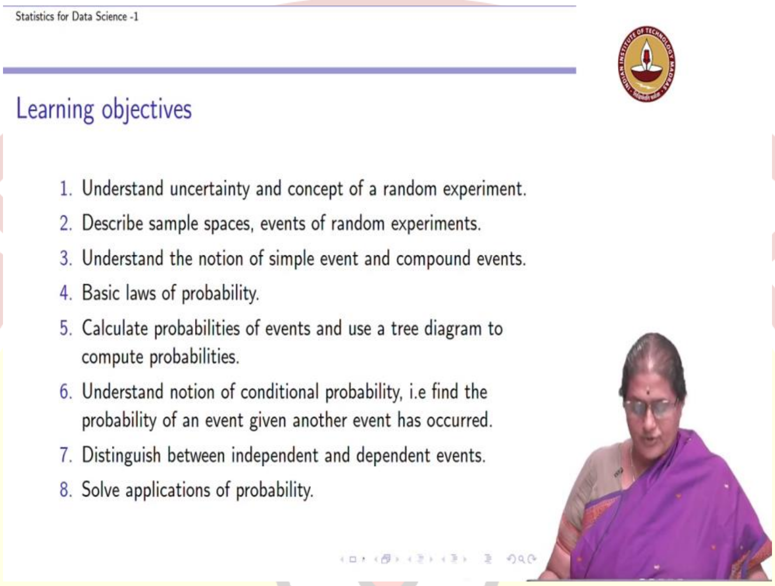
IIT Madras

ONLINE DEGREE

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)



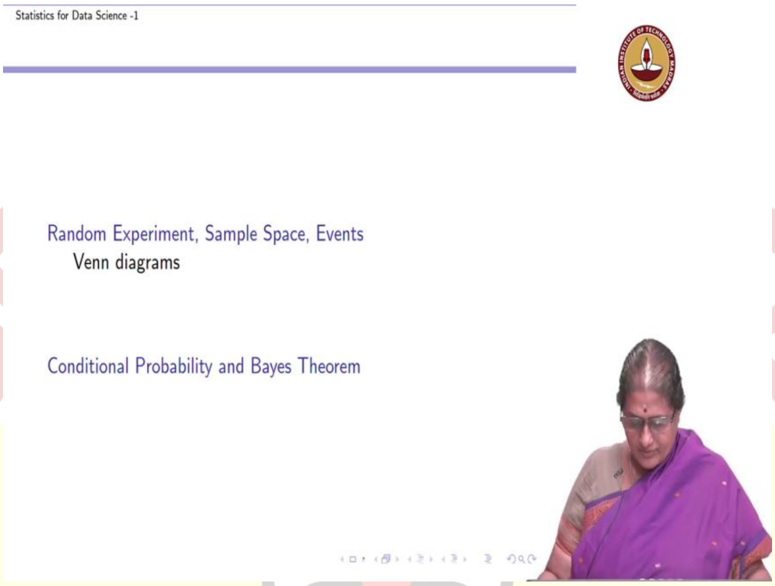
So having learned about permutations and combinations, we now move on to learn about the fundamentals of probability. So what are we expected to learn in the next two weeks is, we are going to first understand what we mean by uncertainty. And we will understand about a very important concept that is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

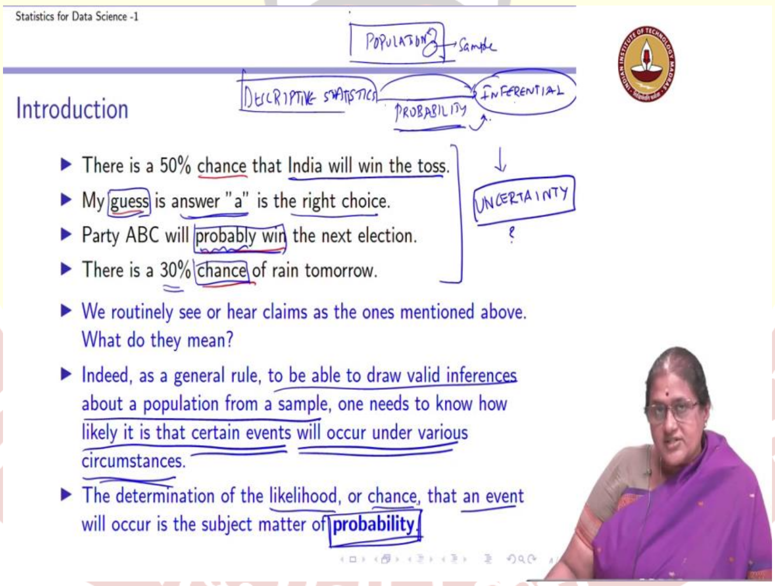
Now, the concept of random experiment and what are events, sample spaces, events these are extremely important in our understanding of probability. We then move forward to understand what is the notion of simple event and compound event, we will introduce the basic laws of probability.

