Probability and statistics

Prof. Dr. Somesh Kumar
Department of Mathematics
Indian Institute of Technology, Kharagpur

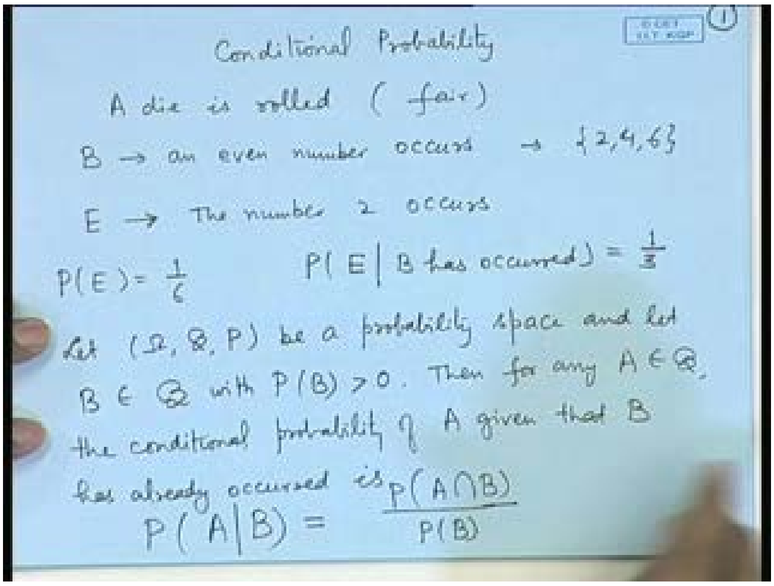
Module No. #01

Lecture No. #05

Laws of Probability-II

Today, we are going to introduce the concept of conditional probability. So, sometimes after conducting a random experiment or after observing certain random experiment, we have some partial information about the outcome of the experiment. Now, in the light of that partial information about the random experiment or about the outcome, if we want to find out the probability of certain event, then the probability gets modified.

(Refer Slide Time: 00:49)

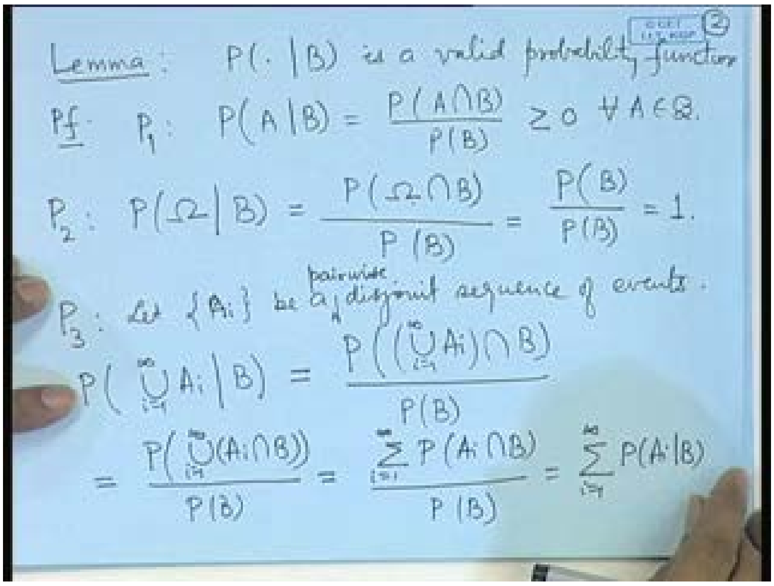


Let us take a very simple example, suppose we say, a die is rolled, suppose it is a fair die and let us say B is the event that an even number occurs and if I say E is the event that the number 2 occurs. Now, if I want to find out probability of E it is equal to 1 by 6. However, if I know that an even number has occurred, then probability of E given that B has occurred is 1 by 3 because the B has elements 2, 4 and 6,that means my sample space has been restricted to consist of three elements 2, 4 and 6. Now, out of this, if we say 2 occur, then it is one of the three possibilities and therefore, the probability of that is equal to 1 by 3.

So, you can see that the effect of the partial information makes us to make, helps us in making better probability statements, so let us define formally the conditional probability. So, if I say that omega, B, P is probability space and let B be an event with positive probability. Then for any event A, the conditional probability of A given that B has already occurred, it is defined by probability of A given B. So, this is the notation, probability of A given B that is defined as probability of A intersection B divided by probability of B.

Now, given any event B if we are defining the probability of A, then first of all we should check whether it is a proper probability function satisfying the axioms of the probability,let us check that.

(Refer Slide Time: 04:05)



