PROBABILITY THEORY FOR DATA SCIENCE

Prof. Ishapathik Das

Department of Mathematics and Statistics

Indian Institute of Technology Tirupati

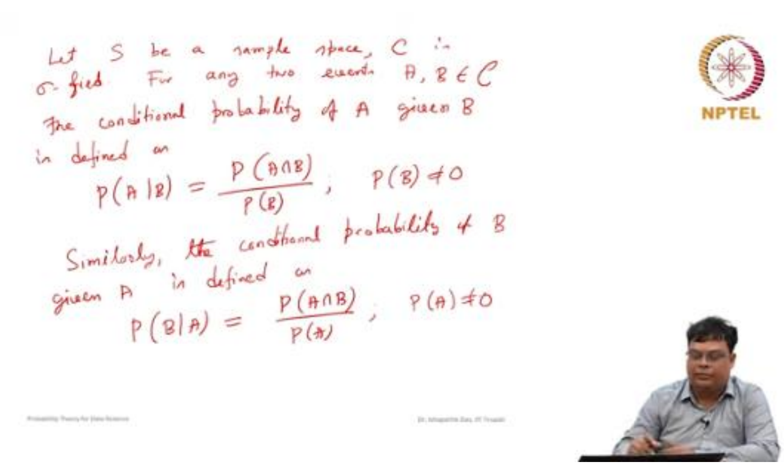
Week - 01

Lecture - 06

Definition of Conditional Probability and Independence

Now, here is the definition. We will formally define conditional probability. Suppose we are interested in computing the probability of an event A, and we have been told that event B has occurred. The conditional probability of A given B is defined as follows: Let S be a sample space, and C be a σ-field.

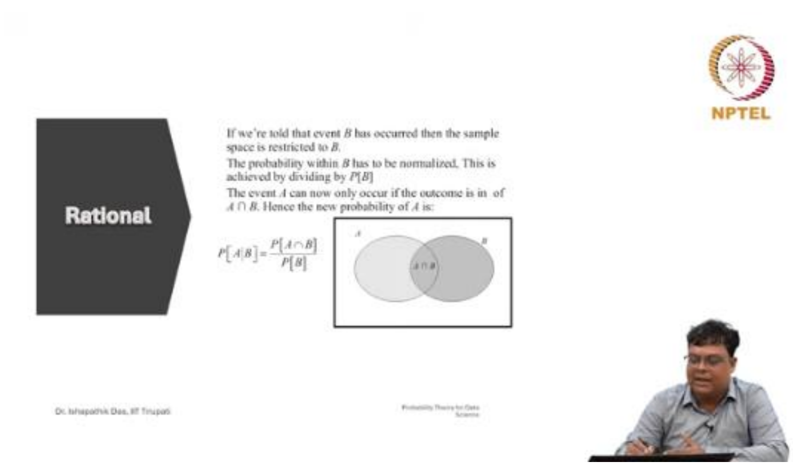
For any two events A and B, since they belong to the σ-field C, the conditional probability of A given B is defined as $P(A \mid B) = P(A \cap B) / P(B)$, assuming that $P(B) \neq 0$. So, if $P(B) = 0$, then we can't define this conditional probability. It will be undefined. Similarly, the conditional probability of A given B, which we've already defined, and the conditional probability of B given A is defined as $P(B \mid A) = P(A \cap B) / P(A)$, where it's assumed that $P(A) \neq 0$.



So, if you are told that event B has occurred, then the sample space is restricted to B. Like the example we discussed, whenever you throw a die and it is informed that an even number has occurred, we need not consider the sample space as {1, 2, 3, 4, 5, 6}. Since we know that only even numbers are the possibilities, the sample space now becomes restricted to B: the even numbers {2, 4, 6}. The probability within B has to be normalized. This is achieved by dividing by the probability of B.

