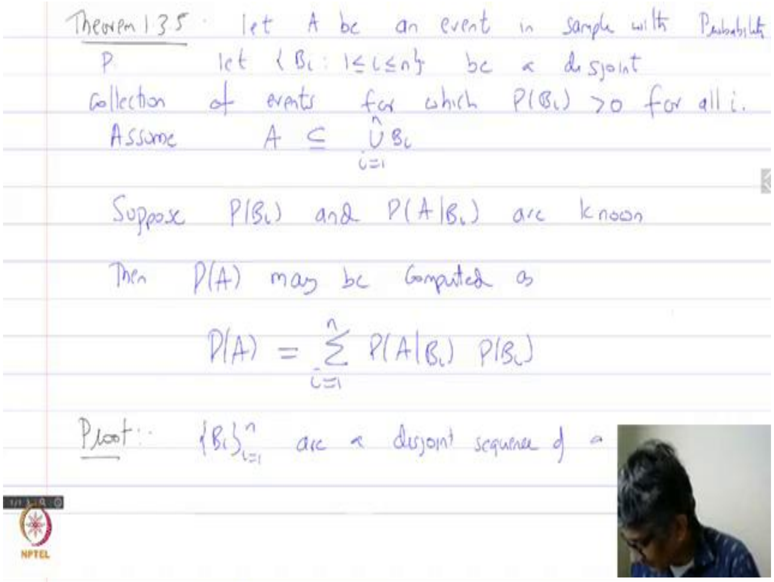
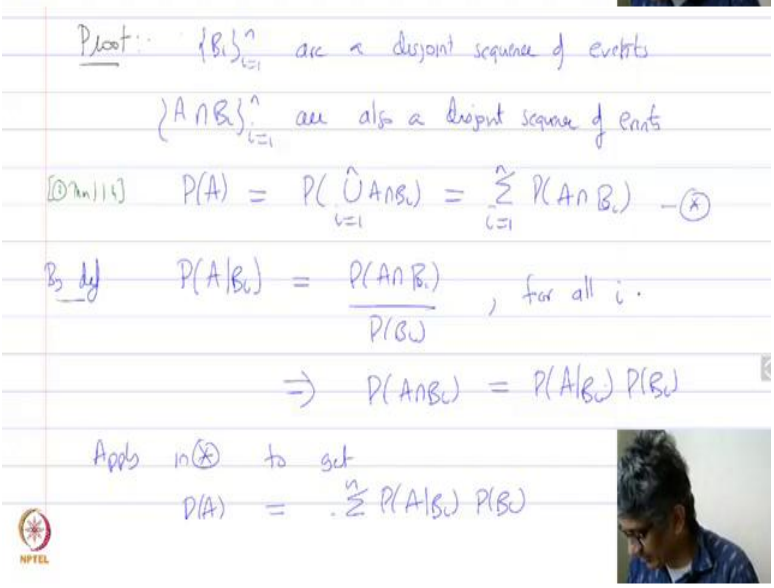
Introduction to Probability - With Examples Using R
Professor Siva Athreya
Theoretical Statistics & Mathematics Division
Indian Statistical Institute, Bangalore
Bayes Theorem

(Refer Slide Time: 00:14)





So, we will need the following result, we are discussing condition probability so far. So, we will need the following result. So, this is theorem 1.3.5. I do not need this, so, maybe I could change the text to black. Theorem 1.3.5 here, the idea is that I may need, I may have many events in my portfolio. So, I may have one event of interest. So, let A be an event, so, you think of A as the event of interest in sample space S with probability P, and then you let B sub I, 1 less than equal to I less than equal to N.

Yeah, a disjoint collection of events, events for which probability of B sub I is positive for all, that is the setup we have. And let us assume also further, that the event A is contained in the union of I equal to 1 to N of the BI. So, your all your BI's serve continually. So, you are partitioning the sample space in some way with BI and your A is contained in it. Suppose, probability of BI and probability of A given BI are known.

So, I know that, like in the previous example, I could compute probability of A given B and probability of B. Once, you know this, then there is a nice way, this going the reverse way. So, then probability of A may be computed using these two conditional problems. So, the probability of A is the same as the sum from I equal to 1 to N, probability of A given BI times probability of B.

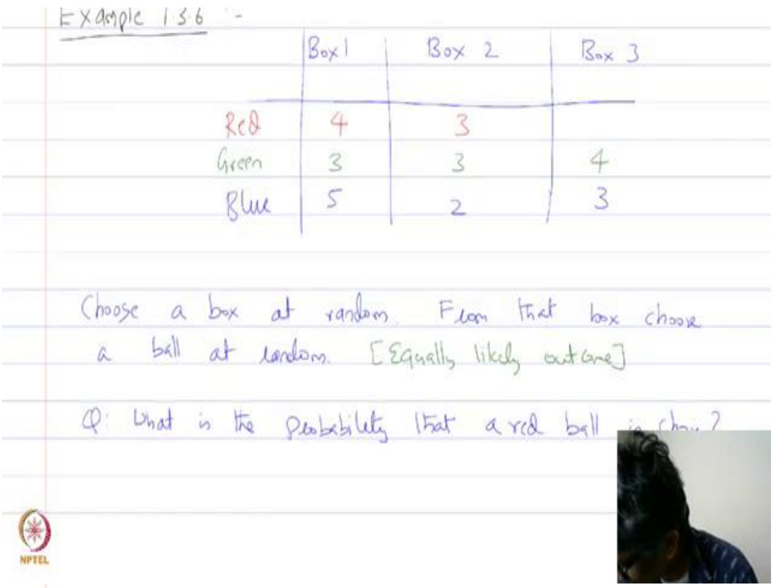
So, idea is that to go reverse. Suppose, I know probability of probability of B and BI and probably A given BI, can I compute the probability of A, this is a reverse idea. You just pause a minute and think about the case N equal to 1, it will be obvious because A is contained in B, and this obviously definition. But otherwise, let us just let us do it step by step. So, all you do is, so, I have BI, I equal to 1 to N or a disjoint sequence of events, you already know that.

