Communication Networks
Prof. Goutam Das
G.S. Sanyal School of Telecommunication
Indian Institute of Technology, Kharagpur

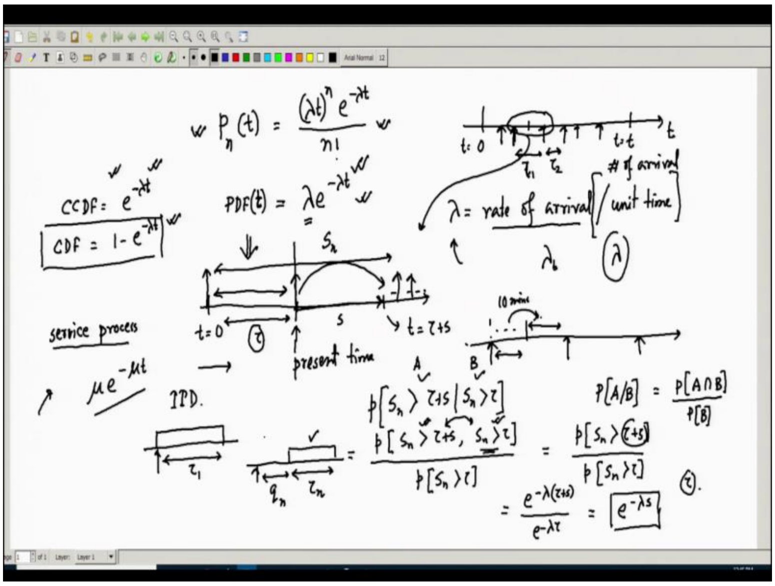
Module - 04
Queuing Theory
Lecture - 19
Memorylessness

So, far we have discussed a process called the Poisson process it's a random process of course and we have tried to characterize it we have been told that it's a stationary process of course. So, therefore, the stationarity property remains intact, and we have given two postulates actually one is nonsimultaneity and the other one is independent increment.

So, by taking those two things we could actually derive the whole Poisson process the underlying stochastic nature of it. So, and then we have got the poison formula also. Now, this independent increment that something we have discussed. So, that is actually where it's like it indicates that the future will not be dependent on the past. So, that is what is called Memorylessness.

So, what we will try to see over here is that because of that assumption of independent intervals, basically this memorylessness things that come into the picture. So, today what we will try to do is we will try to actually explain this particular term memorylessness in more detail or we will try to prove also the property of memorylessness and in that process, you will see what we mean by this property called memorylessness so, that should be our target. So, let us see what we have discussed so far let us try to understand that.

(Refer Slide Time: 02:01)



$$P_n(t) = \frac{(\lambda t)^n e^{-\lambda t}}{n!}$$

$$\text{PDF}(t) = \lambda e^{-\lambda t}$$

$$\text{CCDF} = e^{-\lambda t}$$

$$\boxed{\text{CDF} = 1 - e^{-\lambda t}}$$

λ = rate of arrival [# of arrival / unit time]

service process

$\mu e^{-\mu t}$

IID.

$$p[S_n > \tau + s \mid S_n > \tau] = \frac{p[S_n > \tau + s, S_n > \tau]}{p[S_n > \tau]} = \frac{p[S_n > \tau + s]}{p[S_n > \tau]} = \frac{e^{-\lambda(\tau+s)}}{e^{-\lambda\tau}} = \boxed{e^{-\lambda s}}$$

$$P[A/B] = \frac{P[A \cap B]}{P[B]}$$

So, what we have done so far is we have told that this probability that n arrivals are happening so, this is the counting process if you remember so, up to time t starting from t equal to 0 to t equal to t ok. So, up to this how many arrivals have occurred is a random thing, but it's just trying to count for any value of t, and any value of 1 means n what is the associated probability that something we have characterized and we have told it is actually following Poisson distribution right.

So, lambda t whole to the power n e to the power minus lambda t divided by factorial n this is something we have already characterized so, this particular formula we have derived. From there we have also derived another thing which is called this distribution we have talked about that also so, let us call that as suppose this is tau 1, this is tau 2. So, this is the inter-arrival time the time between two successive arrivals. That is a continuous distribution and we have also from here we could derive that distribution also and that follows exponential distribution.

