# RELIABILITY ESTIMATION CONSIDERING CUSTOMER USAGE RATE PROFILE & WARRANTY CLAIMS

A Thesis
Submitted to the Graduate Faculty
of the
North Dakota State University
of Agriculture and Applied Science

By

Shah Mohammad Limon

In Partial Fulfillment of the Requirements
for the Degree of
MASTER OF SCIENCE

Major Department:
Industrial & Manufacturing Engineering

September 2014

Fargo, North Dakota

# North Dakota State University
## Graduate School

**Title**

RELIABILITY ESTIMATION CONSIDERING CUSTOMER USAGE RATE PROFILE & WARRANTY CLAIMS

**By**

Shah Mohammad Limon

The Supervisory Committee certifies that this ***disquisition*** complies with North Dakota State University's regulations and meets the accepted standards for the degree of

**MASTER OF SCIENCE**

SUPERVISORY COMMITTEE:

Dr. Om Prakash Yadav
Chair

Dr. Val Marinov

Dr. AKM Bashir Khoda

Dr. Yildirim Suzen Bora

Approved:

September 12, 2014 — Date

Dr. Canan Bilen-Green — Department Chair

# ABSTRACT

Providing more realistic reliability prediction based on small proportion of failed population or test data has always been a challenging task. Manufacturers rely heavily on reliability prediction for designing warranty plan. Further, to predict warranty claims for the remaining warranty period, it is important to have more realistic reliability assessment by considering a larger proportion of the population or the maximum possible information on the remaining population. However, generally this information is not readily available and is very difficult to gather on the scattered population. In this work, we propose to use customer usage rate profile to generate censored usage data on the remaining population that do not have any failure and warranty claim yet. We intend to use field data available such as warranty claims, field failures, recall data, and maintenance data to develop usage rate profile and subsequently estimate censored usage time. Finally, reliability estimation methodology is developed considering both censored data and field failure data to provide more reasonable reliability prediction for the remaining warranty period. The proposed methodology is demonstrated considering real life data from a big manufacturing company.

**ACKNOWLEDGEMENTS**

Foremost, I would like to express my humble and sincere gratitude to my advisor Professor Dr. Om Prakash Yadav for his continuous support on my study and research, for his patience, motivation, and immense guidance towards this research and writing of the thesis.

Besides my advisor, I would like to thank the rest of my thesis committee: Professor Dr. Val Marinov, Professor Dr. AKM Bashir Khoda, and Professor Dr. Yildirim Suzen Bora, for their encouragement and insightful comments. Also, my sincere thanks to Jason Muscha and Russell Honeyman for their continuous support on providing data and insightful comments on writing. This research cannot be conducted without their never ending significant technical cooperation.

Last but not the least; I would like to thank all of my family members and friends for their support and encouragement, and especial thanks to my beloved wife, Iffat Toufique, for her support, patience, and unlimited trust.

## DEDICATION

Dedicated to my parents Shah Mohammad Mijan & Monowara Begum, and my lovely sister Mafruha Akter Lia.

## TABLE OF CONTENTS

## LIST OF TABLES

## LIST OF FIGURES

# CHAPTER 1. INTRODUCTION

In today's competitive marketplace, appropriate product performance is one of the key quality factors to business success. In engineering perspective, there exists uncertainty due to many reasons such as: design, materials, manufacturing, and others fault into product performance. The measure of this product performance is denoted by product reliability. According to O'Connor (2002), reliability is "the probability that an item will perform a required function without failure under stated conditions for a stated period of time." As quality assurance become a popular strategy to capture the market share, the manufacturer started to provide warranty heavily from 1980s. The warranty represents the liability of the premature failure under customer use.

The warranty packages provided by manufacturers differ depending on product types, geographic area, and several other factors. For example, in USA automobile manufacturers generally provide 3 years or 36000 miles bumper to bumper warranty whichever comes first; but in most of Asian countries, unlimited miles are provided with few years of warranty (Alam and Suzuki 2009). The warranty policy also differs on repair upon failure, replacement upon failure, full free by manufacturer, or pro-rate cases. Sometimes, a manufacturer provides an extended warranty that customers can purchase. Regardless of packages and policies, warranty is always a liability as it incurs cost for the manufacturers by means of money and goodwill. For example, in the United States, manufacturers spend more than $25 billion per year to resolve warranty related issues (Manna et al. 2007). To reduce this huge amount of warranty related costs, manufacturers have always been interested in knowing their product reliability and failure causes so that this knowledge can be utilized to improve the design of critical components as well as the product. Further, the knowledge of product reliability will help manufacturers to designing warranty policies, budgeting for warranty, and managing spare parts effectively.

During product design stage, components, sub-system, or total product is asses by accelerated testing for reliability estimation. In accelerated testing, product is undergo with higher operating condition to expedite the failure and by appropriate mathematical modeling time to failure and reliability in normal condition is estimated. One of the drawback of this accelerated test result is they do not capture the actual product performance under real usage environment. In a contrary, field failure data provide more reliable information about actual distribution compared to laboratory data (Suzuki 1985b, Karim and Suzuki 2007, Oh and Bai 2001). Field data capture actual usage profile and the combined environmental exposure that are difficult to simulate in laboratory (Rai and Singh 2003). Among the readily available field data, warranty claims reflect the actual product performance in customer's hand.

Though readily available warranty data capture the actual product performance, but this data also has several drawbacks such as the sparse nature of data, incomplete, unclean, and delays and mistakes in reporting (Rai and Singh 2003). Incompleteness of data refers to unavailability of actual usage information of both failed and un-failed population. Additionally, early warranty failure data represent a very small fraction of the entire population, whereas most of the warranty related decisions are made based on reliability estimates derived considering this small fraction of the failed population. If someone estimates the reliability solely based on failure data then inferences drawn based on this estimate will usually be biased. It is, therefore, important to develop a more effective reliability assessment method that captures maximum possible information related to both failed and un-failed populations and provide more realistic reliability estimation.

There are many researches have attempted to estimate reliability and lifetime distribution considering part of un-failed population. For example, one of the most prominent efforts is using follow-up or supplementary survey data of un-failed population. A fraction of un-failed data is collected from follow-up studies and a pseudo-likelihood estimation approach is developed considering these follow-up data. Both parametric and non-parametric methods were attempted to estimate reliability (Suzuki 1985a, Suzuki 1985b, Kalbfleisch and Lawless 1988, Hu et al. 1998). Suzuki (1987) used non-homogeneous Poisson process (NHPP), and Alam and Suzuki (2009) used only failure data while considering censored usage time as unknown. Oh and Bai (2001) proposed to incorporate after warranty field failure data to estimate reliability and lifetime distribution. Wu (2013) provided a very informative review on coarse warranty data analysis that covers approaches and methods used to estimate product reliability considering warranty claims and supplementary data. Kalbfleisch and Lawless (1988) provided some guidelines for collecting follow-up (supplementary) data. However, in many cases the follow-up data collection is not an easy task. It costs money and time, and provides partial (incomplete) information regarding un-failed population. In some cases, it is impossible to collect the follow-up data. Therefore, life estimation for un-failed population is not an easy task and presents a major challenge in obtaining more realistic reliability estimates.

To overcome these problems, this work proposes a usage rate based approach to capture the relevant data (accumulated usage) related to the un-failed population. Hu et al. (1998) also argued that the accumulated usage of the product is more relevant for engineering analysis purposes than the age. It is also assumed that customer usage rate is independent of failure for both failed and un-failed population. The product usage rate is estimated using field data that includes warranty claims data, maintenance data, and other follow-up data such as recall data, survey data, and other supplementary data if available. However, warranty claims and maintenance data are the most dominating among all available data as these are readily available with dealers and other service stations. The censored data (accumulated usage) is generated for un-failed population considering usage rate and age of the product. For censored data generation purposes, the proposed approach considers actual age of the individual (un-failed) units currently in service as well as age distribution of the population. Finally, a reliability estimation approach is suggested considering two different cases; where in the first case both failure and censored data follow the same distribution (Weibull) and in the second case these data sets follow two different (Weibull and lognormal) distributions.

The rest of the work is organized as follows. In chapter 2, an extensive literature review is given on warranty data analysis. In chapter 3, proposed methodology on usage rate profile development is described. In chapter 4, reliability estimation model based on maximum likelihood method (MLE) is discussed in detail. In chapter 5, a case study with real warranty data is analyzed by proposed methodology. In chapter 6, conclusion of the work is given based on the case study results and future research guidelines are provided.

## CHAPTER 2. LITERATURE REVIEW

Warranty data analysis has been used extensively for early detection of reliability problems, finding opportunities for design improvement, and estimating field reliability. Estimation of field reliability is extremely important to manufactures for selecting appropriate warranty policy, establishing maintenance infrastructure, and designing spare parts inventory system. Since warranty data reflect real operating environment and usage rate, they are richer in information content then test data collected from laboratories.  The warranty data analysis approaches can be categorized as one-dimensional approach and two-dimensional approach. The one dimensional approach consists of age-based and usage-based analysis techniques where warranty limit is defined by either age or usage only. On the other hand, the two-dimensional approach considers both product age and accumulated usage simultaneously for reliability analysis purpose.