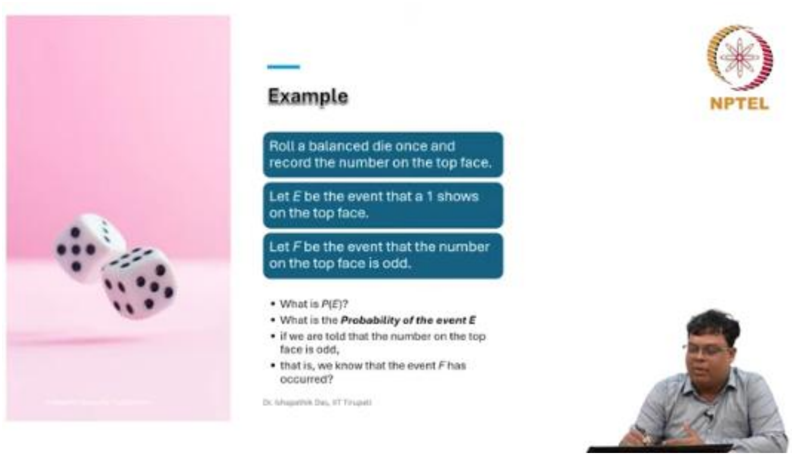
So, in this example, suppose this is the graph where we can see the sample space, A, and B. Whenever B is observed, that means those sample points inside B are the possibilities. Now, what is the proportion of A out of B? The probability of A given B is $P(A|B) = P(A \cap B) / P(B)$. To understand this, event A can now occur if the outcome is in $A \cap B$.

Hence, the new probability of A is $P(A|B) = P(A \cap B) / P(B)$. So, let us do some numerical examples. Suppose that A out of B is the probability now for A given B, which is $P(A|B) = P(A \cap B) / P(B)$. So, this is something to understand: the event A can now occur if the outcome is in $A \cap B$. Hence, the new probability of A is $P(A|B) = P(A \cap B) / P(B)$.



So, let us do some numerical examples. Suppose we roll a balanced die once and record the number on the top face. So, what is a balanced die? A balanced die means we assume that all the faces are equally likely. The probability that one of the observations may come is 1/6.

Let E be the event that 1 shows on the top face. So, E = {1}, and let F be the event that the number on the top face is odd. So, F = {1, 3, 5}. These are the odd numbers. From the classical approach, what is the probability of E? Since all are equally likely, the sample space S = {1, 2, 3, 4, 5, 6}.



Let F be the event that the number on the top face is odd. So, F = {1, 3, 5}. These are the numbers that are odd. So, from the classical approach, what is the probability of E? Because all are equally likely, the sample space S = {1, 2, 3, 4, 5, 6}.

So, the probability of E will be 1/6. The probability of F, because it contains 3 equally likely points out of 6 equally likely points by the classical approach, means the number of ways F can occur is 3 out of the total number of ways, which is 6. This is 3/6 = 1/2.

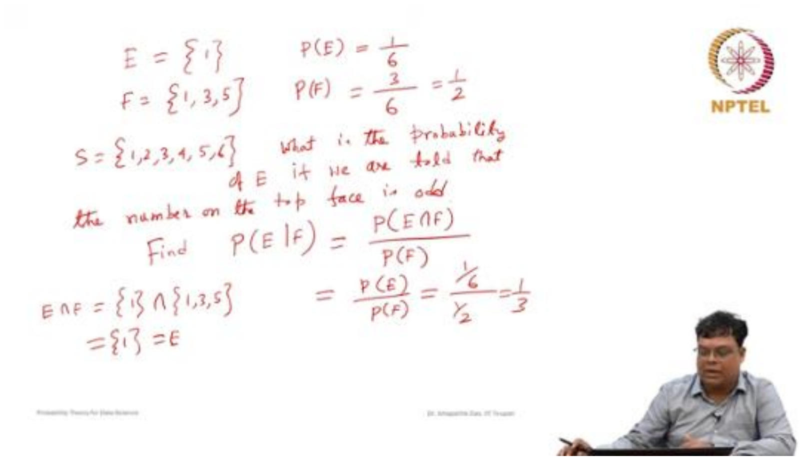
Now, this is the question. What is the probability of E? We have already found that it is 1/6. What is the probability of the event E if you are told that the number on the top face is odd?

So, the question is: what is the probability of E, given F? If we are told that the number on the top face is odd, the sample space is now restricted to F = {1, 3, 5}. Out of these three, only 1 happens, so this probability, from intuition, should be 1/3.

Now, we will apply the definition. The definition says that this probability is P(E | F) = P(E ∩ F) / P(F). So, what is E ∩ F? The probability of F, we know, is 1/2. Now, E ∩ F is nothing but {1} ∩ {1, 3, 5}, which is {1}. This is just E, so we have:

$P(E \mid F) = P(E \cap F) / P(F) = P(E) / P(F) = (1/6) / (1/2) = 1/3.$

This also matches our intuition, where 1 out of 3 possible outcomes can occur. So, it's consistent with both intuition and the definition. This definition is consistent with our intuition and what we understood. Now, we will discuss the Prisoner's Dilemma. So, the Prisoner's Dilemma says that A, B, and C are in jail, and one will be executed.



