Essential of Data Science with R Software-1
Probability and Statistical Inference

Professor Shalabh
Department of Mathematics and Statistics
Indian Institute of Technology Kanpur

Lecture No. 10

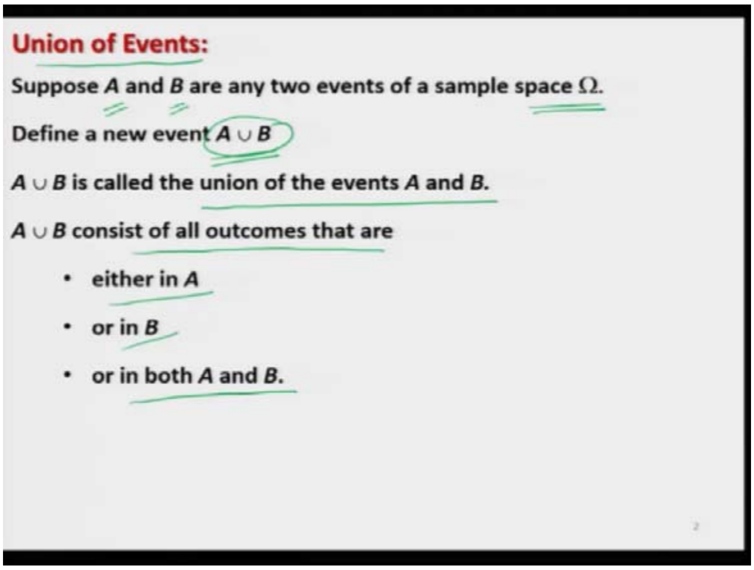
Set Theory and Events Using Venn Diagrams

Hello friends, welcome to the course Essential of data science with R software in which we are going to understand the topics of probability theory and statistical inference. Now you can recall the last lecture we had understood the basic elements of probability theory. That how to define an experiment event et cetera. Once you have defined the events then we are also interested in different combinations of those events. Like I have if this event and this event occur together or only an intersection of those events occur together and so on. And this type of feature is very common when you are trying to use the data science.

So, my objective in this lecture is simply to introduce you with these type of events where we are trying to take a combination of the events in different ways. Now when I am trying to do, it then obviously means at the end we are going to find out the probability of such an event. But before that we have to understand that how we can define such events. And when we are trying to define such event the Venn diagrams that we have studied in mathematics they help's us a lot. So, it is possible that when you trying to deal with the complicated problem that many times this Venn diagrams and set theory day helps us. So, in this lecture I will aim that I define such an event.

And I try to represent those events in the format of Venn diagram using the concepts of set theory. I believe that is the theory you have done in classist then on class 10 also. So, I am just going to use their and I will try to define different types of events. And after that definitely our basic objective is that is how to compute the probabilities of such an event. And then how to compute them in the R software, that will be our final aim. So but at this moment, so that let us try to prepare the foundations first. So, let us begin our lecture. And I belief that you have understood the meaning of an experiment or a simple event.

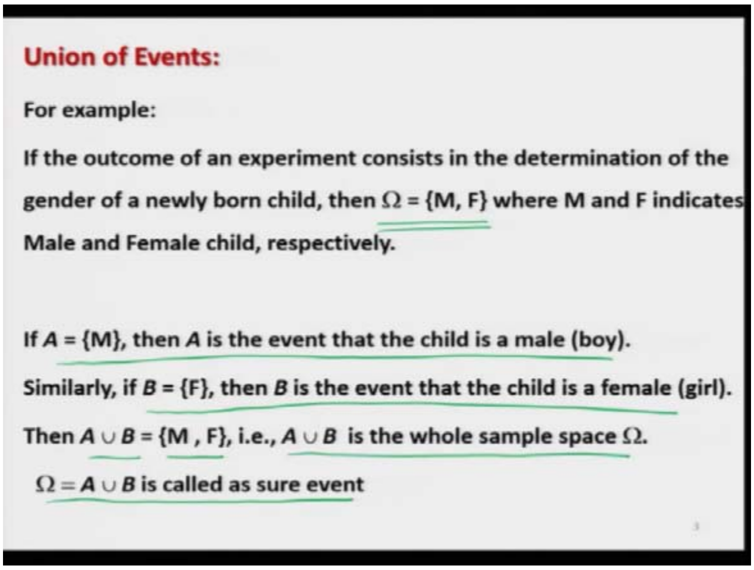
(Refer Slide Time: 2:37)



So, now we try to define first an event in the format of union of events. What is this mean? Suppose there are two events which I am trying to indicate here by capital A and capital B and obviously these two events are from the sample space, Ω. Now I define a new event which is A∪B and the meaning of A∪B is that A∪B consist of all the outcomes that are either in A, or in B, or in both A and B.

So, now onwards we will call this symbol A∪B as the union of events A and B. So, this you can see that this is trying to define an event which is based on the combinations of the outcomes of events A and B.

