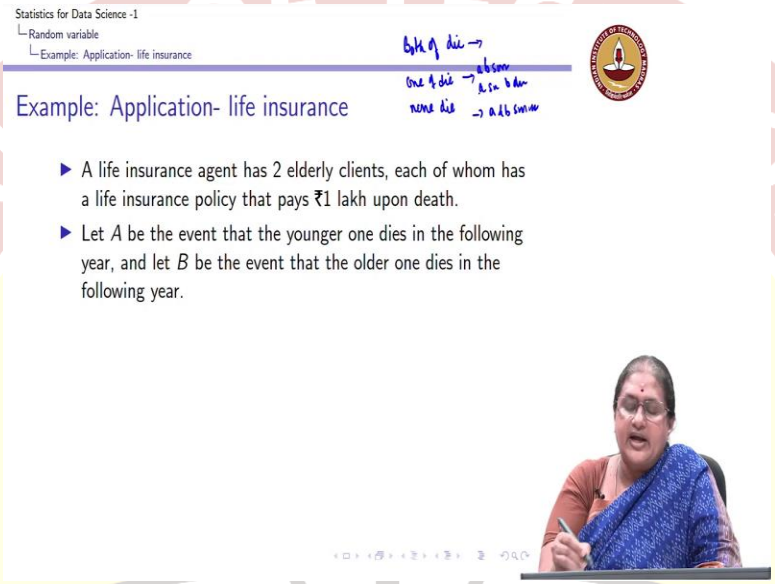


Or if you say test the value 0, you should qualify by saying that what do you mean by Y taking the value 0, by taking the value 0 means that my outcome, the head never appears first in any of the tosses. So, probability Y = 1 is 4/8, Y = 2 is 2/8, Y = 3 is 1/8 and Y takes the value in nil is 1/8. So, what we have seen in the earlier two examples is given a random experiment and sample space I am associating some numerical quantity to each outcome and, and using this concept of a numerical quantity to answer some things about the experiment.

For example, it could be some of the dice or the lesser of the 2 dice or the number of heads that appear in each outcome or what is a count the number or the order of the outcome for which the head appears for the first time.

Statistics for Data Science 1
Professor. Usha Mohan
Department of Management Studies
Indian Institute of Technology, Madras
Lecture No. 8.2
Random Variables - Application

(Refer Slide Time: 00:15)

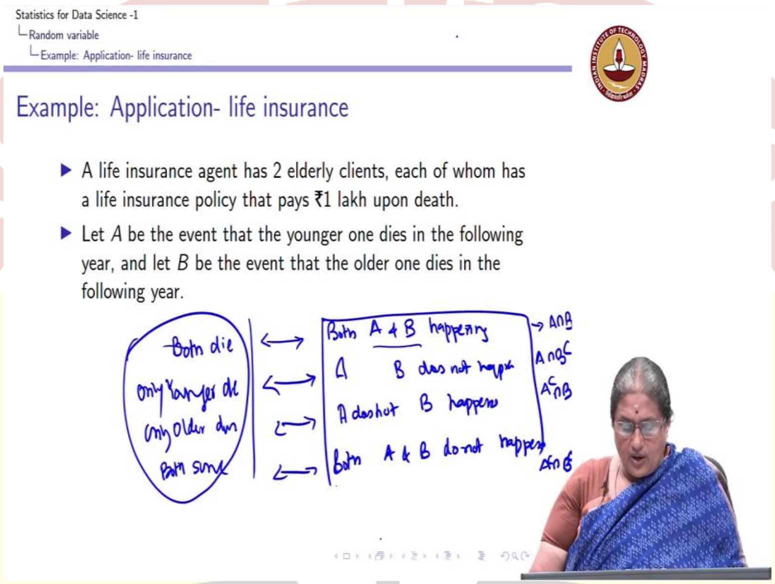


We just looked at a few generic examples of how we can define a random variable, we talked about mapping the random variable to a particular event. And talking also now, given the probability of an event we associated probability of the random variable taking a particular value. So now, let us look at an application wherein we define a random variable and we get the probability of these random variables through the events they correspond to the application is application from life insurance.

So, I have a life insurance agent who has two elderly clients. Both of them have a life insurance policy that pays 1 lakh upon death rupees 1 lakh upon death. So, this is a hypothetical case, I have two people, and both of them have insured life insurance and they will get a pay-out that is the people the survivor, the nominees should get a pay-out of 1 lakh upon the death. So, I am defining two events. Now, what are what is a random experiment here?

So, the random experiment is I want to know, in the next year, in the following year, what are the possible things that could happen, the possible things that could happen is both of them die in the following year, both of them die, or one of them die, or none of them die, these are the possible things that could happen in the following year. Now, when both of them die, I know both die one of them die. So, if I am calling A and B, A dies, B survives or A survives and B dies, or both A and B survive, these are the possible outcomes if I am talking of this in terms of a random experiment.

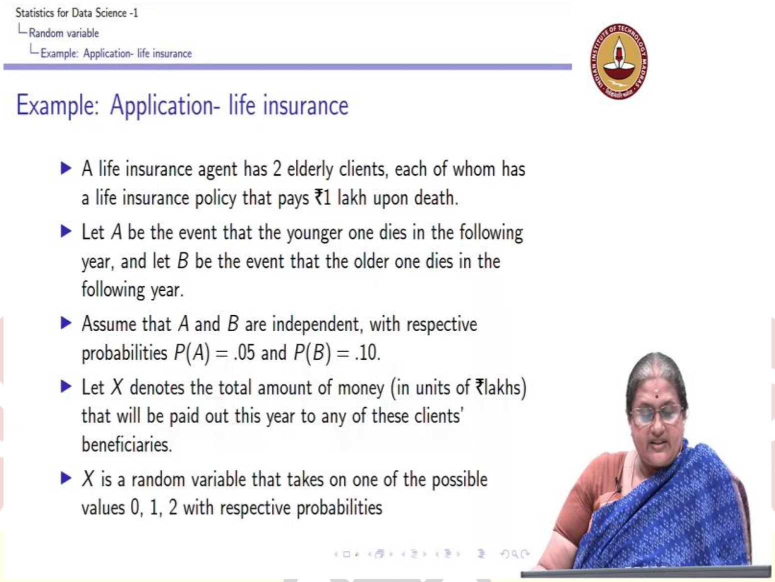
(Refer Slide Time: 2:30)



So now, let A be the event that the younger one dies. So, between the two of them I assume one is older, and one is younger, so I am calling A I am letting A to be the event that the younger one dies in the following year, and B to be the event that the older one dies in the following year. So, I said what are the outcome, the outcome is both die. So, this corresponds to both event A and B happening, younger one dies older one survives only A happens B does not happen. Older one dies only older one only younger one dies, only older one dies so, here A does not happen B happens both survive both A and B do not happen. I hope this is clear. So, I have just first these are the possible outcomes.

And these I have listed or mapped each one of the possible outcomes to in terms of the event, when both die I say both events A and B happens and I know mathematically I can express this as $A \cap B$ only the younger one dies, but the older one does not die, it is $A \cap B^c$. A does not die A^c happens, but and B happens $A^c \cap B$. Both A and B do not die, So, $A^c \cap B^c$. So, in terms of events, we can translate whatever we are expressing in plain English in the mathematical abstraction.

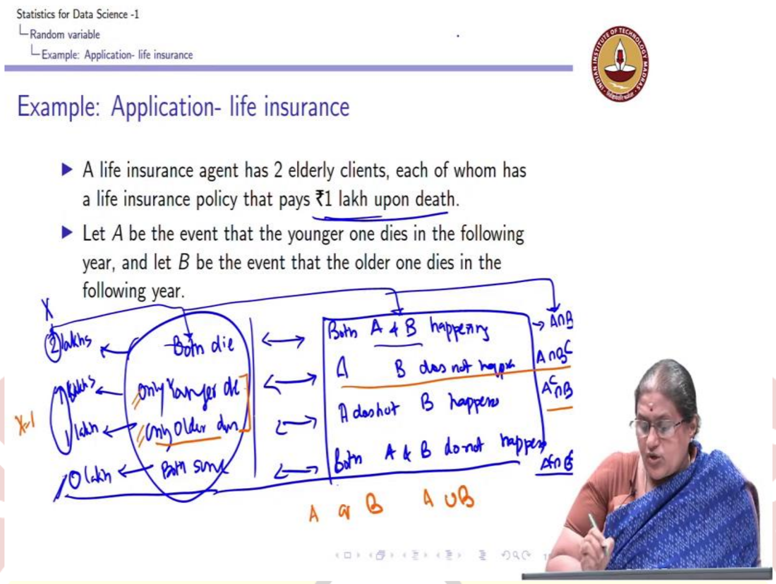
(Refer Slide Time: 04:42)

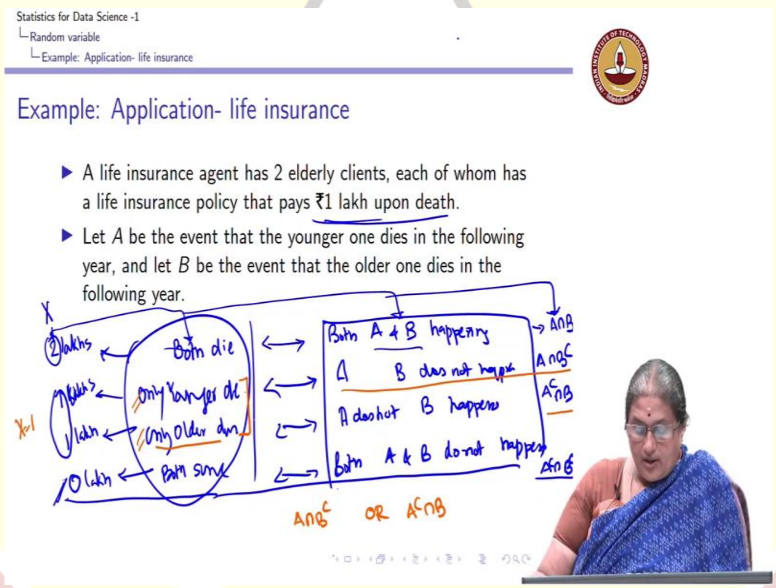


Now, we want to associate so now the next thing is, I assume A and B are independent. This is absolutely a very valid assumption because when I am talking about two people dying, it need not be the case that once death affects anybody so the events are independent and I am associating some subjective probabilities perhaps I am associating a probability A which is very low with the younger one and probability B which is point one with the older one.

Now, let X denote the total money that the company pays out in this year to its clients beneficiaries in amount of lakhs. So, what is that the company note will pay out money only if there is a death, let us assume that the company pays out the money only if there is a death. So, x is a total amount of money that is paid out to any one of the clients beneficiaries.

(Refer Slide Time: 05:50)





So again, let us go back to our earlier thing I had both dying. So, when both die, the company has to give money. When both die it gives money to both the beneficiaries, when the younger one dies it gives and both of them have insured for the same amount, so when both die both of them have insured for 1 lakh upon death, so I will have to pay 2 lakhs to both the beneficiaries only the younger one dies, I pay 1 lakh to the younger ones family, older one dies, I pay 1 lakh to the older one family, when both of them survive, I do not have to pay anything. So, my payment is 0 lakhs.

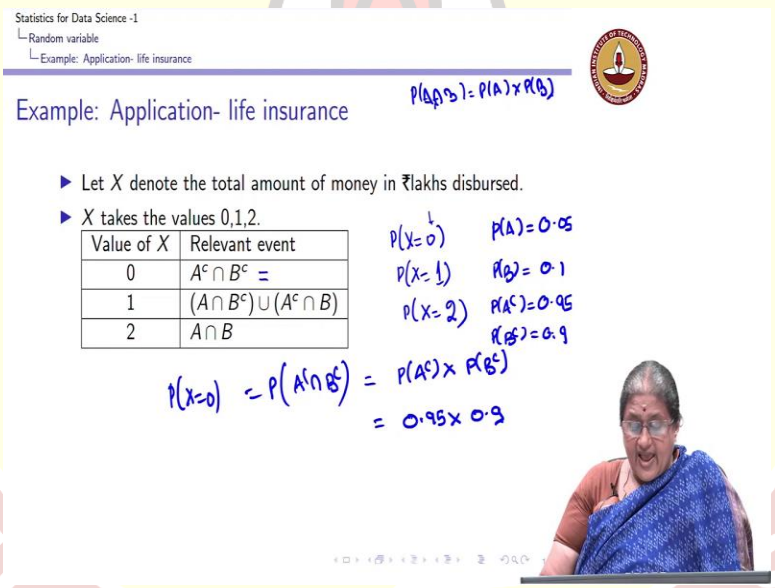
So, if you look at the value this X is taking, it is taking the value 0, it is taking the value 1, and it is taking the value 2. I repeat the value X is taking X is taking the value 0, X is taking the value

1, and X is taking the value 2. And you can see that X taking the value 2 is corresponding to the event that both of them die which again corresponds to both A and B happening which is written as $A \cap B$.

Similarly, the value of 0 corresponds to both surviving which is same as both A and B do not happen, which is $P(A^c \cap B^c)$. But you can see that the probability of X taking the value 1 can happen if only the younger one dies, or only the older one dies only the younger one dies corresponds to $A \cap B^c$, only the older one dies corresponds to $A^c \cap B$.

So, if one of these two events happen, I know X takes the value 1 we have already seen earlier, A or B happens is represented by $A \cup B$ here I am not talking about A or B, I am talking about $A \cap B^c$, or $A^c \cap B$ for the value of X to be equal to 1.

(Refer Slide Time: 08:23)

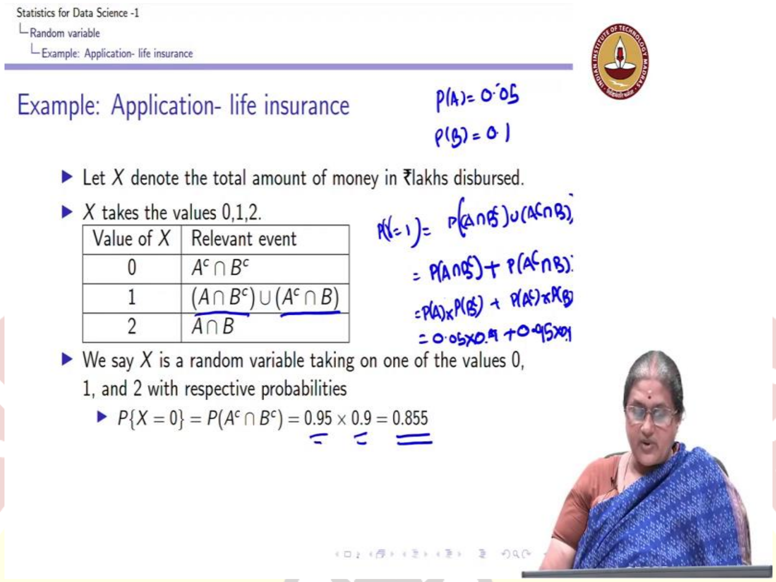


So we can summarise the discussion by saying that X takes the values what are the values X takes? X takes the value 0, the relevant event is A compliment intersection B complement that is what we just established a couple of minutes before that this A X takes the value 0 is with A compliment intersection B complement takes the value 1 with $(A \cap B^c) \cup (A^c \cap B)$.

And X takes the value 2 with $P(A \cap B)$ it is equivalent to A intersection B so X takes the value 0 relevant event is $A^c \cap B^c$ 1 it is $(A \cap B^c) \cup (A^c \cap B)$, it takes the value 2 these are the only 3 values X can take is $(A \cap B)$. Again as in the previous case, I can associate probabilities to I want to know what is the probability with which X does not give out any money probability with it X gives out a pay out of 1 lakh probability with which X gives out a pay out of 2 lakhs. See these are the numerical values associated with the outcomes of the experiment.

So, you can see that probability X equal to 0 is same as $P(A^c \cap B^c)$, which I know if A and B are independent, A^c is also independent of B^c. And we also know$(A \cap B)$ is $P(A) \times P(B)$, this is a multiplication rule. So, I have this is $P(A^c \times B^c)$. What is given to us is $P(A)$ is 0.05. $P(B)$ is 0.1, which gives us $P(A^c)$ is 0.95 and $P(B^c)$ is 0.9. Hence, I have $P(A^c)$ is 0.95 into a 0.9.

(Refer Slide Time: 11:08)

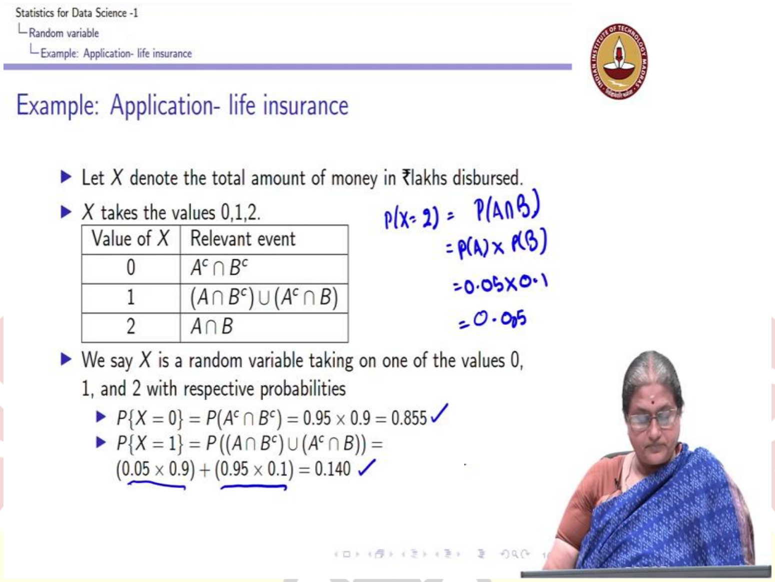


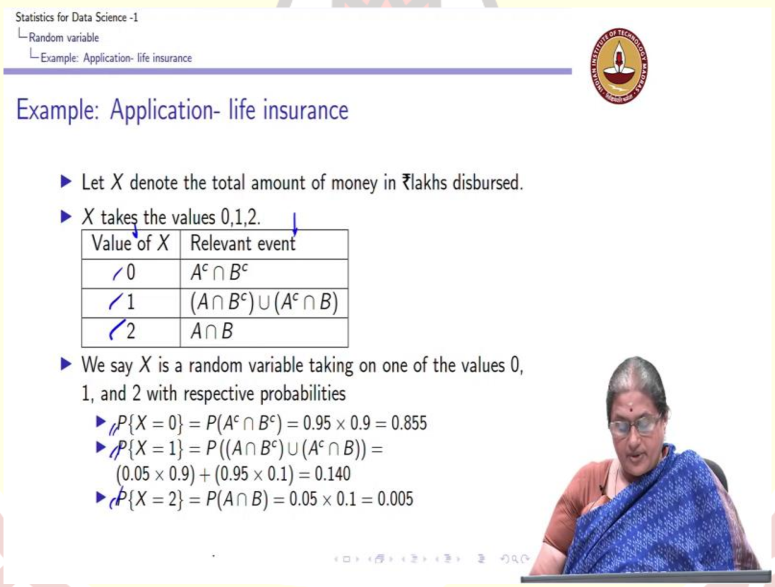
So, I have probability X takes the value 0 is 0.95 into 0.9, which is given by 0.855. Similarly, $P(X) = 1$, is again $P((A \cap B^c) \cup (A^c \cap B))$., we can check that these two are disjoint sets. And we know for any two disjoint sets $P(A \cup B)$ is $P(A) + P(B)$.

This is my addition rule for disjoint sets, I can use that I know probability X equal to 1 is $P((A \cap B^c) \cup (A^c \cap B))$, which is $P(A \cap B^c) + P(A^c \cap B)$. Again, if A and B are independent, A is independent of B^c. So, this would be $P(A) \times P(B^c) + P(A^c) \times P(B)$.

Again, we know $P(A)$ is 0.05 $P(B)$ is 0.1. So, I have this as $0.05 \times 0.1 + 0.95 \times 0.9$ here 0.9 here into 0.1 0.05 into 0.9 plus 0.95 into 0.1 that is my probability X equal to 1 which is $0.05 \times 0.9 + 0.9 \times 0.1$ which is 0.140.

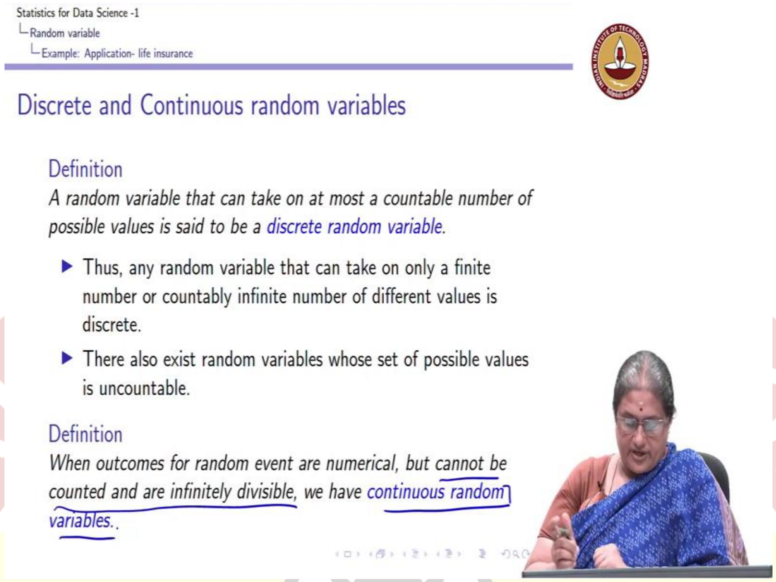
(Refer Slide Time: 12:59)





Finally, I want to know what is probability X equal to 2, this is $P(A \cap B)$ again A and B are independent. So, this is $P(A) \times P(B)$, we know $P(A)$ is 0.05 $P(B)$ is 0.1. So, I can get the probability of A X equal to 2 is 0.005. So, what we have seen is, I have from given the events A and B and the respective probabilities I am associating the relevant events to the values X can take to get what is the $P(X = 0)$, $P(X = 1)$, and $P(X = 2)$, . So, this is a simple application of how we define a random variable and associate probabilities to that random variable.

(Refer Slide Time: 14:02)

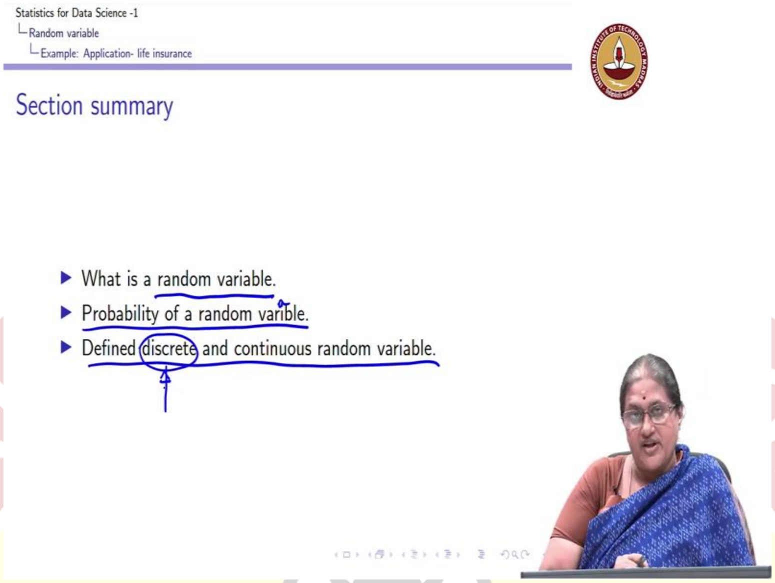


So, what is next, so if you look at what we are going to do in the first portion of the next few weeks is we are going to look at discrete random variables. So, what is a discrete random variable, a discrete random variable can be formally defined as a random variable that can take on at most a countable number of possible values. In other words, we will look at examples where a random variable can take on a finite number of values.

For example, X takes the value 0, 1, 2, 3 finite number of values or X could take a value x 1, 2, 3, 4. So, for countably infinite but it has to be countable. You have already learned about what is countable sets or what is properties of countability in your mathematics courses, but what I want you to understand is a random variable which can take on a finite number or a countably infinite number is referred to as a discrete random variable.

Now, there could be random variables whose set of values are uncountable. Now, I am not going to formally define a continuous random variable, but then, when the outcomes of a random variable are numerical, but cannot be counted on and infinitely divisible, we have continuous random variables. The formal definition of a continuous random variable is slightly more involved. I will come to it when we specifically look at the continuous random variable case. But for now, we are going to focus on discrete random variables.

(Refer Slide Time: 16:01)

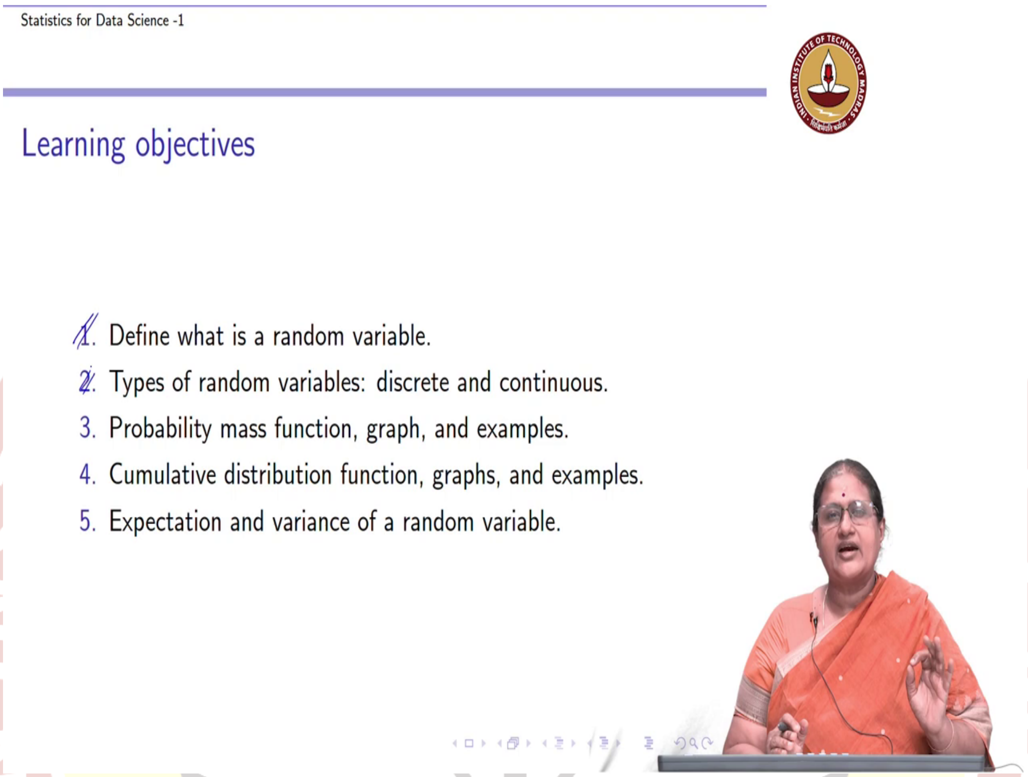


So, in summary, what we are learned in this module is, we introduce the extremely important concept of what is a random variable, namely a numerical quantity associated with outcomes of a sample space. And then through actually mapping it to relevant events from the sample space, we computed what was the probability of a random variable, this is what we did. Then, we define what are discrete and continuous random variables, we did not give a very rigorous definition of a continuous random variable. But whenever a random variable takes countable number of values, we typically refer to them as discrete otherwise, and we refer to them as continuous random variables.

So, moving forward, we would focus on a discrete random variable. And how do we describe these discrete random variables, in terms of its distribution is what is going to be of interest next.

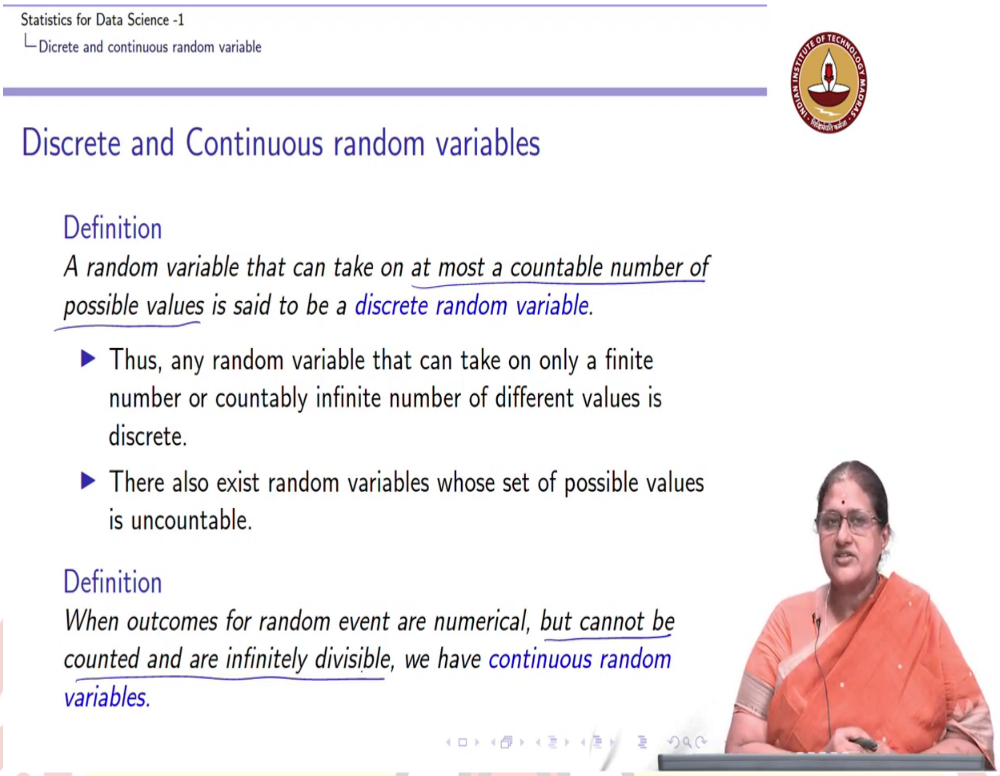
Statistics for Data Science 1
Professor Usha Mohan
Department of Management Studies
Indian Institute of Technology, Madras
Random variables - Discrete and continuous random variable

(Refer Slide Time: 00:14)



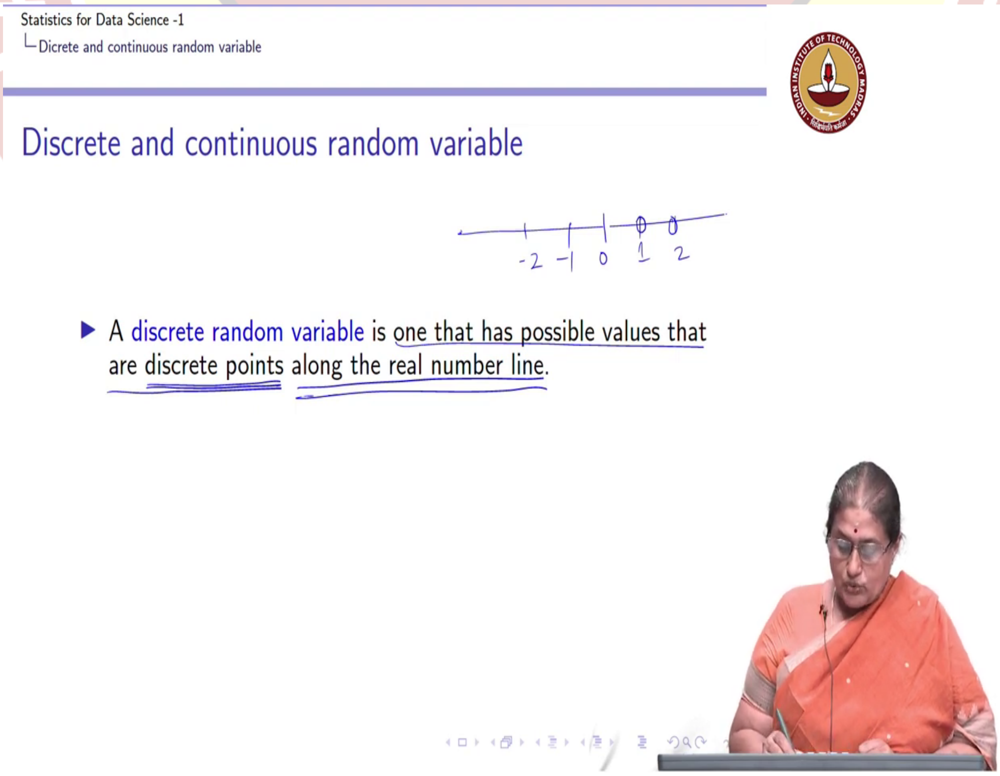
So, in the last lecture, we understood what was a, we defined what was a random variable. So, now we are going to look at, and towards the end of last lecture, we defined what was a discrete and continuous random variable.