In age-based analysis, the product age (calendar time), also known as time-in-service, is consider for estimating product reliability. Several researchers have proposed age-based warranty data analysis approaches (Kalbfleisch et al. 1991, Lawless 1998, Karim et al. 2001, Karim and Suzuki 2007).  More recent works include estimate lifetime distribution of warranty claims such as fitting Weibull distribution on small number of failure claims (Ion et al. 2007), estimating mixed distribution (Majeske 2003), and estimating life distribution considering sales delays (Wilson et al. 2009).  The "age" refers to calendar time since the product is delivering to customers. In many cases, the warranty data are available in different aggregated groups and in this situation exact age of product is difficult to found. To overcome these issues, three kinds of research found related to aggregated warranty claims: age, claims date, and sales date related aggregated data. In case of age also known as type I aggregated claims, total claims are aggregated based on different age interval and this age interval can be constant or variable. Kalbfleisch et al. (1991) proposed a non-parametric estimator for expected number of claims of age aggregated data considering NHPP and Kalbfleisch and Lawless (1996) extend the work for variable age interval. In cases of claims date also known as type II aggregated claims, exact date of claims is not known rather total number of claims is aggregated for a specific period of time. Suzuki et al. (2000, 2001) and Karim et al. (2001) uses NHPP model for repairable items and employed expectation maximization (EM) algorithm for estimates number of claims. In case of sales delay also known as type III aggregated claims, exact sales date is not known rather total number of sales is aggregated for a specific period of time. Lawless and Kalbfleisch (1992) introduced an estimator using NHPP for expected number of claims of sales delay aggregated data and Wang et al. (2002) introduced parametric and non-parametric MLE of the claims for repairable and non-repairable cases respectively.

The usage-based approach considers accumulated usage time or accumulated mileage (for automobile) as a measure of failure time. The major challenge in using usage-based approaches is obtaining censoring time for the surviving population that has not reported any failure. This causes difficulty in estimating the life distribution in the absence of censored population. Moreover, Wu (2012) claims that usage time distribution of non-failed products different than failed products, which makes reliability estimation task even more difficult. Nevertheless, the usage time is more useful and important for engineering analysis and reliability improvement (Hu et al. 1998). One of the most common approaches to deal with unknown censored data is supplementary data analysis (Suzuki 1985a, Kalbfleisch and Lawless 1988). Oh and Bai ( 2001) proposed to estimate lifetime distribution with additional field data and Attardi et al. (2005) introduced mixed Weibull regression model to estimate failure time of incomplete data. Suzuki (1987) proposed NHPP while usage time of un-failed product cannot observe and Suzuki et al. (2008) proposed both parametric and semi-parametric method to estimate product field reliability without including un-failed product. Vintr and Vintr (2007) surveyed to customers for analyzing their usage behavior and intensity.

For automobile, many researches consider both the age and usage time into their analysis that known as two-dimensional approach. The rational for considering two-dimensional approach is that automobile warranty coverage considers both age and mileage limits, and it is therefore important to develop methods capturing both age and usage time. Two-dimensional warranty data analysis literature can be classified into three different categories: marginal approach, bivariate approach, and composite scale approach. The marginal approach considers usage rate as random variable, which can be modeled either as discrete variable or as continuous variable with a density function. For example, Lawless et al. (1995) considered the occurrence of warranty claims for automobile when both age and mileage affect failure. Their model assesses the dependence of failures on age and mileage and estimates survival distributions and rates from warranty claims data. Kleyner and Sanborn (2006) present a model where the usage time is a primary variable and the mileage accumulation is estimated from field return data. Their approach accounts for an observed reduction in the number of warranty claims in the second half of the warranty period. The bivariate approach directly estimates a joint bivariate distribution from warranty data. Singpurwalla and Wilson (1993) develop a bivariate failure model for automobile warranty data indexed by time and mileage. Several other researchers consider age and usage time together into the field reliability estimation such as, Yang and Zaghati (2002), Jung and Bai (2007), Lawless et al. (2009), and many others. The composite scale approach integrates the two scales (age and usage) to create a single composite scale and failures are modelled as a counting process using this approach (Gertsbakh and Kordonsky 1998; Duchesne and lawless 2000). Ahn et al. (1998) and Iskandar and Blischke (2003) used power law process with the new time scale as a model for the reliability analysis of a repairable system. Moreover, early warranty data also used to detect the reliability issues of the product. For example, Lu (1998) uses early failure data to estimate and asses the product reliability. Wu and Meeker (2002) also propose to use early warranty data to identify reliability problems. Authors suggest stratifying and monitoring data more frequently so that, it increases the chance to detect manufacturing or other reliability problems.

Though warranty data represents product usage under real environmental condition, there are several issues related to warranty data such as, aggregate claims, delays in reporting and sales, or incomplete censored data that introduce more uncertainty in reliability analysis approach. The aggregated claims have already been discussed in age-based analysis section of this chapter in above. The delays are mainly refers two kinds: reporting delays and sales delays. Both the reporting and sales delays are divided into two categories: type I and type II. The type I reporting delays is a delay by the manufacturer to report it after failure occurs and mostly it delays for verifying the claims. There are two approaches to deal with type I reporting delay. According to first one eliminate the reported cases (1992) and according to second approach incorporating reporting delay probabilities into the analysis. Lawless and Kalbfleisch (1992) and Kalbfleisch et al. (1991) proposed estimation of expected number of claims considering given reporting delay probabilities. Also, a NHPP model is used to estimate reporting lag distribution and expected claims number (Kalbfleisch et al. 1991, Suzuki et al. 2000, Kalbfleisch and Lawless 1991). The type II reporting delays is a delay by the customers not to report failure to the manufacturer immediately but while reported it is updated to claims immediately. Rai and

Sing (2006) proposed a non-parametric approach to estimate hazard rate functions for type II reporting delay warranty claims.

The sales delays occur when the exact date of sales is unknown and this makes difficult to find the product or time-in-service. The larger sales delay also increases the chances of warranty claims (Robinson and McDonald 1991.). In type I sales delay, the manufacturer do not know the exact date of sales after its production date but only for failed items the failure times and censoring time may be obtained through warranty claims verification process (Suzuki et al. 2001, Hu et al. 1998, Ion et al. 2007). Hu et al. (1996) proposed non-parametric estimation and Karim and Suzuki (2004) proposed NHPP model to estimate lifetime distribution for type I sales delay. Among the parametric approaches, Ion et al. (2007) and Karim (2008) introduced the Weibull and the lognormal distribution respectively to fit type I sales delay. In type II sales delay, both failed and un-failed items might not have exact censoring time because of unknown sales date and this situation occurs from type II aggregated claims (Mohan et al. 2008). Baxter (1994) introduced a non-parametric approach for lifetime distribution and Crowder and Stephens (2003) introduced moment based estimator for sales delay data. Lim (2003) and Karim and Suzuki (2004) proposed to estimate the distribution of the sales delay considering multinomial and poission model respectively. Wilson et al. (2009) proposed parametric approach to estimate lifetime distribution considering both sales and claims reporting delay. Rai and Singh (2006) consider the customer behavioral factor into the warranty claims that makes soft failure into reporting delay. To deal with censored data, additional follow-up data is incorporated with warranty claims.

For usage based analysis, though usage time is more useful, however, to get the usage information for non-failed population is a challenging task. To overcome this problem, a follow-up study of the non-failed population is proposed by Suzuki (1985a) and Kalbfleisch & Lawless (1988). A random survey is conducted to collect the usage time and other relevant information for a portion of non-failed population. To conduct this survey, total number of population should be known. Lawless and Kalbfleisch (1992) reported few issues about follow-up survey data and guidelines for collecting survey and follow-up data. Suzuki (1985a) conducted a follow-up studies to collect non-failed usage information and proposed a modified Kaplan-Meier estimator for reliability analysis. In another work, pseudo-likelihood function has been developed to estimate the lifetime distribution from follow-up data, and both parametric ((Suzuki 1985b, Kalbfleisch & Lawless 1988) and nonparametric (Hu et al 1998) approach were used to estimate lifetime distribution from pseudo-likelihood function. This pseudo-likelihood method that uses follow-up data also extended to covariate analysis, where a regression model is developed between lifetime and dependents explanatory variables. For example, Karim & Suzuki (2007) took region, type of products, and failure modes as covariate with age based lifetime analysis and assume Weibull as a lifetime distribution. One of the major problems with the follow-up study data is it takes time, costs money, and sometimes it becomes impossible to collect information through customer survey. Yang and Zaghati (2002) also mention that survey data is expensive in many cases, and therefore, warranty claims is a solution for mileage accumulation model.

In follow-up studies it consider warranty claims along with a portion of non-failed population data, however, total population is not considered either in age-based, usage time based, or two-dimensional approaches. Park (2005) considers non-failed censored information as missing data and use popular expectation maximum (EM) algorithm to estimate the ML function. Alam and Suzuki (2009) proposed a method to estimate the lifetime distribution considering non-failed population usage is unknown. To incorporate the all non-failed censored population, we propose a usage rate based warranty analysis. In our approach, though we also consider non-failed population usage time is unknown, but, usage time is then estimate by usage rate profile. Usage rate profile develops from field data where majority information came from warranty claims and procedure is describe in next section. Known failure usage and estimated non-failed usage then utilize for lifetime parameter and reliability estimation. For more model and approaches, two review papers on warranty analysis are suggested to read (Wu 2012, Wu 2013).

# CHAPTER 3. CUSTOMER USAGE RATE PROFILE

In this section, proposed methodology of getting usage time of un-failed population is described. Customer usage profile is estimated by field data and as field data collected from different sources, a procedures is proposed for improve data quality. From the usage profile, it is also described the usage rate distribution and accumulated usage data for un-failed population. Different scenarios are described based on censored usage time of un-failed population.

Customer usage rate profile provides more relevant information about the usage behavior of the entire population of the product. The term “usage” might vary from product to product, such as for an automobile, mileage is used to capture usage; for a copy machine, the number of copies is termed as usage; and for utility equipment, operation hour is used to capture the usage. It is, however, difficult to get the usage rate information for the entire population. Several researchers have assumed that accumulated usage for an automotive product is different for failed and censored population, and in majority of cases the accumulated usage of censored population is unknown (Alam and Suzuki 2009). However, in reality the usage rate is independent of failure and hence it is fair to assume that the usage rate will be the same for failed and surviving populations. The reason behind this assumption is that usage rate depends on user behavior and not on failure of the product. Although occasionally severe failures might affect the usage rate to some extent, it does not have a major impact on the average usage rate. Lawless et al. (1995) also used a similar approach where failure time is assumed to be independent of mileage accumulation rate.