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.

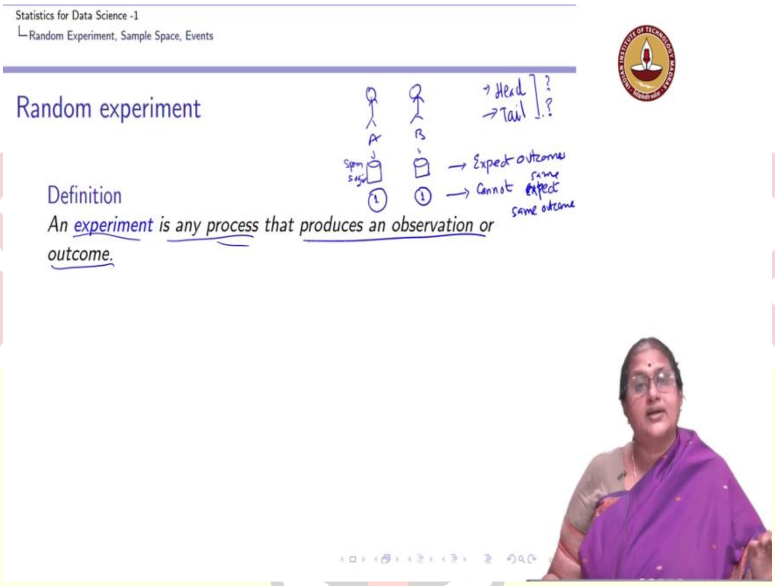
Similarly, I am just telling that there could be a chance of rain. Am I certain it would rain tomorrow? The answer is no. I am just saying there could be a chance of rain and I am quantifying that chance with a 30% chance. So, we routinely keep hearing claims of this kind, where we are addressing this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

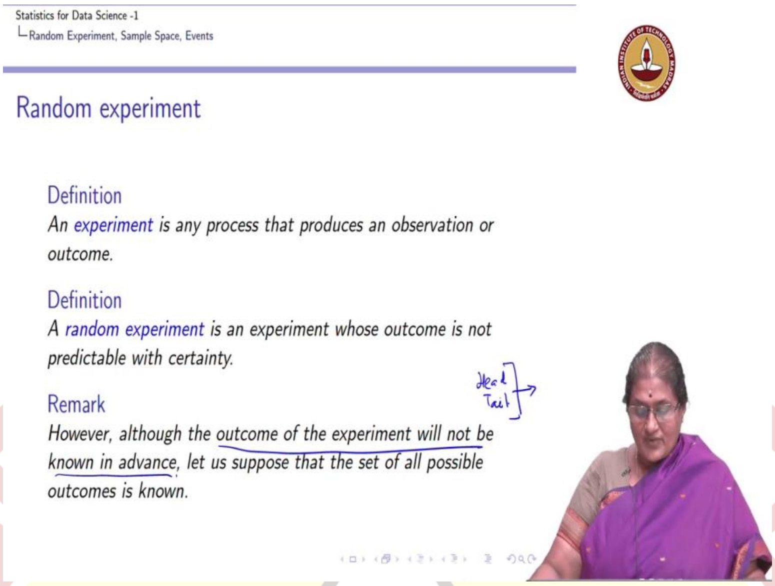
So I am seeing what is the outcome, it is going to be a mixture of sugar and water this is what I am mixing. Now suppose I give both of them 1 rupee coin, okay I am giving both of them a 1 rupee coin and I am asking them to toss the coin, that is each one of them gets identical 1 rupee coin and both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

