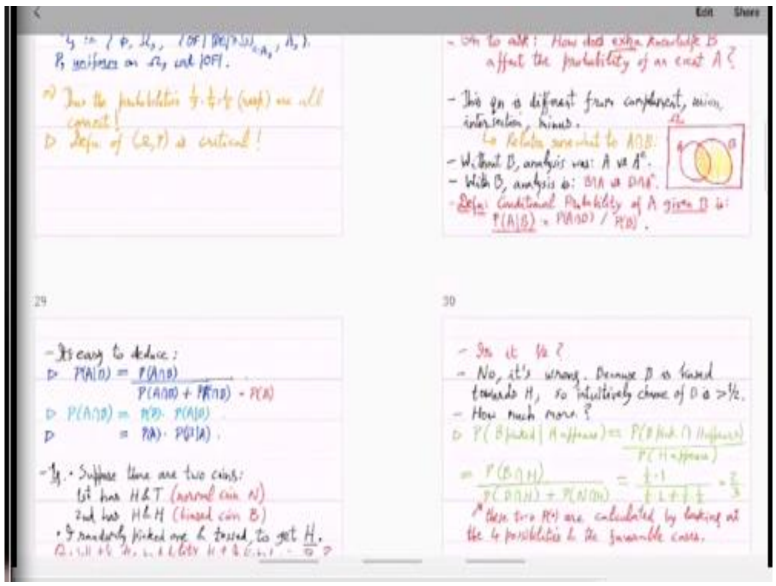
Probability for Computer Science
Prof. Nitin Saxena
Department of Computer Science & Engineering
Indian Institute of Technology-Kanpur

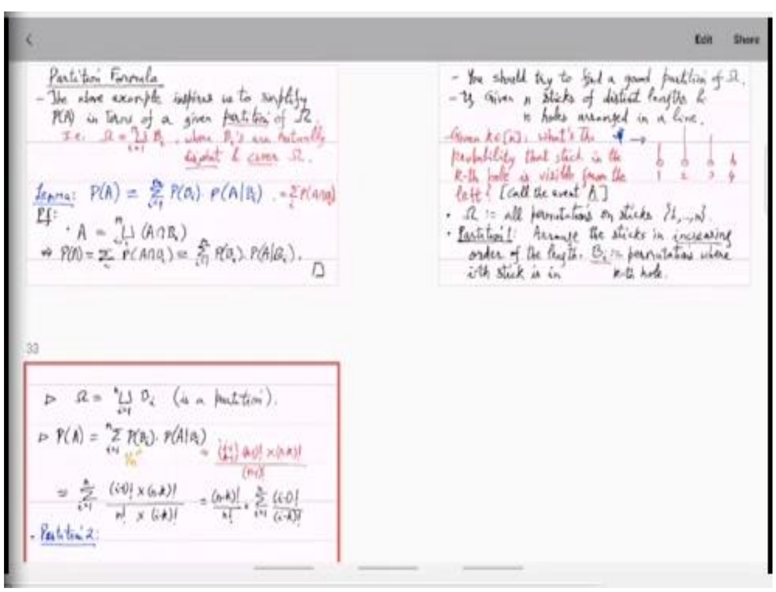
Lecture-07
Independent Events, Bayes theorem

(Refer Slide Time: 00:15)



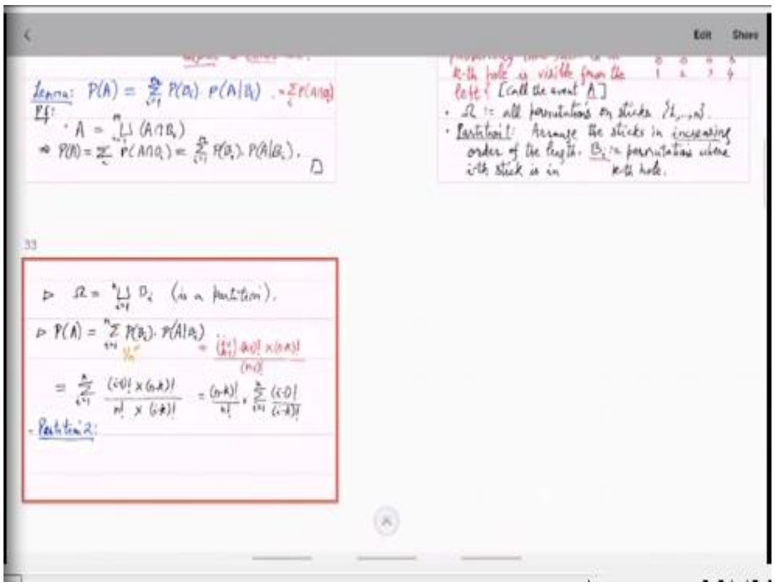
So, we were doing conditional probability. And so, after given this definition that probability of A given B is intersection over probability of B.