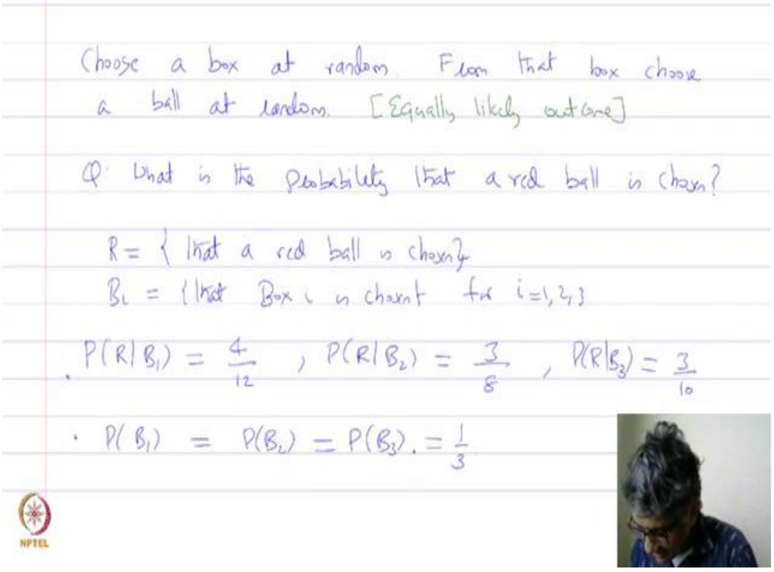
So, this implies, what does this imply? This will imply A intersection BI, I equal to 1 to N are also a disjoint sequence of events. Because, if BI are disjoint, there is no common element in the BI's, there will be no common element in there. So, now, this will imply the following, you apply that by that maximum probability or more or less the, I think it is equation 2 in this theorem 1.1.4 let us go 1.1.4 and we prove this fact that probability is finitely additive for that probability of A is the probability of union I equal to 1 to N, A and BI and that is the same as the sum from I equal to 1 to N probability of A and BI because they are disjoint sequence of events.

But now, you look at it this way. So, now, you know the following, probability of A given BI, by definition was going to be probability of A and BI. So, now A and BI divided by probability of BI, this we know by definition but that is same as saying probability of A and BI is equal to probability of try let me just do step by step let us go back, that is same as saying multiply probability of A and BI is the same as probability of A given BI times probability of BI.

So, therefore, what do we obtain, you apply this in star, let us call this a star. So, you apply in star to get what, to get probability of A is the same as the sum from I equal to 1 to N, probability of A given BI times probability of BI, and that is what we want. So, this is true for all that, that is right. So, simple idea, that if you have if you know conditional probability, you can recover the original probability as well and vice versa. This is kind of useful idea to have.

(Refer Slide Time: 08:22)





Okay, so, let me do an example, to illustrate this idea. Suppose, we have three boxes and in the three boxes, we have let us say, a collection of balls. So, we have three boxes, let us say three or three boxes, box 1, box 2, and box 3, there are three boxes and in each of these three boxes, let us say I have three sets of balls. Let us say, I have a blue set of blue balls, let us say I have, let us have blue balls here. So, I have blue balls, let us say I have 5 blue balls in one box, 2 blue balls in the second box, 3 blue balls in the third box.

And, let us say I have a bunch of red balls. So, red balls I have, let us say I have 4 red balls, I have 3 red balls, and 3 red balls. And let us say, I have green, let us say I have 3 green balls in this box and the three in this box, and in this, I have 4 that is going to put the red in this one, let us have put three, 4 3 3, 3 3 4, and 5 2 3, that is what I have in the box. Now, what I do is, the experiment I do is, I choose a box at random that means I choose one of the boxes in an equally likely manner, and from that box choose a ball at random.

So, when I mean this, I just mean that I am is equally likely out constructed, so that is all. This means that we have equally likely outcomes in both cases because each box is equally be chosen and each ball is equally chosen. So, once I have it, so, then the question I ask you is the following.

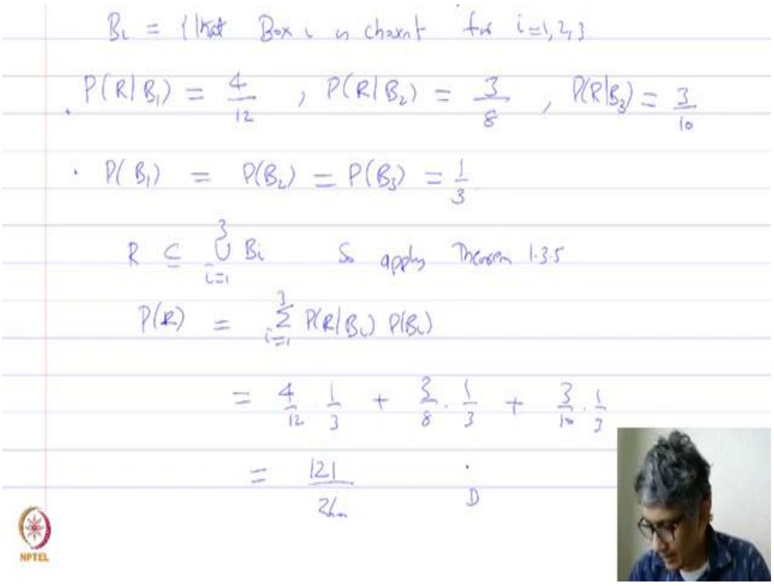
What is the chance or the probability that a red ball is chosen? So, how does one do this? So, you are choosing a box first and then choosing a ball next and then I want the end result that I have a red ball that is chosen? So, how does one do this? So, what one does is, you let us say set up notation first, let us say R be the event that a red ball is chosen.

