PROBABILITY THEORY FOR DATA SCIENCE

Prof. Ishapathik Das

Department of Mathematics and Statistics

Indian Institute of Technology Tirupati

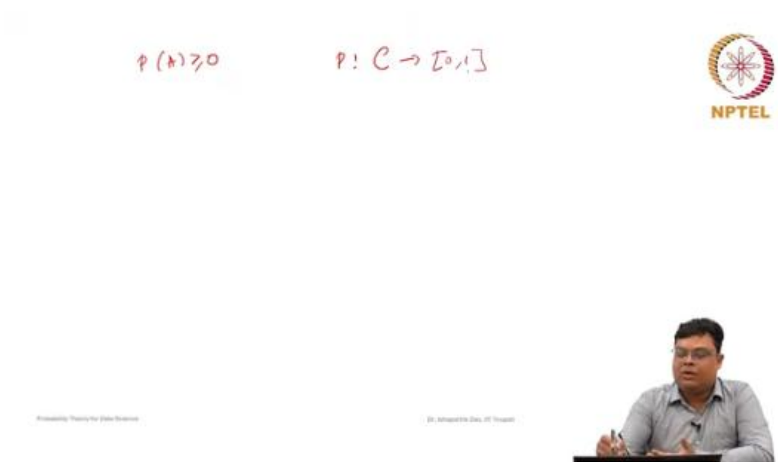
Week - 01

Lecture - 04

Important Theorems

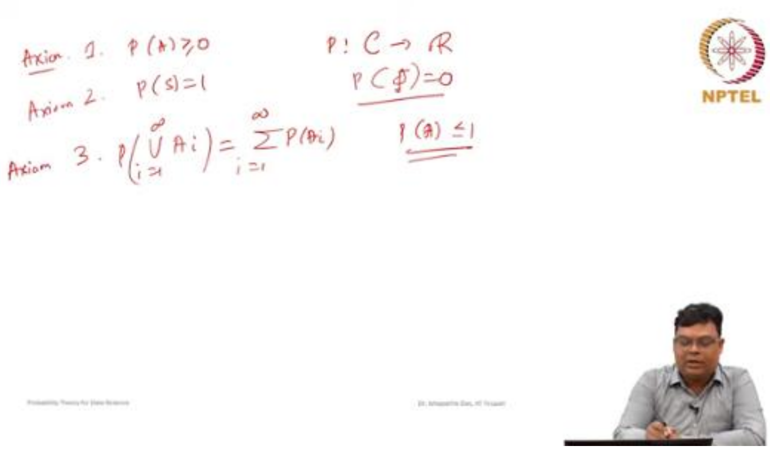
Note that we just discussed that $P(h) \geq 0$. But you know that probability, initially, is defined as a function from C to R. It is a real number. But you know that it is not only a real number; it is actually between 0 and 1 only.

The probability is always between 0 and 1. But initially, we only discussed it as a function from C to R. For any A, the probability of A is a real number. But one by one, we provided the axioms. Axiom 1 states that $P(A) \geq 0$.

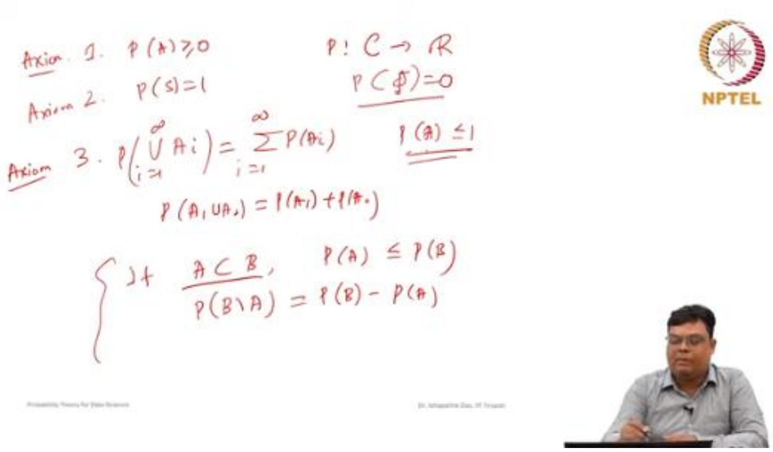


It is a positive real number, including 0, meaning it is ≥ 0. The probability of S is equal to 1. But we know that for the impossible event, $P(\emptyset) = 0$. However, here, we did not explicitly mention it. We stated $P(A) \geq 0$.

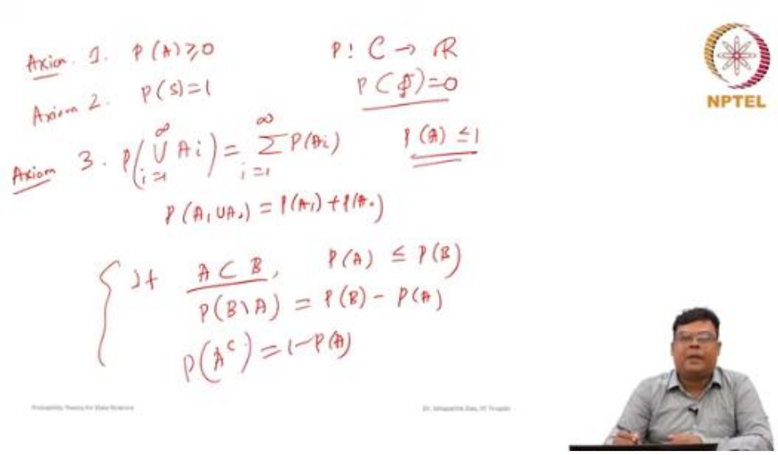
But $P(A) \leq 1$ was not mentioned in the axiom. Also, the third axiom states that the probability of the union ∪ (from i = 1 to ∞) of A_i is equal to the sum Σ (from i = 1 to ∞) of P(A_i). So, these are axiom 1, axiom 2, and axiom 3. In particular, we discussed that it is given as $P(A_1) + P(A_2)$.



