2016

# Application of Complex Lifetime Models for Analysis of Product Reliability Data

Ruhi, Sabba

University of Rajshahi

# Application of Complex Lifetime Models for Analysis of Product Reliability Data

*A Dissertation*
*Submitted to the*
*University of Rajshahi in*
*Fulfillment of the Requirements*
*for the Degree of Doctor of Philosophy*

**By**

**SabbaRuhi**

**Department of Statistics**
**University of Rajshahi**
**Bangladesh**
**2016**

# Application of Complex Lifetime Models for Analysis of Product Reliability Data

*A Dissertation*
*Submitted to the*
*University of Rajshahi in*
*Fulfillment of the Requirements*
*for the Degree of Doctor of Philosophy*


**By**

**SabbaRuhi**


**Under the supervision of**


**Dr. Md. RezaulKarim**
Professor
Department of Statistics
University of Rajshahi
Bangladesh



**Department of Statistics**
**University of Rajshahi**
**Bangladesh**
**2016**

**Dedicated**

**To**

**MyFamily**

# CERTIFICATE

This is to certify that this dissertation entitled "Application of Complex Lifetime Models for Analysis of Product Reliability Data" is an authentic work carried out by SabbaRuhi, Research Fellow, Department of Statistics, University of Rajshahi, Bangladesh, for the award of the degree of Doctor of Philosophy (Ph.D.) in Statistics under my supervision and guidance.

To the best of my knowledge, the work of the dissertation in part or in full has not been submitted to any other university or institute for the award of any degree or diploma.

I also certify that I have read the draft and final version of the dissertation and approve it for submission.

Supervisor

Rajshahi
June 26, 2016

Dr. Md. RezaulKarim
Professor
Department of Statistics
University of Rajshahi
Bangladesh.

# DECLARATION

I do hereby declare that this dissertation entitled "Application of Complex Lifetime Models for Analysis of Product Reliability Data" submitted by me (SabbaRuhi) in the Department of Statistics, University of Rajshahi, Bangladesh for the award of the degree of Doctor of Philosophy (Ph.D.) is based on my research work carried out under the supervision of Professor Dr. Md. RezaulKarim, Department of Statistics, University of Rajshahi.

To the best of my knowledge, this work neither in part nor in full has been submitted to any other university or institute for the award of any degree.

Rajshahi
26/06/2016

SabbaRuhi
Ph.D. Research Fellow
Roll No.: 11314, Session: 2011-2012
Department of Statistics
University of Rajshahi
Bangladesh.

# ACKNOWLEDGEMENTS

I would like to thank those people who, during the several months in which this thesis was realized, provided me with all lot of assistance. First of all I would like to thank my PhD supervisor Dr. Md. RezaulKarim, Professor, Department of Statistics, University of Rajshahi for his constant guidance, advice, encouragement and every possible helps and the planning of this research.

Second, I would like to thank all of my respectable teachers of my department and all of my friends for their suggestions and academic helps during the course of the work.

Third and equally important in any regard I would like to thank my husband, kid, parents and family member for their supporting and sacrificing during my research.

Rajshahi

June, 2016 The Author

# Contents

# List of Figures

# List of Tables

# ABSTRACT

Proper data collection and analysis are very important for effective investigation of product reliability. Data is critical for building and selecting suitable statistical models and model provides new insights for improvements to maintenance and management operations in manufacturing industries. Regarding variations in product quality and reliability, component nonconformance and assembly error are two important problems which occur frequently in manufacturingindustries.In such situations, the complex lifetime models are required for analyzing product reliability data. This thesis proposes a general model for modeling the effects of quality variation. This model includes the mixture model and competing risk model as the special cases.

The thesis applies these models for analysis of three sets of product reliability data - Aircraft windshield failure data, Battery failure data and Hydraulic pump failure data. A set of competitive 2-fold and 3-fold mixture models are considered for modeling the data sets. The maximum likelihood estimation method via the Expectation Maximization (EM) algorithm is applied mainly for estimating the parameters of the models and reliability related quantities.

For the Aircraft windshield failure data, results indicate that the method of estimation with the EM algorithm procedure is betterthan theWeibull Probability Paper(WPP) plot procedure. For Battery failure data, based on the measures of lifetime quantities, it can be concluded thatdata without maintenance information provides approximately similar results with the data having maintenance information. According to the graphical representation and estimated values of different model selection criterions, we found that the 3-fold Weibull-Normal-Exponential mixture model can be selected as the best model for the Hydraulic pump failure data. The selected distribution for pumps with assembly errors failure mode is Normal and the distribution for pumps without assembly error failure mode is Weibull. According to the optimization of the proposed objective function, the 3-fold Weibull mixture model gives a bit larger optimal maintenance period, however the Weibull-Normal-Exponential model shows a reduction in the maintenance cost for the pump.

Simulation studies are conducted for investigation the performances of the proposed models and methods. The simulation results indicate that the proposed models and methods of estimation are applicable for analyzing 2-fold and 3-fold mixture models for censored product reliability data with incomplete information.

The results presented in this thesis would be useful for managerial implications in assessing and predicting the reliability and maintenance cost of the products.

**Keywords:** Product reliability, Data analysis, EM algorithm, Mixture model, Competing risk model, Simulation.

# Chapter 1
# Introduction and Background

## 1.1 Reliability and Reliability Function

Due to the increasing global marketplace and the resulting enhanced competition, now a days, manufactures need to develop new, higher technology products with improved quality, reliability, and productivity. With increasing product reliability, the responsibility of engineering organizations also increase to insure that reliability requirements are met. As a result, engineers and manufacturesbecome aware to calculate and reporton a product's reliability. One of the important part to improve the quality of a product is to improve it's reliability. High quality and high reliability products can have a strong competitive advantage in the market. So, manufacturers who have had little experience with life data analysis or applied statistics are work on improving their reliability. In today's technological world nearly everyone depends upon the continued functioning of a wide array of complex machinery and equipment for their everyday health, safety, mobility and economic welfare. Consumers expect their cars, computers, electrical appliances, lights, televisions, etc. to function day after day, year after year whenever they need them. They expect purchased products to be reliable and safe.

It takes a long time for a company to build up a reputation for reliability, and only a short time to drop their reputation after supplying a defective product. Continual assessment of new product reliability and ongoing control of the reliability of everything transported are critical necessities in today's competitive business arena. Reliability theory developed apart from the mainstream of probability and statistics, and was used primarily as a tool to help nineteenth century maritime and life insurance companies compute profitable rates to charge their customers.

Reliability of a product (or system) conveys information about the absence of failures.The purpose of reliability analysis is to specify the probability of success for a specified time. As suggested by Condra (1993), reliability can be defined as ‘quality over time’. Blischke and Murthy (2000) mentioned that, reliability of a product (system) conveys the concept of dependability, successful operation or performance, and the absence of failures. On the other hand, unreliability (or luck of reliability) conveys the opposite. Since the process deterioration leading to failure occurs in an uncertain manner, the concept of reliability requires a dynamic and probabilistic framework. According to Meeker and Escober (1998), the reliability is the probability that, a system, vehicle, machine, device, and so on will perform it’s intended function for a specified time period when operating under normal (or stated) environmental condition. The decisions, taken during the design, available knowledge of component reliability (often supplied by vendors), development and in the manufacturing of the product help to determine the reliability of the product, and depends on a number of factors, including manufacturing quality, operating environment (e.g., heat, humidity, dust and chemical solvents), usage intensity (frequency and severity), maintenance activities (e.g., frequency and depth of preventive maintenance), and operator’s skills (e.g., Murthy, 2010).

Reliability depends on operating conditions. In other words, a device is reliable under given conditions but can be unreliable under more severe conditions. Reliability usually varies with time and being calculate in a quantitative way. It is a numerical valuebetween zero and one. Warranty data provide a valuable source of information for assessing the reliability of an item in operation (called the ‘field reliability’) and to make decisions regarding the reliability improvements needed to control the consequences of unreliability.

Reliability is always associated with a given time. That is, the given percentage representing the probability of success is a function of time and is essentially paired with an associated time. For example, a specification may call for a 90% reliability at 100 hours of operation. This means that the product has a 90% probability of running for 100 hours without failure. It can also be interpreted as 90% of a population of such products will run for 100 hours, while the other 10% will have failed before 100 hours.

For a continuous random variable $T$, the reliability function or survival function gives the probability of surviving beyond time $t$ or a unit survives to time $t$.The reliability function denoted by $R(t)$ and defined as:

$$R(t) = \Pr(T > t) = \int_t^{\infty} f(t)dt\ ,\ \ 0 \le t \le \infty$$

Note that, $R(t)$ is monotone decreasing continuous function with $R(0)=1$ and $R(\infty) = \lim_{t\to\infty} R(t) = 0$. The reliability function is also known as the survival function, which is denote by $S(t)$.

The reliability function also can be expressed as:

$$R(t) = 1 - \Pr(T \le t) = 1 - F(t) = 1 - \int_0^t f(t)dt$$

Here $F(t)$ is the cumulative distribution function (cdf) of $T$ , gives the probability that, a component will fail before time $t$. Alternatively, $F(t)$ can be expressed as the proportion of units in the population that will fail before time $t$. $R(t)$ is the complement of the cdf of $T$. And $f(t)$ is the probability density function (pdf) of the random variable $T$. The pdf for a continuous random variable $T$ is defined as the derivative of $F(t)$ with respect to $t$. i.e., $f(t) = \dfrac{dF(t)}{dt}$ . The pdf can be used to represent relative frequency of failure times as a function of time. Although the pdf is less important than the other functions for applications in reliability, it is used extensively in the development of technical results.

We may also represent the reliability function as:

$$R(t) = \int_t^{\infty} f(t)dt$$

Reliability function also can be expressed as:

$$R(t) = \exp\left[ -\int_0^t h(t)dt \right]$$

Here $h(t)$ is the hazard function of $T$. The hazard function is also known as the hazard rate, the instantaneous failure rate function, etc. It is defined by:

$$h(t) = \lim_{\Delta t\to 0} \frac{\Pr\left(t < T \le t + \Delta t \middle| T > t\right)}{\Delta t} = \frac{f(t)}{1 - F(t)} = \frac{f(t)}{R(t)}$$

The hazard function expresses the propensity to fail in the next small interval of time, given survival to time *t*. That is, for small $\Delta t$

$$h(t) \times \Delta t \approx \Pr\left(t < T \le t + \Delta t \middle| T > t\right)$$

The hazard function can be interpreted as a failure rate in the following sense. If there is a large number of times [say, $n(t)$] in operating at time *t*, then $n(t) \times h(t)$ is approximately equal to the number of failures per unit time [or $h(t)$ is approximately equal to the number of failures per unit time per unit at risk]. The hazard function has units of fraction failed per unit time. Because of it's close relationship with failure process and maintenance strategies, some reliability engineers think of modeling failure time in terms of $h(t)$. During much of the useful life of a product, the hazard may be approximately constant, because, failures are caused by external shocks that occur at random. Late-life failures are due to wear out.

## 1.2 Reliability Data Sources

Because of rapid advances in manufacturing technology, consumers expect to purchase highly sophisticated, reliable and long lasting products. In recent years many manufacturers are collecting and analyzing field failure data to boost the reliability of their products and to improve goodwill and customer satisfaction (Blischke et al. 2011).Traditional reliability data have consisted of failure times for failed units and service times or censored times for censoring units. Laboratory life tests, field tracking studies, and warranty claim databases are the three main sources of reliability data. Accelerated life tests are often conducted to gain information in a timely manner for a component must last for years or even decades. Components are tested at high levels of cycling rate, voltage, temperature, stress, or another accelerating variable to get reliability information quickly. Then a physically-motivated model is used to extrapolate to usage conditions. See Nelson (2009) for more details on the statistical aspects of accelerated testing. Although laboratory life test data are often used to make decisions about the design of the product reliability, the 'real' reliability data comes from the field, often in the form of warranty returns or specially-designed field tracking studies. The collection of field data is typically costly and time consuming but careful field tracking provides a good quality of reliability data. For example, good field data are often available for medical devices and a company's fleet of assets.Warranty databases also deliver a rich source of reliability information. Warranty data provide at least a partial alternative to obtaining field data. With

warranty periods becoming longer, tracking products through this longer time frame provides much additional information that may be of significant value in the new product development process.

## 1.3 Some Common Difficulties with Reliability Data

There are three closely related problems that are typical with reliability data and not common with most other forms of statistical data. These are:

- Censoring (when the observation period ends, some observations does not meet fail i.e., some are survivors, are known as censored). More detail discussion on censored data can be found in Section 1.7.
- Lack of Failures (if there is too much censoring, even though a large number of units may be under observation, the information in the data is limited due to the lack of actual failures).
- Information missing in the database (sometimes some important information are not collected and reported in the reliability database).

These problems cause extensive practical difficulty when analyzing data for assessing productreliability. Typically, the solutions of these problems need to make additional assumptions and useof complicated statistical models and methods.

## 1.4 Distinguishing Features of Reliability Data

Reliability data have a number of special features that requires the use of special and complicated statistical methods. For example:

(i) Reliability data are typically censored (i.e., exact failure times are not known). The most common reason for censoring is the frequent need to analyze life test data before all units have failed. More generally, censoring arises when actual response values (e.g., failure times) cannot be observed for some or all units under study. Thus censored observations provide a bound or bounds on the actual failure times.

(ii) Most reliability data are modeled using distributions for positive random variables such as:Weibull, Exponential, Gamma and Lognormal. Applications ofNormal distribution in modeling reliability data are limited.

(iii) Inferences and predictions involving extrapolation are often required. For example, we might want to estimate the proportion of the units in the population that will fail after 1000 hours, based on a test that runs only 500 hours (extrapolation in time). Again we might want to estimate the time at which 10% observations of the population will fail at$50^{\circ}$Cbased on tests at$85^{\circ}$C (extrapolation in operating condition).

(iv) While making a reasonable decision by analyzing product failure data, it is often necessary to use past experiences or other scientific or engineering technologies. This information may take the form of a physical based model and/or the specification of one or more parameters (e.g., physical constants or materials properties) of such a model. This is also a form of extrapolation from the past to the present or future behavior of a system or product.

(v) Usually, the traditional parameters of a statistical model (e.g., mean and standard deviation) are not of primary interest. Instead, design engineers, reliability engineers, manufacturers and customers are mainly interested in specific measure of certain characteristics of product reliabilityof a failure time distribution (e.g., failure probabilities, quartiles of the life distribution, failure rates, mean time to failure).

(vi) Especially with censored data, it is difficult to meet analytical solutions, hence model fitting requires computer implementation of numerical methods, and often there is no exact theory for statistical inferences.

## 1.5 Lifetime Models

A variety class of statistical models have been developed and studied extensively in the analysis of the product failure data. The univariate continuous distributions can be broadly divided into two categories: simple distributions and complex distributions. Hence, lifetime models can be divided into two categories:

1. Standard Lifetime Models
2. Complex Lifetime Models

### 1.5.1 Standard Lifetime Models

In empirical modeling, the type of mathematical formulations needed for modeling is dictated by a preliminary analysis of data available. Some of the standard lifetime distributions that are used in analyzing product reliability data in this thesis are as follows:

- Two parameter Weibull distribution
- Exponential distribution
- Normal distribution
- Lognormal distribution

Details on the standard lifetime models will be discussed in section 2.2.

### 1.5.2 Complex Lifetime Models

Because of quality variation, the lifetimes of the productssometimes do not follow standard distributions and must be modeled by more complex model formulations. Complex models involve two or more standard lifetime models. Various types of complex models have been applied extensively in failure data analysis for manufactured products. Complex lifetime models that have used in this thesis are:

- Mixture Model
- Competing Risk Model
- Model for Effect of Quality Variation in Manufacturing (includes both Mixture & Competing risk model)

#### 1.5.2.1 Mixture Model

In the case of manufactured products, there are situations where some components of a product are produced over a period of time by collecting items from different merchants, using different raw materials, machines, and manpower and in different environmental conditions. The physical characteristics and the reliabilities of such components may be different, but sometimes it is difficult to distinguish them clearly.

In such situations, mixtures of distributions are often used in the analysis of reliability data for these components. In mixture distribution the lifetimes of the components are not independently and identically distributed (iid). A mixture model contains the combination of two or more standard models. Each model is mixed with a certain proportion. Details discussion on Mixture model is given in section 2.3.1.

#### 1.5.2.2 Competing Risk Model

In reliability analysis, the cause of failure of any component is called a failure mode or competing risk. Products may have more than one causes of failure. Competing risk model is applicable in the situation when there is information on failure of the components. In the analysis of lifetimes one is usually dealing with time to an event of interest like failure of a component or system, dead of an individual, or end to a subscription service. In most cases, there are several competing reasons (risks or modes of failure) that cause the event of interest. Say, for example, if we consider the failure of a bicycle, the competing risks for failure include: a flat tire, a broken chain, or the rupture of a brake cable.
The interest in competing risks is old but the formal development and application of the methodology to problems associated with engineering, survival analysis, and other applied areas are relatively new. Details on competing risk modelwill be discussed in section 2.3.2.

#### 1.5.2.3 Model for Effect of Quality Variation in Manufacturing

The quality variation of a product can occureither by the differences in manufacturing machines, raw materials, manpower, environment or by the various causes of error while assembling the components that built the product. In this thesis we have introduced a general complex lifetime model on the variation of the quality of the manufactured products. Instead of using mixture and competing risk model separately, this general model includes both of the mixture and competing risk models.We considered two main causes that effect on the variations of the quality of the manufactured products, they are:

**Assembly error:**Even a simple product consists of several components, that are assembled in production. Errors, caused while assembling the components to construct a product are known as assembly error.These types of error could be detected soon after putting the product in to operation. Failures causing from

assembly errors can be considered as a new mode of failure, which is different from other failure modes that one examines during the design process. For the products with assembly error, failure will occur sooner rather than later, and that the mean time to failure (MTTF) under this new failure mode is much smaller than the design MTTF. The type of assembly operation depends on the product. As for example: for an electronic product, one of the assembly operations is soldering. If the soldering is not done properly which is known as 'dry solder', then the connection between the components can break within a short period, leading to a premature failure.

**Component non-conformance:** Because of variations in quality or manufacturing process or lack of ability, some components cannot be able to perform their required intended functions. Due to the lacking's in their aptitude, when products do not meet the required design specifications, they are called nonconforming component. Items that are produced with such nonconforming components results in some items having lower reliability, means such nonconforming components will also tend to have an higher failure rate, shorter MTTF, etc. than the intended design value.The performance of nonconforming items is usually lower to the performance of conforming items. As a result, nonconforming items are less reliable than conforming items in terms of reliability measures such as MTTF.

Details on model for effect of quality variation in manufacturing will be found in section 2.3.3.

## 1.6 Failures and Failure Modes

Literally failure means lacking or falling short in something expected, attempted or desired.Also failure means the termination of the ability of an item to execute a required function.From an engineering point of view, it is useful to define failure in a different and broader sense. Failure is an event when machinery/equipment is not capable of perform scheduled operations to specification, means device cannot perform its function satisfactorily. Witherell (1994) elaborates as follows: 'It (failure) can be any incident or condition that causes an industrial plant, manufactured product, process, material or service to degrade or become unsuitable or unable to perform its intended function or purpose safely, reliably and cost-effectively'.

Product failures are depend on how reliable the product is and this, in turn, is influenced by several factors, some under the control of the manufacturer (decisions

made during the design and production stages) and others under the control of the consumer (operating environment, usage mode and intensity, and so forth).

The causes of failure of any product are known as failure mode. A failure mode is a description of a fault. It is sometimes referred to as fault mode. Failure modes are identified by studying the (performance) function of a product. A brief description of the different failure modes are as follows:

(i) *Intermittent failures*: Failures that last only for a short time. A good example of this is software faults that occur intermittently.

(ii) *Extended failures*: Failures that continue until some corrective action rectifies the failure. They can be divided in the following two categories:
   a. *Complete failure*: This results in total loss of function.
   b. *Partial failure*: This results in partial loss of function.

   Each of these can be further subdivided into the following:
   a. *Sudden failures*: Failures that occur without any warning.
   b. *Gradual failures*: Failures that occur with signals to warn of the occurrence of a failure.

A complete and sudden failure is called a *catastrophic failure*, and a gradual and partial failure is designated a *degraded failure*.

## 1.7 Censoring

The idea of censoring is associated with a sample. In case of censoring, we analyze a part of sample values. Censoring means that, in a group of individuals, a known number of observations is missing at either one end (in case of single censoring) or at both ends (in case of double censoring). Censoringoccurs when exact lifetimes are known for only a portion of the individuals under study; the lifetimes of the remainder individuals are known only to exceed certain values.

For example, in a life testing experiment, it may not be practical to continue analysis until all items under study have failed because of time limits and other restrictions on data collection. This limits the number of individuals to be considered for the inferential study which is known as censoring. If the experiment is terminated before

all items have failed, then for the items which are still unfailed at the time of termination only a lower bound on lifetime of these items are available. That is, the exact values of the unfailed observations are not known except, they are greater than or equal to a predetermined value. Suppose in a life test experiment, $n$ items may be placed on test, but a decision made to terminate the test after time $T$. Let $r$ ($r<n$) items have failed from starting to the time period $T$, so the exact lifetimes of $c=(n-r)$ items are not known but their initial time period is available. These $c$ items are called *censored* for $n$ items.

Censoring arises for a variety of reasons. There are several types of censoring, for example:

- Left, Right and Interval censoring
- Type-I, Type-II and Randomly censoring
- Single and Multiple censoring

## 1.7.1 Left, Right and Interval Censoring

A left censored value is one that is known only to be less than some specified value, e.g., $X<10$hrs. A right censored value is one that is known only to be more than some specified value e.g., $X>$ 5 hrs. A value is interval censored if it is stated as being within a specified interval, e.g. 5 hrs$<X<$10 hrs. Any observation of a continuous random variable could be considered interval censored, because its value is reported to a few decimal places. A value of $X$ = 25 hrs might be interpreted as 24.5 hrs $\leq X<$ 25.5 hrs. This sort of fine-scale interval censoring is usually ignored and the values are treated as exactly observed. When the intervals are large and the range of the data is small, e.g., 10 or fewer intervals over the range of the data, it is better to consider values as interval censored (Dixon and Newman, 1991).

## 1.7.2 Type-I, Type-II and Random Censoring

Sometimes experiments are run over a fixed time period in such a way that, lifetime of an item will be known exactly if it is less than some predetermined value. In such situation the data are said to be 'Type-I censoring' or 'time censoring'. A sample is Type-I censored when the censoring levels are known in advance. The number of

censored observations *c* (and hence the number of uncensored/failure observations *r*) is a random outcome, even if the total sample size*n*, is fixed. For example, let, 50 items may be placed on a test, but a decision made to terminate the test after 200 hours, either 50 items failed or not. Then this type of censoring is known as Type-I or Time censoring.

A sample is Type-II censored if the sample size *n* and number of censored observations *c* (and hence the number of uncensored/failure observations *r*) are fixed in advance. The censoring level(s) are random outcomes. Type-II censored samples most commonly arise in time-to-event studies that are planned to end after a specified number of failures, and Type-II censored samples are sometimes called failure-censored samples (Nelson, 1982, p.248). For example, let, 50 items may be placed on test, but a decision made to terminate the test when 30 items will fail. Then this type of censoring is known as Type-II or Failure censoring.

A sample is Randomlycensored when both the number of censored observations and the censoring levels are random outcomes. This type of censoring commonly arises in medical time-to-event studies. A subject who moves away from the study area before the event of interest occurs has a randomly censored value. The outcome for a subject can be modeled as a pair of random variables, (*X,C*), where *X* is the random time to the event and *C* is the random time until the subject moves away. *X* is an observed value if $X \leq C$ and right censored at *C* if $X > C$.

### 1.7.3 Single and Multiple Censoring

A sample is singly censored (e.g., singly left censored) if there is only one censoring level *T*. (Technically, left censored data are singly left censored only if all *r* uncensored observations are greater than or equal to *T*, and rightcensored data are singly right censored only if all *r* uncensored observations are less than or equal to *T* (Nelson, 1982, p.7); otherwise, the data are considered to be multiply censored.)

A sample is multiply censored if there are several censoring levels,$T_1, T_2, \ldots, T_n$, where $T_1 < T_2 < \cdots < T_n$. Multiple censoring commonly occurs with environmental data because detection limits can change over time (e.g., because of analytical improvements), or detection limits can depend on the type of sample or the background matrix. The distinction between single and multiple censoring is mostly

of historical interest. Some older statistical methods are specifically for singly censored samples. Most currently recommended methods can be used with either singly or multiply censored samples, but the implementation is often easier with one censoring level.

## 1.8 Maintenance

Every object (product, plant or infrastructure) is designed and built to some performance requirement and included of several components (or elements). The performance of the object depends on the performance of its components that may degrade with age and/or usage intensity,whicheffecton the performance of the object. A component is deemed to have failed when its performance falls below a pre specified level. The failure of an object is due to the failure of one or more of its components. We expect the object/product to perform in a same way and same intensity for long time without any disturbance. To ensure the product to be reliable and safe we need to maintain it in a proper manner after a specific time period. Maintenance is a combination of technical, administrative, and managerial activities carried out during the lifecycle of an object. Maintenance actions are of two types:'Preventive maintenance' (PM) actions to control the degradation processes and reduce the likelihood of failure of an item (component or object) and 'Corrective maintenance' (CM) actions to restore a failed item to a specified operational state, involving either repair or replacement of the item.

### 1.8.1 Maintenance Outsourcing

Usually, maintenance was done inhouse by the owner of the object and also dealt with the data management issues. Over the last few decades, there has been an increasing tendency in the outsourcing of maintenance where some or all the maintenance is carried out by an external service agent under a maintenance service contract (MSC). Effective maintenance requires proper data management – collecting, analyzing and using appropriate models for making decisions. In maintenance outsourcing data is needed for different purposes, e.g. contract formulation, monitoring the quality of maintenance provided by the service agent, improvements to maintenance, etc.

## 1.8.2 Maintenance Service Contract

A maintenance service contract (MSC) is a legal manuscript that is binding on both parties (the business or customer and the service agent) and it needs to deal with technical, economic and legal issues.

**Classification of contracts**

•**Standard contracts:** Mainly in the form of extended warranties for consumer products and service contracts for commercial and industrial products (e.g., lifts in buildings). The terms of the contract are determined by the service provider taking into account the marketing aspects.

•**Customized contracts:** For complex plants and infrastructures where the contract is often initiated by the owner and the terms decided jointly.

**Technical Issues**

- Types of maintenance tasks (PM and/or CM) to be carried out
- The details of the tasks to be carried out
- Types of the component/piece parts used for maintenance (standard part, Part manufacturing approved part, etc.)
- Turnaround time
- Documentations

**Economical/Financial Issues**

- Payments
- Penalties
- Risks
- Insurance

**Legal Issues**

- Terms of contract
- Contract duration
- Dispute resolution
- Guaranty/warranty
- Force major issues

Unless the contract is written properly and relevant data (relating to the object and collected by the service agent) are analyzed properly by the customer the long-term costs and risks will increase.

## 1.9 Optimum Maintenance Cost

Every engineered object (product, plant or infrastructure) needs preventive and corrective maintenance. The cost of maintenance can vary from 5% to 30% (Campbell, 1995) of the operating budget depending on the industry sector. This implies that businesses need to manage maintenance effectively to ensure minimum costs. This requires proper data management to assist in building models for effective decision making.Obtaining the solution to the problem involves building a model and deciding on the optimal age for PM action requires an objective function. The objective function is the asymptotic expected cost per unit time. Note that every time instant an exchanged product is put into operation can be viewed as a renewal point for a renewal process characterizing the replacements of products over time. The time between two successive renewal points defines a cycle. The asymptotic expected cost per unit time can be obtained as the ratio of the expected cycle cost (ECC) and the expected cycle length (ECL).

The time to failure for a product, $X$, is a random variable with distribution function $F(x)$. A PM action results if $X \geq T$ in which case the cycle length is $T$ with probability $R(T)$. A CM action results when $X < T$ and the cycle length is $X$. As a result ECL is given by

$$ECL = \int_0^t tf(t)dt + TR(T) = \int_0^T R(t)dt \tag{1.1}$$

Let us consider the following additional notations:

$C_f$ : Average cost of a CM replacement

$C_p$ : Average cost of a PM replacement

$C_n$ : Sale price for new item (Given by the owner)

$C_r$ : Cost (charged by the service agent) for reconditioning an item under CM or PM action

$\xi$ : Additional cost (due to downtime, loss in revenue, etc.) resulting from CM action

Now the value of ECC consists of the cost of preventive maintenance in addition to the cost of corrective maintenance, which is given by

$$ECC = C_f F(T) + C_p R(T) \tag{1.2}$$

From (1.1) and (1.2) we have the asymptotic average cost per unit time given by

$$J(T; F(.)) = \frac{C_f F(T) + C_p R(T)}{\int_0^T R(t)dt} \tag{1.3}$$

Let we denote the optimal of$T$ by$T^*$, this is the value that yields a minimum for $J(T; F(.))$.The optimal $T$ depends on the average cost of each CM and PM.

A maintenance action involves replacement by a new item or a reconditioned item with probabilities $q$ and $(1-q)$ respectively. As a result, the average cost of a PM action is

$$C_p = qC_n + (1-q)C_r \tag{1.4}$$

And of a CM action is

$$C_f = C_p + \xi \tag{1.5}$$

The optimal $T^*$ is obtained using (1.3) with thecdf,$F(t)$ and the optimal expected cost per unit time is given by $J(T^*; F(.))$.Here we can see that, the optimal $T^*$ depend on the additional cost $\xi$. The optimal $T^*$ and optimal expected cost per unit time $J(T^*)$ on various values of$\xi$ for the different models can be estimated.

## 1.10 Mean Time to Failure

Mean time to failure (MTTF) describes the expected time to failure for a non-repairable product. That is, MTTF is the average time that an item will function before it failswith the modeling assumption that the failed system is not repaired. It is the mean lifetime of the item.With censored data, the arithmetic average of the data does not provide a good measure of the center because at least some of the failure times are unknown. The MTTF is an estimate of the theoretical center of the distribution that considers censored observations.The MTTF can be used in several ways; for example:

- To determine whether a redesigned system is better than the previous system in demonstration test plans.
- As a measure of the center of the distribution when the distribution fits the data satisfactorily.

Suppose we tested 3 identical systems starting from time 0 until all of them failed. The first system failed at 10 hours, the second failed at 12 hours and the third failed at 13 hours. The MTTF is the average of the three failure times, which is 11.6667 hours. If these three failures are random samples from a population and the failure times of this population follow a distribution with a probability density function (pdf) of $f(t)$, then the population MTTF (denoted by $\mu$ or $E(T)$) can be mathematically calculated by:

$$MTTF = E(T) = \int_0^\infty tf(t)dt = \int_0^\infty [1 - F(t)]dt = \int_0^\infty R(t)dt$$

## 1.11 Fractile

Fractiles are numbers that divide an ordered data set into equal parts. Fractiles are of various types:

- Quartiles divide a data set into 4 equal parts
- Deciles divide a data set into 10 equal parts
- Percentiles divide a data set into 100 equal parts

The $p$-fractile of a sample is denoted by the value $t_p$ such that at least a proportion $p$ of the sample lies at or below $t_p$ and at least a proportion $1–p$ lies at or above $t_p$.Where $p$ takes the value from 0 to 1, that is $0 < p < 1$. If the continuous cdf $F(t)$ is strictly increasing, then there is a unique value $t_p$ that satisfies $F(t_p) = p$, and the estimating equation for $t_p$ can be expressed as

$$t_p = F^{-1}(p)$$

where$F^{-1}(p)$ denotes the inverse function. For example, if$t$ denotes the lifetime of an item, $t_p$ is the time at which $100p\%$ of the units in the product population will have failed. $t_p$is also known as $B100_p$(e.g., $t_{.10}$ is also known as B10). The median is

equal to item, $t_{0.5}$. The fractile for $p$= 0.25 is called '1$^{st}$quartile' or '25$^{th}$ percentile' and $p$ = 0.75 is called '3$^{rd}$quartile' or '75$^{th}$ percentile'.