(Refer Slide Time: 00:30)



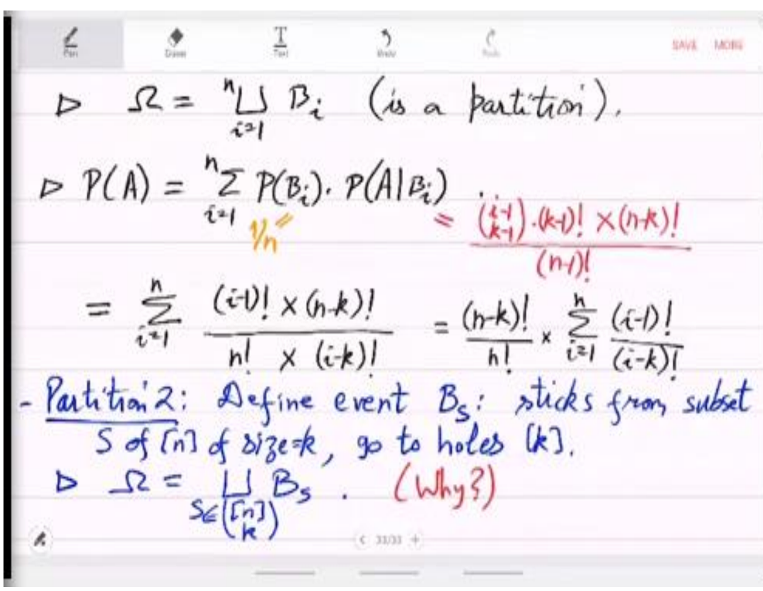
After giving this definition, we started looking at this partition formula, which is helpful in case you have a good partition of the sample space given by the B i's. And so we were working out this example where you are given n sticks of distinct lengths to be put randomly in n holes, one stick in one hole. So, it is equivalent to choosing a random permutation on n elements. Such that from the left the kth hole with the stick is visible, the stick in the kth hole is visible. What is the probability of this event? We are calling this event A.

(Refer Slide Time: 01:16)



So, we worked this out with one partition, we partition the sample space which is essentially the omega set of all permutations.

(Refer Slide Time: 01:27)

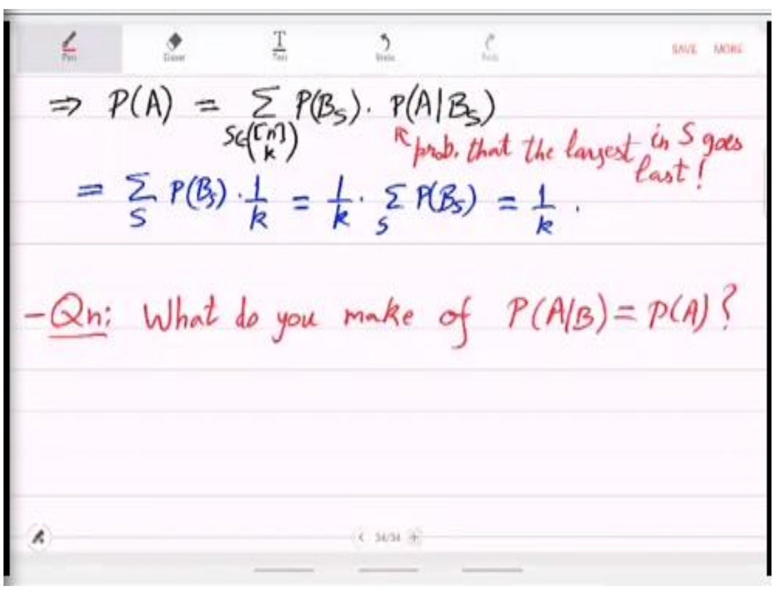


And then we got a complicated formula, now we will try something else to get an easier expression. So, here what we will do is? We will define B sub S as the event where sticks from subset S of n of size k, they go to holes 1 to k. So, the first k holes they will be mapped to sticks from a k subset of 1 to n. These are the sticks but then how they will go into first k holes? That could be a permutation of these sticks from subset s.

