Statistical Methods for Scientists and Engineers

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

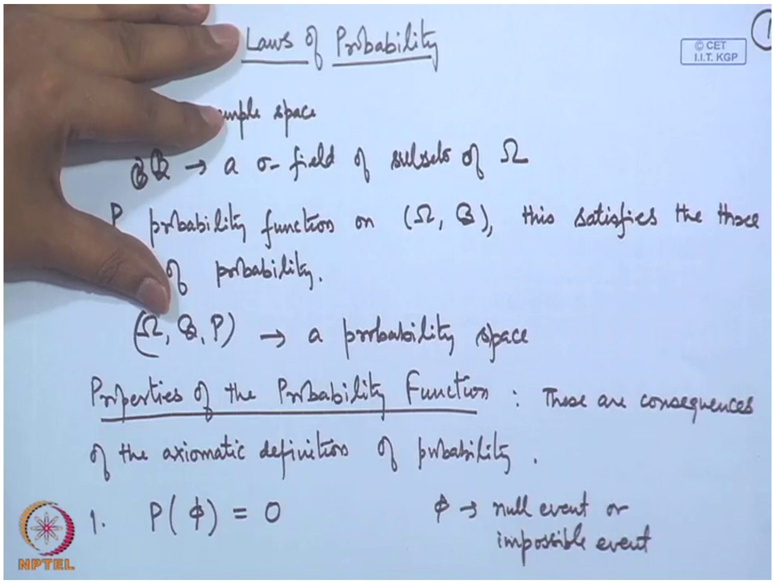
Lecture - 02
Laws of Probability

Friends, in the last class, I traced the origins of the term Probability, and we saw the historical development of the definition of probability, the methods of calculation of the probability. We have seen that historically the problems that were considered they were of nature in which we can consider a mathematical definition or we can say so called classical definition, we had games of chance and dice throwing, tossing of a coin and so on.

Where we could use the definition which was classical definition where we have a finite number of elements in the sample space and the outcomes are considered to be equally likely. Later on it was discovered that the definition is quite restricted and it cannot be used for several kind of phenomena for which we need the problemistic statements. A more practical definition was so called empirical definition or a statistical definition of probability which is based on the evidence.

That means how many times over a period of time a particular event occurs so looking at the proportion, if this proportion is stabilizing to a certain number we defined that as the empirical definition of probability. However, it was found that even in this definition has certain inadequacies and therefore, later on for measure theoretic definition of probability was given by A N Kolmogorov, it was called the axiomatic definition.

(Refer Slide Time: 02:15)

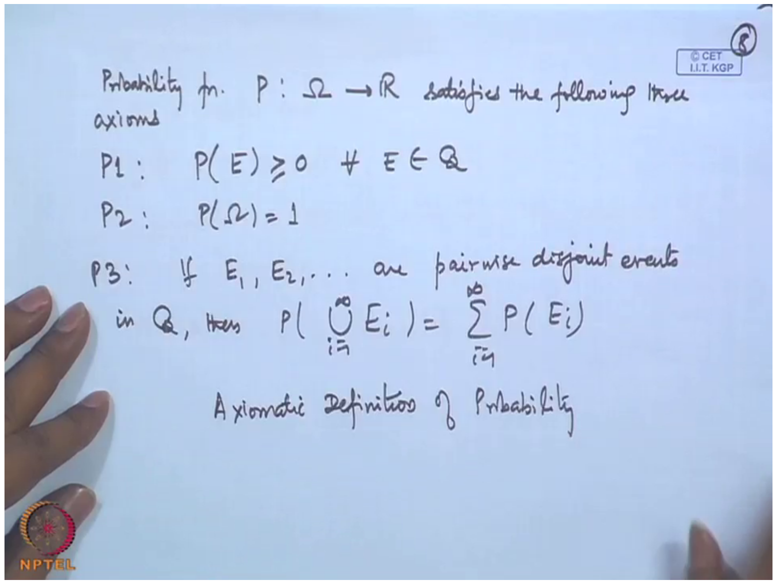


And in the last lecture towards the end I introduced this definition, just to recall the terms here we have a random experiment which results in a sample space omega, and we consider a class of subsets of omega which is satisfying certain conditions and we called it a sigma field of subsets of omega, based on this we have defined P as the probability function this is defined as the probability function on omega, B, P.

And this is satisfying, this satisfies the 3 axioms of probability, so we can say the 3 axioms of probability are satisfied here. So we will call omega, B, P as a probability space, so let us look at our structure we have a random experiment on the basis of the random experiment the possible outcomes which are of interest towards we are collected in a sample space called omega. And we consider the class of subsets of omega which satisfies certain conditions as we recollect from our last lecture it is a sigma field.

And then on this we define P as the probability function satisfies the three axioms of probability just to recollect what we did in the last class, let me show of the slide for the definition of probability it is as follows.