(Refer Slide Time: 00:35)



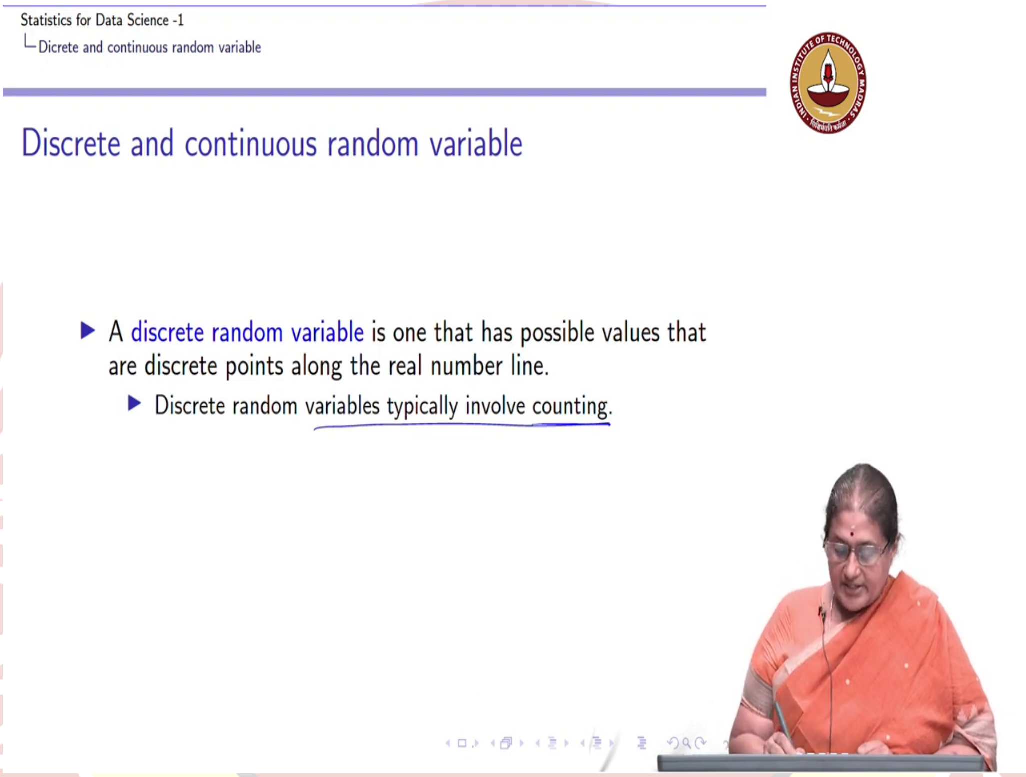
So, we will spend some time to understand more about what is a discrete and continuous random variable. So, these were the definitions which we check, which we gave in the last session, which was we define the discrete random variable as a random variable that can take on at most a countable number of possible values. Whereas we said a continuous random variable is it cannot be counted, or it is infinitely divisible. Now, let us understand what is a discrete random variable and a continuous random variable through a example.

(Refer Slide Time: 01:25)



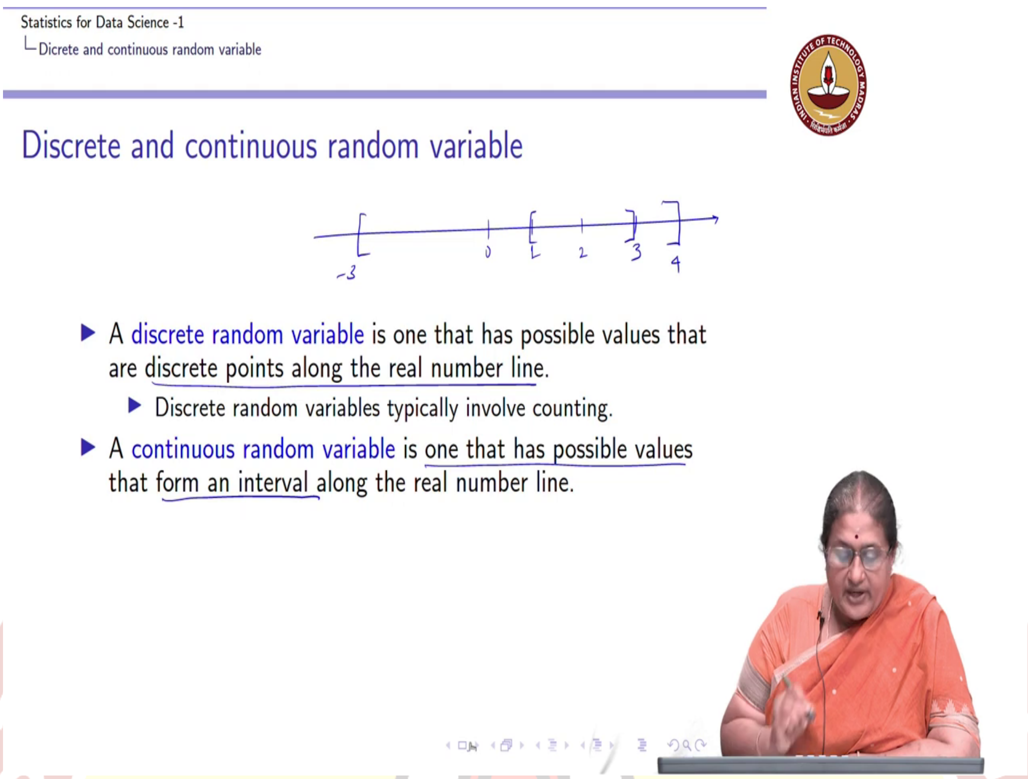
So, instead of going and understanding a definition, I can also say a discrete random variable is a random variable that can take possible values that are discrete points, I am again, emphasising that are discrete points. So, if I have my real number line here, so I have a 0 here, I have a 1 here, I have a 2 here, it could take a -1, it could take a -2. So, these are isolated discrete points along the real number line. I repeat, a discrete random variable is a random variable that can take possible values, along a real number line that are discrete points, the key idea is they are discrete points.

(Refer Slide Time: 02:22)



So, typically, when we talk about discrete random variables, you involve the operation of counting number of heads, you are counting.

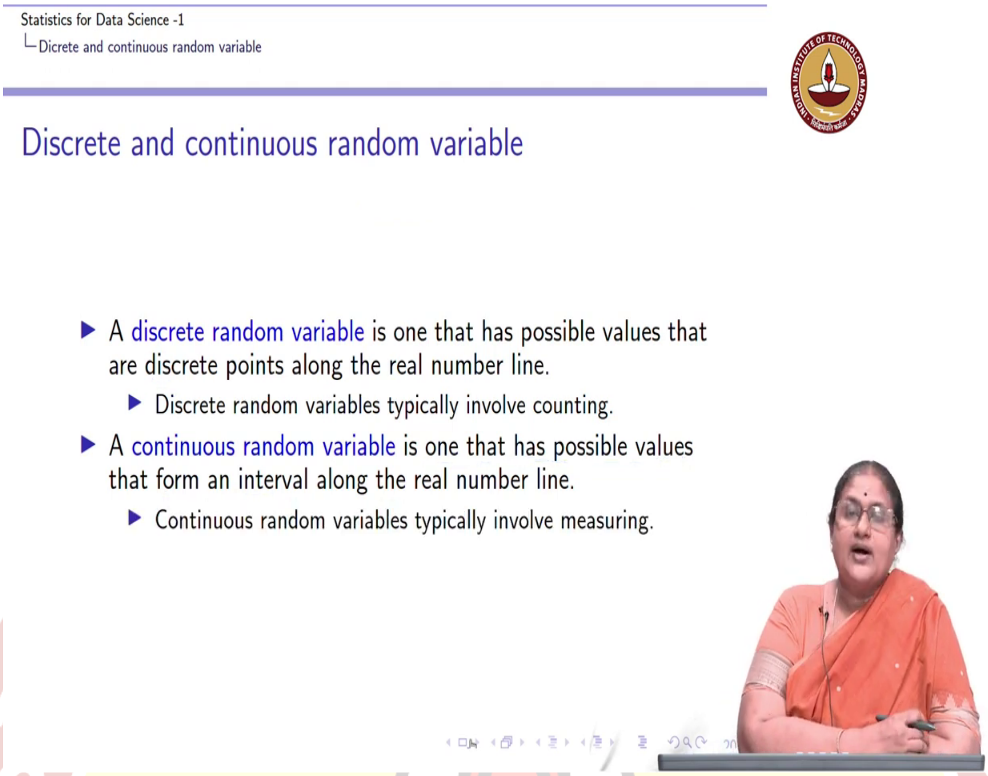
(Refer Slide Time: 02:38)



Whereas a continuous random variable is the random variable that has possible values. So, if I again go back to my real number line, it will not take. So, I have it will not take isolated discrete values, but it would take values that might form an interval along the real number line, can it be this interval, it could be, it could be -3 to $+4$, it could be, but it takes the values in an interval.

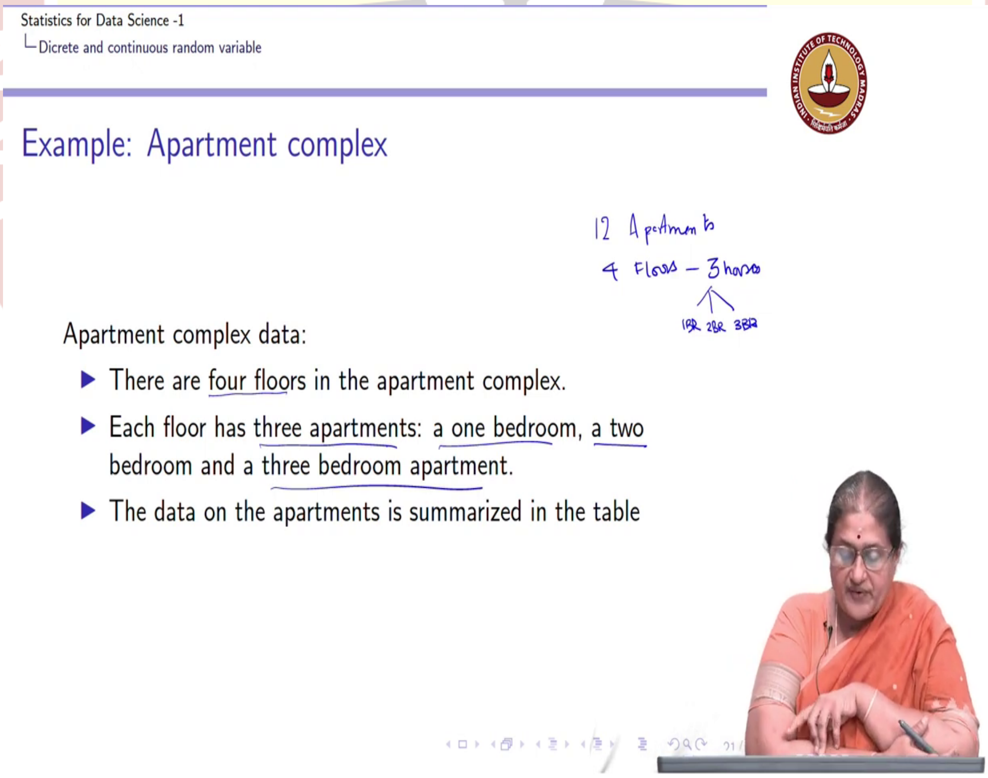
A more rigorous definition of what is a continuous random variable would be offered in advanced courses. But at this point of time, you need to understand a discrete random variable, will take discrete points along a real number line, whereas a continuous random variable would take values in an interval along a real number.

(Refer Slide Time: 03:41)



So typically, a continuous random variable would involve a measurement or something whereas a discrete random variable involves a counting.

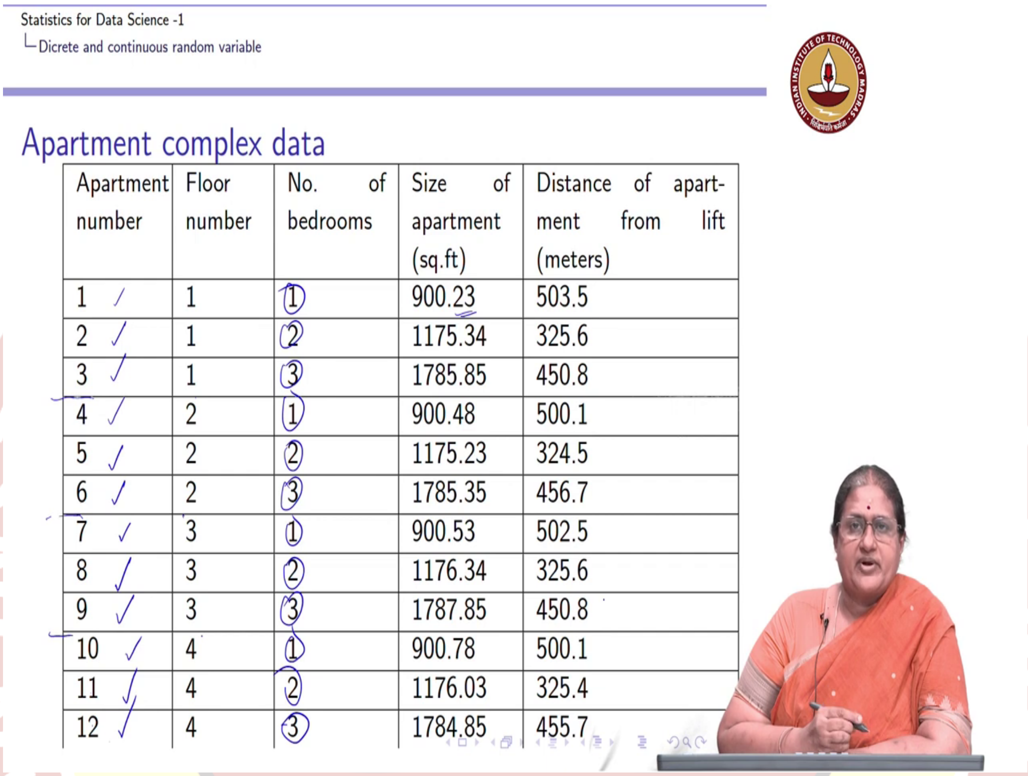
(Refer Slide Time: 03:55)



So, now let us motivate and understand this through an example. Let us look at an apartment complex. So, what is this apartment complex? In apartment complex there are 4 floors in an apartment complex. Each floor has 3 apartments. So, if I am looking at this apartment complex, each floor has a 3 apartments in so it has a 1-bedroom apartment, a 2 bedroom apartment and a 3 bedroom apartment.

So, there are totally 12 apartments. There are 4 floors, there are 3 houses per floor. And in these 3 houses or 3 apartments. First is a 1 bedroom I have a 2 bedroom and a 3 bedroom apartment. This is how this apartment complex has been structured.

(Refer Slide Time: 04:51)



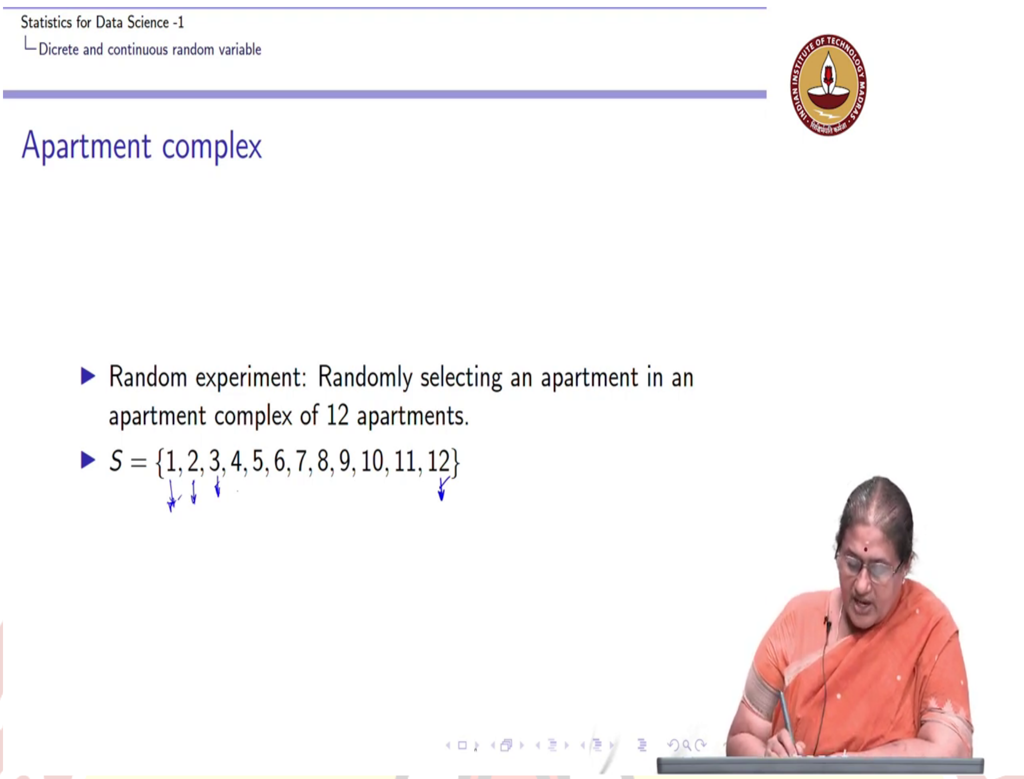
Apartment number	Floor number	No. of bedrooms	Size of apartment (sq.ft)	Distance of apartment from lift (meters)
1	1	1	900.23	503.5
2	1	2	1175.34	325.6
3	1	3	1785.85	450.8
4	2	1	900.48	500.1
5	2	2	1175.23	324.5
6	2	3	1785.35	456.7
7	3	1	900.53	502.5
8	3	2	1176.34	325.6
9	3	3	1787.85	450.8
10	4	1	900.78	500.1
11	4	2	1176.03	325.4
12	4	3	1784.85	455.7

So, let us look at the data which is summarised in the following table. So, if you look at the data, I have 12 apartments that is given by the cases 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 these are my 12 apartments. So, you see the first 3 apartments are in the first floor 4, 5, 6 are in second floor 7, 8, 9 are in third floor and 10, 11, 12 are in fourth floor. When you look at the number of bedrooms, the first apartment, fourth apartment, seventh and tenth have 1 bedroom each.

The second, fifth, eighth and eleventh have 2 bedrooms each. Whereas the third, 6, 9 and 12 have 3 bedrooms each. So, this is the data I have collected. Now, even though typically you might expect the sizes of apartments to be the same, but there is some error and I am allowing I am correct, computing the sizes of the apartment, correct to 2 decimal points in square feet.

So, the first apartment is 900.23 $sq.feet.$, the fourth is 900.48, the seventh is 900.53 $sq.feet.$ So, all the measurement of the sizes of the apartment in $sq.feet$ for each of the apartment is given. In addition, we also are capturing how many metres from the left is each one of the apartment. This is just fictious data, this is just data from an apartment complex. And we are capturing it in this following way. Now, I can look at this as a random experiment.

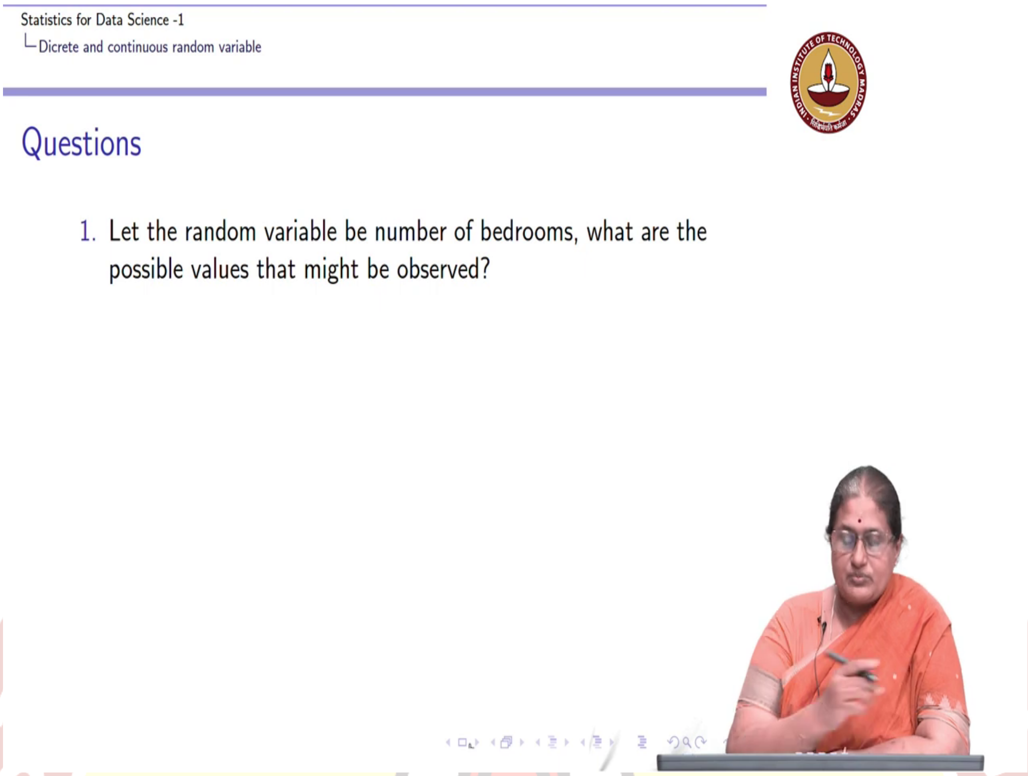
(Refer Slide Time: 06:52)

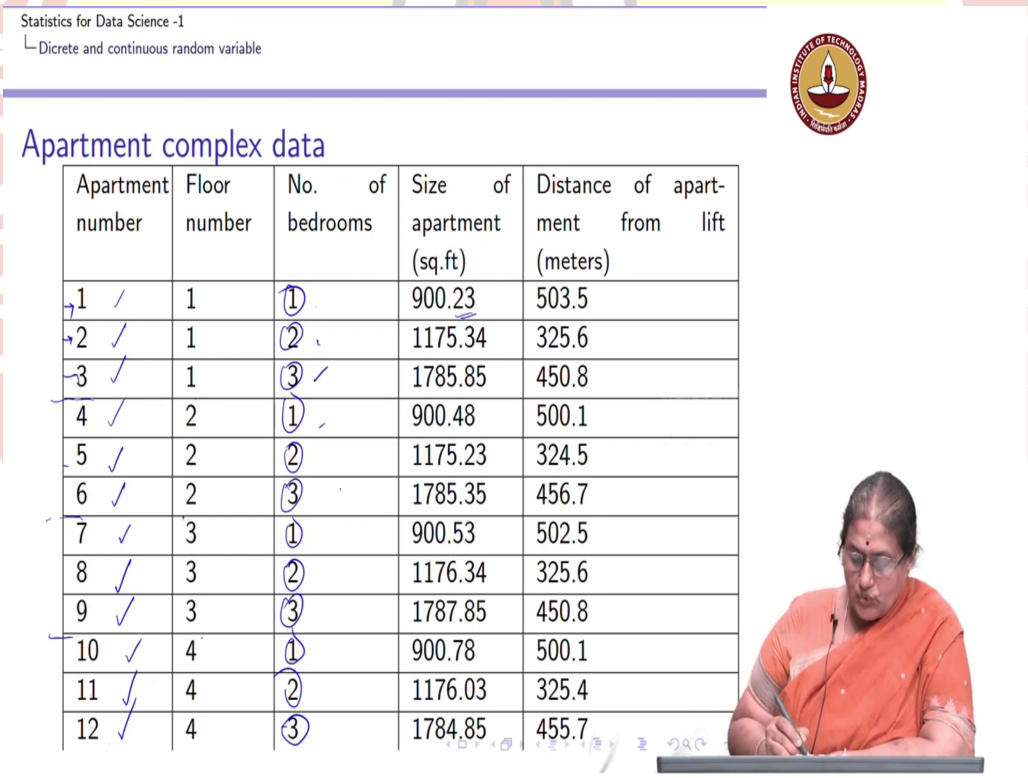


Now, what is the random experiment in this case? Now, let me define the random experiment is I am randomly selecting an apartment from these 12 apartments, I am selecting one apartment from these 12 apartments. So, the sample space of my random experiment is any one of the 12 apartments. So, the outcome is I could either have selected the first or the second or the third, so forth up to the 12th.

So, this constitutes the sample space of my experiment. The outcomes is one, one represents the first apartment, two represents second, third and so forth, 12 represents the twelfth apartment. Now, recall when we talked about a random variable, we said that a random variable associates some quantity with every outcome of the sample space.

(Refer Slide Time: 08:05)

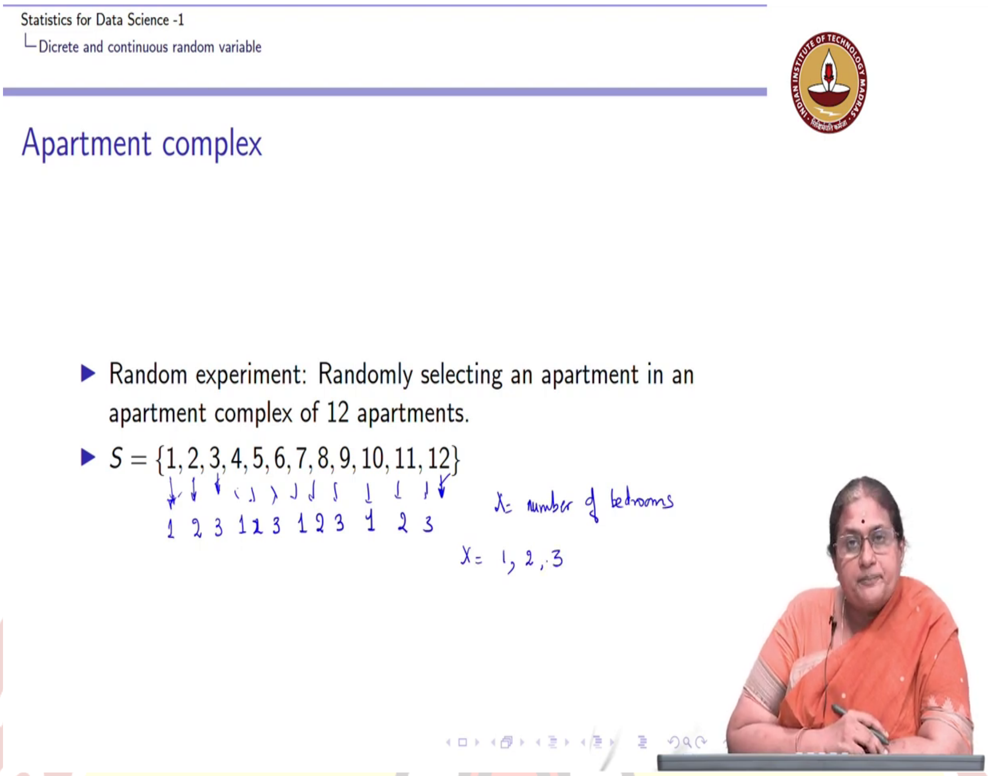




Apartment number	Floor number	No. of bedrooms	Size of apartment (sq.ft)	Distance of apartment from lift (meters)
1	1	1	900.23	503.5
2	1	2	1175.34	325.6
3	1	3	1785.85	450.8
4	2	1	900.48	500.1
5	2	2	1175.23	324.5
6	2	3	1785.35	456.7
7	3	1	900.53	502.5
8	3	2	1176.34	325.6
9	3	3	1787.85	450.8
10	4	1	900.78	500.1
11	4	2	1176.03	325.4
12	4	3	1784.85	455.7

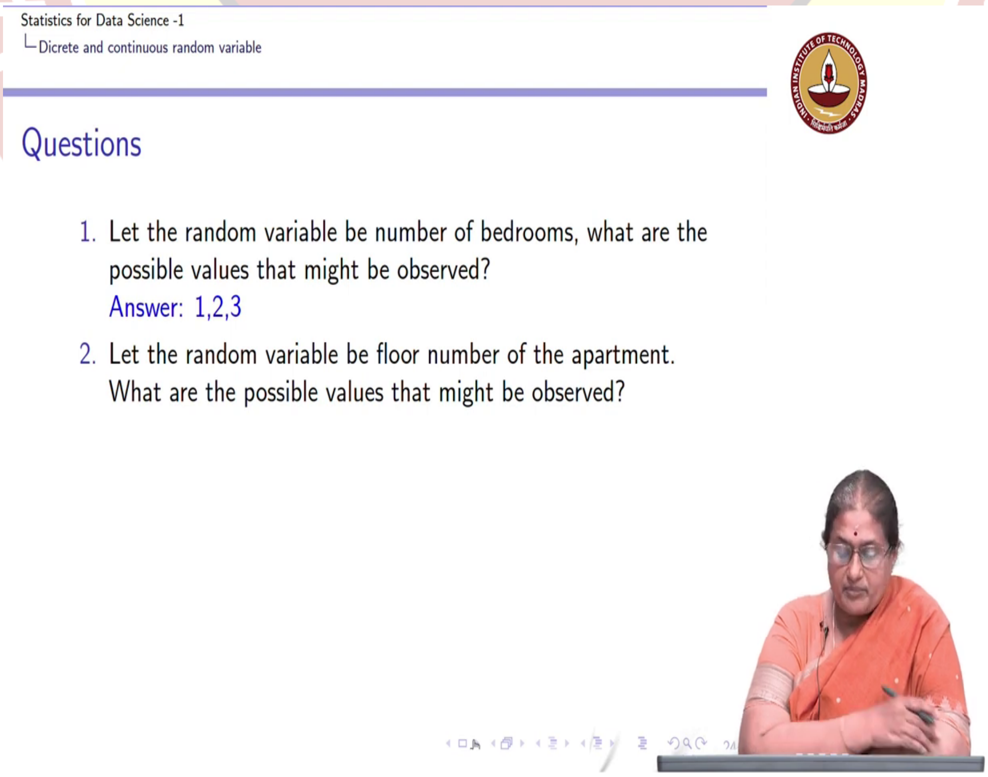
Now, suppose I am asking the following questions. Let a random variable be the number of bedrooms. So, let us go back to our data, I can see that the random variable that is associated with my first outcome is 1, the random variable that is associated with the second is 2, third is 3, fourth is again 1, fifth is 2.

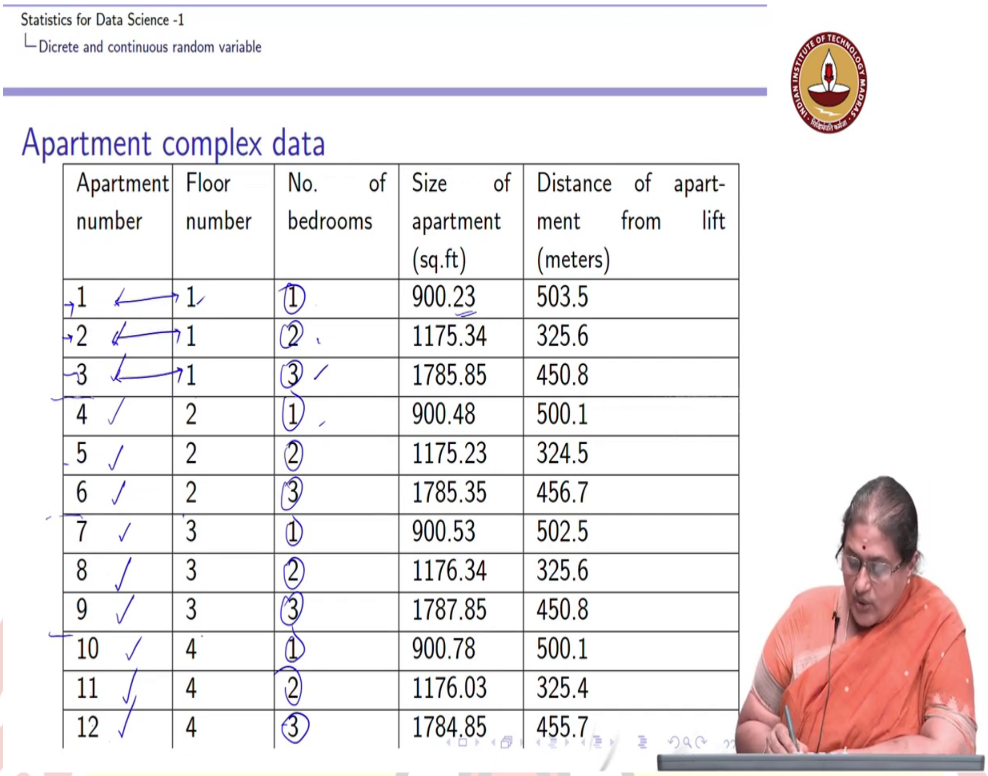
(Refer Slide Time: 08:38)



So, if I am looking at random variables associated with each outcome, I have a 1, 2, 3; a 1, 2, 3; a 1, 2, 3; a 1, 2, 3. These are the random variables. Where what is the random variable? The random variable X is the number of bedrooms in the apartment. And I see that this random variable, every outcome of my sample space has a random variable or a quantity associated with it. What are the possible values this random variable can take now? It can take the possible values the random variable can take are 1, 2, and 3.

