PROBABILITY THEORY FOR DATA SCIENCE

Prof. Ishapathik Das

Department of Mathematics and Statistics

Indian Institute of Technology Tirupati

Week - 01

Lecture - 03

Special Events and Various Approaches to Defining Probability

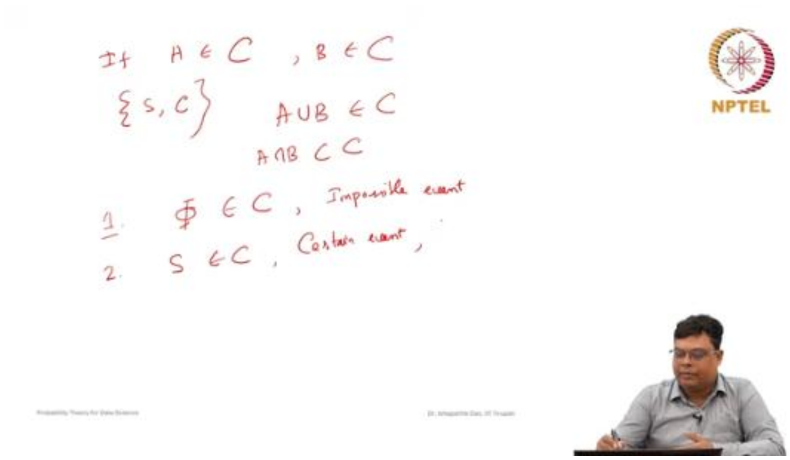
Since these are all subsets, if A is an event, it means that $A \in \sigma$-field. If B is also an event, it means that $B \in \sigma$-field. Suppose, let us consider in general, suppose S and C. So, S is the sample space, and C is some collection of subsets, like a class of subsets of S that is a σ-field that we discussed. Then $A \cup B$, by the definition of a σ-field, will be in C.

So, it will be an event again because any element in C will be defined as an event. Then, $A \cap B$ will also be in C. Now, it is also an event. Now, some of the special events—one event is called an impossible event, a null event, some of the names—suppose $\emptyset$. $\emptyset$ is a null set, so it does not contain any elements.

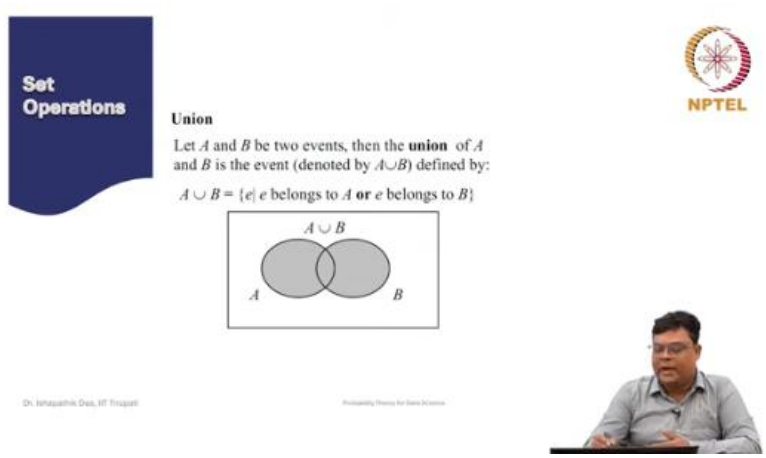
You are throwing a die: 1, 2, 3, 4, 5, 6; something will happen. But $\emptyset$ does not contain any of that, so $\emptyset$ cannot occur anytime. That is why it is known as an impossible event or a null event sometimes. The second is the sample space, S. Whatever observation comes in any run will be inside S.

Then we say that S occurs. This means whatever happens, whatever the random experiment, whenever you do it, S must happen all the time. So, this is also in C, so it is an event. It is known as an impossible event or a null event. Some other names are special event, null event, or empty event.

The event contains no elements. And S is known as a certain event, or it is known as the entire event. Entire event always occurs. Sometimes it is referred to as an entire event or a certain event. It is called an entire event, so we also say entire event.

