PROBABILITY THEORY FOR DATA SCIENCE

Prof. Ishapathik Das

Department of Mathematics and Statistics

Indian Institute of Technology Tirupati

Week - 01

Lecture - 05

Numerical Examples and Introduction to Conditional Probability

Let us discuss one numerical example here. Chennai and Mumbai are two of the cities competing in the IPL. There are also many others. The organizers are narrowing the competition to the final five cities. There is a 20% chance that Chennai will be among the final five, a 35% chance that Mumbai will be among the final five, and an 8% chance that both Chennai and Mumbai will be among the final five.

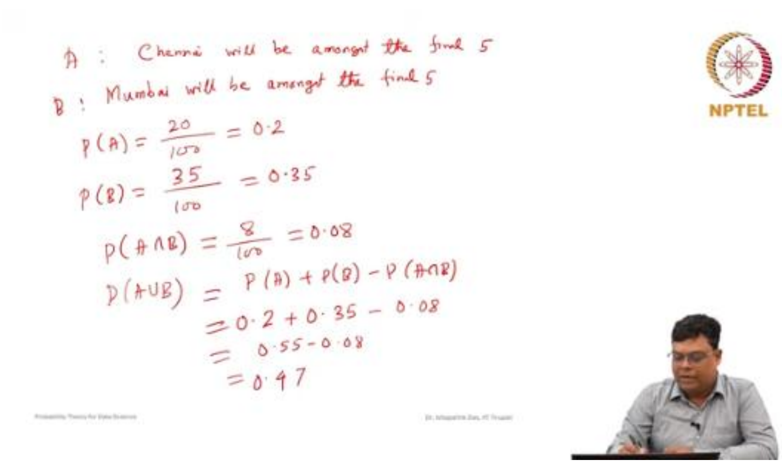
What is the probability that Chennai or Mumbai will be amongst the final five? So, whenever we see 'OR', that is nothing but $P(A \cup B)$. If we show 'AND', then this is $P(A \cap B)$. So, how can we solve this problem? This problem will be represented by the events A and B.

A is the event that Chennai will be among the final five. Let A be the event defined as Chennai being among the final five. B is the event that Mumbai will be among the final five.

So, according to the given problem, what is the probability of A? There is a 20 percent chance that Chennai will be among the final five. So, 20 percent means 20 out of 100, which is 0.2. And the probability of B is nothing but the chance for Mumbai. There is a 35 percent chance that Mumbai will be among the final five. So, this is 35/100, or 0.35. It is also given that there is an 8 percent chance that both Chennai and Mumbai will be among the final five. So, whenever it is both, that means it's the intersection. So, Chennai and Mumbai being among the final five, $P(A \cap B)$, is given as 8 percent, or 8/100, which is 0.08.

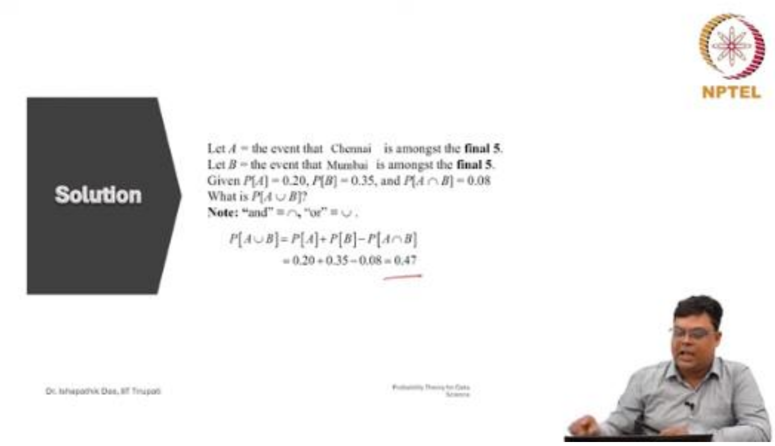
Now the question is, what is the probability that Chennai or Mumbai will be amongst the final five? So, basically, it is 'or'. We will find the probability of $A \cup B$. It is asked to find the probability of $A \cup B$. Now, we will apply Theorem 1.7, which we just proved. According to Theorem 1.6, this is $P(A) + P(B) - P(A \cap B)$.

All these values are given: this is $0.2 + 0.35 - P(A \cap B)$, which is 0.08. So, we just have to simplify this. So, $0.2 + 0.35 - 0.08 = 0.47$. I think 0.47. So, you can find this solution here. The event A is that Chennai is amongst the final five.