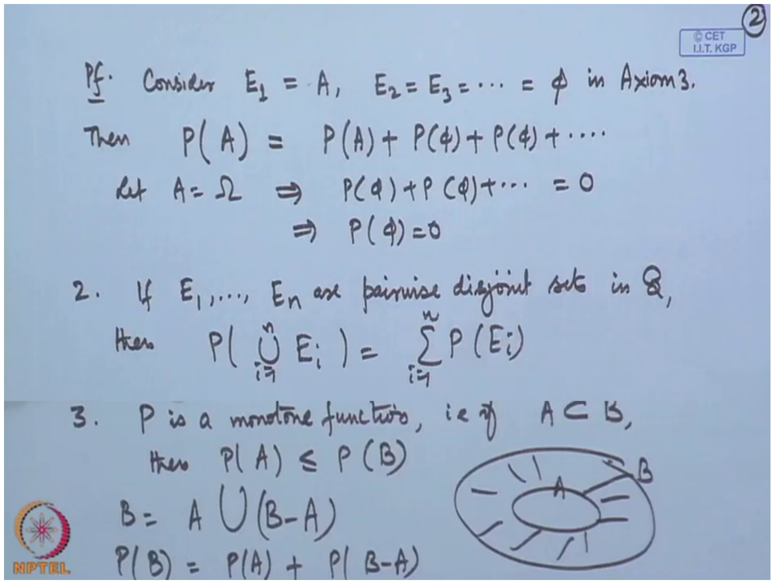
(Refer Slide Time: 04:15)



We have probability to be a non-negative function, the probability of the fully space=1, and if we are considering a pairwise disjoint events then the probability of their union is = some of the probabilities, so these are the 3 basic axioms of the axiomatic definition of probability, so this is the so called axiomatic definitions of probability. So now when we say that omega, B, P is a probability space this is satisfying the 3 axioms that were given there.

Now let us look at certain laws or certain properties that this probability function will have, so let us look at certain properties of the probability function, these are actually consequences of the axiomatic definition of probability. The first consequences that the probability of the empty set is 0, phi denotes the null event or the impossible event as we defined earlier. Note that we must expect that the probability of impossible event should be 0.

(Refer Slide Time: 06:12)



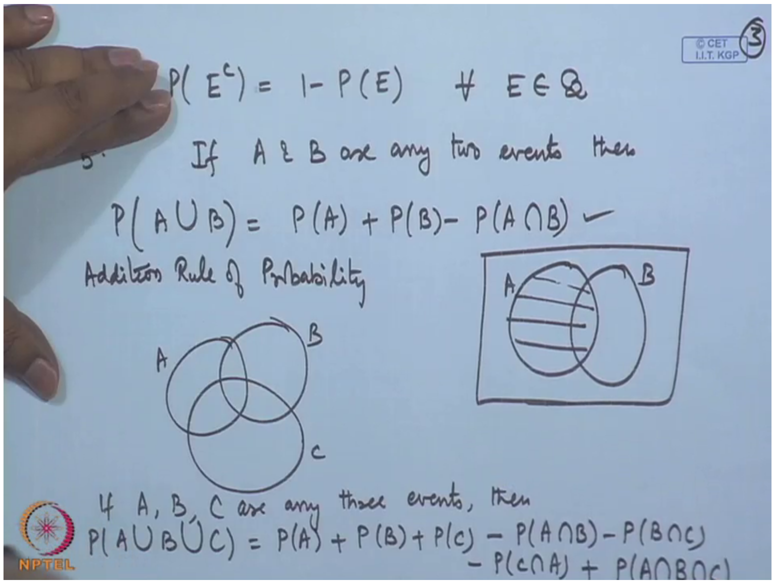
Now what we are saying is that this is consistent with our axioms, so let us see here let me give a sketch proof of this here, consider say E1= say certain set here say A and E2, E3 and so on, this is= say 5 in axiom 3. If we do that we get probability of E1 union E2 union so on that will become A on the left hand side, on the right hand side it will be the probability of E1+ probability of phi+ probability of phi and so on.

Now we can this probability of A is the non-negative number, so we are saying that so we can actually remove this number from here by considering say A=omega then what will happen? You can have P omega on both the sides you can cancel, so you get P phi + P phi=0, this means P phi is actually 0. Let us look at the second consequence if E1, E2, En are pairwise disjoint sets in B, then probability of union Ei i=1 to n= sigma probability of Ei.

So let us compare it with the third axiom here if we have any collection of pairwise disjoint sets, then the probability of union is = sum of the probabilities, now in this axiom if we take only a finite number here and the remaining one I substitute as phi then this is reducing to this statement. So let us look at the third statement P is a monotone function that is if A is a subset of B, then probability of A<= probability of B.

Once again you can easily construct a proof here, let us consider say this set as B and A as a subset, then basically what you are having that B is actually = A union B-A, now this is a disjoint thing so probability of B= probability of A+ probability of B-A, so since this is a non-negative term you can conclude that probability of A<= probability of B, so it is a monotone function.