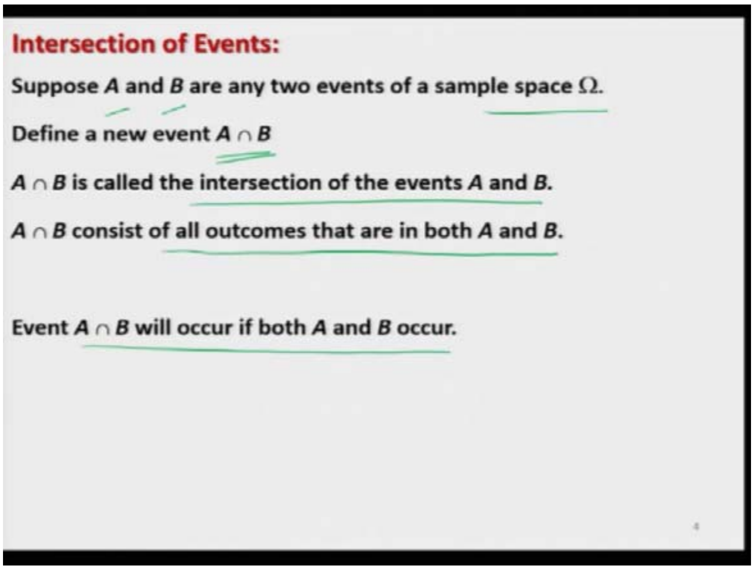
(Refer Slide Time: 3:28)



For example, in case if I take a very simple example. Suppose an experiment is conducted on the simple experiments that is baby is born in a hospital and we try to determine the gender of the baby. That can be either male or female. So, now the sample space will consist of two events capital M and capital F where capital M is indicating the male child and capital F is indicating the female child. And now let us try to define here the event A to B that the child is a male, that is a boy.

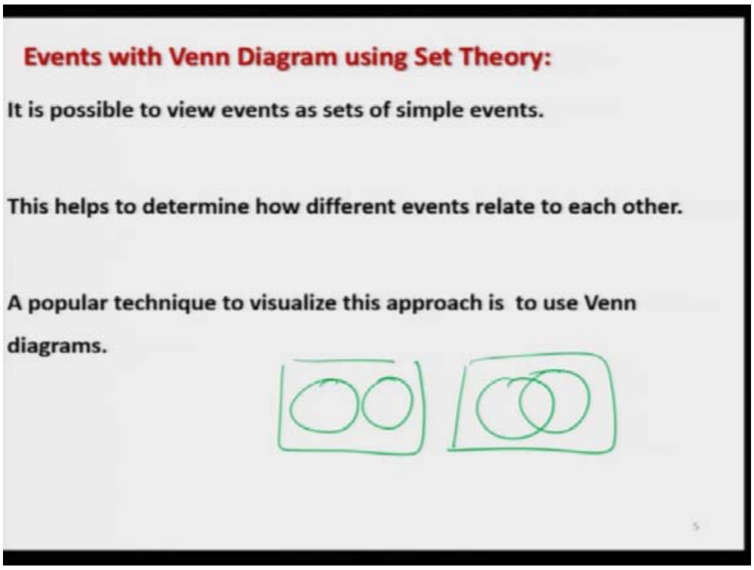
And similarly, we define another event B which is F that is B is the event that the child is female girl. Now since they are only two simple events here A and B. So, this A∪B will be the collection of M and F. So, in this case you will see that this A∪B is the whole sample space omega. And we had learnt in the earlier lecture that this event omega is equal to A∪B in this case is called as sure event. That definitely the child will be either male or female.

(Refer Slide Time: 4:45)



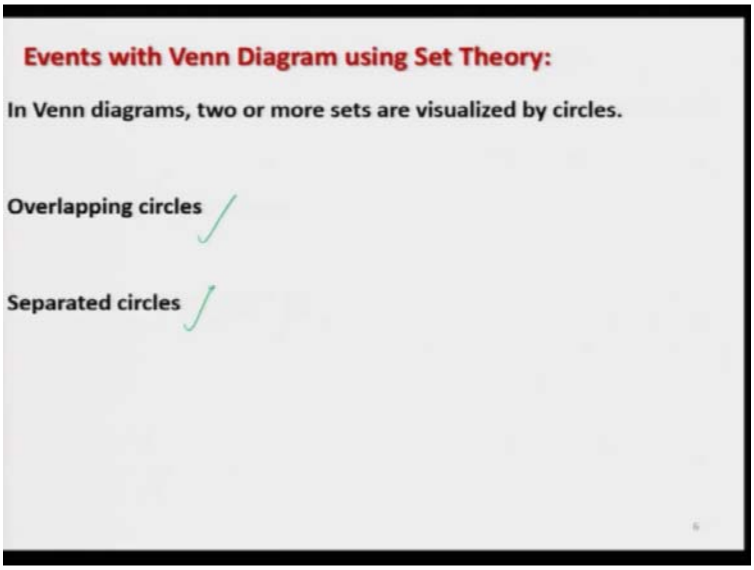
And now similarly we can also define the intersection of event. So, suppose A and B are the two events which belongs to the sample space Ω. So, if we define here our new event A∩B. So, this is called the intersection of the events A and B. This A∩B consist of all such outcomes that are in both A and B. So, obviously the intersection A and B will occur only if both A and B occur. So, this is how we can define the A∪B and A∩B.

