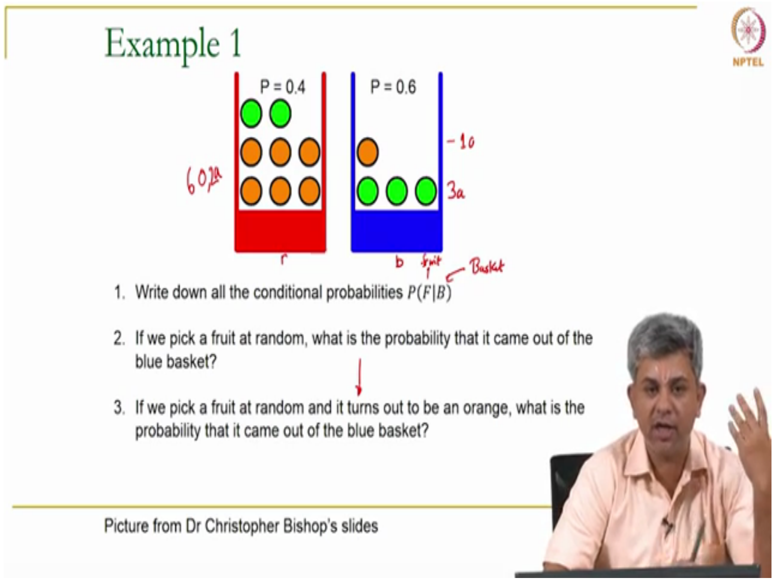
Machine Learning for Engineering and Science Applications
Professor Dr. Balaji Srinivasan
Department of Mechanical Engineering
Indian Institute of Technology, Madras
Bayes' Theorem - Simple Examples

(Refer Slide Time: 0:21)

In the last video we had seen quick derivation of Bayes' Theorem and in this video we will be looking at couple of very simple examples and both of these are discrete although in the course we will be using certain continuous distributions with Bayes' Theorem. These are very very elementary examples. So in case you are already comfortable with Bayes' Theorem, this is just supposed to be a review, so you can easily skip this video.

(Refer Slide Time: 0:38)

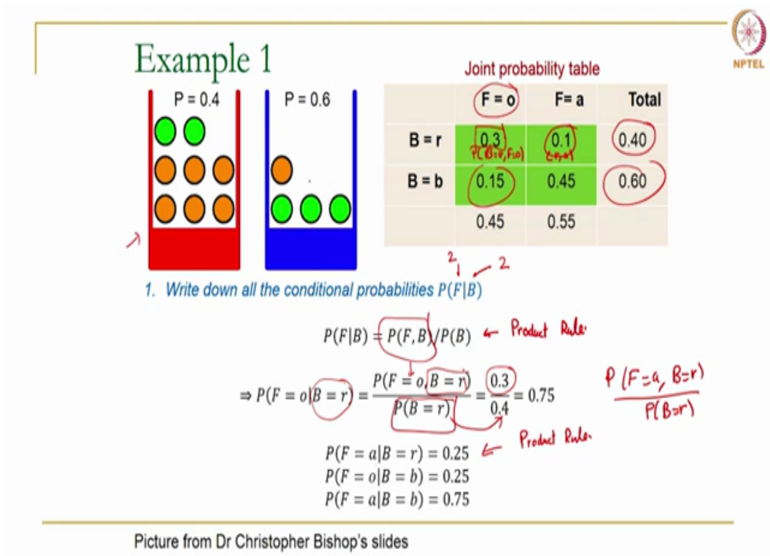


So let us return to the old problem that we were looking at, there were two baskets. One of them which was red and one of them which was blue and we were looking at the case where there are oranges and apples. And 6 oranges and 2 apples here; and 3 apples and 1 orange in the blue basket. And we randomly pick a fruit out of one of these baskets. What we know is that the blue basket is picked slightly more times, so we pick the blue basket with probability of 0.6 and we pick the red basket with the probability of 0.4.

So this was the background of the problem. And let us go a little bit further then what we did in the last video and let us ask a few questions. So the questions illustrated here are find out all the conditional probabilities. All the possible conditional probabilities for the random variable F, represented fruit and the random variable B representing the basket. Then we have a couple of more questions. If you pick a fruit at random, what is the probability? So let us say you dip hand in and you close your eyes and you do not know which basket you are putting your hand in.