## 1.12 Review of the Literature

Mathematical models have been used in solving real-world problems from many different disciplines. This requires building a suitable mathematical model.Over the last several years, a number of new models have been proposed that are either derived from, or some way related to the distributions like Weibull, Exponential, Normal etc. Meeker and Escober (1998b), discussed various models and methods for estimating product reliability.Titterington et al. (1985) mention that mixture distribution models have been used for a long time and give a comprehensive reference list of the applications of such models. The earliest reference (dating to 1886) involves a normal mixture. Earliest Weibull mixture models can be traced to the late 1950s [see, Mendenhall and Hader (1958) and Kao (1959)]. Since then the literature on Weibull mixture models has grown at an increasing pace and with many different applications of the model. Jiang and Kececioglu (1992) proposed the principle of the maximum likelihood estimate through the EM algorithm, applied to both postmortem and non-postmortem censored data, grouped, as well as suspended data. They also indicated that some of the log-likelihood functions of the mixed-Weibull distributions have multiple local maxima; therefore, the algorithm should start at several initial guesses of the parameters set. The searching of the largest local maximum can stop when a good fit has been found. They recommended the graphical examination method for this purpose.Tarum (1999) presented a method of bathtub equation that allow for analysis and prediction of failure rates where infant mortality, chance, and wear out failures are combined. The bathtub curve help to model mixed failure modes. They used rank regression and maximum likelihood estimation method to fit the curve. They proposed to use either a competing risk mixture or a competing risk, when there is an apparent mixture of two failure modes.

Marín, Bernal and Wiper (2003) applied the Bayesian method using birth-death MCMC algorithm, to fit the Weibull mixture model for unknown number of components to heterogeneous, possibly right censored survival data. They proposed it as appropriate models for the analysis of clinical trial data with several sub-populations showing different behavior and for the observed data consist of both complete and right censored lifetimes. Jiang and Murthy (2003) presented an *n*-fold

Weibull competing risk model and applied the WPP (Weibull probability paper) plot to estimate the model parameters. They also presented different possible shapes for the density and failure rate functions.

.According to Murthy, Xie, and Jiang (2004), many standard probability distributions have been used as models to model data exhibiting significant variability.A variety class of statistical models have been developed and studied extensively in the analysis of the product failure data. Literatures on mixture models including the graphical method of estimation based on the WPP plot can be found in Murthy et al. (2004).

Bucar, Nagode and Fajdiga (2004) discussed that the reliability of an arbitrary system can be approximated well by a finite Weibull mixture with positive component weights only, without knowing the structure of the system, on condition that the unknown parameters of the mixture can be estimated. It can be concluded that the suggested Weibull mixture with an arbitrary but finite number of components is suitable for lifetime data approximation. They considered the data modeling and parameter estimations when a set of grouped complete data is available. All described methods for parameter estimation of the Weibull mixture distribution are applied in five examples, four simulated and one from literature. They presented four different numerical methods for the estimation of unknown parameters of a Weibull mixture: EM algorithm, Alternative algorithm, Minimax algorithm and Multivariate regression and suggested that the EM algorithm is the most suitable method for determining the Weibull mixture distribution of failure times. Though, the EM algorithm has some weaknesses as well. The result and convergence of the iterative procedure for an estimation of unknown parameters of the mixture distribution of failure times frequently depends on the initial conditions of iteration. It turned out that the application of multivariate regression method for estimation of unknown mixture weights provides a useful model for representing failure data if the specific form of the mixture model is determined in advance.

Bertholon et al. (2004) proposed a simple competing risk distribution corresponds to the minimum between exponential and Weibull distributions in analysis of both accidental and aging failure lifetime data. The distribution parameters are estimated through MLE and Bayesian inference. They observed that the competing risk model fits both the real life and simulated data well, than that of the Exponential and Weibull models.

Park and Kulasekera (2004) developed maximum likelihood estimators for the competing risks analysis of data from multiple groups, with both failure time and failure cause censorings under multivariate exponential distributions. Kundu and Sarhan (2006) presented competing risks among several groups to analyze incomplete data. They extended the work of Park & Kulasekera (2004); considered the same latent failure times model formulation assuming Weibull distribution failure times, rather than the exponential distribution, and it is assumed that the latent failure times are independent Weibull random variables with the same shape parameter within a particular group, but different scale parameters. It is observed that, instead of the exponential distribution, the Weibull distribution may be used in this case. Asymptotic distributions of the maximum likelihood estimators of the different parameters are obtained and asymptotic confidence intervals are also proposed.

Li et al.(2007) proposed a hierarchical mixture of software reliability models (HMSRM) for software reliability prediction to develop general prediction models in current software reliability research. They illustrated that their approach performed quite well in the later stages of software development, and better than single classical software reliability models. They also indicated that the method can automatically select the most appropriate lower-level model for the data and performances are well in prediction.Elfaki et al. (2007) applied EM Algorithm on Cox's model with Weibull distribution and Cox's with exponential distribution and found that with a large sample size based on expectation maximization (EM) algorithm, both models give similar results and the modification of the two models showed better results compared with Crowder et al., especially for the second causes of failure.

Li and Lin (2009) studied a semi-parametric mixture model for the two-sample problem with right censored data using simulation and proposed EM algorithm for the semi-parametric maximum likelihood estimates of the parametric and nonparametric components of the model that provided a useful alternative to the Cox (1972) proportional hazards model for the comparison of treatments based on right censored survival data.Alwasel (2009) presented competing risks model for incomplete and censored data when the causes of failures follow modified Weibull distributions. Two cases were tested: The first was, when the causes of failure follow exponential against modified Weibull and the second was, when the causes of failure follow Weibull against modified Weibull distribution. Method of maximum likelihood was applied to obtain the point estimations and asymptotic confidence intervals of the parameters.

Castet and Saleh (2010) suggested to use mixture of Weibull distributions, to analyze the satellite reliability data with censored observation. They compared the results obtained from mixture model with a single Weibull distribution and found that the mixture Weibull distribution provides significant accuracy in capturing all the failure trends in the failure data for nonparametric satellite reliability. The model parameters were estimated by MLE method.

Erişoğlu et al. (2011) proposed a mixture of two different distributions such as Exponential-Gamma, Exponential-Weibull and Gamma-Weibull to model heterogeneous survival data and indicated that, mixture of the different distributions are appropriate for the heterogeneous survival times.Nielsen (2011) discussed three different estimation methods: maximum likelihood estimation (MLE), method of moments estimation (MME) and median rank regression (MRR). Overall, he found that MLE is the best among these three. Both MLE and method of moments required iterative algorithms, which was not necessary for the median rank regression method. Again for any true values of the parameters, all the three estimators provide approximately the same accuracy.

Lee and Scott (2012) presented expectation–maximization (EM) algorithms for fitting multivariate Gaussian mixture models to data that are truncated, censored or truncated and censored. They indicated that these two types of incomplete measurements are naturally handled together through their relation to the multivariate truncated Gaussian distribution.Bordes and Chauveau (2012) discussed several iterative method based on EM and stochastic EM methodology, that allow to estimate parametric or semi parametric mixture model for randomly right censored lifetime data, provided they are identifiable. They consider different levels of completion for the (incomplete) observed data, and provide genuine or EM-like algorithms for several situations. The censored semi parametric situations, a stochastic step is the only practical solution allowing computation of nonparametric estimates of the unknown survival function. The effectiveness of the new proposed algorithms is demonstrated in simulation studies.Assuming dependence for failure modes, Ancha and Yincai (2012) used copula as the dependence link function to assess competing risk models in accelerated life testing. They used the simulated data and found that the usual independence assumption would have a crucial effect on the reliability assessment for constant stress ALTs when the failure modes of the series system were dependent.

Razali and Al-Wakeel (2013) used two-fold and three-fold mixture of two and three parameter Weibull distribution to analyze failure time data. They found that, for the two-fold Weibull mixture distribution, the values of $R^2$ and $R^2_{Adj}$ were high while the values of SSE and MSE were low but for three-fold Weibull mixture distribution the values of $R^2$ and $R^2_{Adj}$ decrease. Hence they deduced that as the number of parameters increase, the accuracy of the model decrease. They also indicates that as two-fold and three-fold mixture Weibull distributions has uni-modal, bi-modal and tri-modal shapes, they provide better fits than that of the single Weibull distribution with uni-modal shape.Barabadi (2013) used the WPP plot in order to select the appropriate Weibull distribution for a historical data of power transformer. Through a flow chart,they provided a guideline for the selection of an appropriate model from the Weibull family of distributions for failure data.Zhang and Dwight (2013) used the graphical approach, WPP plot to choose an optimal model for a given data set and to model the data. Their proposed model selection procedure was based on the shapes of the fitting plots. They discussed the characteristics of Weibull-related models with more than three parameters including sectional models involving two or three Weibull distributions, competing risk model and Weibull mixture model. According to Zhang and Dwight (2013), the initial estimate of model parameters through the graphical approach yields certain degree of subjectivity but they think other more accurate statistical methods such as: MLE, Least Squares estimation, etc. are necessary to be applied. While estimating the parameters of the sectional models, they suggested applying the recursive method to solve a set of equations that involve several parameters.He et al. (2013) proposed a new mixed Weibull distribution model to evaluate the overall reliability of paper-oil insulation. The model represents the relationship between the overall reliability and states of eight characteristic parameters of paper-oil insulation. They presented their model as an effective one to evaluate the reliability of paper-oil insulation and also the reliability of insulation of power transformer in service.Noor and Aslam (2013) presented the Bayesian inference for the two Inverse Weibull mixture models for modeling the complex failure data set for type-I censoring. They considered two cases: when the shape parameter is known and when all parameters are unknown. Bayesian analysis was carried out using informative (Gamma) and non-informative (Jeffrey's) priors.Sarhan, Alameri and Al-Wasel (2013) discussed the competing risks model with generalized Weibull distribution for incomplete and censored data. Method of maximum likelihood was used to investigate both the point and asymptotic confidence interval estimators. They studied hypothesis tests for a real data from Lawless (2003) and

found that the generalized Weibull distribution fits the data better that the exponential, generalized exponential and Weibull distributions.

Ateya (2014) analyzed a real data set by using a finite mixture of two Generalized Exponential (GE) distributions and a 2-fold Weibull mixture distribution and found that the GE mixture model fits the data better. MLE of the model parameters were estimated using EM algorithm based on right censored failure observations and indicated that the results are specialized to type-I and type-II censored samples.Ateya and Alharthi (2014) used EM algorithm to find out the MLEs of the model parameters of a finite mixture of modified Weibull distributions for type-I and type-II censored data. They applied a two-fold modified Weibull mixture and a two-fold Weibull mixture distribution to analyze a real life data set to emphasize that the modified Weibull mixture model fits the data better than the other mixture model.Benaicha and Chaker (2014) presented the maximum likelihood algorithm for a two-fold Weibull mixture distribution. The impact of traditional Weibull distribution was compared with 2-fold Weibull mixture distribution for historical failure time data of power transformers of National Society for Electricity and Gas (SONELGAZ) in Western Algeria. Zhang, Hua and Xu (2014) proposed a mixture Weibull proportional hazard model (MWPHM) to predict the failure of a mechanical system with multiple failure modes. They selected the MWPHM over the Weibull proportional hazard model (WPHM), because, MWPHM includes all the contribution of different failure modes and also provides more detailed information on the lifetime distribution. They used a set of simulated data and estimated the model parameters by combining historical lifetime and monitoring data of all failure modes.Considering the effects of failure modes on competing risks model, Yáñez, Escobar and González (2014) introduced two different competing risks models: (a) a model with independent risk (b) a model derived from a bivariate Weibull with dependence. They found that the characteristics of these two models are very similar. They considered two and more than two risks in their work.

Elmahdy (2015) compared the fitted reliability functions of the 3-parameter Weibull, competing risk and 2-fold Weibull (2 parameter) mixture models and indicated as an efficient approach for moderate and large samples with a higher censored and few exact failure times. They proposed to apply this approach for complete, censored, grouped and ungrouped samples. The parameters estimated by both graphical and numerical methods, such as WPP plot, MLE, Bayes estimators, non-linear Benard's median rank regression. The parameters of finite Weibull mixture distributions were

estimated through EM algorithm.Ruhi (2015) applied 2-fold mixture models and estimated the parameters by MLE. Ruhi, Sarker and Karim (2015) applied 2-fold Weibull-Weibull mixture model for analyzing failure data and used the EM algorithm to estimate the model parameters. They indicated the maximum likelihood estimate procedure performs well than the graphical procedure given in Murthy et al. (2004).Feizjavadian and Hashemi (2015) used a Marshall–Olkin bivariate Weibull distribution to investigate dependent competing risks for progressively hybrid censoring (can be produced by combining Type-I and Type-II censoring) condition. Maximum likelihood and approximated maximum likelihood estimators are applied to estimate the unknown parameters.Balakrishnan, So and Ling (2015) analyzed the one-shot devices with two competing risks model (corresponding to the failure of each device) assuming that the lifetime distribution is exponential and that there are no masked causes of failure. EM algorithm had been developed for the estimation of model parameters. While comparing the proposed method with the Fisher-scoring method, they found the robustness of EM algorithm over the Fisher-scoring method.Balakrishnan et al. (2015) extended the work of Balakrishnan and Ling (2013) by introducing the analysis of a one-shot device testing using competing risks model under an ALT setting. They confined their attention to the case of two competing risks corresponding to the failure of each device, assuming that the lifetime distribution is Weibull with no masked causes of failure in the data set and found their model more flexible when compared to the exponential distribution. The EM algorithm was developed to estimate the model parameters. They noticed the robustness of EM algorithm to the choice of initial values than that of the Fisher scoring method.

El-kelany (2015) considered a competing risks model with three independent causes of failure for complete and incomplete observations under Type-I censoring. They assumed the competing risk with two parameter Weibull, exponentiated Weibull and Rayleigh distribution. The MLEs of different parameters with different sampling schemes and their properties are studied under these assumptions.Ünal et al. (2015) discussed statistical inference for Weibull distribution based on competing risks data under progressive Type-I group censoring. The maximum likelihood procedure is used to get point estimates and asymptotic confidence intervals for unknown parameters.

Feroze (2016) was the first, who developed the Bayesian inference of inverse Weibull mixture distribution based on doubly type II censored data. They used both simulated

and real life data for this purpose.Iskandar and Gondokaryono (2016) investigated the application of the Bayesian estimation analyses with Weibull distribution model to competing risk systems. They used the simulation data and indicated that, for perfect information on the prior distribution and with smaller sample sizes, the estimation methods of the Bayesian analyses are better than those of the maximum likelihood. They also showed the robustness of the Bayesian analysis within the range between the true value and the maximum likelihood estimated value lines.Iskandar (2016) expressed the basic concepts that constitute Bayesian approach with Gamma distribution, to analyze the competing risks models in reliability. There are some limitations with their models: causes of failure are independent, only the scale parameter is a random variable, and prior distribution that used is uniform.

Bedford (2005) discussed the problems with dependent competing risk model, specifically in the reliability context. They expressed the problems of identifying the joint distribution or marginal distributions without making nontestable assumptions. They discussed the way in which the assumption of independence usually gives an optimistic view of failure behavior, possible models for maintenance, and generalizations of the competing risk problem to nonrenewal system.

## 1.13 Purpose of the Research

The purpose of this research is to apply some complex lifetime models for analysis of product reliability data in order to select suitable lifetime models and to assess and predict the reliability and maintenance cost of the products which are useful for managerial implications in manufacturing industries.

## 1.14 Objectives

The specific objectives of this research are to:

- Compare graphical (Weibull Probability Paper(WPP) plot) and statistical parametric (Maximum Likelihood Estimation (MLE)) methods for analyzing product reliability data.
- Investigate the effects on the estimated reliability if the information on maintenance action is unknown in the database.

- Find the suitable models for a product reliability data set and to trace the possible hidden sub-populations and their distributions.
- Find the suitable distributions for different failure modes.
- Estimate the optimal maintenance period and cost of a product.
- Investigate the overall performances of the proposed models and methods by simulation study.

## 1.15 Research Questions

In order to fulfill the above-stated objectives, the following research questions (RQ) have been raised:

- **RQ 1.** Which of the methods, graphical (WPP plot) or statistical parametric (MLE), perform well for analyzing product reliability data?
- **RQ 2.** What would be the effects on the estimated reliability if the information on maintenance action is unknown for all observations in the database?
- **RQ 3.** What are the suitable 3-fold lifetime models of hydraulic pump? What are the possible three hidden sub-populations and their distributions for this pump?
- **RQ 4.**What are the suitable distributions for the pumps with assembly errors and the pumps without assembly errors?
- **RQ 5.** What would be the effect on optimal maintenance policy according to the selected models?
- **RQ 6.** What are the overall performances of the proposed models and methods?

## 1.16 Layout of the Study

The chapter wise summary of this thesis is given below:

The present chapter provides some basic concept of reliability, review of the literature and summary of the rest of the chapters of this thesis.

In chapter 2, we introduce the basic concepts and preliminaries of some widely used lifetime models for product reliability data, including Weibull, Exponential, Normal, Lognormal, Mixture models, Competing risk models and Model for effect of quality variation. The important characteristics of these models on reliability are also given.

In chapter 3, we describe various statistical methods for reliability data analysis. This includes both nonparametric and parametric estimation procedure, and different model selection methods.

Chapter 4 introduces three sets of real data (Aircraft windshield failure data, Battery failure data and Hydraulic pump failure data) that are analyzed in this thesis. This chapter also discusses about field reliability data and it's limitations.

Chapter 5 presents, the results of data analysis.

Chapter 6 contains the results of simulation methods for 2-fold and 3-fold mixture models, which is used to investigate the property of the population.

Finally Chapter 7 presents the conclusion of this thesis. This includes the main contributions of the thesis and future research.

R codes of 2-fold and 3-fold mixture models are presented in the Appendix portion.

# Chapter 2
# Lifetime Models for Product Reliability Data

## 2.1 Introduction

In the real world, problems arise in many different circumstances. Models have been playing an important role in problem solving for a long time, beyond the recorded history of the human race. Many different kinds of models have been used,that include: physical (full of scaled) models, pictorial models, analog models, descriptive models, symbolic models and mathematical models. The applications of mathematical models are relatively new (roughly the last 500 years). Initially, mathematical models were used for solving problems from the physical science (e.g., predicting motion of planets, timing of high and low tides), but over the last few hundred years, mathematical models have been used broadly in solving problems from biological and social sciences. This is hardly any discipline where mathematical models have not been used for solving problems.

Different class of statistical models have been established and studied extensively in the analysis of the product reliability data. In this thesis, the lifetime models are divided into two categories:

1. Standard Lifetime Models
2. Complex Lifetime Models

## 2.2 Standard Lifetime Models

In empirical modeling, the type of mathematical formulations needed for modeling is dictated by a preliminary analysis of data available. If high degree of variability exist in the data set, then appropriate models are needed that can capture this variability. This requires probabilistic and stochastic models to model a given data set. Some of

the standard lifetime distributions that are used in analyzing product failure data in this thesis are as follows:

- Two parameter Weibull distribution.
- Exponential distribution.
- Normal distribution.
- Lognormal distribution.

### 2.2.1 Two Parameter Weibull Distribution

When the random variable$T$ has a Weibull distribution with two parameters, we indicate this by $T \sim \text{Weibull}(\beta,\eta)$. We shall refer to this as the 'Standard Weibull Model' with $\eta(>0)$ and $\beta(>0)$ being the scale and the shape parameters, respectively.

The probability density function, $f(t)$ for twoparameter Weibull model is given by

$$f(t;\beta,\eta)=\frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1}\exp\left[-\left(\frac{t}{\eta}\right)^{\beta}\right] \quad t \geq 0\ ,\ \eta,\beta>0 \qquad (2.1)$$

The shape of the density function depends on the model parameters.

The cdf of two parameter Weibull distribution is

$$F(t;\beta,\eta)=\int_0^t f(t;\beta,\eta)dt=1-\exp\left[-\left(\frac{t}{\eta}\right)^{\beta}\right] \qquad ;t\geq 0 \qquad (2.2)$$

The reliability function $R(t)$, of a lifetime distribution is defined as the probability of survival beyond age $t$, and is given by

$$R(t)=1-F(t)=\exp\left[-\left(\frac{t}{\eta}\right)^{\beta}\right] \qquad ;t\geq 0 \qquad (2.3)$$

The hazard function $h(t)$ is given by :

$$h(t)=\frac{f(t)}{R(t)}==\frac{\frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1}\exp\left[-\left(\frac{t}{\eta}\right)^{\beta}\right]}{\exp\left[-\left(\frac{t}{\eta}\right)^{\beta}\right]}=\frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1} \quad ;t\geq 0 \tag{2.4}$$

The $p$-th percentile of two parameter Weibull distribution is given by

$$F(t_p)=p$$

$$1-\exp\left[-\left(\frac{t_p}{\eta}\right)^{\beta}\right]=p$$

$$\exp\left[-\left(\frac{t_p}{\eta}\right)^{\beta}\right]=1-p$$

$$\left(\frac{t_p}{\eta}\right)^{\beta}=-\ln[1-p]$$

$$t_p=\eta[-\ln(1-p)]^{1/\beta} \tag{2.5}$$

The mean and variance of this distribution are respectively,

$$\text{Mean}=E(t)=\int tf(t)dt=\eta\sqrt{\frac{1}{\beta}+1}$$

and

$$\text{Variance}=\eta^2\left[\sqrt{\frac{2}{\beta}+1}-\left\{\sqrt{\frac{1}{\beta}+1}\right\}^2\right]$$

The median is given by

$$t_{median}=\eta(\ln 2)^{1/\beta}$$

When $\beta>1$, there is a single mode, $t_{\text{mod}e}$ ,and it is given by

$$t_{\text{mod}e}=\eta\left[1-\frac{1}{\beta}\right]^{1/\beta}$$

When $\beta\leq 1$, the mode is at $t=0$ .

Weibull distribution can be used to model the minimum of a large number of independent positive random variables from a certain class of distributions. This distribution also can be used to model failuretime data with a decreasing or an increasing hazard rate.

### 2.2.2 Exponential Distribution

When the random variable $T$ has single or one parameter Exponential distribution with scale parameter $\delta > 0$, we indicate this by $T \sim \text{EXP}(\delta)$.

The $f(t)$ for one parameter Exponential model is given by

$$f(t;\delta) = \delta \exp(-\delta t)\ t > 0 \tag{2.6}$$

The cdf of the Exponential distribution is

$$F(t;\delta) = \int_0^t f(t;\delta)dt = 1 - \exp(-\delta t) \tag{2.7}$$

The reliability function, $R(t)$ is

$$R(t) = 1 - F(t) = \exp(-\delta t) \tag{2.8}$$

The hazard function $h(t)$ is given by:

$$h(t) = \frac{f(t)}{R(t)} = \frac{\delta \exp(-\delta t)}{\exp(-\delta t)} = \delta \tag{2.9}$$

The $p$-th quantile $t_p$ of the Exponential distribution is the solution of

$$F(t_p) = p$$

$$1 - \exp(-\delta t_p) = p$$

$$\delta t_p = -\log(1-p)$$

$$t_p = -\frac{1}{\delta}\log(1-p) \tag{2.10}$$

This is the $p$-th quantile of the Exponential distribution.

Mean and variance of Exponential distribution are:

$$E(T) = \frac{1}{\delta} \text{and} \quad Var(T) = \frac{1}{\delta^2}$$

Exponential distribution is the simplest distribution that is commonly used in the analysis of reliability data. This distribution has the important characteristic that its hazard function is constant (does not depend on time t). It is popular for some kinds of electronic components (e.g. capacitors or robust highquality integrated circuits). But this distribution would not be appropriate for a population of electronic components having failure-causing quality defects.

### 2.2.3 Normal Distribution

When $T$ has a Normal distribution, we indicate this by $T \sim \text{NOR}(\mu, \sigma)$. Where $\sigma > 0$ is a scale parameter and $-\infty < \mu < \infty$ is a location parameter. As a model for variability, the Normal distribution has a long history of use in many areas of application.

The $f(t)$ for Normal model is given by

$$f(t;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[ -\frac{1}{2\sigma^2}(t-\mu)^2 \right] \qquad -\infty < t, \mu < \infty \text{ and } \sigma > 0$$

$$f(t;\mu,\sigma) = \frac{1}{\sigma}\Phi\left(\frac{t-\mu}{\sigma}\right) \tag{2.11}$$

where $\Phi(z) = \frac{1}{\sqrt{2\pi}} \exp\left( -\frac{1}{2} z^2 \right)$ is the pdf of the standard Normal distribution.

The cdf of the Normal distribution is

$$F(t;\mu,\sigma) = \int_{-\infty}^{t} f(t;\mu,\sigma)dt$$

$$= \int_{-\infty}^{t} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[ -\frac{1}{2\sigma^2}(t-\mu)^2 \right] dt = \int_{-\infty}^{t} \frac{1}{\sigma}\Phi\left(\frac{t-\mu}{\sigma}\right) dt = \Phi\left(\frac{t-\mu}{\sigma}\right) \tag{2.12}$$

where $\Phi(z) = \Phi\left(\frac{t-\mu}{\sigma}\right) = \int_{-\infty}^{z} \Phi(t)dt$ is the cdf of the standard Normal distribution.

The reliability function is

$$R(t) = 1 - F(t) = 1 - \Phi\left(\frac{t-\mu}{\sigma}\right) \tag{2.13}$$

The hazard function $h(t)$ is given by:

$$h(t) = \frac{f(t)}{R(t)} = \frac{\frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{1}{2\sigma^2}(t-\mu)^2\right]}{1-\Phi\left(\frac{t-\mu}{\sigma}\right)} \tag{2.14}$$

The $p$-th quartile $t_p$ of the Normal distribution is the solution of

$$F(t_p) = p$$

$$\Phi_{nor}\left(\frac{t_p-\mu}{\sigma}\right) = p$$

$$\left(\frac{t_p-\mu}{\sigma}\right) = \Phi^{-1}(p)$$

$$t_p = \mu + \sigma\Phi^{-1}(p)$$

$$t_p = \mu + \sigma(z_p) \tag{2.15}$$

where $\Phi^{-1}(p) = z_p$ is the $p$-th quantile of the standard Normal distribution.

Therefore $t_p = \mu + \sigma(z_p)$ is the $p$-th quantile or 100 $p$-th percentile of Normal distribution.

The mean and variance of Normal distribution are:

$$E(T) = \mu \quad \text{and} \quad Var(T) = \sigma^2$$

In reliability data analysis, the use of the normal distribution is less common. It is useful for certain life data, for example, electric filament devices (e.g., incandescent light bulbs, toaster heating elements) and strength of wire bonds in integrated circuits.

## 2.2.4 Lognormal Distribution

When $T$ has a Lognormal distribution, we indicate this by $T \sim \text{LOGNOR}(\mu,\sigma)$, then $Y = \log(T) \sim NOR(\mu,\sigma)$. The median $t_{.5} = \exp(\mu)$ is a scale parameter and $\sigma > 0$ is a shape parameter. The Lognormal distribution is a common model for failure times.

The probability density function, $f(t)$ for Lognormal model is given by

$$f(t;\mu,\sigma)=\frac{1}{\sigma(t)\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{\log(t)-\mu}{\sigma}\right)^2\right],0<t \tag{2.16}$$

Let, $y = log(t)$ Hence, $dy=\frac{1}{t}dt$ and $-\infty < y < \infty$

Therefore, the pdf of $y$ becomes

$$f(y;\mu,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{1}{2\sigma^2}(y-\mu)^2\right]-\infty<y<\infty$$

Hence, the pdf of the Lognormal distribution is:

$$f(t;\mu,\sigma)=\frac{1}{\sigma(t)}\Phi_{nor}\left[\left(\frac{\log(t)-\mu}{\sigma}\right)\right] \tag{2.17}$$

The cdf of $y$ is

$$F(y;\mu,\sigma)=\int_{-\infty}^{y}f(y;\mu,\sigma)dy$$

$$=\int_{-\infty}^{y}\frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{1}{2\sigma^2}(y-\mu)^2\right]dy=\int_{-\infty}^{y}\frac{1}{\sigma}\Phi\left(\frac{y-\mu}{\sigma}\right)dy=\Phi\left(\frac{y-\mu}{\sigma}\right)$$

where $\Phi(z)=\int_{-\infty}^{z}\Phi(t)dt$ is the cdf of the standard Normal distribution.

Hence,

$$F(t;\mu,\sigma)=\Phi_{nor}\left[\left(\frac{\log(t)-\mu}{\sigma}\right)\right] \tag{2.18}$$

is the cdf of the Lognormal distribution.

The reliability function is

$$R(t)=1-F(t)=1-\Phi_{nor}\left(\frac{\log(t)-\mu}{\sigma}\right) \tag{2.19}$$

The hazard function, $h(t)$ is given by

$$h(t) = \frac{f(t)}{R(t)} = \frac{\frac{1}{\sigma(t)\sqrt{2\pi}} \exp\left[ -\frac{1}{2}\left( \frac{\log(t) - \mu}{\sigma} \right)^2 \right]}{1 - \Phi_{nor}\left( \frac{\log(t) - \mu}{\sigma} \right)} \tag{2.20}$$

The *p*-th quantile $t_p$ of the Lognormal distribution is:

$$F(t_p) = p$$

$$F(t) = \Phi_{nor}\left[ \left( \frac{\log(t) - \mu}{\sigma} \right) \right]$$

$$\Phi_{nor}\left[ \left( \frac{\log(t_p) - \mu}{\sigma} \right) \right] = p$$

$$\left( \frac{\log(t_p) - \mu}{\sigma} \right) = \Phi^{-1}(p)$$

$$t_p = \exp\left[ \mu + \sigma\Phi^{-1}(p) \right] \tag{2.21}$$

Mean of Lognormal distribution is:

$$E(T) = exp\left( \mu + \frac{\sigma^2}{2} \right)$$

Variance of Lognormal distribution is:

$$Var(T) = exp(2\mu + \sigma^2)[exp(\sigma^2) - 1]$$

Hazard function of the lognormal *h*(*t*) start at zero, increases to a point in time, and then decreases eventually to zero. This distribution is often used as a model for a population of electronic components that exhibits a decreasing hazard function. The lognormal distribution is a common model for failure times.The distribution is useful in modeling failure time of a population of electronic components with a decreasing hazard function (due to a small proportion of defects in the population). It is also used to describe the failure-time distribution of certain degradation processes. It can be justified for a random variable that arises from a product of a number of identically distributed independent positive random quantities.

## 2.3 Complex Lifetime Models

Because of quality variation, the failure times do not follow standard distributions and must be modeled by more complex model formulations. Complex models involve two or more simple models. The following Complex lifetime models have used in this thesis:

- Mixture model
- Competing risk model
- Effect of quality variation in manufacturing (includes both Mixture & Competing risk model)

### 2.3.1 Mixture Models

A general *m*-fold mixture model involves *m* subpopulations and is given by

$$F(t;\theta)=\sum_{j=1}^{m} p_j F_j\left(t;\theta_j\right)\; p_j>0\; \sum_{j=1}^{m} p_j=1 \tag{2.22}$$

where $F_j(t;\theta_j)$ is the cdf associated with*j*-th sub-population with parameters $\theta_j$, $j=1,2,\dots,m$and $p_j$ is the mixing probability of the *j*-th sub-population.

Mixture of distributions can model the variability resulting from parts being bought from $m$ different suppliers with $F_j(t)$ denoting the failure distribution for parts obtained from *j*-th supplier, $j=1,2,\dots,m$.

The probability density function of *m*-fold mixture model is given by:

$$f(t;\theta)=\sum_{j=1}^{m} p_j f_j(t;\theta_j) \tag{2.23}$$

where $f_j(t)$ is the density function associated with $F_j(t)$.

The reliability function of *m*-fold mixture model is:

$$R(t)=1-F(t;\theta)=\sum_{j=1}^{m} p_j R_j\left(t;\theta_j\right) \tag{2.24}$$

The hazard function $h(t)$ is given by

$$h(t)=\frac{f(t)}{1-F(t)}=\sum_{j=1}^{m}w_j(t)h_j(t) \tag{2.25}$$

Where $h_j(t)$ is associated with $j$-th subpopulation, and

$$w_j(t)=\frac{p_jR_j(t)}{\sum_{j=1}^{m}p_jR_j(t)} \text{ and } \sum_{j=1}^{m}w_j(t)=1 \tag{2.26}$$

We see that the failure rate for the model is a weighted mean of the failure rate for the subpopulations with the weights varying with *t*.