(Refer Slide Time: 5:23)



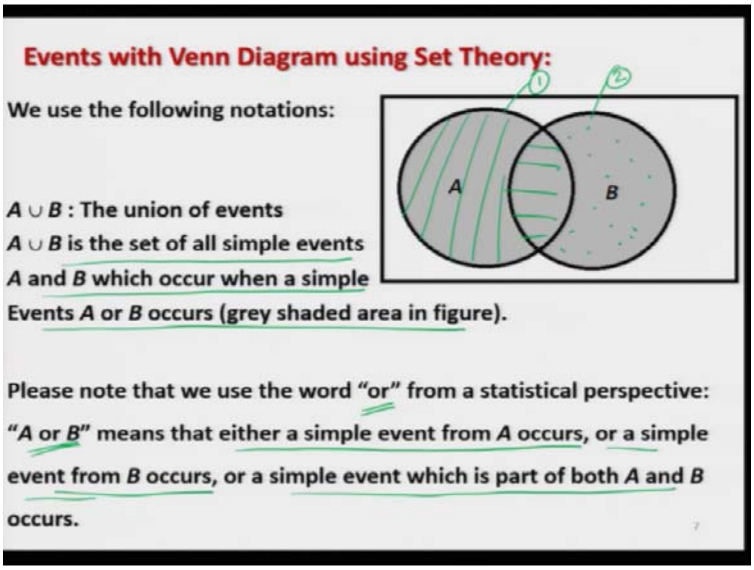
And now we will try to view these events using the Venn diagram from the set theory. So, you know that when we are trying to create the Venn Diagrams we have this types of circles. This merge this merge or this and this like this. So, this Venn Diagram will help us in visualizing the outcomes of different events A and B and in particularly when they are trying to occur together.

(Refer Slide Time: 6:02)



So, now using this Venn Diagram first will let me try to define A∪B and A∩B and after that we will define different types of events. So, what we are going to do that we are going to observer here two types of Venn Diagram in which there will be some overlapping circles and in some cases there will be some separated circles.

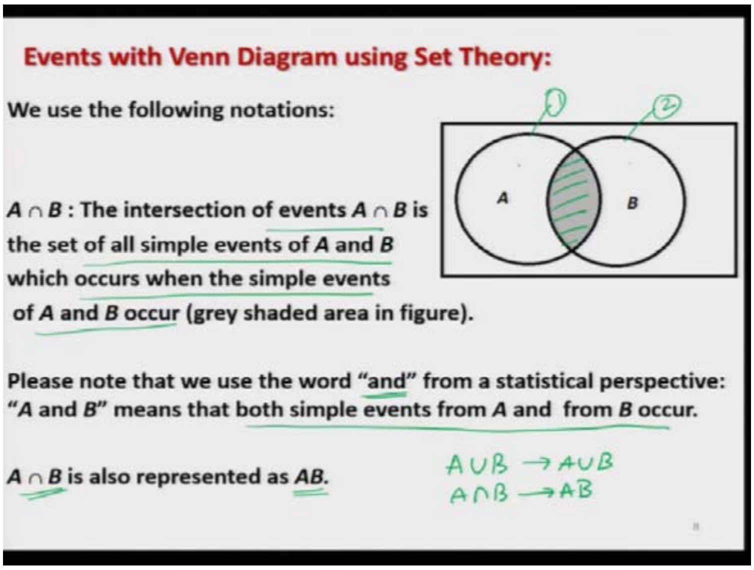
(Refer Slide Time: 6:19)



So, first we try to understand the Venn diagram for A∪B. So, suppose we have two events here A and B and these events are indicated by these two circles. So, circle number 1 is indicating the event A and circle number 2 is indicating the event B. So, now, A∪B is the set of all such events A and B which occur when a simple event A or B occurs. A or B occurs means, if A occurs, if B occurs like this and if both A and B occurs.

So, this entire grey shaded area that is going to indicate the area that is covered by the event A∪B. And you have to always keep in mind that when we are trying to use the word "or", then from the statistical point of view. A or B means events A or event B this means that either a simple event from A occurs, or a simple event from B occurs, or simple event which is a part of both A and B occurs. So, this is how we try to interpret it.

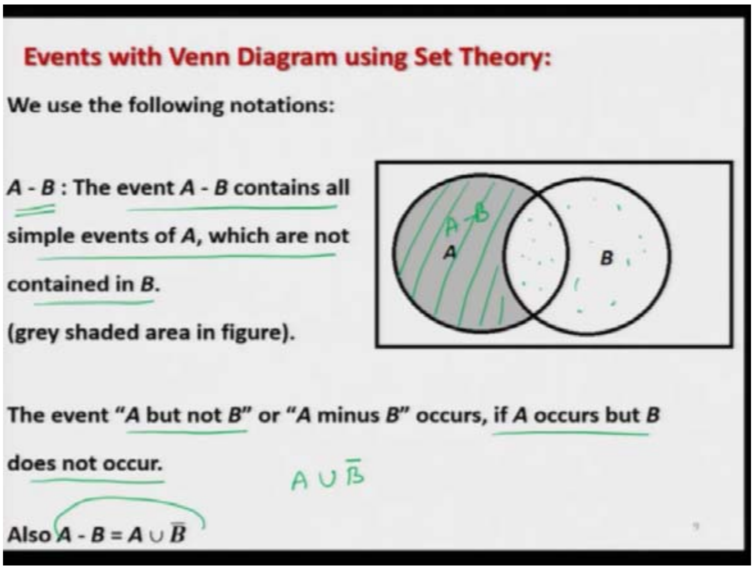
(Refer Slide Time: 7:31)



And similarly, when we try to look at the A∩B so once again I will use the circle number 1 to indicate the event A and circle number 2 to indicate the event B, then A∩B is the set of all simple events of A and B which occurs when the simple events of A and B occur together. So, if you can see here A has occurring here, B is occurring here, and this the part where A and B are occurring together. So, from the statistical point of view when we use the word "and" that means event A and event B occur that means both simple events from A and from B occur together.