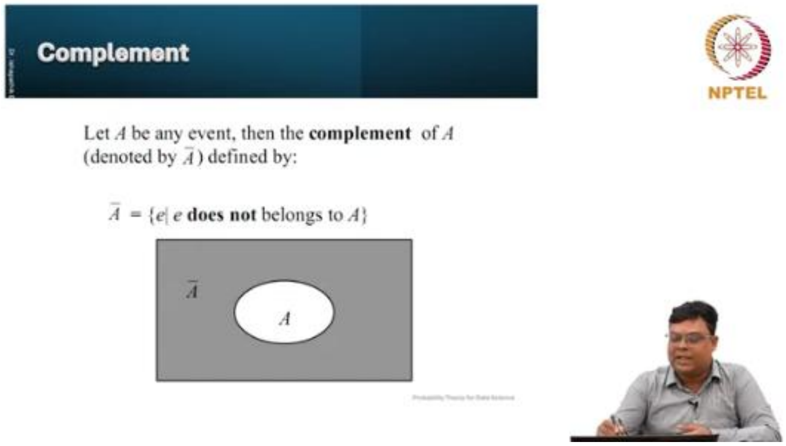


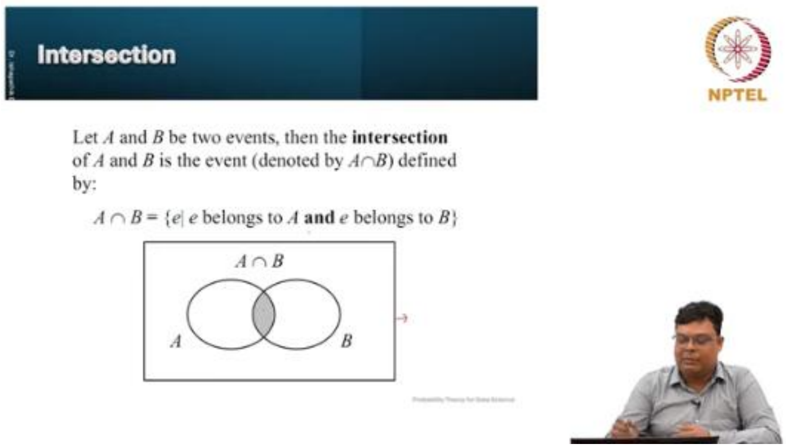
There are some other unions. Union means, in the Venn diagram, it is shown. It contains all the elements of A and B. Any element, either it belongs to A or B, or it may be both. It is known as union.



Intersection means it is in the common space. So, that means this is A and this is B. This is the sample space. This is A and this is B, and this part is known as A ∩ B. That means all the elements of e that belong to A as well as belong to B, i.e., e ∈ A and e ∈ B.

It is known as the intersection. So, another concept is called the complement of an event. So, how is the complement of an event defined? The complement of an event means you consider all the elements which do not belong to A. So, A is any event, suppose let A be the event.

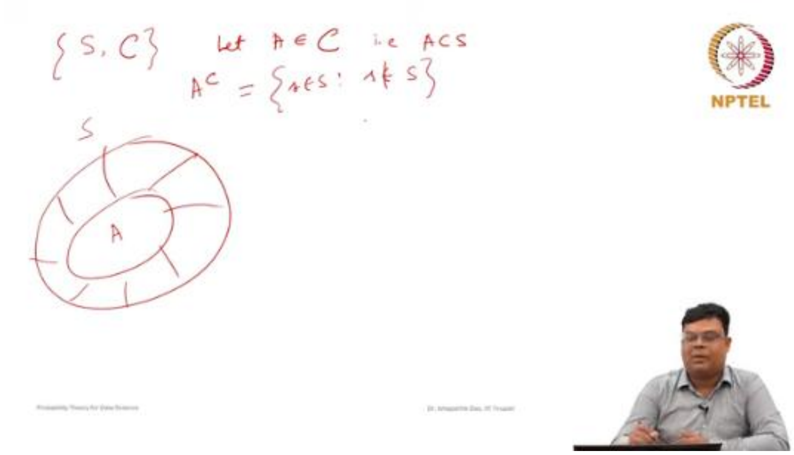




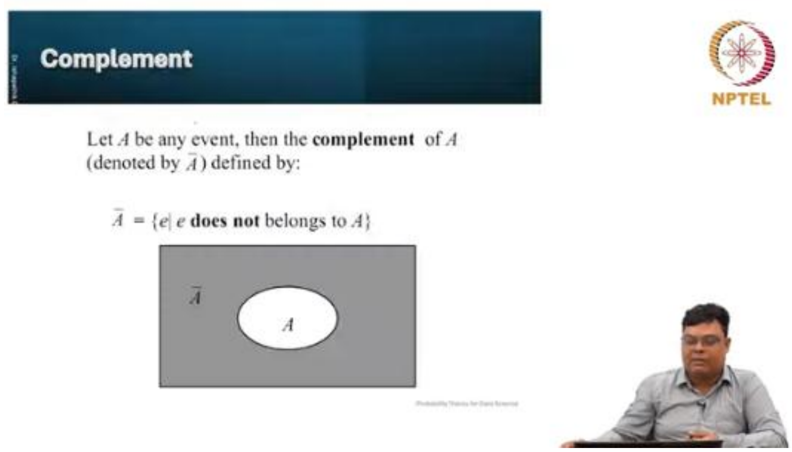
So here, we are denoting S as the sample space, and C is the collection of subsets of S that form a σ-field—that is the event. Let A ∈ C; that means A will be a subset of S. So, what will be the complement? We have already discussed that for a σ-field. So, the complement will be nothing but, basically, it is shown here in the Venn diagram.

So, this is nothing but S, and this is A. All the elements other than A. So, this is A. So, all elements e ∈ S such that e ∉ A. This is nothing but the complement.

So, I assume that you know all these things because, in some basic mathematics courses, you have already learned these set-theoretic concepts. But because it is required here as well, that is why it is repeated in these slides. However, I assume that you already know this. That is why we have already discussed A^c for the discussion of the field and the σ-field. We have already discussed what a complement is.