#### 2.3.1.1 Special Case: 2-fold Weibull Mixture Model (*m=2*)

Let us suppose that,

$$F_1(t)\sim \text{Weibul}(\beta_1,\eta_1)\text{and}F_2(t)\sim\text{Weibull}(\beta_2,\eta_2)$$

Putting *m*=2 in eq.(2.22) we get the cdf of the 2-fold mixture model, which is:

$$F(t)=pF_1(t)+(1-p)F_2(t) \tag{2.27}$$

For the 2-fold two parameter Weibull mixture model, the model is characterized by five parameters: the shape and scale parameters for the two subpopulations and the mixing parameter $p$ ($0<p<1$).

The cdf of Weibull distribution is obtained from eq. (2.2). Putting this value in eq(2.27) we get the cdf of the 2-fold Weibull mixture model, which is:

$$F(t)=\left[p\left\{1-\exp\left[-\left(\frac{t}{\eta_1}\right)^{\beta_1}\right]\right\}+(1-p)\left\{1-\exp\left[-\left(\frac{t}{\eta_2}\right)^{\beta_2}\right]\right\}\right]$$

$$F(t)=\left[1-\exp\left[-\left(\frac{t}{\eta_2}\right)^{\beta_2}\right]\right]+p\left[\exp\left\{-\left(\frac{t}{\eta_2}\right)^{\beta_2}\right\}-\exp\left\{-\left(\frac{t}{\eta_1}\right)^{\beta_1}\right\}\right] \tag{2.28}$$

The density function of 2-fold mixture model is:

$$f(t)=pf_1(t)+(1-p)f_2(t) \tag{2.29}$$

Putting this value of the pdf of Weibull distribution from eq.(2.1), in eq.(2.29) we get the pdf of the 2-fold Weibull mixture model, which is:

$$f(t)=p\left[\frac{\beta_1}{\eta_1}\left(\frac{t}{\eta_1}\right)^{\beta_1-1}\exp\left\{-\left(\frac{t}{\eta_1}\right)^{\beta_1}\right\}\right]+(1-p)\left[\frac{\beta_2}{\eta_2}\left(\frac{t}{\eta_2}\right)^{\beta_2-1}\exp\left\{-\left(\frac{t}{\eta_2}\right)^{\beta_2}\right\}\right] \tag{2.30}$$

The reliability function of 2-fold mixture model is:

$$R(t)=pR_1(t)+(1-p)R_2(t) \tag{2.31}$$

We get the reliability function of Weibull distribution from eq.(2.3). Applying this value in eq.(2.31), we obtain the reliability function of 2-fold Weibull mixture model:

$$R(t)=p\exp\left[-\left(\frac{t}{\eta_1}\right)^{\beta_1}\right]+(1-p)\exp\left[-\left(\frac{t}{\eta_2}\right)^{\beta_2}\right] \tag{2.32}$$

And the hazard function of 2-fold mixture model is:

$$h(t)=w_1(t)h_1(t)+w_2(t)h_2(t)$$

$$h(t)=\frac{pR_1(t)h_1(t)}{pR_1(t)+(1-p)R_2(t)}+\frac{(1-p)R_2(t)h_2(t)}{pR_1(t)+(1-p)R_2(t)} \tag{2.33}$$

From eq.(2.4) we get the hazard function of Weibull distribution.

Hence, $h_1(t)=\frac{\beta_1}{\eta_1}\left[\frac{t}{\eta_1}\right]^{\beta_1-1}$ and $h_2(t)=\frac{\beta_2}{\eta_2}\left[\frac{t}{\eta_2}\right]^{\beta_2-1}$

Now, from equation (3.33),the hazard function of 2-fold Weibull mixture model is given by:

$$h(t)=\frac{p\exp\left[-\left(\frac{t}{\eta_1}\right)^{\beta_1}\right]\frac{\beta_1}{\eta_1}\left(\frac{t}{\eta_1}\right)^{\beta_1-1}}{p\exp\left[-\left(\frac{t}{\eta_1}\right)^{\beta_1}\right]+(1-p)\exp\left[-\left(\frac{t}{\eta_2}\right)^{\beta_2}\right]}$$

$$+\frac{(1-p)\exp\left[-\left(\frac{t}{\eta_2}\right)^{\beta_2}\right]\frac{\beta_2}{\eta_2}\left(\frac{t}{\eta_2}\right)^{\beta_2-1}}{p\exp\left[-\left(\frac{t}{\eta_1}\right)^{\beta_1}\right]+(1-p)\exp\left[-\left(\frac{t}{\eta_2}\right)^{\beta_2}\right]}$$

$$h(t)=\frac{pf_1(t)+(1-p)f_2(t)}{p\exp\left[-\left(\frac{t}{\eta_1}\right)^{\beta_1}\right]+(1-p)\exp\left[-\left(\frac{t}{\eta_2}\right)^{\beta_2}\right]}$$

Finally,

$$h(t)=\frac{f(t)}{p\exp\left[-\left(\frac{t}{\eta_1}\right)^{\beta_1}\right]+(1-p)\exp\left[-\left(\frac{t}{\eta_2}\right)^{\beta_2}\right]} \tag{2.34}$$

### 2.3.1.2 Special Case: Weibull-Normal-Exponential Mixture Model (*m=3*)

Let us suppose that,

$F_1(t)\sim$ Weibul$(\beta,\eta), F_2(t)\sim$Normal$(\mu,\sigma)$ *and* $F_3(t)\sim$Exponential$(\delta)$

Putting *m*=3 in eq.(2.22) we get the cdf of the 3-fold mixture model is given by:

$$F(t)=p_1F_1(t)+p_2F_2(t)+(1-p_1-p_2)F_3(t) \tag{2.35}$$

For the 3-fold mixture model, the model is characterized by the original model parameters for the three subpopulations and the mixing parameters $p_1$ and $p_2$, $(0<p_1 \text{and} p_2<1)$.

The distribution function for Weibull, Normal and Exponential distributions are obtained from eq. (2.2), (2.12) and (2.7), respectively. Putting these values in eq(2.35) we get the cdf of the Weibull-Normal- Exponential mixture model, which is:

$$F(t)=p_1\left\{1-\exp\left[-\left(\frac{t}{\eta}\right)^{\beta}\right]\right\}+p_2\Phi\left(\frac{t-\mu}{\sigma}\right)+(1-p_1-p_2)\{1-\exp(-\delta t)\} \tag{2.36}$$

The density function of 3-fold mixture model is given by:

$$f(t)=p_1f_1(t)+p_2f_2(t)+(1-p_1-p_2)f_3(t) \tag{2.37}$$

The pdf of Weibull, Normal and Exponential distributions are obtained from eq. (2.1), (2.11) and (2.6), respectively. Hence the pdf of the Weibull-Normal- Exponential mixture model is:

$$f(t) = p_1\left[\frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1}\exp\left\{-\left(\frac{t}{\eta}\right)^{\beta}\right\}\right] + p_2\frac{1}{\sigma}\Phi\left(\frac{t-\mu}{\sigma}\right) + (1-p_1-p_2)\delta\exp(-\delta t) \tag{2.38}$$

Again the reliability function of 3-fold mixture model is given by:

$$R(t) = p_1R_1(t) + p_2R_2(t) + (1-p_1-p_2)R_3(t) \tag{2.39}$$

From eq. (2.3), (2.13) and (2.8), we get the reliability functions of Weibull, Normal and Exponential, respectively. Hence the $R(t)$ of the Weibull-Normal- Exponential mixture model is:

$$R(t) = p_1\left[\exp\left\{-\left(\frac{t}{\eta}\right)^{\beta}\right\}\right] + p_2\left[1-\Phi\left(\frac{t-\mu}{\sigma}\right)\right] + (1-p_1-p_2)\exp(-\delta t) \tag{2.40}$$

And the hazard function of 3-fold mixture model is:

$$h(t) = w_1(t)h_1(t) + w_2(t)h_2(t) + w_3(t)h_3(t)$$

$$h(t) = \frac{p_1R_1(t)h_1(t) + p_2R_2(t)h_2(t) + (1-p_1-p_2)R_3(t)h_3(t)}{p_1R_1(t) + p_2R_2(t) + (1-p_1-p_2)R_3(t)} \tag{2.41}$$

From eq.(2.4), eq.(2.14) and eq.(2.9), we get the hazard functions of Weibull, Normal and Exponential, respectively.

Hence, $$h_1(t) = \frac{\beta}{\eta}\left[\frac{t}{\eta}\right]^{\beta-1}$$

$$h_2(t) = \frac{\frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{1}{2\sigma^2}(t-\mu)^2\right]}{1-\Phi\left(\frac{t-\mu}{\sigma}\right)}$$

And $$h_3(t) = \delta$$

Hence, putting the values of $R_j(t)$s and $h_j(t)$s in equation (2.41),we get the hazard function of Weibull-Normal-Exponential mixture model. Where $j$ =1, 2, 3.

## 2.3.2 Competing Risk Models

A Competing risk model involves*m* distributions and is derived as follows. Let *T*denote an independent random variable with a distribution function $F_j(t)$ , $j = 1,2,\dots,m$. Then the cdf for a general *m*-fold competing risk model is given by

$$F(t;\theta) = 1 - \prod_{j=1}^{m}\left[1 - F_j\left(t;\theta_j\right)\right] \tag{2.42}$$

here $F_j\left(t;\theta_j\right)$ are the distribution functions of the *j*-th sub-populations with parameters $\theta_j, j = 1,2,\dots,m$. This model is commonly referred as the *Competing Risk Model*. It has also been called The *Compound Mod, Series System Model* and *Multirisk Model.* The risk model has a long history and Nelson (1982, pp. 162-163) traces its origin to 200 years ago.

This is called a competing risk model because it is applicable when an item (component or module) may fail by any one of *m* failure modes, i.e., it can fail due to any one of the *m* mutually exclusive causes in a set $C_1, C_2, \dots, C_m$.

Note that (2.42) can be written as

$$\overline{F}(t) = \prod_{j=1}^{m}\left[\overline{F}_j(t)\right] \tag{2.43}$$

Where $\overline{F}(t) = 1 - F(t)$ and $\overline{F}_j(t) = 1 - F_j(t)$ for$j = 1,2,\dots,m$ .

The density function of a general *m*-fold competing risk model is given by:

$$f(t) = \sum_{j=1}^{m}\left[\prod_{\substack{k=1\\k\neq j}}^{m}\left[1 - F_k(t)\right]\right] f_j(t)$$

and can be written as

$$f(t) = \overline{F}(t)\sum_{j=1}^{m}\left[\frac{f_j(t)}{\overline{F}_j(t)}\right] = \sum_{j=1}^{m}\left[f_j(t)\prod_{\substack{k=1\\k\neq j}}^{m}\left[R_k(t)\right]\right] \tag{2.44}$$

Again we know, $\overline{F}(t) = R(t)$ and $\overline{F}_j(t) = R_j(t)$ . So, from eq.(2.43), we get the reliability function of *m*-fold competing risk model is:

$$R(t) = \prod_{j=1}^{m}\left[R_j(t)\right] \tag{2.45}$$

The hazard function of *m*-fold competing risk model is given by:

$$h(t) = \sum_{j=1}^{m} h_j(t) \tag{2.46}$$

where $h_i(t)$ is the hazard function associated with*j*-thsubpopulation.

### 2.3.2.1 2-Fold Weibull Competing Risk Model (*m=2*)

Suppose that,

$$F_1(t) \sim \text{Weibul}(\beta_1, \eta_1) \text{and} F_2(t) \sim \text{Weibull}(\beta_2, \eta_2)$$

Putting *m*=2 in eq.(2.42) we get the cdf of the 2-fold competing risk model is given by:

$$F(t) = F_1(t) + F_2(t) - F_1(t)F_2(t) \tag{2.47}$$

Using the cdf of Weibull distribution from eq.(2.2), we get the cdf of 2-fold Weibull competing risk model, as:

$$F(t) = \left[\left\{1-\exp\left[-\left(\frac{t}{\eta_1}\right)^{\beta_1}\right]\right\} + \left\{1-\exp\left[-\left(\frac{t}{\eta_2}\right)^{\beta_2}\right]\right\} - \left\{1-\exp\left[-\left(\frac{t}{\eta_1}\right)^{\beta_1}\right]\right\}\left\{1-\exp\left[-\left(\frac{t}{\eta_2}\right)^{\beta_2}\right]\right\}\right]$$

$$= 2 - \exp\left[-\left(\frac{t}{\eta_1}\right)^{\beta_1}\right] - \exp\left[-\left(\frac{t}{\eta_2}\right)^{\beta_2}\right] - 1 + \exp\left[-\left(\frac{t}{\eta_1}\right)^{\beta_1}\right] + \exp\left[-\left(\frac{t}{\eta_2}\right)^{\beta_2}\right]$$

$$- \exp\left[-\left\{\left(\frac{t}{\eta_1}\right)^{\beta_1} + \left(\frac{t}{\eta_2}\right)^{\beta_2}\right\}\right]$$

$$F(t) = 1 - \exp\left[-\left\{\left(\frac{t}{\eta_1}\right)^{\beta_1} + \left(\frac{t}{\eta_2}\right)^{\beta_2}\right\}\right] \tag{2.48}$$

The probability density function,*f*(*t*) of the 2-fold competing risk model is:

$$f(t) = \overline{F}(t)\left[\frac{f_1(t)}{\overline{F}_1(t)} + \frac{f_2(t)}{\overline{F}_2(t)}\right] \tag{2.49}$$

Using the pdf of two parameters Weibull distribution given in eq.(2.1), we get:

$$\overline{F}(t) = \exp\left[-\left\{\left(\frac{t}{\eta_1}\right)^{\beta_1} + \left(\frac{t}{\eta_2}\right)^{\beta_2}\right\}\right]$$

Hence,

$$\overline{F}_1(t) = \exp\left[-\left\{\left(\frac{t}{\eta_1}\right)^{\beta_1}\right\}\right] \text{ and } \overline{F}_2(t) = \exp\left[-\left\{\left(\frac{t}{\eta_2}\right)^{\beta_2}\right\}\right]$$

Now,from eq.(2.49), the pdf for 2-fold Weibull competing risk model is :

$$f(t) = \exp\left[-\left\{\left(\frac{t}{\eta_1}\right)^{\beta_1} + \left(\frac{t}{\eta_2}\right)^{\beta_2}\right\}\right]\left[\frac{\frac{\beta_1}{\eta_1}\left[\frac{t}{\eta_1}\right]^{\beta_1-1}\exp\left[-\left(\frac{t}{\eta_1}\right)^{\beta_1}\right]}{\exp\left[-\left\{\left(\frac{t}{\eta_1}\right)^{\beta_1}\right\}\right]} + \frac{\frac{\beta_2}{\eta_2}\left[\frac{t}{\eta_2}\right]^{\beta_2-1}\exp\left[-\left(\frac{t}{\eta_2}\right)^{\beta_2}\right]}{\exp\left[-\left\{\left(\frac{t}{\eta_2}\right)^{\beta_2}\right\}\right]}\right]$$

$$f(t) = \exp\left[-\left\{\left(\frac{t}{\eta_1}\right)^{\beta_1} + \left(\frac{t}{\eta_2}\right)^{\beta_2}\right\}\right]\left[\frac{\beta_1}{\eta_1}\left[\frac{t}{\eta_1}\right]^{\beta_1-1} + \frac{\beta_2}{\eta_2}\left[\frac{t}{\eta_2}\right]^{\beta_2-1}\right] \qquad (2.50)$$

Again the reliability function, $R(t)$ of 2-fold competing risk model is:

$$R(t) = R_1(t)R_2(t) \qquad (2.51)$$

Using the reliability function of Weibull distribution from eq.(2.3) in eq.(2.51), we achieve the reliability function of 2-fold Weibull competing risk model as:

$$R(t) = \exp\left[-\left\{\left(\frac{t}{\eta_1}\right)^{\beta_1} + \left(\frac{t}{\eta_2}\right)^{\beta_2}\right\}\right] \qquad (2.52)$$

The hazard function $h(t)$ of 2-fold competing risk model is:

$$h(t) = h_1(t) + h_2(t) \qquad (2.53)$$

$$\text{Now, } h_1(t) = \frac{\beta_1}{\eta_1}\left[\frac{t}{\eta_1}\right]^{\beta_1-1} \quad \text{and} \quad h_2(t) = \frac{\beta_2}{\eta_2}\left[\frac{t}{\eta_2}\right]^{\beta_2-1}$$

Hence, the hazard function of 2-fold Weibull competing risk model is given by:

$$h(t) = \frac{\beta_1}{\eta_1}\left[\frac{t}{\eta_1}\right]^{\beta_1-1} + \frac{\beta_2}{\eta_2}\left[\frac{t}{\eta_2}\right]^{\beta_2-1} \qquad (2.54)$$

### 2.3.3 Effect of Quality Variation in Manufacturing

In the thesis, we propose to use a complex lifetime model for modeling the effect of quality variation inmanufacturing process. It is a general model in the sense that it contains both mixture and competing risk model.

Herefollowing two main causes are considered for making variation on the quality of the manufactured items:

- Assembly error: new mode of failure (detected soon after put in to operation), and
- Component non-conformance:resulting in some items having inferior reliability (higher failure rate, shorter mean time to failure, etc.)

The following notations are used for modeling the effect of quality variation:

- Designed reliability : $R_0(t)$
- Conforming item's cdf : $F_0(t)$
- Reliability function associated with assembly error (new failure mode): $R_1(t)$
- cdf associated with assembly error (new failure mode): $F_1(t)$
- Reliability function associated with non-conforming component item: $R_2(t)$
- cdf associated with non-conforming component item: $F_2(t)$
- Probability that an item has an assembly error : $q$
- Probability that an item has an non-conforming component : $p$
- Probability that the item has no non-conforming component : $(1-p)$

Under assembly errors situation, the reliability of produced items can be modeled by a modified competing risk model (Murthy et al. 2003) given by

$$R_a(t) = R_0(t)[1 - qF_1(t)], \quad t \geq 0 \tag{2.55}$$

Under component non-conformance situation, the reliability of the produced items can be expressed as

$$R_n(t) = (1-p)R_0(t) + pR_2(t), \quad t \geq 0 \tag{2.56}$$

With both assembly errors and problems with component non-conformance, that is, having combined effects, the reliability function of the produced item becomes:

$$R_q(t) = [(1-p)R_0(t) + pR_2(t)][1 - qF_1(t)], \quad t \geq 0 \tag{2.57}$$

The cdf of the model with effect of quality variation in manufacturing is given by:

$$F_q(t) = 1 - R_q(t)$$

$$F_q(t) = 1 - [(1-p)R_0(t) + pR_2(t)][1 - qF_1(t)] \quad (2.58)$$

$$= (1-p)F_0(t) + pF_2(t) - (1-p)qR_0(t)F_1(t) + pqF_1(t)R_2(t) \quad (2.59)$$

Now, with both assembly errors and problems with component non-conformance, the probability density function of the item produced is:

$$f_q(t) = \frac{dF_q(t)}{dt} = \frac{d}{dt}\big[1 - [(1-p)R_0(t) + pR_2(t)][1 - qF_1(t)]\big]$$

$$f_q(t) = (1-p)f_0(t) + pf_2(t) + (1-p)qf_1(t)R_0(t) - (1-p)qf_0(t)F_1(t) + pqf_1(t)R_2(t) - pqf_2(t)F_1(t) \quad (2.60)$$

It can be easily seen that, this pdf contains three distributions with pdfs $f_0(t)$ , $f_1(t)$and $f_2(t)$, and two additional parameters$p$ and $q$.

Based on the reliability function for the combined model effects (2.57), the following three Cases can be considered:

**Case-1:**

If the product has neither component non-conformance error nor assembly error i.e., if we put $p = 0$ and $q = 0$in eq. (2.57), eq. (2.59) and eq. (2.60), respectively,then we get:

$$R_q(t) = R_0(t) \tag{2.61}$$

$$F_q(t) = F_0(t) \tag{2.62}$$

$$f_q(t) = f_0(t) \tag{2.63}$$

These are the equations of the reliability function, cdf and pdf, respectively, when there exists no quality variation in the produced items. In this case, the field reliability remains the same as of design reliability of the product.

**Case-2:**

If the product has no assembly error but only component non-conformance i.e., if we put $q = 0$ in eq.(2.57), eq.(2.59) and eq.(2.60), respectively, then the reliability function, cdf and pdfbecome:

$$R_q(t) = (1-p)R_0(t) + pR_2(t) \tag{2.64}$$

$$F_q(t) = (1-p)F_0(t) + pF_2(t) \tag{2.65}$$

$$f_q(t) = (1-p)f_0(t) + pf_2(t) \tag{2.66}$$

These equations indicate the reliability function, cdf and pdfof a 2-fold mixture model, respectively.

**Case-3:**

If the product has no non-conformance component but only assembly error i.e., if we put $p = 0$ and $q = 1$ in eq.(2.57), eq.(2.59) and eq.(2.60), respectively, then the reliability function, cdf and pdfbecome:

$$R_q(t) = R_0(t)R_1(t) \tag{2.67}$$

$$F_q(t) = 1 - R_0(t)R_1(t) \tag{2.68}$$

$$f_q(t) = f_0(t)R_1(t) + f_1(t)R_0(t) \tag{2.69}$$

These givethe reliability function, cdf and pdfof a 2-fold competing risk model, respectively.

The models under Case-2 and Case-3 are the main models applied in this thesis extensively for modeling the product reliability data.

# Chapter 3
# Statistical Methods for Reliability Data Analysis

## 3.1 Introduction

This chapter discusses how to fit a suitable probability model to a given dataset. It presentsthe following statistical methods to address the issue.

## 3.2 Nonparametric Estimation of cdf

Probability plots are commonly used to identify an appropriate model for fitting a given dataset. To present the data on a plotting paper, the empirical cdf for each failure time must be first estimated using a nonparametric method. Some parameter estimation methods such as the graphical and least square methods also require estimating the empirical cdf.

A nonparametric analysis provides an intermediate step toward a more highly structured model allowing more precise or more extensive inferences, provided that the additional assumptions for such model are valid. Nonparametric analysis allows the analyst to characterize life data without assuming an underlying distribution. There are several methods for conducting a nonparametric analysis; we have used the Kaplan-Meier method discussed below:

### 3.2.1 The Kaplan-Meier Estimate of Reliability Function

Exact failure times arise from a continuous inspection process (or, perhaps, from having used a very large number of closely-spaced inspections). In the limit as the number of inspection increases and the width of the inspection intervals approaches zero, failures are concentrated in a relatively small number of intervals. Most intervals will not contain any failures. $\hat{F}(t)$ is constant over all intervals that have no failures. Thus with small intervals $\hat{F}$ will become a step function with gaps over the interval,

where, there were failures and with jumps at the upper endpoint of the intervals. In the limit, as the width of the intervals approaches zero, the size of the gaps approaches zero and the step function increases at the reported failure times. This limiting case of the interval-based nonparametric estimator is generally known as the Product Limit or Kaplan-Meier Estimator (Kaplan and Meier (1958)).

A useful way of portraying ungrouped univariate survival data is to compute and graph the empirical survivor function or, equivalently the empirical distribution function. This also provides a nonparametric estimate of the survivor or distribution function for the life distribution under study.

If there are no censored observations in a sample of size $n$, the empirical survivor function (ESF) is defined as

$$S(t) = \frac{number\ of\ indeviduals\ surviving\ longer\ than\ t}{total\ number\ of\ individuals\ studied} \tag{3.1}$$

This is a step function that decreases by $1/n$ just after each observed lifetime, if all observation is distinct. More generally, if there are $d$ lifetimes equal to $t$, the ESF drops by $d/n$ just pass $t$.

When dealing with censored data, some modifications of equation (3.1) is necessary, since the number of lifetime is greater than or equal to $t$ will not be generally known exactly. The modification of (3.1) described here has come to be called the "Product-Limit" (PL) estimate of the survivor function or, sometime the Kaplan-Meier estimate, from the authors who first discussed it's properties (Kaplan and Meier, 1958).

- The Kaplan–Meier estimator (also known as the product limit estimator) estimates the survival function from life-time data.
- The Kaplan-Meier estimator can be used to calculate values for nonparametric reliability for data sets with multiple failures and suspensions.
- Product limit estimate of $S(t)$ is the MLE of $S(t)$.

This is an estimator used as an alternative to the median ranks method for calculating the estimates of the unreliability for probability plotting. The Kaplan–Meier estimator estimates the survival functions from life-time data. In medical research, it might be

used to measure the fraction of patients living for a certain amount of time after treatment. An economist might measure the length of time people remain unemployed after a job loss. An engineer might measure the time until failure of machine parts. An ecologist may use it to estimate how long fleshy fruits remain on plants before they are removed by frugivorous.

## Assumptions

- Censored individuals have the same prospect of survival as those who continue to be followed. This cannot be tested for and can lead to a bias that artificially reduces *S*.
- Survival prospects are the same for early as for late recruits to the study (can be tested for).
- The event studied (e.g. death) happens at the specified time. Late recording of the event studied will cause artificial inflation of *S*.

Suppose that, an initial sample of *n* units start operating at time zero. If a unit does not fail in the interval *i*. It is either censored at the end of the interval *i* or it continuous into interval (*i*+1). Information is available on the status of the units at the end of each interval. The intervals may be large or small and need not be of equal length, as long as the intervals for different units do not overlap.

Suppose that, there are observations on *n* individuals and that there are $k(k \le n)$ distinct times $t_1 < t_2 < ... < t_k$ at which deaths occurs. The possibility of there being more than one death at $t_i$ is allowed. Let $d_i$ denote the number of units that died or failed in the *i*-th interval $(t_{i-1}, t_i]$. In additions to the lifetimes $t_1,...,t_k$ ,there are also censoring times $L_i$ for individuals, whose lifetimes are not observed. Also let, $r_i$ denote the number of units that survive interval *i* and are right-censored at $t_i$. The units that are alive at the beginning of interval *i* are called the 'risk set' for interval *i* (i.e., those at risk to failure) and the size of risk set at the beginning of interval *i* is

$$n_i = n - \sum_{j=0}^{i-1} d_j - \sum_{j=0}^{i-1} r_j, i = 1,2,...,m \quad (3.2)$$

Where $m$ is the number of intervals and it is understood that, $d_0 = 0$ and $r_0 = 0$. An estimator of the conditional probability of failing in the interval *i,* given that a unit enters this interval, is the sample proportion failing

$$\hat{p}_i = \frac{d_i}{n_i}\ , \qquad \text{i} = 1,...,m \tag{3.3}$$

Now the Product-limit estimate of the survival function $S(t_i)$ is defined as

$$\hat{S}(t_i) = \prod_{j=1}^{i}[1 - \hat{p}_j]\ , \qquad \text{i} = 1,2,...,\text{m}$$

$$\text{Hence, } \hat{S}(t_i) = \prod_{j=1}^{i}\left[\ \frac{n_j - d_j}{n_j}\ \right]\ ,\text{i} = 1,2,...,\text{m} \tag{3.4}$$

Let $T$ be the random variable that measures the time of failure and let $F(t)$ be its cumulative distribution function. Note that

$$S(t) = P[T > t] = 1 - P[T \le t] = 1 - F(t)$$

$$\text{Hence,} \qquad \hat{F}(t_i) = 1 - \hat{S}(t_i) \tag{3.5}$$

Consequently, the right-continuous definition of $\hat{S}(t)$ may be preferred in order to make the estimate compatible with a right-continuous estimate of $F(t)$.

Here $\hat{p}_i$ is the maximum likelihood (ML) estimator of the conditional probability $p_i$. This implies that, $\hat{F}(t_i)$ is the ML estimator of $F(t_i)$. The nonparametric estimator $\hat{F}(t_i)$ is defined at all $t_i$ values (endpoints of all observation intervals). Additionally, if interval $i$ is known to have zero failures, then $\hat{F}(t_i) = \hat{F}(t_{i-1})$ for $t_{i-1} \le t \le t_i$. If interval $i$ is known to contain one or more failures, $\hat{F}(t)$ increases from $\hat{F}(t_{i-1})$ to $\hat{F}(t_i)$ in the interval $(t_{i-1}, t_i]$ but, as before, $\hat{F}(t)$ is not defined over the interval.

The product-limit estimate possesses several desirable large sample properties, a main one being that, $\hat{S}(t)$ is a consistent estimate of $S(t)$, under suitable assumptions about censoring. A through study of the properties of the PL estimate is rather involved; work in this area has been done by Breslow and Crowley (1974), Meier (1975), Efron (1967), Kaplan and Meier (1958), Peterson (1977), Johansen (1978) and others. We shall merely outline a few pertinent results and refer the reader to these papers for

more details. The important points are that, $\hat{S}(t)$ is a consistent estimate of $S(t)$ under quite broad conditions.

A rigorous approach can be taken by discretizing the time axis and then passing to a limit. Suppose that, the time axis is partitioned into intervals $I_j = \lfloor a_{j-1}, a_j \rfloor$, $j=1,\ldots,k+1$, with $a_0 = 0$, $a_k = T$ and $a_{k+1} = \infty$. Once again $T$ is an upper limit on the observation time; asymptotic results about $\hat{S}(t)$ will refer to the interval $[0,T)$. The PL estimate (3.3) can be viewed as the left continuous limit that the standard life table estimates give in estimating the $S(a_j)'s$ when $k \to \infty$ while $\max\left(\left|a_j - a_{j-1}\right|\right) \to 0$. Here we suppose, as before, that censoring times are never equal to lifetimes.

To study properties of $\hat{S}(t)$, it is necessary to make specific assumptions about censoring. For example, Breslow and Crowley (1974) adopt the independent random censorship model. Meier (1975) assumes that, the censoring times are fixed but that the sequence of censoring times has certain properties as the sample size $n \to \infty$. In a careful treatment of this problem, it is necessary to verify the conditions under which the two limiting operations $n \to \infty$ and $k \to \infty$ can be interchanged. Breslow and Crowley (1974) and Meier (1975) give rigorous discussions of this.

Nonparametric analysis allows the user to analyze data without assuming an underlying distribution. This can have certain advantages as well as disadvantages. The ability to analyze data without assuming an underlying life distribution avoids the potentially large errors brought about by making incorrect assumptions about the distribution. On the other hand, the confidence bounds associated with nonparametric analysis are usually much wider than those calculated via parametric analysis, and predictions outside the range of the observations are not possible. Some practitioners recommend that any set of life data should first be subjected to a nonparametric analysis before moving on to the assumption of an underlying distribution.

## 3.3 Parameter Estimation Method

For a given set of data and a given parametric model, the parameter estimation deals with determining the model parameters. There are several methods to estimate the parameters and different methods produce different estimates. We have used the Maximum Likelihood Estimation (MLE) Method to estimate the model parameters. A

well-known iterative procedure; the ExpectationMaximization (EM) Algorithm is used to estimate the MLE of the parameters.

## 3.3.1 Maximum Likelihood Estimate of Parameter

### 3.3.1.1 MLE for Complete Data

For the case of complete data, the likelihood function $L$ is given by

$$L(\theta) = \prod_{i=1}^{n} f(t_i;\theta) \qquad (3.6)$$

Here $\theta$ are the values of the model parameters. The maximumlikelihood estimate (MLE) of $\theta$ is the value of $\hat{\theta}$, that maximizes the likelihood function given by (3.6). As a result the estimate is a function [say, $\psi(t_1,t_2,...,t_n)$] of the data. The expression $\psi(T_1,T_2,...,T_n)$ is called the maximumlikelihood estimator and plays an important role in study of the estimate.

Under certain regularity conditions,maximumlikelihood estimators are consistent, asymptotically unbiased, efficient and normally distributed. Asymptotic efficiency here means that, as $n \to \infty$, the covariance matrix of the estimator achieves the lower bound of the Cramer-Rao inequality. For a rigorous treatment of the theory of maximum likelihood, asymptotic results and related topics see Stuart and Ord (1991).