(Refer Slide Time: 09:24)

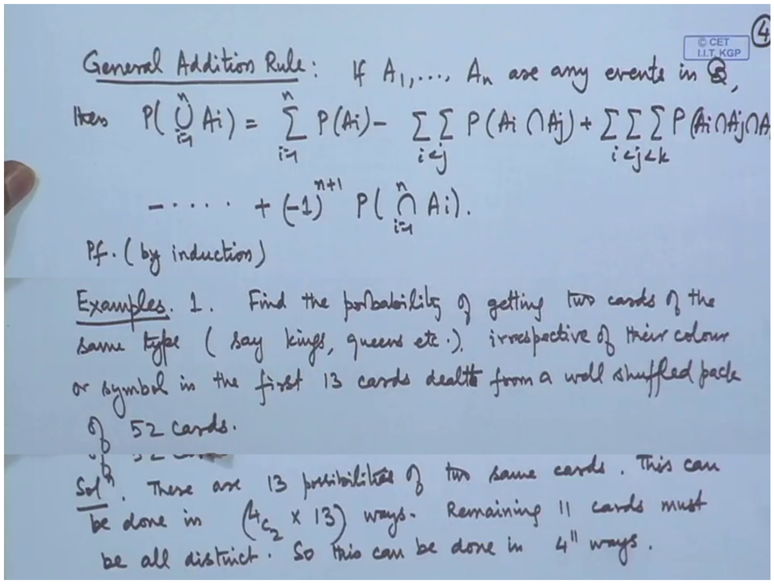


Some simple consequences you can see probability of E complement will be = 1-probability of E for all E that is the probability of complement is 1-the probability of the original set. Similarly, probability of so you can also say that if I consider say A and B are any 2 events then probability of A union B= probability of A+ probability of B-probability of A intersection B, this is actually called addition rule of probability.

The proof is again very simple you can look at sketch by using Venn diagrams, suppose I have a event here and B event here and if I am looking that A union B then A union B can be written as A+ this portion of B which is actually B-A intersection B, so the probability of A union B becomes probability of A+ probability of B-probability of A intersection B. Now this result one can look at generalization, suppose in place of 2 I have 3 events say A, B, C.

Then in that case if A, B, C are any 3 events, then probability of A union B union C= probability of A+ probability of B+ probability of C-probability of A intersection B-probability of B intersection C-probability of C intersection A+ probability of A intersection B intersection C. Now naturally so what does it denote actually as we have explained earlier this is probability of occurrence of either A or B or both.

(Refer Slide Time: 12:18)



Similarly, here it will be probability of occurrence of at least one of A, B, C, so naturally one will be interested in finding out the probabilities of occurrence of at least one of n number of events, so that gives us a general addition rule if A1, A2, An or any events in B, then probability of union Ai i= 1 to n probability of Ai intersection Aj+ probability of Ai intersection Aj intersection Ak and so on+- 1 to the power n+1 probability of intersection Ai i= 1 to n.

So one can actually prove these results by induction, because the proof of this is almost trivial here and extension of this you can use by you considering A union B and C and then splitting and then doing with twice. Similarly, for proving this you can do it by induction, in this particular course I would be skipping the proofs because I have to cover various topics which are to be used by scientists and engineers.

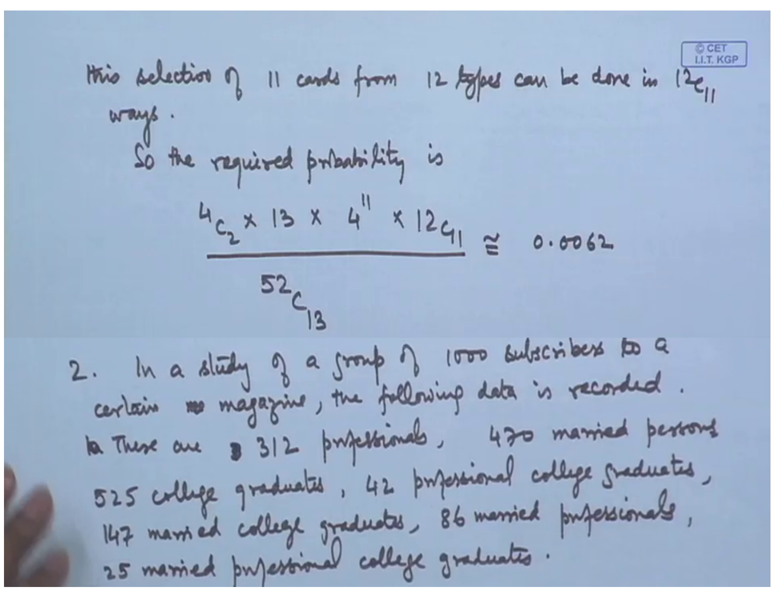
So if any user is interested in the proofs, so one can look at the references which are given here in the syllabus and also you can look at the other lectures which are available on probability and statistics by myself. So those lectures covered the proofs of these facts here I will be only telling the main rules which will be used by the users which are applied scientist and engineer will be making use of the results of probability theory.

Let me give an example here of some elementary probability problems, find the probability of getting 2 cards of the same type say 2 Kings 2 Queens etc. 2 cards of the same type irrespective of their color or symbol in the first 13 cards dealt from a well shuffled pack of cards. So let us look at the conditions of the experiment, we have a well shuffled pack of 52 cards that means we have 13 cards of hearts, 13 parts of a spade, 13 parts of a club and 13 cards of diamond.