And you are picking a fruit out of random. What is the probability that it came out of the blue basket? And the second one is essentially a variation of the previous question which is if you pick a fruit at random and you look at it and now you know that it is an orange that you picked, what is the probability that it came out of the blue basket?

(Refer Slide Time: 2:14)



Let us start with this. Remember that we had drawn the joined probability table. Actually we had drawn the joined distribution numbers but now I have turned it into a probability table. If you review from last time, we had a total number of 100. If you divide each of those numbers by 100, you get the probabilities. Remember what this represent. 0.3 represents the joined probability of the basket being red the fruit being orange, so et cetera.

We have this as the probability of red and apple et cetera, so you can see this. So let us address the first question. You want to write down all the conditional probabilities, P of F given that B is true. So as you can see this takes four different possibilities. We have two possibilities for the fruits and two possibilities for the baskets. So you could have an orange coming out of the red basket, you could have an apple coming out of the blue basket, so on and so forth.

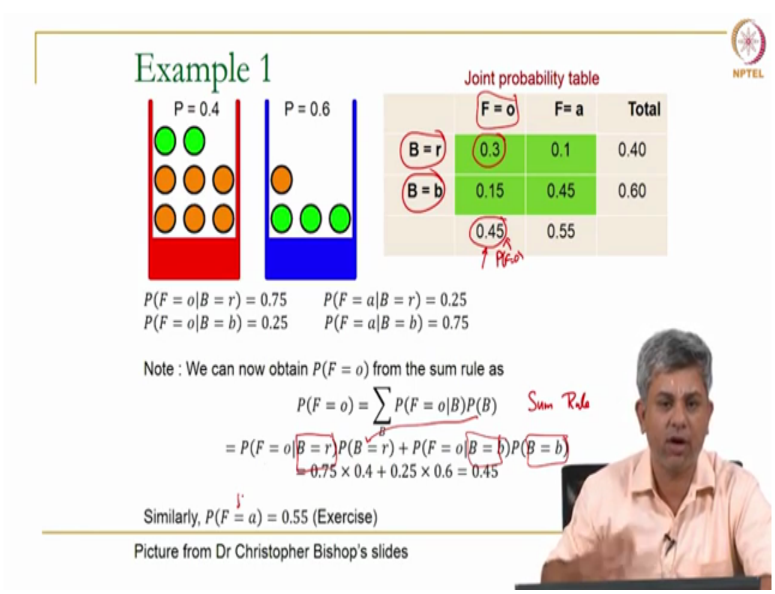
So how do we find this out? We have the elementary expression for conditional probability which we looked at in the last video. This is actually just a rewriting of the product rule. So if you use that, now you can take a specific case, probability that the fruit is an orange given that the basket is red. So what you are told is you have chosen the red basket, now you want to find out the probability that the few fruit is an orange.

Now you can do it in great detail here as is done right now. So the probability is given by P of F, B which is in this case P that the fruit is an orange and the basket is red, divided by the probability that the basket is red. And once you do that, we have this number. Remember this is the joint probability, probability that the fruit is an orange and the basket is red. This is 0.3 and the net probability that the basket is red is 0.4. So 0.3 by 0.4 is 0.75. This is the sort of detailed way of calculating this. We could have also calculated it in a simpler way.

You can kind of do it by inspection. Given that the basket is red, you already know that there are 6 oranges and 2 apples. So the probability becomes 6 over 8, so that is also 0.75. So that is the second way of doing it. Similarly you can now find out all the other cases. Probability that the fruit is an apple given that the basket is red, again you can do it by inspection, it is 0.25. Also, I would recommend that in case you are not yet comfortable with it, do it by product rule also.

So just to guide you through it, this would be probability that fruit is an apple and basket is red, divided by the probability that the basket is red which comes to 0.1 divided by 0.4 which is 1 by 4. So similarly fruit is an orange, basket is blue. Given that basket is blue, if you come here, this is 0.15 divided by 0.6, that is also 0.25. You can again do it by inspection also. Similarly probability that the fruit is an apple, given that basket is blue, 0.45 by 0.6 that is 0.75, again you can see this as 3 by 4 also.

