Advanced Engineering Mathematics

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

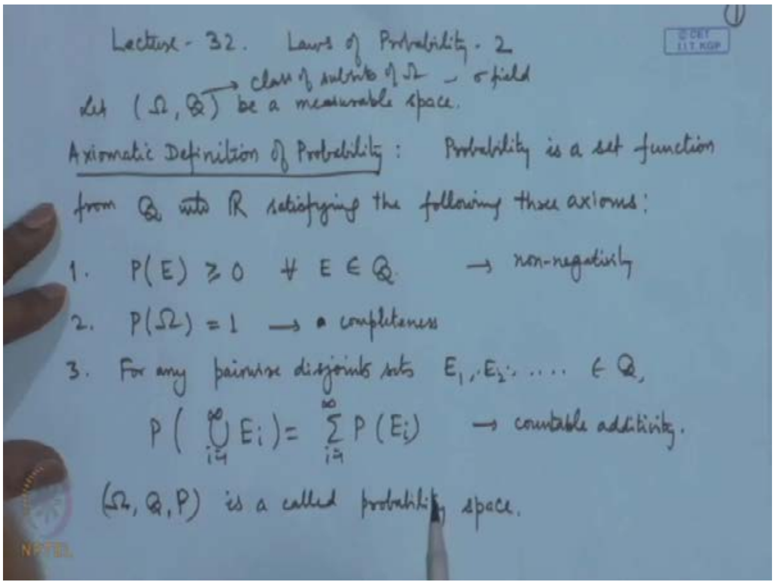
Lecture No. # 33

Laws of Probability- II

Friends, in the previous lecture, we have introduced the basic concepts of probability theory. Then, I gave two definitions of probability one I called as a classical definition, and another one is the relative frequency definition of the probability. And we have seen their relative merits and demerits. So, later on there was a need of a strong, a axiomatic framework under which the probability function will be valid, under the given conditions. So, this development was it is attributed to the Russian mathematician, A. N. Kolmogorov.

Now we have already seen that the basic unit of a random experiment is the sample space. Now, along with the sample space, we consider a subset of the class of, a class of subsets of the sample space. And we impose certain condition on that class, which we call that it is closed under a complementation, and it is also closed under the unions, countable unions. Therefore, this structure, which we call as a sigma field together with the sample space, we call it a measurable space.

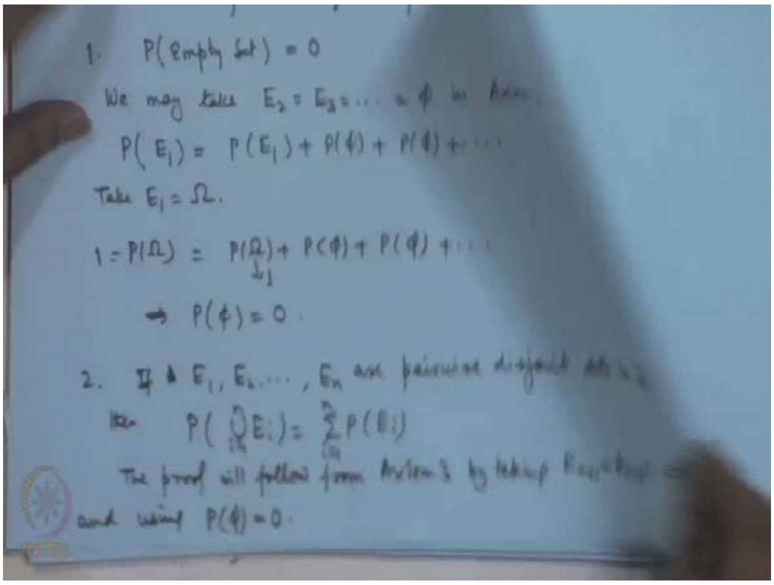
(Refer Slide Time: 01:36)



So, let us consider, let omega, B be a measurable space. So, we have the following axiomatic definition, axiomatic definition of probability. So, we say probability is a set function, because for every event, we are defining a probability; and now what is an event? Event is a subset here, subset of omega; now this B is the class of subsets of class of subsets of omega, and we are assuming that this is the sigma field, it is the sigma field. So, on each set, we are defining or we can say, we are associating a number, which is called P of that, P of a; earlier in classical definition, we have seen P a we are writing as m by n, then similarly in the second case, after relative frequency definition we are writing it as limit of P n by n as n tends to infinity. So, a certain number is associated with the set a.

So, it is a set function. So, probability is a set function, so we call it from B into R satisfying the following three axioms. So, first axiom is that probability of every event is non-negative; second is probability of the whole sample space is 1; third for any pair wise disjoint sets E 1, E 2 and so on belonging to B; probability of union E i is equal to some of the probabilities, i is equal to 1 to infinity. So, the first axiom is called the axiom of non-negativity; the second axiom is the theorem, the axiom of completeness or the axiom of totality, because we are fixing that the total probability cannot be more than 1; so it is equal to 1 here; and this is countable additives . So, now, omega B, P this is called a probability space that means, our measurable space omega, the sample space omega, a class of events, and with the probability function along with that, this is called probability space.

