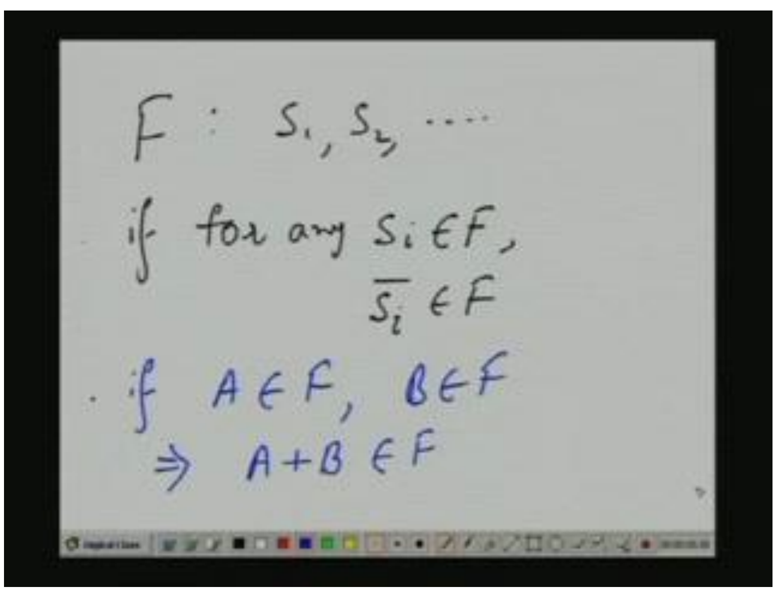
Probability and Random Variables
Prof. M. Chakraborty
Department of Electronics & Electrical Communication Engineering
Indian Institute of Technology, Kharagpur

Lecture - 3
Axioms of Probability (Contd.)

So, in the last class we discussed the axioms of probability their properties. And, then we discussed something called field of sets and in that connection we mentioned Borel set Borel field. I will just touch up on those things today to start with and also explain the motivation for considering these, but this is again only for you know mathematical correctness. But having just done that, I will move on to the motion of conditional probability.

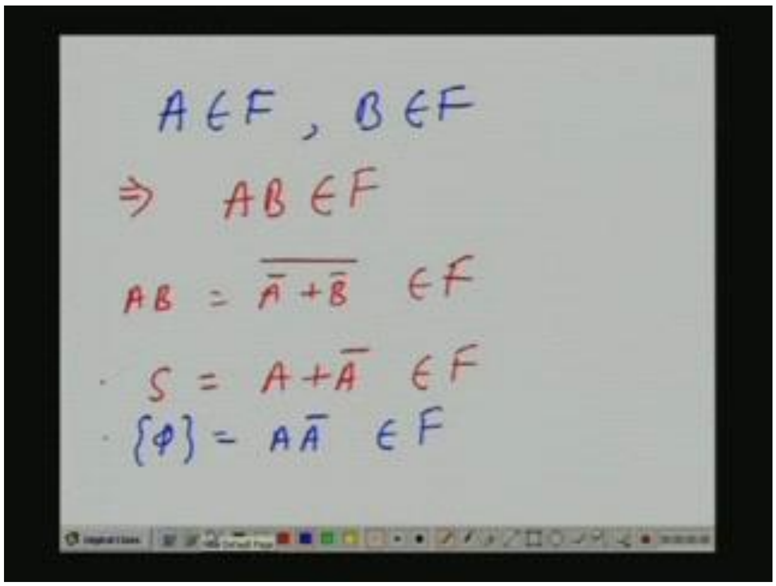
That I will be explaining first notionally and through examples, then using this conditional probability I will find out it is properties. Then we will discuss something called total probability and Bayes' theorem again we will take up some examples, and then we will go for statistically independence of 2 events. And we of course, try to give physical interpretation of these.

(Refer Slide Time: 01:59)



So, as I mentioned in the last class I will consider a field I will consider a field of sets F. It will consist of suppose sets S 1 S 2 and dot dot dot so on and so forth. Now, this will be a field if for any Si, I would say element, because is a field means is a class of sets. So, each set is a element of this class F we should have Si bar also member of F that is, if a set belongs to this field that is complement of mass belong to the field number 1. Number 2, number 2 is if A is an element of F and B is an element of F, then union also is an element of F this 2 conditions must be satisfied. There are some properties which follow these which will be now considering.

(Refer Slide Time: 03:40)

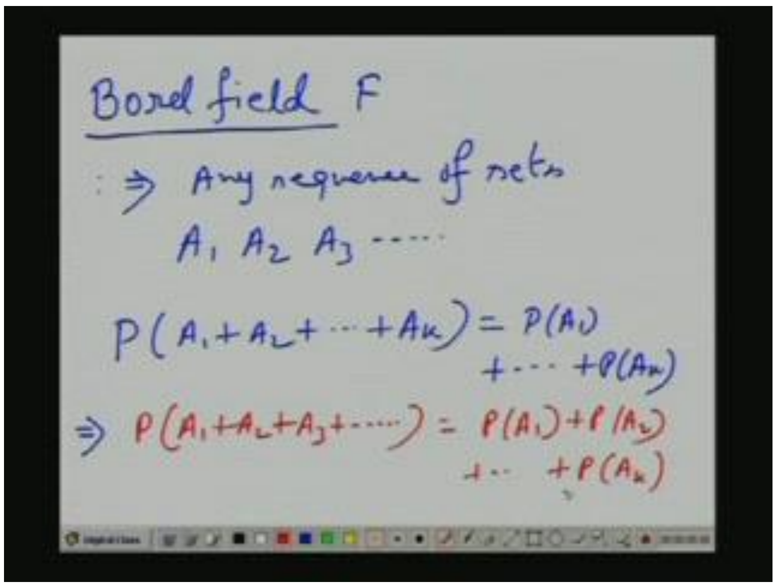


First, let us consider this thing it suppose A is element of F B element of F, then will show that they intersect at AB it is also an element of F it is not very difficult to show. Well. We can always write AB as using De Morgan's law. Now, we know that if A is an element of F, then A bar is an element of F. Similarly, B bar is an element of F and since this is an element of F this is an element F their union is an element of F and complement of that union also is, then an element of F which means this is an element F.