So, this is actually lambda e to the power minus lambda t so, that PDF of let us say this inter-arrival time tau or inter-arrival time t ok. So, this follows exponential distribution whereas lambda we have also characterized that what was lambda? Lambda was the rate of arrival per sorry unit time ok. So, basically, this is per unit of time that is the.

So, this is the rate of arrival or the number of arrivals per unit of time. So, it is actually a number of per unit time ok. So, this is something and this particular thing because it is stationary so, this is not a function of t, it is not lambda t. So, basically, at any position in time, you try to measure lambda it will be the same, this is the property of the stationarity so, that remains lambda. So, this is constant over time.

So, therefore, the rate of arrival remains constant not only because this distribution is also stationary, but the arrival process is also stationary, all higher order statistics also remain stationary over time. So, anytime you take it will have a similar distribution so, it does not depend on where you start counting the 0. So, anywhere you put that 0 and count 40 amount of time you will see similar kinds of statistics ok.

So, that is something we have so far understood. Now, from here let us try to see what we understand by memorylessness so, let us try to characterize this memorylessness. So, memorylessness means let us again put the timeline let us say I had one arrival which has happened. So, I am just letting us say I take this portion and I am inflating it over here ok? So, this is where the arrival happens and I am declaring now as t equal to 0 ok.

Now, what I know is suppose from here up to this I have gone. So, let us say that is some tau we have progressed ok. So, let us say that is some tau we have already progressed, and what we know is no arrival has occurred over here. So, up to tau, t equal to 0, there was one arrival so, I am trying to define the memorylessness property in a probabilistic term.

So, at t equal to 0 I know there was one arrival after that the time has progressed and I have come to this situation so, this is my present time ok. So, that is my present time and in the present time I can see that from t equal to 0 to up to this present time t equal to tau there has not been any arrival that is my observation.

So, that is my history actually, this is called the history means when the last arrival happened and since then what means there was no arrival? So, this knowledge is history this is something which is history. Now, standing over here I am trying to see let us say I am trying to predict the future; that means, let us say I have another amount of time let us call that S ok.

So, therefore, this will be t equal to tau plus s. So, I am trying to predict from sitting here I am trying to predict that within this no arrival will be happening. Now, I know that inter-arrival time has statistics so, from here when the next arrival no arrival will be happening means the next arrival will be beyond this point. So, it might be somewhere here or here or somewhere which is beyond this t tau plus S ok.

So, that is the inter-arrival time; that means, the inter-arrival time has this distribution. Now, what I am trying to do is if I have elapsed some amount of time does it really matter how much time has elapsed with that I will be able to predict the history; that means, let us let me give a very simple example. So, let us say it is a deterministic arrival; that means, let us say I am standing on a platform and every 20 minutes a train arrives so, it is exactly having twenty minutes of these things.

Now, the last arrival you have seen and from there you count a history let us say 10 minutes have elapsed. Now, if I try to give a prediction then what will be my prediction the prediction depends on this one will be a very accurate prediction because it's a deterministic process it's not random, but that will depend on the elapsed time; which means, the history when was the last arrival happened I will take that data 10 minutes so, the next arrival will be happening in another 10 minutes.

So; that means, that when the next arrival will be happening that depends on this history in exponential that is the fundamentally different things in all other distributions this will not be happening. So, if there is an elapsed time you will be able to if it is uniformly distributed again you will be able to tell that if this much amount of time has gone.

So, basically, there is more probability that within the next some amount of time, the arrival will be happening ok. So, all other distributions will be the same matter of fact, but exponential will be able to prove that this history how much time has elapsed does not really matter.

The next arrival when that will be happening if I can stand over here and say maybe my arrival last arrival was only over there where I am standing from there I can predict whatever prediction I will be giving that is an accurate prediction in a probabilistic term right?

So, this is something will be able to prove and this is called the memorylessness property; that means, you are only bothered about the present state from the present state can you can predict something. So, I am over here and I have observed no arrival at this instant that is good enough, and from there you are trying to from this moment onwards want to predict what has happened in history or in the past I really do not care.

Whether the last arrival has happened 1 hour before or 3 hours before or 2 seconds before I do not care. My next arrival prediction will be equally likely ok. So, for all these cases all these cases will have the same similar probability of next arrival which is a very phenomenal thing you will see it has a huge application and implication in at least theoretical queuing analysis it's a very strong tool. This makes the whole queuing analysis probably mathematically possible, let us try to understand this part ok.