(Refer Slide Time: 05:21)

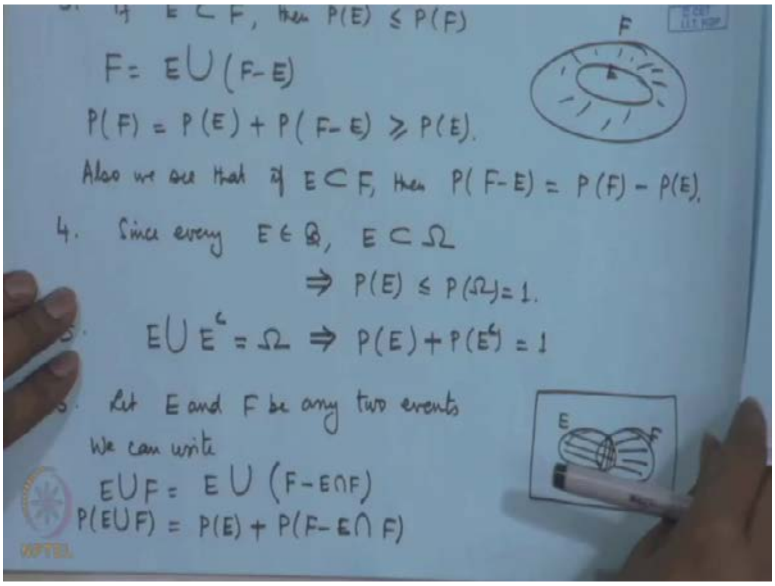


Now certain consequences of or ramification of this definitions are as follows. Certain ramification of the definition, the axiomatic definition; so, which will follow easily. So, for example, you may easily conclude that probability of a empty set is 0; how? See, we will take say in the axiom 3, here we are events E 1, E 2 etcetera, suppose I take one event only, and all other events I take to be phi; then what will happen? E 2, E 3 and so on, these to be phi in axiom 3. Then what we will get? Left hand side will become E 1, the right hand side will become E 1 plus probability of phi plus probability of phi and so on. Now, what we are getting is, we may put E 1 is equal to omega; if we put E 1 is equal to omega, then we will get P omega that is 1 is equal to P omega plus P phi plus P phi and so on; now this is also 1, so this cancels out, and you are getting P phi, which is a non-negative number, it must be equal to 0.

Similarly, you may conclude that if E 1, E 2, E n are pairwise disjoint sets in B, then probability of union E i, i is equal to 1 to n, that is finitely additive, P is finitely additive. The proof is again very simple, because here we can take E 1, E 2, E n, and the remaining E i is to be phi; and then since P phi is 0, this will lead to this. The proof will follow from axiom 3 by taking E n plus 1 is equal to E n plus 2 and so on equal to phi, and using P phi is equal to 0.

(No audio from 08:28 to 08:46)

(Refer Slide Time: 09:05)



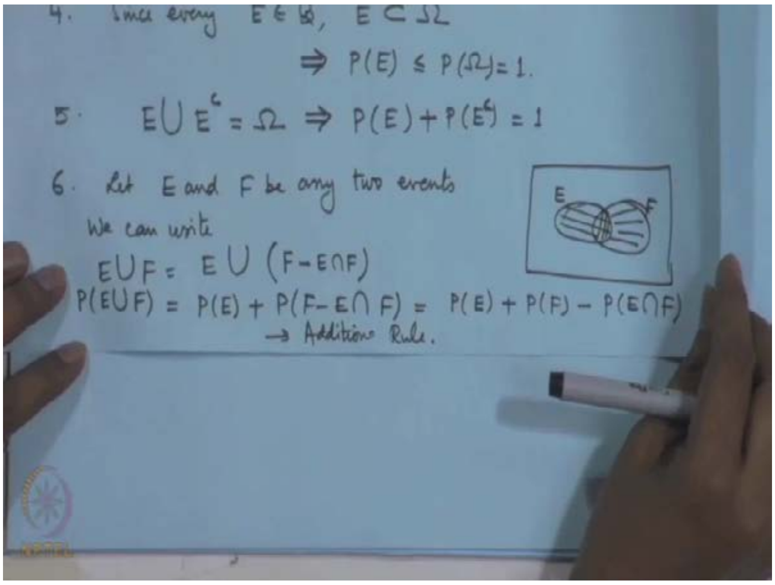
We also have some further things for example, if we say that an event is having more favorable out comes, then the event should have a higher probability of occurrence. So, this type of property should also follow from the axiomatic definition. So for example, if you say, if E is a sub set of F, then probability of E should be less than or equal to probability of F. Now the proof of this is very simple actually; we can look at say, a Venn diagram representation; suppose this is event, F this is event E. So, if we write F as union of E and union of F minus E, we can write F as E union F minus E. Then easily you can say that these are disjoint; so probability of F becomes probability of E plus probability of F minus E.

Now here you can see that this is a probability of certain event therefore, this is going to be non-negative; therefore, this is greater than or equal to probability of E. Also you can see here that a sub corollary of this, if E is a sub set of F, then probability of F minus E is equal to probability of F minus probability of E. Further you can observe that since, every E belonging to B has E subset of omega, this means that the probability of E will always be less than or equal to probability of omega that is 1. In the first axiom, we are having P E greater than or equal to 0; now we are also having that it is less than or equal to 1.

Another easy consequence you can see that if I consider any event E, and if I consider its complement, then this is equal to omega. Now if you apply the additive property, then it will give probability of E plus probability of E complement is equal to 1; that means, if I know the probability of one event, I can easily obtain the probability of the complementary event. Further we can see, let E and F be any two events that means, if I look at, suppose this is my sample space, and A and B are any, E and F are any two general events. Then you can see this E union F if I look at, then E union F I can represent as union of two parts.

We can write E union F is equal to E union this portion; this portion we can express as F minus E intersection F. Again, we have split the E union F into two disjoint regions, E and F minus E intersection F, these two are disjoint; because this portion is E, and this is F minus E intersection F. Therefore, the additive, additivity will be holding, and we get probability of E union F is probability of E plus probability of F minus probability of E intersection F. Now again observe this E intersection F portion, this is actually a subset of F. So, if we apply this that if E is the subset of F, then the probability of different is equal to the difference of the probabilities.