Starting from 1 to 10 numbers and then 11 is Jack, 12 is queen and 13 is king, so what is the probability that we will be drawing 2 cards of the same type in this, so let us consider this. Now there are 2 cards of the same type, so how many such possibilities are there, since we have 13 different numbers for example 1, 2, 3 and so on, so there are 13 possibilities of 2 same cards okay, now these can be done in now for each there are 4 cards so 4C2 is the number of probabilities and out of these 13 any of them can be done.

So this is possible in 4C2*13 ways, now you have 11 more cards so there should all be distinct because we are putting that only will 2 cards are same type, so remaining 11 cards must be all distinct, so this can be done in 4 to the power 11 ways. Now we have chosen one type here, so for which 2 cards are there, now out of remaining 12 type 11 have to be chosen, so that can be done in 12C1 ways.

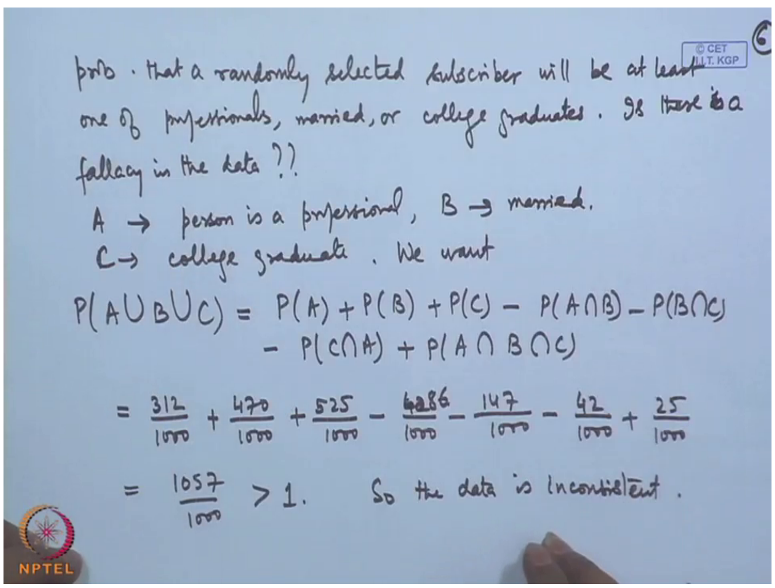
(Refer Slide Time: 18:40)



Now this selection of 11 cards from 12 types can be done in 12C11 ways. So the required probability it is 4C2*13*4 to the power 11*12C11 and total number of ways of choosing 13 cards out of the pack of 52 is 52C13, so this number can be calculated and this is approximately 0.0062. Let us take another example, in a study of a group of 1000 subscribers to a certain magazine, the following data is recorded.

There are 312 professionals, 470 married persons, 525 college graduates, now so we have considered the data with respect to profession marriage status and education out of this, then 42 professional college graduates, 147 married college graduates, 86 married professionals and 25 married professional college graduates.

(Refer Slide Time: 22:09)



Find the probability that a randomly selected subscriber will be at least one of professional married or college graduate that means he will be satisfying at least one by one of the properties. Now is there fallacy in the data? So we are asking an additional enquiry here, let us see suppose I define the event A as that the person is a professional, say B is the event that person is married, say C is the event that is a college graduate.

Then basically we are interested in what is the probability of A union B union C, now according to the addition rule this will be probability of A+ probability of B+ probability of C-probability of A intersection B-probability of B intersection C-probability of C intersection A+ probability of

A intersection B intersection C, now this= number of professionals is 312, so this will become 312/1000+ married persons are 470 so it will become 470/1000.

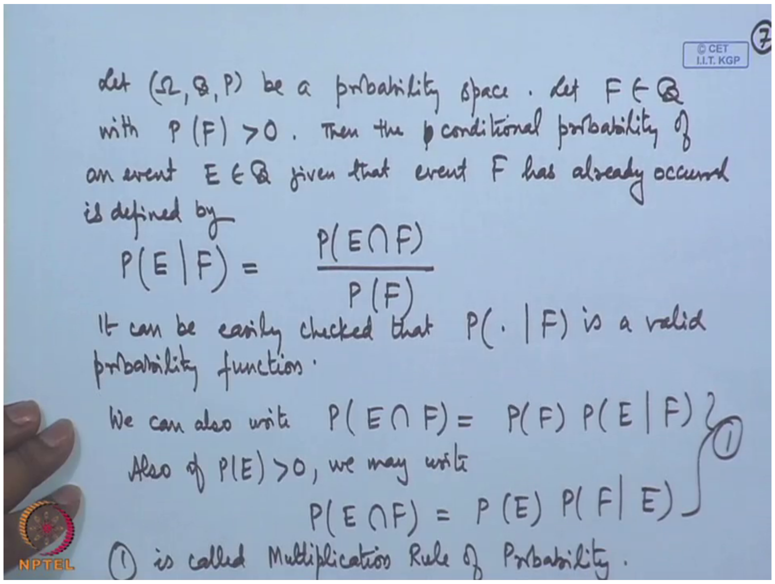
Then the number of college graduates is 525 and like that we substitute all the values A intersection B is the married professionals, so that is we have 42, then we have married college graduates that is this is 86 here, then we have B intersection C that is married college graduates that is 147, then we have professional college graduates, so we have 42 then we have A intersection B intersection C that is married professional college graduates 25/1000.