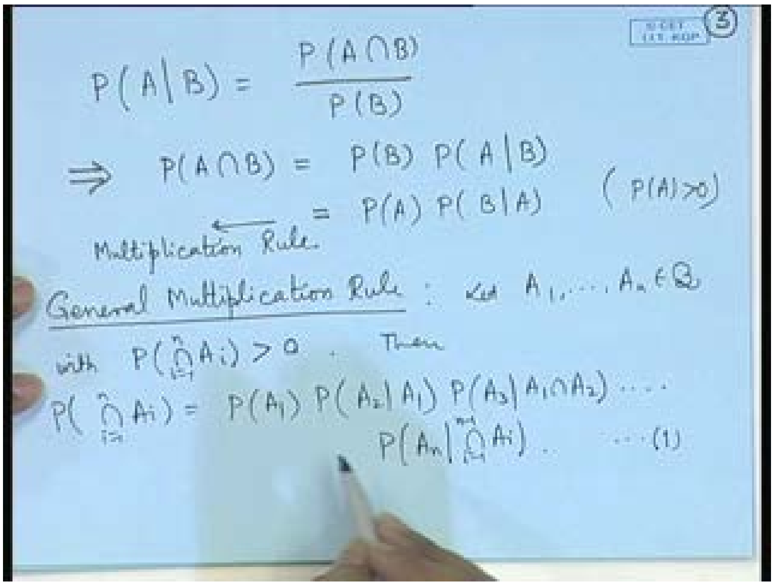
So, we state it as a result that P dot B is a valid probability function. Meaning thereby that it will satisfy the three axioms; that is probability of A given B that is equal to probability of A intersection B divided by probability of B. Now, according to the assumptions of the probability, probability of A intersection B is always greater than or equal to 0, probability of B is strictly positive, therefore this is always greater than or equal to 0 for all events B,so the first axiom is satisfied.

Let us look at the second axiom, probability of omega given B, this is equal to probability of omega intersection B divided by probability of B. Now, note that B is a subset of omega, therefore the numerator will be probability of B divided by probability of B and therefore, it is equal to 1. The third axiom is the countable additive axiom,so let us consider A i be a disjoint sequence of events.

Let us consider probability of union A i given B. Now, by the definition of the conditional probability, this will be equal to probability of union A i intersection B divided by probability of B. Now, in the numerator I can apply the distributive property of the unions and intersection,so this becomes probability of union of A i intersection B divided by probability of B.

Now, here note that A i is where disjoint, basically we can consider them to be pair wise disjoint. So, if they are disjoint, then A i intersection B's will also be pairwise disjoint and therefore, by the axioms of the probability, by the third axiom, I will have probability of union A i intersection B is equal to sigma i is equal to 1 to infinity, probability of A i intersection B divided by probability of B, where each of the term then becomes probability of A i given B. So, you can note here that probability of the union is equal to sum of the probabilities, which is the countable additive axiom. Therefore, the conditional probability function is a well defined probability function.

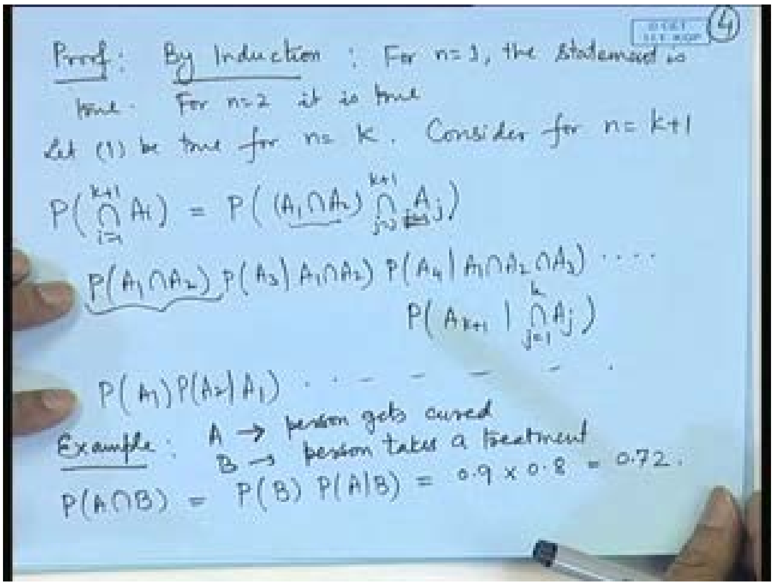
(Refer Slide Time: 07:25)



Now, let us look at some direct consequences of this definition of the conditional probability. Let us consider, say the statement probability of A given B is equal to probability of A intersection Bdivided by probability of B. Now, this statement we can write as, probability of A intersection B is equal to probability of B into probability of A given B. So, an interpretation of this statement is that, the probability of simultaneous occurrence of A and B is equal to the conditional probability of A given B multiplied by the probability of B alone happening. If we interchange the roles of A and B, then we can also write it as probability of A into probability of B given A, of course provided we consider that, probability of A is positive. Now, this representation of giving the expression for the simultaneous occurrence probability in terms of a product, where in the product one term is a conditional probability and another term is a marginal probability is known as multiplication rule,so this is called multiplication rule.

So, the general multiplication rule is, let us consider events A 1, A 2, A n be events with now, in order to define the conditional probabilities, the probabilities must be positive. So, we can take the smallest event which may occur in the definition and we can consider the probability of that to be positive. And then, the probability of intersection A i is equal to 1 to n can be expressed as probability of A 1 into probability of A 2 given A 1 into probability of A 3 given A 1 intersection A 2 and so on. Probability of A n given intersection A i, i is equal to 1 to n minus 1,let me give this is statement number 1.

(Refer Slide Time: 10:18)



So, in order to prove this one we can follow induction. So, by induction if we want to see for n is equal to1, the statement will reduce to simply probability of A 1 is equal to probability of A 1 which is trivially true. So, for n is equal to 1, the statement is true, for n is equal to 2, it is true, as we mentioned by the definition of the conditional probability that probability of A 1 intersection A 2 will become equal to probability of A 1 into probability of A 2 given A 1. Now, the statement for n is equal to 2 will be used for extending from K to K plus 1. So, let 1 be true for n is equal to kand consider for n is equal to k plus 1. So, probability of intersection A i, i is equal to 1 to k plus 1, now this one we can express as probability of A 1 intersection A 2 intersection A j, say j is equal to 3 to j is equal to 3 to k plus 1. So, here if you see in the intersection, these are k minus 1 sets and this is 1 set. So, this is total k sets and we have made the assumption that the statement is true for k.