So, there are many other things we already use in probability, such as if A ⊆ B, then P(A) ≤ P(B). And we know that it is a kind of notation: all the elements in B that are not in A, or B - A, is nothing but P(B) - P(A), if A ⊆ B. So, those properties include that P(∅) = 0 and P(A) ≤ 1. Those properties we have to prove because they are not included in the axioms. So, we will use those properties frequently in the future because of probability.



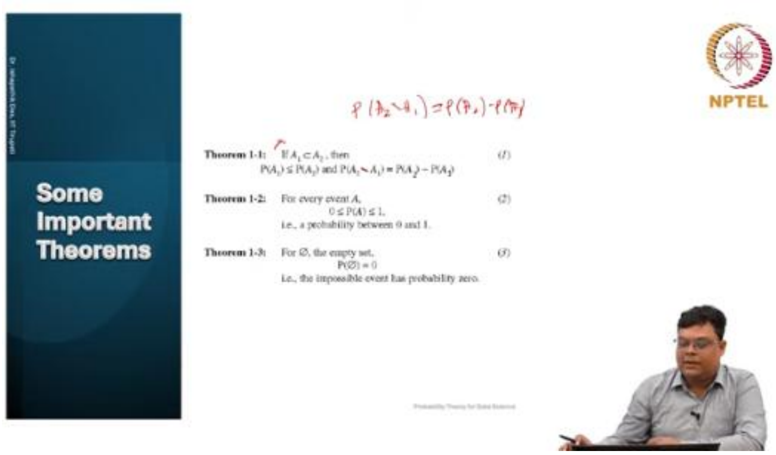
This means that if A ⊆ B, then P(A) ≤ P(B). What about the probability of A^c ? It is nothing but 1 - P(A). So, those basic probabilities you have already learned are frequently used for computing numerical probability problems. So, those results are neither in the axioms nor have we proved them, so we cannot use them because we do not know if they are true or false.



So, we can use it only when we prove it. If it is in the axiom, you can directly use it because it satisfies the axiom, and only then is it considered probability. So, if we say P is a probability, then it must satisfy Axiom 1, Axiom 2, and Axiom 3. For any other property you want to use in the future, we have to prove it. So, we will prove some important theorems here because we will frequently use them in the future.

So, it is important to know that it is truly correct, so we have to prove it first. Let us discuss Theorem 1.1. If $A_1 \subseteq A_2$, then $P(A_1) \leq P(A_2)$. The notation for $A_2 \setminus A_1$ should be written as $A_2 \setminus A_1$, representing all elements in A_2 that are not in A_1. In set theory, this is not like subtracting real numbers, such as $3 - 2 = 1$.

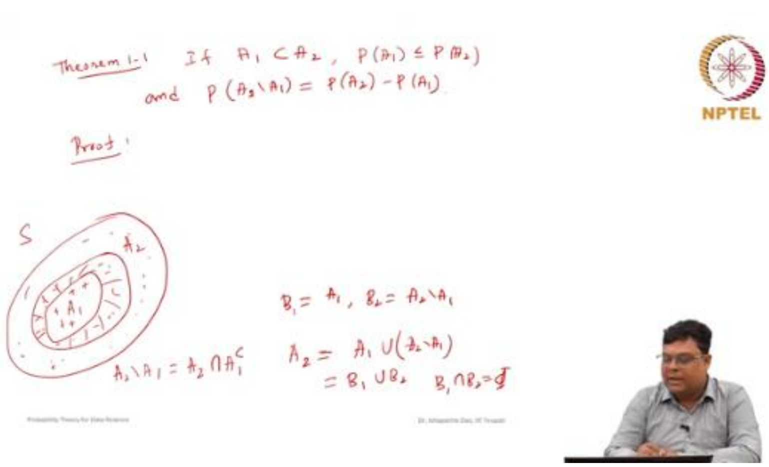
$A_2 \setminus A_1$ means the elements in A_2 that do not belong to A_1. So, this is simply $P(A_2) - P(A_1)$, and it holds under the given condition. So, let us prove this theorem first. So, Theorem 1.1 says that if $A_1 \subseteq A_2$, then $P(A_1) \leq P(A_2)$, and $P(A_2 \setminus A_1) = P(A_2) - P(A_1)$. So, let us prove this.



First of all, let us write down what A2 − A1 is. Suppose this is a sample space S, where A2 ⊆ A1. Now, A2 − A1 means this set. This set can be represented as A2 − A1, which can also be expressed as A2 ∩ $A1^c$. When you intersect A2, the common elements of $A1^c$ will contain all the elements that are in S but do not belong to A1.