Now if you look at this, this turns out to be 1057/1000 which is obviously >1, so that data is inconsistent because you cannot have the probability >1, so using the addition rule we are able to detect fallacy in the given data. Now we look at some further aspects of the rules of probability we define something what is called conditional probability, it is something like this, if we consider here that what is the probability that person is a professional we are giving the answer that is 312/1000.

But suppose I asked the question what is the probability that the randomly selected person is a professional given that he is already married, in that case I will have to look at the number of persons who are already married among them how many are professionals, and therefore, the answer will become 86/470. Similarly, if you want to find out what is the probability that the person is a professional given that he is a college graduate?

So in that case I have to look at all the college graduates and among them how many are professionals, so we have only 42 such things, so we have 42/525 rather than 312 that means if we apply a condition in our random experiment, then the probability of original event may change this is called the concept of conditional probability.

(Refer Slide Time: 27:29)

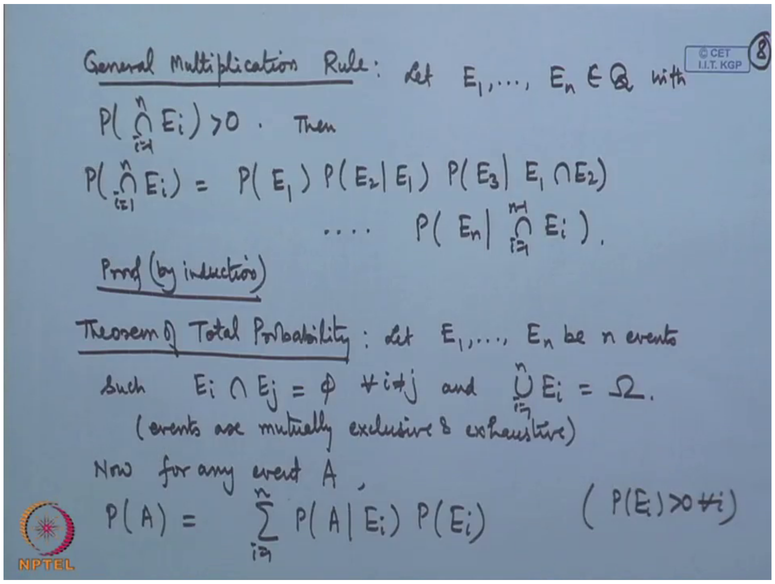


So let me introduce formally let omega, B, P be a probability space, let F be an event with probability of F >0 then the conditional probability of an event E given that event F has already occurred is defined by probability of E given F this is the notation of conditional probability, this is not E/F it is E given F, so probability of E intersection F/probability of F. One can easily check that this is a valid probability function.

It can be easily checked that this conditional function is a valid probability function, now let us look at these consequences of this definitions also, so from here we can see that the conditional probability = the probability of simultaneous occurrence of 2 events/probability of the other event which is conditioning event. So from here we can also write probability of E intersection F= probability of F*probability of E given F.

Also if probability of E is positive we may write probability of E intersection F= probability of E*probability of F given E, so these statements this is called multiplicative probability or multiplication rule of probability, this is the multiplication rule of the probability. Now easily one can see that in place of 2 events if we have 3 events, 4 events and so on, then this type of a statement can be further extended.

(Refer Slide Time: 30:32)



So we have as a consequence a general multiplication rule, let E1, E2, En be n events with probability of each of them actually positive, so I am taking the smallest one to the positive. Then probability of intersection Ei that is the simultaneous occurrence of each of them= probability of E1*probability of E2 given E1*probability of E3 given E1 intersection E2 and so on probability of En given intersection Ei i= 1 to n-1.

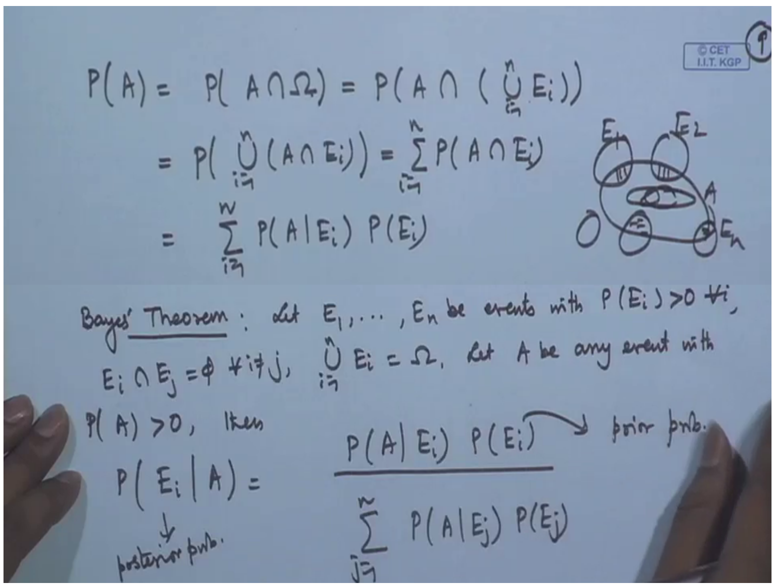
One can again prove by induction this thing, I am skipping the proof here, now let us look at further consequences of these conditional, many times we are looking at the phenomena in the following fashion. Let us consider certain event for example we are looking at the death of a person that means we are looking at the records of the deaths in a particular locality, so for example it could be a city municipality data.