### 3.3.1.2 MLE for Random Censored Data

The general likelihood with failure and left, right and interval censored data, is the total likelihood or joint probability of the data for $n$ independent observations, can be written as

$$\begin{aligned} L(\theta) &= C\prod_{i=1}^{n} L_i(\theta) \\ &= C\prod_{i=1}^{m+1} [F(t_i)]^{l_i} [F(t_i) - F(t_{i-1})]^{d_i} [1 - F(t_i)]^{r_i} [f(t_i)]^{n_i} \end{aligned}$$

where $l_i$, $d_i$, $r_i$, and $n_i$ represent the numbers of left censored, interval censored, right censored and uncensored observations, respectively, and $n = \sum_{j=1}^{m+1} (l_j + d_j + r_j + n_j)$.

$C$ is a constant depending on the sampling inspection scheme but not on $\theta$. So we can take $C$=1.

For Random censored data, the likelihood function is given by

$$L(\theta) = \prod_{i=1}^{n}[f(t_i)]^{\delta_i}[1 - F(t_i)]^{1-\delta_i} \quad (3.7)$$

where $\delta_i = 1$, if the *i*-th observation is uncensored or failure at the time $t_i$ and $\delta_i = 0$ if the *i*-th observation is censored at the time $t_i$. We want to find the parameter vector, $\theta$, so that $L(\theta)$ becomes maximum.

## 3.3.2 Expectation Maximization Algorithm

The general form of Expectation Maximization (EM) algorithm was given in Dempster, Laird, and Rubin (1977), although essence of the algorithm appeared previously in various forms. The EM algorithm is a broadly applicable iterative procedure for computing maximum likelihood estimates in problems with incomplete data. The EM algorithm consists of two conceptually distinct steps at each iteration:

- Expectation-step or E-step and
- Maximization-step or M-step

### 3.3.2.1 Formulation of the EM Algorithm

Suppose we have a model for a set of complete data *Y*, with associated density $f(Y|\theta)$, where $\theta = (\theta_1, \theta_2, \dots, \theta_d)'$ is a vector of unknown parameters with parameter space $\Omega$.

Here the complete data, $Y = (Y_{obs}, Y_{mis})$. That means, *Y* indicates all the observations that we wish to have. $Y_{obs}$ represents the observed part of *Y* i.e., these all are the values of the observations that we have and $Y_{mis}$ denotes the missing values i.e., the incomplete or unobserved observations.

The EM algorithm is designed to find the value of $\theta$, which is denoted by $\theta^*$, that maximizes the incomplete data log-likelihood

$$\log L(\theta) = \log f(Y_{obs}|\theta)$$

That is, the MLE of $\theta$ based on the observed data $Y_{obs}$.

The EM algorithm starts with an initial value $\theta^{(0)} \in \Omega$. Suppose that $\theta^{(k)}$ denotes the estimate of $\theta$ at the *k*-th iteration; then the (*k*+*1*)st iteration can be described in two steps as follows:

- **E-step:**

  Find the conditional expected complete-data log likelihood given observed data and $\theta = \theta^{(k)}$.

$$Q\left(\theta|\theta^{(k)}\right) = E\left(\log L\left(\theta, Y|Y_{obs}, \theta = \theta^{(k)}\right)\right)$$
$$= \int \log L(\theta|Y)\, f\left(Y_{mis}|Y_{obs}, \theta = \theta^{(k)}\right) dY_{mis}$$

  This, in the case of linear exponential family, amounts to estimating the sufficient statistics for the complete data.

- **M-step**

  Determine $\theta^{(k+1)}$ to be a value of $\theta \in \Omega$ that maximizes $Q\left(\theta|\theta^{(k)}\right)$.

  The MLE of $\theta$ is found by iterating between the E and M steps until a convergence criterion is met.

Details can be found in Hartley (1958), Dempster et al. (1977), Little and Rubin (1987) and McLachlan and Krishnan (1997).

### 3.3.2.2 Estimation of Mixing Proportions Using EM Algorithm

Suppose that the pdf of a random vector has a*m*-component mixture form

$$f\left(t|\Theta\right) = \sum_{j=1}^{m} p_j f_j\left(t|\theta_j\right) \tag{3.8}$$

and the reliability function of *T* has the form

$$R\left(t|\Theta\right) = \sum_{j=1}^{m} p_j R_j\left(t|\theta_j\right) \tag{3.9}$$

where the parameters $\Theta = \left(p_1,....,p_m,\theta_1,....,\theta_m\right)$ are such that $p_j > 0$ for (j=1,…,m) and $\sum_{j=1}^{m} p_j = 1$. Constant $p_j$ is called mixing parameters and $f_j$ a component density

function parameterized by $\theta_j$. Generally speaking, a mixture distribution can be composed of $m$ component distributions $f_j$, each of a different type. Estimating unknown parameters of a mixture in its different underlying components is a difficult undertaking.

We let $y = (t_1, ..., t_n)^T$ denote the observed random sample obtained from the mixture density (3.8).

The log-likelihood function for random censoring that can be formed from the observed data y is given by

$$L(t, \Theta) = \prod_{i=1}^{n} [f(t_i|\Theta)]^{\partial_i} [R(t_i|\Theta)]^{(1-\partial_i)} \tag{3.10}$$

where $\partial_i$ is the censored indicator .i.e., if $\partial_i$ =1, then the observation is failure & if $\partial_i$ =0, then the observation is censored. The log-likelihood function can be expressed as:

$$\ln L = \sum_{i=1}^{n} [\partial_i \ln f(t_i) + (1-\partial_i) \ln R(t_i)] \tag{3.11}$$

We now introduce as the unobservable or missing data the vector

$$z = (z_1^T, ..., z_n^T)^T$$

where $z_i$ is a $m$-dimensional vector of zero-one indicator variables and where $z_{ij}$ is one or zero according to whether $t_i$ arose or did not arise from the $j$-th component of the mixture ($i = 1, \ldots, n$ ; $j = 1, \ldots, m$). The EM algorithm handles the addition of the unobservable data to the problem by working with the current conditional expectation of the complete-data log likelihood given the observed data. On defining the complete-data vector $x$ as

$$x = (y^T, z^T)^T$$

The complete-data log-likelihood for $\Theta$ has the form:

$$Q(\Theta, \Theta^{(k)}) = \sum_{i=1}^{n}\sum_{j=1}^{m} z_{ij}\partial_i \ln[p_j f_j(t_i|\theta_j)] + \sum_{i=1}^{n}\sum_{j=1}^{m} z_{ij}(1-\partial_i)\ln[p_j R_j(t_i|\theta_j)]$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{m} z_{ij} \ln p_j + \sum_{i=1}^{n}\sum_{j=1}^{m} z_{ij}\partial_i \ln\left[f_j\left(t_i\middle|\theta_j\right)\right] + \sum_{i=1}^{n}\sum_{j=1}^{m} z_{ij}(1-\partial_i)\ln\left[R_j\left(t_i\middle|\theta_j\right)\right] \tag{3.12}$$

As eq.(3.12) is linear in the unobservable data $z_{ij}$, the Estep (on the $(k+1)$th iteration) simply requires the calculation of the current conditional expectation of $Z_{ij}$ given the observed data $y$, where $Z_{ij}$ is the random variable corresponding to $z_{ij}$. Now

$$E_{\Theta^{(k)}}\left(Z_{ij}\middle|y\right) = z_{ij}^{(k)}$$

Where by Bayes Theorem

$$z_{ij}(k) = \frac{p_j(k)\left[\partial_i f_j\left(t_i\middle|\theta_j(k)\right) + (1-\partial_i)R_j\left(t_i\middle|\theta_j(k)\right)\right]}{\sum_{j=1}^{m} p_j(k)\left[\partial_i f_j\left(t_i\middle|\theta_j(k)\right) + (1-\partial_i)R_j\left(t_i\middle|\theta_j(k)\right)\right]} \tag{3.13}$$

The evaluation of this expectation is called the E-step of the algorithm. The second step (the M-step) of the EM algorithm is to maximize the expectation we computed in the first step with respect to the parameters to obtain new parameter estimations $\Theta^{(k+1)}$. To maximize eq. (3.12), we can maximize the term containing $p_j$ and the term containing $\theta_j$ independently since they are not related. To find the expression for $p_j$; we introduce the Lagrange multiplier $\lambda$ with the constraint that $\sum_{j=1}^{m} p_j = 1$; and taking the derivative of eq. (3.12) with respect to $p_j$ equal to zero:

$$\sum_{i=1}^{n} \frac{1}{p_j} z_{ij}^{(k)} + \lambda = 0 \tag{3.14}$$

Summing both sizes over $j$ and using $\sum_{j=i}^{m} z_{ij}^{(k)} = 1$; we get that $\lambda = -n$ resulting in:

$$p_j^{(k+1)} = \frac{1}{n}\sum_{i=1}^{n} z_{ij}^{(k)} \tag{3.15}$$

For some distributions, it is possible to get analytical expressions for $\theta_j$ as a function of everything else. For example, if we assume $\theta_j$ contains the scale parameter $\eta_j$ &

shape parameter $\beta_j$, of subpopulation $j$; $\eta_j > 0$, $\beta_j > 0$. i.e., $\theta_j = (\eta_j, \beta_j)$. Then eq (3.8) & eq (3.9) become

$$f(t|\Theta) = \sum_{j=1}^{m} p_j f_j(t|\eta_j, \beta_j) \tag{3.16}$$

and

$$R(t|\Theta) = \sum_{j=1}^{m} p_j R_j(t|\eta_j, \beta_j) \tag{3.17}$$

Taking the natural logarithm of eq. (3.16) & (3.17) and substituting into the right side of eq. (3.12), we get:

$$Q(\Theta, \Theta^{(k)}) = \sum_{i=1}^{n}\sum_{j=1}^{m} z_{ij} \ln p_j + \sum_{i=1}^{n}\sum_{j=1}^{m} z_{ij}\partial_i \ln\left[f_j(t_i|\eta_j, \beta_j)\right] + \sum_{i=1}^{n}\sum_{j=1}^{m} z_{ij}(1-\partial_i)\ln\left[R_j(t_i|\eta_j, \beta_j)\right] \tag{3.18}$$

Taking the derivative of eq (3.18) with respect to $\eta_j$ & $\beta_j$; set them to 0, we get:

$$\frac{\partial Q(\Theta, \Theta^{(k)})}{\partial \eta_j} = \sum_{i=1}^{n} z_{ij}\partial_i \frac{\partial f_j(t_i|\eta_j, \beta_j)/\partial \eta_j}{f_j(t_i|\eta_j, \beta_j)} + \sum_{i=1}^{n} z_{ij}(1-\partial_i)\frac{\partial R_j(t_i|\eta_j, \beta_j)/\partial \eta_j}{R_j(t_i|\eta_j, \beta_j)} = 0 \tag{3.19}$$

$$\frac{\partial Q(\Theta, \Theta^{(k)})}{\partial \beta_j} = \sum_{i=1}^{n} z_{ij}\partial_i \frac{\partial f_j(t_i|\eta_j, \beta_j)/\partial \beta_j}{f_j(t_i|\eta_j, \beta_j)} + \sum_{i=1}^{n} z_{ij}(1-\partial_i)\frac{\partial R_j(t_i|\eta_j, \beta_j)/\partial \beta_j}{R_j(t_i|\eta_j, \beta_j)} = 0 \tag{3.20}$$

Eq. (3.15), (3.19) & (3.20) for the estimation of the new parameters $\Theta^{(k+1)}$ in terms of the old parameters $\Theta^{(k)}$ perform both the expectation step and the maximization step simultaneously. The algorithm proceeds by using the newly derived parameters as the guess for the next iteration. The E and M steps are iterated until the algorithm converges.

Finally, the algorithm for the MLE of the parameters of a mixture distribution with censored data can be summarized:

**1.** Begin with an initial guess of $p_j^{(0)}, \eta_j^{(0)}$ & $\beta_j^{(0)}$.

**2.** Using the initial values of $p_j^{(0)}$, $\eta_j^{(0)}$ & $\beta_j^{(0)}$ to calculate the $k$ -th conditional expectation of $z_{ij}$, i.e., $z_{ij}^{(k)}$ using eq. (3.13)

**3.** In iteration $k$, find the MLEs of $p_j^{(k+1)}$, $\eta_j^{(k+1)}$ & $\beta_j^{(k+1)}$ as follows:

**a.** Find the MLE for $p_j^{(k+1)}$ using eq. (3.15).

**b.** Use eq.(3.19); calculate $\eta_j^{(k+1)}$.

**c.** Use eq.(3.20); calculate $\beta_j^{(k+1)}$.

**4.** Repeat steps 2 & 3 until the desired accuracy is reached.

Although the principal reasons for the popularity of the EM algorithm are its easy implementation and stable convergence, various attempts have been made to speed it up, for it can converge quite slowly in some applications.Details on the application of EM algorithm for mixture models with censored data can be found in Ateya (2012), Bordes and Chauveau (2012) and Ruhi, et al. (2015).

### 3.3.2.3 Applications of EM Algorithm

EM algorithm is frequently used for –

- Data clustering (the assignment of a set of observations into subsets, called clusters, so that observations in the same cluster are similar in some sense) used in many fields, including machine learning, computer vision, data mining, pattern recognition, image analysis, information retrieval, and bioinformatics.
- Natural language processing (NLP is a field of computer science and linguistics concerned with the interactions between computers and human (natural) languages).
- Psychometrics (the field of study concerned with the theory and technique of educational and psychological measurement, which includes the measurement of knowledge, abilities, attitudes, and personality traits.)
- Medical image reconstruction, especially in positron emission tomography (PET) and single photon emission computed tomography (SPECT).
- Multivariate Data analysis with Missing Values.
- Analysis of Least Squares with Missing Data.
- Multinomial with Complex Cell Structure.

- Analysis of PET and SPECT Data.
- Analysis of Mixture distributions.
- Analysis of Grouped, Censored and Truncated Data.

#### 3.3.2.4 Advantages of EM Algorithm

- The EM algorithm is numerically stable, with each EM iteration increasing the likelihood
- Under fairly general conditions, the EM algorithm has reliable global convergence (depends on initial value and likelihood!). Convergence is nearly always to a local maximizer
- The EM algorithm is typically easily implemented, because it relies on complete data computations
- The EM algorithm is generally easy to program, since no evaluation of the likelihood nor its derivatives is involved
- The EM algorithm requires small storage space and can generally be carried out on a small computer (it does not have to store the information matrix nor its inverse at any iteration)
- The M-step can often be carried out using standard statistical packages in situations where the complete-data MLEs do not exist in closed form
- By watching the monotone increase in likelihood over iterations, it is easy to monitor convergence and programming errors
- The EM algorithm can be used to provide estimated values of the missing data

## 3.4 Model Selection Criterion

A statistical hypothesis test is a method using observed samples to draw a statistical conclusion. Generally, it involves a null hypotheses and an alternative hypothesis about the distributions of the observations or about some statistical property (e.g., trend or independence).

In this section we focus on a number of model selection criterions for selecting the suitable models for a data set among a set of competitive models. Most statistical methods assume an underlying distribution in the derivation of their results. However, when we assume that our data follow a specific distribution, we take a serious risk. If our assumption is wrong, then the results obtained may be invalid.

There are two main approaches to checking distribution assumptions. One involves empirical procedures, which are easy to understand and implement and are based on intuitive and graphical properties of the distribution that we want to assess. Empirical procedures can be used to check and validate distribution assumptions.

There are also other, more formal, statistical procedures for assessing the underlying distribution of a data set. These are the goodness of fit (GoF) tests. They are numerically convoluted and usually require specific software to perform the lengthy calculations. But their results are quantifiable and more reliable than those from the empirical procedure. Here, we are interested in Akaike Information Criterion (AIC), Anderson-Darling (AD) test statistic, the Kolmogorov-Smirnov (KS) test statistic and the root mean squareerror (RMSE).

### 3.4.1 Akaike Information Criterion

Akaike's information criterion (AIC), developed by Hirotsugu Akaike under the name of ‘an information criterion’ (AIC) in 1971, is a measure of the goodness of fit of an estimated statistical model . This is the most widely used criterion for selecting the best model for a given data set.In the general case, the AIC is

$$AIC = 2k - 2\ln(L) \quad (3.21)$$

where$k$ is the number of parameters in the statistical model, and $L$ is the maximized value of the likelihood function for the estimated model.

AIC not only rewards goodness of fit, but also includes a penalty that is an increasing function of the number of estimated parameters. This penalty discourages overfitting. The preferred model is the one with the lowest AIC value. The AIC methodology attempts to find the model that best explains the data with a minimumoffree parameters. By contrast, more traditional approaches to modeling start from a null hypothesis. The AIC penalizes free parameters less strongly than does the Schwarz criterion.

The AIC is not a test of the model in the sense of hypothesis testing; rather it is a test between models-a tool for model selection. Given a data set, several competing models may be ranked according to their AIC, with the one having the lowest AIC being the best. From the AIC value one may infer that e.g. the top three models are in

a tie and the rest are far worse, but it would be arbitrary to assign a value above which a given model is 'rejected'.

### 3.4.2 Anderson-Darling Test Statistic

The Anderson-Darling (AD) test (Stephens, 1974) is used to test if a sample of data came from a population with a specific distribution. The Anderson-Darling test makes use of the specific distribution in calculating critical values. This has the advantage of allowing a more sensitive test and the disadvantage that critical values must be calculated for each distribution. Currently, tables of critical values are available for the normal, lognormal, exponential, Weibull, extreme value type I, and logistic distributions.

The Anderson - Darling test statistic is defined by

$$A^2 = -N - S$$

where

$$S = \sum_{i=1}^{N} \frac{2i-1}{N} \left[\ln F(Y_i) + \ln\left(1 - F(Y_{N+1-i})\right)\right]$$

$F$ is the cumulative distribution function of the specified distribution. Note that the $Y_i$ are the ordered data.

The critical values for the Anderson-Darling test are dependent on the specific distribution that is being tested. Tabulated values and formulas have been published (Stephens, 1974, 1976, 1977, 1979) for a few specific distributions (normal, lognormal, exponential, Weibull, logistic, extreme value type 1). The test is a one-sided test and the hypothesis that the distribution is of a specific form is rejected if the test statistic, $A$, is greater than the critical value.

### 3.4.3 Adjusted Anderson Darling Test Statistic

For a given distribution, the Anderson-Darling statistic may be multiplied by a constant (which usually depends on the sample size, n). These constants are given in the various papers by Stephens. In the sample output below, this is the adjusted Anderson-Darling (AD*) statistic. This is what should be compared against the critical values. Also, be aware that different constants (and therefore critical values) have been published. One just needs to be aware of what constant was used for a

given set of critical values (the needed constant is typically given with the critical values).

Minitab calculates an AD* statistic for the distribution ID plot and for reliability/ survival analysis. These AD* statistic are equivalent and are represented in the output as Anderson-Darling (adj) or AD*.

AD* is used because *p*-values for the Anderson-Darling statistic could not be calculated for multiply censored or arbitrary censored data. Unlike the standard Andersn-Darling statistic, the AD* is generalized to account for different plot-point methods the user can choose for constructing the probability plot.

Using the plot points and the probability integral transformation described in D'Agostino and Stephens (1986), Minitab calculates the AD* as:

$$AD^* = n\sum_{i=1}^{n+1}(A_i + B_i + C_i) \tag{3.22}$$

Here $A_i = -Z_i - \ln(1 - Z_i) + Z_{i-1} + \ln(1 - Z_{i-1})$

$$B_i = 2\ln(1 - Z_i)F_n(Z_{i-1}) - 2\ln(1 - Z_{i-1})F_n(Z_{i-1})$$

$$C_i = \ln(Z_i)F_n(Z_{i-1})^2 - \ln(1 - Z_i)F_n(Z_{i-1})^2 - \ln(Z_{i-1})F_n(Z_{i-1})^2 +$$

$$\ln(1 - Z_{i-1})F_n(Z_{i-1})^2$$

$Z_i$is the fitted estimate of the cdf for the *i*-th plot point

$F_n(Z_i)$is the non-parametric estimate of cdf for the *i*-th plot point

$$Z_0 = 0$$

$$F_n(Z_0) = 0$$

$$Z_{n+1} = 1 - (1E - 12)$$

The AD* test statistic provides arelative measure of GoF. When comparing the GoF of multiple distributions for a given data set, the distribution with the smallest AD* offers the best fit. This comparative technique is valid only when comparing the fit of multiple distributions for a single data set.

### 3.4.4 Kolmogorov–Smirnov Test Statistic

Consider a complete dataset. The Kolmogorov–Smirnov test statistic (KS test) is the maximumdifference between the empirical cdf and theoretical cdf given by

$$D_n = \underset{1\le i\le n}{max}\left\{max\left(\left|F(t)-\frac{i}{n}\right|,\left|F(t)-\frac{i-1}{n}\right|\right)\right\}$$

If the sample comes from distribution $F(t)$, then $D_n$ will be sufficiently small. The null hypothesis is rejected at level $\alpha$, if $D_n > k_\alpha$, where $k_\alpha$ is the critical value at significance level of $\alpha$. According to Jiang (2015) the critical value of the test statistic can be approximated by

$$k_c = \frac{a}{\sqrt{n}}\left(1-\frac{b}{n^c}\right) \tag{3.23}$$

The coefficient set (*a,b,c*) is given in Table 3.1.

Table 3.1: Values of *a, b* and *c* at different values of $\alpha$

| $\alpha$ | 0.1 | 0.05 | 0.01 |
|---|---|---|---|
| a | 1.224 | 1.358 | 1.628 |
| b | 0.2057 | 0.2593 | 0.3753 |
| c | 0.6387 | 0.7479 | 0.8858 |

The AD statistic is a modification of the KS test statistic and gives more weight to the tails than does the KS test. The KS test is distribution free in the sense that the critical values do not depend on the specific distribution being tested (note that this is true only for a fully specified distribution, i.e. the parameters are known). But the AD test makes use of the specific distribution in calculating critical values. The AD test is an alternative to the chi-square and KS GoF tests.

### 3.4.5 Root Mean Square Error

The rootmeansquareerror (RMSE) or rootmeansquare deviation (RMSD) is a frequently used measure of the differences between values (sample and population values) predicted by a model or an estimator and the values actually observed. The RMSE represents the sample standard deviation of the differences between predicted values and observed values.

The RMSE of an estimator $\hat{\theta}$with respect to an estimated parameter $\theta$ is defined as the square root of the mean square error:

$$RMSE(\hat{\theta}) = \sqrt{MSE(\hat{\theta})} = \sqrt{E\left((\hat{\theta} - \theta)^2\right)} \tag{3.24}$$

For an unbiased estimator, the RMSE is the square root of the variance, known as the standard deviation.

# Chapter 4
# Product Reliability Data

## 4.1 Introduction

In recent years many manufacturers have collected and analyzed field reliability data to assess the quality and reliability of their products and to improve customer satisfaction. There are many sources of reliability- related data of a product. To analyze and model, reliability data are mainly collected from the laboratory or field, and sometimes from the published literature and experts' judgments. The test data are observed and obtained under controlled conditions and the field data are usually recorded and stored in a management information system. Warranty claim data is used as an important source of field failure data which can be collected economically and efficiently through repair service networks and therefore, a number of procedures have been developed for collecting and analyzing warranty claim data (e.g. Karim and Suzuki, 2005; Karim et al., 2001; Lawless, 1998; Murthy and Djamaludin, 2002; Suzuki, 1985a,b; Suzuki et al., 2001.

## 4.2 Field Reliability Data

Manufacturer use the field information as a feedback to learn about the reliability problems of a product and to improve the quality of future generations of the same or similar products. Consumers can be used the field information to optimize the maintenance activities and spare part inventory control policy. Many enterprises use a management information system to store the maintenance-related information. Most of such systems are designed for the purpose of management rather than reliability analysis. As a result, the records are often ambitious and some essential information useful for reliability analysis is missed.

Failure data may be obtained from a reliability or life test led in a controlled environment, the purpose of which is to operate units to failure in order to obtain data for reliability analysis. Preferably, to gain a set of complete data, where all of the units put on the test should be operated until they fail. Sometimes this is not possible due to time and budgetary restrictions and there will be accumulated test time for units that did not fail. But since, in field failure data, the units under analysis were operated under actual use conditions. Hence, while analyzing field failure data, the problem of operating conditions is not a matter of concern. Again field failure data is superior to laboratory test data in the sense that it contains valuable information on the performance of a product in actual usage conditions.

## 4.3 Limitations of Field Reliability Data

One of the drawbacks of field reliability data is that it may consist primarily of suspended data. Another one of the drawbacks of field reliability data is that it may be contaminated or incomplete. For example, many times field data gained for reliability analysis may have originally been collected for another purpose, such as financial warranty purposes. In some cases, the data may not contain all of the necessary information, required to achieve a good reliability analysis. Also, there may be large portions of essential information missing, that is, large segments of the field population which are unaccounted for. Have they failed? How long have they been running? Are they still in operation? The answers to these questions are very important to analyze the field data and if this information cannot be delivered for a large segment of the product's population, a field data analysis may provide erroneous results. It is generally a good idea to involve a reliability professional to develop the field data collection systems in order to avoid some of these drawbacks.

This thesis analyses the following three data sets for estimating and predicting product reliability.

- Data Set 1: Aircraft windshield failure data (secondary data)
- Data Set 2: Battery failure data (primary data)
- Data Set 3: Hydraulic pump failure data (secondary data)

The following three sections of this chapter present these three data sets.

## 4.4 Data Set 1: Aircraft Windshield Failure Data

The windshield on a large aircraft is a complex piece of equipment, comprised basically of several layers of material. Aircraft windshield contains a very strong outer skin with a heated layer just beneath it, all coated under high temperature and pressure. Failures of the items are not structural failures. Instead, they typically involve damage or delamination of the nonstructural outer ply or failure of the heating system. These failures do not damage the aircraft but need replacement of the windshield.

All the windshield data are routinely collected and analyzed. At any specific point in time, these data will include failures to date of a particular model as well as service/censored times of all items that have not failed. These types of data are known as incomplete data in the sense that not all failure times have as yet been observed. Data on failure and service times for a particular model windshield are given in Table 4.1taken from Murthy, Xie and Jiang (2004), originally given in Blischke and Murthy (2000). The data consist of 153 observations of which 88 are classified as failed windshields, and the remaining 65 are service time (censored time) of windshields that had not failed at the time of observation. The unit for measurement is 1000h.

Table 4.1: Aircraft Windshield Failure Data

| Failure Times | | | | Service Times | | |
|---|---|---|---|---|---|---|
| 0.04 | 1.866 | 2.385 | 3.443 | 0.046 | 1.436 | 2.592 |
| 0.301 | 1.876 | 2.481 | 3.467 | 0.14 | 1.492 | 2.6 |
| 0.309 | 1.899 | 2.61 | 3.478 | 0.15 | 1.58 | 2.67 |
| 0.557 | 1.911 | 2.625 | 3.578 | 0.248 | 1.719 | 2.717 |
| 0.943 | 1.912 | 2.632 | 3.595 | 0.28 | 1.794 | 2.819 |
| 1.07 | 1.914 | 2.646 | 3.699 | 0.313 | 1.915 | 2.82 |
| 1.124 | 1.981 | 2.661 | 3.779 | 0.389 | 1.92 | 2.878 |
| 1.248 | 2.01 | 2.688 | 3.924 | 0.487 | 1.963 | 2.95 |
| 1.281 | 2.038 | 2.823 | 4.035 | 0.622 | 1.978 | 3.003 |
| 1.281 | 2.085 | 2.89 | 4.121 | 0.9 | 2.053 | 3.102 |
| 1.303 | 2.089 | 2.902 | 4.167 | 0.952 | 2.065 | 3.304 |
| 1.432 | 2.097 | 2.934 | 4.24 | 0.996 | 2.117 | 3.483 |
| 1.48 | 2.135 | 2.962 | 4.255 | 1.003 | 2.137 | 3.5 |

Continued…

| Failure Times | | | | Service Times | | |
|---|---|---|---|---|---|---|
| 1.505 | 2.154 | 2.964 | 4.278 | 1.01 | 2.141 | 3.622 |
| 1.506 | 2.19 | 3 | 4.305 | 1.085 | 2.163 | 3.665 |
| 1.568 | 2.194 | 3.103 | 4.376 | 1.092 | 2.183 | 3.695 |
| 1.615 | 2.223 | 3.114 | 4.449 | 1.152 | 2.24 | 4.015 |
| 1.619 | 2.224 | 3.117 | 4.485 | 1.183 | 2.341 | 4.628 |
| 1.652 | 2.229 | 3.166 | 4.57 | 1.244 | 2.435 | 4.806 |
| 1.652 | 2.3 | 3.344 | 4.602 | 1.249 | 2.464 | 4.881 |
| 1.757 | 2.324 | 3.376 | 4.663 | 1.262 | 2.543 | 5.14 |
| 1.795 | 2.349 | 3.385 | 4.694 | 1.36 | 2.56 | |

## 4.5 Data Set 2: Battery Failure Data

This data set represents the failure of the battery used in two different products, IPS and private cars,consist of both failure and censored lifetimes of the battery. This data set is primary data and collected from the users in Rajshahi region. The IPSs were used in residence or offices. Some batteries were maintained regularly and some were not maintained. The batteries of this data set are from different manufacturing companies broadly characterized into three categories denoted by ‘R’, ‘L’ and ‘O’.The information regarding the names of the manufacturing companies are not disclosed here to protect proprietary nature of the information.

The data set gives 192 observations of the battery and presented in Table 4.2. As can be seen the data consists of 107 failure data and 85 censored data. The column, characterized ‘Time’ indicates the age (in months) of the item at the data collection period. The column ‘Type’ specifies whether the data is a failure (denoted by 1) or censored data (denoted by 0). The column ‘Maint.info’ indicates that whether the battery was maintained regularly or not, here ‘1’ means the battery was maintained and ‘0’ means the battery was not maintained. Among the 192 observations 150 are found as maintained regularly by the user and 42 observations are found as non-maintained. The column labeled ‘Product’presents the products in which the battery was used and finally the column labeled ‘Brand’indicates the name of the manufacturing company of the battery.