(Refer Slide Time: 09:30)

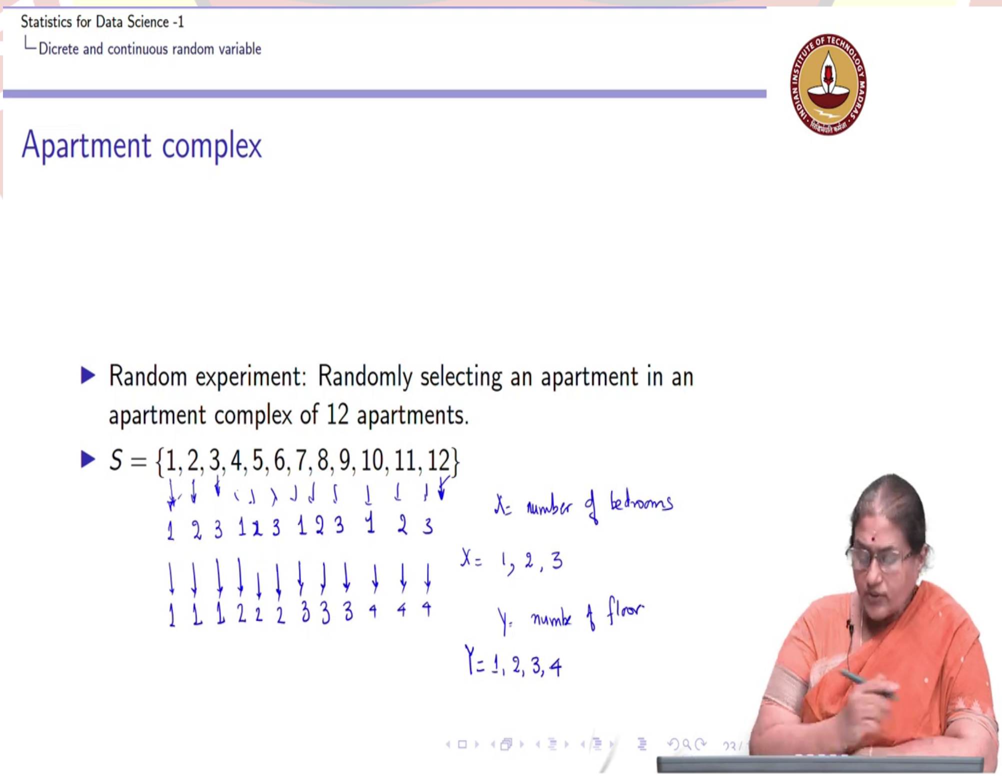




Apartment number	Floor number	No. of bedrooms	Size of apartment (sq.ft)	Distance of apartment from lift (meters)
1	1	1	900.23	503.5
2	1	2	1175.34	325.6
3	1	3	1785.85	450.8
4	2	1	900.48	500.1
5	2	2	1175.23	324.5
6	2	3	1785.35	456.7
7	3	1	900.53	502.5
8	3	2	1176.34	325.6
9	3	3	1787.85	450.8
10	4	1	900.78	500.1
11	4	2	1176.03	325.4
12	4	3	1784.85	455.7

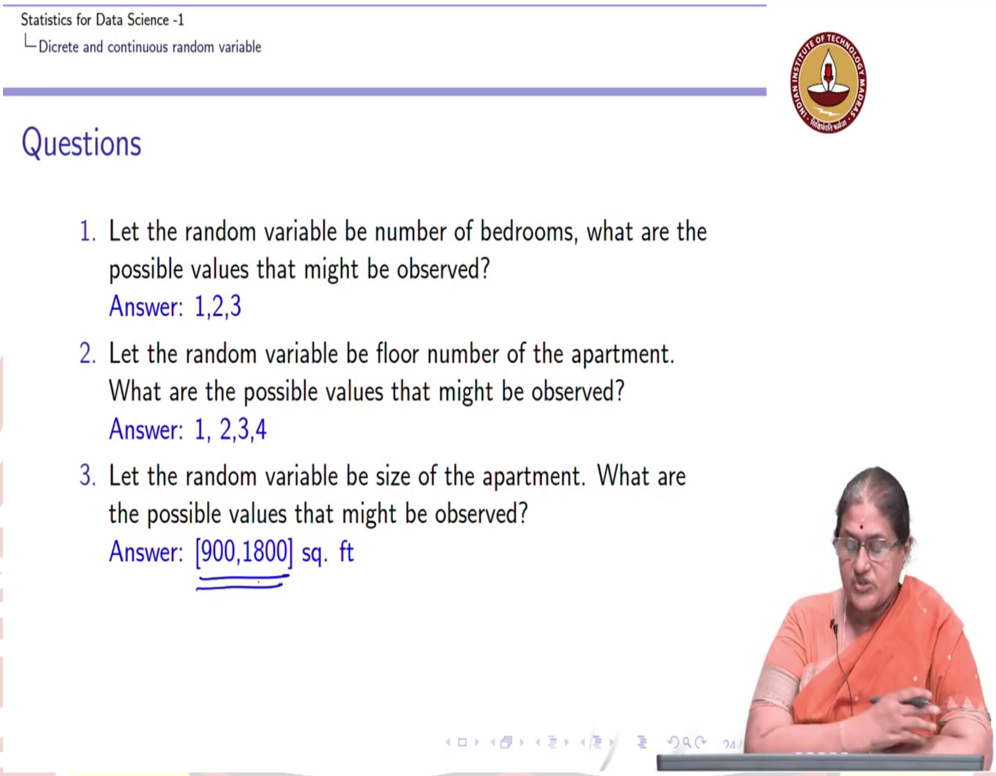
So, you can see that the possible values the random variable can take are 1, 2, and 3. Now, let us go to the second question. The second question says that let the random variable be the floor number of the apartment. Now, again let us go back to our data. The first for with the first apartment. I know that the floor number associated with my first number is 1 with a second apartment is 1, third is 1.

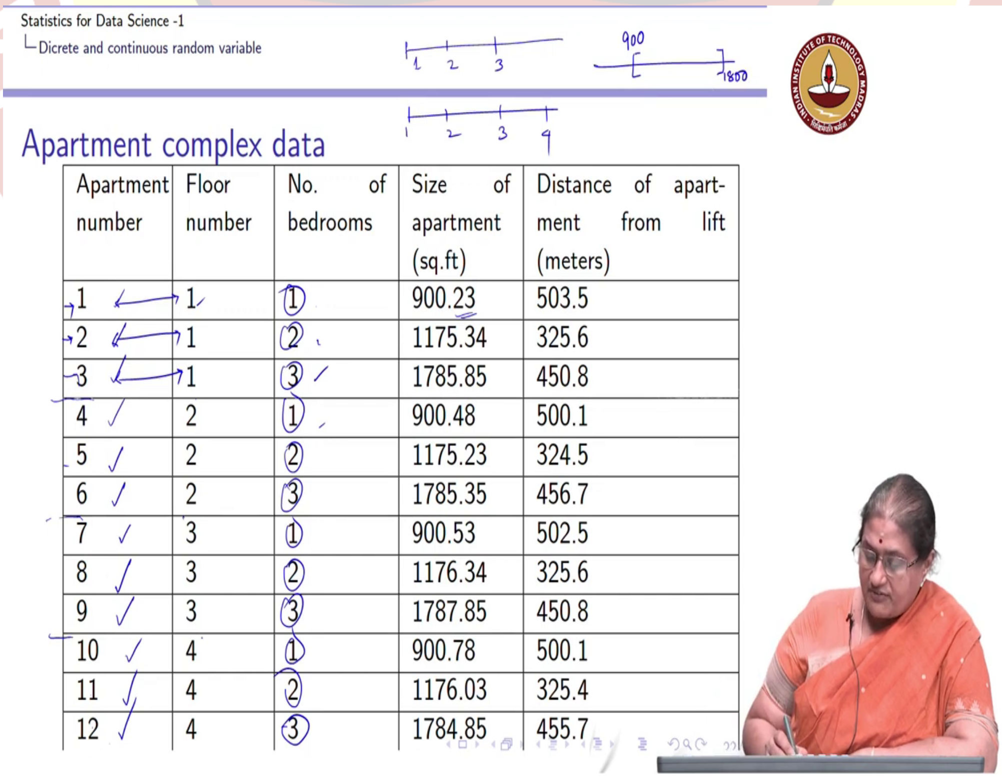
(Refer Slide Time: 10:07)



So, in terms of my sample space, when I come back here, and I am defining say Y to be the number of the floor or the floor number with 1 I am associating variable 1 with 2 again 1 with 3 again 1, 4 it is a 2, 5 it is a 2, 6 is a 2, 7 is a 3, 8 is a 3, 9 is a 3, 10 is a 4, 11 is a 4, and 12 is a 4. So, the possible values the random variable, which is the floor number, every outcome is associated with a quantity, the possible values, this random variable can take are 1, 2, 3 and 4.

(Refer Slide Time: 11:02)



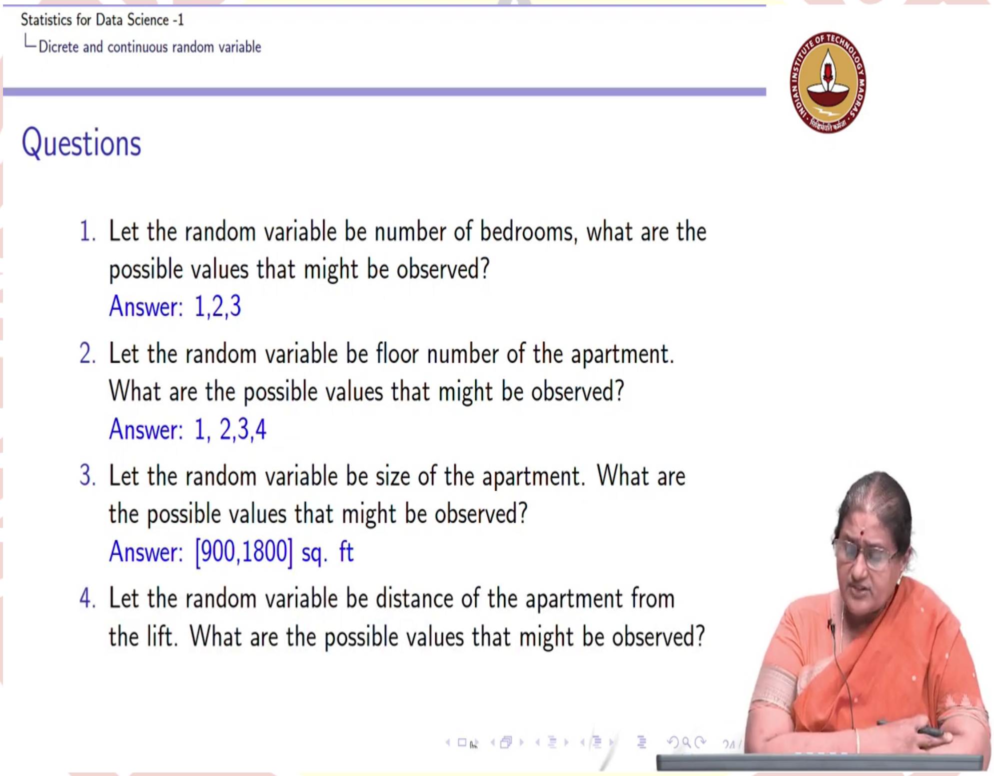


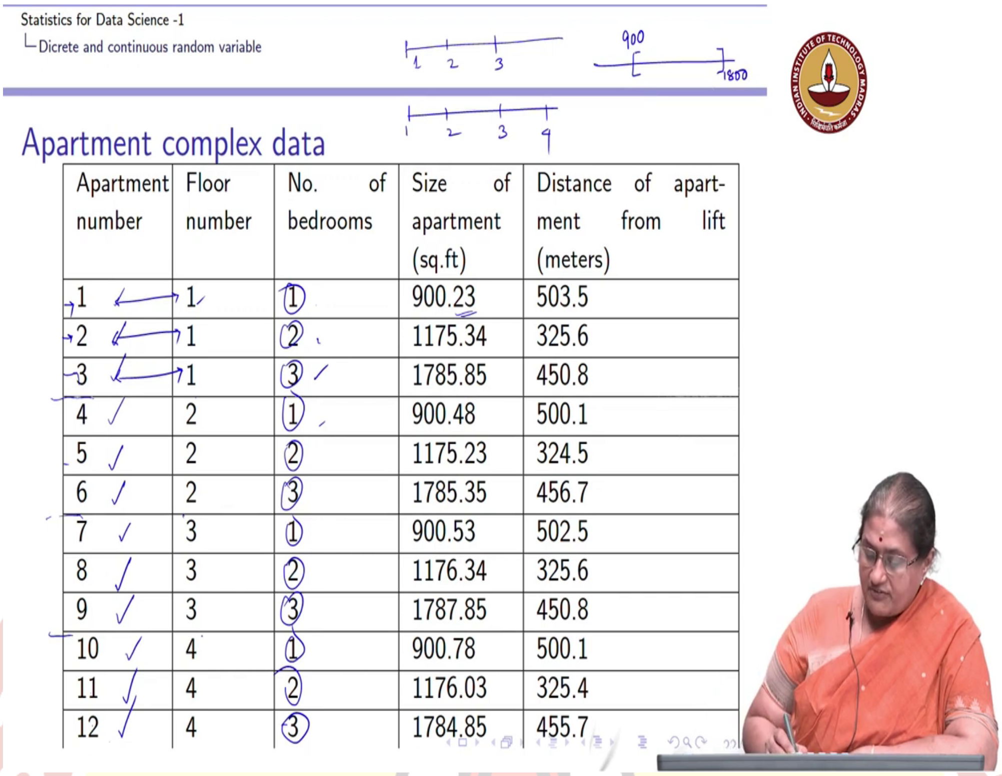
So, now you can see that the possible values this random variable can take are 1, 2, 3 and 4. Now, let me define the random variable to be the size of the apartment. Earlier, we looked at number of bedrooms, then we looked at the floor number, now I am looking at the size of the apartment. Again, let us go back.

Now, when we go back to the data and look at the size of the apartment, we do not see, with that it can be mapped to discrete points on the real number line. Whereas here I could have, so in the first case, I could map it to 1, 2, and 3, these are discrete isolated points. In the second case, it is 1, 2, 3, 4.

Whereas in the third case, when I talk about a real number line, I see that I can have that the data can take any value between 900 square feet to say 1780 or 1800 square feet. So, in a sense, I can give the values that possible values, the size of the apartment can take, the possible value, the size can be in an interval, and this interval is 900 to 1800 square feet.

(Refer Slide Time: 12:35)

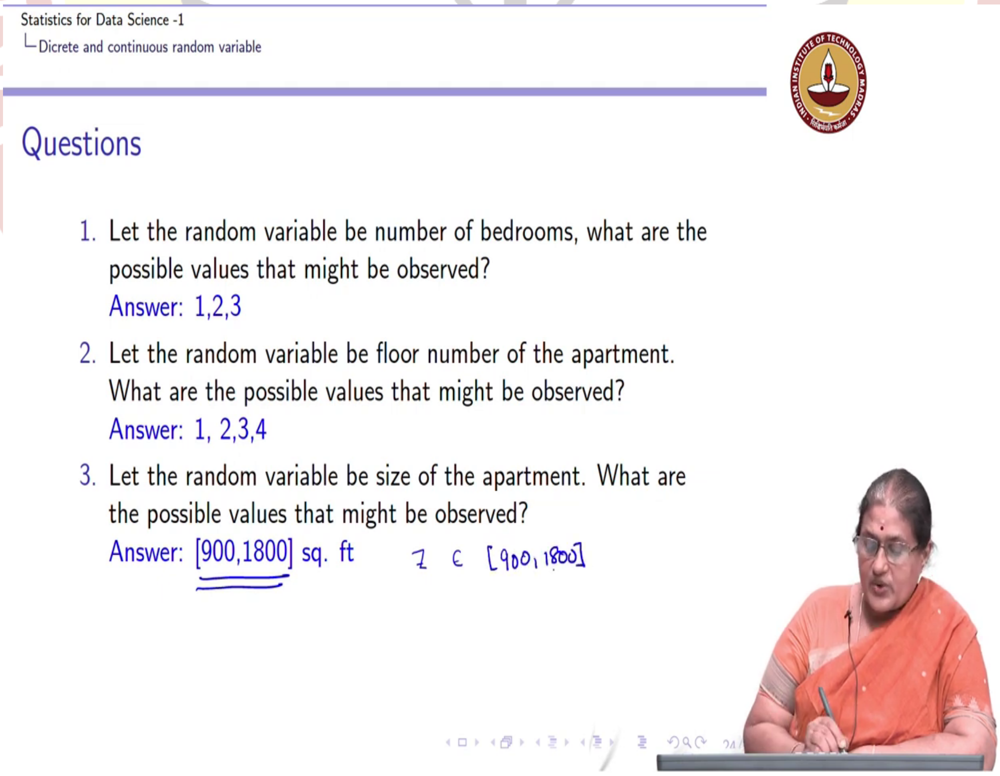


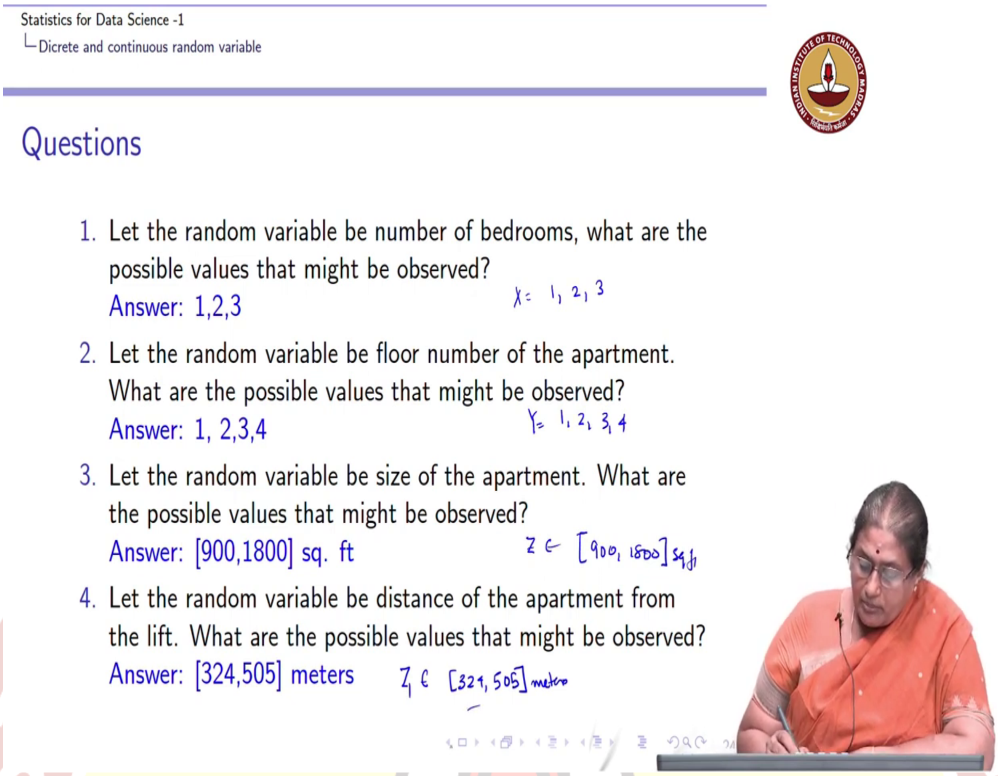


Apartment number	Floor number	No. of bedrooms	Size of apartment (sq.ft)	Distance of apartment from lift (meters)
1	1	1	900.23	503.5
2	1	2	1175.34	325.6
3	1	3	1785.85	450.8
4	2	1	900.48	500.1
5	2	2	1175.23	324.5
6	2	3	1785.35	456.7
7	3	1	900.53	502.5
8	3	2	1176.34	325.6
9	3	3	1787.85	450.8
10	4	1	900.78	500.1
11	4	2	1176.03	325.4
12	4	3	1784.85	455.7

Now, let us define the fourth random variable to be the distance of the apartment from the left. Again, we go back to our data, when we go back to our data, we see that again here there are not discrete points, but it is again a continuum. And when you look at a continuum, I can again write that the distance of the apartment from the feet.

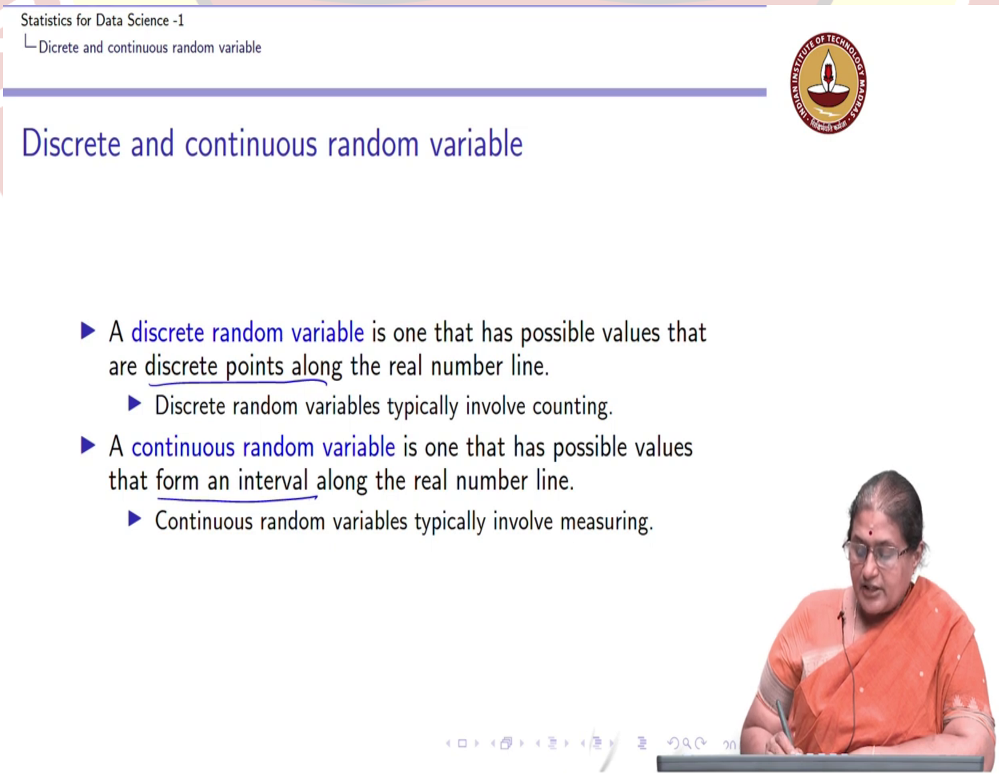
(Refer Slide Time: 13:01)

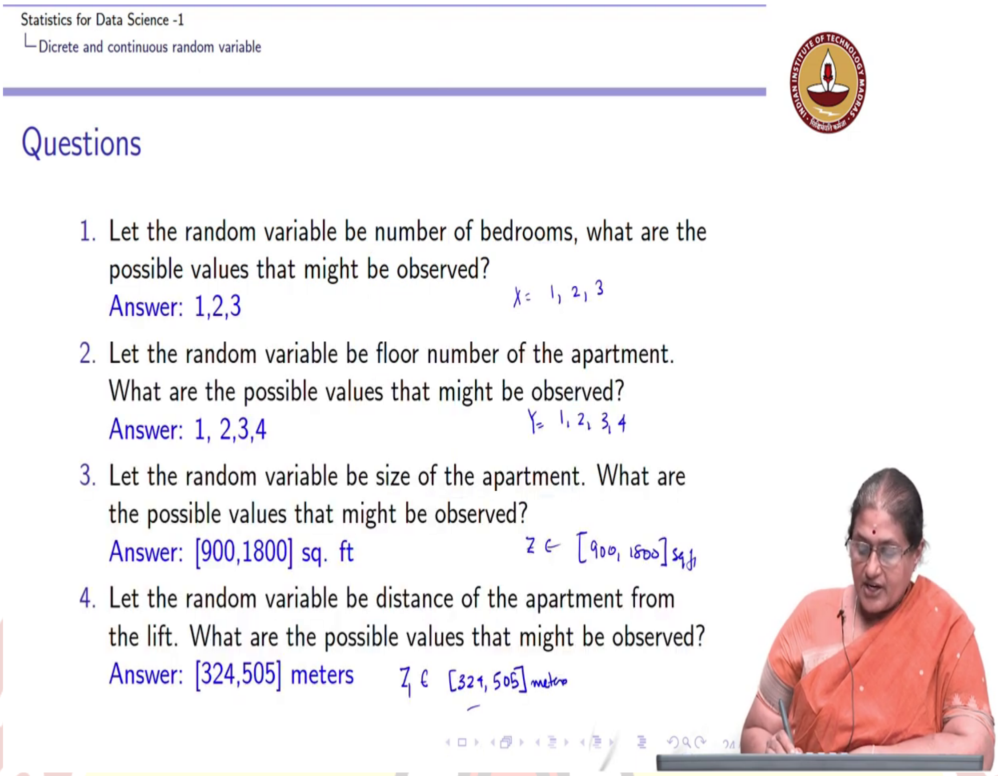


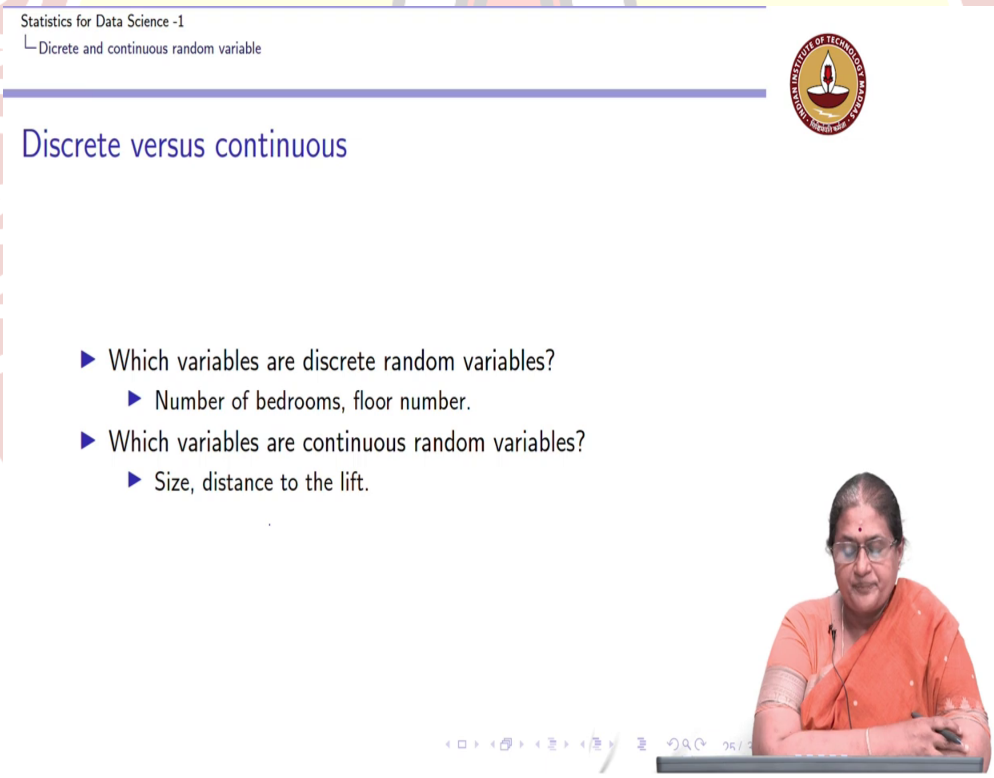


So here, if I define this random variable in the earlier case, the size to be some Z, I know that this random variable Z would belong to 900 to 1800 square feet. Now, I could call this say Z_1 and I know that this belongs to 324 to 505 metres. So, we have seen the example of a random variable here, which takes values 1, 2, 3 here it takes the values 1, 2, 3, 4, here it belongs to a interval and in the last case, it also belongs to a interval.

(Refer Slide Time: 14:05)

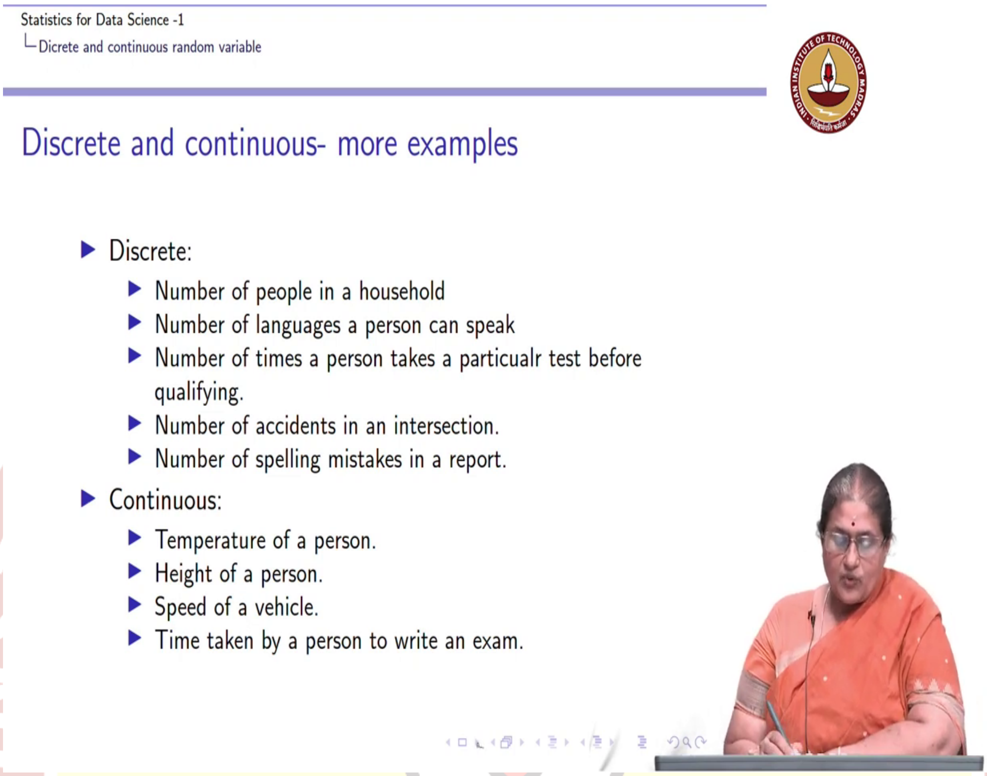






So, going back to the definition, which we proposed is of definition the understanding which we proposed for a random variable, the discrete and continuous, we said a discrete random variable typically would take discrete points along the real number line, and here it forms an interval, I can now classify the 2 random, the 4 random variables which we have given as examples here as the first 2 that is the discrete random variables in the data are nothing but the number of bedrooms and floor number, whereas the continuous random variables in my example data set, the apartment data set are size of the apartment, and distance of the apartment to the left.

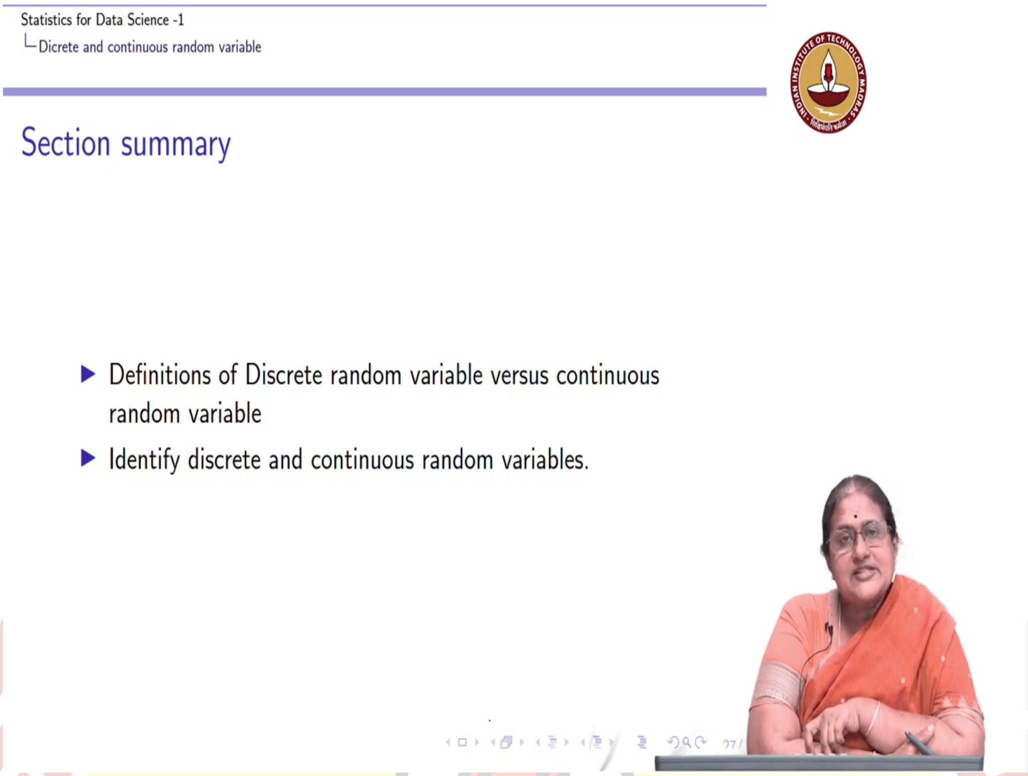
(Refer Slide Time: 14:48)



So, let us look at a few more examples where we naturally come across discrete random variables. Again, remember discrete random variables involves counting. So, whenever you are having instances where for example, you count the number of people in a household, a number of languages a person can speak, number of times a person takes a particular test before qualifying or before passing, number of accidents that happen in an intersection, number of spelling mistakes in a report all these count, you can see that everything represents the operation of counting, you can discuss or you can think of this as a discrete random variable.

Whereas the instances where I have, I am computing or measuring the temperature of a person, measuring the height of a person, measuring the speed of a vehicle, or measuring the time taken by a person to write an exam, time taken by a person to reach office. All these instances which involve the measure of a quantity, I can think of the variable or the random variable that is involved as a continuous random variable.

