Introduction to Queueing Theory

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

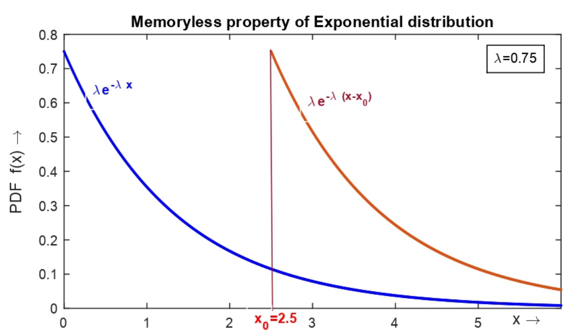
Lecture - 09
Exponential Distribution and its Properties, Poisson Process

Hi and hello, everyone; what we will see next is the "Continuous Time Markov Chain" or discrete-state Markov processes, continuous-time discrete-state Markov process, which we are calling as a continuous-time Markov chain. Before we go to the continuous-time Markov chain, first, we will see a simple version of it which is what is the Poisson process that is what we are going to see first. So, for the Poisson process to start with like, we will start with what you already know as the exponential distribution and some of its properties because this is more important to understand the properties first because this is what is going to be used throughout the continuous-time Markov chain theory. In our course on queueing theory like, what we are going to see is the first, at least a major portion of the course relies on the Markovian models. So, and hence this is the one distribution that is crucial. So, it is a widely used distribution in queueing theory because of its nature and connection with Markov chains. And to get familiarized with notation and other stuff like what we are using, $Exp(\lambda)$ within bracket lambda means an exponential distribution with parameter λ, but $\lambda(> 0)$.

And the PDF is $f(x) = \lambda e^{-\lambda x}, x \geq 0$, of course; this is 0 otherwise, is what is the form. Since there are many different forms of this exponential distribution, so, it is better that we familiarize ourselves with what is our form. So, sometimes some books would call negative exponential and so on, but it does not matter. We will call simply as an exponential for $\lambda e^{-\lambda x}$ form with parameter λ, which we denote by $Exp(\lambda)$, which has its distribution function given by $F(x) = P\{X \leq x\} = \begin{cases} 0, & x < 0 \\ 1 - e^{-\lambda x}, & x \geq 0. \end{cases}$
as you know already. And the complementary CDF or complementary distribution function $\bar{F}(x)$ basically represents $P\{X > x\} = 1 - F(x)$. And $E(X) = \frac{1}{\lambda}$ and $Var(X) = \frac{1}{\lambda^2}$. Sometimes in order to you know make it a little bit clear in the general context, it will be written as exponential distribution with mean $1/\lambda$. Exponential distribution with mean $1/\lambda$ we mean its support is 0 to ∞ and the form is $\lambda e^{-\lambda x}$ then only you will get mean $1/\lambda$. So, that is also another terminology that is used if that is convenient sometimes to represent. Rather than exponential distribution with mean $1/\lambda$ is what would be written to express this particular thing. And because of $E(X) = \frac{1}{\lambda}$ and $Var(X) = \frac{1}{\lambda^2}$, the coefficient of variation for exponential distribution is unity, it is equal to 1, and that is the important one of the properties of this exponential distribution. For us, in our analysis of our queueing systems, this random variable X would represent the quantity of time, and the parameter λ would then represent a rate that has units of events per time. So, that is sometimes also called "**rate**" because of the interpretation that one does here because whenever this represents a quantity in time, for example, the time between two arrivals, the time it takes to serve a particular customer, and so on.

So, X represents time, and in that case, λ would represent the rate. Now for the continuous positive random variable X having distribution function F_X and the PDF f_X, the failure rate or hazard rate is d actually. The hazard rate function is defined by $r(t)$; it is $\frac{f_X(t)}{1-F_X(t)}$ because we know what F_X is and what is $1 - F_X$ and what is f_X for the exponential. So, for the exponential case, this will be a constant. So, the constant hazard rate function would represent the exponential distribution case. What it does is that it helps us to compute the probability that there will be a failure by time $t + dt$, given that there has been no failure up to time t. So, that is what $r(t)$ will help us to compute, and that is what you will see later. Now, this exponential distribution satisfies $P\{X > t + s|X > t\} = P\{X > s\}$, for any $t, s \geq 0$. So, you can check this easily, and this property is what is called a memoryless property; again exponential random variable satisfies this memoryless property, and it can be shown that the exponential distribution is the only continuous distribution with this property.

