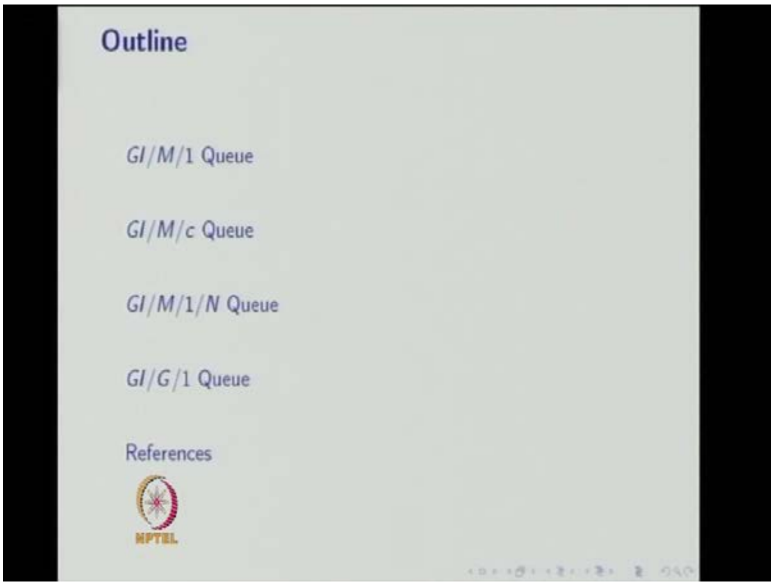
Stochastic Processes
Prof. Dr. S. Dharmaraja
Department of Mathematics
Indian Institute of Technology, Delhi

Module - 8
Renewal Processes
Lecture - 5
Non Markovian Queues

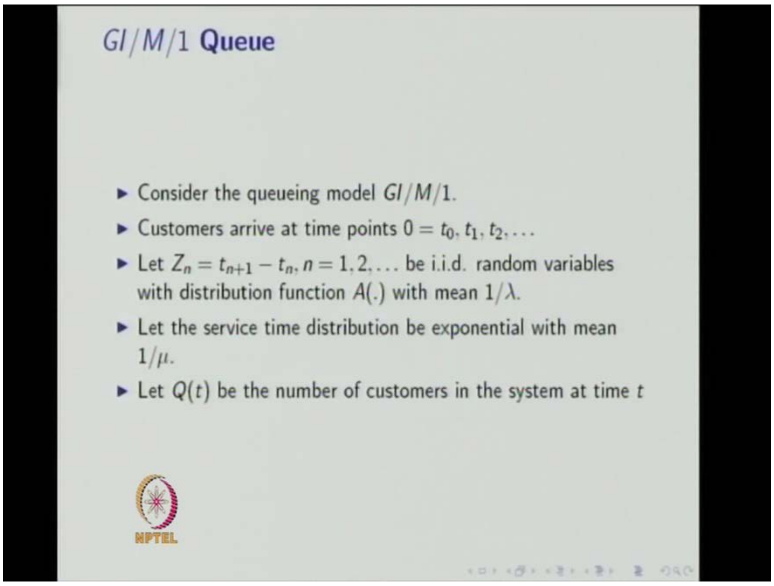
This is a stochastic processes module 8, Renewal process. In the lecture 1, we have discussed the renewal function and renewal equation. In the lecture 2, we have discussed the generalize renewal processes and renewal limit theorems. In lecture 3, we have covered Markov renewal and regenerative processes. In lecture 4, we have discussed non Markovian queues such as M G 1 queue, M G 1 and queue, M G C C (( )) systems.

(Refer Slide Time: 01:20)



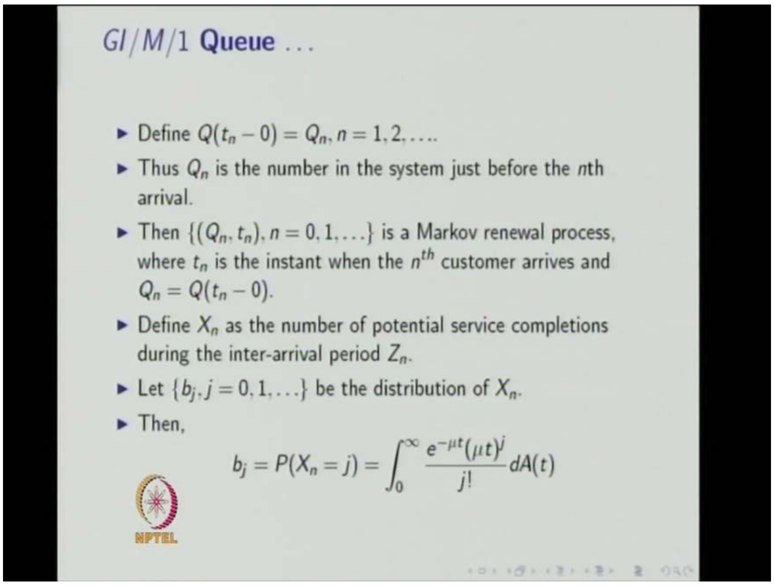
This is a lecture-5, non Markovian queues, in this lecture we are going to discuss GI/M/1 Queue, GI/M/c Queue then GI/M/1/N Queue and finally, GI/G/1 Queue. What is GI/M/1 Queue? It means that the inter arrival time follows non-exponential distribution, which are independent. Therefore, the GI some books, they use only G as a notation; M stands for the service time is a exponential distribution only one server in the system with infinite capacity. So, consider the customers arrive at a time points t 0, t 1, t 2 and so on. Let Z n is equal to t n plus 1 minus t n; be the i.i.d.; random variables with the distribution function with the CDFA with the mean 1 by lambda.

(Refer Slide Time: 01:34)



Therefore, as a special case if you assume that the inter arrival time is exponential distribution with the mean, 1 by lambda. Then, the arrival follows a poisson process with the parameter lambda. But in this G/M/1 model the inter arrival time is non exponential distribution with the CDF function, A with the mean 1 by lambda. Let the service time distribution be exponential with the mean 1 by mu. Let Q (t) be the number of customers in the system at time t.