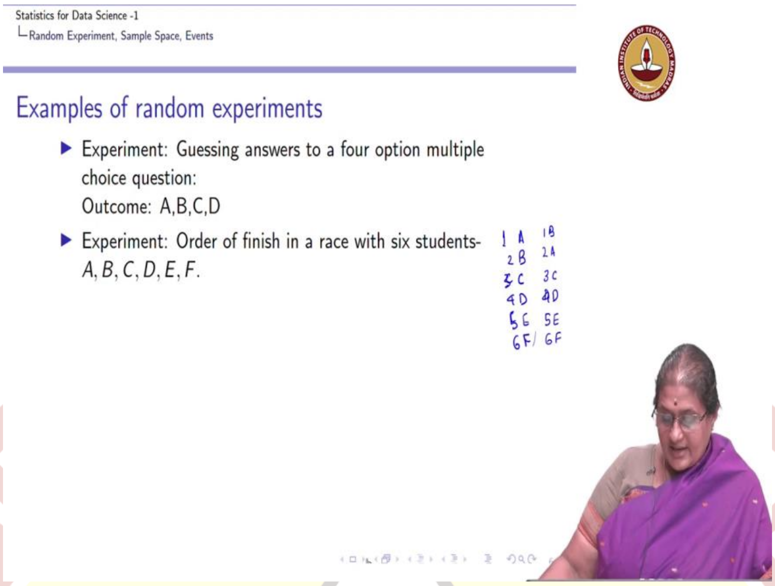
So the uncertainty is which of the outcomes is actually going to happen. It is not a random experiment. It is not an experiment with unknown outcomes. Here, I know what are my outcomes. In this experiment, I know my outcome is either a head or a tail but the uncertainty is whether it is a head or tail.

(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)

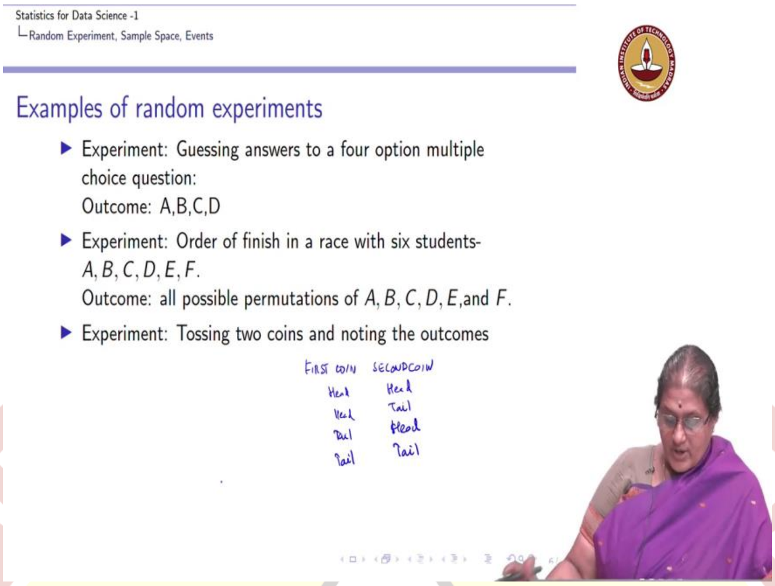


So, now let us look at examples of a few random experiments. Suppose, my experiment is guessing answer to a four option multiple choice question. Let my options be A, B, C, D. The possible outcomes of this experiment are I have A, B, C, and D. Those are the possible outcomes because the experiment is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

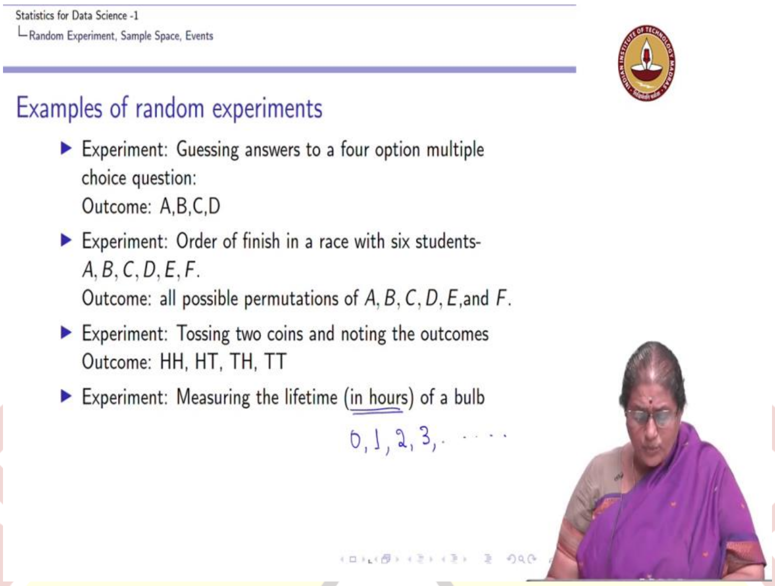


So my total number of outcomes is nothing but I can list all possible permutations of A, B, C, D, E, and F. Those are my possible outcomes of the experiment where I need the order of finish. Now, why is this a random experiment? Because when six people are competing in a race, I do not know who is going 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.

