الجمهورية الجزائرية الديمقراطية الشعبية

Ministry of Higher Education and Scientific Research

UNIVERSITY SAAD DAHLEB DE BLIDA

Faculty of Sciences

Mathematics Department

MASTER's THESIS

In mathematics

Option : Stochastic and Statistical Models

THEME :

Statistical inference of Step-Stress Partially Accelerated Life Tests Using the Chen Distribution With Progressive Censoring Data Type-II

Realised by

BENRELKIA Asma

Argued before the Jury:

TAMI Omar	University Blida 1	Chairmann
FRIHI Redhouane	University Blida 1	Examinator
RASSOUL Abdelaziz	ENSH of Blida	Advisor

July 2023

DEDICATION

To my dear father and mother, whom I can never thank enough,
for everything
To my beloved siblings : Abderrahmane, Mohamed Ali, Afef and Amira
To my close friend Yousra,the best I have ever had
And to you dear reader.

ACKNOWLEDGMENTS

Gratitude,
First of all elhamdoulilah, gratitude to Allah who grant me to stand here with my dissertation.

I would like to express my deepest gratitude to the following individuals and administration who have played a significant role in the completion of this dissertation:

First and foremost, I am profoundly grateful to my supervisor,Professor RASSOUL Abdelaziz, whose unwavering guidance, invaluable insights, patience and constant support have been instrumental in shaping this research. Your expertise and dedication have truly inspired me throughout this process.

I would like to extend my sincere appreciation to the members of my dissertation committee,the director Dr. TAMI Omar and The examiner Dr.FRIHI Redouane , for their constructive feedback, thought-provoking discussions, and suggestions that greatly enhanced the quality of this work. Your collective expertise and commitment to academic excellence have been invaluable. I am grateful to the staff and faculty of Science for providing a conducive academic environment and access to resources necessary for conducting this research. Their support, administrative assistance, and library services have been immensely helpful.My deepest thanks go to my dear parents, my sisters and brother, my family and my precious friends who have encouraged and supported me throughout my studies. And finally, thanks to all the people who have contributed in any way to the elaboration of this work.

Contents

LIST OF FIGURES

LIST OF TABLES

ملخص

المذكرة تركز على اختبارات الحياة المسرَعْة جزئيًا بخطوات التحميل (SS-PALT) المُطبّقة على المنتجات ذات توزيع الحياة على شكل (bathtub) بمعاملين، وتحديدًا توزيع تشين (Chen) الهدف هو تقدير معاملات التوزيع وعامل التسارع باستخدام تقدير الاحتمال الأقصى (MLE) استنادًا إلى الختم التدريجي من النوع الثاني. توفر المذكرة أيضًا التباين التأويلي ومصفوفة التباين المشترك للمقدرات. لتحديد مجالات الثقة (CIs) للمعاملات، يتم استخدام طريقتين : التقريب العادي للتوزيع التأويلي لـ (MLEs) وطريقة البوتستراب. يتم الحصول على المقدرات عدديًا باستخدام حزمة **R** من خلال إجراء تكراري. لتوضيح طريقة التقدير المقترحة، يتم تقديم مثال عددي. علاوة على ذلك، يتم توفير مثال من الحياة الواقعية لتوضيح النهج المقترح بشكل أكبر.

كلمات مفتاحية: توزيع تشين؛ خطوة اجهاد اختبارات الجودة المتسارعة جزئيا؛ رقابة تدريجية نوع ٢؛ تقدير الاحتمالية القصوى؛ مجالات الثقة

Abstract

This memory focuses on Step-Stress Partially Accelerated Life Tests (SS-PALT) applied to products with a two-parameter bathtub-shaped lifetime distribution, specifically the Chen distribution. The objective is to estimate the distribution parameters and acceleration factor using maximum likelihood estimation (MLE) based on progressive Type-II censoring. The thesis also provides the asymptotic variance and covariance matrix of the estimators. To establish confidence intervals (CIs) for the parameters, two approaches are employed: the normal approximation to the asymptotic distribution of the MLEs and the bootstrap method. The estimators are numerically obtained using the Mathematica Package through an iterative procedure. To demonstrate the proposed estimation method, a numerical example is presented. Furthermore, a real-world example is provided to illustrate the suggested approach.

Keywords: Chen distribution; Step-stress partially accelerated life tests; Progressive Type-II censoring; Maximum likelihood estimation; Asymptotic confidence intervals

Résumé:

Ce mémoire porte sur les tests de durée de vie accélérée partiellement par étapes (SS-PALT) appliqués aux produits présentant une distribution de durée de vie en forme de bathup à deux paramètres, plus précisément la distribution de Chen. L'objectif est d'estimer les paramètres de la distribution et le facteur d'accélération à l'aide de l'estimation du maximum de vraisemblance (MLE) basée sur une censure progressive de Type-II. La thèse fournit également la variance asymptotique et la matrice de covariance des estimateurs. Pour établir des intervalles de confiance (IC) pour les paramètres, deux approches sont utilisées : l'approximation normale à la distribution asymptotique des MLE et la méthode du bootstrap. Les estimateurs sont obtenus numériquement à l'aide du package Mathematica par une procédure itérative. Pour illustrer la méthode d'estimation proposée, un exemple numérique est présenté. De plus, un exemple concret est fourni pour illustrer davantage l'approche suggérée.

Mots clés: Distribution de Chen ; Essais de durée de vie partiellement accélérés ; Progressif ; Censure de type II ; Estimation de vraisemblance maximale ; Intervalles de confiance asymptotiques Type-II ; Estimation du maximum de vraisemblance ; Intervalles de confiance asymptotiques.

Abbreviations

CDF	Cumulative Distribution Function
PDE	Probability Distribution Function
MTTF	Mean Time to Failure
MTBF	Mean Time Befor Failure
MTTR	Mean Time To Repair
ALT	Accelerated life test
PALT	Parially Accelerated life test
CSALT	Constant-Stress Accelerated life test
PSALT	Progressive-Stress Accelerated Life Test
SSALT	Step-Stress Accelerated Life Test
SSPALT	Step-Stress Parially Accelerated Life Test
ML	Maximum Likelihood
MLE	Maximum Likelihood Estimaton
CIE	Confidence Intervals Estimation
CD	Chen D istribution
RF	Reliability Function

GENERAL INTRODUCTION

In order for the researchers to enhance the credibility, validity, and usefulness of their statistical analyses and research findings,Reliability is required. Reliability is the extent to which a system, process, or object can be trusted to consistently perform with accuracy and dependability. It measures the capability of consistently delivering the intended or expected outcomes over time and across different circumstances, also in the realm of technology and engineering, which means reliability primarily relates to the likelihood of a system or component operating without failure during a specified time frame and under predefined operational conditions without encountering unforeseen malfunctions or errors. overall, Reliability plays a vital role in guaranteeing the consistent and dependable performance of systems, products, or processes. Its importance lies in minimizing disruptions, reducing downtime, and mitigating any potential negative consequences for users or stakeholders. By focusing on reliability, organizations can enhance user satisfaction, maintain operational continuity, and safeguard against adverse impacts on productivity or reputation therefore it is crucial for safety, productivity, and customer satisfaction. It also minimizes failures, accidents, and downtime, ensuring consistent performance, It is vital in mission-critical applications where failure can have severe consequences.

Reliability analysis uses failure data to estimate parameters, predict performance, and evaluate maintenance strategies. It enables engineers to optimize reliability, mitigate risks, and enhance overall system performance. The uses of reliability analysis are diverse and span across different stages of product development and operation. During the design phase. It aids in selecting appropriate materials and components with desired reliability characteristics, and one commonly used method in life testing of the products under a process of reliability test is the Accelerated Life Testing (ALT) method. ALT involves subjecting the product to higher stress levels than it would typically encounter in normal operating conditions. By accelerating the testing process, engineers can obtain failure data more quickly and estimate the product's reliability characteristics within a shorter time frame with less money it consists of a variety of test methods for shortening the life of test items or hastening the degradation of their performance. The aim of such testing is to obtain data quickly, which, when properly modeled and analyzed, yields the desired information on product life or performance under normal use conditions, This allows for faster decision-making regarding design improvements, materials selection, and product release, also to reduce the costs to their minimum.ALT (Accelerated Life Testing)is a general term that refers to any testing approach that accelerates the stress levels experienced by the test units to shorten the time required to observe failures. It involves subjecting the units to higher temperatures, increased voltage, intense vibration, or other stress factors to simulate real-world operating conditions, The accelerated conditions in ALT help to shorten the testing duration, enabling engineers to obtain reliable information within a reasonable time frame. By intentionally accelerating the failure mechanisms, engineers can observe failures more frequently, allowing them to collect sufficient data for statistical analysis and reliability estimation, In ALT the test items are tested only at accelerated conditions. and there are mainly three ALT methods. The first method is called the constant-stress ALT, the stress is kept at a constant level throughout the life of test products. The second one is referred to as progressive-stress ALT, the stress applied to a test product is continuously increasing over time. The third is the step-stress ALT, in which the test condition changes at a given time or upon the occurrence of a specified number of failures, the fourth step-stress partially ALT, the stress levels are intentionally kept lower than what would be required in traditional ALT, making it a partially accelerated testing approach this type has been studied by several authors. These ALT methods, including CSALT, PSALT, SSALT, and SSPALT, are valuable tools in assessing product reliability, identifying potential failure modes, and making informed decisions about design improvements, material selection, and product enhancements.

Progressive censoring is a type of censoring scheme used in survival analysis and reliability studies. In progressive censoring, the observation or failure times of the test units are not fully observed or known. Instead, the failure times are only known to lie within specific intervals or windows. they are employed in situations where complete failure time information is difficult or expensive to obtain, or when the test units cannot be continuously monitored until failure. This type of censoring allows for more efficient data collection and analysis, as it reduces costs and minimizes the required sample size.

There are different progressive censoring schemes, one of which is Type I Progressive Censoring, In this scheme, the failure times of test units are censored within specific intervals. The intervals may be predetermined or determined based on the observed failure times. Type I censoring allows for efficient data collection and reduces costs compared to complete failure time observation. the second, Type II Progressive Censoring, This scheme involves progressively narrowing the intervals in which failure times are censored. Initially, the intervals are wide, and as failures occur, the intervals are narrowed to obtain more precise failure time information. Type II censoring provides more accurate estimates of reliability parameters compared to Type I censoring. The advantages of progressive censoring schemes include efficient data collection, minimized sample size, flexibility in study design, and improved reliability estimation.

