**Essentials of Data Science with R Software-1**
**Professor Shalabh**
**Department of Mathematics and Statistics**
**Indian Institute of Technology Kanpur**
**Lecture 17**
# Conditional Probability

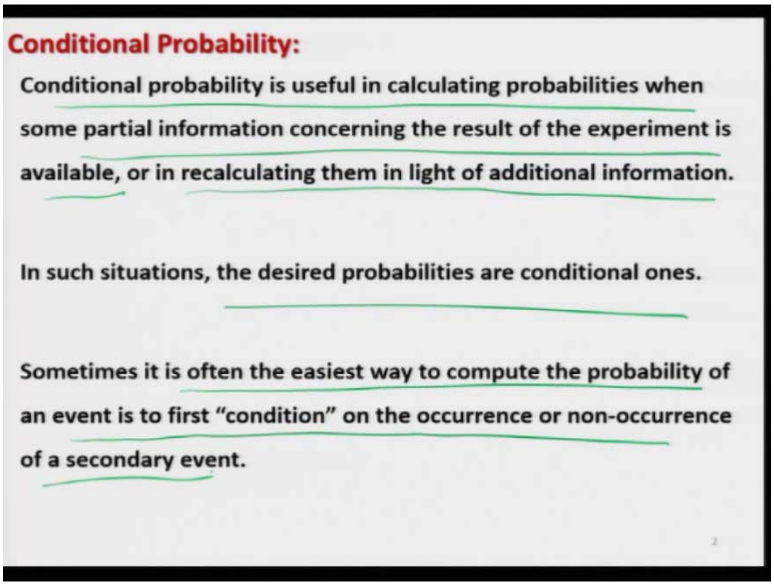
Hello friends. Welcome to the course Essentials of Data Science with R Software-1, in which we are trying to understand the topics of probability theory and statistical inference. So, now, you see, up to now, you have understood the entire concepts and theory behind the simple probability. The probability which you call by a nickname m upon n. So, now, I am sure that there should not be any problem in the understanding of the basic concepts of probability and how to interpret the value of the probability, and even you can compute it very easily.

So, now I want to move into another direction that how to compute different types of probabilities. And one among them is the conditional probability. You know, when in real life something is happening, we want to compute the probability and the probability depends on some conditions also.

Suppose, if there is a shopping website and that shopping website wants to compute the probability of something, given some condition. For example, I want to give a condition that I want to know the probabilities of shopping a clothing when the customer is a female or the customer is a child in the age group of say this, 7 to 12 years. So, what is that? You have all sorts of customers in your sample space from children, young children, young people, older people, etc. But you are trying to put a condition over there. And then you want to compute the probability.

So, whenever we are trying to handle more than two events, and we want to compute the probability and the value of the probability is conditioned on one of the event, then the concept of conditional probability comes into picture and we try to use it to estimate different types of probabilities. So, now, instead of telling you these stories, why not we take a very simple example and try to understand what is the meaning of conditional probability that what it interprets? And how it is useful for us?

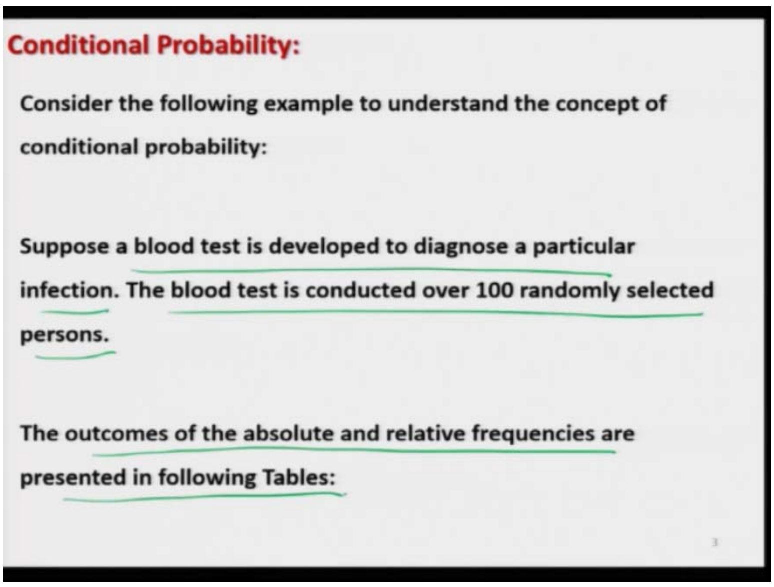
(Refer Slide Time: 02:50)




So, let us begin our lecture and we try to take a very simple example on the conditional probability. So, the conditional probability is useful in calculating the probabilities when some partial information concerning the results of the experiment is available with us, or when we try to recalculate those probability in the light of some additional information. In such situation the desired properties are the conditional ones.

