UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN
FACULTAD DE INGENIERGÍA MECÁNICA Y ELÉCTRICA

TESIS

FITTING PHASE TYPE DISTRIBUTION TO SERVICE PROCESS WITH SEQUENTIAL PHASES

PRESENTA
ANKITA NANDA

COMO REQUISITO PARA OBTENER EL GRADO DE MAESTRÍA EN CIENCIAS EN INGENIERÍA DE SISTEMAS

ENERO, 2015

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FACULTAD DE INGENIERGÍA MECÁNICA Y ELÉCTRICA

SUBDIRECCIÓN DE ESTUDIOS DE POSGRADO

TESIS

FITTING PHASE TYPE DISTRIBUTION TO SERVICE PROCESS WITH SEQUENTIAL PHASES

PRESENTA

ANKITA NANDA

COMO REQUISITO PARA OBTENER EL GRADO DE MAESTRÍA EN CIENCIAS EN INGENIERÍA DE SISTEMAS

COMITÉ TUTORIAL

DIRECTOR: DR. CESAR EMILIO VILLAREAL RODRÍGUEZ

CODIRECTOR: DR. SARAT KUMAR ACHARYA

ENERO, 2015

UNIVERSIDAD AUTONOMA DE NUEVO LEON

FACULTAD DE INGENIERIA MECANICA Y ELECTRICA

SUBDIRECCIÓN DE ESTUDIOS DE POSGRADO

Los miembros del Comité de Tesis recomendamos que la Tesis "**FITTING PHASE TYPE DISTRIBUTION TO SERVICE PROCESS WITH SEQUENTIAL PHASES**" realizada por la alumna "**ANKITA NANDA**", con número de matrícula 1655143, sea aceptada para su defensa como opción al grado de "MAESTRA EN CIENCIAS EN INGENIERÍA DE SISTEMAS"

El Comité de Tesis

DR. CESAR EMILIO VILLAREAL RODRÍGUEZ
Asesor

DR. SARAT KUMAR ACHARYA
Revisor

DR. ROMEO SÁNCHEZ
Revisor

Vo. Bo.

Dr. Simón Martínez Martínez
Subdirección de Estudios de Posgrado

San Nicolás de los Garza, Nuevo León, 16 de enero, 2015

Table of contents

List of figures

List of tables

AGRADECIMIENTOS

A mi familia por apoyarme y estar conmigo en todo momento.

Al Posgrado en Ingeniería de Sistemas de la Facultad de Ingeniería Mecánica y Eléctrica de nuestra máxima casa de estudio por ser el lugar donde he logrado mi formación académica.

Al Consejo Nacional de Ciencia y Tecnología (CONACYT), México por apoyar me con la beca durante mi estudio de maestría en la Universidad Autónoma de Nuevo León.

Al Dr. Cesar Emilio Villareal Rodríguez, por su apoyo y enseñanza durante el desarrollo de esta investigación.

Al Dr. Sarat Kumar Acharya, por su guía y sus sugerencias durante desarrollo de esta tesis.

Al Dr. Romeo Sánchez por sus importantes observaciones que me ayudaron a mejorar esta investigación.

A todos los personales del PISIS y, amigos quienes estuvieron conmigo en todo momento.

ABSTRACT

The work of this thesis is concerned with fitting Hypo-exponential and Erlang phase type distributions for modeling real life processes with non-exponential service time. There exist situations where exponential distributions cannot explain the distribution of service time properly. This thesis presents the application of two traditional statistical estimation techniques to approximate the service distributions of processes with coefficient of variation less than one. It also presents an algorithm to fit Hypo-exponential distribution for complex situations which can't be handled properly with traditional estimation techniques. The result shows the effect of variation of sample size and other parameters on the efficiency of the estimation techniques by comparing their respective outputs. Furthermore it checks how accurately the proposed algorithm approximates a given distribution.

CHAPTER 1

INTRODUCTION

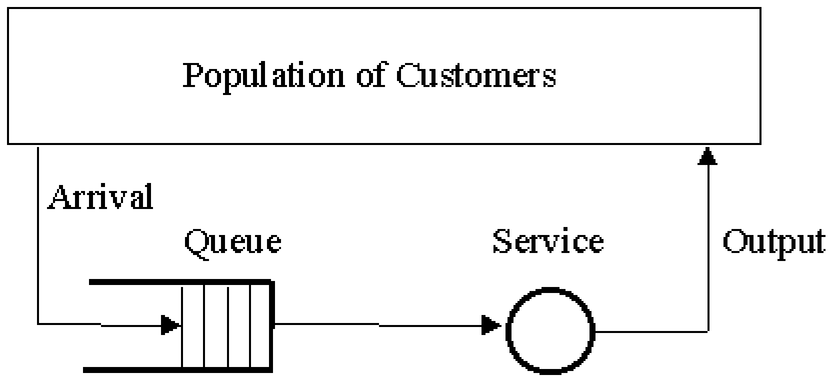
1.1 Queuing theory

In day to day life we very often encounter situations where we need to wait in a line in order to receive service. Such waiting lines are called queues. Some real life examples of queues are:

- Waiting line for medical check-up at hospital
- Waiting for seats at restaurants
- Security checking at airport
- Waiting to receive service at banks
- Traffic lights

A queue is formed when arrival of customers is faster than they are served. Diagram of a simple queuing system is given below.

Figure 1.1 Components of a simple queuing system



In order to provide quality service i.e. to reduce the waiting time of customers, more number of servers have to be planted. On the other hand, this increases cost of maintenance increasing the service cost and with large number of servers the utility of the system could be less resulting in less efficient system. So an optimum number of servers have to found in terms of quality and efficiency. Here comes the role of queuing theory.

Queuing theory is the mathematical study of systems including queues. "Queuing theory was developed to provide models to predict the behavior of systems which attempt to provide service for randomly arising demands. The earliest problems studied where those of telephone traffic congestion. The pioneer investigator was Danish mathematician A. K. Erlang, who in 1990 published "The Theory of Probabilities and Telephone Conversations" (Gross and Harris, 1985). It analyzes congestions and delays in the waiting line by examining different components of the queuing system such as arrival rate, service rate, number of servers and number of customers in the system, etc. Queuing theory is used to develop more efficient systems that reduce customer waiting times, increase number of customers to be served, and reduce the service costs.

1.2 Queuing system characteristics

There are six characteristics that specify a queuing system.

Arrival process: This defines the manner in which customers arrive to the queuing system. Arrival process is defined in terms of probability distributions of inter arrival time i.e. the time difference between two consecutive arrivals. Many possible assumptions exist for arrival process to a queuing system. Some of them are bulk arrival, balking, reneging etc. Common inter arrival distributions include Poisson, Erlang, Deterministic, General distribution with known mean and variance etc.

Service process: It is defined in terms of probability distribution of the service time. The Service time is the time spent by the customer while receiving service. Like the arrival process, there are many possible assumptions for service time distribution. Most common assumptions are independent identically distributed random variables and exponential service time distribution. The other possible distributions of service time can be Erlang, Coxian, Hyper-exponential, Hypo-exponential, Deterministic, General distribution with known mean and variance etc.

Number of Servers: It is the number of parallel service channels that can provide service simultaneously. These Servers may or may not be identical. The service discipline determines the allocation of the customers to the servers.

System Capacity: It is defined as the maximum number of permitted customers in the system (including those in service). System capacity can be finite or infinite. Some queueing processes puts a restriction on the maximum number of allowed customers in the system. Such queuing systems are called finite length queuing system. If there is no restriction on the number of incoming customers, the queuing system is referred as an infinite queuing system.

Queue Discipline: The queue discipline defines the order in which waiting customers are provided service. The most common disciplines are:

- First in first out (FIFO): Customers are served according to the order they arrive to the system.
- First in last out (FILO): Customers are served in the reverse order of their entry. So, the ones who joins last will get service first.
- Served in random order (SIRO): Under this rule customers are selected for service at random, irrespective of their arrivals in the service system. Every customer in queue is equally likely to be selected.
- Priority Service: Under this rule, customers are grouped in priority classes on the basis of some attributes.

- Processor Sharing: The server is switched between all the queues for a predefined slice of time in a round-robin manner.

Number of service phases: A queuing system may contain a single or multiple service phases. An example of a multi-phase queuing system can be medical checkup in a hospital where each patient has to pass through several phases such as medical history: ear, nose, eye checkup; blood test and so on. In some multi- phase queuing process, recycling or feedback can take place (Gross and Harris, 1985:6).

In general a queuing process is described by a series of symbols and slashes such as A/B/X/Y/Z, where A indicates arrival pattern, B the service pattern, X the number of service channels, Y the restriction on system capacity and Z the queue discipline.

1.3 Problem description

Analyzing a queuing system requires a clear understanding of the appropriate service measurements. To understand real life queuing systems and to obtain a mathematical model for them, it is necessary to make some assumptions. To model a queuing service process, the most commonly made assumption is that the time of service is exponentially distributed. The wide of use of the exponential distribution lies in its memory less property. The memory less property of the exponential distribution states that the remaining time left of a service process at any point of time is unpredictable. In other words, if the life time of an item possesses an exponential distribution, then at any point of the use of that item one can make the statement that the item is as good as a new item in terms of its failure time.

Exponential distribution is a very convenient choice for modeling service and inter-arrival time distributions. It gives very acceptable results in characterizing system behavior. Nevertheless there exist situations where this could not be a good choice of modelling. One important example of this situation is when the service times are completely predictable (i.e. deterministic). There also exist processes which possess coefficient of variation greater than or less than one. So, an exponential distribution could not explain these processes well as it has coefficient of variation equal to one.

The thesis focusses on the processes where the server provides services in sequential phases where each customer has to start from the first phase and goes on covering all of them one by one. If the holding time in each phase is exponentially distributed then the total process service time will have a distribution which is the sum of exponential distributions of all the phases. In such cases comes the necessity of fitting Hypo-exponential phase type distribution and Erlang phase type distribution to better explain the service time distribution of such processes. Here some parameter estimations techniques and approximation algorithm will be described which can be used to fit Hypo-exponential and Erlang phase type distributions real life processes with non-exponential service time having coefficient of variation less than one.

The present research has been divided into six chapters. Rest of this chapter gives some related work in this direction and the objectives of the current research. **Chapter 2** describes some commonly used phase type distributions. **Chapter 3** deals with characterizations of Erlang and Hypo-exponential distributions. This includes how to calculate the mean, variance and other characteristics of these two distributions following **Chapter 4** which gives some real life applications of Hypo-exponential and Erlang distributions. In **Chapter 5** we will be describing types of statistical method that will be used for estimation of parameters. This includes two traditional methods of estimation i.e. method of method (MOM) and maximum likelihood estimation (MLE). Furthermore an approximation algorithm is given to approximate distributions where information is known in terms of mean and variance only. **Chapter 6** provides in detail how to use these methods to estimate the parameters of Erlang and Hypo-exponential distributions for different situations. At final, **Chapter 7** simulates the results obtained by applying proposed method on randomly generated data using RStudio software and analyzes the result.

1.4 Literature review

Queueing theory was developed to provide models to predict behavior of systems that attempt to provide service for random demands. The earliest problems studied were those of telephone traffic congestion. For the first time A.K. Erlang in 1909 published "The Theory of Probabilities and Telephone Conservations" in this connection. Later on he observed that a telephone system was generally characterized by either M/M/c or by M/D/1 system. Later on a number of works have been done in this area. The works of Molina, Felix Pollaczek, Lindley, Bailey, Linderman, Reuter, Kendall, Benes, Bhat, Conway, Little, Maxwell, Neuts, Prabhu, Satty are worth mentioning.

Normally queueing theory is thought of as an applied probabilistic analysis of waiting times mainly focused on minimum waiting time under standard conditions. But statistical inference plays a major role in any use of queueing models in decision making. The earliest work in this direction seems to be that of Clarke (1957), who obtained the maximum likelihood estimator (MLE) of the parameters of M/M/1 queueing equilibrium. Acharya et.al (2013) analyzes the derivation of maximum likelihood estimates for the arrival rate and service rates in a stationary M/M/c queue with heterogeneous servers. Cox (1955) and Wolff (1965) carried out the investigation with several ideas. Wolff (1965) discussed maximum likelihood estimation for a class of ergodic queueing models which give rise to birth and death processes. The paper by Benes (1957) and Gross and Harris (1985) are also worth mentioning. Bhat and Rao (1987) have discussed the problem in detail. They have studied the asymptotic inference for single server queues. Basawa and Prabhu (1981) have studied the asymptotic inference for single server queues and have proved consistency and asymptotic normality of the MLE of the parameters in a G/G/1 queue, while Acharya (1999) has studied the rate of convergence of the distribution of the maximum likelihood estimators of the arrival and the service rates in a G/G/1 queueing system.

Statisticians generally divide the statistical problems into two types. Those are parameter estimation and distribution selection. A short history of the parameter estimation case is given above. But for the distribution selection case, appropriate data are to be examined as a basis for determining a choice of model.

Several studies have been done related to fitting phase type distribution to service time of a process. For example; Maode (2009) tried to fit Coxian distribution to actual service time and Hypo-exponential distribution to the interrupted service time of the optical burst switching network. In 2011, Nigel Thomas in his book: Computer Performance Engineering modelled the replenishment time of an Inventory as Hyper-exponential and Hypo-exponential distribution respectively and found that the previous case gives the optimal policy. Similarly, Khalid et.al (2001) has modelled the input-output operation time of a single bus multiple processor system as 3 phase Hypo-exponential distribution and 3 phases Erlang distribution respectively to measure system performances. Goldberg and Whitt (2007) fitted exponential, Hypo-exponential and Hyper-exponential distribution to service time of two phase inspection process for calculating the last departure time from an M/G/1queue with a terminating arrival process.

This thesis work tries emphasizes queuing process with non-exponential service time having coefficient of variation less than one. It presents the use of traditional estimation techniques (MOM & MLE) for fitting Erlang and Hypo-exponential distribution to real life service process with coefficient of variation less than one. Bakoban(2012) presents Bayesian and non-Bayesian estimation of Erlang distribution under progressive censoring. In practice it often occurs that the only information of random variables that is available is their mean and standard deviation, or if one is lucky, some real data. To obtain an approximating distribution it is required to fit a phase-type distribution based on the mean and the coefficient of variation of a given positive random variable, by using the given estimation techniques. Traditional estimation techniques are also very simple to implement and it can be used with any kind of data set. On the other hand, it explains curve fitting algorithm for approximation of Hypo-exponential distribution with more than two phases that can't be handled properly by traditional estimation techniques. This approximation also algorithm tries to approximate a given distribution based on the mean and the coefficient of variation of a given sample of data and hence can be used widely.