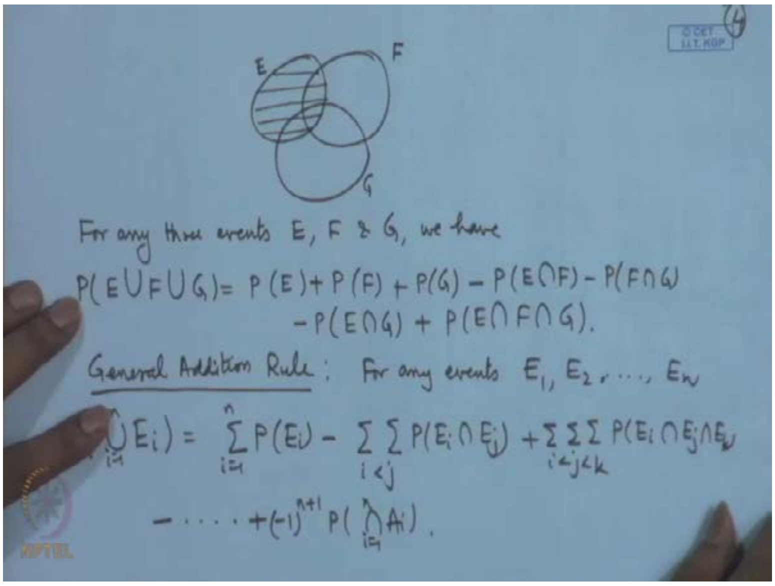
(Refer Slide Time: 09:05)



Then using these we can say that probability of F minus E intersection F should be equal to probability F minus probability E intersection F. So, we are getting probability of E union F is equal to probability E plus probability F minus probability of E intersection f.

So, this is called addition rule. Roughly you can explain it like this that if there are any two events, then the probability of the union will be equal to the sum; and since in the sum, the common portion is added twice. So, we have to remove it once.

(Refer Slide Time: 13:54)



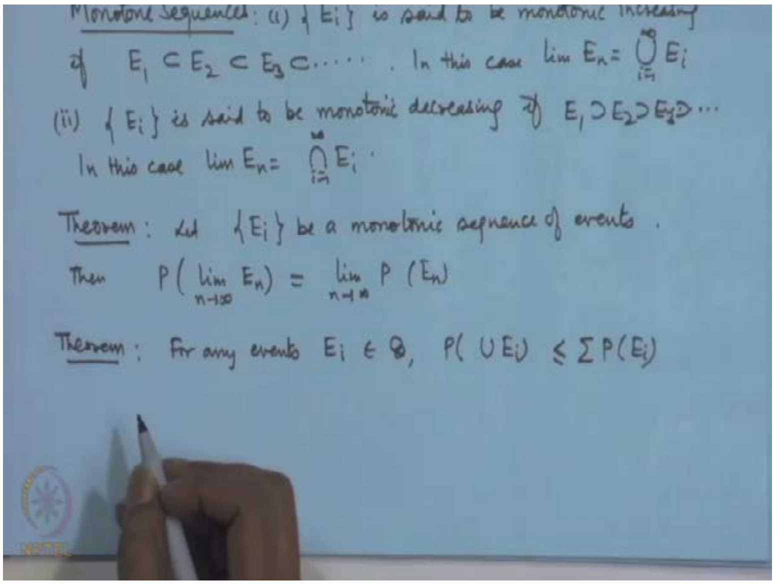
Now, this fact has an easy generalization for example, in place of 2, if you have 3. Suppose we have 3 events A, B, C; suppose this is say E, F and G; and we wanted to find out the probability of E union F union G, then if I consider this full thing E, then in F, I will have to remove E intersection F; and then in G both of these portions are removed. Now, consequently this portion is removed twice, which is E intersection F intersection G. So, for 3 events, then we will have the formula. For any 3 events E, F, and G, we have probability of E union F union G is equal to probability E plus probability F plus probability of G minus probability of E intersection F minus probability of F intersection G minus probability of E intersection G plus probability of E intersection F intersection G.

So, based on this, we have the general addition rule; for any events E 1, E 2, E N; probability of union E i, i is equal to 1 to n is equal to sigma probability of E i; subtracting 2 at a time, then adding 3 at a time and so on; next you will be subtracting 4 at a time and so on minus 1 to the power n plus 1 probability of intersection i is equal to 1 to n. This is the general addition rule; the proof can be easily obtained by using induction, because we have already seen that the result is true for 2. So, for n is equal to

1, it is trivial, trivially true; for n is equal to 2, we have already seen. So, if we write for n plus 1, you split in 2 upto n and then take one additional that is union A i, i is equal to 1 to n union E n plus 1. Apply the formula for 2, and in the previous portion you apply the formula, which you have assumed for n, then the result will come. I am skipping the proof here; however, it is available in the standard text books, and also in my other lectures on probability and statistics.

Now, we look at certain other important ramification of the symmetric definition. Since, we are we can talk about large number of events, we can talk about the sequence of the events also. So, when we talk about the sequence of events, there is the concept of the limit. For example, we may have sequence of set such that E 1 is subset of E 2, E 2 is a sub set of E 3 and so on; these are known as monotonically increasing sequence of events. Similarly we may have monotonically decreasing sequence like E 1 could contain E 2, E 2 may contain E 3 and so on; these are known as monotonically decreasing.