Therefore, we can write it as probability of A 1 intersection A 2 into probability of A 3 given A 1 intersection A 2 into probability of A 4, given A 1 intersection A 2 intersection A 3 and so on, to probability A k plus 1 given intersection A j, j is equal to 1 to k. And now the first statement we can express as probability of A 1 into probability of A 2 given A 1 and all the remaining terms will be there.

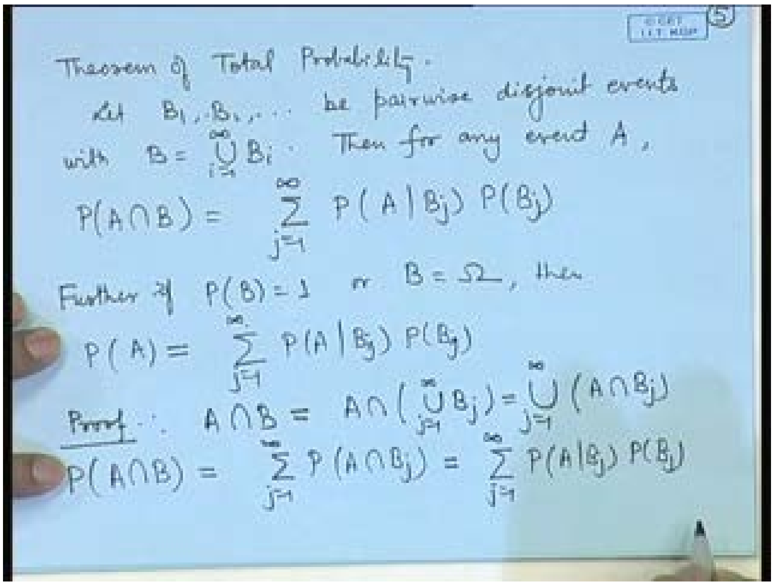
Hence, the statement 1 is true for n is equal to k plus 1 also and therefore, by the principle of mathematical induction the general multiplication rule is true. The significance of this result lies in that many times we are not aware of the probability of the simultaneous occurrence. It may be somewhat difficult because what may happen that the events are conditional that is after occurrence of one something may happen for example, if a person is sick with a certain disease and he takes a treatment what is the probability that he will be cured?

Now, the probability of simultaneous occurrence of this event may not be so simple. However, if we know that the percentage of people who get cured by taking that particular treatment and how many persons eventually get recovered then, this probability can be calculated that means, if I say, A is the probability of getting cured and B is suppose A is the event of getting cured and B is the event of taking a particular treatment.

Now, if we look at the percentage of people actually getting cured and the percentage of people who take the treatments and they are cured or in the reverse way, if we look at the percentage of people who take the treatment and the out of the percentage of the people who take the treatment how many get cured, then this probability can be evaluated.

Let me write it as an example; let A be the event person gets cured; B is person takes a treatment. So, now if I look at, suppose probability of A intersection B is to be determine then we can write it as probability of B into probability of A given B. Suppose, we know that how many people take that particular treatment, suppose 90 percent of the people take the treatment and of the number of people who take the treatment, suppose 80 percent get recovered, then this probability turns out to be 0.72, which may be might have been difficult otherwise. Now, this representation that a certain event can be considered as a consequence of certain other event or happening after a certain event, this leads to some interesting possibilities and one of the major things is the so called theorem of total probability.

(Refer Slide Time: 15:55)



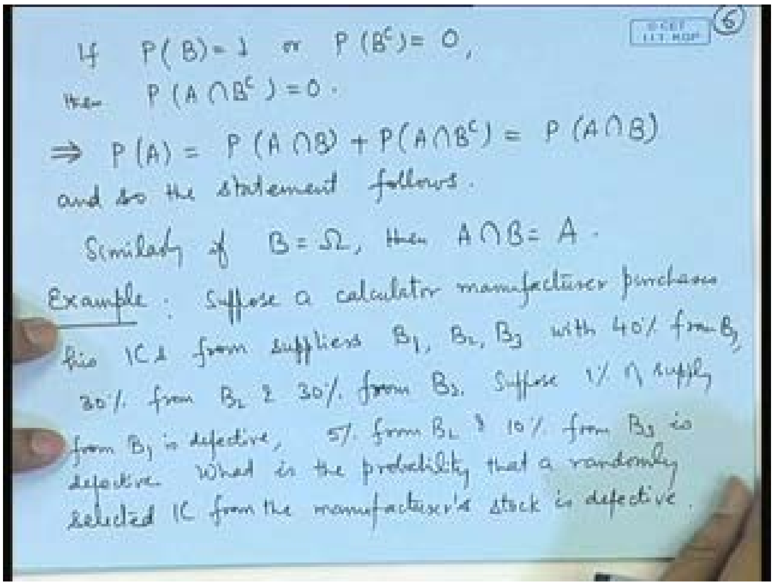
So, let B 1, B 2 etcetera be pairwise disjoint; that means, mutually exclusive events with B is equal to say union B i; i is equal to 1 to infinity here it could be a finite union also it will not make any difference to the actual statement of the theorem. Then, for any event A, probability ofA intersection B can be decomposed as probability of A given B j into probability of B j; j is equal to 1 to infinity.