The probability that A will be executed is 1/3, and the complement is 1 − 1/3 = 2/3. The probability that B will be executed is also 1/3, and the probability of B^c (B not being executed) is 2/3. The probability that C will be executed is 1/3, and the probability of C^c (C not being executed) is 2/3, which we know.

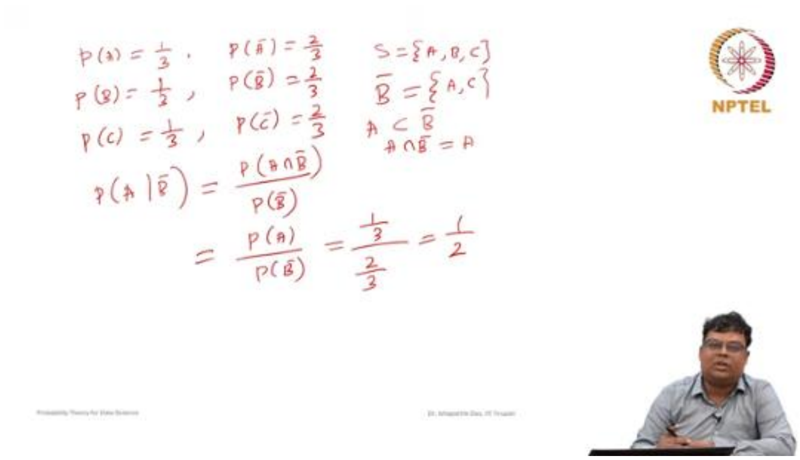
Now, we have to find the conditional probability. Suppose the guard says that B^c (B will not be executed). Now, what is the conditional probability of A given B^c? So, this probability will be the probability of $A \cap B^c$ divided by the probability of B^c. The probability of B^c is already known to be 2/3. Now, we need to understand what $A \cap B^c$ is.

Now, what is B^c? B^c means that B will not be executed. So, it implies that in the sample space S = {A, B, C}, if we consider B^c, it means we are left with A and C. So, that means only A and C can be executed. Therefore, $A \subseteq B^c$. Hence, $A \cap B^c = A$.

So, this means that the probability of A given B^c is the probability of A divided by the probability of B^c. The probability of A is 1/3, and the probability of B^c is 2/3. This is 1/3 ÷ 2/3 = 1/2. So, we have found that $P(A \mid B^c) = 1/2$.

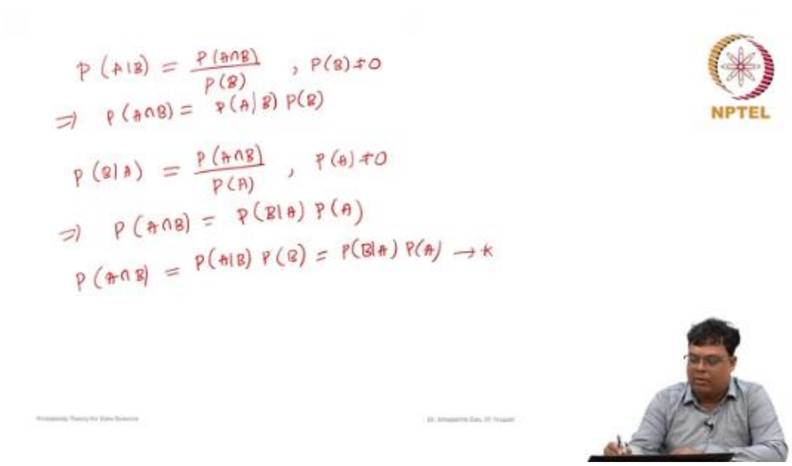
From intuition, if it is informed that B will not be executed, then only two remain: A and C. Out of A and C, one will be executed, and they are equally likely. So, the probability is 1/2. From the definition, we also get 1/2.

But this is to understand using the simple examples we discussed. So, we will discuss more examples and some additional theorems as well. What we found is that the probability of A given B is nothing but the probability of $A \cap B$ divided by the probability of B, assuming that $P(B) \neq 0$.

