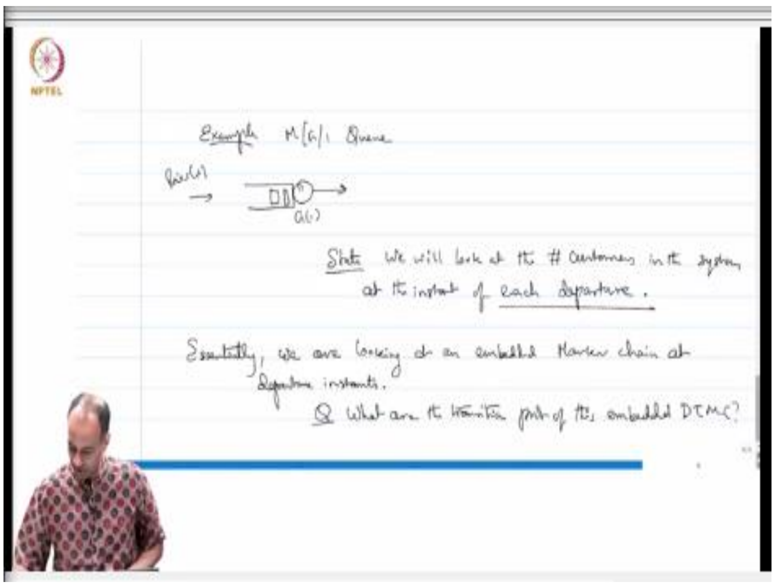
Stochastic Modeling and the Theory of Queues
Prof. Krishna Jagannathan
Department of Electrical Engineering
Indian Institute of Technology-Madras

Lecture-81
Semi Markov Processes-Part 2

Now semi-Markov process also has several applications, when you do not have exponential holding times but there is an underlying DTMC, you can use semi Markov processes.

(Refer Slide Time: 00:29)



A typical probably the most famous example of a semi-Markov process is an M/G/1 queue. So, M/G/1 queue, you have poisson lambda arrivals and some general server whose service time distribution for each job is some G. You have independent service between different jobs or different customers. And you have independence between the poisson arrival process and the service process; this is what an M/G/1 queue is.

Now in this case it is not see since this G is not an exponential distribution, G could be anything general. We cannot say that this is not a Markov process, so you have generic holding time in various states but there is an embedded Markov chain as we will see. So, the way to think about the M/G/1 queue I will just tell you how to think about this as a semi-Markov process. I will not get through the entire derivation because it is a bit tedious, I will just point you to a reference.

So, let us consider this semi-Markov example the M/G/1 example. Let us say the state of what is the state? So, I am going to look at what is the state? So, we will look at number of customers in the system at the instant of each departure this is the standard trick, each departure instant. So, for an M/G/1 queue you cannot just capture the state of the system by telling me the number of customers in the system at any given time.

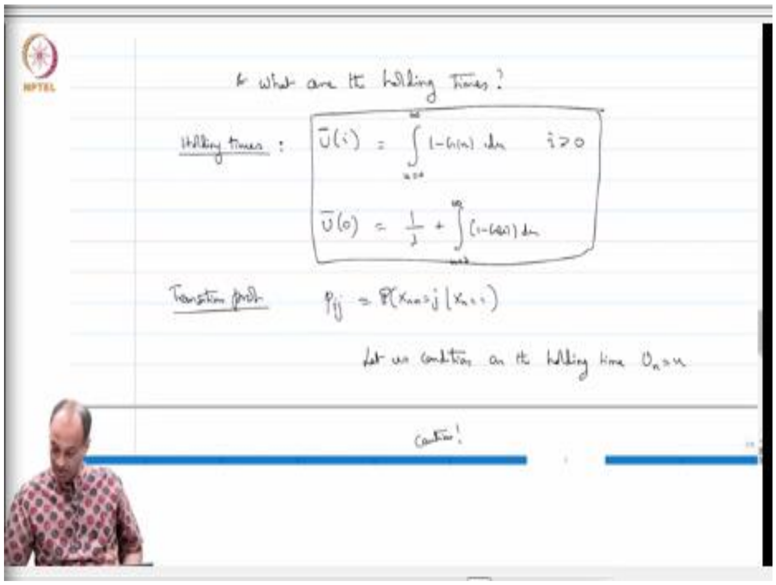
In an M/M/1 queue that is enough because both the arrival process and the service process are memory less. However for an M/G/1 queue this luxury is not available to us. What happens? If you just tell me that there are 17 customers in the system. That is not enough for me because you also have to tell me how long the customer in service has completed. Because that is not a memory less exponential, you see so if it really makes a difference to the queuing system whether that the 17 customers in the system did the customer in service.