Progressive censoring reduces the cost and effort of obtaining failure time information by censoring data within intervals, making data collection more practical. It also reduces the required sample size, leading to more efficient data analysis and cost savings. Researchers can tailor censoring intervals based on resources and objectives, allowing for flexible study design. Progressive censoring provides valuable information about failure times, enabling more accurate estimation of reliability measures in varying conditions.

our work is presented as follows: in Chapter 1, we tackle Reliability and it's measures.also commonly used distributions and the main estimation method. After that, in Chapter 2, we present the Accelerated Life Test, its types and different categories of censored data ,then Progressive Censoring types, also an overview of the Chen distribution And statistic inference for step stress partially ALT using Chen distribution are presented. Finally, we depict the solution to how to estimate the parameters of this distribution with Maximum Likelihood Estimation. In the last Chapter, We give the simulation study, we have list of R packages popularly known for fitting distributions. There might be more packages. But, we decided to focus on these ones. we have tried to explain the concepts with application in 2 real data set. We complete this work with a general conclusion.

CHAPTER 1

RELIABILITY SYSTEMS

1.1 Introduction

Reliability refers to the ability of a system or component to perform its intended function without failure over a specified period of time under given operating conditions. It is a measure of the system's ability to consistently operate and meet performance requirements and is a crucial aspect in various fields, including engineering, manufacturing, and quality control. It plays a significant role in ensuring the safety, efficiency, and durability of products and systems. Reliability analysis involves studying the failure behavior of components or systems, identifying potential failure modes, and estimating the probability of failure within a specified time frame. To describe the Time to Failure as a statistical distribution, Reliability Distribution Analysis is required which is usually characterized by a specific pattern where different distribution types are supported to finally estimate the entire process by using the maximum likelihood as well as confidence interval as estimation methods, the same distributions are going to be tested under some fit techniques .

1.2 Reliability

Reliability is a crucial aspect of product or system performance that focuses on their ability to function without failure or degradation over a specified period under given operating conditions. It plays a significant role in various industries, including manufacturing, engineering, aerospace, automotive, electronics, and many others. Reliability analysis provides insights into the probability of failure, expected lifetime, and performance under different conditions, allowing businesses to optimize design, maintenance, and decision-making processes.

Mathematically, reliability $R(t)$ is the probability that a system will function properly between time 0 and time t, (also called the survival function),

$$R(t) = P(T > t), \quad t \geq 0 \tag{1.1}$$

where T is a random variable that represents the failure time.
Unreliability $F(t)$, a measure of failure, is defined as the probability that the system will fail by time t,

$$F(t) = P(T \leq t) \quad , t \geq 0$$

In other words, $F(t)$ is the failure distribution function. If the failure time random variable T has a density function $f(t)$, then ,

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

or, equivalently,

$$f(t) = -\frac{d}{dt}[R(t)].$$

Mathematically speaking, the density function can be expressed in terms of T:

$$\lim_{\Delta t \to 0} P(t < T \leq t + \Delta t)$$

This can be seen as the likelihood that the failure time T will happen between the operating time t and the subsequent operating period, $t + \Delta t$.

Consider a new and successfully tested system that operates well when put into service at time $t = 0$. The longer the time period, the less probable it is that the system will continue to work. Naturally, there is no chance of success for an infinitely long period of time.
As a result, the system operates with a chance of one and gradually drops to zero.
Clearly, If the time to failure is described by an exponential failure time density function, then

$$f(t) = \frac{1}{\theta} \exp\left(-\frac{t}{\theta}\right), \quad t \geq 0, \theta > 0, \tag{1.2}$$

and this will lead to the reliability function is,

$$R(t) = \int_t^{\infty} \frac{1}{\theta} \exp\left(-\frac{s}{\theta}\right) ds = \exp\left(-\frac{t}{\theta}\right), \quad t \geq 0 \tag{1.3}$$

Consider the Weibull distribution where the failure time density function is given by,

$$f(t) = \frac{\beta t^{\beta-1}}{\theta^{\beta}} \exp\left(-\left(\frac{t}{\theta}\right)^{\beta}\right) \quad t \geq 0, \theta > 0, \beta > 0$$

Then the reliability function is,

$$R(t) = \exp\left(-\left(\frac{t}{\theta}\right)^{\beta}\right), \quad t \geq 0$$

Hence, the reliability function can be directly determined given a specific failure time density function or failure time distribution function. For certain distributions, (see1.3) for additional explanation.

1.2.1 Reliability Measures

reliability measures are utilized to assess the performance and durability of products or systems. These measures provide insights into the reliability characteristics and help make informed decisions regarding product design, maintenance, and quality control,Some commonly used measures of reliability include:

1.2.1.1 System Mean Time to Failure(MTTF)

Suppose that the reliability function for a system is given by $R(t)$, MTTF It is a reliability metric that represents the average time elapsed between the initial operation of a system, component, or device and its first occurrence of failure. **Mathematically**, is given by

$$MTTF = \int_0^{\infty} t f(t) dt \tag{1.4}$$

Substituting ,

$$f(t) = -\frac{d}{dt}[R(t)]$$

into equation 1.4 and performing integration by parts, we obtain

$$\begin{aligned} \text{MTTF} &= -\int_0^{\infty} t d[R(t)] \\ &= [-tR(t)]|_0^{\infty} + \int_0^{\infty} R(t) dt \end{aligned} \tag{1.5}$$

The first term on the right-hand side of equation 1.5 equals zero at both limits, since the system must fail after a finite amount of operating time. Therefore, we must have $tR(t) \to 0$ as $t \to f$. This leaves the second term, which equals,

$$\text{MTTF} = \int_0^{\infty} R(t) dt \tag{1.6}$$

Thus, MTTF is the definite integral evaluation of the reliability function,and is typically used for non-repairable systems or components, meaning that once they fail, they cannot be restored to their original state.

It is important to note that MTTF does not account for repair or restoration time, as it assumes that failures are not repaired and the system or component is replaced after failure.

1.2.1.2 **Mean Time Between Failures** (MTBF)

MTBF represents the average time elapsed between two consecutive failures of a system or components. It is a widely used metric for assessing the reliability of systems with a repairable nature. A higher MTBF indicates a longer average duration between failures, indicating greater reliability.
MTBF is an expected value of the random variable time between failures. **Mathematically**,

$$\mathrm{MTBF} = \mathrm{MTTF} + \mathrm{MTTR}$$

A system where faults are rapidly diagnosed is more desirable than a system that has a lower failure rate but where the cause of a failure takes longer to detect, resulting in lengthy system downtime. When the system being tested is renewed through maintenance and repairs, E(T) is also known as MTBF.

1.2.1.3 **Failure Rate Function**

The probability of a system failure in a given time interval $[t_1, t_2]$ can be expressed in terms of the reliability function as,

$$\begin{aligned} \int_{t_1}^{t_2} f(t)dt &= \int_{t_1}^{\infty} f(t)dt - \int_{t_2}^{\infty} f(t)dt \\ &= R(t_1) - R(t_2) \end{aligned}$$

or in terms of the failure distribution function (or the unreliability function) as,

$$\begin{aligned} \int_{t_1}^{t_2} f(t)dt &= \int_{-\infty}^{t_2} f(t)dt - \int_{-\infty}^{t_1} f(t)dt \\ &= F(t_2) - F(t_1) \end{aligned}$$

The failure rate is the frequency with which failures take place during a given period of time $[t_1, t_2]$. It is described as the likelihood that a failure will occur once every unit of time in the interval, assuming no failures have occurred before time zero (t_1). the failure rate is a result,

$$\frac{R(t_1)-R(t_2)}{(t_2-t_1)R(t_1)}$$

Note that the failure rate is a function of time. If we redefine the interval as $[t, t+\Delta t]$,the above expression become to,

$$\frac{R(t)-R(t+\Delta t)}{\Delta t R(t)}$$

The rate in the above definitions is expressed as failures per unit of time when in reality, the time units might be in terms of miles, hours, etc. A lower failure rate indicates higher reliability.

The hazard function is defined as the limit of the failure rate as the interval approaches zero. Thus, the hazard function $h(t)$ is the instantaneous failure rate and is defined by

$$\begin{aligned} h(t) &= \lim_{\Delta t \to 0} \frac{R(t)-R(t+\Delta t)}{\Delta t R(t)} \\ &= \frac{1}{R(t)}\left[-\frac{d}{dt}R(t)\right] \\ &= \frac{f(t)}{R(t)} \end{aligned}$$

A device's chance of failing in the brief period from time t to $(t+dt)$is represented by the amount $h(t)dt$.

The hazard function is significant because it shows how the failure rate changes over the lifespan of a population of components by showing each one's hazard function along a single axis. For instance, two designs might be equally reliable at a certain point in time, but their failure rates prior to this point might be different.

In statistical theory, the force of mortality is equivalent to the hazard function, and the death rate is equivalent to the failure rate. As a result, the ratio of the probability density function (pdf) to the reliability function is known as the hazard function, hazard rate, or failure rate function.

1.2.1.4 Maintainability

When a system fails to perform satisfactorily, the repair is normally carried out to locate and correct the fault. The system is restored to operational effectiveness by making an adjustment or by replacing a component.