Further, if probability of B is equal to 1 or say B is equal to omega, then probability of A is equal to sigma probability of A given B j into probability of B j. So, you can see here, if we treat the event A as a consequence of either of B 1, B 2, etcetera then, the eventual probability that A occurs can be represented in terms of that it has been caused by B 1, it has been caused by B 2 etcetera and the corresponding probability of those causes themselves. We will look at some example here and but firstly, let me give the proof of this.

So, consider A intersection B, nowA intersection B can be written as A intersection and the representation for B is union of B j; j is equal to 1 to infinity. By applying the distributive property of the unions and intersections, we can write it as union of A intersection Bj, j is equal to 1 to infinity.

Now, we notice here that, B j's were pair wise disjoint sets, therefore A intersection Bj's will also be pair wise disjoint and therefore, the countable additively axiom of the probability applies here. And you can write probability of A intersection B as summation of the probabilities of A intersection B j. Now, by applying the multiplication rule on this simultaneous probability, this becomes probability of A given B j into probability of B j, which is actually the statement of the theorem of total probability.

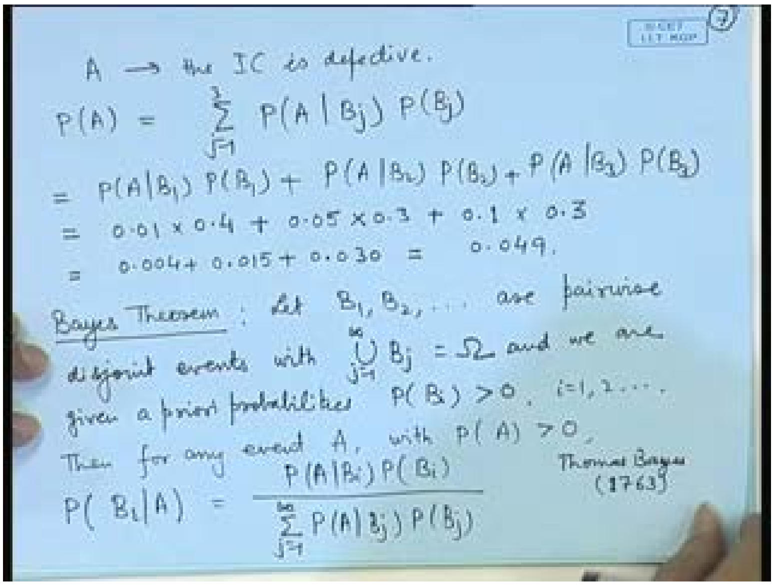
(Refer Slide Time: 19:31)



Now, we look at if probability of B is 1 or B is omega case also here. So, if probability of if we take probability of B is equal to 1 or probability of B complement is equal to 0, then probability of A intersection B complement is 0 and therefore, probability of A can be written as probability of A intersection B plus probability of A intersection B complement as probability ofA intersection B and so the statement follows.

Similarly, if we are saying that B is omega, thenA intersection B is simply A and the statement is true. So, let me give an example here, suppose a calculator manufacturer purchases his ICs from suppliers say B 1, B 2, B 3 with 40 percent from B 1, say 30 percent from B 2, and 30 percent from B 3. Suppose, 1 percent of supply from B 1 is defective, 5 percent from B 2 and 10 percent from B 3 is defective, what is the probability that a randomly selected IC from the manufacturer’s say, stock is defective.

(Refer Slide Time: 22:35)



Now, here if we consider the event say A, that the IC is defective, then probability of A can be represented as sigma probability of A given B j into probability of B j, j is equal to 1 to 3. That means, since the IC could have come from either first manufacturer or from second manufacturer, second supplier or from the third supplier and therefore, the consequence that it is defective is also the probabilities dependent upon who supplied that. So, this becomes probability of A given B 1 into probability of B 1 plus probability of A given B 2 into probability of B 2 plus probability of A given B 3 into probability of B 3.

So, now probability that it is defective given that it came from the first one is only 0.01 and the probability that it was supplied by the first supplier are 0.4. The probability of being defective from the second supplier is 0.05 and the probability that supplier two supplied is 0.3, probability that it was it is defective provided it was supplied from a supplier 3 is 0.1, and the probability of getting the supply from the third supplier is 0.3. So, eventually this turns out to be 0.004 plus 0.015 plus 0.03 which is equal to 0.049. So, the eventual probability of the IC to be defective is point approximately 0.05. Although, individually if we see from the first supplier only 1 percent, from second supplier 5 percent and from the third supplier it is 10 percent but since, the procurement is mixed with the different percentages getting supplied from different suppliers, therefore the actual probability turns out to be 0.049.

Further, you can say consequence of this cause effect relationship is that sometimes we know the final outcome, now what is the probability that this outcome was caused by something? Now, this representation or you can say this way of looking at the probability is called posterior probability, because here we can consider the events B 1, B 2, B 3 that the supplies came from first supplier, second supplier, third supplier as prior events or something which was happening before. Now, being defective or non defective is an event which is occurring after that because it is after procurement.