(Refer Slide Time: 03:56)



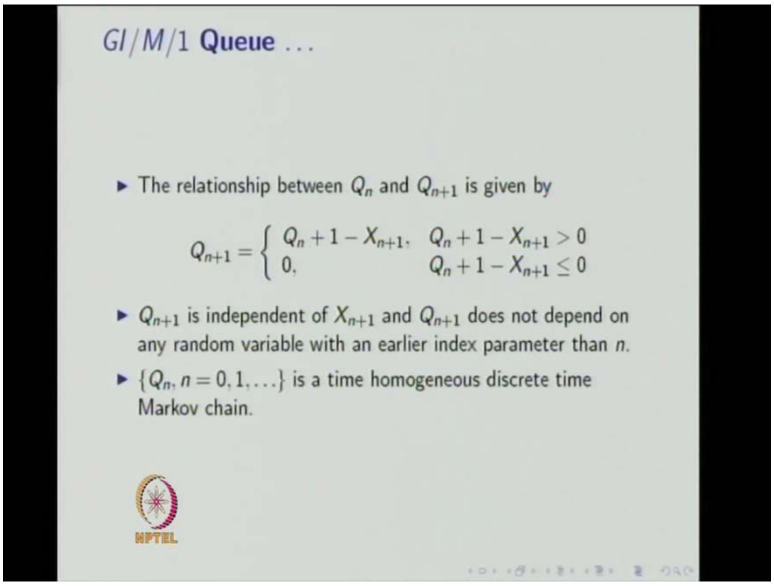
So, Q (t) for t greater than or equal to 0 is a stochastic process. Since, the Q (t) is the number of customers in the system at any time t. Therefore, the corresponding stochastic process is a discrete state continuous time stochastic process. So, the underlying stochastic process in the G/M/1 Queue is a Q (t), t greater than or equal to 0.

Now, define Q (t n minus 0) has a Q n. Thus, Q n is a number of customers in the system just before the nth arrival. Therefore, the Q n for n is equal to 1, 2 and so on. This follows a discrete state, discrete time stochastic process. So, this is a embedded stochastic process from Q (t), in the Q (t) is a discrete state continuous time stochastic process. Whereas, Q (n) is a discrete state discrete time stochastic process because the Q n is the number of customers in the system just before the n th arrival. The bivariate random variables Q n, t n for different values of n forms a Markov renewal process; where, tn is the instant when the nth customers arrives and Q n is defined Q t n minus 0.

Since, inter arrival time is a non exponential distribution, with the mean 1 by lambda and the service time is exponential distribution with the mean 1 by mu. Single server in the system and infinite capacity. Therefore, the Q n, t n form a Markov renewal process and the t n is a time instant in which the arrival occurs. Now, define the random variable X n as a number of potential service completions during the inter-arrival period Z n. Z n is nothing but t n plus 1 minus t n; that is a inter arrival time. X n is a number of potential service completions; during the inter arrival period Z n and b j be the distribution of X n. Obviously, X n is the discrete type random variable for fixed n and the b j is a probability mass function for the random variable X n.

Since, the number of potential service completion could be 0, 1 and so on. So, the probability mass function for different values of j, b j is nothing but the probability of X n is equal to j that is nothing but the integration 0 to infinity e power minus mu t; mu t power j divided by j factorial and the integration with respect to A (t); where A (t) is the distribution function of inter arrival time with the mean 1 by lambda.

(Refer Slide Time: 07:44)

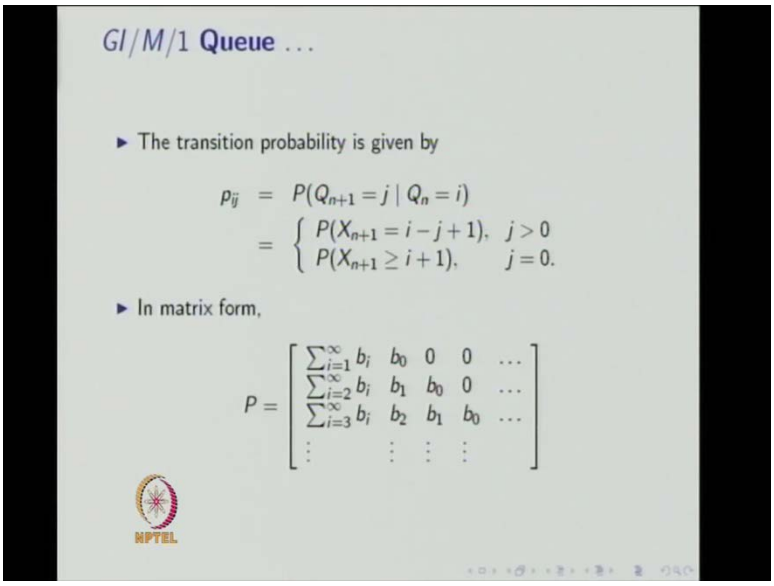


And, mu is the parameter for exponential distribution of service time. The way we defined Q n is nothing but the number of customers in the system just before the nth arrival 1 can relate the Q n plus 1 in terms of Q n. We can find the relationship between Q n and Q n plus 1, that is Q n plus 1; will be Q n plus 1 minus X n plus 1; whenever Q n plus 1 minus X n plus 1 is greater than 0. Otherwise, it is 0, for Q n plus 1 minus X n plus 1 is less than 0. The reason is the number of customers in the system just before the nth arrival n plus 1 eth arrival is same as the number of customers in the system just before the nth arrival plus the n plus 1 eth; customer who arrived that is plus 1 minus. How many customers are served during the inter arrival period? That is a X n plus 1.

So, if you subtract Q n plus 1; the X n plus 1; you will get Q n plus 1. Whenever, Q n plus 1 minus X n plus 1 is greater than 0; if it is less than or equal to 0; then the number of customers will be again 0. Just before the n plus 1 eth arrival also will be 0. We know that the X n plus 1 is independent of Q n plus 1. Hence, the Q n plus 1 depends only on Q n and Q n plus 1 is independent of X n plus 1. Therefore, the Q n forms a time homogeneous discrete time Markov chain. The Q t is a discrete state continuous time stochastic process and the Q n is the discrete time discrete state stochastic process.

Since, Q n plus 1 is equal to Q n plus 1 minus X n plus 1 or 0; and Q n plus 1 is independent of X n plus 1; as well as Q n plus 1 depends only on Q n. Therefore, Q n for n is equal to 0, 1, 2 and so on, form a time homogeneous. That means, time invariant and we also satisfy the Markov property. Therefore, this discrete time discrete state stochastic process is called a discrete time Markov chain satisfying the time homogeneous property. Therefore, it is called a time homogeneous discrete time Markov chain.

