Introduction to Queueing Theory

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

Lecture - 37

Regenerative Processes, Semi-Markov Processes

Hi and hello, everyone. What we have seen in the previous lecture were these renewal processes and the related ideas. So, what we will see next what we are going to do next is we want to consider or get the idea of a Semi Markov Process. But before going that, what we will define as an intermediate process or a generalization of this renewal process ideas is what we call a regenerative process. So, what is that? So, this is what we define first; we will come back to the specifics.

Definition. *[*Regenerative Process*]*
A stochastic process $\{X(t), t \geq 0\}$ *is called a regenerative process if there exists time points* $0 = T_0 < T_1 < T_2 < \ldots$ *such that, for* $n \geq 1$*,*

- *the process* $\left\{X\left(T_n + t\right), t \geq 0\right\}$ *is independent of the process* $\{X(t), 0 \leq t < T_n\}$.
- *the processes* $\left\{X\left(T_n + t\right), t \geq 0\right\}$ *and* $\{X(t), t \geq 0\}$ *have the same joint distribution.*

The T_n*'s are called regeneration epochs (times or points) and the lengths* $T_1 - T_0, T_2 - T_1, \ldots$ *are called regeneration cycles.* T_n *'s are IID random variables and* $\{T_n, n \geq 0\}$ *defines a renewal process. The renewal process is said to be embedded in* $\{X(t)\}$ *at the epochs* $T_1, T_2 \ldots$*. Every time a renewal occurs a cycle is said to be completed.*

Thus, a regenerative process is a stochastic process with time points starting from which the process is a probabilistic replica if the whole process starting at 0.

Example.

1. So, a renewal process, for example, is a regenerative process, and that is why the renewal points or renewal times are also called regeneration points because it is one and the same. But T_n has a much more generic meaning; the T_n represents the time of the nth renewal, which corresponds to S_n there.

2. So, a recurrent Markov chain is a regenerative process with T_n being the time of nth recurrence, an nth visit to the state any particular state you pick it up. So, you are looking at it because it is a recurrent Markov chain; the process will keep visiting itself, and then the nth visit is what is you would call as T_n here.

3. In an $M/G/c$ queue, for example, whenever the queue is empty, then all servers are idle, then only the arrival process has an effect on the future. So, thus the system process, if you look at it carefully, regenerates at those

time points at which the system becomes empty, and those time points could be, nth time empty, it becomes empty, so that is what will be these T_n's here. So, the durations $T_{n+1} - T_n$ between the successive empty system that you would find are IID, and hence this $M/G/c$ system process is a regenerative process.

Example. *[Alternating Renewal Process]*
Another example of a regenerative process is an alternating renewal process. Such a process can be envisaged by considering that a system can be in one of two possible states - say, 0 and 1. Initially, it is at state 0 and remains at that state for a time Y_1, and then a change of state to state 1 occurs in which it remains for a time Z_1, after which it again goes to state 0 for a time Y_2 and then goes to state 1 for a time Z_2 and so on. That is, its movement could be denoted by $0 \to 1 \to 0 \to 1 \ldots$. (For the initial state of 1, the movement sequence is $1 \to 0 \to 1 \to 0 \ldots$)

- *Suppose that $\{Y_n\}, \{Z_n\}$ are two sequences of IID random variables and that Y_n and Z_n need not be independent, Let*

$$T_n = T_{n-1} + Y_n + Z_{n,} \quad n = 1, 2, \ldots$$

  Then at time T_1 the process restarts itself, and so also at times $T_2, T_3 \ldots$. The interval $T_n - T_{n-1}$ denotes a complete cycle, and the process restarts itself after each complete cycle.