When maintenance is carried out in accordance with recommended processes and resources, maintainability is defined as the likelihood that a failed system will be returned to specific conditions within a certain amount of time. In other terms, maintainability is the likelihood of finding and fixing a problem with a system in a specific amount of time. To guarantee that the system product can be efficiently and economically maintained by the client, maintainability engineers must collaborate with system designers. In order to calculate the amount of time needed to complete the operation, the level of expertise required, the kind of support equipment needed, and the documentation needed, this function necessitates an analysis of the part removal, replacement, tear-down, and build-up of the product. Let T denote the random variable of the time to repair or the total downtime. If the repair time$T$ has a repair time density function $g(t)$, then the maintainability, $V(t)$, is defined as the probability that the failed system will be back in service by time t, i.e.,

$$V(t) = P(T \leq t) = \int_0^t g(s)ds$$

For example, if $g(t) = \mu e^{-\mu t}$ where $\mu > 0$ is a constant repair rate, then

$$V(t) = 1 - e^{-\mu t},$$

which represents the exponential form of the maintainability function.
An important measure often used in maintenance studies is the mean time to repair (MTTR) or the mean downtime. MTTR is the expected value of the random variable repair time, not failure time, and is given by

$$MTTR = \int_0^\infty tg(t)dt$$

When the distribution has a repair time density given by $g(t) = \mu e^{-\mu t}$, then, from the above equation,MTTR = $\frac{1}{p}$. When the repair time T has the log-normal density function $g(t)$, and the density function is given by,

$$g(t) = \frac{1}{\sqrt{2\pi}\sigma t} e^{-\frac{(\ln t - \mu)^2}{2\sigma^2}}, \quad t > 0$$

then it can be shown that,

$$MTTR = me^{\frac{\sigma^2}{2}}$$

1.2.1.5 Availability

Availability is a measure of the system's ability to undergo repairs after a failure,It is representing the probability of successful operation at a specific time t,**Mathematically**,

$$\begin{aligned}\text{Availability} &= \frac{\text{System up time}}{\text{System up time} + \text{System down time}}\\ &= \frac{\text{MTTF}}{\text{MTTF}+\text{MTTR}}\end{aligned}$$

Availability is a measure of success used primarily for repairable systems:
▶ For non-repairable systems, Availability $A(t)$ equals reliability $R(t)$.
▶ In repairable systems, $A(t)$ will be equal to or greater than $R(t)$.
Higher availability signifies better reliability and the ability of a system to fulfill its intended function when required.

1.3 Common Distribution Functions

A statistical absolutely continuous distribution can be fully characterized by its probability density function. It plays a crucial role in understanding the behavior of life-testing data in reliability engineering and various other fields. Additionally, it aids in deriving other essential functions used in reliability analysis, such as the reliability function or hazard rate function, mean time to failure function, mean residual life function, and median life function. In this discussion, we will explore some important lifetime distributions along with their significant characteristics [1].

1.3.1 Exponential Distribution

The Exponential distribution is a commonly used statistical distribution, especially in the fields of reliability and survival analysis. It has gained popularity due to its simplicity, even in cases where its suitability may not be fully justified. Several researchers, including Davis (1952)[2], and Epstein (1958)[3] have provided arguments in favor of using the Exponential distribution.
The Exponential distribution is closely associated with the Poisson process, which is a process where events occur continuously and independently at a constant failure rate. The Exponential distribution is the probability distribution of the time between successive events in a Poisson process. In addition to its relevance in the context of the Poisson process, the Exponential distribution appears in various other scenarios, such as waiting time problems. For example, Maguire et al. (1952)[4] conducted a study on mine accidents and demonstrated that the time intervals between accidents followed an Exponential distribution.

A random variable T is said to follow an exponential distribution with scale parameter θ if it has the following CDF,

$$F(t) = 1 - e^{-\frac{t}{\theta}} = \lambda e^{-\lambda t}, \quad t \geq 0 \tag{1.7}$$

where $\theta > 0$ is the mean time to failure and the p.d.f. is,

$$f(t) = \frac{1}{\theta} e^{-\frac{t}{\theta}} = \lambda e^{-\lambda t}, \quad t \geq 0 \tag{1.8}$$

The Reliability function is,

$$R(t) = e^{-\frac{t}{\theta}} = e^{-\lambda t}, \quad t \geq 0$$

where $\theta = \frac{1}{\lambda} > 0$ is an MTTF's parameter and $\lambda \geq 0$ is a constant failure rate. The hazard function or failure rate for the exponential density function is constant, i.e.,

$$h(t) = \frac{f(t)}{R(t)} = \frac{\frac{1}{\theta} e^{-\frac{1}{\theta}}}{e^{-\frac{1}{\theta}}} = \frac{1}{\theta} = \lambda$$

One of the most prominent features of the exponential distribution is the fact that it is the only continuous distribution with a constant failure rate. It is a particular case of gamma distribution. It is the continuous counterpart of the geometric distribution, and it has the key property of being memoryless, i.e. an old unit and a new unit have the same probability of failure at a future time interval.

Example: A manufacturer performs an operational life test on ceramic capacitors and finds they exhibit a constant failure rate with a value of $3x10^-8$ failure per hour. What is the reliability of a capacitor at 104 hours?
Solution: The reliability of a capacitor at 10^4 hour is

$$R(t) = e^{-\lambda t} = e^{3x10^{-8}t} = e^{3x10^{-4}t} = 0.9997$$

The resulting reliability plot is shown in Figure 1.1.

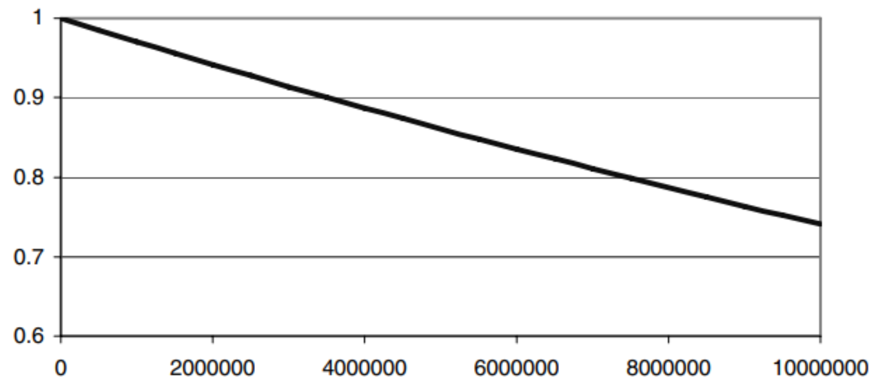

Figure 1.1: Reliability function vs time

1.3.2 Weibull Distribution

The exponential distribution is often limited in applicability owing to the memoryless property. A generalization of the exponential distribution, the Weibull distribution (Weibull 1951)[5] is frequently used to characterize fatigue life, ball bearing life, and vacuum tube life. The Weibull distribution is ideal for modeling component lifetimes with variable hazard rate functions and for representing a variety of engineering applications since it is very adaptable. The probability density function with three parameters is,

$$f(t) = \frac{\beta(t-\gamma)^{\beta-1}}{\theta^\beta}\exp\left(-\left(\frac{t-\gamma}{\theta}\right)^\beta\right), \quad t \geq \gamma \geq 0$$

where θ and β are known as the scale and shape parameters, respectively, and γ is known as the location parameter. These parameters are always positive. By using different parameters, this distribution can follow the exponential distribution, the normal distribution, etc. It is clear that, for $t > \gamma$, the reliability function $R(t)$ is,

$$R(t) = \exp\left(-\left(\frac{t-\gamma}{\theta}\right)^\beta\right), \qquad \text{for } t > \gamma > 0, \beta > 0, \theta > 0 \tag{1.9}$$

hence,

$$h(t) = \frac{\beta(t-\gamma)^{\beta-1}}{\theta^\beta}, \quad t > \gamma > 0, \beta > 0, \theta > 0 \tag{1.10}$$

It can be shown that the hazard function is decreasing for $\beta < 1$, increasing for $\beta > 1$, and constant when $\beta = 1$.

Note that the Rayleigh and exponential distributions are special cases of the Weibull distribution at $\beta = 2$, $\gamma = 0$, and $\beta = 1,\gamma = 0$, respectively. For example, when $\beta = 1 \text{and} \gamma = 0$ the reliability of the Weibull distribution function in equation 1.9 reduces to

$$R(t) = e^{-\frac{t}{\theta}}$$

and the hazard function given in equation 1.10 reduces to $1/\theta$, a constant. Thus, the exponential is a special case of the Weibull distribution. Similarly, when $\sigma = 0$ and $\beta = 2$, the Weibull probability density function becomes the Rayleigh density function. That is

$$f(t) = \frac{2}{\theta}e^{-\frac{t}{\beta}} \quad t \geq 0, \alpha, \beta > 0$$

The resulting reliability function is shown in Figure 1.2.

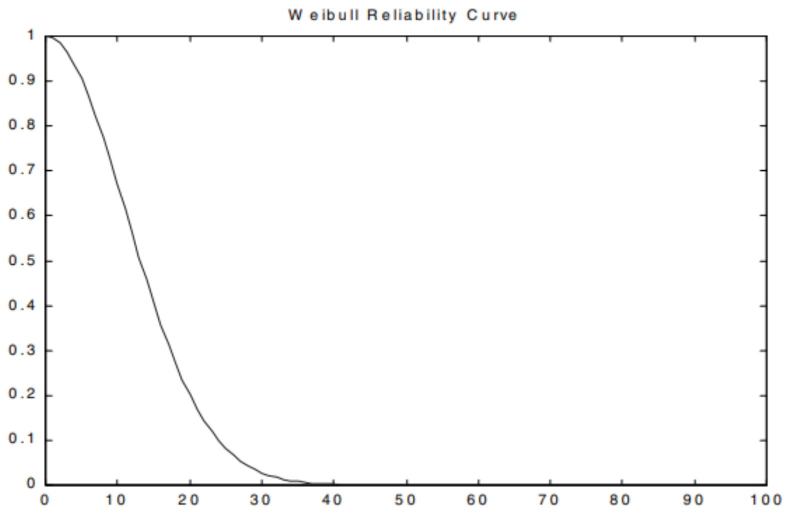

Figure 1.2: Weibull reliability function vs time

1.3.3 Binomial Distribution

One of the discrete random variable distributions that are most frequently utilized in reliability and quality control is the binomial distribution. It can be used in reliability engineering, for instance, when dealing with scenarios where an event can either be a success or a failure. The pdf of the distribution is given by,

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \quad x = 0, 1, 2, \ldots, n$$

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$

where n= number of trials; x = number of successes; p = single trial probability of success.

The reliability function, $R(k)$, (i.e., at least k out of n items are good) is given by

$$R(k) = \sum_{x=k}^{n} \binom{n}{x} p^x (1-p)^{n-x}$$

1.3.4 Poisson Distribution

Although the Poisson distribution can be used in a manner similar to the binomial distribution, it is used to deal with events in which the sample size is unknown. This is also a discrete random variable distribution whose pdf is given by,

$$P(X = x) = \frac{(\lambda t)}{e^{-\lambda t}} \text{ for } x = 0, 1, 2, \ldots,$$

where λ = constant failure rate, x = is the number of events.
In other words, $P(X = x)$ is the probability of exactly x failures occurring in time t. Therefore, the reliability Poisson distribution, $R(k)$ (the probability of k or fewer failures) is given by,

$$R(k) = \sum_{x=0}^{k} \frac{(\lambda t)^x e^{-\lambda t}}{x!}$$

This distribution can be used to determine the number of spares required for the reliability of standby redundant systems during a given mission.