Table 4.2: Battery Failure Data

| Time | Type | Maint. info | Product | Brand | Time | Type | Maint. info | Product | Brand |
|---|---|---|---|---|---|---|---|---|---|
| 2 | 0 | 0 | IPS | L | 24 | 0 | 1 | IPS | L |
| 3 | 0 | 1 | IPS | O | 24 | 1 | 1 | IPS | L |
| 3 | 0 | 1 | IPS | O | 24 | 1 | 1 | IPS | L |
| 4 | 0 | 0 | Car | R | 24 | 0 | 1 | IPS | L |
| 5 | 0 | 1 | Car | R | 24 | 0 | 1 | IPS | R |
| 6 | 0 | 0 | IPS | O | 24 | 0 | 1 | IPS | O |
| 6 | 0 | 1 | IPS | R | 24 | 0 | 1 | IPS | O |
| 6 | 1 | 1 | IPS | L | 24 | 1 | 1 | IPS | O |
| 6 | 0 | 0 | IPS | L | 24 | 0 | 1 | IPS | R |
| 6 | 0 | 0 | Car | R | 24 | 0 | 1 | IPS | R |
| 6 | 1 | 0 | Car | O | 24 | 0 | 1 | IPS | R |
| 6 | 0 | 0 | Car | O | 24 | 0 | 1 | IPS | L |
| 6 | 0 | 0 | Car | R | 24 | 1 | 1 | IPS | O |
| 6 | 1 | 0 | IPS | L | 24 | 0 | 1 | IPS | R |
| 7 | 0 | 1 | IPS | L | 24 | 1 | 1 | IPS | R |
| 7 | 0 | 1 | IPS | O | 24 | 0 | 1 | IPS | R |
| 7 | 0 | 0 | IPS | L | 24 | 1 | 1 | IPS | O |
| 7 | 0 | 0 | Car | R | 24 | 1 | 1 | IPS | L |
| 7 | 0 | 0 | IPS | L | 24 | 0 | 1 | IPS | O |
| 7 | 0 | 0 | Car | R | 24 | 1 | 1 | IPS | O |
| 7 | 1 | 0 | IPS | L | 24 | 0 | 1 | IPS | R |
| 8 | 1 | 1 | IPS | R | 24 | 1 | 1 | Car | R |
| 8 | 0 | 1 | IPS | R | 25 | 0 | 1 | IPS | L |
| 8 | 0 | 1 | IPS | O | 27 | 0 | 1 | Car | R |
| 8 | 0 | 0 | IPS | L | 29 | 1 | 1 | IPS | L |
| 8 | 0 | 1 | Car | R | 30 | 1 | 1 | IPS | L |
| 8 | 0 | 1 | Car | O | 30 | 1 | 1 | IPS | O |
| 8 | 1 | 0 | IPS | L | 30 | 1 | 1 | IPS | L |
| 9 | 0 | 0 | IPS | O | 30 | 1 | 1 | IPS | L |
| 9 | 0 | 1 | IPS | L | 30 | 0 | 1 | IPS | L |
| 9 | 1 | 1 | Car | R | 30 | 1 | 1 | IPS | R |
| 10 | 0 | 1 | IPS | O | 30 | 0 | 1 | IPS | L |
| 10 | 0 | 1 | IPS | R | 30 | 0 | 1 | IPS | R |
| 10 | 0 | 1 | IPS | L | 30 | 1 | 1 | IPS | R |
| 10 | 0 | 0 | IPS | L | 30 | 0 | 1 | IPS | O |
| 10 | 0 | 1 | Car | R | 30 | 1 | 1 | IPS | O |
| 10 | 1 | 0 | IPS | L | 30 | 1 | 1 | IPS | R |
| 10 | 1 | 0 | IPS | L | 30 | 0 | 1 | IPS | R |

Continued…

| Time | Type | Maint. info | Product | Brand | Time | Type | Maint. info | Product | Brand |
|---|---|---|---|---|---|---|---|---|---|
| 10 | 0 | 0 | Car | R | 30 | 0 | 1 | IPS | O |
| 11 | 0 | 1 | Car | R | 30 | 1 | 1 | IPS | L |
| 11 | 0 | 0 | Car | R | 30 | 1 | 1 | IPS | O |
| 11 | 1 | 0 | IPS | L | 30 | 1 | 1 | IPS | O |
| 11 | 1 | 1 | Car | R | 30 | 1 | 1 | IPS | R |
| 11 | 1 | 0 | IPS | L | 30 | 1 | 1 | IPS | L |
| 11 | 1 | 0 | IPS | L | 30 | 1 | 1 | Car | R |
| 12 | 0 | 1 | IPS | R | 31 | 1 | 1 | IPS | L |
| 12 | 1 | 1 | IPS | O | 31 | 1 | 1 | IPS | O |
| 12 | 0 | 1 | IPS | O | 33 | 0 | 1 | IPS | L |
| 12 | 1 | 0 | IPS | O | 33 | 1 | 1 | Car | R |
| 12 | 0 | 1 | IPS | R | 34 | 1 | 1 | Car | R |
| 12 | 1 | 1 | IPS | L | 34 | 1 | 1 | IPS | L |
| 12 | 0 | 1 | IPS | L | 36 | 1 | 1 | IPS | L |
| 12 | 0 | 1 | IPS | R | 36 | 0 | 1 | IPS | R |
| 12 | 0 | 0 | IPS | O | 36 | 1 | 1 | IPS | O |
| 12 | 1 | 0 | IPS | R | 36 | 1 | 1 | IPS | R |
| 12 | 0 | 0 | IPS | L | 36 | 1 | 1 | IPS | O |
| 12 | 0 | 1 | IPS | L | 36 | 1 | 1 | IPS | O |
| 12 | 0 | 1 | IPS | O | 36 | 0 | 1 | IPS | R |
| 12 | 0 | 1 | IPS | R | 36 | 1 | 1 | IPS | R |
| 12 | 1 | 0 | Car | R | 36 | 1 | 1 | IPS | R |
| 12 | 0 | 1 | Car | R | 36 | 1 | 1 | IPS | L |
| 12 | 0 | 1 | IPS | L | 36 | 0 | 1 | IPS | R |
| 12 | 1 | 0 | IPS | L | 36 | 1 | 1 | IPS | L |
| 13 | 1 | 0 | IPS | L | 36 | 1 | 1 | IPS | L |
| 13 | 1 | 0 | Car | R | 36 | 0 | 1 | IPS | O |
| 13 | 0 | 0 | IPS | L | 36 | 1 | 1 | IPS | R |
| 13 | 1 | 0 | IPS | L | 36 | 1 | 1 | IPS | O |
| 13 | 1 | 0 | IPS | L | 36 | 1 | 1 | IPS | R |
| 14 | 0 | 1 | IPS | L | 36 | 1 | 1 | IPS | R |
| 14 | 0 | 1 | IPS | L | 36 | 1 | 1 | IPS | R |
| 14 | 1 | 0 | Car | R | 36 | 1 | 1 | Car | R |
| 15 | 1 | 0 | IPS | L | 38 | 0 | 1 | Car | O |
| 15 | 0 | 1 | IPS | L | 40 | 1 | 1 | Car | R |
| 15 | 1 | 0 | IPS | O | 42 | 0 | 1 | IPS | R |
| 15 | 1 | 0 | IPS | L | 42 | 1 | 1 | IPS | R |
| 17 | 0 | 1 | Car | R | 42 | 1 | 1 | IPS | R |
| 18 | 0 | 1 | IPS | R | 42 | 1 | 1 | IPS | L |
| 18 | 1 | 0 | IPS | L | 42 | 1 | 1 | IPS | R |
| 18 | 1 | 1 | IPS | R | 42 | 1 | 1 | IPS | O |

Continued…

| Time | Type | Maint. info | Product | Brand | Time | Type | Maint. info | Product | Brand |
|---|---|---|---|---|---|---|---|---|---|
| 18 | 1 | 1 | IPS | O | 42 | 0 | 1 | IPS | R |
| 18 | 1 | 1 | IPS | O | 42 | 1 | 1 | IPS | R |
| 18 | 1 | 1 | IPS | O | 42 | 1 | 1 | IPS | O |
| 18 | 0 | 1 | IPS | O | 42 | 1 | 1 | IPS | R |
| 18 | 1 | 1 | IPS | R | 42 | 1 | 1 | IPS | R |
| 19 | 1 | 1 | Car | R | 42 | 1 | 1 | IPS | L |
| 19 | 1 | 1 | Car | R | 42 | 1 | 1 | IPS | L |
| 20 | 0 | 0 | IPS | O | 42 | 1 | 1 | IPS | O |
| 20 | 1 | 1 | IPS | O | 44 | 1 | 1 | IPS | L |
| 20 | 1 | 1 | IPS | L | 45 | 1 | 1 | IPS | L |
| 20 | 1 | 1 | Car | R | 50 | 1 | 1 | Car | R |
| 21 | 0 | 1 | IPS | L | 52 | 1 | 1 | IPS | L |
| 22 | 0 | 1 | IPS | R | 53 | 1 | 1 | IPS | O |
| 23 | 1 | 1 | IPS | R | 53 | 1 | 1 | IPS | L |
| 23 | 0 | 1 | IPS | R | 55 | 1 | 1 | Car | O |
| 23 | 1 | 1 | Car | R | 56 | 1 | 1 | Car | O |
| 24 | 0 | 1 | IPS | O | 76 | 1 | 1 | Car | O |

## 4.6 Data Set 3: Hydraulic Pump Failure Data

This data set deals with the maintenance of Hydraulic pumps used in Excavators by a mining company. The data had been collected by the owner (mining company) and carried out an analysis and building models for pump failures. The data given in Murthy, Karim and Ahmadi (2015) and Karim, Ahmadi and Murthy (2015) contain of both failure and censored lifetimes of the pump.

The hydraulic pumps considered here used in excavators by a mining company. In open cut mines, coal and overburden are transported using excavators and dump trucks. Hydraulic system is one of the important among the several systems that included in an excavator. The hydraulic system is comprised of several hydraulic pumps (for linear and rotational motions), hydraulic oil filters and several hydraulic lines. A pump is known as failure, if it cannot provide the required flow rate at the specified pressure. The data have recorded by the maintenance department, consist of the failure times (for units that have failed and required Corrective Maintenance action) and service times (for units that have not failed yet and sent for Preventive Maintenance action) for 102 units and presented in Table 4.3. As can be seen the data

consists of 45 failure and 57 censored observations. The column, labeled ‘Age’ means the age (in hours) of the item at the end of the data collection period and the column labeled ‘Type’ indicates whether the data is a failure data (denoted by 1) or censored data (denoted by 0). More detail description of the data can be found in Murthy, et al. (2015) and Karim, et al. (2015).

### 4.6.1 Pump Failures

A pump is considered to have failed if it cannot provide the required flow rate at the necessary pressure. Pump failure is detected by sensors and relayed to the operator. The pump failure is happened because of failure of one or more components of the pump. There can be one or more failure modes for each component and several causes leading to the failure.

### 4.6.2 Pump Maintenance

The mine operates 3 identical excavators on site with 2 engines per excavator and 4 hydraulic pumps (variable displacement axial piston pumps) per engine. The mine has a small maintenance department which carries out the PM and manages the outsourcing of pumps for CM actions (when a pump failure occurs) and PM actions involving the overhaul of pumps.

The mining company used an age based policy for pump maintenance. Under this policy a pump is subjected to a replacement (PM action) after being in operation for specified period ($T$ hours) or on failure (CM action) should it occur earlier. The pump used in the replacement may be either new or reconditioned. Based on the condition of the pump removed (under either PM or CM) it is either scrapped or subjected to an overhaul which results in a reconditioned pump. The general accepted notion is that a reconditioned pump is as-good-as a new pump. The maintenance was outsourced to a maintenance service agent (Karim, et al., 2015).

Table 4.3: Hydraulic Pump Failure Data

| Age (hrs) | Type | Age (hrs) | Type | Age (hrs) | Type | Age (hrs) | Type |
|---|---|---|---|---|---|---|---|
| 81 | 0 | 3333 | 1 | 9334 | 1 | 12198 | 0 |
| 149 | 1 | 3569 | 1 | 9368 | 1 | 12198 | 0 |
| 245 | 1 | 3837 | 0 | 9729 | 1 | 12198 | 0 |
| 340 | 1 | 3837 | 0 | 9751 | 0 | 12198 | 0 |
| 407 | 1 | 4150 | 0 | 10299 | 1 | 12236 | 0 |
| 461 | 1 | 5123 | 1 | 10389 | 0 | 12236 | 0 |
| 629 | 1 | 5258 | 1 | 10413 | 0 | 12236 | 0 |
| 856 | 0 | 5662 | 0 | 10557 | 1 | 12236 | 0 |
| 947 | 0 | 5923 | 1 | 10944 | 1 | 12236 | 0 |
| 1460 | 1 | 6333 | 1 | 10970 | 1 | 12236 | 0 |
| 1513 | 1 | 6717 | 1 | 11647 | 0 | 12394 | 0 |
| 1670 | 1 | 7207 | 1 | 11678 | 1 | 12459 | 0 |
| 1688 | 0 | 7265 | 1 | 11686 | 1 | 13097 | 0 |
| 2093 | 0 | 7624 | 1 | 11798 | 0 | 13497 | 0 |
| 2242 | 0 | 7625 | 0 | 11869 | 0 | 13497 | 0 |
| 2242 | 0 | 7973 | 1 | 11869 | 0 | 13497 | 0 |
| 2242 | 0 | 8183 | 1 | 11923 | 0 | 13497 | 0 |
| 2242 | 0 | 8217 | 1 | 12005 | 0 | 13497 | 0 |
| 2242 | 0 | 8390 | 1 | 12082 | 0 | 13497 | 0 |
| 2607 | 1 | 8462 | 1 | 12090 | 0 | 13497 | 0 |
| 2668 | 1 | 8728 | 1 | 12136 | 0 | 14407 | 1 |
| 2806 | 1 | 8817 | 1 | 12141 | 0 | 15536 | 1 |
| 3132 | 0 | 8870 | 1 | 12143 | 0 | 16289 | 1 |
| 3132 | 0 | 8884 | 0 | 12163 | 0 | 17517 | 1 |
| 3132 | 0 | 9055 | 1 | 12198 | 0 | | |
| 3132 | 0 | 9182 | 1 | 12198 | 0 | | |

# Chapter 5
# Data Analysis

## 5.1 Introduction

As discussed before, product reliability data collected from the field can have a number of special features related to censoring, non-homogeneity of the population, and lack of information on explanatory variables and multiple failure modes. These special features lead to difficulties in applying ordinary statistical models and methods for data analysis and thus they require the use of complex statistical models and special statistical methods.

This chapter analyses the three data sets (presented in previous Chapter) for estimating and predicting products reliabilities.

## 5.2 Aircraft Windshield Failure Data Analysis

Murthy, Xie and Jiang (2004) proposed the 2-fold Weibull mixture model for the Windshield data. They indicated that WPPPlot method give good estimate of the model parameters. We have applied the EM algorithm, to find the maximum likelihood estimates of the parameters for the 2-fold Weibull mixture model and investigate the performance of the proposed method over the method of Murthy et al. (2004).

The aims of the analysis are to check whether the graphical estimation method or maximum likelihood estimation method fit the data well and to estimate the reliability of the Windshield. R programming codes are written for the computations of the MLE of the model parameters by using EM algorithm of this data set. Programming codes for analyzing the data with 2-fold Weibull mixture model are given in section A.1 in the Appendix. The given codes can be used for other 2-fold mixture models after simple modifications.

### 5.2.1 Nonparametric Estimate of Reliability Function

Figure 5.1 is the reliability (or survival) plot for the component. The plot appears to be reasonable; it shows the estimated MTTF is 3.03549 thousand hours or approximately 127 days. The nonparametric estimate of median lifetime is 2964 h, indicates that 50 % of the Windshield fails at 2964 h. The nonparametric estimate of cdf, known as empirical distribution function (edf) is one minus the estimated reliability function.

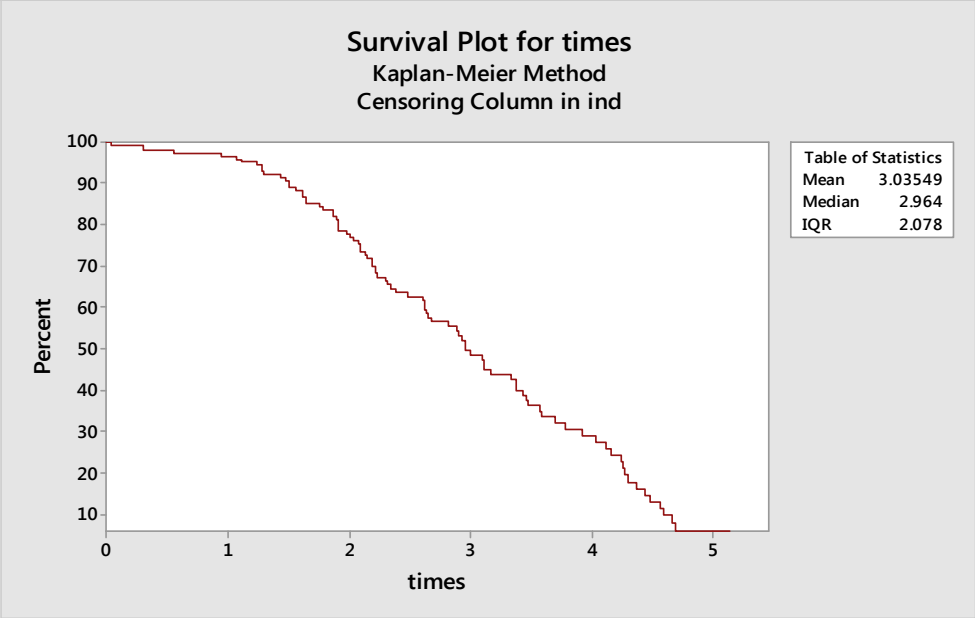


Figure 5.1: Non-parametric reliability plots of Windshield failure data

### 5.2.2 Parametric Estimate of Reliability Function

Murthy et al. (2004) assumed the 2-fold Weibull mixture model for this data set and estimated the model parameters based on WPP plot method. In this thesis we have applied the EM algorithm, to find the maximum likelihood estimates of the parameters$\theta = (\beta_1, \eta_1, \beta_2, \eta_2, p, (1-p))$ for the 2-fold Weibull mixture model. 2-fold Weibull mixture model is discussed in Section 2.3.1.1. Also this model can be derived in Case-2 from the general model for quality variation (2.3.3).We investigate the performance of the proposed method over the method of Murthy et al. (2004). A comparison between the estimates of the parameters obtained by two different methods is given in Table 5.1.

Table 5.1: Estimates of parameters of 2-fold Weibull mixture model

| Parameters | Estimates based on WPP | Estimates based on EM algorithm |
|---|---|---|
| $\hat{\beta}_1$ | 0.429 | 1.2390 |
| $\hat{\eta}_1$ | 8.230 | 0.2481 |
| $\hat{\beta}_2$ | 2.990 | 2.7787 |
| $\hat{\eta}_2$ | 3.210 | 3.4852 |
| $\hat{p}$ | 0.136 | 0.0176 |
| $(1-\hat{p})$ | 0.864 | 0.9824 |

Table 5.1 indicates that two methods give reasonably different estimates for the model parameters.

### 5.2.3 Model Selection

This section applies the graphical approach for selecting the best fitted model for the data set. We have estimated the cdfs and reliability functions of 2-fold Weibull mixture model based on both nonparametric (by Kaplan–Meier method) and parametric (by EM Algorithm) approaches. The cdf and reliability function are also estimated by using the WPP plot method [estimates are taken from Murthy et al. (2004)]. Figures 5.2 and 5.3 compare the estimated reliability functions and cdfs, respectively to find out the best approach for the data set.

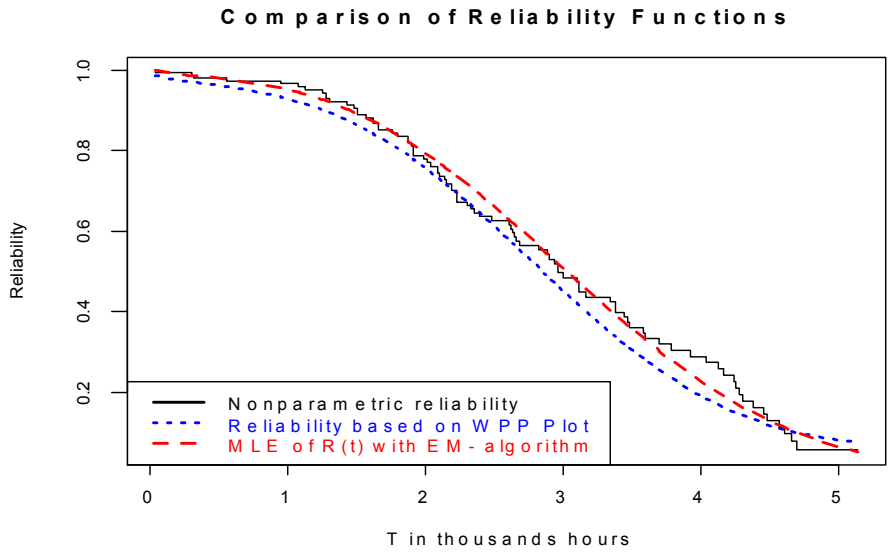


Figure 5.2: Comparison of reliability functionsof Windshield failure data

From Figure 5.2, we observe that the reliability function obtained by the EM algorithm method is much closer to the Kaplan-Meier estimate than that of the reliability function estimated by the WPP plot method. The plots of cdfs shown in Figure 5.3 conclude the same. These indicate that the method of maximum likelihood estimation with the EM algorithm procedure is better than the WPP plot procedure.

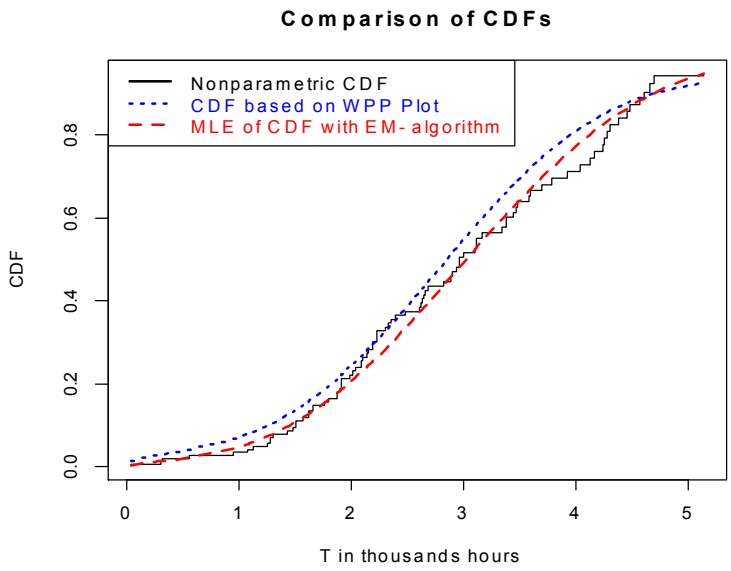


Figure 5.3 Comparisons of cdfs of Windshield failure data

### 5.2.4 Reliability Characteristics

Some of the reliability related important characteristics such as mean time to failure (MTTF), B10 lifetime, B50 (median) lifetime of the Windshield obtained by two methods are displayed in Table 5.2:

Table 5.2:Estimates of reliability characteristics of Windshield

| Quantities | EM algorithm method | WPP Plot method |
|---|---|---|
| MTTF | 3.0519 | 5.5782 |
| B10-Lifetime | 1.5240 | 1.3125 |
| B50-Lifetime | 3.0040 | 2.9298 |

Table 5.2 indicates that the estimates of MTTF obtained from maximum likelihood method via the EM algorithm and from WPP plot method are 3.0519 (thousand hours) and 5.5782 (thousand hours), respectively. Estimate of MTTF obtained by EM

algorithm is very close to the nonparametric estimate of MTTF (3.03549 thousand hours) given in Figure 5.1. The WPP method overestimates the MTTF in this case. From the estimates of B10-lifetime and B50-lifetime, we may conclude according to EM algorithm method that, 10% of the total Windshield fail at 1524 hours and 50% of Windshield fail approximately at 3004 hours. These estimates would be useful to the manufacturer and users for fixing the proper warranty period and/or replacement period of Windshield.

## 5.3 Battery Failure Data Analysis

In the battery failure data set, there is a variable named 'Maint. info' which indicates whether the battery was maintained regularly or not. This information is known for all observations and among 192 observations 150 are found as maintained regularly by the user and 42 observations are found as non-maintained.

To analyze this data set, let us assume following two cases:

- **Case-1:**Information on Maintenance action is known for all observations in the database.
- **Case-2:**Information on Maintenance action is unknown for all observations in the database.

The aim of analysis is to compare the estimated reliabilities of the batteries for two different Cases. That is, what would be the performance of the method if it is assumed that information on maintenance action is not given in the database?

### 5.3.1 Nonparametric Estimates of Reliability Functions

First we estimate the reliability functions for the batteries that were maintained regularly and for the batteries that were not maintained regularly. Figure 5.4 shows the nonparametric estimate of reliability function for batteries that were maintained regularly by the users. The figure shows the nonparametric estimate of MTTF is 35.0772 months (or approximately 1052 days or 2.88 years). The nonparametric estimate of median lifetime is 36 months, indicates that 50 % of the battery fails at approximately 1080 days.

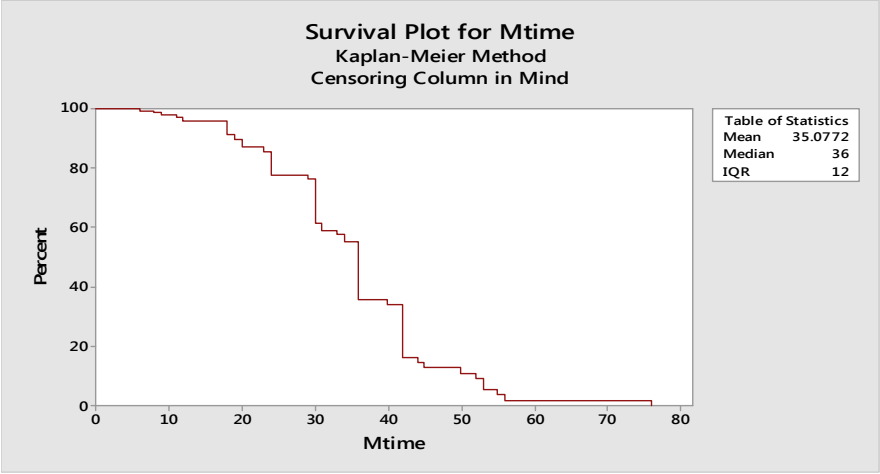


Figure 5.4: Non-parametric reliability plot for regularly maintained batteries

Figure 5.5 shows the nonparametric estimate of reliability function for batteries that were not maintained regularly by the users. The figure shows the estimated nonparametric MTTF is 12.7807 months (or approximately 383 days or 1.05 years). The nonparametric estimate of median lifetime is 13 months, indicates that 50% of the battery fails at approximately 390 days.

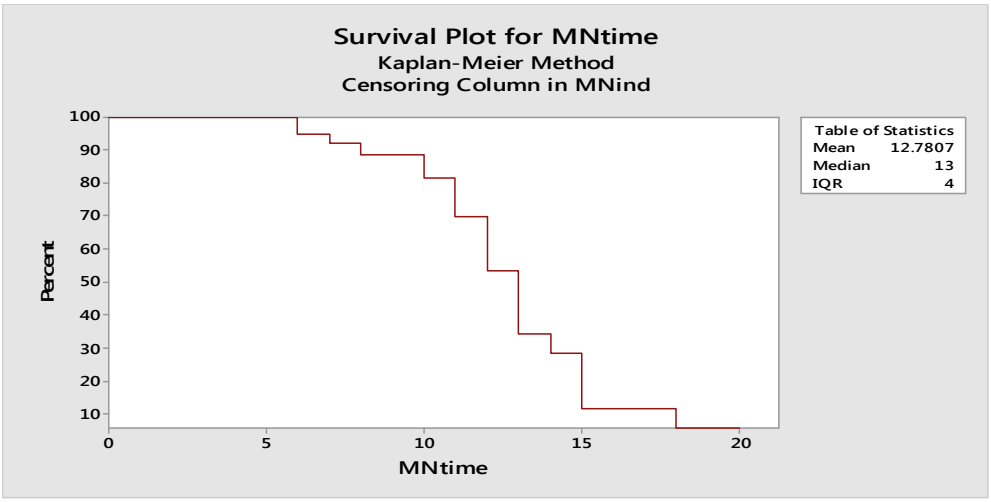


Figure 5.5: Non-parametric reliability plot for non-maintained batteries

As expected, Figures 5.4 and 5.5 indicate that the reliability of the batteries that were maintained regularly is much higher than that of the reliability of the batteries that were not maintained regularly.

### 5.3.2 Parametric Model Selection

This section applies the graphical and statistical approaches for selecting the best fitted models for the data set for two cases, Case-1 and Case-2. First we present the results for Case-1, when maintenance information is known. Under this Case, since the maintenance information is known, the observations can be grouped into two groups – maintained and non-maintained items. Different lifetime models are applied and corresponding cdfs are estimated for both groups. It is found that single Normal and single Weibull distributions give comparatively better fit among all the other models, for both the groups: maintained batteries and non-maintained batteries. The results of these two groups are displayed in Figures 5.6 and 5.7, respectively.

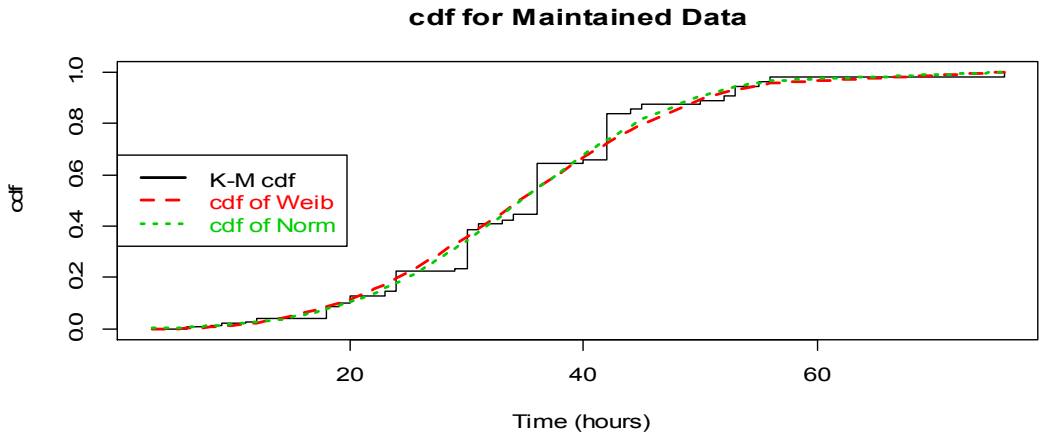


Figure 5.6: Comparison of cdfs for regularly maintained batteries

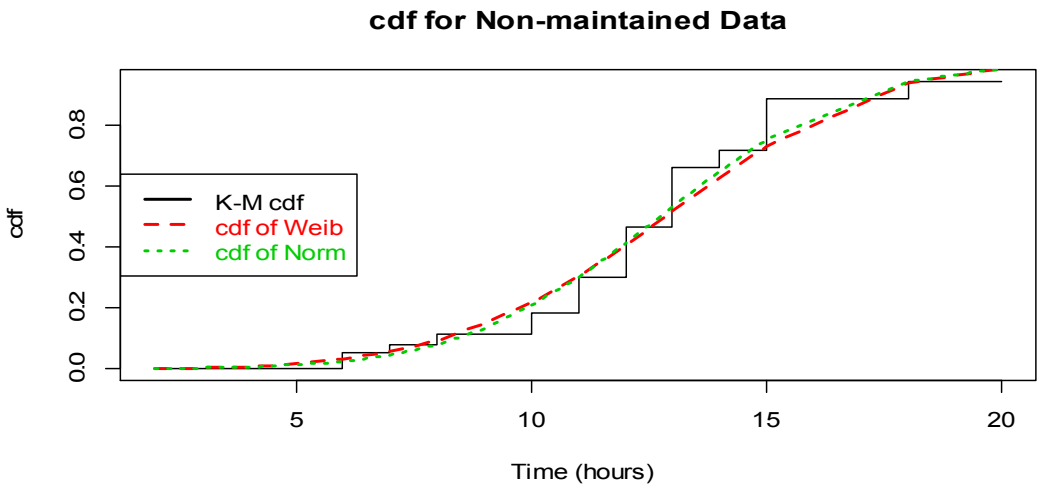


Figure 5.7: Comparison of cdfs for non-maintained batteries

Based on the comparisons with non-parametric cdfs, both of the Figures 5.6 and 5.7 indicate that Weibull or Normal can be considered as the best fitted models.

The estimated values of different model selection criterions, such as, Akaike Information Criterion (AIC), adjusted Anderson Darling (AD*) value and root mean square error (RMSE) for maintained and non-maintained groups are given in Tables 5.3 and 5.4, respectively.

Table 5.3: Estimates ofAIC, AD*and RMSE for maintained items

| Models | AIC | AD* | RMSE |
|---|---|---|---|
| Normal | **699.674** | **1.624** | **0.0339** |
| Weibull | 700.336 | 1.714 | 0.0378 |

From Table 5.3, we found that the Normal distribution contains the lowest values for AIC, AD* and RMSE. Hence Normal distribution can be selected as the best fitted model for the batteries that were maintained regularly.

Table 5.4: Estimates ofAIC, AD*and RMSE for non-maintained items

| Models | AIC | AD* | RMSE |
|---|---|---|---|
| Normal | **131.706** | **5.604** | **0.0539** |
| Weibull | 132.482 | 5.638 | 0.0628 |

Table 5.4 indicates that Normal distribution contains the lowest value for AIC, AD* and RMSE and hencethe Normal distribution can be selected as the best model for the batteries that were not maintained regularly.