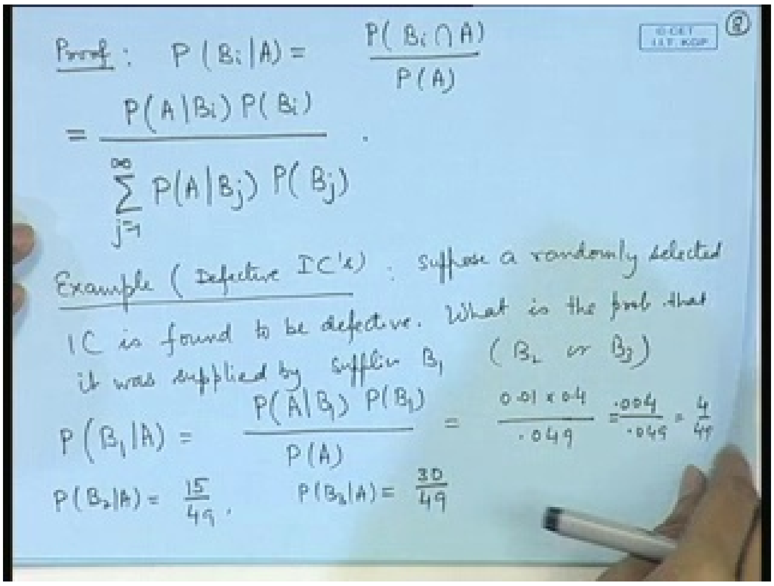
Now, suppose it says that, supposewe take a suppose the I C is supplied and somebody takes it and it is found to be defective then what is the probability that it came from first supplier or it came from a second supplier or it came from the third supplier. This way of looking at this probability is called posterior events or the probabilities of the posterior events.

So, the result for this is called Bayes theorem; let me state the result first, suppose B 1, B 2, etcetera are pair wise disjoint events with union of B j is equal to say omega, that means they are exhaustive also. And we are given a priori probabilities; probability of Bi as positive then for any event Aof course, since conditional probabilities are involved we must say probability of A is positive, then we define probability of B i given A that is equal to probability of A given B i into probability of B i divided by sigma probability of A given B j into probability of B j, j is equal to 1 to infinity.

So, let us look at the representation of this, this means that we know the final outcome A, what is the probability that it was caused by the i th cause? So, this is the posterior probability and this Bayes theorem actually helps us to calculate this, it is calculated in terms of the prior probabilities and the consequent probabilities which are called like the cause effect relationship that what is the probability that the event was caused by the i th cause?

So, the eventual posterior probability can be represented in terms of this, this reverse relationship it is due to Bayes reverent Thomas Bayes and it was published in his (())published in 1763 posthumously.

(Refer Slide Time: 29:32)



The proof of this statement is Let us looking at the proof of this statement. So, we will just apply the definition of the conditional probability; probability of B i given A is equal to probability of B i intersection A divided by probability of A by the definition of the conditional probability. Now, what we can do is that, in the numerator we can apply the multiplication rule in the reverse way. So, this becomes probability of A given B i divided into probability of B i. In the denominator you have probability of A, applying the theorem of total probability we can write it as sigma probability of A given B j into probability of B j; j is equal to 1 to infinity which proves the statement of the Bayes theorem.

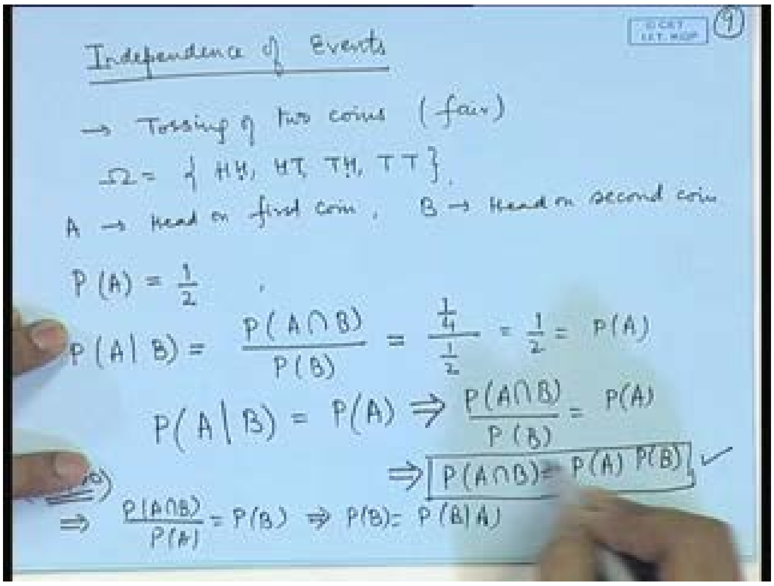
One thing which one can notice here is that, the proofs of the statements of the theorems are quite simple. This is happening because we have given an axiomatic structure to the to the probability and therefore, all the statements are simply consequences of certain relationships in the set theory. When the axiomatic definition was not known, all of these statements had complicated proofs because, if you use say relative frequency definition, then you have to look at the various relationships between the limits, if you use the classical definition then you have to represent all the events in terms of favorable number of cases and the total number of cases etcetera. So, all of these statements used to be quite complicated, the major contribution of the axiomatic foundation is that it made all the things extremely straight forward and also valid, because everybody can verify that the statements are true.

As a Let us look at the previous example of the IC defective problem; suppose a randomly selected IC is found to be defective. What is the probability that it was supplied by say, supplier B 1 or B 2 or B 3? Let us look at the first one; in the first one we are interested in finding out the probability of B 1 given A. So, this is by the definition of Bayes theorem, it will become probability of A given B 1 into probability of B 1 divided by probability of A. We already evaluated the probability of A as 0.049. So, it becomes simply probability of A given B 1 that is 0.01 into 0.04; that is, a probability of B 1 divided by 0.049 which is equal to 4 by. So, it is 0.004 by 0.049 which we can write as 4 by 49.