Did he just get into service or did he is going to finish service? I do not know. So, the number of customers in the system is not a complete specification of a M/G/1 queue. So, what we are going to do is, we are going to embed a Markov chain. Essentially we are looking at an embedded Markov chain at departure instance. Why is this in fact a Markov chain? So, if I have a departure right now and there is a bunch of customers, so when I have a departure a new customer is getting the service and then that customer will complete service after some G amount of time some general amount of time.

Otherwise look at the number of customers in the system after that customer departs. Now this is a Markov chain that is because the arrival process is memory less. So, when that customer is in service the number of arrivals I will have will be poisson distributed with parameter lambda u, where u is the service time of that customer in service. And using this you can see that there is an embedded Markov chain.

So, after all there is independence between the service process and the arrival process and the arrival process is memory less. So, during the service of any customer, the number of the state change in the Markov chain which is the embedded Markov chain which is simply the number of arrivals minus the customer one customer who departs will not depend on other things. It is have this Markov property, it will only depend on how many the number of customers at the departure of one customer will depend on the number of arrivals during that customer service that is all. So, for this Markov chain, question, what are the transition probabilities of this embedded DTMC? This is one question you want to answer and what are the holding times?

(Refer Slide Time: 06:06)



So, this is a semi-Markov. So, if I tell you that the transition probability of the embedded Markov chain is some P ij if I explicitly characterize that for you. And if I also characterize the holding times in each state then I have specified as SMP for your semi-Markov process, characterization of this M/G/1 queue. Now, so let us calculate these things. So, first of all, so what are the holding times? Maybe let us look at the holding times first.

So, let us say U bar of i, so let us say there are i customers in the system, how long do I remain in the state i? So, you will be in state i, so if let us say i is non zero first, let us say there are non zero number of customers in the system, how long do I remain in state i? So, you will remain in state i as long as that customer completes service. So, this is just the U this is just integral 1 – G u du U = 0 to infinity.

It is just the expected service time of that customer, this is true for i strictly greater than 0. But when U is 0, I mean where there is nobody at service, how long do you stay in this empty state?

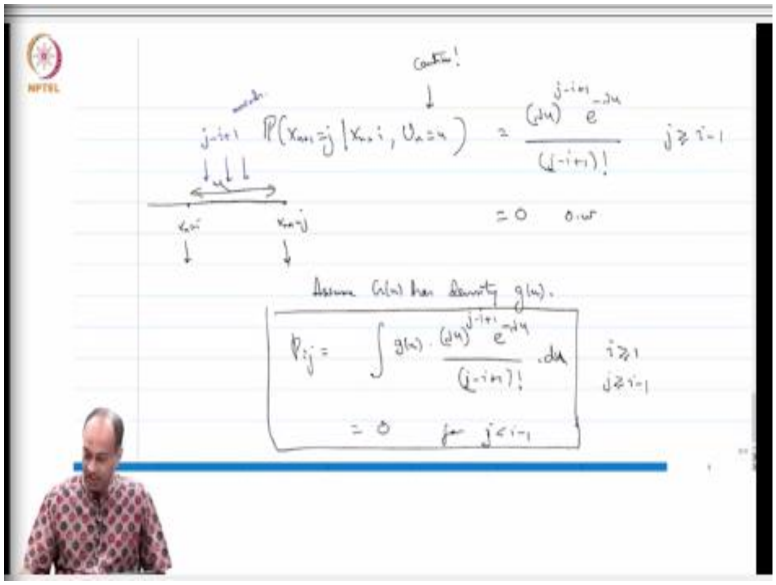
You will have to wait for the expected first arrival time which 1 over lambda of the poisson process plus for that person to complete service which is u = 0 to infinity same thing 1 - G u du.