By utilizing the above information and based on the ideas of mixture distributions, the cdf of the battery, say $G_1(t)$, can be estimated as

$$G_1(t) = p_1 F_1(t) + p_2 F_2(t) \tag{5.5}$$

Here, the cdf for maintained batteries $F_1(t)$ follows Normal distribution with parameters $(\mu_1, {\sigma^2}_1)$,the cdf for non-maintained batteries$F_2(t)$also follows Normal distribution with parameters $(\mu_2, {\sigma^2}_2)$, $p_1$ is the probability of batteries belong to

maintained subpopulation (=150/192) and $p_2$be the probability of batteries belong to non-maintained subpopulation (= 42/192).

Next we consider modeling for the Case-2, where information on maintenance action is unknown for all observations in the database.To do this, we put $q = 0$ in the general model in eq.(2.60) and obtain the pdf of 2-fold mixture model as given in eq.(2.66). To select the best fitted model for this data, we have applied a group of mixture models (combination of different standard lifetime models). The EM algorithm, discussed in section 3.3.2, is applied for finding the MLEs of the parameters of the 2-fold mixture models. It is observed that the 2-fold Weibull and Weibull-Normal mixture models give comparatively better fit than all the other mixture models. Results are displayed in Figure 5.8 to identify the model that fits best for the data set for Case-2.

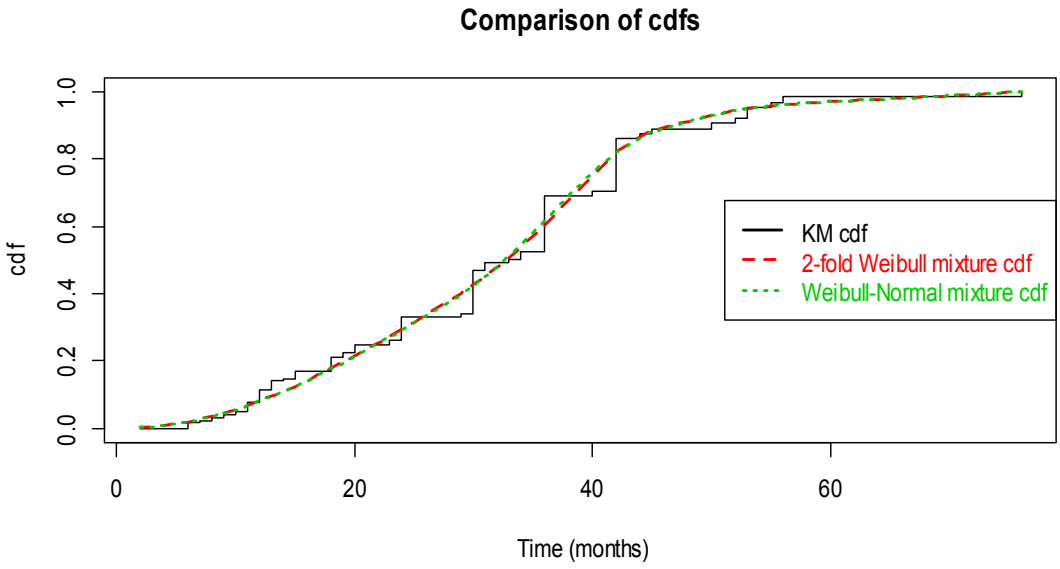


Figure 5.8: Comparison of cdfs when maintained information is unknown

From the above figure 5.8 we can see that both the mixture models (2-fold Weibull & Weibull-Normal mixture) give approximately same result and give a good fit. So, both the models (2-fold Weibull mixture and Weibull-Normal mixture) can be selected as the best fitted model for this data set, when information on maintenance action is unknown.

The estimated values of different model selection criterions are given in Table 5.5 for selecting the best fitted mixture models for Case-2.

Table 5.5: Estimates ofAIC, AD* and RMSE when maintained information unknown

| Models | AIC | AD* | RMSE |
|---|---|---|---|
| 2-fold Weibull mixture | **907.4917** | 0.8585 | 0.0350 |
| Weibull-Normal mixture | 908.0093 | **0.8349** | **0.0344** |

Table 5.5 shows that the 2-fold Weibull mixture distribution has the lowest AIC value and Weibull-Normal mixture model contains the lowest values for AD* and RMSE. Hence, we may conclude that any of the two mixture models can be selected as the best fitted model for Case-2.

### 5.3.3 MLEs of the Parameters

The MLEs of the parameters of the models for Case-1 are displayed in table 5.6. Note that, in this Case, the estimates of $p_1$ and $p_2$ are (150/192= 0.7812) and (42/192= 0.2187), respectively.

Table 5.6: MLEs of the Parameters for Case-1(Maintenance information known)

| Models | MLEs of Parameters |
|---|---|
| *Normal* $(\mu_1, \sigma_1)$ | $(\mu_1, \sigma_1) = (34.6410, 11.6502)$ |
| *Normal* $(\mu_2, \sigma_2)$ | $(\mu_2, \sigma_2) = (12.7456, 3.3523)$ |

Similarly, the MLEs of the parameters of the models for Case-2 are displayed in table 5.7. Note that, in this Case, the MLEs of the parameters are obtained via the EM algorithm.

Table 5.7: MLEs of the Parameters for Case-2 (Maintenance information unknown)

| Models | MLEs of Parameters |
|---|---|
| $Weibull(\beta_1, \eta_1) - Weibull(\beta_2, \eta_2)$ | $(\beta_1, \eta_1, \beta_2, \eta_2, p_1, p_2) = (9.5796,$ $39.3209, 2.1019, 33.0784, 0.2696, 0.7304)$ |
| $Weibull(\beta, \eta) - Normal(\mu, \sigma)$ | $(\beta, \eta, \mu, \sigma, p_1, p_2)$ $= (2.0920, 33.1085, 37.1120, 4.7752,$ $0.7261, 0.2739)$ |

Therefore, the fitted model for Case-1 is:

$$\hat{G}_1(t) = 0.7812\ \text{Norm}(\mu_1 = 34.6410, \sigma_1 = 11.6502) +$$

$$0.2187\ \text{Norm}(\mu_2 = 12.7456, \sigma_2 = 3.3523)$$

and for Case-2 is:

$$\hat{G}_2(t) = 0.2696\ \text{Weib}(\beta_1 = 9.5796, \eta_1 = 39.3209) +$$

$$0.7304\ \text{Weib}\ (\beta_2 = 2.1019, \eta_2 = 33.0784)$$

For 2-fold Weibull mixture model, or

$$\hat{G}_2(t) = 0.7261\ \text{Weib}(\beta = 2.0920, \eta = 33.1085) +$$

$$0.2739\ \text{Norm}(\mu = 37.1120, \sigma = 4.7752)$$

For Weibull-Normal mixture model.

## 5.3.4 Measures of Lifetime Quantities

To compare the Case-1 and Case-2, this section presents the measures of some lifetime quantities (such as, MTTF, median, B5-lifetime and B10-lifetime) of the batteries for Case-1 and Case-2. These results are shown in Table 5.8.

Table 5.8: Comparison of lifetime quantities for Case-1 and Case-2

| | **Maintenance information known (Case-1)** | **Maintenance information unknown (Case-2)** | |
|---|---|---|---|
| | Normal-Normal Mixture | 2-fold Weibull Mixture | Weibull-Normal Mixture |
| MTTF | 29.85 | 31.47 | 31.46 |
| Median | 29.85 | 30.50 | 30.34 |
| B5 life | 13.67 | 13.66 | 13.83 |
| B10 life | 17.25 | 16.66 | 16.69 |

Table 5.8 shows that all the measures of lifetime quantities for Case-1 and Case-2 are very close to each other. So, from this analysis we can conclude thatdata without maintenance information provides approximately similar result with the data having maintenance information.

## 5.4 Hydraulic Pump Failure Data Analysis

Karim, et al. (2015) applied single Weibull, 2-fold Weibull mixture and 3-fold Weibull mixture models for this data set and suggested the 3-fold Weibull mixture model as the best fitted model on the basis of various graphical and statistical approaches. While analyzing this data set, first we assume that the data contain only mixture but no failure mode information. So, we put $q = 0$ in eq.(2.60), which gives the form of the pdf of a mixture model. Again assuming the data do not contain any mixture but only failure mode information, we put$p = 0$ and $q = 1$in eq.(2.60) and obtain the pdf of a competing risk model. Various graphical and statistical methods have been applied to find out the best fitted models.

Firstly let us assume thatthere exists only mixture in the data. In addition to 3-fold Weibull mixture model, here we assume two other 3-fold mixture models (Weibull-Normal-Exponential and Normal-Lognormal-Weibull) for the data. Our aim is to find out whether any other 3-fold mixture models fit this data set better than the 3-fold Weibull mixture model or not. And if the distribution changed, what would be its effect on optimal maintenance policy. R programming codes are written for the computations of thisdata set. Programming codes for analyzing the data with Weibull-Normal-Exponential mixture model are given in section A.2 in the Appendix. The

given codes can be used for other two 3-fold mixture models after simple modifications, mainly related to the functions dweibull(), pweibull(), dnorm(), pnorm(), dexp() and pexp() and the parameter vector theta.

Secondly, assuming the data contain only failure mode information, we apply a set of 2-fold competing risk models to find out the distributions for two different failure modes - existenceof assembly errorsand no assembly errors.

### 5.4.1 Model Selection

This section applies the graphical and statistical approaches for selecting the best fitted model for the data set among three competitive 3-fold mixture models listed in Table 2. A relatively straightforward approach to select a tentative model is to utilize the plotting methodology where the cdfs obtained from parametric estimates are compared with the empirical distribution function. More detail about this comparison can be found in Blischke, Karim and Murthy (2011). The cdfs of 3-fold Weibull, Weibull-Normal-Exponential and Normal-Lognormal-Weibull mixture modelsare compared with the empirical distribution function (nonparametric estimate of cdf from Kaplan-Meier (KM) estimate) and the resultsare displayed in Figure 5.9.

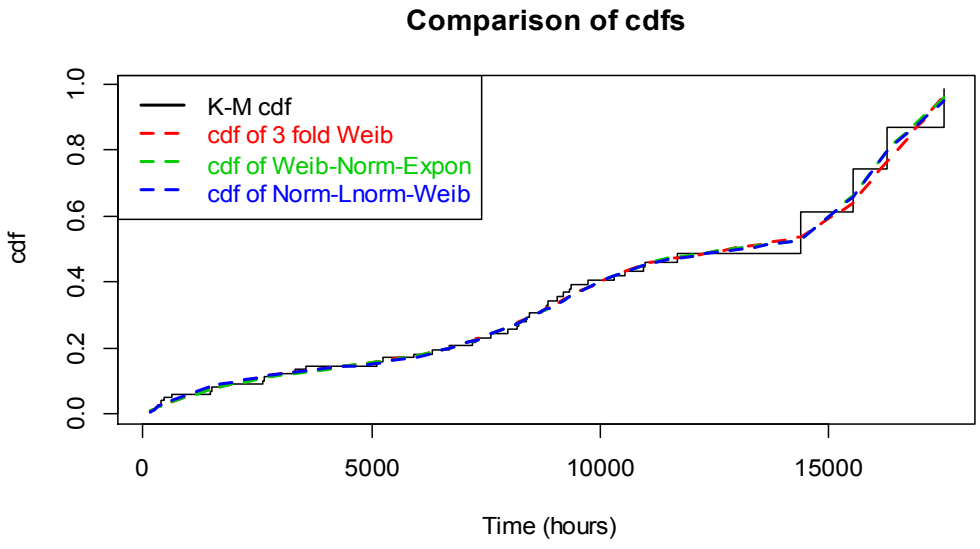


Figure 5.9: Comparison of parametric and nonparametric estimates of cdfs

Figure 5.9 indicates that all the cdfs obtained from the three different mixture models give approximately same result, except at the right tail of the figure of cdfs, where the cdfs of Weibull-Normal-Exponential and Normal-Lognormal-Weibull mixture models

belong slightly closer to the nonparametric estimate of cdfthan that of the cdf of 3-fold Weibullmixture model. Hence we may consider both theWeibull-Normal-Exponential and Normal-Lognormal-Weibull mixture models for the data set.

The statistical approaches provide a more rigorous method for model selection and validation. Various statistics (such as AD, KS test statistic, AICand RMSE) are applied here for model selection and validation. The estimates of AIC, AD*, KS test statistic and RMSE for the three competitive models are given in Table 5.9.

Table 5.9: Estimates ofAIC, AD*, KS test statistic and RMSE for the models

| 3-fold Mixture Models | AIC | AD* | KS test statistic | RMSE |
|---|---|---|---|---|
| 3-fold Weibull | 965.5942 | 0.6272 | 0.1068 | 0.0247 |
| Weibull-Normal-Exponential | **963.2531** | 0.5278 | 0.0876 | **0.0209** |
| Normal-Lognormal-Weibull | 964.6492 | **0.4781** | 0.0877 | 0.0217 |

From Table 5.9, we found that the Weibull-Normal-Exponential mixture model contains the smallest values of AIC& RMSE and the Normal-Lognormal-Weibull mixture model contains the smallest value of AD* test statistic among all of the mixture models. Hence, it can be concluded that, among these mixture models, Weibull-Normal-Exponential mixture model can be selected as the best model for hydraulic pump failure data according to the values of AIC andRMSE.

We have also applied the Kolmogrov-Smirnov (KS) test statistic as a goodness-of-fit test for these3-fold mixture models.At the 5% level of significance, with $n$ =102, the critical value of the Kolmogorov-Smirnov one-sample test can be estimated from eq. (3.23), which is 0.1333. Since the observed value of the KS test statistic for all the 3-fold mixture models (given in Table 5.9) are less than the critical value, we cannot reject the null hypothesis, $H_0$, that the observed data are from a population specified by these 3-fold mixture distributions.But we may consider that among all these three mixture models the Weibull-Normal-Exponential mixture model gives the smallest value for the KS test statistic.

Secondly, let us assume that the data contain only failure mode information. Aset of 2-fold competing risk models are applied and it is found that, the reliability functions of Weibull-Normal, Weibull-Lognormal and Normal-Lognormal competing risk modelsgo relatively close through the empirical distribution function (nonparametric estimate of cdf from Kaplan-Meier (KM) estimate), see Figure 5.10.

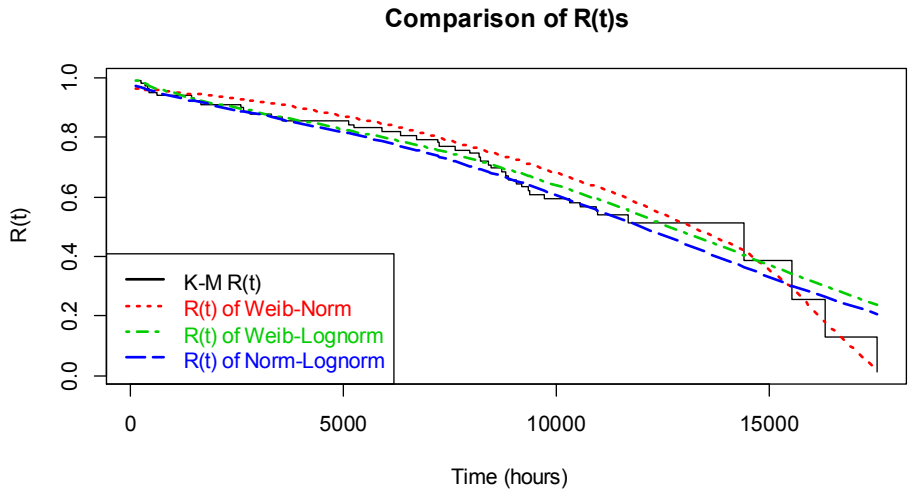


Figure 5.10: Comparison of $R(t)$s of competing risk models

The estimates of AIC, KS test statistic and RMSE for the three competitive models are given in Table 5.10.

Table 5.10: Estimates of AIC, KS test statistic and RMSE for competing risk models

| Competing risk Model | AIC | KS test statistic | RMSE |
|---|---|---|---|
| Weibull-Normal | 985.4092 | 0.0924 | 0.0294 |
| Weibull-Lognormal | 971.5454 | 0.2228 | 0.0459 |
| Normal-Lognormal | 975.0428 | 0.1953 | 0.0421 |

From Table 5.10, we found that the Weibull-Normal competing risk model contains the smallest values of RMSE and the Weibull-Lognormal competing risk model contains the smallest value of AIC among all of the models.

At the 5% level of significance, with $n$ =102, the critical value of the Kolmogorov-Smirnov one-sample test can be estimated from eq. (3.23), which is 0.1333. Since the observed value of the KS test statistic for only Weibull-Normal competing risk model (given in Table 5.10) is less than the critical value, we cannot reject the null hypothesis, $H_0$, that the observed data are from a population specified by the Weibull-Normal competing risk model.

Hence according toFigure 5.10 and Table 5.10we may select the Weibull-Normal competing risk model as the best model.

Figure 5.11 shows the estimated reliability functions of Weibull and Normal models separately to identify the distribution of two different failure modes for this data set.

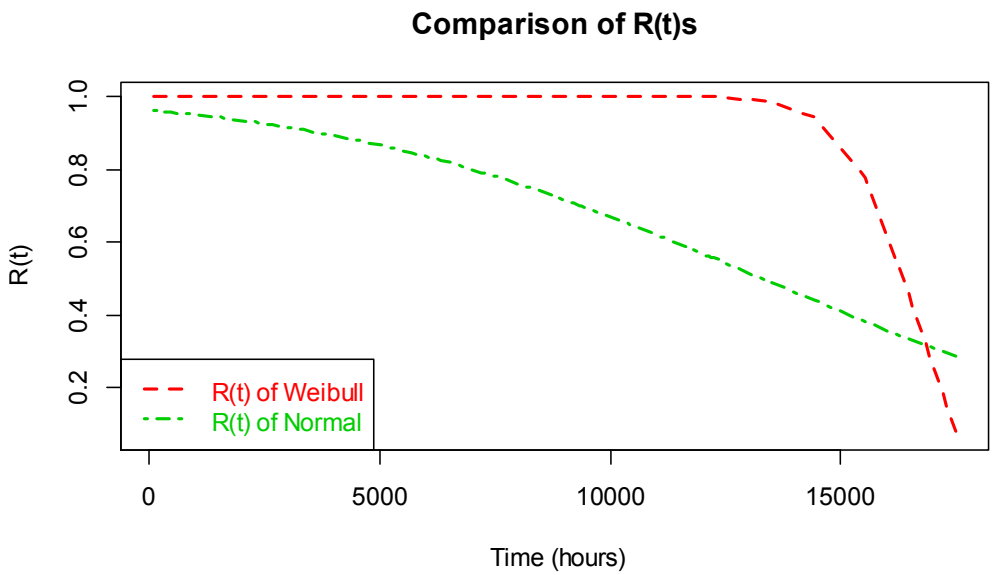


Figure 5.11: Comparison of $R(t)$s of failure mode distributions

Since, $R(t)$ of Normal distribution shows lower values than that of $R(t)$ obtained from Weibull distribution, hence we may conclude thatpumps with the problem of assembly error follow Normal distribution and the pumps with no assembly error follow Weibull distribution.

## 5.4.2 MLEs of the Parameters of Mixture Models

This section gives the MLEs of the parameters of different mixture models. The parameters of the three mixture models are estimated by applying maximum likelihood method via the Expectation-Maximization (EM) algorithm. The MLEs of the parameters are displayed in Table 5.11. In Table 5.11, the parameters $p_1$, $p_2$and$p_3$ represent the mixing probabilities of the 1st, 2nd and 3rd sub-populations, respectively.

Table 5.11: MLEs of the Parameters of Assumed Mixture Models

| 3-fold Mixture Models | MLEs of Parameters |
|---|---|
| Weibull($\beta_1,\eta_1$)- Weibull($\beta_2,\eta_2$)- Weibull($\beta_3,\eta_3$) | $\{\beta_1,\eta_1,\beta_2,\eta_2,\beta_3,\eta_3,p_1,p_2,p_3\}$ = {1.0191,2364.0191,5.5758,9481.8351, 16.6426,16535.5039,0.1659,0.3220,0.5120} |
| Weibull($\beta,\eta$)-Normal($\mu,\sigma$)-Exponential($\delta$) | $\{\beta,\eta,\mu,\sigma,\delta,p_1,p_2,p_3\}$ = {5.5391, 9527.83, 15991.11, 1073.821, 0.0004, 0.3249, 0.5076, 0.1674} |
| Normal($\mu_1,\sigma_1$)-Lognormal($\mu_2,\sigma_2$)-Weibull($\beta,\eta$) | $\{\mu_1,\sigma_1,\mu_2,\sigma_2,\beta,\eta,p_1,p_2,p_3\}$ = {15992.0308, 1072.7513, 7.5063, 1.3759, 5.4782, 9497.0899, 0.4947, 0.1872, 0.3180} |

## 5.4.3 Mean Time to Failure (MTTF)

- For Weibull($\beta_1,\eta_1$)-Weibull($\beta_2,\eta_2$)-Weibull($\beta_3,\eta_3$) mixture model, the mean for $F_3(t;\beta_3,\eta_3)$ = 16018.005> mean for $F_2(t;\beta_2,\eta_2)$ = 8760.457> mean for $F_1(t;\beta_1,\eta_1)$ = 2345.628.
- For Weibull($\beta,\eta$)-Normal($\mu,\sigma$)-Exponential($\delta$) mixture model, the mean for $F_2(t;\mu,\sigma)$= 15991.110> mean for $F_1(t;\beta,\eta)$= 8799.642> mean for $F_3(t;\delta)$ = 2500.000.
- For Normal ($\mu_1,\sigma_1$)-Lognormal($\mu_2,\sigma_2$)-Weibull ($\beta,\eta$) mixture model, the mean for $F_1(t;\mu_1,\sigma_1)$ =15992.031> mean for $F_3(t;\beta,\eta)$ =8765.749> mean for $F_2(t;\mu_2,\sigma_2)$ = 4688.418.

## 5.4.4 New Intuitions

Note that the data is a mixture from three sub-populations. Each of these sub-populations can be interpreted in terms of the characterisation of the real world relevant to the pump. Here we discuss a new situation associated with this pump.This situation is based on the following assumptions (Murthy, et al., 2015):

1. All new pumps are statistically identical.

2. Some of the items replaced during PM and CM action (or service exchange) are scrapped (as they are deemed to be repairable) and others reconditioned.
3. All reconditioned pumps are also statistically identical.
4. The reliability characteristics of a new pump are different from that of a reconditioned pump.
5. A pump used during service exchange (under PM or CM action) can be either correctly or incorrectly installed.

We use the following notations:

$q$ : Probability that the pump is scrapped and replaced by a new one under service exchange.

$1-q$ : Probability that the pump is not scrapped and reconditioned under service exchange.

$p$ : Probability that the item used in service exchange is installed correctly.

$1-p$ : Probability that the item used in service exchange is not installed correctly.

$F_N(t)$ : Failure distribution of new item installed correctly.

$F_R(t)$ : Failure distribution of reconditioned item installed correctly.

$F_I(t)$ : Failure distribution of incorrectly installed item (new or reconditioned).

As a result, the probabilities of the different outcomes after a service exchange are as indicated in Table 5.12.

Table 5.12: Probabilities of different outcomes

| | | | Installation | |
|---|---|---|---|---|
| | | | Correct | Incorrect |
| | | | $p$ | $(1-p)$ |
| Scrap/repair | Scrap (new) | $q$ | $qp$ | $q(1-p)$ |
| | Not scrap (recondition) | $1-q$ | $(1-q)p$ | $(1-q)(1-p)$ |

It is easily seen (using the conditional approach) that the time to failure of an item used in service exchange is given by a distribution function

$$G_3(t) = (1-p)F_I(t) + (1-q)pF_R(t) + qpF_N(t) \qquad (5.1)$$

The distribution function given in eq.(5.1) can be treated as the cdf of a 3-fold mixture model.

As discussed in table 5.12, we also get the following from eq. (5.1):

$(1-p)$ : is the probability that the item was incorrectly installed.

$(1-q)p$ : is the probability that a reconditioned iten installed correctly.

$qp$ : is the probability that a new item installed correctly.

Note that the MTTF (mean time to failure) for a new item installed correctly > MTTF for a reconditioned item installed correctly > MTTF for an item (new or reconditioned) installed incorrectly.

If we select the Weibull($\beta,\eta$)-Normal($\mu,\sigma$)-Exponential($\delta$) mixture model as the best model for the data, then for the Weibull-Normal-Exponential mixture model we found MTTF of Normal distribution = 15991.110> MTTF of Weibull distribution =8799.642> MTTF of Exponential distribution = 2500.000.

Again for the best fitted mixture model (Weibull-Normal-Exponential), we estimate the reliability function of Weibull, Normal and Exponential models, respectively for this data and the results are presented in figure 5.12:

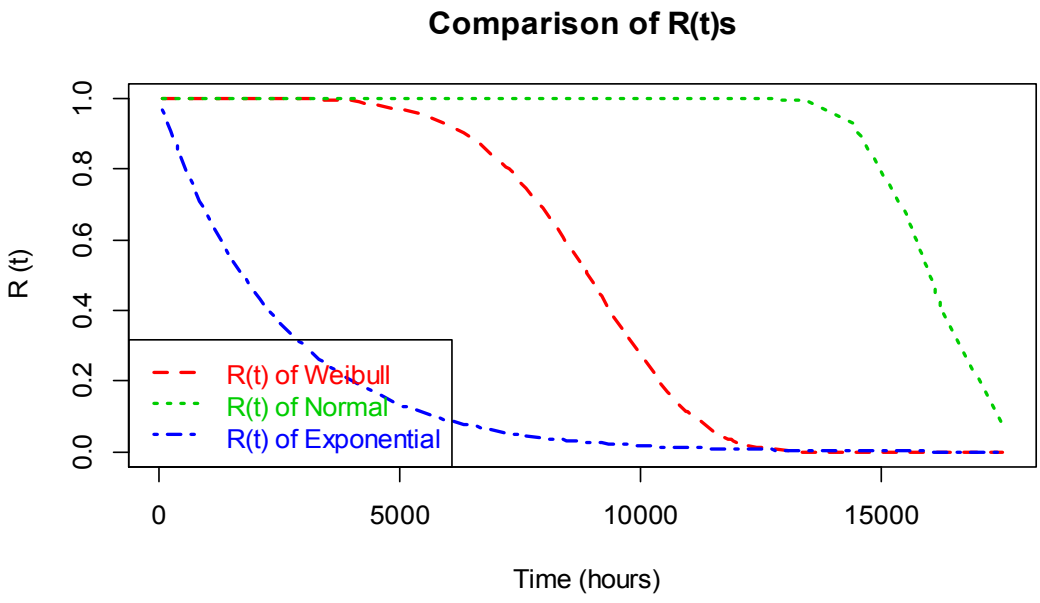


Figure 5.12: Comparison of R(t)s for Weibull, Normal and Exponential model

This figure indicates that R(t) of Normal distribution > R(t) of Weibull distribution > R(t) of Exponential distribution. This means, new and correctly installed items follow Normal, recondition and correctly installed items follow Weibull and incorrectly installed items follow Exponential distributions, respectively.

From eq. (2.36) we found that the cdf of 3-fold Weibull-Normal-Exponential distribution is

$$F(t) = p_1\left\{1-\exp\left[-\left(\frac{t}{\eta}\right)^{\beta}\right]\right\} + p_2\Phi\left(\frac{t-\mu}{\sigma}\right) + (1-p_1-p_2)\{1-\exp(-\delta t)\}$$

Which means

$$F(t) = p_1F_1(t;\beta,\eta) + p_2F_2(t;\mu,\sigma) + p_3F_3(t;\delta) \tag{5.2}$$

Hence comparing eq (5.1) and eq (5.2), and from the above discussion we can write

$$p_3 = (1-p),\ \ p_1 = (1-q)\,p \text{ and } p_2 = qp \tag{5.3}$$

$$F_3(t;\delta) = F_I(t),\ \ F_1(t;\beta,\eta) = F_R(t) \text{ and } F_2(t;\mu,\sigma) = F_N(t) \tag{5.4}$$

Using the estimates of $p_1$, $p_2$ and $p_3$ from Table 5.11 in equation (5.3), we get the estimates of $p$ = 0.8326 and $q$ =0.6096. i.e., the probability of installation an item correctly is 0.8326 and the probability of installation of a new item is 0.6096.

### 5.4.5 Optimum Maintenance Cost

Like Karim et al. (2015), we use the followingadditional notations and assumptions.

$C_n$ : Sale price for new pump is \$80,000 (Given by the owner).

$C_r$ : Cost (charged by the service agent) for reconditioning a pump under CM or PM action (\$60,000).

$\xi$ : Additional cost (due to downtime, loss in revenue, etc.) resulting from CM action. We look at values of $\xi$ = \$70,000, \$90,000,\$110,000 and \$130,000.

The optimal $T^*$ is obtained using eq. (1.3) with 3-fold mixture cdf $F(t)$ and the optimal expected cost per unit time is given by $J\left(T^*;F(.)\right)$.

Here we can see that, the optimal $T^*$ depend on the additional cost $\xi$. The optimal $T^*$ and optimal expected cost per unit time $J(T^*)$ on various values of $\xi$ for the three different 3-fold mixture models has been estimated.These results are given in Table5.13, from where it can be seen,for every model, the optimal $T^*$ decrease and optimal $J(T^*)$ increasing with $\xi$ increases, as to be expected.

Table 5.13:Optimal $T^*$ and $J(T^*)$ for different values of $\xi$

| Model | Optimal Values | Additional Cost | | | |
|---|---|---|---|---|---|
| | | $\xi$=70000 | $\xi$=90000 | $\xi$=110000 | $\xi$=130000 |
| 3-fold Weibull | $T^*$ | 14631 | 14484 | 14377 | 14295 |
| | $J(T^*)$ | 10.4048 | 11.4324 | 12.4516 | 13.4656 |
| Weibull-Normal-Exponential | $T^*$ | 14466 | 14360 | 14285 | 14228 |
| | $J(T^*)$ | 10.3174 | 11.3151 | 12.3066 | 13.2943 |
| Normal-Lognormal-Weibull | $T^*$ | 14476 | 14368 | 14291 | 14234 |
| | $J(T^*)$ | 10.3261 | 11.3187 | 12.3186 | 13.3078 |

This table indicates that the 3-fold Weibull mixture model gives a bit larger optimal maintenance period $T^*$ than other two models, however the Weibull-Normal-Exponential model shows a reduction in the maintenance cost than the other two models for all $\xi$.

Now for the additional cost $\xi = \$90000$, the figure of time $T$ and cost $J(T)$ for these three mixture models are presented in Figure 5.13:

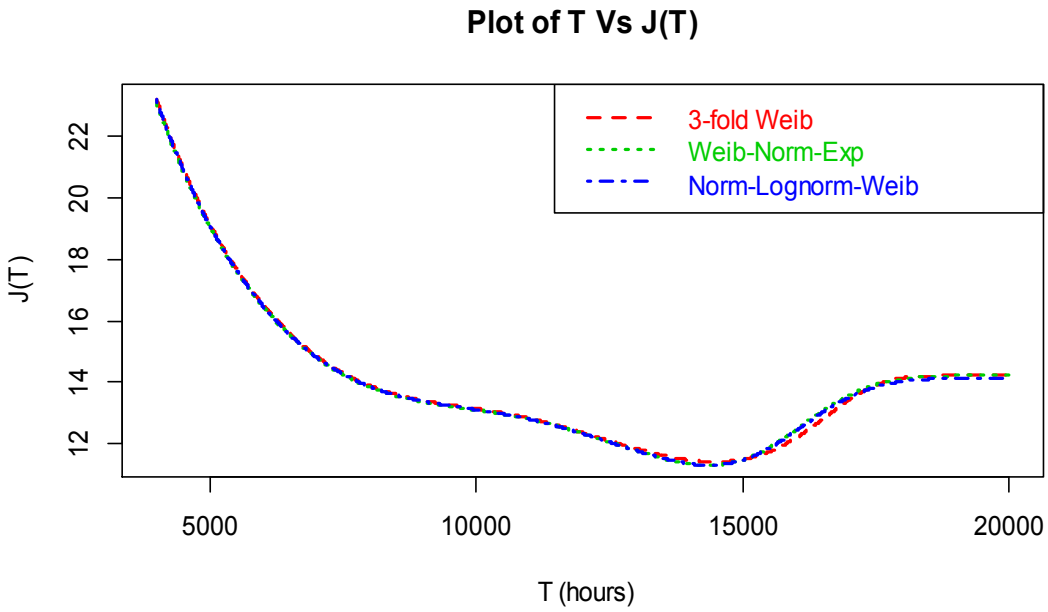


Figure 5.13: Comparison of $T$ Vs $J(T)$ at$\xi = \$90000$

We may also calculate and represent the results of $T$ Vs $J(T)$ at the other additional costs.