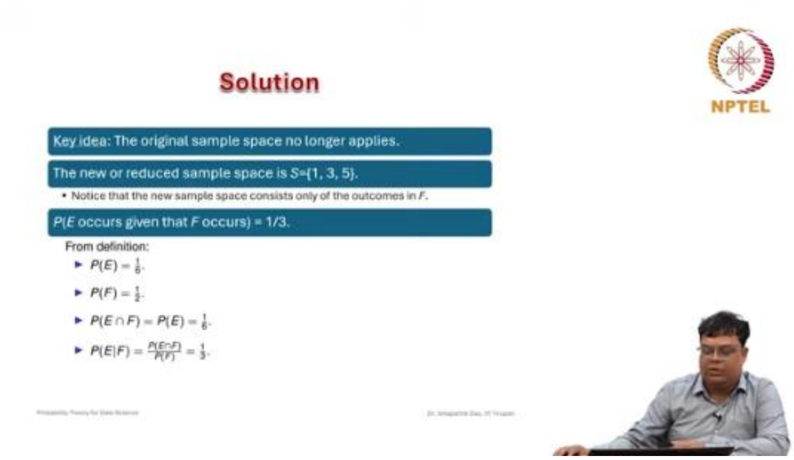


This implies that the probability of A ∩ B can be written as P(A | B) × P(B). Note that this notation "A | B" is just a vertical line, not a backslash or forward slash. Now, the probability of B | A, by definition, is P(A ∩ B) ÷ P(A), provided that P(A) ≠ 0. From here, we get that P(A ∩ B) = P(B | A) × P(A). Hence, the probability of A ∩ B can be represented in two ways: as P(A | B) × P(B), or as P(B | A) × P(A).

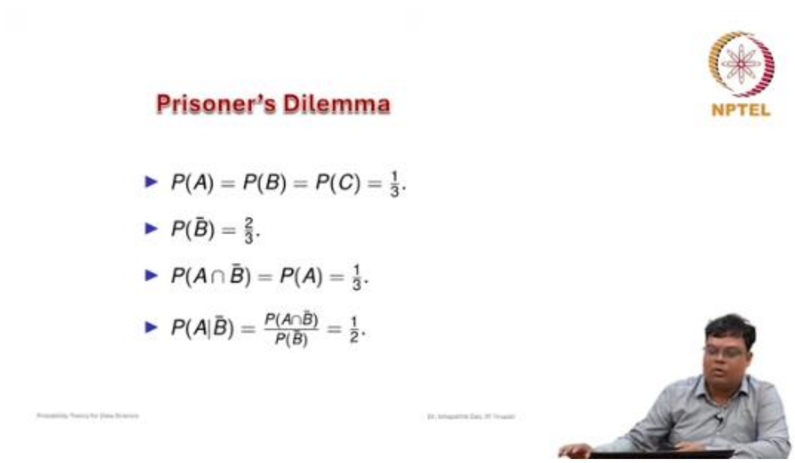
Sorry, not 'plus'—this is equal to P(A | B) × P(B), equal to P(B | A) × P(A). So, this is one relationship. Okay. Now, next, we will discuss. So, this is the solution that has already been given, including the Prisoner's Dilemma. You see that P(E) = 1/6, P(F) = 1/2, and P(E ∩ F) = 1/6.



All of these things we have already discussed. So, you can go through these slides as well. And here, in the Prisoner's dilemma, P(A) = P(B) = P(C) = 1/3. P(B^c) = 2/3. The intersection of A and B^c is nothing but P(A) = 1/3.



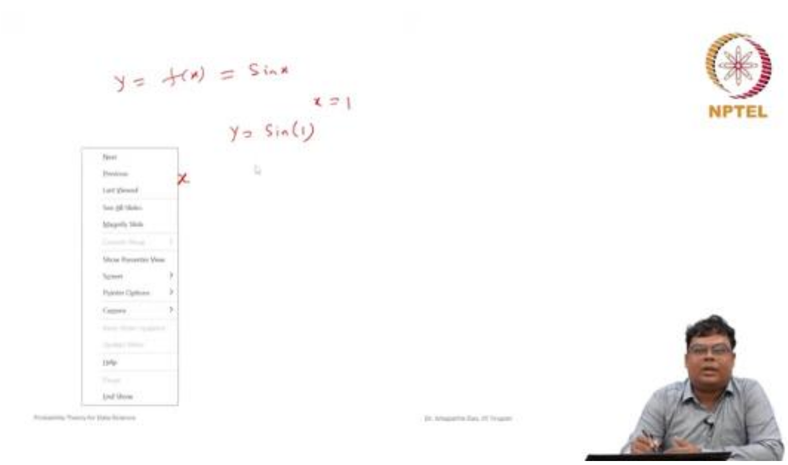
So, the probability of A given B^c becomes 1/2. So, those things are given in the slides also. Now, we will discuss one important concept. This is independence. Usually, how do we understand independence?