(Refer Slide Time: 18:21)

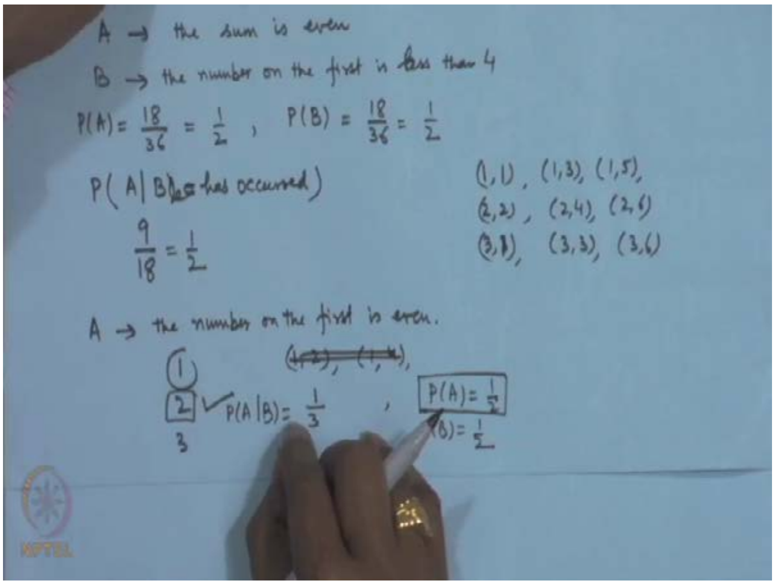


Now, whenever we have a monotonic sequence, the limits exist; that means, limit of a monotonically increasing sequence that is equal to the union; the limit of the decreasing sequence will be its intersection. So, we define what is known as monotone sequences. So, E i is said to be monotonic increasing if E 1 is a subset of E 2 is a sub set of E 3 and so on; in this case, limit of the sequence is defined to be union of E i. Similarly if E i is said to be monotonic decreasing, if E 1 contains E 2 contains E 3 and so on; in this case the limit of the sequence turns out to be intersection of all the sets.

Then we have the following result; let E i, be a monotonic sequence of events; then probability of limit E i is equal to the limit of the probability, that is known as the continuity assumption, I have continuity result for the probability that is the probability is also a continuous function. I am skipping the proofs of these results, because of the lack of time in this particular course. The probability we have proved that if I consider two events, then probability of a union B is equal to probability A plus probability B minus probability a intersection B. So, we are subtracting from the sum, a certain number; so that means, it is less than or equal to probability A plus probability B.

So, in generally if I consider any number of events, then for any events E i is belonging to B, probability of union E i is always less than or equal to sigma probability of E i. No matter, what this union is; this union may be finite, it could be infinite union also. This is known as the (( )) property of the probability function

(Refer Slide Time: 21:57)



Next, we proceed to the concept of conditional probability conditional probability. So, let us consider an example. Let two dice, let me assume they are fair, be tasked; let us consider say A to B an even that the sum is even, and we put an event say B that the number on the first is less than 4. Let us look at the probabilities of these events; what is the probability of A? In the sample space, there will be 36 possibilities; out of which, 18 possibilities would lead to the even sum; for example, 1 1 1 3 that means, if both are add or both are even; if that is happening, then the sum will be even. So, the probability of A will be half.

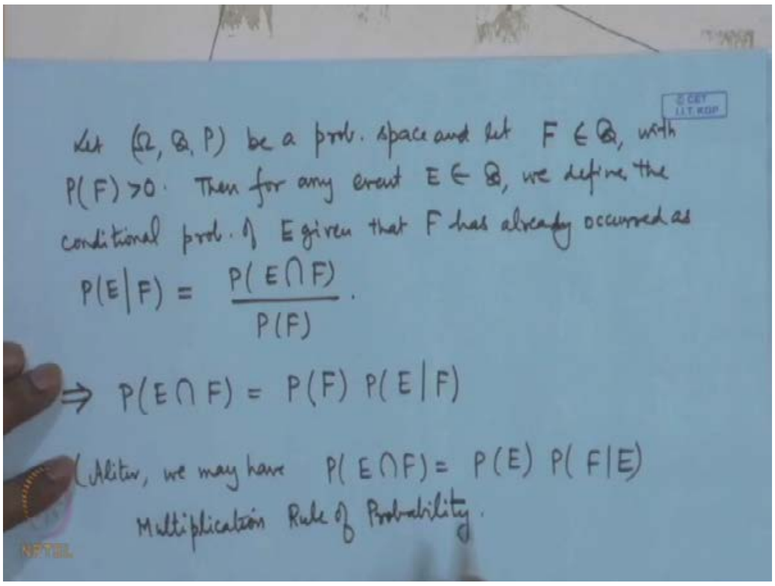
Let us look at what is the probability of B; that the sum on the first is less than 4 sorry the number on the first is less than 4. Now number on the first; first means that first dice has the possibilities 1, 2, 3, 4, 5, 6. So, 1, 2, 3 are the favorable outcomes here; along with that all the outcomes of the second one are associated. Once again there are 18 outcomes that are possible, that is equal to half. Now, let us look at what is the probability of A given B; I am considering this is something different; that means, given that B has occurred. Now, what does it mean? It means that we know already that the number on the first is less than 4; then what is the probability that the sum is even.

Now, let us look at the possibilities. The first number is less than 4; that means, number could be 1, 2 or 3. Now, what is the possibility for the second one? It could be (1,1), (1,3), (1,5); for 2, it will be (2,2), (2,4), (2,6); for 3, will be 3, 3 (3,1), (3,3) and (3,6) that is 3; there are total 9 possibilities. There are total 9 possibilities, out of the total possibilities of B; that is equal to 18. So, this number turns out to be 18, that is equal to half.

