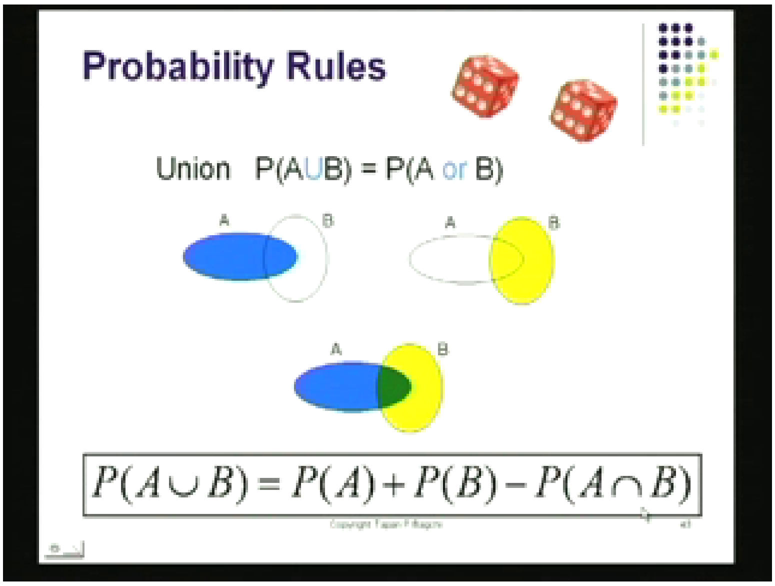
Six Sigma
Prof. Dr. T. P. Bagchi
Department of Management
Indian Institute of Technology, Kharagpur

Module No. # 01
Lecture No. # 07
Review of Probability and Statistics –III

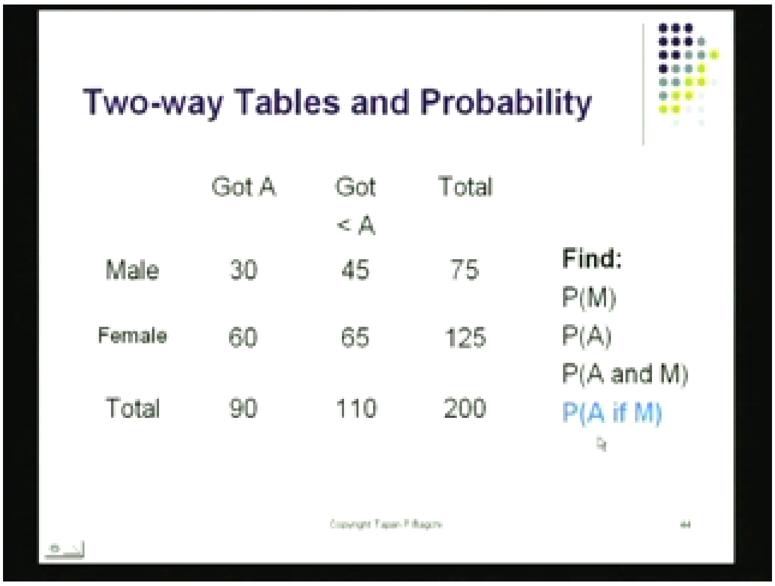
Good afternoon, welcome back again, I we continue to our without discussion of probability theory and this time, I am going to begin a topic called Conditional Probability in a couple of minutes, we will be looking at examples, which are directly taken from probability theory and these are all brand on what we call conditional probability, a very very important concept in the theory of probability.

(Refer Slide Time: 00:43)



If you look at my screen here, I have the union defined as I have defined here, the union is basically probability of A union B turns out to be any object or any outcome that falls either in A or in B is also in A a and in this case union B and therefore, if you look at the colored dimensions here, you look at the blue and the yellow; everything that is either in blue or in yellow or of course in both. They all turn out to be having the property that belong to A union B, this is like one and you already know about the intersection, which is the green area right in the middle that is the intersection part. And as I told you the formula here is the probability of A union B which is like finding an outcome either in A or in B that turns out to be P A plus P B minus P intersection A B, this we have seen before.

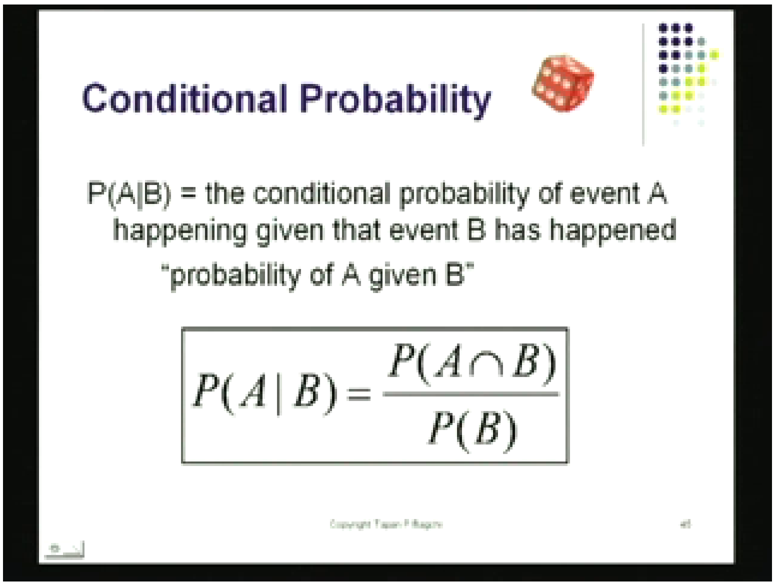
(Refer Slide Time: 01:39)



Let us move into an example now, that the example goes as follows, again I have got kids writing tests, when I had given them a test and I have counted how many got A and how many got less than A. And also I have kept track of how many were male and how many were females, how many girls were there, how many boys were there. The question that are being raised again is, what we had before which is like, what is the probability that are randomly pick student is going to have is going to be male or a randomly picked student is going to be scoring A, he would have scored A, he or she would have scored A and what is the probability that the person I picked is male, and he scored A, what is that probability this and immediately it turns as, this is the matter of intersection.

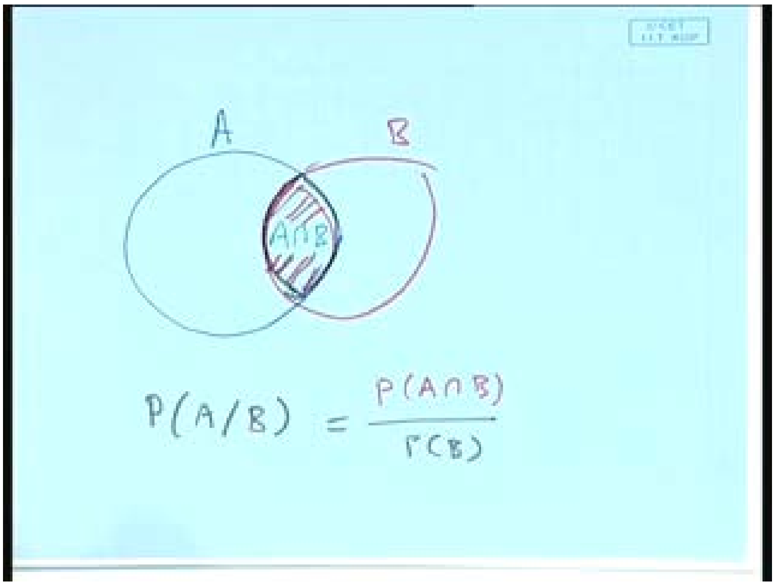
Now, so far there is not such of a problem I can probably find that out, but in addition the probability, I am giving another question which says, what is the probability that the person has scored A, if he is a boy, if he is male, that means given M, given that the person is male, what is the probability, that he would have scored A, this could require me to defines our additional theory.

(Refer Slide Time: 02:58)



And let see, how you move into that? We define what we call the conditional probability what we say is probability A given B, probability of event A is happening given that, B has happened before it; this the probability of A occurring given that B as occurred. And this is given by this formula here, probability A given B and it is all is mark with the slash that will be there, you see the slash there, it is equal to probability A intersection B divided by P B, let me give you a pictorial idea, let me give you a little idea of how this is done, how did I did I end up with this thing.

(Refer Slide Time: 03:43)



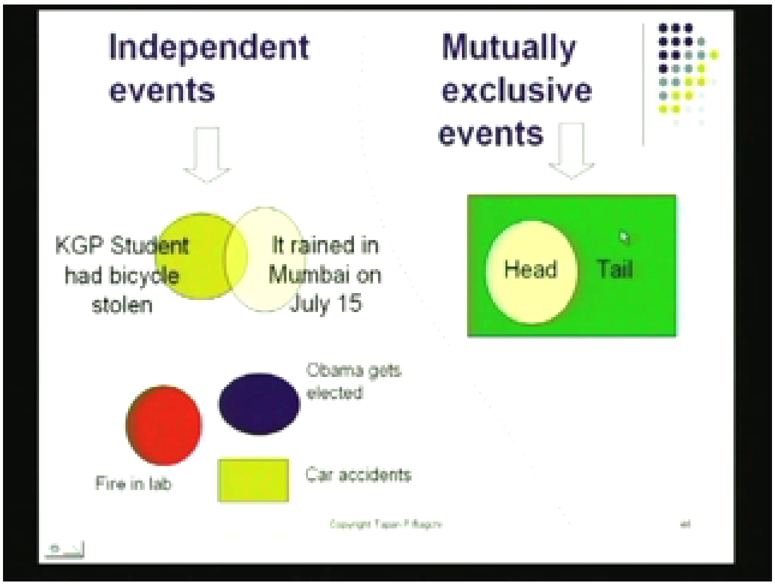
I have what you see, I have the set A or the event A, and I have what we call, the set B which is the other event, I am going to call it give it a different colour, this is set B; A intersection B is this part, that is going to be A intersection B is going to be this part A intersection B.