Another interesting observation is that this field consists of both the total set S and the empty set phi, because total set S is nothing but S is A plus A bar and A is non empty. Because, the field F was defined to be a non empty class which means there has to be at least 1 non empty set say A. And therefore, it is complement when put under union they give rise to S and since, A is a member of F A bar is a member of F and their union is a member of F this means that S also is a member of F.

Similarly, empty set phi can be written as AA bar and once again A is a non empty set, because there is at least 1 non empty set A member of F and AA bar. That is intersection between A and it is complement is empty set and this is both A and A bar are member of F their intersection also is member of F.

(Refer Slide Time: 06:04)



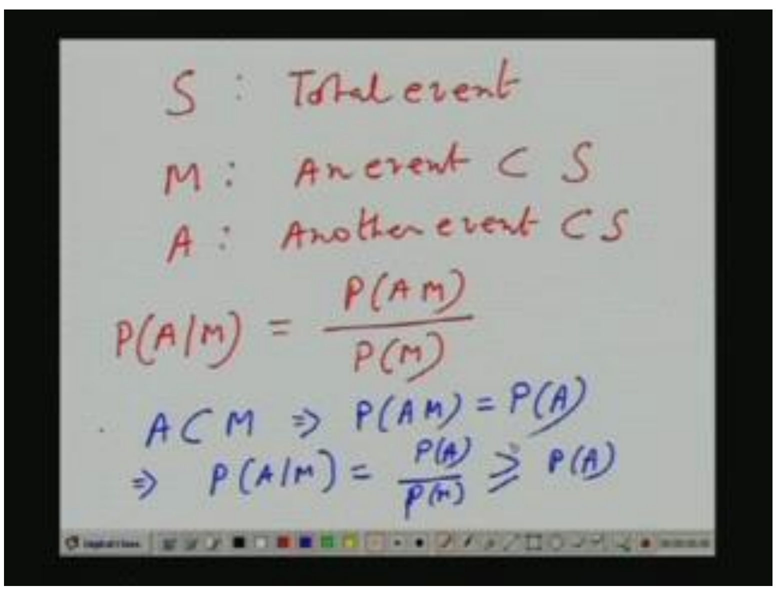
So, these are the properties of a field of sets. Next we consider what is called Borel field. The Borel field F it is such it implies that if you take any sequence any sequence of sets. As you see, the sequence consists of infinite an element which is a set, then not only A 1, A 2, A 3 and other members of this sequence are member of F. But, their union intersections for instance A 1 union A 2 intersection A 3 union A four and like so on and so forth. Any finite or infinite union or intersection they are also member of F this is called a Borel field.

Why we are doing this is, because so far when we have defined this probability axioms we have considered basically finite sets and we said that the set consist of the set actually is the total event it consist of experimental outcomes. And we formed various subsets using those outcomes and we call them events. But, why we are dealing with infinite sets, then all possible subsets really do not qualify for events, because these probability axioms cannot be applied to all the subsets there will be some mathematical difficulty.

Then we say that in such cases those subsets which formed a Borel field they would qualify for events and these probability axioms can be applied there. For example, we know that if we had some events. So, total k events A 1, A 2 Ak and k is finite and they are mutually exclusive then; obviously, we can write this as summation of the individual probabilities question is question is, if it is an infinite union can you write like this.

So, mathematically we cannot just jump at this you know that this fact that the statement that we can write actually for that to happen A 1, A 2, A 3 and other members must form or must come for a Borel field. Then we can we should able to write is an infinite sum alright. So, this we are discussing just to I mean remain mathematically correct for that purpose only I introduced you to the concept of field of sets, and then Borel field. So that, our definition of this in the basic axioms they are more complete and more correct. So, that is all as per as this axioms are concerned.

(Refer Slide Time: 09:45)



We now move to a very interesting topic that is called conditional probability. Conditional probability means that, suppose you are dealing with total event S as before that is S is the total event S is the total event and M an event basically it is a subset of S and A is another event again subset of S. First I give you definition, then we will come to its physical interpretation. Conditional probability we write like this P of A given M is defined as PAM is a joint probability.

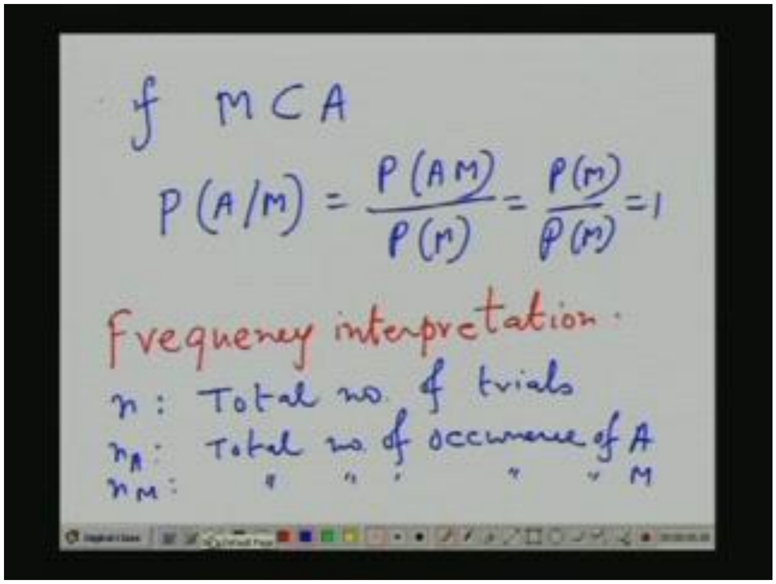
That is the probability of the intersection between A and M that occurring that event being the intersection between A and M that probability of that event divided by PM. Actually, by this conditional probability what we are trying to say is this that look P of A means probability of the event A occurring. That is, if event A has various possibilities that is various elements forming the subset. Then for any of this elements occurring not being true is mean by P of A P of A by M means, suppose we already know that event M has taken place.

Then subject to that knowledge what is the probability that event A also has taken place. Now, how that comes from this definition and all will see soon, but certain things followed from here. Firstly, suppose it is given that this set A is nothing, but a subset of M that is the event A is already content in event. M in that case, we know PAM what is the intersection between A and M intersection between A and M is A itself this is equal to PA.

