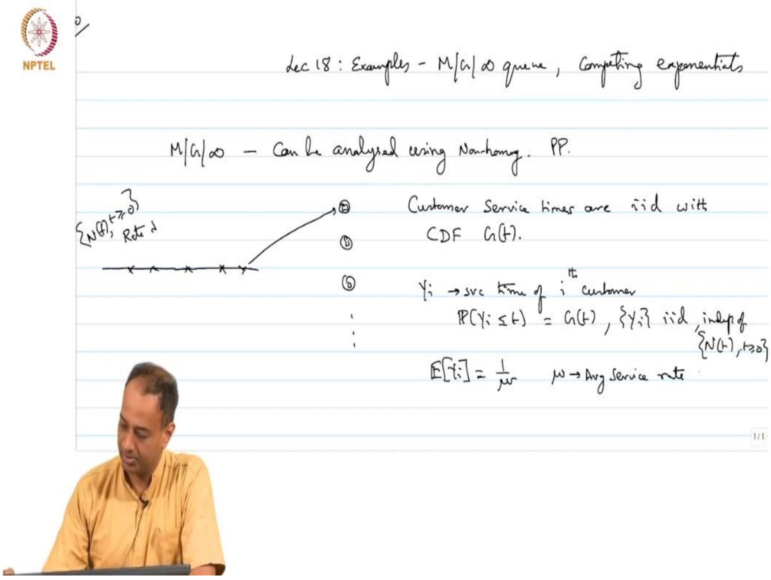
Stochastic Modeling and the Theory of Queues
Prof. Krishna Jagannathan
Department of Electrical Engineering
Indian Institute of Technology - Madras

Module - 3
Lecture - 20
Examples: Non-Homogeneous Poisson Process

(Refer Slide Time: 00:13)



Lec 18 : Examples - M/G/∞ queue, Competing exponentials

M/G/∞ - Can be analysed using Non-homog. PP.

Customer Service times are iid with CDF $G(t)$.

$Y_i \to$ svc time of i^{th} customer

$\mathbb{P}(Y_i \leq t) = G(t)$, $\{Y_i\}$ iid, indep of $\{N(t), t \geq 0\}$

$\mathbb{E}[Y_i] = \frac{1}{\mu}$ $\mu \to$ Avg Service rate

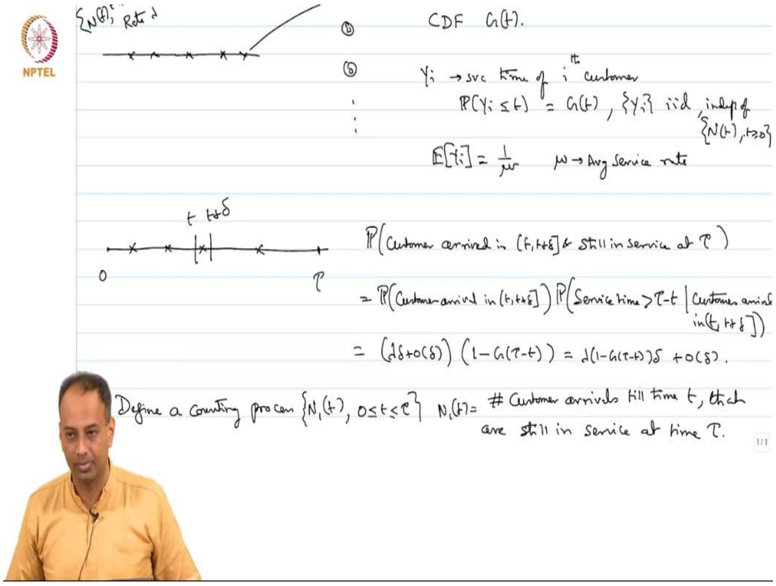
Welcome back. Yesterday we studied the non-homogeneous Poisson process, also developed queueing theory notation, the slash notation. We will take a look at this M G infinity queue as an application which can be analysed using non-homogeneous Poisson process. So, the arrival process to an M G 1 queue is of course a Poisson process of rate lambda. So, let us call this N of t.