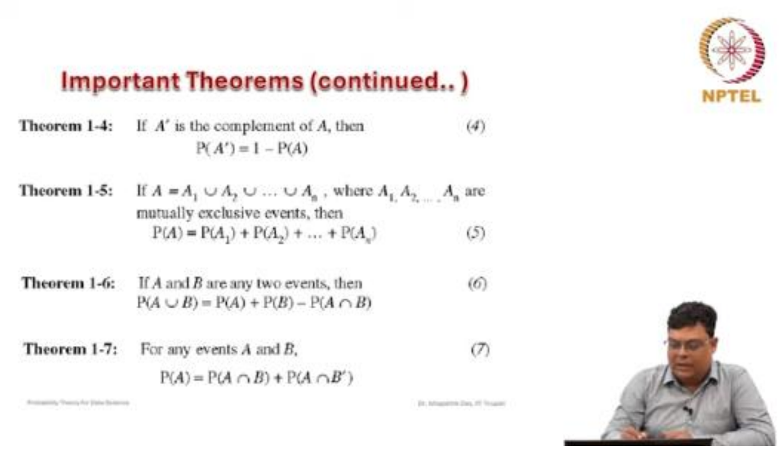


Let B be the event that Mumbai is amongst the final five. The given probability of A is 0.2, and for B, it is 0.35. The question is, what will be the probability of $A \cup B$? 'AND' means intersection, and 'OR' means union. So, the probability of $A \cup B$ is equal to $P(A) + P(B) - P(A \cap B)$.

This is $0.20 + 0.35 - 0.08$, which equals 0.47. So, this is the answer: either Chennai or Mumbai will be among the final five. This represents $A \cup B$. So, this is one numerical example related to these theorems. Now, this is the last theorem. There may be many results, and some of the basic results are ones that we think will be frequently used in the future.



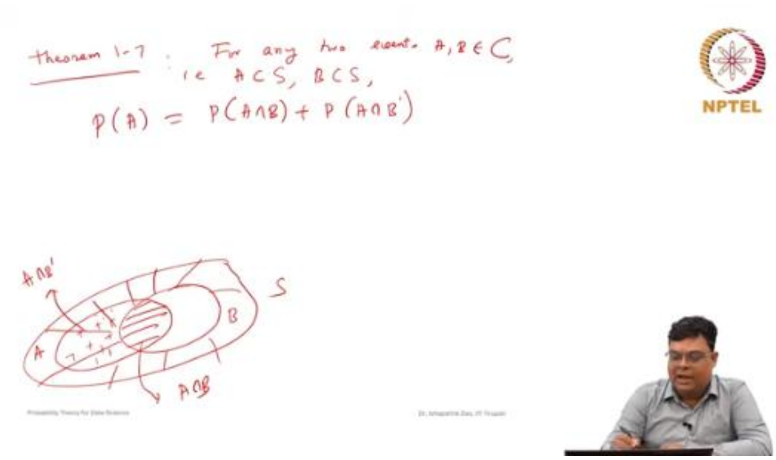
Other results, whenever required, we will discuss and prove again. So, here is the last theorem, 1.7. For any events A and B, $P(A) = P(A \cap B) + P(A \cap B^c)$. For A, this is Theorem 1.7. For any two events, A and B, we can say that $A \subseteq S$, and $B \subseteq S$.



The probability of A can be represented as $P(A) = P(A \cap B) + P(A \cap B^c)$. If you show this with a Venn diagram, suppose this is S, this is A, and this is B. So, this is A, and this is B. Now, A can be represented as $A \cap B$, which is the part of A that overlaps with B. This is $A \cap B$, plus the disjoint part, which is the remaining part. So, what is this?

This is simply B^c, or B' which includes all the values other than B. The intersection of A with B^c is this part. So this part is $A \cap B^c$. So this is the union of $(A \cap B^c) \cup (A \cap B)$. How can we prove this analytically?

Graphically, we can see this. So, any event B is a subset of S. Since B is a subset of S, $B \cup B^c = S$, and $B \cap B^c = \emptyset$. So, B is a subset of S. That is why $B \cup B^c = S$, and $B \cap B^c = \emptyset$.