Now, let me modify this example. I say that the event A, I modify; my event A is now say that the here I am considering the B to be the number on the first is less than 4, and I consider A to be the event that the number on the first is even. Now let us see; what are the possibilities of the A now? Out of the possibilities of B; the B is giving the possibilities that numbers are 1, 2, 3; and the first one is also even, then (1,2), (1,4) sorry so this will not be considered here, because the number on the first is less than 4 is 1, 2, 3, but the even possibilities only, one possibility here.

So, total number of possibilities will become only 1 by 3; what is the probability of A given B that will be equal to 1 by 3, whereas what is the probability of A? The probability of A was half; what is the probability of B that was equal to half. Now, you see the probability of A is actually half here; what is the probability that the number on the first die is even, the probability is half. But if we know that the event B that is the number on the first is less than 4 has already occurred, then the probability of A gets modified, because out of the possibilities of B, only three possibilities are there; here only one possibility that is to the occurrence of A. So, the probability of A given B becomes 1 by 3, this is the concept of conditional probability that means, if we are having certain prior knowledge about the occurrence of event, then that may have an effect on the probability of occurrence of another event. So, the ultimate probability of that may get changed basically, it means that the sample space has reduced.

(Refer Slide Time: 27:46)

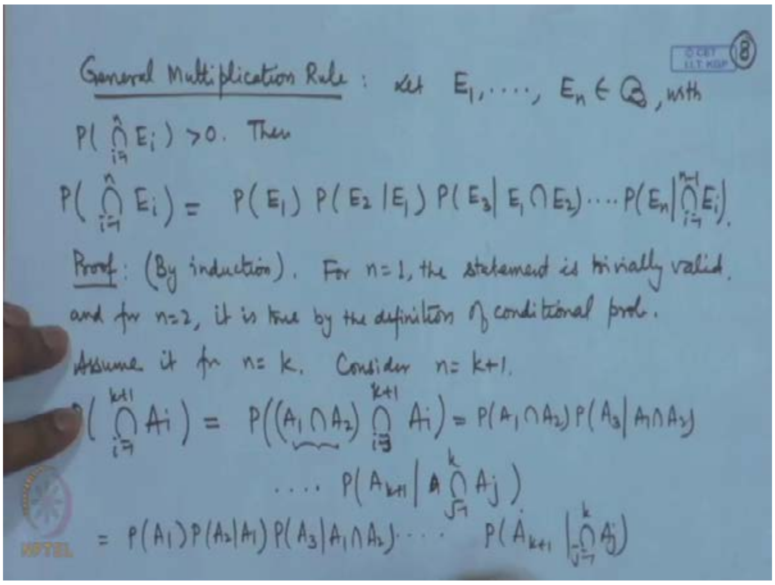


So, let us formally define the conditional probability that let omega, B, P be a probabilities space, and let F be an event with probability of F positive. Then, for any event E, we define the conditional probability of E given that F has already occurred as probability of E given F this is the notation, and that is defined as probability of E intersection F divided by probability of F. One can check that this concept is satisfying the axioms of the probability; that means, it is a non-negative function, the probability of omega, suppose I take omega here, then this will come omega intersection F divided by F; now omega intersection F is F, so this ratio will become 1.

And if I consider countable sequence E 1, E 2 and so on, then if I consider the probability of union E i given F, then probability of union E i intersection F will become distributed here, and then it will become equal to this sum. So, the basic axioms of probability are valid for this new definition, which we call conditional probability; so, this definition is justified. Now, let us look at the further ramifications of this definition also. So, we can write like this. So, this implies probability of E intersection F, we can write as probability of F into probability of E given F.

See as alternative, we may have probability of E intersection of F also as probability of E into probability of F given E; provided this conditional probability is well defined that means, if probability of E is positive, in that case we may write like this, in place of this. Now, what does this this statement of this event a statement represents? This represents that probability of a simultaneous occurrence of two events E and F is equal to the probability of one of them into the probability of conditioning on the other one. If we look at the second one that is also the same; this is the probability of the simultaneous occurrence, and this is equal to the probability of occurrence of E into the probability of conditional occurrence of F. This simple statement is known as the famous multiplication rule of probability, multiplication rule of probability. Now, naturally this can be further generalized in place of two events, we may have three events, we may have n events.

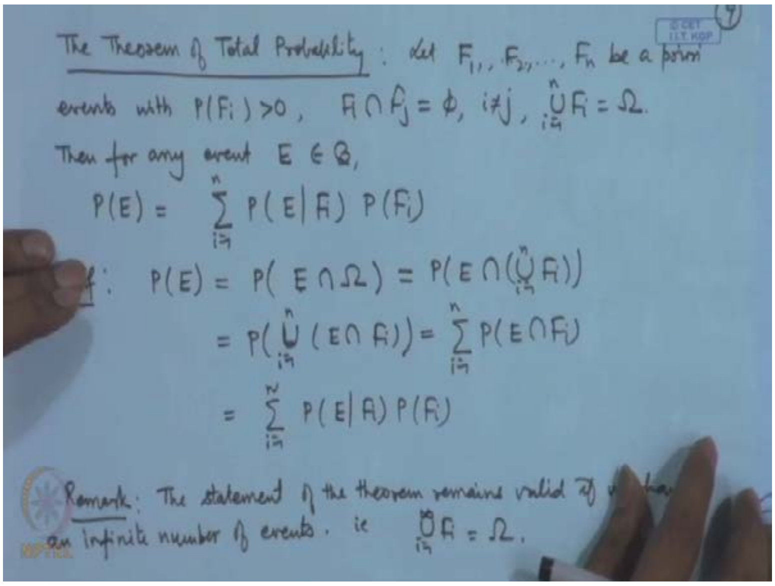
(Refer Slide Time: 31:42)



So, we write the general multiplication rule; the general multiplication rule; let us consider events E 1, E 2, E n, n events are there. And we assume, because if we have to define the conditional probabilities, then the probabilities of those conditioning event should be positive. So, we take the smallest one that probability of say intersection E i, i is equal to 1 to n to be positive; then the probability of simultaneous occurrence of this can be represented as probability of say E 1 into probability of E 2 given E 1 into probability of say E 3 given E 1 intersection E 2 and so on, probability of E n given intersection E i, i is equal to 1 to n minus 1.