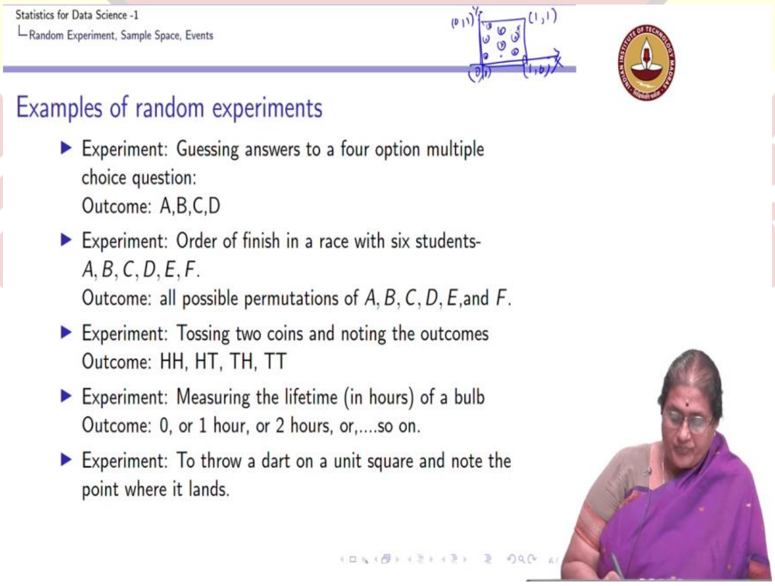
So, the possible outcome is I have a head on my first coin a head on my second coin. I have a head on the first coin, a tail on my second coin. A tail in my first coin, a head in my second coin. A tail in my first coin and a tail in the second coin. These are the possible outcomes of the experiment of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



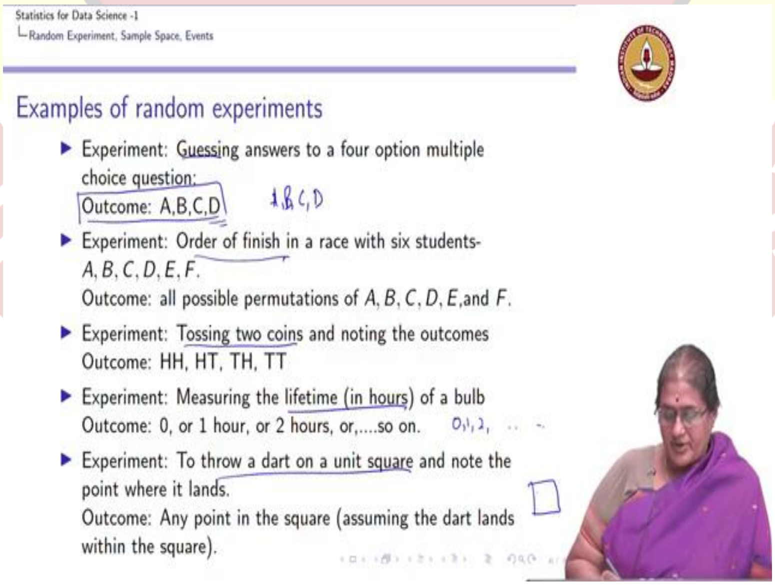
It could be any amount of time. So, it is 0 or 1 hour or 2 hours or so on. These are the possible outcomes. I do not know how long a bulb is going to last. It could last for 30 minutes, 45 minutes, in which case my outcome is, I am going to record it as a 0 hour. If it lasts for or it depends, if it,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

