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. 09
Sample Space and Events

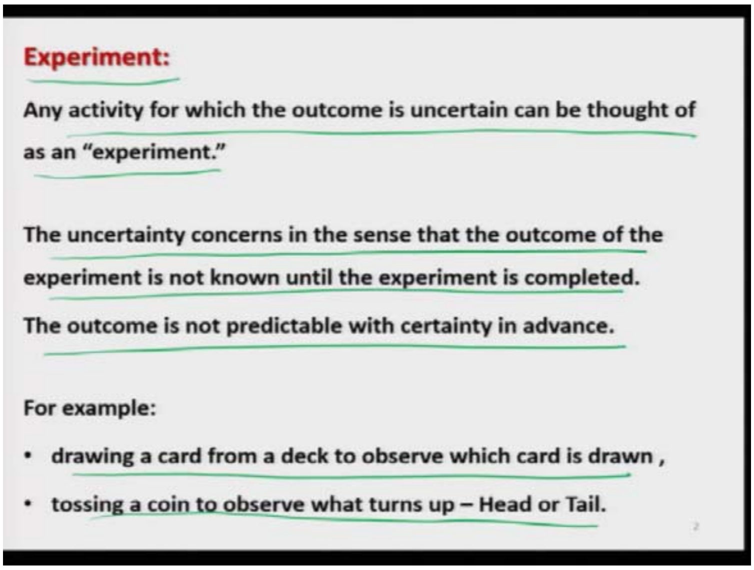
Hello friend, welcome to the course Essentials of Data Science with R Software-1 where we are trying to learn the topics of probability theory and statistical inference. And now we have begin with the topics of probability theory. So, can you recall that in the last lecture that was a very story telling type lecture where I had tried my best to give you some idea about the probability theory and I tried my best to convince you that probability is not a new thing for you that you are going learn. You already know it the only thing is that I want to now define it in a mathematical terminology.

The advantage will be once you are able to show someone that this is how you are trying to compute the probability not you friend but this word will agree with you. That means you will be able to develop a very good probability model. And that is the basic objective of our data science. So, now from this lecture I will try to take up very small issue, very small ingredients and I will try to move towards the construction of a probability model. I will try to take very small topic elementary topics and I will try my best to give you or to explain you in a very simple language.

But definitely when you are trying to learn it. You had to keep in mind there is a very strong theory behind these concepts. My objective here is very different. I am trying to connect the statistics with the data science. And data science is heavily depended on the computer. And for that we also need a software. So, now at this stage I will be not that much interested in giving you the theoretical construct, but I will be more interested in the application part of the tools. And for that whatever are the basic ingredients, I will try to explain you. That is my objective.

In case if you want to know the hard code theory behind these topics, I will simply suggest you pick up any book of statistics and try to learn it. Now in this lecture we are going to talk about the first basic ingredients of probability theory which is sample space and events. So, let us begin our lecture.

(Refer Slide Time: 2:47)



So, now if you I ask you a very simple question you know the answer. What is an experiment? How do you define an experiment? This is the first basic terminology which is needed for the probability theory. So, I can say that any activity for which the outcome is uncertain can be thought as an experiment. You always try to toss a coin or roll a dice. Do you note that before tossing the coin, what you will get head or tail? Certainly not. That you come to know only after the experiment is completed.

Do you know that if you start from your home at certain time at what time you will reach to your college, exactly? You always tell only the approximate value. Means if it takes about 20 minutes, then if you start from home at 10 0'clock in the morning. You may reach to your college at 10:18, or 10:19, 10:20, 10:21, 10:22 and so on. So, what is that? That is also an experiment that before reaching over there, you do not know how much exactly the time you are going to take.

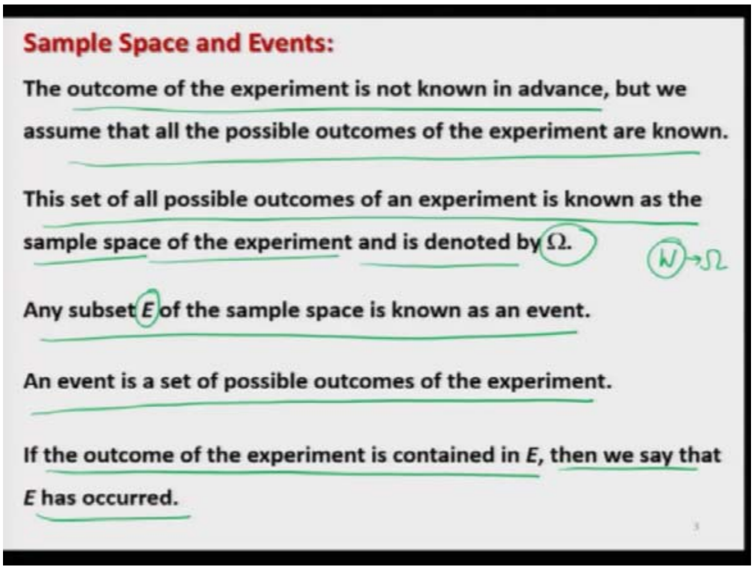
And the uncertainty in such events concern in the sense that the outcome of the experiment is not known unless and until the experiment is completed. And the outcome is not predictable with certainty in advance. That is why they are called as experiment. You are trying to conduct an experiment then you understand the meaning of it. You will never say that you are going to conduct an experiment to see whether the sun rises in the east or in the west; that you know.