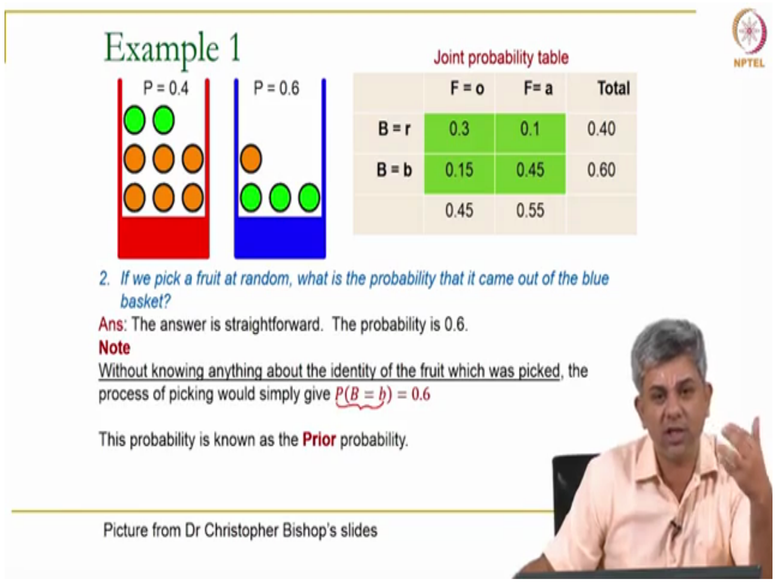
(Refer Slide Time: 5:49)



Now it is worth noting that you can obtain this number here which is essentially probability that the fruit is an orange using the sum rule. So recall that the sum rule was the marginal probability that the fruit is an orange can be given as P of F given B multiplied by PB, summed over all possible values of the basket. So in this case the orange fruit could have come out of one of two possibilities. Either it could have come out of the red basket or it could have come out of blue basket.

You can now write this in detail. If you got an orange, it could have either come out of the red basket and that probability we know, P of F equal to o given that B equal to r, multiplied by P equal to r. Similarly, the other case, basket is blue. So if you sum this up, it again retrieves the same number 0.45. Then why do this? Because I just wanted to show that you can recreate all these cases from joined probabilities or from conditional probabilities. You can similarly calculate the fact that P of F equal to a. Once again if you are not quite comfortable with this calculation, I would recommend that you do it in detail.

(Refer Slide Time: 7:15)

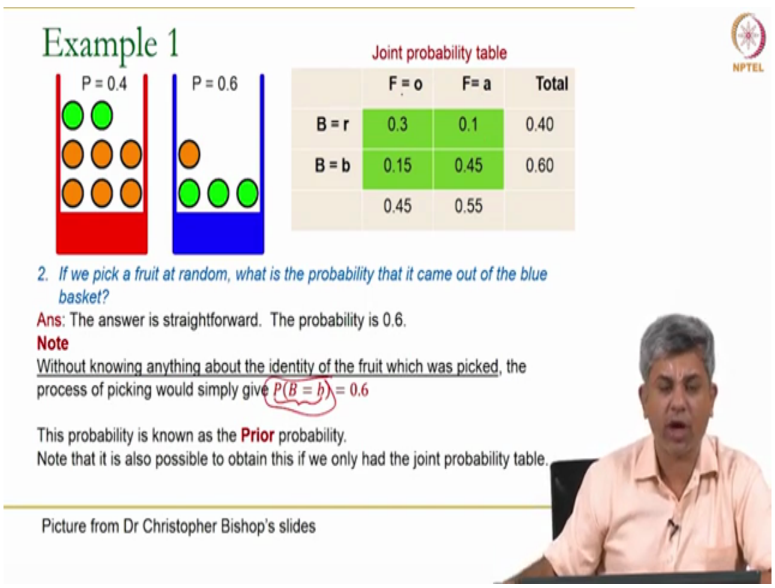


So let us move on to the second question. If you pick a fruit at random, so you just dip your hand and pick a fruit at random, you do not have any information about the fruit but you just want to know what is the probability that the fruit came out of the blue basket. Now this is more or less a trivial question. You know that the basket blue, blue basket is picked out at a probability of 0.6, so the probability is indeed 0.6.

There are a few things that I would like to point out here. The first thing is if you do not know the identity of the fruit which is basically this example, the process of picking would simply give you probability, this number, probability that the basket is blue which is 0.6. This probability is called the prior probability. So please note this term, prior probability. Prior here refers to the fact that you are telling a probability for picking that fruit or that which basket that fruit came out before knowing which basket it actually came out, before knowing sorry, before knowing the identity of the fruit.