- *Let $E[Y_n] = E[Y], \quad E(Z_n) = E(Z)$. Then the long-run proportions of time that the system is at states 0 and 1 are, respectively, are*

$$p_0 = \lim_{t\to\infty} P\{X(t) = 0\} = \frac{E(Y)}{E(Y) + E(Z)}$$
$$\textit{and} \quad p_1 = \lim_{t\to\infty} P\{X(t) = 1\} = \frac{E(Z)}{E(Y) + E(Z)} = 1 - p_0.$$

If you are looking at the long-term fraction of time or long-term proportion of the time the system spends in upstate or downstate is given by this expression. So, this is what you need to know.

- The results of the alternating renewal process example given above have an important application in queueing theory.

- Consider a single-server queueing system such that an arriving customer is immediately taken for service if the server is free, but joins a waiting line if the server is busy.

- The system can be considered to be in two states (idle or busy) according to whether the server is idle or busy.

- The idle and busy states alternate and together constitute a cycle of an alternating renewal process. A busy period starts as soon as a customer arrives before an idle server and ends at the instant when the server becomes free for the first time.

- The epochs of commencement of busy periods are regeneration points. Let I_n and B_n denote the lengths of nth idle and busy periods, respectively, and let

$$E(I_n) = E(I) \quad \text{and} \quad E(B_n) = E(B).$$

- Then the long-run proportion of time that the server is idle equals

$$p_0 = \frac{E(I)}{E(I) + E(B)} \tag{1}$$

and the long-run proportion of time that the server is busy equals

$$p_1 = \frac{E(B)}{E(I) + E(B)}. \tag{2}$$

So, we have already used this when we talked about the busy period analysis when we consider the birth-death queueing models. We wrote this we used this scenario or we showed that if you obtain from the distribution whatever mean that you are obtaining you can obtain through different argument, but for this is what is one can see.

- *Remark:* If the arrival process is Poisson with mean λt, then it follows (from its lack of memory property) that an idle period is exponentially distributed with mean $1/\lambda$, i.e, $E(I) = 1/\lambda$. Then when p_0 or p_1 is known, $E(B)$ can be found.

So, this is how one can use that alternating renewal process idea in queueing theory as well. So, now having known what a regenerative process is, we will come to the semi Markov process and the Markov renewal process. These are connected; that is why we write this together whenever these are treated together. So, this is a special class of regenerative processes, and this is very important for the analysis of many of our queueing systems. In fact, if you want to study it, the real queueing theory starts with this kind of semi Markov model, not just the Markovian model. Because there you do not feel the nature of the queueing problems. But when you consider the semi Markovian queues, you will feel the nature of these queueing models and queueing problems associated with that and so on; that is why for many people like they say that it just starts with the semi-Markovian thing only they will consider this as a queueing theory part. So, the Markovian queues give enough ideas for us to generalize to semi Markovian. It is; after all, ultimately, it is just the Markov CTMC's that then you are studying. Of course, here also, all processes are stochastic processes in that sense. So, there is no difference. So, we can as well take the Markovian process itself as the starting point and this particular class, this semi Markov process or Markov renewal process. This class generalizes the Markov processes and renewal processes at the same time. This is a generalization of both the Markov process as well as the renewal process. But it is a special class of regenerative processes. Of course, there are further processes one can go further, which we are not going we will stop at semi Markov process only because that is what you will need.

So, let us first define what this process is all about.

Definition.
Let S denote a countable state space. For every $n = 0, 1, 2, \ldots$, let X_n denote a RV on S and T_n a nonnegative RV such that $0 = T_0 < T_1 < T_2 < \ldots$ and $\sup_{n\to\infty} T_n = \infty$ almost surely. Define the process $\{Y(t), t \geq 0\}$ by

$$Y(t) = X_n \quad \textit{for} \quad T_n \leq t < T_{n+1}$$

for all $t \geq 0$. If

$$P\{X_{n+1} = j, T_{n+1} - T_n \leq u | X_0, \ldots, X_n, T_0, \ldots, T_n\} = P\{X_{n+1} = j, T_{n+1} - T_n \leq u | X_n\}$$

hods for all $n = 0, 1, 2, \ldots, j \in S$, and $u \geq 0$, then $\{Y(t), t \geq 0\}$ is called a semi-Markov process on S.