And let us say BI be the event that box I is chosen for I equal to 1 2 3. And now, you observe a little bit carefully. So, now, if you look at this about calculation, what happened, our calculation see there are three boxes and so on so for. Now, if you look at this, if you look at, if I go to box 1 and I try to choose a ball and I want it to be red, I have how many 4 plus 3 plus 5 that is like 7 and 12 and the chance between a red ball is 4 by 2. So, that means the obligation is that probability of red given B1, that means there are 12 balls there and there is 12 in total and I have 4 likely red balls.

So, that means 4 by 12, probability of red given B2, I go to box 2, I have 3 plus 3 plus 2 that is 8 at the bottom and I have 3 favorable guys, probability of red given B3, is going to be again equal to, I have 3 plus 4 plus 3 that is 10 and I have 3 is my, I have 3 at all my favorite event. And, I also know this is one possibility and I also know the chance of B1 is equal to chance of B2 is equal to chance of B3 is equal to 1, this both these things I know for sure.

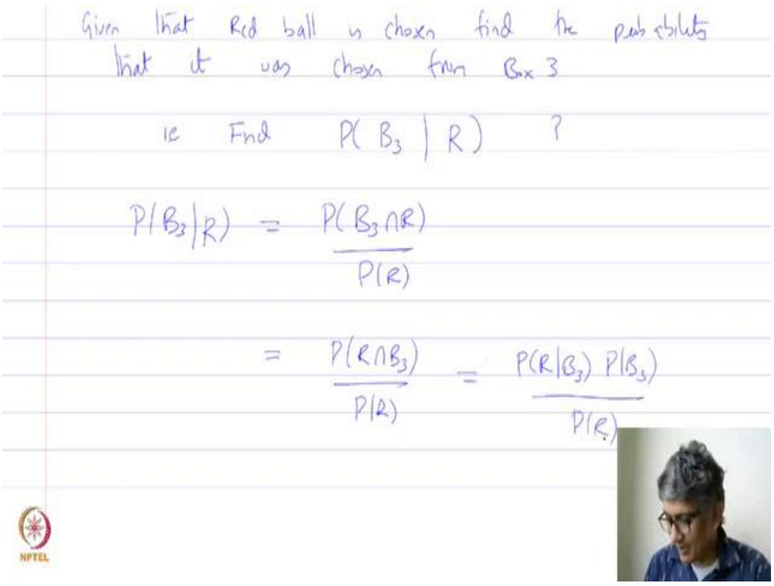
So, that is just from the model. So, I have three boxes once I pick a box, I know these conditional probabilities, and once I and you know I pick each box with these things. So, now, we are in the center of this theorem.

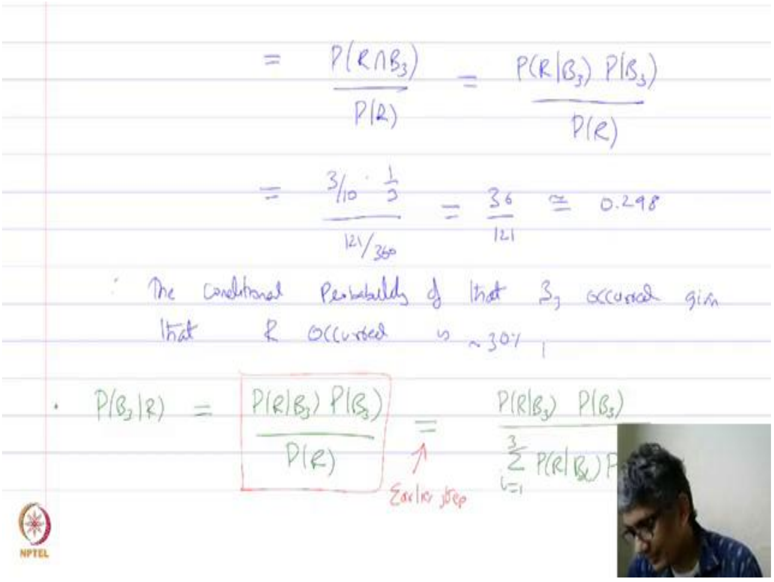
(Refer Slide Time: 14:13)



So, now, we also know the following. The event R is also contained in the event of union I equal to 1 to 3 BI's because this once a box is chosen, you choose a ball from that. So, red ball is chosen is contained in that. So, you apply theorem 1.3.5, what you can get probability of R is equal to the sum I equal to 1 2 3, probability of R given BI times the probability of BI and that turns out to be 1, that turns out to be, you go and plug in all the calculations. This sum is 4 by 12 into one-third plus 3 by 8 into one-third plus 3 by 10 into one-third and give up on calculation and then you will end up getting something like, it is 36 24 30 I think 360 is l, I think it is 121. Transferring a red ball is 121 by 360. So, that is one simple calculation one can do this.

(Refer Slide Time: 15:41)





Now, the same example. Suppose, I twist a little bit. So, suppose, I take my favorite event R. So, let us say, I want to understand, given that the red ball is chosen, can I find the fact that the third box was, it came from the third box. So, given the red ball is chosen, can we find the probability that it was chosen from box 3. So, what do I want to find? So, the in-probability term, that is you want to find the conditional probability of B3, given that a red ball was chosen.

So, this is sort of an interval, because now, I am going reverse. So, here I went forward, I said I had these boxes, add these collections, I knew the R given B1, R given B2, R given B3 because that is the conditional probability of an equal likelihood experiment. So, I know exactly what is going on and then I want to know what the probability of R is, now, that is easy to do. But now, I want to go to the reverse, I want to find probability of B3 given R. So, what you do is, you do a little trick.