Now we may like to look at the causes of the death, so a person might have died due to a disease, he might have died due to an accident and so on so forth, and even in the diseases one may have died due to a lung disease, one may die due to a kidney disease, or one may die due to a liver disease etc. so in that case if you are looking at the event death of a person what is the probability of death among the total population at a certain age for example.

Then we can classified it according to the different causes, so this is called the you can say the total probability that means the total probability of a particular event can be split into the probability of that event caused by something caused by another cause and so on. So we have theorem of total probability, so let E1, E2, En be n events such that Ei intersection Ej=phi for all i != j, and union of Ei= omega.

So which basically what we are saying that the events are mutually exclusive and exhaustive okay, so now for any event A, probability of A can be written as sigma probability of A given Ei*probability of Ei, here we are assuming that the probability of Ei is positive for each i. Let me explain the proof here at least for this statement here.

(Refer Slide Time: 34:51)



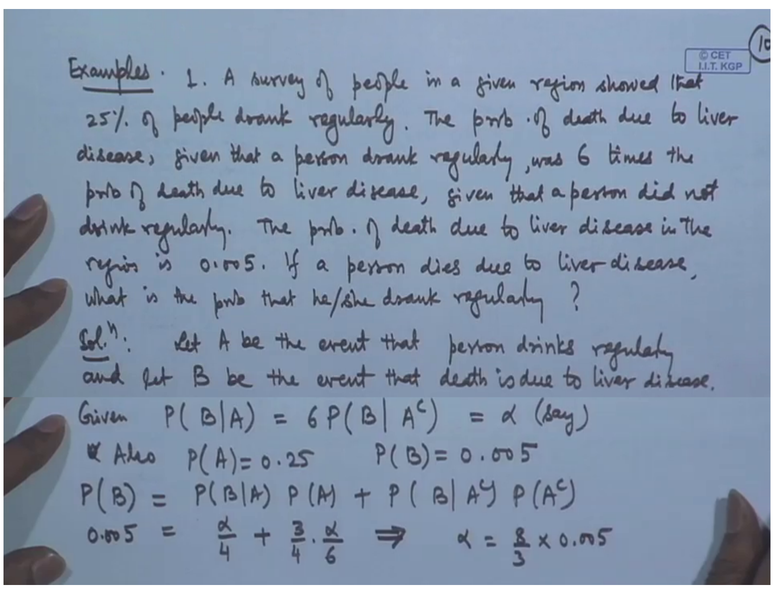
See let us write down probability of A, we can consider it as probability of A intersection the fully space because any event is the subset of the fully space and this you can write as A intersection, since Ei's are exhaustive event we can write omega as union of Ei's, so this is becoming probability of union A intersection Ei i=1 to n. Now this you can consider see you have mutually exclusive events say E1, E2, En etc. okay.

And this is some event A and if A is disjoint with if Ei's are disjoint and I consider say A is a set then A intersection E1, A intersection E2, A intersection E3 and so on they will also be disjoint, so this will become probability of because the probability is negative, now on this we can apply the multiplication rule, so this becomes probability of A given Ei*probability of Ei i=1 to n. Now a further consequence of this is the famous Bayes Theorem which is named after Reverend Thomas Bayes.

So we have the same set up here, let E1, E2, En be events with probability of Ei is positive, Ei intersection Ej=phi for all i !=j and union Ei= omega. Let A be any event with probability of A positive, then probability of Ei given A it is = probability of A given Ei*probability of Ei/sigma probability of A given Ei*probability of Ej j= 1 to n, so this is called then posterior probability, this is called then prior probability that means we have the prior probability of the causes.

And also the conditional probability of an event given those causes, then we can calculate suppose we know that what is the event that has actually occurred, then what is the posterior probability that means if we know the effect then what was the cause we can find out the probability of the cause, so this is so called Bayes theorem because it is named after Reverend Thomas Bayes and it was published in 1763. There are very interesting consequences of this result I can explain through certain example let us consider.

(Refer Slide Time: 38:45)



A survey of people in a given region showed that 25% of people drank regularly, the probability of death due to liver disease, given that a person drank regularly was 6 times the probability of death due to liver disease, given that a person did not drink regularly. The probability of death due to liver disease in the region is 0.005, if a person dies due to liver disease, what is the probability that he or she drank regularly?