And many times, whenever you are dealing with real life or real-life data, that it is often the easiest way to compute the probability of an event is to first condition on the occurrence or the non-occurrence of the secondary event and then try to compute it.

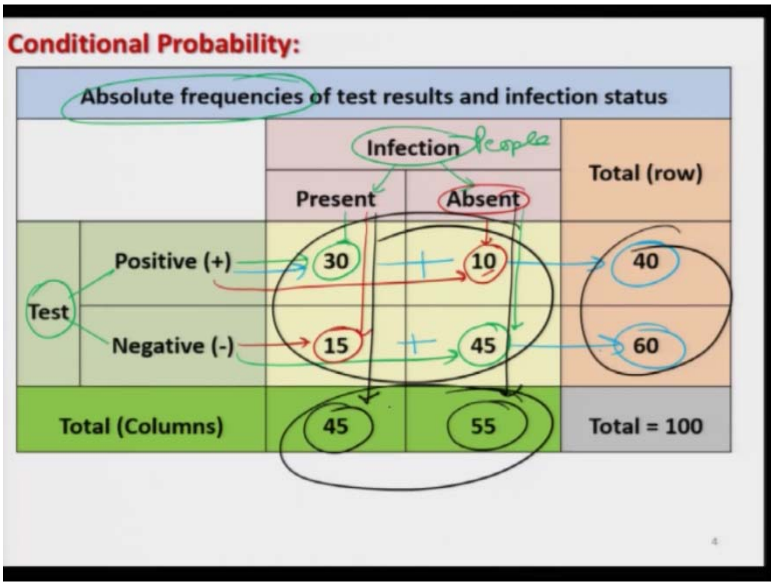
(Refer Slide Time: 03:28)



So, look let me try to consider an example here, which will try to show you that how this concept of conditional probability can be used very easily and how it is going to be different from the simple probability that we consider earlier up to now. So, suppose there is a blood test and this blood test is developed to diagnose a particular type of infection and blood tests is conducted over 100 randomly selected persons.

So, now, you can see in such a situation, that whenever the test is administered to a group of people, then there are four possibilities that the patient has a problem or there is no problem and that test is giving you the correct result or test is giving you a wrong result. So, based on that, what happened that these outcomes are recorded in terms of the absolute and relative frequencies, and we try to present them in the following 2 by 2 contingency table.

(Refer Slide Time: 04:24)



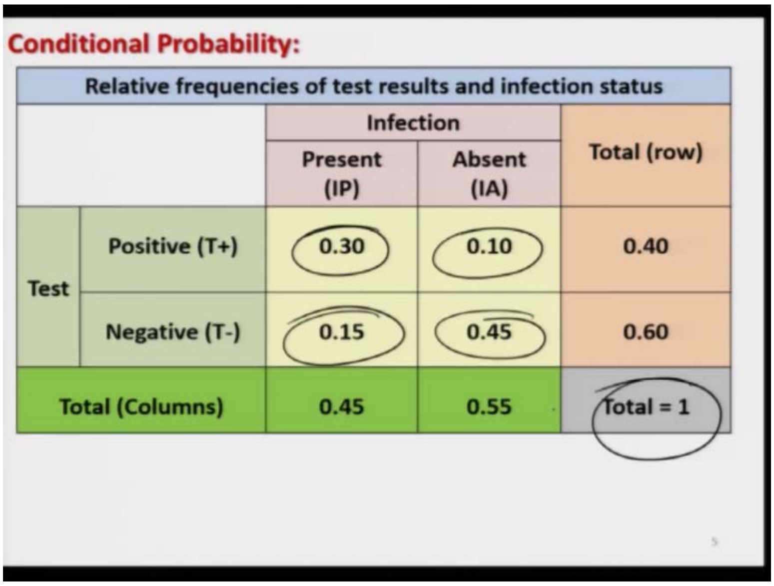

So, if you try to see here that in this table, we are trying to indicate the absolute frequencies of the test results that the blood test is given to 100 randomly selected people. So now, there are people here and we try to classify them with respect to the infection, whether the infection is present in the blood of those people or it is absent. And then we try to take the outcome of the test, the test is indicating the positive or negative. Positive means yes, the infection is present and negative means that infection is absent.