So, this is a homogeneous Poisson process of rate lambda; customers arrive to this M G infinity queue at rate lambda. There are infinitely many servers, so, there is no waiting, no buffering. So, as soon as a customer arrives, she is just directed to a server that is available. There are always many servers; there are so many servers that there is no waiting; there is always an available server for any arriving customer.

Now, the customers; so, customer service times are assumed to be independent identically distributed with CDF G of t. So, that means that if Y i is the service time of ith customer, then we have probability Y i less than or equal to t is just G of t, and Y i are assumed to be IID and independent of the arrival process N of t. The expected service time is of course the same for all customers, is taken to be 1 by mu, where mu is called the average service rate.

So, no matter which server the customer goes to, his service time, the time he will take to get service is some random variable Y i which is IID. This is the same across all servers, identically distributed across all users, independent of other customers and independent of the arrival process. Is the setup clear?

(Refer Slide Time: 03:56)



$\{N(t)\}$; Rate λ

① CDF $G(t)$.

② $Y_i \to$ svc time of i^{th} customer

$\mathbb{P}(Y_i \le t) = G(t)$, $\{Y_i\}$ iid, indep of $\{N(t), t \ge 0\}$

$\mathbb{E}[Y_i] = \frac{1}{\mu}$ $\quad \mu \to$ Avg Service rate

$t \quad t+\delta$

$0 \quad \tau$

$\mathbb{P}($Customer arrived in $(t, t+\delta]$ & still in service at $\tau)$

$= \mathbb{P}($Customer arrived in $(t, t+\delta])\, \mathbb{P}($Service time $> \tau - t \mid$ Customer arrival in $(t, t+\delta])$

$= (\lambda\delta + o(\delta))(1 - G(\tau - t)) = \lambda(1 - G(\tau - t))\delta + o(\delta)$.

Define a Counting process $\{N_1(t), 0 \le t \le \tau\}$ $N_1(t) =$ # Customer arrivals till time t, that are still in service at time τ.

Now, it turns out that you can look at this queueing process, the number of customers in the system at any given time, using this non-homogeneous Poisson process viewpoint. So, you proceed as follows: You take some time tau, because you have all these arrivals. You take some time t and t plus delta. So, you look at the probability of having a customer arrived in t, t plus delta and still in service at some time tau; tau is bigger than t, in this picture.

So, this is nothing but the probability that there is an arrival t, t plus delta times the probability that the service time of the customer being greater than tau minus t, because the customer arrived here and he still is in service; so, his service time must be greater than t minus tau, given customer arrival in t, t plus delta. This is just, I am writing P of A intersection B as P of A times P of B given A.

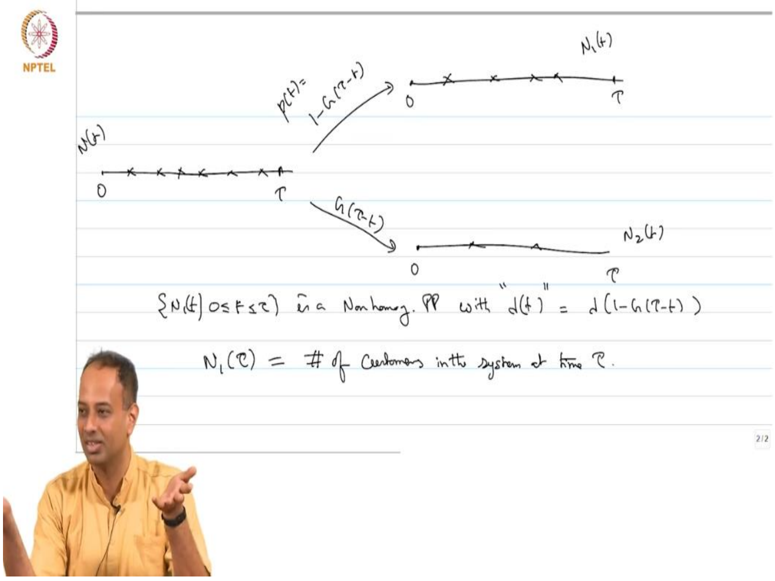
So, the first, this guy we know; this is simply; what is the probability of a customer arrival in t, t plus delta? It is a Poisson process. So, we know, this is equal to lambda delta plus little o delta. And the service time of the customer is independent of the arrival process. So, conditioned on the customer arrival being in t, t plus delta, the service time distribution is the same as the unconditional distribution.