If you look at probability of say B 2 given A, then in the same logic this will become equal to 15 by 49 if we calculate probability of say B 3 given A, then by the same logic it will become 30 by 49, you can see here the difference in the posterior probabilities. Although only 30 percent of the product comes from manufacturer third but, if the finally selected IC is found to be defective then it is most likely to have come from the third supplier, the reason being that in the hypothesis of the problem it has been assumed that more number of product from the third supplier are defective, which is actually 10 percent whereas, from the first supplier it is only 1 percent. So, although more supply comes from the first manufacturer; however, the probability that it came from him provided it is defective then the probability is much smaller. So, this is an important you can say consequence of the Bayes theorem that the posterior probabilities are quite different from the prior probabilities.

Here, you can see probability of B 1 is 0.4 whereas, probability of B 1 given A is only 4 by 49 which is much smaller, probability of B 2 is 0.3 whereas, probability of B 2 given A is 15 by 49, probability of B 3 is 0.3 whereas, probability of B 3 given becomes 30 by 49, which is much higher. So, the probabilities get revised in the light of the knowledge about the outcome, sothis is a good way of looking at the Bayes theorem as a cause and effect relationship.

(Refer Slide Time: 35:48)



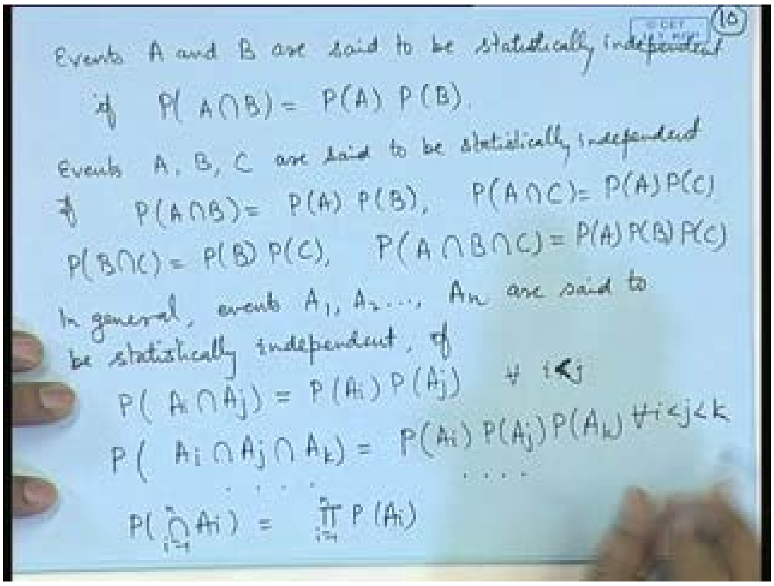
Now, we have another concept called independence of events. Let us consider suppose tossing of 2 coins, suppose they are fair coins, so my sample space here is H H, H T, T H and T T. Suppose, I consider my event A as head on first coin and B is a event head on second coin, if I look at the probability of say A, head on the first coin is in H H and H T, so it is half. Suppose, if I consider probability of A given B then it is equal to probability of A intersection B divided by probability of B. Now, probability of B is simply equal to half because we are saying head on the second coin, head on the second coin is occurring in this and this. So, it is half and probability of A intersection Bmeans head on the first coin and head on the second coin which is happening only in one case, sothe probability is one by 4. So, this is turning out to be half which is same as probability of A.

Now, this is different from the statements which I gave in the beginning of the conditional probability definition that if I consider probability of E and probability of E given B; these two were different, it means that happening of B definitely affects the happening of E whereas in this one, probability of A and probability of A given B is the same; that means, happening of B does not affect happening of A.This means the happening of A is independent of the happening of B; this is the concept of independence of the events.

So, if we proceed from the definition of the conditional probability we must define event A to be independent of event B if we have probability of A given B is equal to probability of A. Each translates into probability of A intersection B divided by probability of B is equal to probability of A, which is translating to probability of A intersection B is equal to probability of A into probability of B.

Now, if we also make assumption that probability of A is positive, then this will translate to probability of A intersection B divided by probability of A is equal to probability of B provided we have assumed that probability of A is positive. So, this will again imply that probability of B is equal to probability of B given A; that means, B is independent of happening of A. So, provided I consider that probabilities of A and probabilities of B are positive, the concept of independence of A with B or B with A is symmetric in nature. Therefore, to avoid unnecessary complications we can consider this as the definition of probability as the independence of the events A and B.

(Refer Slide Time: 39:58)