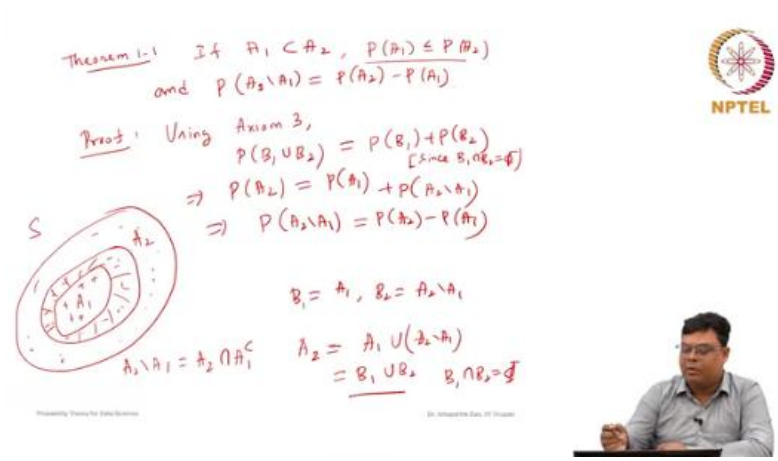
If you take the intersection of A1 and A2, this is nothing but A2 ∩ A1^c. Now, A2 and A2 − A1 can be written as the disjoint union of A1 and A2 − A1. So, the disjoint union represents A1 ∪ (A2 − A1). Suppose this set is B1 and this set is B2. Then B1 ∪ B2 = A2, and not only that, B1 ∩ B2 = ∅.

Here, we consider B1 = A1 and B2 = A2 − A1. For simplicity, we are just using this notation. So, now we can see that B1 ∩ B2 = ∅. Now, just with the Venn diagram, we can understand it very quickly; that is why we wrote it. For set theory, you can show that if S ∈ (A1 ∪ (A2 − A1)), then S ∈ A2.



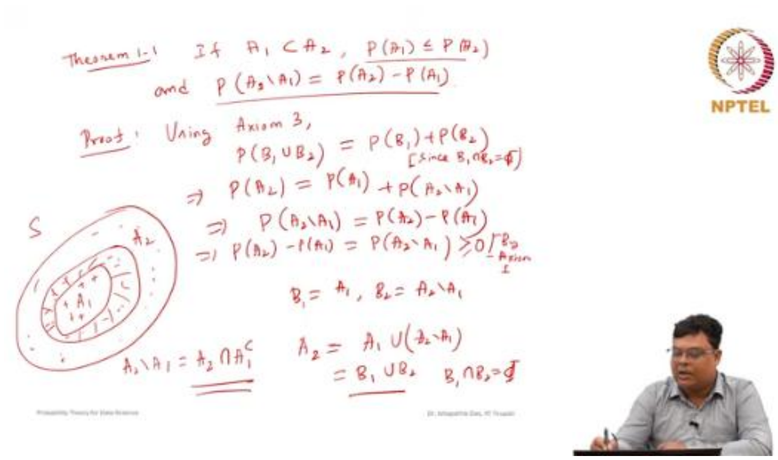
This is also a null set, which you can demonstrate because B1 is simply the set A1, and B2 is all the elements that do not belong to A1 but do belong to A2. That is why these are disjoint. So, now using Axiom 3, we know that because $B1 \cap B2 = \emptyset$, the probability of $B1 \cup B2$ implies that $P(B1 \cup B2) = P(B1) + P(B2)$. Since $B1 \cup B2 = A2$, this implies that $P(A2) = P(B1)$, which is the same as $P(A1)$. Since $B1 \cap B2 = \emptyset$, this is required by Axiom 3.

Now, $P(B1) = P(A1) + P(B2)$, which is $P(A2) - P(A1)$. This implies that $P(A2 \setminus A1) = P(A2) - P(A1)$. So, this is the first result: if $A1 \subseteq A2$, then $P(A1) \leq P(A2)$. This is the theorem: if $A1 \subseteq A2$, then this holds. Now, next, we have to prove that $P(A2 \setminus A1) = P(A2) - P(A1)$.

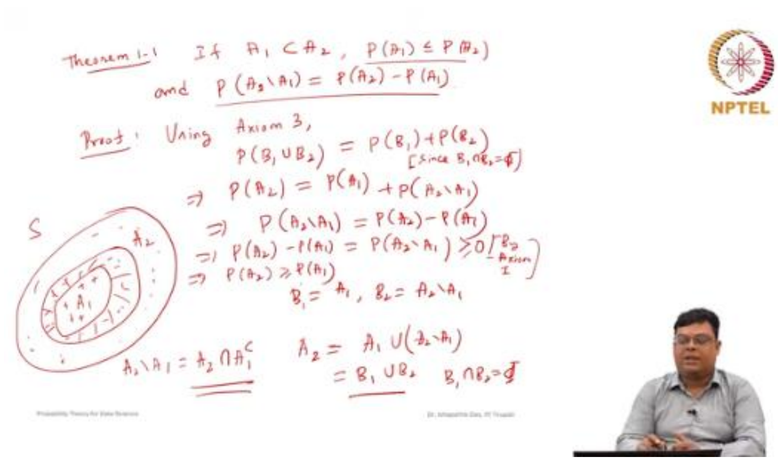


Sorry, we have already proven this result. But we did not prove this result: that $P(A1) \leq P(A2)$. So, now this is the probability of $A2 - P(A1)$. We found that $P(A2 \setminus A1)$. Now, by Axiom 1, $A2 \setminus A1$ is also an event because $A2 \setminus A1$ can be written as $A2 \cap A1^c$. So, it is an event.