1.3.5 Normal Distribution

Normal distribution plays an important role in classical statistics owing to the Central Limit Theorem. The normal distribution is largely used in reliability engineering for evaluations of product susceptibility and external stress. For many mechanical systems, this two-parameter distribution is used to represent systems in which a failure occurs as a result of some wear-out event, The pdf is given by,

$$f(t) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2\right),$$

where μ is the mean value and σ is the standard deviation. The cumulative distribution function (cdf) is,

$$F(t) = \int_{-\infty}^{t} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{s-\mu}{\sigma}\right)^2\right) ds,$$

The reliability function is,

$$R(t) = \int_t^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{s-\mu}{\sigma}\right)^2\right) ds,$$

The hazard function for a normal distribution is,

$$h(t) = \frac{f(t)}{R(t)}$$

The Normal distribution is a versatile empirical model, commonly applicable to various failure mechanisms such as corrosion, migration, crack growth, and chemical reactions. While it may not always be the ideal choice, its empirical success can be attributed to its theoretical derivation under matching assumptions for these mechanisms.

1.3.6 Log Normal Distribution

The log-normal lifetime distribution is a very flexible model that can empirically fit many types of failure data. This distribution, with its applications in maintainability engineering, is able to model failure probabilities of repairable systems and to model the uncertainty in failure rate information. The log-normal density function is given by,

$$f(t) = \frac{1}{\sigma t\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{\ln t - \mu}{\sigma}\right)^2\right), \quad t \geq 0 \tag{1.11}$$

where μ and σ parameters such that $_\infty < \mu < \infty$, and $\sigma > 0$. Note that μ and σ are not the mean and standard deviations of the distribution.
Mathematically, if a random variable X is defined as $X = \ln(T)$, then X is normally distributed with a mean of μ and a variance of σ^2, that is,

$$E(X) = E(\ln T) = \mu$$

and

$$V(X) = V(\ln T) = \sigma^2$$

Since $T = e^t$, the mean of the log-normal distribution can be found by using the normal distribution. Consider that,

$$E(T) = E\left(e^X\right) = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(x - \frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) dx$$

and by rearrangement of the exponent, this integral becomes,

$$E(T) = e^{\mu+\frac{\sigma^2}{2}} \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}\left[x-\left(\mu+\sigma^2\right)\right]^2\right) dx$$

Thus, the mean of the log-normal distribution is,

$$E(T) = e^{\mu+\frac{\sigma^2}{2}}$$

Proceeding in a similar manner,

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

thus, the variance for the log normal is,

$$V(T) = e^{2\mu+\sigma^2}\left(e^{\sigma^2}-1\right)$$

The cumulative distribution function for the log-normal is,

$$F(t) = \int_0^t \frac{1}{\sigma s\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{\ln s-\mu}{\sigma}\right)^2\right) ds$$

can be related to the standard normal deviate Z by,

$$\begin{aligned} F(t) &= P[T \leq t] = P(\ln T \leq \ln t) \\ &= P\left[Z \leq \frac{\ln t-\mu}{\sigma}\right] \end{aligned}$$

Therefore, the reliability function is given by ,

$$R(t) = P\left[Z > \frac{\ln t-\mu}{\sigma}\right] \tag{1.12}$$

and the hazard function would be,

$$h(t) = \frac{f(t)}{R(t)} = \frac{\Phi\left(\frac{\ln t-\mu}{\sigma}\right)}{\sigma t R(t)}$$

The log-normal lifetime model, like the normal, is flexible enough to make it a very useful empirical model. Figure 1.3 shows the reliability of the log-normal vs time. It can be theoretically derived under assumptions matching many failure mechanisms. Some of these are corrosion and crack growth, and in general, failures resulting from chemical reactions or processes.

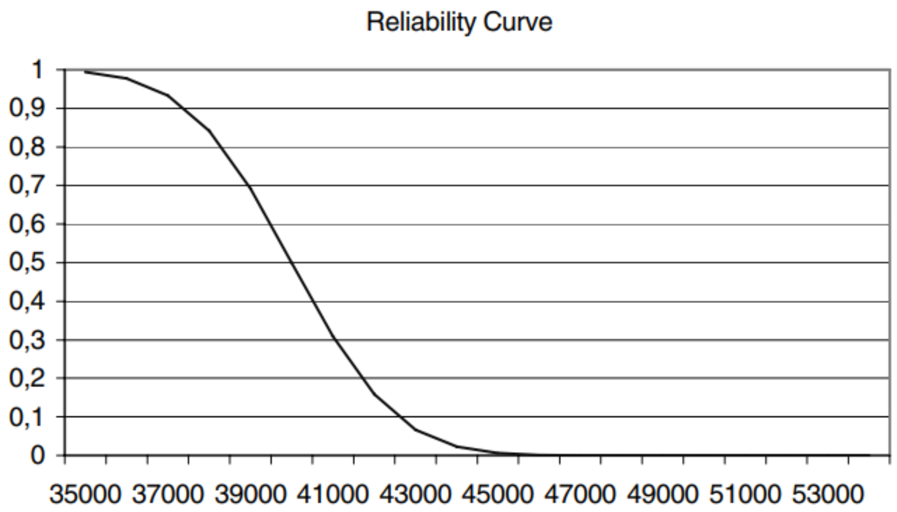

Figure 1.3: log-normal reliability function vs time

1.3.7 Gamma Distribution

A gamma distribution can be used as a failure probability function for components whose distribution is skewed. The failure density function for a gamma distribution is,

$$f(t) = \frac{t^{\alpha-1}}{\beta^{\alpha}\Gamma(\alpha)}e^{-\frac{t}{\beta}} \qquad t \geq 0, \alpha, \beta > 0 \tag{1.13}$$

where α is the shape parameter and βs the scale parameter. Hence,

$$R(t) = \int_{i}^{\infty} \frac{1}{\beta^{\alpha}\Gamma(\alpha)} s^{\alpha-1} e^{-\frac{s}{\beta}} ds$$

If α is an integer, it can be shown by successive integration by parts that

$$R(t) = e^{-\frac{t}{\beta}} \sum_{i=0}^{\alpha-1} \frac{\left(\frac{t}{\beta}\right)^{i}}{i!} \tag{1.14}$$

and

$$h(t) = \frac{f(t)}{R(t)} = \frac{\frac{1}{\beta^{\alpha}\Gamma(\alpha)} t^{\alpha-1} e^{-\frac{1}{\beta}}}{e^{-\frac{t}{\beta}} \sum_{i=0}^{\alpha-1} \frac{\left(\frac{t}{\beta}\right)^{i}}{i!}}$$

The gamma density function has shapes that are very similar to the Weibull distribution. At $\alpha = 1$, the gamma distribution becomes the exponential distribution with the constant failure rate $\frac{1}{\beta}$, It is a flexible lifetime model that may offer a good fit to some sets of failure data and it is frequently used in Bayesian reliability applications, and the resulting reliability plot is shown in Figure1.4

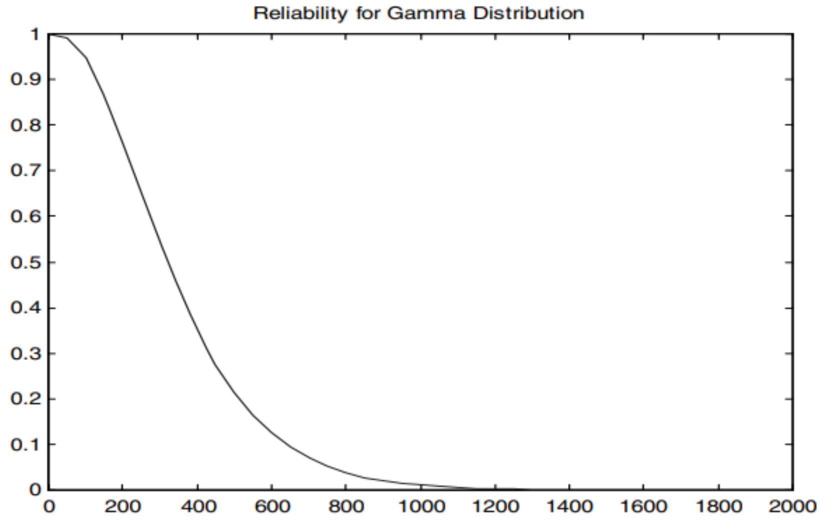

Figure 1.4: Gamma reliability function vs time

1.3.8 Beta Distribution

The two-parameter Beta density function, $f(t)$, is given by

$$f(t) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} t^{\alpha}(1-t)^{\beta}, \quad 0 < t < 1, \alpha > 0, \beta > 0$$

where α and β are the distribution parameters. This two-parameter distribution is commonly used in many reliability engineering applications.

1.3.9 Pareto Distribution

The Pareto distribution was initially created as a population income model. The population density in cities, stock market swings, and individual incomes are examples of phenomena having unusually long right tails. The Pareto distribution's probability density function is given by

$$f(t) = \frac{\alpha k^{\alpha}}{t^{\alpha+1}} \quad k \leq t \leq \infty$$

The reliability of the Pareto distribution are,

$$R(t) = \left(\frac{k}{t}\right)^{\alpha}$$

The Pareto and log normal distributions have been commonly used to model the population size and economic incomes. The Pareto is used to fit the tail of the distribution, and the log-normal is used to fit the rest of the distribution.

1.3.10 Lomax Distribution

The Lomax distribution is also known as the Pareto-II distribution, Lomax (1954) [6]first presented this to model data on business failure. Some of its applications include income analysis; reliability modeling and accelerated life testing (Hassan and Al-Ghamdi, 2009) [7]; fitting firm size data; receiver operating characteristic (Campbell and Ratnaparkhi, 1993)[8]; distribution properties and recurrence relation moments of record values (Balakrishnan and Ahsanullah, 1994)[9]. Researcher Bryson (1974) [10] stated that the Lomax distribution offers a useful alternative to conventional life distributions such as the Exponential, Weibull, or Gamma distributions in scenarios where it is assumed that the population distribution is heavily tailed.

