Probability and Statistics

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

Module No. #01

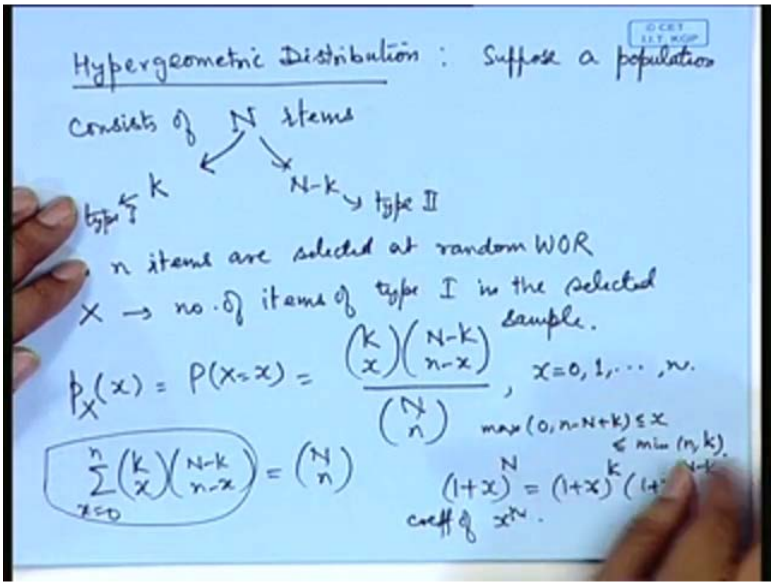
Lecture No. #11

Special Distributions-II

In the Bernoullian trials that we have considered, we are assuming that each trial has been conducted in identical conditions; that means the probability of success or failure remains unchanged in each trial. However, in the finite sampling situations, this type of probability being constant is not applicable. Suppose, there are five people out of which three are say males and two are females, the probability that in the first trial, we select a person, then the probability that it is a male is 3 by 5.

Now, suppose in the first trial a male has been selected, and then now, we are left with two males and two females. And therefore, the probability that a male is selected in the second trial, it will be 2 by 4 that is half, so it is, it has changed, it is not the same. Now, to describe the distribution of number of successes or probability of selecting of one type of persons or one type of items in finite population models is described by hyper geometric distribution.

(Refer Slide Time: 01:31)



So, let us consider this hyper geometric distribution. Suppose, a population consists of N items, N items, N persons or N characteristics, now out of this, k things are of say type I and remaining N minus k things are of type II. So, you can think of like in a population how many males and how many females are there, how many are smokers, how many are non-smokers, how many are infected with a certain disease, how many are not infected with a certain disease, how many are of a particular educational level, how many are below that educational level, etcetera.

So, like the Bernoullian trials, where we are looking at everything with only a two sided reason that success or failure, here also the same thing is there, but the model is that we are considering finite population model. Now, out of this n items are selected at random, so here Bernoullian model will be applicable if we select one item, keep it back that is with replacement, then select another one, then the probability of success or failure in each trial will be same.

However, here we do without replacement; that means, all of them are taken and that means, one is taken and then it is not kept back. So, the probability of selecting an item of type I or type II changes at the next stage.

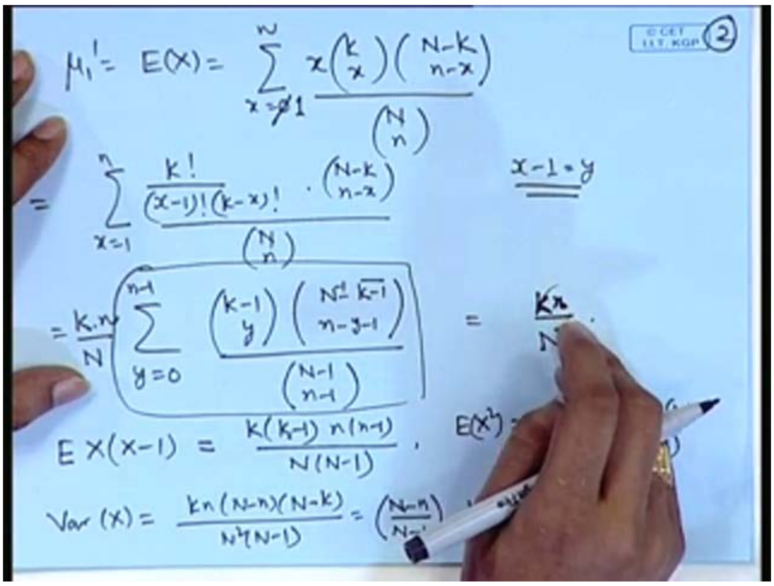
So, X is the number of items of type I in the selected sample, then what is probability of X is equal to say small x that is, so it is equal to k c x N minus k c n minus x divided by N c n. Obviously, x will take values from 0, 1 to N, however in order to write down this expression, this binomial coefficients must be valid; that means, x must lie between 0 and k, n minus x must lie between 0 to N minus k, etcetera. Therefore, the actual restriction for x will be from maximum of 0 to n minus N plus k to minimum of n and k.

So, x takes integral values 0, 1 to N where x lies between this two. So, this is called hyper geometric distribution, see the fact that sigma k c x N minus k c n minus x, x equal to 0 to n under this restrictions, it is equal to N c n, this can be proved by considering the expansion of 1 plus x to the power N in to (()), we can express it as 1 plus x to the power k and 1 plus x to the power N minus k.