So now, this blood test is given to 100 people and out of 30 people the infection is really present and the test is saying positive, that means the test is correctly diagnosing the infection. And similarly, there are 45 people, in which the infection is really absent and this test is also giving the negative result that when the infection is absent and test is also saying negative. So, that means, that test is good.

Beside those things, we have here, two more values that the infection is actually absent and the test is saying positive that means, the test is telling that the infection is present, but whereas the infection is really absent. So, there are 10 such outcomes and similarly, there are 15 outcomes in which the infection is, infection is present in the people and the test is saying negative So, that means, that this is saying that there is no infection whereas the infection is really present. So, this is how we can summarise the outcome of these 4 possible combinations.

And if you try to see here, so, there are here 40 people here, which is 30 plus 10 is equal to 40 for which the test is saying that it is positive. And there are here 15 plus 45, this is here there are 60 persons with whom the test is saying negative and similarly, there are here 45 people in which the test is present and there are 55 people in with the test is absent.

(Refer Slide Time: 06:55)

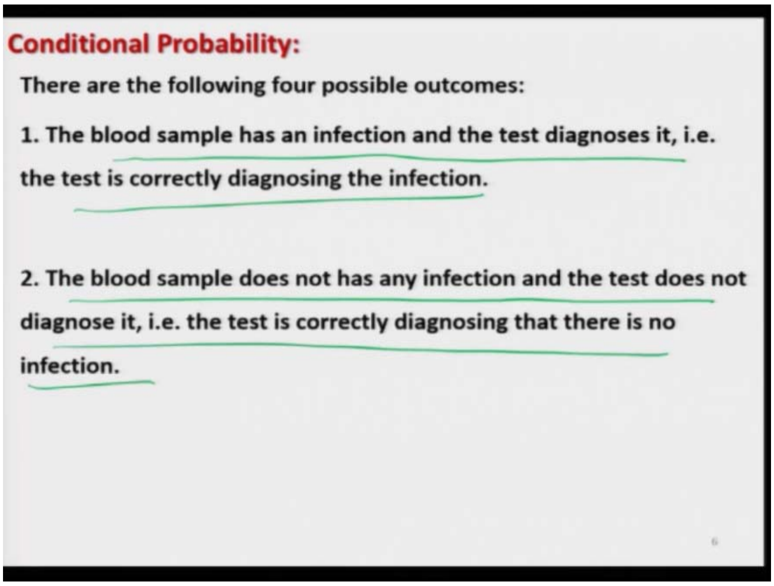


**Conditional Probability:**

| Relative frequencies of test results and infection status | | | | |
|---|---|---|---|---|
| | | Infection | | Total (row) |
| | | Present (IP) | Absent (IA) | |
| Test | Positive (T+) | 0.30 | 0.10 | 0.40 |
| | Negative (T-) | 0.15 | 0.45 | 0.60 |
| Total (Columns) | | 0.45 | 0.55 | Total = 1 |

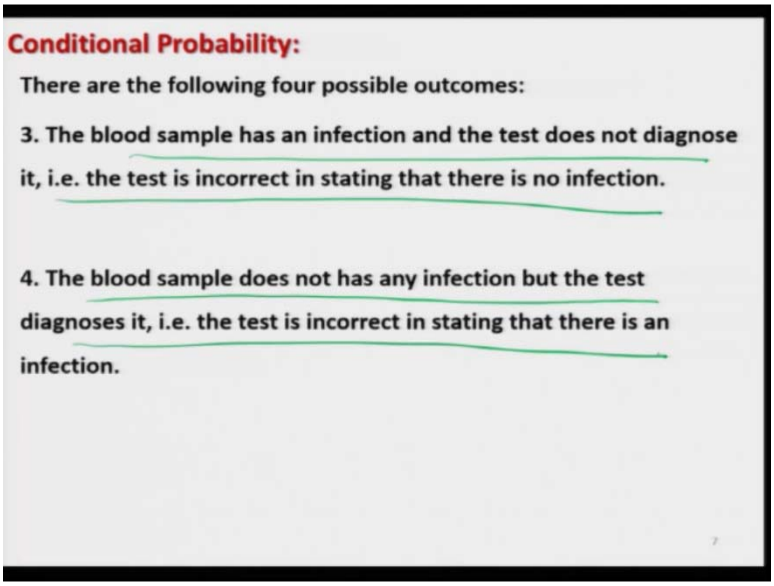
Similarly, in case if I try to translate these results in terms of relative frequency, then what I have to do, I simply have to just divide these 4 values and these values just by 100. And then I can present these things in terms of relative frequency. So, this relative frequency will become 0.30, 0.10, 0.15, 0.45 and the total frequency which was here 100 this now becomes here 1.