And you will see that when we are trying to use this symbol A∩B. So, in many books it is also written as AB. A∪B will be written always as A∪B. But A∩B at many places you will find that this is written as AB. So, that is what you have to understand that is the simple terminology or symbolic notations when we are trying to deal with probability theory.

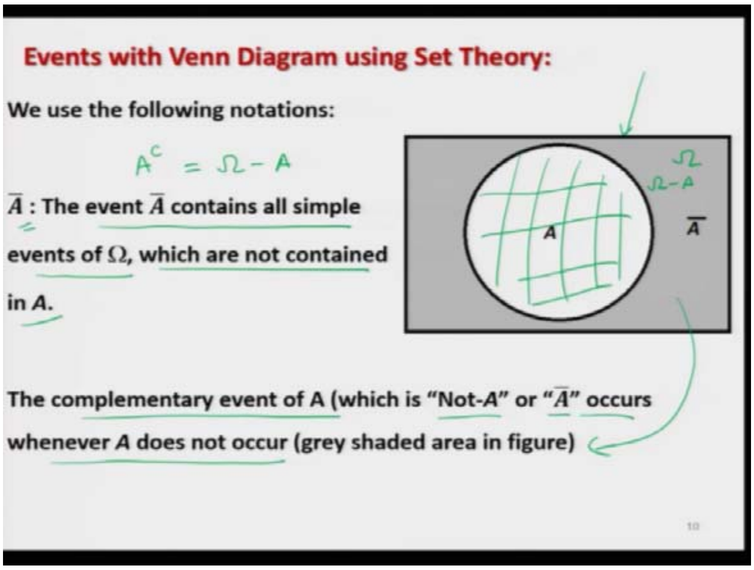
(Refer Slide Time: 8:39)



Now similarly we have defined some other events for example just like you have use the events A and B to defined their joint occurrence together or separately. So, similarly we can define some other type of events. For example, we know from the set theory when we are write A - B that means all the events of A which are not in B. So similarly I can translate this definition into the event and I can define this A - B as an event that contain all simple events of A, which are not contained in B.

You can see here this is the area grey area which is containing all the events of A, which are not containing in the grey. So, this shaded area is going to indicate the event A - B. So, this event A - B or A but not B if it occurs, if A occurs but B does not occur. And if you try to represent this phenomenon using the set theory, then it can be represented by like this A - B is equal to A∪B compliment. Well, that is clear from the Venn diagram also. And I am sure that you have done this thing in your earlier classes also.

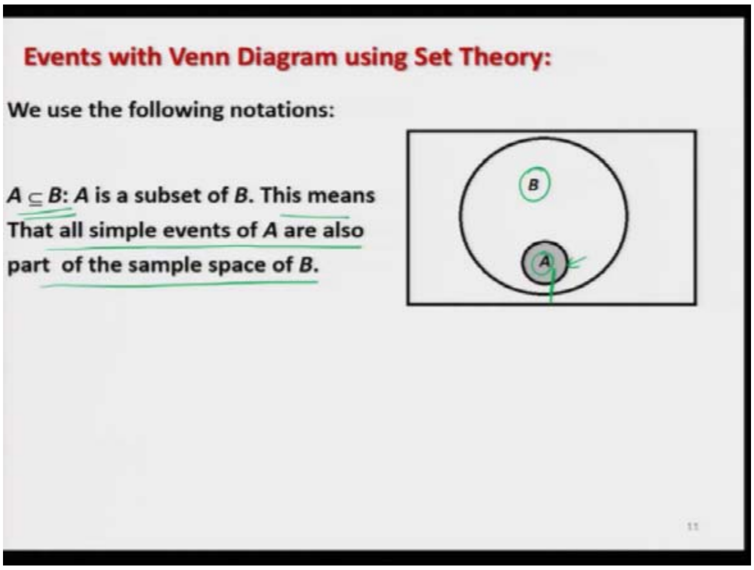
(Refer Slide Time: 9:56)



And similarly, I try to define here another type of event which is A compliment. A compliment mean whatever is occurring in A except that everything occurs. So, this A compliment event we are trying to denote by A bar. Yes, sometime you will see that I some books it is also defined at A^c, the super script. Of course this event may compliment contains all simple event of Ω which are not contained in A.

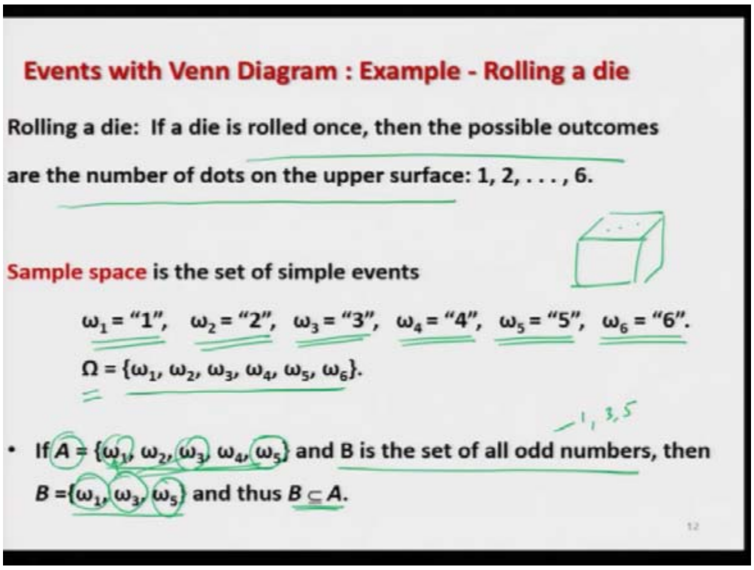
So, in case if you try to represent this phenomenon using the Venn diagram, you can see here this is your Ω. Now you are trying to say this part here is A. So, if you try to ignore this part A then whatever is remaining this is Ω - A, and Ω – A is your simply your A compliment. So, A compliment is actually Ω - A. So, this means the complimentary event of A which is not A or A compliment occurs whenever A does not occur. And this entire grey area that is indicating all the events which are covered under the category of A compliment.