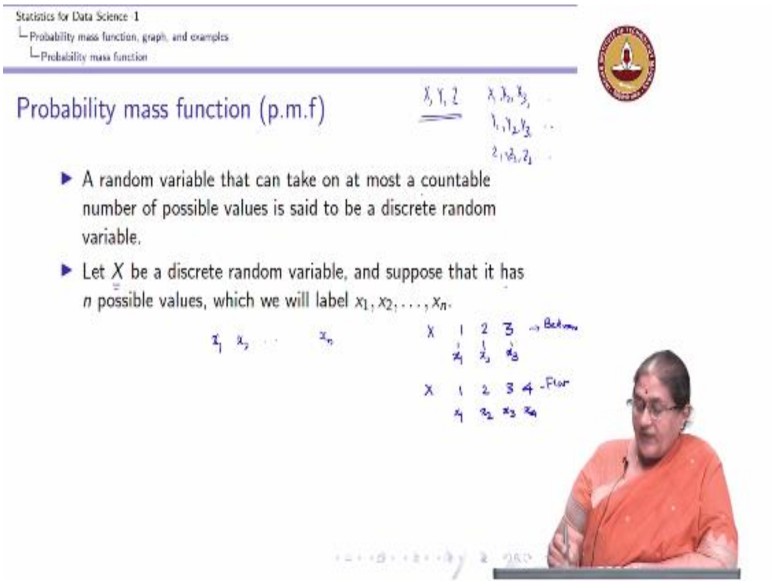
(Refer Slide Time: 16:16)

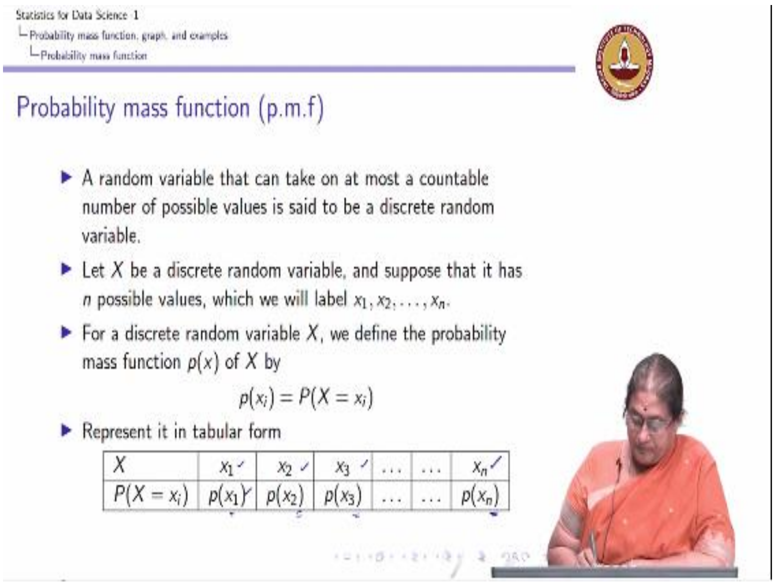


So, in summary, what we have learnt in this section was, we first defined formally what was a discrete random variable and a continuous random variable, a discrete random variable takes countable values, and it can be mapped onto discrete points on a real number line. Whereas, a continuous variable, it actually can be mapped on to an interval on a real number line. For now, we are going to focus on discrete random variables.

Statistics for Data Science-1
Professor. Usha Mohan
Department of Management Studies
Indian Institute of Technology, Madras
Lecture No. 8.4
Discrete Random Variables - Probability Mass Function Properties

(Refer Slide Time: 00:14)





So, we continue our discussion on discrete random variables. Again recall a discrete random variable is a random variable that can take on at most a countable number of possible values. We refer to it as a discrete random variable we are now going to focus only on discrete random variables. Again when we look at discrete random variables let I denote a random variable with the upper case $X, Y, Z, X_1, X_2, X_3, Y_1, Y_2, Y_3$ and so forth or Z_1, Z_2, Z_3.

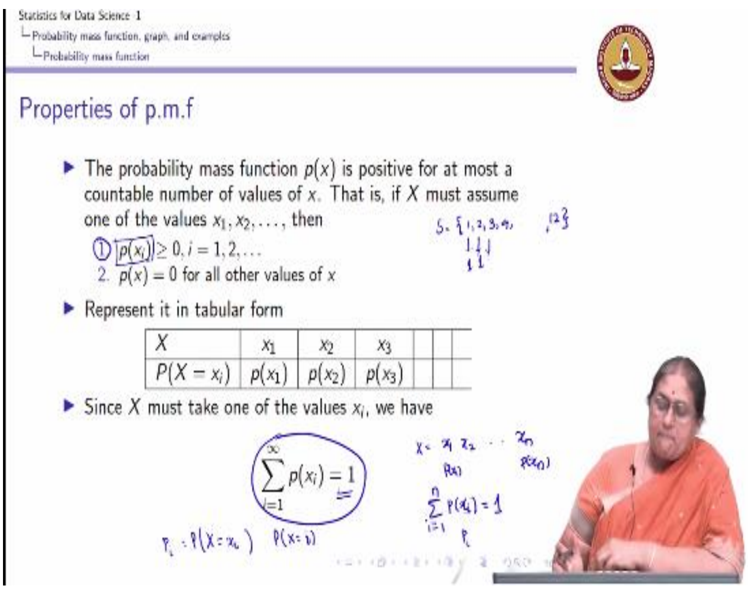
This is typically we represent a random variable through upper case alphabets capitals X, Y, Z. So, now let X be a random variable let it take a finite number of values. Now what do I mean by it takes a finite number of values. It takes n possible values. Let me represent that by $x_1, x_2, \ldots\ x_n$. For example, if X takes the values 1, 2, 3 my x_1 would have been 1, x_2 would have been 2, x_3 would have been 3.

If X takes the value 1, 2, 3, 4 as in the case of the number of floors. This was the number of bedrooms, this is the floor, then x_1 is $1, x_2$ is $2, x_3$ is taking the value 3 and x_4 is 4. So, in general I can talk of this random variable X taking values $x_1, x_2, \ldots\ x_n$ that is what I mean by X is taking n finite values $x_1, x_2, \ldots\ x_n$. Once I know x_1, X takes these values we define the probability mass function. How do I define it?

This is the function for all the values what are the values X is taking X is taking values $x_1, x_2, \ldots x_n$. So, associated with every value X takes I know there is a probability associated with it. So, I have what is a probability of X taking the value x_1 I know what is the probability of X taking the value x_2 and X taking the value x_n. This function the probability of X equal to x_i for each of these values. This function is referred to as the probability mass function of the random variable.

A nice way to represent it is in tabular form. So, X takes the values x_1, x_2, x_3 up to x_n the probability is with probability X equals to $x_1\ is\ P(x_1),\ P(x_2),\ P(x_3)$ up to $P(x_n)$.

(Refer Slide Time: 03:43)



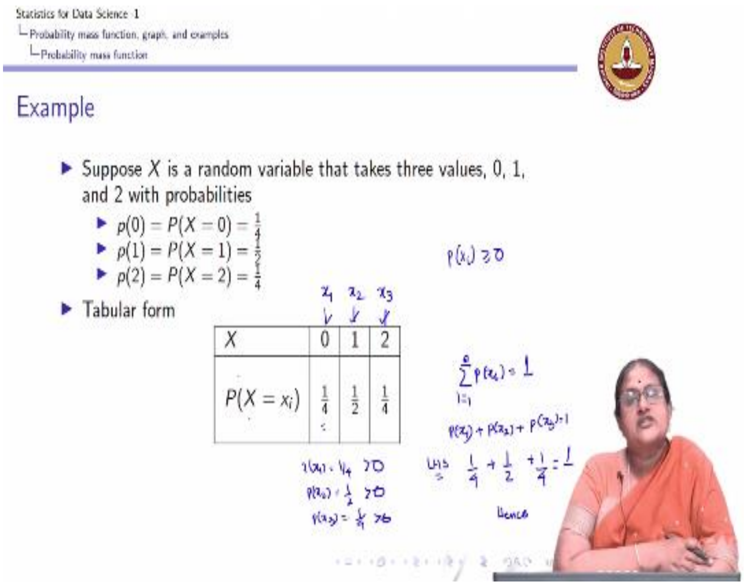
So, let us look at a very simple example. So, here we assume that this X takes finite number of values so X takes $x_1, x_2, \ldots x_n$. possible values, but I could also have the case that X assumes value x_1, x_2, x_3 countable, but infinite. Again associated with each of these exercise I will a $P(x_1)$, I will have a $P(x_2)$, I will have a $P(x_3)$ so forth. So, whenever I talk about the discrete random variable, I could either have countably finite or countably infinite number of values.

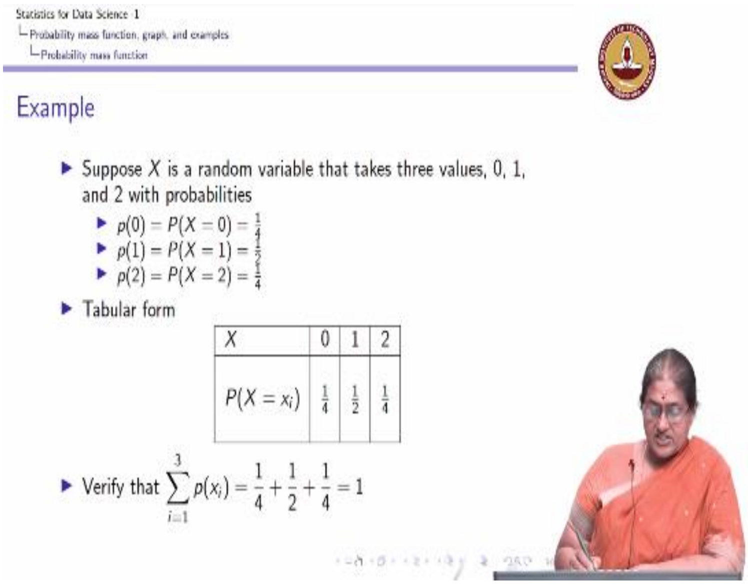
But nevertheless whatever it is there are two key properties of the probability mass function. I repeat there are two key properties of the probability mass function. They are namely $P(x_i) \geq 0$. In other words $P(X = x_i)$ is always non negative and the second property is that some probability of X is equal to 0 for all values of x. So, if I represent it in tabular form I have $P(x_i) \geq 0\ \forall i$ and the next property is since x I know that every point in my sample space.

For example, if my sample space in the apartment complex was 1, 2 up to 12. Every point was mapped to a random variable we also know from the axioms of probability the $P(S) = 1$. Since every point is mapped on to a random variable, it makes sense for us to say that the summation over all possible values X can take the summation of the probabilities should add up to 1. So, the two key properties are probability of x_i is now negative and the summation of over all possible values of x should be equal to 1.

Now if x takes only finitely many values with probability $P(x_1)$, $P(x_n)$, then I know $\sum_{i=1}^{n} P(x_i)$ should be equal to 1. Some books refer $P(x_i)$ with just P_i, but we need to understand that whenever you see a P_i this could be probability X takes the value x_i or probability X takes a value i. You need to understand how the probability mass function is defined, but nevertheless however you are defining the probability mass function the probability of x taking a particular values is always now negative and the sum of probabilities over all possible value should be equal to 1. These are the key properties of the probability mass function.

(Refer Slide Time: 07:30)

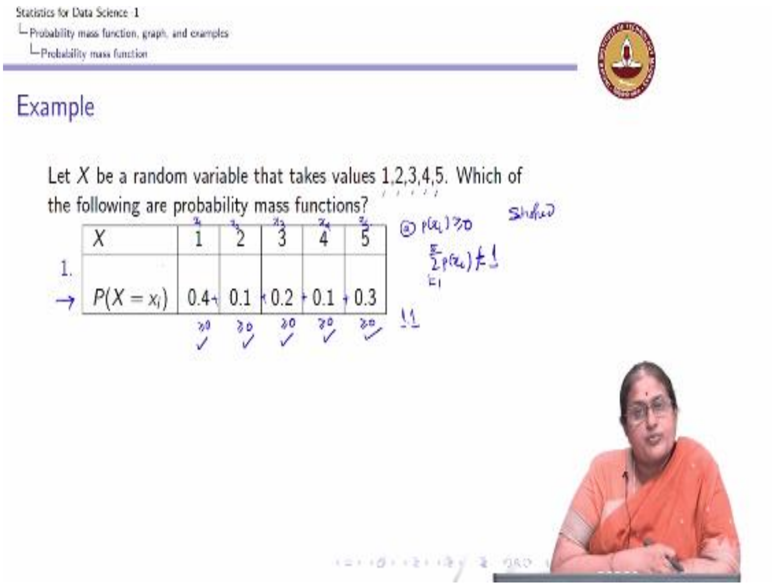




Now let us look at an example. Suppose X is a random variable that takes 3 values. So again I have a random variable which takes only 3 values it is finite 0, 1, 2 and what is the probability? $P(X = 0)$ is $\frac{1}{4}$, X is equal to 1 is half and X equal to 2 is $\frac{1}{4}$. Tabularly I can represent it as x takes the value 0, 1 and 2 so my x_1, x_2, x_3 x_1 takes 0, 1, 2 with $P(X = x_1)$ which is a 0 is $\frac{1}{4}$. X equal to x_2 which is a probability x equal to 1 is $\frac{1}{2}$ and X equal to x_3 which is X equal to 2 is $\frac{1}{4}$

So, what is the first property I need to see. $P(x_i) \geq 0.$ I know $P(x_1)$ is $\frac{1}{4}$ this is greater than 0 $P(x_2)$ is $\frac{1}{2}$ which is greater than 0, $P(x_3)$ is $\frac{1}{4}$ which is also greater than 0. The second property is $\sum_{i=1}^{n} P(x_i)$ should be equal to 1. In other words I need to verify whether $P(x_1) + P(x_2) + P(x_3) = 1$. This is what I need to verify. I can see that $P(x_1)$ is $\frac{1}{4}$, $P(x_2)$ is $\frac{1}{4}$, $P(3)$ is $\frac{1}{4}$ I can verify that this is equal to 1. Hence, what we have here is a probability mass function of the random variable.

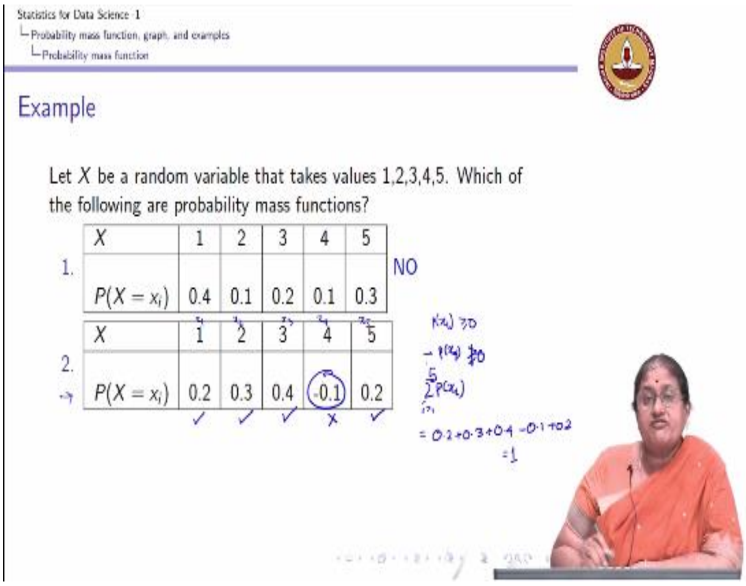
(Refer Slide Time: 09:26)



Let us look at certain more examples to understand the properties of the probability mass function. Now suppose I have X is a random variable that takes values 1, 2, 3, 4, 5. So I have this is my x_1, x_2, x_3, x_4, x_5. Again it takes finite number of values with the given probabilities. So, is this a probability mass function? So, the first condition is I need to check $P(x_i) \geq 0$. Yes, for this done this is greater or equal to 0 this greater or equal to 0, this greater or equal to 0 this is also greater or equal to 0. The next condition so the first condition is satisfied.

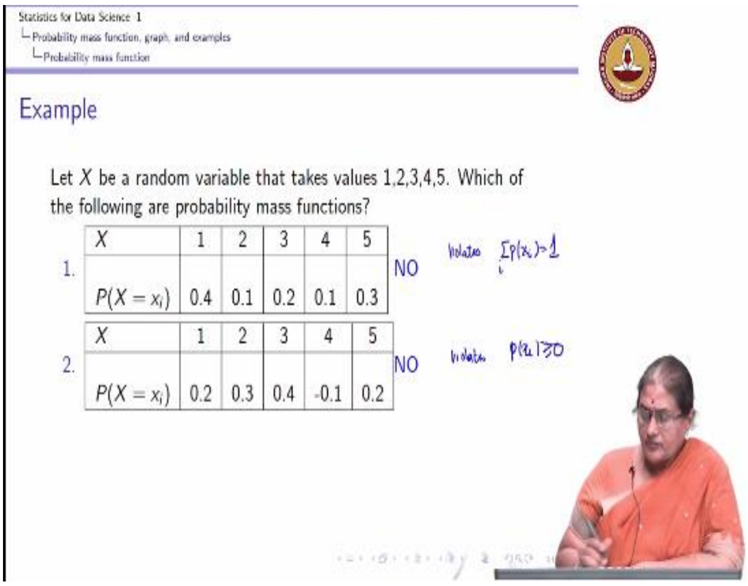
Now the second condition is I need to check whether $\sum_{i=1}^{5} P(x_i)$ is equal to 1. So, I have a 0.4 + 0.1 which is a 0.5. 0.5 + 02 which is a 0.7, 0.7 + 0.1 which is a 0.8, 0.8 + 0.3 which is 1.1 so I get the $\sum_{i=1}^{5} P(x_i)$ is not equal to 1. Hence, this does not satisfy the probability properties of a probability mass function. Hence, the first table is not a probability mass function.

(Refer Slide Time: 11:03)



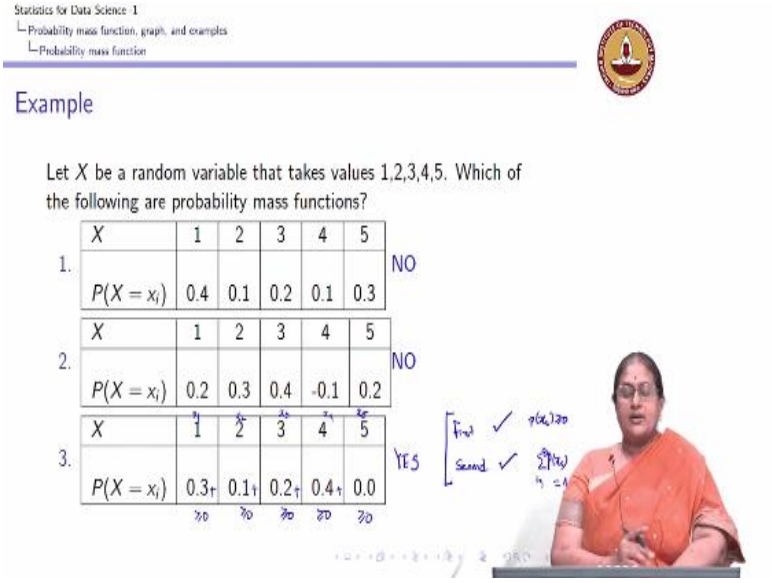
Now, let us look at the second example again X is taking the value 1, 2, 3, 4 and 5. So, the first property $P(x_i) \geq 0$ here it is yes, yes, yes. This is a no, this is a yes. So, you can see that $P(x_4)$ is not greater or equal to 0. However, if you notice $\sum_{i=1}^{5} P(x_i)$ which is equal to 0.2 + 0.3 + 0.4 - 0.1 + 0.2. You can see that this is 0.9 which is equal to 1. So, I have a situation where the function satisfies the second property, but not the first property. Hence, because of this it is not a probability mass function.

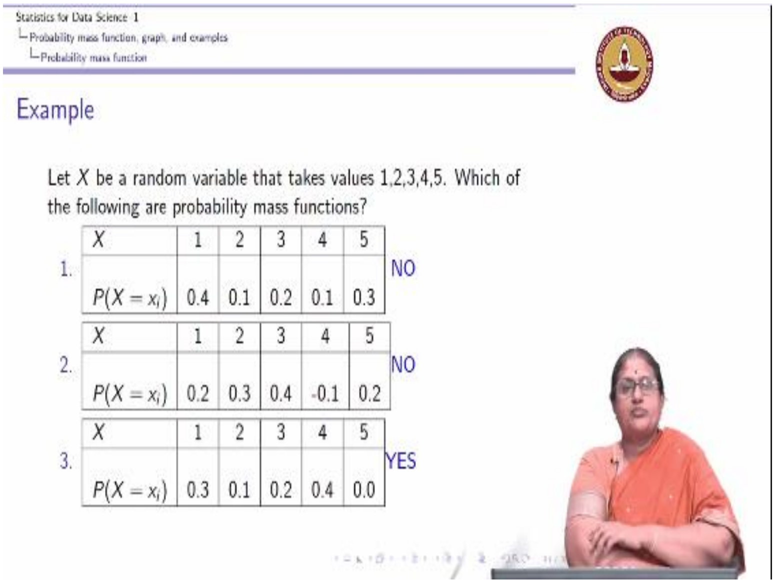
(Refer Slide Time: 12:07)



So, here it violates $\sum_{i=1}^{n} P(x_i)$ =1 over all values of i. Here it violates the $P(x_i) \geq 0$.

(Refer Slide Time: 12:26)



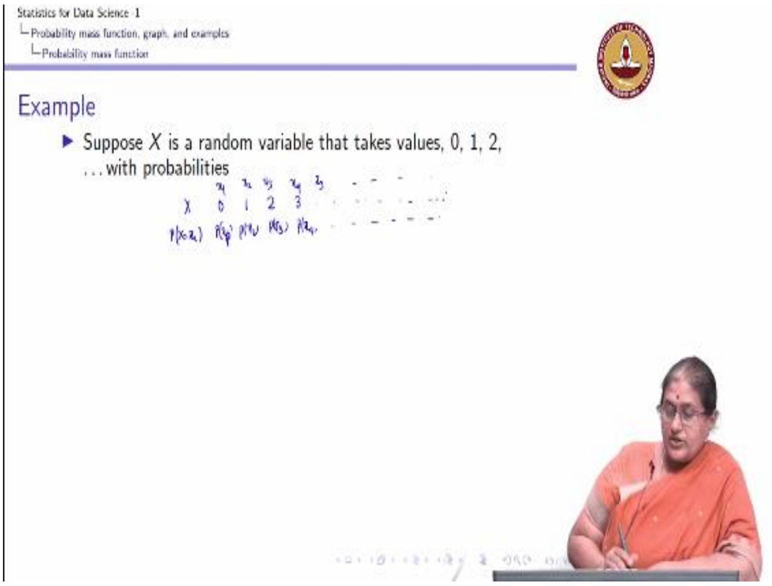


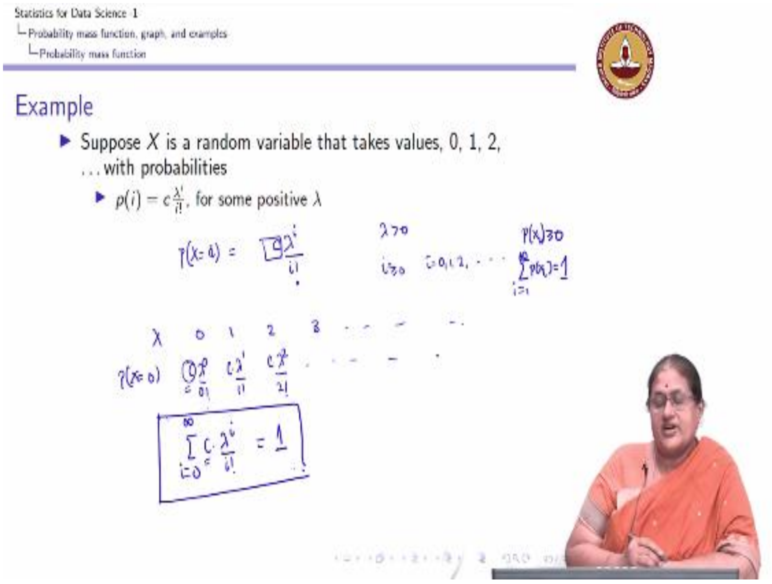
Now, let us look at a third example. In this example again X takes the value $x_1, x_2, x_3, x_4,$ and x_5 with these probabilities. Again, let me check the first condition greater or equal to 0, greater or equal to 0, greater than 0, greater or equal to 0, greater or equal to 0. First condition is satisfied first condition is $P(x_i) \geq 0$ satisfied. Now let us check the second condition which is $\sum_{i=1}^{5} P(x_i)$ 0.3 + 0.1 0.4, 0.4 + 0.2 0.6, 0.6 + 0.4 1, 1 + 0 is 1. Yes it is equal to 1. So, both the conditions are satisfied hence this is a probability mass function.

So, the first thing which we need to understand is given a random variable again I specify we are looking only at discrete random variables. So, given a random variable X that takes the values $x_1, x_2,$ up to x_n. For the first case I considered countably finite number of values if I am talking about a probability mass function I need the two properties. And what are the two properties we are looking at when we are talking about a probability mass function?

The first is $P(x_i) \geq 0$ and the second is summation probability of x_i over all possible values of x should be equal to 1. Now, let us look at another example where x does not take finite number of values. It takes countable number of values, but not finite.

(Refer Slide Time: 14:30)





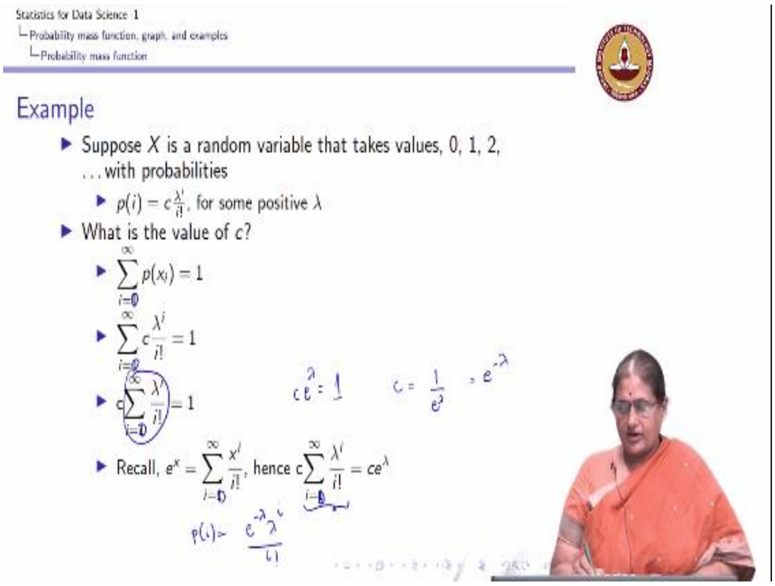
So, let us look at this example. So, what are the values this random variable is taking? X is taking the value 0, 1, 2, 3 it keeps taking these values. Let me write $P(X = x_i)$, probability of x_0 or x_1, x_2 probability of x_3, probability of x_4 so this is my x_1, x_2, x_3, x_4 so X is taking these values I am not having finite number of values. Now what is it the $P(X = i)$ is given by $\frac{c\lambda^i}{i!}$.

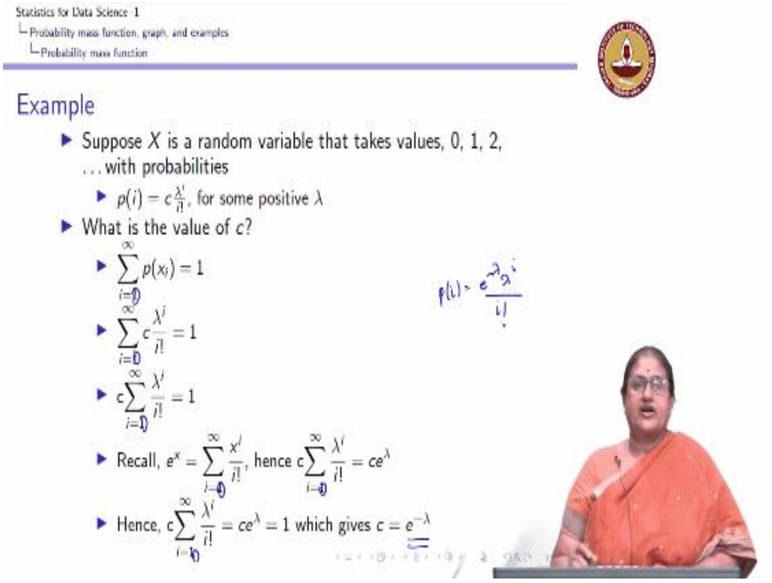
Now $\lambda$is positive $i \geq 0$ because i takes the value 0, 1, 2 all of it. So, I want to know for what value of c will this be a probability mass function? Again what are the values $X$is taking Xtakes the value 0, 1, 2, 3 so forth $P(X = 0)$ would be $\frac{c\lambda^0}{0!}$. This would be $\frac{c\lambda^1}{1!}$, this should be $\frac{c\lambda^2}{2!}$and so forth.

So, what value of c would make this a probability mass function? So, the first thing is I know λ is positive and i is positive. So, c has to be greater or equal to 0 because again what are the conditions? The conditions of $P(x_i)$ should be greater or equal to 0 and $\sum_{i=1}^{\infty} P(x_i) = 1$.

For the first condition I need c to be non negative because everything else is going to be now negative. For the second condition I need to check $\sum_{i=1}^{\infty} \frac{c\lambda^i}{i!} = 1$. So, what value of c would give this, this is what we need to check because for this to be a probability mass function I need this condition to be satisfied.

(Refer Slide Time: 17:12)



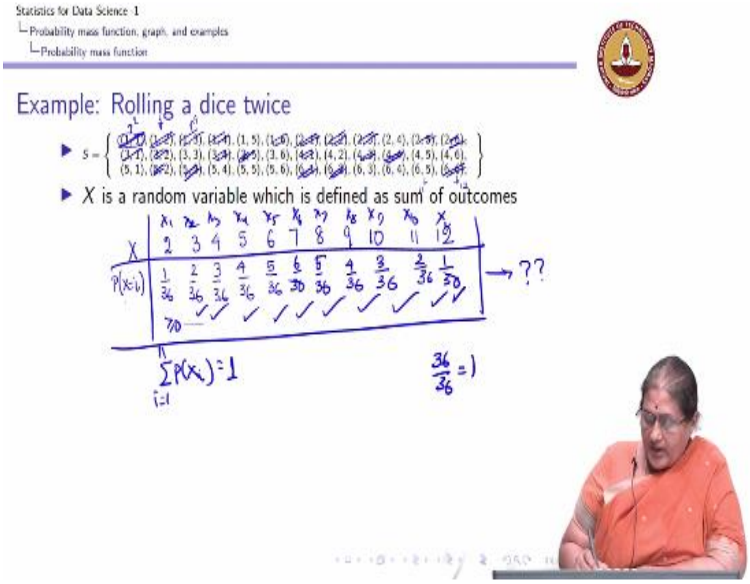


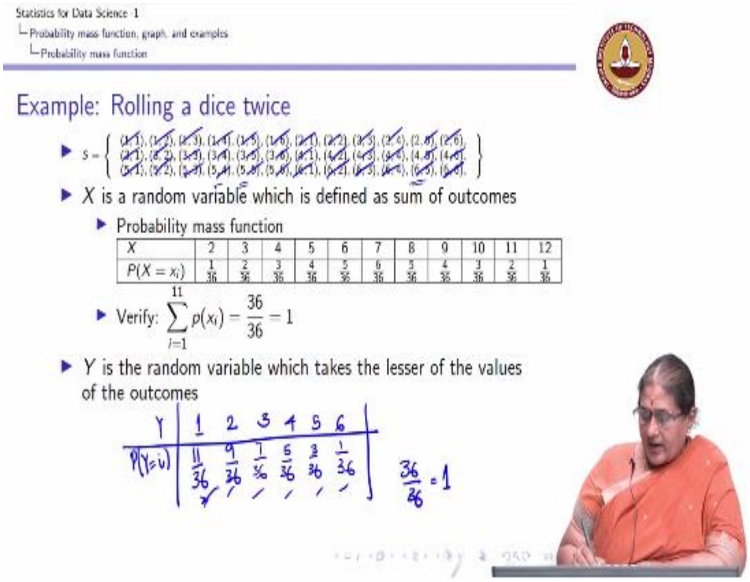
So, the question is what is the value of c for which this is a probability mass function. So, this is i goes from 0 to infinity. This is again i goes from 0 to infinity. So I know it is again 0 to infinity here so i goes from 0 to infinity $c\lambda^i$ because X takes the value 0, 1, 2. I need to look at i goes from 0 to infinity. Now we all know the following that e^x is so we all know the following that e^x is $\sum_{i=1}^{\infty} \frac{x^i}{i!}$.

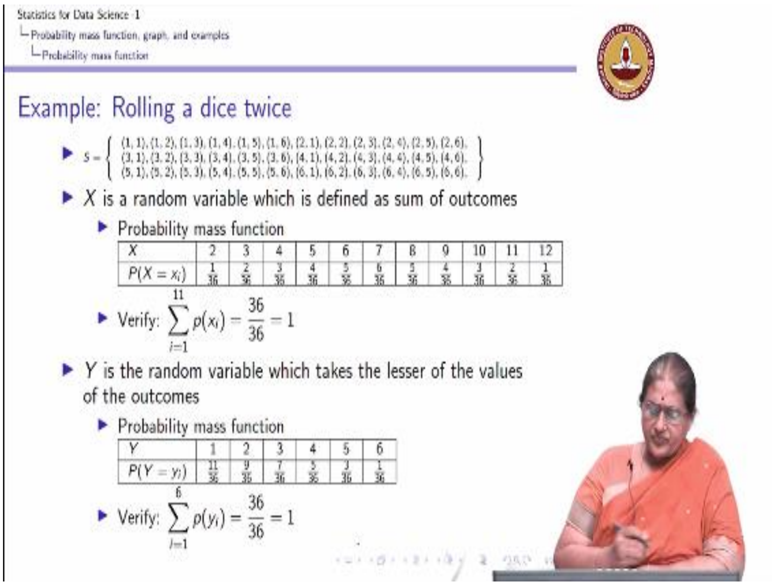
Hence, I have summation $\sum_{i=1}^{\infty} \frac{c\lambda^i}{i!}$ is ce^λ because this would be e^λ. So, now I have from here ce. So, this term is going to be e^λ that is what this tells us that $\sum_{i=1}^{\infty} \frac{x^i}{i!}$ is e^x. So, this term would be e^λ. So, I get $ce^\lambda = 1$ which should give me $c = \frac{1}{e^\lambda}$ or $e^{-\lambda}$.

