Essentials of Data Science with R Software-1
Professor Shalabh
Department of Mathematics and Statistics
Indian Institute of Technology Kanpur
Lecture-19
Bayes' Theorem

Hello friends, welcome to the course essentials of data science with R software. And now you can see that we have learned the concepts of probability theory in different ways, simple probability, conditional probability and we have used different types of theorems in the form of some results, additive theorem, multiplicative theorem and so on. And you have seen that these theorems results concept, they are going to be applied in different types of conditions, situations. And you can draw various types of conclusions from there, that depends on your capability, how much you can dig out the information through these tools.

Now, continuing on the same lines, today, I am going to talk about a very important result in the probability theory, this is Bayes' theorem. Believe me, this Bayes' theorem is very important and this is the foundation of the Bayesian inference. And you know that whenever we are trying to compute some probability, there are two possibilities that the event is going to happen for the first time, or the event has already occurred in the past. For example, if I ask you that, how much time do you take from going from your home to your college? If you are going for the first time, then you might not be knowing, but now since, you are going to the same college for a long time, so you will say okay, I will reach in say 20 minutes or so.

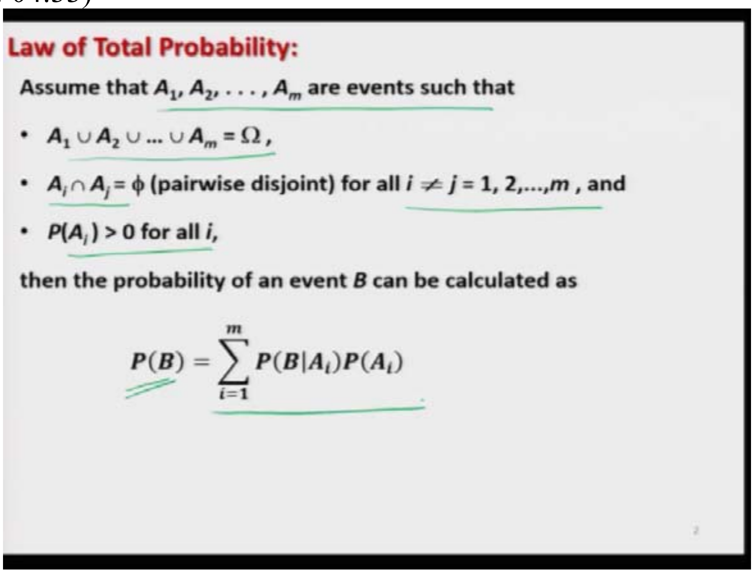
How do you come to this idea, or if, in fact, you can also say if there is no traffic then 15 minutes, if there is heavy traffic then 30 minutes, if and if there is a reasonable traffic possibly 20 minutes and so on. Now, if you try to see, if I ask you a simple question today that how much time are you going to take, then will you compute the probability only on one observation or you will try to use all this information somewhere inside your mind? Somewhere in the background, you will try to capture that information and somehow you will try to compute the probability and you will tell me that what is the probability.

For example, if I asked you that, how much time I am going to take, going from my hotel to the airport or railway station in a new city, means I am going there for the first time, but you are staying there for a very long time. So, you will ask me, at what time you would like to go? I will say, this is a 5 o'clock in the evening. So, that is going to be office time, people will be coming back. So, it is going to be crowded. So, you will so, whatever is the time, you will give me some more time, more time in the sense, the time taken at 4 o'clock in the morning. When there is no traffic usually, what is that you automatically computed that time taking into account the prior information. That prior information was there with you, and that information was collected either from you or your friends or somehow you came to know that at 5 o'clock, there is going to be a traffic, and it may take longer time and then you are trying to tell me that how much time I am going to take.

So, under these types of events if you try to compute various types of probabilities, they can also utilize this type of prior information. And once you are trying to use the prior information, things are expected to improve, because you are trying to incorporate more information into the system. If you try to incorporate more information, well, that information has to be correct, then you expect that your statistical inferences are going to be improved.

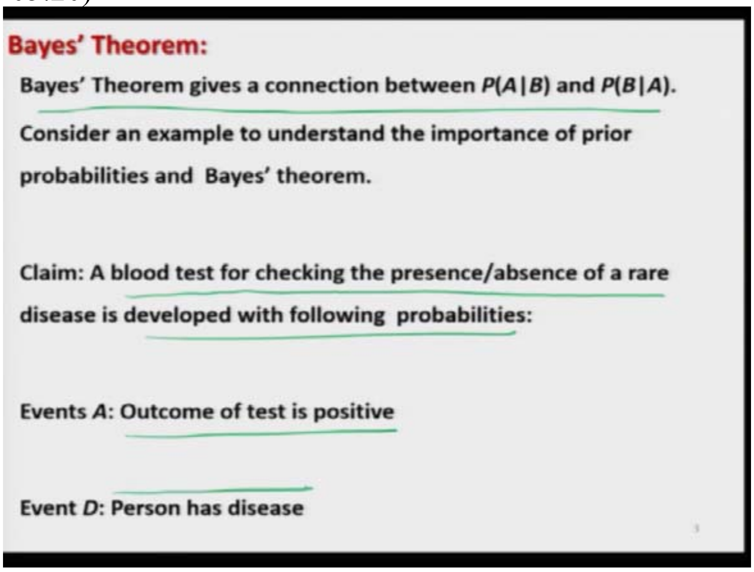
So, with this point of view, let me try to explain you here the concept of Bayes' theory and then I will try to take couple of examples, to make you understand better, I will try my best and you also try your best. So, with this objective, let us begin our lecture.

(Refer Slide Time: 04:33)



So, now you already have learnt one thing, what was the law of total probability. And this rule says that if, $A_1, A_2, \ldots, A_m$ are the events or that $A_1 \cup A_2 \cup A_m$ is equal to capital Ω and A_i, and A_j, they are pair wise disjoint, that is $A_i \cap A_j$ is equal to ϕ for all i not equal to j 1, 2 m. And P(A_i is positive for all i that is obvious, then the probability of an event can be calculated as P(B, which is equal to summation i goes from 1 to m P(B, given A_i into P(A_i).