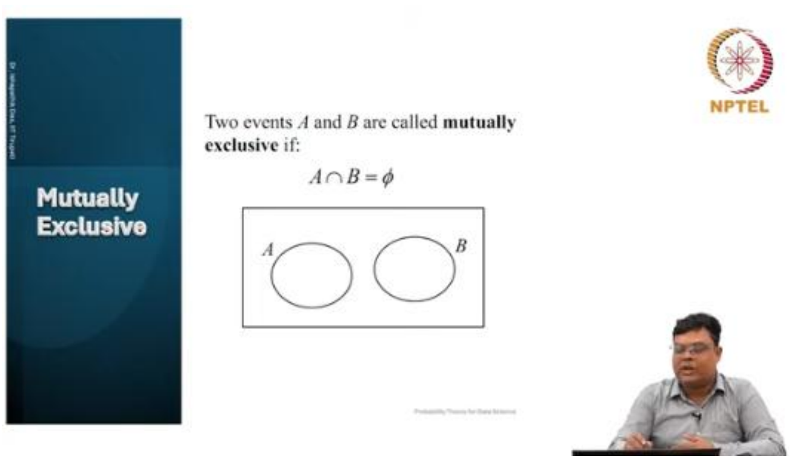


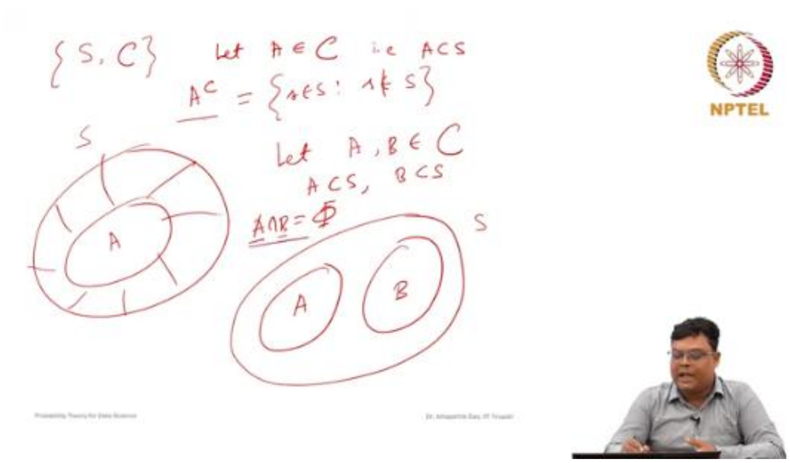
Complement—we have already used this concept. This is A^c. And also, whenever two sets, suppose, belong to C — let A, B ∈ C — that means it is an event. That means A ⊆ S, and B ⊆ S. So, suppose this is S, this is A, and this is B, such that A ∩ B = ∅.



So, what is A ∩ B? We have already discussed. So, all e ∈ S belongs to A and B. So, here you can see that if it is A and it is B, none of the points belong to both A and B. In that case, if A occurs—suppose you are doing the experiment and the element e ∈ A—then B cannot occur.

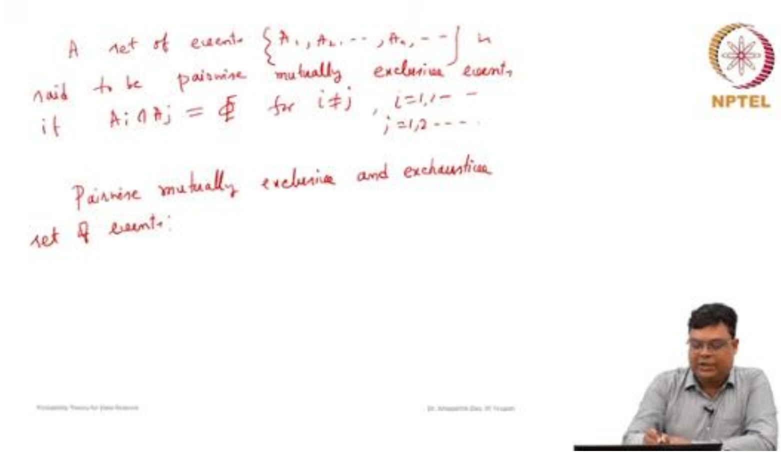
That element cannot be in B because the intersection is ∅. So, if A occurs, then B cannot occur. If B happens, then A cannot happen. So, that is why whenever A ∩ B = ∅, it is known as mutually exclusive event. So, here it is written: A ∩ B.





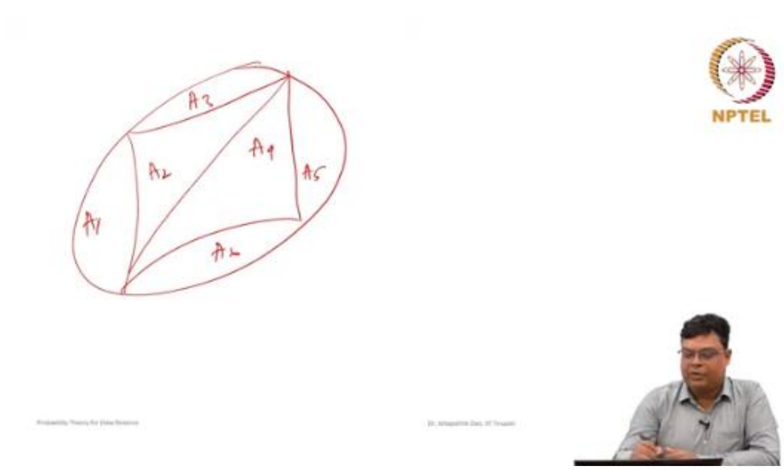
Two events, A and B, are called mutually exclusive if $A \cap B = \emptyset$. So, the extension of this concept is known as a pairwise mutually exclusive set of events. A set of events A_1, A_2, ..., A_n may be a countable collection. It is said to be a pairwise mutually exclusive set of events if $A_i \cap A_j = \emptyset$ for any $i \neq j$, where $i = 1, 2, \ldots$ and $j = 1, 2, \ldots$ This is called the pairwise mutually exclusive set of events.