A random variable T has Lomax distribution with two parameters λ and θ, if it has the following CDF given as

$$F(t) = 1 - \left(1 + \frac{t}{\theta}\right)^{-\lambda}, \quad t > 0, \theta > 0, \lambda > 0.$$

where λ and θ is the shape and scale parameter, respectively. The p.d.f. is

$$f(t) = \frac{\lambda}{\theta}\left(1 + \frac{t}{\theta}\right)^{-(\lambda+1)}, \quad t > 0, \theta > 0, \lambda > 0.$$

The reliability function is

$$R(t) = \left(1 + \frac{t}{\theta}\right)^{-\lambda}, \quad t > 0.$$

The hazard rate function is

$$h(t) = \frac{\lambda}{\theta}\left(1 + \frac{t}{\theta}\right)^{-1}, \quad t > 0, \theta > 0, \lambda > 0.$$

In the context of lifetime distribution, the Lomax distribution belongs to the class of decreasing failure rate distribution (Chahkandi and Ganjali, 2009)[11].

1.3.11 Rayleigh Distribution

The Rayleigh function is a flexible lifetime distribution that can apply to many degradation process failure modes.It is a special case of Weibull distribution with wide applicability. For more details, one may refer to Johnson et. al. (1994)[12]. The cumulative distribution function of Rayleigh distribution with scale parameter σ is,

$$F(t) = 1 - \exp\left(\frac{-t^2}{2\sigma^2}\right), \quad t > 0, \sigma > 0.$$

The Rayleigh probability density function is

$$f(t) = \frac{t}{\sigma^2} \exp\left(\frac{-t^2}{2\sigma^2}\right), \quad t > 0, \sigma > 0.$$

The hazard rate function and reliability of the Rayleigh function are, respectively,

$$h(t) = \frac{t^2}{\sigma^2}$$

and

$$R(t) = \exp\left(\frac{-\sigma t^2}{2}\right), \quad t > 0, \sigma > 0,$$

Some important remarks of this distribution:
• The Chi-squared distribution with two degrees of freedom is equivalent to Rayleigh distribution with a scale parameter equal to one.
• The Weibull distribution is a generalization of the Rayleigh distribution

1.4 The Estimation Methods

The estimation processes are crucial for determining the precision of the estimated model parameters and for describing the best test strategies in Accelerated life testing. Thus, in order to find precise estimates that may improve the precision of the test plans, the researchers proposed various types of estimation procedures in SSPALT, such as Maximum Likelihood (ML) method.

1.4.1 Maximum Likelihood Method

The method of maximum likelihood estimation is one of the most popular and important methods of estimation. Fisher (1912)[13] investigated several of its optimality properties and compared this with its competing methods. In order to introduce this method, let us assume $T_1, T_2, ..., T_n$ to be a random sample drawn from a population with p.d.f. (or p.m.f.) $f(t;\tilde{\theta})$, $\tilde{\theta} \in \Theta \subseteq \mathbb{R}^k$. The joint distribution of sample observations, in this case, is given by

$$
\begin{aligned}
f(t;\tilde{\theta}) &= \begin{cases} \prod_1^n f\left(t_i;\tilde{\theta}\right); & \text{if T is discrete ,} \\ \prod_1^n P_{\tilde{\theta}}\left(t_i \leq T_i \leq t_i + dt_i\right); & \text{if T is continuous} \end{cases}, \\
&= \prod_1^n f\left(t_i;\tilde{\theta}\right) dt_i
\end{aligned}
$$

where dt_i denotes the length of the interval.
The joint density in the equation above is a function of t for given $\tilde{\theta}$, i.e., the joint density is a function of n observations. Whenever this is denoted as a function of $\tilde{\theta}$ given t, we call it a likelihood function and denote it by $L(\tilde{\theta}, t)$. The likelihood function is a function of parameters $(\theta_1, ..., \theta_k)$. Generally, k is either 1 or 2.

1.4.2 Definition

A value of θ in Θ, say θ_0, is said to be a maximum likelihood estimator (MLE) of θ, given a set of sample observations t, if

$$L(\theta_0, \mathbf{t}) = \max_{\tilde{\theta}\in\Theta} L(\tilde{\theta}; \mathbf{t}).$$

Here the range of MLE $\theta_0(t)$ is the same as it is for the parameters. $\theta_0(t)$ must be in θ; if not, then it is not the MLE of θ. Since the log is a monotonic function. The function logs$L(\tilde{\theta}; t)$is called log-likelihood. Likelihood-based statistical inference uses the information as to how the likelihood or log-likelihood changes by varying θ. A function that represents this change,

$$S(\mathbf{t}, \tilde{\theta}) = \frac{\partial}{\partial\tilde{\theta}} \log L(\tilde{\theta}; \mathbf{t})$$

Is known as a score function. If log $L(\tilde{\theta}; t)$ (i) is a function of $\tilde{\theta}$ which is twice differentiable and (ii) the extrema of $L(\tilde{\theta}; t)$ does not occur on the boundary, i.e., extrema is an interior point of the domain of the function log $L(\tilde{\theta}; t)$ i.e., Θ, then the MLEs of θ are defined as a solution of the equations,

$$
\begin{aligned}
\frac{\partial}{\partial\theta_i} \log L(\tilde{\theta}; t) &= 0; i = 1, \ldots, k \\
&\text{or} \\
S(t, \tilde{\theta}) &= \left(\frac{\partial}{\partial\theta_1} \log L(\tilde{\theta}), \ldots, \frac{\partial}{\partial\theta_k} \log L(\tilde{\theta})\right)' = (0, \ldots, 0)'
\end{aligned}
$$

hence, the $k \times k$ matrix is given by

$$\frac{\partial}{\partial \theta} S(t,\tilde{\theta})\bigg|_{\theta=\theta_0(t)} = \left[\frac{\partial^2}{\partial \theta_i \partial \theta_j} \log L(\tilde{\theta};t)\right]_{\theta=\theta_0(t)} = \left|-J_{ij}(\theta_0)\right| \tag{1.15}$$

is negative definite $\mathbf{J}_{ij}(\theta_0)$ is positive definite, where is the solution of equation (1.15). The matrix $\mathbf{J}_{ij}(\theta_0)$ is called the observed information matrix. It measures the amount of information about θ available in the experiment. However, the matrix

$$\begin{aligned} I(\theta) &= E_\theta[S(T,\theta)S(T,\theta)'] \\ &= \left|E_\theta\left(\frac{\partial}{\partial \theta_i}\log L(\theta)\right)\left(\frac{\partial}{\partial \theta_j}\log L(\theta)\right)\right| \\ &= \left|E_\theta\left(-\frac{\partial^2}{\partial \theta_i \partial \theta_j}\log L(\theta;t)\right)\right| \\ &= \left|E_\theta J_{ij}(\theta)\right| = E_\theta[J(\theta)] \end{aligned}$$

is called the Fisher Information matrix. A large value of $|I(\theta_0)|$ indicates that the likelihood function is more curved; thus one can easily calculate the maximum of likelihood functions over θ to get the MLE.

1.4.3 Confidence Intervals of Estimates

The asymptotic variance-covariance matrix of the parameters (λ and β) is obtained by inverting the Fisher information matrix

$$I_{ij} = E\left[-\frac{\partial^2 L}{\partial \theta_i \partial \theta_j}\right], \quad i,j = 1,2$$

where $\theta_1, \theta_2 = \lambda$ orβ (Nelson 1990)[14]. This leads to

$$\begin{bmatrix} \mathrm{Var}(\hat{\lambda}) & \mathrm{Cov}(\hat{\lambda},\hat{\beta}) \\ \mathrm{Cov}(\hat{\lambda},\hat{\beta}) & \mathrm{Var}(\hat{\beta}) \end{bmatrix} = \begin{bmatrix} E\left(-\frac{\partial^2 \ln L}{\partial^2 \lambda}\Big|_{\hat{\lambda},\hat{\beta}}\right) & E\left(-\frac{\partial^2 \ln L}{\partial \lambda \partial \beta}\Big|_{\hat{\lambda},\hat{\beta}}\right) \\ E\left(-\frac{\partial^2 \ln L}{\partial \beta \partial \lambda}\Big|_{\hat{\lambda},\hat{\beta}}\right) & E\left(-\frac{\partial^2 \ln L}{\partial^2 \beta}\Big|_{\hat{\lambda},\hat{\beta}}\right) \end{bmatrix}^{-1}$$

We can obtain an approximate (1-α)100% confidence interval on parameter O and β based on the asymptotic normality of the MLE (Nelson 1990)[14] as follows:

$$\hat{\lambda} \pm Z_{\alpha/2}\sqrt{\mathrm{Var}(\hat{\lambda})} \text{ and } \hat{\beta} \pm Z_{\alpha/2}\sqrt{\mathrm{Var}(\hat{\beta})}$$

where $Z_{\alpha/2}$ is upper percentile of standard normal distribution.

1.5 Goodness of Fit Techniques

The goodness of fit techniques is statistical methods used to assess how well a statistical model fits observed data. These techniques help evaluate the adequacy of a model in describing the underlying data and can be applied in various fields, including regression analysis, hypothesis testing, and probability distributions. Some commonly used goodness of fit techniques include:

1.5.1 Chi-squared Test

The chi-square test is used to compare observed frequencies with expected frequencies based on a specific model. It assesses whether the observed data significantly deviates from the expected values, indicating a lack of fit. The following statistic

$$\chi^2 = \sum_{i=1}^{k} \left(\frac{x_i - \mu_i}{\sigma_i} \right)^2$$

has a chi-squared χ^2 distribution with k degrees of freedom. The steps of the chi-squared test are as follows:

1. Divide the sample data into mutually exclusive cells (normally 8-12) such that the range of the random variable is covered.

2. Determine the frequency, f_i, of sample observations in each cell.

3. Determine the theoretical frequency, F_i, for each cell (area under density function between cell boundaries X_n - total sample size). Note that the theoretical frequency for each cell should be greater than 1. To carry out this step, it normally requires estimates of the population parameters which can be obtained from the sample data. log-likelihood

4. Form the statistic

$$S = \sum_{i=1}^{k} \frac{(f_i - F_i)^2}{F_i}$$

5. From the χ^2 tables, choose a value of χ^2 with the desired significance level and with degrees of freedom ($= k - 1 - r$), where r is the number of population parameters estimated.