So, this is for U bar 0 alone is a little bit different because you have to wait for the first arrival to come and then that person to complete service and exit. When you are already in the 0 state, you have to wait for the first arrival to come and then the first arrival has to complete service, then the next transition will happen. To the next transition will be to a state where it depends on the number of arrivals during the service of the first customer.

These are the holding times which are easy to calculate. Next we will address the issue of P ij's, transition probabilities P ij. So, this is simply probability that X n + 1 = j given X n = .i So, this is a little bit easier to calculate, if I tell you that, so let us say you are in state i when the nth departure happened there are n customers in the queue. Now what is the probability that there are at the next departure instance there are j customers in the queue? That is what the question is.

Now this is a little easier to calculate if you condition on the nth holding time. Now let us condition on the holding time U n = u. This you have to be little careful here U n could be a continuous random variable in general. So, you have to do it with the appropriate caution that U n = u I really mean U n is between u and u + delta. So, with that understanding that we always have we can write this out.

(Refer Slide Time: 10:05)



Probability, what is probability that X n + 1 = j given X n = i and U n = u? I will just put caution here, you know what I am talking about, you have to really put U n between u and u + delta. So, what is this? So, let me just write this out, maybe a picture will help us here. So, let us say this is my here X n = i, I am going to what is the probability of going to X n + 1 = j? Given that this holding time is some u.

So, remember that the customer in service is going to leave, this is 1 departure instant and this is the next departure instant. So, the 1 customer whose in service is going to leave, so you are going to have if there are no arrivals in this interval you are going to have i - 1 customers j will be i - 1 if there are no arrivals. But of course there could be arrivals coming in here. In this there will be poisson arrivals typically coming in this u duration of time conditioned on this u there will be poisson number of arrivals with parameter lambda u.

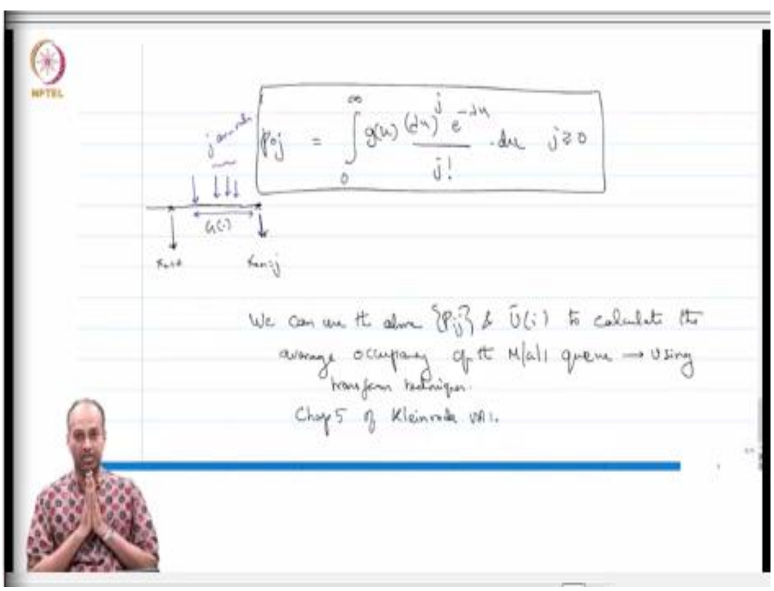
So, how can you have j customers at the next departure instant? That can happen if there are j - i + 1 arrival, is that correct? Did I do that correctly? So, if there are j - i + 1 poisson arrivals in this instant. So, I will have totally j + 1 jobs but the next one the job in service leaves, so I will be left with j customers. So, what is the probability of there being j - i + 1 poisson arrivals in this interval, I know the answer. This is simply lambda u to the number of arrivals which is j - i + 1 e power -lambda u over j - i + 1 factorial.

This is true for j greater than or equal to i - 1 and it is equal to 0 otherwise. See because j cannot be smaller than i – 1, if there are no arrivals between 2 consecutive departures you would have reduced 1 customer if there are no arrivals. Between 2 consecutive departures, the state cannot go down by more than 1 that is simply not possible by definition. So, this is what it is, this is the conditional distribution.