So, pairwise mutually exclusive and exhaustive sets of events — this is another definition. This is very important; later on, we will discuss this again. So, a pairwise mutually exclusive and exhaustive set of events, such as a set of events A_1, A_2, ..., A_n is said to be pairwise mutually exclusive and exhaustive set of events if $A_i \neq A_j$. There are two conditions because pairwise mutually exclusive means $A_i \cap A_j = \emptyset$ if $i \neq j$, where $i = 1$ to ∞ and $j = 1$ to ∞. Sorry, not n; it will be infinite countably.



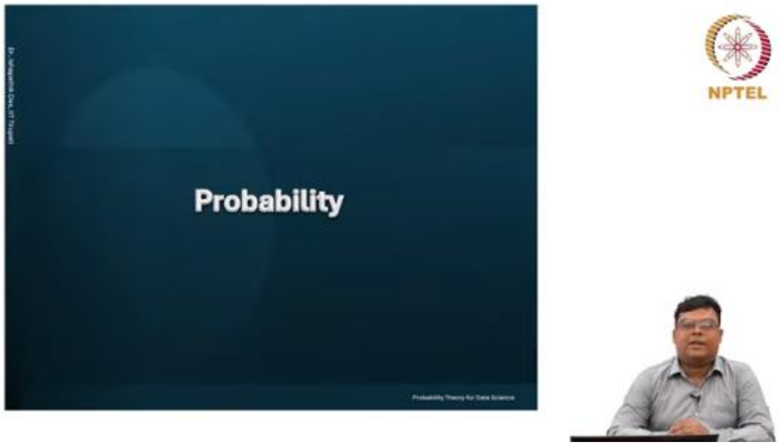
The second condition is that — so the first condition is that they are pairwise mutually exclusive events. The second condition is that if you take the union, $A_1 \cup A_2$, like this, so basically we denote it as shorthand notation: $\cup$ (i = 1 to ∞) A_i. This will be nothing but S. So this is nothing but A; it is like a partition of a set. So, that was S, and you just take a disjoint collection.

So this is, suppose, A_1, A_2, A_3, A_4, A_5, A_6. So they are pairwise mutually exclusive and if you take the union of them, then this will be nothing but S. So, it may be countably infinite or it may be some finite number of sets also, two sets also, just here in general I discussed. So, this may be considered as finite collections also. It is pairwise mutually exclusive and exhaustive set of events.



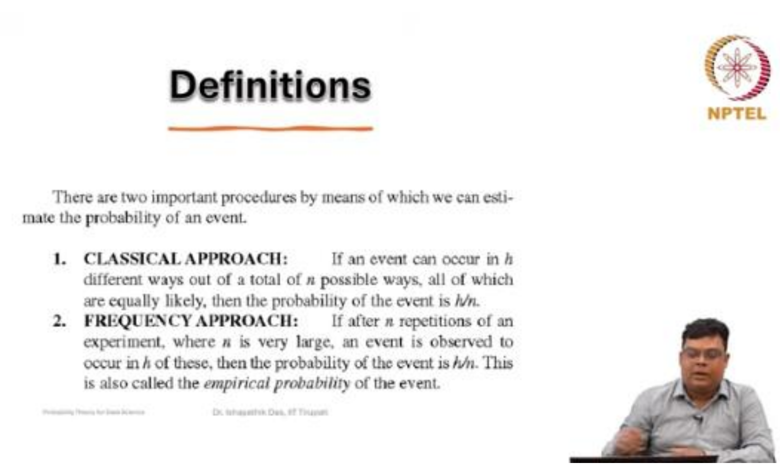
So, whenever it is a finite collection, then we will discuss just this is a set of events A_1, A_2, ..., A_n is said to be pairwise mutually exclusive or exhaustive set of events. If $A_i \neq \emptyset$ and then $\cup$ (i = 1 to n) A_i = S. This is the example actually I gave here. This is a finite collection only. So, this is pairwise mutually exclusive and exhaustive set of events.

So, these are some of the important concepts we discussed, definition of sample space (S) and also definition of events. Whenever we learned it, we understand it clearly. Next, we will discuss what is probability. Now, we are in the position of defining the probability. So, next, let us start the discussion of probability. When scientists started defining probability, there are several approaches came out because some of the approaches for defining probabilities were not widely accepted by all the scientists.



Some of the approaches we will discuss. Finally, the probability will be defined by an axiomatic approach where mathematicians always define it by the axioms. Here at the beginning, some of the approaches we will mention here to understand it how probability it was defined at the beginning. It is known as the classical approach. If some of the assumptions are correct and it is assumed or it can be considered then this definition works well.

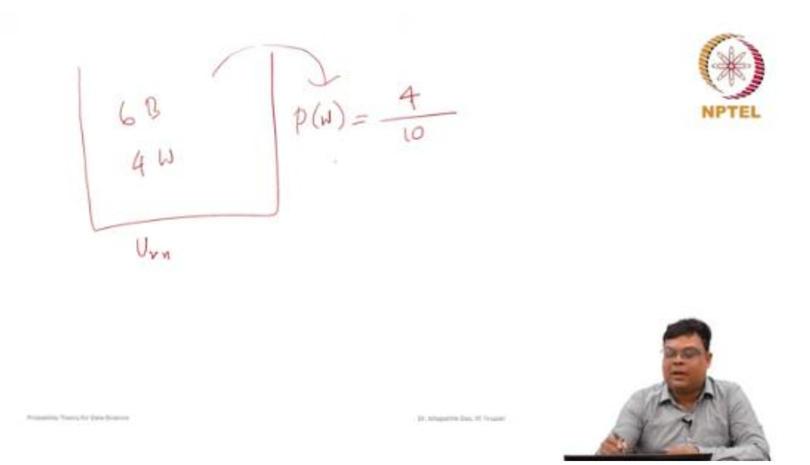
But problem is with those assumptions we will discuss slowly. First let us discuss classical approach. What is classical approach? So this is usually we understand what is probability. Usually, so whenever we want to find a probability like this.