So, if we look at the coefficient of x to the power small n in this expansion, then in this one we will get N c n and on this side, since it is product of two terms x to the power N can be obtained as x to the power 0 and to x to the power N, x to the power 1 into x to the power n minus 1 and so on. So, if we collect all the coefficients, we will get these terms, therefore the sum of this is equal to 1 and it is a valued probability distribution.

Let us look at the moment structure of these distribution, since the factorials are involved like in the binomial distribution, if we want to calculate say expectation x. Now, here the x into something will come, therefore this factorial has to be adjusted, therefore it suggest that like in the binomial distribution, the moments of the hyper geometric distribution can also be calculated easily using the factorial moment structure.

(Refer Slide Time: 06:42)



So, we may consider mu 1prime that is expectation x that is sigma x k c x N minus k c n minus x divided by N c n, x from 0 to N and as you can see the corresponding to x equal to 0, this term vanishes. So, this will be from 1 to N and then if we expand in the factorials, this can be written as k factorial divided by x minus 1 factorial k minus x factorial, N minus k c n minus x divided by N c n. Now, in order to adjust the terms, we may put x minus 1 is equal to say y, so this becomes sigma y is equal to 0 to n minus 1.

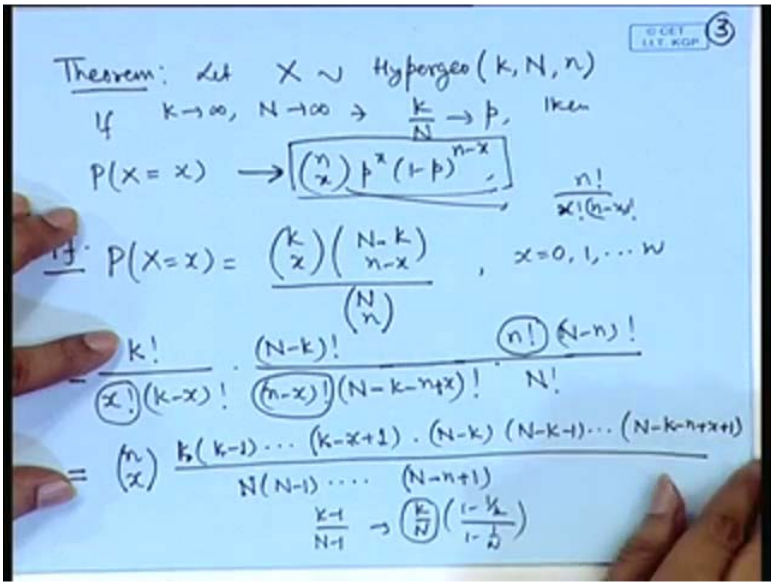
Now, accordingly all other terms have to be adjusted, therefore we can write it as k minus 1 c y N minus k c n minus, so we can write it as N minus 1 minus k minus 1 divided by N minus y minus 1 divided by N minus 1 c n minus 1. So, the terms that have been adjusted will give us k n by N. So, this sum is 1, therefore you get k N by n as the expected value of X.

Now, if we explain this, there are n objects out of which k objects are of type I, so the proportion of the type I objects is k by 1. So, in n draws, what is the expected number of type I that will be n into the proportion? So, in a similar way if we want to calculate say expectation X square, we will need to calculate the second factorial moment, doing the calculations in a similar way, this one will become X is equal to 2 to N. So, we adjust the terms in the form k minus 2 c y where y will be from 0 to n minus 2, the final expansion will give us k into k minus 1, n into n minus 1 divided by N into N minus 1.

This gives us expectation X square which is equal to expectation of X into X minus 1 plus expectation X as k n into k n minus k minus n plus N divided by N into N minus 1, therefore variance of x will become equal to k n N minus n, n minus k n square into N minus 1 is equal to N minus n by N minus 1, k n by N into 1 minus k by N. Notice here that, if we consider large population where N becomes large, k becomes large. So, such that k by N goes to a fixed proportion say p, then this converges to n p. Similarly, here this will converge to n p 1 minus p; this will go to 1 that is going to n p q. So, the mean and variance of the hyper geometric distribution will converge to the mean and variance of a binomial distribution.

This suggests a wider phenomena that is in a hyper geometric distribution, if the population size is considered to be infinitely large such that the proportion of type I items is p, then if we are considering a sample of size n, then the number of type I items must follow a binomial distribution.

(Refer Slide Time: 11:07)



So, this fact can be theoretically proved, let X follow a hyper geometric distribution with parameters k N and n. If k tends to infinity, N tends to infinity such that k by N turns to p, then probability of X is equal to say x converges to n c n p to the power x 1 minus p to the power n minus x which is the binomial probability distribution.