Collecting the customer usage rate information has been a major challenge. However, with the advancement in communication networks and service data management, now it is relatively easier to gather usage rate related data from several sources. These sources include recall data, maintenance data, warranty data, and online connect data. Partial surveys and recall data have been used as a part of censored data to estimate the parameters for field performance (Suzuki 1985b, Kalbfleisch and Lawless 1988). Regular maintenance data can be collected from dealers or maintenance departments where customers bring their product for regular maintenance during and after the warranty period. It is important to keep in mind that this information is for the un-failed product because it is collected during regular maintenance only. Another source of usage data is warranty claims database. Since, warranty claims represent only failure data during the warranty period, it provides relatively better customer usage information, the number of hours accumulated, and other types of usage data that may be utilized to estimate usage rate of the product. Though warranty data has several shortcomings, it contains a great source of information regarding the actual performance of the product. Another source of customer usage data is online connecting data. As technology grows, it is possible to track the usage of product utilizing microchip to capture real time utilization. This approach is expensive and sometimes the customer might not allow tracking of their usage behavior, but it provides a possible mode to collect usage information in many possible cases (Hong and Meeker 2010, Meeker and Hong 2013). Production and sales data can be utilized to capture the time that a product is in service also known as product age. In order to estimate usage rate, it is essential to gather both the accumulated usage and the time in service data accurately.

Since accumulated usage data is collected from multiple sources, the quality and uncertainty in the data will vary significantly from one source to another. For example, survey data might have higher uncertainty and more quality related issues as compared to warranty claims and maintenance data. Also the possibility of human error in data collection is much higher if data is gathered through a survey. This variation in the quality of data collected from multiple sources poses the greatest challenge in estimating usage rate by combining data from several sources.  We, therefore, recommend using appropriate tools such as fuzzy logic or a neural network model to combine data coming from different sources and estimate usage rate. Fig. 1 shows the conceptual data filtering and processing model.

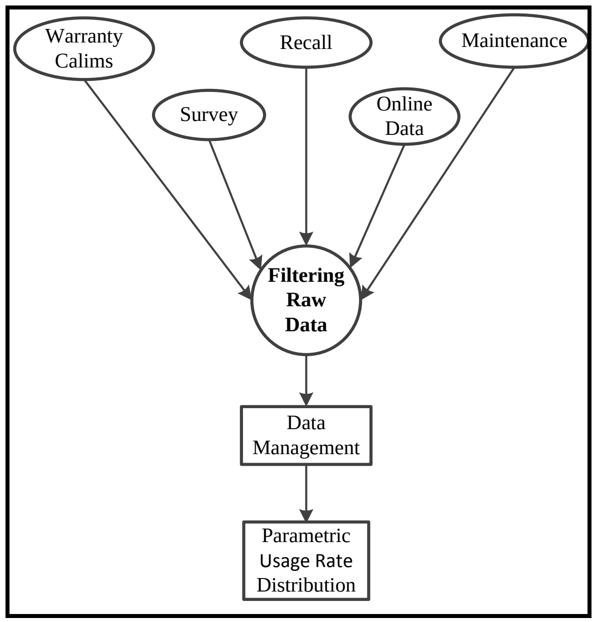



Fig. 1: Processing the different quality of data

In order to process the data collected from several sources, it is important to identify the variables of interest such as product age, accumulated usage, product model, and other related variables. The outliers (extremely large or small and infeasible data points) in each category of variables should be removed. Also we need to screen out all other data points or variables that do not match with the product model under consideration. After gathering all relevant data, the usage rate is calculated considering the accumulated usage from the available data and product age (or time in service). If needed, the usage rate data is also filtered to take out infeasible data points. For example, if usage rate is calculated as actual use per day, then any usage rate data showing more than twenty four hours per day should be removed as it represents infeasible data points.  Any other kinds of outliers observed in the data should be analyzed and removed from data set if necessary. This filtering process will improve the quality of data and present the final data set as if it is collected from a single source.

Though data processing and filtering takes care of removing outliers and ensuring uniformity in data, there is still a possibility of having some discrepancy as most of these data come from multiple sources with varying levels of quality. We, therefore, strongly recommend establishing a mechanism to ensure the quality of data such as statistical process control method suggested by Jones-Farmer et al. (2014). The intent is to develop a data quality control system similar to the manufacturing process, where refined data are treated as a final product and raw data coming from various sources are considered as input. Once all the variables of interest are in hand, the usage rate is calculated. The usage rate $u_i$ for the $i^{th}$ vehicle out of *n* vehicle is obtained as

$$u_i = \frac{w_i}{t_i}, \; i = 1, 2, \ldots, \mathrm{n} \tag{1}$$

where $w_i$ and $t_i$ represent total usage and time in service for the $i^{th}$ vehicle, respectively. It is important to note that total usage for any given vehicle can be obtained from warranty claim, maintenance and service records, and any other source available to manufacturers for getting vehicle related information. The time in service information can be obtained from the sales records of dealers. The following section discusses the estimation of usage rate distribution parameters.

### 3.1. Usage Rate Distribution

In this work, a parametric distribution analysis is applied for usage rate. Since usage rate varies from customer to customer, it is important to treat it as a random variable and establish an appropriate usage rate distribution. Earlier studies (Hu et al. 1998, Lu 1998) show that for automobiles the usage rate is generally linear over time and follows the lognormal distribution.

The usage rate data histograms of two different industrial utility equipment show the lognormal distribution fit (see Fig. 2), which essentially supports the earlier assumption on usage rate distribution. Further, the usage rate also differs from market segment to market segment as depicted in these two different distribution fits supporting our argument of treating usage rate as a random variable.

We, therefore, model usage rate with the lognormal distribution and estimate the model parameters. Considering a random variable *U* that follows a distribution with probability distribution function (pdf) *f(u)*, the likelihood function will be given as:

$$L(\theta) = \prod_{i=1}^{n} f(u_i, \theta) \quad (2)$$

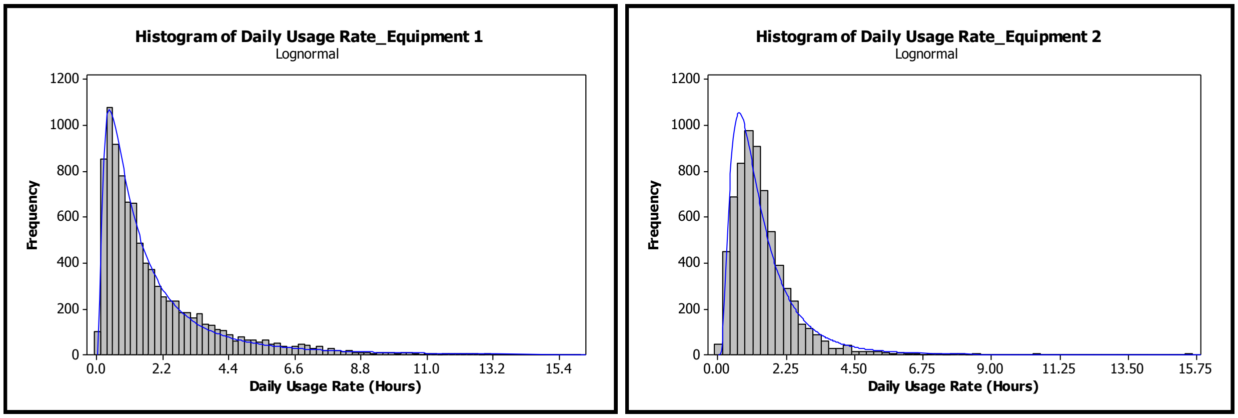
By taking the natural logarithm on both sides of Eqn. (2), the log likelihood function is written as;

$$log\, L(\theta) = \sum_{i=1}^{n} log\, f(\, u_i, \theta) \quad (3)$$

where θ represents the parameters of interest that need to be estimated. For the lognormal distribution, these parameters are the location and the scale (standard deviation) θ = (μ, σ), respectively. The maximum likelihood estimates of these two parameters are given as:

$$\hat{\mu}_u = \frac{\sum_{i=1}^{n} log(u_i)}{n} \quad (4)$$

$$\hat{\sigma}_u^2 = \frac{\sum_{i=1}^{n} (log(u_i) - \hat{\mu}_u)^2}{n} \quad (5)$$




(a) (b)

Fig. 2: Lognormal fit usage rate for utility (a) equipment 1(b) equipment 2

Generally, the distribution model is fitted based on the available data to estimate model parameters. However, the usage rate data gathered usually represents a small fraction of the surviving population and the estimation of model parameters based on this small fraction of the population will have higher uncertainty. We, therefore, propose to use a parametric bootstrap resampling method (Efron 1979, Meeker and Escobar 1998) for estimation of model parameters. The bootstrap resampling method provides robust estimation of model parameters with tighter confidence intervals. The bootstrap method performs resampling of the same sample size and range as original sample, estimates the model parameters for each sample, and then provides the final estimation of model parameters by taking an average of all the sample parameters. Fig. 3 shows the graphical representation of the bootstrap method. The final estimates of model parameters are then used to generate censored usage time data for the surviving population. The next section provides a detailed discussion on generating censored usage time.