Now, if I have to figure out the probability of what I said conditional probability on that I am going to write as probability A given B, let us look at the total possibilities given B means, all of B would have occurred, so I divide that in the bottom, so I just call it P B in the bottom, that is the denominator. In the numerator I only have to carry that part which is common between A and B that is this part that is this part this is the part, this is the part that would have occurred if B had occurred, now a is occurring and also b is occurring this the intersection part.

So, what have I got here, what have I got to write here, I have to write here probability of A given B is going to be probability A intersection B, which is this part, which is this part really, this is the part that is A intersection B divide by this whole probability place. Now, A ratio of this little thing here, divided by this whole thing there, that is going to be my conditional probability; I am not interested in this part, because there B is not occurring, I am interested in only those parts where B is occurring, B is occurring all over this place and this is the fraction of times when A and B occur together.

So, the probability of A occurring, when B has occurred B has occurred all over the place and the probability of A has occurred in this area, so it is going to be this ratio that actually gives me the conditional probability of probability of A given B, does this; this is a very very important formula, this will be using many many times when you get into probability calculations, you will be doing this many many time.

(Refer Slide Time: 06:01)



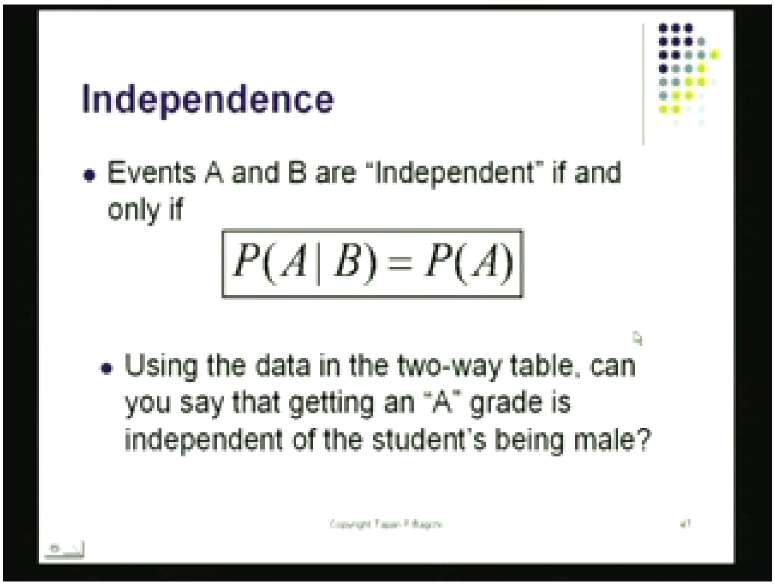
Let us see how we use them, what we have to remember of course, it is again a little reminder of what we done before, we have what we call independent events, an independent events are once that occur independent of regardless support by they are occurred somewhere else, so for example, a Kharagpur I I T student have his bicycle stolen and by chance on July 15th it rained in Mumbai, these eventually we have nothing to do with each other, they are independent of each other, that is like an example of events being independent.

And look at this side I have got you know fire in the lab, in the chemistry lab there was a fire and you know Obama got elected of a couple hundred days back Obama got elected in the US, probably these have nothing to do each other they are probably independent of each other, and there was a car accident at the (()) gate of I I T, Kharagpur there was a car accident. Again this has nothing to do with Obama, not did it have anything to with the lab action, and these are examples of independent events.

So, I see here in independent events thrown together that is all, on this other side I have got mutually exclusive events what are those, I am tossing a coin and I get with a certain probability I get heads, certain number of times I get heads, and at other times I get tails and only because these two, these are the only two outcomes can possible head and tail these some of the probability of there being a head and there being a tail, it is going to be 1, probability of head plus probability of tail is going to be 1.

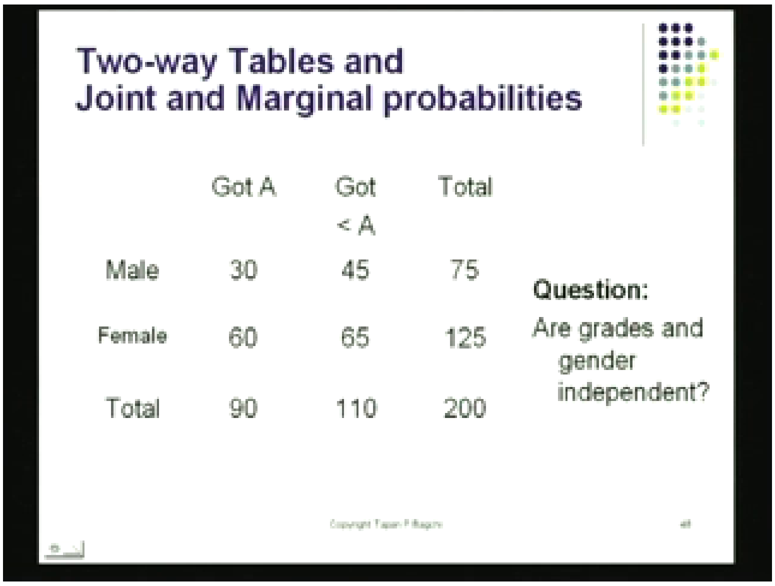
And this a really, if head occurs tail will not occur, if tail occurs head will not occur therefore, these are mutually exclusive events, I hope you are cleared now, about independent events which occur individually regardless support by they are happening somewhere else, and mutually exclusive events are such, if one of them occurs, if one of them occurs the other will not the once that are mutually exclusive, they pre include the occurrence of the other one, we have to remember this when we are combining events, we are combining probabilities.

(Refer Slide Time: 08:19)



So, events A and B are independent, if and only if, I do this conditional probability and I find it is unchanged from the probability of A; so whether B had occurred and B had not occurred, it would not really matter for the probability of occurrence of A. If this is the situation I say A and B are independent and of course, this will also imply the probability B given A would be equal to P B, probability B given A but, b equal to P B. And now, we can go back to the two way table that we did, and would like to take a look at what exactly is the situation.

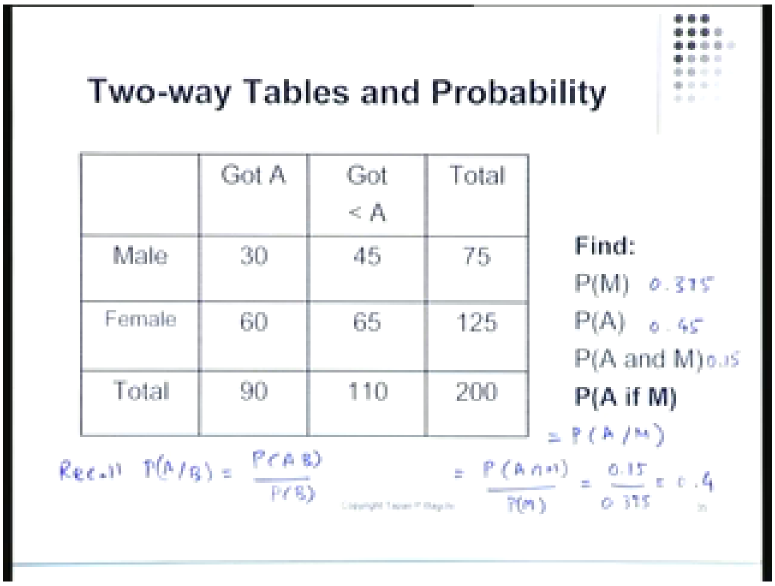
(Refer Slide Time: 08:58)



So, I bring up the two way table, and I have some numbers there and what I have done is, all the creates independent and before that I would like to do a calculation, I would like to go back to that thing there, and I have this problem set which is given to me, and the question that are being asked in this case, what is the probability that the student is male, what is the probability that he or she scored A? This we randomly pick student, what is the probability that the randomly pick student scored A and this is a boy and what is the probability that if he was a boy he scored A, this is the conditional probability, probability of A given B, B has occurred now want to see B is given to me and I want to see the probability of A being there.

Let us see, how we calculate these things I have the solution here on the on this little piece of paper here.

(Refer Slide Time: 10:00)



Let us start by looking at this two way table, we start by taking a look at this two way table, what the two way table does it allows me from counting it allows me to calculate the different probabilities in a very simple way, which is just by counting, so here I am using counting, I am doing really object evaluation, I am not going by opinion, I am going here by hard data.

What is the probability of the randomly picked student is a male, just take a look at that, how would will I get 0.375, the chance of he is being male is those who got A they were 30 and those who did not get A they were 45 this is the total is 75 I divide 75 by 200, and I use by machine and it turns out that the machine gives me a number 0.375 that is the probability of I have randomly picked student being male.