There are many different contexts in which we talk about independence. For example, Independence Day refers to when a country becomes independent. Sometimes we say that two variables are independent. So, whenever we say that two variables are independent, we mean that one variable is not dependent on the other. For example, if we discuss two variables, X and Y, and say they are independent, that means $P(X \cap Y) = P(X) * P(Y)$.

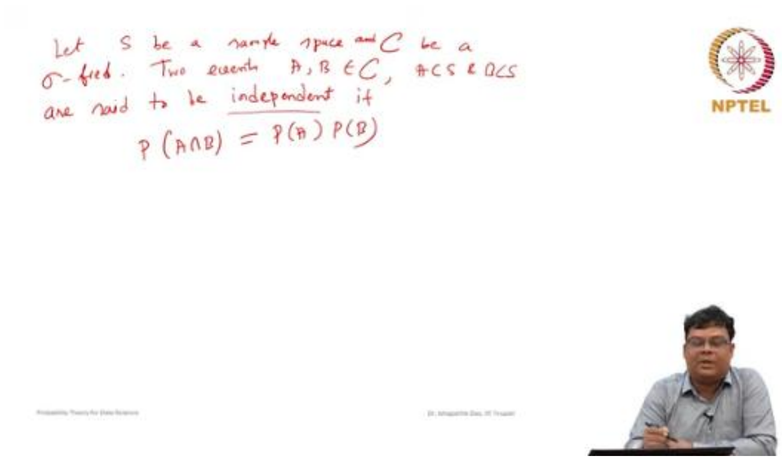
This means there is a function between them, like y = f(x). If we suppose this function is equal to sin(x), then if you give x a value of 1, y has to be a particular value, which is sin(1). There is no other value. Suppose two friends go to the market and buy some items. They decided that one friend will follow the other.

They agreed that one friend, whatever he buys—let's say his expenditure is x—then the other friend will double that expenditure. So, if one friend's expenditure is x, then the other friend's expenditure will be 2x. They are dependent because if you know one friend's expenditure in the market, then you can know the other value as well. This is known as functional dependence. Now, in the context of probability theory, when we say two events are independent, we refer to them as stochastically independent.



This means that the probability of one event depends on the probability of another event. So, how do we define that? Here, it's defined in this way: let S be a sample space and C be a sigma field, which is a class of subsets of S. Two events, suppose A and B, belong to C, meaning A ⊆ S and B ⊆ S. They are said to be independent if P(A ∩ B) = P(A) * P(B).

So, that’s how we say two events are independent. So, from this definition, the terminology of independence can be understood in a specific way. It means that P(A) is not dependent on P(B). So, how can we explain this? Now, we found that the probability of—what is it—conditional probability.