The only other distribution that has this property is the geometric distribution which is the discrete analog of this exponential distribution. So, what does it tell? Suppose if you are looking at our queueing context. Suppose an arrival from the previous arrival to time up to now like if it is time t. So, if this is there is no arrival, then what is the probability that the arrival will happen after s time units from now, the current time is t that if you think it. This will be the same as if you are starting it is from this time and the probability that the arrival will happen after s time units. That is what $P\{X > t + s|X > t\} = P\{X > s\}$ would represent; if you have to put a context on this particular case. Now to understand or to appreciate this memoryless property a bit more, let us look at this figure that we have here.



- ***Characterizations***:
  ▶ X has exponential distribution if and only if $E(X|X > y) = y + E(X)$ for all y.

  You can intuitively you can understand this must be true because of this memoryless property because it does not matter whether you have crossed up to time y the expectation because the distribution is going to be the same, so the expectation is going to be the same. So, it is simply adding up to $y + E[X]$. Or you can think that expectation of $X - y|X > y$ is $E[X]$ for all y; that is what you can show it here also. So, this is one characterization. The other one is there are many characterizations, but these are a couple of things that might be of some help at some point of time.

  ▶ Two independent continuous variables X_1, X_2 are exponential if and only if $Z = \min\{X_1, X_2\}$ and $W = X_1 - X_2$ are independent. Now, let us look at some more properties.

Now, let us look at some more properties.

- The LT or LST of an exponential RV is $\lambda/(s+\lambda)$.

- If $X_1, X_2, \ldots, X_n$ are i.i.d. $Exp(\lambda)$, then $\sum_{i=1}^{n} X_i \sim Gamma(n, \lambda)$.

  Recall: PDF of $Gamma(n, \lambda)$ is: $\frac{\lambda^n}{\Gamma(n)} x^{n-1} e^{-\lambda x}$

- If $X_1, X_2 \ldots, X_n$ are independent $Exp(\lambda_i)$ then $\min_i X_i \sim Exp(\sum_i \lambda_i)$.

- If $X_1, X_2 \ldots, X_n$ are independent $Exp(\lambda_i)$ then $P\{X_i = \min_j X_j\} = \dfrac{\lambda_i}{\sum_j \lambda_j}$.

- If X_1 and X_2 are independent exponential random variables with parameters λ_1 and λ_2, respectively, then the CDF of $W = \max\{X_1, X_2\}$ is

$$F_W(w) = P\{W \leq w\} = \begin{cases} 0, & w < 0 \\ 1 - e^{-\lambda_1 w} - e^{-\lambda_2 w} + e^{-(\lambda_1+\lambda_2)w}, & w \geq 0. \end{cases}$$

- If $X_1, X_2 \ldots, X_n$ are independent $Exp(\lambda_i)$, then a mixture of them having PDF

$$f(x) = \sum_{i=1}^{n} a_i \lambda_i e^{-\lambda_i x}, \quad 0 \leq a_i \leq 1, \sum_{i=1}^{n} a_i = 1, \ \ (x \geq 0)$$

  is called a n-stage **hyper-exponential distribution** (denoted as H_n).

- If $X_1, X_2 \ldots, X_n$ are independent $Exp(\lambda_i)$ with $\lambda_i \neq \lambda_j$ for $i \neq j$, then the PDF given by

$$f_{X_1+X_2+\ldots+X_n}(x) = \sum_{i=1}^{n} C_{i,n} \lambda_i e^{-\lambda_i x},$$

  where $C_{i,n} = \prod_{j \neq i} \dfrac{\lambda_j}{\lambda_j - \lambda_i}$, is known as the PDF of a n-stage **hypo-exponential** distribution.

- the coefficient of variation of this hyper-exponential is always greater than 1 one can show, and it is always less than 1 for a hypo-exponential, whereas it is exactly equal to 1 for an exponential distribution.

  So, whenever you have a model where this is being exhibited, you can think of which distribution to pick from the family of exponentials or related to exponentials; one can think of hyper-exponential and hypo-exponential because the coefficient of variations of this one is greater than 1, the other is less than 1, and the third one is equal to 1 which is exactly for the exponential distribution.

- Hyper-exponential corresponds to n-stages in *parallel* and hypo-exponential corresponds to n-stages in *series*.

Now, what we will now do is we will define what we call a counting process.

- A (continuous-time) random process $\{N(t), t \geq 0\}$ is said to be a counting process if $N(t)$ is the number of events occurred from time 0 up to and including time t. For a counting process, we assume

  (i) $N(0) = 0$.

(ii) $N(t) \in \{0, 1, 2, \dots\}$ for all $t \in [0, \infty)$.

(iii) $N(t) - N(s)$ for $0 \le s < t$ shows the number of events that occur in the interval $(s, t]$.

So, a counting process is a stochastic process with state-space as the set of all nonnegative integers, and that starts at 0, and it has as the time progress, the state only increases; that is what we will come out to be here because it is counting the number of events so-called what we call whatever we may define as an event. So, that is the number of events it is counting. So, it is a counting process that then you are looking at.