Hence, I get my $P(i)$ is $\frac{e^{-\lambda}\lambda^i}{i!}$. So, I can get my c to be $e^{-\lambda}$ which is going to give me the solution that $P(i)$ is $\frac{e^{-\lambda}\lambda^i}{i!}$.

(Refer Slide Time: 19:22)







Now let us go back to the examples which we have discussed earlier. So, we looked at the first example we looked at was rolling a dice twice. We know that the sample space has 36 outcomes and I have listed down the 36 outcomes here which are (1, 1) to (6, 6). Now let us define the random variable which is defined as a sum of outcomes. Again, we have seen this in the earlier lecture that every outcome is mapped to a particular values.

So, you can see that (1, 1) is mapped to 2; (1, 2) is mapped to 3; (1, 3) is mapped to 4; (6, 6) is matched to 12; (6, 5) is mapped is 11 the sum of outcomes. So, I can see that this random variable X takes the values 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 and 12. These are the values this random variable take X takes the value 2, (1, 1) is the outcomes which maps to it X takes the value 3 (1, 2) and (2, 1) are the outcomes that give the value 3. (4, 1), (3, 2), (2 , 3), 1 so each one of them you can see is mapped 12 is from the outcome (6, 6).

Now again associated with each one of these values $P(X = i)$. So, associated what is the $P(X = 2)$? I know only outcome gives this value so it is $\frac{1}{36}$ (1, 2) and (2, 1) here so this is $\frac{2}{36}$, X equal to 4 comes from (1, 3); (2, 2) and (3, 1) $\frac{3}{36}$, X equal to 5 I have (1, 4), I have (2, 3) I have (3, 2), I have (4, 1). So, it is $\frac{4}{36}$ this should we can check is $\frac{5}{36}$.

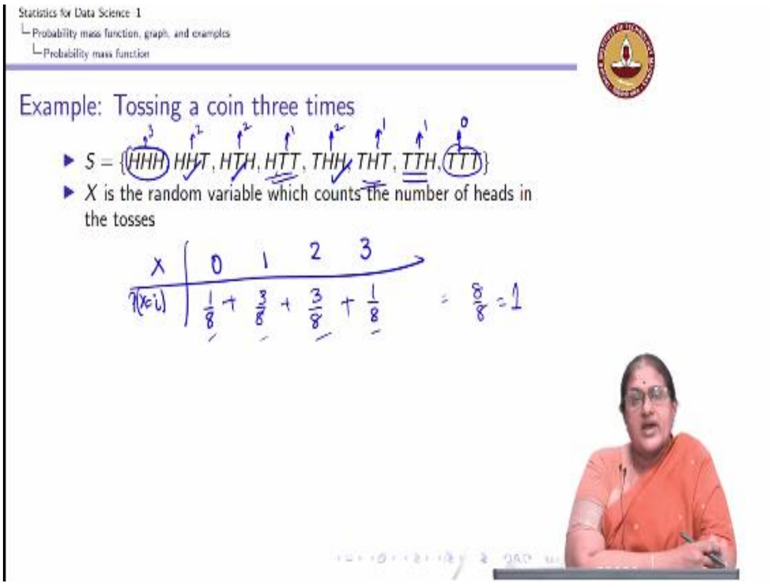
X equal to 7 is (1, 6); (2, 5); (3, 4); (4, 3); (5, 1) and (5, 2) and (6, 1) which is $\frac{6}{36}$. Similarly, this would again 8 I know 8 would come from a (6, 2); (5, 3); (4, 4); (3, 5) and a (2, 6). So, again this is $\frac{5}{36}$ you can verify for all other values this is going to be my probability mass function. So, I know X takes the values 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 with the probability this one.

Now is this a probability mass function? Again what are the properties of the probability mass function? This should be greater or equal to 0 I can see each one of them is greater or equal to 0. The second thing we need to verify is the sum of the all probabilities i going from all x_i so this is $x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}$ and x_{11} i going from 1 to 11 should be equal to 1. 3 + 3 6, 10, 15, 21, 26, 26 plus again you can see + 10 is 36 so you can see that this is $\frac{36}{36}$ which is equal to 1 and I can see that this indeed is a probability mass function.

Now let us go to the second random variable which we defined. Again we defined the second variable which takes the lesser of the values of the outcomes. So, again I have Y takes the value 1 for these outcomes. So for all these outcomes Y takes the values 1, Y takes the value 2 for these outcomes, it takes the value 3 for these outcomes, it takes the value 4 for these outcomes, it takes the value 5 for these outcomes and it takes the value 6 for this outcome.

So, this is $\frac{1}{36}$ 5, 1, 2, 3 so this is a $\frac{3}{36}$ 4 would be $\frac{5}{36}$ for 3 it was $\frac{7}{36}$, 2 it was $\frac{9}{36}$ and 1 11 out of these outcomes $P(Y = i)$. So, I know Y takes these values with the respective probabilities. Again is this a probability mass function. Let us verify I know all of them are greater than or equal to 0. So, the first property is satisfied the second property 11 + 9 20, 20 + 7 27, 32 35, 36. It is actually $\frac{36}{36}$ which is equal to 1. So, I have both the properties are satisfied and hence this also defines a probability mass function.

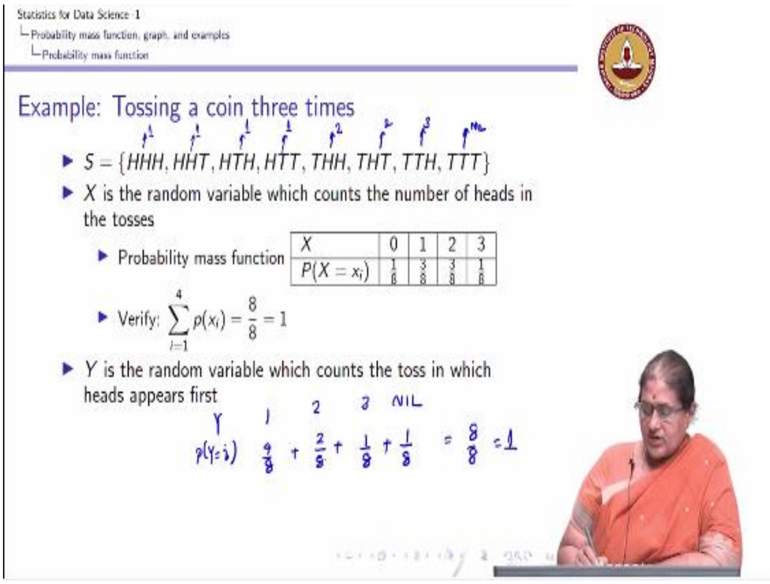
(Refer Slide Time: 25:43)

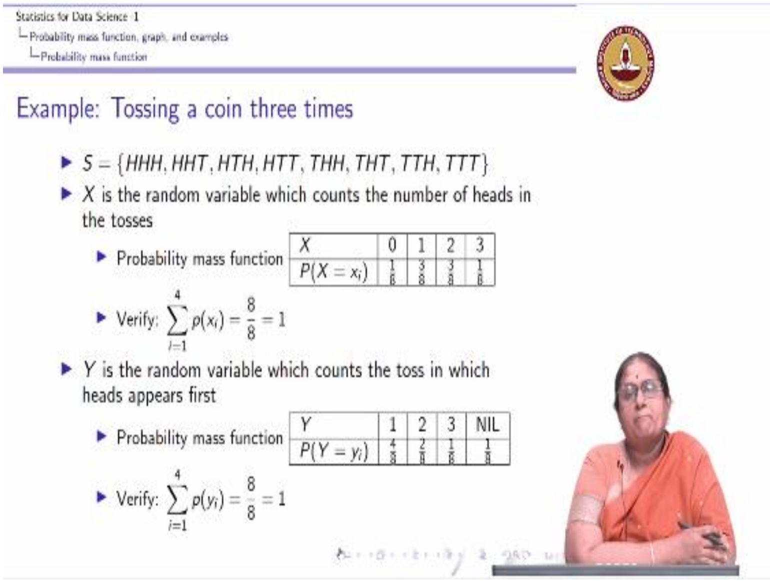


Now let us look at the next example of tossing a coin 3 times. Now when I toss a coin 3 times again this is my sample space. Again we define X to be the random variable that counts the number of heads in the toss. Again this outcome it is 3 here I have 2, here I have 2, here I have 1 again 2 I have 1 head, 1 head and no head. So, X takes the value 0, 1, 2 and 3 with what probabilities.

The probability with X takes no head it corresponds to this outcome it is $\frac{1}{8}$, 3 heads corresponds to this outcome which is again $\frac{1}{8}$. One head it corresponds to this, this and this so it is $\frac{3}{8}$, 2 corresponds to this, this and this outcome which is again $\frac{3}{8}$. Now is this a probability mass function? All of them are greater or equal to 0 and the sum of the probabilities which is equal to 1 + 3 + 3 + 1 is $\frac{8}{8}$ which is equal to 1. So, hence it is indeed a probability mass function.

(Refer Slide Time: 27:06)

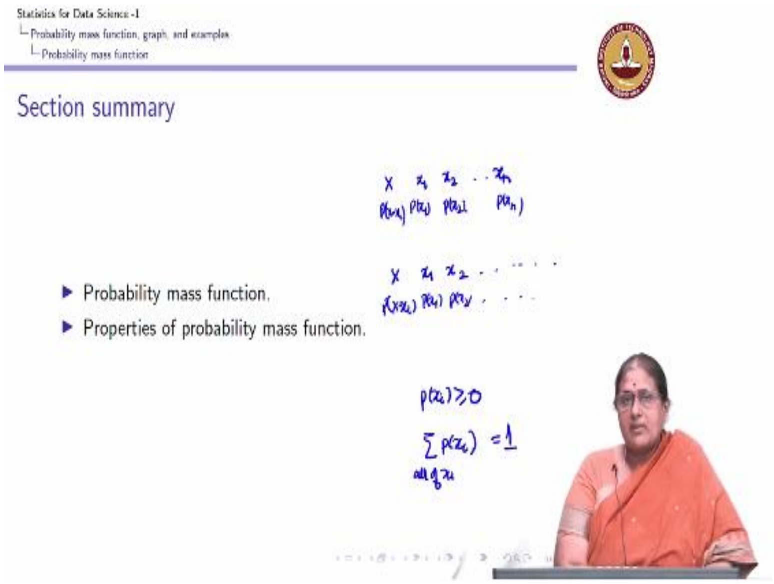




So, the next thing which we can look at is the random variables which counts the toss in which head appears first. Again, here head appears first in the first toss again here it appears in the first toss here it again appears first in the first toss, here it appears in the first toss, here it appears in the second toss this is again second toss, this is third toss and this is the nil toss. Remember, we defined it as nil we did not define it as 0.

So, the values Y takes are again 1, 2, 3 and nil and the probability with Y takes those values. It takes the value 1, 4 out of 8, 2 out of 8, 3, 1 out of 8 and nil 1 out of 8 all of them are greater or equal to 0. The second property I need to check whether they add up to 1 which is $\frac{8}{8}$ which is equal to 1 and I can see that this indeed is again a probability mass function.

(Refer Slide Time: 28:24)

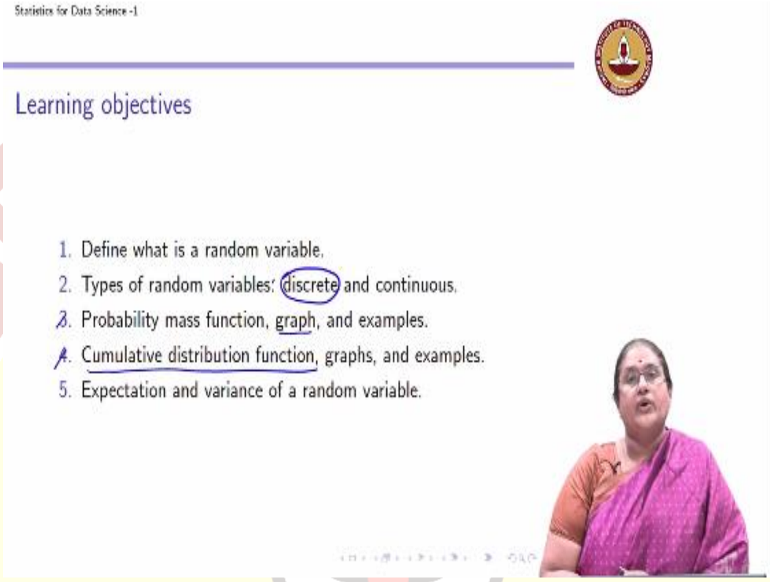


So, by this time what you have to understand is what is the probability mass function. Again remember we are talking only about a discrete random variable. So, what we are looking at is we have defined a discrete random variable which can take finite number of values or it can take countably infinite number of values with $P(X = x_i)$ which is $P(x_1),\ P(x_2), \dots\ P(x_n)$.

Again, same thing when it takes countably infinite number of values $P(x_1),\ P(x_2)$ say it is a probability mass function if each of the $P(x_i) \geq 0$ and summation over all possible values of x_i should be equal to 1, then it is a probability mass function. So, the next is can I graph this probability mass function. So, that is what we are going to look next.

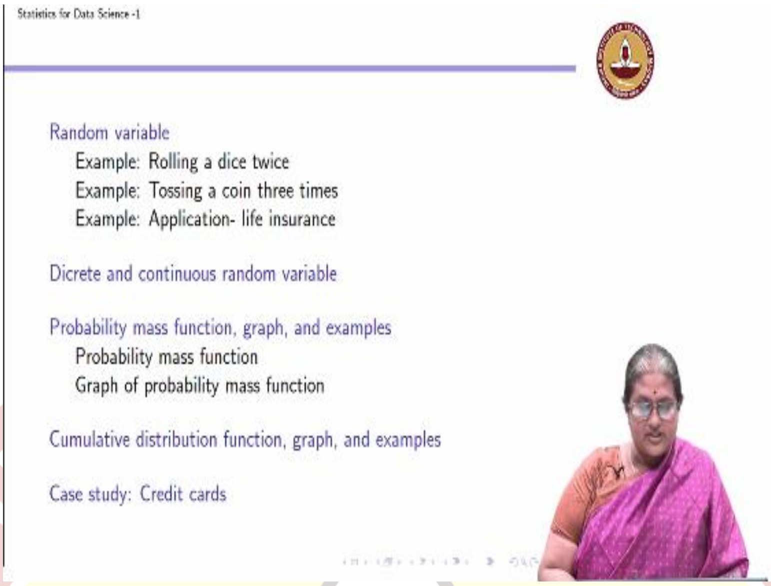
Statistics for Data Science-1
Professor. Usha Mohan
Department of Management Studies
Indian Institute of Technology, Madras
Lecture No. 8.5
Discrete Random Variables - Graph of Probability Mass Function

(Refer Slide Time: 00:14)



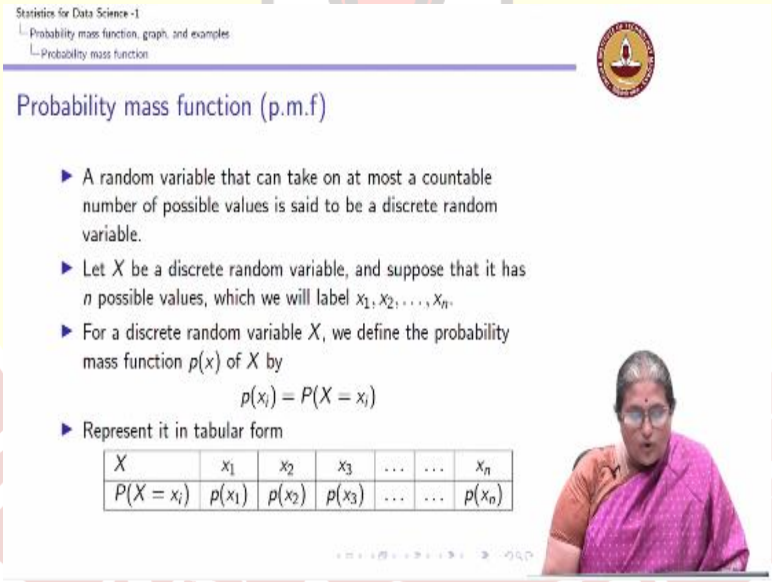
So, so far what we have seen is we defined what is a random variable and then we discussed about discrete and continuous random variables and we defined what was a probability mass function. So, today we are going to continue to understand about what is a probability mass function, we will see what do we mean by a graph of a probability mass function and then we will introduce what is a cumulative distribution function all of this when the variable is a discrete variable that is the learning objective.

(Refer Slide Time: 00:50)



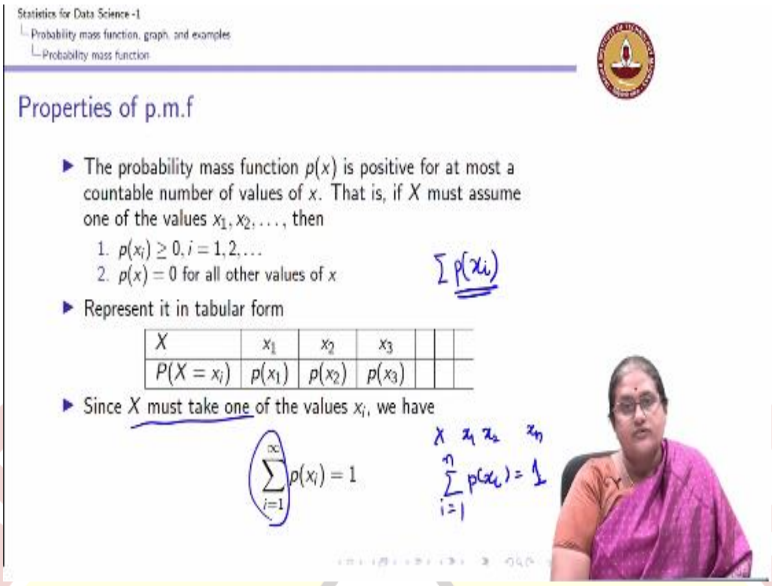
So, let us go and review what we have learned about a probability mass function so far.

(Refer Slide Time: 00:57)



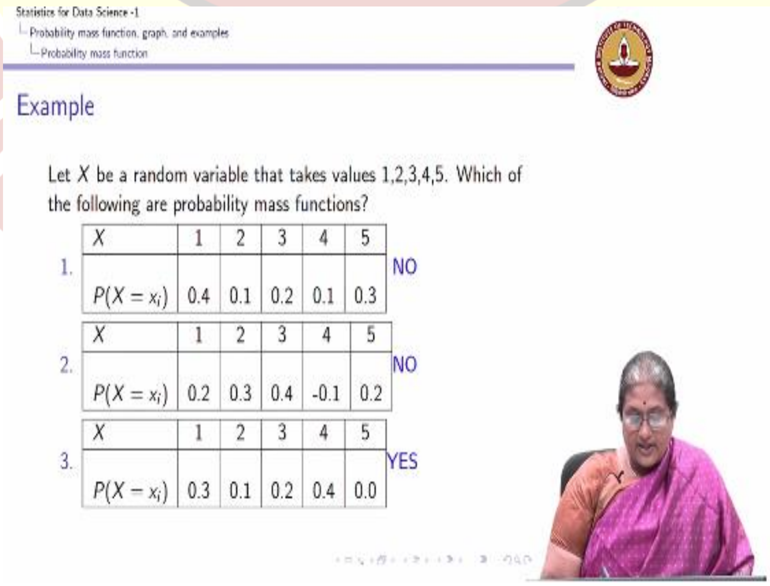
We saw that a probability mass function of a discrete random variable is a random variable which can take at most countable number of possible values. Now if each of these random variables which we are labeling them as x_1, x_2 that is X is taking these values the probability mass function is a probability with which the random variable X takes a particular value x_i.

(Refer Slide Time: 01:35)



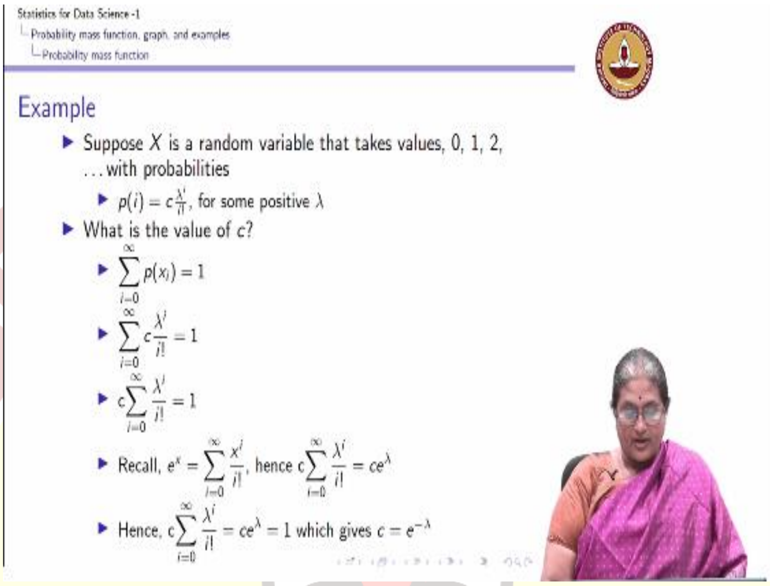
Now since this random variables takes these values, then we know the properties of a probability mass function is that probability since we are talking about a probability, $P(X_i)$ in other words probability of X equal to xi is always non-negative and the next important property is since X must take one of the values summation over all possible values of x. If X takes the values x_1, x_2 up to $\sum_{i=0}^{\infty} P(x_i) = 1$if X takes the value x_1, x_2, x_n finite number of values, then it would be $\sum_{i=0}^{n} P(x_i) = 1$. So, these are the two important and critical properties of a probability mass function.

(Refer Slide Time: 02:32)



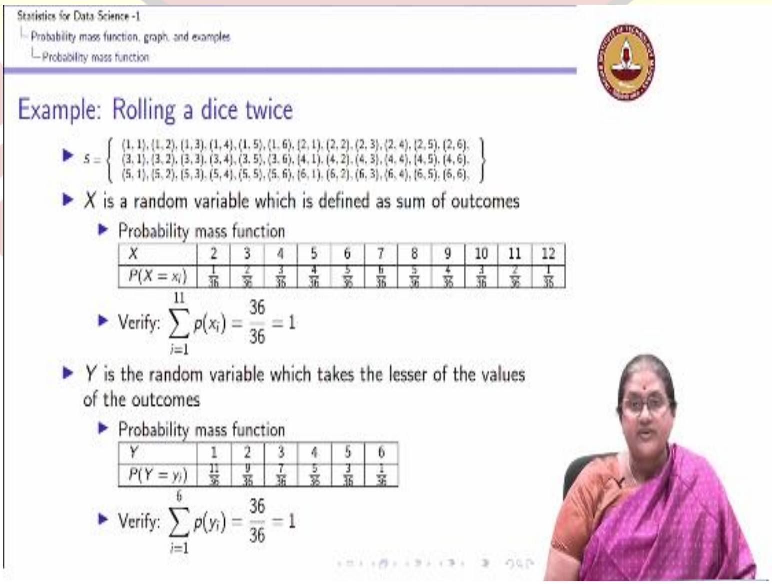
What should we be able to answer? So, we should be able to answer given a probability mass function we should be able to verify whether something is a probability mass function. This we achieve by checking out all the properties.

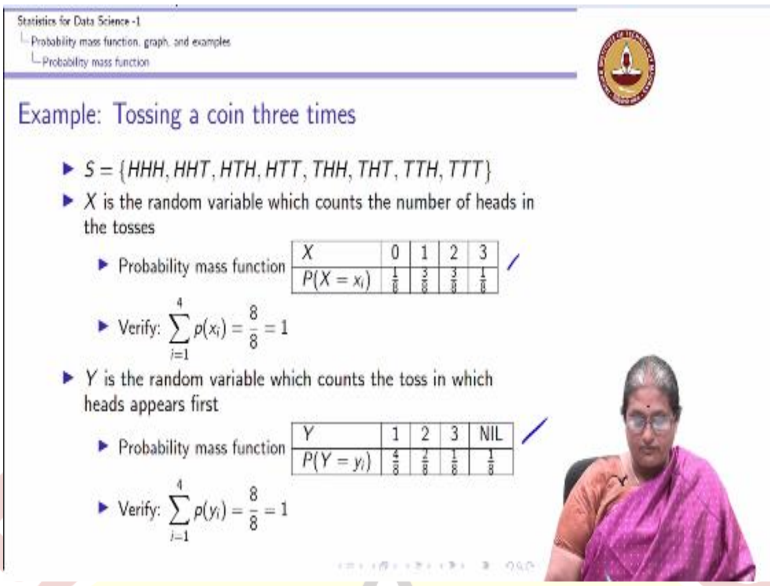
(Refer Slide Time: 02:49)



And then afterwards suppose I have given a probability mass function where I have to find out a constant then we also saw that by equating this to one we can find out what is the value of the constant for which this would define a probability mass function.

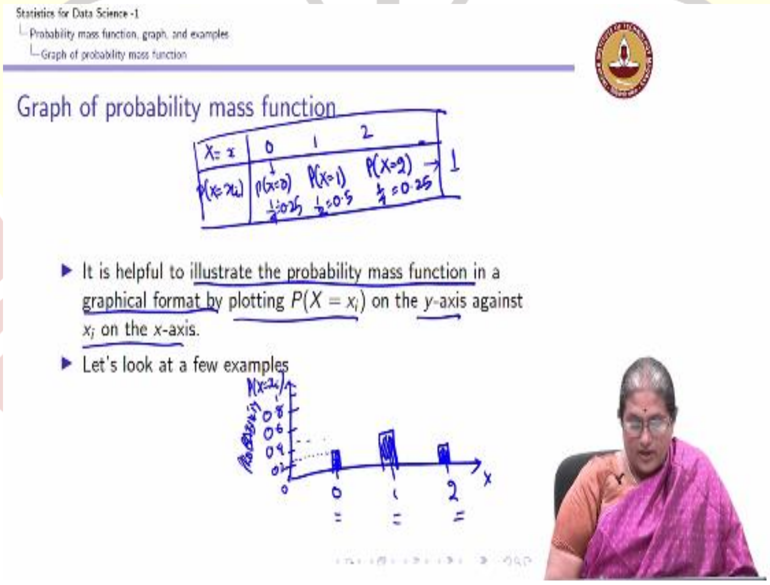
(Refer Slide Time: 03:14)





So, this is what we have done in our earlier case. We went back to the examples that is rolling a dice twice to see what is the tabular form of the probability mass function for all the random variables we have defined. Both in rolling a dice twice and tossing a coin thrice. So, this is what we have seen we have seen these were the tabular form of actually illustrating or presenting my probability mass function.

(Refer Slide Time: 03:44)



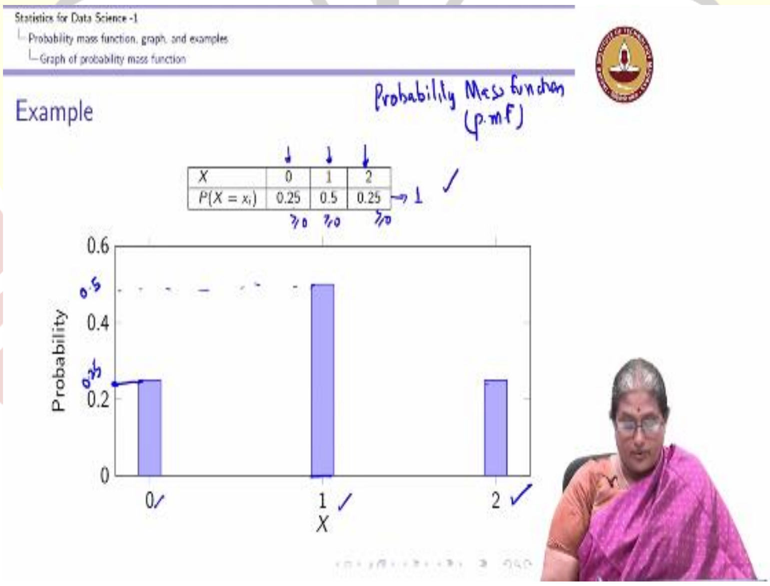
Now we go to the next thing of how can I graph the probability mass function? Now why should I graph the probability mass function? It is always helpful to illustrate the probability mass function in a graphical format. What do I mean by a graphical format? Recall, a probability mass function has the values of X for example if X takes the value 0, 1, 2 then what I have in a tabular form is this would give me what is the probability of P(X=0).

This would give me what is the P(X=1) and this will give me P(X=2) I know the sum of all these probabilities would be 1 and since they are probabilities all of them are greater or equal to 0. So, the question is can I represent or illustrate this information in the form of a graph. So, what do we mean by that? I can plot the probabilities on the y axis against x_i on the x axis.

So, a simple way to discuss this is what are the values X takes here? X takes the value 0, 1 and 2. Remember that these are just values that is taken if they are ordinal then the order has to be maintained on the y axis I am going to have the probability of X taking a particular value. I just write it as probability. So, this is what is my probability mass function. So, suppose probability X equal to 0 is a one fourth, X equal to 1 as a half, I know that the y axis takes values say 0.2, 0.4, 0.6, 0.8 and 1 because it is a probability it would start with a 0 and end with a 1.

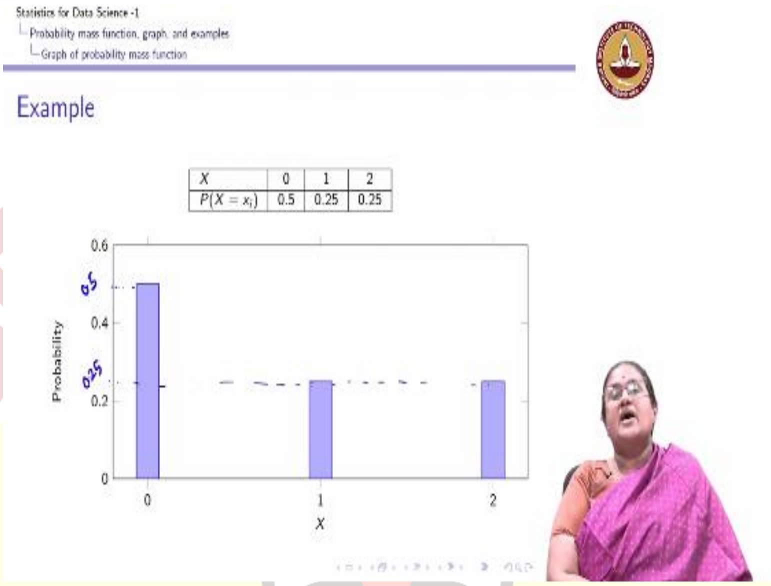
So, P(X=0)=0.25 which is 1/4, 1/2 is a 0.5 again 1/4 which is again a 0.25. I can construct the probability mass function I plot probability so this is a 0.25 so I can go here I construct a bar which is 0.25. Similarly, this is 0.5 P(X=1)=0.5 again P(X=2) = 0.25.

(Refer Slide Time: 06:31)



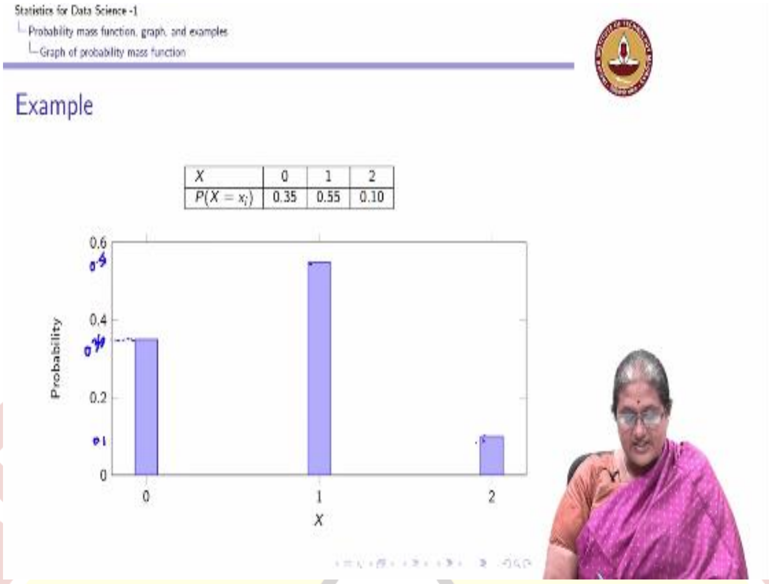
So, let us look at this example so I have this is a tabular form of my probability mass function. I have X equal to 0 is 0.25, 0.5 and 0.25. So, X takes the value 0, 1 and 2. This would be 0.25 so this corresponds to 0.25 this height corresponds to 0.5 this is again corresponding to 0.25. So, this is what we refer to as the probability mass function. It is abbreviated as pmf and the reason it is a probability mass function as we can imagine this as a discrete points that x takes. I have a weight associated and this is what is at every discrete point 0.25, 0.5 and 0.25. Again I can verify that this adds up to 1 so it is a probability mass function all of them are non negative so I do not have a problem.