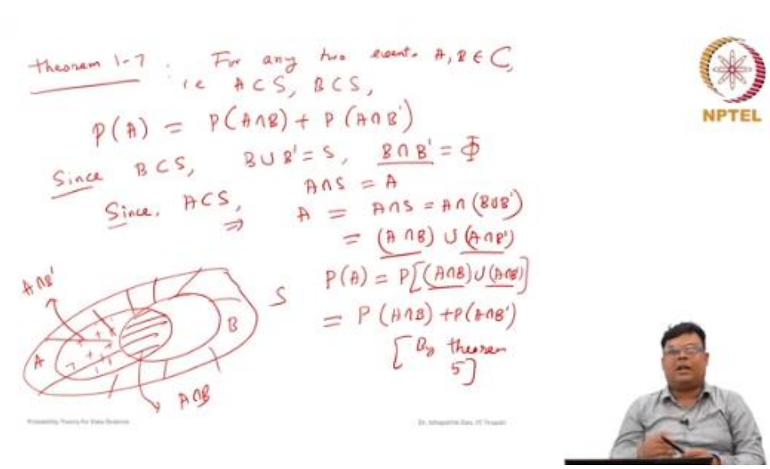


The sets $A \cap B$ and $A \cap B^c$ will be disjoint because $B \cap B^c$ is already disjoint. Any element in $A \cap B$ means it belongs to B. So, any element in $A \cap B^c$ means it belongs to B^c. Now, any element cannot be part of both because if it is in both, then it will be in both B and B^c, which is a null set ($\emptyset$). So, that is why $A \cap B$ and $A \cap B^c$ will be disjoint.

Now, since A is a subset of S, what will $A \cap S$ be? It is nothing but A. This means A can be represented as $A \cap S$, which is equal to $A \cap (B \cup B^c)$. Now, if you use the distributive property, it becomes $(A \cap B) \cup (A \cap B^c)$.

So, now, the probability of A can be represented as the union of two disjoint sets: $A \cap B$ and $A \cap B^c$. These sets are disjoint, as we mentioned in Axiom 3, and we proved this in Theorem 1.5. For a finite collection of mutually disjoint sets, we can add these two probabilities: $P(A \cap B) + P(A \cap B^c)$. This is supported by Axiom 3, or you can refer to Theorem 5, which states that $P(A) = P(A \cap B) + P(A \cap B^c)$.

So, this is one of the important theorems. Later, we will see that there are many applications. These are some extensions. So, suppose B_1, B_2, ..., Bn are pairwise mutually disjoint and mutually exclusive sets of events.

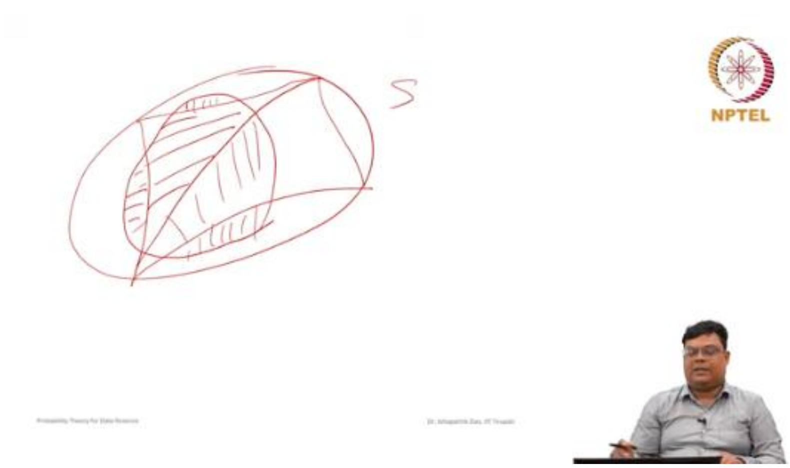


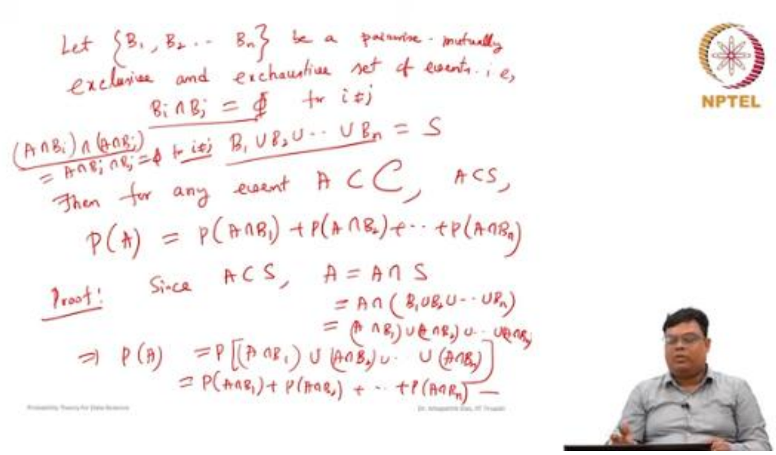
This means that, by definition, they are mutually exclusive, which implies that the intersection of B_i and B_j is pairwise mutually exclusive for $i \neq j$, where i and j range from 1 to n. If you take the union of B_1, B_2, ..., Bn, this will cover the whole set S, forming a partition of the set S. So, for any event A that belongs to C, it means A is a subset of S. The probability of A is the probability of $A \cap B_1$ plus the probability of $A \cap B_2$ plus the probability of $A \cap Bn$.

