Introduction to Queueing Theory

Prof. N. Selvaraju
Department of Mathematics
Indian Institute of Technology Guwahati, India

Lecture - 36
Renewal Process

Hi and hello, everyone. What we have seen so far are Markovian Queueing models, which are basically nothing but queueing systems modeled by a continuous-time Markov chain. We have seen that there are enough of such models themselves to actually model many of the related systems in a more realistic manner, especially when you generalize, to say, a phase-type distribution or something like that, and all features can also be studied. But still, what it is the case that they can be viewed as if they are the special cases of some Markov processes, analysis of some Markov processes, though the analysis that we carry out are within the context of queueing systems. So, that is that is what we have seen so far, whether it is BDP-based queueing models or more general Markovian queueing models, or Markovian network-based models. All these models are all basically built on are based on the Markov processes, or Markov chains, especially the CTMC, except the discrete type model was basically the DTMC model. Now, then all the theory that you developed in Markov chain were basically used to study this. But now, what we are going to do is that we are going to go a little bit beyond purely Markovian models to what are we call as semi Markovian models, where we still retain some amount of Markov property there in the whole process, but it is not that a fully Markov process. So, we will see I mean, as we go along, you will understand what this exactly means. To understand that, we need to extend our knowledge of Markov process to beyond something called, as you know, semi Markov processes. Now, that is what we are going to do as a stepping stone; what basically we define or consider for our analysis are what are called as renewal processes. And this renewal processes, in fact, played a good role in proving certain results of Markov process as well. Say, for example, you have seen that whenever the stationary distribution exists, the probability that you find the system state is n is $1/mean\ recurrence\ time$; the mean recurrence time is what then you have. So, that result basically follows immediately from a result of this renewal process.

So, like that, there are some books like Ross, for example; they first treat renewal processes, then they come to Markov process so that the proofs become easier and so on. So, this is also an important process, but in our context, this is an important process with respect to queueing models, in the way of generalizing our queueing models from the Markovian queueing models to something more general. So, we will give some definitions and some basic ideas about these processes before we actually jump into the semi Markovian queues. Let us start with what we call as renewal process. What is a renewal process? It is; basically, we are first defining it before we look at anything else. So, this is what is the definition.

Definition. *Let $\{X_n, n = 1, 2, \dots\}$ be a sequence of non-negative independent and identically distributed random variables with distribution function F and finite mean μ. Define the sequence $\{S_n, n \geq 0\}$ by*

$$S_0 = 0, \;\; S_n = S_{n-1} + X_n = X_1 + X_2 + \cdots + X_n, \quad n \geq 1$$

The random variable S_n is called the nth renewal time, while the time duration X_n is called the nth renewal interval. Further, define the random variable of the number of renewals until time t by

$$N(t) = \sup\{n : S_n \leq t\}$$

Then the continuous-time process $\{N(t), t \geq 0\}$ is called a renewal process with distribution F (or generated or induced by the distribution F).

So $\{N(t), t \geq 0\}$ is what is called the renewal process; here, the interest is mainly in the number of events that are happening. You can think you can associate this with number of certain kinds of an event happening. And this X_n is basically you can think of it as if the time between the consecutive events is what this X_n's denote and this S_n will give you the time of the nth event. And this $N(t)$ would define the number of events happened by the time.

Now, you can associate, for example, with respect to our queueing model itself if you think of it as an arrival. If you think of it as an arrival that, X_n is the inter-arrival time, S_n is the time for the nth arrival, and $N(t)$ is the number of arrivals by time t is what then you can think. So, in general, the word which is used is these renewals. We may also say that $\{X_n\}$ defines a renewal process, and some authors also say that $\{X_n\}$ itself is the renewal process.

Because in this definition, once X_n's are specified, then $S_n = X_1 + X_2 + \cdots + X_n$, and then $N(t)$ you are defining in terms of S_n, which is basically in terms of X_n. So, one can define this process $\{N(t)\}$ as a derivation of that, so $\{X_n\}$ itself one can call it as a renewal process.