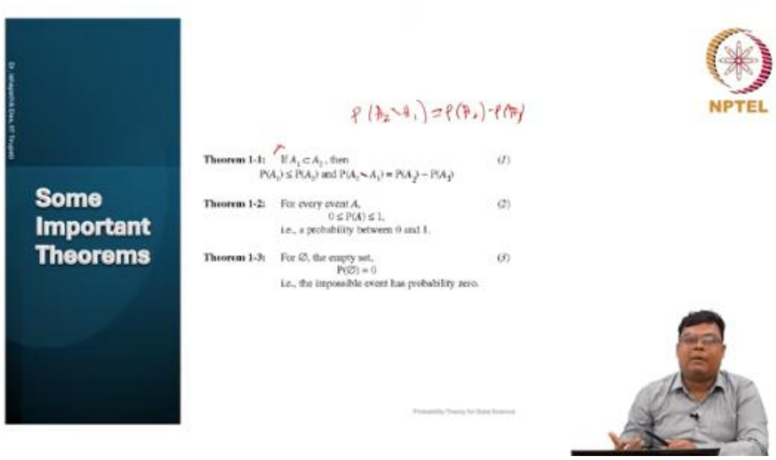
So, this is ≥ 0 by Axiom 1. So, you can see what Axiom 1 says. Axiom 1 says that $P(A) \geq 0$ for any event A. So, A is because all the other set elements are—A2 also belongs to C, and A1 also belongs to C because they are the events we are talking about. $A2 \cap A1^c$ will also be in C, so it will be an event.



So that is why this will be ≥ 0. So, from that, because $P(A2) - P(A1) \geq 0$, we say that $P(A2) \geq P(A1)$. So, this is the proof of Theorem 1. This proof is not very complicated or difficult, but still, we need to prove it because, in the future, we will write it down and use this theorem, so we need to know that we will frequently use it. $A \subseteq B$, so $P(A) \leq P(B)$.



The probability of $B - A = P(B) - P(A)$. Whenever $A \subseteq B$, we can use this result. If $A \not\subset B$, we cannot use this result. So, that is why it is important. Now, Theorem 1.2 says that for every event A, $0 \leq P(A) \leq 1$.

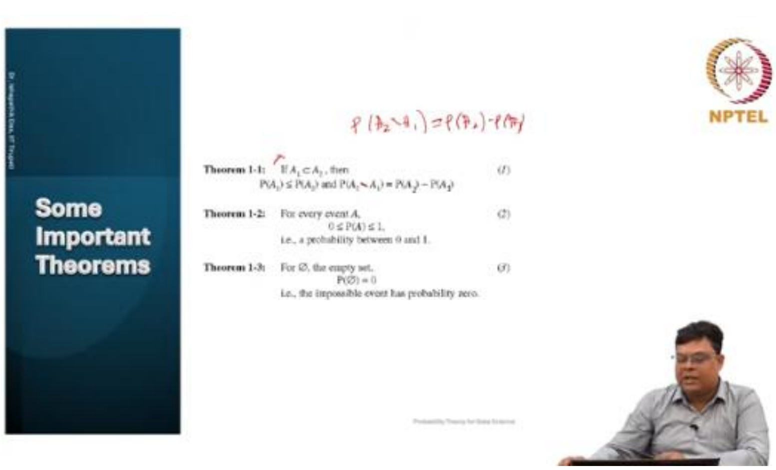


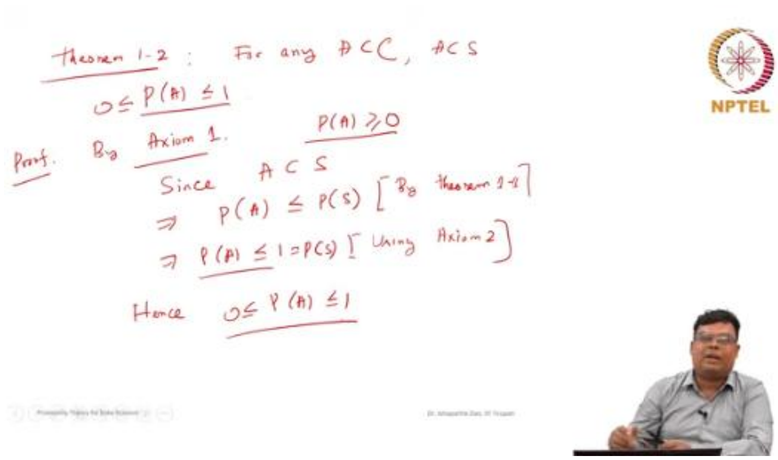
In Axiom 1, we already know that P(A) ≥ 0. However, we do not yet know that P(A) ≤ 1; that is what we have to prove. So, let us prove it here for Theorem 1.2. For any A ∈ C, which means A ⊆ S—any event, actually—the probability of A is ≤ 1 and ≥ 0. By Axiom 1, P(A) ≥ 0.

Now we have to prove that P(A) ≤ 1. How can we prove that? Since A ⊆ S, this implies that P(A) ≤ P(S). We have already proved this, so we can use it here by Theorem 1.1. This implies that P(A) ≤ 1, since P(S) = 1.

Using Axiom 2, which states that P(S) = 1, we can conclude that combining these two results gives us P(A) ≤ 1 and P(A) ≥ 0. So, these are very straightforward proofs.

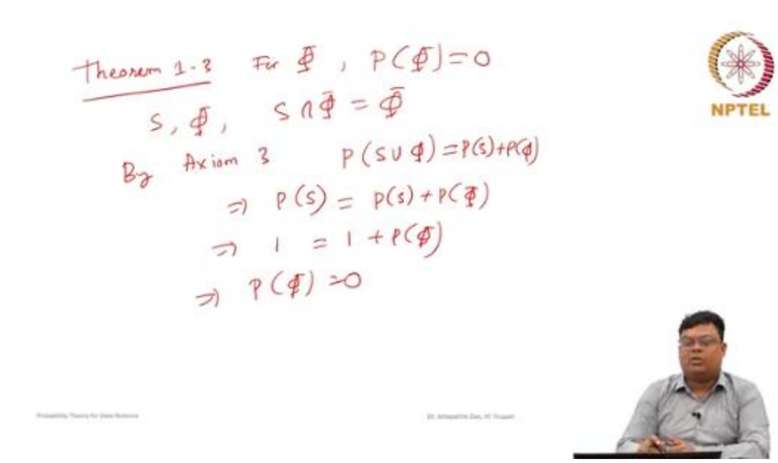
The next theorem states Theorem 1.3: for the empty set, P(∅) = 0. This is not mentioned in the axioms, so we have to prove it. So, Theorem 1.3: for ∅, since ∅ ∈ C (it is an event), P(∅) = 0.