So, how can we prove that? We can use the previous concept. Since A is a subset of S, A can be represented as $A \cap S$. Now, $A \cap S$ can be represented as $B_1 \cup B_2 \cup Bn$. By the distributive property, this becomes $A \cap B_1 \cup A \cap B_2 \cup ... \cup A \cap Bn$. Now, what will be the probability of A? By the left-hand side, the probability of A will be, this is nothing but the probability of $A \cap B_1$, $\cup A \cap B_2$, $\cup A \cap Bn$.

Now, since B_1, B_2, and B_j are pairwise mutually exclusive events, $A \cap B_1$ and $A \cap B_2$ will also be pairwise mutually exclusive events. This can be shown like this: if you take any i and j, the intersection of $A \cap B_j$ is nothing but $A \cap B_i \cap B_j$, which is equal to the null set ($\emptyset$) for $i \neq j$. So, that is why they are pairwise mutually exclusive events. Now, using Theorem 5, because they are pairwise mutually exclusive events, this is nothing but the probability of $A \cap B_1$ plus the probability of $A \cap B_2$ plus the probability of $A \cap Bn$.

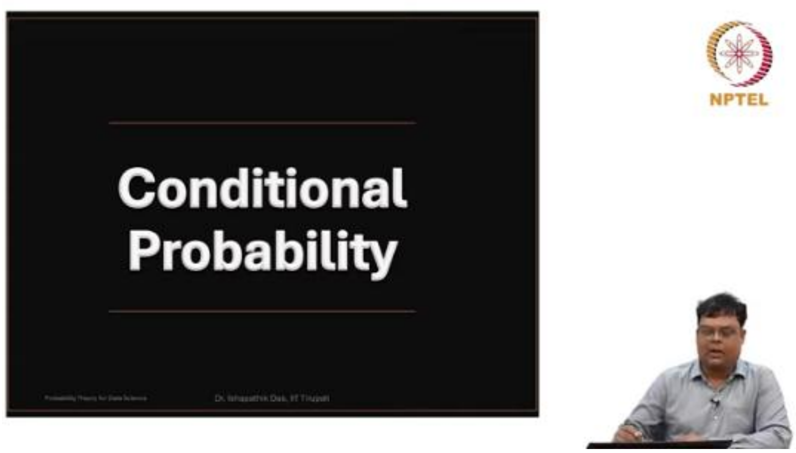
So, this result we will discuss again whenever we talk about how to find the probability of a total, which is called the total probability of an event. So, that means if you have this kind of partition. So, then we can write the total probability. Graphically, we can represent it like this: it is like a set S. Suppose you have some kind of partition, where for any event A, all are pairwise disjoint. Any event is actually part of this partition, with some belonging to B_1, some to B_2, some to B_3, and some to B_4, like this. So, this is the representation. So, we have completed some of the important theorems now.





Next, we will discuss some other results and new topics. If you require any other important results, we will prove them again. These are basic concepts; you may have learned them earlier, but we wanted to prove them formally to ensure there are no doubts about whether we are just assuming them. We have already proved this from the axioms. So, that is why we proved all these results.

We have already discussed this example. Now, next, we will discuss the concept of conditional probability. Now we will discuss the new concept of conditional probability. Sometimes we are given information in the form that some event has already happened. Initially, we have a random experiment associated with it. Suppose the random experiment is rolling a die.