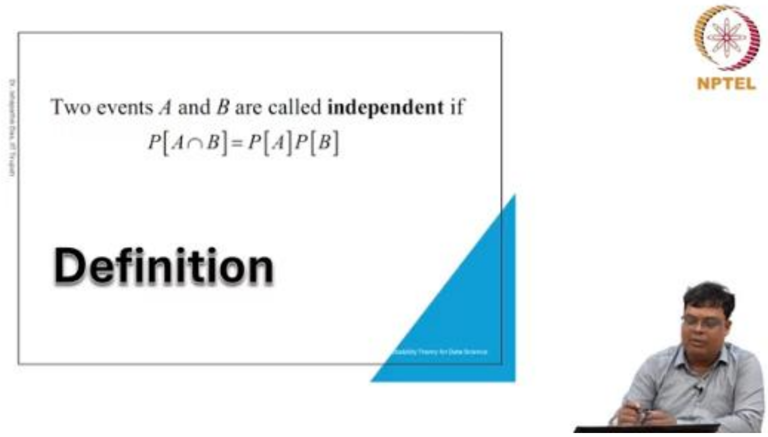


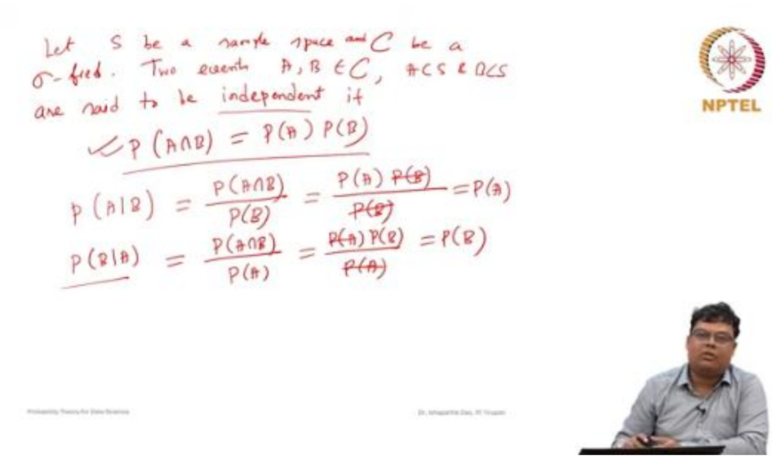
Suppose A and B are independent. This means it satisfies this relationship: P(A ∩ B) = P(A) * P(B). We will give some examples now, but first, let's try to understand the definition of how this meaning of independence is actually associated with the definition. Now, let’s find the conditional probability of A given B. By definition, this is P(A ∩ B) / P(B), assuming that P(B) ≠ 0.

So now, if A and B are independent, then P(A ∩ B) can be represented as P(A) * P(B), divided by P(B). Then, it will cancel out, and this is nothing but P(A). Similarly, if you want to find the probability of B conditioned on A, the conditional probability of B given A is P(A ∩ B) / P(A). This is P(A) * P(B) / P(A), assuming that P(A) ≠ 0; otherwise, it will be undefined. So, this is nothing but P(B).

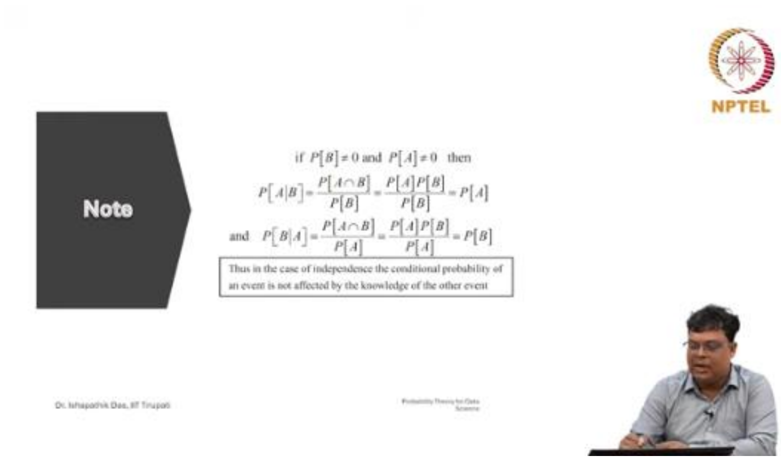
Thus, the conditional probability of A given B is nothing but P(A). This means that even if you are informed that B has occurred or you are not informed, this probability does not change. So, P(A) actually does not depend on whether B has happened or not. Similarly, you can see that if this condition is satisfied, then P(B | A) does not change whether this information is provided or not. So, P(A) does not depend on the conditional probability of A given B; it's the same as P(A).

This means P(A) does not change if you provide this information about B or not, or whether B has happened or not. That's why, in this sense of probability, A and B will be independent if these conditions are satisfied. So, this is called independence. So, this definition of independence is given here. So, you can see that two events, A and B, are called independent if P(A ∩ B) = P(A) * P(B), which we just discussed.