(Refer Slide Time: 10:46)

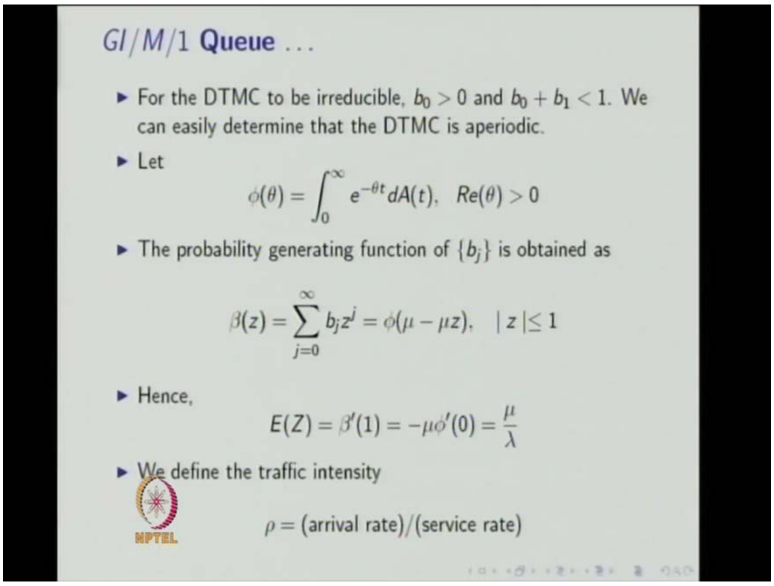


Once, you know that Q n is a discrete time Markov chain, you can find the one step transition probability matrix, whose elements are p i j. So, the pij are nothing but what is the probability that Q n plus 1 will be j given that Q n was i; that is same as what is the probability that? i plus j minus 1; customers are served during the inter arrival time period, that is probability of X n plus 1 is equal to i minus j plus 1. Whenever, j is greater than 0; if j is equal to 0 then, it is nothing but probability that X n plus 1 is equal to i plus 1. In matrix form you can write it P as a matrix whose elements are p i j.

So, the first element will be p 1 plus p 2 and so on, and the second element in the first row will be b naught, b 0. Since, b j is are nothing but the probability mass function for the random variable X n that is for all n, for all n, it is identically distributed. Therefore, the probability mass function of X n is a b j is and the running index of j is is 0, 1 and so on. Therefore, if you make the row sum b naught plus b1 plus b 2 and so on, that will be 1, whereas in the second row the first element will be b 2 plus b 3 and so on. The second row second element will be b 1, second row third element will be b naught and so on.

Substitute i and j in the above equation you substitute i and j accordingly, you will get the these values summation of b is starting from 2 and b 1, b 2, b 0 and so on.

(Refer Slide Time: 13:36)



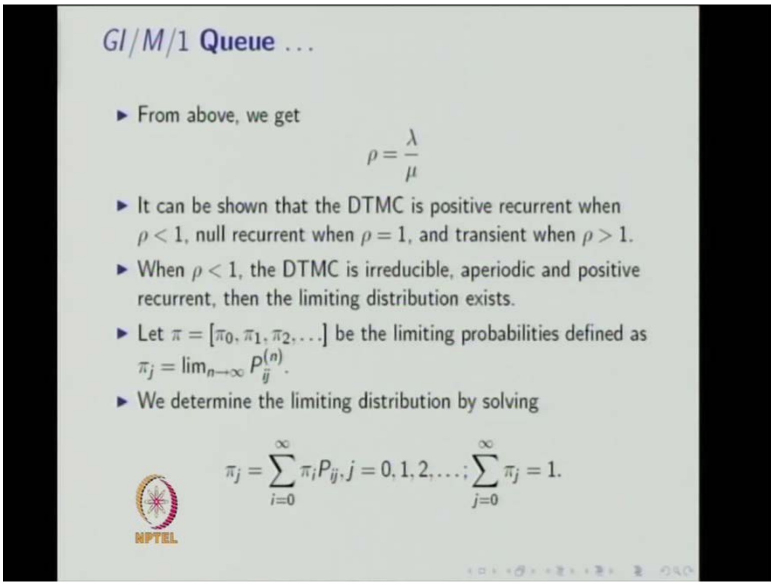
Similarly, you can get the 3rd row you can verify that each element of a b B matrix will be lies between 0 to 1 and the row sum will be 1. So basically stochastic matrix. These are the one step transition probability matrix P. Our interest is to find out the study state or limiting distribution. For that, we need a irreducible and positive recurrent and the aperiodic also. So, for the irreducible you need b 0 has to be greater than 0 as well as b 0 plus b 1 has to be less than 1. If this is satisfied then you will get the conclusion the given time homogeneous discrete time Markov chain will be a irreducible. That means, each state is a communicating with each other states with the condition of b 0 is greater than 0; b 0 plus b 1 is less than 1. We can easily determine that the DTMC is a periodic.

We have discussed the aperiodic in the discrete time Markov chain. So, we can verify that this discrete time Markov chain is a periodic also. That means, if the period is 1; now, we find out the Laplace transform of the CDF of inter arrival time distribution that is a phi naught. So, phi naught is equal to integration 0 to infinity e power minus theta times t d A (t); where A (t) is the CDF; of inter arrival time distribution with the real of theta as to be greater than 0. Now, we are finding the probability generating function for p j is that is a distribution of X n. So, that is beta (z); that the summation j is equal to 0 to infinity b j z power j; that you can write down in terms of Laplace transform. So, this

Laplace strange transform so, Laplace strange transform of A (t) so, it is a psi of mu minus mu times z.

Now, we can find out the expectation of Z that is nothing but, the beta dash of 1; if you differentiate probability generating function then substitute Z is equal to 1; will be the mean number of arrivals that is Z. Mean number of inter-arrival time so, that is nothing but, the mu by lambda. Now we can define the traffic intensity that is nothing but rho, rho is equal to arrival rate divided by the service rate.