In that case is, PA by PM which is surely less than equal to PA, because PM is what it is not I mean I am assuming that PM is non zero. So, this division is allowed alright it should be sorry it should be greater than equal to that is PM is less than 1; obviously, PA by PM is greater than equal to PA. When PM is 1, then only it is equal to otherwise it is greater than PA and I am of course, assuming that PM is non 0 otherwise this division is not allowed.

It means that if we are considering the occurrence of the total set total event which is bound to occur always. And subject to that, we will see the probability of A, then of course, probability of A is less, because I am really considering all possible events. And in the background of that I am trying to tell you something about the probability of A. But, when I am not considering the total event S, but I only considering the event M and subject to the occurrence of event M, then I am trying to tell you the probability of occurrence of A. Obviously, chances will be more, because since I am restricting myself to M first I am basically not considering all other possibilities as a result chance of observing A higher alright.

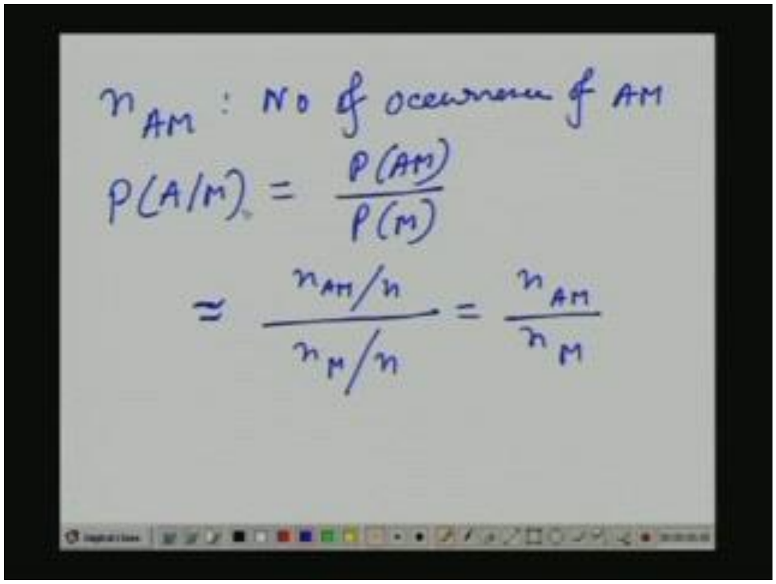
(Refer Slide Time: 14:23)



Secondly, if M is contend in A this event M is basically a subset of event A which means P we had this. But intersection between A and M in that case is M and this is PM by PM which is 1 or obviously, if M is contend in A, then the fact that M has occur itself means A has occurred. So obviously, the probability is 1.

So, what we are trying to do we first enforce some definition and finding out some properties and tying to correlate those properties with some physical observation physical facts applied. But still a better interpretation will come up if you consider the frequency interpretation. That is suppose, you are conducting an experiment and suppose you are conducting an experiment and event. Total number of trial is n or first let me write down n is the total number of trials that is, you are conducting an experiment total number of times you conducted is small n. NA and of course, small n is very large nA total number of occurrence of event A.

(Refer Slide Time: 17:06)



Similarly, nM same as this, but of event M and nAM is the number of occurrence of AM that is the intersection between A and M. In that case, if you apply our formula, then PA by M is PAM there is the probability of the joint occurrence, there is the intersection between the A and M divided by PM. And what happens is M AM what is PAM we are having total trial small n out of which M AM times we are getting, this intersection that is both A at M simultaneously.

That means, the corresponding probability I put an approximate sign here, because once I get into the frequency based interpretation. I have to have some approximation and approximation becomes better and better if total number of trials is larger and larger. So, this is divided by n this is the probability standing for PAM divided by. And what is capital M again divided by n out of n trials I got nM times this event M occurring, which leads to this and try to understand the physical meaning of this.

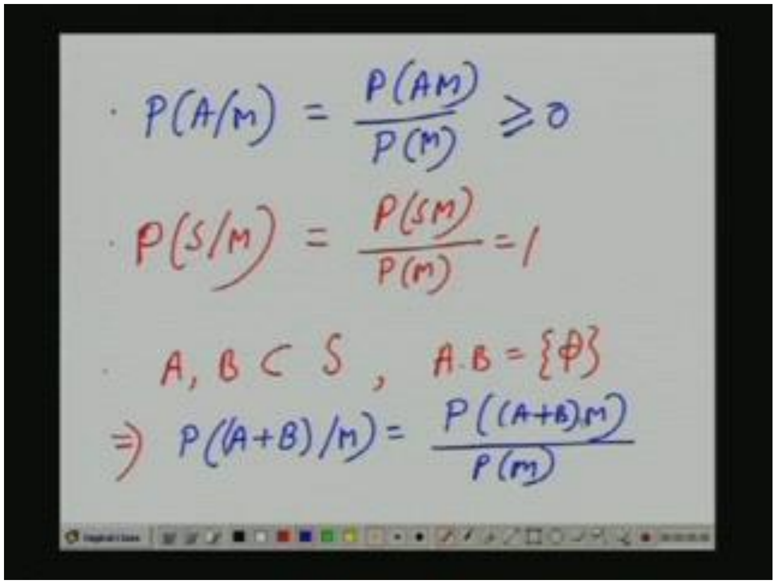
What it means, that is it is that we are conducting experiment and you are having total very large number of trials for n. But, then we are not as far as this result is concerned. We are now not looking at all the n trials we are confining ourselves to only those trials. Where the event M had occurred and how many such trials we have nM out of this nM trials. On how many occasions did you have A occurring? Obviously, I have to see the intersection between A and M, because that is when both A and M occurs and that is nAM.

So, what is this giving this gives you the conditional probability of the event A occurring subject to the fact that M has already occur. I repeat again nAM is the n was the total number of trial, but out of which if you just consider those occasions only, where M has occurred, then total number of such cases is nM. So, out of this nM cases where on each occasion M has occurred we find out on how many occasions A has occurred.