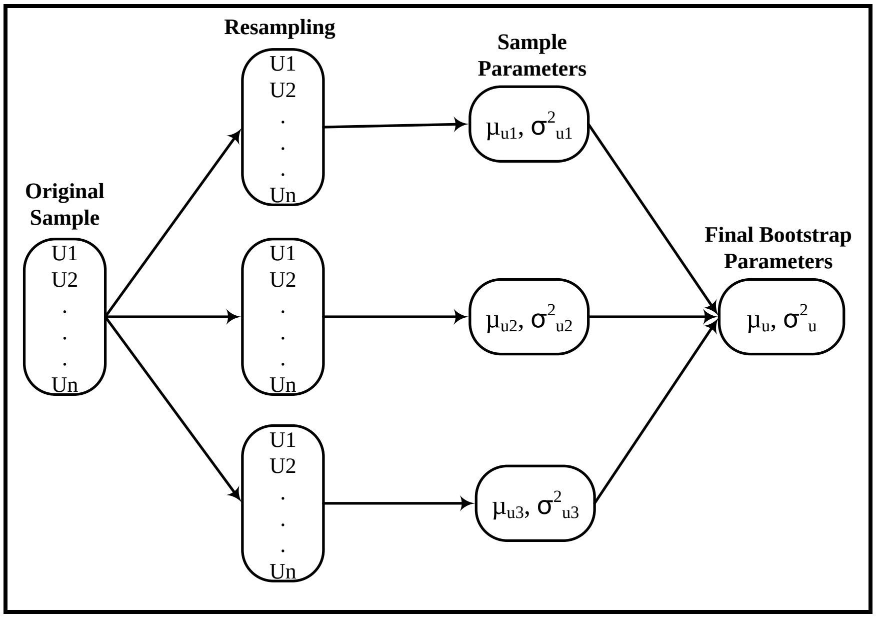



Fig. 3: Bootstrap resampling method for parametric estimation

### 3.2. Censored Accumulated Usage Data

To getting the censored accumulated usage time for all un-failed population is one of the main focuses of this research. As we assume usage rate is indifferent of failure so, it is same for failed and un-failed population. Thus, once the usage rate distribution parameters are estimated, the next step is to generate censored data for the surviving population using these parameter estimates and age of the product. Assuming at any given point of time, the age of the product (or time in service) is also defined as a random variable, then the accumulated total usage is given as:

$$T_c = U \times A \tag{6}$$

Where $T_c$ denotes the censored usage time, *U* and *A* represent random variables of usage rate and product age, respectively. The product age can be estimated by using product manufacturing and sales related information, which is easily available in the warranty database or with dealers. The measurement unit of *A* is calendar time such as days, weeks, or months, whereas usage rate is measured as usage per calendar time such as mileage per day or usage hours per day.

Since the accumulated usage (or censored) time $T_c$ is a product of two random variables, it is also treated as a random variable. Considering usage rate and product age (time in service) as two independent variables, the expected value and variance of the accumulated usage time $T_c$ can be determined by the following equations (Kapadia et al. 2005):

$$E(T_c) = E(U) \times E(A) \;=\; \mu_U \;\mu_A \tag{7}$$

$$Var(T_c) = Var(U)Var(A) + Var(U)(\mu_A)^2 + Var(A)(\mu_U)^2 \tag{8}$$

If the distribution types and distribution model parameters of usage rate and product age are known, it becomes easier to generate censored data for the surviving population. To generate censored data for un-failed population, we consider two different scenarios as given below:

Scenario 1: the actual age ($A_i$) of each unit surviving in the field is known and

Scenario 2: the product age $A$ of the surviving population is treated as a random variable.

**Scenario 1:** In this case, the product age $A_i$ represents the actual calendar time of each unit within the warranty period. The actual age can be derived from product manufacturing details, sales date, and total time spent in the field with some level of certainty but individual units might have a different age or time in service period. This allows us to consider the age of the product population in field as a variable. Considering the actual age of individual units $A_i$, the accumulated usage time is calculated as:

$$T_{c\,i} = \; U_i \times A_i \tag{9}$$

where the random variable usage rate follow is assumed to follow lognormal distribution $U_i \sim \; logn(\hat{\mu}_u, \hat{\sigma}_u^2)$ and $A_i$ is the actual age for the $i^{th}$ unit. The censored data for the surviving population can be generated using Eqn. (9), which requires random data generation on usage rate for the surviving population. The expected value and variance of usage rate data for the surviving population following the lognormal distribution can be determined by considering the following Eqns. (Hogg et al. 2012):

$$E(U) = e^{\left(\mu_u + \frac{\sigma_u^2}{2}\right)} \tag{10}$$

$$Var(U) = e^{(2\mu_u + \sigma_u^2)}\left(e^{\sigma_u^2} - 1\right) \tag{11}$$

One can also estimate the expected value and variance of the random variable *A* from the available product age data of individual units. Once these two parameters of both variables *U* and *A* are known, Eqns. (7) and (8) can be used to estimate the parameters of censored time distribution and subsequently generate censored time for the surviving population.

**Scenario 2:** In the second scenario, we treat the population age as a random variable that follows a specific probability distribution function. Generally, units are produced based on product demand or production capacity and subsequently end up in the field after selling them to customers. For example, auto companies produce a certain number of units every month and sell those units to customers in the market. At the same time, a certain amount of units will have spent enough time in the market and will be getting out of the warranty period. If we visualize this continuous process of a certain number of units being produced and getting into the market per unit time (week or month) and also almost a similar number of units are going out of the warranty period, it almost represents a steady flow process where on average there are a similar number of units moving through the system. We believe that this scenario can be reasonably modeled as a uniform distribution function where one parameter represents warranty time length and the other parameter captures the average number of units entering and/or leaving the warranty period.

Considering that the distribution of usage rate follows the lognormal distribution $U_i \sim\ logn(\hat{\mu}_u, \hat{\sigma}_u^2)$ and the product age follows the uniform distribution $A_i \sim\ unif(b, c)$, Eqn. (9) can be used to generate censored time data. The distribution parameters for the usage rate distribution can be estimated using Eqns. (10) and (11), whereas the distribution parameters for the uniform distribution are given as:

$$E(A) = \frac{b+c}{2} \tag{12}$$

$$Var(A) = \frac{(c-b)^2}{12} \tag{13}$$

Using the distribution parameters of these two distributions, random data sets can be generated for both random variables. The number of data points in each random data set should be equal to the surviving population size. These two random number data sets can then be used to generate censored time for the surviving population using Eqn. (9). Alternatively, given that the distribution parameters for both random variables are known, Eqns. (7) and (8) can be used to estimate the parameters of the accumulated usage (censored) time. These estimated parameters can then be used to generate censored data for the surviving population.

Once censored data are available, the maximum likelihood method can be used to estimate model parameters of the combined data set for estimating product reliability. There exist few difficult to getting usage rate distribution but proposed data refinement and bootstrap method reduces the variation and statistical biasness. Moreover, different scenarios will help to understand the actual accumulated usage data and to estimate the overall reliability that is discussed in next section.

# CHAPTER 4. RELIABILITY ESTIMATION

From chapter 1, we know that reliability is a time dependent function. To estimate the reliability it is necessary to know about the usage hours of each failure product and usage hours of un-failed product. The usage hours for failure product can readily available from warranty claims and usage hours for un-failed product is getting by using the methodology described in chapter 3. In this section, we also proposed to model maximum likelihood function considering different and same distribution of un-failed population.

For reliability estimation, we assume that field failure data follow the Weibull distribution as suggested in the literature (Meeker and Escobar 1998). The probability distribution and the survival (reliability) function of a Weibull random variable are given as follows:

$$f(t) = \frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1} e^{-\left(t/\eta\right)^{\beta}} \tag{14}$$

$$S(t) = \bar{F}(t) = 1 - F(t) = e^{-\left(t/\eta\right)^{\beta}} \tag{15}$$

where $\beta$ and $\eta$ are the shape factor and the characteristics life, respectively. To assess product reliability, the estimation of distribution parameters is the most critical step in warranty data analysis, especially when the data is acquired from different sources and censored data follows different distributions. In this work, we propose to look into those possibilities and use the maximum likelihood method for estimating model parameters.

The simplest scenario is when there are only failure data and no censored data. The maximum likelihood function for failure data only is given as:

$$L(\theta) = \prod_{i=1}^{r} f(t_i, \theta) \tag{16}$$

Here $r$ is the number of failures. However, the necessity to provide a more realistic reliability estimate requires not only failure data but also to capture the maximum possible information related to the surviving population. The inclusion of this information as censored data results into a more complex data analysis problem, especially when we consider the different distribution function for both data sets. For our investigation purpose, we developed two different likelihood functions to include both failure and censored data in reliability estimation efforts. In the first case, both failure time and censored usage time data are assumed to follow the Weibull distribution whereas in the second case we consider the lognormal distribution for censored data and the Weibull distribution for failure data.

**Case 1:** When both the failure time and the censored time follow the Weibull distribution, the likelihood function is given as:

$$L(\theta) = \prod_{i=1}^{r} f(t_i, \theta) \prod_{j=r+1}^{n} S\left(t_{cj}, \theta\right) \tag{17}$$

where $r$ is the number of failures, $n$ is the total number of data points, θ is the parameter of interest, and $f(t_i, \theta)$ and $S(t_{cj}, \theta)$ are the probability distribution function and the survival function, respectively. Using the probability density function and survival function of Weibull distribution given in Eqns. (14-15), Eqn. (17) can be written as:

$$L(\theta) \quad = \prod_{i=1}^{r} \frac{\beta}{\eta}\left(\frac{t_i}{\eta}\right)^{\beta-1} e^{-({t_i}/{\eta})^{\beta}} \prod_{j=r+1}^{n} e^{-({t_{cj}}/{\eta})^{\beta}} \tag{18}$$

After taking the logarithm of both sides of Eqn. (18), the log –likelihood function is given as:

$$log\, L(\theta) = r\, log\, \beta - r\, log\, \eta + (\beta - 1) \sum_{i=1}^{r} log\left(\frac{t_i}{\eta}\right) - \sum_{i=1}^{r} \left(\frac{t_i}{\eta}\right)^{\beta} - \sum_{j=r+1}^{n} \left(\frac{t_{cj}}{\eta}\right)^{\beta} \tag{19}$$

**Case 2:** In the second case, we consider that censored usage time follows the lognormal distribution and failure data follow the Weibull distribution. Both failure time $T$ and censored time $T_c$ are continuous random variables and are treated as independent variables. For failure time *T*, the probability distribution and survival functions are given in Eqns. (14) and (15) respectively. For censored usage time data $T_c$, the probability distribution and the survival functions are written as:

$$g(t_c) = \frac{1}{\sigma t_c \sqrt{2\pi}}\ e^{\left[-\frac{1}{2}\left(\frac{ln(t_c)-\mu}{\sigma}\right)^2\right]} \quad (20)$$

$$\bar{G}(t_c) = 1 - \ \Phi\left[\frac{ln(t_c)-\mu}{\sigma}\right] = \bar{\Phi}\left[\frac{ln(t_c)-\mu}{\sigma}\right] \quad (21)$$

where $\Phi(.)$ is the cumulative distribution function of the standard normal distribution.