Examples.

$N(t) =$ The number of persons who enter a particular store upto time t — Counting process.

$N(t) =$ Total number of people who were born upto time t — Counting process.

$N(t) =$ The number of persons in a store at a time t — NOT a counting process.

- A counting process $\{N(t), t \ge 0\}$ is said to have independent increments if the number of events that occur in disjoint time intervals are independent.
  That is, for any $0 \le t_1 < t_2 < t_3 < \cdots < t_n$, the random variables $N(t_2) - N(t_1), N(t_3) - N(t_2), \dots, N(t_n) - N(t_{n-1})$ are independent.

- A counting process $\{N(t), t \ge 0\}$ is said to have stationary increments if the distribution of $N(t+s) - N(t)$ depends only on s, for all $s, t \ge 0$.

So, it means that over the time point where the interval is placed here, the interval is $(t, t+s)$ where the interval is placed is not relevant; only the length of the interval is relevant. So, they said to have stationary increments if the distribution depends only on s, not on t, then it said to have stationary increments.

Now, with these definitions of the basic three processes, we will try to see what is a Poisson process.

- The Poisson process is one of the most widely-used counting processes and usually used in scenarios where we are counting the occurrences of certain events that appear to happen at a certain rate, but completely at random (without a certain structure).

  But what is important is completely at random if it is something happening in a deterministic fashion that every 1 hour 1 customer comes or every 1 hour comes, that is not really a Poisson process model. But something that occurs completely at random without any structure, then that is a scenario where this Poisson process could be a potential model to represent that behaviour. For example, the following scenarios could be considered:

  ▶ The number of car accidents at a site or in an area.
  ▶ The location of users in a wireless network.
  ▶ The requests for individual documents on a web server.
  ▶ The outbreak of wars.
  ▶ Photons landing on a photodiode.
  ▶ Arrival of customers at a post office.

There could be n number of situations where it has been shown using empirical studies that when you have a large number of unknown sources from which the arrivals can happen, or the incidents can happen, the events can happen, then the superposition or all these things put together would behave more or less like a Poisson process is what has been shown in the literature. So, it is a quite good approximation for most of the situations, and it is a quite good first cut approximation if you have to say so. Because you are extracting the very fundamental property of randomness only through this Poisson process, and once you get then, of course, you can improve upon these models further. So, in that sense, this serves as a first model for many situations. So, what is a Poisson process, and how do we define a Poisson process?

- A counting process $\{N(t), t \geq 0\}$ is said to be a (homogeneous) Poisson process (PP) with rate (or intensity) $\lambda > 0$ if
  1. $N(0) = 0$,
  2. it has independent increments, and
  3. the number of events in any interval of length $t > 0$ has $Poi(\lambda t)$ distribution.

We are calling it homogeneous because the parameter here is independent. It is basically the rate or intensity of the events to occur according to this particular process now; we are assuming that to be a constant rate; of course, this is only an approximation; whenever that is there, you tend to say this is a homogeneous Poisson process. It starts at 0 because it is a counting process. So, it is at time 0; it is no event is happening, then the event starts to occur, and it has independent increments, which means the increments in non-overlapping intervals are independent. And the number of events in any interval of length t has a Poisson distribution with parameter lambda t.

- This (homogeneous) Poisson process also has stationary increments we can show; of course, sometimes you know this is also given as part of the definition, but one can also show from this that it has stationary increments in this particular case.

So, this definition then fixes all finite-dimensional distributions of the stochastic process. Because if you want to know completely about a stochastic process, you need to know the finite-dimensional distributions. So, in this particular case, this definition fixes all the finite-dimensional distributions of the stochastic process.

- Fix any $T > 0$. Define a new process $N_T(\cdot)$ by $N_T(t) = N(T + t) - N(T)$. Then $\{N_T(t)\}$ is again a Poisson process with rate λ. Thus a Poisson process probabilistically restarts itself at any point of time (Markov property).

- **Alternative Definition**: A counting process $\{N(t), t \geq 0\}$ is said to be a (homogeneous) Poisson process (PP) with rate (or intensity) $\lambda > 0$ if
  1. $N(0) = 0$,
  2. it has independent increments, and
  3. we have the orderliness property:
     $P\{1 \text{ event between } t \text{ and } t + \Delta t\} = \lambda \Delta t + o(\Delta t)$, and
     $P\{2 \text{ or more events between } t \text{ and } t + \Delta t\} = o(\Delta t)$,
     where $o(\Delta t)$ denotes a quantity that becomes negligible when compared to Δt as $\Delta t \to 0$, i.e, $\lim_{\Delta t \to 0} \frac{o(\Delta t)}{\Delta t} = 0$.