For that you have to see the intersection between A and M and see how may times we really got this intersection when these corresponding event coming up that was this. So, what is the corresponding probability? That is total trial with M only (( )) with M known to be occurring divided by number of times A occurs with the fact that M has occurred. So, this is called conditional probability. So, it means that subject to the fact that M has occurred what is the probability that A also occurs this is this goes for P of A by M these are physical interpretation.

Now, 1 can ask me this question that I have defined 1 particular ratio as conditional probability, but still is a probability. So, does it satisfy those basic axioms of probability for it to be acceptable as some set of probability? So, we can take a look into that and see that these are all satisfied.

(Refer Slide Time: 21:16)



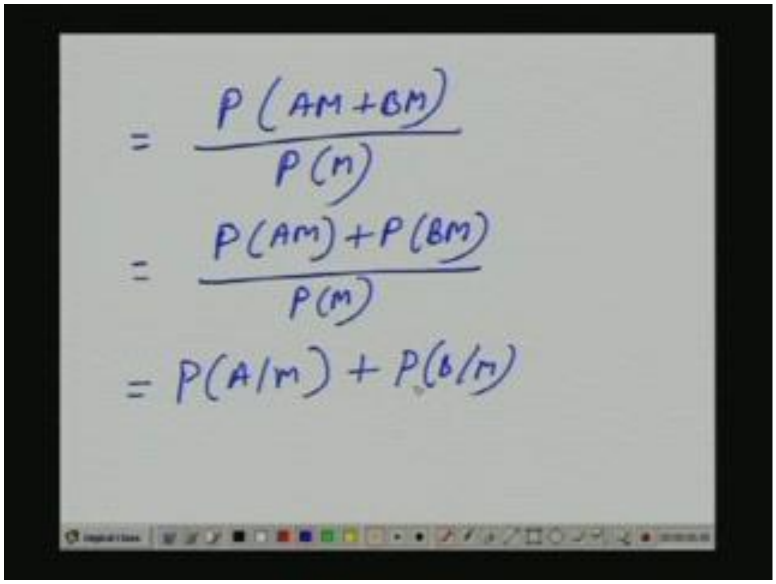
First PA by M this was now both these probabilities are non negative and I am as of course, assuming this is nonzero, because M is not an empty event. When you consider this conditional probability with your quote I mean the event here to be the impossible event. Suppose, this probability it is not 0, but neither of them is negative which means this is always greater than equal to 0. Of course it will be 0, if this is 0 and this becomes 0 if A at M are such that mutually exclusive. So, when A occurs M does not occur and vice versa only, then it is 0.

So, this satisfies axiom number 1. Then if S was the total event how about P S by M what is the probability of the total event. That is P SM divided by PM and as I told you earlier S intersection with M is M itself, because M is fully contend in S, because this is the total event total set. That is S is a certain event which means PM by PM which is equal to 1.

So, axiom number 2 also is satisfied, and now consider axiom number 3 suppose A B they are 2 events that is they are 2 subsets of the total S and they are mutually exclusive. AB is phi, then how about A union B you can put a bracket around this subject to M condition to M. Now, we know first we will use the definition which is nothing, but A union B intersection with M divided by P of M.

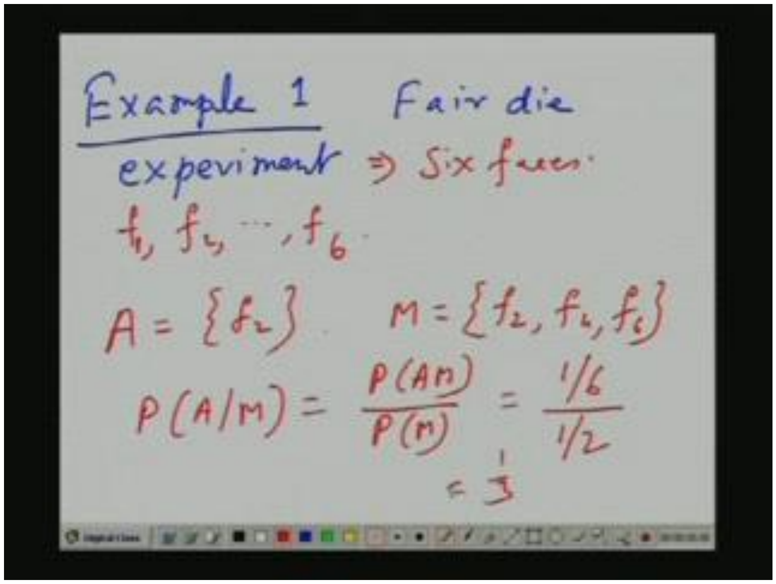
Now, we know we have already seen the distributive property of this sets I mean set operations union and intersection. So, A union B intersection with M is nothing but A intersection with M union B intersection with M.

(Refer Slide Time: 24:32)



So, this leads to, but A and B they are mutually exclusive. So, AM which an intersection between A is and M. So, this contend in A also and BM which is an intersection between B and M and (( )) part of B also. They are 2 are mutually exclusive, because A and B they are mutually exclusive. Which means, this probability can be written as and PAM PAM after all what is PAM by PM that is by our definition for conditional probability, that is nothing, but PA by M and what is P BM by PM similarly PB by M. So, M axiom number 3 satisfied. So, this is indeed a probability a motion of probability it is called conditional probability. We now consider some examples to have some further insight about this.

(Refer Slide Time: 26:08)