# Chapter 6
# Simulation

## 6.1 Introduction

In this chapter, we have used computer simulation to evaluate the performance of the methods numerically. 2-fold and 3-fold mixture data were generated numerically are used to develop two different special cases of 2-fold and 3-fold mixture models, to find the ML estimates of model parameters under right censored data. Using simulated data, the ML estimates of the model parameters, the mean squared errors (MSEs) and the amount of bias of estimates are computed. Simulation programming codes are written using statistical software package R.

## 6.2 Steps of Simulation Study

Here we describe the step-by-step algorithm for simulation of the mixture model and estimation of model parameters via the EM algorithm.

**Step 1:** We consider a set of true value for the parameters, say, $\theta$ of the mixture model. Under this set of parameter, we generate $n = \sum_j n_j$ samples from the mixture model using the software R-Language (version-3.2.2). Here $j$ is the sub-populations of the mixture model. A desired percent of the largest generated sample out of $n$, are considered as the right censored observations and remaining are assumed as failed lifetime. Again considering a certain observations as right censored, a set of different sample sizes were generated.

**Step 2:** Based on the generated right censored data, we estimate the parameters via the EM algorithm assuming that the mixing sub-populations are unknown. The methodology of EM algorithm is discussed in section 3.3.2.2.

**Step 3:** The above Steps 1 and 2 are repeated 1000 times under two Cases:

**Case (i):** for a variety percent of censored observations and
**Case (ii):** for different sample sizes.

We compute the mean squared errors (MSEs) and the amount of Bias of the estimates for the both Cases (i) and (ii).

**Steps 4:** Summarize and discuss the simulation results based on 1000 repetition.

## 6.3 Bias, Variance and MSE

Bias is defined as the difference between the true parameter value and the parameter estimate. Mathematically this is defined to be:

$$\text{Bias} = E(\hat{\theta}) - \theta \tag{6.1}$$

where$\hat{\theta}$ is our estimated parameter and $\theta$ is the true parameter value. This is computed by finding the average of each of the 1000 parameter estimates and subtracting the true value, which was used to generate the data.

Variance is defined to be the deviation about the mean. In this case, it is the sample variance for each of the parameters and sample sizes. This is computed by finding the sum of squared deviances of each data point from the data's average value and dividing by ($n$ – 1)

$$\text{Variance} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \tag{6.2}$$

Once we have the measures for bias and variance we can compute the mean square error as defined below:

$$MSE = Bias^2 + Variance \tag{6.3}$$

## 6.4 Simulation Output Analysis

The simulation studies are performed for two different mixture models:

1. Simulation for 2-fold mixturemodel
2. Simulation for 3-fold mixturemodel

### 6.4.1 Simulation for 2-fold Mixture Model

We consider a set of true values for the 5 parameters $\theta=\{\beta_1,\eta_1,\beta_2,\eta_2,p\}$ of a 2-fold Weibull mixture model. Under this set of parameters, we generate $n=n_1+n_2$ samples from the 2-fold Weibull mixture model. The steps described in section 6.2are repeated1000 times for a variety percent of censored observations (10%, 20% and 30%) andfor different sample sizes (*n* = 200, 400 and 600).

Tables 6.1, 6.2, 6.3 and 6.4 represent the summary results of the simulations based on 1000 repetitions under the given true values.

Table 6.1 presents the Mean Square Error (MSEs) at different percent of censored observations (10%, 20%, 30%) when sample size is 200. And table 6.2 shows MSEs at 20% percent of censored observation for different sample sizes (200, 400, 600). In these tables, the first column shows the parameters of the model and second column shows the true values of the parameters. Tables 6.1 and 6.2 give the MSEs of the MLEs of parameters obtained by the EM algorithm.

Table 6.1: MSEs for *n*= 200 at different percent of censored observations

| **Parameters** | **True values** | **Mean squared errors (MSEs) of MLEs** | | |
|---|---|---|---|---|
| | | **10% cens obs.** | **20% cens obs.** | **30% cens obs.** |
| $\hat{\beta}_1$ | 3.50 | 2.17953 | 2.6223 | 4.2156 |
| $\hat{\eta}_1$ | 700.00 | 3232.01855 | 4873.0835 | 9777.0312 |
| $\hat{\beta}_2$ | 1.20 | 0.15011 | 0.0619 | 0.0711 |
| $\hat{\eta}_2$ | 850.00 | 87452.76431 | 1614229.0410 | 269062.7034 |
| $\hat{p}$ | 0.30 | 0.00114 | 0.0391 | 0.0500 |
| $(1-\hat{p})$ | 0.70 | 0.03382 | 0.0391 | 0.0500 |

Table 6.2: MSEs at 20% percent of censored observation for different sample sizes

| Parameters | True values | Mean squared errors (MSEs) of MLEs | | |
|---|---|---|---|---|
| | | ***n*=200** | ***n*=400** | ***n*=600** |
| $\hat{\beta}_1$ | 3.50 | 2.6223 | 1.9544 | 1.5033 |
| $\hat{\eta}_1$ | 700.00 | 4873.0835 | 1907.8766 | 1574.6524 |
| $\hat{\beta}_2$ | 1.20 | 0.0619 | 0.0294 | 0.0210 |
| $\hat{\eta}_2$ | 850.00 | 1614229.0410 | 25655.13287 | 10622.3852 |
| $\hat{p}$ | 0.30 | 0.0391 | 0.0242 | 0.0159 |
| $(1-\hat{p})$ | 0.70 | 0.0391 | 0.0242 | 0.0159 |

From the above tables 6.1 and 6.2, we found that for all of the sets, if the percent of censored observations decrease (i.e., if number of failures increase), the MSEs of the MLEs of the parameters are also decrease for all most all parameters, as expected. Similarly, the MSEs of the MLEs of parameters decrease for increasing sample sizes.

The results obtained from Table 6.1 and Table 6.2 has been expressed in Figure 6.1 and Figure 6.2, respectively:

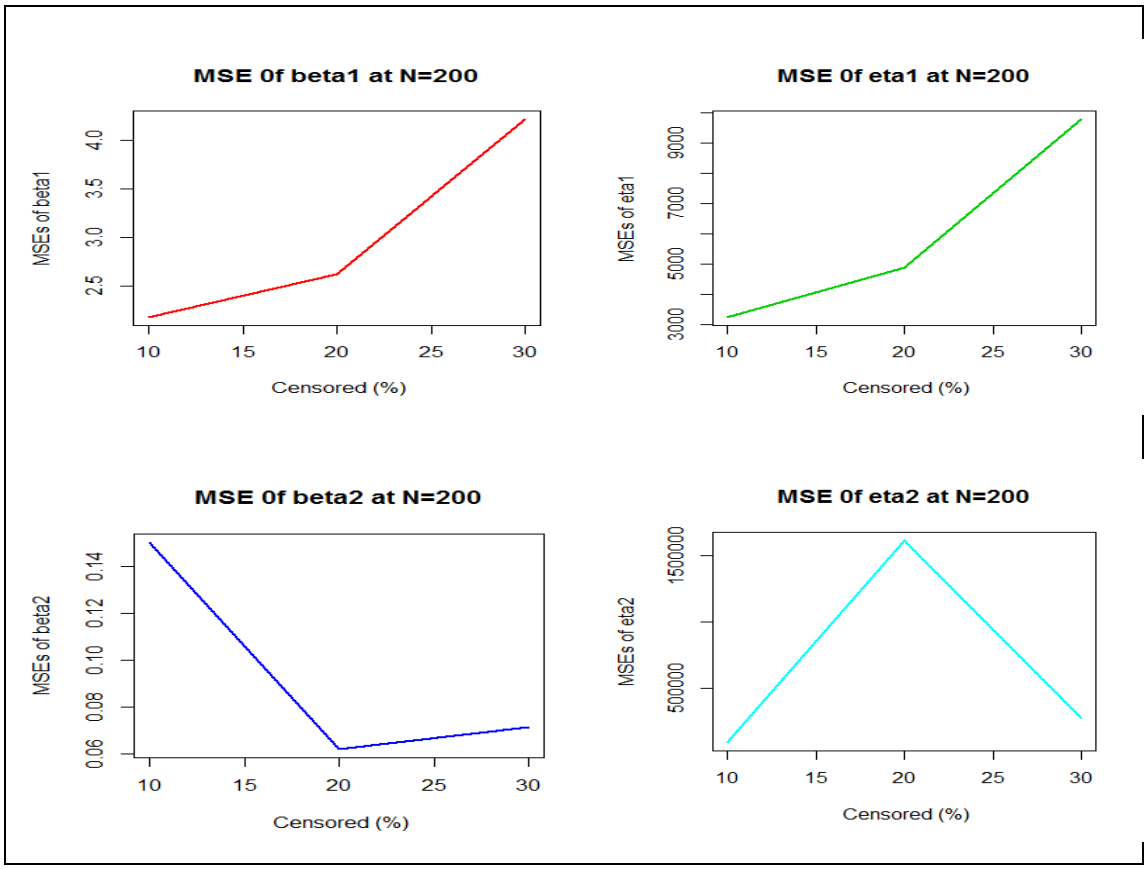


Figure 6.1: MSEs for *n*= 200 at different percent of censored observations

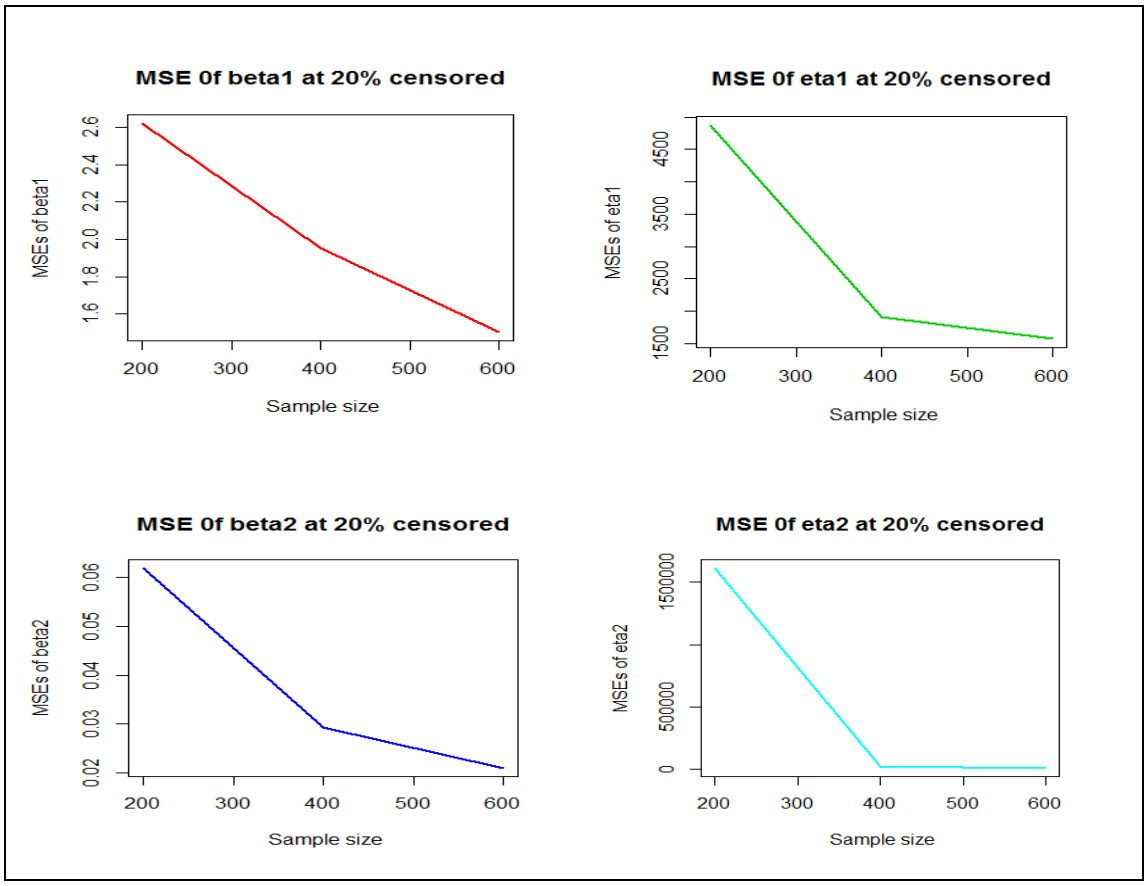


Figure 6.2: MSEs at 20% percent of censored observation for different sample sizes

Again Table 6.3 presents the amount of bias at different percent of censored observations (10%, 20%, 30%) when sample size is 200. And Table 6.4 shows the amount of bias at 20% percent of censored observation for different sample sizes (200, 400, 600).

Table 6.3: Amount of Bias for $n$= 200 at different percent of censored observations

| **Parameters** | **True values** | **Amount of Bias of MLEs** | | |
|---|---|---|---|---|
| | | **10% cens obs.** | **20% cens obs.** | **30% cens obs.** |
| $\hat{\beta}_1$ | 3.50 | 0.2945 | 0.4587 | 0.6968 |
| $\hat{\eta}_1$ | 700.00 | -0.6464 | 0.1768 | 5.3863 |
| $\hat{\beta}_2$ | 1.20 | -0.0399 | -0.0546 | -0.0291 |
| $\hat{\eta}_2$ | 850.00 | 94.5603 | 205.9218 | 62.0363 |
| $\hat{p}$ | 0.30 | 0.0737 | 0.0692 | 0.0732 |
| $(1-\hat{p})$ | 0.70 | -0.0737 | -0.0692 | -0.0732 |

Table 6.4: Amount of Bias at 20% percent of censored observation for different *n*

| Parameters | True values | Amount of Bias of MLEs | | |
|---|---|---|---|---|
| | | ***n*=200** | ***n*=400** | ***n*=600** |
| $\hat{\beta}_1$ | 3.50 | 0.4587 | 0.3619 | 0.3365 |
| $\hat{\eta}_1$ | 700.00 | 0.1768 | 3.0983 | 2.2486 |
| $\hat{\beta}_2$ | 1.20 | -0.0546 | -0.0336 | -0.015 |
| $\hat{\eta}_2$ | 850.00 | 205.9218 | 41.5291 | 23.1364 |
| $\hat{p}$ | 0.30 | 0.0692 | 0.0398 | 0.0192 |
| $(1-\hat{p})$ | 0.70 | -0.0692 | -0.0398 | -0.0192 |

The amount of bias of the MLEs of parameters for different percent of censored observations and for different sample sizes are given in Tables 6.3 and 6.4, respectively. The amount of bias decrease for decreasing of the percent of censored observations (i.e., for increasing the number of failures) for all most all of parameters.Similarly, the amount of bias decrease for increasing of the sample sizes.

All of these comparisons from Tables 6.1, 6.2, 6.3 and 6.4 indicate that the proposed method of estimation is applicable for analyzing 2-fold mixture model for censored data.

## 6.4.2 Simulation for 3-fold Mixture Model

In this Section, we conduct simulation studies with a 3-fold Weibull-Normal-Exponential mixture model under right censored data.

We consider a set of true values for the 7 parameters $\theta = (\beta, \eta, \mu, \sigma, \delta, p_1, p_2)$ of 3-fold Weibull-Normal-Exponential mixture model. Under this set of parameters, we generate $n = n_1 + n_2 + n_3$ samples. The steps of simulation analysis described in section 6.2are repeated 1000 times for a variety percent of censored observations (0%, 10%, 20% and 30%) andfor different sample sizes (300, 600 and 900).

Table 6.5 presents the Mean Square Error (MSEs) at different percent of censored observations (0%, 10%, 20%, 30%) when sample size is 300. Table 6.6 shows MSEs at 20% percent of censored observation for different sample sizes (300, 600, 900).

Table 6.5: MSEs for $n$ = 300 at different percent of censored observations

| Parameter | True Value | Mean Square Error (MSEs) of MLEs | | | |
|---|---|---|---|---|---|
| | | 0% cens obs. | 10% cens obs. | 20% cens obs. | 30% cens obs. |
| $\beta$ | 1.5 | 1.8408 | 3.6161 | 8.6121 | 6.1823 |
| $\alpha$ | 16500 | 40312754 | 34915543 | 36182067 | 52402705 |
| $\mu$ | 12500 | 4652303 | 13747737 | 18942191 | 25029653 |
| $\sigma$ | 6500 | 2080092 | 5102398 | 8610976 | 9672087 |
| $\delta$ | 0.0009 | 1.3717e-05 | 5.5011e-05 | 3.4801e-05 | 0.0007 |
| $p_1$ | 0.35 | 0.0547 | 0.0531 | 0.0536 | 0.0447 |
| $p_2$ | 0.25 | 0.1163 | 0.1021 | 0.0567 | 0.0631 |
| $p_3$ | 0.40 | 0.0629 | 0.05854 | 0.0426 | 0.0473 |

Table 6.6: MSEs at 20% percent of censored observation for different sample sizes

| Parameter | True Value | Mean Square Error (MSEs) of MLEs | | |
|---|---|---|---|---|
| | | $n$=300 | $n$=600 | $n$=900 |
| $\beta$ | 1.5 | 8.6122 | 2.7287 | 0.7183 |
| $\alpha$ | 16500 | 36182067 | 32752424 | 21827666 |
| $\mu$ | 12500 | 18942191 | 11046977 | 3828475 |
| $\sigma$ | 6500 | 8610976 | 3657922 | 1365859 |
| $\delta$ | 0.0009 | 3.4802e-05 | 3.8230e-06 | 3.9624e-06 |
| $p_1$ | 0.35 | 0.0536 | 0.0460 | 0.0425 |
| $p_2$ | 0.25 | 0.0567 | 0.1026 | 0.0951 |
| $p_3$ | 0.40 | 0.0426 | 0.0656 | 0.0712 |

From the above Tables 6.5 and 6.6, we found that for all most all of parameters, if the percent of censored observations decrease (i.e., if number of failures increase), the MSEs of the MLEs of the parameters are also decrease, as expected. Similarly, the MSEs of the MLEs of parameters decrease for increasing sample sizes.

The results obtained in Tables 6.5 and 6.6 are presented in Figures 6.3 and 6.4, respectively.

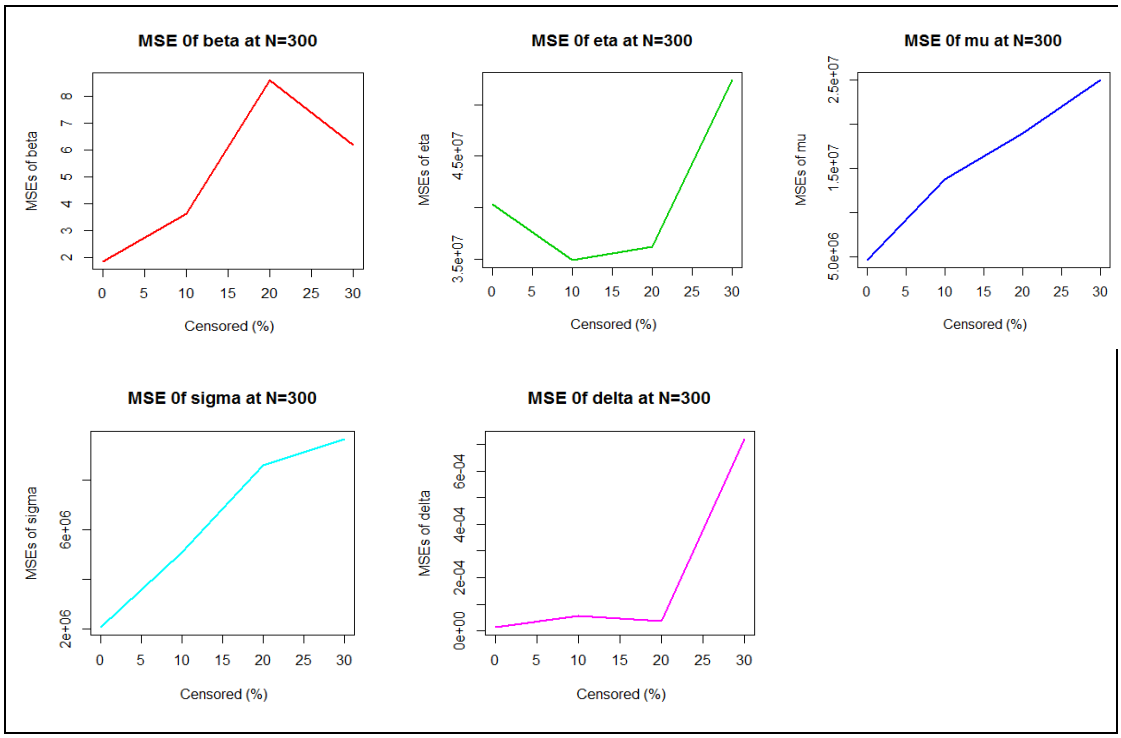


Figure 6.3: MSEs for *n*= 300 at different percent of censored observations

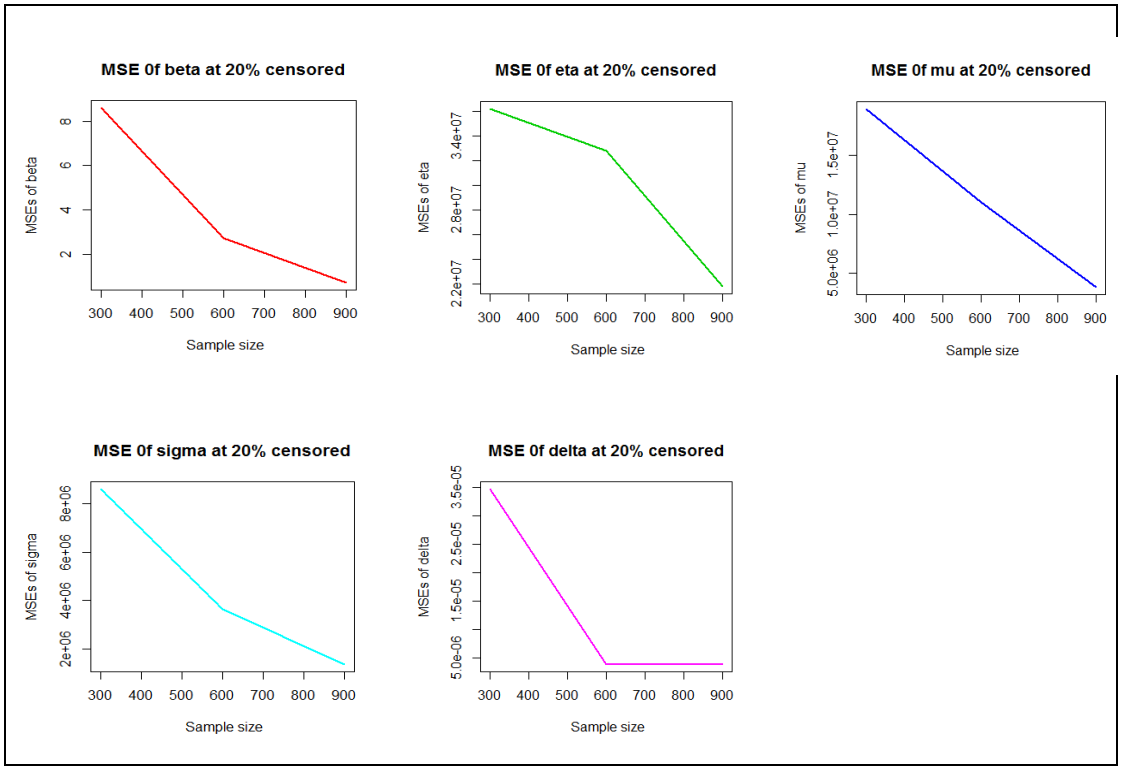

Figure 6.4: MSEs at 20% percent of censored observation for different sample sizes

Again Table 6.7 presents the amount of bias at different percent of censored observations (0%, 10%, 20%, 30%) when sample size is 300 and Table 6.8 shows the amount of bias at 20% percent of censored observation for different sample sizes (300, 600, 900).

Table 6.7: Amount of Bias for *n*= 300 at different % of censored observations

| **Parameter** | **True Value** | **Amount of Bias of MLEs** | | | |
|---|---|---|---|---|---|
| | | **0% cens obs.** | **10% cens obs.** | **20% cens obs.** | **30% cens obs.** |
| $\beta$ | 1.5 | 0.2612 | 0.8131 | 1.1043 | 1.0014 |
| $\alpha$ | 16500 | -591.0491 | -680.8534 | -501.8605 | -461.0324 |
| $\mu$ | 12500 | 367.1699 | 209.2456 | 241.6512 | 525.2763 |
| $\sigma$ | 6500 | -348.3239 | -513.345 | -804.4014 | -545.7877 |
| $\delta$ | 0.0009 | 0.00049 | 0.00081 | 0.00053 | 0.0011 |
| $p_1$ | 0.35 | -0.0606 | 0.00019 | 0.0362 | 0.0011 |
| $p_2$ | 0.25 | 0.2508 | 0.1945 | 0.1089 | 0.1180 |
| $p_3$ | 0.40 | -0.1902 | -0.1947 | -0.1451 | -0.1517 |

Table 6.8: Amount of Bias at 20% percent of censored observation for different *n*

| **Parameter** | **True Value** | **Amount of Bias of MLEs** | | |
|---|---|---|---|---|
| | | ***n*=300** | ***n*=600** | ***n*=900** |
| $\beta$ | 1.5 | 1.10434 | 0.6080 | 0.2288 |
| $\alpha$ | 16500 | -501.8605 | -511.7068 | -698.2374 |
| $\mu$ | 12500 | 241.6512 | 214.2128 | 114.9425 |
| $\sigma$ | 6500 | -804.4014 | -169.5012 | -100.7139 |
| $\delta$ | 0.0009 | 0.00052 | 0.00033 | 0.00029 |
| $p_1$ | 0.35 | 0.0361 | 0.0035 | 0.0261 |
| $p_2$ | 0.25 | 0.1089 | 0.2089 | 0.2092 |
| $p_3$ | 0.40 | -0.1451 | -0.2124 | -0.2354 |

For all most of all situations, the amount of bias decrease for decreasing of the percent of censored observations (i.e., for increasing the number of failures). Also the amount of bias decrease for increasing of the sample sizes, as expected. Therefore, we may conclude that the proposed method of estimation is applicable for analyzing 3-fold mixture model for censored data.

# Chapter 7
# Conclusion

## 7.1 Main Contributions

There are situations where variations in product quality and reliability can be occurred for a variety of reasons. Two of such reasons are the component nonconformance and assembly errors, which occur frequently in manufacturing. In such situations, the complex lifetime models are required for analyzing product reliability data. Proper data collection and analysis are very important for effective analysis of product reliability. Data is critical for building and selecting suitable statistical models and model provides new insights for improvements to maintenance and management operations in manufacturing industries.

This thesisproposes a general model for modeling the effects of quality variation. The mixture model and competing risk model are the special cases of this general model. The thesis applies these models for analysis of three sets of product reliability data.

According to the research questions, the main contributions of the thesis are as follows:

**RQ 1.**Which of the methods, graphical (WPP plot) or statistical parametric (MLE), perform well for analyzing product reliability data?

- For the Aircraft windshield failure data, we observed that the reliability function obtained by the MLE method (via the EM algorithm) is much closer to the nonparametric Kaplan–Meier estimate of reliability function than that of the reliability function estimated by the graphical (WPP plot) method.
- Therefore, the results indicate that the method of estimation with the EM algorithm procedure is betterthan the WPP plot procedure.

**RQ 2.**What would be the effects on the estimated reliability if the information on maintenance action is unknown for all observations in the database?

- For Battery failure data, we considered two-fold mixture models assuming two unknown sub-populations and EM algorithm is applied for estimating the models parameters. Based onthe measures of lifetime quantities, we can conclude thatdata without maintenance information provides approximately similar results with the data having maintenance information.

**RQ 3.** What are the suitable 3-fold lifetime models of hydraulic pump? What are the possible three hidden sub-populations and their distributions for this pump?

- According to the graphical representation and estimated values of different model selection criterions, we found that the 3-fold Weibull-Normal-Exponential mixture model can be selected as the best model for the hydraulic pump failure data.
- The possible three sub-populations are (i) new item installed correctly, (ii) reconditioned item installed correctly, and (iii) item (new or reconditioned) installed incorrectly.
- The distributions for (i) new item installed correctly is Normal $(\mu = 15991.11, \sigma = 1073.821)$, (ii) reconditioned item installed correctly is Weibull $(\beta = 5.5391, \eta = 9527.83)$, and (iii) item (new or reconditioned) installed incorrectly is Exponential$(\delta = 0.0004)$.

**RQ 4.**What are the suitable distributions for the pumps with assembly errors and the pumps without assembly errors?

- The selected distribution for pumps with assembly errors is Normal and the distribution for pumps without assembly errors is Weibull.

**RQ 5.**What would be the effect on optimal maintenance policy according to the selected models?

- Based on the optimization of the proposed objective function, we see that the 3-fold Weibull mixture model gives a bit larger optimal maintenance

periodthan other two models, however the Weibull-Normal-Exponential model shows a reduction in the maintenance cost than the other two models.

- That is, the proposed model suggests the optimum maintenance period for the pump that reduces the maintenance cost.

**RQ 6.** What are the overall performances of the proposed models and methods?

- The simulation studies indicate that the proposed models and methods of estimation are applicable for analyzing 2-fold and 3-fold mixture models for censored product reliability data.

Finally, the results presented in this thesis would be useful for managerial implications in assessing and predicting the reliability and maintenance cost of the product more accurately.

## 7.2 Future Research

A number of research areas listed below have been identified for future research that might be stimulated by extending the present research.

- Estimation of the parameters of the general model for the effect of quality variation in manufacturing that includes both assembly errors and problems with component non-conformance jointly would be useful. This requires a suitable data set and the extensions of the EM algorithm with two types of posterior probabilities – one for assembly error items and another for component non-conformance items.
- A scope of the futureresearch with various types of censored data would beinteresting.
- Development of a R-package with the written R-codes will be useful to manufacturers and product reliability researchers.