What is probability that the randomly picked student has scored A, A could have occurred two ways either he was male or she was female or she was other student was female therefore, 90 is the probability of any student of any sex scoring A; therefore, 90 divided by 200 that is the probability, that the randomly pick student is going to be scoring would have received A. Let us try to do that, so what I do is I take 90 and I divide that by 200 and I look at the number that turns out to be 0.45, and in this 0.45 turns out to be answer there, see the 0.45 written there.

Now, let us took a look at this third possibility probability of A and M, what is A and M, A is the scored A, the person scored A and the person was male; that is going to be now

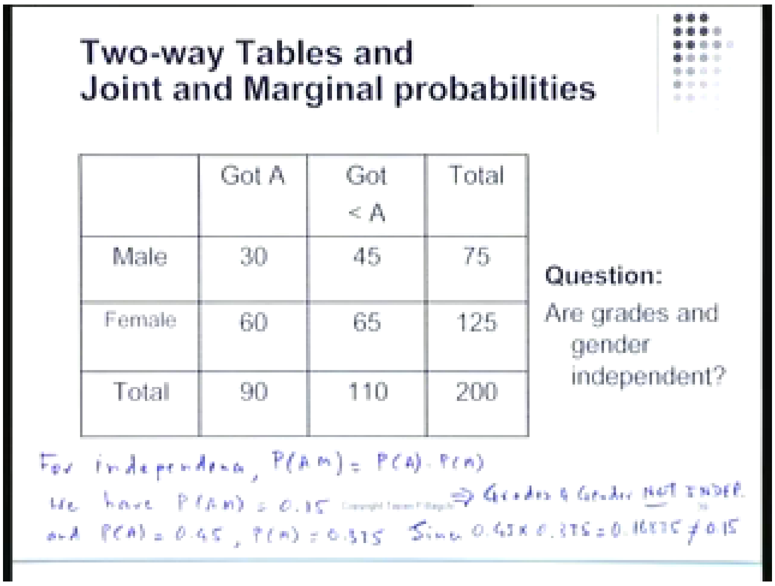
30 divided by 200, again by counting and 30 divided by 200 it turns out I do not need the machine now; it turns out to be 0.15, that is the probability that the randomly picked student was male and also had scored A.

Now, this condition probability of A given M, he scored A given that he was a boy, how do I do that, I bring up the probability calculations and I go go to my conditional probability formula and the conditional probability formula is given as this A intersection B divided B. And let us take a look at, just look at this, these are all joint probability, these are all going to be joint occurrence as of things therefore, if I have to find the joint occurrence A and M, A and M is this cell, A and M is this cell, this is the cell that is gives me the count of scoring A and also being male.

So, it turns out 30 divided by 200 that is the probability of A and M, A and M that is 0.15 and then the probability of the person was, because when I am using my my formula, when I am using my conditional probability formula, I have probability A given B that is equal to probability A B divided by probability B, which turns out to be probability A intersection M divided by P M. P M in this case, I have already calculated it is 0.375, so 0.15 divided by 0.375 this a very complicated question, let us try to work it out, 0.15 divided by 0.375 equal to 0.4, so that is the answer, that is the answer right there. So, I have solved this problem now, and I have also solved a question that involved conditional probability.

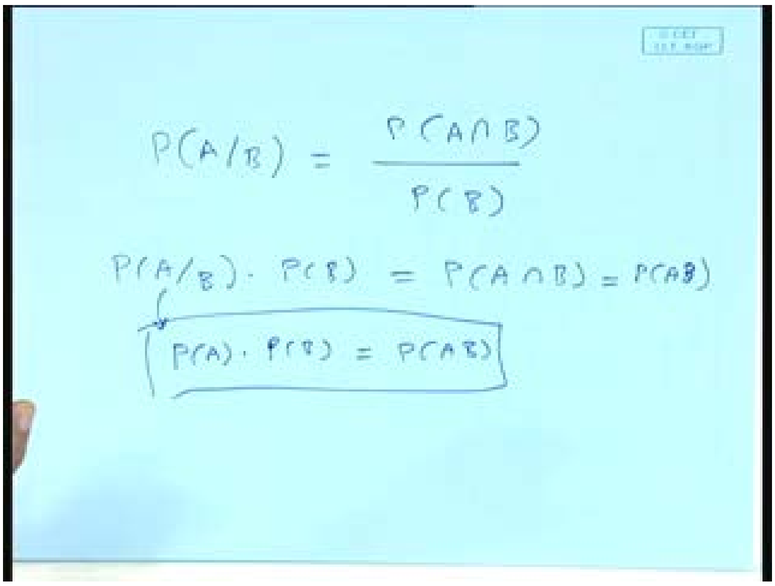
Let us go now, to the next question, what is the next question, let us try to get the question now, the question is this are genders and grades independent, what is the test for independence remember now recall test of independence is P A given B is the same as P A or P B given A is independent of A, which is it is equal to P B. So, lets we pick up all those things out, what I am going to do is, I am going to work out those probabilities and if that is so it also turns out there is a simple test for probability for the independence of probability; then and that I am going to show you, just by flipping around a little bit and I am going to bring that sheet there, the sheet is here now.

(Refer Slide Time: 14:40)



Are genders and grades independent, and it turns out then remember now the intersection, remember the formula I had let me remind you of that formula, let me remind you of that formula here, remember the conditional probability formula which is this, and this formula I could also rewrite, let me rewrite that for you on a sheet of paper, there nothing it will become a little more clear.

(Refer Slide Time: 15:18)



The conditional probability formula says P A given B is equal to P A intersection B divided by P B, this I can also write as P A intersection A given B multiplied by P B is equal to P A intersection B, which I can also write loosely as P A B, no problem so far. Now, is A and B are independent then this guy turns out to be P A and multiplied by P B this is equal to P A B, this is the result, this result I want to utilize.

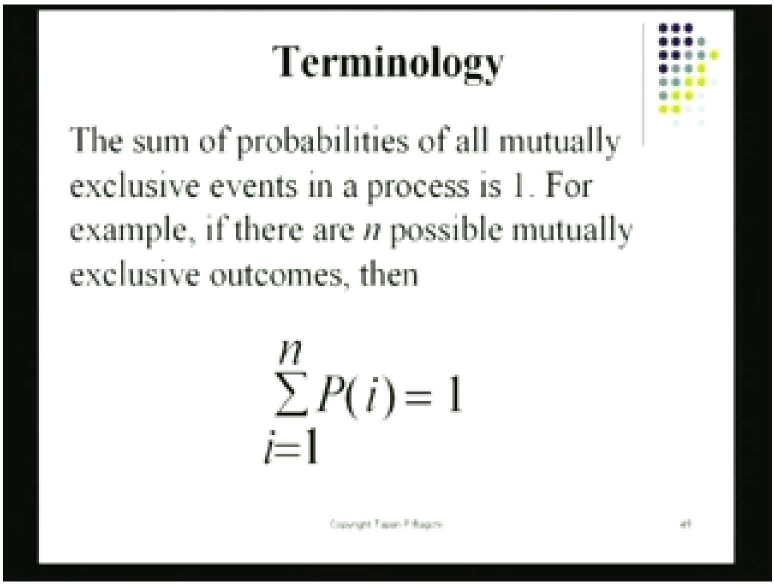
I look at P A b is that equal to P A times P B are this should be true for all outcomes by the way, you cannot just check this in a special case and say that I have got A and B declared to be independent, they are ready to be declared independent, you cannot do that. You have to check out all the outputs, all all the outcomes and from that you should make a general statement are very fine that I had in no situation this condition fall at it then of course, A and B are independent.

Let us see how you utilize that, we go back to the problem are grade are grades are grades independent of gender, and I have the solution here, what what do I have to check now, I only have to check this condition P A M, A is scoring A, M is male, so I have got sex here, I have got score here, is that equal to this this is what I would like to check. So, really this is being forced as a question, this being forced as a question I have already calculated P A and M in the previous question, I have already calculated P A and M and I have also calculated P A and I have also calculated P M.

So, I have got all the things I have got this guy calculated separately, this calculated separately, this calculated separately I just have to see check is the left hand side equal to the right hand side, that is what I have to check; and what I did was, I took the value of P A and I took the value of P M multiplied them. And let me just do it for you, so that you will believe that I have done the job, I have 0.45 multiplied by 0.375 equal to and I get 0.16875, but my god you know, we should have got we should have got 0.15, because that is P M.

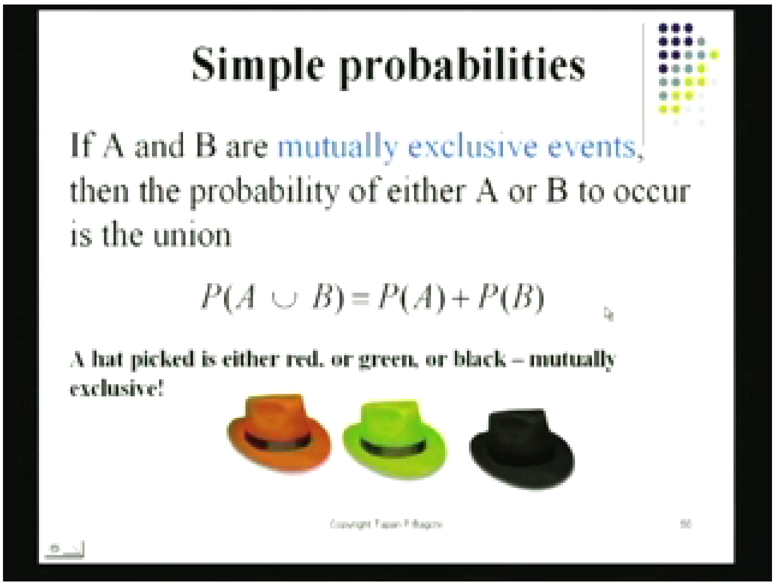
So, it turns out this number is not the same as this therefore, what is happening here, this condition is being violated what does it mean grades and gender are not independent. So, it depends for there a boy or a girl why going to be doing in the example, that is like the this little test here is suggesting that.