So, what we have here, suppose if you look at the timeline here. So, this is 0. So, this is time I can call this as S_1; this is S_2, this is S_3, this is S_4, and so on; then this interval is basically what X_1, this is X_2, and this is X_3, and this is X_4 and so on. So, that is what you have here. Now, you see here this is what is the given X_1, X_2, X_3, X_4, which are all IID and positive; S_1, S_2, S_3, S_4 are the partial sums of this sequence you can say; then $N(t)$ is basically sup$\{n : S_n \leq t\}$ Suppose if I pick a t here, say, for example, here. Suppose if my t is somewhere here; suppose if I call this as my t, then the $N(t)$ is here. Suppose if my t is here, then the maximum of $\{n : S_n \leq t\}$, this set n, such that $S_n \leq t$ is 0, it is empty set, so the max supremum of that is 0 with the convention that we follow. So, $N(t)$ is 0. So, you will get if t is here, it is 0. If t is here, then you can see the maximum of n; see this is $S_1 \leq t$ is satisfied. So, the set is nonempty, and n is 1, but n 2 is not there, so it is only 1. So, $N(t) = 1$. So, here $N(t) = 2$; because for S_1 also this will be true, and for S_2 also this will be true; for S_3, this will not be true. So, my $N(t) = 2$ in this interval for any t in this interval. Now, in this interval, suppose if since I am taking this t here. So, my $N(t) = max\{1, 2, 3\}$. So, if you want to include 0, you can include 0 no problem; otherwise, you can see this thing here. So, $N(t) = max\{1, 2, 3\} = 3$ is what then you will get here. For any t in this interval, my $N(t)$ will be the maximum of n's. So, n is $1, 2, 3$. So, $N(t) = max\{1, 2, 3\} = 3$. So, basically, $N(t)$ gives you the counts the number of renewals that have happened by time t see this is what is the number of renewals has happened; because the fourth renewal happens or the fourth arrival happens after time t which is what you have taken as this point here. So, this is what is the renewal process. So, this is what it is again; you can think about it. Now, you can easily correlate this with your Poisson process; if I take my X_i's to be IID exponential, this is nothing but your Poisson process. Rather than taking any nonnegative IID with distribution function F, if I take an exponential, then what you are going to get or what you have got is basically the

Poisson process. So, this is, in a way generalization of this Poisson process; by removing that, we will just see that now.

Now, if $S_n = t$ for some n, then renewal is said to occur at time t, and hence S_n gives the time or epoch of the nth renewal and is called the renewal epoch, a regeneration epoch; mainly renewal epoch, regeneration because of some other reason, but I mean called so, which we will see little later, we will see why this word is also used. So, this $N(t)$ gives the number of renewals occurring in $[0, t]$. This X_n is the inter-event time or waiting time between the $(n-1)st$ and nth renewal. So, inter-event times are IID; that is what now we have seen.

Now, the Poisson process is the unique renewal process with Markov property; because we have seen from this kind of picture, we are drawn for the Poisson process as well, you can see that you can easily relate to it. This generalization of the Poisson process means that the renewal process is a generalization of the Poisson process. And if the renewal process is obtained, if you remove the restriction of exponentially distributed holding times and by considering that the inter-event times are any general IID nonnegative random variables, if you consider so, then what you get is renewal. So, between the Poisson process and the renewal process, this is the difference you have here.

So, any example that we have considered so far Poisson process arrivals happen, departure happens; wherever you got this exponential, if you remove, if you remove that exponential assumption and you replace with any IID positive random variable with the distribution function F and mean μ, then what you are going to get is a renewal process, that is the model that we have it in mind.

Example.
Consider a stage in an industrial process relating to production of a certain component in batches. Immediately on completion of production of a batch, that of another batch is undertaken. Suppose that the times taken to produce successive batches are IID random variables with distribution F. We get a renewal process with distribution F.

If you take arrivals, the inter-event times arrivals, if it is the case, then you can think of that as in the arrival context, in the queueing context that as a renewal process. If you take the service process; the duration of the service, the service times are all any positive IID random variable, then the service process is a renewal process. Any other process that we have considered, be it working vacation or anything, that wherever exponential is there if you can replace this, then what you get is the corresponding renewal process models. Of course, one has to look at how one can analyze such models, which we will come to a bit later; but as a way of generalizing the model, if you can think of a Markovian model, where the Poisson process is exponential distribution is there if you replace with, then what you get the corresponding renewal processes, that is the way. How to handle such a process and how to analyze it is another question that we have to look at.

- We will always assume that $P\{X_i = 0\} = 0$. The strong law of large numbers implies that $S_n/n \to \mu$ with probability one as $n \to \infty$. Hence $S_n < t$ cannot hold for infinitely many n and thus $N(t)$ is finite with probability one.