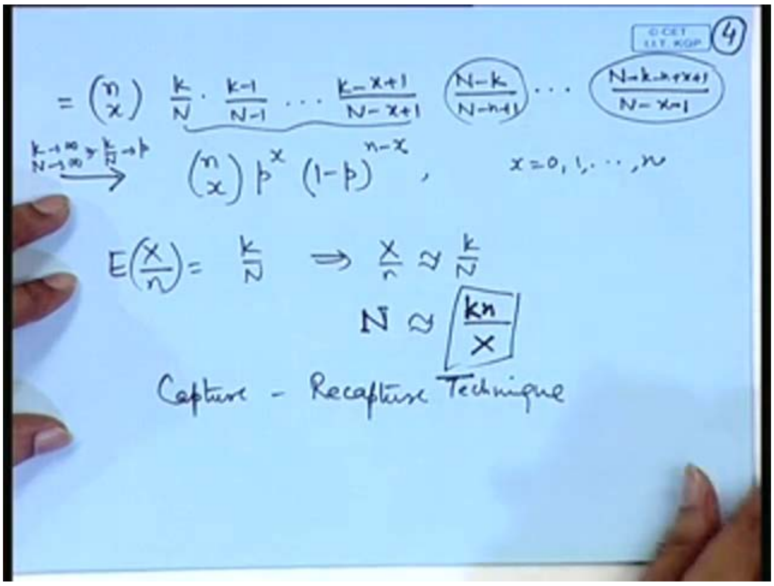
So, let us look at the proof, so in a hyper geometric distribution, probability that X equal to x is written as k c x n minus k c n minus x divided by N c n, here the values of x are from 0, 1 to N subject to a restriction that x is between maximum of 0 n minus N plus k to minimum of n k (Refer Slide Time: 12:24). Now, here if you are assuming that N and k are infinitely large where k by N is a fixed proportion p, then naturally the term k minus N is going to be infinitely large and negative, therefore n minus that will be negative, therefore here maximum will be 0.

Similarly, as k tends to infinity, n will become the minimum value; therefore effectively speaking the range of x is from 0 to N. Now, if we want to take the limit of this as k tends to infinity and N tends to infinity such that k by N tends to a fix number, look at the expansion of these binomial coefficients, because here we cannot take the limit. So, let us look at this, it is equal to k factorial divided by x factorial k minus x factorial N minus k factorial divided by n minus x factorial N minus k minus n minus, so this becomes plus factorial divided by N c n. So, this is N factorial n factorial N minus n factorial, if we want to take the limit as capital N tends to infinity, k tends to infinity, etcetera, we have to further simplify these coefficients (Refer Slide Time: 13:44).

Now, here if we look at the term in the binomial probability mass function n c x is coming which is having n factorial x factorial and n minus x factorial. So, naturally these terms can be seen here, n factorial, x factorial and n minus x factorial. So, this term is readily available here, now this k c x (Refer Slide Time: 12:28). So, this term we can write as k into k minus 1 up to k minus x plus 1 into k minus x factorial. So, this becomes k into k minus 1 up to k minus x plus 1.

The second term is N minus k N minus k minus 1 and so on, up to N minus k minus n plus x plus 1. And in the denominator, N factorial can be written as N into n minus 1 and so on, up to N minus n plus 1. Notice here, k k minus 1 up to k minus x plus 1,these are x terms and here if we look at the first term here, k in the numerator and denominator it is k by N which converges to p. So, likewise if you look at k minus 1 by N minus 1, this can be written as k by N 1 minus 1 by k divided by 1 minus 1 by n; as k tends to infinity, N tends to infinity, this goes to 1 and k by N tends to p. So, in a similar way, we can consider up to k minus x plus 1.

(Refer Slide Time: 16:10)



So, after adjustment of the terms, we can see that this is equal to m c x k by N k minus 1 by N minus 1 and so on, up to k minus x plus 1 divided by N minus x plus 1. Now, x is up to N, therefore this term is coming, N minus x plus 1 is coming somewhere here, N minus n plus 1 will come afterwards. So, there are N minus x terms still left to work which we need to adjust with other terms. So, for example, here next term will become k minus N minus x N minus x minus 1 and so on, up to N minus n plus 1. Now, once again we need to adjust these terms (Refer Slide Time: 16:57), so how will we adjust?

If we look at N minus k divided by N minus n plus 1, then this is equal to if we take common n here, this is 1 minus k by N divided by 1 minus n minus 1 by N. So, as n tends to infinity, this goes to 1 minus p, n minus 1 by capital N goes to 1, this goes to 0. So, 1 minus this goes to 1, therefore we can adjust the terms like N minus k divided by N minus n plus one and so on up to N minus k minus n plus x plus 1 divided by N minus x minus.

So, now if I take limit as k tends to infinity N tends to infinity such that k by N tends to p, then this is converging to n c x, these terms as we explained, each of them converge to p, these are x terms. So, these goes to p to the power x and these are N minus x terms, each of them are going to 1 minus p. So, these goes to 1 minus p to the power n minus x which is nothing but, the probability mass function of a binomial distribution.

So, this finite population sampling, in case the population size is large is same as a binomial sampling. Some peculiar application of the hyper geometric distribution are as follows, suppose we want to estimate the number of tigers in a reserve forest, number of suppose we want to estimate the number of fish in a fresh water lake. For example, a a fishing company wants to estimate the number of fish which are available in a particular area of the lake.