6. Reject the hypothesis that the sample distribution is the same as theoretical distribution if

$$S > \chi^2_{1-\alpha, k-1-r},$$

where $\alpha$is called the significance level.

1.5.2 Kolmogorov-Smirnov d Test

The Kolmogorov-Smirnov test evaluates the similarity between the cumulative distribution function (CDF) of the observed data and the CDF of the model. It measures the maximum difference between the two distributions and provides a statistical test for assessing the goodness of fit.

Both the χ^2 and “d” tests are non-parameters. However, the χ^2 assumes large sample normality of the observed frequency about its meanwhile the “d” only assumes a continuous distribution. Let$X_1 \leq X_2 \leq X_3 \leq ... \leq X_n$ denote the ordered sample values. Define the observed distribution function, $F_n(x)$, as follows:

$$F_n(X) = \begin{cases} 0 & \text{for } x \leq x_1 \\ \frac{i}{n} & \text{for } x_i < x \leq x_{i+1} \\ 1 & \text{for } x > x_n \end{cases}$$

Assume the testing hypothesis

$$H_0 : F(x) = F_0(x)$$

where $F_0(x)$is a given continuous distribution and $F(x)$ is an unknown distribution. Let

$$d_n = \sup_{-\infty<x<\infty} |F_n(x) - F_0(x)|$$

Since $F_0(x)$ is a continuous increasing function, we can evaluate $|F_n(x) - F_0(x)|$ for each n. If $d_n \leq d_n$, D then we would not reject the hypothesis H_o; otherwise, we would reject it when $d_n > d_{n,\alpha}$, D. The value $d_{n,\alpha}$ can be found in Table of" Kolmogorov-Smirnov d " where n is the sample size and a is the level of significance.

1.5.3 Anderson-Darling Test(AD)

The Anderson-Darling test is another method for evaluating the fit of a model to data. It assesses the discrepancy between the empirical distribution function of the observed data and the theoretical distribution function of the model.
The Anderson-Darling test statistic is calculated using the following formula:

$$A^2 = -n - S,$$

where

A^2 is the Anderson-Darling test statistic.

n is the sample size.

S is the sum of the terms calculated as:

$$S = \sum [(2i-1)(\ln(F(x_i)) + \ln(1 - F(x_{n+1-i})))],$$

where

i represents the $i^t h$ ordered observation in the sample.

$F(x_i)$ is the cumulative distribution function (CDF) evaluated at the $i^t h$ ordered observation.

$F(x_{n+1-i})$ is the CDF evaluated at the $(n+1-i)^t h$ ordered observation.

The test calculates a test statistic, denoted as A^2, which measures the discrepancy between the observed data and the expected values based on the chosen distribution. A larger A^2 value indicates a greater discrepancy between the observed data and the expected distribution.

1.5.4 Information Criteria

Information criteria, such as the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), are used to compare different models based on their goodness of fit and complexity. Lower values of these criteria indicate better-fitting models.

1.5.4.1 Akaike Information Criterion (AIC)

The Akaike information criterion (AIC) is a mathematical method for evaluating how well a model fits the data it was generated from. In statistics, AIC is used to compare different possible models and determine which one is the best fit for the data.AIC is calculated using the formula:

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

where k is the number of parameters in the model and L is the maximum likelihood estimate of the model.

1.5.4.2 Bayesian Information Criterion (BIC)

he Bayesian Information Criterion (BIC), also known as the Schwarz criterion, is a statistical measure used for model selection among a set of candidate models. It provides a way to balance the goodness of fit of a model with its complexity, penalizing models with more parameters.

The BIC is calculated using the following formula:

$$BIC = -2\log(L) + k\log(n)$$

where

$\log(L)$ is the log-likelihood of the data given the model.

k is the number of parameters in the model.

n is the sample size or the number of observations.

The BIC is a criterion that balances model fit and complexity by penalizing models with more parameters. A lower BIC indicates a better model, helping in model selection among alternatives.

1.5.5 Cramer–von Mises criterion (CVM)

In statistics the Cramer–von Mises criterion is a criterion used for judging the goodness of fit of a cumulative distribution function F^* compared to a given empirical distribution function F_n. It is defined as,

$$\omega^2 = \int_{-\infty}^{\infty} (F_n(x) - F^*(x))^2 dF^*(x)$$

CHAPTER 2

INFERENCE FOR STEP STRESS PARTIALLY ACCELERATED LIFE TEST IN THE CHEN DISTRIBUTION

2.1 Introduction

companies in the manufacturing sector is forced to provide reliable products that meet customer's expectations However, testing of products under usual stress conditions takes a long period of time, often in years or decades. Moreover, these experiments are expensive, which makes it difficult or even impossible to obtain the failure information under usage conditions for such products. Therefore, Accelerated Life Testing (ALT) is frequently used to induce more failures by subjecting the product to extreme conditions of stress, strain, temperature or pressure. Step-stress testing is a special case of ALT, in which the stress applied to the test units that can increase at pre-specified times or at a prefixed number of failures. in this study : SSPALT (Step Stress Partially Accelerated Life Testing) with a model based on progressive-time censored data with all its types and schemes using the model of the Chen distribution and Maximum Likelihood estimation .

2.2 Accelerated Life Testing (ALT)

Life testing is a procedure used to determine the lifespan of a product or subject by observing when it fails. However, in some cases, the lifespan of a product can be extremely long, making traditional life testing procedures costly and time consuming.
ALT introduced by Chernoff (1962)[15] and Bessler et al. (1962)[16], is commonly used to provide information about the life distribution of a product or material.All or some of the test units are subjected to more severe conditions than the typical conditions.

If the experimenter includes all test units under such stresses,the test is called ALT, but if he/she includes some of them then the test is called partially accelerated life test (PALT). In ALT test units are run only at accelerated conditions, while in PALT they are run at both normal and accelerated use conditions,The major assumption in ALT is that the mathematical model relating the lifetime of the unit and the stress are known or can be assumed. In some cases, such life–stress relationships are not known and cannot be assumed, i.e ALT data cannot be extrapolated to use condition. So, in such cases, PALT is a more suitable test to be performed for which tested units are subjected to both normal and accelerated conditions.Partially accelerated life tests studied under step-stress scheme by several authors, for example, see Goel [17],DeGroot and Goel [18], Bhattacharyya and Soejoeti [19], Bai and Chung [20].

For more detailed information about the analysis of ALT, one may refer to the works of Bagdonavicius (1978)[21], Viertl (1988)[22], Nelson (1980)[23], Meeker and Escobar (1998)[24]. These references provide further insights into the statistical analysis and methodologies used in the field of ALT.

2.2.1 ALT Types

According to Nelson [23] There are several types of ALT methods that can be used depending on the specific objectives and constraints of the testing process. Here are some commonly used ALT types:

2.2.1.1 Constant-Stress ALT (CSALT)

The stress that is given to the test goods in CSALT is independent of time. A continuous, higher than normal degree of stress is applied to the testing units until either all units fail (full case) or the test is stopped, producing censored test results(see [25],[26]).

2.2.1.2 Progressive-Stress ALT (PSALT)

In PSALT, the stress applied to a test product continuously increases over time (see [27],[28]).

2.2.1.3 Step-Stress ALT (SSALT)

In a step-stress accelerated life test, all test unites are first subjected to a predetermined low stress level and then at a pre-specified time, the stress level is increased and then the test is continued until a next pre-determined time. In this test,if the stress is changed only once it is known as a Simple SSALT and if the stress is changed

multiple times, then it is called a multiple SSALT. One of the major issuesin dealing with data from SSALT is the explanation of the effect of the stress change on the remaining lifetime of the test units (see [20],[29]).
There were some proposed solutions for this problem from which the model of tampered random variable model (TRVM) which was introduced by DeGroot and Goel (1979)[18] for partially accelerated life test, this test is similar to Simple SSALT, except that the low stress level is actually the normal working condition of the item/machine.

2.2.1.4 Step-Stress Partially ALT(SSPALT)

Step-Stress Partially Accelerated Life Testing. It is a specific type of ALT method that combines the concepts of step-stress testing and partially accelerated life testing.
In SS-PALT, all of the n test units are initially tested under normal conditions. If a unit does not fail within a pre-specified time period, it is then subjected to accelerated conditions until failure. The purpose of this approach is to determine the effects of switching from normal to accelerated stress levels on the remaining lifetime of the test units.

Advantages

- It provides a cost-effective approach by reducing the testing time compared to full testing at each stress level.
- It allows for the evaluation of multiple stress levels, including the highest level, which can be useful in understanding the product's reliability profile.
- It provides a way to estimate the product's life expectancy and reliability at stress levels beyond its normal operating conditions.

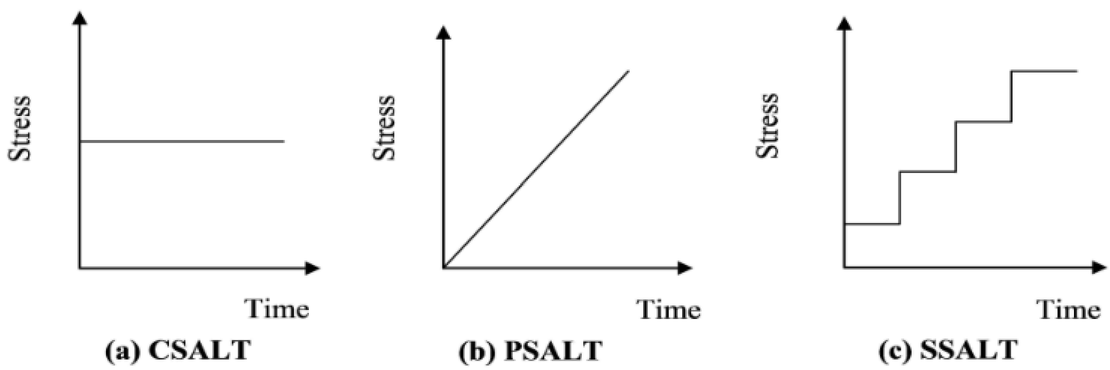

Figure 2.1: Different forms of stress loading

2.3 Censored data

Censored data refers to observations or measurements in a dataset that are incomplete or partially known. It occurs when the exact value of an event or outcome of interest is unknown or unobserved. Censoring can occur in various types of studies, including survival analysis, reliability engineering, clinical trials, and time-to-event analysis. Censored data typically arise in situations where the event of interest has not occurred for some or all of the study participants within the study period. This can happen due to several reasons, such as the study being ongoing, participants being lost to follow-up, or the event of interest not yet happening by the end of the study.