(Refer Slide Time: 07:27)



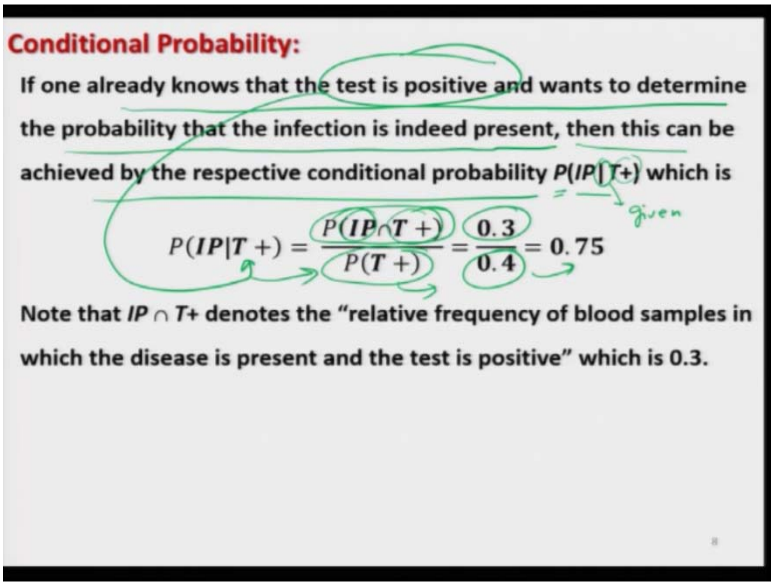

So, now, under this type of condition, we have here four types of possible outcomes. The first blood sample has an infection and that test diagnosis, that is the test is correctly diagnosing the infection. The blood sample does not has any infection and the test does not diagnose it, that is the test is correctly diagnosing that there is no infection.

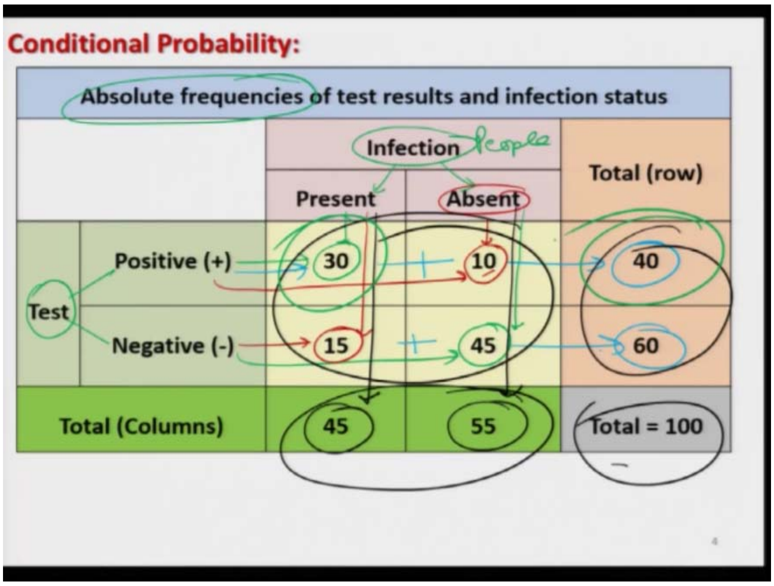
(Refer Slide Time: 07:48)

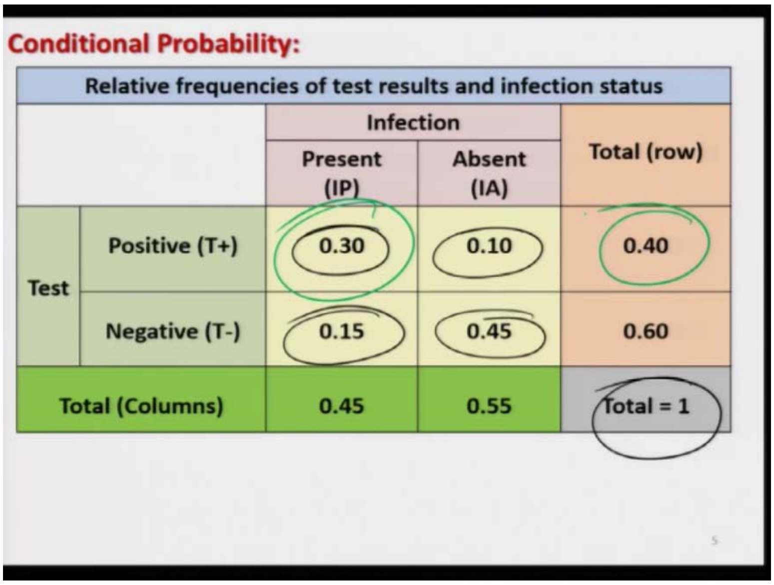



And the third possibility is that the blood sample has an infection and the test does not diagnose it, that is the test is incorrect in stating that there is no infection. And the fourth option is that the blood sample does not has any infection, but the test diagnosis it, that the test is incorrect in stating that there is an infection.