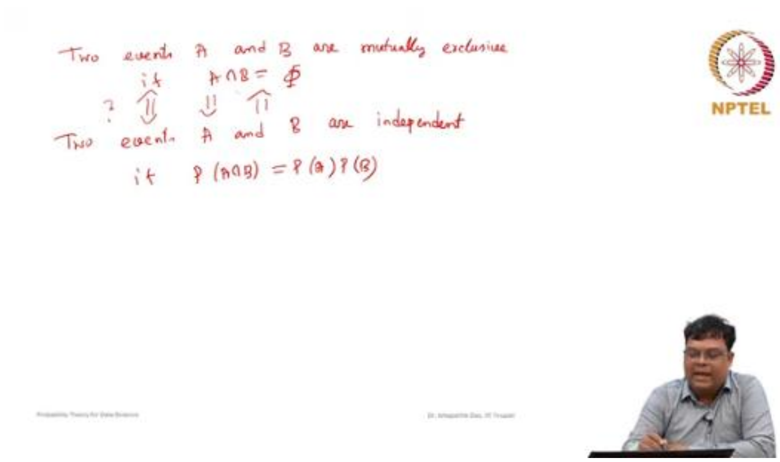


So, now this is the explanation we just discussed. If you know that P(B) ≠ 0 and P(A) ≠ 0, then P(A | B) = P(A), and P(B | A) = P(B). Thus, in the case of independence, the conditional probability of an event is not affected by the knowledge of the other event. So, in that sense, A and B are independent. Now we have already discussed some of the concepts.



We say mutually exclusive. So, suppose two events are mutually exclusive. Two events, A and B, are mutually exclusive if $A \cap B = \emptyset$. Just now, we learned that two events A and B are independent if $P(A \cap B) = P(A) \times P(B)$. Now, the question is that sometimes, when we learn new topics, like mutually exclusive events and independent events, we may wonder whether one definition is equivalent to another or if one implies the other. For example, we can ask whether mutually exclusive events are independent or if two independent events are mutually exclusive.

So, are any of these relationships true, or are they not? So, let us discuss some examples. Suppose $A \cap B = \emptyset$. So, suppose there is a sample space like this. A and B are mutually exclusive.

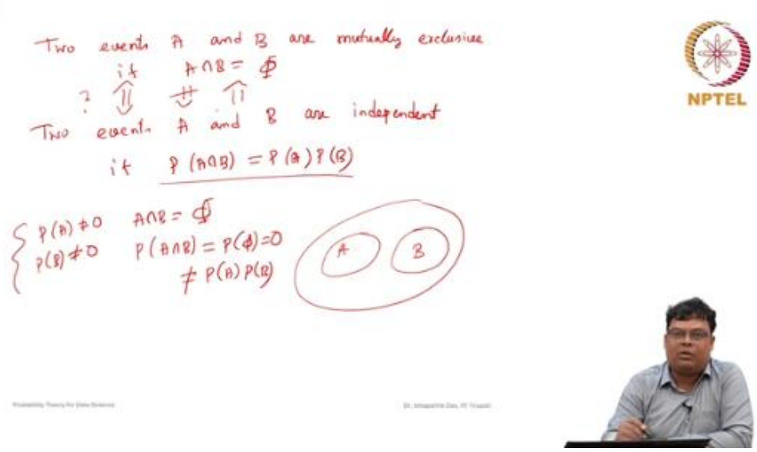


So this is A, and this is B, and also, $P(A) \neq 0$ and $P(B) \neq 0$. Now, as they are mutually exclusive, $P(A \cap B)$ will be the empty set $\emptyset$, and $P(\emptyset) = 0$. Now, it cannot be possible that $P(A) \times P(B)$ is non-zero because both are non-zero. It can only be possible when one of them has to be zero. In other words, if both are non-zero and they are mutually exclusive, then they are highly dependent.

This means that if one event happens, then the other event cannot happen. For example, consider two students going to school. If one student comes to school, then the other does not come; they are very much dependent. If you find one student in the school, then you know that the other student will not be there but somewhere else. So, this is a highly dependent situation.

Independence can only occur when one of them has a probability of zero. That is the only possibility. Therefore, mutually exclusive events imply that one event has a probability of zero, which means they do not imply independence. In general, mutually exclusive events do not imply independence. We will also discuss whether the reverse is true or not.

Let us discuss one numerical example. Here is the numerical example: Let us consider a random experiment of throwing a die. The sample space is {1, 2, 3, 4, 5, 6}. Here, the sample space S consists of the numbers one through six, and we consider it to be an unbiased, balanced die.



Now, let's define the events: event A = {1}, event B = {2, 4, 6}, and event C = {3, 6}. So these are the three events we are considering: A = {1}, B = {2, 4, 6}, and C = {3, 6}.

Now let us consider events A and B. You can see that A ∩ B has no common points. A contains the number 1, and B contains the even numbers. So, A ∩ B = ∅, which means they are mutually exclusive.