- We have the distribution function of S_n, for $n \geq 1$ as $F_n(x) = P\{S_n \leq x\} = F^{n*}(x)$, where F^{n*} is the n-fold convolution of F with itself.
  ► The above follows from the fact that, if X and Y are independent and distributed according to CDFs F and G, respectively, it has a different distribution function; just for the sake of consideration, you consider that these are

F and G. Then

$$P\{X+Y \leq t\} = F * G(t) = \int_0^t G(t-u)dF(u), \quad \text{for all } t \geq 0.$$

In discrete random variables, it is summation say p and r are the PMFs of these random variables; then we know $\sum_i p(n-i)r(i)$ is what is the quantity that you will have, which is exactly the same thing in this particular case with respect to the distribution function which you write, this is what we call it as the convolution. Now, what is $F^{n*}(x)$? This is n-fold convolution; $\int_0^t G(t-u)dF(u)$ is two-fold in a way, so this n-fold convolution is what then you get. Of course, if you are working in the transform domain, in this particular case, this is nothing but the nth power of the transform of the random variables X_i's. But when you are working in the time domain, this will become the n-fold convolution; that is what it is.

- It can be shown that $\lim_{t\to\infty} \frac{N(t)}{t} = \frac{1}{\mu}$ holds with probability one. Therefore, the quantity $1/\mu$, here you have to be careful, this mu is different from in Poisson process λ or μ that we have used, (i.e., the inverse of the mean length of a renewal interval) is called the **rate** of the renewal process.

If you want to call say something as the rate of the renewal process, much like the case that you had for in the case of the Poisson process, then you would call this $1/\mu$, where μ is the mean of the inter-event times or in that renewal times. So, that $1/\mu$ is what you call the rate of the renewal process. So, we have to remember that the distribution function this F_n, which is basically of S_n, is basically given by the n-fold convolution of the distribution function of X because X_n's are IIDs. So, that is what we are going to construct, and we will see what that is and so on.

- Observe that $\{N(t) \geq n\} \Leftrightarrow \{S_n \leq t\}$ (or equivalently $\{N(t) < n\} \Leftrightarrow \{S_n > t\}$).
  Therefore, the distribution of $N(t)$ is given by

$$\begin{aligned} p_n(t) = P\left\{N(t)=n\right\} &= P\left\{N(t) \geq n\right\} - P\left\{N(t) \geq n+1\right\} \\ &= P\left\{S_n \leq t\right\} - P\left\{S_{n+1} \leq t\right\} = F_n(t) - F_{n+1}(t) \\ &= F^{n*}(t) - F^{(n+1)*}(t). \end{aligned}$$

- The function $M(t) = E(N(t))$ is called the renewal function of the renewal process with distribution F. The renewal function plays a fundamental role in renewal theory. The expected number of renewals in $[0,t]$ is given by

$$\begin{aligned} M(t) = \sum_{n=0}^{\infty} n p_n(t) &= \sum_{n=0}^{\infty} n\left\{F^{n*}(t) - F^{(n+1)*}(t)\right\} = \sum_{n=1}^{\infty} F^{n*}(t) \\ &= F(t) + \sum_{n=1}^{\infty} F^{(n+1)*}(t). \end{aligned}$$

- Now, observe that $\sum_{n=1}^{\infty} F^{(n+1)*}(t) = \sum_{n=1}^{\infty} \int_0^t F^{n*}(t-x)dF(x) = \int_0^t \left\{\sum_{n=1}^{\infty} F^{n*}(t-x)\right\} dF(x)$, assuming that interchange of summation and integration is valid.

- Substituting the above in $M(t)$ above, we get the fundamental equation of renewal theory or renewal equation given by

$$M(t) = F(t) + \int_0^t M(t-x)dF(x).$$

So, the renewal equation is nothing but the integral equation; here, this is an integral equation because $M(t)$ is the one you need to determine, and this $M(t)$ appears inside this integral and also outside. So, this is an integral equation. So, this is the integral equation satisfied by the renewal function, what is called the renewal equation or the fundamental equation of renewal theory.

In the study of the renewal process, the major portion is concerned with the properties of this $M(t)$, and from there, what is the inference that you can give with respect to the renewal process also.

- Renewal theorems (elementary renewal theorem, Blackwell's theorem, key renewal theorem) involving limiting behaviour of $M(t)$ are powerful results in renewal theory and are important from the point of view of applications (Refer to any standard text, like Ross).

Of course, we require that much, so we just start going into details and also for lack of time. For interested people, you can always refer to this.