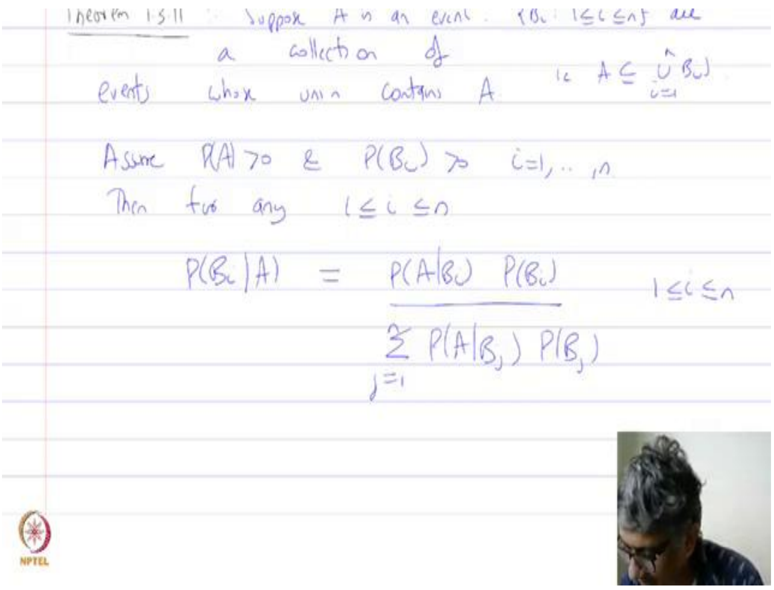
So, probability of B3 given R is the same as what is definition is B3 and R divided by probability of R. This you already found out probability of R, the top here to find out, but top you just flip it a little bit right, you flip it as probability of R and B3, the same event divided probability of R, but that is the same as probability of R given B3 times probability of B3 divided by probability of R.

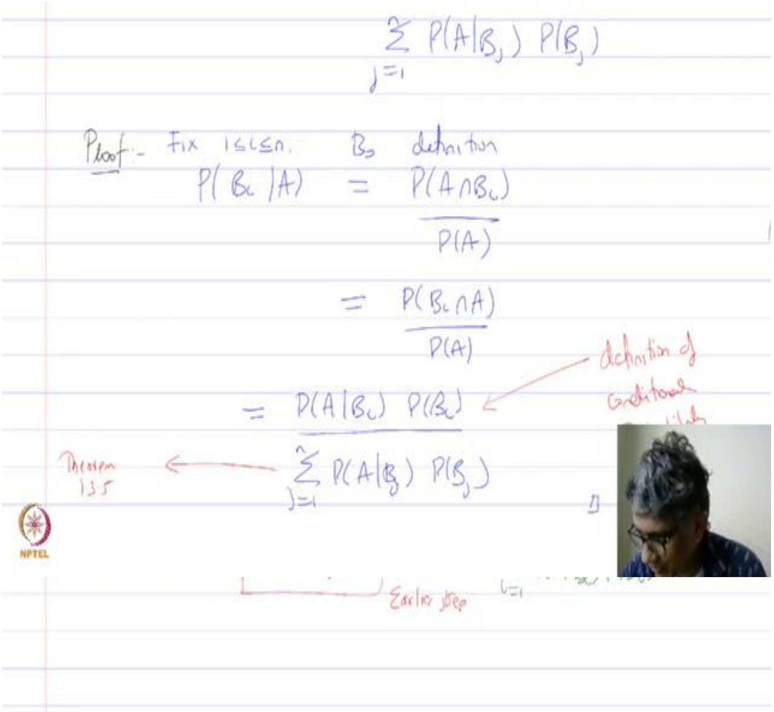
So, you know both these things, but that you know exactly wrong, you know R given B3 is 3 by 10, that is what we did before. Let us go back and check, R given B3, there are 10 balls and three of them are black, red, so, 3 by 10 come back down, probability of B3 is point 1 over 3 and probability of R which is calculated as 121 by 360. So, you do this and then you get 36 by 121. And that is the same as something like point 298. Therefore, the conditional probability of choosing box three by the different probability that, that box three was chosen or B3 occurred, given that R occurred, that is the red ball was chosen is around 30 percent.

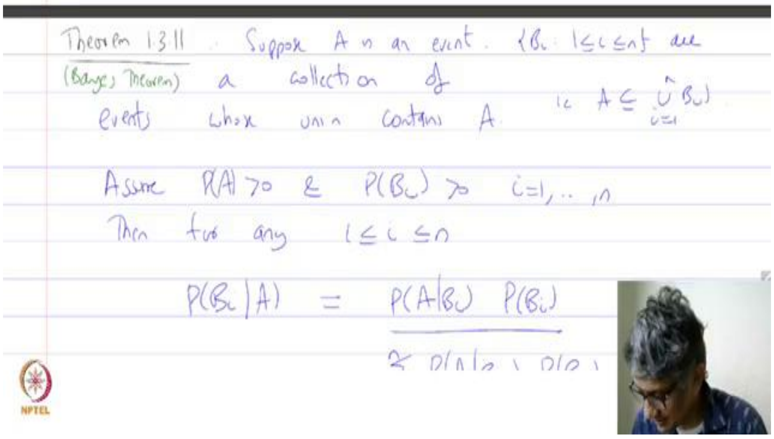
So, just one observation before I do this, we will go on to the general setup of this idea. The idea is quite general. So, the idea was that we took, we wanted to know probability of B3 given R and then we did some calculations and then we turned out that we got it as probability of R given B3 times probability of B3 divided by probability of R. This is just a standard sort of calculation, then if you notice in the earlier step, what we did was we already done this in the top, we had probability of R given B3, that will be there, that we knew and probability of B3.