In case if you try to combine two chemical known compounds in a given quantity, you know that what is going to be the outcome unless and until you are trying to create an experiment. Because in experiment you will not be knowing that what is going to be the outcome. Whatever the experiments, chemical experiments we try to do in our class 10 and class 12, we know their outcome and whenever you are trying to combine two chemicals and if you are not getting the same outcome possibly you try to repeat the experiment. Because you know the outcome but that is not really an experiment.

The experiment will be something like people are trying to experiment for the Covid vaccine on different types of patients where they do not know before the conduct of the experiment that whether the vaccination is going to work or not. So, similarly I can take here some more example. For example, when you are trying to draw a card from a deck to observe which card is drawn. You do not know unless and until the card is in your hand or the card is drawn.

Similarly, if you toss a coin you do not know what is going to turn up head or tail. That is an experiment. But remember one thing probability theory is beyond the tossing of coin rolling a die and drawing a card from a deck. Probability theory is now happening in you day to day life.

(Refer Slide Time: 6:00)



Now you should be convince about it. Well, we are trying to take very simple example to explain you at least you know that how to toss a coin or how to roll a dice. That is the reason we always take the example of where tossing a coin or rolling a dice. But you should not laugh or you should not feel that why probability theory is always happening with the tossing a coin. So, now we try to introduce or try to understand two terminologies. One is sample space and another is event.

Whenever you are trying to conduct an experiment, the outcome of the experiment is not known to us in advanced. But in most of the cases you will see that we assume that all possible outcomes of the experiment are known. For example, in case if you try to toss a coin, what will happen? Either there will be a head or there will be a tail. Similarly, if you try to roll a dice, what will happen? Either there will be 1, 2, 3, 4, 5 or 6.

Even if you want to go from your home to your college where you know that approximately it takes 20 minutes under the general condition. Then what will happen? You know that it will take time say 17 minutes, 18 minutes, 19 minutes, 20 minutes, 21, 22, 23, 24, 25 minutes. Unless and until there is some major issue. So, we do not know at what time I am going to reach to my college but I know with the very high possibility that in any case I will be reaching between 15 minutes to 25 minutes.

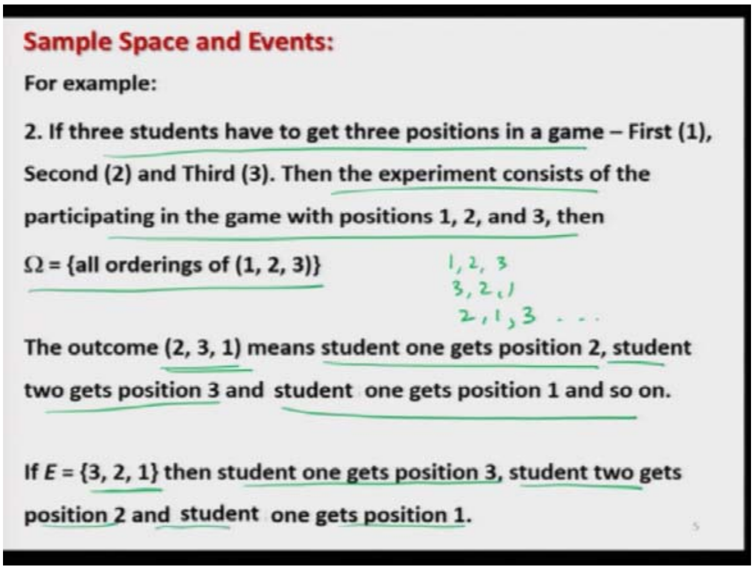
Similarly, when you try to go to your exams you do not know before the conduct of exam that what type of grade you are going to get A, B, C, D or F. But definitely you know for sure that what are the possible outcomes. A student can get a grade A, B, C, D or F. So, the set of all possible outcome of an experiment is known as the sample space of the experiment and it is indicated by the symbol Ω. Ω is Greek letter just like we have capital A, B, C, D so we have capital W and the capital W in the Greek letter is Ω.

And now when you try to take a subset from this sample space, this is called as an event. How? Suppose if I say you are trying to toss a coin. The possible outcomes are head and tail. But now can you say that there are two possible events; one can get a head or one can get a tail. Similarly, if you take about 15 minutes to 25 minutes to reach your college from your home. Can I say that I want to know that what are the chances that you will reach there between 18 minutes and 21 minutes?