And of course we know the distribution of U n, what is U n? It is the service time of the customer in service. So, if let us say assume let us say that G u has, so that some density g u what I mean is that if you differentiate G u you get g u. I do not need this but life becomes a little easier if this is the case. So, in that case I can write P ij = integral g of u, I am just averaging over the distribution of that probability.

This is just total probability, - j - i + 1 e power -lambda u over j - i + 1 factorial du. So, this is what it is P ij, what is the probability that I have j customers or in the next state between successive departures if I started with i customers? This is valid only for i greater than 0, so i = 0 case I have to treat differently. This is valid for i greater than 0 or maybe I should write it as i greater than or equal to 1 and j greater than or equal to i - 1 and it is 0 for j less than i - 1. So, maybe I should point that out also equal to 0 for j less than i - 1.

(Refer Slide Time: 14:59)



Finally the case P 0j has to be treated separately. So, what is the issue now? I do not have i, I have not 0 customers when this departure happened the queue is empty. Then I want the next customer to leave behind j customers in the system. So, that can happen if and only if there are j customers arrive that during the service time of the first customer. So, in the case when maybe I should draw a separate picture to make things this P 0j case is perfectly clear.

Let us say at this departure instant the system is empty X n = 0 and the next consecutive departure instant I want to have j arrivals, I mean j customers. How is that possible? So, if you have the system is empty first of all you need a 1 person arrivals to come, 1st poisson arrival to come here which will take exponential 1 over lambda time because it is a memory less. And this guy is going to leave here, so this customer blue customer gets into service and let us say this is FCFS this will be the service time.

And in that service time you want j arrivals during that person service time. What is the probability of that? So, P 0j will be equal to the probability of having j poisson arrivals during that person's service time which is like u. Which is this is distributed according to G, the service time of the first arrival. So, this can be calculated as u = 0 to infinity g of u lambda u to the j e to the -lambda u over factorial j, this is true for j greater than or equal to 0 du for j greater than or equal to 0.

So, these are the transition probabilities of the embedded Markov chain of the M/G/1 queue and the holding times are whatever we found out, it just depends on whether there are non zero number of customers or 0 customers, this also we found out. So, this sort of this completely characterizes the semi-Markov process, the holding times are like G in all non zero states, the holding times are 1 exponential lambda + G in the 0 state.

And my P ij’s and P 0j are given here, so this completely characterizes the semi-Markov characterization of M/G/1 queue. Now that you have the P ij’s of the embedded Markov chain and the holding times, you can go ahead and calculate pi i etcetera. So, using these transition probabilities and U i's you can calculate the fraction of time spent in state j, the usual things you can do.

You can write out balance equations and find out the pi where then from pi you can get the p’s and all that. See you can do all that but as you can see the expressions are a bit messy, so it is hard to do it in closed form, I do not think you can do it in closed form. However a transform approach if you use z transforms or probability generating functions you can actually obtain the fraction of time spent with j customers in the M/G/1 queue.

You can actually go ahead and derive the waiting time with an M/G/1 queue with the waiting time distribution and all that. We will not get into all the derivation, I will just make a remark that we can use transform technique we can use the above P ij and U i to calculate the average occupancy of the M/G/1 queue. See in principle you have all these P ij’s and U i's, so you can calculate pi and then calculate average occupancy that is the fraction of time spent in state j and all that the M/G/1 queue.

This will give you have to use transform techniques. So, the transform of the number of customers in the M/G/1 queue or the transform of the waiting time of the system time of a customer in M/G/1 queue these are all very well known. Slightly messy calculations, I am not going to get into those. If you are interested in finding out these transforms of the waiting time or the occupancy distribution I refer you to chapter 5 of Klienrock's book volume 1.

Klienrock is a very famous queuing systems book, chapter 5 deals with these transforms M/G/1 queue, chapter 5 is entirely about the M/G/1 queue. And you can get the transforms of waiting times and occupancy distribution of the M/G/1 queue using transform techniques; you can look at this reference. So, that concludes our discussion on semi-Markov process and this also concludes the course. So, thank you very much for watching, the course is finished.