(Refer Slide Time: 07:56)



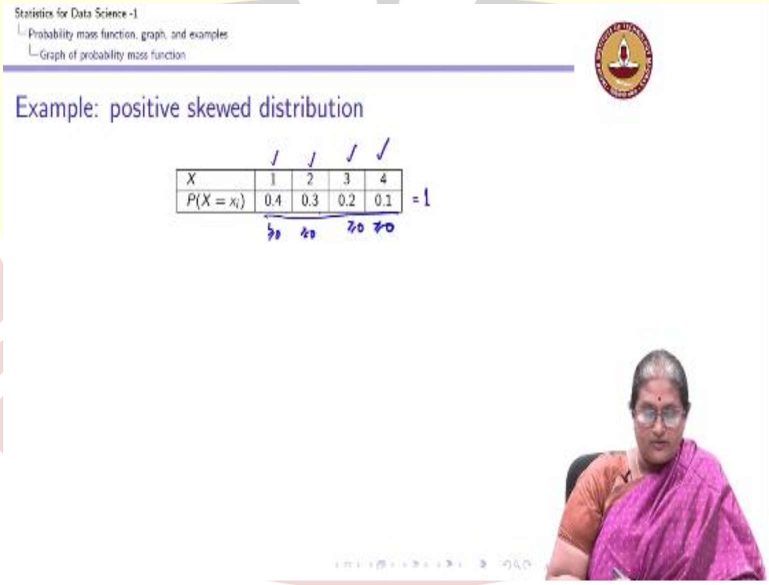
Now let us look into another example. Here again I verify this is 0.5 plus 0.25 plus 0.25 it adds up to 1 all of them are greater or equal to 0 so it is a probability mass function. So, in this case I can see that probability X equal to 0 takes the value 0.5, X equal to 1 takes the value 0.25 so is X equal to 2 so this is my probability mass function. What you can notice here is in the earlier example the distribution looks symmetric whereas here I do not see a symmetricity in my distribution.

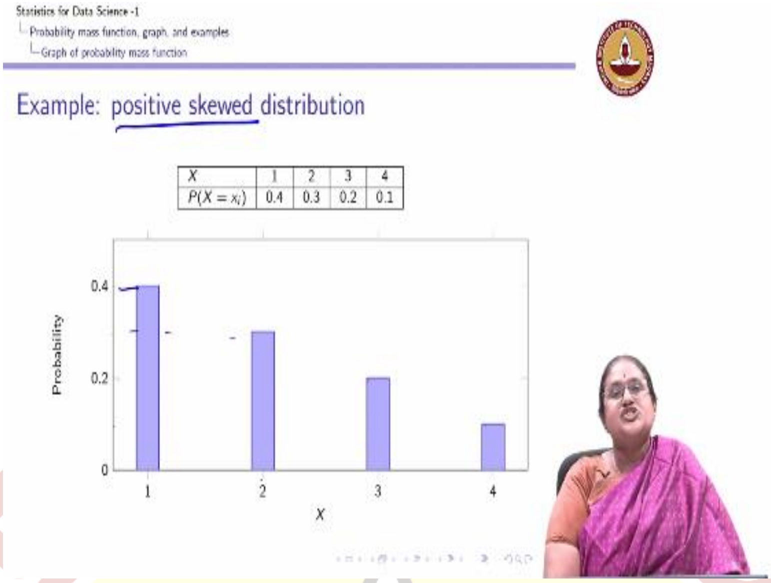
(Refer Slide Time: 08:41)



I could also have a case where I do not have any pattern I have here X equal to 0 is 0.35, X equal to 1 is 0.55 and X equal to 2 is a 0.1 and that is demonstrated or illustrated by this graph. So, why is a probability mass function important or useful?

(Refer Slide Time: 09:15)

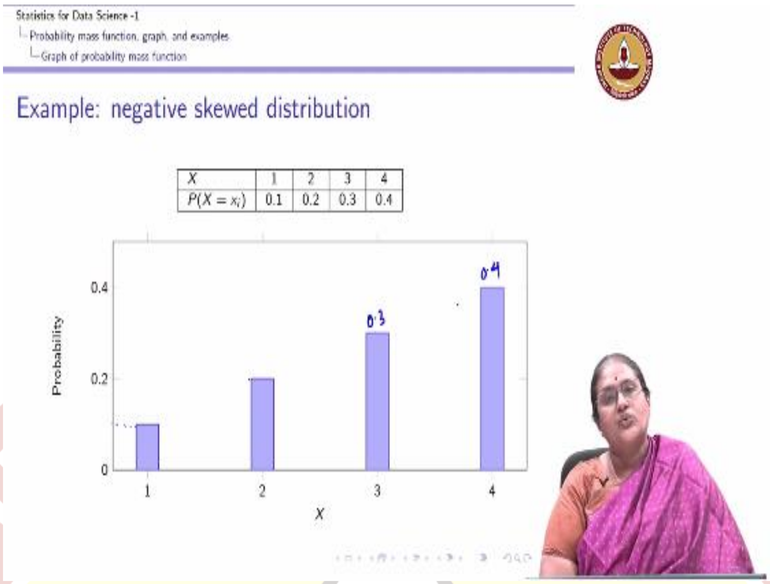




Sometimes we can see that this gives us the shape of the distribution of the random variable. What do we mean by this? For example, if I have this case where X is again is a discrete random variable which takes the value 1, 2, 3, 4 with these probabilities again 0.4 + 0.3 is 0.7, 0.7 + 0.2 is 0.9, 09 + 0.1 all of them add up to 1 and all of them are non negative. So, I have a probability mass function.

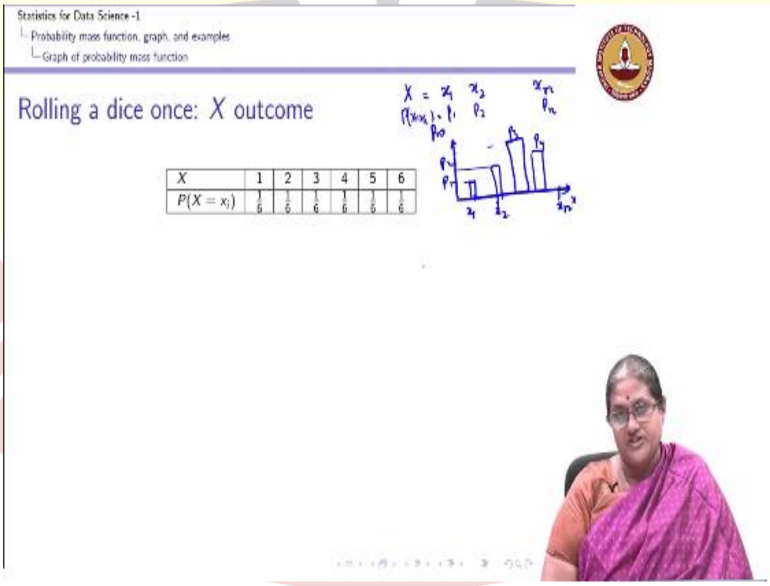
Now if I plot this I can see that this is a 0.4, X equal to 2 is a 0.3, X equal to 3 is a 0.2 and X equal to 4 is a 0.1. What you can see is this distribution exhibits a skewness and these are referred to as positive skewed distributions. So, in advanced courses in addition to the center and variability the shape of a distribution is very important. So, you can see that this is a skewed distribution.

(Refer Slide Time: 10:36)



Now if you look at the other example for example here my 1 probability X equal to 1 is a 0.1, X equal to 2 is a 0.2, X equal to 3 is a 0.3 and X equal to 4 is a 0.4 you have what we refer to as a negative skewed distribution.

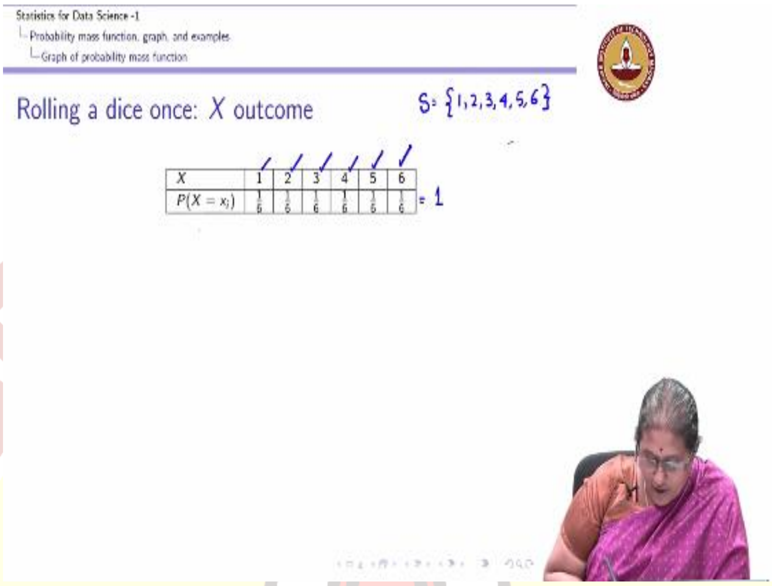
(Refer Slide Time: 10:59)



So, in summary you can see that whenever I have a distribution of x taking values x_1, x_2, x_n with probability of X equal to x_i this p_1, p_2, p_n. For now I am assuming X takes finite number of values, then the graph which plots x_1, x_2, and x_n on the x axis and the probabilities that is this would be corresponding to p_1 this could be p_2 this could be p_3 and this could be my p_4 and so forth this is referred to as a graph of a probability mass function.

So that is what we have seen so far. So, now let us go back to the examples that we have considered and look at how the PMF can be illustrated using a graph.

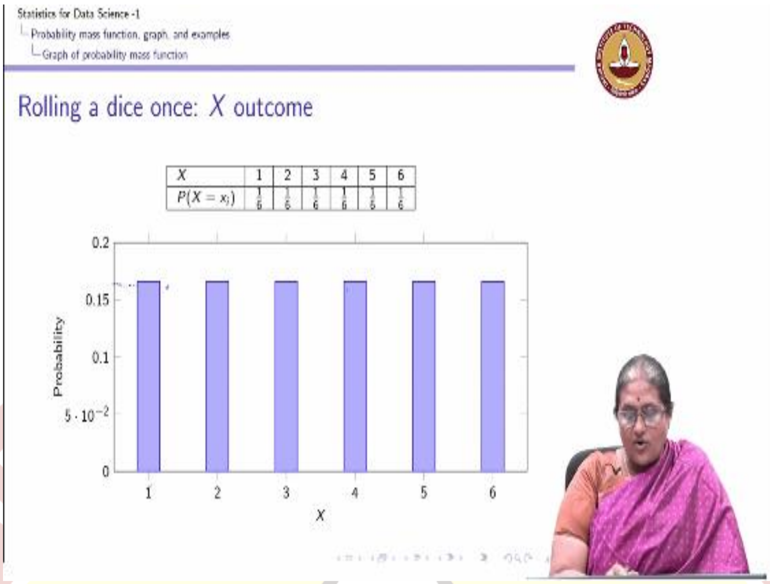
(Refer Slide Time: 12:08)



So, let us start with rolling a dice 1 I know the sample space for this is going to be the following. I am rolling a 6 sided dice it is a fair dice so I am rolling it only once. I know that the outcomes could be any one of the 6 outcomes it is a fair dice. So, the random variable takes the values 1, 2, 3, 4, 5 and 6 with the probabilities because it is a fair dice the chance of getting a 1 is the same as the chance of getting a 2 and so forth.

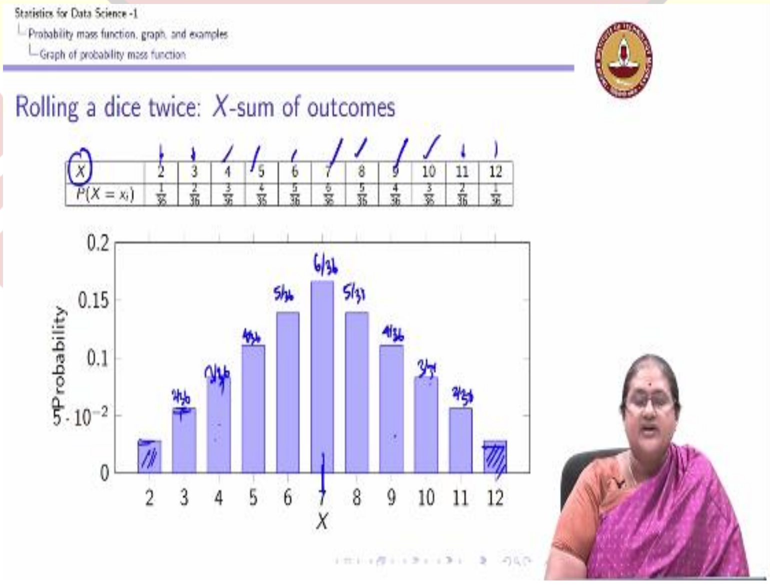
So, I have listed the probability of X taking the values 1 to 6 is the same which is 1 by 6, 1 by 6, 1 by 6 so forth. We can again verify that the sum is equal to 1 all of them are greater than or equal to 0 all the probabilities are now negative. Hence it defines a probability mass function.

(Refer Slide Time: 13:17)



Now if I plot this probability mass function I can see that it is a constant you can see that for every value X takes the height of the bars are the same which is around 1.16. So, this is one shape so the graph of a probability mass function illustrates that there is some degree of uniformity in my distribution. So, immediately you can see that this looks like a uniform distribution. We will talk about distributions later, but what you can see from the graph is immediately you can see all the bars are of the same height.

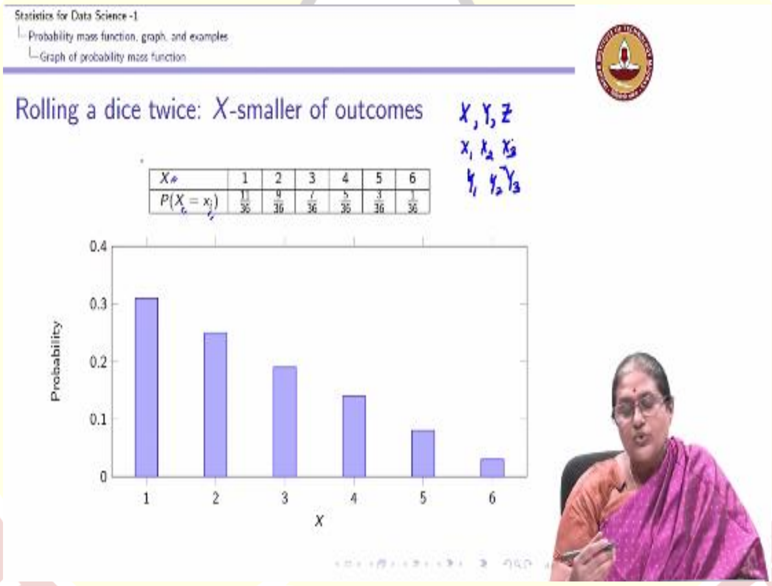
(Refer Slide Time: 14:06)



Now let us go back to rolling a dice twice. Again here what is a random variable? Random variable is the sum of outcomes so we have already seen that this is the tabular form X takes values 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 with the respective probabilities as shown in the table. Now if I come and look at the probability mass function you can see that X takes the value 2 with a very, very low probability. X takes the value 3, but X takes the value 2 and 12 with the same probability as shown here by the bars.

Similarly, X X takes the value 3 and 11 with the same probability which is 2 by 36 it takes the value 4 and 10 with the value 3 by 36 it takes the values 5 and 9 with the value 4 by 36 probability. It takes the value 6 and 8 with the probability is 5 by 36 and it takes the value 7 with the probability 6 by 36. So, again you can see that the distribution of this random variable is symmetric about a particular point. You can see that there is a element of symmetricity or you can see a symmetric behavior in the distribution.

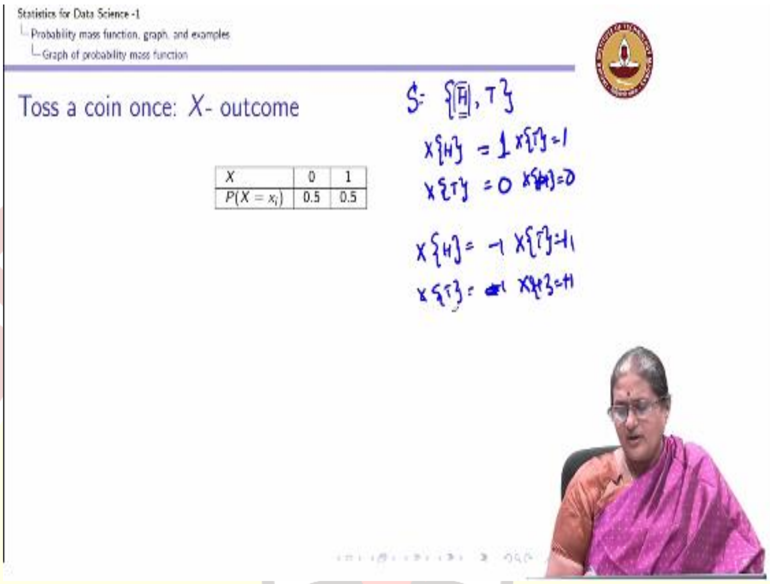
(Refer Slide Time: 15:50)



Now let us look at smaller of the outcomes again we know that X takes values 1, 2, 3, 4, 5, 6 the sum adds up to 1 once I plot this I know X takes these values you can see that it is a skewed distribution. Again recall, both these random variables were defined on the same random experiment and sample space. It depends on what is it you are actually mapping each of the outcomes to and you can see that the distribution in one case was a symmetric distribution in the other case it was a skewed distribution.

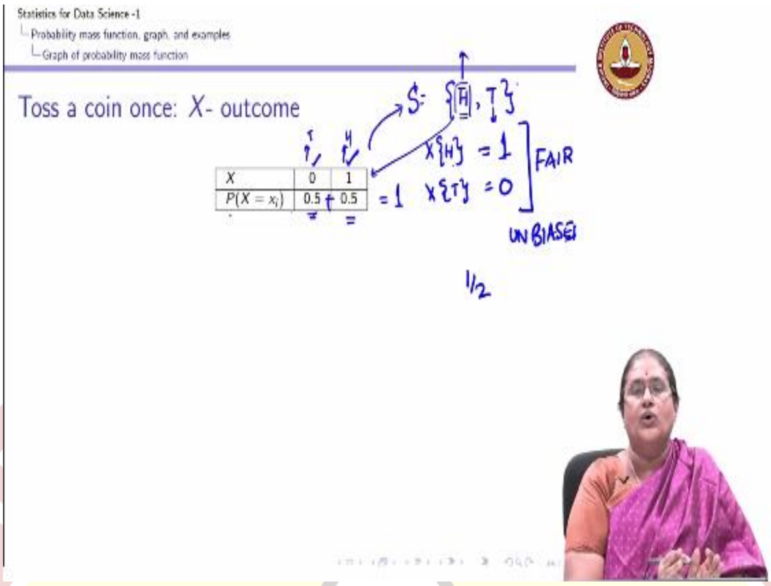
I just want to make a small point here when we discuss this random variable I had defined it as a y and I said probability Y equal to yi here I am referring it to as X and I am looking at X equal to xi. It really does not make any difference whether you are referring it to X or Y only thing you need to understand that what represents the random variables and what represents the value they are of the random variable. It could be X, Y, Z. So, typically we also saw that random variables are typically expressed with upper case alphabets X, Y, Z or X_1, X_2, X_3 or Y1, Y2, Y3 these are typically what are used to represent random variables.

(Refer Slide Time: 17:37)



Now let us look at the tossing a coin once. Now again what is a sample space for my experiment? When I am tossing a coin once I can get a head or a tail. Now again if I want to just note the outcomes I would associate a value to these outcomes. There are two ways I can do this mapping, either I can map head to 1, tail to 0 or I can map a tail to 1 or head to 0. So, you can see that a random variable just maps the value of an outcome to a number. Can we map head to a minus 1 and tail to a plus 1 or tail to a minus 1 and head to a plus 1 we can do it.

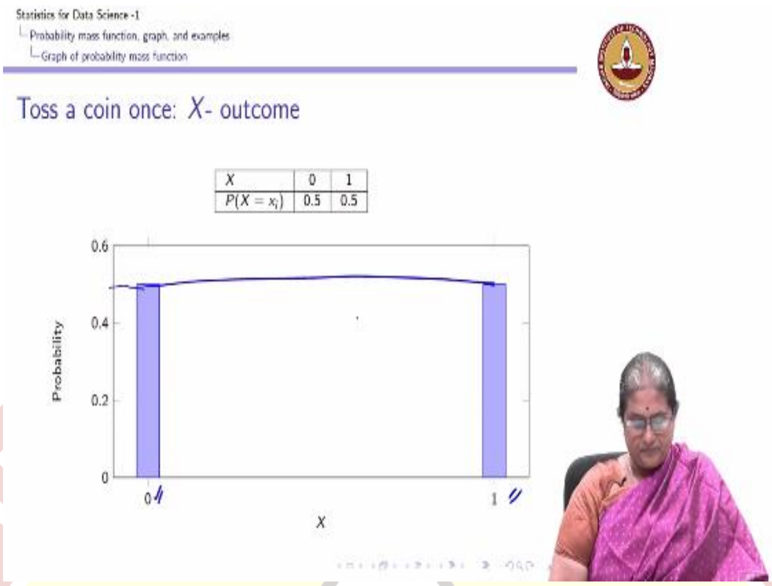
(Refer Slide Time: 18:41)



But for convention sake let us always for just our understanding now we will map head to 1 and tail to 0. Whatever I am mapping I have to define it very clearly. So, if head takes the value 1, I know the random variable again takes the value 0 and 1. Again, I assume that my coin is a fair coin or an unbiased coin. If I have a fair coin or an unbiased coin, the probability of getting a head is the same as probability of getting a tail which we know is half.

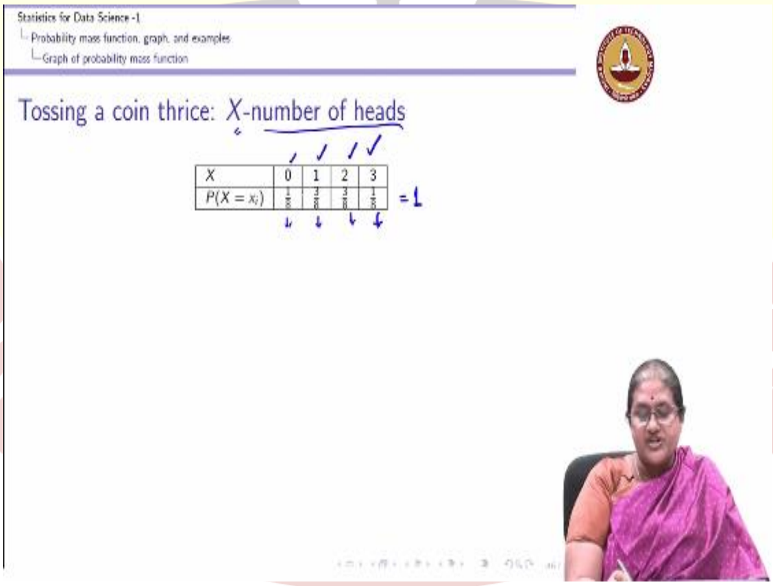
This is something which I know the probability of getting a head is the same as probability of getting a tail which is half so I am just putting here probability X equal to xi is 0.5, 0.5 both of them I add that becomes 1. This for now is a tail and this is a head. Whenever we are doing this mapping we need to understand what was our original experiment and what has been the mapping or how we have associated every outcome to a random variable that is something which we need to know.

(Refer Slide Time: 20:09)



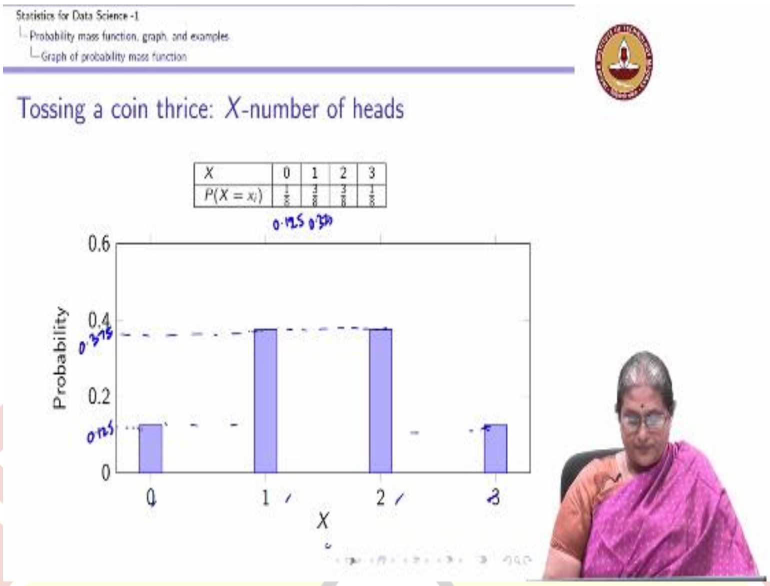
Now given this x takes only two values 0 and 1 with the same probability again we can see that this probability since both of them have the same probability there is some sort of a constant or a uniformity which is illustrated here.

(Refer Slide Time: 20:31)



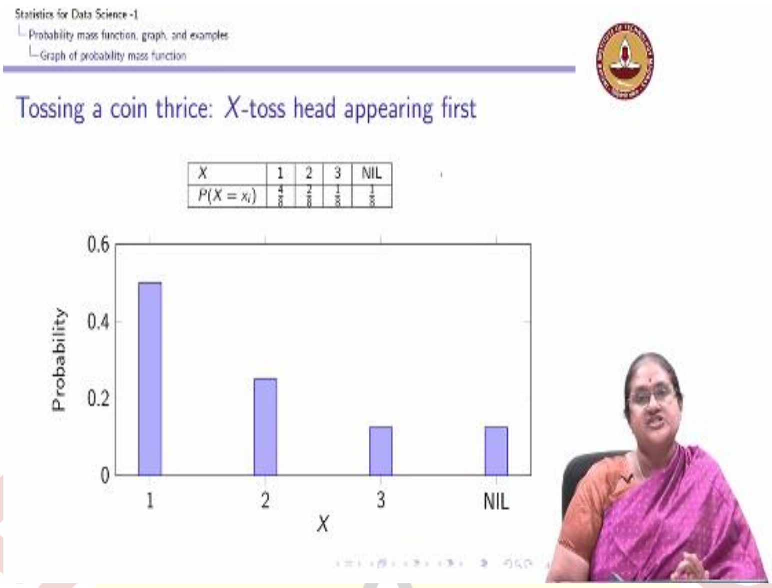
Let us go to the example of tossing a coin 3 times again my random variable was counting the number of heads we knew that the value this random variable takes is 0, 1, 2 or 3 with the probability 1 by 8, 3 by 8, 3 by 8 and 1 by 8. We again verified this is a probability mass function because the sum of the probability is equal to 1.

(Refer Slide Time: 20:59)



Now again if I plot this I can see that X takes the value 0, 1, 2, 3 with probabilities 1 by 8 is 0.125. This is 0.375 so 0.125 is here, 0.375 is here again 2 takes 0.375, X equal to 3 is 0.125 this is 0.375 again you see that this distribution is a symmetric distribution around the values X takes. You can see from a graph there is some amount or it indicates symmetric distribution.

(Refer Slide Time: 21:45)

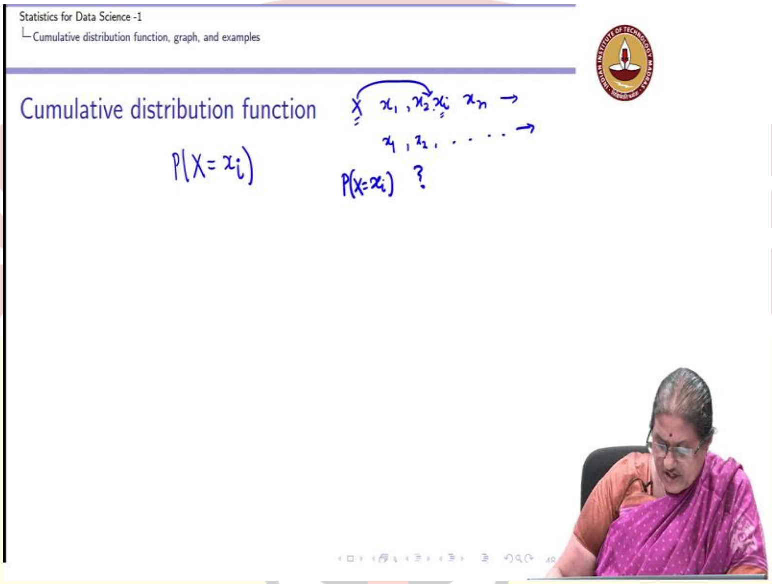


Now let us look at the next example of tossing a coin thrice where I am seeing the toss where head appears first. Again we have seen already that head can take the values, X can take or 1, 2, 3 in other words a head can appear first in the first toss or the second toss or the third toss we defined nil. Technically, this nil is not a real number we should map this nil to a real number we also mentioned about this earlier, but I am just going to retain the nil now.

And then you can see that the distribution of this random variable again exhibit some sort of skewness as we have already discussed earlier what is a skewed distribution. So, you can see that going back to the examples that we have already learned whenever we are talking about the probability mass function we can describe the distribution through the shape of a distribution. So, the next thing which we are interested in knowing is what we mean by a cumulative distribution function.

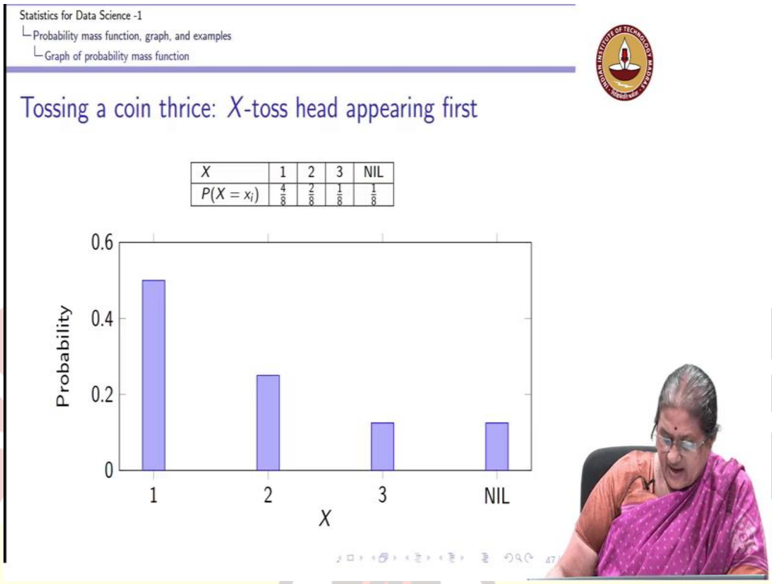
Statistics for Data Science-1
Professor Usha Mohan
Department of Management Studies
Indian Institute of Technology, Madras
Random variables - cumulative distribution function

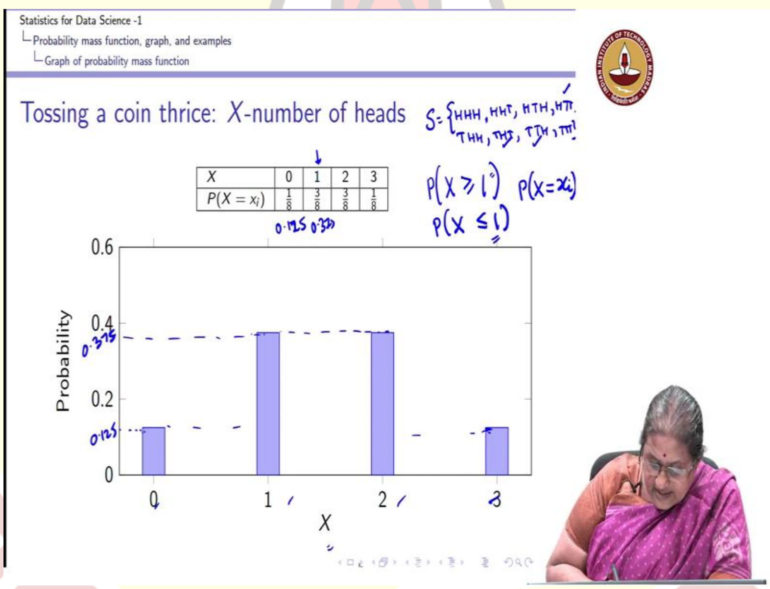
(Refer Slide Time: 0:15)



So, the next concept we are going to learn about is, what we understand by a cumulative distribution function. A probability mass function gave us a $P(X)$ taking a particular value xi. So, we assume that X takes values x1, x2, xn if they take finite number of values, it takes the values x1, x2, so forth if it takes countably infinite number of values and the probability mass function or distribution just told us what is the probability with which X takes a particular value xi. That is what a distribution function tells us. In other words, it tells us that what is the chance of this random variable taking a particular value.