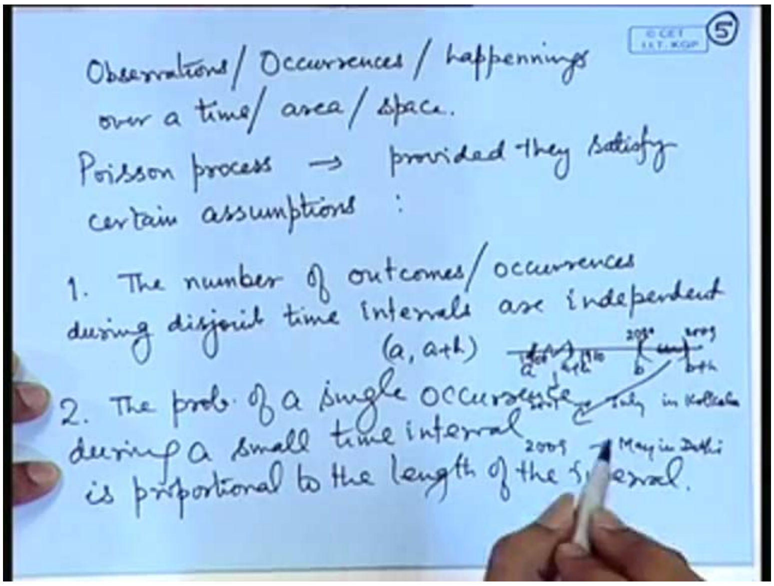
Now, it is impossible to take out the water or go inside the lake and check how many fish area there, one can calculate this by certain sampling procedure. So, we may take a sample of k fish from the, suppose the total number of fish are capital N. So, we take a sample of small k fish and tag them and float them back in the lake. Now, you take a sample of size small n and see how many of them are tag which is we call x, then expectation of X by n is equal to k by N as we proved just now. So, this means that X by n is approximated by k by N, so N can be approximated by K n by X.

Since, we have considered sample of capital X, the initial k type of fish were tagged and small n is also known to us, this number is known to us. So, we can estimate the number of fish in the pond or in the lake of a particular type this is known as capture-recapture technique.

Next, let us consider the phenomenon of the type where we are considering occurrence of events at shorter duration. For example, we will look at a telephone operators desk, how many telephone calls are received during half an hour period. Suppose, we are looking at a traffic at particular crossing, how many vehicles of a particular time pass, how many accidents occur there during a one hour period, how many people arrive at a ticket counter at a railway reservation counter, how many people enter through a door of a shopping mall during a particular period; if we consider such type of events, they cannot be putted into the success failure model or type I model, type II category, etcetera.

We describe in a slightly different way, so one thing that we can observe here is that we are observing the events happening over a period of time or a certain area for example, if we are looking at number of vehicles passing thorough a certain portion of a road, it could be in space also for example, we may look at the number of say, comets observed in a particular portion of say, universe or a galaxy.

(Refer Slide Time: 22:19)



So, we are looking at observations or occurrences or happenings observed over time, area or space. So, if you are observing over time, it is the length of the time or time interval, area as a specified unit, space as a specified unit.

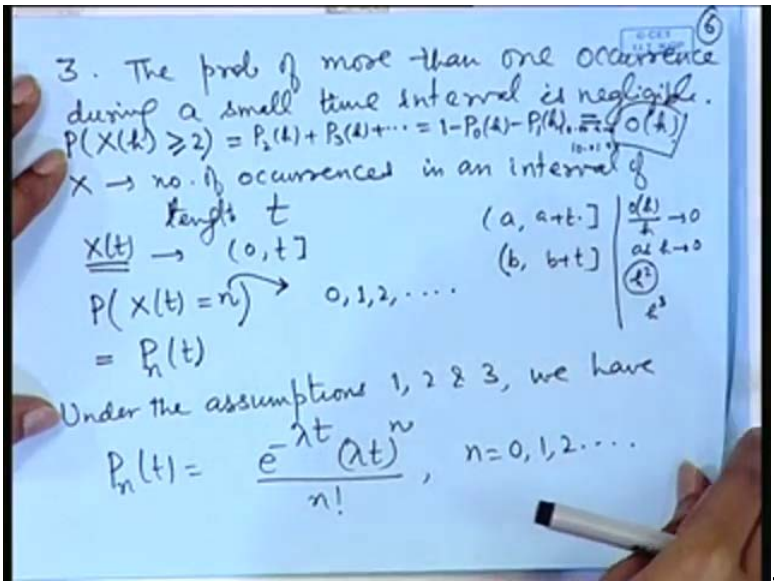
We can consider them as happening under a Poisson process, provided they satisfy certain assumptions provided they satisfy certain assumptions. Let us look at the assumptions, so the number of outcomes or number of occurrences, so for convenience you will consider number of occurrences during disjoint time intervals are independent.

So, for example, we may consider time from a to a plus h, suppose on the time scale, the time a and a plus h are here, suppose the time b to b plus k is here, then how many occurrences are here, or how many occurrences are here, or (( )). For example, the number of earthquakes recorded during say, year 1900 to 1910 is independent of the number of earthquakes recorded during 2000 to 2009.

If we are looking at the number of say, accidents occurred in say year 2001 in month of July in a particular city say, Kolkata and we look at the number of accidents occurring in 2009 in the month of say, May in Delhi. So, this must be independent, this is the first assumption that we are making here, now in place of time interval one may consider area, one may consider space (Refer Slide Time: 24:58). So, that means, different portions of the area, different portions of the space etcetera can be considered.

Now, one words for convenience, we will restrict attention to the phenomena which are observed over time. So, the derivation of the distribution we will do with respect to that, but it is obvious that it can be changed to area or to the space also. The second assumption is that, the probability of a single occurrence during a small time interval is proportional to the length of the interval.

(Refer Slide Time: 26:08)



The third assumption is the probability of more than one occurrence during a small time interval is negligible; let us look at the physical significance of this. So, if we say an earthquake occurs at a moment, say at 10 am on a particular date in a particular place, and then it is highly unlikely that say at 10 past 1 minute past 10 am, again there will be earthquake of the similar nature at the same place.

If we are looking at a train accident at 9 o'clock at a particular place, then it is unlikely that at 9 hours and say 30 seconds, there is another train accident there. So, this probability is negligible and the probability of a single occurrence in a small time interval is proportional to the length of the interval; that means this is something like a rate of occurrence of the event.