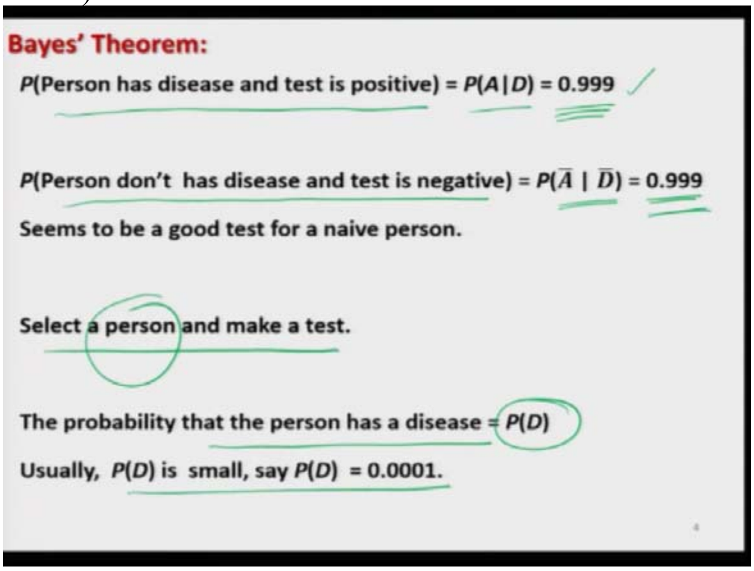
(Refer Slide Time: 05:20)



So, now, if you try to see, I will try to use this theorem in getting the Bayes' theorem. The Bayes' theorem gives a connection between the two conditional probabilities, properties of A|B and P(B|A). And in order to understand what this Bayes' theorem tries to give you, what type of probabilities it gives us, let me try to take a very interesting example and try to show you that, how it is used and what type of information it is going to give you and how it is different from the simple probability.

So, suppose, there is a blood test, and that blood tests claim that it can check the absence or presence of a rare disease. And there will be some probabilities, this I will tell you, and there are two possible events related with this claim. The first event can be, that the outcome of a test is positive and the event D, we are trying to define as person has disease.

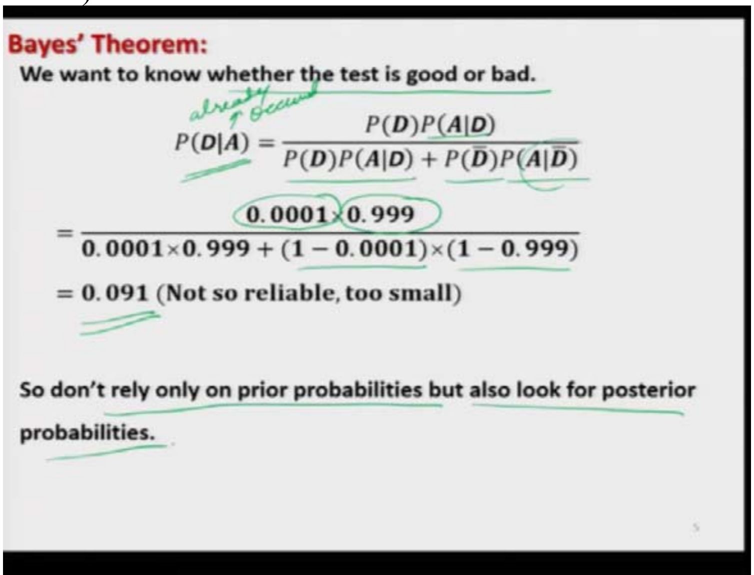
(Refer Slide Time: 06:30)

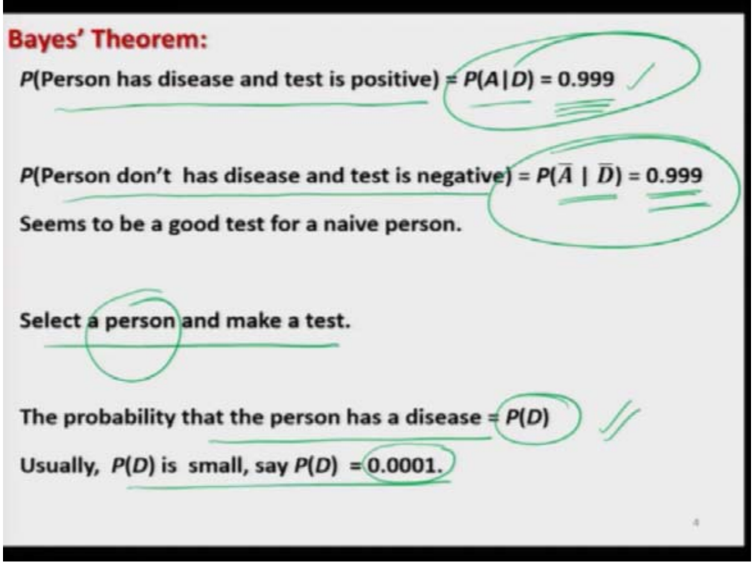


Now, I try to do here one thing, that suppose I have some information, that the probability that the person has disease and test is positive that the P(A|D) that is 0.999. And we have one more information, that probability, that person do not has disease and test is negative, that is $P(A^C \mid D^C)$, and this is also 0.999. Now, as a common man, if you try to see, you are trying to say that, in case if the person has disease and test is also saying that the person has the disease, that the test is positive, this probability is 0.999, that is almost 99.9 percent.

And on the other hand, in case if the person does not has any disease, then that test is saying negative, this probability is once again 99.9 percent. So, obviously, you will assume or you will feel that the test is excellent, test is very good. Now, what we try to do? We select a person and we make a test. And now this person, who is selected for the test, this person may or may not have that disease. So, we assume that the probability, that the person has a disease is P(D). So usually, this P(D) is going to be small. Why should I expect that every person has a disease? And suppose this probability is very small 0.0001.