So, what are these, these are the subsets of the all possible values. So, any subset of the sample space is known as an event and this is denoted by say here capital E. That is the symbol what we are going to use in our course. So, now if you try to think that whenever you try to reach such definitions in the book, many time you get badly confused that what are they trying to see or what are they trying to show you? But if you try to understand them just spend couple of minutes and try to see. What are they trying to explain you the things becomes very simple.

So, an event is a set of possible outcome of the experiment. And if the outcome of the experiment is contained in E that is the subset, then we say that E has occurred. For example, if a coin is tossed and the sample space has two points head and tail. Now in case if I try toss the coin and suppose it comes head. So head you can see that this is one of the point that was available inside the sample space and then we can say that an event has occurred. And what is that event? The event head has occurred, that can also be that a event scale has occurred. So, any outcome that has to be contained in E, then we say that the event has occurred.

(Refer Slide Time: 11:11)



For example, if there is a hospital where people are trying to record the gender of the newly born baby, newly born child. So, there are only two possibility that whether the child is male or the female. So, in case if you try to collect these two possible outcomes, then together they will try to constitute the sample space Ω. So, here for example M is indicating that the child is male and F is indicating that the child is female.

So, now when somebody is going to the hospital, the person will say that the sample space consist of only two points, male and female. Now whenever there is a baby born, then what so ever be the gender of that baby that can be either male or female. So, in case if the baby is male, then we say that E consist of M and then E is the event that the child is a male boy. And similarly if E consist of only F that is the baby is female, then we say that the event E is that the child is a female or a girl.

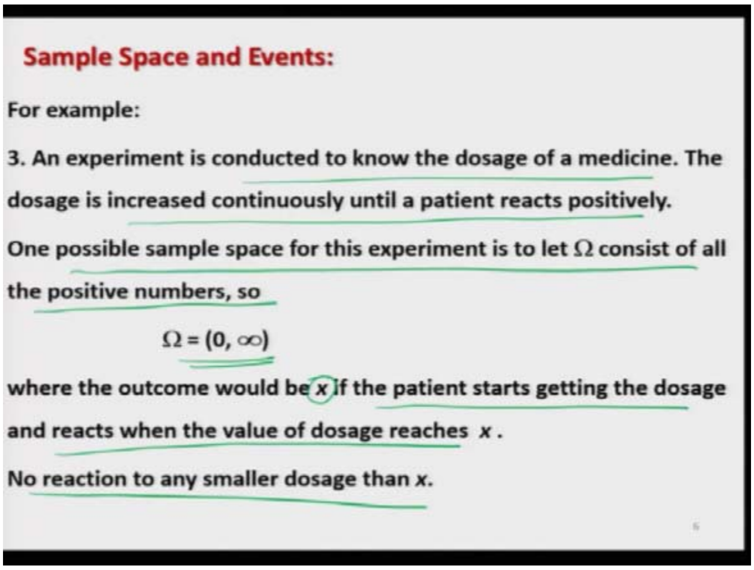
Similarly, if I take another example if there are three students who have got three positions in the game. Means if there are three students and they are trying to participate in a game there are only three possibilities that the student can get the first position, second position, or third position. Now the experiment is conducted and those three students try to occupy the positions 1, 2, and 3. Then what will be the sample space?

Sample space will consist of all ordering of the number 1, 2, and 3. Why? Because there are three students and those 3 students can occupy the places in different orderings. For example, the most simple one will be first student get first positions, second student gets second position, third student get third position. But do you think is it the only possibility. No, first student can get the third position, second student can get the first position and so on.

So, in case if you try to see there will be different combination possible 1, 2, 3; 3, 2, 1; 2, 1, 3 and so on. And in case if you try to collect all such ordering they will try to create the sample space. Now in case if you try to understand what is an event or how are you going to interpret it. Suppose the outcome is here 2, 3, 1, this means student one has got the position number 2 student two has got the position number 3 and student one has got the position number 1.

So, now in case if you are getting an outcome like 3, 2, 1, then this is an event. This means what? That student 1 gets position number 3, student 2 gets position number 2, and student 1 gets position number 1. This is the meaning of this 3, 2, 1. So, this is an event.