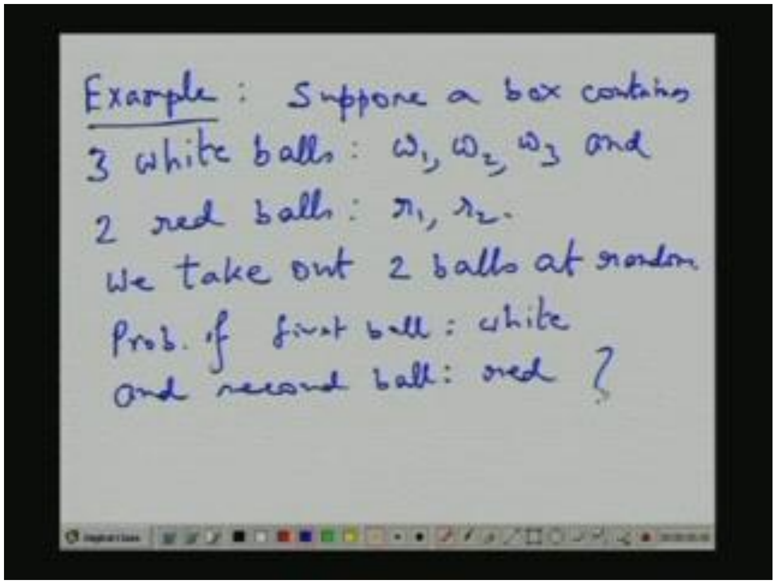


Example 1: Suppose, we consider this experiment the fair die experiment fair means unbiased it has got six faces f 1 f 6 f 1. Means the face which is marked as 1 f 2 is the face which is marked as 2 and so on and so forth up to f 6. And since, it is unbiased I mean the probability of each face showing up is just 1 by 6. Suppose, this is the case and we are crossing this die question is if you have given this if you are given this event A which is nothing, but F 2 coming out.

Obviously, the probability of f 2 as such is 1 by six giving all possibilities that is giving the total event meaning 1 of the face is standing up always, then meaning the background of that the probability of A event A is nothing but, 1 by six. But suppose, I consider another event M which stands for even faces either f 2 or f 4 or f 6 that is, faces which are marked as f 2 or f 4 or f 6. Then subject to the fact that M has occurred, but you were only considering those cases where the face that can source up on top that is an event 1.

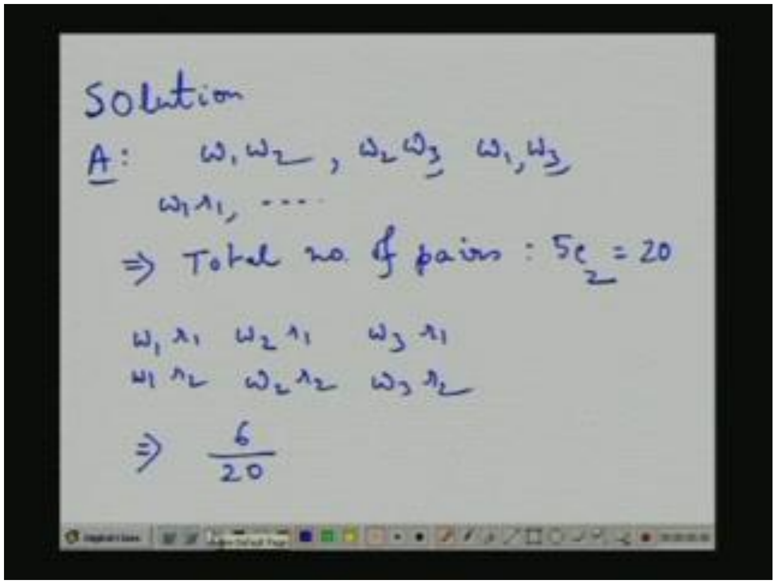
Subject to that or under that circumstance what is the probability that particular face is neither f 4 nor f 6, that is neither 4 not 6, but f 2. So, that will be by your formula and this P of A by M will be P AM by PM. In this case you see intersection between A and M A is already a subset of M. So, intersection between A and M is nothing but A. So, P of AM A intersection M is nothing, but P of A which is 1 by 6 and P of M. We have got 1 1 1 3 elements. So, probability of even faces standing up is nothing, but 1 by 2 half for odd faces half for even faces. So, it transferred to be 1 by 3.

(Refer Slide Time: 29:50)



This example number 1 (( )) … one more example: suppose, a box contains 3 white balls w 1 w 2 w 3 and 2 red balls r 1 r 2 we take out 2 balls at random we take out 2 balls at random. What is the probability of first ball white and second ball red what is the probability is a question? We can approach these problems in 2 ways we can approach this problem directly within the basic axioms of probability. That is, we will consider all possible I mean events out of which we will find out on how many occasions an event like this will take place and take a ratio and that will give us a I mean the solution, but that is what approach. So, you can also approach this problem in a smarter way using the motion of conditional probability. So, we follow both.

(Refer Slide Time: 32:03)

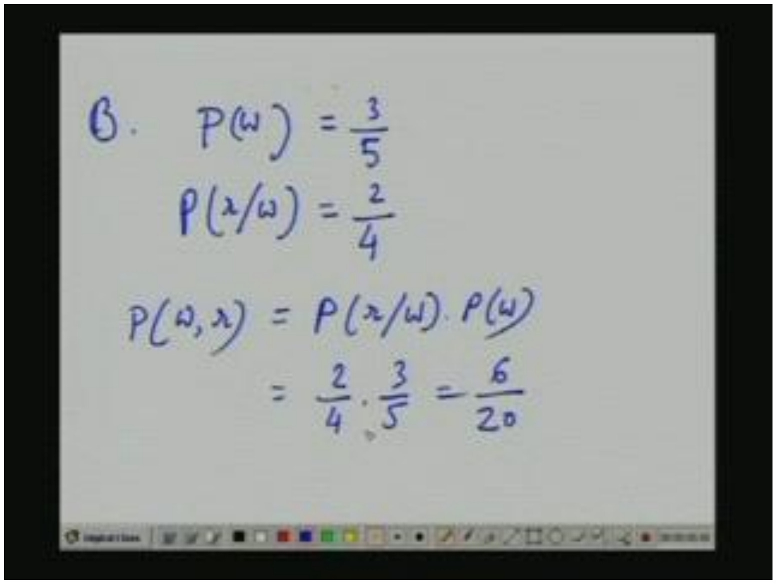


First that is solution, A approach A here first let us find out how many pairs we have that is pairs likes w 1 w 2 is 1 pair w 2 w 3 1 pair w 1 w 3 1 pair, and then w 1 r 1 one pair and like that. So, total number of pairs is how many five c 2 there are five elements we are just combining taking 2 at a time. So, that is equal to 5 into 4 20 out of this twenty events each event is a possible pair.