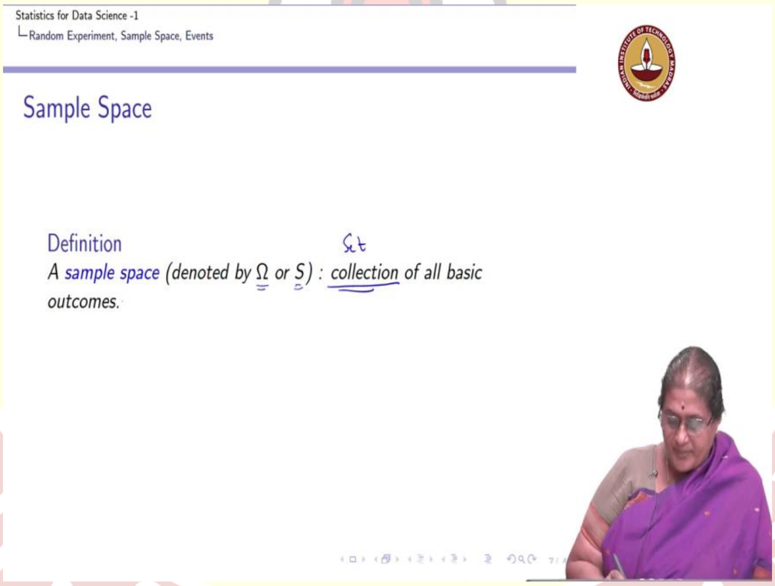


So, any point in this square is a possible outcome to the random experiment. Why is it random? I know it would be any point in the square but I really do not know which of these points in this square would be the actual outcome. So, these are examples of random experiments. You can see that they are 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.

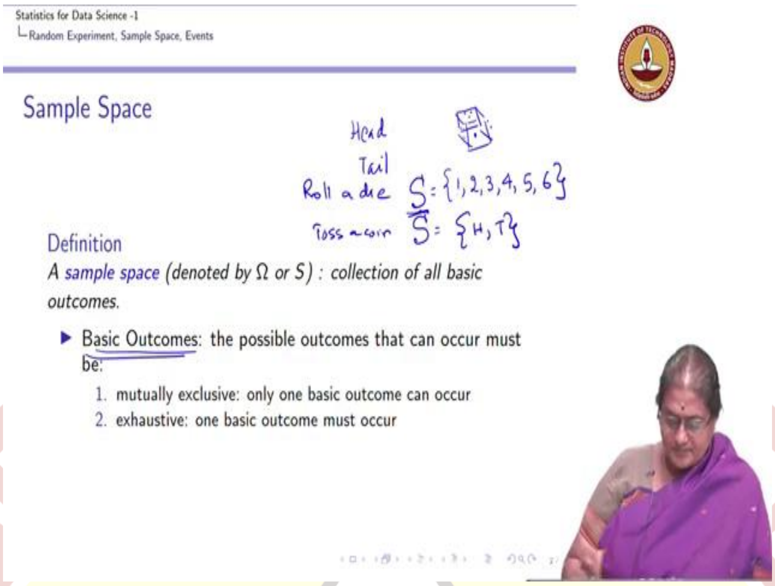
Now again, when I look at this point in the square, I am recording it as a coordinate. Again, there are infi... countable, but there are many points in this square. So immediately, you see that I need to have a way to represent all the outcomes of this random experiment.

(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)