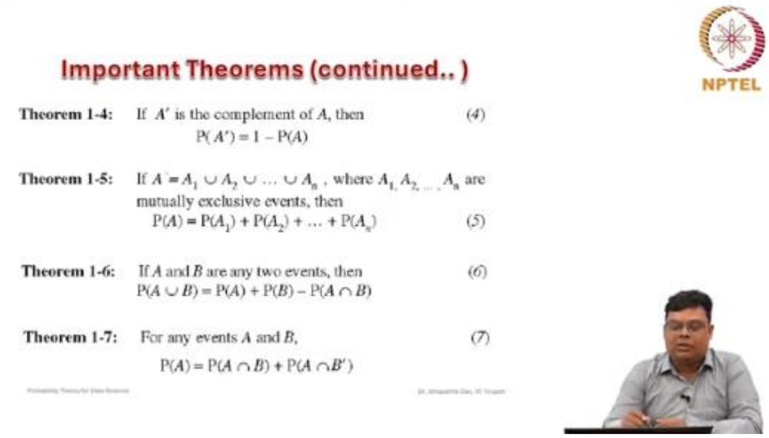
So, $P(\emptyset) = 0$. $\emptyset \in C$. Now, because $S^c \in C$, $P(\emptyset) = 0$. How can we prove that? So now, if you consider S and $\emptyset$ as two events, $S \cap \emptyset = \emptyset$.

From Axiom 3, $P(S \cup \emptyset) = P(S) + P(\emptyset)$. Now, on the left side, it is nothing but S, so this is equal to $P(S) + P(\emptyset)$. The probability of S, we already know, is 1. This is $1 + P(\emptyset)$, which implies $P(\emptyset) = 0$. So, these are very straightforward, but we want to prove them because, without proof, we may not be able to use them in the future.

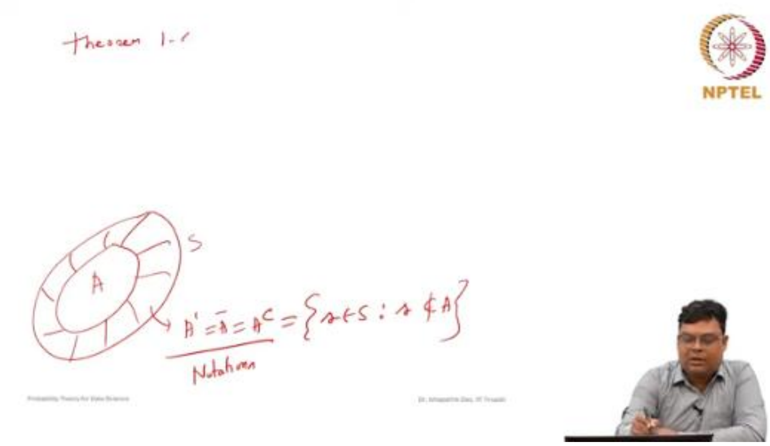


Whenever we use this, we need to have a proof. So now, the next theorem, Theorem 1.4, states that for any event A, the complement of A, A^c, is also an event, and $P(A^c)$ is defined. So sometimes A^c is nothing but if S is the sample space, then $A^c = S \setminus A$.

Now, A^c includes all the elements that do not belong to A. This is A^c.