(Refer Slide Time: 14:28)



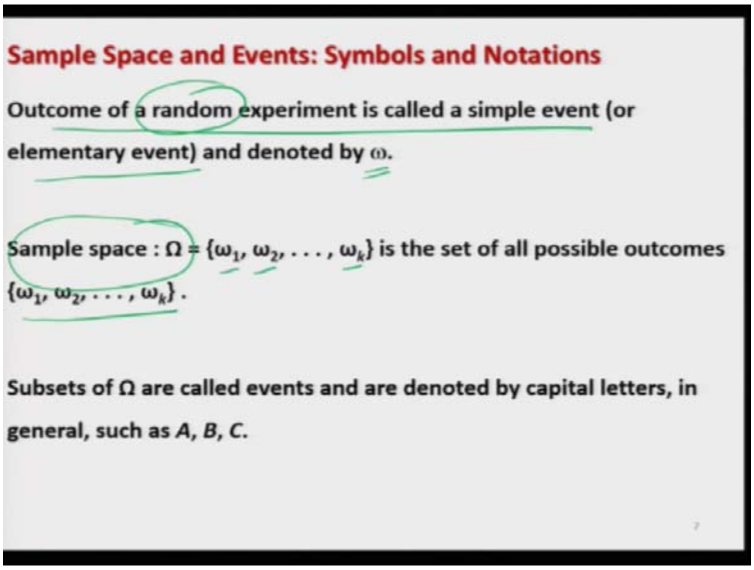
Similarly, if I try to take one more example. Suppose an experiment is conducted and we want to know the dose of a medicine. Suppose if somebody wants to conduct such an experiment, what the person has to do? The person has to give a certain amount of dose to the patients and the quantity of that medicine or the dose is increase continuously and it is continued until the patients, patient reacts positively.

You have seen that somebody has got a body temperature, first the doctor says give half tablet then if there is no effect, they try to increase it to 1 tablet. And if there is no effect, they try to increase to 1.5 and 2 tablet and so on. Definitely they will try to do up to only that level up to which the medicine is not harming the patient. So, in this case what will happen? One possible sample space for this event is to assume that let omega consist of all the positive numbers.

We got dose can be any value. Between 0 and infinity. Well infinity is something like a symbolic. But it is going to be positive. And what will be the outcome of such an event. The outcome will be, can be denoted by some value here is small x which is the value of the doses that is going to help us or that is going to have patients that if the patient start getting this dose and react positively when the value of dose, doses reaches to x.

And as long as the dose was smaller than x, there was no reaction, there was no improvement, the medicine was not working. This is what we mean. So, now you can see here that understanding the concept of sample space an event is not difficult.

(Refer Slide Time: 16:36)



The only thing is this you have to think and decide yourself for a given experiment that you are trying to conduct. That how are you going to define the sample space and objective. One thing I would like to clarify here very clearly. That the definition of a sample space and events are related to the objective of your study. It depends what you really want to do. For example, in case if you are trying to roll two dice, then there can be different objective in which you might be interested.

For example, you want to know the some of the numbers on the upper faces of the dice or you want to know the difference or you want to know the some of their square, whatever it is. So, it depends on your objective what you want to do but my advice is that try to understand the experiment. Try to understand the need of the experiment that why the experiment has been conducted. What you really want to know? Based on that you try to define the objective.

And based on that you try to take the definition and try to decide what is your sample space and events. So, now we have understood this concept but we need to indicate it by some mathematically terms so that anybody can understand it. So, now whenever we are trying to conduct an experiment the outcome of a random experiment is called as an, as a simple event or elementary event and it is denoted by small omega.

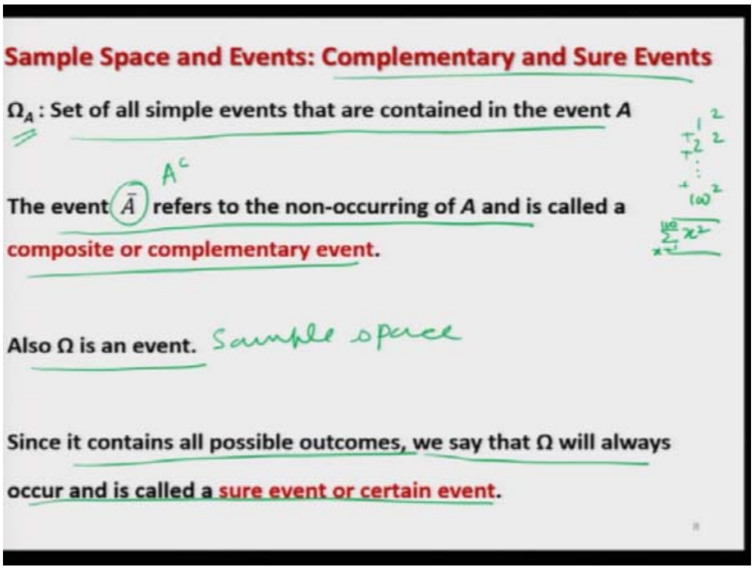
You can see here I have introduced here a word random experiment. You see whenever you are trying to conduct an experiment as we discussed earlier, we do not know what is the outcome. What is the possible outcome that you are going to get well that total possible outcomes are known to us, but out of them which of the outcome is going to come. That we do not know that is in that sense we call the experiment to be a random experiment.

So now whenever we are trying to conduct an experiment there can be more than one outcomes and suppose there are k possible outcomes which are indicated by say ω_1, ω_2, say ω_κ. So, now in case if I try to collect the set of all possible outcomes of ω_1, ω_2, sω_κ then they are going to constitute the sample space Ω. So, now you see here whatever you have explained a concept of Ω in these 3 examples.