So, there are 20 events. Out of these 20 events let us find out on how many occasions you can have first ball being white and second ball being red; that means, you can have combinations like either this w 1 r 1 or w 1 r 2 or again w 1 r 3. Similarly, for w 2 r 1 w 2 r 2 w 2 r 3 and w 3 r 1 w 3 r 2 w 3 r 3. Total number of such combinations we do not have r 3 let me erase. So, w 1 either r 1 or r 2, then w 2 again r 1 or r 2 and w 3 again r 1 or r 2; obviously, we have 6 combinations. Any of these combinations would give us white ball first or red ball next.

So, probability of this happening is, then 6 by 20, because twenty is the total probability total number of ways total number of pairs out of which on 6 occasions you can get a pair like this so, probability 6 by 20.

(Refer Slide Time: 35:14)

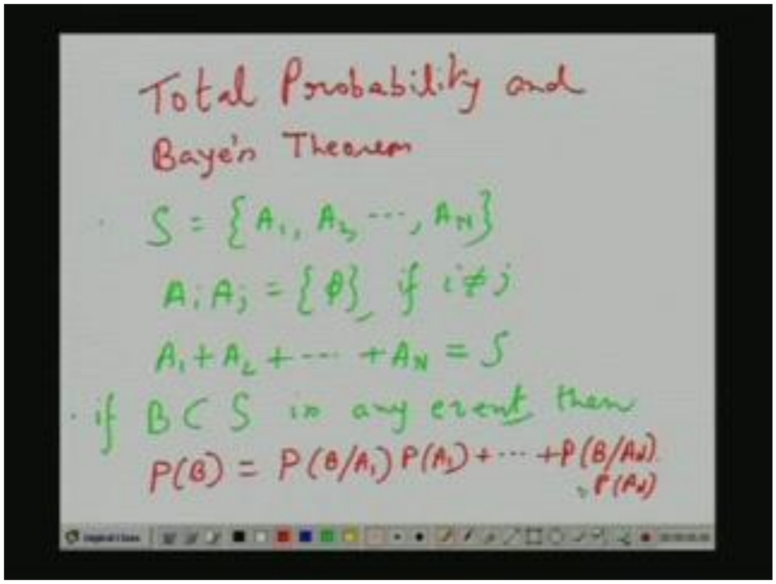


But, using conditional probability how to approach this the solution b. This we use conditional probability first we find out P W by S. S means here we are just taking 1 ball first we are finding out on how much this is not P W by S (( )) for it. Suppose, you first find out that when you take out the first ball what is the probability of that ball turning out to be white 1. That is we find out P w P w this will be how much if there are 5 balls out of which and you and I am picking up any 1 at random. So, on out of each 5 occasions only on 3 occasions a white ball can come up.

So, this is 3 by 5, and then after I have taken out 1 ball which is a white 1 what is the probability that next ball is red, that is r w that is giving the first ball is white. What is the probability that the next ball is red? Now, I have taken out 1 ball at random it could have been red could have been white. But suppose of course, the probability of it turning out to be white is 3 by 5, but suppose it is now known to be white. That means, how many balls I am now left with I am left with four balls.

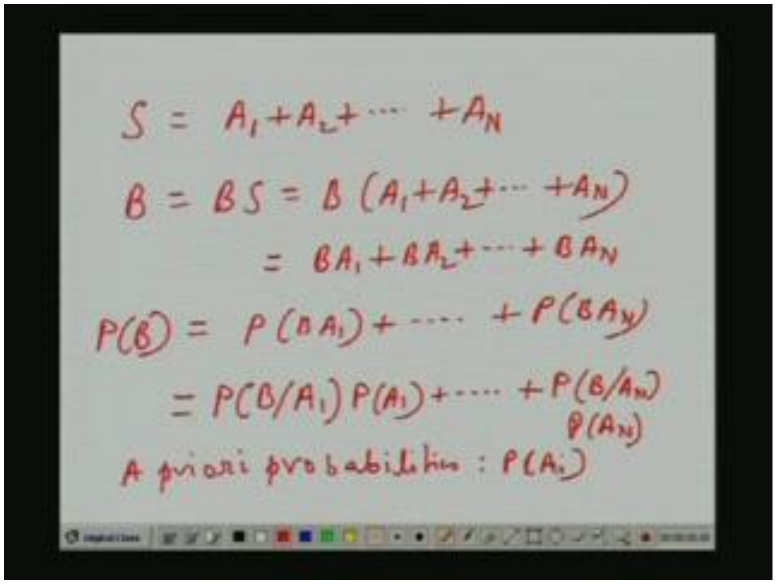
Out of these 4 balls that is subject to the condition that the first 1 is white, it coherently means that, that I am now left with 4 balls out of which only 2 are white not 3 and 2 are red as usual. Then what is the probability of I mean next ball being red? That will be 2 by 4. In that case, what is the probability that, first we have white followed by red? That will be since conditional probability is joint probability divided by P w, the other way I can write this joint probability is nothing but which is 6 by 20 that is the result we got earlier.

(Refer Slide Time: 38:43)



We now come to something called total probability. Suppose, S that is the certain event that is the total set it has a partition in this form A 1 A 2 up to say A N means A 1 A 2 up to AN. They are subsets of S which are mutually exclusive and their union equals S that is, Ai Aj which is phi if I not equal to j and A 1. If this is equal to S and that is a partition. That is each is an event and they are mutually exclusive events, but union of this events is a total event I mean certain event of the total set S. Then question is if B a subset of S is any event, then PB there is the probability of this event B can be written and will prove it can be written as we will prove it is called total probability theorem it is very simple.