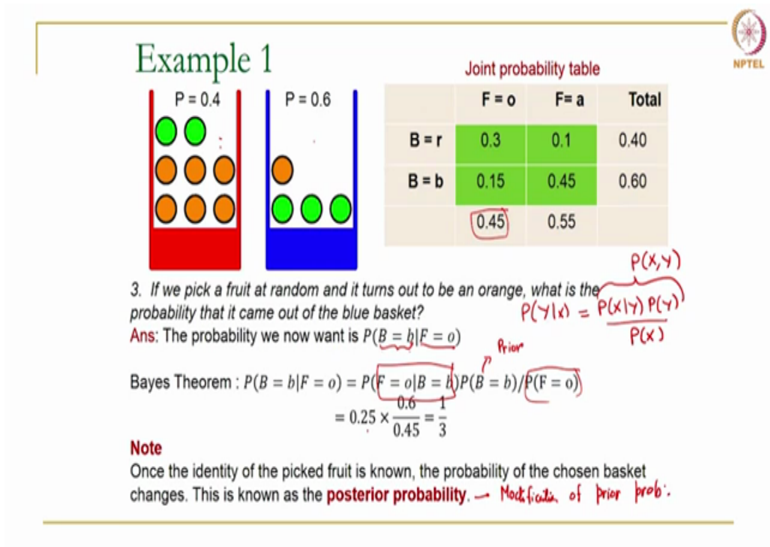
So this is prior to knowing the identity of the fruit, you are identifying the basket out of which it came and that probability is quite simple. In fact, it was given to you before. And this probability simply 0.6, it might seem like I am saying too much about something that does not need explanation but you will see that this is useful as we move on forward.

(Refer Slide Time: 8:39)



So once again you could have obtained this probability, P of B equal to b simply using the joint probability table.

(Refer Slide Time: 8:50)



Now let us move onto the next part of the same question which is, if you pick a fruit at random, so let us say you close your eyes, pick a fruit at random, look at it and now you know that it is an orange. Now you have some extra information before the previous case, before case 2 where you did not know the identity of the fruit or you had not yet looked at the identity of the fruit. Now if you look and see that this is actually an orange, now what is the probability that it came out of the blue basket?

The subtle thing here is that the probability changes because you have more information now. So to give you an example, for example I ask you, if you are walking on the road, what is the probability of meeting an Indian? Now if I give this question randomly to a person in the world, we know that Indians are about one-fifth of the population. So randomly walking on the road without knowing the information about where you are walking, your probability of meeting an Indian, a person of Indian origin is about 0.2, about one-fifth.

However if I tell you that I am walking on the road and I am walking in India, so now I have given some extra information, if I tell you that I am walking in India, what is the probability of meeting a person of Indian origin on the road? This becomes really really really high because majority of the population of India is basically in people of Indian origin. So similarly even though you pick the fruit, if you did not know what the identity of the fruit was, then the probability of the fruit coming from a blue basket is simply 0.6. This is the prior probability.

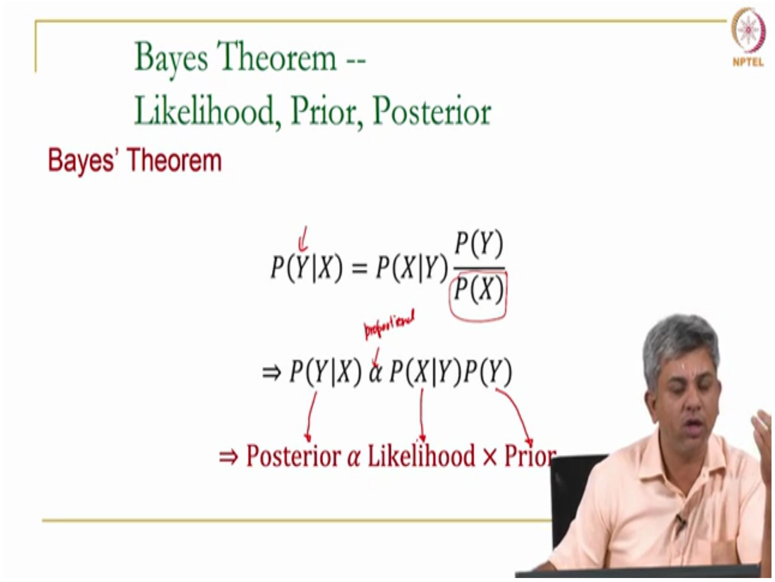
Now we are asking for what is known as the posterior probability, that is knowing the identity of the fruit and we know that the fruits are distributed differently in both these baskets. Now you want to know what the probability is that it came out of the blue basket. So mathematically we are asking for the probability that the basket is blue given that the fruit is an orange. So this is a classic case for Bayes' Theorem.