Why? Because, we have assumed this, that the service random variables are IID across customers and independent of arrivals process. So, this conditioning can be dropped, that is the basic point. So, unconditional probability of service time greater than tau minus t, which is equal to what? I know; see, the probability of a service time less than or equal to little t is G of t. So, a probability that service time greater than tau minus t is 1 minus G of tau minus t.

This can simply be written as lambda 1 minus G of tau minus t delta plus little o delta. Now, let us define a counting process N 1 t, and this is defined from 0 to tau. So, for each t, N 1 t is the number of customer arrivals till time t, that are still in service at time tau. So, N 1 of t for t going between 0 and tau is the number of arrivals. So, if you look at some time t here, you look at all the arrivals that have come till time t.

So, of course, in the original process, that is, you know what that is; that is the total number of Poisson arrivals in 0 to t; but in N 1, you are only looking at those arrivals which still remain in the system at time tau. So, you are throwing away those arrivals which have left the system. So, the key argument is that, from this above calculation, if you look at this, this is saying that the process N 1 t is like a non-homogeneous Poisson process, whose lambda t is given by that expression, lambda times 1 minus G of tau minus t. Is that clear? It is like, we can draw a splitting picture, which I will do. Let me just do that.

(Refer Slide Time: 10:00)



$\{N_1(t)\ 0 \le t \le \tau\}$ is a Nonhomog. PP with "$\lambda(t)$" = $\lambda(1-G(\tau-t))$

$N_1(\tau)$ = # of Customers in the system at time τ.

So, you have; it is like splitting; so, this is the original process, this is N of t. You are splitting that to keep those arrivals. So, we are keeping those arrivals which; this is my N 1 t; everything is going from 0 to tau, the consideration is from 0 to tau. So, let us call this N 1 t; this is some N 2 t. So, at each time t, you look at the number of arrivals, you look at all those arrivals that have come to the system.

And if that arrival has left the system already, after being served, you send it down; if the arrival is still in the system, you send it up. That is what you are doing, and you do, you can look at this as the delta time slot picture also. So, in each delta time slot, if the arrival has come, you send it up with probability 1 minus G of t minus tau; that is what is happening. And send it down with probability; this is 1 minus G of tau minus t and this is G of tau minus t.

And you do this independently across all the arrivals, because the service times are independent of the arrival process. Now, so, this is like a Bernoulli split, independent Bernoulli split, except it is not a IID Bernoulli split; the P is changing across time; that is all that is changing. So, what can be easily argued is that, while you get this expression for the probability of an arrival still being in service, it is also independent, this N 1 t also has the independent increment property, like we just argued.

So, it just follows from the definition of a non-homogeneous Poisson process, is that N 1 t is a non-homogeneous Poisson process of rate lambda times 1 minus G tau minus t. So, this N 1 t for, N 1 of t for 0 less than or equal to t less than or equal to tau is a non-homogeneous Poisson process with the rate, the lambda t equal to lambda 1 minus G tau minus t. In all this, you can think of this tau as being fixed.

You are fixing some tau, and you are looking only at 0 to tau, and you are looking at this interval 0 to tau and the counting process N 1 of t in this interval 0 to tau. So, I am putting lambda t within quotes because that is the lambda t of that non-homogeneous process. **"Professor - student conversation starts"** Yes. Essentially, what I am looking at; yeah, so, you are looking at the splitting probability being a function of t.

See, tau is fixed, let us look at tau as being fixed. So, I am just thinking of this as some P of t. So, you fix a tau; you are looking at some time t, let us say t, t plus delta. Now, the coin toss is going up with probability P of t equal to all that, 1 minus G of tau minus t. So, if there is an arrival in that little interval, you will send it up with probability P of t, and send it down with probability 1 minus P of t.

So, this coin toss is an independent Bernoulli coin toss, except that the probability of sending up or down is a function of time. If you want, you just call this P of t equal to. At each time t, little interval t, t plus delta, I toss a coin independently, not with probability P, but probability P t; and if it throws head, if there is an arrival, I will send it up; if there is no arrival, of course, I do nothing; if there is an arrival and my coin shows tail, I will send it down.