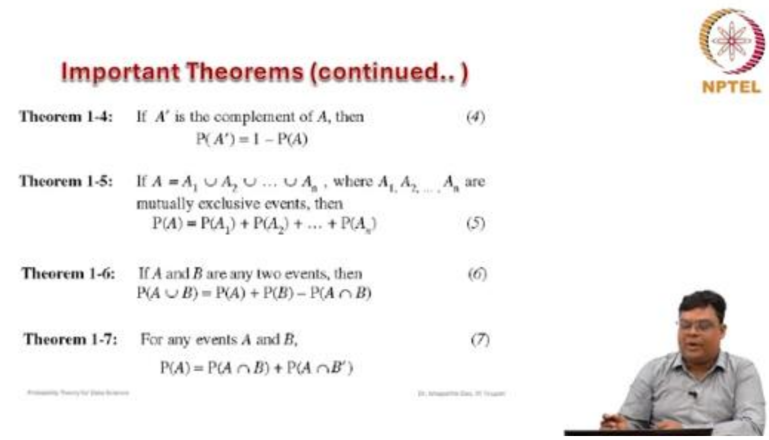


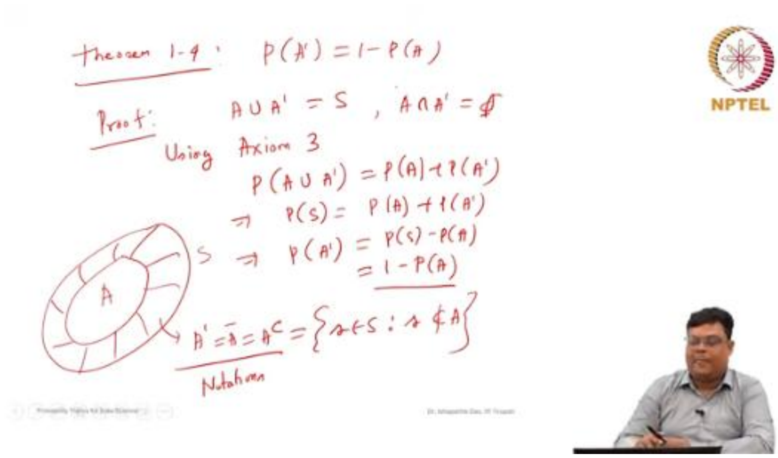
We sometimes denote A complement with A', $\bar{A}$, or simply A^c; it's just a notation. This is defined as all elements in S such that $S \notin A$. Now, suppose we use A'. It is Theorem 1.4 that says P(A') or $P(A^c)$ is nothing but $1 - P(A)$. So, how do we prove that?



So, we can use this relation because $A \cup A^c$ is nothing but S. Also, $A \cap A^c = \emptyset$. So, using Axiom 3, $P(A \cup A^c) = P(A) + P(A^c)$ by Axiom 3 because they are disjoint, which implies that $P(A \cup A^c) = P(S)$. This is nothing but $P(A) + P(A^c)$, which implies that $P(A^c) = P(S) - P(A)$. Since $P(S) = 1$, we get $P(A^c) = 1 - P(A)$. So, actually, it is a typo.

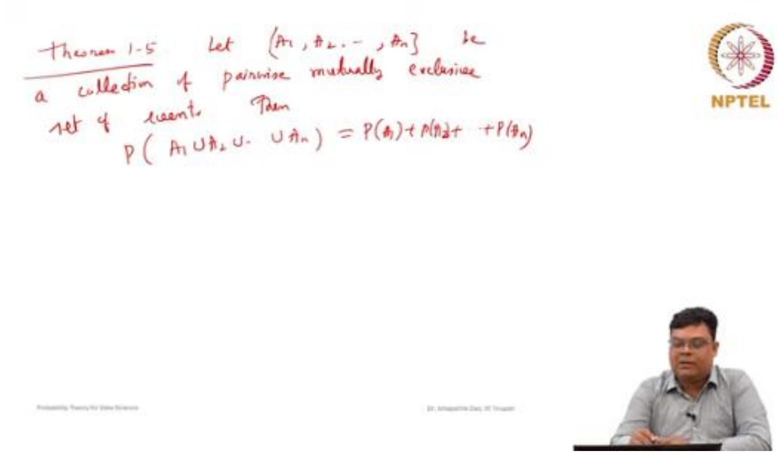
Sorry, it is a mistake. $P(A^c) = 1 - P(A)$. So, this is Theorem 1.4. Now, let's move on to Theorem 1.5. So, Theorem 1.5 says that for a finite number of mutually exclusive events, suppose $A_1, A_2, ..., A_n$ are pairwise mutually exclusive events.





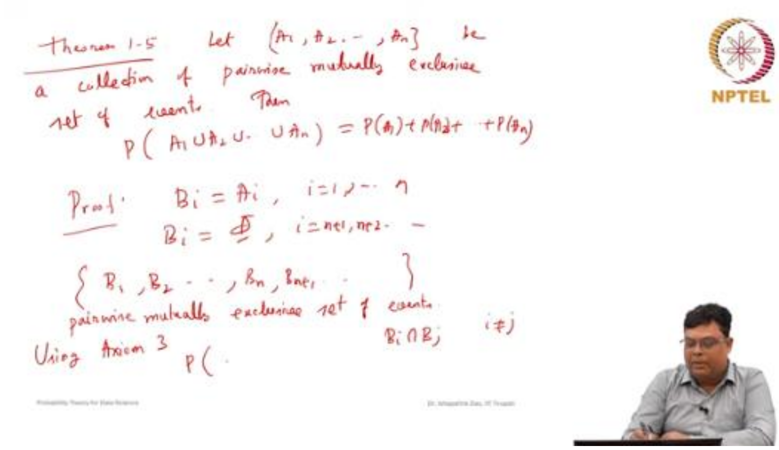
We can write that $P(A_1 \cup A_2 \cup \ldots \cup A_n) = P(A_1) + P(A_2) + \ldots + P(A_n)$. It seems that this is very obvious from Axiom 3. Axiom 3 already mentions it for an infinite number of events, but it also covers two events. Therefore, we need to prove it for the finite case as well. So, first, let us prove this.

So, Theorem 1.5 says that let A_1, A_2, A_n be a collection of pairwise mutually exclusive sets of events. Then, $P(A_1 \cup A_2 \cup \ldots \cup A_n) = P(A_1) + P(A_2) + \ldots + P(A_n)$ is for any finite case. Suppose it is infinite; it is given. So then, we have to prove that any finite union will also satisfy because they are mutually exclusive.