1.5 Objective

The main objective of this thesis is to explain the different statistical estimation techniques and curve fitting algorithm for better approximation of Erlang and Hypo-exponential phase type distribution to model real life processes with non-exponential service time. Thus, the specific objectives of this research are:

- To estimate the parameters of Erlang distribution for fitting distribution to service process getting service by the server in multiple phase where each phase is assumed to have an exponential distribution with same rate.
- To estimate parameters of Hypo-exponential distribution for fitting distribution to service process getting service by the server in multiple phase where each phase is assumed to have an exponential distribution with different rate.
- To simulate these estimation techniques using randomly generated data set and analyze the outcome.

CHAPTER 2

PHASE TYPE DISTRIBUTION

2.1 Introduction

A probability distribution resulting from the convolution of two or more exponential distributions is known as phase type distribution. If a system consists of one or more inter-related service phases occurring in sequence, then the entire process is said to possess a continuous phase type distribution or simply phase type distribution. The order of execution of these phases can again be either deterministic or stochastic in nature. The absorption time in Markovian processes are known to have this kind of distribution. There also exists a discrete time equivalent of the phase type distribution popularly known as discrete phase type distribution. It is the probability distribution resulting from the convolution of two or more geometric distributions. The absorption time in a discrete Markov chain possesses a discrete phase type distribution. Phase type distributions are generally used to approximate any positive-valued distribution.

2.2 Continuous phase type distribution

There exist several different types of continuous phase type distributions. The difference between them lies in the order of execution of the sequential phases of the process. All of them are widely used to model real life processes.

2.2.1 Exponential distribution

It is the simplest and most commonly used among all the phase type distribution. The time between occurrences of two consecutive events of a Poisson process is described by an exponential distribution. Mathematically an exponential distribution is defined by one parameter λ called rate of the distribution.

A random variable X is said to have exponential distribution with rate λ and denoted as $X \sim exp\ (\lambda)$, if its probability density function is given by

$$f(x;\lambda) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

The cumulative probability function is given as

$$F(x;\lambda) = \begin{cases} 1 - e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Where mean is

$$\mathrm{E}(T) = \frac{1}{\lambda}$$

And variance is

$$\mathrm{Var}(T) = \frac{1}{\lambda^2}$$

2.2.2 Erlang distribution

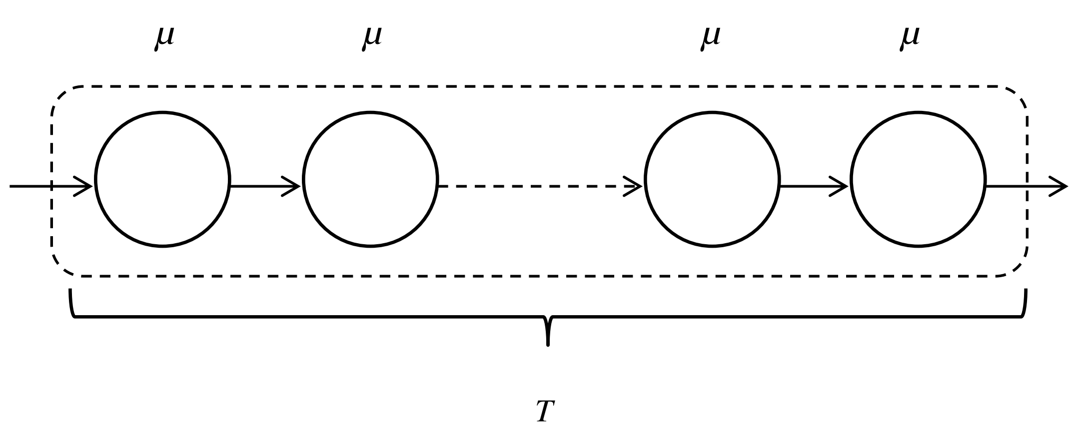
An Erlang distribution was first introduced by Danish mathematician Agner Krarup Erlang to examine number of telephone calls made at the same time to the same station. After his name the distribution was named as Erlang distribution. This distribution is used to model service time of a process which consists of a number of phases to be executed sequentially starting at phase one. If a process consists of several phases in sequence and the time spent in each phase is exponentially distributed with same rate, then the distribution of the total service time of the process will be sum of several exponential distributions with same rate. The resulting distribution is known as Erlang distribution.

An Erlang distribution is a continuous probability distribution that is defined by two parameters called shape parameters and rate parameter. The number of phases in the process is called the shape of the distribution. Erlang distribution is a special case of gamma distribution when shape parameter k is a positive whole number. When k is equal to 1, Erlang distribution is reduced to an exponential distribution. The initial probability distribution (Grinstead, 1997:406) of an Erlang distribution is given as the vector: $\pi = (1,0,\dots,0)$.

It signifies that the process starts at first phase with a probability 1 and at any other phase with probability 0. So, the entry point of a client to the process is always first phase.

Following figure depicts the diagram of a process having service time T with Erlang distribution.

Figure 2.1 State transition diagram of Erlang distribution



A random variable T having Erlang distribution with shape k and rate μ is denoted as: $T \sim Erlang(k, \mu)$.

Probability density function of *T* is given by (Rodriguez, 2010:14)

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

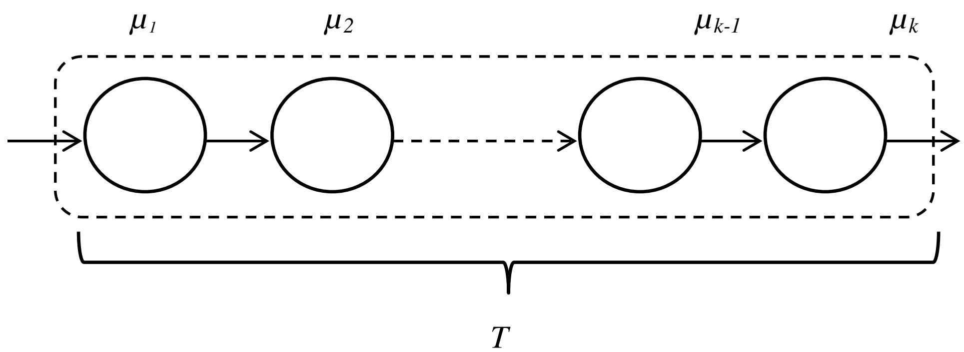
Cumulative probability function of T is given as (Rodriguez, 2010:14)

$$F(t; k, \mu) = 1 - \sum_{n=0}^{k-1} \frac{(\mu t)^{n}}{n!} e^{-\mu k}, for\ t \geq 0$$

2.2.3 Hypo-exponential distribution

Hypo-exponential distribution is a generalization of Erlang distribution. It is used to model processes with many sequential phases where time spent in each phase is exponentially distributed with different rates unlike an Erlang distribution where all the phases have the same rate of distribution. In other words, a Hypo-exponential distribution is a convolution of *k* exponential distributions, each phase *'i'* with their own rate μ_i. An Hypo-exponential distribution is defined by a vector of rates indicating rates of different phases. A random variable *T* with an Hypo-exponential distribution with k phases and rates $(\mu_1, \mu_2 \ldots \mu_k)$ is mathematically denoted as: $T \sim Hypo(\mu_1, \mu_2 \ldots \mu_k)$. The initial probability distribution (Grinstead, 1997) of an Hypo-exponential distribution is same as that of an Erlang distribution given by: $\pi = (1,0, \ldots, 0)$. So, a process with Hypo-exponential distribution always starts at first phase. The following figure depicts the diagram of a process having service time T with an Hypo-exponential distribution with shape parameter k and rates $(\mu_1, \mu_2 \ldots \mu_k)$.

Figure 2.2 State transition diagram of Hypo-exponential distribution



The probability density function of T is given by (Bolch, 1998:33-34)

$$f(t; , \mu_2, \ldots, \mu_k) = \sum_{i=1}^{k} \left(\frac{1}{\mu_i}\right) e^{-t/\mu_i} \left(\prod_{j=1, j \neq i}^{k} \frac{\mu_i}{\mu_i - \mu_j} \right) \quad for\ t > 0\ and\ \mu_i > 0$$

The cumulative probability distribution of T is given as

$$F(t) = P(T \leq t) = \sum_{i=1}^{k} \frac{\left(\prod_{j=1, j \neq i}^{k} \mu_i \right)}{\prod_{j=1, j \neq i}^{k} (\mu_j - \mu_i)} e^{-\mu_i} \ for\ t > 0\ and\ \mu_i > 0$$

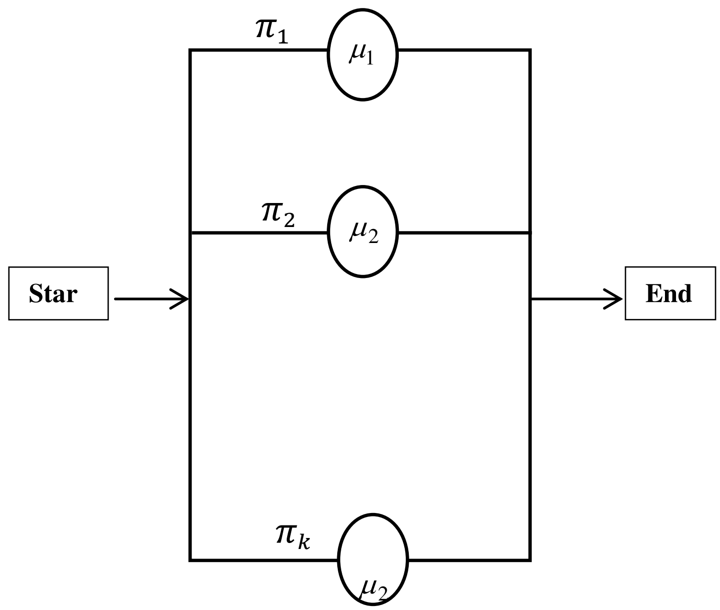
2.2.4 Hyper-exponential distribution

"A random variable T is said to have Hyper-exponential distribution if T is with probability p_i, where i =1, 2… k an exponential random variable T_i" (Adan, 2001:15). This kind of distribution is generally used to model situation where several parallel but mutually exclusive processes occurs with different probability and each process is exponentially distributed with respect to their own rates. A random variable T distributed Hyper-exponentially is mathematically represented as:

$T \sim Hyper\left((\pi_1 \dots \pi_k), \left(\mu_1 \dots \mu_k\right)\right)$ *where* π_i *is the probability of execution and* μ_i *is the rate of of* i^{th} *phase respectively.*

This is pictorially depicted below.

Figure 2.3 State transition diagram of a Hyper-exponential distribution



Unlike the Hypo-exponential and Erlang distribution, a Hyper-exponential distribution can have either of its phases as the initial phase with a specified probability and then follow an exponential distribution with a specified rate. The initial probability distribution (Grinstead, 1997) of a Hyper-exponential distribution with k phases is given as: $\pi = (\pi_1 \dots \pi_k)$. It means the process will start at phase i with probability π_i where i = 1, 2 … k.

Its probability density function is given by (Rodriguez, 2010:15) as:

$$f\left(t;(\pi_1,\pi_2,\ldots,\pi_k),\left(\mu_1,\mu_2\ldots\mu_k\right)\right)\sum_{i=1}^{k}\pi_i\,\mu_i\,e^{-\mu_i t}$$

$$for\ t>0,\ \mu_i>0,\pi_i>0\ and\sum_{i=1}^{k}\pi_i=1$$

The cumulative probability distribution is given by (Rodriguez, 2010:15)

$$F(t)=\sum_{i=1}^{k}\pi_i\left(1-e^{-\mu_i t}\right),\qquad t\geq 0$$

Where the mean is

$$\mathrm{E}(T)=\sum_{i=1}^{k}\frac{\pi_i}{\mu_i}$$

And the variance is

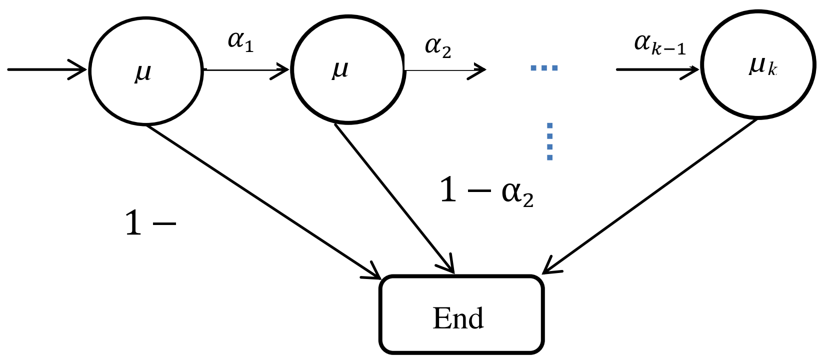
$$\mathrm{Var}(T)=\left[\sum_{i=1}^{k}\frac{\pi_i}{\mu_i}\right]^2+\sum_{i=1}^{k}\sum_{j=1}^{k}\pi_i\pi_j\left(\frac{1}{\mu_i}-\frac{1}{\mu_j}\right)^2$$

This distribution is used in the field of telephony to enhance the service of telecommunication network.

2.2.5 Coxian distribution

Coxian distribution is a generalization of a Hypo-exponential distribution. In Coxian distribution, the process starts from the first phase with probability one, but unlike a Hypo-exponential distribution, it can reach to the absorbing state from either of its phases. It is depicted by the following diagram.

Figure 2.4 State transition diagram of a Coxian distribution



As shown in the diagram, the process starts at phase one. Once it is in some phase i, where i = (1, 2 … k-1), it can move either to the next phase i+1 with probability α_i or directly to the absorbing state with probability (1- α_i). Once the process is in state k, it can only move to the absorbing state with probability one. More over the time spent in phase i is exponentially distributed with rate μ_i.

A random variable having a Coxian distribution with k phases is mathematically denoted as: $T \sim Coxian\left((\alpha_1 \ldots \alpha_{k-1}), \left(\mu_1, \mu_2 \ldots \mu_k\right)\right)$. The initial probability vector is same as that of a Hypo-exponential distribution which is given as: $\pi = (1, 0 \ldots 0)$.

The probability density function is given as: $f(t) = \pi * e^{\{St\}q}$ (Azarbayejan, 2011).

The cumulative probability function is given as: $F(t) = \pi * e^{\{St\}1}$ (Azarbayejan, 2011).

Where $S = \begin{bmatrix} -\mu_1 & \alpha_1\mu_1 & \cdots & 0 & 0 & 0 \\ 0 & -\mu_2 & \alpha_2\mu_2 & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & -\mu_{k-2} & \alpha_{k-2}\mu_{k-2} & 0 \\ 0 & 0 & 0 & 0 & -\mu_{k-1} & \alpha_{k-1}\mu_{k-1} \\ 0 & 0 & 0 & 0 & 0 & -\mu_k \end{bmatrix}$

$q = \left(\mu_1, \mu_2 \dots \mu_k\right)^T$ and 1 is $k \times 1$ one vector

The mean is: $\mathrm{E}(T) = -\alpha S^{-1}1$

Coxian distribution is used prominently in the theory of networks of queues. Any distribution function can be approximated arbitrarily closely by a Coxian distribution. There are many other kinds of phase type distribution which are constructed by the convolution of one or more of the above mentioned distributions. Those distributions are used for analyzing more complex processes. For example a mixture of phase type distribution results from the combination of two or more Erlang distribution. Also when the number of phases tends to infinity, the Erlang distribution becomes a deterministic distribution i.e. the total time of the entire process becomes a constant. In the category of discrete phase type distribution lie geometric distribution, binomial distribution and negative binomial distribution.