Exercise.
Using the alternative definition and starting from the basic principles, obtain the differential-difference equations satisfied by $p_n(t) = P\{N(t) = n\}$ *as*
$p_0'(t) = -\lambda p_0(t)$
$p_n'(t) = -\lambda p_n(t) + \lambda p_{n-1}(t), \;\; n \geq 1.$
Solve these and show that $N(t) \sim Poi(\lambda t)$ *distribution.*

So, that you can show that this alternate definition implies the previous definition, but what is our interest not just not showing this particular thing, but the way how one arrives at the differential-difference equations satisfied by the state probabilities in a continuous-time Markov chain in general and in this particular case it is a Poisson process.

So, you need to understand that process very well in order to understand the Markov chain process completely, and then one can look at how one can solve $p_0'(t) = -\lambda p_0(t)$
$p_n'(t) = -\lambda p_n(t) + \lambda p_{n-1}(t), \;\; n \geq 1.$
once you get this. So, you know already that $N(t) \sim Poi(\lambda t)$ which is what is already given in the definition; this is the alternative definition; by showing equivalence between these two, you will do this process that makes you familiar with the process of how to obtain this kind of differential-difference equation satisfied by the state probabilities $p_n(t)$ in a general setup as well. So, this will give you a stepping stone. So, pay attention to how we are doing this in this case; this is an exercise you can do, but let me also give a hint of a few steps on how one can do that.

Consider $p_0(t)$, the probability of no event by time t. Then,

$$\begin{aligned} p_0(t + \Delta t) &= P\{0 \text{ arrivals in } [0,t] \text{ and } 0 \text{ arrivals in } (t, t+\Delta t]\\ &= p_0(t)\,[1 - \lambda \Delta t - o(\Delta t)] \end{aligned}$$

This gives $p_0'(t) = -\lambda p_0(t)$ whose solution is $p_0(t) = e^{-\lambda t}$ (using the fact that $p_0(0) = 1$).
For $n \geq 1$, we consider

$$\begin{aligned} p_n(t + \Delta t) = \;& P\{n \text{ arrivals in } [0,t] \text{ and } 0 \text{ arrivals in } (t, t+\Delta t]\\ &+ P\{n-1 \text{ arrivals in } [0,t] \text{ and } 1 \text{ arrivals in } (t, t+\Delta t]\\ &+ P\{n-2 \text{ arrivals in } [0,t] \text{ and } 2 \text{ arrivals in } (t, t+\Delta t]\\ &\vdots\\ &+ P\{0 \text{ arrivals in } [0,t] \text{ and } n \text{ arrivals in } (t, t+\Delta t]\\ = \;& p_n(t)\,[1 - \lambda\Delta t - o(\Delta t)] + p_{n-1}(t)\,[\lambda \Delta t + o(\Delta t)] + p_{n-2}(t)\,[o(\Delta t)] + \cdots + p_0(t)\,[o(\Delta t)]. \end{aligned}$$

Rewriting as $\quad p_n(t + \Delta t) - p_n(t) = -\lambda \Delta t p_n(t) + \lambda \Delta t p_{n-1}(t) + o(\Delta t),$
we obtain $p_n'(t) = -\lambda p_n(t) + \lambda p_{n-1}(t), \;\; n \geq 1.$

For $n = 1$, we have $p_1'(t) + \lambda p_1(t) = \lambda p_0(t) = \lambda e^{-\lambda t}$ and the solution is given by

$$p_1(t) = ce^{-\lambda t} + \lambda t e^{-\lambda t}.$$

Using $p_1(0) = 0$ gives us $c = 0$ and hence $p_1(t) = \lambda t e^{-\lambda t}$.

Using the method of induction, it can now be shown that

$$p_n(t) = e^{-\lambda t}\frac{(\lambda t)^n}{n!}, \qquad n = 0, 1, 2, \ldots.$$

Aliter: (Method of generating functions)

Define $P(z,t) = \sum_{n=0}^{\infty} p_n(t) z^n = E(z^{N(t)})$. The differential-difference equations can be reduced to

$$\frac{\partial}{\partial t}P(z,t) = \lambda(z-1)P(z,t) \quad \Rightarrow \quad P(z,t) = e^{\lambda(z-1)t}. \quad \text{(using } P(z,0) = 1\text{)}$$

So, this is another way of doing it through generating functions. The method of induction is one way; generating function is another way; at least two methods then you will be familiar with. So that you can use any of them depending upon the requirement so; in this particular case, you would need the theory of a little bit of differential equation ideas to see; of course, we are assuming calculus knowledge. So, obviously, this is doable in some sense; otherwise, it is not a big deal because you can always use the basic idea that you will have with respect to this only. You will have alternative methods, always. So, let us stop here, and we will continue in the next lecture with further properties of this Poisson process.

Bye.