(Refer Slide Time: 11:10)



Similarly, we try to define here one more event A⊆B and which is indicated by this ⊆. That is standard symbol. So, the meaning of A⊆B is that all simple events of A are also part of the sample space B. So, you can see here that in this case this white circle this is the event B and this grey circle this is the event A. So, this area that is define here that is the part where the simple event of A this is also a part of B. So, this is how you can see that different types of events are occurring and you can correlate this events with what is really happening in the real word.

(Refer Slide Time: 11:55)



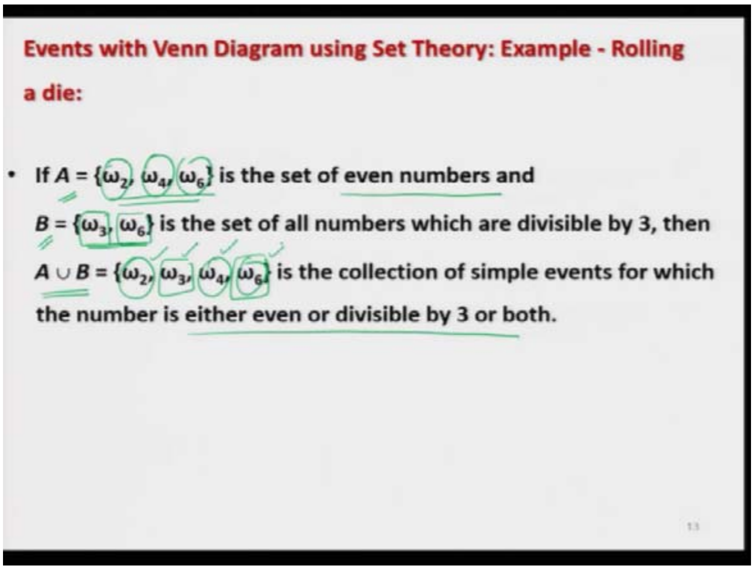
And in order to give you an idea, let me try to take a very simple example to explain you that how this events can be defined. So, let me take a very popular example of rolling a die. So, we know that if our die is roll once then the possible outcomes are going to be the numbers of dots on the upper surface of the dice like this one. This is your here dice then whatever is occurring here. Suppose 3 occurring that mean this is our outcome.

And there are six such possibilities of getting numbers 1, 2, 3, 4, 5, and 6. So now when we try to define here the sample space. So, sample space will consist of all the simple events. And we have indicated that the simple events are going to be indicated by the symbol ω. So, we try to take here ω_1 is equal to 1 that means number 1 occurs on the upper surface. ω_2 is the number 2 occurs, ω_3 is the number 3 occurs, ω_4 is the number 4 occurs, ω_5 means the number 5 occurs and ω_6 is the number 6 occurs. So, this omega Ω the sample space will consist of 6 points ω_1, ω_2, ω_3, ω_4, ω_5, and ω_6.

Now in case if you try to define here an event. Which consist of five numbers 1, 2, 3, 4, 5 then this event A can be defined as the collection of points ω_1, ω_2, ω_3, ω_4, and ω_5 and we defined one more event here B which is the set of all odd numbers. So, you can see from this Ω the odd numbers are occurring with 1, 3, and 5. So, that means B is going to contained the simple events ω_1, ω_3, and ω_5. And you can see here that this elements ω_1, ω_3, and ω_5 they are here.

So, obviously B is a subset of A. So, this is how you define such an event. And in case if you try to think about a real life scenario suppose a shopping website is being excess by both male and females and they have the entire observations and they want to study that which of the clothing's are going to be sold more for male or for female. So, you have to collect all the data and then you have to take out the subset of the females or male from the entire dataset. So, in those types of events, this type of events and their probability are going to help you.

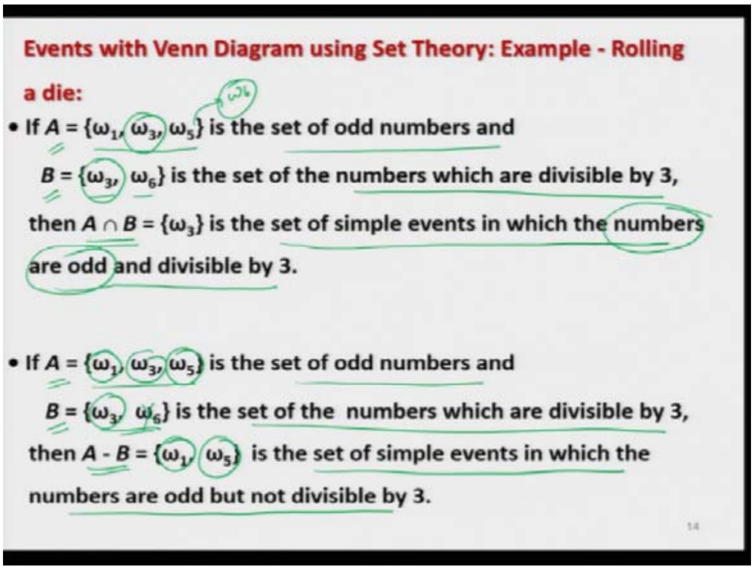
(Refer Slide Time: 14:32)