2.3.1 Censored Types

2.3.1.1 Right censored data

Right censored data is data for items that have not yet failed. They are considered “still alive” as their failure time has not yet occurred, though it is expected to occur at some point in the future. Right censored data is also referred to as “suspensions” or “survivors”.

2.3.1.2 Left censored data

In left-censored data, the failure event of interest has occurred before the start of the test or data collection process. It is similar to interval-censored data, where the exact failure time is unknown, but in the case of left-censored data, the lower interval is fixed at 0. This means that we only know that the failure event happened before or at the start of the observation period.We rarely see this type of data in reliability engineering.

2.3.1.3 Interval censored data

Interval-censored data refers to a type of data in which the exact failure time of an event of interest is not known, but we have information about the lower and upper bounds of an interval within which the failure occurred. In other words, we have knowledge that the failure event happened sometime between two specific points in time.

Generalization We could consider all censored data to be interval censored data, with the following bounds:

►**Right censored data**: lower bound = end of observation time, upper bound = infinity
►**Left censored data**: lower bound = 0, upper bound = start of observation time
►**Interval censored data**: lower bound = last observation time before failure, upper

bound = first observation time after failure

2.3.1.4 Type I censoring

Type-I censoring occurs when the study is terminated at a specific time, and any event occurring after that time is considered censored. It is a form of right censoring.

2.3.1.5 Type II Censoring

Type-II censoring occurs when a fixed number of events have occurred, and the study is terminated. Any remaining participants without observed events are considered censored. It is a form of right censoring.

2.4 Progressive Censoring Schemes

In general life-time test, it takes a lot of time and money for the experimenters to wait until all units fail. In order to save the experimental cost, the experiment can be stopped before all units have failed. The data obtained from experiments in this way are called censored data. The conventional Type-I and Type-II censoring schemes can only remove units at the terminal point of the experiment. Removal of units at other points of the experiment is not allowed, so they are limited and not flexible enough.
Therefore, we consider a more general censoring scheme named progressive type I and Type-II censored data (is proposed by Cohen 1963)[30],Subsequently, the relevance and importance of the PC were discussed in various disciplines by several authors like Mann 1971[31], Balakrishnan and Cohen [32], Balakrishnan and Aggarwala 2000 [33], and Balasooriya and Low 2004 [34].

2.4.1 Progressive Type-I censorings

Type I progressive censoring is based on predetermined time points $T_1, T_2 ...... T_m$ The failure times are recorded one by one till the experiment is completed at time point T_m. As a result, the intervention times $T_1, T_2 ...... T_m$ have fixed values, however, the sample size as well as the censoring scheme that was utilized were chosen at random (and may vary from the originally planned censoring scheme at a certain time owing to the lack of enough surviving units to complete the requisite censorship), and the period of the test is constrained by T_m whereas it is not fixed (random) in the case of progressive Type II censoring.

Figure 2.2 is an explanation of the type I progressive censoring sample where the n and R are random variables while the completed time of the experiment is a fixed value.Balakrishnan and Cramer,2014 [35]

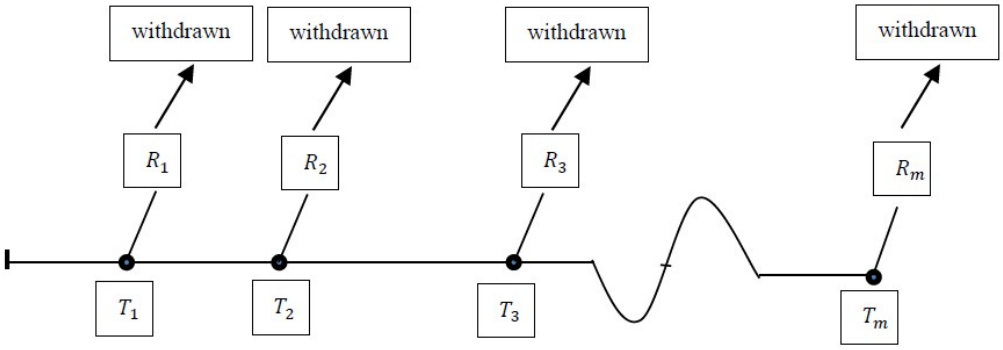

Figure 2.2: llustration of Progressive Type-I Censoring

If $R_1 = R_2 = ... = R_m = 0$, So, $n = m$ that represents the complete sample.

If $R_1 = R_2 = ... = R_{m-1} = 0$, we have $R_m = n - m$ that represents to the conventional Type.

2.4.2 Progressive Type-II Censorings

Deletions are carried out during failure periods that have been identified in the progressive Type-II censoring procedure. When a failure is detected, the number of units, that are pre-established, is instantly taken from the remaining units. As a result, the number of observations is predetermined, but the duration of the experiment is unpredictable. The following approach can be used to generate Type-II censored order statistics in a progressive manner.
For example, to test the reliability of products, n units are put down at the same time. When the first failure has occurred, R_1 surviving units are randomly selected and removed from the experiment. After the second failure, immediately R_2 units are withdrawn, and so on. After the m^{th} failure, the process is repeated until all R_m remaining units are removed. The R_i 's are specified before starting the experiment. The investigator determines the number of items in the sample n, the number of failures m, and the scheme of the progressive censoring samples. before starting a life experiment. $(R_1, R_2, ..., R_m)$, with $n = m + \sum_{i=1}^{m} R_i$.

Figure below is an explanation of a Type II progressive censoring sample where the n, m and R are fixed values while the completed time of the experiment is a random variable. (Siyi Chen and Wenhao Gui, 2020) [36]

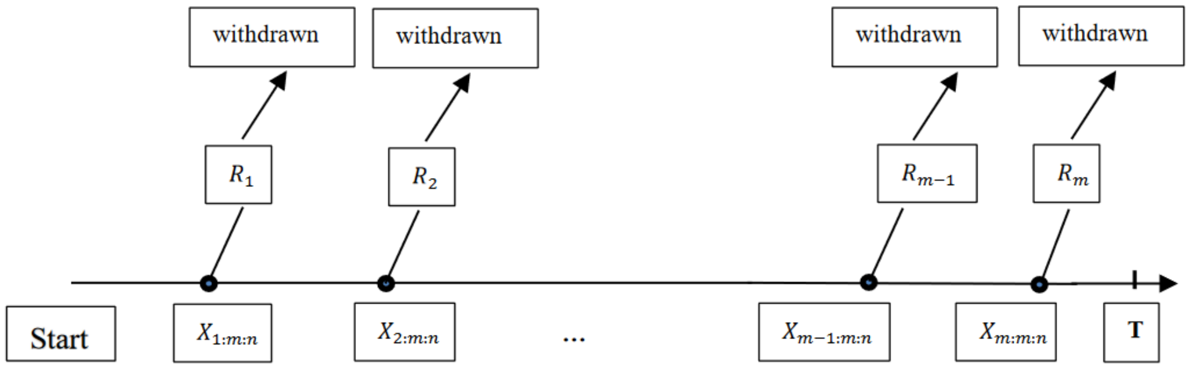

Figure 2.3: llustration of Progressive Type-II Censoring

2.5 General Progressive Type II censoring

We can generalize this by assuming that the first$r$ failed units are not observed. When $(r+1)^{th}$ failure occurs, R_{r+1} units are removed randomly. When the failure occurs next, we randomly remove some units as before,and so on at the m^{th}failure. This is General Progressive Type-II Censoring. It was first popularized by Balakrishnan and Sandhu [37]. When $r = 0$, this data reduce to Progressive Type-II Censoring. Schematic diagrams are presented in Figures 2.3.

General progressive Type-II censored data have a broad applied value, such as in the cold standby series system (see in [38],[39]). Some scholars have also done some research on this kind of censored data. Balakrishnan and Lin [40] developed exact interval estimation for the location and scale parameters of an exponential distribution based on general progressively Type-II censored samples. Balakrishnan and Jie [41] proved the existence and uniqueness of the MLEs of the parameters for normal population, based on general progressively Type-II censored samples. Ma and Gui [42] obtained an analytic and explicit estimator of the scale parameter and proposed two new goodness of fit tests for the scale family based on general progressively Type-II censored samples.

Suppose n units are placed on a life testing experiment and let $T_1, T_2, ..., T_n$ be their corresponding lifetimes. We assume that T_i, $i = 1, 2, ..., n$ are independent and identically distributed with pdf $f(t)$ and cdf $F(t)$. With progressively Type II censoring,n units are placed on test. Consider that $T_{1;m,n} < T_{2;m,n} < ... < T_{m;m,n}$ is the corresponding progressively Type II censored sample, with censoring scheme $R = (R_1, R_2, ..., R_m)$. Since the joint pdf of $T_{1;m,n} < T_{2;m,n} < ... < T_{m;m,n}$ is given by Balakrishnan and Aggarwala [33] as:

$$f_{1,2,...,m}(t_{1:m:n}, t_{2:m:n}, \ldots, t_{m:m:n}) = c\prod_{i=1}^{m} f(t_{i:m:n})\left[1 - F(t_{i:m:n})\right]^{R_i}$$

$$0 < t_{1:m:n} < t_{2:m:n} < \cdots < t_{m:m:n} < \infty,$$

where

$$C = n(n - R_1 - 1)(n - R_1 - R_2 - 2)\cdots$$

$$c = \prod_{i=1}^{m}\left(n - m + 1\sum_{i=1}^{m-1} R_i\right), m \le n$$

$f(\cdot)$ is the probability distribution function of the underlying distribution, $F(\cdot)$ is the cumulative distribution function of the underlying distribution.

2.6 Model Description and Basic Assumptions

2.6.1 Chen distribution

In 1999 Chen introduced The Chen distribution,which is relatively a new distribution in lifetime distributions , It is commonly used in reliability analysis and survival studies. It is a flexible distribution that can model various shapes of failure rates, including both increasing and bathtub-shaped failure rates.(see[43])

This distribution has a two parameters: the scale parameter (α) and the shape parameter (λ). The scale parameter determines the spread or duration of the distribution, while the shape parameter influences the shape of the hazard rate curve.