The proof is very simple in fact, it can be proved by induction; once again, let me demonstrate this proof here. So, for n is equal to 1, the statement is trivially true, and for n is equal to 2, it is true by the definition of conditional probability. So, assume it for n is equal to k, and then let us consider n is equal to k plus 1. So, we look at the probability of intersection A i, i is equal to 1 to k plus 1. What we do? We represent it as A 1 intersection A 2 intersection A i, i is equal to 1 to i is equal 3 to k plus 1. So, we are considering this as 1, and these as another k minus 1 events, so this total number of events becomes k because this I am treating as 1 into t. So, this becomes probability of A 1 intersection A 2 into probability of A 3 given A 1 intersection A 2 and so on, probability of A k plus 1 given intersection A jj, j is equal to 1 to k. Now, the first one again we can write as probability of A 1 in to probability of A 2 given A 1 into probability of A 3 given A 1 intersection A 2. So, this proves the general multiplication rule.

(Refer Slide Time: 35:28)



Now, next we look at the theorem of total probability.

(No audio from 35:31 to 35:47)

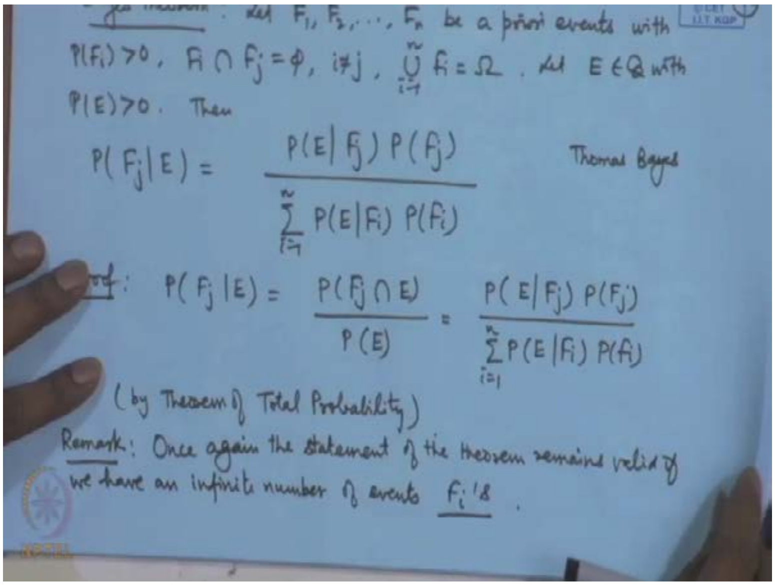
Let say F 1, F 2, F n be apriori events with probability of F i positive, and F i intersection F j be equal to phi for i naught is equal to j, and say union of F i, i is equal to 1 to n be equal to omega; that means, the events are mutually exclusive each with positive probability, and they are exhaustive events. Then for any event E, probability of E can be written as probability E given F i into probability of F i. So, actually the meaning of this apriori means that we are considering something like that event E is observed after F i (s) have occurred. So, as a consequence of either of F i(s) E could have occurred. In that case the probability of the final occurrence of E can be considered as a some of the probabilities of individual occurrences of E through F i(s) each of that means, what is the probability that F would have been cost by F 1, what is the probability that E could have been caused by F 2 and so on, and then we multiply by the individual probabilities of F 1 F 2 also.

So, let us look at the proof of this; proof is not difficult; in fact, we simple use that these are mutually exclusive and exhaustive events. So, we can write E as E intersection omega, where omega is the sample space, and then this omega we can expressed as union F i, i is equal to 1 to n now here we can apply the distributive property as of unions and intersection. So, you get union i is equal to 1 to n E intersection F i. Now if F i(s) are disjoint, then E intersection F i(s) are also going to be disjoint. So, this can be written as sigma probability of E intersection F i, i is equal to 1 to n; now on this we can apply the multiplication rule, so we get this as sigma probability of E given F i in to probability of F i, i is equal to 1 to n.

Now easily you can see that in place of n that is the finite number of events, if the same thing was true for an infinite number of events that means, your pairwise disjoint and exhaustive, then the same statement would have been true, if n is replaced by infinite. So, let me write it as a remark. The statement of the theorem remains valid, if we have an infinite number of events that is union F i, i is equal to 1 to infinity is equal to omega. So, the result will be still true.

Now one may look at in a slightly different way, here we are looking at E as a consequence of F I; so we want to know that what is the probability or what is the chance of the consequence E, and this E could have been caused by either of F i(s). Now, suppose we observe the outcome, then, what is the probability that it was caused by say F 1 or by F 2 that means, we look at it in the reverse way. Here I am calling a F i(s) to be apriori events; therefore, and this apriori probabilities of these are known, and these conditional probabilities of these are also known, and we are able to calculate the probability of E. But suppose we know that E has occurred, what is the probability of say F 1 or what is the probability of F 2? This reverse looking at this result is known as the famous Bayes theorem.