And then for every subset you have this event B sub s. So, this is a partition of all permutations, why is that? So, this is because any permutation in omega will be mapping exactly k sticks to the first k holes. So, those k sticks define the s, for that permutation, so every permutation is covered. Now notice that a permutation from B s and a permutation from B s prime some other subset.

Those permutations will be different, they cannot be the same because they are mapping different sticks at least one stick is different in the first k holes. So, this is really a permutation, so you can work this out but it is really straightforward. So, let us now see how this partition is better than what you saw in the last class.

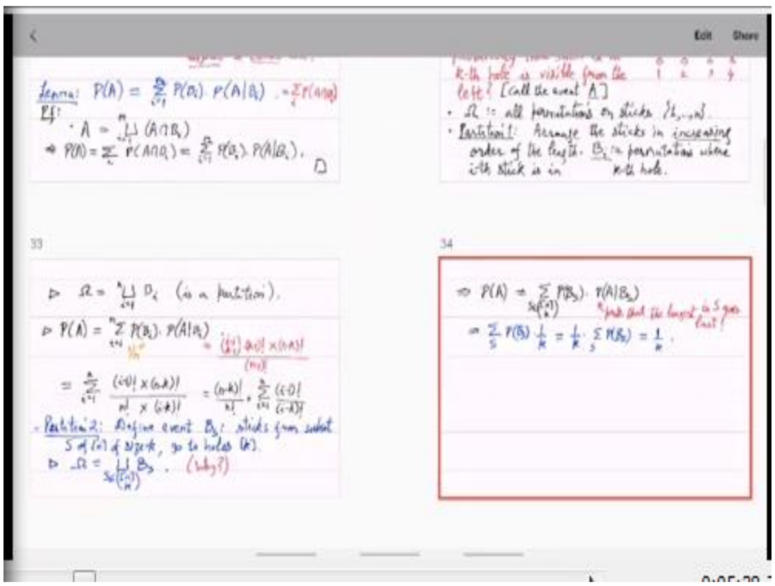
(Refer Slide Time: 03:53)



So, probability of A which is the kth stick been visible in the kth hole. So, this is probability of B s times probability A given B s, S all the subsets these are n choose k many subsets. And remember that this is the probability that the largest in S goes last. So, keep in that in mind, you can simplify the probability, so this is sum over the subsets probability of B s. Now, what is the chance that largest of S goes in the last hole, the kth hole?

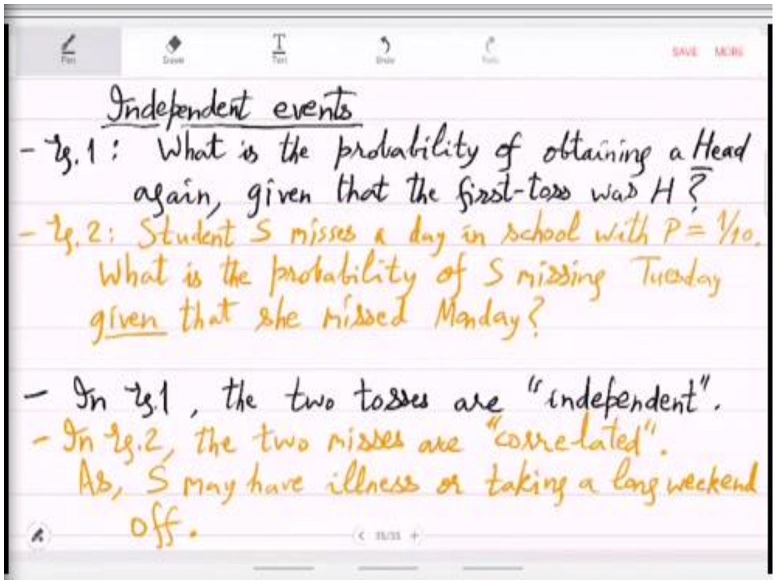
So, there are k (()) (04:58) of size k, so largest is unique because the length for distinct. So, that unique stick going in the kth hole is 1 by k probability, so which is 1 by k times probability of B s over these S. But that you know is sum over the partition, so this is 1, so you get 1 by k straightforwardly. Hence the chance that the kth hole stick is visible, this is simply 1 by k, you can see the difference.