In this thesis we will be focusing on the service processes that possess coefficient of variation less than one which can be approximated well with two continuous phase type distributions i.e. Erlang and Hypo-exponential distributions.

CHAPTER 3

CHARACTERIZATION OF PHASE TYPE DISTRIBUTION

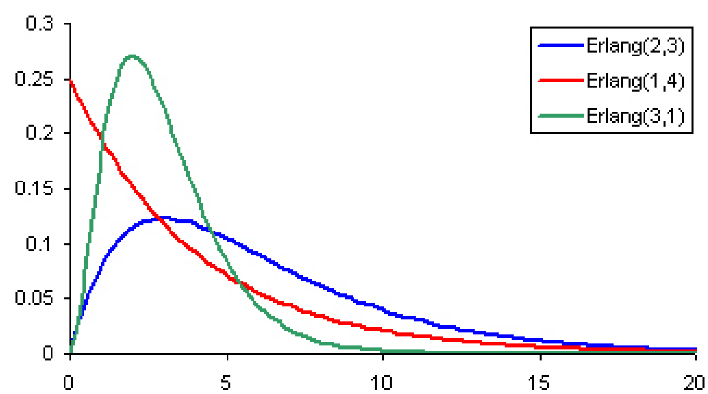
3.1 Erlang distribution

As discussed earlier an Erlang distribution is the convolution of more than one exponential distribution with same rate.

The probability density function of a random variable T with an Erlang distribution with shape parameter k and rate parameter μ denoted as $T \sim Erlang(k, \mu)$ is given by

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

Figure 3.1 Probability density plot of Erlang distribution



3.1.1 Phase transition matrix

The rate transition matrix between phases of an Erlang distribution is given as

$$S = \begin{bmatrix} -\mu & \mu & \cdots & 0 & 0 & 0 \\ 0 & -\mu & \mu & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & -\mu & \mu & 0 \\ 0 & 0 & 0 & 0 & -\mu & \mu \\ 0 & 0 & 0 & 0 & 0 & \mu \\ & & & & & \end{bmatrix}$$

3.1.2 Initial probability vector

The initial probability distribution (Grinstead, 1997) of an Erlang distribution is given by: $\pi = (1,0,\dots.,0)$.

This indicates that the process starts at the first phase with probability one and at the other phase with probability 0. So, only the first phase can be the entry point of the process.

3.1.3 n^{th} Moment

n^{th} Moment of an Erlang distribution with parameter k and μ is given by

$$\mathrm{E}(T^n) = \int_0^\infty t^n * f_T(t; k, \mu) \ for\ t, \mu \ \geq 0$$

$$= \int_0^{\infty} t^n * \frac{\mu^k t^{k-1} e^{-\mu t}}{(k-1)!} \; for\ t,\ \mu \ \geq 0$$

$$= \frac{\mu^k}{(k-1)!} \int_0^{\infty} t^{k+n-1} e^{-\mu t} \; for\ t,\ \mu \ \geq 0$$

Substituting $k + n = l$ in the above equation we will get

$$\mathrm{E}(T^n) = \frac{\mu^k}{(k-1)!} \int_0^{\infty} t^{l-1} e^{-\mu t} \; for\ t,\ \mu \ \geq 0$$

Multiplying and dividing the above equation by $\frac{\mu^l t^{l-1} e^{-\mu t}}{(l-1)!}$, we get

$$\mathrm{E}(T^n) = \frac{\mu^k}{(k-1)!} * \frac{(l-1)!}{\mu^l} \int_0^{\infty} \frac{\mu^l t^{l-1} e^{-\mu t}}{(l-1)!} \; for\ t,\ \mu \ \geq 0$$

$\frac{\mu^l t^{l-1} e^{-\mu t}}{(l-1)!} \; for\ t,\ \mu \geq 0$ is probability density function of $Erlang(l,\ \mu)$

which integrates in the interval $[0,\infty)$ to give 1 (Walck, 2007). So, the above equation is reduced to

$$\mathrm{E}(T^n) = \frac{\mu^k}{(k-1)!} * \frac{(l-1)!}{\mu^l} \; for\ t,\ \mu \ \geq 0$$

$$= \frac{\mu^k}{(k-1)!} * \frac{(k+n-1)!}{\mu^{n+k}} \; for\ t,\ \mu \ \geq 0$$

$$= \frac{\mu^k}{(k-1)!} * \frac{(k+n-1)!}{\mu^{n+k}} \; for\ t,\ \mu \ \geq 0$$

$$= \frac{(k+n-1)!}{(k-1)! * \mu^n} \; for\ t,\ \mu \ \geq 0$$

$$= \frac{(k+n-1)(k+n-2)(k+n-3)\dots k(k-1)!}{(k-1)! * \mu^n} \; for\ t,\ \mu \ \geq 0$$

$$(\because (k+n-1)(k+n-2)(k+n-3)\dots k(k-1)!)$$

$$= \frac{(k+n-1)(k+n-2)(k+n-3)\dots k}{\mu^n} \; for\ t,\ \mu \ \geq 0$$

3.1.4 Mean & Variance

$$\text{Mean}(T) = \text{E}(T) = \frac{(k+1-1)}{\mu} = \frac{k}{\mu}$$

The second raw moment of an Erlang distribution is given as

$$\text{E}(T^2) = \frac{(k+2-1)(k+2-2)}{\mu^2} = \frac{(k+1)k}{\mu}$$

So the variance will be given as:

$$\text{Var}(T) = \text{E}(T^2) - (\text{E}(T))^2$$

$$= \frac{(k+1)k}{\mu^2} - \left(\frac{k}{\mu}\right)^2 = \frac{k}{\mu}$$

3.1.5 Coefficient of variation

The coefficient of variation is defined only for continuous variables. It is the ratio between the mean standard deviation and the mean value of the variable. Its discrete analogue is known as Lexis ratio which is the ration between variance and mean.

The coefficient of variation for Erlang distribution with shape k and rate μ is calculated as:

$$c = \frac{\sqrt{k/\mu^2}}{k/\mu} = \frac{1}{\sqrt{k}}$$

As $\mathrm{k} \geq 2, \mathrm{c} < 1$ for an Erlang distribution.

3.2 Hypo-exponential distribution

A Hypo-exponential distribution is the convolution of more than one exponential distribution but with different rates.

3.2.1 Phase transition matrix

The rate transition matrix of a Hypo-exponential distribution is given as

$$S = \begin{bmatrix} -\mu_1 & \mu_1 & \cdots & 0 & 0 & 0 \\ 0 & -\mu_2 & \mu_2 & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & -\mu_{k-2} & \mu_{k-2} & 0 \\ 0 & 0 & 0 & 0 & -\mu_{k-1} & \mu_{k-1} \\ 0 & 0 & 0 & 0 & 0 & -\mu_k \end{bmatrix}$$

3.2.2 Initial probability vector

The initial probability vector Grinstead, 1997) of a Hypo-exponential distribution is given by: $\pi = (1,0, \dots.,0)$. This indicates that the process starts at the first phase with probability one and from the other phase with probability 0. So, only the first phase can be the entry point of the process.

3.2.3 Cumulative distribution function

The cumulative distribution function of a random variable *T* with Hypo-exponential distribution with the phase transition matrix *S* and initial probability vector π is given according to theorem (Rodriguez, 2010: 10) as: $F(T) = 1 - \pi e^{tS} * 1$

Theorem: *If* $\tau \sim PH(\pi, T)$, *the distribution function of* τ *is given by*

$$F(s) = 1 - \pi\, exp(Ts)\, e.$$

The probability density function of a random variable T with Hypo-exponential distribution with the phase transition matrix S and initial probability vector π is given by theorem (Rodriguez, 2010: 9) as: $f_T(t) = -\pi e^{tS} S * 1$

Theorem: *If* $\tau \sim PH(\pi, T)$, *its density function of is given by*

$$f(s) = \pi\, exp(Ts)\, t,$$

$where\ t = -Te$

3.2.4 n^{th} Moment

By theorem of (Rodriguez, 2010: 10), n^{th} moment of a Hypo-exponential distribution is given as: $E(T^n) = (-1)^n n!\ \pi S^{-n} 1$

Theorem: *Let* $\tau \sim PH(\pi, T)$. *The* n^{th}*moment of* τ *is given by*

$$E(\tau^n) = (-1)^n n!\, \pi T^{-n} e.$$

3.2.5 Mean & Variance

$\mathrm{Mean}(T) = \mathrm{E}(T) = (-1)^1 1!\ \pi S^{-1} 1$

$$= -\ \pi S^{-1} 1\ = \sum_{i=1}^{k} \frac{1}{\mu_i}$$

$\mathrm{E}(T^2) = (-1)^2 2!\ \pi S^{-2} 1$

$$= 2\,\pi S^{-2}1$$

$$= 2\sum_{i=1}^{k} 1/\mu_i \sum_{l=1}^{i} 1/\mu_i$$

Therefore, $\mathrm{Var}(T) = \mathrm{E}(T^2) - \mathrm{E}(T)^2$

$$= 2\sum_{i=1}^{k} 1/\mu_i \sum_{l=1}^{i} 1/\mu_i - \left(\sum_{i=1}^{k}\frac{1}{\mu_i}\right)^2$$

$$= \sum_{i=1}^{k}\frac{1}{\mu_i^2}$$

3.2.6 Coefficient of variation

The coefficient of variation for Hypo-exponential distribution with parameters $\left(\mu_1, \mu_2 \dots \mu_k\right)$ is calculated as:

$$c = \frac{\sqrt{\sum_{i=1}^{k}\left(\frac{1}{\mu_i^2}\right)}}{\sum_{i=1}^{k}\left(\frac{1}{\mu_i}\right)}\ where\ \mu_i > 0\ \forall\, i$$

Now $\left(\sum_{i=1}^{k}\left(\frac{1}{\mu_i}\right)\right)^2 = \sum_{i=1}^{k}\frac{1}{\mu_i^2} + 2\sum_{i<j}^{k}\left(\frac{1}{\mu_i}\right)\left(\frac{1}{\mu_j}\right)$

$\mu_i > 0\ \forall\, i \Rightarrow \frac{1}{\mu_i} > 0\ \forall\, i \Rightarrow 2\sum_{i<j}^{k}\left(\frac{1}{\mu_i}\right)\left(\frac{1}{\mu_j}\right) > 0.$ Let us denote it as p.

Therefore, $\left(\sum_{i=1}^{k}\left(\frac{1}{\mu_i}\right)\right)^2 = \sum_{i=1}^{k}\frac{1}{\mu_i^2} + p\ where\ p > 0$

$$\Rightarrow \left(\sum_{i=1}^{k}\left(\frac{1}{\mu_i}\right)\right)^2 > \sum_{i=1}^{k}\frac{1}{\mu_i^2}$$

$$\Rightarrow \left(\sum_{i=1}^{k}\left(\frac{1}{\mu_i}\right)\right)^2 > \sum_{i=1}^{k}\frac{1}{\mu_i^2}$$

$\Rightarrow \sqrt{\left(\sum_{i=1}^{k}\left(\frac{1}{\mu_i}\right)\right)^2} > \sqrt{\sum_{i=1}^{k}\frac{1}{\mu_i^2}}$ (taking square root on bothe sides of the equation)

$$\Rightarrow \sum_{i=1}^{k}\left(\frac{1}{\mu_i}\right) > \sqrt{\sum_{i=1}^{k}\frac{1}{\mu_i^2}} \Rightarrow \frac{\sqrt{\sum_{i=1}^{k}\frac{1}{\mu_i^2}}}{\sum_{i=1}^{k}\left(\frac{1}{\mu_i}\right)} < 1$$

and therefore $c < 1\ for\ a\ Hypoexponential\ distribution.$

CHAPTER 4

APPLICATIONS OF ERLANG AND HYPO-EXPONENTIAL DISTRIBUTIONS

4.1 Applications of Erlang distribution

Fitting an Erlang distribution demands the assumption that the rate of each phase of a process is same. If the number of phases of a process is known in advance or rates of all the phases are assumed to be equal, then the coefficient of variation *c* of the total time of a process satisfies:

$$\frac{1}{\sqrt{k}} \leq c < \frac{1}{\sqrt{k-1}} \; for \; k \in \mathbb{Z}^{+}$$

Then Erlang-k distribution gives a better fit to the total time of the process (Adan, 2001). Following are the examples of some of the situations those can be modelled as an Erlang distribution.

4.1.1 Machine repairing system

In machine reparation processes, certain machines require several steps to be completed sequentially. If the time of repair spent in each phase is exponentially distributed with same parameters, then the total repairing time of the machines can be modelled as an Erlang distribution.

4.1.2 Compilation of the computer programs

Compilation of a computer program consists of several blocks that are processed sequentially, one after the other. If time spent is each block follows an exponential distribution with same rate independent of other blocks, then total compilation time will be sum of several exponential distributions with same rate. So, the total compilation time will have an Erlang distribution.

4.1.3 Hits on a web page

If the waiting time until next hit to a webpage is modelled as an exponential distribution with some rate $\mu > 0$, then the total waiting time till the n^{th} hit will be the sum of n exponential random variables with the same μ. Let $Y_1, Y_2 \dots Y_n$ are the waiting time for successful hits. Here $Y_i \sim exp(\mu)$ where $i = 1,2 \dots n.$ Then, the total time until n^{th} hit to the web page will be $Y = Y_1 + Y_2 \dots Y_n$ which is sum of n exponential random variables with the same rate μ and hence $Y \sim Erlang(n, \mu)$.

4.1.4 Traffic congestion in telecommunication network

Suppose arrival of calls to a customer care service follows exponential distribution with rate μ. We want to find the probability that it takes at least t minutes for n people to call. Let T_i denote the inter arrival time between $(i-1)^{th}$ and i^{th} call, then $T_i \sim exp(\mu)$. So, the total time T until receiving n successive calls is the sum of 'n' exponential random variables which can be denoted as: $T = T_1 + T_2 \dots T_n$. In other words, $Y \sim Erlang(n, \mu)$.