(Refer Slide Time: 08:09)





**Conditional Probability:**

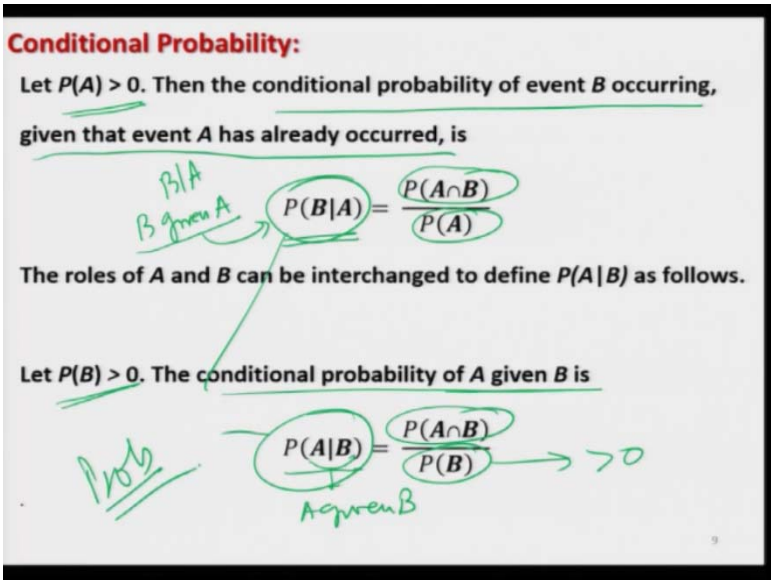
| Relative frequencies of test results and infection status | | | | |
|---|---|---|---|---|
| | | Infection | | Total (row) |
| | | Present (IP) | Absent (IA) | |
| Test | Positive (T+) | 0.30 | 0.10 | 0.40 |
| | Negative (T-) | 0.15 | 0.45 | 0.60 |
| Total (Columns) | | 0.45 | 0.55 | Total = 1 |

Now, based on this let me try to solve this issue, that if one already knows that the test is positive, and wants to determine the probability that the infection is indeed present, then how to know this thing. So, this type of probability can be achieved by the conditional probability concept. And this is indicated by a probability of IP and this vertical line is called as given. So, this is IP event T plus IP means the infection is present and T plus is test is positive. So, this is the symbol of conditional probability of IP event T plus.

So, this is obtained in as the ratio of the probability of intersection of IP and T plus divided by probability of T plus because we know that the test is positive. If you try to see here test is positive. So, this is indicated here in T positive and it is coming here as a ratio. So, in case if you try to see here that the number of people in which the test is positive, and the infection is present, both this event are occurring in 30 number of people or their relative probability is 0.3.

So, this is what we have written here 0.3. And the total number of people in which the test is positive, this you can see here, this is obtained here. Test is positive or in the terms of relative frequency 0.40. So, this is written over here, and if you try to solve it here this is 0.75, right. So, this is how we try to compute this type of conditional probability, that the based or certain condition that given the condition that the test is positive.

(Refer Slide Time: 10:18)