So, the renewal function means $M(t) = E(N(t))$, and the renewal function satisfies $M(t) = F(t) + \int_0^t M(t-x)dF(x)$ renewal equation. So, once I have this renewal equation, you can write the solution of this also; the solution will be in terms of again renewal function and so on, so that is a different matter. Now, the two quantities here that are of interest and will have to be handled carefully compared to our earlier models are called residual life types and excess lifetimes. So, basically, what do we have?

- We consider two random variables of interest in renewal theory. For any given $t > 0$, there corresponds a unique $N(t)$ such that

$$S_{N(t)} \le t < S_{N(t)+1} \quad \Big\{\text{i.e. } t \text{ falls in the interval } \ X_{N(t)+1}\Big\}$$

- **The residual (or excess) lifetime** at time t is given by the time from t to the next renewal epoch, i.e.

$$Y(t) = S_{N(t)+1} - t \quad \{\text{It is also called forward recurrence time at } t\}$$

- **The spent (or current) lifetime or age** time t is given by the the time to t since the last renewal epoch, i.e.

$$Z(t) = t - S_{N(t)} \quad \{\text{It is also called backward recurrence time at } t\}$$

So, if I look at the diagram here, basically, this is my $S_{N(t)}$, and this is my $S_{N(t)+1}$, and if I call this as t, then this quantity is what you call it $Z(t)$, and this is you call it as $Y(t)$. So, that is what we are saying here. So, this is also called age; because this is what is happening, also called backward recurrence time.

- The total lifetime at t (or length of the lifetime containing t) is given by

$$Y(t) + Z(t) = S_{N(t)+1} - S_{N(t)} = X_{N(t)+1}.$$

You can think about this in the queueing context as well. Suppose if you are thinking about this arrival, the same thing with service or anywhere even, but let us stick to arrivals here. So, if the events that we are talking about here are arrivals, that means there is an arrival here at this point, there is an arrival at this point, this is your current time. So, this

$Z(t)$ would then denote that it is the elapsed time since the last arrival, and this is the time to the next arrival from the current time point. Now, in the Poisson process case or in the Markovian queueing models' case, when we associate this relationship for the time interval, the first one the arrivals at any given point of time; because of this memoryless property, this distribution is that whether you are looking at is $Z(t)$ or $Y(t)$, they are all like simply exponentials. So, this mainly we will be interested about time for the next arrival, so that will be exponential because of the memoryless property. So, that is the reason there you had, but here since it is exponential distribution is the only distribution that has the memoryless property. So, you do not expect; that when you assume a general distribution for inter-arrival times or service times, you do not expect that exponential property to hold here. So, we need to consider this exclusively if we are interested in, say, residual lifetime or an excess lifetime, and so on. So, this will play a critical role wherever, for example, wait in time computation; at whatever time it is you are arriving, now the time when the last customer arrived and time now, like in the earlier case what we did? You just counted in front of you how many customers are there; you just summed up them, all of them are basically exponential. Suppose if it is not, if it is service is happening with respect to general distribution, then you have to compute the remaining service time for the customer who is currently in service and add up to the remaining ones who are waiting in the queue. So, you have to split that into this case. Now, to compute that, you need the distribution of the remaining service time, which is basically this $Y(t)$ process. So, that is where the modification will come, and hence this is important in the analysis; if you are using a renewal process as opposed to a Poisson process for arrival or service and so on. Now, before we talk about its distribution, we define this notion of what is called as a random variable or its distribution function being lattice. What do we mean?

- These random variables $Y(t)$ and $Z(t)$ arise naturally in queueing contexts (e.g., arrivals, departures).
- *Definition:* A non-negative RV X (and also its CDF F) is called **lattice** if there is a positive number $d > 0$ with $\sum_{n=0}^{\infty} P\{X = nd\} = 1$. If X is lattice, then the largest such number d is called the period of X (and F).
- **The distribution of** $Y(t)$ can be obtained as

$$P\{Y(t) \leq x\} = F(t+x) - \int_0^t [1 - F(t+x-y)]dM(y), \quad x \geq 0 \quad \text{[and 0 for } x \leq 0].$$

  If F is non-lattice, then the limiting distribution of $Y(t)$ is

$$P\{Y \leq x\} = \lim_{t\to\infty} P\{Y(t) \leq x\} = \frac{1}{\mu}\int_0^x [1 - F(y)]dy, \quad x \geq 0.$$