So remember Bayes' Theorem is P of y given x is equal to P of x given y, multiplied by P y, divided by P x. In fact, you might remember that this is just the joint probability of P of x, y. Okay. So we will use this. Here y is, basket is blue and x is essentially fruit is an orange. So if you write the expression out and evaluate it, recall that the probability that the fruit is an orange if the basket was blue, is simply 1 by 4 which we have written here. Next, the probability that the basket is blue, this is remember the prior probability. That is, before knowing anything what would you have said was the probability of the basket being blue, this is simply 0.6.

And the probability that the fruit is an orange here was something that we calculated earlier. This is 0.45, if you calculate this whole number, this comes out to one-third. So there are several ways of calculating it and again if you find something else is a little bit more intuitive, please do it that way just to reassure yourself that the calculation is correct. Most importantly this is a very very elementary example of the use of Bayes' Theorem. We can easily say what this probability is, probability that the fruit is an orange given that the basket is blue is clear and very easy to say is one-fourth.

The other one is a little bit harder and we usually use Bayes' Theorem in that direction. So the difference between question 2 here and question 3 is simply this difference of not knowing the identity of the fruit in which case the prior probability was 0.6 and knowing the identity of the fruit in this case actually reduced the probability to one-third. So you can see posterior probability as a modification of the prior probability. So we will use this viewpoint as we move on in the course. So before you knew anything, the probability was 0.6. After you knew something that comes out of the process, the result of the process, your probability is little bit modified.

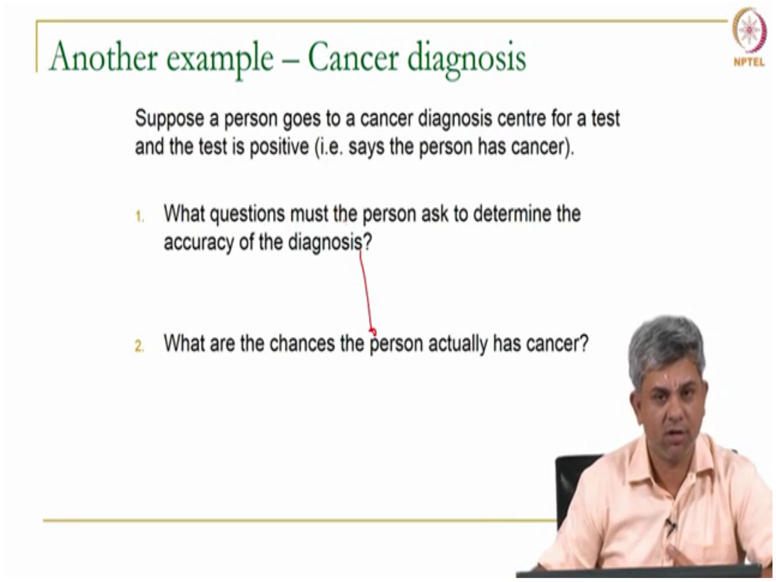
(Refer Slide Time: 13:17)



So here are some terminologies that we will be using for the rest of the course. Remember just a repetition of what Bayes' Theorem is, P of Y given X is P of X given Y, multiplied by PY divided by PX. Now notice that usually as we will see later on in the course, maybe about the fourth of fifth week of the course, we are more interested in the numerator in many many cases. So this is simply, this stands for proportional to as you might know.

So we know that the posterior probability is proportional to P of X given Y, multiplied by PY. Now each of these terms has a name. P of Y given X as I said earlier is called the posterior probability. P of Y, before you knew anything, before you knew about the state of particular X that you are looking at is called the prior probability. P of X given Y is called the likelihood. So this is some standard terminology that we will be using. So likelihood multiplied by prior is proportional to the posterior probability.