So, what let us try to see this thing mathematically let us capture. So, what we are trying to capture is that probability let us say this is my S n. So, which is the inter-arrival time; that means, what I am trying to do is my probability S n; which means, the inter-arrival time is definitely greater than tau plus S given I have a condition that I already know from my past experience that it is at least greater than tau because no arrival has happened before tau.

So, therefore, inter-arrival time I already know that it is greater than tau so, given that S n is greater than tau. So, this conditional probability I will have to now evaluate that will give me the probability that within this time next S amount of time no arrival will be happening given that up to tau no arrival has happened. So, this is what I am trying to evaluate.

So, this is a joint sorry this is a conditional probability so, you can apply Bayes theorem P A given B ok. So, that you can write you can write it this way. So, I will similarly write over here the probability that S n is greater than tau plus S and S n is also greater than tau divided by probability B which is this event. So, these are the two events right A and B, this is A and this is B.

Now these two joint events actually mean the same thing if something is greater than tau that will be greater than tau plus S. So, I can actually eventually write it because these two means if this condition is happening this will be automatically happening something is greater than a bigger number it will be; obviously, greater than a smaller number. So, we do not have to even specify that if this happens this will be happening.

So, this is eventually nothing but S n greater than tau plus S. Now what do I have to do? I just have to put this distribution I know this is the PDF ok? What was the associated CDF? You remember we have CCDF was e to the power minus lambda t and CDF was 1 minus e to the power minus lambda t ok. So, basically, this was the CDF; that means, the particular thing is between 0 to t and this tells it is beyond t.

So, that beyond t formula I will have to put over here so, exactly that I will be putting over here so; that means, this I will be putting CCDF because this should be beyond this tau plus S. So, it will be e to the power minus lambda e to the power minus lambda tau plus S divided by e to the power minus lambda tau what it is so, tau gets canceled. So, as you can now see this probability is not dependent on tau.

So, whatever the history it does not depend on that and that proves the memorylessness property and that is the phenomenal property of only exponential distribution, that is the only distribution that happens to exist in a continuous domain and for which this memoryless property is observable.

This is a very fundamental property of exponential distribution and this is something that will always happen as long as things are of this nature memoryless nature this will be always true. One of the memoryless processes we have already characterized is the arrival process you can take that so, in the arrival process this is what happens. If your service process is let us now see the beauty of this particular criteria.

So, let us say my service process. So, what is the service process? So, for every arriving customer how much time of the system he brings in for which the system will be occupied to serve him ok. So, this is if it is a packet how much time it takes for the transmitter to transmit the packet? So, this is the time it depends on the packet size of course, as we have told and the data rate.

So, if you have a 10 Gbps data rate and you have some let us say 1000 bytes of data or 1000 bit of data accordingly you can you can calculate how much time it will take to transmit the whole data. So, that is when a server is actually occupied to serve him. So, if you go to a fast food center you order something then you get the delivery so, between your ordering and then getting the things this is the service time actually.

So, this service time we have already talked about that is also a random thing. Now, if the service process is also following exponential and we call it independent exponential; that means, everybody is bringing exponentially distributed service time. So, basically, it is taken from an exponential distribution, but it is independent; that means, the first customer how much service he will be bringing its really not dependent on whether his previous customer or later to him whoever is coming what service time they will be bringing.