- Noting that $\{Y(t) > x\} = \{Z(t-x) > x\}$, the distribution of $Z(t)$ can be deduced as

$$P\{Z(t) \leq x\} = \begin{cases} 0, & x \leq 0 \\ F(t) - \displaystyle\int_0^{t-x} [1 - F(t-y)]dM(y), & 0 < x \leq t \\ 1, & x > t \end{cases}$$

  If F is non-lattice, then the limiting distribution of $Z(t)$ is

$$P\{Z \leq x\} = \lim_{t\to\infty} P\{Y(t) \leq x\} = \frac{1}{\mu}\int_0^x [1 - F(y)]dy, \quad x \geq 0.$$

- When these exist, the two limiting distributions Y and Z are identical. It can be easily verified that for exponential X_i, the distributions of $Y(t)$ and $Z(t)$ are again exponential with the same mean $\mu = E(X_i)$.

- The mean of Y and Z can obtained as $E(Y) = E(Z) = \dfrac{E(X_i^2)}{2E(X_i)}$.
- If F is a lattice distribution, then the distributions of $Y(t)$ and $Z(t)$ have no limits for $t \to \infty$ except in some special cases.

Now, some generalizations of this renewal process. So, let us look at here what is called first a delayed or modified renewal process.

So, if we start with the process at time 0, if there is a renewal that happens, and if the process is continuing, then it is the ordinary renewal process, which is the case. For example, if you have a tube light or bulb which you are installing for the first time and then, from then you are counting the time of its failure, the next renewal, renewal, renewal so and so on it goes on. Rather if you are counting, for example, the bulb was installed at some point of time; you do not know when, but now you are starting your count at this point of time. So, the bulb was in operation for some time before. So, you are starting at this point of time; the bulb will fail after some time. So, basically, this the first one then is not the same as the subsequent ones; because this is not the point of renewal, this is the point of non-renewal; like renewal happened sometime before, which you do not know, you did not capture. So, at this point, only you are observing.

- **Delayed (modified) Renewal Process:** First, suppose that the first inter-arrival time X_1 (i.e. time from the origin upto the first renewal) has a distribution G which is different from the common distribution F of the remaining inter-arrival times $X_2, X_3 ....$ i.e. the initial distribution G is different from subsequent common distribution F. We then get what is known as a modified or delayed renewal process. Such a situation arises when the component used at $t = 0$ is not new. When $G \equiv F$, the modified process reduces to the ordinary renewal process.
- **Alternating renewal processes.** Consider a stochastic process $\{X(t), t \geq 0\}$ with state space $\{0, 1\}$. Suppose the process starts in state 1 (also called the 'up' state). It stays in that state X_1 amount of time and then jumps to state 0 (also called the 'down' state). It stays in state 0 for Y_1 amount of time and then goes back to state 1. This process repeats forever, with X_n being the nth up time, and Y_n the nth down time. The nth up time followed by the nth down time is called the nth cycle.

Example.
We consider the working of a component, the lifetime (or time to failure) being given by a sequence $\{X_n\}$ of IID random variables, on the assumption that the detection of failure and repair or replacement of the failed component take place instantaneously. Here the corresponding system has only one state-the working state and a renewal occurs at the termination of a working state (ar failure af a component).

Consider now that the detection and repair or replacement of a failed item are not instantaneous and that the time taken to do so is a random variable. The system then has two states-the working state and the repair state (during which repair of the failed component or search for a new one is under way). Here the two sequences of states-the working states and the repair (failed) state alternate. Suppose that the duration of the working states (or lifetimes or limes to failure) are given by a sequence IID random variables and the duration of repair states (times taken to repair or search) are given by a sequence of IID random variables. We have then an alternating renewal processes or two-stage renewal process.

So, this is a what all about renewal process that we want to get some idea about; basically, what is the renewal process, how this is connected with other processes that we have seen earlier. A Poisson process is a Markov process; whenever you have this, the renewal intervals or IID exponential, it becomes a Markov process and Poisson process especially. So, this is a generalization of what we have considered. And then the additional notion required in the excess lifetime mainly, its remaining lifetime or time for the next arrival in our context; a time for the service completion of the ongoing service, when you look at an arbitrary time is what these quantities are in effect. So, this is what one has to consider and how we will consider everything that we will do as we go along. So, this is about the renewal process. So, the semi Markov process thing like we will take it up in the next lecture.

Thank you, bye.