If we introduce the notation say x is the number of occurrences in an interval of length t. So, we will denote it by x t, now interval of length t can be considered a to a plus t, it may be considered some b to b plus t, etcetera. So, for convenience, we will consider the interval from 0 to t.

Since, we are making the assumption that the numbers of occurrences in disjoined time intervals are independent, therefore the starting point does not make any difference. So, the number of occurrences in an interval of length t will be denoted by x t, we are interested in probability that x t is equal to some number say n. Obviously, since it is number of occurrences n will be able to take values 0, 1, 2 and so on. So, like a geometric distribution, this is an infinite valued probability distribution and the number of values taken are countably infinite, we will use the notation for probability x t is equal to n as P n t.

So, our aim is to derive an expression for P n t under the assumptions 1, 2 and 3. So, under the assumptions 1, 2 and 3, we have P n t is equal to e to the power minus lambda t, lambda t to the power n by n factorial. Now, before going to the proof of this, let us understand, what is this coefficient lambda coming from.

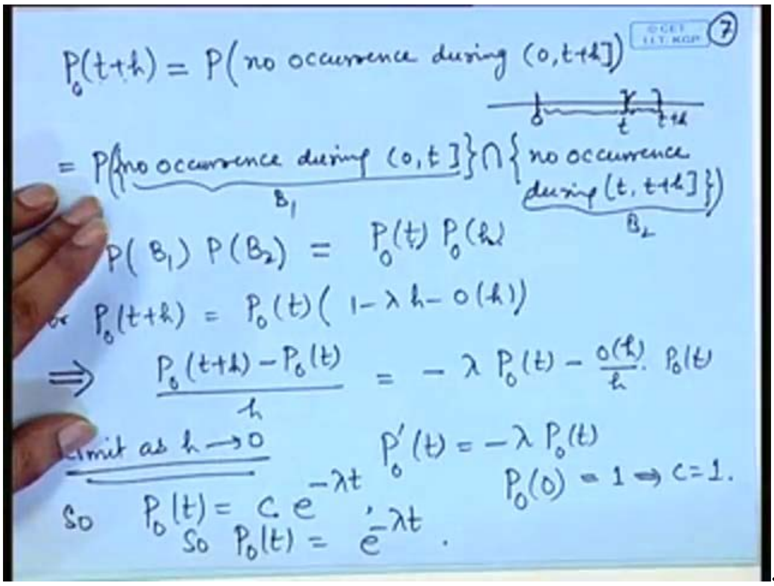
So, we made certain assumption here, the probability of a single occurrence during a small time interval is proportional to the length of the interval. That means, we are saying probability of x during a small time interval which we can right as epsilon or h, probability that there is a single occurrence in the interval of length h is proportional to the length of the interval. So, this is P 1 h is equal to lambda h, so lambda is the constant of proportionality here, this lambda is reflected here in the expression for probability x t is equal to n.

Now, if we make this assumption, then the second, the third assumption can also be expressed in terms of this, probability of more than one occurrence. Now, more than one occurrence means probability of X h small time interval is greater than or equal to 2. That means, it is equal to P 2 h plus P 3 h and so on, which we can write as 1 minus P 0 h minus P 1 h.

So, the assumption is that this is negligible, so by negligibility, we can express it in terms of o h, which is small order h, this small order h is explained by small o h by h goes to 0 as h goes to 0. For example, h square this is an o h term, for example, h cube is an o h term.

To prove this statement which gives the probability distribution of the number of occurrences in the interval 0 to t, let us call it statement number 1. To prove this statement, we can make use of the induction process, now we will firstly prove it for n is equal to 0 and n is equal to 1, and then we will make assumption for n is equal to k and proceed to n is equal to k plus 1 case. To proof for n is equal to 0, we need to prove P 0 t is equal to e to the power minus lambda t.

(Refer Slide Time: 32:35)



So, let us consider the derivation of this P 0 t we consider P 0 t plus h. Now, this means probability of no occurrence during the interval 0 to t plus h. So, if we are looking at the timeline t to t plus h. So, we can consider this as splitting into two portions, 0 to t and t to t plus h if there is no occurrence from 0 to t plus h, it means there is no occurrence here and there is no occurrence here (Refer Slide Time: 33:13). So, this can be represented as probability of no occurrence during 0 to t, if we consider this as an event, then it is probability of simultaneous occurrence of these two events, no occurrence during t to t plus h.

Now, we make use of the first assumption that is the number of outcomes or occurrences during disjoined time intervals are independent. So, if we look at this event say B 1 and this event as say B 2, then events B 1 and B 2 are independent. So, probability of B 1 intersection B 2 will become probability of B 1 into probability of B 2.