(Refer Slide Time: 05:35)



So, in the previous calculation, you got a complicated formula of factorials and sums, here you directly get 1 by k, so you have to be creative with partitioning**.** So, let us now move to the next property that disconditional probability gives rise to. So, the question to ask now is what do you make of when the conditional probability does not change? Given the conditional probability, so A conditioned on B is does not change it remains probability of A. So, B has kind of no effect? So, this somehow means that A and B are independent.

(Refer Slide Time: 06:39)



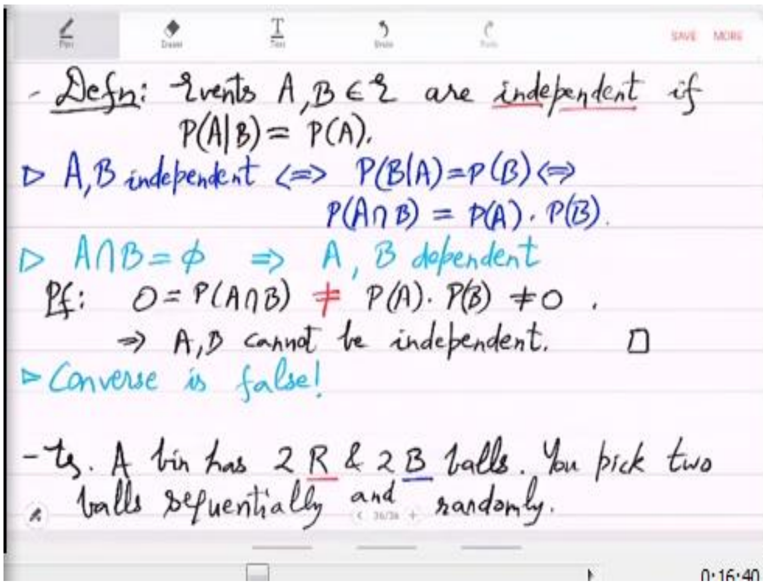
So, that is our next topic, well not really topic but next property to study, so independent events. So, the examples are, what is the probability of obtaining a head? Again, given that the first toss was head. So, given that the first toss was head conditioned on that event, what is the chance that second will also be a head? That is one let us say example of A and B, second example is say a student misses student S misses day in school with probability equal to one tenth.

So, what is the probability of S missing Tuesday given that she missed Monday? So, there is an extra information, just the probability of missing Tuesday would have been 10%. But on top of that you are given this information extra information that she also missed school on Monday. So, then does the probability remain 10%, is it less than 10% or more than 10%? So, these examples are actually very different, why?

Because in example 1, it does not matter what happened in the first toss, the second toss probability does not change. Because as you can imagine it is an independent event, so in example 1 the tosses are independent. While in example 2, the events are actually dependent because once student miss Monday, the chance of missing Tuesday should be slightly more than 10%. Because the reasons you have to go into the reason maybe the student was taking a long weekend off, student was travelling or the student is sick.

Because of those reasons actually missing Tuesday the probability will increase, with the knowledge that she missed on Monday, so these two events are actually dependent. The two misses are correlated or dependent. As S may have illness or taking a long weekend off. So, those reasons actually creep in once you know that, the student has missed Saturday, Sunday and Monday. So, how do you formalise this mathematically? So, actually it is very easy.

(Refer Slide Time: 11:36)



The definition is simply by the probability distribution function. So, events A, B in this sigma algebra are independent. If the probability of A given B does not change, if extra information does not change the probability, the probability distribution function value. Then you call them independent otherwise you call them dependent. So, immediately you deduce from the definition that A, B independent means that the opposite is also true.

So, for B given A will not change, and why is that? Well, because you actually using the independent event definition, you can write this formula for intersection. So, you would write down this formula, and then from this you learn that A given B is independent. A is independent of B and B is independent of A, this is a symmetric property because of the symmetricity of intersection that is all.