The sequence of random variables $\{(X_n, T_n), n \geq 0\}$ is called the embedded Markov renewal chain (or simply the Markov renewal process).

- The semi-Markov process is *homogeneous* if

$$Q_{ij}(t) = P\{X_{n+1} = j, T_{n+1} - T_n \leq u | X_n = i\}$$

  is independent of n. We consider only this case.

- By definition, a semi-Markov process is a pure jump process.
  Because its sample paths look like the Poisson process's sample paths. So, there are not many things that you would find in this particular, but the Poisson process is increasing.

  So, this is a pure jump process. So, it could be up or down; that is what you will have; the only thing is the duration, and everything would depend on the state you are going to visit next and the current state you are in, and so on.

- So, the sample paths are step functions; they could be up and down.

- By construction, the semi-Markov process is determined by the embedded Markov renewal chain and vice versa.

- Let $\{Y(t), t \geq 0\}$ is a homogeneous Markov process with S and parameters $\lambda_i, i \in S$ for the exponential holding times. The embedded Markov chain $\{X_n\}$ has the transition matrix $P = (p_{ij})$. Then $\{Y(t)\}$ is a semi-Markov process with $Q_{ij}(t) = p_{ij}\left(1 - e^{-\lambda_i t}\right)$ for all $i, j \in S$.

  Thus, for a Markov process, the distribution of $T_{n+1} - T_n$ is exponential and independent of the state entered at time T_{n+1}. These are the two features for which the semi-Markov process is a generalization of the Markov process on a discrete state space (i.e., CTMC).

So, that is what explicitly it models that it is also j, and $T_{n+1} - T_n$ distribution need not be exponential; it could be any general distribution is what then you will get as. Because this T_n's you just assumed this way, and it is a renewal process T_n.

- It can be shown easily that, for a semi-Markov process $\{Y(t), t \geq 0\}$ with embedded Markov renewal chain $\{(X_n, T_n), n \geq 0\}$, the chain $\{X_n, n \geq 0\}$ is a Markov chain (called embedded Markov chain). And, we denote its TPM by $P = (p_{ij})_{i,j \in S}$. Then the following relation holds for all $i, j \in S$:

$$p_{ij} = P\{X_{n+1} = j | X_n = i\} = \lim_{t \to \infty} Q_{ij}(t).$$

- According to its embedded Markov chain $\{X_n\}$ we call a semi-Markov process irreducible, recurrent or transient.

  If the embedded Markov chain is irreducible, we say the semi Markov process is reducible. If the embedded Markov chain is recurrent, then the semi Markov process is recurrent; if the embedded Markov chain is transient, then the semi Markov process is also transient; much like when we generalized the DTMC to CTMC, we call in a similar fashion, it is the same way we are doing it here.

  An irreducible recurrent semi-Markov process is regenerative, as one can fix any initial state $i \in S$ and the times of visiting this state to be a renewal process.

- Define $F_{ij}(t) = Q_{ij}(t)/p_{ij}$ for all $t \geq 0$ and $i, j \in S$ if $p_{ij} > 0$, while $F_{ij}(t) = 0$ otherwise. Then, this can be interpreted as

$$F_{ij}(t) = P\{T_{ij} \leq t\} = P\{T_{n+1} - T_n \leq t | X_n = i, X_{n+1} = j\},$$

  i.e, this is the distribution function of T_{ij}, the conditional sojourn time at state i given that the next transition is to state j.

  Then, the unconditional sojourn time at state i equals $\tau_i = \sum_j p_{ij} T_{ij}$.

It is not the example but how you can put other processes in terms of this Markov renewal process.