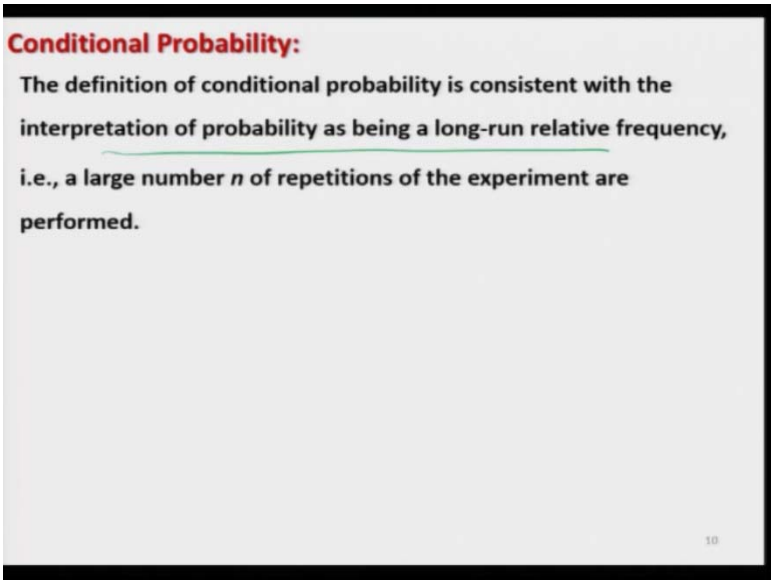



So, now, based on this example, now, I can extend it and try to give you a formal definition of the conditional probability. So, suppose that there is an event A whose probability we are assuming it to be greater than 0 that is obvious that is the assumption, then the conditional probability of an event B occurring given that the event A has already occurred is given by P(B|A), this is the indication mean probability B given A. This the meaning of this symbol. What is B vertical line A? So, this is defined here as a P(A∩B)/P(A).

And now, if you want to interchange the roles of events A and B that you are trying to find out here the probability of B given A, now if you are trying to find out the P(A|B). So, we simply have to assume that probability of B is greater than 0 and the conditional probability of A given B is P(A|B), this is the meaning and the meaning of this is A given B, this is P(A∩B)/P(B).

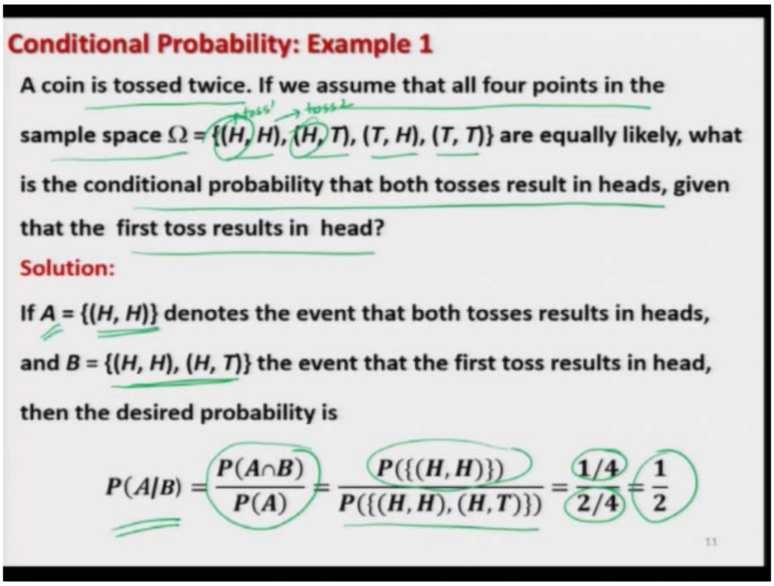
And you see, because you are trying to use a probability of B in the denominator that is why you have to assume that it is a strictly greater than 0. Yeah, and this definition what you have given here by P or P(A) given B, these are also the probabilities. And whatever you had made, while defining the probability, those all those axioms which were valid for a simple probability that would be also valid in such a case with some modifications. So, you have to understand that this is also a probability just like what you have computed the simple relative frequency-based probability.

(Refer Slide Time: 12:12)




And actually, this definition of conditional probability is consistent with the interpretation of the probability as being a long run relative frequency that is the large number n of repetitions of the experiments are performed. And actually, this concept will help you when you are trying to compute this probability or trying to simulate this probability on the basis of a software say R.

(Refer Slide Time: 12:35)




So, now, let me try to take some examples to explain you that how are you going to compute this conditional probability algebraically and after that, I will try to show you that how you can compute the conditional probability, or you are how you can simulate it also. But now, let me try to consider here a very simple example.