This medicine example, this students example and this male and female child example. That omega is now well defined. For example, if you try to take here this example of say male and female. This ω_1 can capital M and ω_2 can be capital F. And similarly if you try to take here this example of medicine means any value of the dose of the medicine that is reacting to help the patient in curing him or her that is your omega value.

In this case the value of Ω can be in some interval form also. So, now the subsets of Ω are called event and they are denoted by capital letters in general capital A, capital B, capital C and so on. But that is our convention. That is our understanding and at least in this course we are going to follow the, this convention.

(Refer Slide Time: 20:09)



Now I try to make the things more clear and we try to define it in a mathematical symbol. Because once you try to define the things in a mathematical symbol the things become very easy to understand. Now suppose if you want to compute the some square of the first 100 natural numbers, then I can write down here 1, 2, 3, 4 … 100 then I try to square them then I try to sum them and so on. All those things I can write down very easily here $\sum_{x=1}^{x} x^2$.

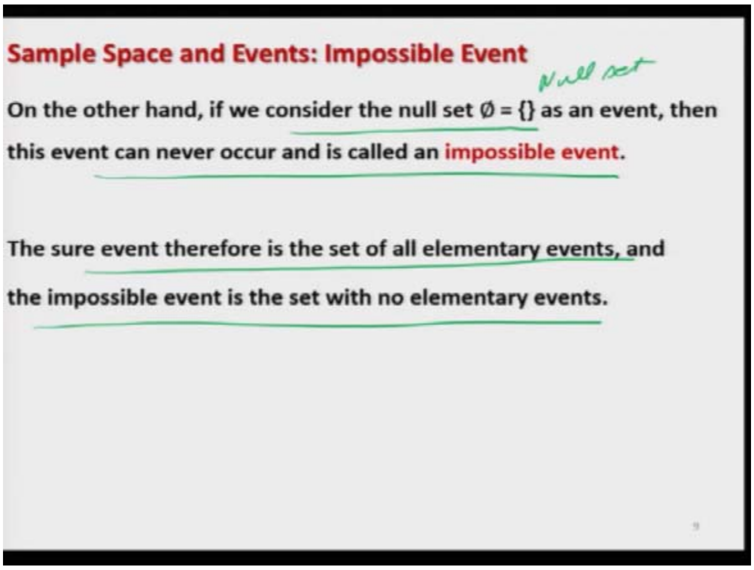
So, you can see here mathematically you can express these thing very compactly which anybody can understand very easily. With this point of view, I try to express whatever we have discuss, whatever we have understood in a mathematical symbol. So, now I try to define here two things, complementary event and sure events. So, let this Ω_A this is the set of all simple event that are contained in the event A. So, that means anything can happen from this set.

Whatever cannot happen that is also another set and that event which cannot occur in this this is called as complementary event or a composite event. And that is indicated by $\overline{A}$. Sometime you write down here A complement also whatever you want to write different books followed different symbols and notation that is not a big deal but here we are going denote Ω_A it by $\overline{A}$.

So this is also an event and this refers to as non-occurring of the event and it is called as composite event or complementary event. So, you can see here that Ω this is the entire sample space. Sample space consist all possible outcomes. So, do not you think that omega is also an event? Yes, Ω is an event. And since it contains all possible outcome we can say that omega will always occur and it is called as a sure event or certain event.

Suppose there is a coin, there are two possibilities head or tail. And if I say well what is the probability of getting head and tail or head or a tail. Can you tell me? This is joint event means I just want to know that if head comes or tail comes I will be very happy the probability is always one because there are only two possibilities where that either head will occur or tail will occur. So, when a event is comprising of both the events the probability of this event will always be equal to one.

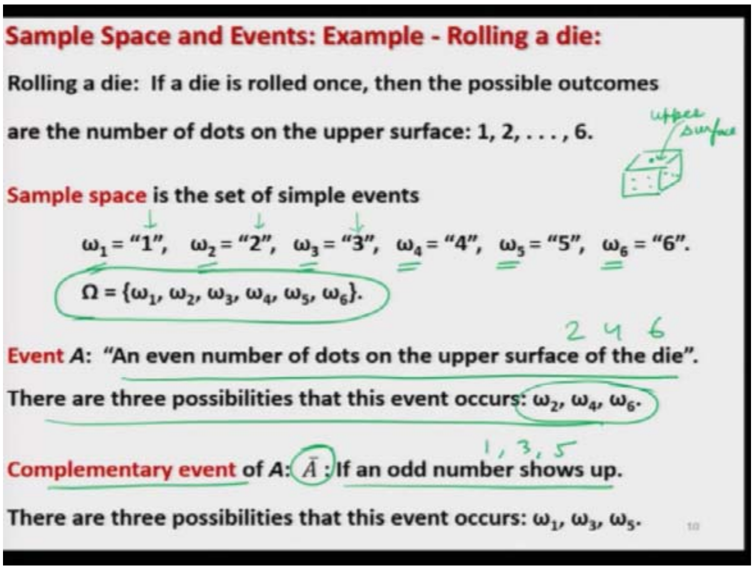
(Refer Slide Time: 23:06)



