Markovian Queues - Erlang Models

- We have so far studied queueing models with Poisson process arrivals (exponential interarrival times) and exponential service times, or simple variations thereof.
- In many practical situations, the exponential assumptions may be rather limiting, especially the exponentially distributed service times assumption.
- We will now try to go beyond systems having the Poisson-exponential assumptions.
- How do we handle queues where the interarrival time and/or service time distributions are not exponential?
  ▶ Look to express these distributions in terms of exponential distributions (and use the method of stages) and be within the Markovian framework (so that our approach can remain the same).
  ▶ If the above is not possible, then try to handle with the actual distributions (we will do later) or use approximation methods with mean and variance (may not be with the actual distributions) (and this is beyond our scope).
- Within our framework, Erlang distributed interarrival/service times form the starting point of dealing with more general distributions.

The Erlang Distribution (E_k)

- Recall that $T \sim Gamma(\alpha, \theta)$ with PDF

$$f_T(t) = \frac{\theta^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\theta t}, \quad t > 0; \quad \alpha, \theta > 0.$$

  - Here, $E(T) = \alpha/\theta$ and $Var(T) = \alpha/\theta^2$.

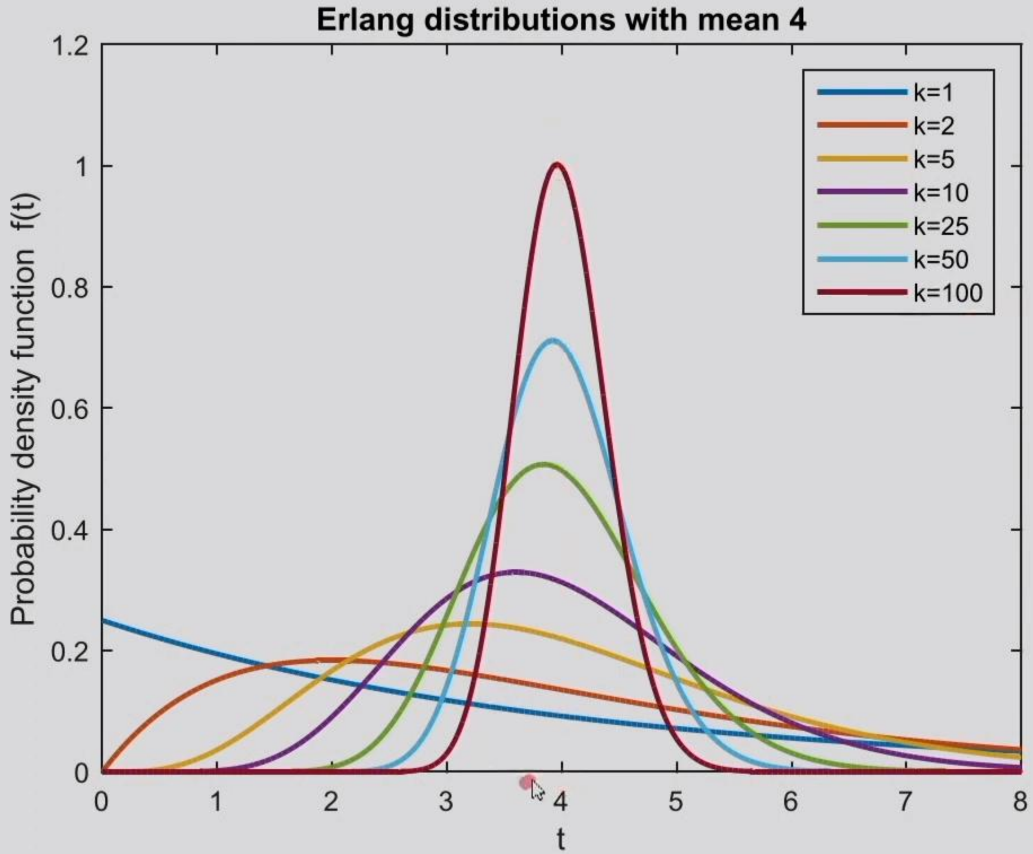
- Now, consider a special class of these distributions where α is restricted to positive integer. Specifically, put $\alpha = k$ and $\theta = k\mu$, where k is any arbitrary positive integer and μ is any arbitrary positive constant. This, gives the Erlang family of probability distributions with PDF and CDF (with parameters k and μ)

$$f(t) = \frac{(\mu k)^k}{(k-1)!} t^{k-1} e^{-k\mu t}, \quad t > 0,$$

$$F(t) = 1 - \sum_{n=0}^{k-1} e^{-k\mu t} \frac{(k\mu t)^n}{n!}, \quad t \geq 0 \qquad \text{[and } F(t) = 0 \text{ for } t \leq 0\text{]}$$

  with the mean and the variance as $E(T) = \dfrac{1}{\mu}$, and $Var(T) = \dfrac{1}{k\mu^2}$.