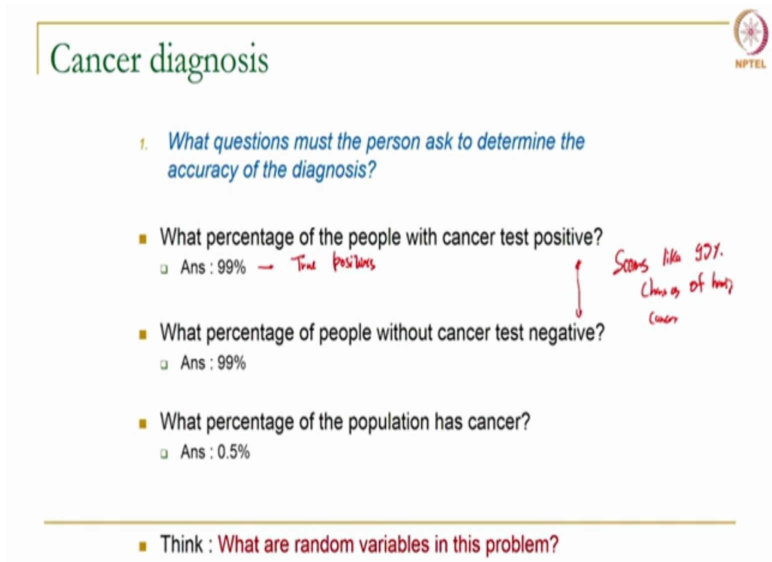
(Refer Slide Time: 14:25)



So let us take another example, again a very very simple example. If you have not seen this before, the results can be a little bit non-intuitive. So let us say a person goes to a cancer diagnosis center for a test and the test turns out to be positive, in this case positive actually means that the person has cancer. At least the test says that the person has cancer. Now obviously this person is going to be worried. Now what are the mathematical questions or at least numerical questions that the person can ask or must ask in order to determine the accuracy of the diagnosis.

And given that the test turned out positive, what are the chances that the person actually has cancer? Obviously the numbers here depend on the answers to the previous question. So let us look at this.

(Refer Slide Time: 15:10)



So one question that anybody would ask is how accurate is the test, but this is kind of a big question. So you would need little bit more specification in order to, specificity in order to answer the question. So you ask a more specific question, and let us say the person asks what percentage of people with cancer test positive? So this is an intelligent question to answer. And the person doing the test, the test agency tells this person that this is 99 percent of the people who have cancer do test positive.

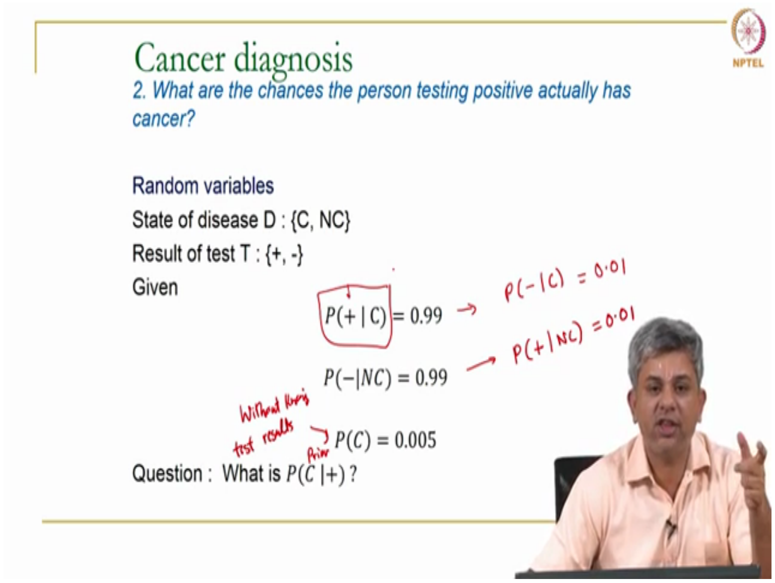
But this is not sufficient because the test could always say the person has cancer and you could still test positive most of the times. So instead you also need to ask. So these are true positives that is people who have cancer, who are testing positive. Now you also ask the other way, that is the flip side of this question which is how many percentage of people without cancer test negative. Now these two informations do not flow from each other. So please notice this and please think about this.

It is not obvious how many people without cancer will test negative. Like I said your machine could be broken and it could be always saying cancer in which case all the people without cancer will still test negative even if 99 percent people are accurate here. And the answer that the person says is that 99 percent of the people without cancer also test negative. Now this could be very worrying to the person because now more or less if you just go by intuition it seems like that seems like 99 percent chances of having cancer if you test positive.

