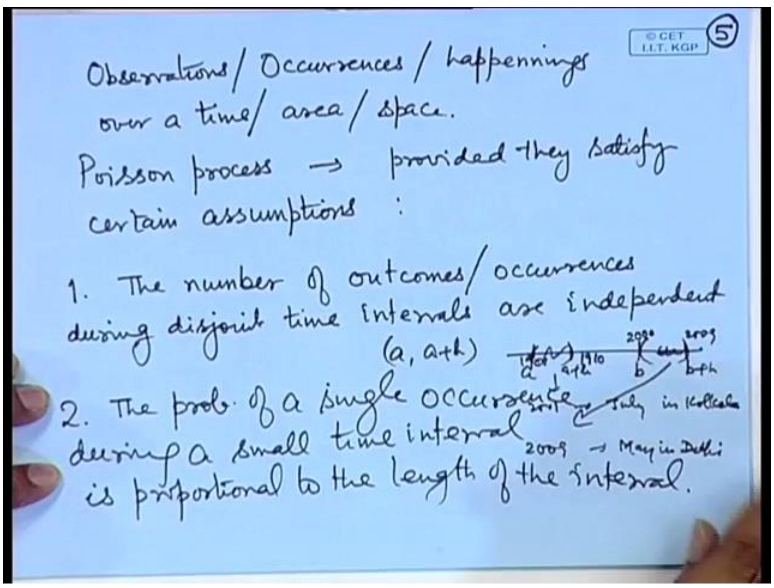
Probability and Statistics
Prof. Somesh Kumar
Department of Mathematics
Indian Institute of Technology, Kharagpur

Lecture - 22
Poisson Process – I

Next let us consider the phenomenon of the type where we are considering occurrence of events at shorter duration. For example, we look at a telephone operator's test how many telephone calls are received during half an hour period. Suppose we are looking at traffic at a particular crossing, how many vehicles of a particular kind 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 inter through a door of a shopping mall during a particular period. If we consider such type of events they cannot be put it into the success failure model or type 1 model type 2 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 period of time or certain area. For example if you are looking at number of vehicles passing through a certain portion of a road; it could be in a space also. For example, we may look at the number of say comments observed in a particular portion of say universe or a galaxy.

(Refer Slide Time: 02:01)



So, we are looking at observations or occurrence or happenings observed over time area or a space. So, if you are observing over time it is a 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 the satisfy certain assumptions. Let us look at the assumptions.

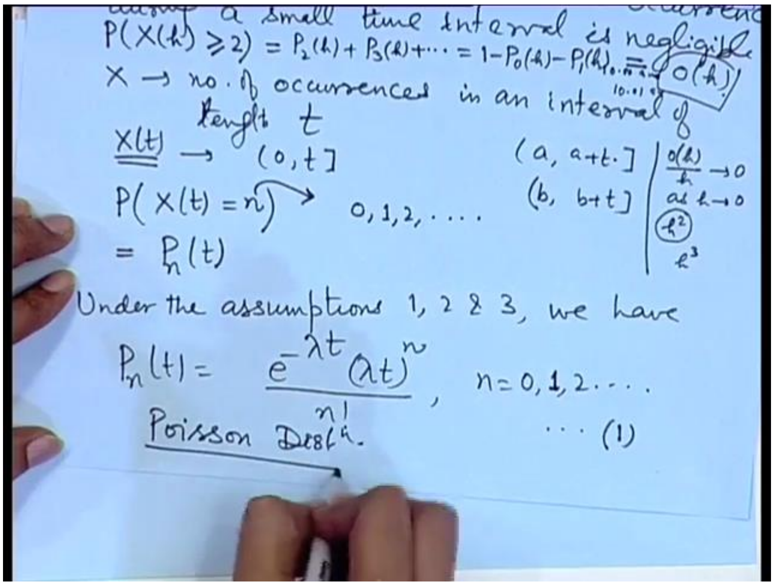
So, the number of outcomes or number of occurrences, so for convenience we will consider number of occurrences during disjoint time intervals are independent. So, for example, we may consider time from a to a plus h; suppose on the time scale the time a and a plus h are here. Suppose the time b to b plus k is here. Then how many occurrences are here or how many occurrences are here are materials?

For example; the numbers of earthquakes recorded during say year 1900 to 1910 is independent of the number of earthquakes recorded during 2002 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, these 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 so that means, different portions of the area different portions of the space etcetera can be considered. Now onwards 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 change 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: 05:51)



The third assumption is; the probability of more than 1 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, then it is highly and likely that say at 10 pass 1 minute pass 10 am and then there will be an 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 at 9 hours and say 30 seconds that 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 number of occurrence in disjoint time intervals are independent, therefore the starting point does not make any difference.