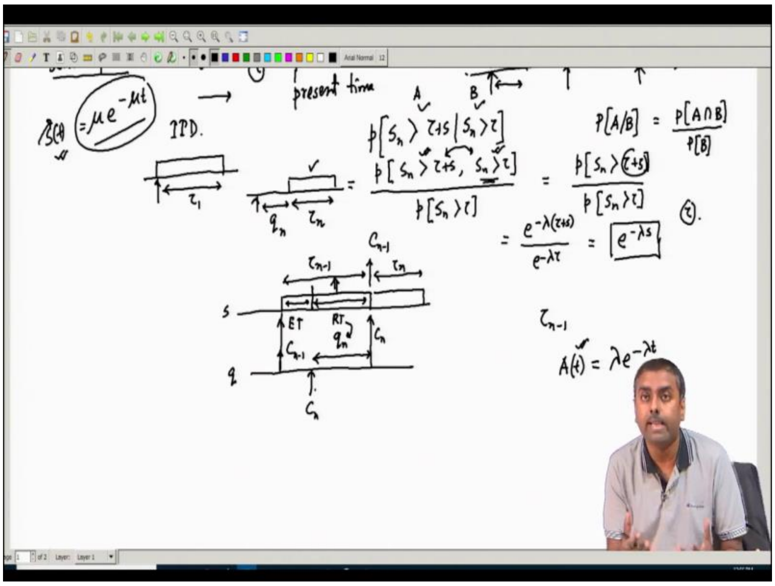
So, if these are all independent statistically independent and that service time actually follows an exponential distribution. So, it also follows something like this we can give a different parameter mu e to the power minus mu t where mu is the service rate now, earlier lambda was the arrival rate mu is the service rate; that means, in per unit time our server on an average how many he can serve ok.

So, this is also a stationary process, and it's taken from this all the services are taken from this. So, basically the same exponential distribution, but they are all independent. So, that is why we call IID ok, independently and identically distributed, or identically and independently distributed ok.

So, all these services actually take some amount of time. So, you make an arrival, and then from the server you request some amount of time, if you are immediately going to the server some amount of time that let us say the one you demand. If you do not immediately to service you wait for some amount of queuing let us call that as q n and then you get a service which is again some tau n for the nth customer ok. So, this is now IID exponentially distributed ok?

If it is exponentially distributed now the beauty comes due to memorylessness, let us try to understand this part.

(Refer Slide Time: 18:57)


present time
$S(t) = \mu e^{-\mu t}$
IID.
$P[S_n > \tau+s \mid S_n > \tau]$
$P[A/B] = \frac{P[A \cap B]}{P[B]}$
$= \frac{P[S_n > \tau+s, S_n > \tau]}{P[S_n > \tau]} = \frac{P[S_n > \tau+s]}{P[S_n > \tau]}$
$= \frac{e^{-\lambda(\tau+s)}}{e^{-\lambda\tau}} = \boxed{e^{-\lambda s}}$
τ_1
q_n
τ_n
C_{n-1}
τ_{n-1}
s
q
ET
RT
C_n
$A(t) = \lambda e^{-\lambda t}$

So, what happens is, if suppose there is some service which is going on we have shown that timeline. So, let us say n minus 1st customer has entered into the server ok. So, this is C n minus 1 and he goes out from the system C n minus 1 from the server and this has entered into the this one from the q this is q and this is server.

Now, in between somebody has arrived. Now if he starts counting his count of delays how much time does he have to wait? So, suppose this is the C nth arrival and he waits for this amount of time, this is his waiting time so, this is his q n and then at this instance only he will be joining the server and then he will be served for let us say this was tau n minus one and this is tau n. So, he will be served for this amount of time ok?

Now, if I try to see how much time he will have to wait he will have to wait his this queuing delay which is the amount of service left from the time he enters the system. As you can see this is the overall service time tau n minus 1 out of this some amount of time has already elapsed before he joins the queue.

So, always whenever you come to a system you will see somebody in the queue because it's a continuous process that is going on the services are going on continuously it is not that when you arrive then only somebody will be joining the queue that never happens you join independently.

So, you can actually cut a customer in between his service that is exactly what has happened over here. The nth customer has cut the customer the n minus 1st customer who is in service in between him he has joined the queue. So, basically, there is an elapsed time and there is a residual time.

So, this residual time is basically contributing towards the queuing ok. So, if there are multiple customers what will happen just this residual time will be there plus their service time will be added because after this the next customer will be served he will have to wait for that also then the next then the next. So, all the service time will be added.

So, that is very good, all service times are getting added, but first, whenever you join whoever is in service for him it is not the whole service time it is a fractional service time that is being added over here. Now, if we really have to characterize this system you have to keep track of how much time has elapsed which is really complicated.

You have to then characterize the system by how much time has elapsed for the customer who was being served plus additionally how many customers are waiting and what their service time distribution so, this information you will have to keep and that is why the analysis later on will see will appreciate also that the analysis becomes very complicated. Whereas, for exponential what is happening see the beauty of it exponential what I have just told.