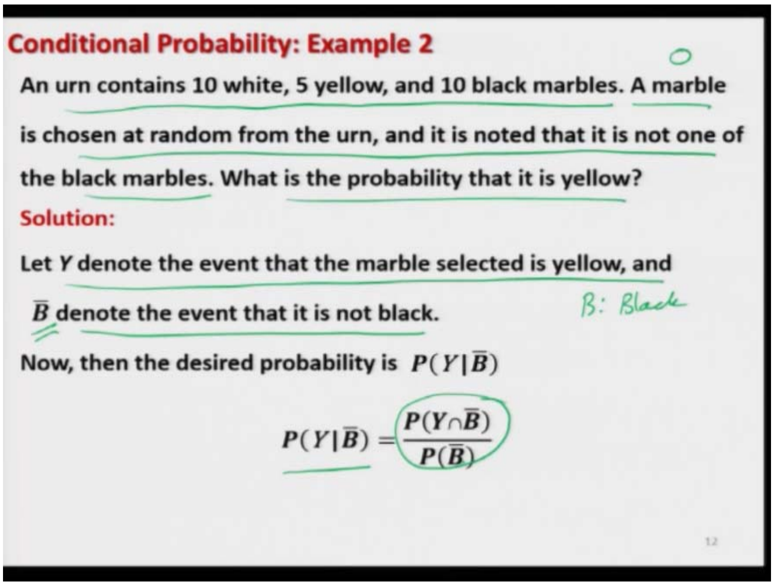
Suppose our coin is tossed two times and if he assumed that all the 4 points in the sample space, they are equally likely that means, they are obtained as head, head, head, tail, tail, head and tail, tail. That is the, this is you are here toss number one and this is here toss number two, and so, on. So, we have here four such equally likely event. And we want to find out the conditional probability that both tosses results in heads, given that the first toss results in head.

So, now you can see here the events in which we have the first toss as a head, these are here two events, this one and this one out of four. So, now in case if we define here two events A and B such that A is the event that both the tosses results in head, and it is indicated by H, H and B is the event that the first toss results in head.

Now, the desired probability is P(A|B). So, now, in case if you want to compute this P(A|B) then it is going to be here P(A∩B)/P(A). and which is here like probability of H, H divided by probability of H, T. So, you can see here probability of getting H, H is simply 1 upon 4.

There is one event and there are four possible event and there are two possibilities in which the head can come in the first toss. So, this will be 2/4. So, this probability is going to be here 1/2. So, this is the basic concept of the probability theory over here.

(Refer Slide Time: 14:30)

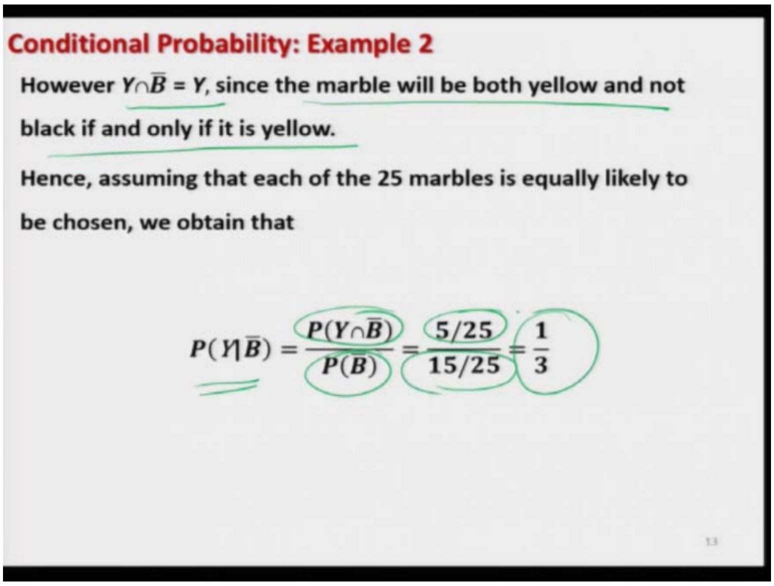



So, now, we consider one more example and, in this example, suppose there is an urn which has 10 white, five yellow and 10 black marbles. Marbles, you know these are the small glass balls in which we used to play in our childhood. So, our marble is chosen at random from the urn and it is noted that it is not one of the black marbles. And we want to know, what is the probability that it is a yellow marble.

So, in order to do these things, which we know that now there is condition that it is not one of the black marble. So, we try to define here two events, one for the yellow marble and another for not being the black marble. So, I can take here suppose Y indicate that even there the marble selected is yellow, and suppose B is the event that the marble is black. So, obviously, this B compliment will indicate the event that the marble is not black.