And we have done this following thing, that to calculate probability of R, we had done this as summation I equal to 1 to 3 probability of R given BI times probability of BI. That is what, this is crucially we use this angle. So, this exactly, we did not use chance that okay, we could have skipped this step. So, this step is going to skip, so, even if you did not want to do this step, you could have done, you could just do the last step and this is what we missed, and that is sort of a, so, we do all the conditional probabilities and the probability of BI's to get the reverse conditional problem that is of BI given B3 given R. So, this is a part of a good theorem probability called Bayes theorem, which I will state next.

(Refer Slide Time: 21:41)







So, here is the theorem. So, the theorem is the following, let me write in black as before, let us keep convention straight. Theorem 1.3.1, suppose, A is an event and BI's such that 1 less than equal to I less than equal to N are a collection of these joint events, oh sorry I have to use blue, let us go back and fix this, the joint events whose union contains A. So, here, so let me just make it blue for consistency. Suppose, A is an event and BI 1 less than equal to I less than equal to N are a collection of events, eventual union continues, that is A is contained in a union I equal to 1 to N of BI.

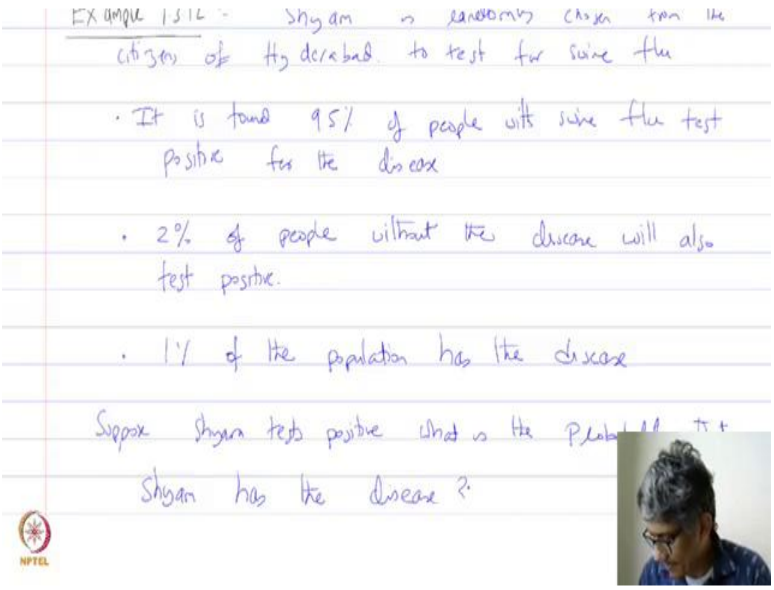
Let us assume, A is probability of, A is positive and probability of BI's are positive I equal to 1 to N, then for any 1 less than equal to I less than equal to N probability of BI given A is the same as probability of A given BI times the probability of BI divided by the sum from, I equal to 1 to N on still equal to the J, so index is clear, J equal to 1 to N probability of A given BJ times the probability of BJ, this is true for all I believe.

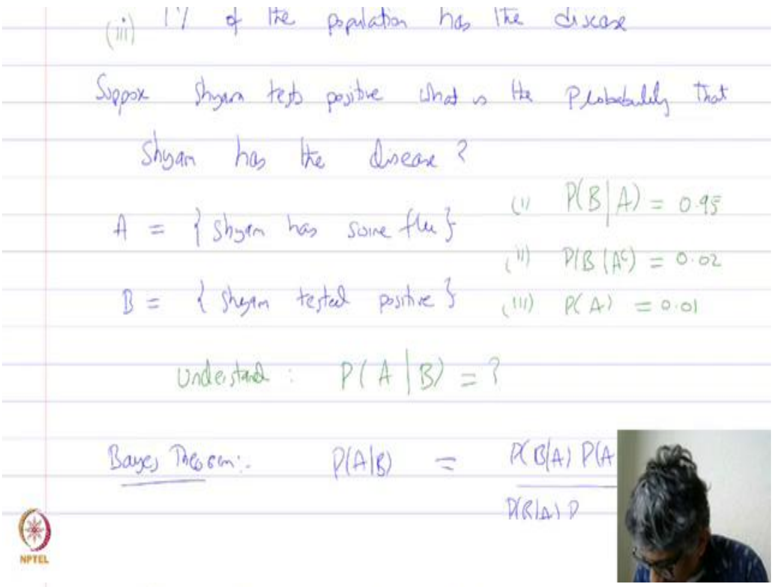
So, if you know probability A given BI's and BI you get already BI to N. So, prove already, we have done in some sort in the previous step. I will just rewrite the group properly, the proof is the following. So, probability of BI is given A, again we know the same as probability of A and BI divide by probability of A, this is just conditional probability. So, you fix an I between 1 and N by definition. So, by definition, is true.