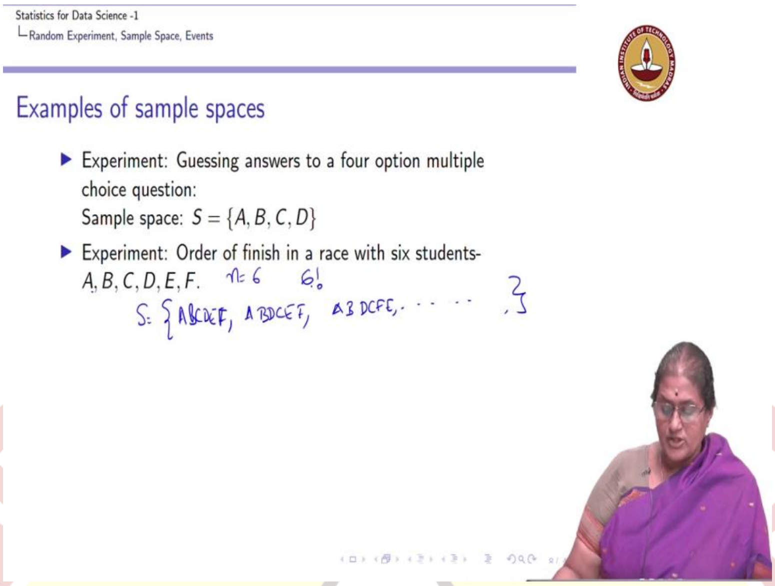


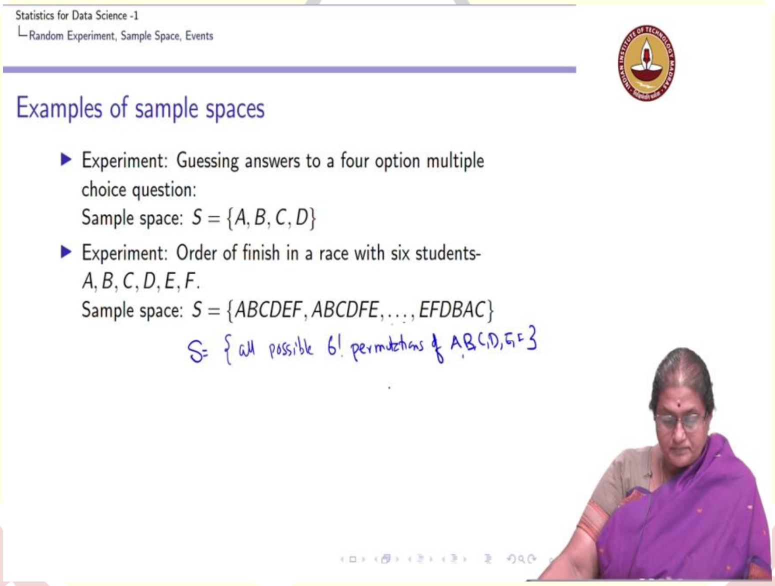
Now, what do we mean by basic outcomes? The basic outcomes are all possible outcomes of a random experiment which should be mutually exclusive. By mutually exclusive I mean that only one not the outcome can occur. So when I am tossing a coin, the outcomes are head or a tail, these are the 2 possible 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)

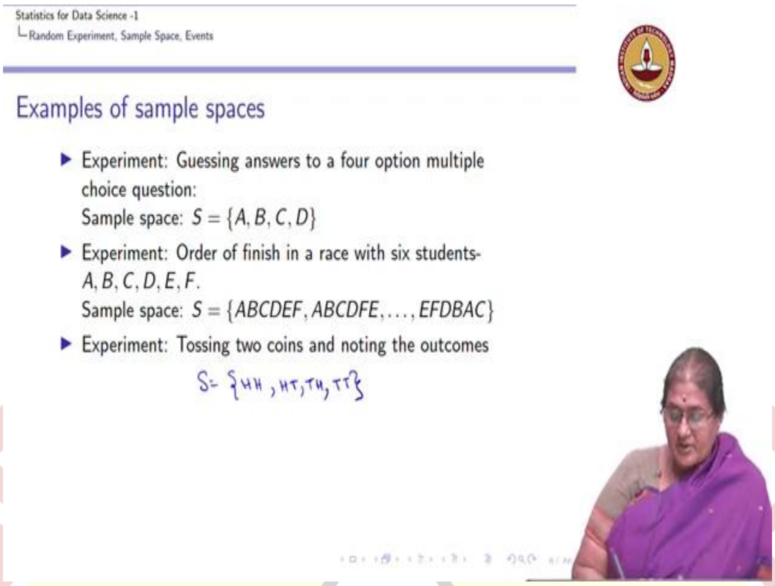


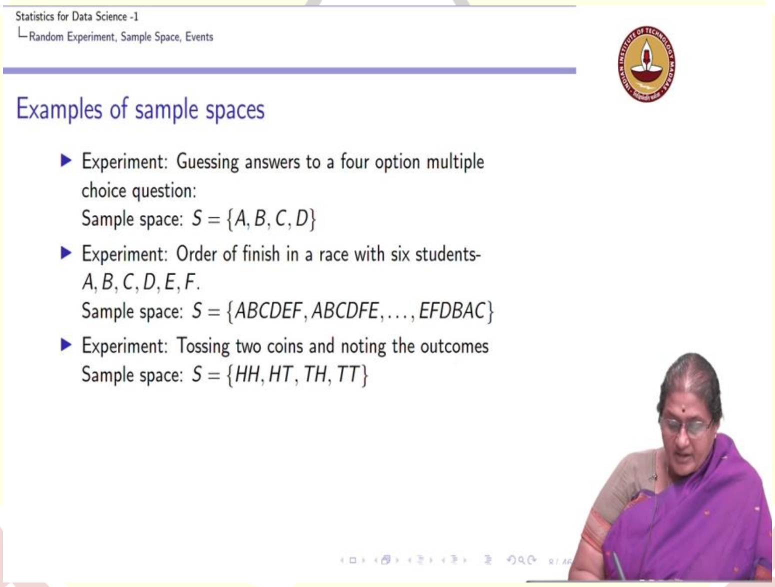


So, now let us go back to the experiments which we have seen earlier and represent them now as sample space. Recall sample space is a set of all outcomes of my experiment. So guessing answers, the sample space is going to be I have 4 outcomes so A, B, C, D I represent S equal to A, B, C, D. Order of 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.