The likelihood function when both failure time and censored time follow two different distributions is given as (Alam and Suzuki 2009):

$$L(\theta) = \prod_{i=1}^{r} f(t_i,\theta)\ \bar{G}(t_i,\theta)\ \ \times \left[\int_0^\infty F(t_c)\ g(t_c)\ dt_c\right]^{n-r} \quad (22)$$

This model considers censored time as an unknown and hence the integral part of the likelihood function makes the estimation of model parameters extremely difficult. Since the proposed approach generates censored usage time for the surviving population and hence considers censored time as known, the likelihood function given in Eqn. (22) can be re-written as (Lawless 2003):

$$L(\theta) = \prod_{i=1}^{r} f(t_i,\theta)\ \bar{G}(t_i,\theta) \prod_{j=r+1}^{n} g\left(t_{cj},\theta\right)\ \bar{F}(t_{cj},\theta) \quad (23)$$

Using probability distributions and survival functions of both Weibull and lognormal distributions, Eqns. (23) can be written as:

$$L(\theta) \ \ = \prod_{i=1}^{r} \frac{\beta}{\eta}\left(\frac{t_i}{\eta}\right)^{\beta-1}\ e^{-({t_i}/{\eta})^\beta}\ \ \bar{\Phi}\left[\frac{ln(t_i)-\mu}{\sigma}\right]\ \ \prod_{j=r+1}^{n} \frac{1}{\sigma t_{cj}\sqrt{2\pi}}\ e^{\left[-\frac{1}{2}\left(\frac{ln\left(t_{cj}\right)-\mu}{\sigma}\right)^2\right]}\ \ e^{-({t_{cj}}/{\eta})^\beta} \quad (24)$$

After taking the logarithm of both sides of Eqn. (24), the log –likelihood function can be obtained as:

$$\log L(\theta) = r \log \beta - r \log \eta + (\beta - 1) \sum_{i=1}^{r} \log \left(\frac{t_i}{\eta}\right) - \sum_{i=1}^{r} \left(\frac{t_i}{\eta}\right)^{\beta} + \sum_{i=1}^{r} \log \overline{\Phi} \left(\frac{\log t_i - \mu}{\sigma}\right) - \sum_{j=r+1}^{n} \left(\frac{t_{cj}}{\eta}\right)^{\beta} - (n-r) \log \sqrt{2\pi} - (n-r) \log \sigma - \sum_{j=r+1}^{n} t_{cj} - \sum_{j=r+1}^{n} \frac{1}{2} \left(\frac{\log t_{cj} - \mu}{\sigma}\right)^2 \qquad (25)$$

To find the estimate of model parameters, $\beta, \eta, \ \mu, and\ \sigma$, we need to maximize the log-likelihood function of Eqns. (19) and (25). However, for both equations it is impossible to achieve closed form solutions; it is, therefore, necessary to solve the log-likelihood function by using an appropriate numerical method. Modern statistical software *R* is used to find the MLEs by numerical method. A non-linear built-in optimization function *optim* that is based on an algorithm provided by Nelder & Mead (1965) is used to maximize the log-likelihood function. The *R* code is written for each log-likelihood function and then the *optim* function is used for maximizing the likelihood function. The *optim* solution also provides the *hessian* matrix at the maximum point, which can further be used to determine the fisher information matrix. It develops a 95% confidence interval for each parameter estimate by using the fisher information matrix. In the *optim* solution, the initial value of the estimate is important because it affects the convergence of the solution. Therefore, historically known values or close estimates based on some initial input needs to be used to address this problem.

Once the distribution parameters are estimated, the field reliability is estimated using Eqn. (15). It is important to note that censored usage time information is to be used to update the distribution parameters for providing a more realistic reliability estimation of the product in the field. Fig. 4 depicts the proposed framework for estimating usage rate and product reliability considering both failure and censored populations.

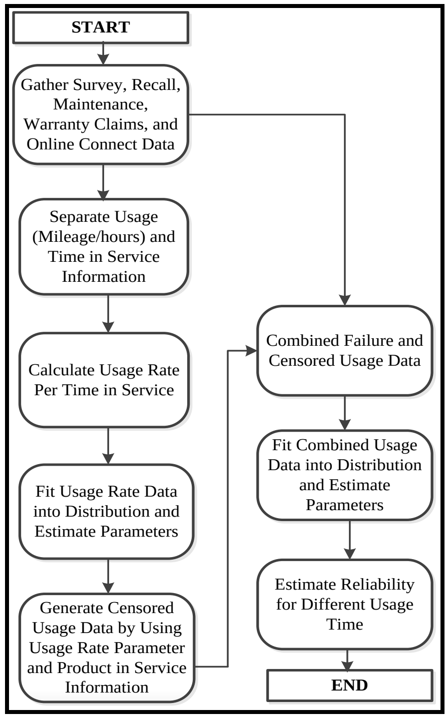



Fig. 4: Framework to estimate reliability form field data

# CHAPTER 5. A CASE EXAMPLE: UTILITY EQUIPMENT

To demonstrate the applicability of the proposed approach, we consider the field data of real utility equipment. The product is used in construction, maintenance, agriculture, and other application areas. The manufacturer provides a twelve (12) month warranty with unlimited usage hours. The product is launched into market starting in 2009 and all the claims up to August 2013 are recorded assuming that all failures within the warranty period have been reported and non-reported items are considered as censored population. The product enters into the market in a staggered way as shown in Fig. 5. To protect the proprietary nature of the information, product details regarding product name and failure modes are not disclosed. Further, the actual failure data have been modified for demonstration purposes.

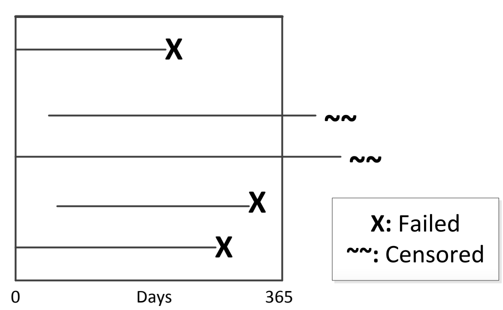



Fig. 5: Failed and censored data with staggered entry

The available field data include warranty claims, maintenance, and recall data. Each type of field data contains the date of sale, date of failure (maintenance or recall), accumulated machine hours, and name of the failed component. Table 1 shows a sample of the field database. Some of the included claim data was out of the warranty period but considered in estimating the usage rate to increase the sample size and provide better estimates. Further, the field data came from different sources with some level of variability in quality of data. Data screening was performed by removing outliers, infeasible data points, or any data points that showed negative time in service. The final data set included a total of 9004 field claims with 6570 of the claims within warranty period, 1120 claims beyond warranty period, and 1314 recall and maintenance data. These refined field data were then used to calculate the usage rate for each individual unit using Eqn. (1) considering the accumulated usage (machine) hours and age of the product. The usage rate data provided a good fit to the lognormal distribution (see Fig. 6), which supports our initial assumption regarding usage rate distribution. The usage rate of this small fraction of the population is then used to estimate the usage rate of the entire population. The bootstrap resampling is used to get robust parameters of the usage rate distribution. Table 2 shows the estimated parameters of usage rate distribution for the entire population. These estimated parameters of the lognormal distribution are then used to generate censored usage time data for the surviving population.

Table 1: A sample of field database

| Model No | Serial No | Accumulated Usage Hours | Failure Code/Mode (if applicable) | Delivery Date | Failure/Other Record Date | Comments |
|---|---|---|---|---|---|---|
| AB | AB 123 | 120 | Mechanical | 1/10/2013 | 4/5/2013 | Under warranty |
| CD | CD 456 | 35 | --- | 3/10/2013 | 3/25/2013 | Recall |
| . | . | . | . | . | . | . |
| . | . | . | . | . | . | . |
| . | . | . | . | . | . | . |

Table 2: Estimated usage rate parameters value

| Usage rate | Parameters | Estimate | Lower 95% CI | Upper 95% CI |
|---|---|---|---|---|
| Lognormal | $\hat{\mu}_U$ | 0.1821 | 0.1607 | 0.2034 |
| | $\hat{\sigma}_U$ | 1.0361 | 1.0221 | 1.0500 |