- For a particular value of k, the distribution is referred to as an Erlang type-k or E_k distribution.



- When $k = 1$, the Erlang reduces to an exponential distribution with mean $\frac{1}{\mu}$.
- As k increases, the Erlang becomes more symmetrical and more closely centered around its mean.
- For $k \to \infty$, the Erlang becomes deterministic with value $\frac{1}{\mu}$.
- In practice, the Erlang family provides more flexibility in fitting a distribution to real data than the exponential family.

- The Erlang distribution is useful in queueing analysis because of its relationship to the exponential distribution.
- Fact: The sum of k IID exponential random variables with mean $\frac{1}{k\mu}$ is an E_k distribution.
  ▶ This makes it possible to take advantage of the Markovian property of the exponential distribution (though Erlang is not Markovian).
- The Erlang distribution is suitable in many applications, for example, when the service consists of many phases (or steps) performed by the same server with each step following an IID exponential distribution.
- The Erlang distribution can also be used even when the physical system does not contain any phases.
  ▲ The use of phases helps from a mathematical perspective (benefitting from the properties of the exponential).
  ▲ The drawback is that using phases increases the size of the state space and the model complexity.

Phase-Type Distributions (PH)

- The idea of phases (of exponentials) to construct distributions can be genealized to include a much wider class of distributions than just the Erlang. These are known as **phase-type** distributions.
- The distributions that uses the concept of phases are: Erlang, hyperexponential, hypoexponential (or generalized Erlang), Coxian and a general phase-type distribution.
- The set of phase-type distributions is dense in the field of all positive-valued distributions, i.e., it can be used to approximate any positive-valued distribution.
- During the last four decades, large number of studies have been done on phase-type distributions and their applications in queueing theory.

Erlang as a Phase-Type Distribution

- We look at another method to generate an Erlang distribution.

  Consider the following three-state continuous-time Markov chain (with an absorbing state at 3).

  The associated Q matrix is $Q = \begin{bmatrix} -\mu & \mu & 0 \\ 0 & -\mu & \mu \\ 0 & 0 & 0 \end{bmatrix}$.

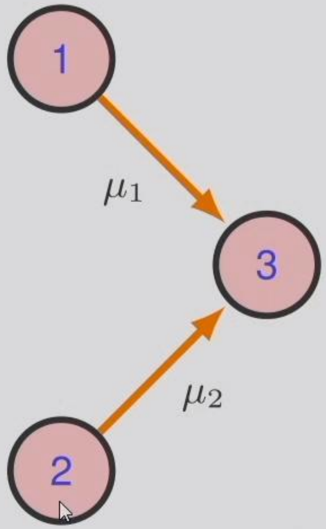
- Suppose that the process starts in state 1 (i.e. $\mathbf{p}(0) = (1, 0, 0)$) and let T be the time to absorption. Then, T follows an E_2 distribution with mean $\frac{2}{\mu}$ because
  - ▶ The sojourn time in state 1 and state 2 are exponentials with mean $\frac{1}{\mu}$ each.
  - ▶ T is the sum of these two IID exponentials with mean $\frac{1}{\mu}$ and hence E_2.

Basic Idea of PH

- The phase-type distribution in question is defined as the time to absorption in a specified CTMC.
- By choosing different CTMCs and by choosing different initial probability distributions, we can construct different phase-type distributions.
- These distributions are not exponential in general, but they can be analyzed using the CTMC theory, taking advantage of the underlying exponential distributions.

Hyperexponential Distributions (H_k)

- Let us now look at generating a hyperexponential distribution. Consider the following CTMC and its Q matrix.



$$Q = \left[\begin{array}{cc:c} -\mu_1 & 0 & \mu_1 \\ 0 & -\mu_2 & \mu_2 \\ \hdashline 0 & 0 & 0 \end{array}\right]$$

- Let the chain starts in state 1 with probability q and in state 2 with probability $1-q$, i.e., the initial state vector is $\mathbf{p}(0) = (q, 1-q, 0)$.
- Then T, the time to absorption, follows a hyperexponential (H_2) distribution (being a mixture of two independent exponentials).
  - Here with probability q, the time to absorption is exponential with mean $\frac{1}{\mu_1}$ and with probability $1-q$, the time to absorption is exponential with mean $\frac{1}{\mu_2}$.