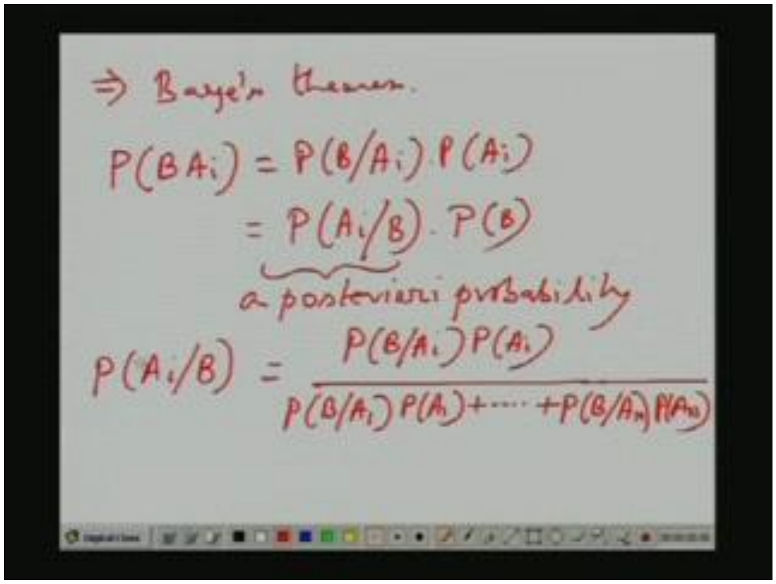
(Refer Slide Time: 41:42)



So, you know that S is given by this, this is given by this; that means, you can write B as nothing, but B intersection S first after all, B is a member of S these are s. And using the distributive property of the set operations we can write it simply as BA 1 plus BA 2 plus dot dot dot B AN. Remember since A 1 A 2 up to AN they are mutually exclusive. So, So, are these sets after all this is nothing, but inter, but an intersection between B and A 1 which is contend in A 1. Then again B and A 2 this intersection is also B bar of A 2 and similar this is the member of AN. And since they are mutually exclusive this sets also are mutually exclusive.

Which means PB is nothing, but P of this and since they are mutually exclusive you can write like. And, then using, then using the conditional probability theorem I would say or definition you can write each of them like this. That means, if we know the probabilities of these events A 1 to AN and statistics you know these are called a priori probabilities a priori probability. And if you know this conditional probability that if A 1 has occur what is the probability of B occurring. And likewise, then using this summation we can compute the total probability of occurring of B. This is the total probability theorem. I mean we extend this little bit and we get for what is called what is an important theorem is called Baye's theorem.

(Refer Slide Time: 45:15)

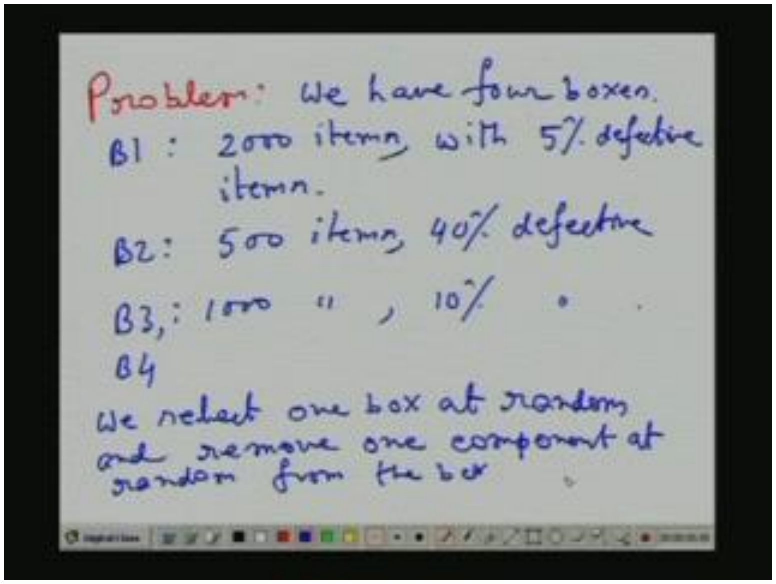


This takes us to Baye's theorem for that we see 1 thing that we can write say BAi in 2 ways, either you can write as before that is conditional probability of B subject to the fact that Ai does occur multiplied the positive probability of Ai occurring. Alternatively we can write like this, this is called the posteriori probability. That if B has occurred if that is known, then what is the probability that Ai had occurred that could have triggered B.

So, here we are looking I mean we are we have already obtained the result B and from there we are looking back and try to try to speculate. What is the chance that Ai must have occur? So, that B has resulted now. So, this is called a posteriori probability. Now, we can write this a posteriori probability in this way. Then that is I take this below this divided by PB and PB is nothing, but what you obtained last time for the total probability PB that is let me say it is.

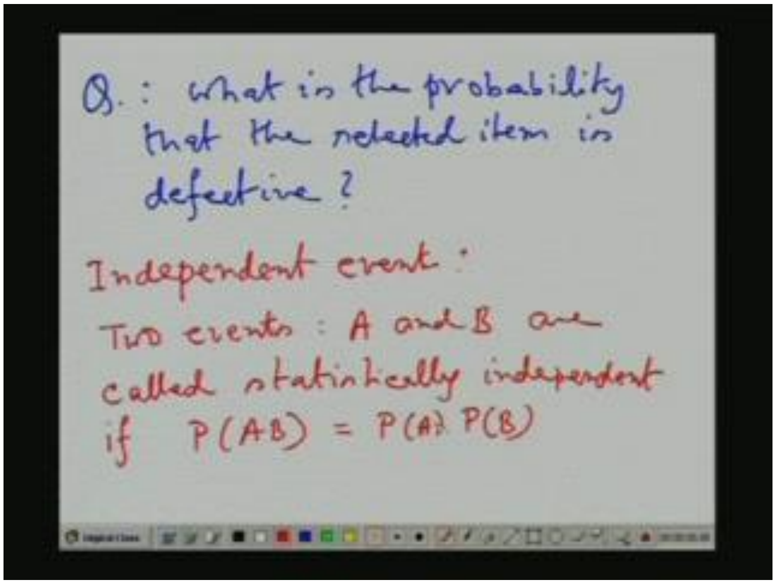
This is called Baye's theorem that is given the (( )) probabilities and the probability of the individual events. We can find out the corresponding a posteriori probability is called Baye's theorem. I will give the problem based on these, but that I will not solve now I will solve in the next class.

(Refer Slide Time: 48:12)



We have we have 4 boxes B 1 it contains 2000 items with five percent defective items. B 2 contains 500 items and 40 percent defective. B 3 it contains thousand items B 3 B 4 both B 3 B 4 both thousand items and 10 percent defective. We select 1 box at random. So, whether we pick up B 1 or B 2 or B 3 or B 4 probability of that box probability of a particular box being selected B 1 B 4 there is obvious.