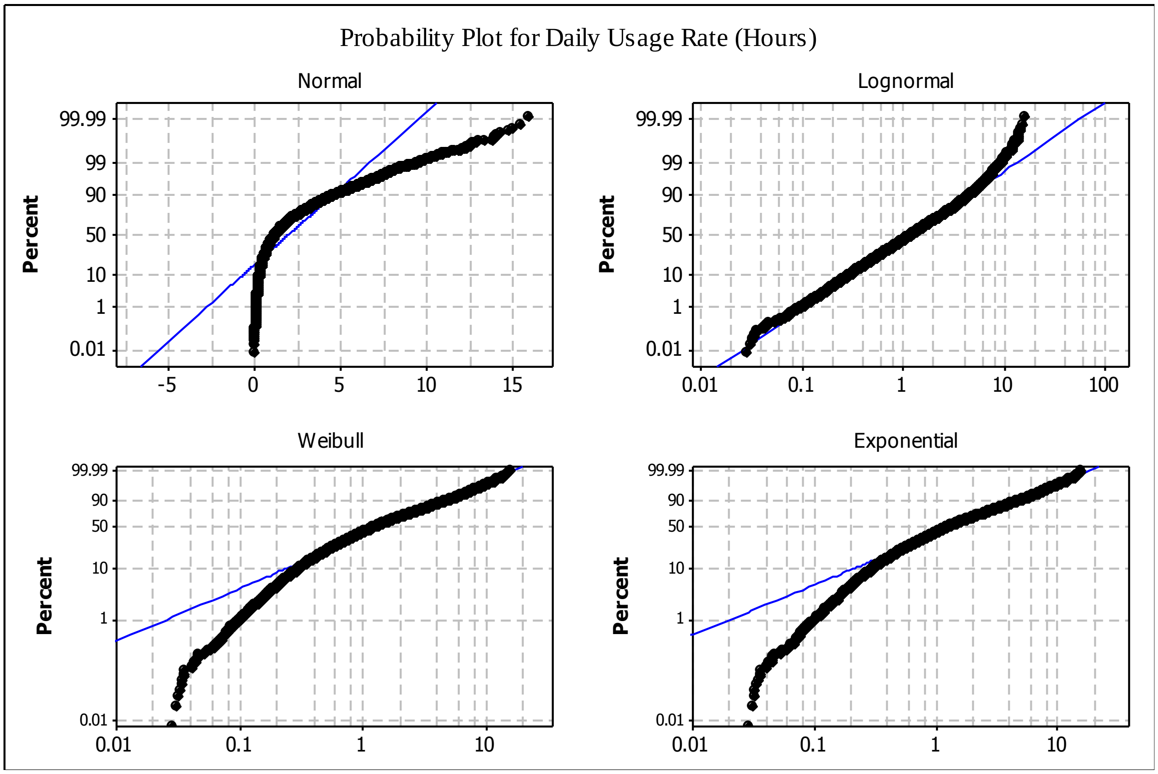



Fig. 6: Probability plot for usage rate distribution

For generating censored usage time for the surviving population, we considered the units manufactured and sold during the year 2011. Table 3 shows a sample of the product built database where product manufacturing information and warranty end date are reported. For the units manufactured and sold during the year 2011, there were 780 claims recorded in the warranty data out of a total 2636 individual units in the field. It is important to note that only the first failure for each product is consider in this analysis. For generating censored time for the remaining un-failed population within warranty period, we considered two different cases as discussed earlier. In the first case, the actual age or time in service is extracted from product built information, and the censored usage time is generated for the remaining un-failed population using Eqns. (7-11). In the second case, it is assumed that the population age follows the uniform distribution. Considering both the population age and usage rate as two random variables following different distributions, the censored usage time data are generated using Eqns. (7-8) and (10-13). Once the censored data set for the un-failed population is available, lifetime parameters are estimated considering both failure and censored data. Again two different scenarios were considered with respect to the censored data distribution where in one case we assumed both censored and failure data follow the Weibull distribution and in the second case we assumed that censored data follow the lognormal distribution. Realizing that both MLE equations (Eqns. (19) and (25)) do not provide closed form solutions, the statistical software *R* is used for ML estimation. Tables 4-5 show the estimated parameters for different scenarios and cases. Using these parameters and the survival function equation, system reliability is estimated for different usage times. Fig. 6 shows the reliability behavior of the system for the different scenarios discussed in the paper.

Table 3: A sample of product built database

| Model No | Serial No | Manufacturing Info. | Retailed or Not | Dealer Info. | Built Data | Delivery Data | Warranty End |
|---|---|---|---|---|---|---|---|
| AB | AB 123 | Plant A | Yes | Dealer G | 12/15/2012 | 1/10/2013 | 1/10/2014 |
| CD | CD 456 | Plant B | Yes | Dealer K | 12/20/2012 | 3/10/2013 | 3/10/2014 |
| . | . | . | . | . | . | . | . |
| . | . | . | . | . | . | . | . |
| . | . | . | . | . | . | . | . |

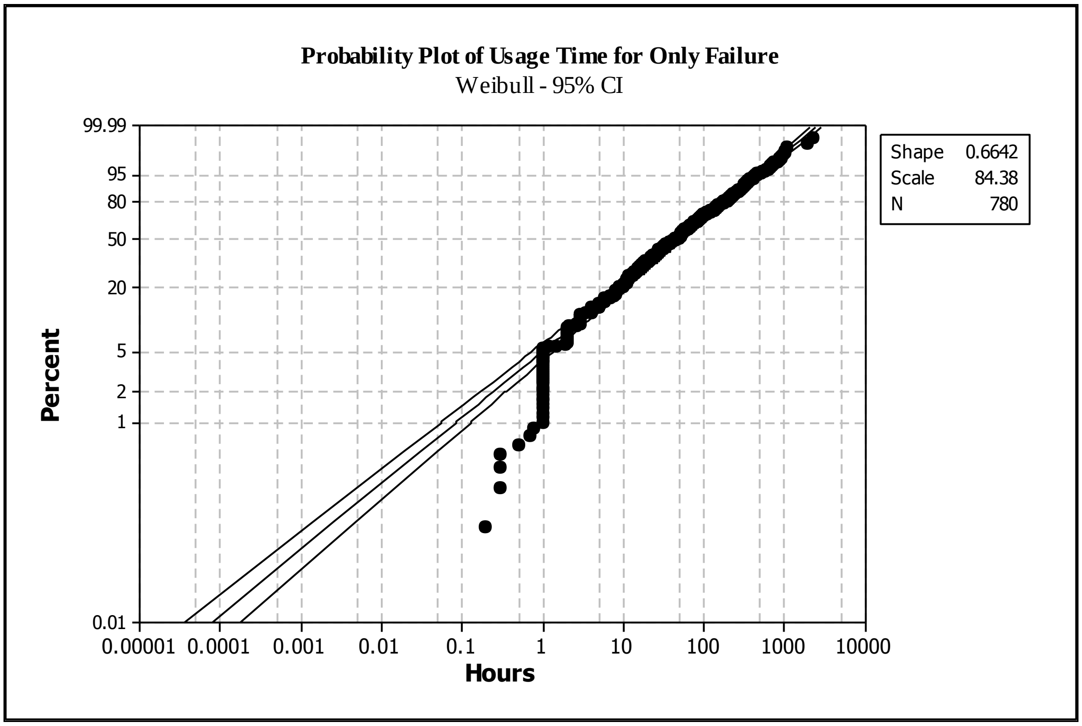



Fig. 7: Probability plot considering only failure population

Table 4: Estimated lifetime parameters considering Weibull distribution

| Weibull Analysis | Parameters | Estimate | Lower 95% CI | Upper 95% CI |
|---|---|---|---|---|
| Only Failure | $\hat{\beta}$ | 0.6645 | 0.6289 | 0.7000 |
| | $\hat{\eta}$ | 84.39 | 74.97 | 93.81 |
| Scenario 1 Case 1 | $\hat{\beta}$ | 0.4874 | 0.4574 | 0.5163 |
| | $\hat{\eta}$ | 2229.99 | 1810.95 | 2649.03 |
| Scenario 2 Case 1 | $\hat{\beta}$ | 0.4852 | 0.4561 | 0.5143 |
| | $\hat{\eta}$ | 2330.49 | 1887.45 | 2773.53 |

Table 5: Estimated lifetime & usage parameters considering Weibull and lognormal distribution

| Weibull-Lognormal Analysis | Parameters | Estimate | Lower 95% CI | Upper 95% CI |
|---|---|---|---|---|
| Scenario 1 Case 2 | $\hat{\beta}$ | 0.4883 | 0.4594 | 0.5172 |
| | $\hat{\eta}$ | 2214.36 | 1799.94 | 2628.78 |
| | $\hat{\mu}_{T_c}$ | 5.35 | 5.30 | 5.40 |
| | $\hat{\sigma}_{T_c}$ | 1.18 | 1.14 | 1.21 |
| Scenario 2 Case 2 | $\hat{\beta}$ | 0.4829 | 0.4547 | 0.5111 |
| | $\hat{\eta}$ | 2242.58 | 1830.59 | 2654.56 |
| | $\hat{\mu}_{T_c}$ | 5.38 | 5.33 | 5.43 |
| | $\hat{\sigma}_{T_c}$ | 1.17 | 1.14 | 1.21 |