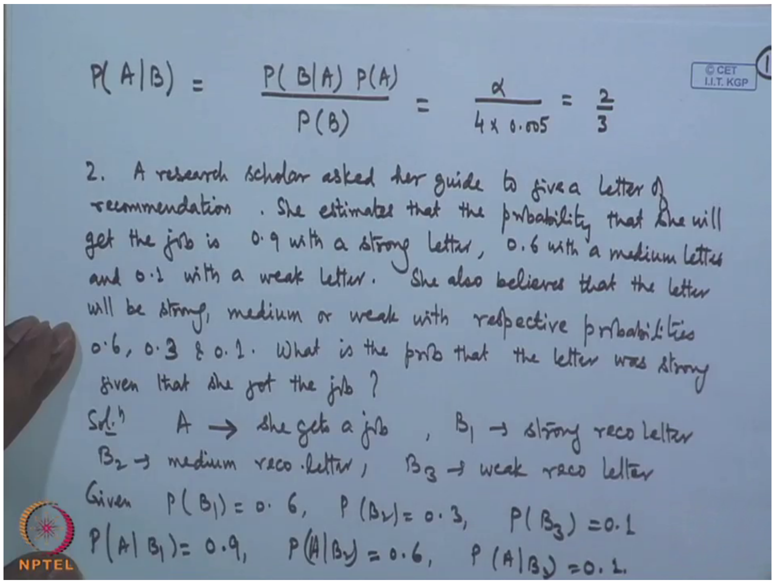
So now you can see that the death may be caused by a liver disease, and for the liver disease the cause maybe drinking regularly or not, now we know that the person dies due to a liver disease, so we are looking at the posterior probability of whether he was a regular drinker or not. So let us define we will apply the Bayes Theorem here, let us define the events here, suppose I consider let A be the event that person drinks regularly, and let us consider B be the event that death is due to liver disease.

Now what is given? It is given that probability of B given A that is the death due to liver disease is 6 times that is due to liver disease given that the person is a regular drunker is 6 times the probability of death due to liver disease given that he is not a regular drinker, let us put say this is =alpha. Also it is given that probability of A person drinks regularly is because it is given that 25% of the persons drink regularly so 0.25.

And also the probability of death due to liver disease is 0.005, so we can apply the theorem of total probability, so probability of B= probability of B given A*probability of A+ probability of B given A complement*probability of A complement. Now we have assumed that the probability of B given A= alpha, so this becomes alpha/4+ 3/4 probability of A complement is 3/4 and B given A complement= alpha/6, and probability of B= 0.005.

So from here you can get alpha= 8/3*0.005 that means the probability of death due to liver disease given that he is a regular drinker that is = 8/3*0.005.

(Refer Slide Time: 44:13)

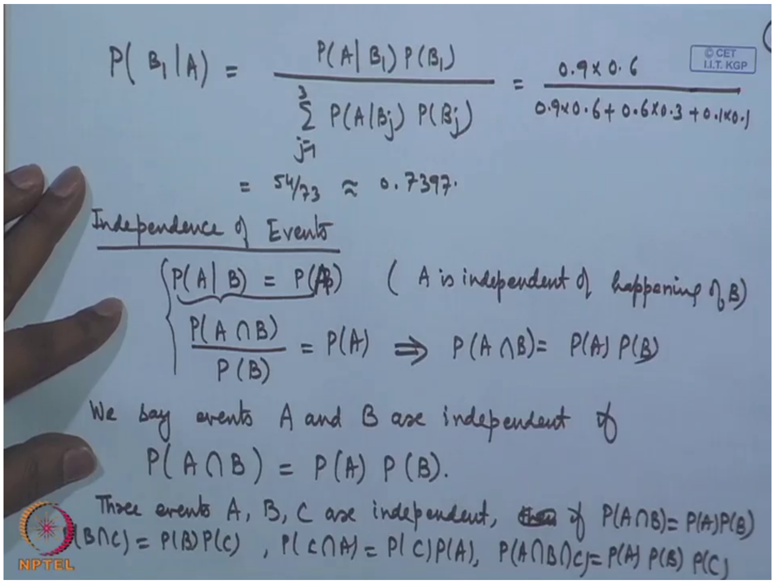


Now let us look at our event what is the probability of that he drank regularly given that he dies due to a liver disease, so here we can apply Bayes Theorem, so that is= alpha/4*0.005 that is= 2/3, because alpha we have already calculated= 8/3*0.005, so if you substitute it here we will get 2/3. So if a person dies due to liver disease then there is a high chance that actually he was a regular drunker. Let us take couple of more problems of similar nature here.

A research scholar asked her guide to give a letter of recommendation, she estimates that the probability that she will get the job is 0.9 with a strong letter, 0.6 with a medium letter and 0.1 with weak letter. She also believes that the letter will be strong, medium or weak with respective probabilities 0.6, 0.3 and 0.1. What is the probability that the letter was strong given that she got the job? So let us define here A to be the event that she gets a job.

Let us consider B1 is the event that the letter is a strong recommendation letter, B2 is the event that medium recommendation letter and B3 is the event a weak recommendation letter, so it is given that probability of B1=0.6, probability of B2=0.3, probability of B3=0.1, also it is given that probability of A given B1=0.9, probability of A given B2=0.6, and probability of A given B3= 0.1.

(Refer Slide Time: 48:34)



We are asked to find out what is the probability of B1 given A that is the probability that letter was a strong given that she got the job, so if we apply the Bayes theorem here probability of A given B1*probability of B1/summation probability of A given Bj*probability of Bj j= 1 to 3, so this is = 0.9*0.6/0.9*0.6+0.6*0.3+0.1*0.1, so if we evaluate this turns out to be simply 54/73 or 0.7397. So there is a 74% of chance that actually the letter was a strong if we already know that she got a job here.