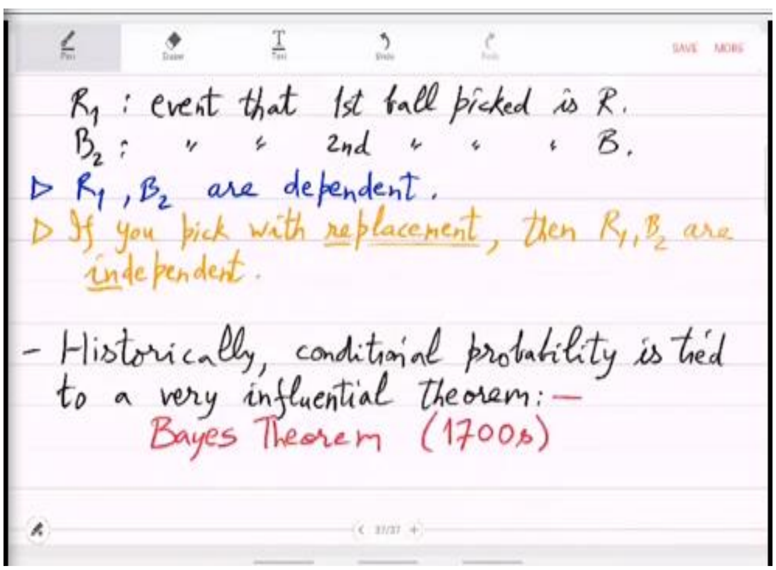
There is another nice property to remember if A and B are disjoint, then what can you say about their independence or dependence? So, you can show that A and B are dependent, why is that?

So, this is because again if you look at the intersection formula, you get that obviously probability of the intersection is 0 because it is empty, while on the other hand you get P A times P B, which is not zero, so these two are not zero.

So, this is the deduction, so probability product is non zero. This you can assume, you can assume that both A and B were events that happen with positive chance. So, only in that case we are talking about. So, in that case RHS is non zero while LHS that is probability of intersection, this is 0. There is no case actually where both A and B happened, because there is no intersection, sample space, elements does not exist which means that they cannot be independent and this we save dependent.

So, disjoint events means that they are dependent, it seems to be a fallacy, converse however**.** Which should also be clear because if A and B are dependent then it does not really mean that they are disjoint, dependent may happen because of many, many other reasons, not because just because of as you saw in the previous example. Another example, so let us say a bin has 2 red and 2 blue balls, there are 4 balls in a bin, you pick 2 balls sequentially and randomly.

(Refer Slide Time: 16:42)



Now R 1 is the event, that first is red and B 2 is the event that second ball is blue, these are the two events that we are interested in. Now, the question is, are they dependent or independent? So, since in the given R 1 since you have picked first ball as red, so now the red balls are fewer in the bin, so the chance of drawing out a blue is higher. So, hence extra information helps, so B 2 and R 1 they are dependent.

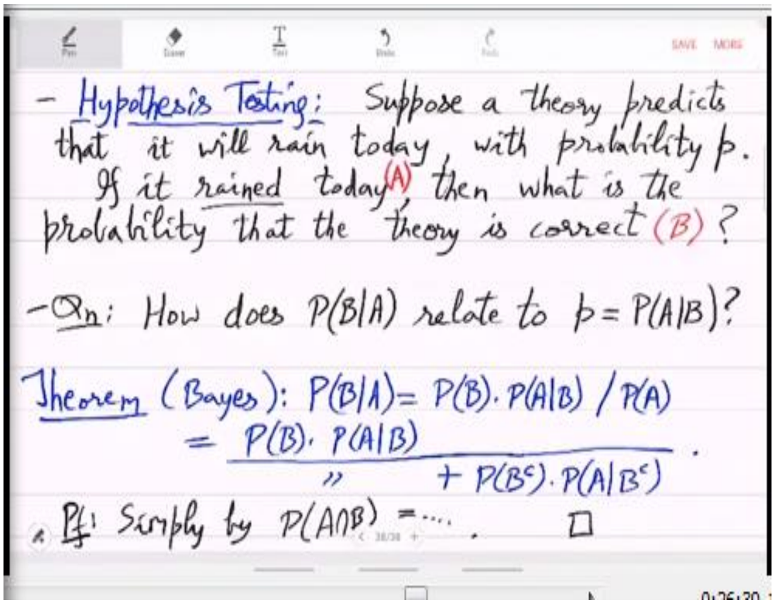
However, if you change the rules of the game, so which is to say that you allow replacement. See the ball that you have picked the first ball you put it back in the bin, then nothing has changed, the bin is the same. So, second ball being blue it will not matter what happened in the first drop, so they will be independent. So, if you pick with replacement then they are independent. So, this is a example where both things can happen. So, you have to specify how the experiment is done. Is the experiment done with replacement or without replacement and then the dependents correlation changes?