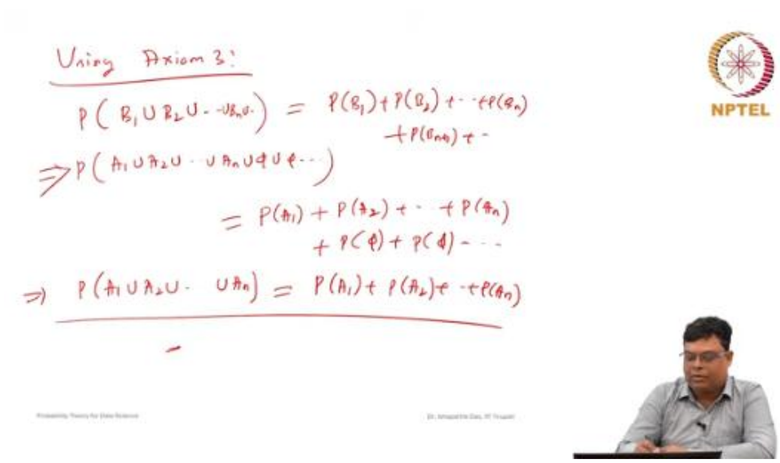
Let us consider $B_i = A_i$ for i = 1 to n, and $B_i = \emptyset$ for i = n+1, n+2, and so on. Hence, we have an infinite collection, and they are pairwise mutually exclusive as well. So, this is actually a countably infinite collection of pairwise mutually exclusive sets of events. They are pairwise mutually exclusive because if you consider any $i \neq j$, then the intersection of B_i and B_j will be $\emptyset$. If i and j are both $\leq n$, since A_1, A_2, A_n are pairwise mutually exclusive, the intersection of A_i and A_j will be $\emptyset$ as well.

For any i or j that is greater than n, one of B_i or B_j will be $\emptyset$, leading to an intersection that is also $\emptyset$. Thus, they are pairwise mutually exclusive events. Hence, using Axiom 3, $P(B_1 \cup B_2 \cup ... \cup B_n) = \Sigma P(B_i)$ for i = 1 to n. So, we can write it like this: $P(B_1 \cup B_2 \cup ... \cup B_n \cup B_{n+1} \cup ...) = \Sigma P(B_i)$ for i = 1 to ∞.



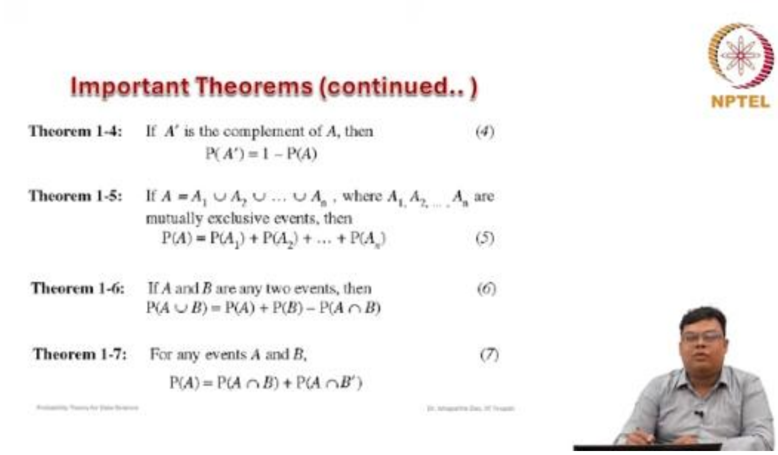
Note that, up to B_1, B_2, and B_n, it is exactly the same as A_1, A_2, and A_n. For any index greater than n, B_i is equal to the empty set ($\emptyset$), which implies that this is equal to $P(A_1)$. Since we have an infinite collection of B_1, B_2, ..., B_n, we can apply Axiom 3 here. By Axiom 3, the probability of the union of B_1, B_2, ..., B_n is equal to the sum of the probabilities of the individual events, i.e., $P(B_1 \cup B_2 \cup ... \cup B_n) = \Sigma P(B_i)$ for i = 1 to n. Now, if we replace the representation of B_i with A_i, we get: $P(A_1 \cup A_2 \cup ... \cup A_n \cup B_{n+1} \cup ...) = P(A_1) + P(A_2) + ... + P(A_n) + P(B_{n+1}) + ...$ Since $P(\emptyset) = 0$, we can conclude that $P(A_1 \cup A_2 \cup ... \cup A_n) = P(A_1) + P(A_2) + ... + P(A_n)$.

Union with $\emptyset$ will yield the same set, and on the right-hand side, it will be $P(A_1) + P(A_2) + P(A_n)$, with the remaining terms being all zeros $(P(\emptyset) + P(\emptyset) + P(\emptyset))$. That is why we have proved this result for the finite case as well. So, this is proved. Now, this is why we can say that in Axiom 3, we have used it for two events. It is not required for Axiom 3 to apply only to infinite events; we can also apply it here.



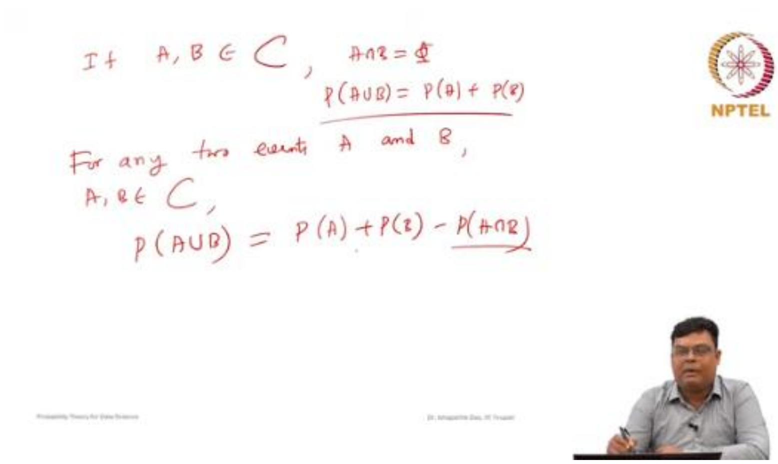
So, we have proved this finite case. Thus, we can conclude that it is true for finite cases as well. Here, we have written this specifically for two finite disjoint sets, stating that $P(A_1 \cup A_2) = P(A_1) + P(A_2)$. This has been established in Theorem 1.5. The next important result is that for any two events A and B.