Example. A pure-birth process is a special type of Markov renewal process with

$$\begin{aligned} Q_{ij}(t) &= 1 - e^{a_i t}, \quad j = i+1, \\ &= 0, \quad \textit{otherwise} \end{aligned}$$

Then

$$\begin{aligned} p_{ij} &= 1, \quad j = i+1, \\ &= 0 \quad \textit{otherwise} \\ F_{ij}(t) &= Q_{ij}(t), \quad \tau_i = T_{ij}, \quad j = i+1 \end{aligned}$$

- A Markov renewal process becomes a Markov process when the transition times are independent exponential and are independent of the next state visited.

- It becomes a Markov chain when the transition times are all identically equal to 1.

  There is everything is 1; that means, each step that it moves so, it becomes a Markov chain, a DTMC to say.

- If the state space S is trivial, i.e., there is only one state, then the increments are IID. And, in this case, $\{T_n, n \geq 0\}$ defines a renewal process.

  So, that is how this is generalizing both the Markov process and renewal process to the Markov renewal process or semi Markov process.

- Semi-Markov processes are used in the study of certain queueing systems.

Now, we will just state asymptotic behavior though we are not going to prove this.

- Let $p_k = \lim_{t \to \infty} P\{Y(t) = k\}$ for $k \in S$.

- So, we said that semi Markov process inherits the properties from its embedded Markov chains Markov chain and the embedded Markov chain. Suppose that the embedded Markov chain $\{X_n\}$ is irreducible and positive recurrent with stationary distribution $\{\nu_j, j \in S\}$; that means that ν_j is the long and proportion of the time you find in the state j, and also it is the proportion of the time you spend in state j that is what you remember. That is, $\nu_j = \lim_{n \to \infty} p_{ij}^{(n)}$ exists and are given as the unique nonnegative solution of

$$\nu_j = \sum_{k \in S} \nu_k p_{kj}, \quad j \in S, \quad \sum_j \nu_j = 1.$$

Then, we would expect that p_k to be proportional to $\nu_k\mu_k$, i.e.,

$$p_k = \frac{\nu_k\mu_k}{\sum\limits_{j\in S}\nu_j\mu_j},$$

where $\mu_k = E(\tau_k)$ is the expected sojourn time in state k until the next transition happens at time T_{n+1}.

So, this is the duration μ_k, and $\nu_k\mu_k$ is the proportion. So, this is then the p_k would then be proportional to $\nu_k\mu_k$ and hence my p_k would be given by $\frac{\nu_k\mu_k}{\sum_{j\in S}\nu_j\mu_j}$ because now this p_k's must be sum to 1. So, this is what is the expression that you would get here. So, one can rigorously prove this for which you have to start from regenerative process some proofs, and then you have to apply that one here to get to this, but it is an intuitive idea that is clear.

So, if p_k if I want to look at it, which is the long-term probability of finding the semi-Markov process $Y(t)$ in state k. Then this will be connected to the embedded Markov chain stationary distribution and the expected sojourn time in state k in that embedded Markov chain, which is $\nu_k\mu_k$ proportion out of the total proportion of this is what then you will be getting it here.

So, this is what you would get as a asymptotic behavior; of course, one can study a lot, but we will not need that much. So, whatever little bit relevant, what is this process. So, what we are going to do when we consider semi Markov process is that we are going to consider the embedded Markov chains and then study. Because the underlying process is a semi Markov process and we look at the embedded Markov chains for the analysis of the queues. So, what is this embedded Markov chain mean, and what is the semi Markov process. We need to get an idea, and that is why we have now given this idea of simple ideas from what is a semi-Markov process, Markov renewal process, how it generalizes from Poisson and renewal processes or a regenerative process means, and so on is what we have seen in this case. So, now, with these ideas in mind, though it is not very elaborate, it is very brief. It is sufficient enough we will go ahead with the analysis of semi-Markov queues in the subsequent lectures. We will see next in the following lectures.
Thank you, bye.