(Refer Slide Time: 1:20)



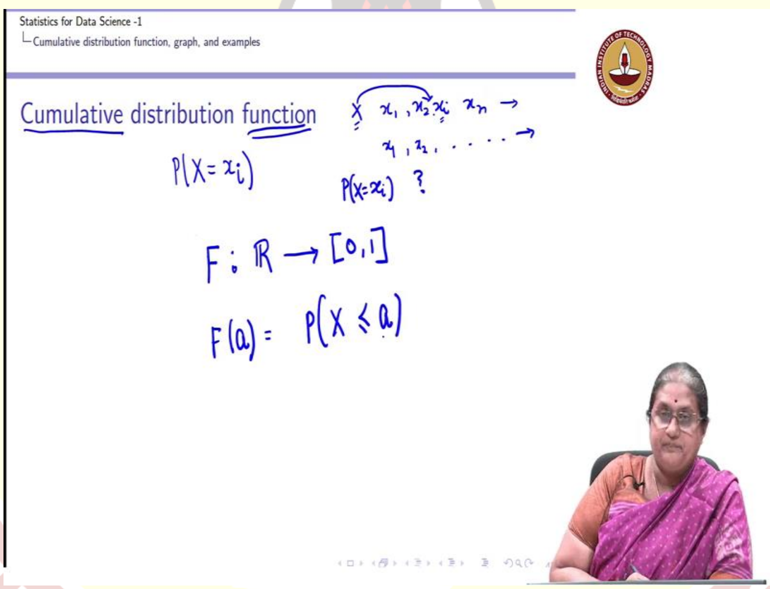


But sometimes we might be interested in knowing for example, we go back to this tossing the coin thrice and look at the solution here. So, you can see that I am counting the number of heads, so I either have, so recall my sample space here was a head, head, head; head, head, tail; head, tail, head; head, tail, tail; tail, head, head; tail, head, tail; tail, tail, head and a tail, tail, tail. Suppose I am interested in knowing what is the chance, so this probability gave me what is the chance of me having 1 head?

I can see that, that corresponds to this outcome, this outcome and this outcome and which is $\frac{3}{8}$. This is how we got a probability mass function. But now suppose I am interested in asking, what is the chance that I have got at least 1 head? So, the way I am translating this is what is the chance that $P(X \geq 1)$? The chance of me getting at least 1 head is equal to this or I could also ask, what is the chance of me getting at most 1 head in 3 tosses? So, in other way, I am asking $X \leq 1$.

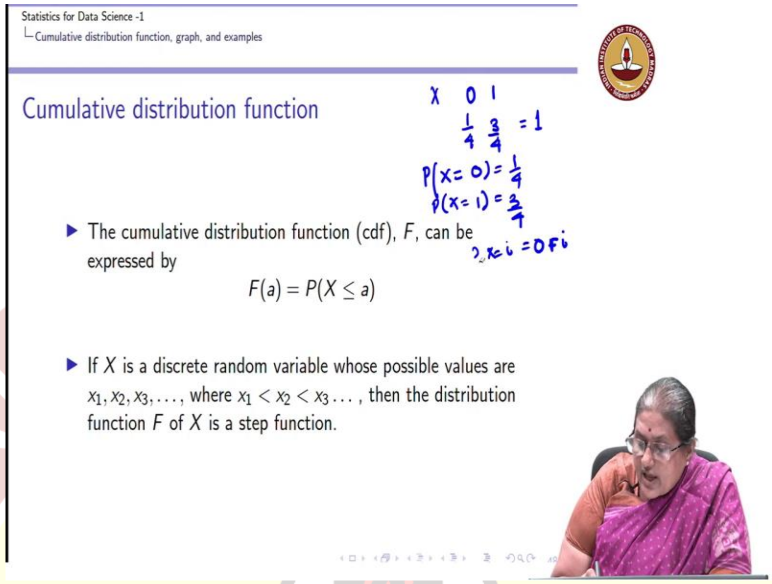
So, instead of looking at what is the chance of X taking a particular value, we might be interested in knowing what is the chance of X is less than or equal to a particular value or greater than or equal to a particular value?

(Refer Slide Time: 3:11)



To answer this question we introduce what is known as a cumulative distribution function. As the name suggests a cumulative distribution is that function which accumulates the probabilities at different points. So, this is a function, so the minute we refer to a function, I need to understand the function is defined on what, this cumulative distribution function which is given by capital F, typically it is referred to as capital F is defined for every real value and it takes values in the closed intervals 0, 1 and how do we define it? For every real value, $F(a)$ is $P(X \leq a)$.

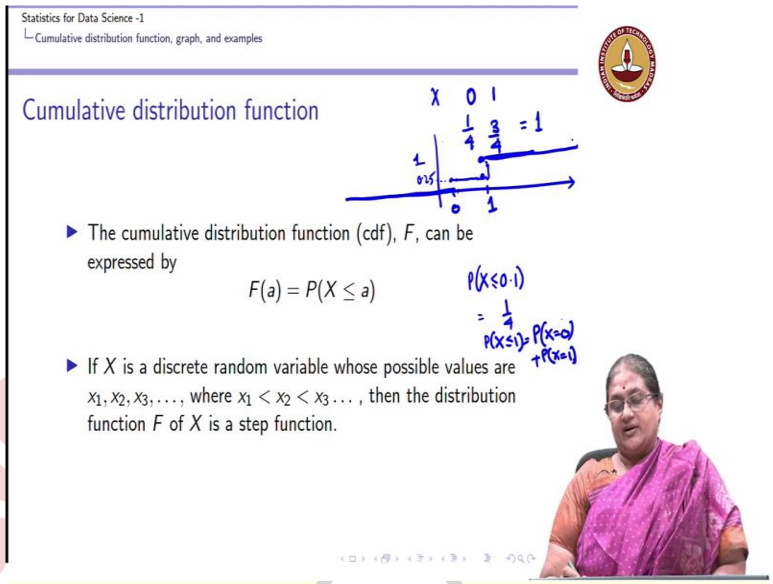
(Refer Slide Time: 4:16)



I repeat, a cumulative distribution function F can be expressed as $F(a)$ is $P(X \leq a)$. So, given this $F(a)$ is defined for every value of a on my real line. So, let us look at an example. For example, if X is taking values, whose possible values are X_1, X_2, X_3, let us start with a very simple example.

X takes finite value 0 and 1 with probability 1 by 4 and 3 by 4. I know that the way my pmf is defined is $P(X = 0)$ is $\frac{1}{4}$, $P(X = 1)$ is $\frac{3}{4}$, it is a probability mass function because both of them add up to 1. And in addition, $P(X = 1) = 0$ for all other i. So, this is equal to 0 for everything else.

(Refer Slide Time: 5:44)



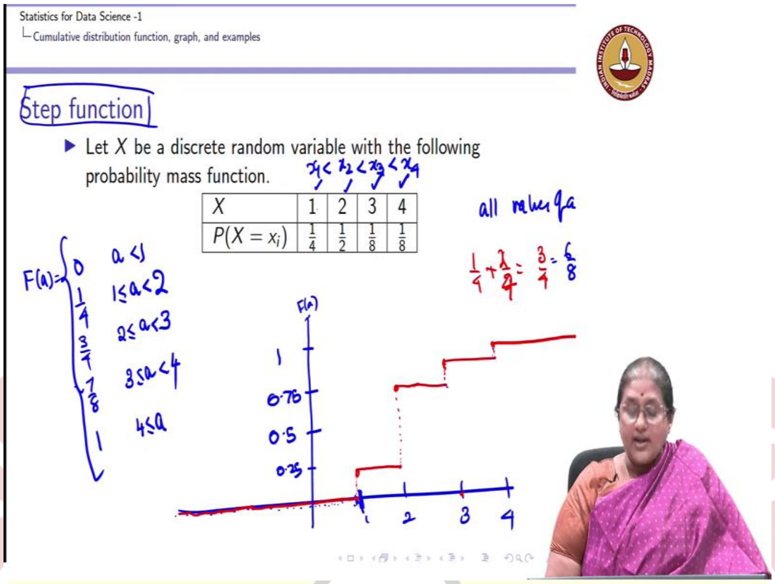
So, what is my chance so if I am going to map every point on my real line to this, then the way I can start is I look at my real line, I have a 0 here so till the point I have touch the 0, my probability is going to be 0 because I have defined $P(X)$ is equal to 0 for all points other than 0.

At the point 0, I have a probability of $\frac{1}{4}$ which is 0.25. At 0.1 so if I am looking at $P(X \leq 0.1)$, I know that this can be $P(X \leq 0) + P(X = 0.1)$, is 0, but probability X is less than or equal to 0 plus probability X is equal to 0.1, I can see that this I already know is $\frac{1}{4}$, this is 0, so this would also be $\frac{1}{4}$.

At 0.2, it will again be $\frac{1}{4}$. At 0.5, it will again be 1 $\frac{1}{4}$. At 0.6 it will again be $\frac{1}{4}$. So, it continues to be 0.4 till the time it hits 1. At X equal to 1, it takes a probability $\frac{3}{4}$. So, what happens? $P(X.1)$, is the same as $P(X \leq 0) + P(X = 1)$, , so it jumps and it takes so this, if this value is 0.25, this value is going to be 1, it is jumping at this value and you can see that it continues to take the value 1 after that. So, this is what is called a cumulative distribution function.

So, what is a distribution function? In the case of a discrete random variable which takes values X_1 which is less than X_2 which is strictly less than X_3. This is a step function. So, now let us look at another example.

(Refer Slide Time: 8:15)



For example, I have X which takes values 1, 2, 3, 4, this is my x_1 which is strictly less than x_2, which is strictly less than x_3, which is strictly less than x_4. This is the value. First step, let us check if it is a probability mass function, yes. 1 by 4, 3 by 4, 3 by 4 is nothing but 4 by, 3 by 4 is nothing but 6 by 8, 6 plus 1 7, 7 plus 1 8, 8 by 8 so it is a probability mass function, all of them are non-negative. So, it is a probability mass function.

So, now again recall, I define $F(a)$ to be $P(X \leq a)$. Now, let us start with the following. I need to define $F(a)$ for all values of a. So, for what is it, so let me start with $(a < 1)$. So, if I go back, I have 1 here, I have 2 here, I have 3 here, I have 4 here. For all a is less than 1, I know probability of x taking any of the values here is equal to 0, because x takes only discrete values 1, 2, 3, 4. So, it is going to be 0. So, I know that let me use a different colour here.

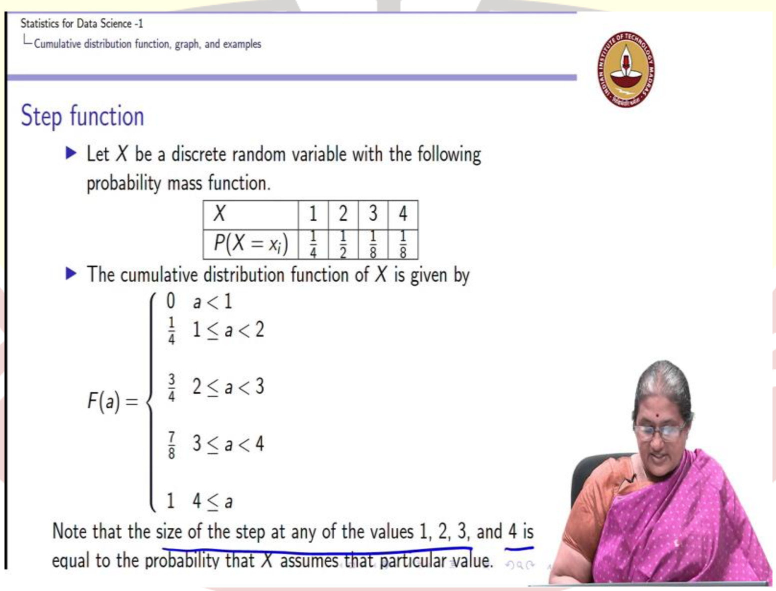
So, I know that till it, so I know that the probabilities or $F(a)$ which is given by the red line is going to be 0 till it hits 1. So, this would be 0. So, $F(a)$ is going to be 0 as long as a is less than 1. Now let me have this portion here which is $F(a)$. So, I am going to write down my $F(a)$ here, so I know $F(a)$ is 0 as long as a is less than 1. Now let us look at the interval, once it hits a, it takes the value 0.25, so let me have a 0.25 here, this is a 0.5, this is a 0.75 and this is 1. There is a 0.25 and it continues till X goes to 1. So, this is what is my value, so it takes the value 1 by 4 for a is between 1 and 2.

So, again, I will plot it using my red line. So, I know that what is it? Between 1 and 2, it is going to be, now there is a discontinuity here and that I am just going to show by dotted lines. Now when I hit 2, the probability of X less than or equal to, so it is again it goes from 1 by 4 Plus 1 by 2 which is nothing but 1 by 4 plus 1 by 2 is this is 2 by 4, so I have a 3 by 4 which is 0.75. So, I have a $P(X < 2)$, it goes here, it continues with the same probability till it hits 3.

So, I have a probability which I write as 3 by 4, for 2 is less than or equal to a is less than 3. Once it hits 3, so this 6, this is nothing but 6 by 8. So, it becomes 7 by 8 so again what happens to my graph from 3 to 4, it is 7 by 8 which is very close to this. Again there is a discontinuity here, after X equal to 4, it continues with 1. There is a discontinuity here.

So, this is how the cumulative, so what do I have here for a so I will have 7 by 8 for 3 is less than or equal to a is less than 4 and 1 for 4 is less than or equal to a and this is what I have plotted here, I know that the plot of this function is what we, it resembles a step, hence it is called a step function We can see that there are discontinuities which I have shown by the dotted line here.

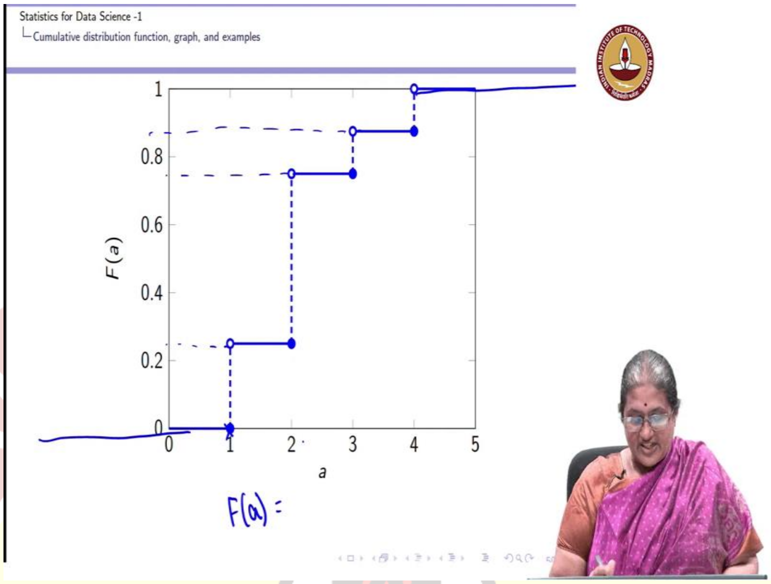
(Refer Slide Time: 13:21)



So, you can see that this is my F of a which was 1 by 4, 3 by 4, 7 by 8 and 1 another thing which you should notice is the size of the step, what is the size of a step here? The size of the step here was 0.75 to 0.25 which is 0.5 which is $P(X = 2)$, the size of the step here is 0.25 which is the $P(X = 1)$, the size of the step here is 1 by 8, the size of the step here is 1 by 8 which are the

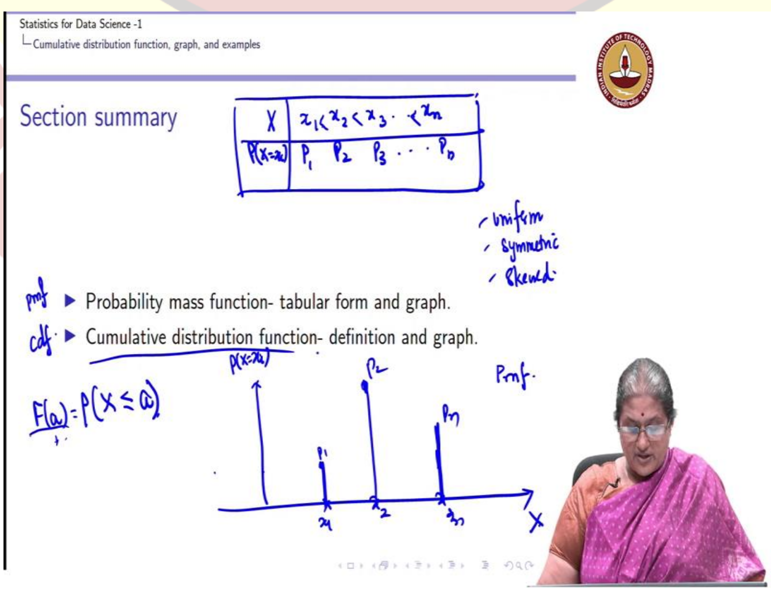
$P(X = 3)$ and $P(X = 4)$ and that is what we have here. The size of the step at any of the values is equal to the probability that X assumes at that particular value and what are the values that X assumes? X assumes 1, 2, 3, 4.

(Refer Slide Time: 14:21)



So, that is how you can see that this is the step function, again for X is less than or equal to 1, it was 0. This is 0.25, this is 0.75, this is 1, 3 by 6, 7 by 8 and after X equal to 4, it continues. So, this is defined for all values of my a. So, the cumulative distribution function is defined for all real values of a and the graph of a cumulative distribution function in the case of a discrete random variable which takes values X_1, X_2, X_n such that X_1 is less than or equal to X_2 is less than, less than X_2 is less than X_n, the distribution, the cumulative distribution is a step function.

(Refer Slide Time: 15:29)



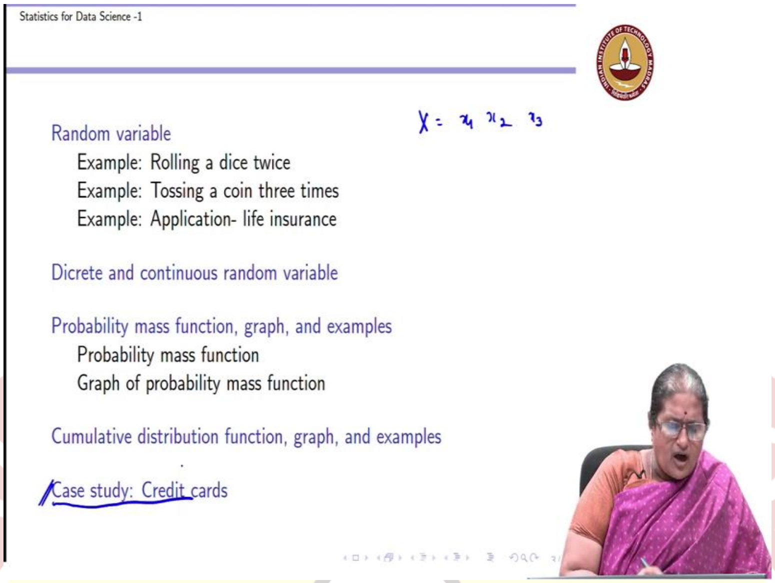
So, in summary what we have seen is, we have seen so far what is a probability mass function. So, given a random variable which takes countably finite or infinite values associated with each one of the values I define what is the probability of X taking a particular value.

I can represent it in a tabular form or I can illustrate it by the values of X on the x axis and the probability of X on the y axis. This is referred to as the graph of the pmf for example, if it take x_1, x_2, xn with probabilities p_1, p_2, and p_n, I know that this is the graph of the probability mass function.

From the graph I can describe the distribution, while describing the distribution I can see whether the distribution is uniform, whether it is symmetric, whether it is skewed. These are the things which we can see from the distribution. And then we further define what was the probably cumulative distribution function, this we define for every value or real value of a which is nothing but probability X is less than or equal to a and in the case the random variable takes x_1, x_2, x_n, with x_1 strictly less than x_2 is strictly less than x_n, then we saw that the graph of this cumulative distribution function is a step function.

Many books refer to the probability mass function as pmf and cumulative distribution function as cdf. We need these concepts to understand the bigger inferential statistics part but this at a conceptual level you need to understand what is a probability mass function and what is a cumulative distribution function.

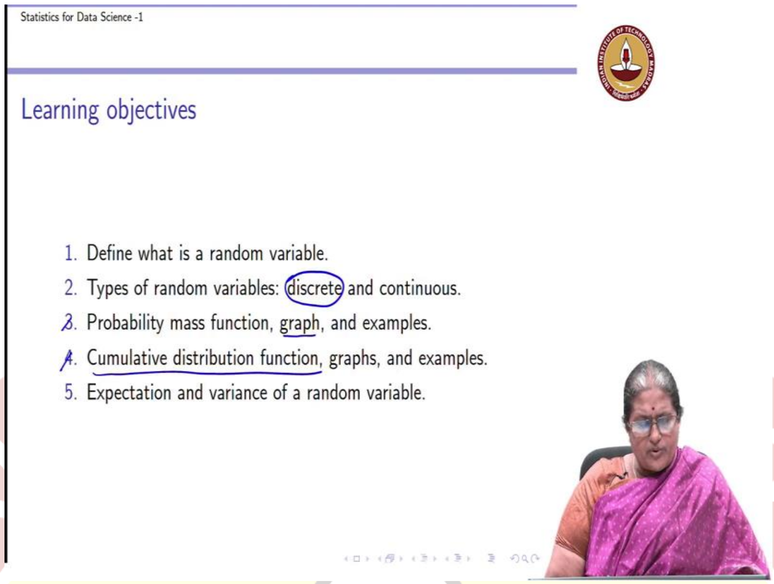
(Refer Slide Time: 17:59)



The next question we are going to answer is again I said we are interested in answering or we are interested in knowing questions about typically we would be interested in answering questions about so, I know now what is a random variable, I know again what is the variable which the values this variable takes.

So, now the next thing which we are going to understand is suppose I am interested in knowing from a population I take a random sample of people and the question I ask them is how many credit cards they own? Again, there is a count of the number, so I can model the response which is the number of credit cards owned by a person as a random variable in particular I can model it as a different random variable.

(Refer Slide Time: 18:55)

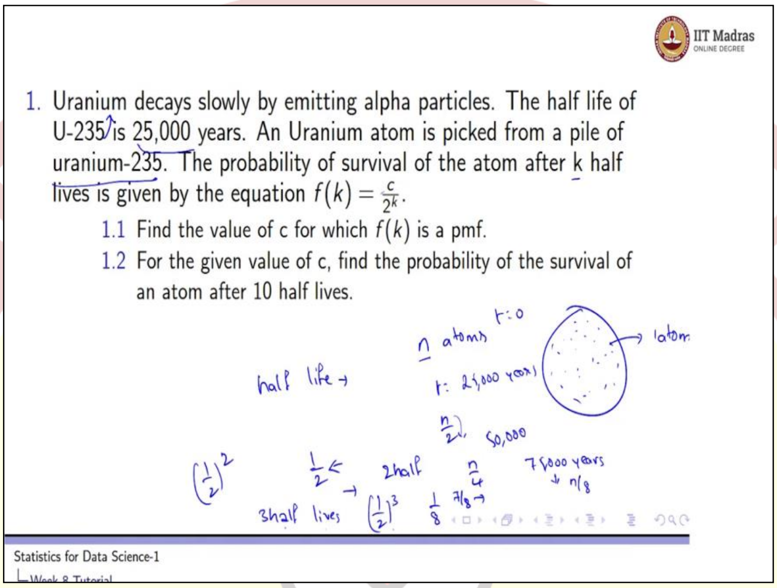


So, we will next look at an application of how to use the concept of both the probability mass function and the probability cumulative distribution function to answer questions from a real application which is about the number of credits cards a person uses. This is what we will be doing next.

Statistics-1 for Data Science
Prof. Usha Mohan
Department of Management Studies
Indian Institute of Technology - Madras

Week 8 - Tutorial 1
Statistics for Data Science-1

(Refer Slide Time: 00:15)



Hello Statistics students in this tutorial will be looking at the week 8 tutorial questions. Let us look at the first question. Uranium decays slowly by emitting Alpha particles. The half life of U-235 this is this nomenclature for uranium U-235 is 25000 years and uranium atom is picked from a pile of uranium 235 or let us assume that you have some so many infinite number of uranium atoms. And from that we picked one particular atom.

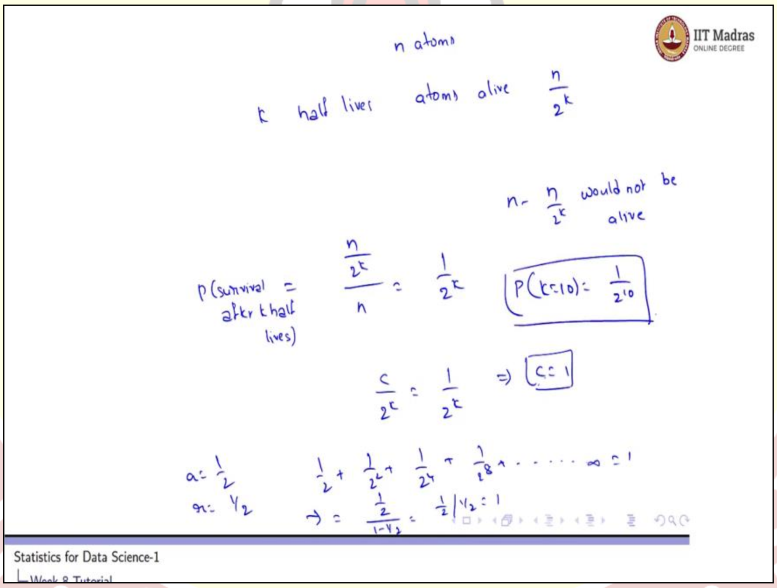
The probability of survival of property of the atom after K half lives is given by the equation $f(k) = \frac{c}{2^k}$. Find the value of c for which $f(k) = 10$. First of all if you look at this question they were saying that half life of U-235 is 25000 years. The first question would be what is half life? Half life is a time taken for uranium for half of its atom to be decaying that means that the number of uranium atoms will be reduced to half after this half life of 25000 years.

Suppose there were n atoms of uranium initially let say time t = 0. So when time t = 25000 years number of atoms of this uranium would be equal to n/2. After 50000 years this n/2 still reduces to n/4 further reduces to n/4. And after 75000 years, it will reduce to n/8; n/4 divided by 2 which is n/8. So this is what half life means. So what they are asking is?

Probability of survival of the atom after k of lives is given by the $f(k) = \frac{c}{2^k}$. So let us look at this expression. So, initially we have n atoms and after one half lives it will be n/2. So that means that only half of the atoms so remained. So probability for an atom to be surviving after one half lives would be 1/2. Since only half of the elements were left over.

And if you want to look at after like 2 half lives, it would be $\left(\frac{1}{2}\right)^2$. After three half-lives would be $\left(\frac{1}{2}\right)^3$ because after 3 half lives only 1/8 th of the atoms are remaining. So for this item to be surviving the probability would be equal to 1/8 because already 7/8 of the atoms after three half lives will be decaying. So the probability after three half lives should be equal to $\left(\frac{1}{2}\right)^3$.

(Refer Slide Time: 03:33)



So you can see pattern here after k half lives let assume initially equal to n atoms. After k half lives atoms left over would be or atoms alive would be equal to $\frac{n}{2^k}$ because after each half life they will be reduced by half. So after k half lives they will be reduced to $\frac{n}{2^k}$. Only these many atoms would be alive. The remaining $n - \frac{n}{2^k}$ atoms would not be alive.

So for the probability; if you look at the probability of survival of the atom after k half lives is given as $\frac{c}{2^k}$. So the problem would be equal to the number of atoms left, $\frac{n}{2^k}$ divided by total number of atoms initial that is n. So the probability would be equal to $\frac{1}{2^k}$, the probability of

Survival after k half lives. So, if you look at the expression there this is $\frac{c}{2^k}$ will be equating to this.

So $\frac{c}{2^k} = \frac{1}{2^k}$ this implies c is equals to 1. Find the value of c for which $f(k)$ is pmf. The probability mass function for it to be pmf; the sum of all these should be equal to 1. That is $\frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^4} + \cdots \ldots = 1$. So let us see since you are substituting c = 1 here, if this is valid.

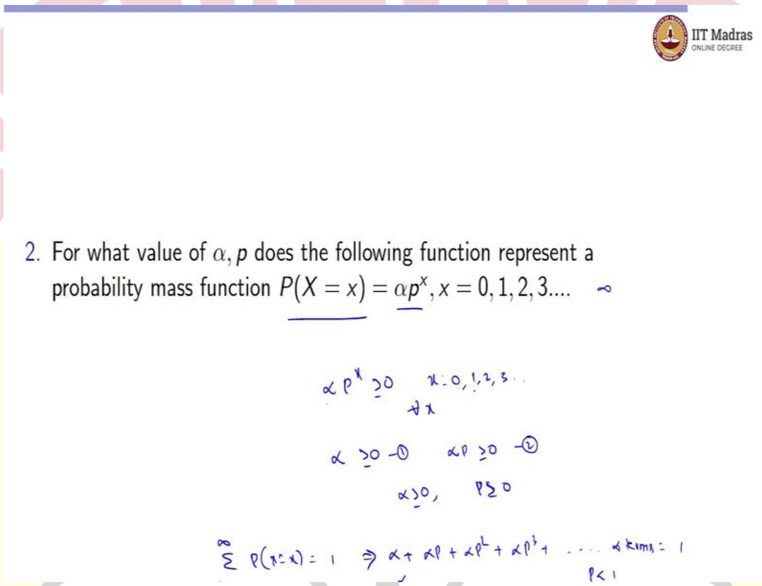
As you have seen from the geometric progression, this is similar to that where $a = \frac{1}{2}$ and $r = \frac{1}{2}$. So, the sum of this should be equal to $\frac{a}{1-\frac{1}{2}} = \frac{\frac{1}{2}}{\frac{1}{2}} = 1$. So we can see that the sum of all these terms all this infinite terms equals to 1. So this is satisfying pmf and for c = 1.

So, the first question is for what value of c for which $f(k)$ is pmf? So c = 1 it would be equals to pmf. So for the given value of c find the probability of survival of atom after 10 half lives? That we have discussed here. The probability of survival after k half lives is given by $\frac{1}{2^k}$. They were asking probability of survival for k = 10 equals that would be equals to $\frac{1}{2^{10}}$. This is the second question answer $\frac{1}{2^{10}}$. Thank you.

Statistics-1 for Data Science
Prof. Usha Mohan
Department of Management Studies
Indian Institute of Technology - Madras

Week 8 - Tutorial 2
Statistics for Data Science-1

(Refer Slide Time: 00:15)



Let us look at the second question. They were saying that for what value of α, p does the following function represent the probability mass function $P(X = x) = \alpha p^x$ for x is equal to 1, 2, 3, and so on infinite. So what are the conditions for this to be happening? The first thing would be each of the values for each of x is equals to 0, 1, 2, 3 and so on.

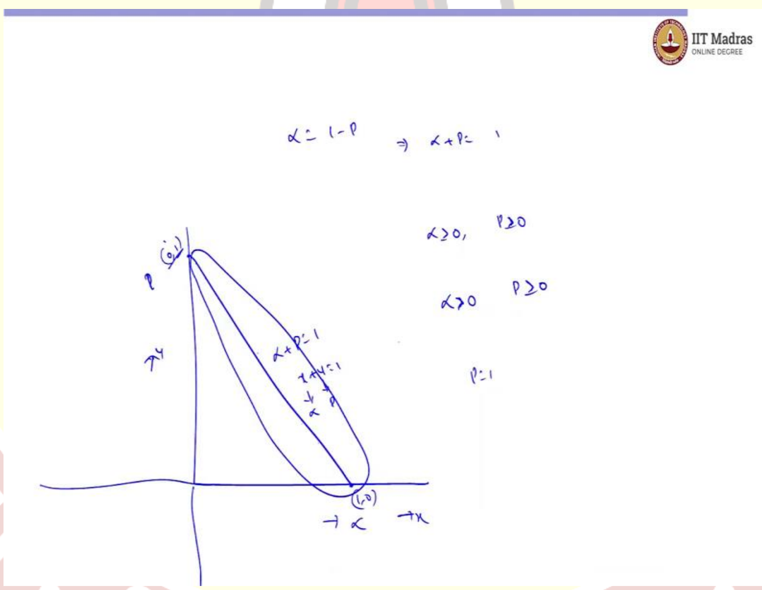
And this value should be greater than zero for all x. So for this to be happening, so, let us take values for x there which is x equal to 0. It will be equals to alpha. So α should be greater than 0 if we put x=1 then it would be equal to α × p. So α p should be greater than 0 from this 1 and 2 we can see that α should be greater than 0, greater than or equal to 0 that is also possible.

P should be greater than or equals 0 both of these conditions are valid for it to be satisfying the probability mass function. There is one more condition that a sum of all the probabilities of x is equals to 0, x is equals to 1, x is equals to 2, x is equals to so on and x equal to infinite should be equal to 1 that is Sigma of x is equals to 1 to infinite probability of x equals to x should be equal to 1. This implies we substitute the values of; sorry here it is x is equal to 0.

So if we substitute the values of x there so the expression would be Alpha since x equals 0 in the first case + α P + αp^2+ αp^3 + some so on infinite terms should be equal to 1. For this to be happening, if you look at at there is only one way this will be possible P should be less than 1. Only in this case this should be converging this is infinite series. It would be converging as we have seen from the geometric progression.

Only for P is less than 1 this is possible and we will be assuming P is less than 1 and solving this problem. So if you assume P is less than 1 then this infinite series sum would be equals to a by 1 – r here a is equal to α and r is equals to p. So a to be equals to α × (1 – P) that is equal to 1.

(Refer Slide Time: 03:10)



So, if you solve this α is equals to 1 – P implies α + P is equals to 1 and we have seen that alpha should be greater than or equal to 0 or P is greater than or equal to 0. We will be remembering the case of α is equal to 0 because; if alpha is equal to 0 this entire series should be equal to 0. Sum would be equals to 0 and we would not be having this condition of sum is equal to 1.

So α is equal to 0 is eliminated so α should be greater than 0. And for this P it could be 0 because let us say for x is equals to; let us say in the case of alpha is equals to 1. α is equal to 1 this implies P would be equal to 0. What does this say is that only for x is equal to 0 we will be having P of x is equal to 0 is equals to 1 and 0 for all the remaining cases. That would still be the probability mass function.