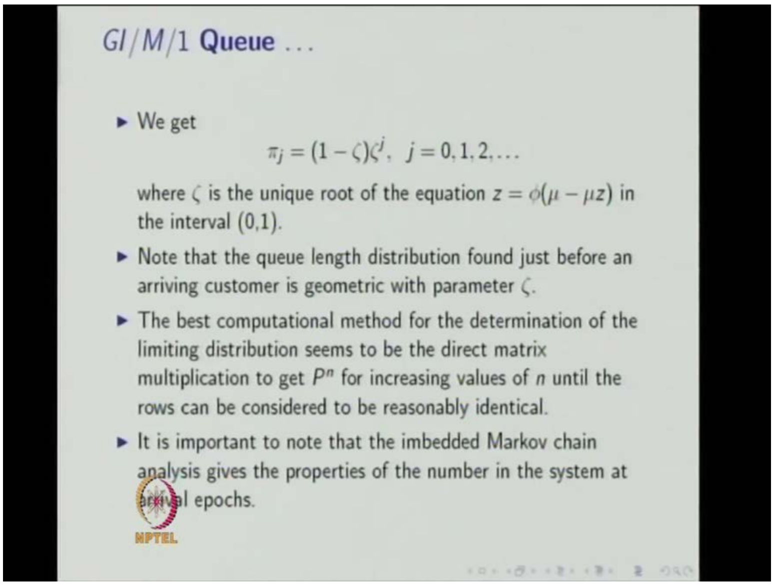
(Refer Slide Time: 17:09)



From, above you can get rho is nothing but lambda by mu; it can be shown that the DTMC is a positive recurrent when rho is less than 1. We have already made it has a irreducible and aperiodic. Now we are giving the condition for a positive recurrent when rho is less than 1 the given DTMC is a positive recurrent. If rho equal to 1 it is a null recurrent; if rho is greater than 1 then it will be a transient; that means when rho is less than 1; all the states are positive recurrent therefore, the DTMC is called a said to be a positive recurrent. Similarly, when rho is 1 all the states are null recurrent; therefore, the DTMC will be a null recurrent and similarly, for rho is greater than 1. So, our interest is to find out the limiting distribution so, when rho is less than 1 the DTMC is a irreducible aperiodic and positive recurrent. Along with the condition b0 is greater than 0; and b 0 plus b 1 is less than 1; with this condition it is a irreducible, with rho is less than one; it is a positive recurrent.

We can easily verify it is a periodic. Hence, the limiting distribution exist and it is unique and that will be a independent of a initial state i therefore, P i j will be a limit n tends to infinity; the probability of i to j in n steps. So, define the limiting distribution probability vector has a pi it consists of pi naught, p i 1, p i 2 and so on, where p i j is are define in this form; limit n tends to infinity p of i 2 j in n steps. We determine the limiting distribution by solving pi is equal to pi p and the summation of p i; i is equal to 1; so, the first one is a homogeneous equation. Including this normalizing condition, we will have a system of non-homogeneous equation. Solve this non-homogeneous, solve this system of non-homogeneous equation you get the limiting probabilities.

(Refer Slide Time: 19:55)



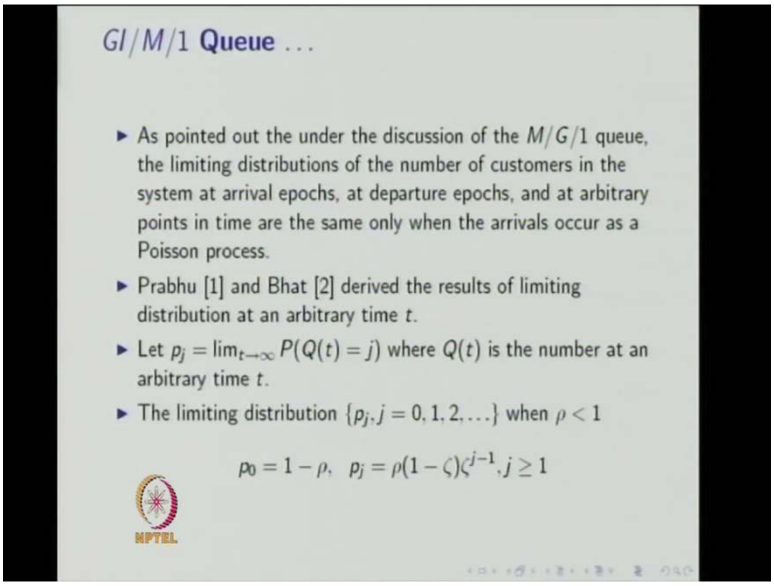
We get the limiting probabilities are 1 minus psi; psi power j; where psi is the unique root of the equation; z is equal to the Laplace strange transform of the inter-arrival distribution with the variable mu minus mu z; in the interval 0 to 1. That means first you have to solve the equation z is equal to psi of mu minus mu z. And you know what is the psi of mu minus mu z from the beta (z) solve the equation in the interval 0 to 1. So, the unique root you have to substitute as a psi; then substitute psi in the p i j; and that will be the since the unique root is in the interval 0 to 1. Therefore, the pi j 1 minus psi; psi power j; that will form a probability mass function for the limiting distribution.

Note that the Queue length distribution found just before the arriving customer is geometric distribution with the parameter psi. The limiting distribution which we got it that is just before the arriving customer. The Queue side, the Queue length distribution formed just before an arriving customer which is geometric distribution with the parameter psi. The best computational method for determination of the limiting distribution since, to be the direct matrix multiplication to get P power n for increasing values of n; until the rows can be consider to be reasonable identical. Whenever, for larger n, P power n has the identical rows it means a limiting distribution exist.

Therefore, the best computational method is find the P power n for a larger n until the rows can be consider to be a reasonable identical. It is important to note that the embedded Markov chain analysis gives the properties of the number of number in the system at arrival epochs. The Q (t) for t greater than or equal to 0; that is a discrete state continuous time stochastic process whereas, a Q n for n; n is equal to 0, 1, 2 and so on.

That is a embedded time homogeneous discrete time Markov chain and that is Q n is nothing but Q t n minus 0 that is nothing but number of customers in the system just before the nth arrival. Hence, it is important to note that the embedded Markov chain analysis gives the properties of the number in the system at arrival epochs not the departure epochs or not at the arbitrary time instance. It gives in the embedded Markov chain analysis gives the properties of the number in the system at only at the arrival epochs.

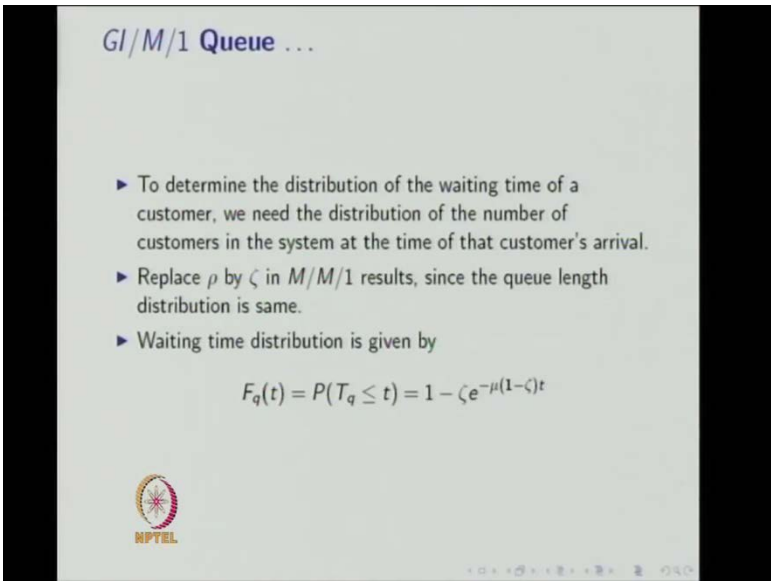
(Refer Slide Time: 23:43)