But it is not true because you have to ask one more question which is the non-intuitive part here, which is what percentage of the population actually has cancer? Now this might not like I said, seem intuitive but you will shortly when we come to Bayes' Theorem, this is a very important quantity. So just as estimate, let us say that about half a percent of the population actually has cancer.

So let us go ahead and move and try and calculate via Bayes' Theorem what the chances are of having cancer. Before that we should ask the question, what are the actual random variables in this problem? Remember the basket example, in the basket example the random variables were which is the basket I picked and which is the fruit I picked.

(Refer Slide Time: 17:58)



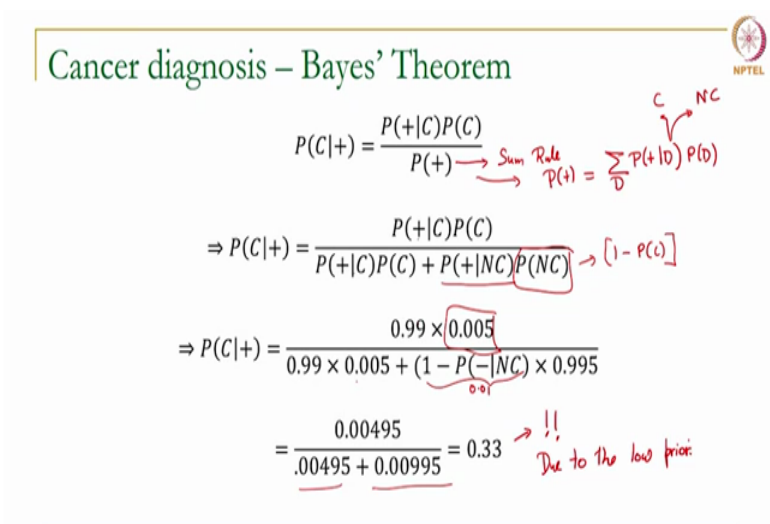
So similarly in the cancer diagnosis question the random variables are, does this person actually have cancer or not? So which I will call state of disease, D. D is a random variable, the person might have cancer or it might be a person without cancer. So C stands for cancer, NC stands for no cancer. The second random variable is the result of the test. The result of the test could either be positive, or it could be negative. So you have two random variables, the state of the disease and the actual result of the test.

So what are we given about the test? We are given the probability that you be test positive given that you have cancer is 0.99. So this is the first number that you have been given. The second number you have been given is that the probability that you will test negative, given that you do not have cancer is still 0.99. You can of course, find out the compliments of these two. So what is the probability that you will test negative, given that you have cancer? This is 0.01. What is the probability that you will test positive given that you do not have cancer? This is also 0.01.

This is 1 minus-, P minus- given NC. So the third thing that we asked for is PC which is what is the probability that a random person has cancer. Let us go back to the example that I told about finding a person of Indian origin while walking on the street. So finding the probability of finding an Indian person while walking down the street when you have given no context at all about where this person is walking is what is called prior probability and that would be about one-fifth.

Similarly without telling the origin of the basket from which the fruit came, if I just said I picked the fruit, what is the probability that it came out of the blue basket, that is the prior probability. That was 0.6. In this case without telling you what the result of the test is, without knowing test results, the prior probability of a person having cancer randomly out of the population is 0.5 percent which is 0.005. So the question we are actually interested in is the flip side of this. We know the probability of testing positive given that you have cancer. What we want to ask is the reverse question which is, what is the probability of having cancer given that I tested positive?

(Refer Slide Time: 20:41)