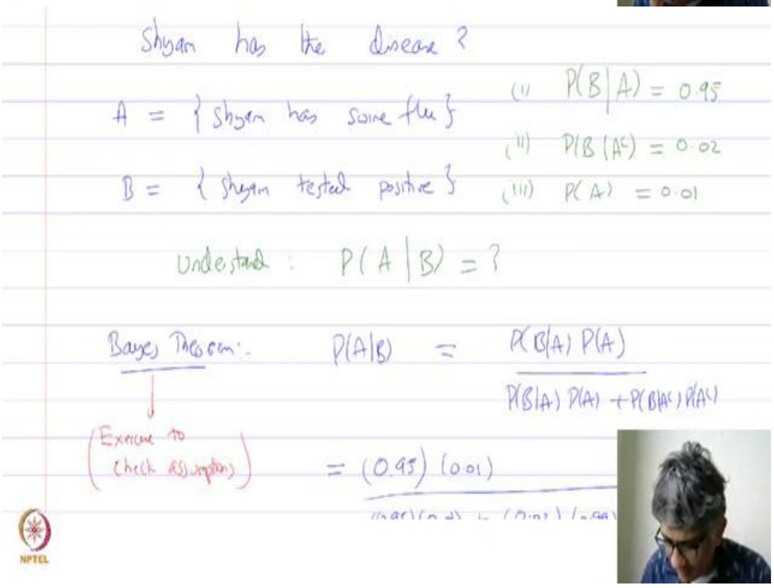
Now, that is the same as again like we did before, this BI and A divided by probability of A. Now, you do two steps. So, think of this as probability of A given BI times probability of BI that is just by definitely probability and the bottom you use the theorem 1.3.5, it says that the same as J equal to 1 to N probability of A given BJ times probability of BJ, so, it is a simple idea. So, in the bottom, we have used this theorem 1.35.

So, what have we done? We have done two things, in the top part, we will use the definition of conditional probability, and in the bottom part, we have used this theorem. That finishes off the proof, that is simple proof. But this is kind of a powerful theorem. So, we will illustrate this original, you do is called Bayes theorem. So, we should write down in green, it is a mathematician called Thomas Bayes, it is called Bayes theorem. And I think in the early 1700s, okay if not in this notation but probability notation the formal probability notation, so much later.

(Refer Slide Time: 27:24)







Let me do an example and then I will close it, example of 1.3.12 is the following. So, let us say, this is very relevant in the current date terms, what is happening in the world. So, Shyam let us say, Shyam is randomly chosen, chosen from let us say, from citizens of Hyderabad let us say and let us say he is given a test for swine flu, so to test for swine flu. So, we have three facts, let us say it is found 95 percent of people with swine flu test positive. So, he give a test for the disease, he give a test, the test has this rate of 99 percent, but it directs correctly.

We also found, that 2 percent of people without the disease, will also test positive. This is a false or positive (())(29:33), then we also know, 1 percent of the population, let us say has the disease. This is what we have. So, then I do this, I tell you that Shyam's test positive. So, let us say, Shyam test positive. Suppose, Shyam test positive, what is the probability that Shyam has the disease? Is that clear? So, I have this example, where I have to do this test to people. The test is 99 percent in sense true, if the 99 percent people, it will test positive for the disease with the who have the disease test positive, 2 percent without retest positive in the population is 1 percent prevalence, what is the probability that Shyam has disease?

So, yes, it is a very, very applicable problem. So, what you do, is define two ways. You define an event A, that the event A is that Shyam has Swine flu, that is the event A. B is the event that Shyam tested positive. So, what all do we know from the top, let us go to the top, let us see what all do we know in terms of A and B. So, we know that 98 of people with Swine flu test possible disease, it means from let us call this observation an assumption, as 1 assumption 1, assumption 2, and assumption 3.

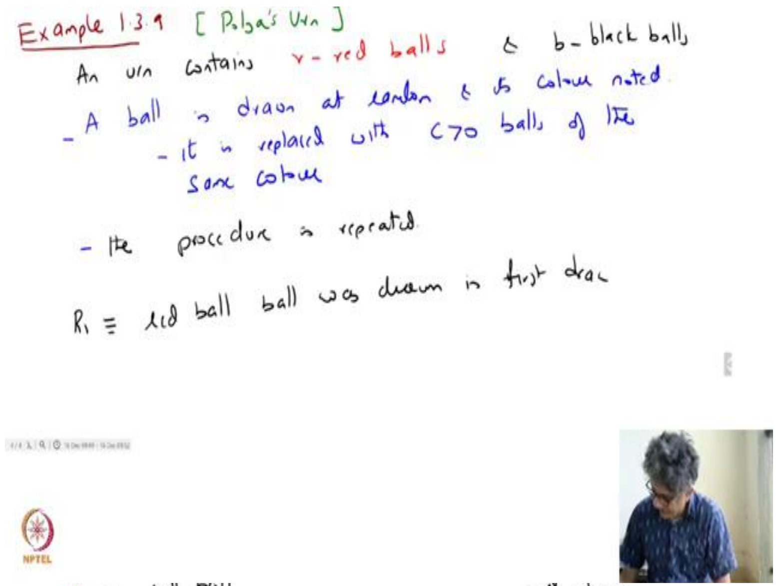
So, from 1, I know, what do I know? I know the chance that Shyam tested positive given that, he has a disease that is given A is going to be 0.95, that is what one tells me. What does 2 tell me? 2 tell me, that 2 percent of people without disease will have positive, it means the chance that B, If Shyam test positive, given that he does not have the disease, that is going to be point zero, zero 2.

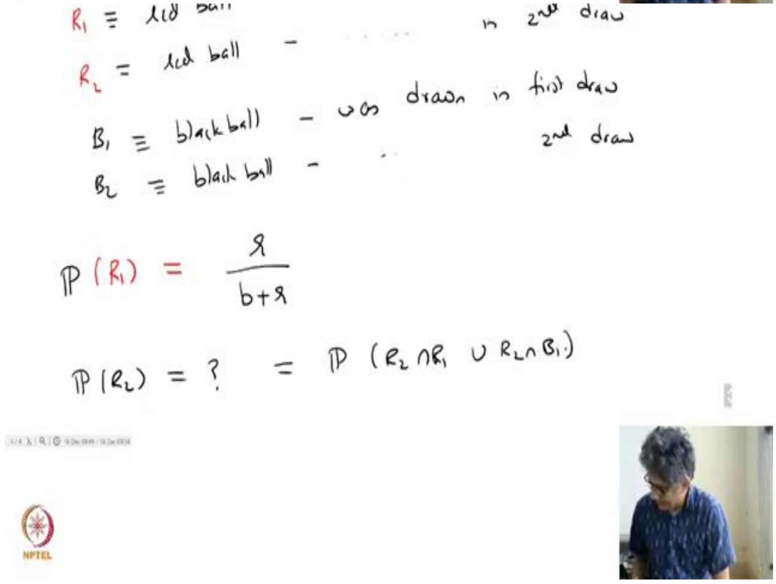
And what does 3 tell me, it does mean the chance that Shyam has a disease is point 1 point 0, and what do I want to compute, I want to understand the fact that Shyam tested positive. What chances it is, that means what is the chance that Shyam has a disease, given that he has tested positive. So, he is applying Bayes theorem, for the two cases. So, you apply Bayes theorem and finish the problem.