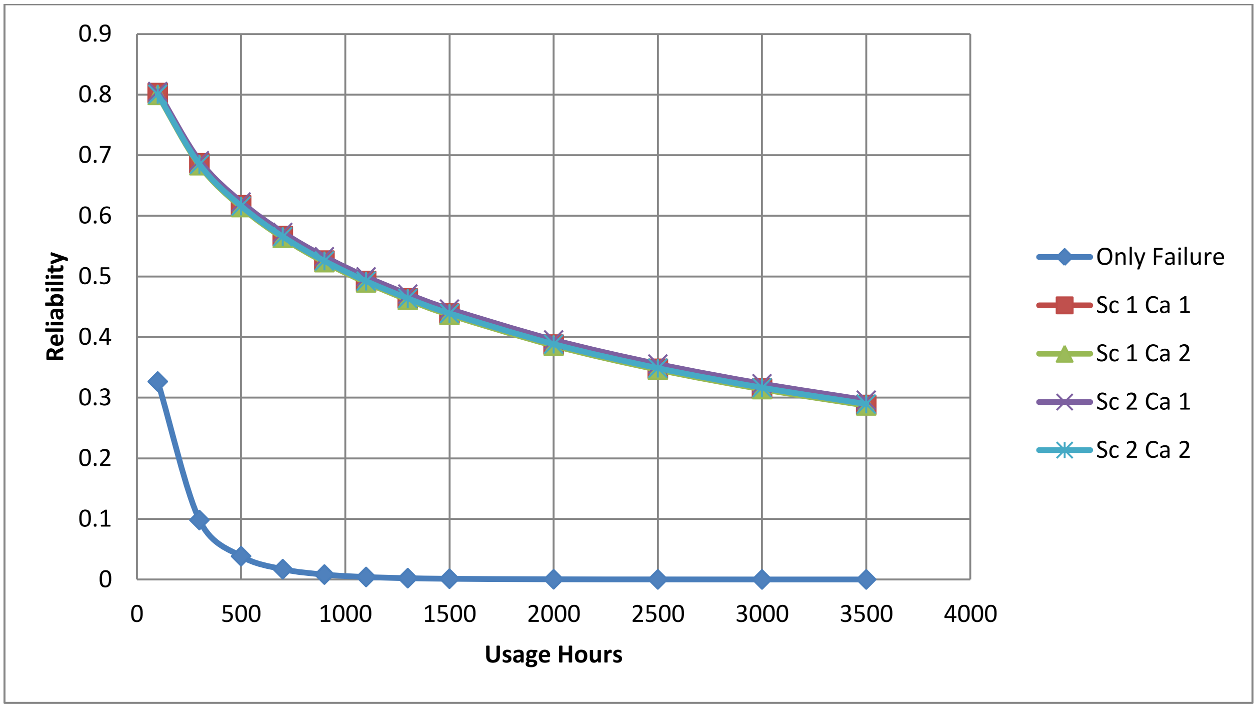



Fig. 8: Reliability comparison among different approaches

Fig. 6 and Tables 4-5 show the Weibull parameters for both failure data as well as the combined data set indicating early failure issues. It is found from the estimated results that for all cases the Weibull shape parameter value is approximately 0.50, which implies infant mortality rate (β~0.50<1). One of the reasons for infant mortality or decreasing failure rate is an immature product design. As this analysis was carried out within the second year of product launch, there is very high possibility that design will still has some deficiencies. The detailed failure analysis input will help the design community to further improve the product design by eliminating current failure modes. Other factors that might play a significant role in early failures are manufacturing and quality related issues. The in-depth analysis of the fundamental root causes of early failure problems is important to improve the reliability of the product. As given in Tables 4-5, the inclusion of the censored population information into the analysis increases the other Weibull parameter, characteristics life, significantly. This indicates the importance of incorporating information related to the surviving population for obtaining more realistic reliability estimates. However, the inclusion of additional information into the reliability analysis does not change the shape parameter of Weibull distribution, which confirms with the shape parameter property.

When generating censored data for surviving populations, two scenarios were considered to capture the age of the population. First we considered the actual age of all individual units and in the second case age is considered as a random variable that follows the uniform distribution. Further, in each given scenarios, two different cases were considered where in the first case both failure and censored data follow Weibull distribution and in the second case it is assumed that the censored data follow the lognormal distribution. The estimated values of the characteristic life show significant change from scenario one to scenario two (see Tables 4-5). In scenario one where we considered the actual age of all surviving units, the estimated characteristic life values were 2229.99 for case one and 2214.36 for case two. However, the estimated characteristic life value slightly increased to 2330.49 for case one and 2242.58 for case two when we treat the population age as a random variable following the uniform distribution. This clearly indicates that the appropriate consideration of the population age while generating censored data for the surviving population is an important criterion.  It further highlights that a consideration of an appropriate distribution for the population age provides more realistic estimates and hence, it is not worth spending time and energy in computing the actual age of each individual unit in the field. Further, the chances of making errors or extracting unrealistic age data are very high if one attempts to estimate the age of each individual unit in the population. This insight certainly helps avoid putting unnecessary efforts in extracting actual age related information from the product built database or field warranty database. On the other hand, our analysis shows that there is not much difference in the estimated values of the characteristic life from case one to case two. This essentially shows that consideration for different distributions of the censored data set does not have a significant impact on the parameter estimation.

The careful analyses of results clearly emphasize that providing a reliability assessment purely based on failure data is unrealistic and biased as it excludes useful information related to a larger proportion of the population, namely the surviving population. Fig. 7 shows a big gap between the reliability estimates based on pure failure data and a combined data set that includes censored time for the surviving population. This certainly supports the concern raised by several researchers and practitioners in the past on sole dependence on failure data analysis in the decision making process. Our further investigation on the process of generating censored data highlights the impact of random variable product age (A) on the reliability assessment. As shown in Fig. 7, the scenario two, wherein product age variable follows the uniform distribution, provides almost same reliability estimates as compared to scenario one wherein we considered the actual age of each individual unit in the field. This certainly cautions practitioners not to spend critical resources in extracting actual age data from the database but motivates them to provide a more realistic distribution of population age to obtain a more accurate reliability assessment. Our analysis did not show significant differences in the parameter estimates and reliability assessment from case one to case two when we considered the two different distributions, Weibull and lognormal distributions, for the censored data. Although consideration that the censored data follow the Weibull distribution makes the analysis process simpler, but it does not alter final outcome significantly. On the other hand, considering that the usage rate follows the lognormal distribution is important for generating censored time for the surviving population, which makes the parameter estimation process somewhat more complex. However, the final reliability assessment results do not show a significant departure from each other, which give us some level of confidence to conclude that the distribution type of censored time is not a significant factor in reliability assessment based on warranty data.

# CHAPTER 6. CONCLUSION

The proposed framework provides a more realistic reliability assessment methodology based on warranty data that includes failure time as well as censored time of the surviving population. The censored time for the surviving population is estimated considering usage rate that essentially captures the customer usage behavior. The inclusion of censored time of the surviving population in the parameter estimation has improved the reliability estimates significantly. The proposed approach considers different scenarios and cases to study the effects of product age distribution and censored data distribution. Our analysis shows that the censored data distribution assumption may not have much impact on the final reliability assessment results but consideration of the appropriate age distribution is important. The incorporation of the surviving population related information in reliability assessment provides more accurate and unbiased estimates that could be very helpful to manufacturers in managing spare parts production and inventory.

The major challenge for the proposed approach is to gather more accurate information related to the surviving population and to generate censored time using this information. However, the availability of this information from different sources having various forms and the support of advanced communication and data management technology has been a great motivator to carry this work forward.

In our future research work, we propose to develop more refined methodology for generating censored data of the surviving populations, estimating the remaining life of the surviving populations, and developing a framework for spare parts management using the remaining life information. As we have the distribution of only un-failed population also, this can be used to estimate the remaining life to the component or total product. This remaining life can be synchronizing with spare sprats production and inventory management. Other methods such as, regression and expectation maximization (EM) algorithm can be utilize for un-failed population estimation and estimate the efficiency of each methods.

**REFERENCES**


Alam MM, Suzuki K. Lifetime Estimation Using Only Failure Information From Warranty Database. IEEE Transactions on Reliability 2009; 58(4): 573–582.

Ahn CW, Chae KC, Clark GM. Estimating parameters of the power law process with two measures of failure time. Journal of Quality Technology 1998; 30:127–132.

Attardi L, Guida M, Pulcini G. A mixed-Weibull regression model for the analysis of automotive warranty data. Reliability Engineering and System Safety 2005; 87:265–273.

Baxter LA. Estimation from quasi life tables. Biometrika 1994; 81(3):567–577.

Crowder M, Stephens D. On the analysis of quasi-life tables. Lifetime Data Analysis 2003; 9(4):345–355.

Duchesne T, Lawless J. Alternative time scales and failure time models. Lifetime Data Analysis 2000; 6:157–179.

Efron B. Bootstrap Methods: Another Look at the Jackknife. The Annals of Statistics 1979; 7(1): 1–26.

Gertsbakh IB, Kordonsky KB. Parallel time scales and two-dimensional manufacturer and individual customer warranties. IIE Transactions 1998; 30:1181–1189.

Hong Y, Meeker WQ. Field-Failure and Warranty Prediction Based on Auxiliary Use-Rate Information. Technometrics 2010; 52(2):148–159.

Hogg RV, McKean JW, Craig AT. Introduction to Mathematical Statistics. 7$^{th}$ Edition. New York: Pearson; 2012.

Hu XJ, Lawless JF. Estimation of rate and mean functions from truncated recurrent event data. Journal of the American Statistical Association 1996; 91(433):300–310.

Hu XJ, Lawless JF, Suzuki K. Nonparametric Estimation of a Lifetime Distribution When Censoring Times Are Missing. Technometrics 1998; 40(1):3–13

Ion RA, Petkova VT, Peeters BHJ, Sander PC. Field reliability prediction in consumer electronics using warranty data. Quality and Reliability Engineering International 2007; 23:401–414.

Iskandar BP, Blischke WR. Reliability and warranty analysis of a motorcycle based on claims data. In: Blischke WR, Murthy DNP, editors. Case Studies in Reliability and Maintenance, New Jersey: John Wiley & Sons, Inc.; 2003, p. 623–656.

Jones-Farmer L.A, Ezell JD, Hazen BT. Applying Control Chart Methods to Enhance Data Quality. Technometrics 2014; 56(1): 29–41.

Jung M, Bai DS. Analysis of field data under two-dimensional warranty. Reliability Engineering and System Safety 2007; 92: 135–143.

Kalbfleisch JD, Lawless JF. Estimation of Reliability in Field Performance Studies. Technometrics 1988; 30(4): 365–378.

Kalbfleisch J D, Lawless JF, Robinson JA. Methods for the analysis and prediction of warranty claims. Technometrics 1991; 33:273–285.

Kalbfleisch J, Lawless J. Truncated data arising in warranty and field performance studies, and some useful statistical methods. Research Report, RR-91–02; 1991.

Kalbfleisch JD, Lawless JF. Statistical analysis of warranty claims data. In: Blischke WR, Murthy DNP, editors. Product Warranty Handbook, New York: Marcel Dekker; 1996, p. 231–259.

Karim MR, Yamamoto W, Suzuki K. Statistical analysis of marginal count failure data. Lifetime Data Analysis 2001; 7 (2):173–186.

Karim MR, Suzuki K. Analysis of field failure warranty data with sales lag. Pakistan Journal of Statistics 2004; 20(1):93–102.

Karim MR. Modeling sales lag and reliability of an automobile component from warranty database. International Journal of Reliability and Safety 2008; 2(3):234–247.

Karim MR, Suzuki K. Analysis of warranty data with covariates. Proc. IMechE Part O: J. Risk and Reliability 2007; 221: 249–255.

Kapadia AS, Chan W, Moye L. Mathematical Statistics with Applications. Florida: Taylor & Francis Group; 2005, p. 195.

Kleyner A, Sandborn P. Forecasting the cost of unreliability for products with two-dimensional warranties. Proceedings of the European Safety and Reliability Conference 2006; 1903–1908.

Lawless JF. Statistical Analysis of Product Warranty Data. International Statistical Review 1988; 66(1): 41–64.