(Refer Slide Time: 08:10)





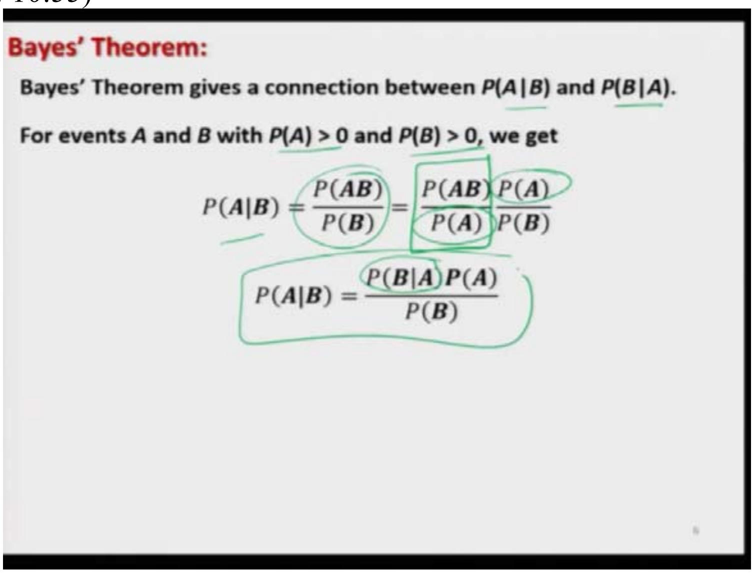
Now, we want to know based on the outcome, that whether the test is good or bad. So now, you know means the condition is this, you already know that P(D), that is the probability of disease is very small. Now, you try to compute this probability, P(D | A), that A has already occurred, or A is true. So now, this probability can be written as a P(D) X P(A|D)/P(D) into P(A|D) + P(D^C) X P(A|D^C).

So now, we have all these values and we try to substitute it here, we have this P(D) to be 0.0001. And the probability of this test is 0.999. So, we try to substitute it here and then P(D^C) and we will be simply one minus 0.0001. And $P(A/D^C)$ will be 1 - 0.999. And if you try to compute it, this probability comes out to be 0.091.

What do you think now? You are saying that test is nearly 9.1 percent that the test is good or bad. Now, if you are saying that the test is nearly 9 percent reliable only, what do you think that test is good or bad? And if you try to see, what you have said earlier, in these two probabilities here, and here, you said the test is 99.9 percent good. But when you are trying to compute the probability that whether the test is good or bad, given that the P(D is very small, this probability is coming out to be very less, nearly 9.1 percent.

So, that means test is not dependable. So, what we have seen here, that if you are trying to use some prior information, some prior probability, then this is going to change the outcome, this is going to change the result. And these types of probabilities which are known to you a priori, they are called as prior probabilities. And up to now, we were looking only on the prior probabilities, and the probabilities which you have now, computed they are posterior probabilities. They are called as posterior probabilities. So, now, the moral of the story is do not rely only on prior probabilities, but also look for the posterior probabilities.

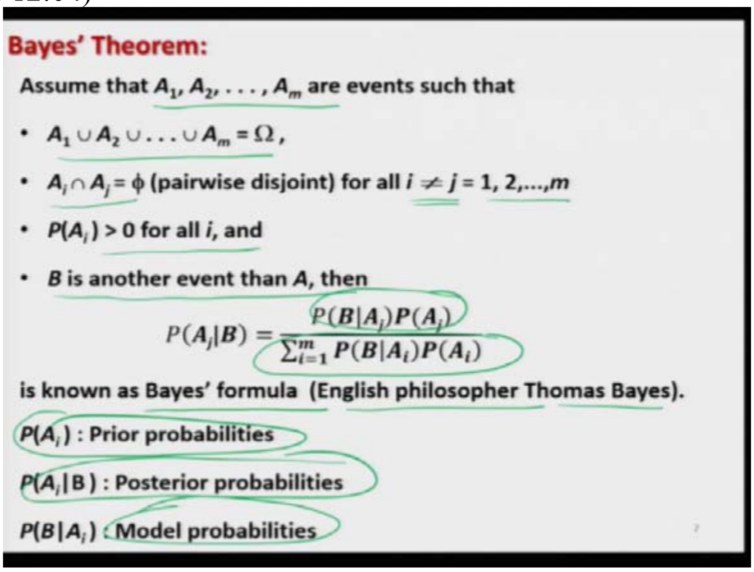
(Refer Slide Time: 10:55)



And this Bayes' theorem will help us in the computation of such prior probabilities. So, this Bayes' theorem gives us a connection between the conditional probabilities of A|B and B|A and for two events A and B with P(A) greater than 0 and P(B) greater than 0, we know that P(A|B) is P(A B)/P(B.

And now I can just multiply and divide this P(A and this expression, and you can see here what is this thing, P(A B)/P(A, this is simply your here the conditional P(B|A). So, now, you can see here that P(A|B) is equal to P(B|A) X P(A) / P(B). Well, you also had found such a result earlier by computing the P(A B).

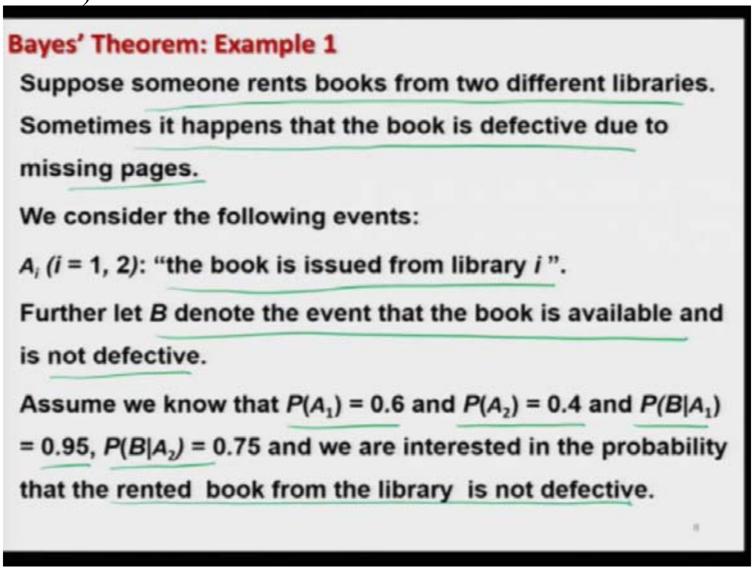
(Refer Slide Time: 12:04)



So now, in case if I try to extend it, then suppose we have here m events A_1 A_2… A_m like this, so that $A_1 \cup A_2 \cup \ldots \cup A_m$ that is equal to sample space Ω and A_i, and A_j pair wise disjoint, that is i for not, for i not equal to j goes from 1 to m, $A_i \cap A_j$ is ϕ. And $P(A_i)$ is, is positive for all i. And suppose B is another event then A, then in case if I want to find out the conditional $P(A_j |B)$, then using this result, which was only for two events, this can be extended to m events.