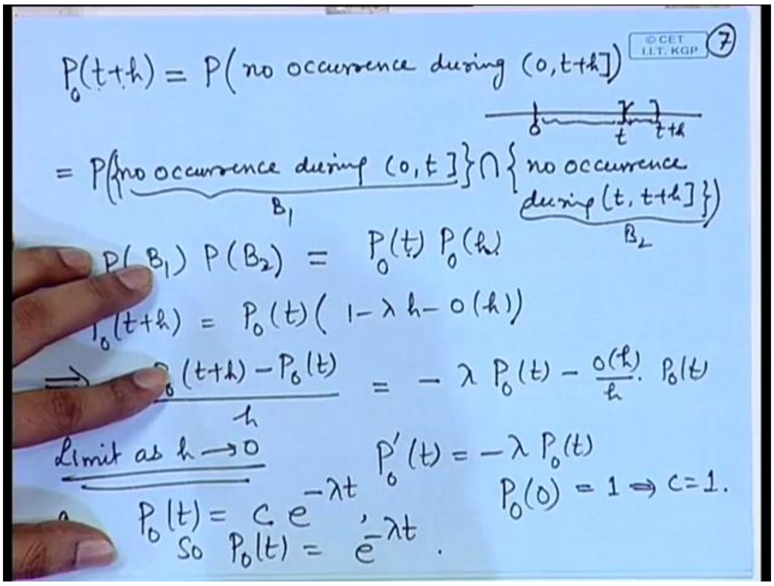
So, the number of occurrence in an interval of length t will be denoted by x t. We are interested in probability that x t is equal to some number say n; obviously, since it is number of occurrences n will be able to take values 0, 1, 2 and so on. So, like a geometric distribution this is an infinite valued probability distribution and the numbers 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 assumptions 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 write 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 a 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 third assumption can also be expressed in terms of this. Probability of more than 1 occurrence; now more than 1 occurrence means probability of x h is 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 a small order h. This is small order h is explained by a 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 proof this statement we can make use of the induction process. Now you will firstly prove it for n is equal to 0 and n is equal and then we will make assumption for n is equal to k and proceed to n is equal to k plus 1 ks.

(Refer Slide Time: 12:19)



To proof for n is equal to 0 we need to proof P 0 t is equal to e to the power minus lambda t. 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 time line t to t plus h. So, we can consider this is 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. So, this can be represented as probability of no occurrence during 0 to t, if we consider this as an even then it is probability of simultaneous occurrence of this two events; no occurrence during t to t plus h.

Now, we make use of the first assumption that is a number of outcomes or occurrences during disjoint 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 in to 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 and 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 occurrences during an interval of length h.

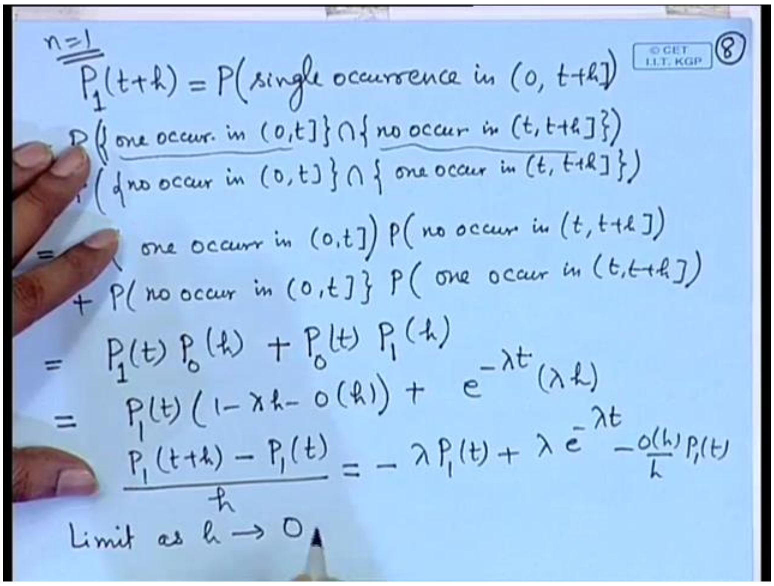
So now, we have the expression P 0 t plus h is equal to P 0 t into P 0 h. Now, we can make use of the assumptions here. We have 1 minus P 0 h minus P 1 h as a small o h and P 1 h is equal to lambda h. If we make use of these two statements we will get P 0 h is equal to 1 minus lambda h minus o h. So, 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 Itake 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 is 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 t to the power minus lambda t. So, the statement 1 is true for n is equal to 0.