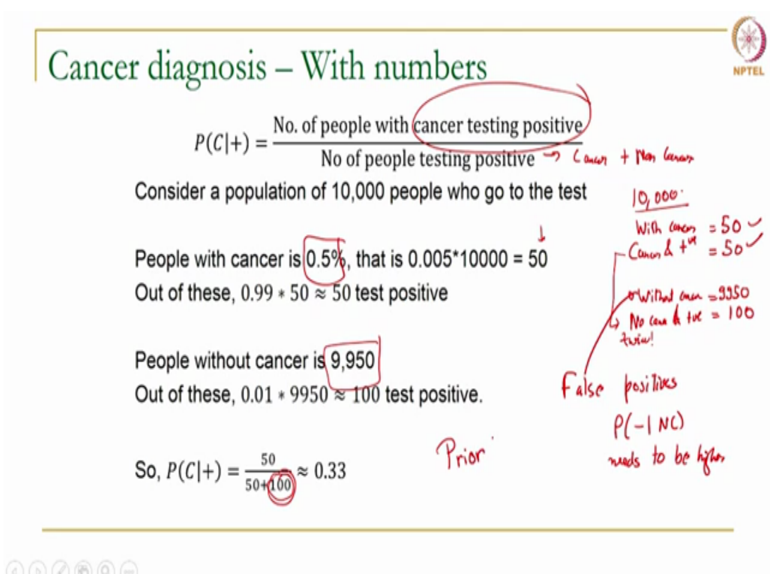
So let us use Bayes' Theorem as before. So probability of C given Plus+ is the same old Bayes' Theorem, plus given C multiplied by PC divided by P Plus+. Now this can be opened out using the sum rule of probability which we have seen a couple of times. SO we know that this is going to be all the possible states of the disease, okay. So if we open it out, diseases either cancer or the state of disease is you either have cancer or you do not have cancer.

So if you open that off, you get P of plus+ given C multiplied by PC plus+ P of plus+ given no cancer multiplied by P no cancer. So let us write these numbers out. This number we know to be 0.99. PC is the probability that a random person has cancer. This is the prior which was 0.005. P plus+ given C 0.005 notice that P of positive without having cancer was, as we saw in the previous slide, 0.01.

And the probability of not having cancer is of course, 1 minus- probability of having cancer. So this comes to 0.995. If you calculate these numbers, you will notice something, numerator is 0.00495. The first number here is 0.00495, the second number is 0.00995 and surprisingly enough you get probability of cancer given that you tested positive is just 33 percent. Even though the test was seemingly 99 percent accurate, we got a probability of only 33 percent that you have cancer given that you tested positive.

This is remarkable. And why does this come? If you look at the numbers here, you can find out that basically it comes due to the low prior. What do I mean by this?

(Refer Slide Time: 23:13)



So let us see this with numbers instead of probabilities. So if you use a numerical, sorry a ratio, frequentist approach to just finding out what this probability is. So probability of cancer even that you tested positive would be if you calculate it using number of people, it is the number of people with cancer who test positive, divided by the number of people who are actually testing positive. So what we want is this actually includes both cancer and non-cancer cases. So you go for a test. Some people who test positive, would have had cancer. Some people who tested positive would not have had cancer.

We want only those cases who tested with positive and that is the probability that we want. So let us consider a population of 10,000 people who go to the test. Out of these, if it is a random selection of the population who go to the test, out of these the people with cancer is going to be 0.5 percent which is only 50 people. So remember this net population is 10,000, people with cancer is 50. Now out of these people, people with cancer, the number of people who will test positive in the test, the test is extremely accurate, 99 percent of 50 is going to give you approximately 50. All 50 of them will test positive.

So cancer and positive is going to be 50. Test is nearly 100 accurate. So let us look at the other case. So out of the 10,000 people, 9,950 do not have cancer. They also go for the test and out of them 1 percent, not a large percentage, only 1 percent test positive. But since the number is large, this comes to 100 people will test positive. So please notice this. If you have cancer, only 50 people with cancer are testing positive and 100 people without cancer are testing positive.

So if somebody gives you the information that you tested positive, you can now see that that is not very significant information because 100 people without cancer tested positive. Twice as many people without cancer tested positive than people with cancer. So this basically the number comes out as this is one-third because this is twice. Now it tells several things. One is this number was this large because this number was so small.

If 50 percent of the population had cancer, both these numbers would balance out and you would get very reasonable numbers here. The other thing you notice is what affects the test really is false positives. That is, the true positives are very good. All the cases with cancer testing positive is very good but this data is being contaminated by all the false positives, all the people without cancer who are still testing positive. So if you want to improve your test, you need to change this number.

If this number is much higher, this number will become much lower. So this is another example of Bayes' Theorem. In our future videos, we will be looking at more direct applications of Bayes' Theorem through a continuous distributions et cetera. This was just supposed to be a simple review of how to use Bayes' Theorem. The important takeaway here is the importance of the prior. So as we will see later on in many cases priors are assigned arbitrarily but they can actually be a nice nob to turn in order to get the kind of results that you want. We will see from the fifth week onwards for this course. Thank you.