(Refer Slide Time: 18:46)



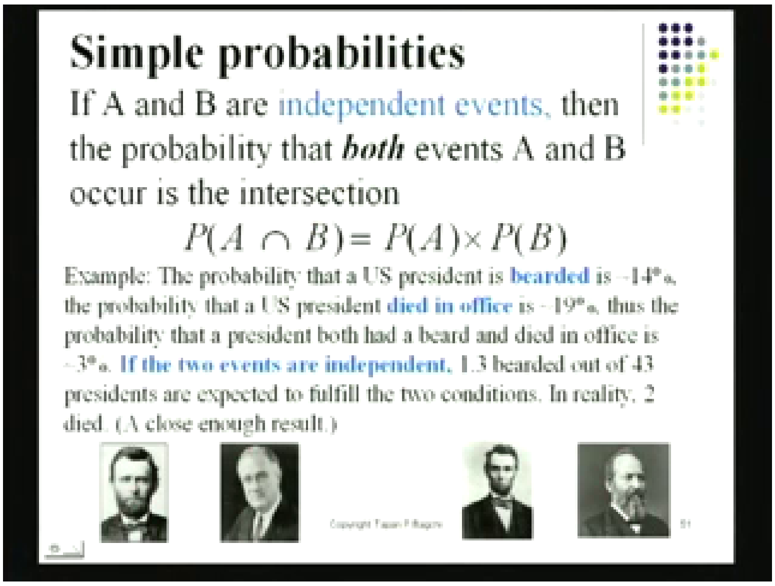
So, it is a question of probability what we have done this, you already know about this condition here, which is like some of all the probabilities they should be equal to 1.

(Refer Slide Time: 18:51)



If we talk about mutually exclusive events, and it is sort of like this you go and pick couple of hats, and there are a certain number of red hats, and the certain number of green hats certain number of black hats, what is the chance that you will be picking up a green or a red or a black, they are now mutually exclusive; it will all depend on the fraction of red being there or the fraction of black being there or fraction of green hats being there, in the in the in that basket where all the hats are put there. And in fact it turns out whenever you got mutually exclusive events the union will turn out to be the sum of the probability of these two, so this is again a very useful formula, which we utilized and we will be also using them, as we go in to problem solving.

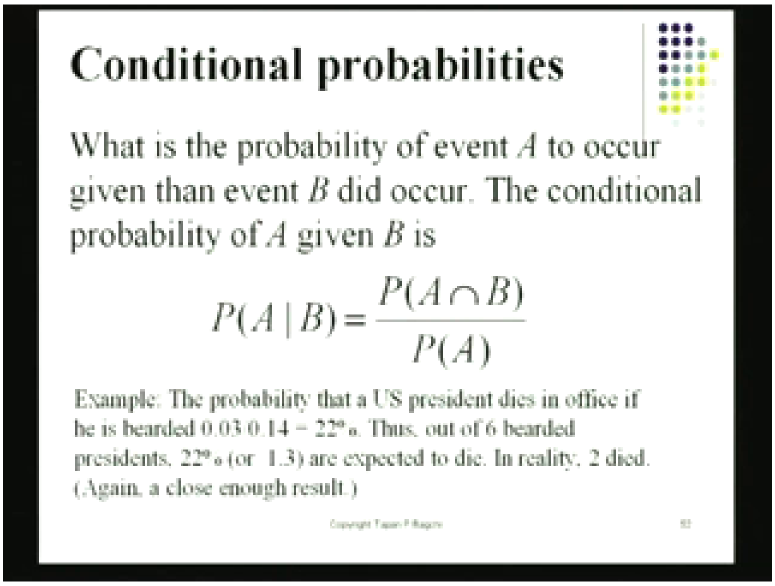
(Refer Slide Time: 19:44)



Now, I have a I have a situation here and I would just like you to read this and verify that this is so, I would like you to just read this and verify that you know being short in the office which is a a president being short in the office, dying out being killed in the office and having beard are these two situations independent of each other, being short in the office and having beard are these independent of each other.

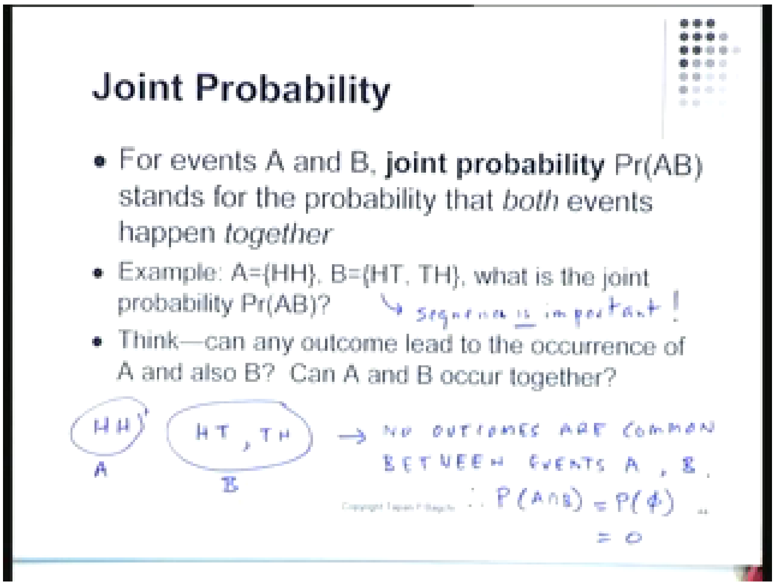
This we can verify by going into back, going going to history and looking into all the data, and at the data turns out to be something different then we will say indeed they are independent of each other having beard has nothing to do with getting short in the office, then of course, I will ask you to read I will ask you to read this slide and work it out from that.

(Refer Slide Time: 20:33)



Now, conditional probability this is something I used already and I told you that conditional probability is something that is there, and what they have done here they tried to use the conditional probability formula, and they have tried to show that P A times P B in this case, turns out to be the intersection and so on so forth, in fact it turns out P A and P B, if you multiply them you will end up P A and P, but I would like you to work it out on your own.

(Refer Slide Time: 21:06)



So, I will not solve that particular problem instead all this solving some other problem, let us take a look at all the other problem that I would like to what solve for you, I have this motion of joint probability what is joint probability, it is basically saying if I have two events A and B. The probability that they will occur together there will be observe to occur together; they might not be causing each other no, all they are saying is they are observed to occur together, it rains and I have poor marks in the in the test.

It is just joint occurrence nothing more than that, it is not that one is causing the other, that will require a lot of lot of verification that will require a designed experiment to really you know sprinkled water and see of grades can affected on that (()) or observe the situation when it is really cause an effect at of type of relationship; if that can be verified then only all see A is causing B otherwise, they just observed to occur together, that is all.

So, turns out the when were, whenever we talk for joint probability events are just seen to occur together nothing more, what kind of events are we talking about, we have got two events here, that we are playing with, one is H H two tosses and both are heads; that is the first event. First event is I toss the coin twice and I get two heads, the second event is I get one head, but first I get a head, then I get a tail, that is one outcome and the second outcome is I get a tail and then I get a head.

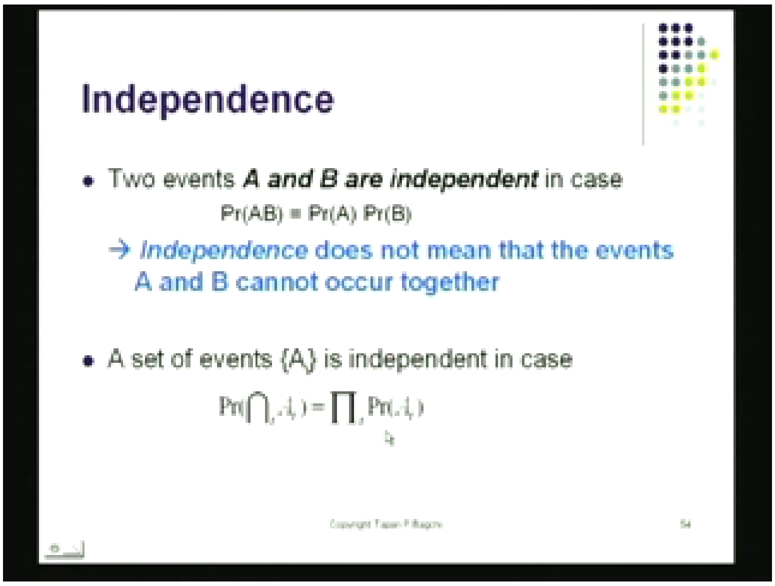
So, there are two out comes now, that constitute the event A, A event B it is head and tail head and then tail or tail and then head. What is the joint probability between A and B, can you really workout the joint probability, can you really workout the joint probability, there is one easy way to do this and that is what I have tried to do here, that is what I have trying to do here what I have done is I have sketched here event A and I have sketched here event B, notice something the outcomes have nothing in common, the outcomes have really these sequence of two events, one is the observation of head, then the observation of tail.