As pointed out the under the discussion of M/G/1 Queue, the limiting distribution of the number of customers in the system at arrival epochs at departure epochs and at arbitrary points in time are the same only in the arrivals occurs as a Poisson process. So, in the M/G/1 Queue the limiting distribution of the number of customers in the system at arrival epochs at departure epochs, and at arbitrary time points are same, because the arrival occurs in the Poisson process are inter-arrival follows independent exponential distribution with the same parameter. Whereas, in the GI/M/1 Queue model the embedded Markov chain results gives the limiting distribution of the number of customers in the system at the arrival epochs only that is not same as the limiting distribution at the departure epochs and that is also not same as the arbitrary time points.

Now, we are finding, now we are going to discuss the limiting distribution at the arrival at the arbitrary time points. Prabhu and Bhat derived the results of limiting distributions. at arbitrary time t. Let p j is a probability that limit t tends to infinity the probability Q t is equal to j. So, this is nothing to do with the embedded Markov chain Q n. We are finding limit t tends to infinity probability that Q t is equal to j; where Q t is the number of customers in the system at arbitrary time t. The limiting distribution exist whenever, the rho is less than 1 and the probability that no customer in the system in the long run or the limiting in a long run that p naught is equal to 1 minus rho; and the p j is equal to rho times 1 minus psi; psi power j minus 1 for j is equal to 1, 2 and so on. So, this is the limiting distribution at arbitrary time.

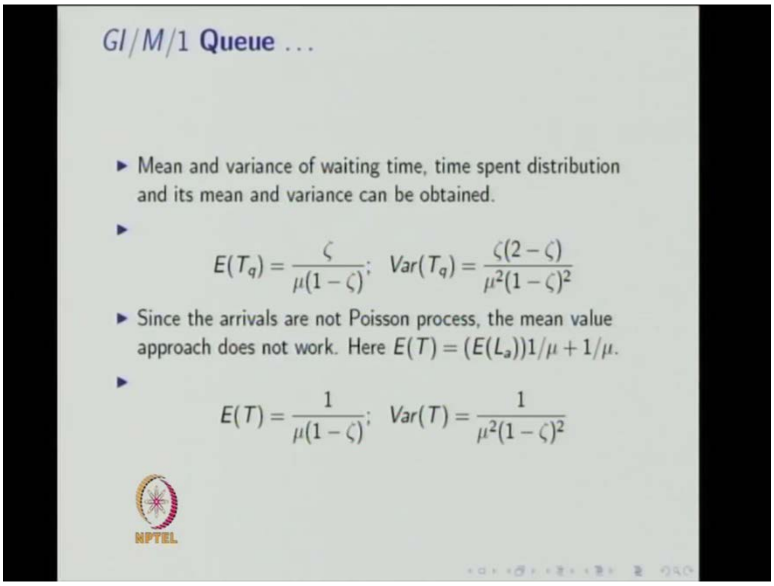
(Refer Slide Time: 26:20)



To determine the distribution of waiting time of a customer we need a distribution of a number of customers in the system at the time of that customers arrives. If you want to find out the distribution of waiting time, replace rho by psi in the M/M/1 results, since the Queue length distribution is same. Our interest is to find out the waiting time distribution. Since, the limiting time distribution at arrival epochs is the same as the limiting distribution of M/M/1 Queue model therefore, you can replace rho in the M/M/1 results by psi to get Queue length distribution if to get the waiting time distribution.

Therefore, by replacing rho by psi in the M/M/1 results of waiting time distribution you get the waiting time distribution for GI/M/1 queue as a the CDF of the waiting time distribution is a 1 minus psi times e power minus mu times 1 minus psi times t; for t greater than or equal to 0; for t is less than 0 it will be 0. So, this is the waiting time distribution for GI/M/1 queueing system.

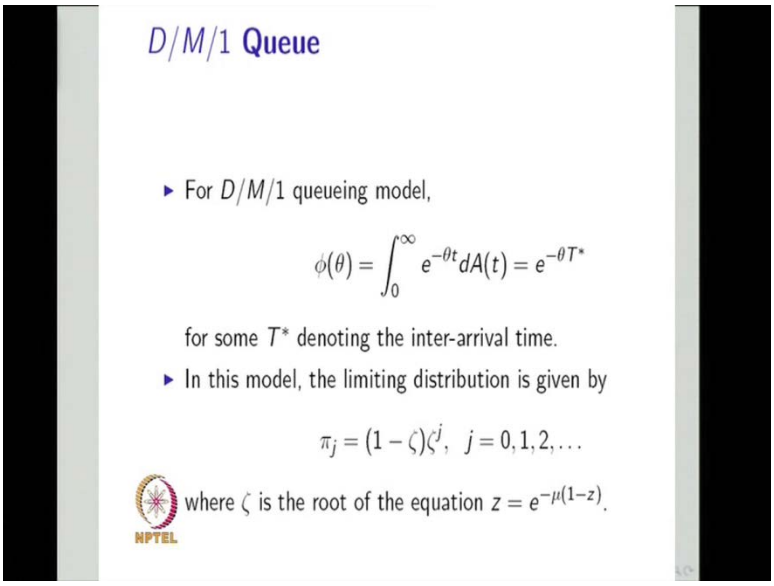
(Refer Slide Time: 28:09)



First we found the limiting distribution at the arrival epochs, next we find the limiting distribution, next we discuss the limiting distribution at the arbitrary time points from the Prabhu and Bhat results, then we have discussed the waiting time distribution, now we are discussing the moments of Queue size. The mean and the variance of waiting time, time spent is to distribution and it is a mean and variance can be obtained. Once, we know the limiting distribution at the arrival epochs as well as that we know the waiting time distribution, you can find the mean and variance of waiting time by adding mean you can find the mean of time spent also. You cannot use the mean value approach because the arrivals are not Poisson process. So, you can find out the average time spent in the system that is E (T) will be average number of customers seen by at the arrival epochs that is E (La) multiplied by 1 by mu plus the average service time that is 1 by mu will give the average time spent in the system.