For example, if I roll a die and then there are six possible option 1, 2, 3, 4, 5, 6 and if I say what is the probability of getting 1, 2, 3, 4, 5 or 6 that is always going to be one. Well definitely one of the point is going occur. So, such points are called as sure event. Now when you have a sure event then definitely there has to be an impossible event also. Because sure event is occurring always. So, impossible event will never occur.

So, in case if we consider the set here ϕ as null set that mean there are no elements and this is also is an event. So, then this event can never occur and it is called as an impossible event. So, if you try to see the sure event is the set of all elementary events and the impossible event is the set with no elementary events. And sure and impossible they are complement to each other.

(Refer Slide Time: 24:18)

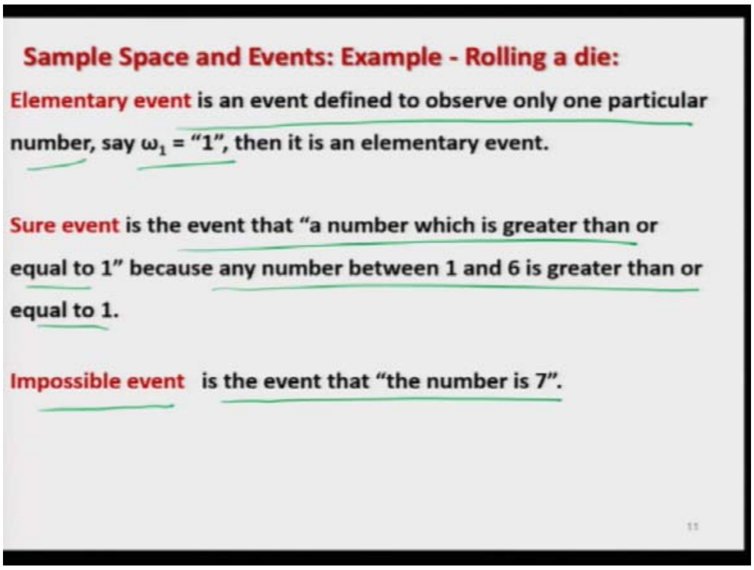


So, now I try to take a very simple example and I try to explain you this concept. Suppose I try to roll a die and suppose the die is roll only once. So, there are six possible outcome. For example, you know that if there is die like this one there is a points like here 1, 4, 1, 3 and so on. So, if they say dice is roll. Then what we try to do? We try to observe what is appearing on the upper surface. You must have played very means different type of the board games from your childhood.

So, now in case this if you want to define the sample space, then there are 6 possibilities. Getting 1, 2, 3, 4, 5 and 6. And I have denoted all this six values by event ω_1, ω_2, ω_3, ω_4, ω_5 and ω_6. So, ω_1 is the event that one appears on the upper surface. ω_2 is the event that 2 appears. ω_3 is the event that 3 appears and so on. And this Ω is going to be the collection of all possible events. So, ω_1, ω_1 up to ω_6.

Now I can define the event according to our objective. Suppose if I define the event A as an even number of dots on the upper surface of the die. Then there are three possibilities that the event occurs with this event. Because the number are going to be even then either a 2 can occur, 4 can occur, or 6 can occur. So, that means they are corresponding to ω_2, ω_4 and ω_6. And what will be the complementary ev2,nt for this event A? If event A is asking that it is trying to consider all even numbers, then obviously the complementary will be consideration of all the odd numbers.

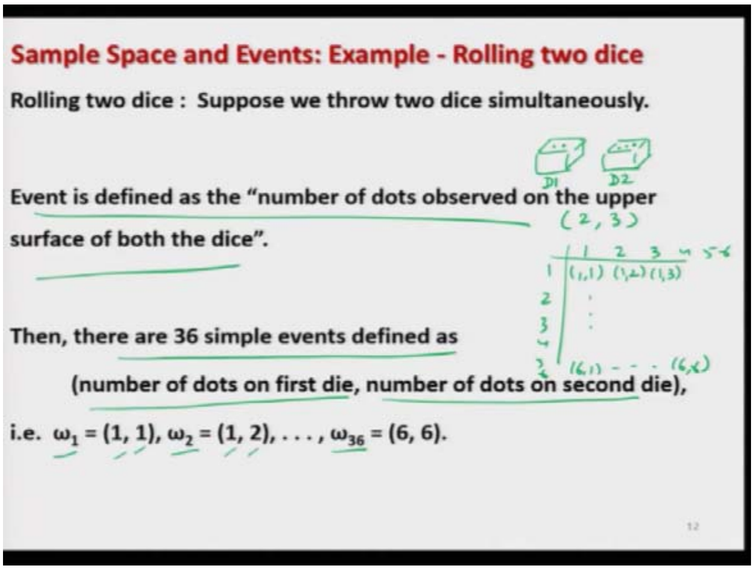
So, the complementary event of A will be denoted by $\overline{A}$. And this will be if an odd number shows up then we have a complementary event. And there are three possibilities. What are those possibilities? Whether you observe 1, 3 or 5. And this are 3 events which are indicated by ω_1, ω_3 and ω_5. (Refer Slide Time: 26:47)