And we can write down here this $P(B|A_j)$ into $P(A_j)$ divided by summation of i goes from 1 to m, $P(B |A_i)$ into $P(A_i)$. And this formula is known as the Bayes' formula or this is the called the Bayes' theorem. And this is actually based on the name of English philosopher, Thomas Bayes, who had given this formula. And here these probabilities, what we have written here, this P(Ai, they are called as prior probability, $P(A_i)$ given B, they are called as posterior probabilities and $P(B|A_i)$, they are called as model probabilities.

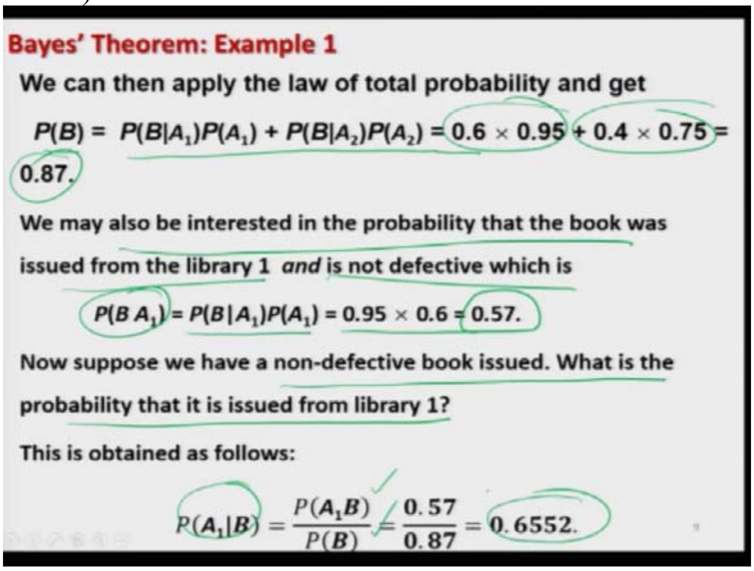
(Refer Slide Time: 13:32)



So now, we take some examples to understand the application of this law of total probability and Bayes' theorem. So, the first example is like this, suppose someone rents books from two different libraries. And sometimes it happens that the book is defective due to missing pages. So here, we consider the two events A_1 and A_2. So, A_1 is indicating that the book is issued from the library one and A_2 is indicating the event that the book is issued from the library 2.

And suppose B is the event that the book is available, and is not defective. So, now we assume that we know that $P(A_1)$ is 0.6, $P(A_2)$ is 0.4. And $P(B \mid A_1)$ is 0.95. And $P(B|A2)$ is 0.75. And based on this information, suppose we are interested in the probability of, probability that the rented book from the library is not defective, we want to know this probability.

(Refer Slide Time: 14:31)

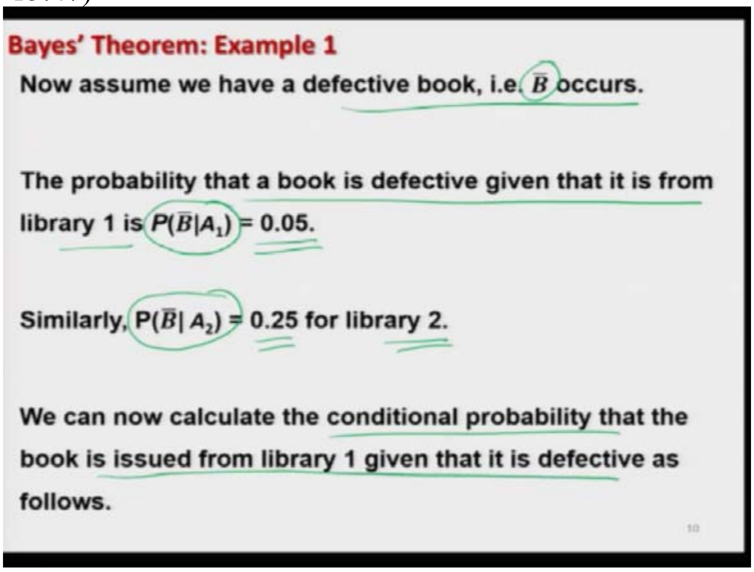


So, now if you try to apply here, the law of total probability, then we can use the formula here. That P(B is going to be $P(B|A_1)$ X $P(A_1)$ + $P(B|A_2)$ X $P(A_2)$. And if you simply try to substitute all these values over here, you get here the value to be 0.87. So, you can see here that this law of total probability is helping you in computing this type of probabilities.

And we may also be interested in the probability that the book was issued from library 1 and, and is not defective. So, this type of probability can be computed by computing the $P(B \cap A_1)$ which can be expressed in terms of conditional probability, such as $P(B|A_1)$ into $P(A_1)$, which is 0.95 into 0.6, which is 0.57. And now, suppose we have a non defective book issued.

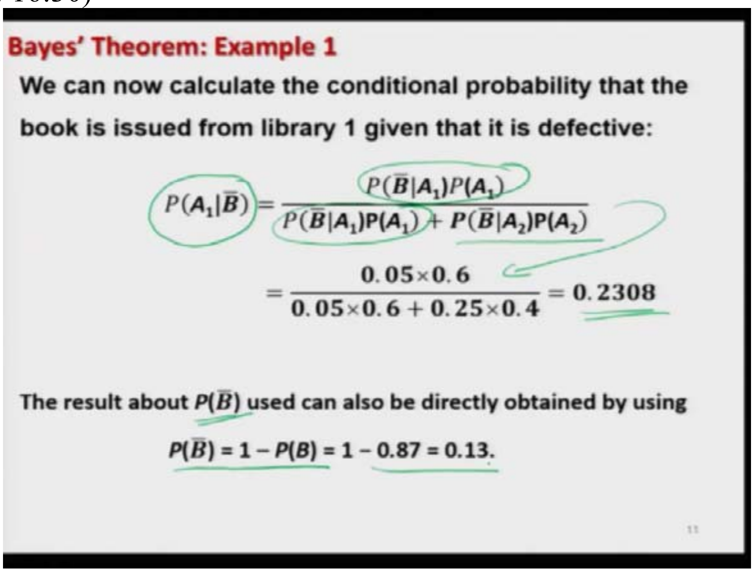
And suppose, we want to know, what is the probability that it is issued from the library one? So, this can be obtained by computing the $P(A_1|B)$. So, that is equal to $P(A_1|B)$ divided by P(B), then this can be computed here is 0.65 after substituting these values. So, you can see here from this example, I am trying to give you different types of events that can be computed.