In order to prove the statement 1 using induction we need to proof it for n is equal to 1.

(Refer Slide Time: 17:10)



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 1 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 1 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 1 occurrence in 0 to t and no occurrence in t to t plus h. So, here we are making using of the addition rule or the theorem of total probability plus probability of no occurrence in 0 to t and 1 occurrence in t to t plus h.

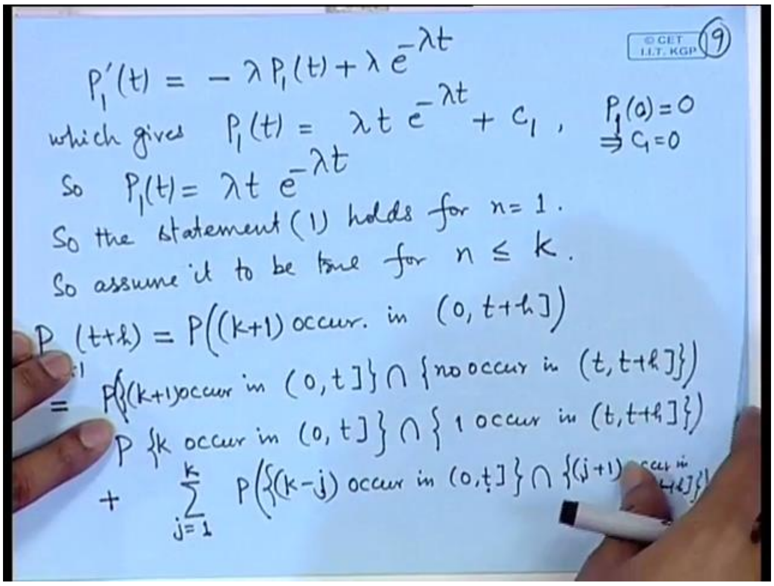
Once again if we make the use of assumption 1; that means, the events occurring in disjoint time intervals are independent then the probability of simultaneous occurrence of this and this is equal to the probability of product of this. That means, it is equal to probability of 1 occurrence in the interval 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 and t is equal to probability that x t is equal to n, then using this probability of 1 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 1 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: 21:17)



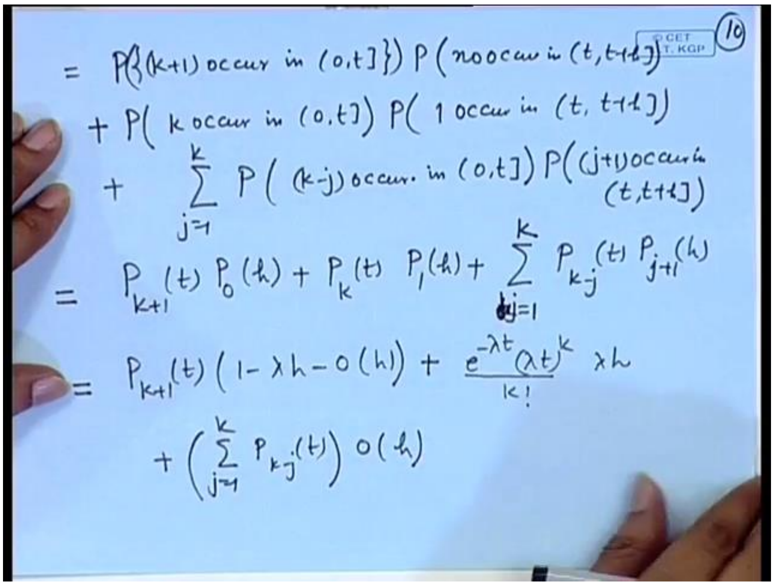
So, we get P 1 prime t is equal to minus lambda P 1 t plus lambda e to the power minus lambda t. If we notice this is again a first order linear differential equation, and therefore the solution can be obtained easily. This gives P 1 t is equal to lambda t e to the power minus lambda t plus a constant. 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 and t if we put n is equal to 1 here that is lambda t into e to the power minus lambda t. So, by 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 proof 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 break up here in the interval 0 to t plus h we have k plus 1 occurrences, now this leads to various possibilities if we have a split the interval into 0 to t and t plus h; t to t plus h. So, all the k plus 1 occurrences may be in the interval 0 to t, no occurrence in t to t plus h, 1 occurrence 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 minus 1 occurrences in the interval t to t plus h.

So, we can express this by using a theorem of total probability of 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 2 occurrence in t to t plus h here j is equal to 2 will give you k minus 2 occurrences here and three occurrences here, j is equal to k will give 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 occurrences 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 disjoint intervals.