(Refer Slide Time: 50:57)



To select 1 box at random and the remove 1 component at random from the box question, what is the probability that the selected item is defective. I would suggest that you are try this and I will take up this problem in the next class. I now continue to move over to a very important topic, I will just touch up on giving the paucity of time, now is called independent events. 2 events A and B are called we call them statistically independent.

If the probability of this intersection, which you called joint probability, because in this case both A and B have to occur jointly. So, joint probability is same as product of the individual probability. This is a very important concept in the next class I will take it up. So, just to sum up what we did today we start we continued with this motion of field of sets. And, then Borel field and just explain why we need those concepts basically for mathematical correctness.

For those who are more interested in the axioms of probability and you know the probability theory they can pick up from their. Then we move toward to a very useful concept called conditional probability we examine its properties. Then we gives some examples to I mean. So, for the lights on it we give a frequency based interpretation of this. Then we come came to some important concept called total probability theorem and Baye's Baye's theorem which gives you which helps you to obtain a posterior probabilities from giving upwardly probabilities. I give some problems to solve in the at home which you can take up in the next class. And, then I just introduced you to the concept of statistical independence which I will try to explain in the next class. So, that is all for today's class.

Thank you.

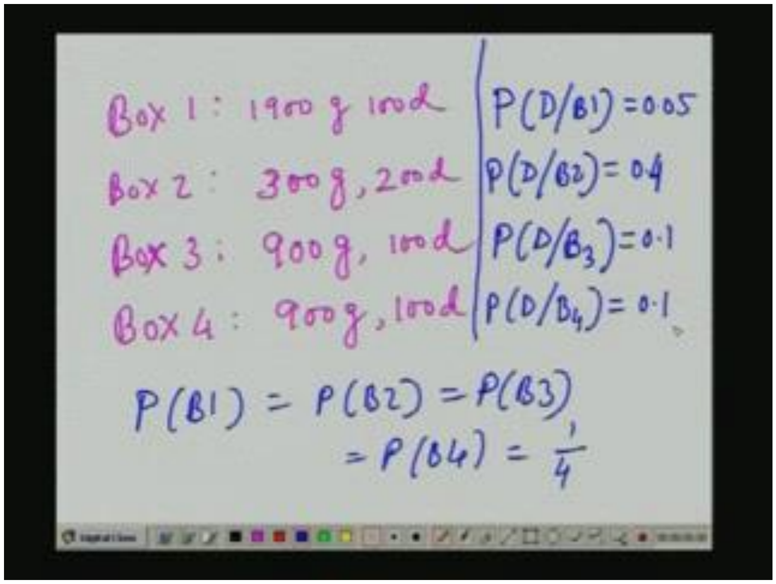
Prof. M.Chakraborty

Lecture – 4

Introduction to Random Variables

So, in the previous class I give a small problem for you to workout. I am not sure whether you have done it or not, but I would like to spend sometime now to solve that problem for your safe. Let me quickly repeat the problem that I gave you said that there is box there are four boxes.

(Refer Slide Time: 54:50)

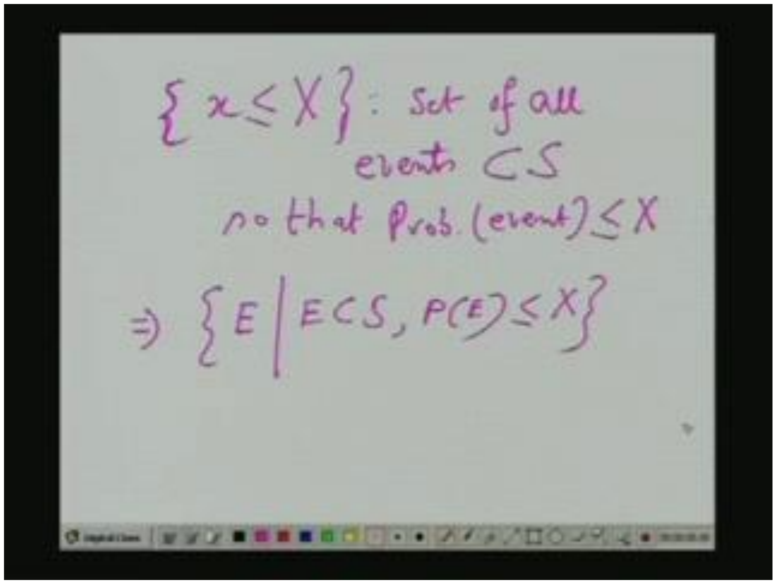


Box 1: it consists of 2000 item out of which 1900 items were good and 100 were defective; box 2: it had 300 good items 200 defective items; box 3: it had a 900 good items 100 defective items. And last box, box four: it also had 900 good items 100 defective items. Question that was ask was that I mean of we select any box in particular box at random first, and then from that box we remove just only 1 element.

Then the first question is, what is the probability that, the particular element which is being removed is a defective 1? So, to answer that first we see that each of the boxes since the boxes they are chosen purely at random and there is no particular preference for a special box, then box 1, box 2, box 3 or box 4. The probability that I choose box 1; so, B 1 or B 2 or B 3 or B 4 that is 1 by 4, because it is uniform.

That means P B 1 is same as P B 2 is same as P B 3 is same as P B 4 is equal to 1 by 4 the uniform. This is very safe assumption, because there is no preference for a particular box. And, then right across I can write that suppose I choose box 1. Then what is the probability that the particular item that is removed from that box is defective 1.

(Refer Slide Time: 57:25)



That is P D by less than equal to x. Do you understand the meaning of this? That is when now we are considering all possible events of x for which I will I show that for each of the event the corresponding probability see we can for events only we have we can define probability. So, now considering those events for which the corresponding probability is having a value less than or equal to X. That said in short we have compacted version is denoted like this.

So, whenever you say random variable is taking values less than equal to X means I am basically considering a set of events whose probability I mean maybe. Maybe since time is up today, I will consider this in the next class in detail basically you know just to tell you nutshell in the nutshell. I am basically trying to get into various part probability and distribution probability density functions.

So, in today's class we started to do that example that we give in the last class I work it out. Then we discussed at length a concept of statistical independence and then again we highlighted that through an example. And then I want do the discussion of random variables. I was just getting into the mathematical definition of a random variable actually certain conditions has to be satisfied, this I will take up in the next class.

Thank you very much.