Now let me also introduce the concept of independence of events, so here we have seen that the conditioning alters the original probabilities let us look at here, for example what is the probability of B1 that is 0.6, but if I consider conditioning by A turns out to be 0.7397 okay. In a similar way we can look at the effect on probability of B2 and so on. Similarly, in one of the previous problems if you look at probability of A was actually 0.25 that means the person is a regular drunker.

But if you know that he died due to liver disease then what is the probability that he was a regular drunker, then that probability is quite high 2/3, so you can see that the effect of conditioning changes the original probability. Now there maybe cases where the conditioning does not change it can happen if the 2 events are quite unrelated, for example if you consider tossing of a coin and tossing of a dice.

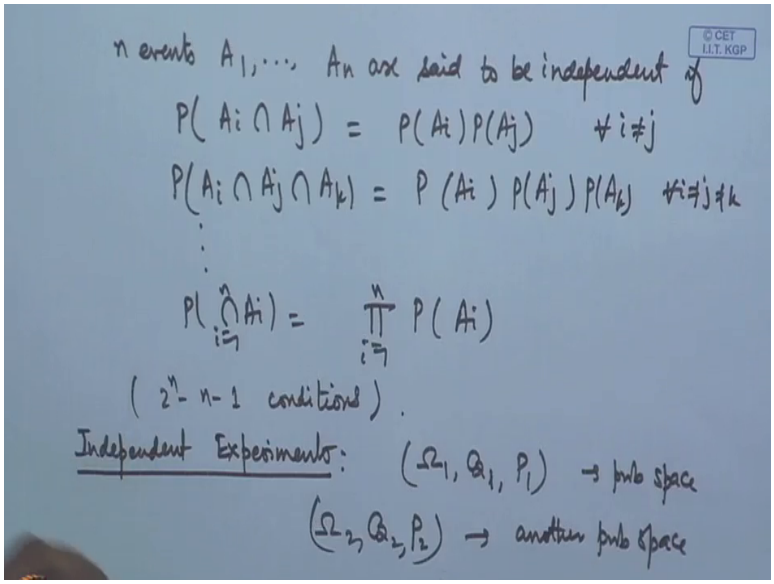
Then what is outcome of the tossing of a coin may not have anything to do with tossing of a that is outcome of a tossing of a dice, so 2 events are totally independent, so using this we define the concept of independence as follows. So as you can see here that we are considering if the event B has no effect on happening of A, then A and B will be independent or you can say A will be independent of B, so we can write like this that probability of A given B= probability of B.

Then we will say that A is independent of happening of B, however, if we write it in elaborate fashion what does it mean? It means probability of A intersection B/probability of B, sorry this is probability of A, = probability of A so this means probability of A intersection B= probability of A*probability of B, now this definition is more symmetric because if we say A is independent of B, we should also say that B is independent of A.

So we define then we say events A and B are independent if probability of A intersection B= probability of A*probability of B, now immediately one can look for the generalization of this for example if we have 3 events A, B, C are independent, then we should have probability of A intersection B= probability of A*probability of B, we should have probability of B intersection C= probability of B* probability of C, probability of C intersection= probability of C*probability of A and probability of A intersection B intersection C= probability of A*probability of B* probability of C.

Actually the condition when we take 2 events at a time then that they are called pairwise independent, and when all of them are satisfied then it is called mutually independent.

(Refer Slide Time: 53:49)



So in general we can define n events A1, A2, An are said to be independent, if I consider every pair A intersection Aj that is= probability of Ai* probability of Aj for all i != j, if we consider probability of Ai intersection Aj intersection Ak that should be= probability of Ai*probability of Aj* probability of Ak, i, j, k must != and so on, probability of intersection Ai i=1 to n must be= the product of all the probabilities, basically these are 2 to the power n-n-1 conditions.

So when this has 2 types of uses, one is that we have a complex experiment and we may like to check whether the events are independent. On the other hand, we consider independent events and then using them we have a complex event for which you want to find out the probability, and then we can make use of the concept of independence in actual calculation of the probabilities of those events. We also consider what is known as independent experiments.

So here we have say omega 1, B1, P1 as one probability space, and another probability space is omega 2, B2, P2 this is another probability space. So for example it could mean that we are considering say we are looking at say longevity of persons in a particular country say India, and then we consider the longevity of persons in say a country in Europe say for example France, and in that case if you have considered sampling of persons in India and in France then that is considered to be independent.

That means the longevity of persons in India and then we calculate various kind of events what is the probability that the age average age or the average longevity <50, >50, between 55 to 60 and so on in India, it will have nothing to do with the similar kind of events in the another countries say France. So this is independent experiments and certainly when we want to join the events in India and France then we can make a complex event.

But by using the independence we can resolve them and find out the probabilities of the corresponding events there. So we have considered the various rules of probability, there are certain other things also for example limiting probabilities, if we have a sequence of events and we consider say limiting probability, what is the probability of the limit of the sequence of subsets?

Then under certain conditions it can be prove that it will be = limit of the probabilities of that sequence, so probabilities satisfies various nice properties, for example we have seen negativity, general addition rule, multiplication rule, total probability, limiting probability and so on. In the following lectures, I will introduce 2 random variables and their distribution, and we will move over to a special discrete and continuous distributions.