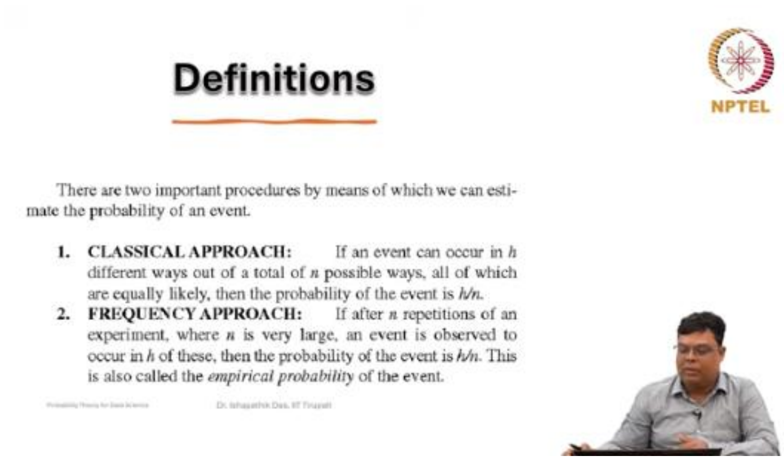
Suppose there is a box or urn, we say. And there are some balls. So, there may be six black balls and four white balls. And this is mixed properly in such a way that the chance of coming out any ball is equal. There is no partiality to one ball, such that that ball can come out more likely than the other balls.

So, that is called equally likely. So, in this approach, it is defined like this: suppose you are drawing a ball. What will be the probability that it will be white? So, classical approach says that if there are n possible ways, those are equally likely to happen. Out of that, the probability we have to consider for the event—that is, the white ball coming out—how many possible ways this event can occur.

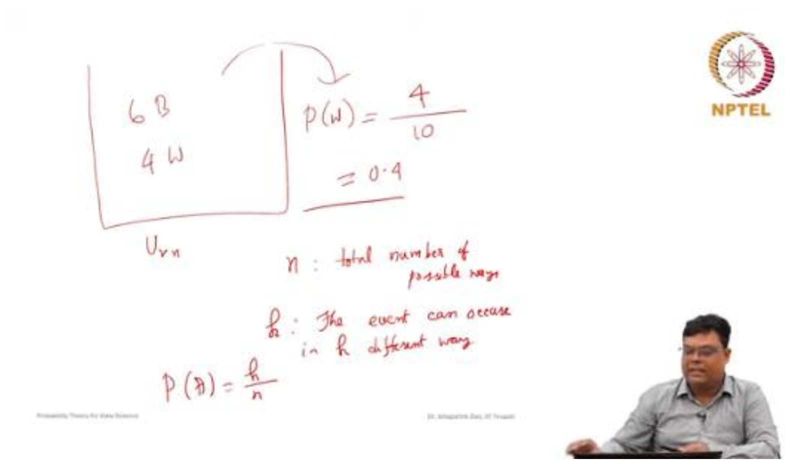
Like, for the white balls, there are four white balls. In total, 10 ways a ball can be drawn. Out of that, four balls can be white. So, four ways that the white ball can come out. Then the probability that the white ball can appear, this is nothing but P(white ball) = 4/10.



So, this is 0.4. This is the classical approach. So, let us see that in the classical approach, if an event can occur in h different ways, out of a total of n possible ways, all of which are equally likely, then the probability of the event is h/n. So, this is n possible ways, n is nothing but it is represented as the total number of possible ways, and h is nothing but the event can occur in h different ways. So, then the probability that this event A is nothing but h/n.



So, here we have already discussed one example that this is 0.4 probability because h = 4, and n is nothing but the total number of possible ways is 10. The problem is that this concept of equally likely. This term, it was not widely accepted by all the scientists. The reason is that when we are defining a probability, the possibility, the probability, equally likely it seems like it is also some name of the probability.



Equally likely, what does it mean? Equally likely means you are saying that equal probable. All the ball can come equal probable. So, that definition itself contain that term. Also, how you know that this is equally likely or not?

If you can assume that, there may be, it is not mixed properly, or some balls are bigger than some balls, smaller. So, in that case, it will not be equally likely, then what you will do? How you will find this probability? So, this equally likely term, It is fake and all scientists could not accept it, this definition. For finite number of balls, it is okay.