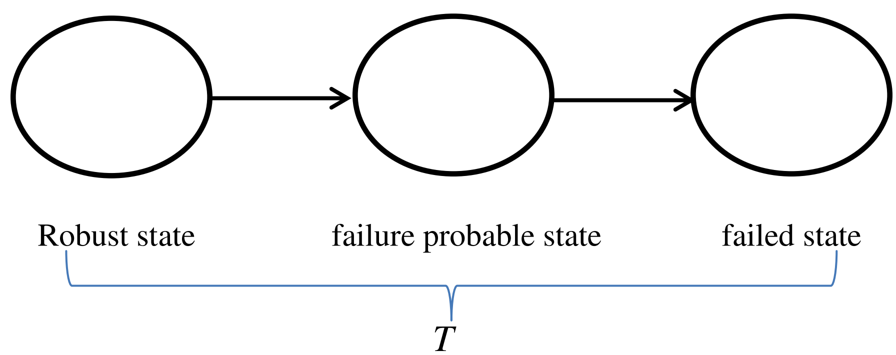
4.2 Applications of Hypo-exponential distribution

Suppose a service process consists of several phases providing different services with different rates, then the Hypo-exponential distribution will be a good choice for distribution of the total service time of the process. A Hypo-exponential distribution being a generalization of the Erlang distribution fits many real life processes and has more versatile use.

4.2.1 Software rejuvenation model

A software system can be modelled as three states as shown in the figure.

Figure 4.1 Software rejuvenation model

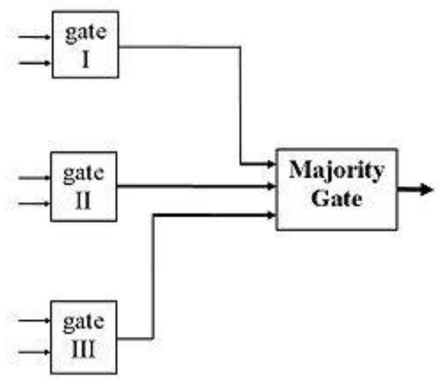


Let T_1 = Time spent in Robust state and $T_1 \sim exp\left(\mu_1\right)$ and T_2 = Time spent in failure probable state and $T_2 \sim exp\left(\mu_2\right)$.Then total life time T of the software can be modelled as a two phase Hypo-exponentially distributed random variable. So, $T = T_1 + T_2$. In other words, $T \sim Hypo\left(\mu_1, \mu_2\right)$.

4.2.2 Triple modular redundant system (TMR)

A triple modular redundant system is composed of three identical components, where two secondary components act as backup. The second component will be powered on only after the primary component fails and the third back up is activated only after the first two components fails. A detector circuit checks the output of primary component in order to identify its failure and a switch is used to configure and power on the secondary component.

Figure 4.2 Triple modular redundant system



Let the lifetimes of the three gates are three independent and exponentially distributed random variables denoted as T_1, T_2 and T_3 with parameter μ_1, μ_2 and μ_3 respectively. Then the distribution of time to failure of the whole system T can be modelled as sum of three exponential variables with different rates which can be calculated as: $T = T_1 + T_2 + T_3$ Alternatively, $T_1 \sim Hypo\left(\mu_1, \mu_2, \mu_3\right)$.

4.2.3 Disk service time

Performance of the hard disk is very important to the overall speed of the system. A slow hard disk hinders the potential of a fast processor like no other system components. The effective speed of a hard disk is determined by the disk service time. Disk access process consists of three sequential sub processes called disk seek, disk latency and disk transfer. If each sub process service time is assumed to be exponentially distributed with its own parameter μ_1, μ_2 and μ_3 respectively, then the disk service time may be modelled as a 3-phase Hypo-exponential distribution as the overall time is the sum of the seek time, the latency time and the transfer time.

4.2.4 Input output operation of a computer

The input-output operation of a computer is defined as transfer information between computer main memory and the outside world. The I/O operations consists of two phases of operations in sequence i.e. control operations and data transfer operations. So, the service time of input output operation of a computer can be modeled as a two phase Hypo-exponential distribution.

CHAPTER 5

METHOD OF ESTIMATION

For the parameter estimation of an Erlang distribution, we will be using two very traditional parameter estimation techniques. They are method of moment **(MOM)** and method of maximum likelihood estimation **(MLE).** For a Hypo-exponential distribution with more than two phases, we will approximate the distribution by using a recently discovered algorithm of Markus Sommereder (Sommereder, 2011) which will be explained in the next chapter.

5.1 Method of moment estimation

The *method of moment* is the oldest statistical technique for constructing estimators of the parameters which is based on matching the *sample moments* with the corresponding *distribution moments.* The method of moments was introduced by Karl Pearson in 1894.

Suppose we have random sample of size *n* of data $x = (x_1, x_2 \dots x_n)$ following certain probability distribution say f with respect to some parameter set $\theta = (\theta_{1,}\theta_{2,} \dots, \theta_k)$

So, the probability distribution function of *X* in terms of θ is given as: $f_X(x; \theta)$. Our objective is to estimate the *k* parameters of the distribution. Now suppose the first k population moments can be expressed as functions of the θs as below.

$$m_1 \equiv E[X] = g_1(\theta_{1,}\theta_{2,} \dots, \theta_k)$$

$$m_2 \equiv E[X^2] = g_1(\theta_{1,}\theta_{2,} \dots, \theta_k)$$

$$\vdots$$

$$m_k \equiv E[X^k] = g_1(\theta_{1,}\theta_{2,} \dots, \theta_k)$$

It is a system of k equations with k unknowns. The solution set of this system of equations will give an estimate of k parameters.

The method of moment estimation is fairly simple to implement. They provide good initial solutions for maximum likelihood estimation. But very often the solutions given by method of moment are biased.

5.2 Maximum likelihood estimation

Given a sample of data possessing certain probability distribution, the maximum likelihood estimation (MLE) tries to estimate the parameter by maximizing the joint probability density function of all the samples. In other words, it will find a parameter that when put to the probability density distribution, makes the observed data ‘‘most likely”. The principle of MLE was originally developed by R.A. Fisher in the1920s.

The maximum likelihood estimation begins with writing a mathematical expression known as the likelihood function of the sample data. After choosing a proper model for the data set, the likelihood of the data set is obtained by multiplying the probability density function of all the data. This expression contains unknown model parameters. Values of these parameters that maximize the sample likelihood are known as the Maximum Likelihood Estimates or MLE's.

Let $x = (x_1, x_2 \dots x_n)$ are 'n' independent and identically distributed random variables following certain probability distribution say f with respect to some parameter set $\theta = (\theta_{1,}\theta_{2,} \dots, \theta_k)$.

So, the probability distribution function of X in terms of θ is given as $f_X(x;\theta)$ and the joint density distribution of the dataset is given as:

$$f_\theta(x_1, x_2, \dots, x_n) = f_\theta(x_1, x_2, \dots, x_n \mid \theta)$$

Given observed values $X_1 = x_1,\ X_2 = x_2 \dots X_n = x_n$, the likelihood of θ is the function

$$\mathcal{L}(\theta) = f_\theta(x_1, x_2, \dots, x_n \mid \theta)$$

$$= \prod_{i=1}^{n} f_\theta(x_i \mid \theta)$$

Rather than maximizing this product which can be quite tedious, we often use the fact that the logarithm is an increasing function. So, it will be equivalent to maximize the log likelihood which is given as:

$$log(\mathcal{L}(\theta)) = l(\theta) = log\left(\prod_{i=1}^{n} f_\theta(x_i \mid \theta)\right)$$

$$= \sum_{i=1}^{n} log\big(f_\theta(x_i \mid \theta)\big)$$

The values of θ that maximize this function will give an estimate of the parameters of the model. The point where this objective function achieves its maxima, the derivative with respect to the parameters will vanish. So the solution can be found by equating the partial derivative of the log likelihood with respect to the parameters to 0 and the by solving a set of equations.

$$\frac{\partial}{\partial \theta_1} l(\theta) = 0$$

$$\frac{\partial}{\partial \theta_2} l(\theta) = 0$$

$$\vdots$$

$$\frac{\partial}{\partial \theta_k} l(\theta) = 0$$

This system of equations has *k* equation with *k* unknown parameters to be estimated which can be solved to get an estimate of the parameters $(\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_k)$.

There are some advantages of the maximum likelihood methods over other methods.

- MLE is consistent, asymptotically normal, and asymptotically efficient under some regularity conditions.
- MLE can be developed for a large variety of estimation situations.
- MLE often has lower variance than other methods.

There are also some supposed disadvantages

- MLE can be sensitive to the choice of starting values, or may not provide a global optimum.
- MLE can be highly biased for small samples.

CHAPTER 6

PARAMETER ESTIMATION OF PHASE TYPE DISTRIBUTION

6.1. Estimation for Erlang distribution when shape parameter is known

Suppose a process consists of k sequential phases to be executed where each phase service time is exponentially distributed with same rate μ. In other words, the service time of the process T has an Erlang distribution with shape k which is known to us and a rate μ which has to be estimated.

6.1.1 Method of moment

As here we need to estimate only one parameter i.e. the rate parameter of the Erlang distribution given that the shape parameter is known, we will need only the first distribution moment and the first sample moment. Let $T_1, T_2 \ldots T_n$ be some samples of data from this population.

First moment of an Erlang distributed variable $T \sim Erlang(k, \mu)$ is given by:

$$\mathrm{E}[T] = \frac{k}{\mu}$$

The sample moment based on the sample is given by

$$\mathrm{E}[T] = \mathrm{Mean}(T) = \frac{1}{n}\sum_{i=1}^{n} T_i = \bar{T}$$

Equating the population moment with the sample moment we get

$$\bar{T} = \frac{k}{\mu}$$

From above equation the rate parameter of the Erlang distribution can be estimated as:

$$\hat{\mu} = \frac{k}{\bar{\bar{T}}}$$

6.1.2 Maximum likelihood estimation

The probability density function of an Erlang distribution with parameters k and μ is given as

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

Let $T_1, T_2 \dots T_n$ are some observed values of total service time in a random sample of size *n*.

The likelihood function of the above sample is given as:

$$\mathcal{L}(k, \mu; t) = f(T_1; k, \mu) * f(T_2; k, \mu) * f(T_3; k, \mu) \dots . f(T_n; k, \mu)$$

$$= \left(\frac{\mu^k}{\Gamma(k)} * T_1^{k-1} * e^{-\mu T_1}\right) \left(\frac{\mu^k}{\Gamma(k)} * T_2^{k-1} * e^{-\mu T_2}\right) \dots \left(\frac{\mu^k}{\Gamma(k)} * T_n^{k-1} * e^{-\mu T_n}\right)$$

$$= \left(\frac{\mu^k}{\Gamma(k)}\right)^n (T_1 T_2 \dots T_n)^{k-1} e^{-\mu (T_1 + T_2 + \dots + T_n)}$$

Taking natural logarithm on both side of the equation:

$$ln\ (\mathcal{L}(k, \mu; T_1 T_2 \dots T_n) = ln\left(\left(\frac{\mu^k}{(k-1)!}\right)^n (T_1 T_2 \dots T_n)^{k-1} e^{-\mu (T_1 + T_2 + \dots + T_n)}\right)$$

$$\Rightarrow l((k, \mu; T_1 T_2 \dots T_n) = n(k\ ln\ \mu - ln(k-1)!) + (k-1) \sum_{i=1}^{n} ln\ T_i - \mu \sum_{i=1}^{n} T_i$$

$$Where\ ln(\mathcal{L}(k, \mu; t) = l((k, \mu; t)$$

The value of μ that maximizes the log likelihood function can be determined by solving the following equations:

$$\frac{\partial}{\partial \mu} l((k,\ \mu;\ T_1 T_2\ \ldots\ T_n) = n\ \frac{k}{\mu} - \sum_{i=1}^{n} T_i\ = 0\ or\ \widehat{\mu}\ =\ \frac{\hat{k}}{\bar{\bar{T}}}$$

This is the same as estimated by MOM. So, from this we can observe that if the shape parameter is known and fixed, we will get the same estimate of the parameter for the rate of an Erlang distribution by MOM and MLE.

6.2 Estimation for Erlang distribution when shape parameter is unknown

When both the parameters of an Erlang distribution are unknown, an approximation can be made to the original distribution. As the shape parameter of an Erlang distribution denotes the number of service phases in a process, so, it can take positive integer values. On the other hand the rate can be any positive real number. Therefore, Erlang distribution is a mixed parameter distribution where one parameter (shape) is a discrete random variable and the other parameter (rate) is a continuous random variable.

We know that the coefficient of variation of an Erlang distribution with *k* phases will be equal to $1/\sqrt{k}$ (Adan, 2001). So, the coefficient of variation of the sample can be taken to approximate the number of phases in a process.

Suppose we have a sample of data whose coefficient of variation is $c < 1$ and let the process consists of *k* phases in sequence. From this the shape parameter (*k*) can be approximated as below.

$$c =\ 1/\sqrt{k}\quad \Rightarrow k = round\left(1/c^2\right)$$

Now once the shape parameter has been estimated, the rate parameter of the distribution can be estimated as:

$$\hat{\mu} = \hat{k}/\bar{T} \quad where\ \bar{T}\ is\ the\ mean\ of\ the\ sample.$$

Finally a local search will be performed in the neighborhood of k to check if our estimate is optimum and if no then, for what values of k and μ maximum likelihood occurs.

6.3 Parameter estimation of Hypo-exponential distribution

For a Hypo-exponential distribution, each phase has its own service rate. As the number of phases increases, the number of rate parameters to be estimated also increases. So, there exists no exact estimation method that can be applied to a Hypo-exponential distribution directly. Rather each Hypo-exponential distribution has to be analyzed separately depending upon the number of phases it consists. But there exists an approximation algorithm to approximate the distribution.

When the number of phase of a Hypo-exponential distribution is two, the traditional method of estimation can be applied but as the number of phases increases, the complexity of the system increases. In this thesis, to approximate such Hypo-exponential distribution, we will use the algorithm provided by Markus Sommerender (Sommereder, 2011).

6.3.1 Estimation using MOM & MLE when shape parameter with two phases

Let the rates of Hypo-exponential distribution are μ_1 and μ_2. With no loss of generality we can assume that $\mu_1 < \mu_2$. Let $T_1, T_2 \dots T_n$ are some observed values of total service time in a random sample of size n. The first two raw moments of the sample m_1 and m_2 are given as:

$$m_1 = \mathrm{E}[T] \; and \; m_2 = \mathrm{E}[T^2]$$

The two measures of moments of *Hypo* (μ_1, μ_2) are given as:

Mean $= \frac{1}{\mu_1} + \frac{1}{\mu_2} = m_1$

$(equating\ the\ first\ distribution\ moment\ with\ the\ first\ sample\ moment\ )$

Variance $= \frac{1}{\mu_1^2} + \frac{1}{\mu_2^2} = m_2 - m_1^2$

$(equating\ the\ second\ distribution\ moment\ with\ the\ second\ sample\ moment\ sample\ moment)$

Let us denote $\frac{1}{\mu_i} = s_i \; where\ s_i\ time\ spent\ in\ i^{th}\ stage.$

$\mu_1 < \mu_2 \;\Rightarrow\; s_1 > s_2.$

Substituting values of $\mu_1 \; and \; \mu_2$, following sets of equations are obtained.