So, if I find head head I say event A has occurred, if I find head and then tail or tail and then head, I say event B has occurred, nothing is common between them, nothing is common between them, if you workout the probability there will be nothing common between them and you would, if you go back to that probability of A and B, A intersection B.

It will turn out to be the product of P A and P B you work it out in details and do that I am just saying right now, there is nothing common between them therefore, there being nothing in that intersection that probability is going to be 0; see in fact I show it here by saying the set A intersection B is null, the probability of that is 0, if this occurs the other will not occur, it turns out if this occurs there is nothing from here, that has also occurred there.

So, the probability at A is occurred it might be some quantity there, this this probability is going to be 0, in those experiments, let us carry on here.

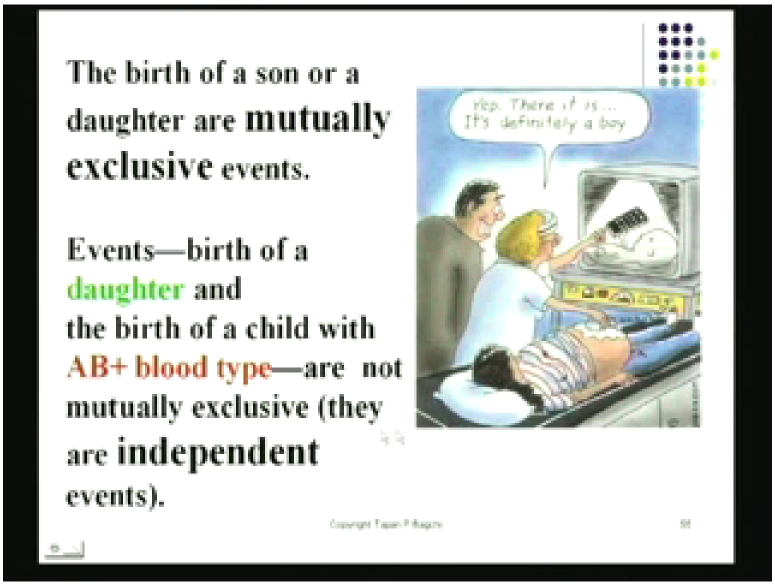
(Refer Slide Time: 25:01)



And let us go to another example, and I am going to be showing another example, which sort of goes like this remember whenever I said A and B are independent, my test was P A B, P A intersection B is going to be P A multiplied by P V and of course, I give you the reminder the independence does not mean that A and B cannot be observed to occur together; they might be occurring together but, it does not mean they are independent of each other, independence does not mean they cannot occur together, they might be occurring together and that is the joint probability, that is the joint occurrence of the two.

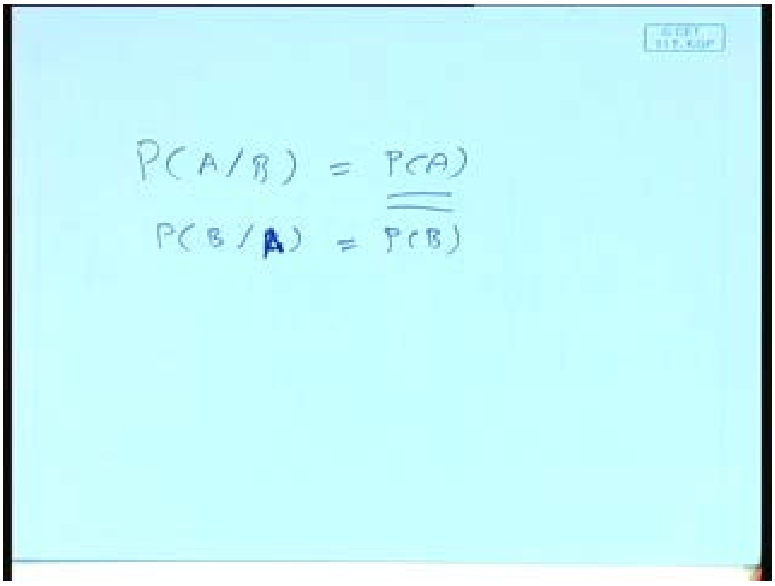
So, they occurring by an accidental also there will be somebody scoring see in an exam, those two events they may occur together or may not occur together, sometimes they occur together, sometimes they occur independent of each other and not really related to each other anywhere at all. If I have a set of events, if I have got A 1, A 2, A 3, A 4 a set of the events, then if I look at their intersection if they are independent of each other then this relationship will hold would hold good, probability of A, A A 1 intersection A 2, intersection A 3 and so on so forth. This is now, whatever is common between A 1, A 2, A 3, A 4, A 5 and so on, that turns out to be the probability of that turns out to be the the product of the probability product of the all the events occurring together that turns out to be this way.

(Refer Slide Time: 26:27)



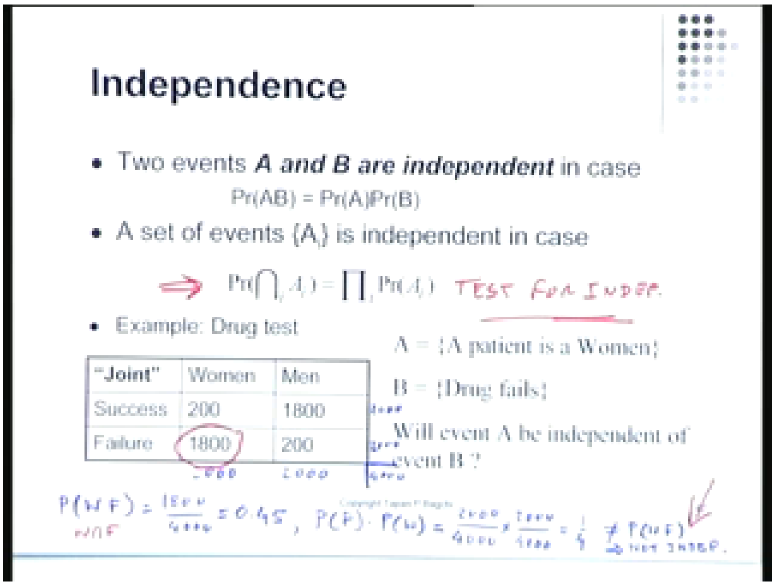
Here is an example, there are two events the birth of a daughter and the birth of a daughter, birth of anyone any child with A B plus blood type. Now, A B plus blood type is determined by biological considerations and different biological conditions lead to the birth of a daughter, these two actually turn out to be independent events, again if you do this test P A B, it will tell out to be P A times P B and that is the test for independence or you could actually say.

(Refer Slide Time: 27:17)



In another way, you could really say if I have test this I will say P A given B this turns out to be P A, on the right hand side I do not have B influencing anything at all and the same thing will be applying if A and B are independent I could say P B given A, B given A this will turn out to be equal to P B that is all; these are these are tests of independence these are tests of independence they actually say that B really has no influence on the occurrence of A or A also has no no occurrence, no no influence on the occurrence of B and let us try to work out one example.

(Refer Slide Time: 27:56)



Let us try to work out this example, we are working out many examples, and obviously we can stop the tape and we can actually rewind and so on, you could do that, but let us take a look at this example here, I have the joint occurrence defined here, how I defined it, I defined it like this a drug is used to treat people, and these people could be men or women and the result of this administrating this drug this drug is either the drug is successful, it heals people or it it fails to heal people.

And some data has been collected, so the drug was given to women and also to men who are sick and in a certain number of times they found to be the drug was found to heal people and certain other cases they were found not to heal people and the fractions are given here, the the the ratios are given here. So, I have here, a total of 200 plus 800 that is 2000 and 1800 plus 200 that is will again, 2000 again, so 4000 people were given this drug, 200 women they got cured, 1800 men got cured, 1800 women could not be cured, and 200 men could not be cured.

And let us talk about the events, that we are going to be testing against each other, A is the event that the patient is a women, what is the chance of the patient being, patient being a women it is going to be if we comes in, if the count comes in this column; this is the chance which is like 50 percent chance of the women of the randomly picked patient being women; and the chance of the drug failing I go to the failure row and I find 1800 plus 200 that is again 2000.

So, what I have here, I have basically got a count of how many women were there in the test, and also how many times the drug failed that count also I have, the question I am asking is are A and B independent, what are A and B, A is the event that the patient is a women and B is the event that the drug failed are the independent of each other, are these independent of each other.