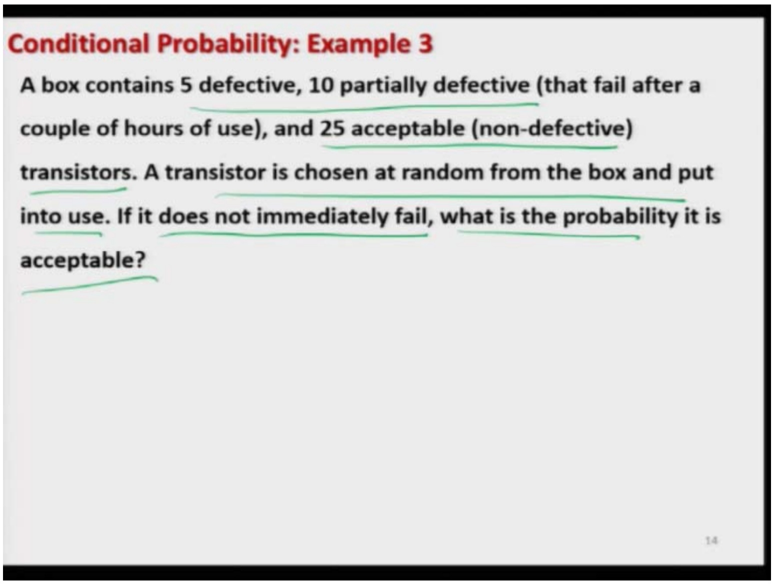
So, now we want to find out the probability. Now, we want to find out the probability. This Y given $\overline{B}$ , so this is going to be $P(Y|\overline{B})$ is equal to $P(Y\cap\overline{B})/P(\overline{B})$.

(Refer Slide Time: 12:47)



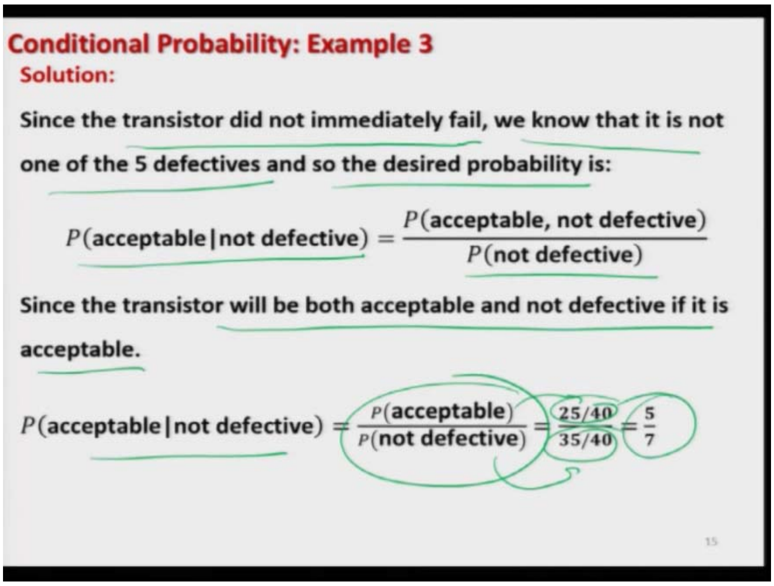

And now, in case if you try to see here in this case $Y\cap\overline{B}$ is simply Y because, the marble will be both yellow and not black if and only if it is yellow. So, assuming that each of the 25 marbles they are equally likely to be chosen then we obtain that probability of Y given B bar is the same as $P(Y\cap\overline{B})$ which will become 5/25 and $P(\overline{B})$ is 15/25 and this will come out to be here 1/3. So, this is how we can compute this probability.

(Refer Slide Time: 16:20)




Now, here I try to take here one more example. That suppose there is a box which contains 5 defective, 10 partially defective mean partially defective means, those say this transistor which fails after some time of use, and there are 25 acceptable which are non-defective transistors. Now, a transistor is chosen at random from the box and it is put into use. And suppose the transistor does not immediately fail, then we want to know what is the probability that it is acceptable. So, you can see here in all these examples, there is always a condition under which we are trying to find out the probabilities.

(Refer Slide Time: 17:00)




So, in order to find such probabilities, we know that since the transistor did not fail immediately, we know that it is not one of those five defective transistors. And so, in this case the desired probability will be probability that the transistor is acceptable, given that it is not defective.

So, that will be probability that it is acceptable intersection not defective divided by probability of not defective. And since the transistor will be both acceptable and not defective if it is acceptable. So, in this case the probability of acceptable given not defective is equal to probability of acceptable divided by probability of not defective and that is equal to probability of acceptable is 25/40 and probability of not defective is 35/40. This will come out to be here 5/7.

So, now we come to an end to this lecture. I have taken here very simple couple of examples, so, that I can indicate you or I can tell you that computation of conditional probability is not difficult and why and what are the situation where this type of probability can be used. Because this is only you who is going to take a call whether you want to find out the simple probability or the conditional probability.

So, now, I would request you that you try to take some examples from the book, try to solve them, practice them that will give you a better idea about this concept. So, you try to practice it and I will see you in the next lecture with more concepts and more topics on the probability theory. Till then, goodbye.