$s_1 + s_2 = m_1 \Rightarrow s_2 = m_1 - s_1$

$s_1^2 + s_2^2 = m_2 - m_1^2$

$\Rightarrow s_1^2 + (m_1 - s_1)^2 = m_2 - m_1^2$

$\Rightarrow s_1^2 + m_1^2 - 2 * m_1 * s_1 + s_1^2 = m_2 - m_1^2$

$\Rightarrow 2s_1^2 - 2 * m_1 * s_1 + 2m_1^2 - m_2 = 0$

This is a quadratic equation in s_1 whose solution is given by:

$$s_1 = \frac{m_1 \pm \sqrt{2m_2 - 3m_1^2}}{2}$$

And putting the values of s_1, s_2 can be obtained as:

$$s_2 = \frac{m_1 \mp \sqrt{2m_2 - 3m_1^2}}{2}$$

Assuming $s_1 > \ s_2$, the final set of estimated values are found as:

$$(\hat{s}_1, \hat{s}_2) = \left(\frac{m_1 + \sqrt{2m_2 - 3m_1^2}}{2}, \frac{m_1 - \sqrt{2m_2 - 3m_1^2}}{2}\right)$$

Finally substituting the values of s_1 and s_2, the rates of the Hypo-exponential distribution can be estimated as:

$$(\hat{\mu}_1, \hat{\mu}_2) = \left(\frac{2}{m_1 + \sqrt{2m_2 - 3m_1^2}}, \frac{2}{m_1 - \sqrt{2m_2 - 3m_1^2}}\right)$$

To find the maximum likelihood estimates of a two phase Hypo-exponential distribution, results obtained from MOM can be taken as initial guess for MLE and can be improved.

The probability density function of a Hypo-exponential distribution with rates μ_1 and μ_2 is given by:

$$f\left(t;\ \mu_1,\ \mu_2\right) = \frac{\mu_1}{\mu_1 - \mu_2} \mu_2 e^{-t\,\mu_2} - \frac{\mu_2}{\mu_1 - \mu_2} \mu_1 e^{-t\,\mu_1}$$

Let $T_1, T_2 \ldots T_n$ are observed values of total service time in a random sample of size *n*.

The maximum likelihood function of this distribution is given as:

$$\mathcal{L}\left(\mu_1,\ \mu_2;\ T_1 T_2\ \ldots\ T_n\right)$$

$$= f\left(T_1;\ \mu_1,\ \mu_2\right) * f\left(T_2;\ \mu_1,\ \mu_2\right) * f\left(T_3;\ \mu_1,\ \mu_2\right) \ldots.\ f\left(T_n;\ \mu_1,\ \mu_2\right)$$

$$= \left(\frac{\mu_1}{\mu_1 - \mu_2} \mu_2 e^{-T_1\,\mu_2} - \frac{\mu_2}{\mu_1 - \mu_2} \mu_1 e^{-T_1\,\mu_1}\right) \ldots \left(\frac{\mu_1}{\mu_1 - \mu_2} \mu_2 e^{-T_n\,\mu_2} - \frac{\mu_2}{\mu_1 - \mu_2} \mu_1 e^{-T_n\,\mu_1}\right)$$

Taking natural logarithm on both the sides

$$ln\,\mathcal{L}\left(\mu_1, \mu_2;\, T_1, T_2\, \dots\, T_n\right) = l\left(\mu_1, \mu_2;\, T_1, T_2\, \dots\, T_n\right)$$

$$= \sum_{i=1}^{n} ln\left(\frac{\mu_1}{\mu_1 - \mu_2}\, \mu_2\, e^{-T_i \mu_2} - \frac{\mu_2}{\mu_1 - \mu_2}\, \mu_1 e^{-T_i \mu_1}\right)$$

The values of μ_1 and μ_2 that maximize the log likelihood function can be determined by solving the following equations:

$$\frac{\partial}{\partial \mu_1} l\left(\mu_1, \mu_2;\, T_1 T_2\, \dots\, T_n\right) = 0$$

$$\Rightarrow \frac{\partial}{\partial \mu_1}\left(\sum_{i=1}^{n} ln\left(\frac{\mu_1}{\mu_1 - \mu_2}\mu_2 e^{-T_i \mu_2} - \frac{\mu_2}{\mu_1 - \mu_2}\mu_1 e^{-T_i \mu_1}\right)\right) = 0$$

and $\frac{\partial}{\partial \mu_2} l(\left(\mu_1, \mu_2;\, T_1 T_2\, \dots\, T_n\right) = 0$

$$\Rightarrow \frac{\partial}{\partial \mu_2}\left(\sum_{i=1}^{n} ln\left(\frac{\mu_1}{\mu_1 - \mu_2}\, \mu_2\, e^{-T_i \mu_2} - \frac{\mu_2}{\mu_1 - \mu_2}\, \mu_1 e^{-T_i \mu_1}\right)\right) = 0$$

Solving these two partial equations gives rise to a set of two equations with two variables as below.

$$\mu_1 \sum_{i=1}^{n} \frac{e^{-\mu_1 T_i}}{e^{-\mu_2 T_i} - e^{-\mu_1 T_i}} - \frac{n\,\mu_2}{\mu_1(\mu_1 - \mu_2)} = 0$$

$$\mu_2 \sum_{i=1}^{n} \frac{e^{-\mu_2 T_i}}{e^{-\mu_2 T_i} - e^{-\mu_1 T_i}} - \frac{n\,\mu_2}{\mu_2(\mu_1 - \mu_2)} = 0$$

Both of these equations are not in closed form. So, they can be solved by any numerical method to estimate the values of μ_1 and μ_2. For our simulation we will be using Newton Raphson method to find optimal solutions of such equations.

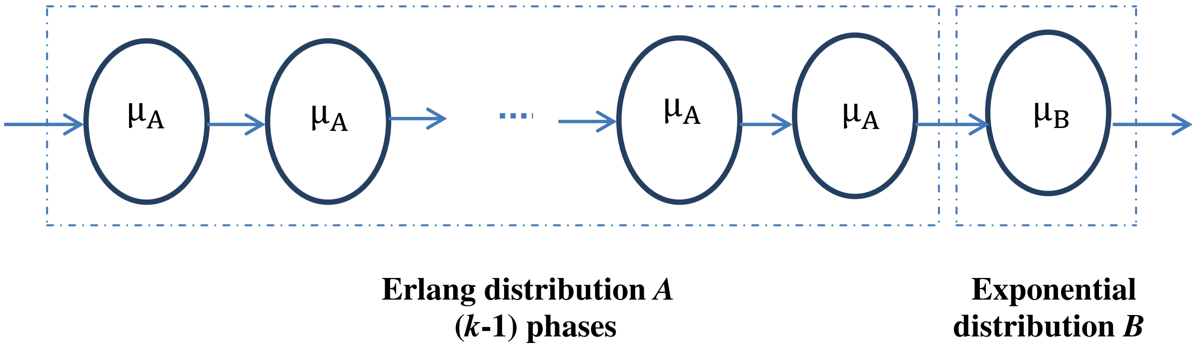
6.3.2 Estimation when number of phases is more than two

As the number of phases becomes three or more, these methods will become very cumbersome to apply directly. Sometimes they give imaginary values. To handle such distributions, Markus Sommereder has proposed a curve fitting algorithm in his book "Modelling of Queueing Systems with Markov Chains" (Sommereder, 2011) for better approximation to this kind of multiphase Hypo-exponential distributions. He used the fact that every Hypo-exponential distribution with more than two phases can be approximated by a combination of an Erlang distribution and an exponential distribution. Based upon this fact we are proposing here two algorithms to approximate such cases.

I. Algorithm-1

Suppose a Hypo-exponential distribution has k number of service phases. This Hypo-exponential distribution can be divided into a combination of k-1 phases Erlang distribution A and an exponential distribution B as denoted by the following diagram.

Figure 6.1 Hypo-exponential distribution as sum of an Erlang and an exponential distribution



Now this can be considered as a two phase Hypo-exponential distribution with rates μ_A and μ_B. The maximum likelihood estimates can be found for these rates using method described above for two phase Hypo-exponential distribution. Let's denote the estimated solutions to be $\widehat{\mu_A}$ and $\widehat{\mu_B}$. But as *A* is an Erlang distribution with *k*-1 phases, so rate of each phase of Erlang distribution will be:

$$(k-1) * \widehat{\mu_A}$$

Thus the final estimated solutions found by MLE will be given as:

$$\left((k-1) * \widehat{\mu_A}, (k-1) * \widehat{\mu_A} \ldots (k-1)\ times, \widehat{\mu_B}\right)$$

II. Mark Sommereder's algorithm

Suppose a Hypo-exponential distribution has *k* phases. This Hypo-exponential distribution can be divided into a combination of *k*-1 phase Erlang distribution *A* and an exponential distribution *B* as denoted in the figure 6.1.

Let us denote c, c_A and c_B are the coefficient of variations of the given Hypo-exponential distribution, Erlang distribution A and exponential variation B respectively. Let μ_A and μ_B are the rates of A and B respectively. An Erlang distribution with k-1 phases will have coefficient of variation (Adan, 2001) equals to:

$$\frac{1}{\sqrt{k-1}}$$

which means

$$c_A = \frac{1}{\sqrt{k-1}}.$$

The coefficient of variation of an exponential distribution is 1. Therefore,

$$c_B^2 = \frac{\mathrm{Var}(B)}{\big(\mathrm{E}(B)\big)^2} = 1 \; or \;\; \mathrm{Var}(B) \; = \; (\mathrm{E}(B))^2$$

Now the total coefficient of variation of the entire Hypo-exponential distribution c is given as

$$c = \frac{\sqrt{\mathrm{Var}(A) + \mathrm{Var}(B)}}{\mathrm{E}(A) + \mathrm{E}(B)}$$

Let us assume that mean of A is 1 or $\mathrm{E}(A) = 1$.

$$c_A = \frac{\sqrt{\mathrm{Var}(A)}}{\mathrm{E}(A)} = \frac{1}{\sqrt{k-1}} \quad or \quad \mathrm{Var}(A) = \frac{1}{k-1}$$

Therefore

$$c = \frac{\sqrt{\frac{1}{k-1} + \mathrm{E}(B)^2}}{1 + \mathrm{E}(B)} \;\; or \;\; c^2 = \frac{\frac{1}{k-1} + \mathrm{E}(B)^2}{1 + \mathrm{E}(B)^2}$$

This is a quadratic equation in $\mathrm{E}(B)$ which can be solved for $\mathrm{E}(B)$ to get the following solution.

$$\mathrm{E}(\hat{B}) = \frac{c^2 \pm \sqrt{\frac{c^2 k - 1}{k - 1}}}{k - 1}$$

But mean time of entire process is sum of mean time spent in *A* and mean time spent in *B*.

$$\mathrm{E}(T) = \mathrm{E}(A) + \mathrm{E}(B) \Rightarrow \mathrm{E}(\hat{A}) = \mathrm{E}(T) - \mathrm{E}(\hat{B})$$

Furthermore, $\widehat{\mu_A} = \dfrac{1}{\mathrm{E}(\hat{A})}$ and $\widehat{\mu_B} = \widehat{\mu_k} = \dfrac{1}{\mathrm{E}(\hat{B})}$

Finally *A* is an Erlang distribution with rate μ_A and *k*-1 phases. So, the rate of each phase is given by:

$$\widehat{\mu_1} = \widehat{\mu_2} = \cdots = \widehat{\mu_{k-1}} = (k-1) * \widehat{\mu_A}$$

Therefore complete solution to the above Hypo-exponential distribution with *k* phases is given by ($\widehat{\mu_1}$, $\widehat{\mu_2}$ ... $\widehat{\mu_{k-1}}$, $\widehat{\mu_k}$).

6.3.3 When shape parameter is unknown

If the shape of a Hypo-exponential distribution is unknown, the estimation has to be performed in two steps. In first step, shape parameter will be estimated. In second step, rate parameter for the corresponding shape will be calculated. This algorithm is described below.

Curve fitting algorithm

Step-1: Estimating number of phases in a Hypo-exponential distribution

For a Hypo-exponential distribution with k phases and each phase with different rate, the relation between coefficient of variation c and number of phase k is given by (Adan, 2001) as:

$$\frac{1}{\sqrt{k-1}} > c > \frac{1}{\sqrt{k}}$$

So from the coefficient of variation of the data, the value of k can be estimated as:

$$\hat{k} = \left\lceil \frac{1}{c^2} \right\rceil$$

In other words when $c > \frac{1}{\sqrt{k}}$, then we need at least k phases to approximate the distribution.

Step-2: Approximation of rates of each phase

Once the k has been estimated, one of the algorithms given in section 6.2.1 for Hypo-exponential distribution with known shape parameter can be used to estimate rates of the phases.

CHAPTER 7

RESULT AND ANALYSIS

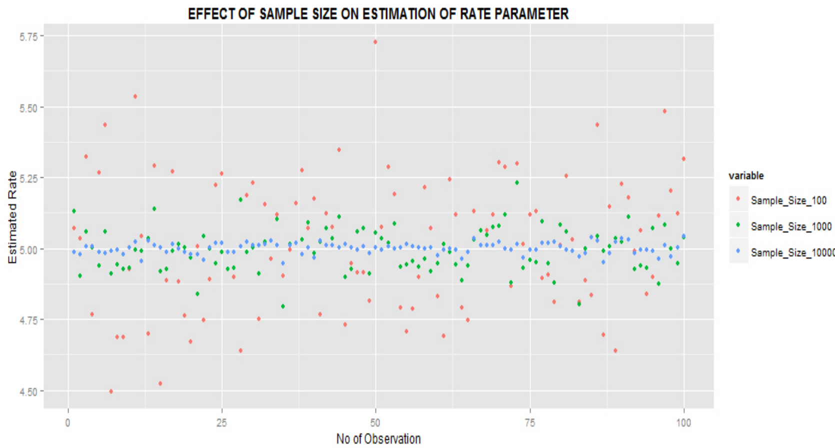
7.1 Parameter estimation for Erlang distribution when shape parameter is known

For this thesis work, all results are simulated using RStudio 0.97.312 version on Windows- XP platform. RStudio is a free and open source integrated development environment (IDE) for R, a programming language for statistical computing and graphics.

7.1.1 Effect of sample size on estimation of rate

It is evident that with an increase in sample size, the estimation becomes more and more accurate. Nevertheless to see how the sample size effects the estimation of rate parameter in an Erlang distribution, we simulate the outcome of our analysis taking different sample size i.e. for n=100, n=1000, n=10000 respectively of same population generated by taking k=5 and μ=5. The obtained result is shown in the following diagram.

Figure7.1 Effect of simple size on estimation of rate



It is clear from the figure that the deviation of the estimated value is more if the sample size is small. We see highest deviation from the original rate for n=100. But for n=10000, estimated rate line is very close to the original rate line. To measure the deviation of the estimated rate from the original rate, we repeat the experiment taking k=5 and μ=5 taking n=100, n =1000 and n =10000. For each sample size, we make 1000 repetitions to get 1000 estimates of rate from which the variance is calculated. The result is listed in the following table.