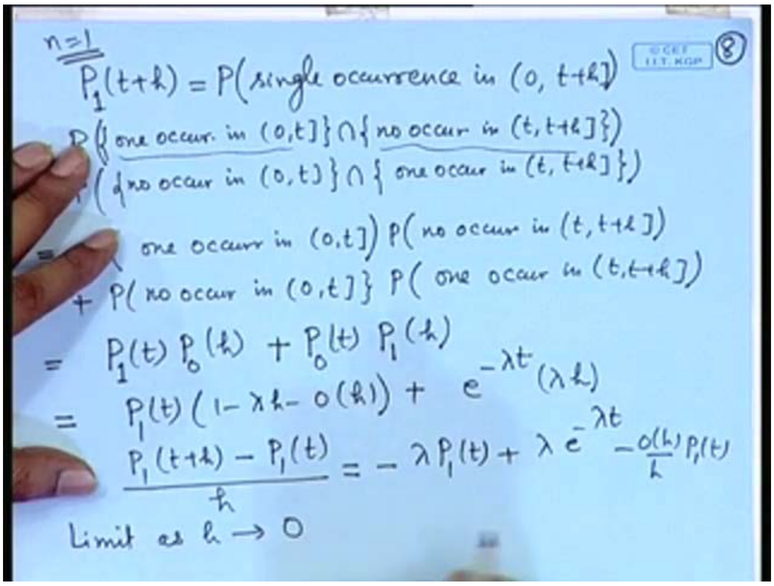
Now, what is probability of B 1, no occurrence during 0 to t. So, if we consider the definition of P n t, it is probability of x t is equal to n, then this probability of B 1 is P 0 t, and similarly probability of B 2 is P 0 h. Because the interval from t to t plus h is of length h, the starting point does not matter, so it is probability of no occurrence during an interval of length h.

So, now we have the expression P 0 t plus h is equal to P 0 t into P 0 h, now we can make use of the assumptions here, we have 1 minus P 0 h minus P 1 h as small o h and P

1 h is equal to lambda h. If we make use of these two statements, we will get P 0 h is equal to 1 minus lambda h minus o h. So, if we substitute this expression here, we get 1 minus lambda h minus o h. So, let us adjust the terms P 0 t plus h minus P 0 t divided by h is equal to minus lambda P 0 t minus o h by h P 0 t.

So, if I take the limit as h tends to 0, then the left hand side is P 0 prime t is equal to minus lambda P 0 t and since o h by h goes to 0, this term vanishes. We are left with a first order linear differential equation for which the solution is P 0 t is equal to c times e to the power minus lambda t where c is a constant of integration. This can be determined by using some initial condition say P 0 0, now P 0 0 stands for the probability of no occurrence in an interval of length 0 which must be 1, this means that c must be 1. So, we get P 0 t is equal to e to the power minus lambda t which is the statement if we consider 1 with n is equal to 0, we will get e to the power minus lambda t, so the statement 1 is true for n is equal to 0.

(Refer Slide Time: 37:27)



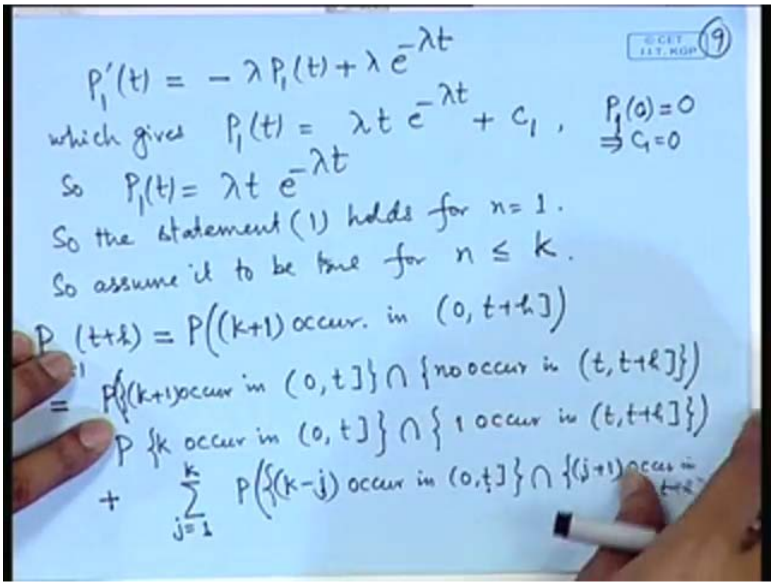
In order to prove the statement 1 using induction, we need to prove it for n is equal to 1. So, for n is equal to 1, let us consider P 1 t plus h. So, P 1 t plus h is probability of single occurrence in the interval 0 to t plus h. So, if we say one occurrence in the interval 0 to t plus h and if we split the interval 0 to t plus h as 0 to t and t to t plus h, then the one occurrence can be either in this interval or in this interval that is 0 to t or t to t plus h.

So, we can express this event as probability of one occurrence in 0 to t and no occurrence in t to t plus h. So, here we are making use of the addition rule or the theorem of total probability, plus probability of no occurrence in 0 to t and one occurrence in t to t plus h. Once again if we make the use of assumption 1, that means, the events occurring in in disjoined time intervals are independent, then the probability of simultaneous occurrence of this and this is equal to the probability of product of this (Refer Slide Time: 39:03). That means, it is equal to probability of one occurrence in the interval 0 to t into probability of no occurrence in t to t plus h plus probability of no occurrence in 0 to t into probability of 1 occurrence in t to t plus h.

So, this is equal to, now if we make use of the notation that P n t is equal to probability that X t is equal to n, then using this probability of one occurrence in interval 0 to t is P 1 t, probability of no occurrence in t to t plus h is P 0 h, probability of no occurrence in 0 to t is P 0 t and probability of one occurrence in t to t plus h is P 1 h.

So, if we substitute the expressions for P 0 h, P 1 h and P 0 t, so P 0 h is 1 minus lambda h minus o h from the assumptions P 0 t, we just now proved as e to the power minus lambda t and P 1 h is lambda h. So, we can express P 1 t plus h minus P 1 t divided by h, this is equal to, in the second term minus lambda h P 1 t is there. So, h will cancel out and we get minus lambda P 1 t and here we get lambda e to the power minus lambda t minus o h by h P 1 t. So, if we again take the limit as h tends to 0, we get the left hand side will yield the derivative of P 1 at t.