# List of Related Publications in International Journals

1. Ruhi, S. and Karim, M.R. (2016). Selecting Statistical Model and Optimum Maintenance Policy: A Case Study of Hydraulic Pump, SpringerPlus. (Accepted)
   http:// DOI: 10.1186/s40064-016-2619-1

2. Ruhi, S., Sarker, S. and Karim, M.R. (2015). Mixture Models for Analyzing Product Reliability Data: A Case Study, SpringerPlus , 4:634.
   http://creativecommons.org/licenses/by/4.0/

3. Ruhi, S. (2015). Application of mixture models for analyzing reliability data: a case study. Open Access Library Journal 2:e1815
   http://dx.doi.org/10.4236/oalib.1101815

# Appendix:Computer Program in R

## A.1 R codes for estimating parameters of 2-fold Weibull mixture model via the EM algorithm

```
# ----- Weibull-Weibull Mixture model - parameters estimation - via EM algorithm
# The program requires package “survival”. In these codes-
# t: Failure time or right censored time
# d: Failure/Censored indicator, 1=failure, 0=censored
# theta: Parameter vector, {beta1, alpha1, beta2, alpha2, p1, p2}
#
# ----------------- Observed data log-likelihood function -----
loglik.obs <- function(t, d, theta){
sum(d*log(theta[5]*dweibull(t, shape=theta[1], scale =theta[2], log = FALSE)
+ theta[6]*dweibull(t, shape=theta[3], scale =theta[4], log = FALSE))
+(1-d)*log(theta[5]*(1-pweibull(t, shape=theta[1], scale=theta[2], lower.tail=TRUE, log.p=FALSE))
+ theta[6]*(1-pweibull(t, shape=theta[3], scale =theta[4], lower.tail = TRUE, log.p = FALSE))))
}
# -- Function for MLEs of the parameters of Weibull-Weibull mixture model ----
# t: Failure time or right censored time
# d: Failure/Censored indicator, 1=failure, 0=censored
# theta:  Initial values of Parameter vector, {beta1, alpha1, beta2, alpha2, p1, p2}
# em.tiny: A small value to stop EM iteration
#-------------------------------------------------------------------
WeibWeibMix <- function(t, d, theta, em.tiny){
n <- length(t)
K <- 2
change.lik <- 0.05
maxi.em.rep <- 500
em.rep <- 1
fjti<- matrix(0, nrow = n, ncol = K)
loglik.obs <- array()
loglik.obs.old <- loglik.obs(t, d, theta)
#--------------------------------- Iteration for EM algorithm ----------------------------------------
while(change.lik > em.tiny && em.rep <= maxi.em.rep) {
# ---------------------------------- E-step: Computation of f(ti|j) as matrix or say pij ---------------
tempj1.f <- dweibull(t, shape=theta[1], scale =theta[2], log = FALSE)
tempj2.f <- dweibull(t, shape=theta[3], scale =theta[4], log = FALSE)
```

```
tempj1.c <- (1-pweibull(t, shape=theta[1], scale =theta[2], lower.tail = TRUE, log.p = FALSE))
tempj2.c <- (1-pweibull(t, shape=theta[3], scale =theta[4], lower.tail = TRUE, log.p = FALSE))
fjti[ , 1] <- (theta[5]*tempj1.f/(theta[5]*tempj1.f + theta[6]*tempj2.f))^d* (theta[5]*tempj1.c/
(theta[5]*tempj1.c + theta[6]*tempj2.c))^(1-d)
fjti[ , 2] <- (theta[6]*tempj2.f/(theta[5]*tempj1.f + theta[6]*tempj2.f))^d*(theta[6]*tempj2.c/
(theta[5]*tempj1 .c + theta[6]*tempj2.c))^(1-d)
#------------------------------------ M-step ------------------------------------------------------
p1 <- sum(fjti[, 1])/n # MLE of p1
p2 <- sum(fjti[, 2])/n # MLE of p2
# -------- To change if f(j|ti) is zero --------------
small.value <- 10^(-8)
for(ii in 1:n){
if(fjti[ii, 1] < small.value){
fjti[ii, 1] <- small.value; fjti[ii, 2] <- 1 - small.value
}
}
for(jj in 1:n){
if(fjti[jj, 2] < small.value){
fjti[jj, 2] <- small.value; fjti[jj, 1] <- 1 - small.value
}
}p
i1 <- fjti[ , 1]
pi2 <- fjti[ , 2]
# ---------------- MLEs of main parameters, except p, via survreg command ---------------
fit1 <- survreg(Surv(t, d) ~ 1, weight = pi1, dist='weibull') # Fit of Weibull for sub-population 1
beta1.hat <- 1/fit1$scale
eta1.hat <- exp(fit1$coefficient)
fit2 <- survreg(Surv(t, d) ~ 1, weight = pi2, dist='weibull') # Fit of Weibull for sub-population 2
beta2.hat <- 1/fit2$scale
eta2.hat <- exp(fit2$coefficient)
theta<- c(beta1.hat, eta1.hat, beta2.hat, eta2.hat, p1, p2) # Updated MLEs
loglik.obs[em.rep] <- loglik.obs(t, d, theta) # Updated observed data log-likelihood
change.lik <- abs(loglik.obs[em.rep] - loglik.obs.old)
loglik.obs.old <- loglik.obs[em.rep]
em.rep <- em.rep + 1
}
if(em.rep >= maxi.em.rep) {print("Algorithm did NOT converge")} # warning message if do not
converge
#------- End of E & M-steps ----------------------------
return(list(beta1=theta[1], alpha1=theta[2], beta2=theta[3], alpha2=theta[4], p1=theta[5],p2=theta[6]))
}
```

## A.2 R codes for analyzing pump failure data with Weibull-Normal-Exponential mixture model.

### A.2.1 Function for estimating parameters ----

```
library(splines)
library(survival)

#  Observed data log-likelihood function -----
#  t: lifetime variable,  d: failure/censored indicator;  theta: parameter vector
loglik.obs <- function(t, d, theta){
sum(d*log(theta[6]*dweibull(t, shape=theta[1], scale =theta[2], log = FALSE)
  + theta[7]*dnorm(t,mean=theta[3], sd =theta[4], log = FALSE)
  + theta[8]*dexp(t, rate=theta[5], log = FALSE))
  + (1-d)*log(theta[6]*(1-pweibull(t, shape=theta[1], scale=theta[2], lower.tail=TRUE, log.p=FALSE))
  + theta[7]*(1-pnorm(t, mean=theta[3], sd=theta[4], lower.tail = TRUE, log.p = FALSE))
  + theta[8]*(1-pexp(t, rate=theta[5], lower.tail = TRUE, log.p = FALSE))))
}

# em.tiny: a small value - convergence criterion for EM algorithm
WeibNormExpMix <- function(t, d, theta, em.tiny){
n <- length(t);  K <- 3;  tiny <- 10^(-8);  change.lik <- 0.05;  maxi.em.rep <- 500;  em.rep <- 1
fjti<- matrix(0, nrow = n, ncol = K);  loglik <- array()

loglik.old <- loglik.obs(t, d, theta)
# ------- Iteration for EM algorithm
while(change.lik > em.tiny && em.rep <= maxi.em.rep) {
# ----- E-step
  tempj1.f <- dweibull(t, shape=theta[1], scale =theta[2], log = FALSE)
  tempj2.f <- dnorm(t, mean=theta[3], sd =theta[4], log = FALSE)
  tempj3.f <- dexp(t, rate=theta[5], log = FALSE)
  tempj1.c <- (1-pweibull(t, shape=theta[1], scale =theta[2], lower.tail = TRUE, log.p = FALSE))
  tempj2.c <-(1-pnorm(t, mean=theta[3], sd=theta[4], lower.tail = TRUE, log.p = FALSE))
  tempj3.c <- (1-pexp(t, rate=theta[5], lower.tail = TRUE, log.p =  FALSE))

p <- c(theta[6], theta[7], theta[8])
fjti[ , 1] <- (p[1]*tempj1.f/(p[1]*tempj1.f + p[2]*tempj2.f+
p[3]*tempj3.f))^d*(p[1]*tempj1.c/(p[1]*tempj1.c + p[2]*tempj2.c+ p[3]*tempj3.c))^(1-d)
```

```
fjti[ , 2] <- (p[2]*tempj2.f/(p[1]*tempj1.f + p[2]*tempj2.f+
p[3]*tempj3.f))^d*(p[2]*tempj2.c/(p[1]*tempj1.c + p[2]*tempj2.c+ p[3]*tempj3.c))^(1-d)
fjti[ , 3] <- (p[3]*tempj3.f/(p[1]*tempj1.f + p[2]*tempj2.f+
p[3]*tempj3.f))^d*(p[3]*tempj3.c/(p[1]*tempj1.c + p[2]*tempj2.c+ p[3]*tempj3.c))^(1-d)

# ----- M-step
p1 <- sum(fjti[, 1])/n;   p2 <- sum(fjti[, 2])/n;   p3 <- sum(fjti[, 3])/n;   p <- c(p1, p2, p3)

t1.new <- array();   t2.new <- array();  t3.new <- array()
d1.new <- array();  d2.new <- array();  d3.new <- array()
fj1.new <- array(); fj2.new <- array(); fj3.new <- array()

# ----- To omit observation from LF if its weight is very small, less than tiny
j <- 0
for(i in 1:n){
if(fjti[i,1] >= tiny){
 j <- j+1;   t1.new[j] <- t[i];   d1.new[j] <- d[i];   fj1.new[j] <- fjti[i,1]
 }
}

j <- 0
for(i in 1:n){
if(fjti[i, 2] >= tiny){
 j <- j+1;   t2.new[j] <- t[i];   d2.new[j] <- d[i];   fj2.new[j] <- fjti[i, 2]
 }
}

j <- 0
for(i in 1:n){
if(fjti[i, 3] >= tiny){
 j <- j+1;   t3.new[j] <- t[i];   d3.new[j] <- d[i];   fj3.new[j] <- fjti[i, 3]
 }
}

# ----- MLEs of main parameters, except p
fit1 <- survreg(Surv(t1.new,  d1.new) ~ 1, weight = fj1.new, dist='weibull')
beta1.hat  <- 1/fit1$scale;  eta1.hat <- exp(fit1$coefficient)
fit2 <- survreg(Surv(t2.new,  d2.new) ~ 1, weight = fj2.new, dist='gaussian')
mu.hat <-  fit2$coefficient;   sigma.hat  <- fit2$scale
fit3 <- survreg(Surv(t3.new,  d3.new) ~ 1, weight = fj3.new, dist='exponential')
```

```
delta.hat <- 1/exp(fit3$coefficient)

theta<- c(beta1.hat, eta1.hat, mu.hat, sigma.hat, delta.hat, p)      # Parameter vector

loglik[em.rep] <- loglik.obs(t, d, theta)
change.lik <- abs(loglik[em.rep] - loglik.old);   loglik.old <- loglik[em.rep];   em.rep <- em.rep + 1
}                                                  # end-loop for EM iteration
if(em.rep >= maxi.em.rep){print("Algorithm did NOT converge")}      # warning message if not converge
return(list(theta.hat=as.numeric(theta), likelihood=loglik.old))
}
```

### A.2.2 Pump failure data in text format as given in Table 4.3 ----

```
Age <- c(81, 149, 245, 340, 407, 461, 629, 856, 947, 1460, 1513, 1670, 1688, 2093, 2242, 2242, 2242,
2242, 2242, 2607, 2668, 2806, 3132, 3132, 3132, 3132, 3333, 3569, 3837, 3837, 4150, 5123, 5258,
5662, 5923, 6333, 6717, 7207, 7265, 7624, 7625, 7973, 8183, 8217, 8390, 8462, 8728, 8817, 8870,
8884, 9055, 9182, 9334, 9368, 9729, 9751, 10299, 10389, 10413, 10557, 10944, 10970, 11647, 11678,
11686, 11798, 11869, 11869, 11923, 12005, 12082, 12090, 12136, 12141, 12143, 12163, 12198,
12198, 12198, 12198, 12198, 12198, 12236, 12236, 12236, 12236, 12236, 12236, 12394, 12459,
13097, 13497, 13497, 13497, 13497, 13497, 13497, 13497, 14407, 15536, 16289, 17517)

Type <- c(0, 1, 1, 1, 1, 1, 1, 0, 0, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 0, 0, 0, 0, 1, 1, 0, 0, 0, 1, 1, 0, 1, 1, 1,
1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 0, 1, 0, 0, 1, 1, 1, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1)

theta.ini <- c(1.522,  16500, 12500.5, 6500, 0.0009, 0.35, 0.25, 0.40)      # Assumed initial values of parameters
WeibNormExpMix(Age, Type, theta.ini, 10^(-6))                                # Execution of estimation-function
```

### A.2.3 Function for estimation of Adjusted Anderson-Darling value ------

```
AD.adj.KM <- function(t, d, theta){
n <- length(t);    Pr <- array();     tt <- array();     new.Pr <- array()
for(i in 1: n){
```

```
 temp.prod <- 1
for(j in 1: i){
   temp.prod <- temp.prod * ((n-j)^d[j]/(n-j+1)^d[j])
 }
 Pr[i] <- 1- temp.prod
}
nf<- 0
for(i in 1: n){
if(d[i]==1){
nf<- nf + 1
tt[nf] <- t[i]
new.Pr[nf] <- Pr[i]                                    # cdf corresponding to failed observations
  }
}
max.t <- max(t)
for(i in 1: n){
if((t[i] == max.t) && (d[i]==1)){
new.Pr[nf] <- new.Pr[nf-1]+(1-new.Pr[nf-1])*0.90    # if largest observation is failure, re-estimate
  }
}
Fnz <- new.Pr                                                  # Nonparametric cdf  (KM estimate)
Fnz[nf+1] <- 1
# ---- Parametric estimates of CDF
beta.hat <- theta[1];    eta.hat <- theta[2];    mu.hat <- theta[3];    sigma.hat <- theta[4];     delta.hat <-
theta[5]
p1.hat <- theta[6];   p2.hat <- theta[7];   p3.hat <- theta[8]
mle.F1 <- pweibull(tt, shape = beta.hat, scale = eta.hat, lower.tail = TRUE, log.p = FALSE)
mle.F2 <- pnorm(tt, mean=mu.hat, sd =sigma.hat, lower.tail = TRUE, log.p = FALSE)
mle.F3 <- pexp(tt, rate=delta.hat, lower.tail = TRUE, log.p =  FALSE)
mle.F  <- p1.hat*mle.F1 + p2.hat*mle.F2 + p3.hat*mle.F3

MLE.CDF <- mle.F
mle.F[nf+1] <- 0.999999999999              # a value close to 1
mle.F.lag <- array()                       # z[i-1]
mle.F.lag[1] <- 0
for(i in 1: nf){
mle.F.lag[i+1] <- mle.F[i]
}
```

```
Fnz.lag <- array()                    # Fnz[i-1]
Fnz.lag[1] <- 0
for(i in 1: nf){
Fnz.lag[i+1] <- Fnz[i]
}
A <- array();   B <- array();   C <- array()
A <- - mle.F- log(1-mle.F) + mle.F.lag + log(1-mle.F.lag)
B <- 2*log(1-mle.F)*Fnz.lag - 2*log(1-mle.F.lag)*Fnz.lag
C <- log(mle.F)*Fnz.lag^2 - log(1-mle.F)*Fnz.lag^2 - log(mle.F.lag)*Fnz.lag^2 + log(1-mle.F.lag)*Fnz.lag^2
C[1] <- 0          # Replace the missing value in the first row of 'C_i' with a zero.
AD.adj <- nf*sum(A,B,C)
return(list("f.time"=c(tt), "KM.F"=c(new.Pr), "MLE.F"=c(MLE.CDF), "AD.adj.value"=AD.adj))
}

theta.hat <- WeibNormExpMix(Age, Type, theta.ini, 10^(-6))$theta.hat    # MLE of parameters
AD.adj.KM(Age, Type, theta.hat)                                         # Estimation of AD value
```

### A.2.4 Function for the estimation of AIC, KS test statistic and RMSE ----

```
AIC.KS.RMSE <- function(t, d, theta.hat){
theta.ini <- c(1.522,  16500, 12500.5, 6500, 0.0009, 0.35, 0.25, 0.40)
loglik<- WeibNormExpMix(t, d, theta.ini, 10^(-6))$likelihood
n.para <- 7                           # change No. of independent parameters
AIC <- -2*loglik + 2*n.para           # AIC value
KS.ts <- max(abs(AD.adj.KM(t, d, theta.hat)$KM.F - AD.adj.KM(t, d, theta.hat)$MLE.F))          # Kolmogorov-Smirnov test statistic
RMSE <- sd(AD.adj.KM(t, d, theta.hat)$KM.F - AD.adj.KM(t, d, theta.hat)$MLE.F)
                  # Root Mean Squared Error
return(list("AIC.hat"=AIC, "KS.hat"=KS.ts, "RMSE.hat"=RMSE))
}

AIC.KS.RMSE(Age, Type, theta.hat)                  # Execution of the function
```

### A.2.5 Codes for creating figure ----

```
plot(AD.adj.KM(Age, Type, theta.hat)$f.time, AD.adj.KM(Age, Type, theta.hat)$KM.F, main="EDF and Weibull-Normal-Exponential mixture cdf", xlab=" Time (hours)", ylab="cdf", col=1, lty= 1, type="s", lwd=1)
lines(AD.adj.KM(Age, Type, theta.hat)$f.time, AD.adj.KM(Age, Type, theta.hat)$MLE.F, col=2, lty=2, type="l", lwd=2)
legend("topleft", c("KM cdf", "Weib-Norm-Expn mixture cdf"), col=c(1,2), text.col=c(1,2), lty=c(1,2), lwd=2)
```

### A.2.6 Codes for estimating optimal maintenance age by minimizing J(T)

```
q <- 0.6096       # value of q in the formula
Cn <- 80000;  Cr<- 60000;  Cp <- q*Cn+(1-q)*Cr
opt.T.JT <- function(zeta){
Cf <- Cp + zeta
TT = seq(4000, 20000, 1)     # Possible values of T, opt T would be within TT
JT <- function(T.value){       # ---- J(T)star function
WNE.F1 <- pweibull(T.value, shape=theta.hat[1], scale=theta.hat[2], lower.tail = TRUE, log.p = FALSE)
WNE.F2 <- pnorm(T.value, mean=theta.hat[3], sd=theta.hat[4], lower.tail = TRUE, log.p = FALSE)
WNE.F3 <- pexp(T.value, rate=theta.hat[5], lower.tail = TRUE, log.p =  FALSE)
WNE.F <- theta.hat[6]*WNE.F1 + theta.hat[7]*WNE.F2 + theta.hat[8]*WNE.F3
WNE.R <- 1-WNE.F
  fun.t <- function(t){
  WNE.f1 <- dweibull(t, shape=theta.hat[1], scale=theta.hat[2], log = FALSE)
  WNE.f2 <- dnorm(t, mean=theta.hat[3], sd=theta.hat[4], log = FALSE)
  WNE.f3 <- dexp(t, rate=theta.hat[5], log = FALSE)
WNE.f  <- theta.hat[6]*WNE.f1 + theta.hat[7]*WNE.f2 + theta.hat[8]*WNE.f3
return(t*WNE.f)
  }

 int.part <- integrate(fun.t, lower=0, upper = T.value)$value
 JT.value <- (Cf*WNE.F + Cp*WNE.R)/(int.part + T.value*WNE.R)
return(JT.value)
 }
TT.no <- length(TT);  JT.out <- array()
for(j in 1:TT.no){
```

```
  JT.out[j] <- JT(TT[j])
}
JT.star.opt <- min(JT.out)
for(i in 1:TT.no){
if(JT.out[i] == JT.star.opt) {T.est <- TT[i]}
}
return(list(zeta.value=zeta, opt.T=T.est, opt.JT=JT.star.opt))
}

opt.T.JT(130000)
```

# Reference


Alwasel I.A., (2009), *Statistical Inference of a Competing Risks Model with Modified Weibull Distributions*, Int. Journal of Math. Analysis, 3(19), 905-918.

Ancha X., Yincai T., (2012), *Statistical Analysis of Competing Failure Modes in Accelerated Life Testing Based on Assumed Copulas*, Chinese Journal of Applied Probability and Statistics, 28(1).

Ateya, S.F. (2014), *Maximum likelihood estimation under a finite mixture of generalized exponential distributions based on censored data*, Statistical Papers, 55(2). 311-325.

Ateya, S.F., Alharthi, A.S., (2014), *Estimation under a finite mixture of modified Weibull distributions based on censored data via EM algorithm with application*, Journal of Statistical Theory and Applications, 13(3), 196-204.

Balakrishnan N., LingM.H., (2013), *Expectation maximization algorithm for one shot device accelerated life testing with Weibull lifetimes, variable parameters over stress*,IEEE Transections on Reliability, 62, 537–551.

Balakrishnan N., So H.Y., Ling M.H., (2015), *EM algorithm for one-shot device testing with competing risks under exponential distribution*, Reliability Engineering and System Safety, ELSEVIER, 137, 129–140.

Barabadi A., (2013), *Reliability model selection and validation using Weibull probability plot-A case study*, Electric Power Systems Research, ELSEVIER, 101, 96–101.

Benaicha H., Chaker A., (2014), *Weibull Mixture Model for Reliability Analysis*, International Review of Electrical Engineering, 9(5).

Bedford T., (2005), *Competing Risk Modeling in Reliability*, Chapter 1, Strathclyde University, Glasgow, UK.

Bertholon H., Bousquet N., Celeux G., (2004), *An alternative competing risk model to the Weibull distribution in lifetime data analysis*, [Research Report] RR-5265, INRIA., pp.25.

Blischke, W.R. and Murthy, D.N.P. (2000), *Reliability*, Wiley, New York, p.18-19.

Blischke, W.R. and Murthy, D.N.P., (2000), *Reliability Modelling*, Prediction, and Optimization. John Wiley & Sons, New York.

Blischke W.R., Karim M.R. and Murthy D.N.P., (2011), *Warranty data collection and analysis*, Springer Verlag, London.

Bordes, L. and Chauveau, D., (2012), *EM and Stochastic EM algorithms for reliability mixture models under random censoring*, Hal-VousConsultezL'Archive.

Breslow, N.E. and Crowley J., (1974), *A large sample study of the life table and product limit estimates under random censorship*. Ann. Stat., 2, 437-453.

Bucar T., Nagode M., Fajdiga M., (2004), *Reliability approximation using finite Weibull mixture distributions*, Reliability Engineering and System Safety, ELSEVIER 84, 241–251.

Campbell, J.D., (1995), *Outsourcing in maintenance management – A valid alternative to self-provision*, Journal of Quality in Maintenance Engineering, 1(3), 18-24.

Castet J-F., Saleh J.H., (2010), *Single versus mixture Weibull distributions for nonparametric satellite reliability*, Reliability Engineering and System Safety, ELSEVIER, 95, 295–300.

Condra, L.W., (1993), *Reliability Improvement with Design Experiments*, New York: Marcel Dekker.

D'Agostino, R.B. and Stepfens M.A., (1986), *Goodness-of-fit Techniques*, Marcel Dekker, New York.

Dempster A.P, Laird NM, Rubin D.B., (1977), *Maximum likelihood from incomplete data via the EM algorithm (with discussion)*. J R Stat Soc B 39, 1–38.

Dixon, P.M. and Newman, M. C., (1991), *Analyzing toxicity data using statistical models for time-to-death: an introduction, In: Mental Ecotoxicology. Concepts and Applications*, edited by M. C. Newman and A. W. Mclntosh, Lewis Publishers, Chelsea, MI, pp. 207-242.

Efron, B., (1967), *The two sample problem with censored data*. Proc. Fifth Berkeley Symp., 4, 831-853.

Elfaki FAM, Daud IB, Ibrahim NA, Abdullah MY and Usman M., (2007), *Competing risks for reliability analysis using Cox's model*, Engineering Computations, 24(4), 373 – 383.

El-kelany G.A.,(2015), *Three Independent Competing Risks Model under Type-I Censoring*, Global Journal of Mathematics, 5(2), 511-523.

Elmahdy E.E., (2015), *A new approach for Weibull modeling for reliability life data analysis*, Applied Mathematics and Computation, ELSEVIER, 250, 708–720.

Erişoğlu Ü, Erişoğlu M. and Erol H., (2011), *A mixture model of two different distributions approach to the analysis of heterogeneous survival data*, International Journal of Computational and Mathematical Sciences 5, 2, 75-79.

Feizjavadian S.H., Hashemi R., (2015), *Analysis of dependent competing risks in the presence of progressive hybrid censoring using Marshall–Olkin bivariate Weibull distribution*, Computational Statistics and Data Analysis, ELSEVIER, 82, 19–34.

Feroze N., (2016), *Bayesian Inference of a Finite Mixture of Inverse Weibull Distributions with an Application to Doubly Censoring Data*, Pak.j.stat.oper.res. XII (1), 53-72.

Hartley H.Q., (1958*), Maximum likelihood estimation from incomplete data*. Biometrics 14, 174-194.

He Z., Wang Y., Li J., Gong S., Grzybowski S., (2013), *New Mixed Weibull Probability Distribution Model for Reliability Evaluation of Paper-oil Insulation.*

Iskandar I., (2016), *Competing risk models in reliability systems, a gamma distribution model with Bayesian analysis approach*, IOP Conf. Series: Materials Science and Engineering 114, 012098.

Iskandar I. and Gondokaryono Y.S., (2016), *Competing risk models in reliability systems, a weibull distribution model with Bayesian analysis approach*, IOP Conf. Series: Materials Science and Engineering 114, 012064.

Jiang R., (2015), *Introduction to Quality and Reliability Engineering*, Springer Series in Reliability Engineering.

Jiang S and Kececioglu D., (1992), *Maximum Likelihood Estimates, from Censored Data, for Mixed-Weibull Distributions*, IEEE TRANSACTIONS ON RELIABILITY, 41(2), 248-255.

Jiang R. and Murthy DNP., (2003), *Study of n-Fold Weibull Competing Risk Model*, Mathematical and Computer Modeling, 38, 1259-1273.

Johansen, S., (1978), *The product limit estimate as a maximum likelihood estimate*. Scand. J. Stat., 5, 195-199.

Kao, J.H.K., (1959), *A graphical estimation of mixed Weibull parameters in life-testing of electron tubes*, Techno metrics, 1, 389-407.

Kaplan, E.L. and Meier, P., (1958), *Nonparametric estimation from incomplete observations*, Journal of the American Statistical Association, 53, 457-481.

Karim, M.R., Ahmadi, A. and Murthy, D.N.P., (2015), *Modeling of maintenance data*, Presented at ICRESH-ARMS Conference, 2015, Lulea, Sweden.

Karim, M.R. and Suzuki, K., (2005), *Analysis of warranty claim data: a literature review*, International Journal of Quality & Reliability Management, Vol. 22, No. 7, pp. 667-686.

Karim, M.R., Yamamoto, W. and Suzuki, K., (2001), *Statistical analysis of marginal count failure data*, Lifetime Data Analysis, Vol. 7, pp. 173-186.

Kundu D., Sarhan A.M., (2006),*Analysis of Incomplete Data in Presence of Competing Risks among Several Groups*, IEEE Transactions on Reliability, 55(2).

Lawless, J.F., (1998), *Statistical analysis of product warranty data*, International Statistical Review, Vol. 66, pp. 41-60.

Lee G. and Scott C., (2012), *EM algorithms for multivariate Gaussian mixture models with truncated and censored data*, Computational Statistics & Data Analysis, 56(9): 2816–2829.

Li G. and Lin C., (2009), *Analysis of two-sample censored data using a semi-parametric mixture model*, Acta Math Sin Engl Ser.; 25(3), 389–398.

Li S, Yin Q, Guo P and Lyu M.R., (2007), *A hierarchical mixture model for software reliability prediction*, Applied Mathematics and Computation, ELSEVIER, 185, 1120–1130.

Little R.J.A. and Rubin D.B., (1987), *Statistical Analysis with Missing Data*. John Wiley & Sons, Inc., New York.

Marin J.M., Rodriguez-Bernal M.T. and Wiper M.P., (2005), *Using Weibull Mixture Distributions to Model Heterogeneous Survival Data*, Communication in Statistics-Simulation and Computation, 34, 673-684.

McLachlan G.J, Krishnan T., (1997), *The EM Algorithm and Extensions*. John Wiley & Sons, Inc., New York.

Meeker, W.Q. and Escober, L.A., (1998), *Pitfalls of accelerated testing*, IEEE Transactions on Reliability, 47, 114-118.

Meeker, W.Q. and Escober, L.A., (1998), *Statistical Methods for Reliability Data*, Wiley, New York.

Meier, P., (1975), *Estimation of a distribution function from incomplete observations*. In Perspectives in Probability and Statistics, J. Gani, Ed. Sheffield, England.: Applied Probability Trust.

Mendenhall W. and Hader R. J., (1958), *Estimation of parameters of mixed exponentially distributed failure time distributions from censored life test data*, Biometrica, 45, 504-520.

Murthy D.N.P, (2010), *New research in reliability, warranty and maintenance*. In: Proceedings of the 4th Asia-Pacific international symposium on advanced reliability and maintenance modeling, pp 504–515

Murthy, D.N.P. and Djamaludin, I., (2002), *New product warranty: a literature review*, International Journal of Production Economics, Vol. 79, pp. 231-260.

Murthy, D.N.P., Karim, M.R. and Ahmadi, A., (2011), *Data management in maintenance outsourcing*, Reliability Engineering and System Safety, 142, 100–110.

Murthy, D.N.P., Xie, M. and Jiang R., (2004), *Weibull Model*, John Wiley & Sons, Inc.

Nelson W., (1982), *Applied Life Data Analysis*, John Wiley and Sons, New York.

Nelson W., (2009), *Accelerated Testing Statistical Models*, Test Plans, and Data Analysis, John Wiley & Sons, Inc.

Noor F., Aslam M., (2013*), Bayesian inference of the inverse Weibull mixture distribution using Type-I censoring*, Journal of Applied Statistics, 40(5), 1076-1089.

Park C., Kulasekera K. B., (2004), *Parametric inference of incomplete data with competing risks among several groups*, IEEE Transactions on Reliability, 53, 11–21.

Peterson, A.V., (1977), *Expressing the Kaplan-Meier estimator as a function of empirical sub survival functions*, Journal of the American Statistical Association, 72, 854-858.

Razali A.M., Al-Wakeel A.A., (2013), *Mixture Weibull distributions for fitting failure times data*, Applied Mathematics and Computation, ELSEVIER, 219, 11358–11364

Ruhi S., (2015), *Application of mixture models for analyzing reliability data: a case study*. Open Access Libr J 2, e1815.

Ruhi, S., Sarker, S. and Karim, M.R., (2015), *Mixture Models for Analyzing Product Reliability Data: A Case Study*, SpringerPlus, 4:634.

Sarhan A.M., Alameri M., Al-Wasel I., (2013), *Analysis of a Competing Risks Model with Generalized Weibull Distributions*, Pakistan Journal of Statistics, 29(3), 271-281

Stephens, M. A., (1974), *EDF Statistics for Goodness of Fit and Some Comparisons*, Journal of the American Statistical Association, 69, pp. 730-737.

Stephens, M. A., (1976), *Asymptotic Results for Goodness-of-Fit Statistics with Unknown Parameters*, Annals of Statistics, 4, pp. 357-369.

Stephens, M. A., (1977), *Goodness of Fit for the Extreme Value Distribution*, Biometrika, 64, pp. 583-588.

Stephens, M. A., (1977), *Goodness of Fit with Special Reference to Tests for Exponentially*, Technical Report No. 262, Department of Statistics, Stanford University, Stanford, CA.

Stephens, M. A., (1979), *Tests of Fit for the Logistic Distribution Based on the Empirical Distribution Function*, Biometrika, 66, pp. 591-595.

Stuart, A. and Ord, J.K., (1991). *Kendall's Advanced Theory of Statistics*, Vol. 2, 5th ed., Oxford University Press, New York.

Suzuki, K., (1985), *Non-parametric estimation of lifetime distribution from a record of failures and follow-ups*, Journal of the American Statistical Association, Vol. 80, pp. 68-72.

Suzuki K., (1985), *Estimation of lifetime parameters from incomplete field data.* Techno metrics 27, 263–271.

Suzuki, K., Karim, M.R., and Wang, L., (2001), *Statistical analysis of reliability warranty data*, in N. Balakrishnan and C.R. Rao (Eds), Handbook of Statistics: Advances in Reliability, Elsevier Science, Vol. 20, pp. 585-609.

Tarum C.D., (1999), *Classification and Analysis of Weibull Mixtures*, Sae Technical Paper Series, 01-0055.

Titterington, M., Smith, A.F.M., and Makov, U.E., (1985), *Statistical Analysis of Finite Mixture Distribution*, Wiley, New York.

Ünal E.A., Wu S.J., Bekci M., Kinaci I., Kuş C., (2015),*Statistical inference for Weibull distribution based on competing risks data under progressive type I group censoring,* Journal of Selçuk University Natural and Applied Science, 4(1), 164-173.

Witherell, C.E., (1994), *Mechanical Failure Avoidance*, McGraw-Hill, New York.

Yáñez S., Escobar L.A., González N., (2014), *Characteristics of two Competing Risks Models with Weibull Distributed Risks*, Acad. Colomb. Cienc, 38(148), 298-311.

Zhang Q., Hua C., Xu G., A., (2014), *mixture Weibull proportional hazard model for mechanical system failure prediction utilizing lifetime and monitoring data*, Mechanical Systems and Signal Processing, ELSEVIER, 43 ,103–112.

Zhang T., Dwight R., (2013), *Choosing an optimal model for failure data analysis by graphical approach*, Reliability Engineering and System Safety, ELSEVIER, 115 111–123.