(Refer Slide Time: 15:47)



And similarly, if you assume that we have a defective book, somebody has got a defective book that is B complement event occurs. Now, we want to know that, suppose the probability that book is defective, given that it is from library one is given $P(\bar{B}|A_1)$ is 0.05 and probability and suppose the probability that book is defective given that it is from library 2, this is probability of, conditional $P(\bar{B}|A_2)$. So, this is 0.25 for library two. Now, we can, now calculate the conditional probability that the book is issued from library one given that it is defective as follows.

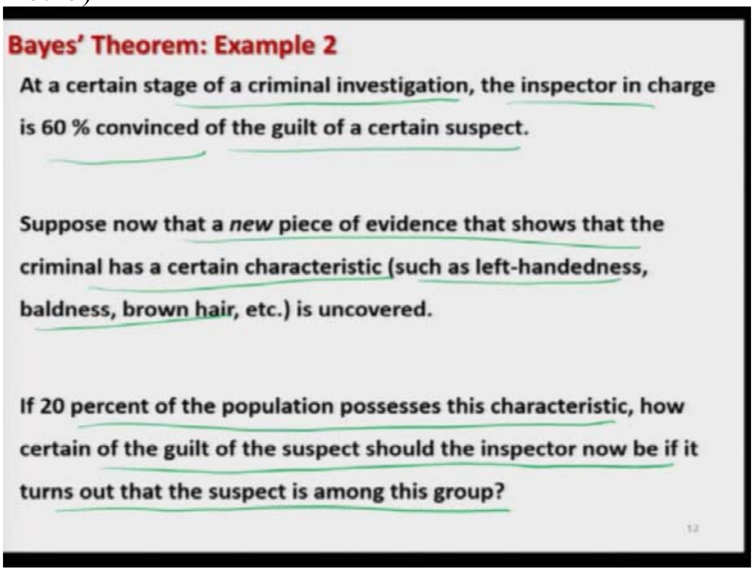
(Refer Slide Time: 16:30)



So, you can see here that we want to find out the $P(A_1|\bar{B}|)$. So, this can be written using the Bayes' theorem that $P(\bar{B}|A_1)$ into $\frac{P(\bar{B}|A_1)P(A_1)}{P(\bar{B}|A_1)P(A_1)+P(\bar{B}|A_2)P(A_2)}$.

And if you simply try to substitute all these respective probabilities over here, then you get the value here in 0.2308. Now, and similarly if you want to know the result about the P(B complement that we have used couple of places, then you can use this result directly P(B complement is equal to 1 - P(B) and you can obtain the value here is 0.13, no issues.

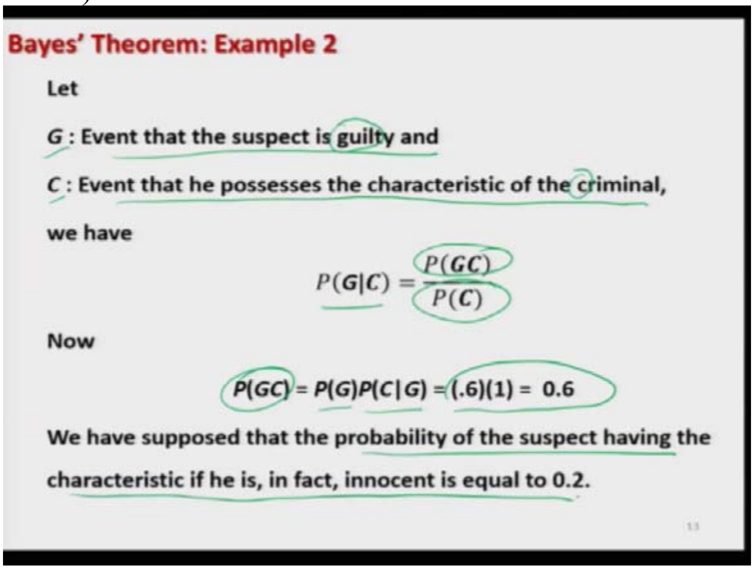
(Refer Slide Time: 17:15)



So, now after giving this, a couple of examples through this example one, let me try to consider one more example. Suppose, at a certain stage of a criminal investigation, the inspector in charge is 60 percent convinced of the guilt of a certain suspect, that somebody has made a crime and the corresponding police officer is 60 percent convince, that possibly that person has done this. So, suppose now, there is a new piece of evidence and that shows that the criminal has a certain characteristic such as the criminal is left-handed that means, he or she works from the left-handed, baldness that the number of hairs on the head, brown hair etc.

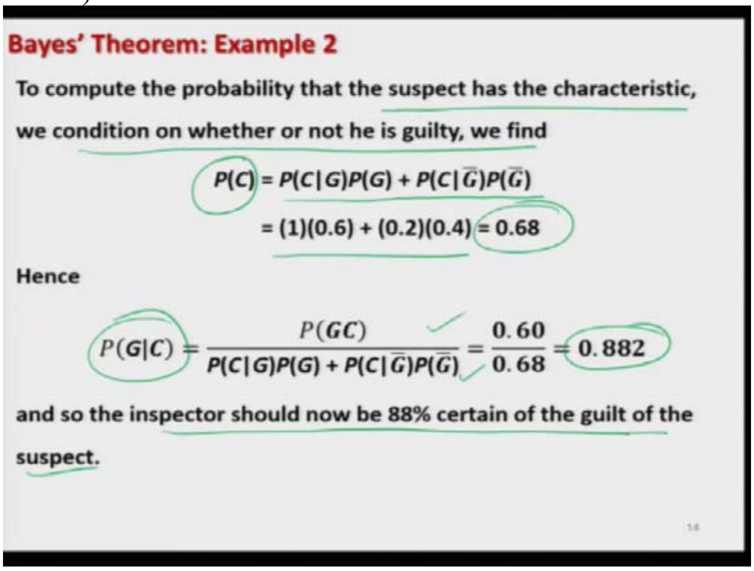
That type of information somehow becomes available from somewhere. And if 20 percent of the population is possessing this type of characteristic, we want to know that how certain of the guilt of the suspect should the inspector now be, if it turns out to be that the suspect is among this group. So, how to compute this type of probability, that you will see, that is a very interesting example, and you will try to see that how this Bayes theorem helps.