(Refer Slide Time: 41:33)



So, we get P 1 prime t is equal to minus lambda P 1 t plus lambda e to the power minus lambda t. If we notice this, this is again a first order linear differential equation and therefore, the solution can be obtained easily, this gives P 1 t is equal to lambda t e to the power minus lambda t plus a constant, the constant of integration can be determined using some initial condition. For example, P 1 0, this means probability of a single occurrence in the interval of length 0 which is obviously 0, so this will give c 1 is equal to 0.

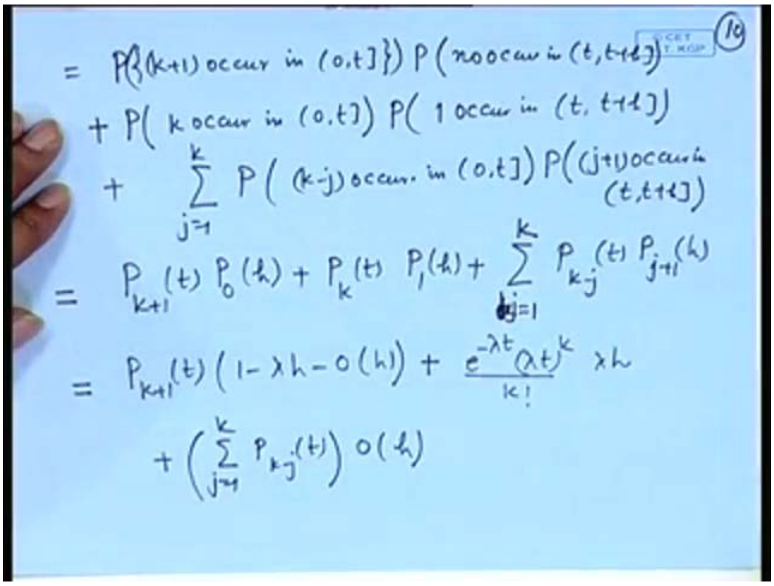
So, P 1 t is equal to lambda t e to the power minus lambda t which is the expression for P n t, if we put n is equal to 1 here that is lambda t into e to the power minus lambda t. So, the statement 1 holds for n is equal to 1, so assume that it to be true for n less than or equal to k. So, now we need to prove for P k plus 1, so P k plus 1 t plus h, this is probability of k plus 1 occurrence in the interval 0 to t plus h. If we look at the breakup here in the interval 0 to t plus h, we have k plus 1 occurrence, now this leads to various possibilities if we have split the interval into 0 to t and t plus h, t to t plus h.

So, all the k plus 1 occurrences maybe in the interval 0 to t, no occurrence in t to t plus h, 1 occurrences in the interval 0 to t or k occurrences in 0 to t and 1 occurrence in t to t plus h k minus 1 occurrences in 0 to t, and 2 occurrences in t to t plus h k minus 2, k minus 3 and so, on, including no occurrence in 0 to t and all k plus 1 occurrences in the interval t to t plus h.

So, we can express this by using the theorem of total probability as probability of k plus 1 occurrences in 0 to t intersection, no occurrences in t to t plus h plus probability of k occurrences in 0 to t, 1 occurrence in t to t plus h, probability of k minus j occurrences in 0 to t and j plus 1 occurrences in t to t plus h, where summation here is from j is equal to 1 to k, this will for j is equal to 1 this gives k minus 1 occurrences in 0 to t and 1 occurrence 2 occurrences in t to t plus h. Here j is equal to 2 will give you k minus 2 occurrences here and 3 occurrences here, j is equal to k will give you no occurrences in 0 to t and k plus 1 occurrences in t to t plus h.

So, we have expressed this event k plus 1 occurrence in 0 to t plus h as the sum of these many probabilities. Once again, we notice here that in individual probability terms here, it is simultaneous occurrence in two disjoined intervals.

(Refer Slide Time: 46:33)

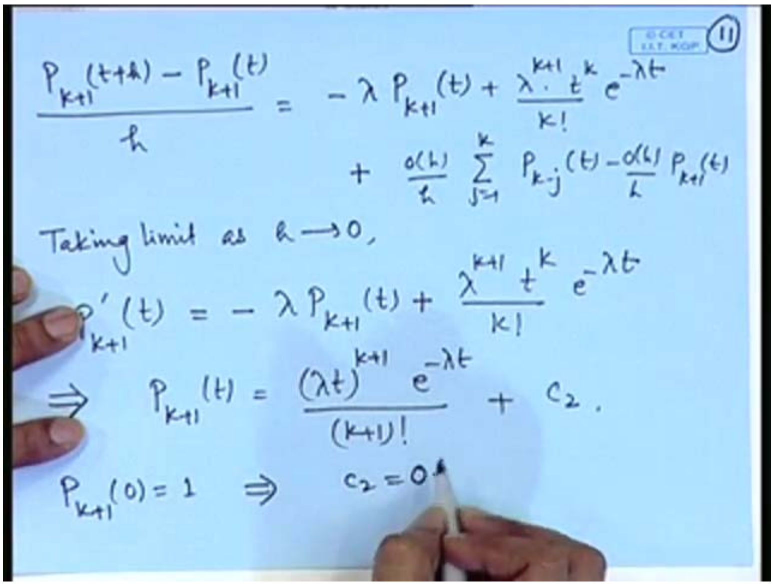


So, by making use of the independence, this becomes probability of k plus 1 occurrences in 0 to t into probability of no occurrence in t to t plus h plus probability of k occurrences in 0 to t into probability of 1 occurrence in t to t plus h, probability of k minus j occurrences in 0 to t into probability of j plus 1 occurrences in t to t plus h.