(Refer Slide Time: 26:14)

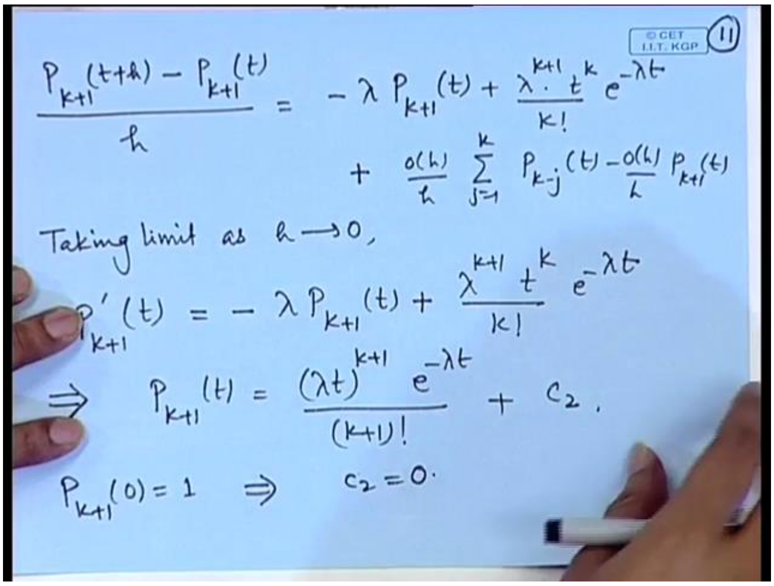


So, by making use of the independence this becomes probability of k plus 1 occurrences in 0 to t into probability of no occurrence in t to t plus h plus probability of k occurrences in 0 to t into probability of 1 occurrence in t to t plus h probability of k minus j occurrences in 0 to t in two probability 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 t k t into P 1 h plus sigma P k minus j t P j plus 1 h, 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 a 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 least 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 o h order. So, this become sigma P k minus j t, of course the expressions (Refer Time: 28:55) 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, t k plus 1 t plus has been expressed as t k plus 1 t into certain term and many other terms.

(Refer Slide Time: 29:21)

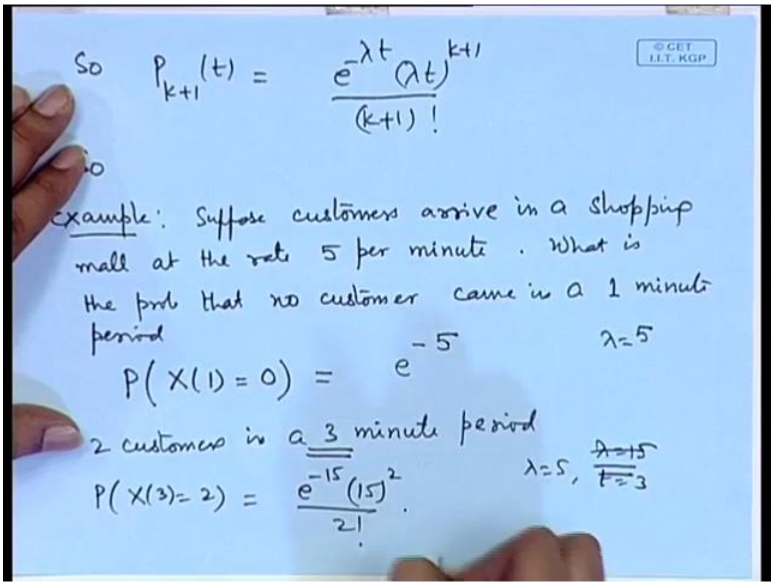


So, we can go for simplification here; t k plus 1 t plus h minus t k plus 1 t divided by h is equal to minus lambda P k plus 1 t plus lambda to the power k plus 1 t to the power k by k factorial e to the power minus lambda t; plus o h by h sigma P k minus j t j is equal to 1 to k minus o h by h P k plus 1 t. So, 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 terms 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.You 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 equals to 0.

(Refer Slide Time: 31:20)



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 and 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 holds.

So, in a Poisson process the distribution of the member of occurrences in a fixed time interval is called a Poisson distribution. So, this particular distribution that we obtain 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 disjoint time intervals are independent, the probability of a single occurrences in a small time interval is equal is proportional to the length of the interval, the probability of more than 1 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 fix time interval during a Poisson process is called Poisson distribution.

In the next lecture we will be discussing the characteristics of this distribution such as its mean variance, moment genetic function etcetera. Now 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 two customers in a 3 minute period; so if you 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 three. 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.