(Refer Slide Time: 40:53)



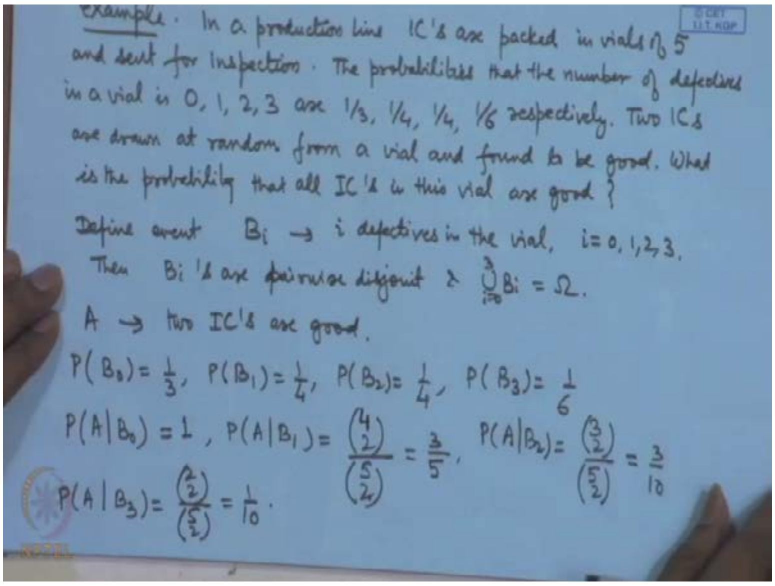
So, the conditions are the same, let F 1, F 2, F n be apriori events with probability of F i greater than 0, F i intersection F j is equal to phi i naught equal to j, union F i is equal to omega that means, mutually exclusive and exhaustive events. Let E be an event with probability E positive, then probability of say F j given E that is equal to probability of E given F j into probability of F j divided by sigma probability of E given F i into probability of F i, i is equal to 1 to n. This result is known as Bayes theorem named after Reverend Thomas Bayes.

The proof is based on the definition of the conditional probability and the theorem of total probability. So, we can express the conditional probability of F j given E as probability of F j intersection E divided by probability of E. Now, the numerator you can just apply the multiplication rule; and in the denominator, apply the theorem of total probability.

(No audio from 42:52 to 43:10)

And once again, we can notice that in place of finite number of events, even if we have infinite number of events in the same condition, then the result will be true. Once again the statement of the theorem remains valid, if we have an infinite number of events F values, and they should satisfy the same criteria. Let me explain this through 1 or 2 examples here.

(Refer Slide Time: 44:11)



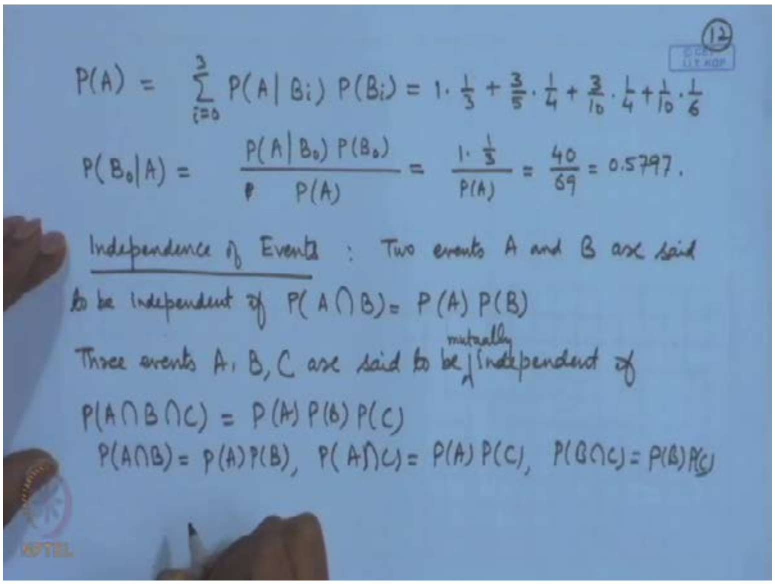
So, in a production line, IC’s are packed in vials of 5, and sent for inspection. The probabilities that the number of defectives in a vial is 0, 1, 2 or 3, these probabilities are 1 by 3, 1 by 4, 1 by 4 and 1 by 6 respectively. So, a vial has 5 IC, s. Now, out of this 5, there can be no defective with probability 1 by 3, there can be one defective with probability 1 by 4, there can be 2 defective with probability 1 by 4, and there can be 3 defectives with the probability 1 by 6. Now 2 IC (s) are drawn at random from a vial and found to be good. What is the probability that all IC’s in this vial are good? So, these events are apriori events that the vial may have 0, 1, 2, 3 defectives.

However, now we are looking at posterior (( )) you are already knowing the outcome that out of one vial which is selected, and we tested two out of that, and they turn out to be good. So, we already know the final outcome. Now what is the probability that my initial event was that it had 0 defective vials. So, this is a clear cut case of the application of the bayes theorem let us see this. So, define event B i saying that i defectives in the vial; this is for i equal to 0, 1, 2, 3, because these are the only 4 possibilities. Then you are having that B i's are disjoint, pairwise disjoint and union of B i is equal to omega.