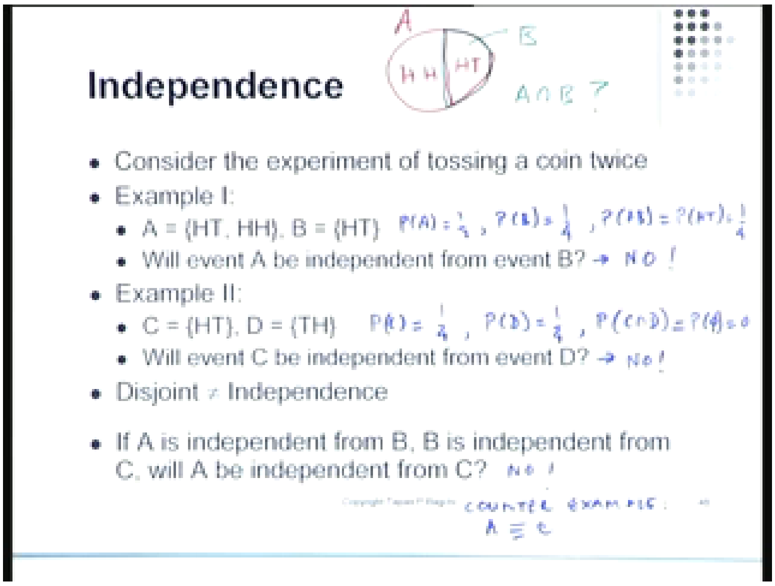
Let us try to see how we do that, how we work it out and for that what I have what I have done again is that taking the, I have taken a print of the slide and I have worked out the examples for you, I have here first of all you notice I have done the totaling, so I found out how many how many women were there in the test and that turns out to be 200 plus 1800 is 2000; how many men were there that again is 2000, how many successes as I had I had 2000 successes and I had also 2000 failures, a total of 400 people 4000 people were involved in this.

Now, let us take a look at this quantity we have, which is like the product of, which is the joint which is the joint probability of women and the drug failing, and that I can find by going to this joint probability table here, and it it turns out, if have to look for that count that is shown here, this is the count of the patient being women and also the drug failing that is 1800. So, 1800 divided by 4000 which is the total number that turns out to be 0.45 that is the that is the probability, that I had women tested and the drug failed, this is that is this probability, this is women and failure that is this probability here.

What I then do is I go into the pieces the parts of it and look at the probability of failure and this really got the marginal probability, the marginal probability of failure, that turns out to be some number, that will be like probability of failure is 2000 divided by 4000 that is this number. And the other part is the patient is women and that again turns out to be if I go to this column, and look at the total 2000 divided by 4000 that is this column here is, I have got 1 by 2 divided by 1 by 2, that is 1 by 4. And look at W F, W F is the joint probability which is 0.45 it is not equal to 0.25 therefore, it turns out, these are not independent, these are not independent.

So, the conclusion is that the test is the the the the situation here is not independent, the the the the drug failing and the patient being women these are these are not independent; this I get by by just applying this little formula here, this formula is the probability is the is the test for independence, this is going to the test for independence, this is the test for independence this is the test, that is the test for independence, that is what I have done.

(Refer Slide Time: 33:26)



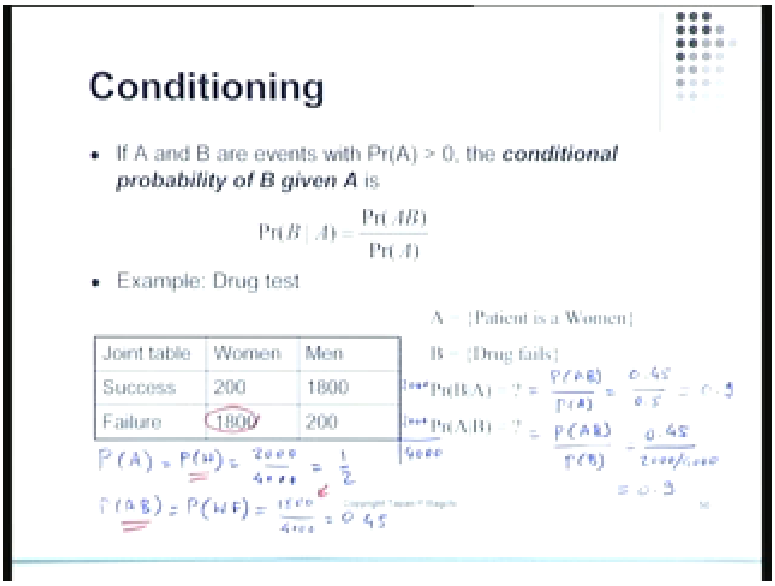
Let us see what else, we are being told about independence and I am going to be coming back again to one of the examples, consider the example when I toss the coin twice, and the outcomes for event A was defined as head and then tail or two heads and B was the event head and tail, will be event B event A B independent of B, this is a question I have, I have this question, what is the test I will be applying here, again P A B, is it equal to P A times P B for that, what I have to do is, have to calculate P A and also I will have to calculate P B and I have done that.

And let me show you, what the calculations look like, it turns out P A, which is like head then tail, which is half times of that turns out to be, 1 by 4 and also I could have the same event occur, if I have 2 heads on the chance of them occurring like this is again 1 by 4. So, 1 by 4 plus 1 by 4, because these are these are two different outcomes and I am talking about this occurring or this occurring therefore, I add the probabilities.

So, P A has the probability of 1 by half 1 by 2, P B is now this case here first getting a head then getting a tail, getting a head is one half and getting a tail is also one half for a multiply there two, because they must be together, that turns out to be 1 by 4. Now, I have got P A and P B what is the chance of my P A B, P A B is what? P A B that is a has occurred and also B has occurred, it is a very funny situation, I have situation like this I have A here and a consists of two things A consists of H H and H T, and guess what B consists of B consists of this part, this is B.

So, what is going to be the intersection of A and B, that is P A B, A intersection B, what is that set, that is this and what is the probability for this, there is 1 by 4, so here again I have got a situation when P A turns P B does not equal this does not equal this, that means again they are not independent of each other. And you workout the other examples on your own, you can work out and I am providing you the solution, and you can come back and you know you check this again, check to make sure that yes indeed there is problem, there there is something we could do.

(Refer Slide Time: 36:33)



Let us take a look at another situation, when I am talking about conditioning, and let us see how I work with the conditioning problem, I start by again having a joint table and the joint table looks like this, if you look at the screen here, the joint table looks like this. And what I am going to be checking is, what is going to be P B given A, I just have to calculate that and for that all we just using I will be using the joint table and I will be looking up for P A B, I will be looking up for P A and I will workout this definition (()). Now, this P B given A will be equal to P B A divided by P A, that is all P A given B will be equal to P A B divided by P B (()), if just see that situations there, this is exactly what I have done here.

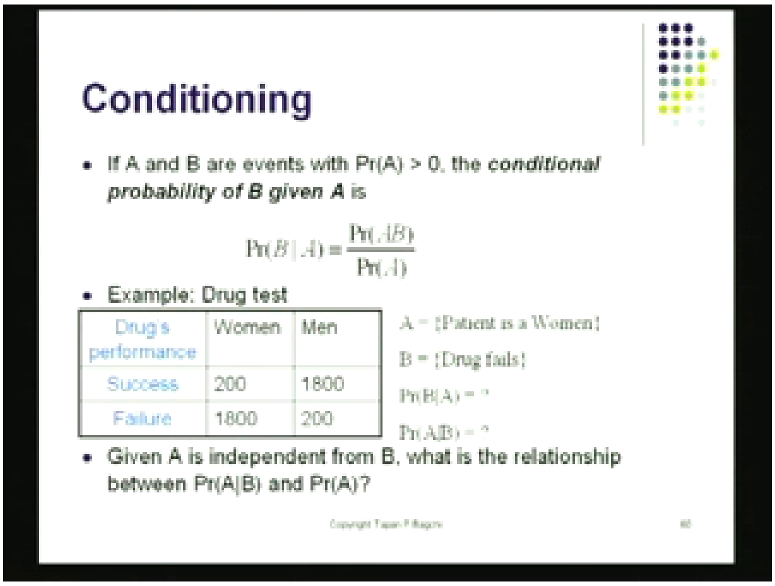
So, conditioning calculations are quiet easy and quiet straight forward, and I what have what have I done here, I have here if you look at my sheet, now P B given A is equal to P A B divided by P A that is P B given A. And I got this P A B already calculated I did the

the fractions before, and so I know that those numbers there I got 0.45 divided by 0.5 that is P A; that turns out to be 0.9, that is one conditional probability, the other conditional probability turns out P A given B there is some symmetry in the data therefore, that also that answer also turns out to be 0.9.

Now, let us just go back and remind you, how I found my P A and P B and so on, what is P A, P A in this case is going to be patient is women patient is women, that is really means how to count up the number of women that is 2000 divided by 4000 that is going to be my patient being women which is this part, that turns out to be halt, they have 50 percent probability. Because half the population is women, this 50 percent probability that they randomly picked person, randomly picked patient is is a women and that turns out to be half. What about P A B, P A B is when the women is when the women is, when the drug fails when a women is using it, and that is this cell here, this is the cell where I have got women using the drug and the drug failing.