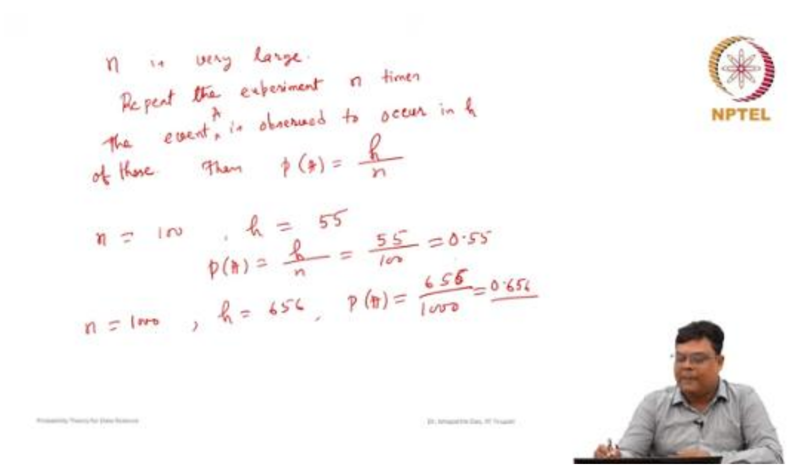
Now, suppose you have a coin. You are tossing a coin and both sides, you do not know if it is equally likely or not. You are tossing the coin 10 times and you see that heads (H) appear more. It will not be exactly 5 and 5. And if you toss the coin 100 times, maybe you are seeing that 80% of the time heads (H) occurring.

So then it is not equally likely. In that case, how can we compute the probability? Then this frequency approach, it came for a frequency approach for defining the probability. The frequency approach says that you repeat the experiment n number of times, where n is very large. So, you repeat the experiment.

So, here it is. You can see that if, after n repetitions of an experiment, where n is very large, an event is observed to occur in h of these, then the probability of the event is h/n. So, n is the—there are two conditions: n is very large, and you repeat the experiment n number of times, n times. The event occurs, so notice that the event is observed to occur in h of these. Whenever n is very large, suppose you denote the event by A.

Then the probability of A is nothing but h/n. For example, you repeat tossing a coin. Suppose n = 100 number of times, and you observe h = 55. So, then your probability of A is nothing but h/n, this is nothing but 55/100, this is 0.55. Suppose you want to repeat your experiment more number of times, 1000.

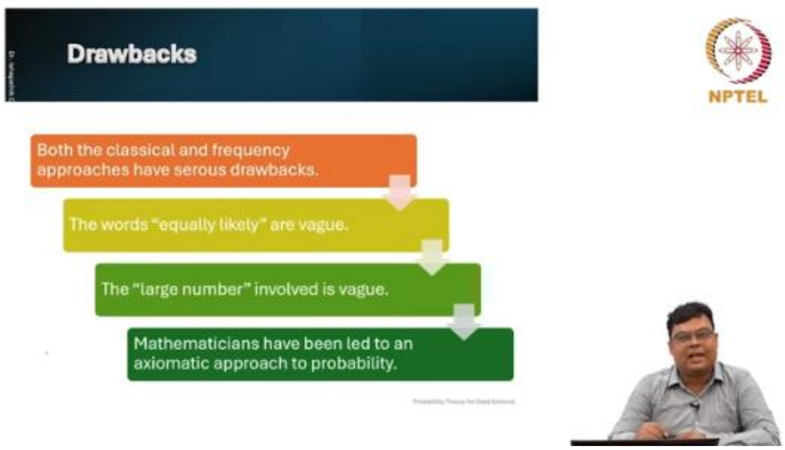
And assume that h = 656. Just I am assuming it. Then your observation is nothing but 1000. This is 0.656. Note that you may not get the 0.55 always.



So then you repeat more number of times. Now the question is, whenever you are saying n is very large, then which n is very large? n = 100 or n = 1000 or n = 5000 or more numbers. So, this n is very large here also, it becomes vague to the scientists. All the scientists could not accept it.

Because when it should be stopped, whether it will be converging or not, we don't know. And in which value we will consider, in which n values it will be considered. Basically, we don't know that which n values it may be considered as very large. That's why this classical approach also, it was not accepted. So, these are the drawbacks that both the classical and frequency approaches have serious drawbacks.

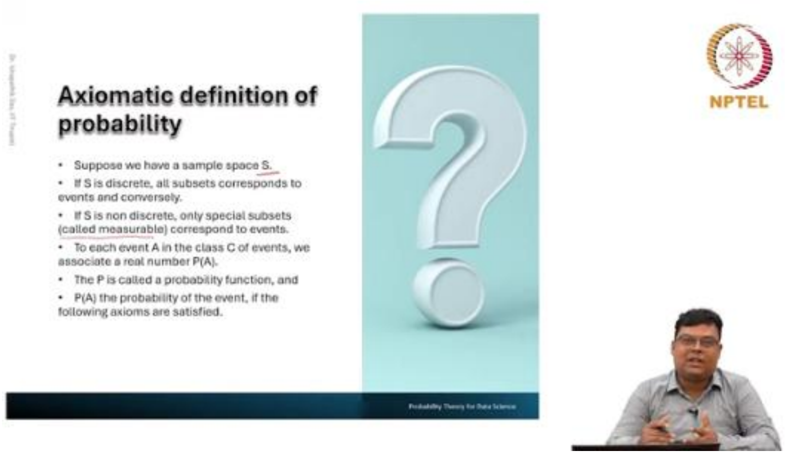
The words equally likely are very vague in the classical approach and the large number, n is very large number also involved is vague. So, mathematicians have been led to an axiomatic approach to probability. So, to defining the probability. Usually, mathematicians define something using an axiomatic approach. The axiomatic approach means we assume a minimum number of rules; if a function satisfies these rules, we define that as probability.



These minimum rules are not proven; we call them axioms. If a function satisfies these axioms, then it is considered a probability. And then some other many other whatever properties if you use by that function we have to prove using the axioms. So, let us discuss axiomatic definition of probability. So, hope you have understood that what is classical approach and what is frequency approach, what are their drawbacks and why you are going to define it the axiomatic approach. So, in the axiomatic approach, suppose we have a sample space S.