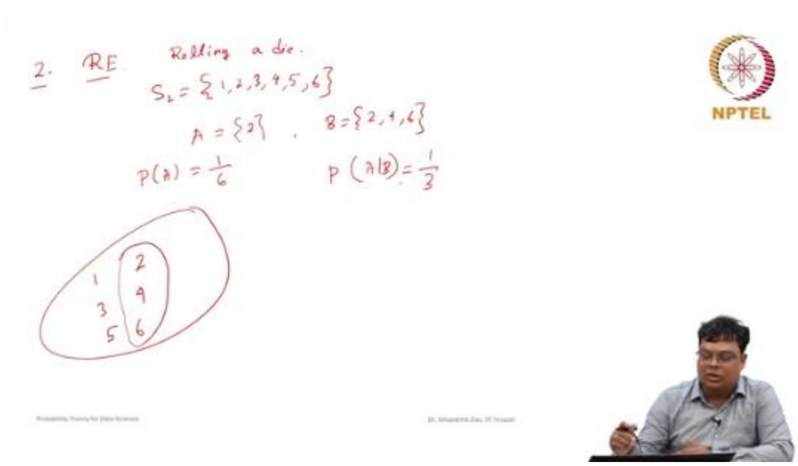
The sample space, you know, corresponds to rolling a die. This is the second example of a sample space we discussed: $S_2 = \{1, 2, 3, 4, 5, 6\}$. Now, suppose in another room, someone rolls the die, and we have no other information. The probability of, let's say, rolling a 2 can be considered. So, let's take one event, A, which is when a 2 is observed on the top face.

Now, suppose someone gave the information that the result is 2, 4, or 6. So, whenever there is no information, we can assume that any of the observations can happen, and by symmetry, the probability will be 1 out of 6 for any observation. Thus, the probability of A will be 1/6 by the classical approach, which assumes that all the observations are equally likely. Now, suppose this information is provided to us: someone says that the die is rolled and they have already observed that the result is an even number. So, if we have this information, we can understand that the probability of any odd number now is 0 because we already know it's an even number.

Therefore, the probability of getting 1, 3, or 5 will be 0. Now, what is the probability that the result will be 2? So, in that case, we denote this kind of notation. We say that the probability of A is 1/6. But the probability of A, given B, means that the information B is given to us. So, this is called conditional probability.

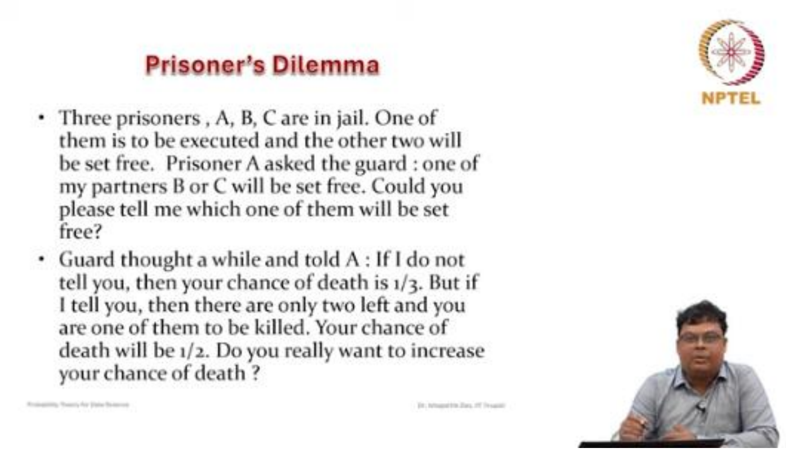
This is different from this. So now, from the intuition, we will discuss the proper definition of conditional probability. Before that, we understand that the probability of any odd number will be 0 if this is an even number. So, for one of the even numbers, what is the probability that 2 may come? Because now the sample space has changed; earlier, the sample space was {1, 2, 3, 4, 5, 6}.

Now, the sample space, whenever someone says that an even number occurs, should be considered as the possible sample space. Given this information, our sample space is nothing but {2, 4, 6}. Out of these three observations, the sample space contains only three equally likely points. So, what is the probability that 2 may appear? This is 1/3. This is from the intuition. So, another example.



This is written here: suppose before observing the outcome of a random experiment, you are given information regarding that outcome. How should this information be used in the prediction of the outcome? Namely, how should probabilities be adjusted to take into account this information? Usually, the information is given in the following form: you are told that the outcome belongs to a given event, or you are told that an event has occurred, like here, where an even number occurred. So, one of the interesting examples here is called the prisoner's dilemma.

There are three prisoners: A, B, and C, in a jail. One of them will be executed, and the other two will be set free. So, in that situation, there are three prisoners: A, B, and C. So, this is the sample space: three prisoners in a jail. Now, one of them will be executed.