So, this how you see that you can define these things. Now in case if you want to define here an elementary event. Then this elementary event is an event defined to observe only one particular number. Say for example, omega 1 equal to 1, then it is an elementary event. Means if you try to take omega 3 equal to 3 then that is also an elementary event. And what is here the sure event which is 100 percent guaranteed to occur?

Sure event is the event that a number which is greater than or equal to 1 occurs. Why? Because any number between 1 and 6 is greater than or equal to 1. So, whatsoever number come the event is always going to occur. So, that is a sure event. When we have our sure event, then we also need to understand what will be the impossible event. So, in this case for example we can define an impossible event like this is the event that the number 7 appears on the upper surface of the die. Do you think that number 7 can occur? There are only 6 numbers at least we have assume that we have die which has only six numbers 1, 2, 3, 4, 5 and 6. So, this event can never happen and this is call as an impossible event.

(Refer Slide Time: 28:15)

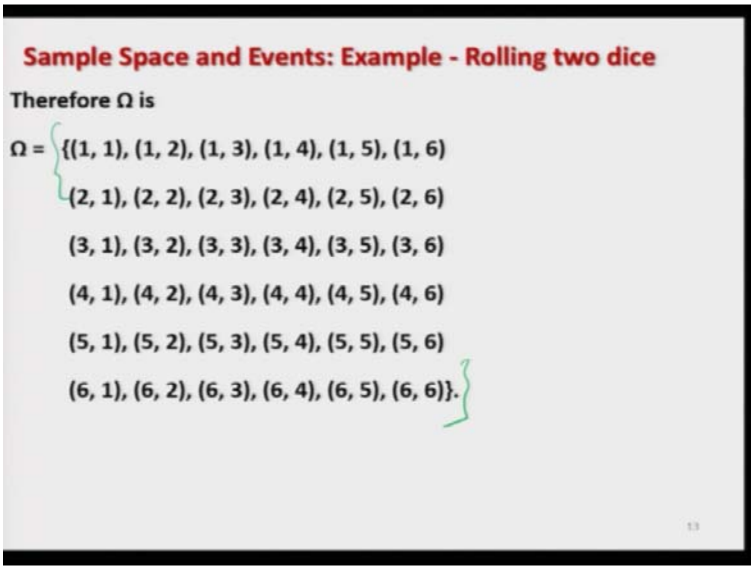


Now I try to change my objective and suppose I try to say that till I am going to roll two dice. And we try to throw both the dice simultaneously like with this one. So, our outcomes are going to be like this one. Suppose if there are two points here and here there are three points on the second die. So, and if I try to write down the number of points on the upper surface. Suppose this is my dice number 1 and this is my dice number 2. So, I can write down here the value here say 2, 3.

Suppose this is an event that whatever numbers we are going to observe on the upper surface of both the dice. That is my event. So, this is what I have defined here. Now, how many possibilities are there. If you try to create here a small table like as 1, 2, 3, 4, 5, 6; 1, 2, 3, 4, 5, 6. You can see here 1 , 1 can come, 1, 2 can come, 1, 3 can come up to here say 6, 1 can come up to here 6, 6 can come. So, there will be all together 36 possible outcomes.

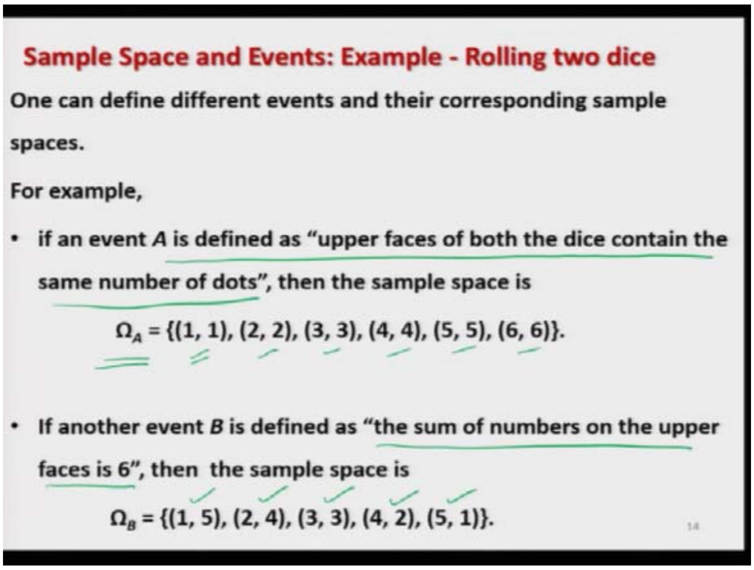
So, there are 36 simple events which are defined as the number of dots on the first die, number of dots on the second die. And they are indicated by ω_1, ω_2 up to ω_{36}. . So, Ω (1, 1) means 1 and 1 occur on the upper surface of both the dice. 1, 2 means 1 occur on the dice one and 2 occurs in the dice two and so on.