This is equal to, so the first term here is obviously, P k plus 1 t into P 0 h plus P k t into P 1 h plus sigma P k minus j t P j plus 1 h, j is k equal to j is equal to 1 to k, we substitute the expressions which are known to us. So, this becomes P k plus 1 t, P 0 h is 1 minus lambda h minus o h plus we made the assumption that the statement 1 is true up to k. So, this becomes e to the power minus lambda t lambda t to the power k by k factorial and P 1 h is lambda h plus, now here j is equal to 1, 2 and so on. So, this is at east P to h, now we have made the third assumption in a Poisson process which says that, the probability of more than 1 occurrence is negligible, we had P 2 h plus P 3 h, etcetera as o h.

So, therefore, all the terms here will be of o h order. So, this becomes sigma P k minus j t, of course the expressions for (()) are available as we have assumed that it to be true up to n is equal to k. However, we need not write them into o h, so P k plus 1 t plus h has been expressed as P k plus 1 t into certain term and many other terms.

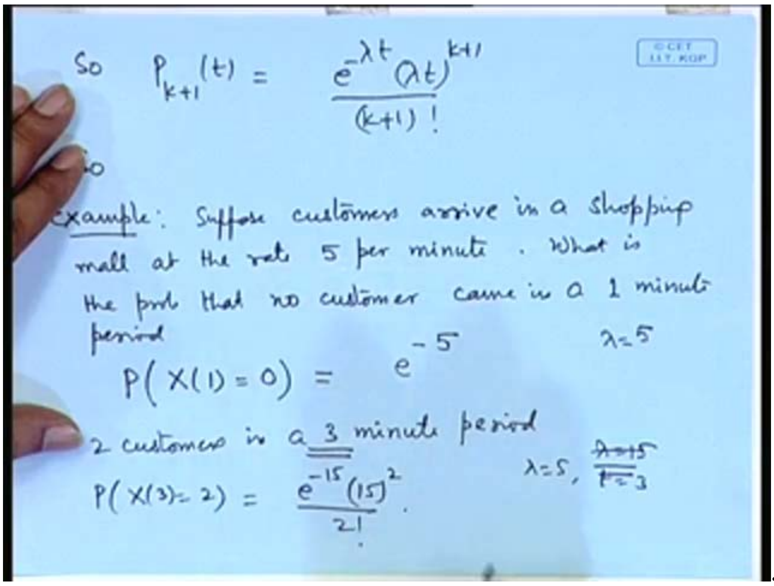
(Refer Slide Time: 49:37)



So, we can go for simplification here, P k plus 1 t plus h minus P k plus 1 t divided by h is equal to minus lambda P k plus 1 t plus lambda to the power k plus 1, t to the power k by k factorial e to the power minus lambda t plus o h by h sigma P k minus j t, j is equal to 1 to k minus o h by h P k plus 1 t.

So, if we take the limit as h tends to 0, we get the left hand side as the derivative of P k plus 1 t is equal to minus lambda P k plus 1 t, this term will be as it is lambda to the power k plus 1 t to the power k divided by k factorial e to the power minus lambda t, the term o h by h goes to 0. So, this entire term goes to 0 and this term also goes to 0, we are again left with a first order linear differential equation and this can be solved. So, we get P k plus 1 t as lambda t to the power k plus 1 e to the power minus lambda t by k plus 1 factorial plus a constant of integration. If we make use of the initial condition say P k plus 1 0 is equal to 1, then this yields c 2 is equal to 0.

(Refer Slide Time: 51:37)



Therefore, the expression for P k plus 1 t is equal to e to the power minus lambda t lambda t to the power k plus 1 by k plus 1 factorial which is the expression for P n t if we put n is equal to k plus 1 here. So, by the principle of mathematical induction, we have proved that the Poisson distribution codes.

So, in a Poisson process, the distribution of the number of occurrences in a fixed time interval is called a Poisson distribution. So, this particular distribution that we obtained, this is known as Poisson distribution and the system that we have described here that is the occurrences in a particular way that is the number of occurrences in disjoined time intervals are independent, the probability of a singular occurrence in a small time interval is equal is proportional to the length of the interval, the probability of more than one occurrence in a small time interval is negligible.

So, this process is called Poisson process and the distribution of the number of occurrences in a fixed time interval during a Poisson process is called a Poisson distribution. In the next lecture, we will be discussing the characteristics of these distributions such as its mean variance, moment generating function, etcetera.

Just to give you an example, suppose customers arrive in a shopping mall at the rate 5 per minute, what is the probability that no customer came in a 1 minute period. So, this means we are asking for probability X equal to 1 is equal to 0. So, here lambda is equal to 5 for unit of time is per minute. So, in a 1 minute period, we are having e to the power minus 5 that is e to the power minus lambda t.

Suppose, we say 2 customers in a 3 minute period, so if we are considering 3 minute period, then lambda will be equal to 15 or lambda t lambda is equal to 5 and t is equal to 3. So, probability of X 3 is equal to 2 that is equal to e to the power minus lambda t lambda t square by 2 factorial. So, in a Poisson process, if we know the rate of the occurrence, then we can find out probabilities of various numbers, for example, probability of a certain number of occurrences, less than a certain number of occurrences, etcetera. So, in the next lecture, we will elaborate further on this, thank you.