So, you know, by Bayes theorem, probability of A given B, is the same as probability of B given A times probability of A divided by probability of B given A times probability of A plus probability of B given A, complement times probability A complement. And that if you work out the mathematics, so, you put the mathematics down it is what you only have, you have B given A is 0.95 and probability of A is point 01, whole thing divided by B given A is point 95 times point 01 plus probability of A compliment is point 02, as A complement is 0.99 and that if you do is comes out to be point 832.

So, that means chance that Shyam has the disease given it is positive is only 0.324, that is not kind of a bad test. So, a couple of things to note. So, this is an exercise of, how I applied Bayes theorem, so, to check assumptions, I did not check that, all the events are positive probability, so, the whole thing works out.

(Refer Slide Time: 34:57)





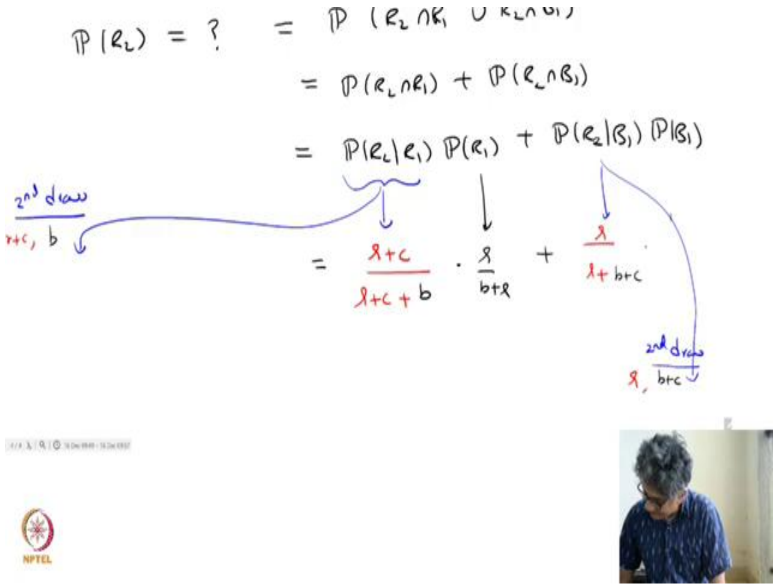
It is by a Polish Mathematician called George Polya. So, it is widely applicable, even though I will not explain the applications in this class. So, these up here, this is a very simple model, it is called the Polya’s Urns, and it is very widely applicable, it is even though the example is very simple. So, suppose, an urn contains R red balls and let us say B black bolls, B blue balls, so, let me B black balls. So, R red balls and B black balls, I forget, R red balls and B black balls.

Now, what you do is, the experiment is the following, A ball is drawn at random, so, it is crucial, A ball is drawn at random and it is color noted and then the crucial part is after this, it is replaced. So, when you draw the ball, but it is replaced with C positive balls of the same color. So, then this is step 1, then you repeat this step, each time you pick a ball you notice color down and then you just replace the ball with C of the same time.

Now, let us say, R1 is the chance that a red ball was drawn in first draw. So, again and let us say, R2 was the chance that a red ball was drawn in the second draw, a lot of the times if it works, then, let us see, then now, let us say B1 and B2 are respectively the same thing. So, that is the black ball was drawn in the first draw, similarly, here a black ball was drawn in the second draw. So, that was the case.

Now, let us do some simple calculations. So, now, the probability of drawing a red ball in the first draw, that is easy, so you have B plus C, B plus R number of balls and you want to choose red, where R is red, R by B that is easy. Now, let us try and do probability of R2. So, now, again the same way. So, the same as the probability of R2 and R1, let us draw first one or R2 and you pick B1, that is what happens in the first draw. Rather, pick a red ball in the first draw or a blue ball, or a black ball on the first draw.

(Refer Slide Time: 40:17)



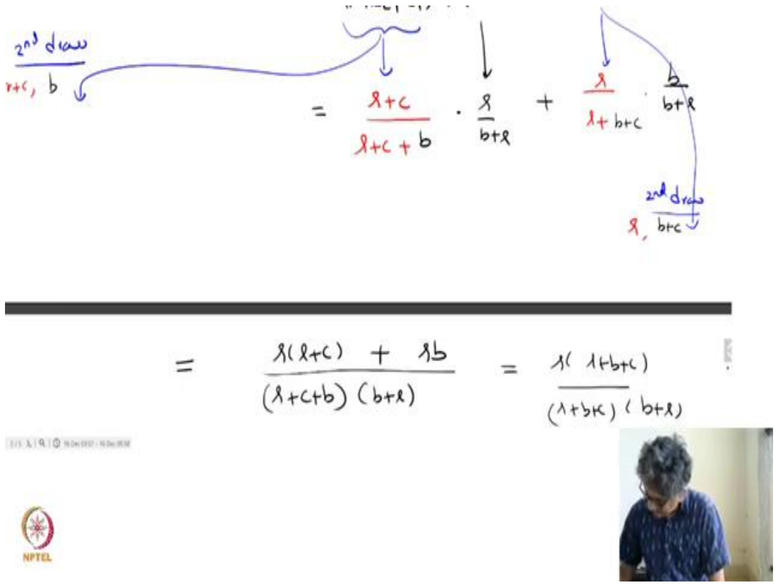
So, get the same as the probability of R2 and R1 plus probability of R2 and B1 and the whole thing we write down, same as R2 given R1 times probability of R1 plus R2 given B1 times probability of B1. Let us go back up here, we know probability of B1 is going to be equal to B by B plus R, that is easy. Now, what is happened here, in these two steps, let us see, let us do two steps, how do you do probability of R2 given R1.