And let us consider say A to be the event that two IC's, which are selected, they are good. Now let us look at the apriori probabilities, probability of B naught is 1 by 3, probability of B 1 is 1 by 4, probability of B 2 is 1 by 4, probability of B 3 is 1 by 6. So, we have the probabilities that are already given to us. Now we can also calculate the following probabilities that is probability of A given B naught. Now this means that what is the conditional probability that the two randomly selected vials are good; given that the vial had no defectives. Certainly if there were no defectives, so whatever you choose it is suppose to be good, so the probability will be 1.

Let us look at what is the probability of A given B 1; B 1 means that the vial had one defective; however, when we select two, both are good; that means, both are form the 4 good ones; there are total 5, there are total 5, out of which, 2 are selected; however, we are selecting from the two good ones, the two from the good ones only; that means, they are selected from 4 only. Now this number of course, even simplify that is 3 c 5, 3 by 5. Similarly, if you look at A given B 2, this event represents that the vial had 2 bad ones, 2 defectives however, we did not choose them. So, the selection is out of 3 good ones, out of the total selection of 2 from the 5; this number simplifies to say 3 by 10. Similarly we can look at what is the probability of A given B 3 that is equal to now the vial had 3 defectives; however, we did not choose them that means, we got the only remaining good ones out of total selection of 5 c 2, this number is 1 by 10.

(Refer Slide Time: 50:32)



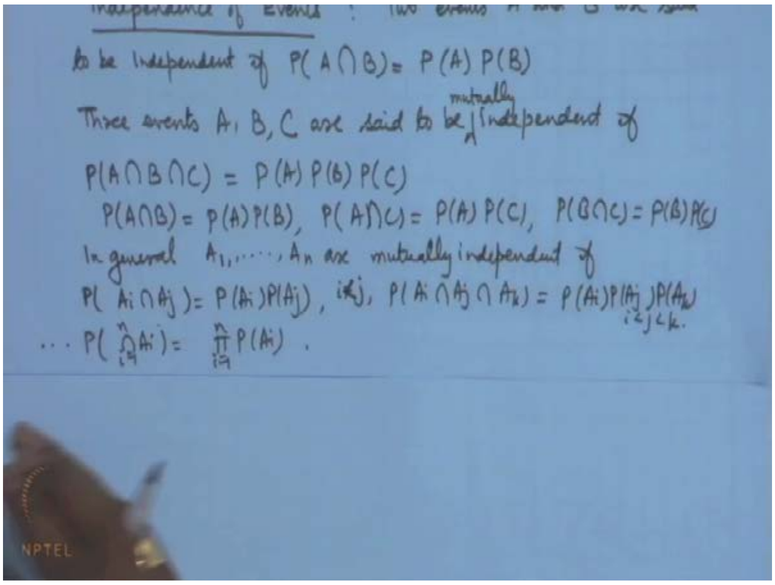
Now let us look at various possibilities for example, what is the probability of A? So, the probability of A, this will be equal to sigma probability A given B i into probability of B i, this is by theorem of total probability; that means, what is the probability that the 2 items, which are selected at random from the selected vial, they are good. So, this will be based on apriori one that means, whether the initial ones, where having 0, 1, 2 or 3 defectives. Now these numbers we have already calculated, so it becomes simply 1 into 1 by 3 plus 3 by 5 into 1 by 4 plus 3 by 10 into 1 by 4 plus 1 by 10 into 1 by 6.

And if you want to calculate what is the probability of say all IC's in this vial are good; that means, we are interested to find out the probability of B naught given A that is equal to probability of A given B naught into probability of B naught divided by probability of A, that is the theorem of total probability. So, this numerator value turns out to be 1 into 1 by 3 divided by the denominator that is probability of A, which after simplification turns out to be 40 by 69.

Let me briefly define the concept of independence of events. We have defined the conditional probability that A given B, it is equal to probability of A intersection B divided by probability of A. Now, if this is equal to probability of A that means, if the conditional probability of A given B is equal to unconditional probability of A, then we should say that B has no effect on A, and we can say A is independent of B. Now if we write down this condition, probability of A intersection B divided by probability B is equal to probability A, it is coming probability of A intersection B is equal to probability A into probability of B. Now that is reducing to that probability of A that is the probability of the simultaneous occurrence is equal to product of the probabilities.

So, we can define two events A and B are said to be independent, if probability of A intersection B is equal to probability A into probability of B. Now likewise, we can generalize definition, if I have three events, three events A, B, C are said to be independent, if probability of A intersection B intersection C is equal to probability A into probability B into probability C. But this is not enough; actually, we need some more conditions; A and B should be pairwise independent, A and C should be pairwise independent, B and C should be pairwise independent. Then this will be said to be independent, and we add the word here mutually independent here; that means, in general if I have n events, then for the independents, I will have to write the condition taking two at a time, three at a time, four at a time, and all of them at a time.

(Refer Slide Time: 55:10)



In general, then we will say A 1, A 2, A n are mutually independent, if probability of A i intersection A j is equal to probability of A i into probability A j for i less than j 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, i less than j less than k probability of and so on, probability of intersection A i, i is equal to 1 to n is equal to product of the probability A i, i is equal to 1 to n. Also it is clear from here that unless all the conditions are satisfied, we cannot say the events are independent. So, for example, if I am talking about three events, it may happen that say this condition is satisfied or this condition is satisfied or this condition is satisfied, but this condition need not be satisfied. Similarly may be this condition is satisfied and this condition is satisfied, but these two conditions are not satisfied.

So, unless all the conditions, which are given for the mutual independents of all the events, then we will not say that the events are independent; because it could happen that A and B are independent, A and C may also be independent, but along with B and C, A may become dependent; that means, B and C together may be able to determine what is A. Now this type of phenomena is observed in various statistical experiments, I will be giving some examples of these phenomena in the next lecture.