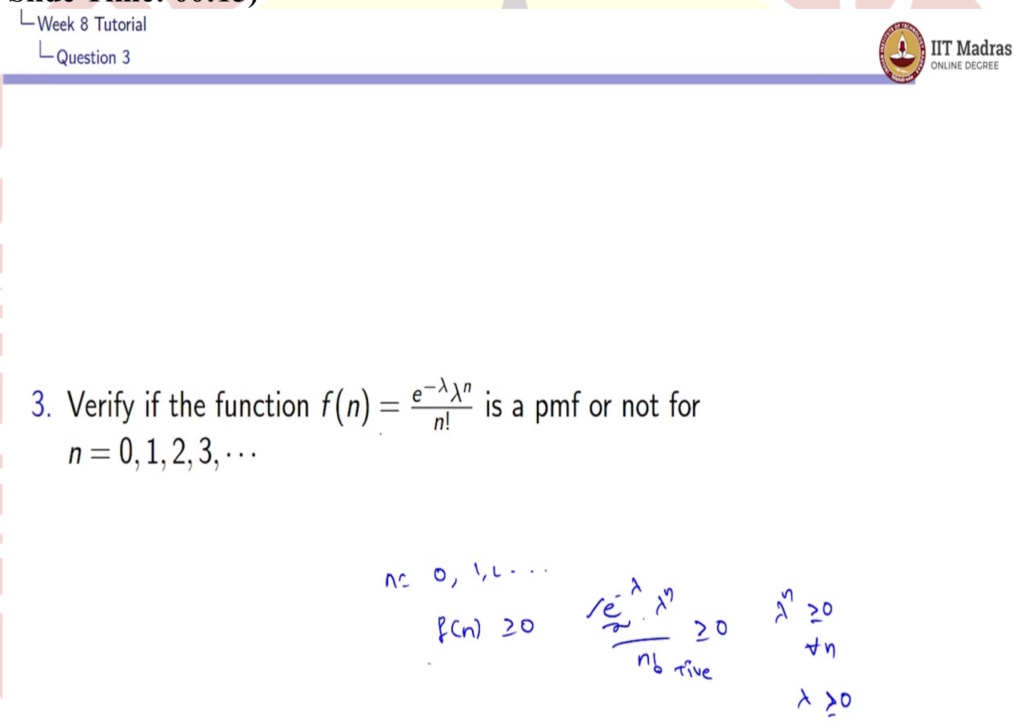
So, the value of all α and P for this it would be possible is; all the values of α, P we draw the line there with α + p equals to 1. It would be a straight line. Let us assume this is α. This is P, this should be equal to 0,1. And this should be equal to 1,0 and all these values falling here except for 0, 1. Except for this P is equal to 1 case for all the points in the straight line this condition will be satisfied.

For any value of α, P which in this pattern this is α + P is equals to 1 and if we consider this as X and this as Y we can write it as X + Y is equal to 1 but X should be equal to α and Y would be equals to P. So all the points that are in this; where $\alpha > 0$and P is greater than 0 for all the elements of α, P this would be possible this condition would be applied so that is second question.

Statistics for Data Science
Prof. Usha Mohan
Department of Management Studies
Indian Institute of Technology - Madras

Week 8 - Tutorial 3
Statistics for Data Science-1

(Refer Slide Time: 00:15)



Let us look at the third question in the tutorial. They have given that; they have asked us to verify if the function $f(n) = \frac{e^{-\lambda}\lambda^n}{n!}$ is a pmf or not for n is equals to 0, 1, 2, 3 and so on. There are 2 cases for this to be pmf as we know. First is for each value of n is equals to 0, 1, 2, 3 and so on infinite. This $f(n)$ should be greater than or equal to 0 that means that $\frac{e^{-\lambda}\lambda^n}{n!}$ should be greater than or equal 0.

We know that the e^{-x} will never be equals to; sorry e^{-x} would never be less than 0. So this should be always positive and $n!$ is always positive. This means that λ^n should be greater than 0 for all n. For this to be possible, $\lambda \geq 0$. Only if $\lambda \geq 0$ we will be having $f(n) \geq 0,$ for every n of 0, 1, 2, 3 and so on.

(Refer Slide Time: 01:36)

$$\sum_{n=0}^{\infty} f(n) = 1$$

$$\Rightarrow \sum_{n=0}^{\infty} \frac{e^{-\lambda} \lambda^{n}}{n!} = 1$$

$$\Rightarrow \quad = e^{-\lambda}\frac{\lambda^0}{0!} + e^{-\lambda}\cdot\frac{\lambda^1}{1!} + e^{-\lambda}\frac{\lambda^2}{2!} + e^{-\lambda}\frac{\lambda^3}{3!} + \dots \infty$$

$$e^{x} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots \qquad = e^{-\lambda}\underbrace{\left(1 + \frac{\lambda}{1!} + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \dots\right)}_{e^{\lambda}}$$

$$= e^{-\lambda}\cdot e^{\lambda} = e^{(-\lambda+\lambda)} = e^{0} = 1$$

So the second condition is the sum of n $\sum_{n=0}^{\infty} f(n) = 1$. This implies $\frac{e^{-\lambda}\lambda^n}{n!}$ plus this implies this expression will be equals to $\frac{e^{-\lambda}\lambda^0}{0!}$, for n=0 plus $\frac{e^{-\lambda}\lambda^1}{1!} + \frac{e^{-\lambda}\lambda^2}{2!} + \frac{e^{-\lambda}\lambda^3}{3!} + \cdots \infty$.

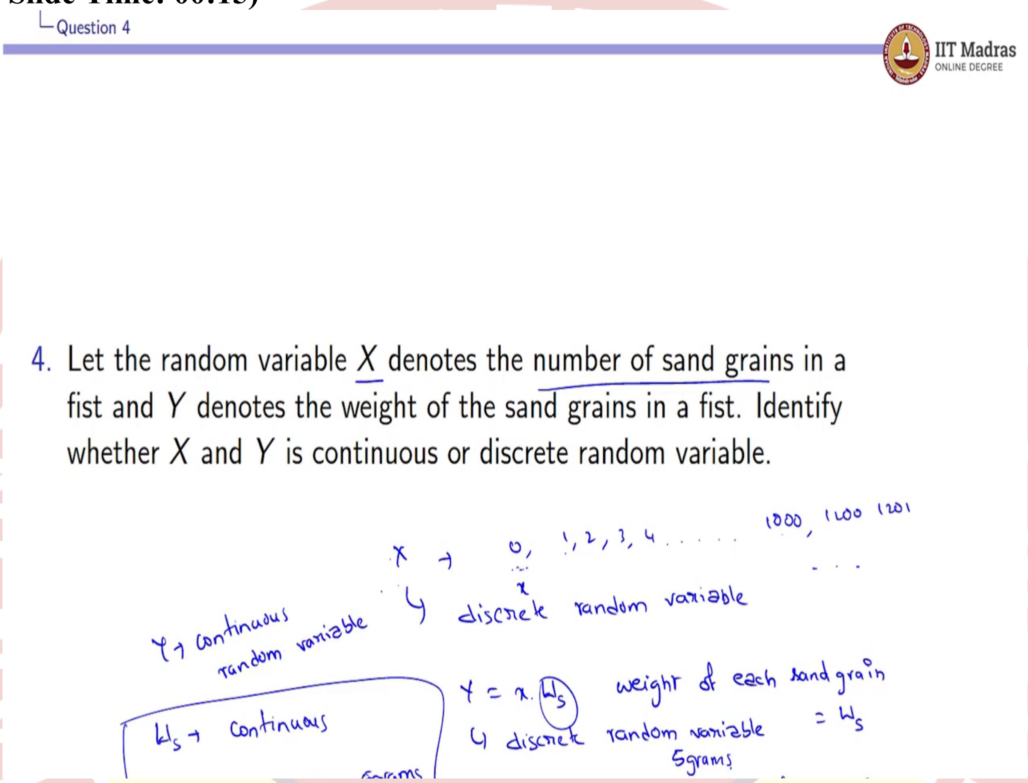
So this value if you look at it this should be $e^{-\lambda}$, if you take it common. This should be $1 + \frac{\lambda}{1!} + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \cdots \infty$. So this is nothing but the value of e^{λ} and we have seen that the expansion of e^{x} is $1 + \frac{1}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots \infty$. This is exactly similar to that; except that in the place of x, we have λ there.

So that would be equals $e^{-\lambda}e^{\lambda} = e^{-\lambda+\lambda} = e^{0} = 1$. So, this function $f(n) = \frac{e^{-\lambda}\lambda^n}{n!}$ is a pmf, so this is correct.

Statistics for Data Science
Prof. Usha Mohan
Department of Management Studies
Indian Institute of Technology - Madras

Week 8 - Tutorial 4
Statistics for Data Science-1

(Refer Slide Time: 00:15)



Hello, so we look at the 4th question in the week 8 tutorial questions. They have given that a random variable X denotes the number of sand grains in a fist and Y denotes the weight of sand grains in the fist. So if you look at it, the number of sand grains that is nothing but the count there. So the given X is to be the random variable. So the possible values for that would be either the number of sand grains in your hand would be 0, 1, 2, 3, 4 and so on.

And it could be some 1000 sand grains since they are like very minute it could be 1000 or like 1200 or 1201 and so on it could be any value. But if you look at it, it could be only the natural numbers. If you consider 0 as your natural number and the values you will take is only this natural numbers 0, 1, 2, 3, 4 and so on any number of values. So clearly it is a discrete since we do not have any values in between 0 and 1.

The values it could take is only discrete. So here X would be discrete random variable. And the next thing is they are asking Y denotes the weight of sand grains in a fist. Identify whether X and Y is a continuous or discrete value. So, for Y there will be 2 cases here, let us look at the first condition, if we assume weight of each sand grain is equal to some W_s then this Y value will be equals to; let this all this value of be it x all the possible values for random variable will be x.

This $\boxed{x}$ could be 0, 1, 2, 3, 4 and so on which ever number but it should be natural number as we saw. So this should be equals to $\boxed{x.W_s}$. Suppose there were $\boxed{x}$ number of sand grains then we will be having $\boxed{x.W_s}$ weight of sand grains. Suppose this $\boxed{W_s}$ equals to 5 grams again this is just hypothetical and if this for different values of $\boxed{x}$ it could take the value either 0 grams there were no sand grains in the fist or 5 grams when there were only 1 sand grain or 10 grams or 15 grams or so on.

It could be for this value of 1000 it would be 5000 weight some 5000 grams. It could gain other values. If you look at the values of $\boxed{Y}$ would be 0, 5, 10, 15 and so on. Here also, it can take any value in between this 0 and 5 it cannot take value of 1, 2, 3, 4. Because there will not be half sand grain like $\boxed{1/4}$th sand grain it would not like that. So that is why the way the possible values for this $\boxed{Y}$ would be 0, 5, 10, 15 and so on up to 5000.

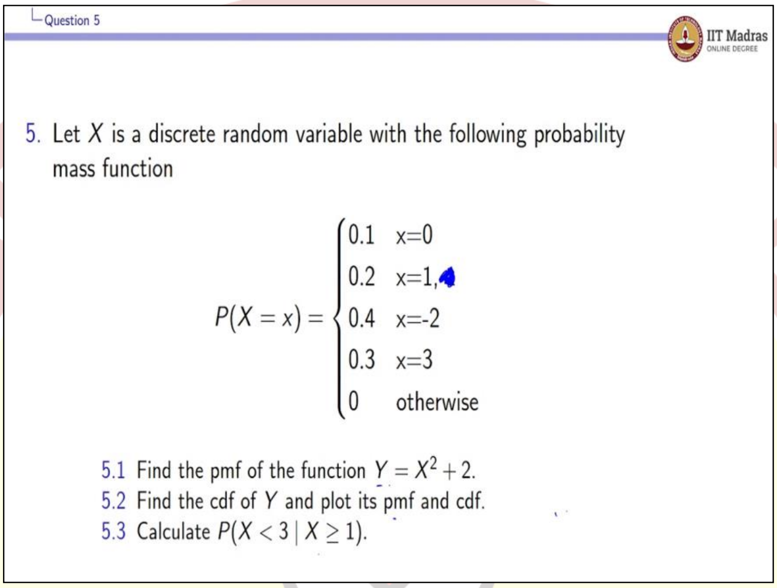
So $\boxed{Y}$ also would be a discrete random variable. This is by the assumption that each sand grains will be having same weight. But if you look at each sand grain of to be different weight then the sum of each sand grain it need not be a particular value, it could be anything. Since this weight as such since $\boxed{W_s}$ is a continuous here I assume $\boxed{W_s}$ to be 5 grams for a particular case. Each of the sand grains having of different weight then the total sum we cannot say to be this particular value.

Suppose the value of sand grains could be ranging from 0 to 5 grams then the total weight of the sand grains $\boxed{x}$number of sand grains, it will be anything between 0 to $\boxed{x}$ into 5, $\boxed{5x}$. It could be anything between $\boxed{0-5x}$. It could be 0.1 gram, 0.2 gram, 0.3 gram and so on. So if you take it in this way. If I assume that each sand grain will be of different weight then this $\boxed{Y}$ would be continuous random variable.

Statistics-1 for Data Science
Prof. Usha Mohan
Department of Management Studies
Indian Institute of Technology - Madras

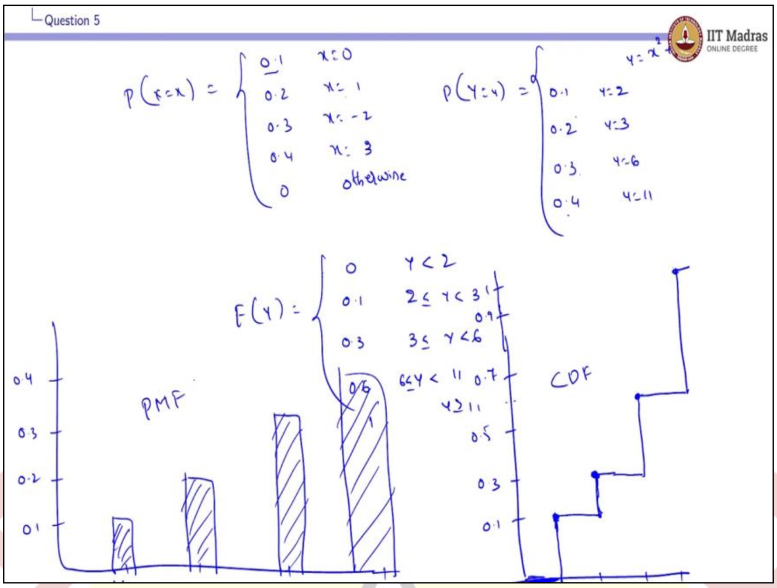
Week 8 - Tutorial 5
Statistics for Data Science-1

(Refer Slide Time: 00:15)



Let us look at the 5th question in the week 8. So they have given that X is a discrete random variable with the following probability mass function $P(X = x)$ for 0 is 0.1 for 1 is 0.2 for -2 it is 0.4 for x is equal to 3 is 0.3. And it is 0 for all the other values of X. They are asking us to find the pmf of the function $Y = X^2 + 2$. This Y is directly dependent on X and they asking us to find the series of 5 and plot its pmf and cdf. And they were asking for $P(X < 3)$ this is conditional probability question.

(Refer Slide Time: 00:55)



They have given that $P(X = x)$ is equal to 0.1 for $X = 0$ that is equals to 0.2 for X $X = 1$, 0.3 for $X = -2$, 0.4 for $X = 3$ and 0 otherwise. To verify that sum is equal to 1 we just add credit 0.1 + 0.2 is 0.3, 0.3 + 0.3 is 0.6, 0.6 + 0.4 is 1, so the sum is equal to 1, it is a probability mass function. So let us look at $P(y)$. We do not know that what this Y would be.

Let us calculate Y for each value of X there, $Y = X^2 + 2$ this is what given to us. So for $X = 0$ it is substituted there so this would be $0^2 + 2$ that would be 2. So for $X = 0$ the Y would be 2, so the probability is same 0.1 and for $X = 1$ the value of Y would be $1^2 + 2$ that is 3. So for $Y = 3$ we have value of 0.2. And for $X = -2$ it will be substituted there. So it would be $-2^2 + 2$ that is nothing but $2^2 + 2$ which is equal to 6.

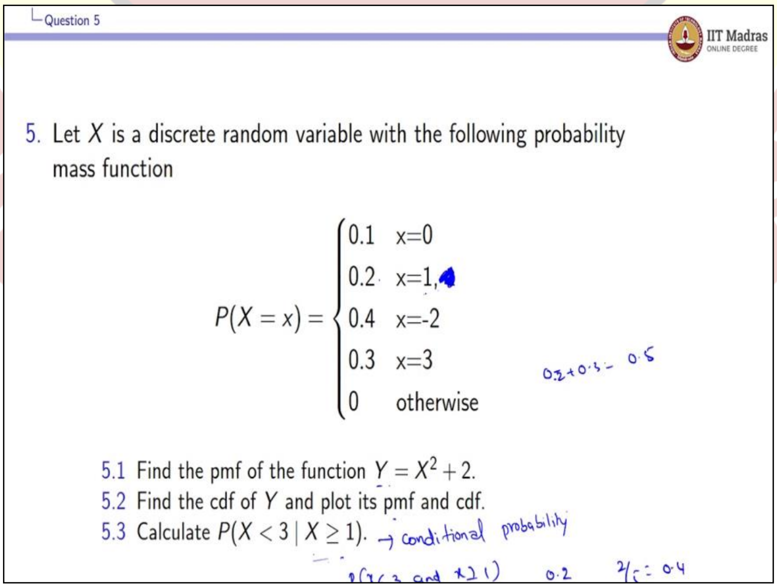
So that it would be 0.3. So for $Y = 6$ probability for that occurring is 0.3. For $X = 3$ we will get a Y of $3^2 + 2$ which is nothing but 9 + 2; 11 we have probability of 0.4. This is the probability mass function for Y. And if you look at the cumulative distribution function for Y, $F(Y)$ could be equals to; if it is 0 $Y < 2$, it could be equal to 0.1 $Y = 2$ until $Y < 3$.
So until $2 \leq Y < 3$ we have probability of 0.1. So for $Y = 3$ it would be 0.1 + 0.2 that is nothing but 0.3. For $3 \leq Y$ is until $Y = 6$. We will have probability of 0.3 cumulative probability of 0.3. So for Y is equals to 6 until Y is equals to 11 it would be 0.3 + 0.3 which is 0.6. So the value would be equals to 0.6 for; $6 \leq Y < 11$. And the latter one is 0.6 + 0.4 which is obviously equal to 1.

For $Y \geq 11$ so the cumulative probability will be equal to 1. For the pmf and cdf we plot it. The pmf would something of this and I am writing the value of Y there $Y = 2$ and this is 3, this is 6 this will be 11 this will not be exactly equal to scale. So values of this would be for $Y = 2$ it is 0.1, 0.2, 0.3, 0.4. So for Y is equals 2 values equal to 0.1. For $Y = 3$ the probability is 0.2.

So for Y =6 the probability is 0.3 and $Y = 11$ the probability is 0.4. And we need to plot the cdf. For that we take the same values 2, 3, 6 so these are the boundary conditions we will be taking this until it is less than to the probability would be 0. Cumulative probability is equal to 0 and once it is equates to 0.1, 0.3, 0.5, 0.7, 0.9 and somewhere around here 1. So this should be for Y equals to you will be having 0.1 probability; until it is equals to 3 you will be having a probability of 0.1 there.

So, exactly at 3 it would be having a jump. For exactly y equals to 3 the cumulative probability would be equals to 0.3. Here also if y is equals to 2 you will be having cumulative probability of 0.1 there and until it is equals to 6 the value will be 0.3. And once you reach 6 the value would be equal to 0.6 so it would be somewhere around here. And until it is equals to 11 it will be 0.6 and once it reaches 7 the value would be 1 and after that continuous one. This is the CDF graph. This is the probability mass function plot.

(Refer Slide Time: 07:38)



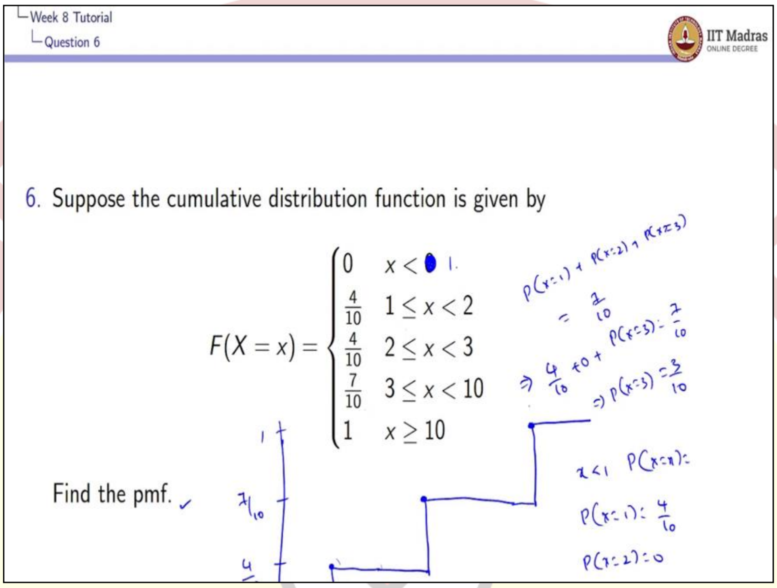
We have one last one that is we are asking for $P(X < 3|X \geq 1)$. So for X to be greater than or equals to 1 the probability would be 0.2 + 0.3 that would be equals to 0.5 this is the conditional probability for this value would be equals to $\frac{P(X<3 \text{ and } X\geq 1)}{P(X\geq 1)}$.

That is nothing but if you look at it $P(X \geq 1)$. We have only 1 case where X is $\frac{1}{2}$ so the numerator would be equal to 0.2 for it to be greater than or equal to 1 if you take two values 1 and 3. So the total probability would be found to be 0.2 + 0.3 is nothing but 0.5. So, this value would be equals to $\frac{2}{5}$ which is nothing but 0.4.

Statistics-1 for Data Science
Prof. Usha Mohan
Department of Management Studies
Indian Institute of Technology – Madras

Week 8 - Tutorial 6
Statistics for Data Science-1

(Refer Slide Time: 00:15)



Hello; again let us look at the final question in the week 8 tutorial. Suppose the cumulative distribution function is given by this function. It is 0 for X is less than 1 the values 4 by 10 for 1 is less than or equal to x is less than 2, 4 by 10 again 2 is less than or equal to x is less than 3. 7 by 10 is less than or equal to x is less than 10 and it is equal to 1 for X is less than or equal to 10. They are asking for the pmf probability mass function.

So if you plot the cumulative distribution function the boundary values here are 0 and 1 x is equal to 1, x is equals to 2, x is equals to 3 so on and x is equals to 10. These are the boundary condition. So until x is equals to 1 it is 0 at x is equal to 1 will be having 4/ 10 so let this be 4/10, 7/ 10 then here it will be 1. At x = 1 you will be having 4/ 10 value so there is a jump there and until it is equals to 2 it will be having 4/10.

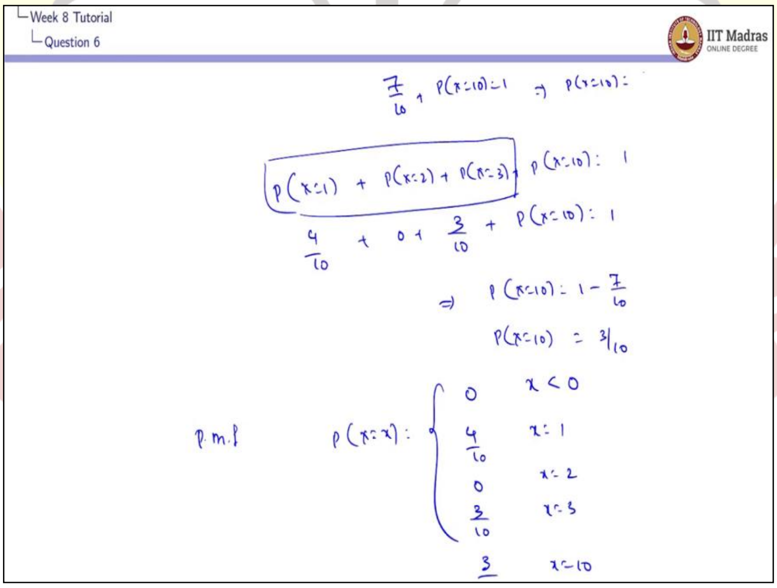
And from 2 to 3 also you will have the value of 4/10 so and that x=3 will have cumulative probability of 7/10 and until it is equal to 10 the value would be 7/10. And once it is equal to 10 or greater than it is equals to 1. So this is the cumulative distribution function. For exactly at x = 1 will be having probability of 4/10 and before that it is 0. So for x < 1 the probability of x is equal to 0.

Since until then until x equals to 1, we do not have any cumulative property from and exactly at 1 you will be having cumulative probability of 4/10 that means that the probability of x equal to 1 is 4/10 and until it is equals to 2 and until it is equal to 3 we do not have any probability there. So, probability of x is equals to 0 and if you look at it from this to here exactly at x = 3 you will be having cumulative probability of 7/10.

That means that $P(X=1)+ P(X=2)+ P(X=3)= 7/10$. So for $P(X<1)=0$, $P(X=1)=4/10$ as you have seen here there is a sudden spike there. And for x is equal to 0 you have probability of 0 since there is no spike there. There is no spike in a graph step function there.

And at x is equal to 3 calculating right now $P(X=1)+ P(X=2)+ P(X=3)=7/10$. For cumulative probability of all until there it is 7/10. So this implies that $P(X=1)$, $4/10 + 0 + P(X=3)=7/10$, implies $P(X=3)$ is $7/10 - 4/10 = 3/10$ after this from 3 to 10, you would not be having any spike and exactly at 10 we are having the spike and that probability is equals to 1.

(Refer Slide Time: 04:33)



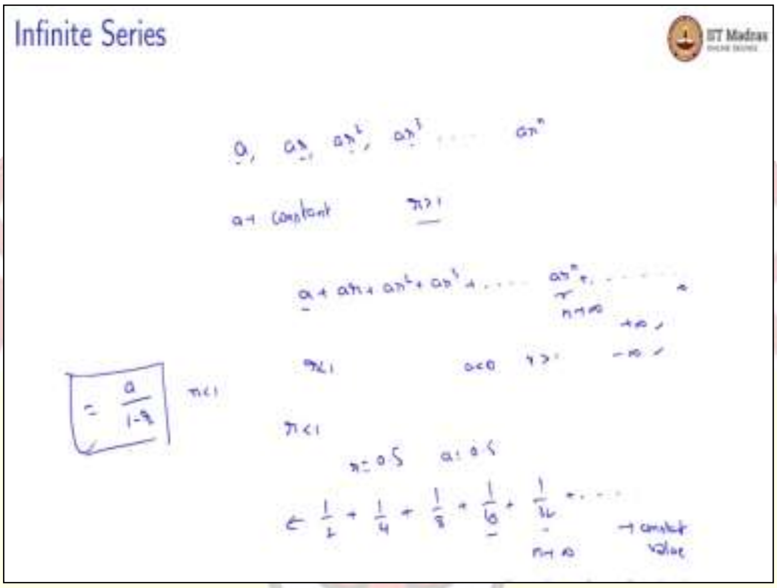
So $P(X=1)+P(X=2)+P(X=3)+P(X=10)=1$, since the cumulative probability for x is greater than or equal to 10 is equals to 1 and so this is equals to 4/10 this is equal to 0 this is equal to 3/10 and this value we do not know we are yet to find out this is equals to 1. This implies $P(X=10)=1-(4/10+3/10)$ it is nothing but 7/10.

This would be equal to 3/10. So that is pmf would be $P(X=x)=0$ for $x<1$, 4/10 for x is equals to 1, 0 for $x=2$, 3/10 for $x=3$ and again 3/10 for $x=10$ and it is equal to 0 for all other values of X. This is the probability mass function of the given cumulative distribution function. And you can solve in different way. We already know that this value is equal to 7/10 from this is. So if you substitute this is nothing but $7/10+P(X=10)=1$ implies $P(X=10)=1-7/10$. You can form it directly there.

Mathematics for Data Science
Prof. Usha Mohan
Department of Management Studies
Indian Institute of Technology - Madras

Infinite Series Tutorial

(Refer Slide Time: 00:15)



Hello, everyone this is Ram from Statistics course support team. In this tutorial we will be learning about infinite series that will be useful for solving some of the problems in week 8. There are 2 types of infinite series just give a brief introduction of infinite series there be 2 types of infinite series one that will be converging where the sum of the Infinite series will be converging to some number some real number and where it will be diverging or like where the sum of the Infinite series would not be converging to a particular number.

It will be leading to a plus infinity or minus infinity. So the first type of infinite series which we were going to look at is geometric progression. I hope you are all aware of the geometric progression. Geometric progression is a series of type $a, ar, ar^2, ar^3 \ldots\ldots ar^n$. So if you look at this, let a be some constant and if you look at the value of r if it is greater than 1.

Then this series will never be converging, in a sense like look at the value of $a + ar + ar^2 + ar^3 + \cdots$. This value will be increasing as long as the n increases number of terms increases. As long as the number of terms is increasing; and since $r > 1$ the value will be becoming bigger and bigger. If you see $a > 1$ and $r > 1$ then it will be leading to $+\infty$ if $a < 0$ and $r > 1$ then it will be leading to $-\infty$.

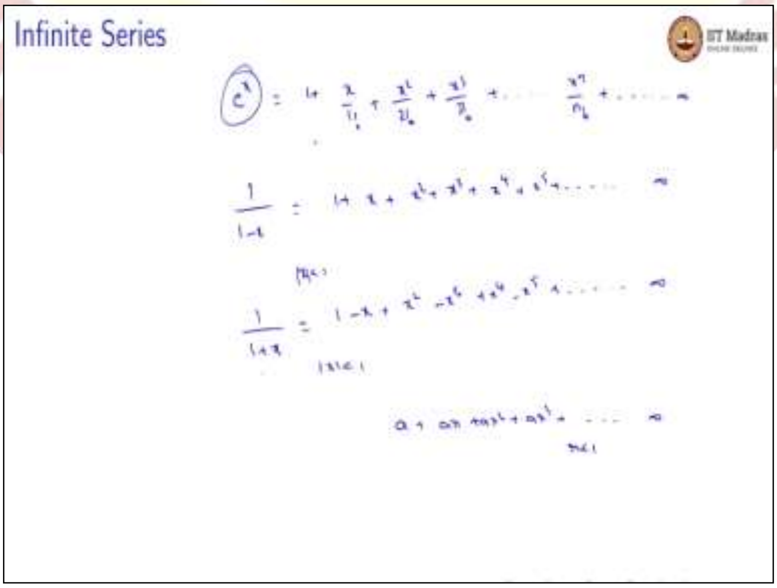
This series and r value is greater than 1. It will be it would not be converging. It will be diverging either to $+\infty$ or $-\infty$ as n tends to infinity as a number of terms over all here we have $n+1$ terms as n tends to infinity. That means that we have like infinite series infinite number of terms. So the whole sum of this expression it will be diverging. But if you take the case of r < 1 suppose r = 0.5 and a = 0.5 you will be having like the first value would be 0.5 = ½ .

Since r = 0.5 ar is nothing but $\frac{1}{2} * \frac{1}{2} = \frac{1}{4}$ and we have $\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \frac{1}{32} + \cdots$. So we can see that this 1/16 this 1/32 this value is reducing. Ultimately we can say that the sum of this value it is approaching to a value. As a number of terms n tends to infinite. So the formula for this infinite or the value of this expression $a + ar + ar^2 + ar^3 + \cdots + ar^n + \cdots$.

As r < 1, if r < 1 than this value will be equals to $\frac{a}{1-r}$. So, some of these terms would be equals to $\frac{a}{1-r}$ given that r < 1 this is the first type of infinite series $\frac{a}{1-r}$. So whenever the series like $a, ar, ar^2, ar^3 \ldots\ldots ar^n$ and if r < 1 and if the number of terms that like they are tending to infinity I just mentioned here ar^n.

If I mention like and so on infinite then the sum of this expression would be equals to $\frac{a}{1-r}$. It will be converging into this particular value $\frac{a}{1-r}$.

(Refer Slide Time: 04:54)



So the next typo of expansion is e^x. $e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots \frac{x^n}{n!} + \cdots$. So if you come across the expression, like $1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots \frac{x^n}{n!} + \cdots$ the sum of that expression would be equals to e^x.

This is one of the infinite series and other infinite series would be the expansion of $\frac{1}{1-x} = 1 + x + x^2 + x^3 + x^4 + \cdots \ldots \ldots . + \infty,\ |x| < 1$. And if you look at the expansion of $\frac{1}{1+x} = 1 - x + x^2 - x^3 + x^4 - x^5 \ldots \ldots \ldots . \infty$.

This is the expansion of $\frac{1}{1+x}$. Here also the same condition $|x| < 1$. So these are the 3 types of infinite series, sorry 4 types of infinite series which were going to look at. One would be $a + ar + ar^2 + ar^3 + \cdots .. + \cdots .. \infty, r < 1$. The other expansion of e^x which could be expressed as the sum of this infinite terms. The other is $\frac{1}{1-x}$ expansion and the last one is $\frac{1}{1+x}$ expansion. These are the infinite series which we will be useful for you to solve the problems. Thank you.

THIS BOOK IS NOT FOR SALE NOR COMMERCIAL USE

Scan QR code to chat with our WhatsApp bot

+91 9840776800

7850-999966
(Mon-Fri 9am-6pm)

Online Degree Office
3rd Floor, ICSR Building, IIT Madras, Chennai - 600036

onlinedegree.iitm.ac.in | support@onlinedegree.iitm.ac.in