Because of the memorylessness property, I do not really care that when he arrived or when he started initiating his service this residual time history has no consequence I can think that his service started only when I joined the queue. Because this distribution will still remain exponential with the same parameter of service time, I can take him statistically being a fresh customer full fresh customer who is being served after my arrival.

So, that reduces the complexity quite a lot it just says you count all the customers and that should be sufficient for your state description which is called the present state only. So, I do not really see what has happened in the past, the all the customers who are in the queue of course, I know they have to fully serve the customer who is in service if there are multiple servers then I have an even complicated scenario.

Then multiple customers will be in service and I have to keep track of all of them and what has happened to everybody I do not have to do that, right now if it is exponential then I do not have to care whenever I enter from there I can start fresh start everything I can forget about the history I do not have to remember how many how much amount of time they have elapsed their service has elapsed I do not have to keep track of all these things. I can statistically I can keep track of him as a fresh customer.

Then it is just customer count which summarizes the whole state that is the beauty of memorylessness. This will only happen you will later on see this will only happen for this kind of memorylessness distribution or memorylessness kind of systems. So, the system has to be memoryless, and that is why there is so, much emphasis on the memoryless system you will see a lot of places not only in queuing theory lot of places in machine learning many of you might have heard about the MDP Markov decision process.

So, everywhere people take advantage of this memorylessness. So, they say this memoryless process is also termed a Markovian process due to the scientist behind it Marco. So, because of him this was introduced this concept was introduced. So, they call the system to be Markovian once it is Markovian now summarizing the system is much easier because the state becomes reduced I do not have to keep track of it. So, many things like you have seen in queuing. So, we have seen we are just seeing one application of it, but this Markovian goes it is a very strong thing that is applicable to many other places.

So, this is what we would like to now focus on the process will be characterizing is Markovian you might be asking if this is a mathematical thing this is good mathematically, but whether this process has any relevance in the practical system fortunately for us is good. We were doing this trunk switch analysis remember we started all these things because we wanted to do the blocking probability of a trunk switch ok.

So, for these things, telephone calls and voice calls the duration of the call which is a service time right, duration of the call means that much amount of time the switch has to be blocked or some switch path has to be blocked for that particular connection or some trunk switch the trunk output has to be blocked for this particular connection ok.

So, this call duration has been seen historically that it actually follows exponential distribution and they are each customer has different, different means they are independent of each other how much time I will be talking it really does not influence others right. My neighbor how much time he will be talking to his let us say his relatives or friends that really does not get influenced by how much time I am talking to my father or my mother right it is not never like that.

So, therefore, they are IID and they are also historically seen that they are exponentially distributed. So, that is the advantage; that means, whatever theory will be developed by these is exactly applicable for designing trunk switches we will see that later on ok. So, therefore, means whatever we are doing is just not a mathematical exercise.

It has a huge practical implication you will see will derive formulas with which the traditionally all the trunk switch has been designed. So, far for almost the last 40 to 50 years, this has been the trend and this is still the play still case there is no other theory that describes it as even better. So, basically, that is what our target will be, but we have some more paths to go till we get to that situation where we can actually analyze the trunk switch.

But so, far what we have understood is very clearly that the arrival process the service process the poison arrivals are poison we have seen what property makes it a poison arrival this poison and exponential have some interrelationship. So, the counting process is any arrival counting process is poison inter-arrival becomes exponential we have also said that maybe our service processes are exponentially distributed which also historically has been seen through data.

So, this is also true, now we have characterized, the arrival and service process after taking this arrival to service process you remember these two are external things that are coming to a system now we are ready to go towards the queuing analysis. So, that is that will be our next step will be targeting the help of these things these understanding of exponential and Poisson processes and then we will be trying to see the interaction between a service process which is described by this.

On arrival process, which is described by this lambda into e to the power minus lambda t ok. So, if we call that as arrival process A t and this we call it as service process S t. So, these two things are also independent of each other because arrival never makes my service dependent.

So, this independence taking into account will be trying to see the timeline we have drawn from there all those queuing inferences what is the delay can I characterize it statistically what is the blocking probability when things will be blocked, can I characterize this statistically? So, those things we will try to see if can we do the average value analysis of them can do the probability calculation of those things. So, that something will be targeted in our next 2 3 classes ok.

Thank you.