The coin toss is a, I mean the head probability, success probability depends on time. That is all that is happening. It is no different from splitting a Poisson process, except that P is P of t now. **"Professor - student conversation starts"** Well, now, see, that is for this calculation, right? See, any t, t plus delta; so, I just calculated the probability that there is an arrival and that arrival is still in the system. That turns out to be like this.

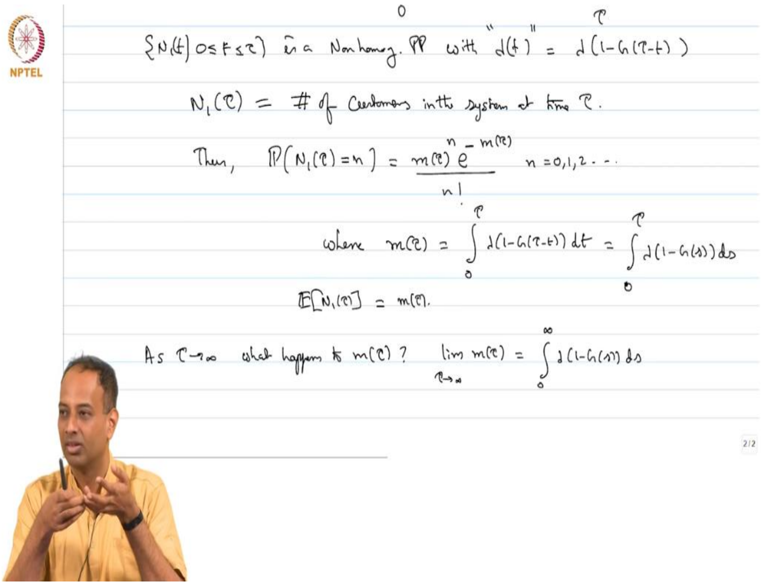
And it is independent across these little micro slots, because the original process has IIP, and these service times are independent across customers and independent of arrival process. Therefore, all I am saying is that, this is a valid way of looking at the system. It is as though I am tossing a coin. So, look at it this way, right? At every t, t plus delta micro slot, I am tossing a coin whose head probability is P of t.

If there is no arrival, there is nothing to do at that time t, t plus delta. If there is an arrival and if your coin toss, this P t coin toss is a head, you send it up; if it is a tail, you send it down. If there is no arrival, there is nothing to send up or down. So, you might as well say, there is no coin toss; or you can say, there is always a coin being tossed with probability head being P t; if there is an arrival, I will act to split; if there is no arrival, I will do nothing.

See, all that is the same, from IID's Bernoulli splitting; here it is independent Bernoulli splitting, but the coin toss probably changes with time; that is only difference. So, this is all there is. **"Professor - student conversation ends"** So, now what is this N 1 tau? N 1 t is the number of arrivals in 0 to t, which are still in the system at tau. So, what is N 1 tau? Is the number of arrivals in 0 to tau, which are still in the system at tau, which means it is the number of customers in the system, at time tau.

Because N 1 t is the number of customer's arriving in 0 to t which are still in system at time tau, so, N 1 of tau is the number of arrivals in 0 to tau which are still in the system at time tau, which is the number of customers in the system at time tau. So, this N 1 tau will be Poisson distributed. So, since N 1 t is a non-homogeneous Poisson process, whose rate we know, we can write this down.

(Refer Slide Time: 19:00)



$\{N_1(t)\ 0 \le t \le \tau\}$ is a Nonhomog. PP with "$\overset{0}{\lambda}(t)$" $= \lambda(1-G(\overset{\tau}{\tau}-t))$

$N_1(\tau) =$ # of Customers in the system at time τ.

Thus, $\mathbb{P}(N_1(\tau)=n) = \dfrac{m(\tau)^n e^{-m(\tau)}}{n!}$ $n=0,1,2\ldots$

where $m(\tau) = \displaystyle\int_0^\tau \lambda(1-G(\tau-t))\,dt = \int_0^\tau \lambda(1-G(s))\,ds$

$\mathbb{E}[N_1(\tau)] = m(\tau)$.