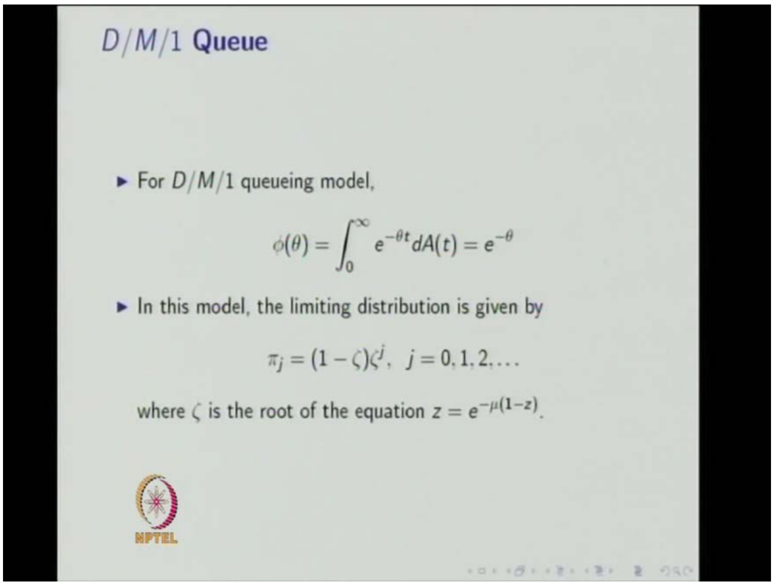
Since, the arrivals are not Poisson processes not Poisson process. You cannot use the mean value approach. By simplification you can get already we know what is the distribution of number of customers seen by seen at the arbitrary already we know the limiting distribution at the arrival epochs. So, we can find the mean from those results then multiplied by 1 by mu plus 1 by mu will give the average time spent in the system. Then, you can find the variance of time spent in the system also. The 1st this one the mean and variance of waiting time since you know the waiting time distribution you can find the mean and variance.

(Refer Slide Time: 31:01)



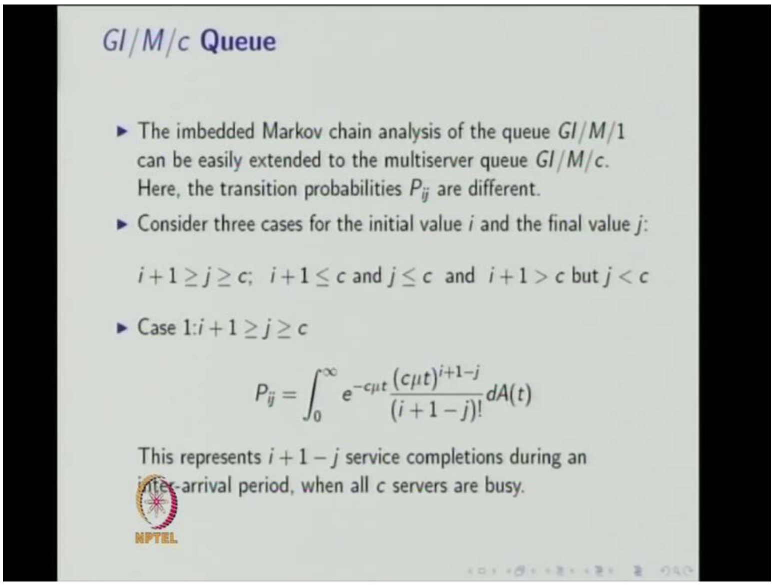
Suppose the inter-arrival time is constant. Then the Laplace strange transform will be e power minus theta; you can find the limiting distribution for the D/M/1 model also with the inter-arrival time is a constant then the corresponding queueing model is a denoted by D/M/1 Queue. So, we can find the limiting distribution at the arrival epochs as 1 minus psi; psi power j; where psi is the root of the equation, psi is equal to e power minus mu times 1 minus z; because psi of theta is equal to e power minus theta. So, to get the root of the equation you have to solve z is equal to psi of mu minus mu z.

(Refer Slide Time: 32:44)



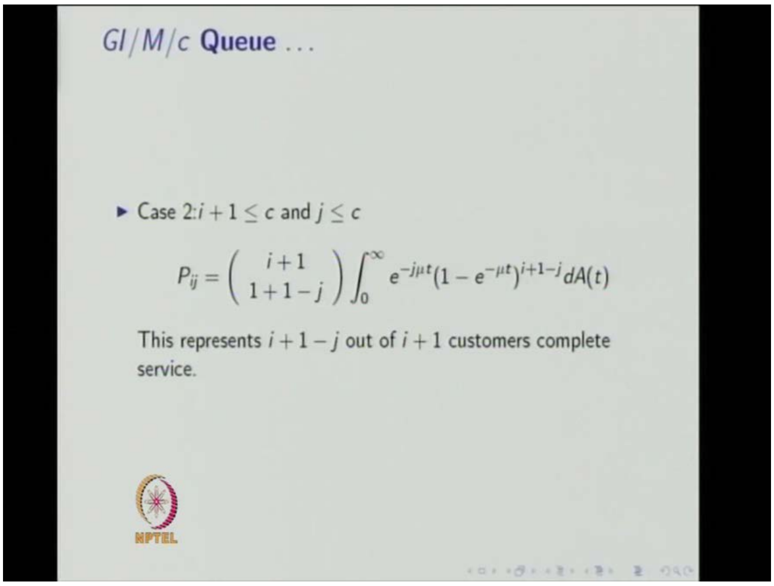
You have to get the root of the equation z is equal to psi of mu minus mu z; so, in the D/M/1 model psi of theta is equal to e power minus theta. Therefore, it will be z is equal to e power minus mu times; 1 minus z. So, if you solve this equation you will get psi; that is a unique root between the interval 0 to 1; substitute i is here, so that is the limiting distribution of the arrival epochs of D/M/1 queueing model.