Suppose A denotes the event that A will be executed, B denotes the event that B will be executed, and C denotes the event that C will be executed. By the principle of equally likely outcomes, the probability of A being executed is 1/3. The probability that B will be executed is 1/3, and the probability that C will be executed is also 1/3. So, the probability of A^c, by definition, is 1 - P(A). We have studied this; it's 1 - 1/3, which is 2/3.

Similarly, we sometimes write it with a bar. So whenever you see different books and references, you will notice different notations. The probability of A^c can be written as $\overline{A}$, or A^c. So, the probability of $\overline{A}$ is 1 - P(A), which is 1 - 1/3, giving us 2/3. The probability of B^c, which we write as $\overline{B}$, is 1 - P(B). This is again 1 - 1/3, which is 2/3. Similarly, the probability of C^c is also 2/3, calculated as 1 - 1/3. This information comes from the fact that we do not have any other information. We know there are three prisoners in jail, one will be executed, and two will be set free. Now, the prisoner A thought that he could get some additional information.

He could ask the guard directly whether he would be executed or not, but the guard may not provide that information. So instead of asking this question directly, A thought he would ask in a different way. He said, "Please don't give any information about me, but among B and C, since we know one will be executed and one will be set free, please tell me who will be set free between B and C." So, A wants this information in terms of whether B^c may happen or C^c may happen. This means one of the events will happen, but it is uncertain.

So, that is the question A asked. Now, the guard thought about how this information may be utilized and how A could benefit from it. The guard thought that if he said the information is that B^c has happened or been observed, it means B will not be executed. So, this means B will not be executed, which is going to happen. This means B^c is given, and we can say that the guard gives the answer that B^c has happened or that B will not be executed.

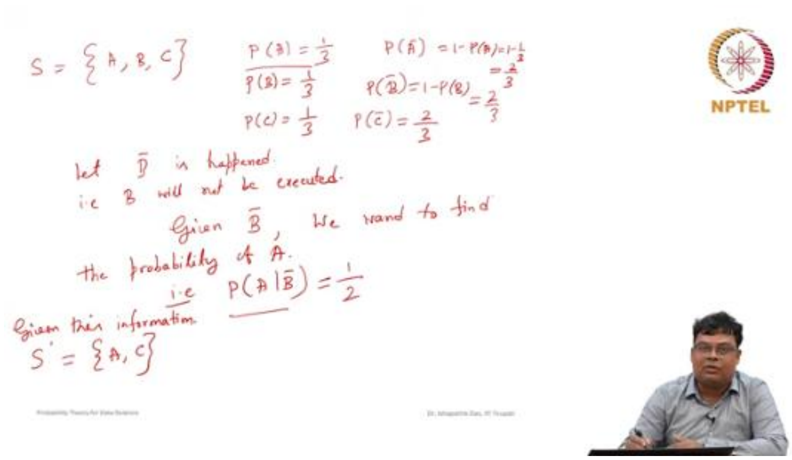
Now, given this information that B^c is true, what is the probability of A? We want to find the probability of A given B^c. So, B^c is given, which means B will not be executed. Then, what is the possibility? Either A or C can be executed.

Now, from the intuition, what will be the probability? Initially, the probability of A is 1/3. Now, the probability of A given B^c—this probability, from the intuition, is that since B^c has already happened, we know that B will not be executed. So, then, among A and C, the sample space changes. Given this information, the sample space now consists only of A and C.

Out of A and C, what is the probability that A will be executed? Since there are two elements and they are equally likely, the probability of A is equal to the probability of C. Thus, out of the two, one will be executed, which gives us a probability of 1/2 for A. From this intuition, we can get this probability, but we need to formally define it. So, how do we find the conditional probability?

Because, from intuition, we may not be able to compute it, or it may not be formally defined. We have to define this formally, and we will do that. But before that, there are some simple examples that can help us understand how given information can be utilized to find the probability of an event. Now, regarding conditional probability, we got the information that earlier, the probability that A would be executed was 1/3. Now, the probability of A being executed, given that B is not executed, is 1/2. So, this means that the probability that A will be executed has increased.

So, the guard asks whether A wants to increase the probability of his death. So, here it is written that three prisoners, A, B, and C, are in jail. One of them is to be executed, and the other two will be set free. Prisoner A asked the guard, "One of my partners, B or C, will be set free. Could you please tell me which one of them will be set free?" The guard thought for a while and told A, "If I do not tell you, then your chance of death is 1/3.



But if I tell you, then there are only two left, and you are one of them to be killed. Your chance of death will be 1/2. Do you really want to increase your chance of death?" So, this conditional probability is utilized to find the probability more appropriately.