Table 7.1 Variance vs. Sample size of Erlang distribution with known shape

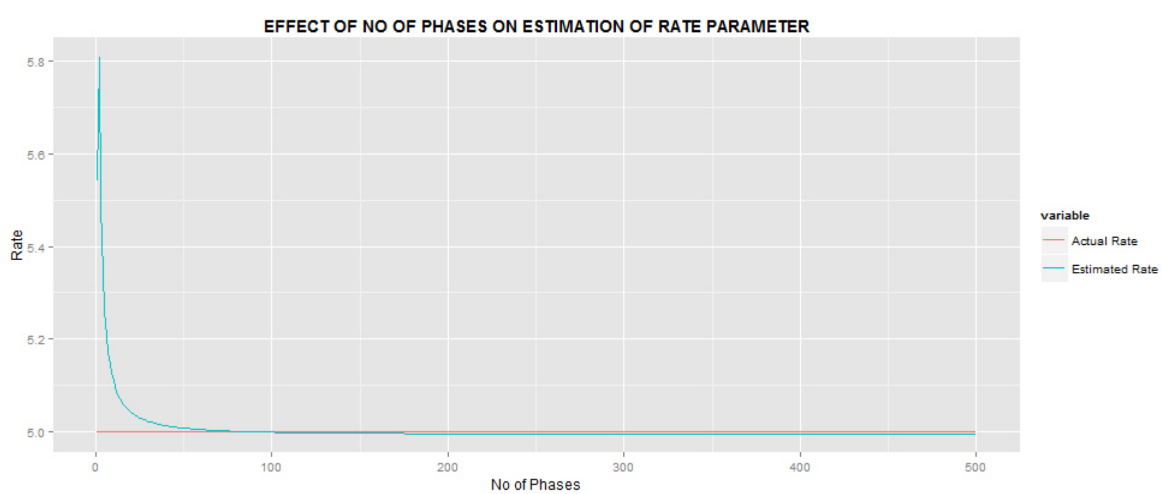
SAMPLE SIZE	ACTUAL RATE	MEAN ESTIMATED RATE	VARIANCE	SAMPLE SIZE RATIO	VARIANCE RATIO
10	5	5.119231736	0.5748992	-	-
100	5	5.002251978	0.05093505	10	$\approx \frac{1}{10}$
1000	5	4.998244856	0.00504865	100	$\approx \frac{1}{100}$
10000	5	5.000357957	0.00051769	1000	$\approx \frac{1}{1000}$

It can be observed from the table, as the sample size increases, the variance of estimation decreases because as sample size increases, mean of the sample becomes more accurate and centered towards its original value which leads to give more accurate estimation of the rate parameter. The percentage of decrease in variance is found to be approximately same as the percent of increase in sample size. This means if the sample size increases by 10^x times, the variance decreases approximately by 10^x times which is evident from the variance ratio.

7.1.2 Effect of number of phases on estimation of rate

To see if number of phases in a process has any effect on the estimation of rate parameter of an Erlang distribution, we fix the value of μ to be 5 and simulate the result obtained by traditional statistical estimation techniques (MOM or MLE) for a sample size of 100 and varying value of k from 1 to 500. The result of this simulation is depicted in the following diagram.

Figure 7.2 Effect of number of phases on estimation of rate parameter of Erlang distribution when shape parameter is known



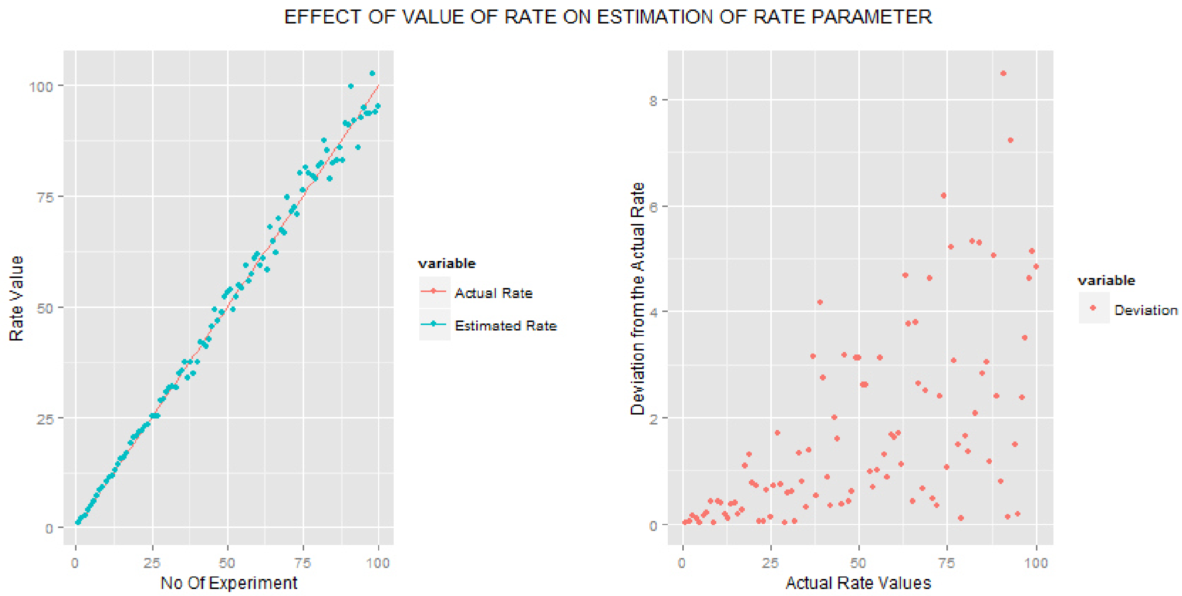
From the above graphic it can be observed that for this simulation, maximum deviation of the estimation of rate occurs at k=2 which gives μ=5.808889041. But as the number of phase increases, the estimated rate gets more and more close to the actual value. From this we can say that as the number of phase increases, then the accuracy estimation of rate parameter increases and for very large value of k, estimated rates becomes equal to actual rate. In other words: $lim_{k\to\infty}\, \hat{\mu} = \mu$. This seems to happen following the law of large[1] numbers. As $k \to \infty$. the number of exponential distributions approximating the Erlang distribution becomes ∞ and this leads to give an exact estimation of rate parameter of the distribution.

7.1.3 Effect of value of rate on estimation of rate

If the k and n is fixed does the value μ has any effect on its estimation, to answer this question we performed simulation taking k=5 and n=100 while varying value of μ from 1 to 100. The result is depicted in the following diagram.

[1] According to the law, the average of the results obtained from a large number of trials should be close to the expected value, and will tend to become closer as more trials are performed.

Figure 7.3 Effect of value of rate on estimation of rate parameter of Erlang distribution when shape parameter is known



From the figure it is clear that for lower rate, the estimation is more accurate than those with higher values. When the rate is less, the estimated rate is more probable to be close to the original rate. But as it increases, the deviation from the original value also increases. In this case the deviation is highest when rate is 91 when the estimated rate becomes 99.48592705. To see how the variance is related to the value of rate, we run the experiment fixing n=100 and k=5 while varying μ =1, μ=100, μ=1000 respectively. The result is shown in the following table.

Table 7.2 Variance vs. Actual rate of Erlang distribution with known shape

ACTUAL RATE	MEAN ESTIMATED RATE	VARIANCE	ACTUAL RATE RATIO	VARIANCE RATIO
1	1.00045	0.002037402	-	-
10	10.00450	0.2037402	10	$\frac{1}{100}$
100	100.04504	20.37402	100	$\frac{1}{10000}$
1000	1000.45040	2037.402	1000	$\frac{1}{100000}$

From the table it can be observed that when the rate increases by a factor of 10,100 and 1000, the variance of estimation increases approximately by 100, 10000 and 1000000 respectively. In other words if the rate increases by 10^x, the variance will be increased approximately by 10^{2x}. This means if rate of a process is high, then variance of estimation will also be high and hence we need large sample of data in order to minimize the variance and to better estimate the distribution.

7.2 Estimation for Erlang distribution when shape parameter is unknown

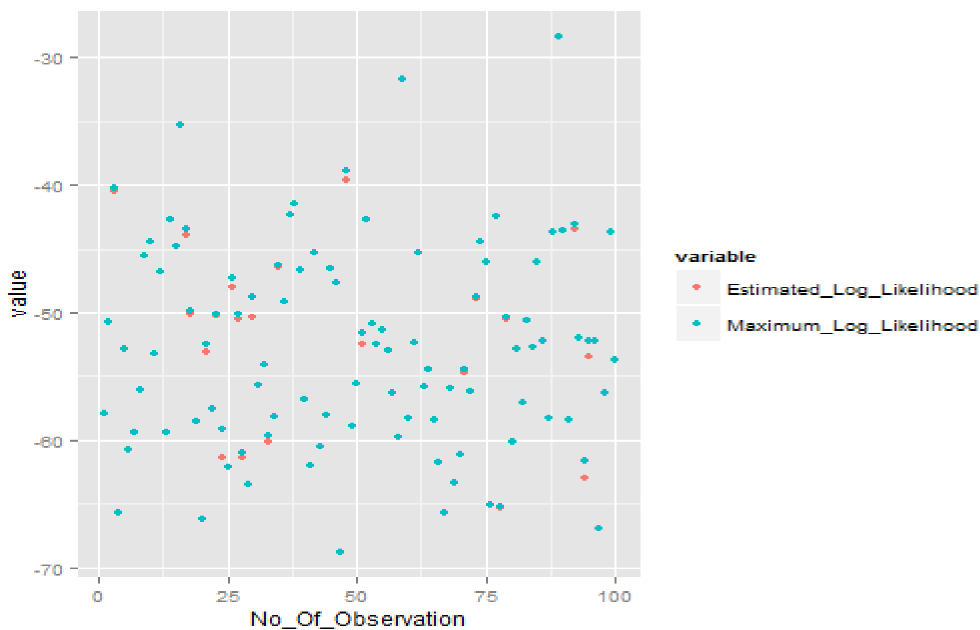
7.2.1 Simulation taking shape and rate both equal to 5

To see an example, the result obtained from our analysis of estimation of shape and rate parameters of Erlang distribution when both are unknown is simulated performing 100 experiments arbitrarily setting shape and rate both to 5. Each experiment consists of 100 samples of data. For each experiment, estimated shape, estimated rate and the log likelihood values are found. To check if (estimated shape, estimated rate) is the point where maximum likelihood occurs for the given data or there is any deviation, the log likelihood values are checked around the estimated shape in an interval:

$$\left[\left(minimum\left((estimated\_shape - 5), 1\right)\right), (estimated\_shape + 1)\right]$$

The result is depicted in the following figure.

Figure 7.4 Log likelihood estimation of an Erlang distribution when both shape and rate parameters are unknown



It can be viewed from the figure that most of the times our estimation gives maximum likelihood estimation. In fact out of 100 experiments, in 76 experiments we get maximum likelihood estimation by our proposed estimation algorithm. In other 24 points where maximum likelihood occurs in the neighborhood of the estimated phase, the difference is very small. Here the maximum deviation between estimated log likelihood and maximum log likelihood occurs during 24th experiments where the shape and rates are estimated to be 6 and 6.020859317 respectively while maximum likelihood occurs at phase 4 with rate

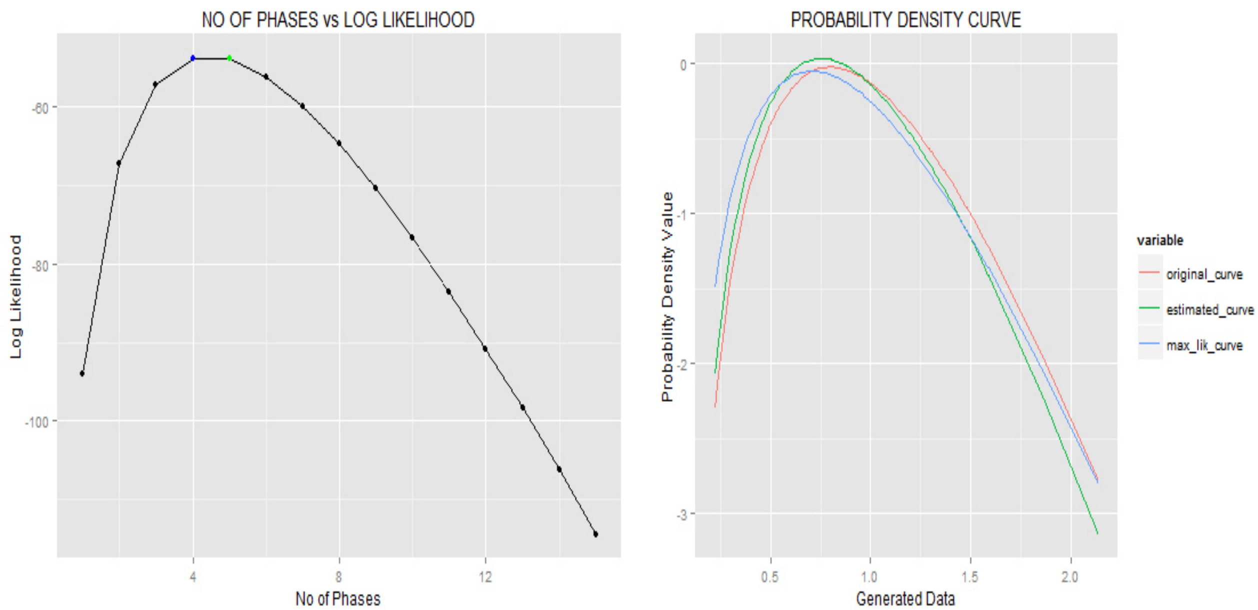
4.013906212. Estimated log likelihood and maximum log likelihood are found to be -61.37416426 and -59.12778444 respectively. So, the difference is only 2.246379822. In other words we can say if we replace our estimated shape and rate by maximum likelihood shape and rate, then the joint probability density function value will increase only by an amount 1.873153e-26 which is very small. To see how accurately we approximate a given curve of Erlang distribution, simulation is done taking k=5 and μ=5 for a random sample of data with 100 samples until we get different values of estimated log likelihood and maximum log likelihood. A screen shot of such case is given below.

What is shape of the Erlang distribution k? 5
What is rate of the Erlang distribution r? 5
What is sample size n? 100

	TITLE	VALUE
1	ORIGINAL SHAPE	5.000000
2	ESTIMATED SHAPE	5.000000
3	MAX-LIKE-SHAPE	4.000000
4	ORIGINAL RATE	5.000000
5	ESTIMATED RATE	5.304504
6	MAX-LIKE-RATE	4.243603
7	ESTIMATED LOG LIKLIHOOD	-53.852889
8	MAX-LOG LIKELIHOOD	-53.755090

The result is depicted in the following two figures.