So, let us continue with conditional probability in a bigger way, which will be more interesting examples and more interesting interpretation. So, historically conditional probability is tied to a very influential theorem which is called Bayes theorem. So Bayes, he was a, I guess a priest in 1700s and this theorem is named after him. It is basically the definition of conditional probability but it is very rich in interpretation. It is applied in many, many places, so it is actually very, very influential name and you will see why?

(Refer Slide Time: 20:40)

So, the starting example is let us start off with, suppose you have multiple theories and multiple observations, obviously you being a human you do not know which theory is correct. And hence you do not know that when an observation is made which theory correctly applies. So, this is called hypothesis testing. And let us see a concrete example of that.

(Refer Slide Time: 21:18)



So, suppose a theory predicts that it will rain today; this is a very practical example. Because monsoons for example they are modelled using different modelling techniques. So, when I say theory, I mean some model. So, suppose you have a model that predicts that it will rain today. And obviously that prediction will come with some probability, the model has some chance of success. So, let us say with probability P, it will rain today.

Now you observed and it actually rained today. So, if it rained today, then what is the probability that the theory is correct? So, there might be many theories, there might be many models, you do not know for sure that your model is the best and is the correct model. You have just done a statistical analysis and saw that whenever you make a prediction using this theory, then it is correct with probability P.

Now you actually see that the prediction for today happened, from that information, what can you deduce about the theory? How good is the theory? And you have want to assign a probability or confidence to that theory or to that model. So, how do you do that? So, let us give these events names, so let us say raining today is event A and theory being correct is event B, so you want to deduce probability of B given A.

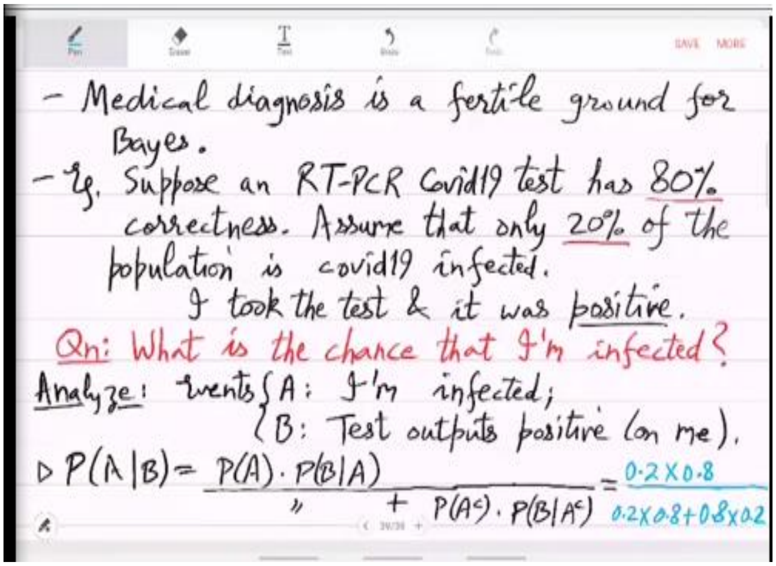
So, you are asking the question, how does probability of B given A relate to the only other thing that you know which is little p? What is that? That is the probability that given the theory is correct. So, given B, what is the chance of there being a rain today? So, that is what you want to relate to probability of B given A. So, do you want to read these 2 opposites, B given A with a given B?

Now actually you have already seen the definition of conditional probability, so you know how they relate and that is the content of Bayes theorem that is all. So, it is very simple to state, there is no problem there. So, you write that probability of B given A is probability of B times probability of A given B divided by probability of A, which is also equal to probability of B times probability of A given B.

And you have actually also seen partition formula, so you can partition A with respect to B. So, you will get this plus the complement of B that is all. The proof is simply by intersection. Just look at the intersection in terms of conditional probability and you will get the first formula and then second formula is just expressing probability of A by partitioning the sample space with respect to B and B complement.

So, this is quite simple, this is a fancy name; it is called Bayes theorem or Bayes theory. So, what is the big deal about this theorem that is in the interpretation?