(Refer Slide Time: 33:06)



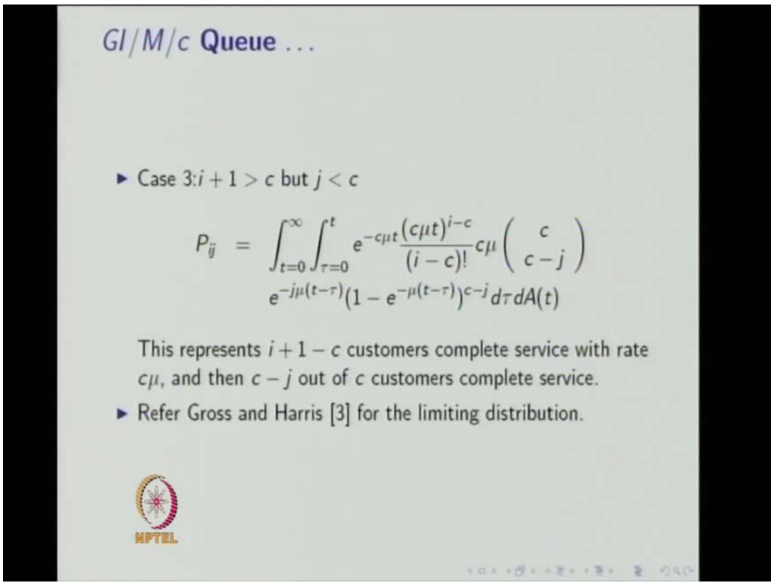
Now, we move into the multiserver queues the embedded Markov chain analysis of the queue GI/M/1 can be easily extended to the multiserver queue GI/M/c, where c is the number of servers in the system. Since the number of servers are greater than or equal to 1. The one step transition probability matrix will be different now that means, each element P i j will be different from the GI/M/1 corresponding one step transition probability, probabilities p i j is so, here you can consider 3 cases for the initial value i and the final value j. So, corresponding to the each cases your p i j can be obtained for the case 1: i plus 1 is greater than or equal to j greater than or equal to c; you will get p i j in this form.

(Refer Slide Time: 34:44)



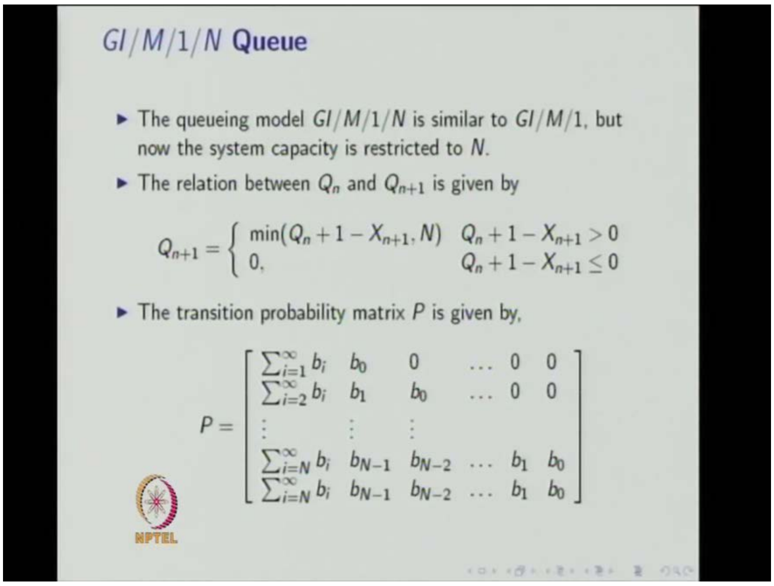
This represents i plus 1 minus j; service completions during inter-arrival period when all c servers are busy. So, the case1 is related to all c servers are busy means, the number of customers in the system is greater than or equal to c that is a situation in which you are getting the one step transition probability for the embedded Markov chain. For the case 2 the i plus 1 is less than or equal to c whereas, j is lesser than or equal to j c also. These cases related to i plus 1 minus j out of i plus 1 customers complete the service.

(Refer Slide Time: 35:07)



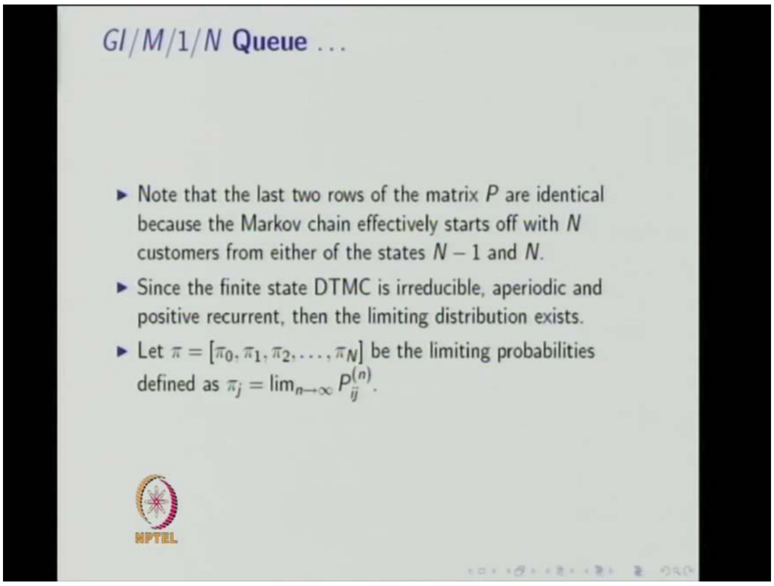
The third case is related to the i plus 1 greater than c as well as j is less than c; this represents i plus 1 minus c customers complete service with the rates c mu then c minus j out of c; customers complete the service. To get the complete limiting distribution at the arrival time the arrival epochs you can refer gross sceneries to get the complete result.

(Refer Slide Time: 36:01)



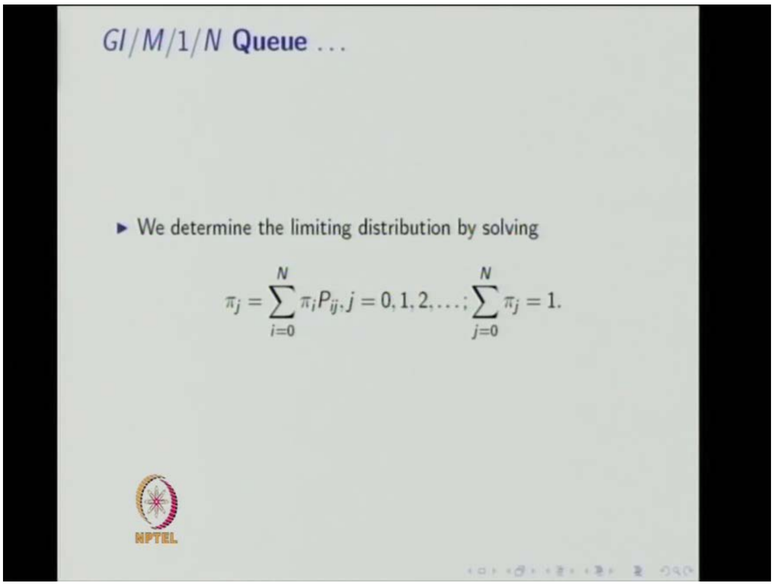
So, here we have discussed only the one step transition probabilities how that is a different from one step transition probabilities of GI/M/1 model. Now, we move into GI/M/1/N Queue; so, this queue is similar to G/M/1, but now the system capacity is restricted to some finite number capital N. Therefore, the relations between Q n and the Q n plus 1 will be a smaller change in it that is a Q n plus 1 will be minimum of Q n plus 1 minus X n plus 1 comma n; where X n plus 1 is the number of potential service completion during the inter-arrival time period. So, once you know Q n plus 1 in terms of Q n. You can conclude since Q n plus 1 is in terms of only Q n and Q n plus 1 is independent of X n plus 1. Therefore, Q n form a time homogeneous discrete time Markov chain and 1 can find the one step transition probability matrix P with the entities p i j. Here also we can verify the row sum will be 1 and each bi is nothing but the probability mass function of each bi is are nothing but the distribution of X n.