Figure 7.5 Probability density curves of Erlang distribution for original, estimated and maximum log likelihood values when both parameters are unknown



For all the phases in the range:

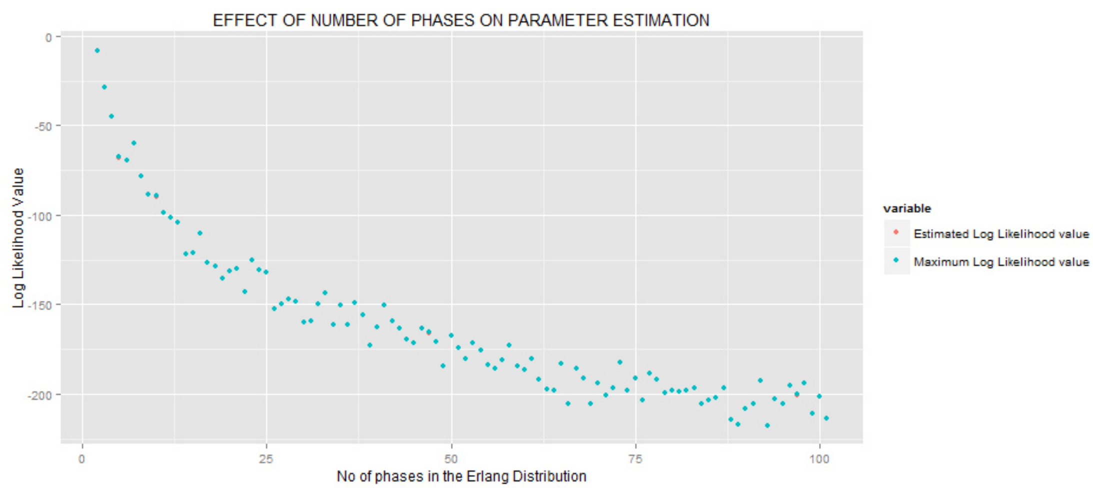
$$\left[\left(minimum\left((estimated\_shape - 10), 1\right)\right), (estimated\_shape + 10)\right],$$

the log likelihood values are calculated and plotted against their corresponding phase. This result is depicted in the first plot of figure 7.5. The second plot consists of three curves i.e. the probability density curve with respect to the actual shape and actual rate, probability density curve with respect to the estimated shape and estimated rate and the probability density curve with maximum log likelihood values. From the second plot we can say that even if our estimated values does not correspond to the maximum likelihood value, but still it gives a very close approximation to the original curve and also the deviation from the maximum likelihood curve is very low.

7.2.2 Effect of variation of shape parameter on estimation

To study the effect of the number of phases on the estimation algorithm, we simulate the result taking n=100, μ=5 and varying value of k from 2 to 100. The outcome is shown as below.

Figure 7.6 Effect of variation of shape parameter on estimation of both parameters of Erlang distribution when both parameters are unknown



Outcome of the simulation shows that, our proposed algorithm gives optimum result indifferent of the values of k. In other words, this algorithm gives maximum likelihood estimate for the set of data even for high values of number of phases. This can be shown as almost in every points, estimated log likelihood of the algorithm coincides with maximum log likelihood.

7.3 Parameter estimation for Hypo-exponential distribution when shape parameter is known with two phases

7.3.1 MOM & MLE estimation

Taking k=2 and n=100, rates are estimated by traditional MOM and then improved by MLE.

Result is depicted in the following diagram.

Figure 7.7 Probability density curve of Hypo-exponential distribution as estimated by MOM & MLE when number of phase is two

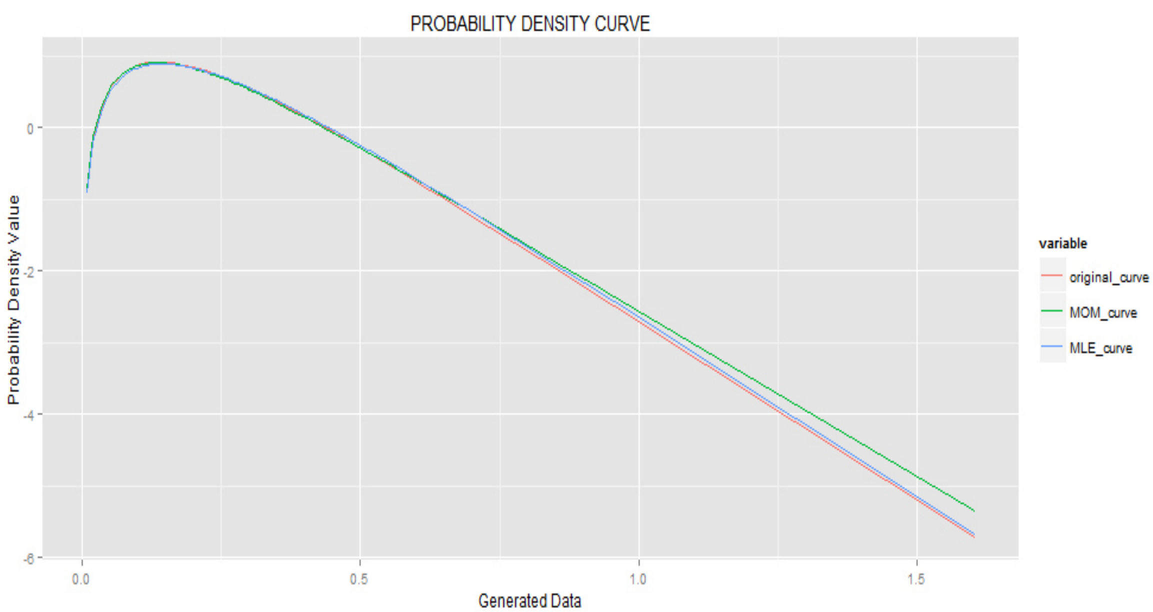


Table 7.3 Log likelihood values and rates of a two phase Hypo-exponential distribution by MOM &MLE

	RATES	LOG LIKELIHOOD
ACUAL VALUES	5,10	27.82216
MOM VALUES	4.38633,11.057331	24.69732
MLE VALUES	5.020418,9.360505	27.86882

For our simulation we got optimum result using MLE as it yields maximum log likelihood 27.86882 which is even better than the original log likelihood value.

7.3.2 Effect of increase in μ_2 on estimation of rate of μ_1

Simulation is done by fixing $\mu_1 = 1$ while varying μ_2 from 2 to 1000. The estimated rates are plotted against original rates to get the following figure.

Figure 7.8 Effect of increase in rate in a Hypo-exponential distribution on estimation of rate parameter by MLE when number of phase is two

From the first figure, we can say that the estimation of rate of a phase depends on the rate of the other phase but up to certain point. This is very evident as μ_2 increases, initially estimated μ_1 goes closer to its actual value and then after some point it again goes away from its original value up to certain point and again with the increase in μ_2, it approaches to its actual value. This way μ_1 oscillates around its original value. But after certain value of μ_2, estimated μ_1 becomes constant irrespective of the value of μ_2. This occurs because here maximum likelihood estimate were found by using Newton Raphson method that uses derivative of the previous guess to improve the guess in the next iteration. At certain high value of μ_2, this derivative vanishes after reaching a particular solution leading the error term to 0 and thus the solution can't be improved anymore. So, estimated μ_1 becomes constant after this value.

In the second figure we don't find any regular pattern. At first the deviation between estimated μ_2 and actual μ_2 seems to increase to certain point around 60 and then estimated μ_2 goes closer to its actual value. This behavior continues and for higher values of μ_2, the estimated value follows a very irregular pattern. This can be said because for some very high values of μ_2, we get good estimation but for some others, the deviation is very much high.

7.4 Parameter estimation for Hypo-exponential distribution when shape parameter is known with more two phases

To find how accurately the approximation is made by our stated algorithm, simulation is performed with n = 100 and varying k from 1 to 5. The corresponding rates are assigned in the multiple of 5 starting from 5. Result of the simulation using MLE and Mark Sommereder algorithm for different values of k and rates are listed in the following two tables.

Table 7.4 Log likelihood values and rates of Hypo-exponential distribution with two or more phases by MOM &MLE

NO OF PHASES	ACTUAL RATES	ESTIMATED RATES	NO OF ITERATIONS PERFORMED	ACTUAL LOG LIKELIHOOD	ESTIMATED LOG LIKELIHOOD
2	5,10	6.574947, 6.573979	17	30.11103	30.67761
3	5,10,15	10.662834,10.662834, 5.334286	11	19.52041	17.45173
4	5,10,15,20	13.961742, 13.961742, 13.961742, 4.653031	16	12.85738	7.299488
5	5,10,15,20,25	17.098868, 17.098868, 17.098868, 17.098868 4.274399	10	9.076633	0.6102097

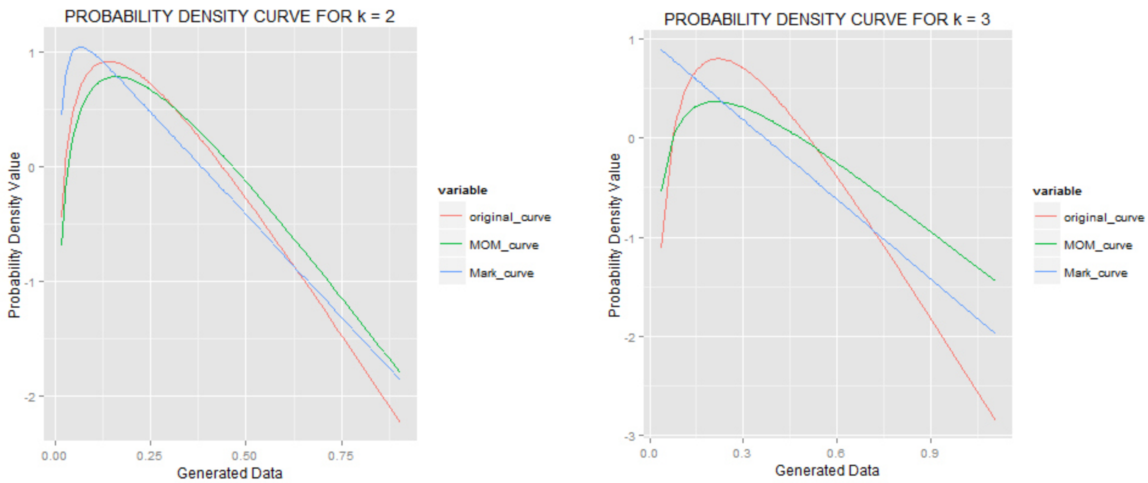
Table 7.5 Log likelihood values and rates of two phase Hypo-exponential distribution by MARK SOMMEREDER algorithm & MLE

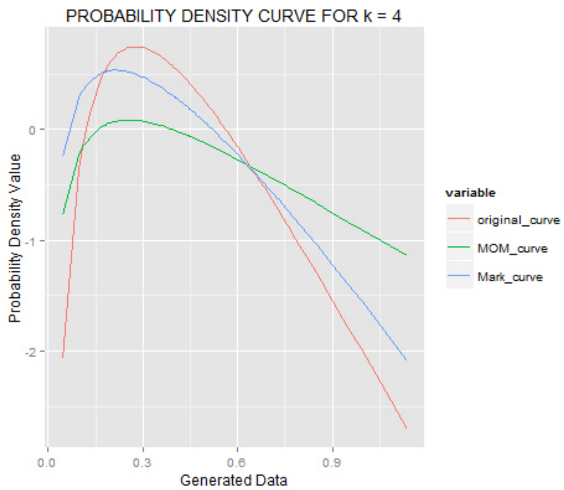
NO OF PHASES	ACTUAL RATES	APPROXIMATED RATES	NO OF ITERATIONS PERFORMED	ACTUAL LOG LIKELIHOOD	ESTIMATED LOG LIKLIHOOD
2	5,10	6.512225, 6.520448	5	30.11103	30.66998
3	5,10,15	10.668783, 10.668783, 5.331363	9	19.52041	17.45173
4	5,10,15,20	13.961025, 13.961025, 13.961025, 4.653876	11	12.85738	7.299488
5	5,10,15,20,25	17.099388, 17.099388, 17.099388, 17.099388, 4.274678	10	9.076633	0.6102099

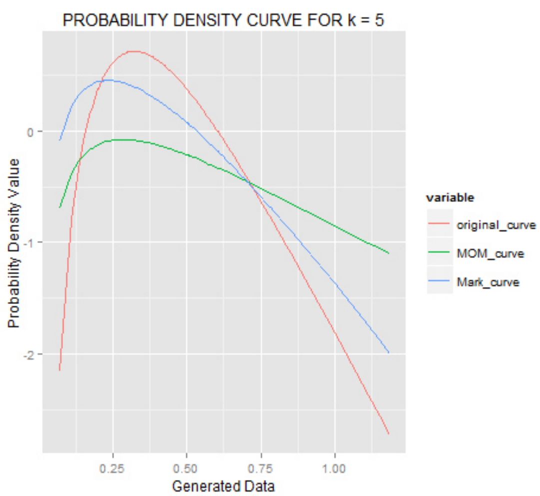
Comparing the last columns of table 7.4 with table 7.5, it can be concluded that Mark Sommereder algorithm always provide equally good or better initial solution for maximum likelihood estimation techniques then that of MOM irrespective of the number of phases. For our data sample, it is yielding better solution for k=2 and k=5. Moreover Mark Sommereder algorithm approaches to the optimal solution in less or equal number of iterations.

Diagrammatic comparisons of these two techniques before applying MLE are given below.

Figure 7.9 Probability density curves of Hypo-exponential distribution for different number of phases as approximated by MOM and Mark Sommereder algorithm







7.5 Estimation when shape parameter is unknown

Result of the simulation using MOM and Mark Sommereder algorithm for different values of shapes and rates are listed in tables 7.6 and table 7.7 respectively.

Table 7.6 Log likelihood values and rates of a Hypo-exponential distribution by MOM &MLE with shape unknown

NO OF PHASES	ESTIMATED NO OF PHASES	ACTUAL RATES	APPROXIMATED RATES OF PHASES	ACTUAL LOG LIKELIHOOD	ESTIMATED LOG LIKLIHOOD
2	3	5,10	13.149894, 13.149894, 6.573979	30.11103	30.67761
3	3	5,10,15	10.662833, 10.662833, 5.334286	19.52041	17.45173
4	4	5,10,15,20	13.961742, 13.961742, 13.961742, 4.653031	12.85738	7.299488
5	4	5,10,15,20,25	12.824150, 12.824150, 12.824150, 4.274399	9.076633	0.6102097

Table 7.7 Log likelihood values and rates of a Hypo-exponential distribution by MARK SOMMRENDER algorithm & MLE with shape unknown

NO OF PHASES	ESTIMATED NO OF PHASES	ACTUAL RATES	APPROXIMATED RATES OF PHASES	ACTUAL LOG LIKELIHOOD	ESTIMATED LOG LIKELIHOOD
2	3	5,10	13.024450, 13.024450, 6.520448	30.11103	30.66998
3	3	5,10,15	10.668783, 10.668783, 5.331363	19.52041	17.45173
4	4	5,10,15,20	13.961025, 13.961025, 13.961025, 4.653876	12.85738	7.299488
5	4	5,10,15,20,25	12.824541, 12.824541, 12.824541, 4.274678	9.076633	0.6102099

Comparing table 7.6 with table 7.7, it is found that, both these algorithm produce almost equal rates for different phases. Although for some cases, the Mark's algorithm gives slightly better solutions and it approaches the optimal solutions in less number of iterations, it has one drawback. For some samples it can't be applied as it is gives negative rate values for different phases[2]. On the contrary, the traditional estimation techniques can be applied to all data samples to find maximum likelihood estimation.