Lawless JF, Kalbfleisch JD. Some issues in the collection and analysis of field reliability data. In: Klein JP, Goel PK, editors. Survival Analysis: State of the Art, Dordrecht: Kluwer Academic; 1992, p.141–152.

Lawless JF, Hu J, Cao J. Methods for the Estimation of Failure Distributions and Rates from Automobile Warranty Data. Lifetime Data Analysis 1995; 1(3):227–240.

Lawless JF. Statistical Models and Methods for Lifetime Data. 2nd Edition. New Jersey: John Wiley & Sons, Inc.; 2003, p. 54–55.

Lawless JF, Crowder MJ, Lee KA. Analysis of reliability and warranty claims in products with age and usage scales. Technometrics 2009, 51(1): 14–24.

Lim TJ. Nonparametric estimation of the product reliability from grouped warranty data with unknown start-up time. International Journal of Industrial Engineering: Theory Applications and Practice 2003; 10(4): 474–81.

Lu M. Automotive Reliability Prediction Based on Early Field Failure Warranty Data. Quality Reliability Engineering International 1998; 14:103–108.

Manna DK, Pal S, Sinha S. A use-rate based failure model for two-dimensional warranty. Computers & Industrial Engineering 2007; 52: 229–240.

Majeske KD. A mixture model for automobile warranty data. Reliability Engineering and System Safety 2003; 81:71–77.

Meeker WQ, Escobar LA. Statistical Methods for Reliability Data. New York: John Wiley & Sons, Inc.; 1998, p. 205.

Meeker WQ, Hong Y. Reliability Meets Big Data: Opportunities and Challenges. Quality Engineering 2013; 26:102–116.

Mohan K, Cline B, Akers J. A practical method for failure analysis using incomplete warranty data. In: Proceedings of Annual Reliability and Maintainability Symposium 2008; 193–199.

Nelder JA, Mead R. A simplex algorithm for function minimization. Computer Journal 1965; 7: 308–313.

Oh YS, Bai DS. Field Data Analysis with Additional After- warranty Failure Data. Reliability Engineering and System Safety 2001; 72:1–8.

O’Connor PDT. Practical Reliability Engineering. 4th Edition. New York: John Wiley & Sons, Ltd; 2002.

Park C. Parameter Estimation of Incomplete Data in Competing Risks Using the EM Algorithm. IEEE Transaction on Reliability 2005; 52(2):282–290.

Rai B, Singh N. Hazard Rate Estimation from Incomplete and Unclean Warranty Data. Reliability Engineering and System Safety 2003; 81: 79–92.

Rai B, Singh N. Customer-Rush Near Warranty Expiration Limit, and Nonparametric Hazard Rate Estimation From Known Mileage Accumulation Rates. IEEE Transactions on Reliability 2006; 55(3): 480–489.

Robinson J, McDonald G. Issues related to field reliability and warranty data. In: Liepins G, Uppuluri V, editors. Data Quality Control: Theory and Pragmatics, New York: Marcel Dekker; 1991, p. 69–90.

Suzuki K. Nonparametric Estimation of Lifetime Distributions from a Record of Failures and Follow-ups. Journal of the American Statistical Association 1985a; 80(389): 68–72

Singpurwalla ND, Wilson S. Warranty problem: its statistical and game theoretic aspects. SIAM Review 1993; 35(1):17–42.

Suzuki K. Estimation of Lifetime Parameters from Incomplete Field Data. Technometrics 1985b; 27 (3): 263–271.

Suzuki K. Analysis of field failure data from a nonhomogeneous Poisson process. Reports on Statistical Applications in Research 1987; 34: 6–15.

Suzuki K, Yamamoto W, Karim MR, Wang L. Data analysis based on warranty database. In: Limnios N, Nikulin M, editors. Recent Advances in Reliability Theory, Boston: Birkhauser; 2000, p. 213–227.

Suzuki K, Karim MR, Wang L. Statistical analysis of reliability warranty data. In: Rao CR, Balakrishnana N, editors. Handbook of Statistics: Advances in Reliability, Amsterdam: Elsevier; 2001, p. 585–609.

Suzuki K, Alam MM, Yoshikawa T, Yamamoto W. Two methods for estimating product lifetimes from only warranty claims data. Proceedings of the 2nd IEEE International Conference on Secure System Integration and Reliability Improvement 2008; 111–119.

Vintr Z, Vintr M. Estimate of warranty costs based on research of the customer's behavior. Proceedings of Annual Reliability and Maintainability Symposium (RAMS) 2007; 323–328.

Wang L, Suzuki K, Yamamoto W. Age-based warranty data analysis without date-specific sales information. Applied Stochastic Models in Business and Industry 2002; 18(3):323–37.

Wilson S, Joyce T, Lisay E. Reliability estimation from field return data. Lifetime Data Analysis 2009; 15(3):397–410.

Wu S. Warranty Data Analysis: A Review. Quality and Reliability Engineering International 2012; 28: 795–805.

Wu S. A review on coarse warranty data and analysis. Reliability Engineering and System Safety 2013; 114: 1–11.

Wu H, Meeker WQ. Early Detection of Reliability Problems Using Information from Warranty Databases. Technometrics 2002; 44(2): 120–133.

Yang G, Zaghati Z. Two-Dimensional Reliability Modeling From Warranty Data. Proc. Annual Reliability and Maintainability Symposium 2002; 272–278.

## APPENDIX

**A.** Product Distribution

Suppose, $U$ and $A$ are two continuous random variable and they are independent. The product of this two random variables is denoted by $T_c$ that will be another random variable. Using the property of independent (Kapadia et al., 2005), Expected value of $T_c$ will be,

$$E(T_c = UA) = E(U) \times E(A) = \mu_U \times \mu_A$$

Similarly, using the property of independent, variance of $T_c$ will be,

$$Var(T_c = UA) = E[(UA)^2] - [E(UA)]^2$$

$$= E[U^2A^2] - [E(U)E(A)]^2$$

$$= E(U^2)E(A^2) - [E(U)]^2[E(A)]^2$$

$$Here, E(U^2) = Var(U) + (\mu_U)^2$$

$$E(A^2) = Var(A) + (\mu_A)^2$$

So, $Var(T_c = UA) = [Var(U) + (\mu_U)^2] \times [Var(A) + (\mu_A)^2] - [(\mu_U)^2] \times [(\mu_A)^2]$

$$Var(T_c) = Var(U)Var(A) + Var(U)(\mu_A)^2 + Var(A)(\mu_U)^2$$

**B.** R code for only failure data: weibull analysis

```
rm(list=ls())
data1<-read.csv(file.choose(), header=T)                # read data from file
x<-c(data1$Hours)
r<-length(x)
LL<-function(theta) {                                   # likelihood function
  b<-theta[1]
  v<-theta[2]
```

```
  loglik<-r*log(b)-r*log(v)+(b-1)*sum(log(x/v))- sum((x/v)^b)
 -loglik
}
fit<-optim(c(runif(1),runif(1)),LL, hessian=T)          # maximizing likelihood function
fisher_info <- solve(fit$hessian)                        # confidence interval estimation
prop_sigma <- sqrt(diag(fisher_info))
prop_sigma<-diag(prop_sigma)
upper<-fit$par+1.96*prop_sigma
lower<-fit$par-1.96*prop_sigma
interval<-data.frame(value=fit$par, upper=upper, lower=lower)
interval
```

**C.** R code for combined failure and censored data: weibull analysis

```
rm(list=ls())
data1<-read.csv(file.choose(), header=T)                 # read data from file
data2<-read.csv(file.choose(), header=T)
x<-c(data1$Hours)                                        # data into vector form
w<-c(data2$Age)
y<-rlnorm(length(w), mu_tc, sigma_tc)                    # generated censored data
r<-length(x)
n<-length(x)+length(w)
LL<-function(theta) {                                    # likelihood function
 b<-theta[1]
 v<-theta[2]
```

```
  loglik<-r*log(b)-r*log(v)+(b-1)*sum(log(x/v))- sum((x/v)^b)- sum((y/v)^b)
 -loglik
 }
fit<-optim(c(runif(1),runif(1)),LL, hessian=T)              # maximizing likelihood function
fisher_info <- solve(fit$hessian)                           # confidence interval estimation
prop_sigma <- sqrt(diag(fisher_info))
prop_sigma<-diag(prop_sigma)
upper<-fit$par+1.96*prop_sigma
lower<-fit$par-1.96*prop_sigma
interval<-data.frame(value=fit$par, upper=upper, lower=lower)
interval
```

**D.** R code for combined failure and censored data: weibull-lognormal analysis

```
rm(list=ls())
data1<-read.csv(file.choose(), header=T)                    # read data from file
data2<-read.csv(file.choose(), header=T)
x<-c(data1$Hours)                                           # data into vector form
z<-c(data2$Age)
y<-rlnorm(length(z), mu_tc, sigma_tc)                       # generated censored data
r<-length(x)
n<-length(x)+length(y)
LL<-function(theta) {                                       # likelihood function
 b<-theta[1]
 v<-theta[2]
```

```
  m<-theta[3]
  s<-theta[4]
   loglik<-(r*log(b)-r*log(v)+(b-1)*sum(log(x/v))- sum((x/v)^b)+ sum(log(1-pnorm((log(x)-
m)/s)))-sum((y/v)^b)-(n-r)*log(sqrt(2*pi))-(n-r)*log(s)
              -sum(log(y))-sum(0.5*((log(y)-m)/s)^2))
     -loglik
 }
fit<-optim(c(runif(1),runif(1),4,1),LL, hessian=T)           # maximizing likelihood function
fisher_info <- solve(fit$hessian)                             # confidence interval estimation
prop_sigma <- sqrt(diag(fisher_info))
prop_sigma<-diag(prop_sigma)
upper<-fit$par+1.96*prop_sigma
lower<-fit$par-1.96*prop_sigma
interval<-data.frame(value=fit$par, upper=upper, lower=lower)
interval
```