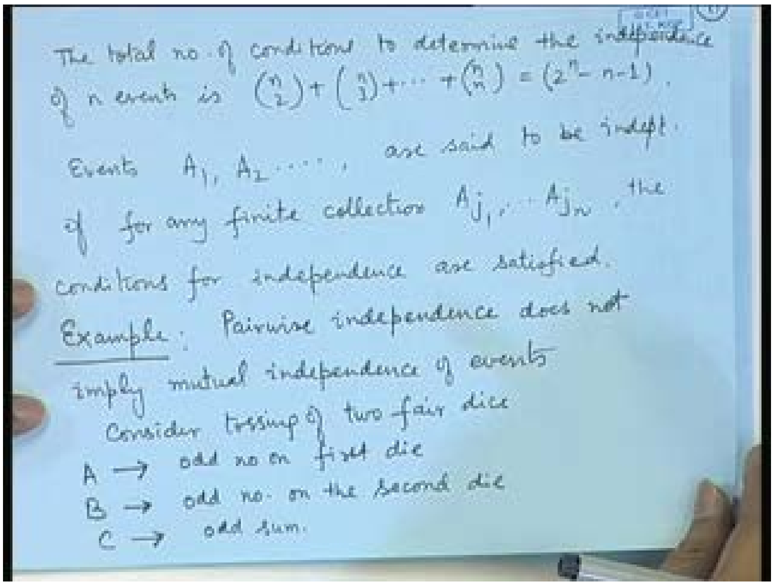
So, we formally define events A and B are said to be statistically independent if probability of A intersection B is equal to probability of A into probability of B. So, you can see that this definition is symmetric in the independence of A with B or B with A and moreover we do not have to write down the condition that probability of A is positive or probability of B is positive, because this definition does not involve the division here.

Naturally a question arises that if I have more than two events then what is the meaning of independence? Let us consider say events A, B and C, then events A, B and C are said to be statistically independent if probability of A intersection B is equal to probability of A into probability of B, probability of A intersection C is equal to probability of A into probability of C, probability of B intersection C is equal to probability of B into probability of C. Now, one may feel that these three conditions are enough for defining the independence of events say A, B and C but, this is only pair wise independence of A and B, A, B and C. We need one more condition namely probability of A intersection B intersection C is equal to probability of A into probability of B into probability of C.

Now, this is required because it may happen that A is independent of B or A is independent of C, but then simultaneous occurrence of B and C may have something to do with A or simultaneous occurrence of A and B may have something to do with C. Therefore, in order to define the statistical independence of three events, we have four conditions namely we take two at a time and then all the three at a time.

In general, then if we have events say A 1, A 2, A n; then they are said to be statistically independent if probability of A i intersection A j is equal to probability of A i into probability of A j for all i not equal to j, you may write in one way only because there is no need to write for i less than j and then for j less than i. So, you may put i less than j then taking three at a time, probability of A i intersection A j intersection A k is equal to probability of A i into probability of A j into probability of A k for all i less than j less than k. We may also consider four at a time and so on.

(Refer Slide Time: 44:27)



Finally, probability of A i intersection i is equal to 1 to n is equal to product of the probabilities A i. The number of conditions, you can see this is n C 2, so how many total number of conditions are there? Total number of conditions will become The total number of conditions to determine the independence of n events is; So, let us look at how many conditions are there here we have n events and we are taking two at a time so, total number of conditions is n C 2. Here, we have 3 from n 3 at a time. So, the total number of conditions is n C 3 and finally, it is n C n or 1. So, the total number of conditions is eventually equal to n C 2 plus n C 3 plus n C n which we can write as 2 to the power n minus n minus 1.

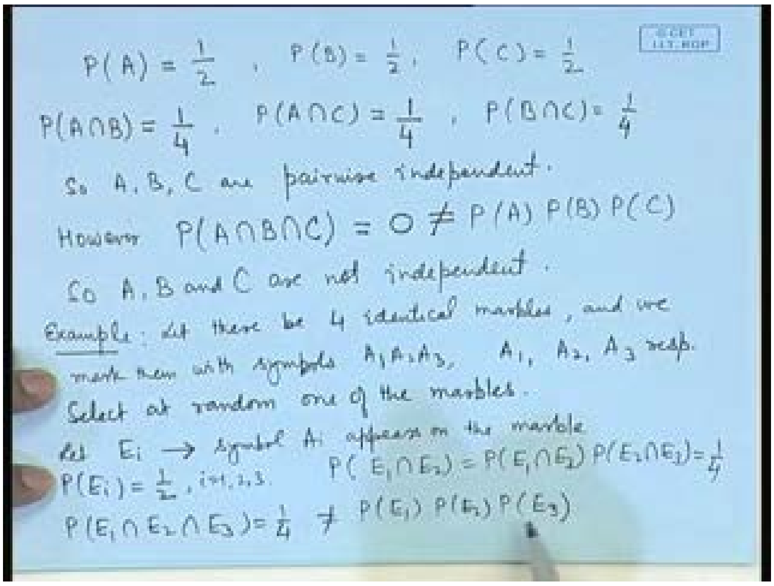
So, a total number of 2 to the power n minus n minus 1 condition are required to determine the statistical independence of n events. Our interest is also to look at infinite number of events and then talk about its independence. Then, logically it means that if we consider any finite subset or any finite collection of that infinite number of events, then that finite collection must satisfy the conditions for the independence.

So, we define events A 1, A 2 and so on; that means, the indexes are belonging to some set are said to be independent. If for any finite collection, let me call it say, A j 1, A j 2, A j n the conditions for independence are,here let me specify that the conditions which are all mentioned must be satisfied in order to have the actual independence of the events.

For example, if I take the case k is equal to 3, then for three eventsthe statistical independence is defined in terms of pair wise independence condition, there are three conditions and one in which all the three are taken. Now, one may say that if the pair wise independence conditions are satisfied then the last one is automatically satisfied, actually it is not true. The satisfying of the first three does not imply the satisfaction of the fourth one or vice versa. This may be satisfied but, one of these may not be satisfied.

Let me give examples of this; that is pair wise independence does not imply mutual independence of events. Consider say, tossing of two fair die, let me consider the event A as say odd number on first die B as the event odd number on the second die and C is the event odd sum, let us consider the probabilities of A, B and C.