A Formal Approach to generate PH

- We now describe an approach (to determine the distribution of time to absorption) that can be applied to more complex CTMCs. We take the hyperexponential case itself.
- The forward Kolmogorov equations for the hyperexponential Markov chain are given by $\mathbf{p}'(t) = \mathbf{p}(t)\mathbf{Q}$ which when written explicitly are:

$$
\begin{aligned}
p_1'(t) &= -\mu_1 p_1(t) \\
p_2'(t) &= -\mu_2 p_2(t) \\
p_3'(t) &= \mu_1 p_1(t) + \mu_2 p_2(t)
\end{aligned}
$$

- Solving the first differential equation, we get $p_1(t) = qe^{-\mu_1 t}$, since $p_1(0) = q$. Similarly, solving the second differential equation, we get $p_2(t) = (1-q)e^{-\mu_2 t}$. Substituting in the third, we get

$$p_3'(t) = q\mu_1 e^{-\mu_1 t} + (1-q)\mu_2 e^{-\mu_2 t}.$$

- Note that, by definition, $p_3(t)$ is the CDF of T, i.e., $p_3(t) = P\{T \leq t\}$ and this means that $p_3'(t)$ is the PDF of T. The PDF given above is a H_2 distribution, as expected.

Alternative Approach:

- Let $\tilde{\mathbf{p}}(t) = (p_1(t), p_2(t))$ (without the absorbing state) and let $\tilde{Q}$ be the corresponding matrix from Q. That is,

$$\tilde{Q} = \begin{bmatrix} -\mu_1 & 0 \\ 0 & -\mu_2 \end{bmatrix}, \quad \text{and} \quad \tilde{\mathbf{p}}'(t) = \tilde{\mathbf{p}}(t)\tilde{Q}.$$

$$Q = \left[\begin{array}{c:c} \tilde{Q} & -\tilde{Q}e \\ \hdashline 0 & 0 \end{array}\right]$$

- The solution to this matrix system of differential equations is

$$\tilde{\mathbf{p}}(t) = \tilde{\mathbf{p}}(0)e^{\tilde{Q}t}, \quad \text{where} \quad \tilde{\mathbf{p}}(0) = (q, 1-q),$$

where

$$\begin{aligned} e^{\tilde{Q}t} &= I + \tilde{Q}t + \frac{(\tilde{Q}t)^2}{2!} + \dots \\ &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} -\mu_1 t & 0 \\ 0 & -\mu_2 \end{pmatrix} + \begin{pmatrix} \mu_1^2 t^2/2 & 0 \\ 0 & \mu_2^2 t^2/2 \end{pmatrix} + \dots \\ &= \begin{pmatrix} e^{-\mu_1 t} & 0 \\ 0 & e^{-\mu_2 t} \end{pmatrix}. \end{aligned}$$

06_MarkovianQ-B.pdf*

Create account Login

- We then obtain finally

$$\tilde{\mathbf{p}}(t) = \tilde{\mathbf{p}}(0)e^{\tilde{Q}t} = (qe^{-\mu_1 t}, (1-q)e^{-\mu_2 t}),$$

which is the same as the one obtained earlier.

- The PDF of T is obtained as before using the last differential equation $p_3'(t) = \mu_1 p_1(t) + \mu_2 p_2(t)$.

- In general, a phase-type distribution is described using the parameters α and $\tilde{Q}$, where $\alpha = \tilde{\mathbf{p}}(0)$ and $\tilde{Q}$ as above (corresponding to the non-absorbing states).
  ♦ A phase-type distribution is denoted by the notation $PH(\alpha, \tilde{Q})$.

- Exercise: Apply this procedure to an appropriate E_2 distribution case and obtain $p_3'(t) = 4\mu^2 te^{-2\mu t}$.

202%

Hypoexponential Distribution

- Consider the following continuous-time Markov chain, where $\mu_1 \neq \mu_2$, and the system starts in state 1.

The associated Q matrix is

$$Q = \left[\begin{array}{cc:c} -\mu_1 & \mu_1 & 0 \\ 0 & -\mu_2 & \mu_2 \\ \hdashline 0 & 0 & 0 \end{array}\right] \qquad \text{[with } \mathbf{p}(0) = (1,0,0)\text{]}$$

- The time to absorption is the convolution of two nonidentical exponential random variables.
  ♦ Note that this is not an Erlang, since an Erlang requires identical exponentials.

- One can derive $\mathbf{p}(t)$ in a similar way and get

$$p_1(t) = e^{-\mu_1 t}$$

$$p_2(t) = \frac{\mu_1}{\mu_2 - \mu_1} e^{-\mu_1 t} - \frac{\mu_1}{\mu_2 - \mu_1} e^{-\mu_2 t}$$

and therefore $$p_3'(t) = \mu_2 p_2(t) = \frac{\mu_1 \mu_2}{\mu_2 - \mu_1} e^{-\mu_1 t} - \frac{\mu_1 \mu_2}{\mu_2 - \mu_1} e^{-\mu_2 t},$$

where $p_3'(t)$ is the probability density function (of the time to absorption) which is a two-term hypoexponential distribution.

- We now have a method for obtaining more general distributions and still use Markovian methods (and this can be of great help in many situations!).

- We will now take up some Erlangian queueing systems to illustrate the ideas.