(Refer Slide Time: 29:58)



So, now if you try to see the sample space Ω. will consist of these possible values. And these all the collection of all events or the sample point that is creating our Ω. So, now you obtain the sample space also.

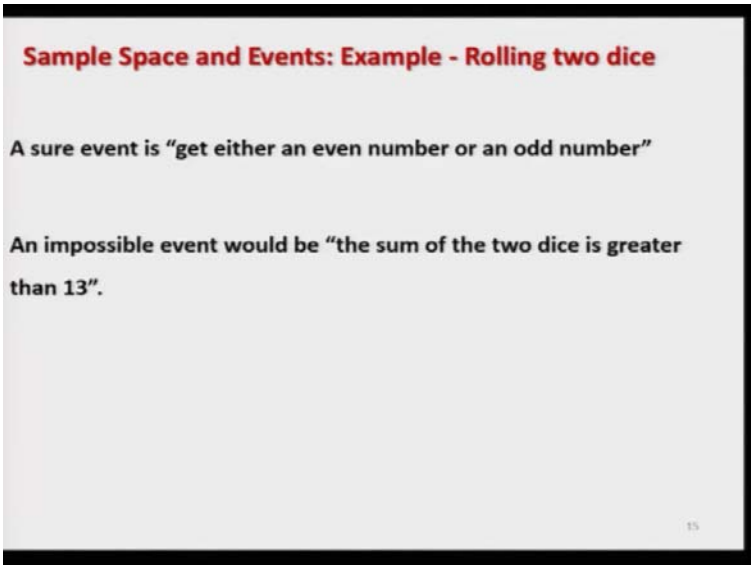
(Refer Slide Time: 30:15)



Now, from this experiment I can have different types of objective and I can define different types of event also. For example, if an event A is defined as upper faces of both the dice contain the same number of dots. Then what do you think, what will be your Ω_A that means the sample space corresponding to this event A. They will have here 6 point that both the dice have either 1, 2, 3, 4, 5, or 6 numbers on their upper faces.

On the other hand, in case if my objective is something else, then I can define another event here B which is the sum of numbers on the upper faces is 6. In that case, the sample space is corresponding to this event is defined as for example. Now this omega B can have six possible points. 1, 5; 2, 4; 3, 3; 4, 2; and 5, 1. And if you try to see if you try to sum any of these, this two numbers their sum is always 6. And similarly if you wish you can define so many events and from this Ω.

(Refer Slide Time: 31:32)



And in this case a sure event is get either an even number or an odd number. Definitely you are going to get 1 out of them. So, this event is 100 percent guaranteed or this is a sure event that will always occur. And an impossible event would be like the sum of the two dice is greater than 13. It is not possible because the maximum value which are a die can take as the number of points on the upper face is 6. So, the maximum value be 6 and 6 and their sum is going to be 12, 13 can never occur , at least in this case. So, that is going to be an impossible events.

So, now we come to an end to this lecture. You see this was a very simple lecture, just trying to introduce with the basic notation of a sample space events et cetera. But definitely they have very important. Because whenever you are trying to work in the decision science. Whenever you want to compute the probability or something. These are the basic ingredients unless and until you define your Ω, ω_A, A, B etc. How can you compute the probability? And this definition are in your control. The person who wants to know the probability of certain event that person knows but that person possibly does not know statistics or data science. That person will come to know only to you. You have to listen and you have to translate what the person is trying to say.

For example, if I take the example of this medicine, you have to decide. What will be the range of the dose of the medicine? Minus infinity to infinity, 0 to infinity, minus infinity to 0, some integer or some real number. What so ever you want you have to decide and one thing is this what so ever decide you have to match with the real thing what is happening in practice. If you try to say the doses are going to be integer value. That means there are only some tablets.

But if there are no tablets but there are some liquid and that liquid can be given in any volume then definitely the dose can be 2.3 or 2.4 ml also. But you have to take a call. And for that you have to understand what is really happening. And I always say that our statistician will be that fellow who is a very good social scientist. You cannot say that I do not know, how the doses of medicines are given.

So, you try to understand this thing try to take very simple example. Try to think and try to construct your own sample space and events and get confident. So, you try to do it and I will see you in the next lecture till then good bye.