Now I try to take to one more example in which I try to define the event A as consisting of three numbers 2, 4, and 6 which are actually even numbers out of this Ω. So, we have a three simple events ω_2, ω_4 and ω_6 in the event A. And they can be try to define here one more event B which consist of all the numbers which are divisible by 3. So, there are 2 numbers 3 and 6 which are divisible by 3. So, we try to create another event here B which is the collection of ω and ω_6.

Now in case if you try to define here A∪B that means all those events which are in both A and B that mean this will consist of ω_2, ω_4, ω_6 this is here. ω_2, ω_4, ω_6, and then you have here ω_3 and ω_6 from event B. So, this ω_3 is here and ω_6 is here which is already here. So, this A∪B will consist of all the points or the simple event which are either even or divisible by 3 or both. So, this will consist of four numbers ω_2 is equal to 2, ω_3 is equal to 3, ω_4 is equal to 4, and ω_6 is equal to 6. So, this is how you can define here the union of the event.

(Refer Slide Time: 15:52)



Now I take one more example and I try to create an event here capital A which is the set of all the odd numbers 1, 3, and 5 out of Ω. So, this will be having 3 sample points ω_1, ω_3 and ω_5 and we try to define here the same event B which consist of those number which are divisible by 3. So, there are two number 3 and 6 which are divisible by 3. So this B is the collection of two points ω_3 and ω_6.

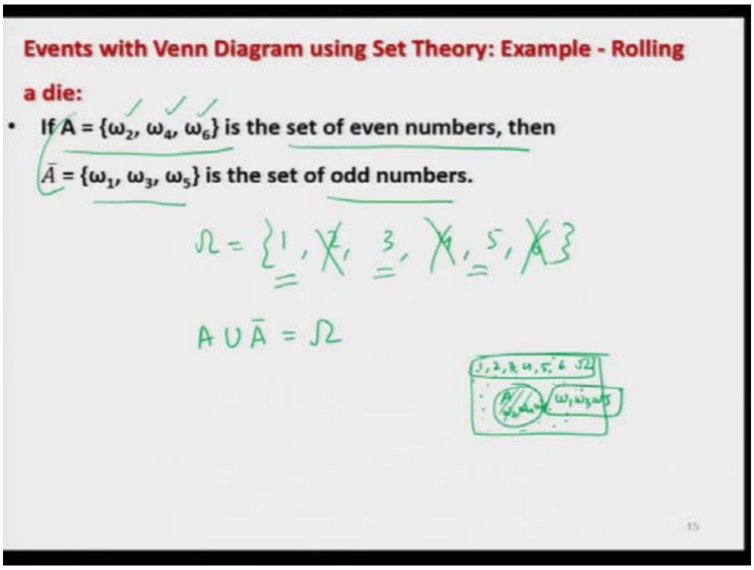
Now I try to define here the event A∩B. So, A∩B is going to contain the common points between A and B. So, you can see here that ω_3 and ω_3 are the two point which are common in both the A and B. So, this A∩B event is the set of the simple event in which the numbers are odd and they are divisible by 3. Well, in case if you try to add here one more number here suppose ω_6 then what will happen then this A∩B will consist of two points ω_3 and ω_6.

But still the ω_6 will not happen here because you have a condition here that the number have to be odd. That is what you have to keep in mind that when you are trying to define different types of combination you have to check and verify whether these things are happening or not. So, now in case if you try to take one more example to define A - B.

So, we consider here an event A which consist of ω_1, ω_3 and ω_5 which are the set of all the odd numbers as earlier and we try to define here this is the same event here B consisting of two point 3 and 6 which is the set of number which are divisible by 3. Then A - B is going to be define by the collection of those point which are in A but not in B. If you try to see here ω_1 which is not in B. So it comes over here. Have you try to see here the omega 3 is common to both. So, ω_3 will not appear in A - B.

And similarly if you try to see here the ω_5, it is also not here in this capital B. So ω_5 will appear in the event A - B. And we are not bothered about ω_6. So, this A - B event will consist of two points ω_1 and ω_5 which is essentially the set of simple events in which the numbers are odd. But not divisible by 3. So, you can see you have obtained what exactly you wanted depending on your objective.