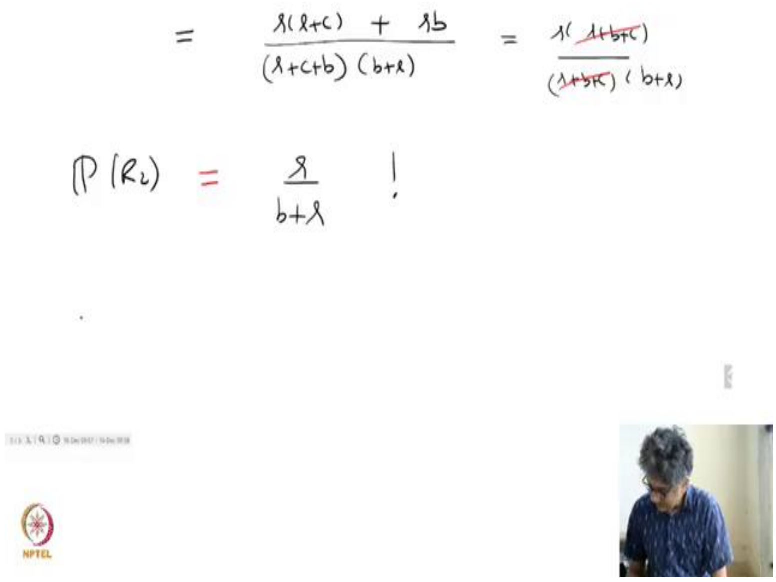
So, in this case, what would you do, you have chosen R1, you have chosen a red ball in the first draw. So, for the second draw in the box, second draw on the box, for second draw, for the second draw in the box there is going to be you have chosen R1 in the first draw, so, it is going to be R plus C red balls and B black balls, that is what is happening in this setup. This probability is going to be equal to, here if I do this calculation. So, I am going to get R plus C, choices I have and I divide by R plus C plus B that is what I get here.

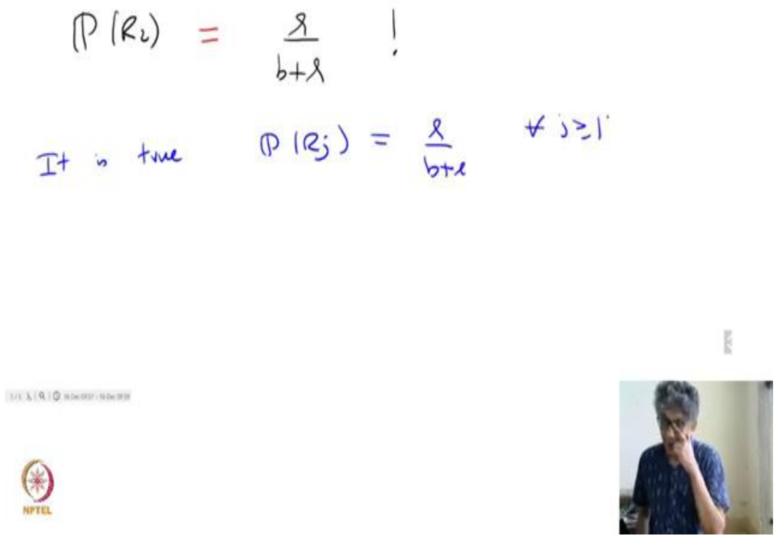
This person right here, we already know from the previous step, that is just probability of there is B by B plus R. Now, if I come here with this person right here, let me come down this way right here, if I come down, it is not a good idea, let me go this side up, if I come all the way down here for this person in the second draw, what will I have in the second draw. I will have, I have chosen B1 the B1 which means, I have chosen a black ball in the first draw.

That means, I have R red balls and I have B plus C black balls, that is what I have in this case. So, this person right here, would just come down to B, I have the only R at the top and I have R plus I have B plus C, but B plus C is numbers really matter. And this is just going to be, so, it is not easy, this is R by B plus R red balls, and this B by B plus R. So, now, I do the (())(43:07) now. I just write the mathematics down.

(Refer Slide Time: 43:11)







So, that is same as, so I have R, I have the denominator the same. So, I have this divided by I have R plus 3 plus B into B plus R, that is on both sides. So, that is easy. The top I have R into R plus C in the first term plus I have R into B in the second term, this is what I get. So. Now, if I remove the R out, it says R into R plus B plus C divided by R plus B plus C and I have another R, another B plus R.

So, that, these two cancel off, and I get R by B plus R. Again, the probability of R2 is also R by B plus R, this is kind of a very very interesting model, that is if you do this experiment where you are, when you keep adding balls the same color, then it so happens that, the probability of choosing a particular color does not change, it is a very interesting sort of aspect of this model it is called Pollera Scheme.

And one can actually check, that this is true for any drop, in fact, one can actually check this. So, let me just write that down so the remarkable feature of this model, is that we are going to actually check that, that one can actually check and one can actually check, it is true that probability of RJ. In fact, is R over B plus R for all J less than equal to 1, leave it on the first one, which I verify.