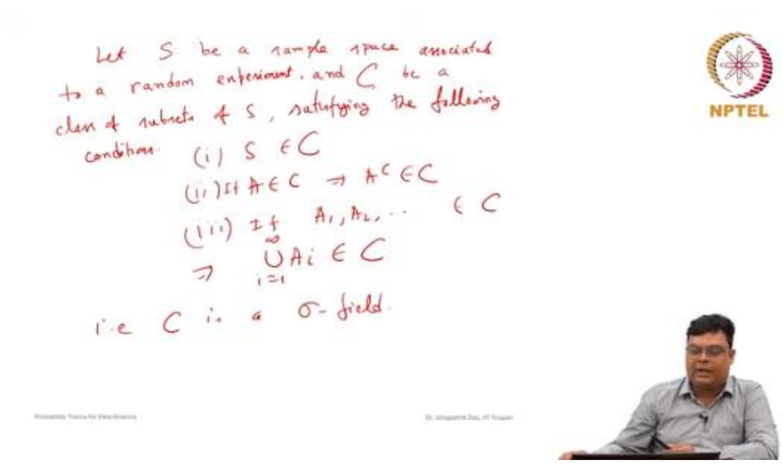
We have a sample space S. If S is discrete, we have already discussed that you can consider all subsets of S as an event. But if S is continuous and non-discrete, then there are some special subsets. In a sigma field, we have to consider that sigma field. Here, it is called a measurable function or measurable sets.

So, all the elements in the sigma field also known as measurable sets. So, it is a part of measure theory. We will not discuss more. We just discuss that sigma field also. So, here we will consider S to be a sample space.



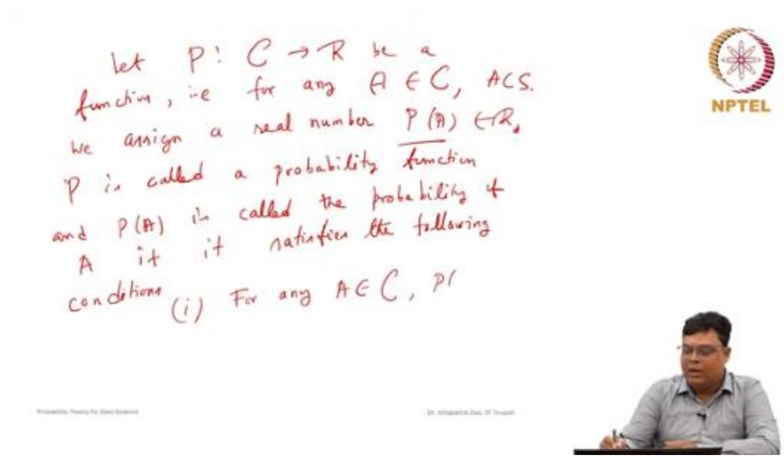
Let S be a sample space associated with a random experiment, and C be a class of subsets of S. C is a class of subsets of S, which is a sigma field. Let us write the class of subsets of S satisfying the following conditions: first of all, $S \in C$. If $A \in C$, it implies that $A^c \in C$. We have already discussed those things.

So, if A_1, A_2, and some countable collection $\in C$, it implies that $\cup A_i \in C$. So, that means C is a sigma field. So, this is the, so we consider here, this is the, in the axiomatic definition. Now, for any $A \in C$, we define a function P. Let $P: C \rightarrow \mathbb{R}$ be a function.

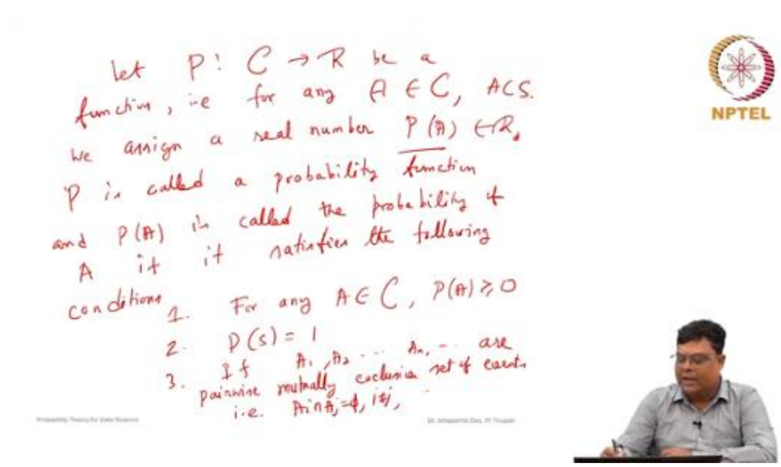


That means for any $A \in C$, basically $A \subseteq S$. We assign a real number P(A) for any $A \in C$. The real number P(A), P is called the probability function, and P(A) is called the probability of the event A if P satisfies the following conditions... First condition is that, so all the things it is written here more clearly. To each A in the class C of events, we associate a real number P(A).

The P is called a probability function, and P(A) is the probability of the event A if the following assumptions are satisfied. So, if the following assumptions are satisfied: first is that, for any $A \in C$, $P(A) \geq 0$. S is always $\in C$. So, for S, $P(S) = 1$. So, let us write 1, 2, 3.