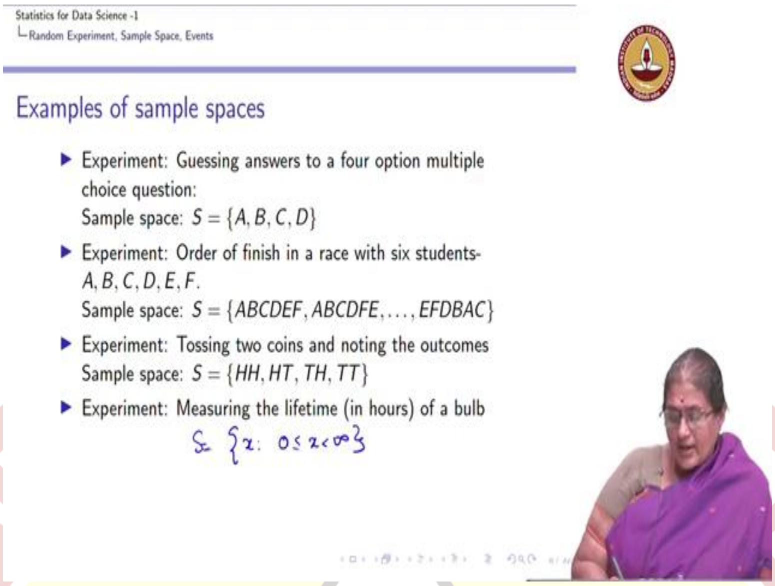
(Refer Slide Time: 25:12)

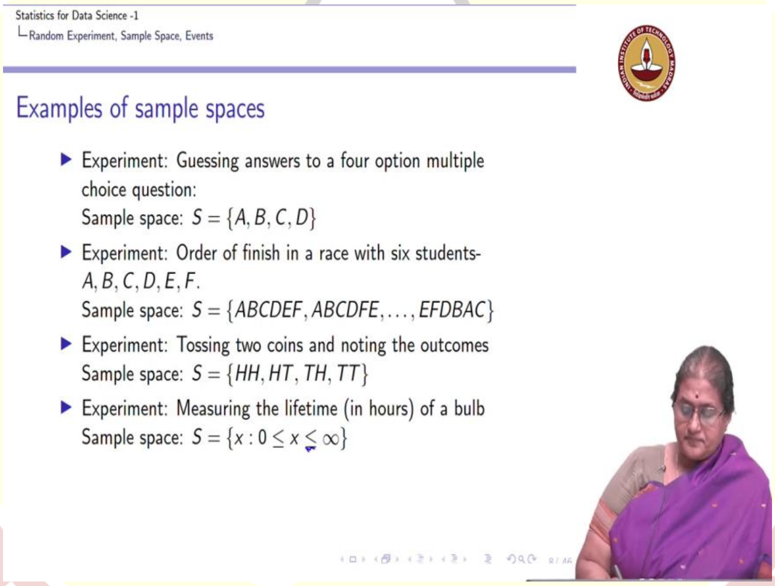




Now, let us go on to the next example. So, I again have 2 coins. So, my sample space is, I can either have a head-head; head in my first toss, head in the second toss or head in the first coin, head in the second coin, head in the first coin, tail in the second coin, tail in the first coin, head in the 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.