As $\tau \to \infty$ what happens to $m(\tau)$? $\displaystyle\lim_{\tau\to\infty} m(\tau) = \int_0^\infty \lambda(1-G(s))\,ds$

2/2

So, you can say, thus probability that N 1 tau equal to n is simply your Poisson formula, which is m tau to the e to the minus m tau m tau to the n over n factorial, for n = 0, 1, 2, ..., where m tau is integral 0 to tau lambda t, whatever lambda t is; in this case, it is lambda times 1 minus G; I am just writing integral lambda t dt, from 0 to tau. So, this just works out to be integral 0 to tau.

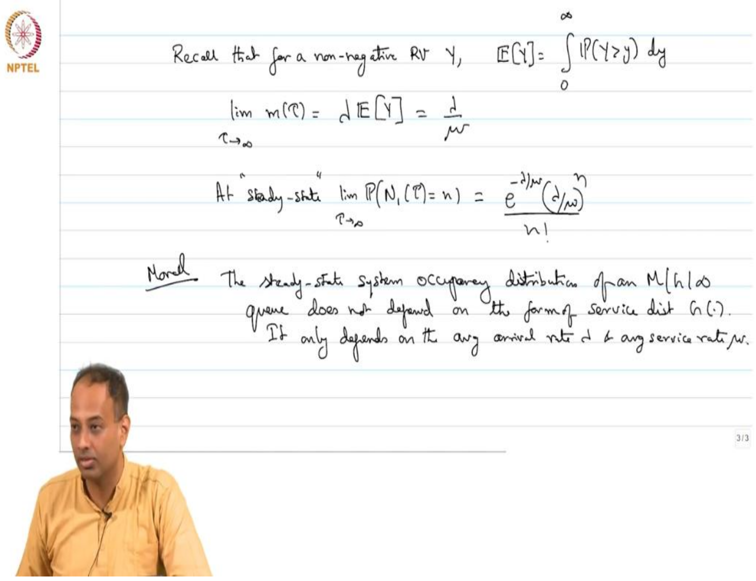
So, you can make tau minus t is equal to s; make that substitution, and this just becomes lambda times 1 minus G of s ds. G is known to you; the CDF of service time is known to you. So, you can calculate this integral. For any time tau, you can calculate m of tau as this integral, because G is known. So, m of tau simply gives you the parameter of the Poisson PMF. Is that clear? And of course, expected; what is expected N 1 tau?

It is the expected number of customers in the system at time tau, which will simply be; what is the expected value of a Poisson PMF? The parameter itself, right? In this case will be what? m tau, which can be calculated like so. Is that clear? So, that is it. So, for any time, nearly we can calculate the PMF of the number of customers in the system, which is also the PMF of the number of servers which are occupied.

There are infinitely many servers, of course. The number of customers in the system is also the number of servers that are occupied. It has PMF, this Poisson PMF with parameter m tau, and expected number of customers is also m tau. Any questions? Now, it is interesting to look at what happens as tau tends to infinity. So, I am starting the system at 0; letting it run for a very long time; and I want to look at what happens to the occupancy of the system; how many, the PMF of the number of people in the system.

It is easy to show that what happens to m; so, what happens to m tau? So, limit tau tending to infinity m tau is nothing but integral 0 to infinity lambda 1 minus G of s ds. Now, integral 0 to infinity 1 minus G s ds is basically the integral of the complimentary CDF of the service time. The service time is of course a non-negative random variable, and the integral of the complimentary CDF of a non-negative random variable is nothing but the expected value.

(Refer Slide Time: 23:42)



Recall that for a non-negative RV Y, $\mathbb{E}[Y] = \int_0^\infty \mathbb{P}(Y > y)\, dy$

$$\lim_{\tau \to \infty} m(\tau) = \lambda \mathbb{E}[Y] = \frac{\lambda}{\mu}$$

At "steady-state" $\lim_{\tau \to \infty} \mathbb{P}(N_1(\tau) = n) = \frac{e^{-\lambda/\mu} (\lambda/\mu)^n}{n!}$

Moral The steady-state system occupancy distribution of an M/G/∞ queue does not depend on the form of service dist G(·). It only depends on the avg arrival rate λ & avg service rate μ.