(Refer Slide Time: 18:30)



So, we try to consider here two events G and C. So, G is the event that the suspect is guilty. So, G is coming from the, G of this word guilty and sees the event that he possessed the characteristic, characteristic of the criminal. So, the C is indicating the criminal. So, that is easy for you to remember. So, now, we have the P(G |C) can be computed by P(G ∩ C) divided by P© and so P(G ∩ C) can be written as P(G) into P(C given G) which is equal to 0.6 into 1 which is equal to 0.6. So, now, we have suppose that the probability of the suspect having the characteristic, if he is in fact innocent is equal to 0.2.

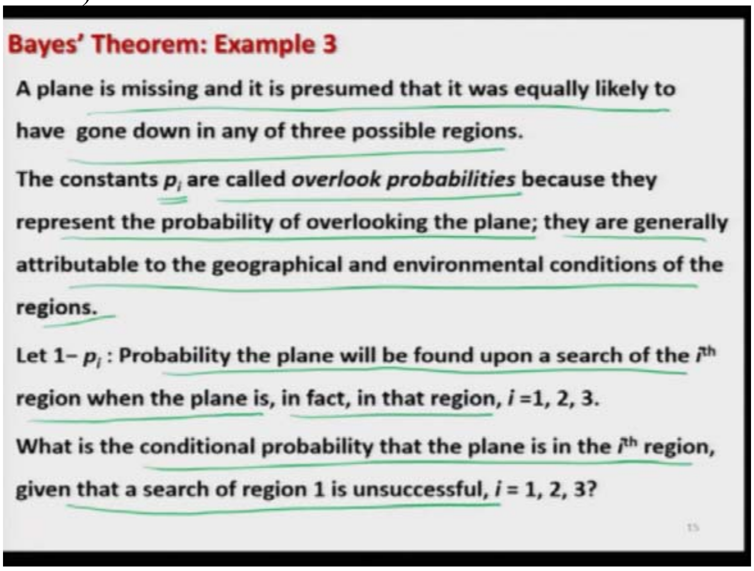
(Refer Slide Time: 19:18)



So, now to compute the probability that the suspect has the characteristic, we condition on whether or not he is guilty, and we try to find out P© which is the probability of being criminal, that can be obtained as $P(C|G)P(G) + P(C|\bar{G})P(\bar{G})$. And if you try to substitute all the respective probabilities, you get here the value 0.68.

Now, using these two values probabilities of C and here the P(GC, what you have obtained here, you can obtain the P(G given C and this comes out to be here, if you try to substitute all the values 0.88. So, now we can conclude that the inspectors should now be at 88.2 percent or nearly 88 percent be confident of the guilt of the suspect, so you can see that how Bayes' theorem has helped police personnel to investigate a crime.

(Refer Slide Time: 20:19)

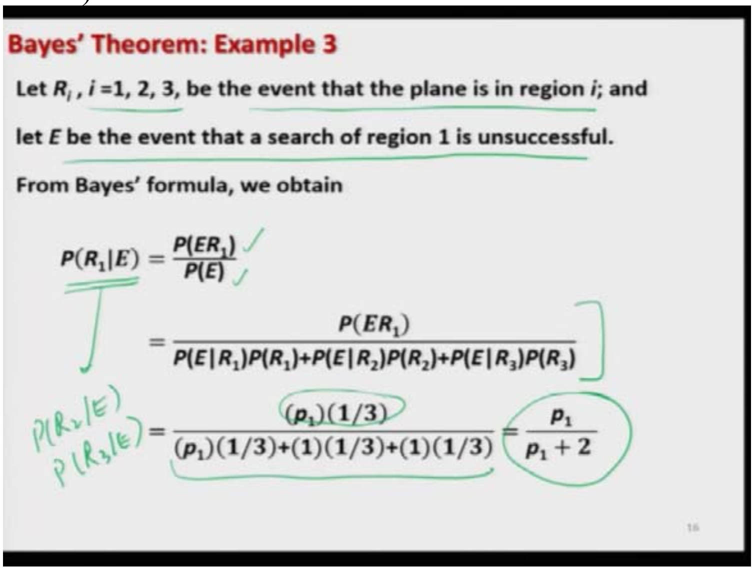


Now, let me try to take here one more example, to make this Bayes' theorem and this application to be more interesting. Suppose a plane is missing and it is pre assumed that it was equally likely to have gone down in any of the three possible regions, may God forbid that this may happen, but if it happens, then people are trying to have a search operation. So, this is based on those things.

So, in these cases, we have the probabilities actually which are the constants, which are p_i they are called as overlook probabilities, because they represent the probability of overlooking the plane and these probabilities are generally attributable to the geographical and environmental condition of the regions. So, this $1 - p_i$ is going to indicate the probability, the plane will be found upon a search of the ith region. And when the plane is in fact in that region, that means, suppose, suppose, we suspect that the plane has crashed in this region and we want to know it.

So, in such a case, we want to know the conditional probability that the plane is in the ith region given that the search of region one is unsuccessful, you can see, I mean these types of things are very common whenever there is an accident people try to look for such a site and they try to investigate from, starting from someplace or they want to compute this type of relativity, they will be very helpful to them.

(Refer Slide Time: 21:52)

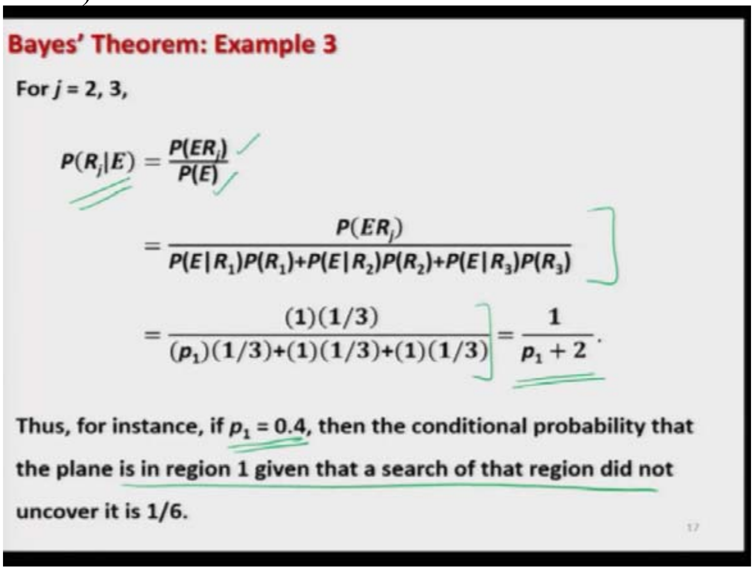


So, they tried to use the Bayes theorem to compute such probabilities and let us try to see what happens. So, let this R_i means R_1 R_2 R_3 these are the three possible regions. So, this R_i be the event that the plane is in the region I, and suppose this E is the event that a search of region one is unsuccessful, that is given to us. So, now using the Bayes' formula, we can obtain $P(R_1 |E)$, probability that the plane is in region one, given that the search of the plane in region one is not successful.