(Refer Slide Time: 37:52)

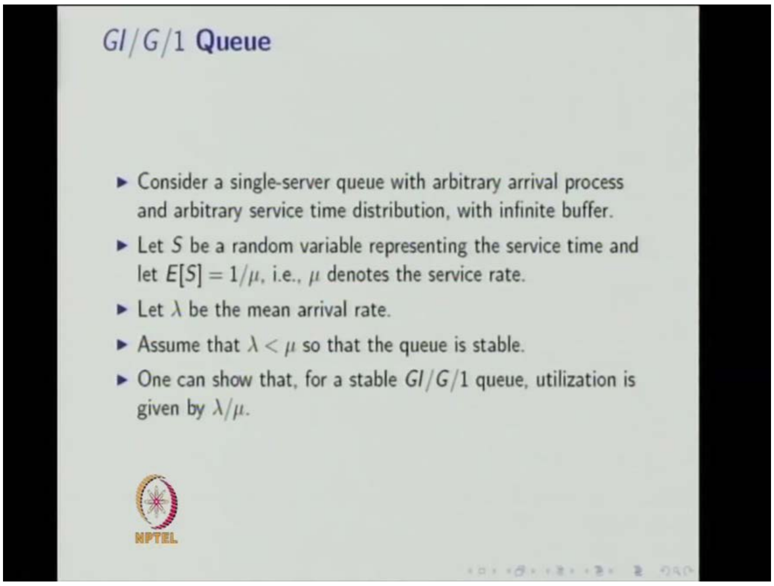


Once, you know the one step transition probability matrix we can go for finding the limiting probabilities. Note that the last 2 rows of the matrix P are identical because of the Markov chain effectively starts off with N customers from either of the states N minus 1 and N. Therefore, if you see the last two rows of the transition probability matrix both the rows will be identical because of the Markov chain effectively starts of with N customers from either of the states N minus 1 and N.

(Refer Slide Time: 38:57)



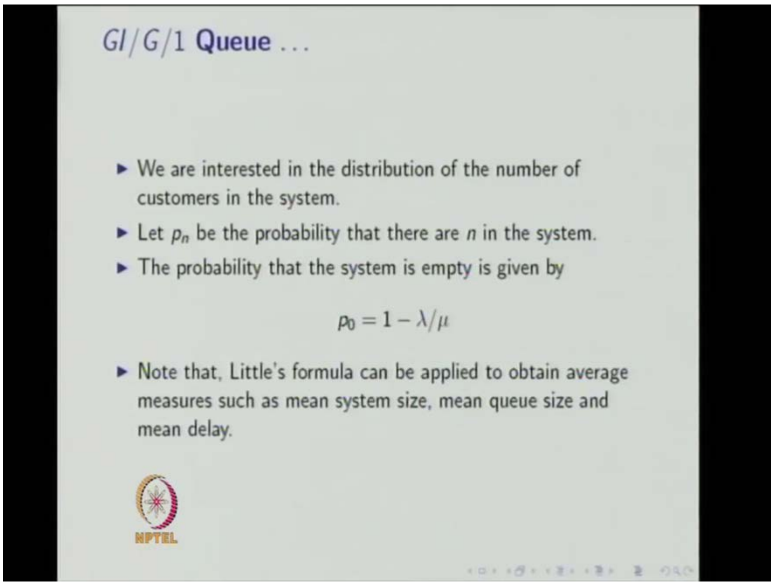
(Refer Slide Time: 39:20)



Since the finite state DTMC is irreducible a periodic and positive recurrent that the limiting probabilities limiting distribution exist. So, let pi be the limiting probabilities defined as p ij is equal to limit n tends to infinity; p i j of n; we can determine the limiting distribution by solving pi is equal to pi p and the summation of p i i is equal to 1. Here there is a restriction on row because it is a finite state Markov chain. At the end we move into GI/G/1 Queue, in this queueing model a single server queue with the arbitrary arrival process that means the inter-arrival time or non exponential distribution and arbitrary service time distribution that means a service time also non exponential distribution with the infinite buffer.

Let S be the random variable representing the service time and expectation of S will be 1 by mu; where mu denotes the service rate. Let lambda be the mean arrival rate; assume that lambda is less than mu so, that the q is stable, for a stable system you need lambda is less than mu; where lambda is related to the mean arrival rate and the mu is related to the service rate.

(Refer Slide Time: 40:58)



One can show that for the stable GI/G/1 Queue the utilization is given by lambda by mu utilization is nothing but what is the probability that the server is busy. So, that is given by lambda by mu; we are interested in the distribution of number of customers in the system in the limiting distribution we are interested in the limiting distribution of a number of customers in the system. Let p n be the probability that there are n customers in the system in a long run. The probability that the system empty is given by p 0 that is nothing but 1 minus utilization; the utilization is given by lambda by mu that is probability, that the server is busy in a long run. Therefore, the probability that the system is empty in a longer run is same has 1 minus utilization; that is 1 minus lambda by mu. Note that Little’s formula can be applied to obtained average measure such as a mean system size, mean queue size, and mean delay.

Since, it is stable system with the lambda is less than mu have the probabilities and once you know the limiting distribution you can use the Little's formula to find out the all other average measures, that means using the limiting distribution you can find out the average number of customers in the system in a long run. Once you know the average number of customers in the long run with the mean arrival rate lambda you can find the average number, average time spent in the system. Once you know the average time spent in the system you can find the average time spent in the queue also.

And, once you know the average time spent in the queue, using the little is formula you can find the average queue size also. So, you using the little formula with the help of limiting distribution we can find the average number of customers in the system, then average time spent in the system, average delay, and average queue size also. So, in this lecture we have covered the non Markovian queues namely GI/M/1, GI/M/1/N, GI/M/c, GI/G/1 Queues with these 5 lectures the renewal process is over. Here is the reference of renewal processes.