(Refer Slide Time: 26:40)



So, let us now see 2 stunning examples. So, the first is in medical diagnosis, medical diagnosis is a fertile ground for Bayes theorem. So, let us go back to this example of COVID tests that we started in the first lecture and it might have confused many of you. So, let us now actually work out the formulas and the numbers. So, suppose an RT-PCR test has 80% correctness.

And assume that when this test was being done, say in the month in last year in 2020, say you were doing this test in June or July, when very few people had COVID virus in them. So, at that time, let us assume that only 20% of the population is actually COVID19 infected. So, that is the data, that 80% correctness of COVID test, which means that somebody having the virus the test will be positive with 0.8 probability, somebody not having the virus, the test will be negative with 0.8 probability.

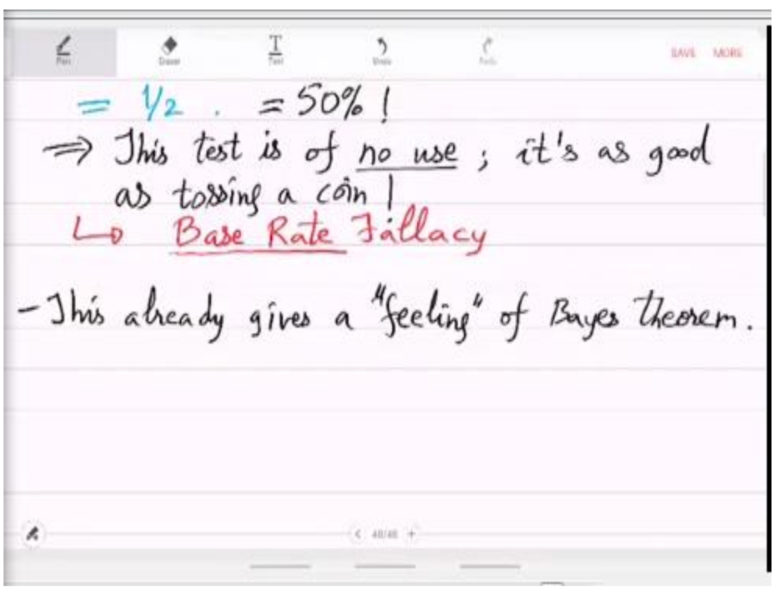
And if you look at the population say of India then only 0.2 are expected to have the virus, the virus has not spread too much. So, now in such a scenario you get the test done and it comes out to be positive. So, I took the test and it was positive, so the question is what is the chance that I am infected with COVID19? So, if you have no background in probability, then you will think 0.8, because you took the test.

And the test has correctness 80%, so you would think that 88% chance you are infected. If you are applying more common sense, then you would think that since the population only 20% are infected, so your probability is also 20%. But both these answers are wrong, they have really no basis. So, what we will use is, we will do this properly by using Bayes theorem, so let us do that analysis.

So, so event A is that I am infected, event B is that test outputs positive. Now, obviously in this scenario you are given this event B, with this extra information you want to find the probability of A. So, let us apply the formula. So, probability of A given B is by Bayes formula, this thing divided by itself plus the compliment, probability of A complement times probability of B given A compliment, that is the formula.

Now let us put the numbers, let me make space here. So, what are the numbers? So, probability of A which is I am infected, that is without any information it is just 0.2. What is probability of test being positive given that I am infected? That 0.8, now the same thing plus, here what is probability of I not being infected? That is 0.8. And what is the probability of test being positive despite I being not infected? That is 0.2. So, let me see whether I can manage it.

(Refer Slide Time: 33:03)



So, this is ultimately half, so it is 50% chance. So, which means that actually this testing that I went through this test is of no use, it is as good as a coin toss. Because this is 50%, so after going through this long process, whether I am infected or not, that probability is just 50%. So, I could have as well just flip the coin, why go through this whole testing? So, it is not 20%, it is not 80%, it is right in the middle.

So, this apparent fallacy is called base rate fallacy. So, this is happening because the base rate is very small, it is 20% of the population is infected. So, it is so small that even 80% correctness is not enough. The correctness should be much higher of the test to give you a decent edge over 50%. So, this should already tell you the strength of, this already give a feeling of Bayes theorem. And we will see even more interesting example in the next slide.