So, now this can be obtained with $P(E \cap R_1)/P(E)$. And now, using the Bayes' theorem, you can rewrite this probability here as like this, a simple extension and now, I try to write down here, these the values that, this probability P1 into 1.3 and I tried to substitute here all the conditional probabilities and the another probabilities of R_1 R_2 R_3 here and we try to compute this thing, this comes out to be p_1 upon p_1 plus 2.

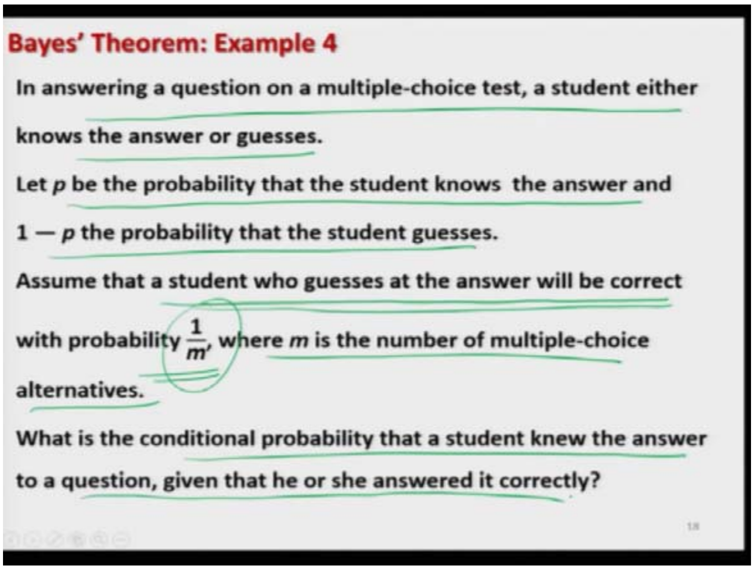
And similarly, we can also find here $P(R_2 |E)$ and $P(R_3$ | E also. So, that will also give us the probability of the event that the pain is in region two given that the search of region one is unsuccessful and similarly, the plane is in region 3, given that the search in the region one is not successful.

(Refer Slide Time: 23:37)



So, you can compute it exactly on the same line, the $P(R_j |E)$ is equal to $P(E R_j)/P(E)$ and if you try to simply substitute all the values as you have done earlier and try to substitute all the value, this will give you here the value 1 upon p1 plus 2. So, for example, if I say, if the value of probability 1 that is the p1 is 0.4, then the conditional probability that the plane is in region one given that the search of the region did not uncover it, is simply 1 by 6. You simply have to just put this value on there. So, you can see here, that this type of application will indicate that how you can apply them and different type of situation, which you are trying to handle in the data sciences.

(Refer Slide Time: 24:26)

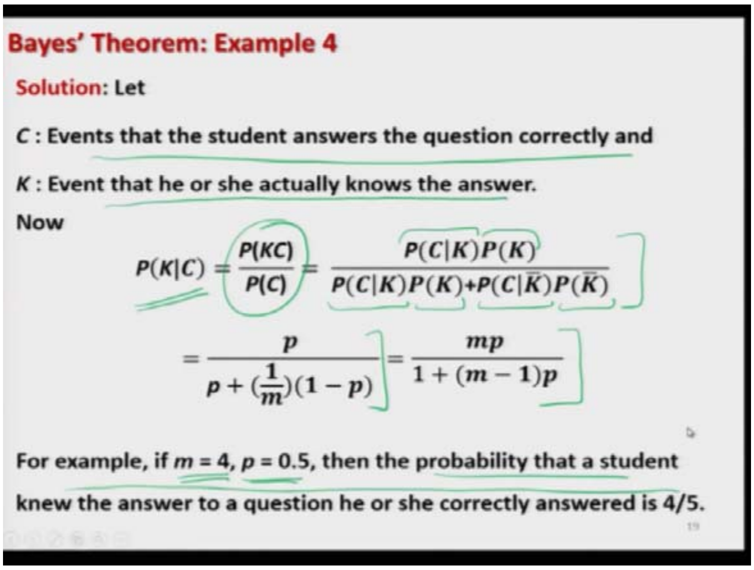


Now, I come to the last example on Bayes' theorem. And that is also an interesting example, that many students try to do it. Well, they have a very different type of belief and now, I am going to show them that their belief is right or wrong, you know that whenever we are trying to attempt a multiple choice question, in a exam, then we then sometimes we try to make a guess, and based on the guess we try to choose the correct answer. So, this example is related to that one.

So, in answering a question on a multiple-choice test, a student either knows the answer or the student makes a guess. So, let a small p is the probability that the student knows the answer. And 1 minus p is the probability that the student makes a guess, that will, that the student doesn't know it, as simple as that. So, assume that a student who makes a guess at the answer will be correct with probability 1 upon m, where m is the number of multiple choice alternatively. You know that as a student at some time, if there are 4 alternatives, and then you say many times that if there are 100 question and if you take at random to them, you will get 25 correct answer approximately.

So, these types of statements are very popular among the students. So now, you will see what really happens. So, this is the same statement that, what if a question has four alternatives, then we are assuming that a student who makes a guess at the answer will be correct with probability 1/4. So now, we want to know the conditional probability that a student knew the answer to a question given that, he or she answered it correctly.