So, A and B are disjoint. Suppose A and B are two events; that means they belong to C, such that $A \cap B = \emptyset$. Then, by Axiom 3, also in Theorem 1.5, this applies to finite cases as well. $P(A \cup B) = P(A) + P(B)$ because they are disjoint. Now, if they are not disjoint for any two events, A and B, which may or may not be disjoint, then any two events, A and B, belong to C, this collection.



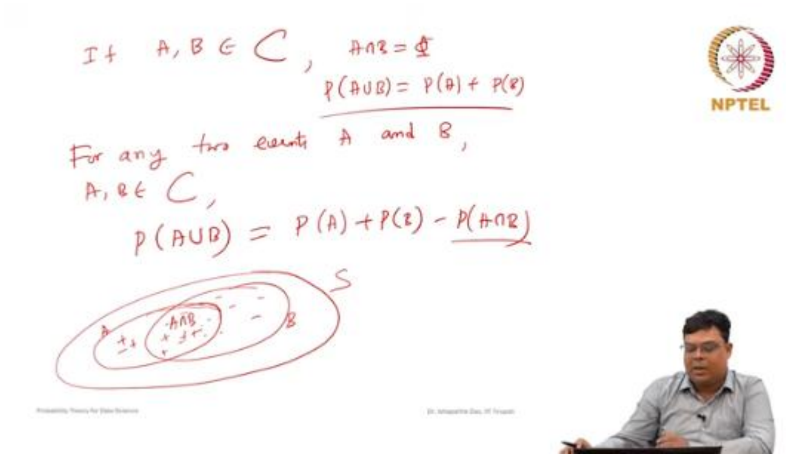
The probability of A ∪ B is such that, because it is a σ-field, A ∪ B will also be in the σ-field. Therefore, it will be considered an event. This is nothing but P(A) + P(B) - P(A ∩ B). This is the theorem. This is one of the important theorems.

You know about the addition of two events. They may be disjoint. If they are disjoint, then P(A ∩ B) = ∅, and P(∅) = 0. So, in that case, this will be true: P(A ∪ B) = P(A) + P(B).



But if we consider general cases, how can we prove it? Suppose we use a Venn diagram. Let’s say this is set A and this is set B. This is the sample space S; this is set A, and this is set B. This set represents A ∩ B. Now, if you add the probability of A, which is P(A), and then add the probability of B, you can see that A ∩ B is counted twice.

Therefore, we need to subtract one instance of A ∩ B to get the probability of A ∪ B. Now, we will also prove it analytically. In a Venn diagram, it may not be exact; we may not get complete clarity. It is simply a graphical representation. How do we prove it analytically?



Let's again take the help of this graph. Suppose this is S, this is A, this is B, and this is $A \cap B$. Now, if you want to represent $A \cup B$, let me clarify. This is S, this is A, and this is B. So, we have S as the sample space, A as one event, B as another event, and $A \cap B$ as the overlap between A and B. Suppose you want to find $A \cup B$; we want to represent it in three different disjoint sets, the union of three different sets.

One set is like this, another set is like this, and the third set is this. The first set can be represented as $A - (A \cap B)$. The second set is $A \cap B$, and the third set is $B - (A \cap B)$. We denote these sets as C_1, C_2, and C_3. Thus, $A \cup B$ can be represented as $C_1 \cup C_2 \cup C_3$, where $C_1 = A - (A \cap B)$, $C_2 = A \cap B$, and $C_3 = B - (A \cap B)$.

Now, $C_1 = A - (A \cap B)$, $C_2 = A \cap B$, and $C_3 = B - (A \cap B)$. You can easily show that C_1, C_2, and C_3 are mutually exclusive events because C_1 is defined by all elements in A that are not in $A \cap B$, while C_2 is exactly $A \cap B$. Therefore, these sets are disjoint. Similarly, C_2 and C_3 will also be disjoint, as will C_1 and C_3. This is because the common part is $A \cap B$.

We are considering all elements in C_1 that are not in $A \cap B$, while C_3 includes all elements in B that are not in $A \cap B$. Therefore, C_1 and C_3 will also be disjoint, and this can be shown easily. Additionally, $C_1 \cup C_2 \cup C_3$ can be shown to equal $A \cup B$. Thus, these sets are disjoint, meaning $C_i \cap C_j = \emptyset$ for $i \neq j$. So, using Theorem 1.5, which applies to finite disjoint, pairwise mutually exclusive events, we can have this result: $P(A \cup B) = P(C_1 \cup C_2 \cup C_3)$.

By using Theorem 1.5, this is the probability of C_1 plus the probability of C_2 plus the probability of C_3. Now, what is the probability of C_1? The probability of C_1 is nothing but $P(A - (A \cap B))$, plus the probability of C_2, which is $P(A \cap B)$, plus the probability of C_3, which is $P(B - (A \cap B))$. So, now we apply Theorem 1.1. Since $A \cap B \subseteq A$, it is nothing but $P(A - (A \cap B)) = P(A) - P(A \cap B)$.

This equals $P(A \cap B) + P(B) - P(A \cap B)$. This is using Theorem 1.1. Now, we simplify and the $P(A \cap B)$ terms cancel out. Thus, we finally have $P(A) + P(B) - P(A \cap B)$. This is one of the important results.

So, there are many applications for finding the probability. Suppose the probabilities of A and B, as well as the probability of $A \cap B$, are known. Then, we can find the probability of $A \cup B$.