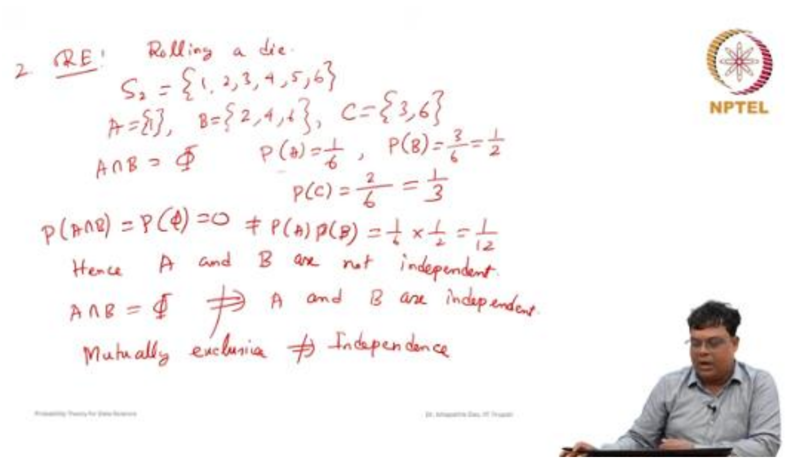
But what is P(A)? P(A) = 1/6. P(B) = 3/6 = 1/2. And P(C) = 2/6 = 1/3.

So, how do we find P(A ∩ B)? P(A ∩ B) = P(∅) = 0. This is not equal to P(A) × P(B). P(A) is also non-zero, and P(B) is non-zero as well. This is (1/6) × (1/2) = 1/12.

Hence, this is not equal to 0. Therefore, A and B are not independent. So, here's one example: A ∩ B = ∅, which shows that mutually exclusive does not imply that A and B are independent. We can say that mutually exclusive does not imply independence. This is one example.

In general, it may be true for some events that if one of the events A or B has a probability of 0, then it may be true that A ∩ B being the empty set implies that A and B could be independent. But in general, we found from this example that this is not the case.

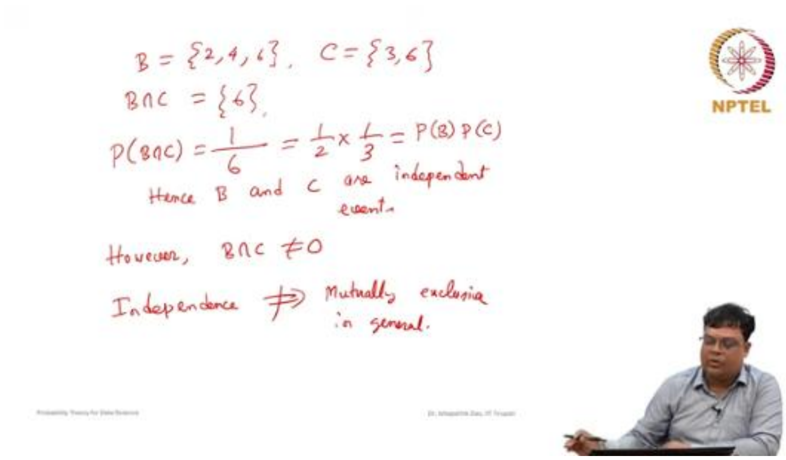
Now, next, let us consider B and C. So, what is B ∩ C? B = {2, 4, 6} and C = {3, 6}.



The intersection of B and C shows that the common point is only 6. So, P(B ∩ C) will be written as follows. Since there is only one point, using the classical approach, there are 6 total outcomes when rolling a die. Out of these, B ∩ C can occur in only one way. Therefore, the probability is 1 out of 6, which is represented as 1/6.

This can also be expressed as (1/2) × (1/3). P(B) = 1/2, and P(C) = 1/3. So, P(B ∩ C) = P(B) × P(C). Hence, by definition, B and C are independent events. However, B ∩ C ≠ ∅, so they are not mutually exclusive.

Hence, independence does not imply mutual exclusivity in general. So, that's what we ask—the question of whether one of the relations is correct. So, there is no relation. Two events may be mutually exclusive but may not be independent. Two events may be independent but may not be mutually exclusive.



So we have learned what conditional probability is and how we can define two events as independent. We've also discussed the relationship between mutually exclusive events and independence. Next, we will discuss one of the important theorems. It is known as Bayes' theorem.