3/3

So, we can say, recall that for a non-negative random variable Y, we know this, integral 0 to infinity, probability Y greater than little y dy is nothing but the expected value. So, if you look at this guy, lambda; if you forget the lambda, pull the lambda out of the integral, it is just a number, you have integral 0 to infinity complimentary CDF of service time. So, you get limit tau tending to infinity; m tau is nothing but lambda times expected Y i, I could just say expected Y, which is the average service time, which is nothing but lambda upon mu.

So, if you let this system run for a very long time, so, at steady state what happens to this? Probability of number of people in the system; so, this is what I mean, probability of N 1 tau is equal to n as n tends to infinity. N 1 tau is this guy, system occupancy. I am looking at the probability that there are n people in the system after a very long time, will simply be e power minus lambda over mu lambda over mu whole to the n over n factorial.

I have just written down what m tau is for after a very long time. So, steady state means, after a very long time, the number of customers in the queue is a Poisson distributed random variable with parameter lambda upon mu; lambda is the arrival rate, mu is the service rate. See, the cool thing about this result is that the distribution of the occupancy, the number of customers in the system depends on the service distribution only through mu.

So, at any finite tau, it depends on the entire distribution. So, at any finite time tau, the number of customers in the system is a Poisson random variable with parameter m tau, given by that integral, which generally depends on the entire service distribution. However, as tau becomes very large; so, you look at the system in steady state, the form of G does not matter at all, what matters is the average service time or average service rate.

So, if you have any M G infinity queue; you have an M G infinity queue with the service time whatever you like, but average service rate mu; and I have a different M G infinity queue, with the same mu, same average service rate, but a totally different service distribution. For any finite tau, our systems will behave differently, but after a very long time, statistically, they will be the same.

So, this is a very non-trivial result, but it comes out very beautifully from this analysis. If you think about it, we did not do anything; we just looked at it the right way. We did not do any very heavy, we did not do anything very heavy analysis, we just did this splitting of Poisson process with this probability being a function of time, and we said this models in M G infinity queue, and we got this very nice result.

So, the moral; the steady state system occupancy distribution of an M G infinity queue does not depend on the form of the service distribution G. So, it only depends on the average arrival rate lambda and average service rate mu. Of course, for any finite tau, it depends on the G, like that integral says, but if you send tau to infinity, this form of G does not matter anymore. Is it clear? Any questions on this?

Interestingly, if you just look at this splitting picture, wherever I drew it, yeah, this guy; if I look at this splitting picture, so, we said this is a non-homogeneous Poisson process, and likewise, this is a non-homogeneous Poisson process. Now, by the same argument we used in the IID Bernoulli splitting, these 2 counting processes, N 1 t and N 2 t will turn out to be independent.

You remember, when you have an IID Bernoulli split, although they are coming from the same process, the up process and the down process were independent. The same argument goes through here, except this P is P of t now. So, here too, the up process N 1 t and the down process N 2 t are independent, same argument. So, what does that mean? In an M G infinity queue, at any given time, the number of customers still in service is independent of the number of customers who have departed.

See, the N 2 t is what? Is a, at any given time t, the number of arrivals in 0 to t, who have departed. N 1 t is the number of arrivals in 0 to t, who have not departed. These two are independent processes by this splitting argument. So, very remarkably, so, the number of customers still in the system is independent of the number of customers who have left the system. It is a very remarkable thing, right?

So, this is not true in general. Generally, if I tell you that a lot of people just left, you may think that the queueing system has fewer people; but not true in an M G infinity queue. It is a very highly non-trivial observation. It comes from the independence of the 2 split processes, which itself is a non-trivial result, as we saw; because of the little o delta business, you get this independence. So, these two processes turn out to be independent.

They are independent non-homogeneous Poisson processes; another beautiful property. So, that finishes our M G infinity queue discussion. **"Professor - student conversation starts"** Yes. So, if you have 2 processes; lambda 1 t is one process, lambda 2; they are independent. So, you work it out. So, the IIP will still follow. Yeah, it should be correct. It is correct, no. So, lambda 1; I think what you are saying is correct; just verify it.

So, you take 2 independent non-homogeneous Poisson processes, lambda 1 t, lambda 2 t. The probability of having arrival in any one of them; yeah, it seems; I think that is correct; what you are saying seems correct. Do verify it. IIP is of course true, SIP is anyway not there. **"Professor - student conversation ends"**