(Refer Slide Time: 18:30)

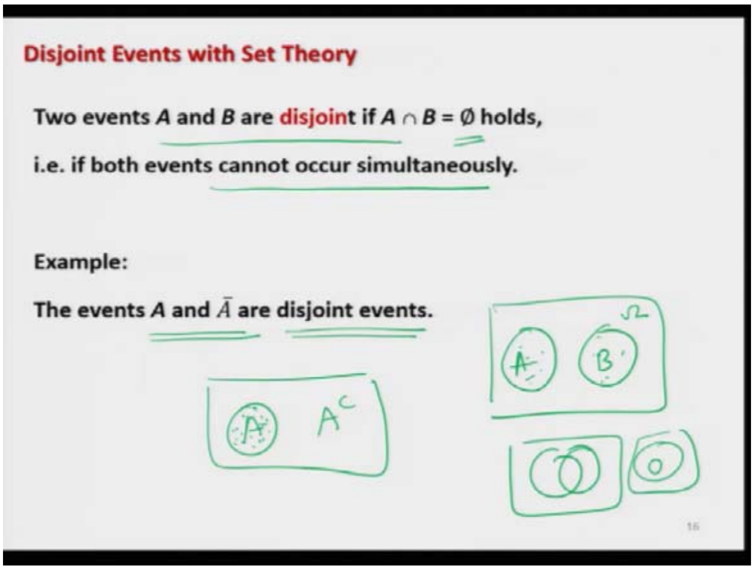


Now similarly, I try to define a complimentary event. So, if I try to define here and event A which is the set of even numbers. So, there are three numbers 2, 4, and 6 in Ω which are divisible by 2. The other even numbers. So, whatever numbers which are in A. There will not occur in A bar. So, omega is 1, 2, 3, 4, 5, and here 6. And now we try to remove this points A that is 2, 4, and here 6. So, now what here left? This is here the numbers 1, 3, and 5.

So, this A compliment is going to have all the 3 numbers ω_1, ω_3, and ω_5. Which are essentially the set of odd numbers. And obviously if you try to see A ∪ A compliment this is going to be Ω. And that you can see here from these two events also. And even if you try to see from this Venn diagram also this is your here Ω consist of the point 1, 2, 3, 4, 5, 6 which is like this for this omega and now you are trying to have here circle indicating the event A which is consisting of ω_2, ω_4, and ω_6.

So, now this A compliment is consisting of this part here which is the doted part here which is here ω_1, ω_3, and ω_5. Now if you try to combine this ω_1, ω_3, ω_5. and this ω_2, ω_4, ω_6. that is going to give you the Ω that is the point 1, 2, 3, 4, 5, and 6. So, whatever you have done here you can verify them through the Venn diagram also.

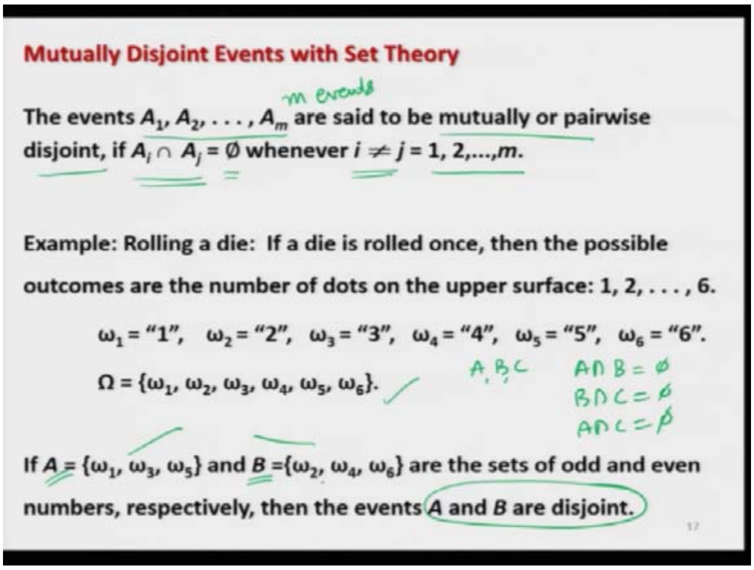
(Refer Slide Time: 20:20)



Now I try to define one more property of this events two events. Two events A and B are said to be disjoint if A∩B is equal to ϕ. That mean there is no element which is common between A and B. And if you try to create the Venn diagram of such an event this will look like this. This is your Ω and so you can see here there are no points which are common between this A and B. And if they are common, the possibly the Venn diagram will look like this. Or that may be something like this also.

So, the interpretation of disjoint event is that both the events A and B. They cannot occur simultaneously. So, you can now correlate this statement with the happening of various types of event you want to compute the probability. And in case if you try to see here from your event point of view; hen the event A and A compliment both are disjoint event. For example, if you want to see in through the Venn diagram this is your here A and this part here is your A compliment. So, if I try to make this A to be here by doted then this A compliment is the remaining part. And you can see here both of them cannot occur together.

(Refer Slide Time: 21:45)



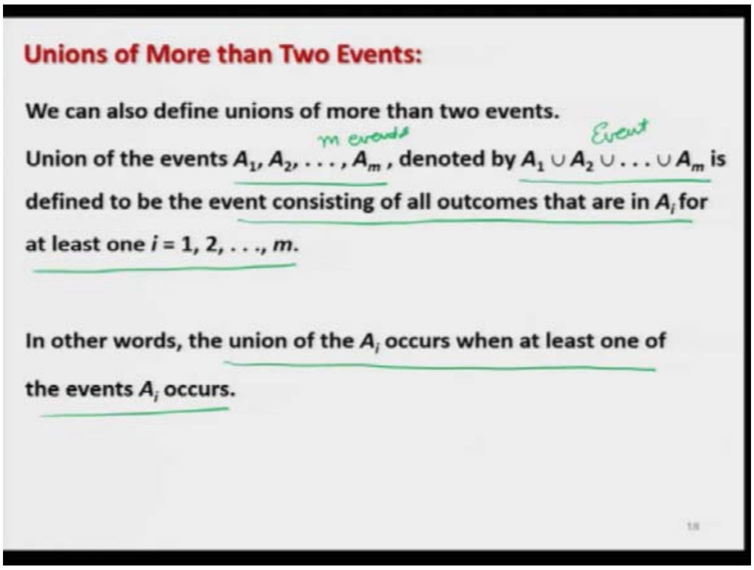
Now whatever I have told you about this events I have explain you by taking the example of only two events. That is for the sake of simplicity so that you can understand it easily. But these events can be defined for more than two events also. These definitions whatever you have done either for union, intersection etc. They can be extended for more than two events also. So, let me try to give you a fair idea about these things.