(Refer Slide Time: 26:07)

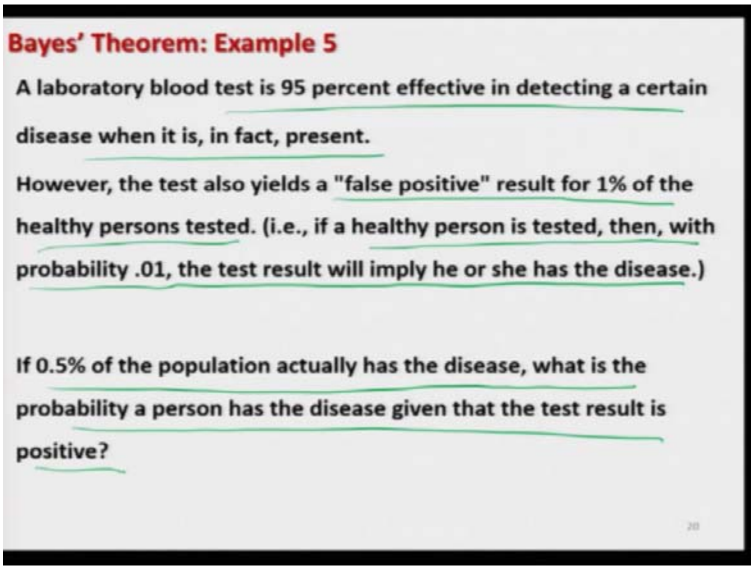


So now, that is a very interesting problem for you, I am sure. So now, suppose there are two events C and K. So, C is the event that the student answered the question correctly, and K is the event that he or she actually knows the answer. Now, we try to obtain here the conditional P(K |C), that mean the conditional probability that he or she actually knows the answer given that the student answered the question correctly.

So now, I can write it down here P(K ∩ C)/ P(C) and using the Bayes' theorem, I can write down here $\frac{P(C|K)P(K)}{P(C|K)P(K)+P(C|\bar{K})P(\bar{K})}$. And if I try to substitute all the values over here, which we have assumed, we get here this answer.

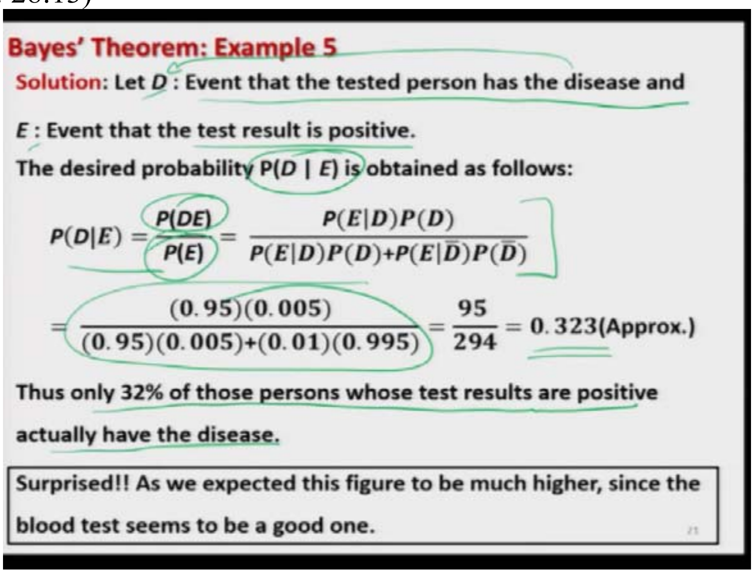
So now, if you try to see here, that if you try to assume that there are 4 alternatives, and if you try to substitute here m equal to 4, and suppose p is equal to 0.5, then the probability that a student knew the answer to a question he or she correctly answered is simply 4 upon 5. So, now you can see that possible, this is a different result than what we expected. So, that is the main use of this Bayes' theorem.

(Refer Slide Time: 27:27)



And now, let me give you here one more example. And suppose laboratory blood test is a 95 percent effective in detecting a certain disease, when the disease in fact is present. However, the test yields a false positive result for 1 percent of the healthy person tested. What is the meaning of false positive? That means if a healthy person is tested, then with probability .01, the test result will imply that he or she has the disease. So, if you assume that 0.5 percent of the population actually has the disease, then we want to know what is the probability, that a person has the disease given that the test result is positive.

(Refer Slide Time: 28:13)



So, this is the same example by which we started and now we are trying to come to this thing. So now, in this last slide, or the last example, we assume that here there are two event D and E. So D is indicating the event that the tested person has the disease, D means disease here, this is how I have defined it. And E is the event that that test result is positive. So, now the probability that we want to compute is given by $P(D|E)$ and using the Bayes' Theorem, we obtain a $P(D|E) = P(D \cap E)/P(E)$ and using the Bayes theorem, we can write it like this.

And we try to substitute here all the values what are given to us and we try to compute the probability, this comes out to be 0.323 approximately. So, it is indicating that only 32 percent of those person who test results are positive, actually have the disease. So, we expected that this figure to be much higher, because blood tests seem to be a good blood test. Good in the sense, that if a person has a problem, it will say yes, there is a problem or if not, the blood test will say no, there is no problem.

But there are 32 percent chances, which is not a small value. So now, I come to an end to this lecture with this Bayes theorem. Well, I have tried to give you the statement of the Bayes theorem. And I have taken couple of example to make you convince that how this Bayes' theorem is very useful in real application. Now, if you try to see, either this is conditional probability or Bayes' probability, anything, the entire probability function is going to be a function of these individual probabilities.

So, that is why it was very important for me to give you the information on different types of probabilities. And now, you can see that as you are going into deeper and deeper you are getting a better result. Now, if I ask you that for a given incident, if you have to choose simply the simple probability or the Bayes' probability or the conditional probability, you know, that how the results are going to be changed and how they are going to impact the outcome.

So, now I will say try to look into your assignments books and try to take some example, try to understand them and more important part for you to understand is how these individual values are coming, how are you going to translate that what is A and what is A complement what is A given B and what is B given A.

That is the job of a data scientist, the data will not raise its hand and say, in this case, you please find out the conditional P(A|B) or conditional P(A|B) complement and so on. Now, one thing more, that this conditional probability is just like the simple probability which is having some nice properties and all those rules about these conditional probabilities are there, but definitely if I tried to take all the topics, who are here, means we will miss some of that important topics.

So, I will request you, why do not you take up a book and try to compensate that part, which I am unable to do here in this course, my objective here is something else. So, I am doing it from that point of view, but definitely by saying this, I cannot say that the topic was 100 percent complete. So, that depends on you how much you now want to learn the statistics, so that you can use it in a much better way in our data sites. So, you try to practice it and I will see you in the next lecture. Till then, goodbye.