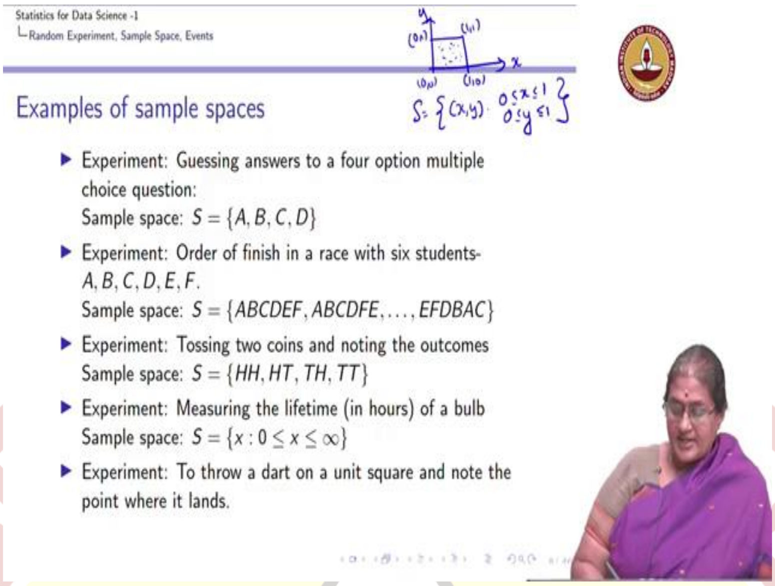
(Refer Slide Time: 25:44)

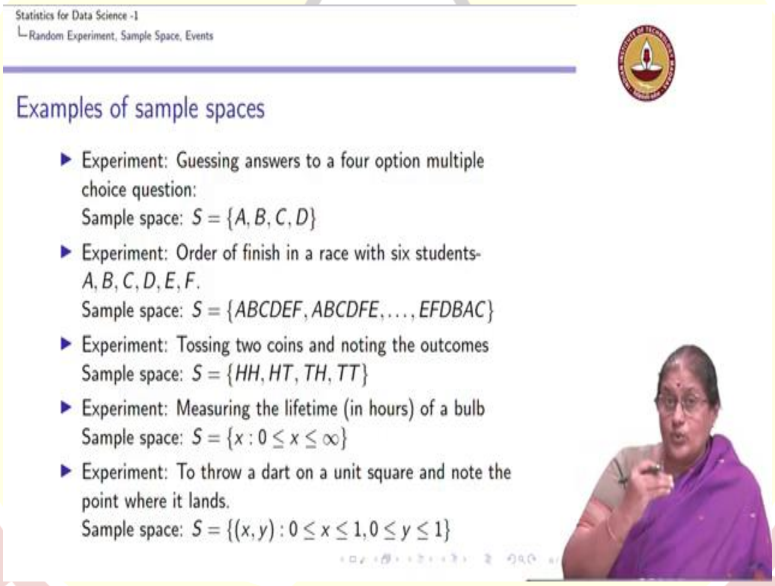




Now measuring the lifetime, I said it could be a 0 hour, it could be a 1 hour. So, I can express this as x where $0 \leq x < infinity$. And, this would actually I can, I know that this is either a 0 hour or a 1 hour or a 2 hour, it is less than infinity, sorry, okay.

(Refer Slide Time: 26:11)

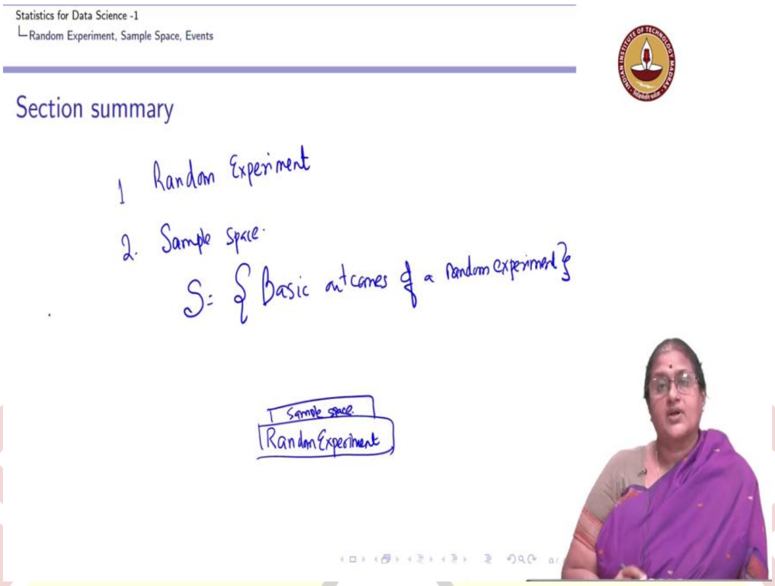




The next thing is when I throw a dart, I have already told you that this is going to be my X axis Y axis, this is going to be the square. If I represent my cord… points in the square using my coordinates of the X, Y could include my boundary points also. So, my S is going to be the set of all ordered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