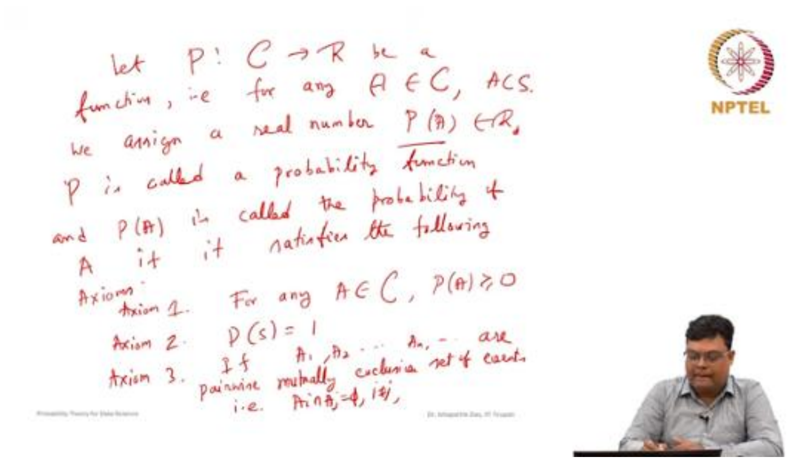
So, this is 1, 2, and 3: If $A_1, A_2, ..., A_n$ are pairwise mutually exclusive sets of events, a countably infinite collection, then it has pairwise mutually exclusive connections. That means $A_i \cap A_j = \emptyset$, for $i \neq j$. Then, whenever $A_1, A_2, ..., A_n$ are pairwise mutually exclusive events, since it is a field, the union from $i = 1$ to ∞ of $A_i = S$.

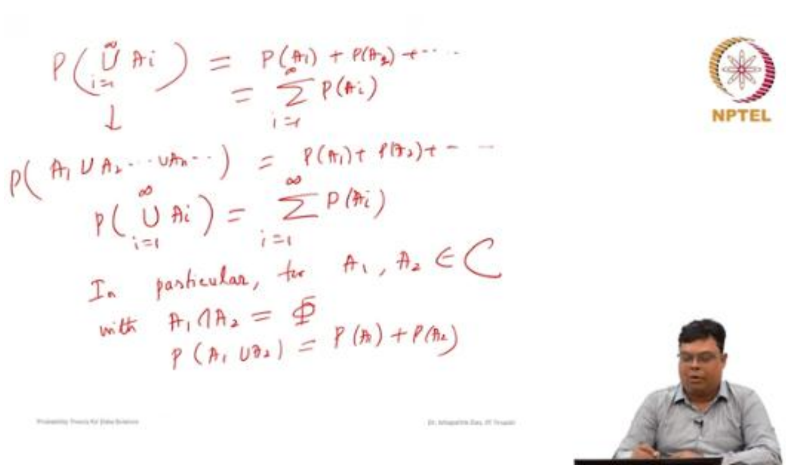


This is an event actually. So, this will belong to C again. So, we can talk about what the probability of this will be. So, this probability will be nothing but $P(A_1)$. So, $P(A_1) + P(A_2)$, like this.

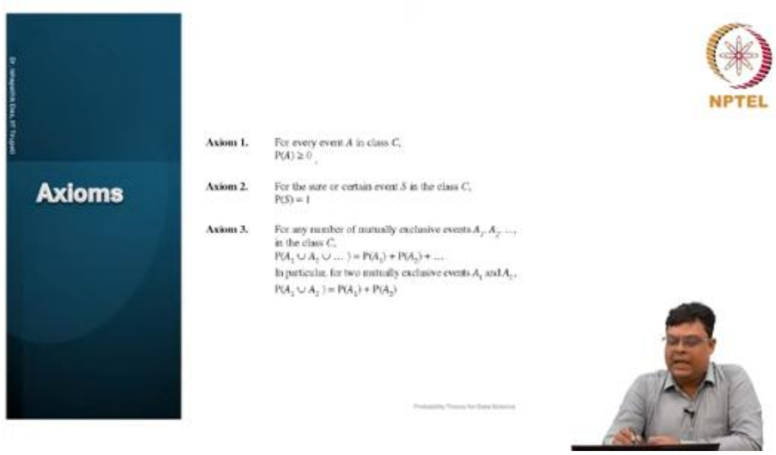
So, in short notation, we write $\sum_{i=1}^{\infty} P(A_i)$. So, here also this union of A_i, we know that $A_1 \cup A_2$, like this. So, the probability of this is equal to $P(A_1) + P(A_2)$, ... So, the probability of the union of A_i from $i = 1$ to ∞ is nothing but the summation $\sum P(A_i)$ whenever they are mutually exclusive. In particular, for two events, suppose $A_1, A_2 \in C$, probability with mutually exclusive with $A_1 \cap A_2 = \emptyset$, then $P(A_1 \cup A_2) = P(A_1) + P(A_2)$.

So this is the third property, third axiom. So this is axiom 1. So if it satisfies the following, suppose let us write axiom. So this is axiom 1, this is axiom 2, and this is axiom 3. So here already it is mentioned.





So axiom 1, axiom 3, it is already mentioned. So that, actually, I have written here slowly so that you can understand clearly. For every element $A \in C$, $P(A) \geq 0$. For the sure or certain event $S \in C$, $P(S) = 1$. For any number of mutually exclusive events $A_1, A_2 \in C$, $P(A_1 \cup A_2 \cup \dots) = P(A_1) + P(A_2) + \dots$.



In particular, for two mutually exclusive events A_1 and A_2, $P(A_1 \cup A_2) = P(A_1) + P(A_2)$. This is the axiom.