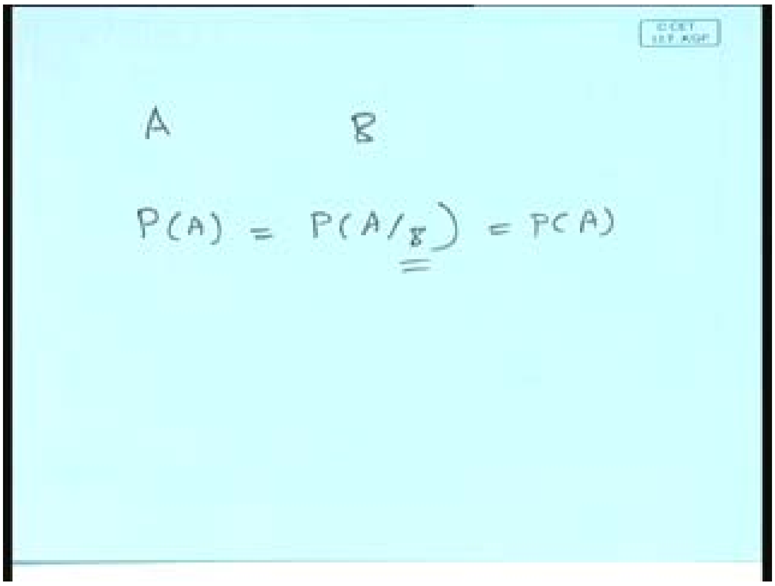
So, that 1800 number I put there, and I have 1800 divided by 4000 and that turns out to be as we have done before with our machine, it turns out to be 0.45, this is I found P A B, which I bring here and I divide that by P B and I end up with my other numbers there.

(Refer Slide Time: 39:45)



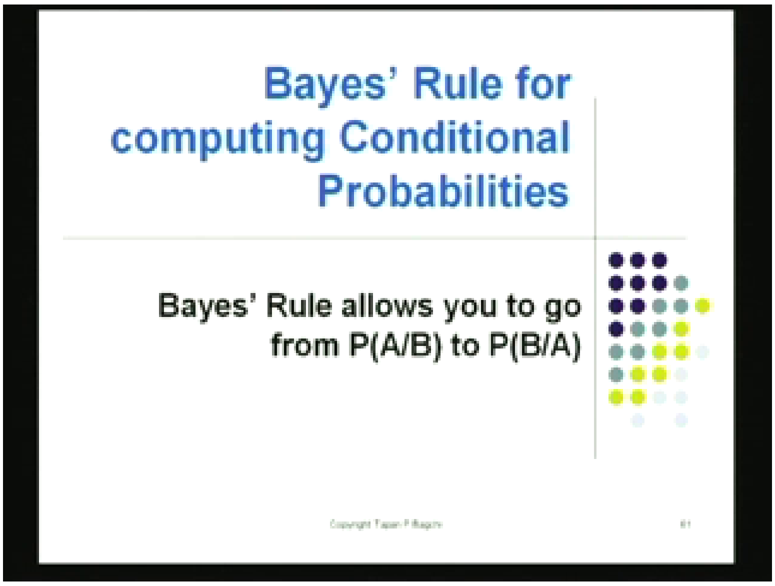
So, conditioning in probabilities are quiet easy to calculate once you have the discipline there with this we can move on and we can actually test, if there is any relationship there. If given A given, given that A is independent of B what is the relationship between P A and P B, this I am sure you can work out, let me ask you the question again and I think you will understand that there are two there are two events.

(Refer Slide Time: 40:12)



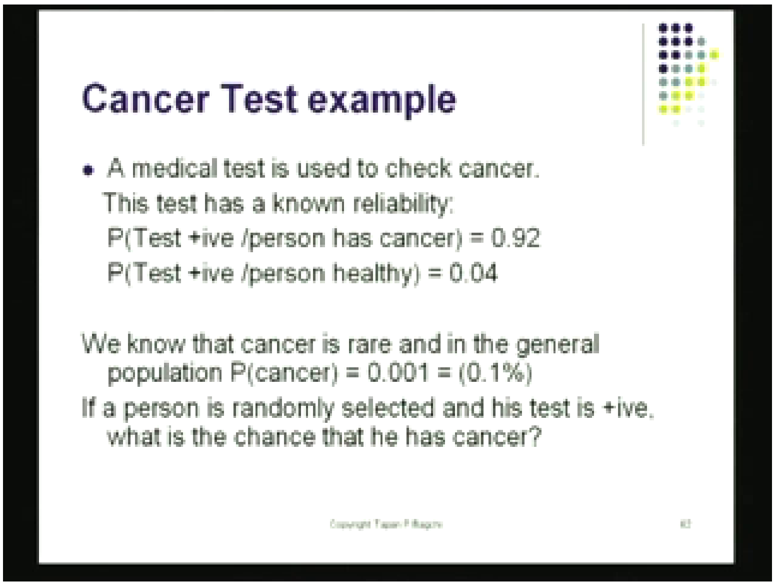
One event is A the other event is B, and what they are saying is A is independent of B, that means P A is the same as P A given B, this is what they are saying they are actually saying that, what is the relationship between P A given B and P A; it turns out they are equal, because this is not dependent on B this is actually equal to P A. If I only know this I do not have to worry about this, because because A in no way is dependent on this, so this conditioning really has no effect at all, whether B has occurred or not it has no impact on the occurrence of A. So, it is a purely straight forward question nothing more just a little tricky question, they just skipped in this little question for you.

(Refer Slide Time: 41:08)



Then we moving to something which is an application of what we done, and this turns out to be a very vital sort of application of conditional probability, and let me tell you that in it also gives us an indication of how good our quality control factor, quality control inspection will be like, and I will illustrate that using some assumptions and I am going to solve it for you, I am going to show you the solution also for this.

(Refer Slide Time: 41:46)



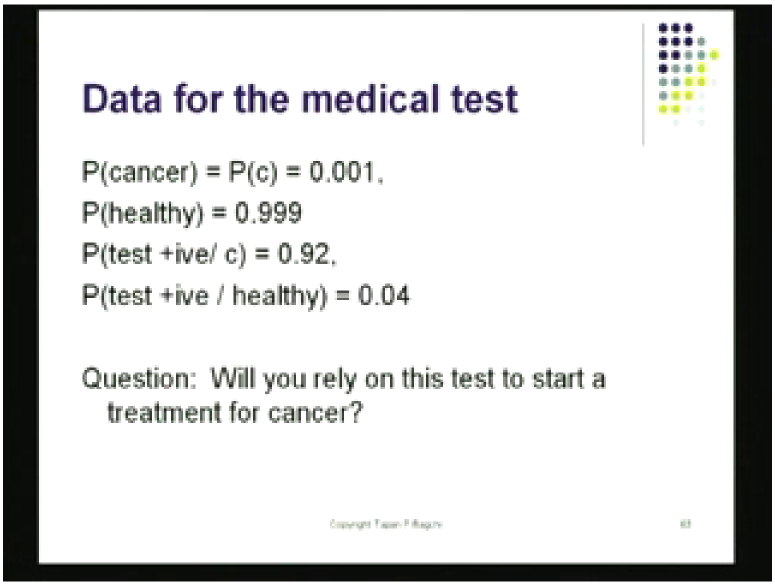
Let me first give you a picture, let me first give you the scenario and to do that what I will be doing is, I will be bringing you situation in a hospital, the hospital is using some device may be mammogram or something it is using to test or some set of a scanning to test to see if a person has a, cancer it turns out that if a if the test says positive, it is not true that the person always has cancer.

In fact the, if the person does have cancer, look at the screen now look at the screen, now look at the display here it says, if the person has cancer the test will be positive, the probability of that is only 0.92, only in 92 percent cases with a real cancer, the test will say that the patient has cancer, really the patient already has cancer, but these only in 92 percent chance that the test will catch it. The reverse situation or bring it a healthy person and he goes there takes the test there is in mammogram, and then the test shows positive 4 percent of the time, so it is a perfectly healthy person he takes a test and because, the instrument is not perfect 4 percent of the time it gives you faulty result.

Now, these numbers 92 percent seems sufficiently high, and 4 percent it seems sufficiently low, so you might say say it is I think I will get that instrument; I will have it installed in our hospital. Now, what we would like to do is, we would like to do this analysis using probability theory, would like to find out is it reasonable for me to purchase this equipment, how often is it going to get me false signals, what kind of false signals we will work that out we will work that out if a person is randomly selected and his test is positive; what is the chance that he really has cancer, think of this again let me repeat the question, the test is positive, what is the chance that the person really has cancer, knowing that the instrument is not perfect.

Let us try to work this out, what had been given of course, is that in general, in the wide population 0.1 percent people have cancer, which is probably true for this area of the geography, it would be quiet different if you are in new jersey or some other place where the chance of having cancer is you know it could be double digits. Because, the there lot of chemicals of the area, lot of garbage dumps and so on, and the air is not clean and so on the forth, it would be true many other places also what new jersey are knowing particular, because I lived there and I saw all these problems there.