(Refer Slide Time: 49:22)



So, if we look at probability of A, odd number on the first die, it means you have 1, 3 and 5 on the first, the total number of possibilities here is 36 and odd number on the first means 1 1, 1 2, 1 3 up to 1 6, 3 1, 3 2, 3 6, 5 1, 5 2, 5 6, so total number of 18 possibilities are there. So, probability of A becomes half, likewise probability of B is half and likewise probability of C; that is the odd sum is also half. If we consider the probability of A intersection B then we are saying odd number on the first and odd number on the second what are the possibilities? 1 1, 1 3, 1 5, 3 1, 3 3, 3 5, 5 1, 5 3, 5 5, there are 9 possibilities out of 36 possibilities, so this becomes 1 by 4.

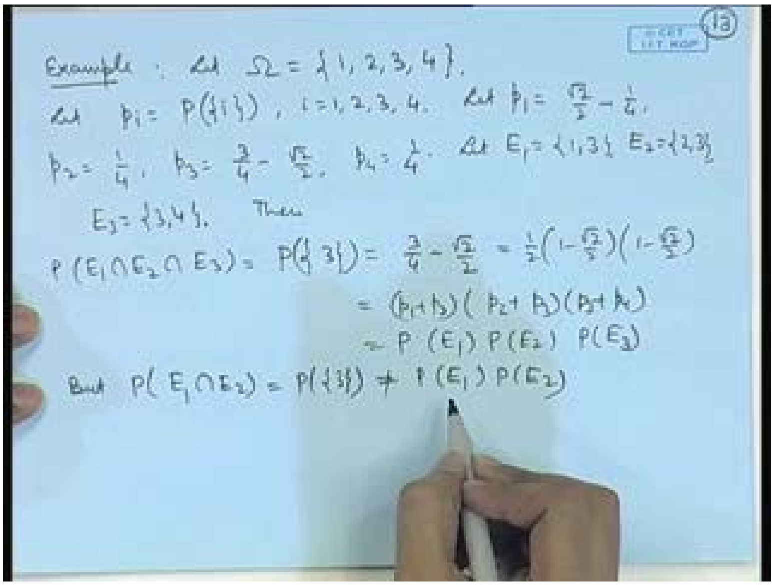
If we look at probability of A intersection C then we are saying that the first number is odd and the second and the sum is odd. It is possible 1 2,1 4, 1 6, 3 2, 3 4, 3 6, 5 2, 5 4 and 5 6. So, this becomes 9 by 36, again 1 by 4. Similarly, probability of B intersection C is 1 by 4. Naturally, you can see that the conditions for the pair wiseindependence are satisfied here. So, A, B, C are pair wise independent.

However, if we look at probability of A intersection B intersection C then if we have odd number on the first die, odd number on the second die, naturally the sum will be even and therefore, there is no case where A intersection Bintersection C is satisfied therefore, this probability is 0, naturally this is not equal to probability of A into probability of B into probability of C. So, A, B and C are not independent although they are pair wise independent.

Let me consider one example in which the condition for the simultaneous that is the third condition may be true but, one of the earlier conditions may not be true. Let us consider the set; let there be four identical marbles and we mark them and we mark them with symbols; that means, on the first one we write A 1, A 2, A 3 and A 1, A 2, A 3; that is, on the first marble we mark A 1, A 2, A 3 on the second one we put A 1, and the third one we put A 2, and the fourth one we put A 3. We draw one of the marble identically, select at random one of the marbles. Let E i denote the event that symbol A i appears on the marble.

Let us look at probability of E i, it will be half for all i's, if I look at probability of say E 1 intersection E 2, probability of E 1 intersection E 3, probability of E 2 intersection E 3;each of this is going to be 1 by 4 because only 1 possibility is there. And therefore, if I look at probability of E 1 intersection E 2 intersection E 3 and that is also 1 by 4 and it is not equal to probability of E 1 into probability of E 2 into probability of E 3. So, this is another example where the events are pair wise independent, but they are not independent.

(Refer Slide Time: 55:11)



Let me take one more example, consider say, omega is equal to 1, 2, 3, 4. Let P i be the probability assigned with the set consisting of I; i is equal to 1, 2, 3, 4. Let us define say P 1 is equal to root 2 by 2 minus 1 by 4, P 2 is equal to say 1 by 4, P 3 is equal to 3 by 4 minus root 2 by 2, P 4 is equal to 1 by 4. Let E 1 be equal to 1, 3; E 2 be equal to 2, 3; and E 3 is equal to 3, 4.

Then, if I look at probability of E 1 intersection E 2 intersection E 3 then it is simply the probability of 3 that is P 3, that is 3 by 4 minus root 2 by 2, which I can write as half into 1 minus root 2 by 2, 1 minus root 2 by 2, which can be written as P 1 plus P 3 into P 2 plus P 3 into P 3 plus P 4. So, this is equal to probability of E 1 into probability of E 2 into probability of E 3. So, the last condition for the independence of E 1, E 2, E 3 is satisfied. But, if I look at probability of E 1 intersection E 2 then that is also equal to 3 and therefore, it is not equal to probability of E 1 into probability of E 2, therefore the conditions for the independence are not satisfied here.

So, in effect it means that, if you want to check the independence of events, then we must check the probability of simultaneous occurrences; by taking all the combinations; by taking 2 at a time, 3 at a time and so on. Thank you.