7.6 Conclusion

An exponential distribution is a very convenient choice to model service times of real world situations. But some processes don't respond to this distribution well. In service processes where a server provides service in phases sequentially, and time taken to provide service in each phase is exponentially distributed, then Erlang and Hypo-exponential distributions will best suit to model such situations. Furthermore, many service processes though consists of a single phase possess coefficient of variation less than one which contradicts the use of an exponential distribution. In such situations, Erlang and Hypo-exponential distribution can be used to study the mathematical behavior of the systems.

In this thesis we have used different estimation techniques to estimate parameters of Erlang and Hypo-exponential distribution. At first, Erlang distribution with known shape parameters is estimated using MOM and MLE where both of these two methods yield the same estimated value for rate parameter. Additionally it is observed that if the sample size increases by 10^x times, the variance of estimation of rate parameter decreases approximately

2 When Mark's algorithm is applied to the sample generated by set. seed (123) operation in R, we get mean time of the entire process E(T) to be 0.3060185 while E(B) was found to be 2.466863. This leads to generate negative mean time in process A i.e. E(A) was found to be -10.06495 which is not possible.

by 10^x times and if rate increases by 10^x, the variance of observed value is increased by a 10^{2x}. It is also found that as the number of phase increases, the accuracy of estimation of rate parameter increases and for some very high value of phase, estimated rates becomes equal to actual rate. This estimation technique can be used to model total time in machine reparation process, to handle traffic congestion in telecommunication networks, etc. In order to fit an Erlang distribution with unknown shape parameter, coefficient of variation is used to estimate the shape parameter first and then from this estimated shape, rate parameter is estimated using traditional method of estimations. Using this algorithm, optimum results were found in 76 out of 100 experiments and we got optimum result even for high values of number of phases of the service process. This algorithm can be used build model for situations where information in known only in terms of Mean and Standard deviation of the service processes with some real data. One important example is model compilation time of a computer, I/O operations of a computer.

Next we deal with Hypo-exponential distribution. To estimate rates of a Hypo-exponential distribution with two phases, MOM and MLE are used. But when the number of phases is more than two, to approximate the distribution, we have used the approximation algorithm by Markus Sommerender. To fit a Hypo-exponential distribution with unknown shape parameter, again coefficient of variation is used to estimate the shape parameters from which rate parameters are estimated. For a two phase Hypo-exponential distribution the original curve was found to be very well approximated using these two traditional estimation techniques MOM and MLE. For Hypo-exponential distribution with more than two phases, it was found that Mark Sommereder algorithm always provide equally good or better initial solution then that of MOM irrespective of the number of phases. Moreover Mark Sommereder algorithm approaches to the optimal solution in less or nearly equal number of iterations.

These estimations techniques can be applied to fit model for TMR systems, software rejuvenation systems etc. For Hypo-exponential distribution with both parameters unknown, it was found that estimated rates as obtained by both of these algorithms are nearly equal for different phases. For some cases, Mark's algorithm gives slightly better solutions and it approaches the optimal solutions in less number of iterations. These algorithms are best suited for situations like modelling hard disk performance time, time for I/O operations of a computer etc.

7.7 Future work

Estimating the change point in queueing models

Let $X_1, X_2, \ldots\ldots, X_n$ be a sequence of independent random variables with probability distribution functions $F_1, F_2, \ldots\ldots, F_n$ respectively. Then, in general, the change point problem is to test the following hypothesis:

$$H_0 : F_1 = F_2 = \ldots\ldots = F_n$$

Versus the alternative

$$H_1 : F_1 = F_2 = \ldots\ldots = F_k \neq F_{k+1} = F_{k+2} = \ldots\ldots = F_n .$$

If the distributions $F_1, F_2, \ldots\ldots, F_n$ belong to the common parametric family $F(\theta)$ where $\theta \in R^p$, then the change point problem is to test the hypothesis about the population parameter $\theta_i, i = 1,2,\ldots.,n$,

$$H_0: \theta_1 = \theta_2 = ... = \theta_n = \theta \text{ (Unknown)}$$

Versus the alternative

$$H_1: \theta_1 = \theta_2 = ... = \theta_k \neq \theta_{k+1} = \theta_{k+2} = ... = \theta_n$$

The problem of testing of parameter change has long been a core issue in statistical inferences. It originally started in the quality control context and then rapidly moved to various areas such as economics, finance, transportation systems, statistical quality control, inventory, production processes, communication networks and queueing, control problems, medicine. Since the paper of Page (1955), the problem has generated much interest and a vast amount of literatures have been published in various fields. For a general review, we refer to Csorgo and Horvath(1997), Chen and Gupta (2000) and the articles therein. The change point problem was first dealt in independent identically distributed samples but it became very popular in time series models since the structural change often occurs in economic models owing to a change of policy and critical social events. For relevant references in independent identically distributed samples and time series models, we refer to Hinkley (1971) , Brown, Durbin, and Evans (1975), Zacks (1983), Picard (1985), Csorgo and Horvath (1988).

The problem of detecting changes in the inter arrival time distribution of customers is a vital concern to various management personnel and system operators in the service planning and health care sectors. Discovering the positions of change point in the inter arrival time distribution would be of great help in planning services to customers. Suppose that for an observed queueing system, the predetermined value for the traffic intensity $\rho < 1$, is fixed (where ρ is the ratio of mean arrival rate to mean service rate). If a shift in traffic intensity from its prior value due to change occurs in the mean inter arrival time distribution, then to keep a similar value of traffic intensity, the number of servers or the service time of servers can be adjusted accordingly.

All of the above procedures require the application of statistical hypothesis testing. However, the problems of estimating change point of the inter arrival time distribution have not been discussed in the queueing literature yet. A comprehensive review for estimating the change points in a sequence of observation $x_1, x_2, ..., x_N$ with distribution functions $F_{1,}, F_2, ..., F_N$ was given by Krishnaiah and Miao (1988). Besides maximum likelihood and least square estimates, the Bayesian method is also a very useful technique for estimating parameters. Chernoff and Zacks(1964) described a Bayesian estimator for a change point in the mean of a sequence of normal random variables based on an arbitrarily specified prior probability distribution. Jain and Das (1993) considered a unified approach for estimating the change point of a sequence of observations and related testing procedures of a sequence of observations from the Bayesian point of view. Jain (2001) has considered a Bayesian approach for estimating at most one change point that occurs in a sequence of random variables whose density functions belong to an exponential family. However, one change point method with some modifications can be applicable for cases with more than one change point. In future one may investigate into the following problems:

To study the problem of estimating the change point in the arrival rate and service rate for certain known queueing models like $M/E_k/1$ and $M/H_k/1$ Hypo-exponential and study the properties of the estimators.

References

1. Acharya, S.K., Rodríguez-Sánchez, Sara Verónica and Villarreal-Rodríguez, César Emilio (2013), Maximum likelihood estimates in an M/M/c queue with heterogeneous servers, *International Journal of Mathematics in Operational Research*, 5(4):537 - 549.

2. Acharya, S.K. (1999), On normal approximation for maximum likelihood estimation from single server queues, *Queueing Systems*, 31(3-4): 207-216.

3. Adan, Ivo and Resing, Jacques (2001*), Queueing Theory*, Department of Mathematics and Computing Science, Eindhoven University of Technology, The Netherlands, February 14.

4. Azarbayejan, Dariush Ebrahimi (2011), *On the use of Phase-Type distributions for modeling general distributed events in manufacturing systems*, Master Thesis, Politecnico Di Milano.

5. Bailey, N.T.J (1952), Study of Queues and Appointment Systems in Out-Patient Departments with Special Reference to Waiting Times, *Journal Royal Statistic Society Series,* B. 14:185-199.

6. Bailey, N.T.J (1954), A Continuous Time Treatment of a Simple Queue Using Generating Functions, *Journal Royal Statistic Society Series,* B. 16:288-291.

7. Bakoban, R.A. (2012), Bayesian and non Bayesian estimation of Erlang distribution under progressive censoring, *International Journal of Research and Reviews in Applied Sciences*, 11 (3) June.

8. Basawa, I.V. and Prabhu, N.U. (1981), Estimation in Single Server Queues, *Naval Research Logistic,* 28:475 - 487.

9. Basawa, I.V. and Prabhu, N.U. (1988), Large Sample Inference From Single Server Queues, *Queueing Systems,* 3:288 - 304

10. Benes, V. E. (1957), A sufficient Set of Statistics for a Simple Telephone Exchange Model, *The Bell System Technical Journal,* 36:939-964.

11. Bhat, U.N. and Rao, S.S. (1987). Statistical Analysis of Queueing Systems, *Queueing Systems,* 1:217 – 247.

12. Bhat, U.N. and Fischer, M. J.(1976), Multichannel queueing systems with heterogenours classes of arrivals. Naval Res. Logist. Quart., Vol. 23, 271-282.

13. Bolch, Gunter, Greiner, Stefan, Hermann de Meer, Trivedi, Kishor S (2006), *Queueing Networks and Markov Chains: Modeling and Performance Evaluation with Computer Science Applications*, Wiley Series, 2nd Edition.

14. Cascetta, Ennio (2001), *Transportation Systems Engineering: Theory and Methods (Applied Optimization)*, Springer, 736.

15. Chernoff, H. and Zacks, S. (1964), Estimating the current mean of a normal distribution which is subjected to change in time. Ann. Math. Statist., Vol.35, 999-1018.

16. Chen, J., Gupta, A.K. (2000), Parametric Statistical Change Point Analysis . Boston: Borkauser.

17. Clerke, A. B. (1957), Maximum Likelihood Estimates in a Simple Queue, *The Annals of Mathematical Statistics*, 28:1036 -1040.

18. Conway, R.W. (1963), Some Tactical Problems in Digital Simulation, *Management Science,* 10:47-61.

19. Cox, D. R. (1955), The Analysis of Non Markovian Stochastic Processes by The Inclusion of Supplementary Variables, *Mathematical Proceedings of the Cambridge Philosophical Society*, 51:443-441.

20. Cox, D.R. (1965), Some Problems of Statistical Analysis Connected With Congestion, *Proceedings of the Symposium on Congestion Theory*, (eds. W.L. Smith and W.E. Wilkinson), University of North Carolina Chapel Hill, N.C.

21. Csorgo, M., Horvath, L. (1988), Nonparametric methods for change point problems. In: P.R. Krishaiah, C.D. Rao (Eds.) Handbook of Statistics (Vol. 7pp. 403-425). New York: Elsevier.

22. Csorgo, M. and Horvath, L. (1997), Limit Theorems in Change-Point Analysis. New York: Wiley.

23. Erlang, A. K. (1909), The Theory of Probabilities and Telephone Conversations, *Nyt Tidsskrift for Matematik,* B. 20:33-39.

24. Grinstead, Charles M and Snell Laurie J (1997) *Introduction to Probability*, American Mathematical Society.

25. Gross D and Harris C.M. (1985), *Fundamentals of Queueing Theory*, second edition. Wiley, New York.

26. Goldberg, D. A. and Whitt, W. (2007), The Last Departure Time from a $Mt/G/\infty$ Queue With a Terminating Arrival Process, *Queueing Systems: Theory and Applications,* 58(2):77 – 104.

27. Hinkley, D.V.(1971), Inference about the change-point from cumulative sums test. Biomertrika, 58,509-523.

28. Jain, S. and Das, K.(1993), Bayesian inference for detecting the shift of a parameter. Working paper #93-05, Department of Industrial Engineering, Univ. of Toronto, Toronto, Canada,.

29. Kendall, D.G. (1951), Some Problems in the Theory of Queues, *Journal Royal Statistical Society Series,* B, 13:151-185.

30. Kendall, D.G. (1964), Some Recent Work and Further Problems in the Theory of Queues, *Theory of Probability and Its Applications,* 9:1-15.

31. Khalid, B. Gunter, B. and Helmut, H. (2001), Practical Performance Modeling: Application of the MOSEL Language, *The Springer International Series in Engineering and Computer Science,* 588(XII):409.

32. Krishnaiah, P.R., Miao, B.Q.(1988), Review about estimation of change points. In: P.R./ Krishnaiah, C.P.Rao(Eds.) Handbook of Statistics (Vol. 7 pp.375-402)., New York: Elservier.

33. Lee, S., Lee, T.(2004), Cusum test for parameter change based on the maximum likelihood estimator,Sequential Analysis,23,239-256.

34. Lee, S., Na, O. (2204a), Test for parameter change in stochastic processes based on conditional least squares estimator. To appear in Journal of Multivariate Analysis.

35. Lee, S. Park, S. (2001). The cusum of squares test for scale change in infinite order moving average processes, Scandinavian Journal of Statistics, 28, 625-644.

36. Lindley, D. V. (1952), The Theory of Queues with a Single Server, *Mathematical Proceedings of the Cambridge Philosophical Society*, 48:277-289.

37. Little, J. D. C. (1961), A Proof For The Queueing Formula $= \lambda W$, *Operation Research*, 9:383-387.

38. Maode Ma (Ed.) (2009), *Current Research Progress of Optical Networks*, XIV:280, Springer, ISBN 978-1-4020-9889-5.

39. Molina, E. C. (1927), Application of the Theory of Probability to Telephone Trunking Problems, *The Bell System Technical Journal,* 6:461-494.

40. Neuts, M. F. (1965), The Busy Period of a Queue with Batch Service, *Operation Research*, 13:815-819.

41. Page, E.S.(1955), A test for change in a parameter occurring at an unknown point. Biometrika, 42, 523-527.

42. Pollackzec, F. (1934), Uber Das Warteproblem, *Mathematische Zeitschrift*, 38:492-537.

43. Rodriguez, Esparza, Luz Judith (2010), *Maximum likelihood estimation of phase-type distributions*, Kongens Lyngby, IMM-PHD-2010-245.

44. Saaty, T. L. (1961), *Elements of Queueing Theory and Applications*, New York: McGraw Hill.

45. Sommereder, Markus (2011), *Modelling of Queueing Systems with Markov Chains: An Introduction to Basic and Advanced Modelling Techniques*, Books on Demand GmbH, 288.

46. Thomas, Nigel (2011), *Computer Performance Engineering*, 8th European Performance Engineering Workshop, EPEW 2011, Borrowdale, UK, October 12-13, 2011. Proceedings.

47. Walck, Christian (2007), *Hand-book on statistical distributions for experimentalists*, Particle Physics Group, University of Stockholm.

48. Wolff, R. W. (1965), Problems of Statistical Inference for Birth and Death Queueing Models, *Operation Research*, 13:343-357.