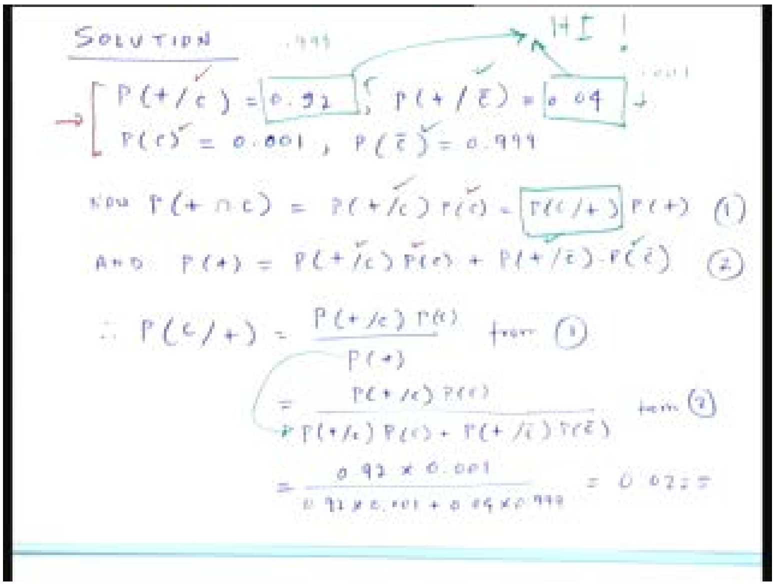
(Refer Slide Time: 45:06)



Let us see, how we work this out, so what is the data that has been given to us, I am first presenting you just the data and the data is this in the wide population 0.1 percent of people have cancer and that means 0.999 is the probability of the person being healthy; no problem there. And the instruments capability is the following it shows the test to be positive, when there is cancer 92 percent of the time, which some people may think is pretty good and also unfortunately for a healthy person also also it says that the person has cancer, 4 percent of the time this is the instruments performance.

The question that is being ask this a manager question, will you rely on this test equipment, and will you start a treatment only because, you start a cancer treatment which is pretty severe, just because the machine said you have cancer, will you do that, given this given this being the record, this being the story. Let us see, how we solve this problem, what all we doing is I will be walking you through some calculations, be just a little patient and I am going to work out the problem for you.

(Refer Slide Time: 46:29)



And we will be concentrating straight on the on the sheet there I worked out for you, it took me little while but, it still could be done on one sheet, so I thought to I I should preserve this for the class here; these are given, these these things are given to us, the probability that the test is positive given that there is cancer is 92, the probability that the test is positive given that there is no cancer that is a healthy body is 4 percent. In the wide population outside, the probability of a person having cancer, persons is having cancer is only 0.001 and the person being healthy that means no cancer is 0.999, these are given to us.

Now, let us try to work out some questions, what is the questions that are I am asking, I am asking this question, what is the chance that the person has cancer given that the test is positive, given that the test is positive what is the chance that the person has cancer, this is what I have to evaluate this is what I have to evaluate, and for that the actually I will have to go back into what we have here the data, I will have to find this, I will have to find this and I will have to find this, how did I find this formula this came straight from our conditional probability conditional probability formula, I will going to show you that formula in just one minute I will show you (Refer Slide Time: 47:48).

In fact what, now we have to do is, we have to look at these things, look at this intersection here, probability that the test is positive and the person has cancer it is equal to the probability test is positive given he has cancer multiplied by the probability his cancer, do I have this data, yes I have this data, so I have this data here, do I have this data, yes I have this data it is also given to me. So, I can calculate this quantity this quantity I can calculate without any (()) and it turns out this is also equal to the probability this way, because I I can flip these two this is a this is, this saying intersection condition.

So, whether I take C first or I take plus first and positive indication first it does not really matter, so here what I have done I have taken I have made C condition of the test being positive and this, so I have actually on this side I have something that is equal to this, on the left hand side I have got the quantities I have got the quantity, I have got this quantity I know it is value it is value is going to be the product of these two; that I can find from these two.

Then I look at this quantity here, which is the probability that the test is indicated to be positive, when can that happen, that can happen by this, probabilities that it test is positive given that the person has cancer multiplied by the probability of that the person has cancer; and that is found from here, so I know this and I know this, no real problem there. And then this part which is like probability that the test is positive given that there is no cancer multiplied by probability that there is there is no cancer, where can I find this I can again look up my given data and it turns out that this I know and also this I know from here.

So, I know all quantities, so I can evaluate this quantity for sure, now the only thing I need to find out this probability of C given that the test was positive, because this is now, this is something that I am going to be finding out, I am going to be calculating that. So, let us try to see, if I could calculate this quantity here, what is this now what is this probability this is the critical question, what is the chance that the person has cancer, given that the test is positive.

Now, look at this, look at this statement here, look at this statement 1, statement 1 says probability that test is positive given that the person has cancer multiplied by this this is the same as probability of test being positive and probability having and and the patient having cancer person having cancer that is also equal to this is like probability A given B multiplied by P B and these probability B given A multiplied by P A that is what it is, so I have really written the same formula, once here another time here just by changing the order of the thing, so it is the same formula.

This is valid from the probability of, these valid from the probability of conditional probability formula, I use this formula and I am after this quantity I am after this quantity what is that quantity just try to put a green bar there, I am after this quantity this is the quantity I am after, I know this and I know this and this quantity I have already evaluated by equation 2; so I know this and I know this and I know this therefore, I can find this and this I write write here, probability of they are being canceled given that the test is positive is equal to this quantity here I just took this, this is what I took and I put that in the numerator and I divided by the denominator there (Refer Slide Time: 51:47).

So, from 1 I found this, then from 2 I find the probability of the test being positive which turns out to be this quantity this quantity, this is the quantity that I put down here, using which equation equation 2, so I have a quantity here, where I have got all the numbers, I have got 0.92, I have got 0.001, I have got 0.92 again, I have got 0.001 and I have got 0.04 given to us and I have got 0.999, this have done. And if I do that, and I use my machine I do (()) whatever and so on, as more than I do some multiplication divide and so on, I get 0.0225, that is the probability believe me, that is the probability of a person having cancer, if the test is shown as positive the instrument says the person has cancer, and he is truly having cancer is this.

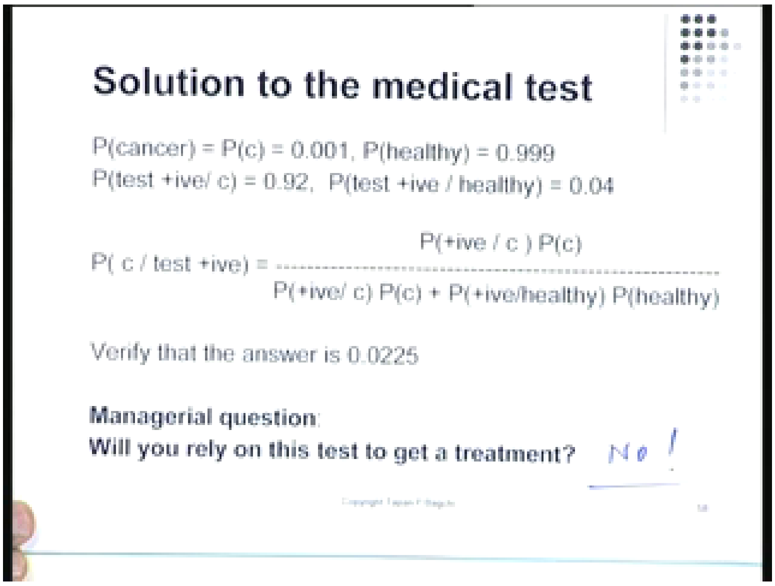
So, the probability having, truly having cancer given that the test is positive is only 2 percent will you trust a machine like this will you trust a machine like this, what we have worked out is not known to many people in medical practice, because they have not gone through this, and just see how serious this matter is, how can we fix this situation, one way is start playing with these numbers here, this 0.92 notice here there is a quantity called 0.92, which is the goodness of the instrument, this is also the goodness of the instrument, these are the factors I need not really worry about population statistics.

But, I must worry about this number and this number these have to be sufficiently high these have to be high, in fact this has to be low, and this has to be high, this has to be low, because this is saying there is cancer when there no cancer, and this is saying there is cancer and it saying positive, this should be as high as possible, what we have done we have taken this two 0.999 and we have reduce this to 0.001, if we do that this result stilt only 50 percent, that means you need a machine that must be much better than what these fellows are supplying to you.

That the performance level of these people are supplying it to, this is something we got to keep in mind, and this is the analysis that you got to go through, this is the analysis you got to go through before you make a purchase, you should not go just by this 0.92, 0.04 they look great, they sound very good. But, when you actually do the calculation, what is the chance of the person having really having cancer, when the test is positive that turns out to be reduce for 2 percent only and obviously, we cannot really send these people for chemotherapy with this sort of probability we cannot sell send people who indicate to be positive to chemotherapy, we just cannot do that, we cannot do that; this just kind of gives you an idea, how important this conditional probability theory is it is very very important.

This gets into quality assurance, this gets into production manufacturing, this gets into any kind of improvement that you looking for like for example, if you want to remove false alarms or bad treatments; then hospital is a test facility and a hospital also is a service facility. It should provide good service, if it is based on this sort of data it is not really possible for this this particular facility to survive for too long and people can take them to the court, if this the can have instrument that is being used and false treatments have been given only there is 2 percent chance of the person really having cancer, when the machine says he has cancer that being so low; people can be taken to court and just be mindful of this. So, when you get into a situation like this please think of this example and this is a pretty famous example, that is use by many different people.

(Refer Slide Time: 56:33)



So, will you really get a treatment and our answer is going to be plain and suppose no, if this is the performance if the, if this kind of performance is what we are getting from the instrument you should not get treated in that hospital; certainly you should not go for treatment, rather you should try to get other test and try to see this number could be improved. And if you not 0.999 is not so good, so that is like a lesson for us, we will continue with this as we move along, thank you very much, thank you.