Suppose if I say I have here more than two events which are indicated by $A_1, A_2, \ldots, A_m$. So, now we have here m events. Now these events are said to be mutually or pairwise disjoint if $A_i \cap A_j$ is equal to ϕ. And in this case, yeah ,whenever i is not equal to j and it goes from 1 to m. And let me try to show this definition by this example also for example the same example in which the die is roll and we have here six points in the sample space.

If I try to take here two events here A and B like as A is the collection of all those points which are odd on and B is the collection of all those points which are even in numbers, then these this both the events A and B both are disjoint. Because either A can occur and when A is occurring then no number which is even can come. And similarly, if B is occurring that means only the even are coming to my space. So, this odd numbers cannot come. So, this is how both these events A and B they are going to be disjoint.

And what I am trying to say here that if you want to know about that more than two events are going to be disjoint or not. Suppose if I say that there are three events A, B and C. And if I want to know whether they are disjoint or not so we have to check their property pairwise. I have to check for A and B, B and C, and A and C. And if all of them are equal to 5 then I can say that all the events A and B and C they are mutually disjoint.

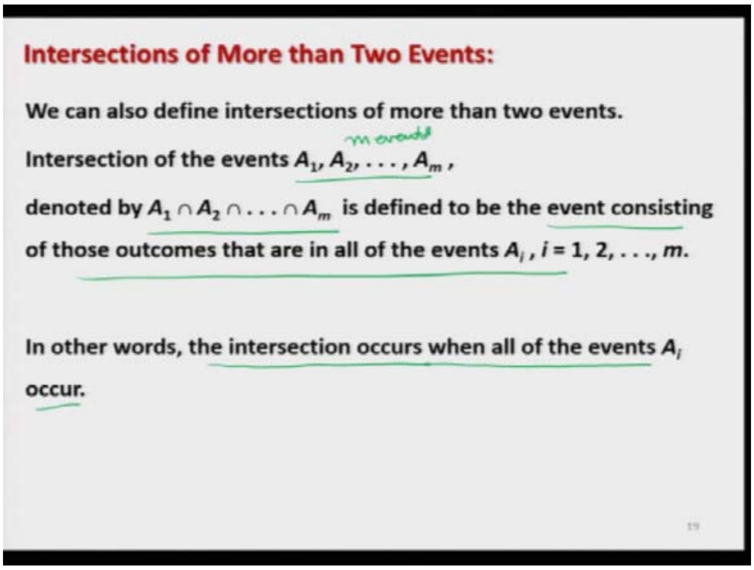
(Refer Slide Time: 23:57)



So, now similarly, we can extent the definition of unions means earlier we have understood what is the meaning of A∪B. So, similarly instead of having only two events suppose I have here more than two events say m events. And they are indicated by A1, A2, …, Am. So, the union of A1, A2, …, Am it is going to be indicated by $A_1 \cup A_2 \cup \ldots \cup A_m$. And this is also and event. And this event is define to be the event consisting of all the outcome that are in A_i for at least one i, i goes from 1 to m.

Or in very simple words I can say that the union of the A_i occurs when occurs at least one of the events A_i occurs that is as simple as that. And this is simply an extension of the concept that you have learnt in the case of A∪B.

(Refer Slide Time: 24:50)



Now as we have define the union of more than two events similarly I can also define the intersection of more than two events. So, suppose if I have here m events say $A_1, A_2, \ldots, A_m$ then the intersection of this event is going to be defined as here $A_1 \cap A_2 \cap \ldots A_m$. And this event is defined to be the event consisting of those outcome that are in all events $A_1, A_2, \ldots, A_m$. And in very simple words I can say that the intersection of these events occurs when all the events Ai occurs.

So, this is how you can define here the intersection of more than two events also. So, now we come to an end to this lecture. And you see that the, this lecture was a very simple lecture and but it was trying to correlate what you have done earlier in your set theory and in Venn diagrams and through this thing I am trying to interconnect the probability theory. Whenever you are trying to compute any probability this is obvious that you need to know that what is the event.

What is your objective? What you really want to find? And based on that you have to define an event in a proper way. And then you have to see how this event is going to occur. There can be different ways of knowing it. Either you can compute the number of events, the number of possible events which are going to happen or not happen. And exactly on the same line this Venn diagram also help you in counting of those numbers which are favorable to us.

And it will be your choice that which of the methodology you are going to opt when you are trying to implement these concepts in real life. So, now I would request you. Why do not you take examples from your book, from your set theory and Venn diagram. And try to just practice and have an idea of what is the meaning of this union intersection compliment minus etc. Once you practice it for simple events like as 2 events or 3 events, it will not be difficult for you to extent this definition for a general case. So, you try to practice it and I will see you in the next lecture with more details on probability theory till then goodbye.