Let random variable T have a Chen distribution (CD) with two positive parameters $\lambda \ge 0$ and $\alpha \ge 0$ has the following probability density function (pdf),

$$f(t;\lambda,\alpha) = \lambda\alpha t^{\lambda-1}\exp\left\{t^{\lambda} + \alpha\left(1 - e^{t^{\lambda}}\right)\right\}, \quad t > 0, \lambda, \alpha > 0 \tag{2.1}$$

The cumulative distribution function(cdf) and the survival function,hazard (failure rate) function , respectively, given by,

$$F(t;\lambda,\alpha) = 1 - \exp\left\{\alpha\left(1 - e^{t^{\lambda}}\right)\right\}, \quad t > 0, \lambda, \alpha > 0 \tag{2.2}$$

and

$$S(t;\lambda,\alpha) = R(t;\lambda,\alpha) = 1 - F(t;\lambda,\alpha) = \exp\left\{\alpha\left(1 - e^{t^{\lambda}}\right)\right\}, \quad t > 0. \tag{2.3}$$

and

$$h(t;\lambda,\alpha) = \frac{f(t;\lambda,\alpha)}{S(t;\lambda,\alpha)} = \lambda\alpha t^{\lambda-1}e^{t^{\lambda}}, \quad t > 0$$

When λ is fixed at some value less than 1 , the relationship between the mean of the

Chen distribution and α is shown in Figure 2.4. The mean decreases when the value of α increases. Additionally, letting the cdf of the Chen distribution equal $\frac{1}{2}$, we can solve for the median

$$x_m = \left[\ln\left(1 + \frac{1}{\alpha}\ln 2\right)\right]^{\frac{1}{\lambda}}$$

Similarly, when λ is fixed and α increases, the median decreases.In addition, the Chen distribution has a great deal of flexibility.

Let T be a nonnegative random variable and follow the Chen distribution. Following Ahmed et al. [44],

$$Y = e^{t^{\lambda}} - 1$$

follows the exponential distribution and

$$Y = \left(e^{t^{\lambda}} - 1\right)^{\frac{1}{\theta}}$$

follows the Weibull distribution.

Furthermore, the Chen distribution reduces to the Gompertz distribution when $\lambda = 1$.

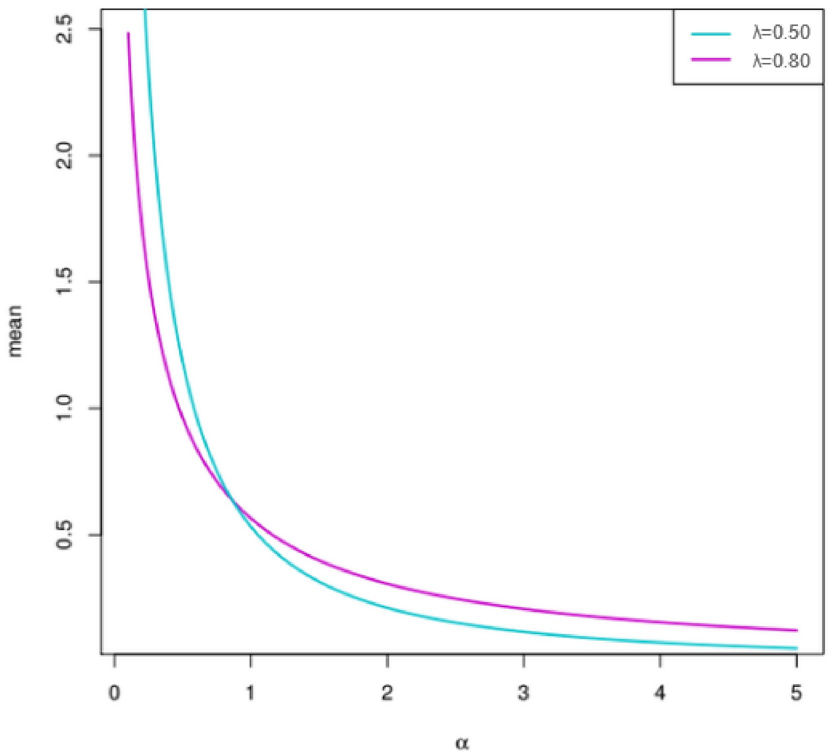

Figure 2.4: The influence of α on the mean of the Chen distribution when λ is fixed.

In the figures below, we present the different functions of the Chen distribution:

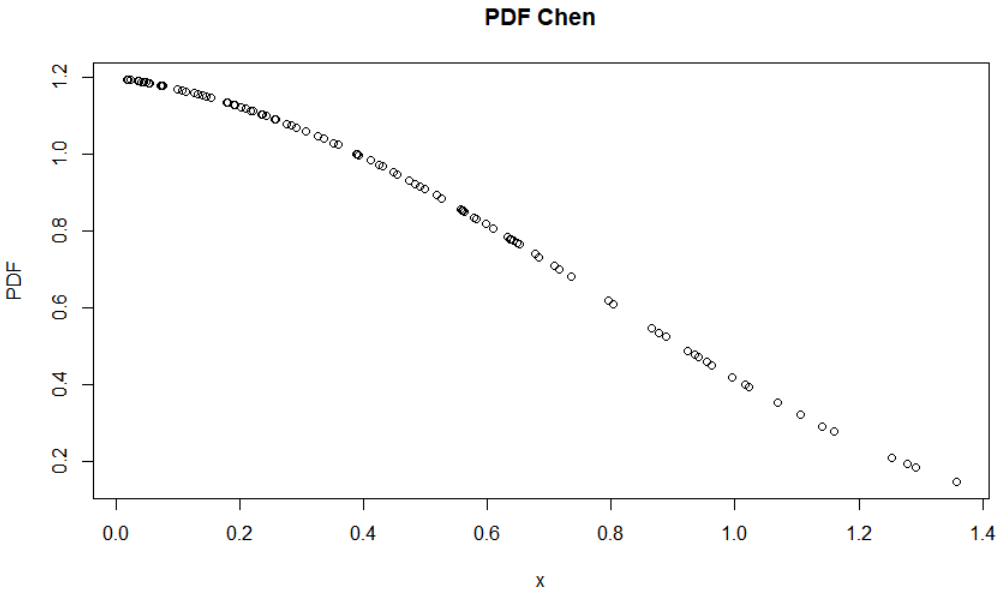

Figure 2.5: Probability density function of Chen distribution.

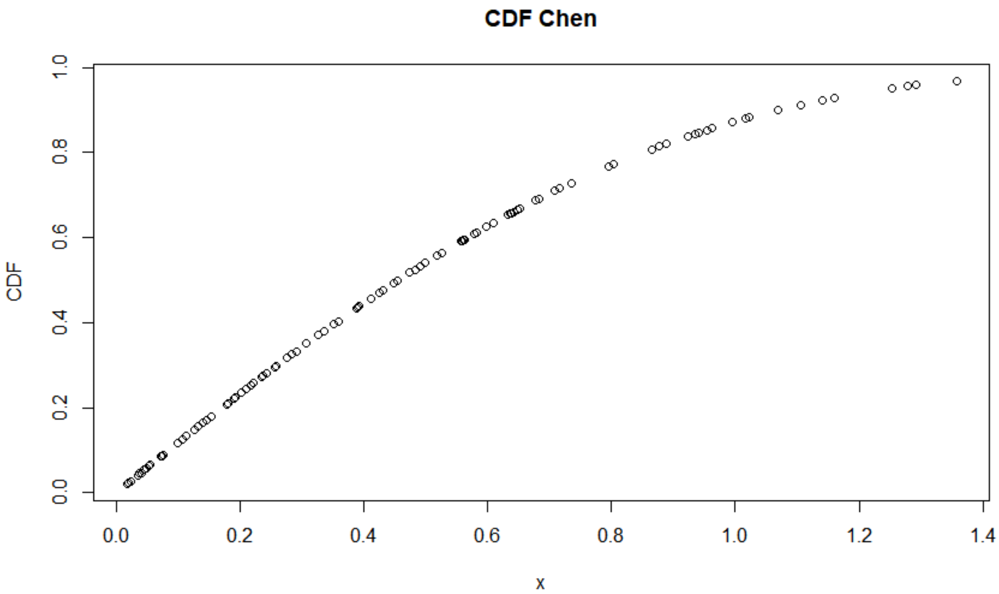

Figure 2.6: Cumulative density function of Chen distribution.

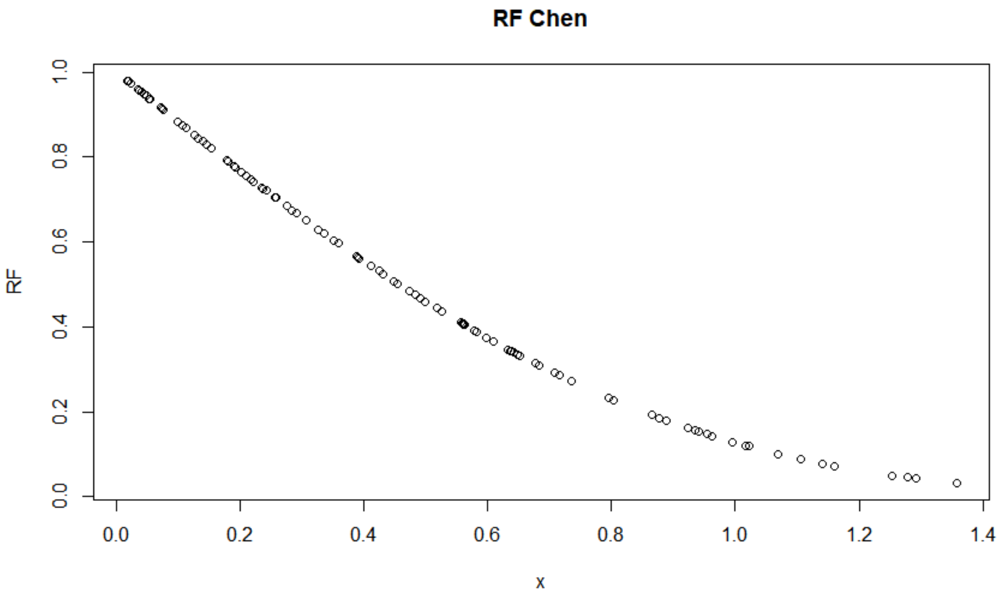

Figure 2.7: Reliability function of Chen distribution.

2.6.2 Assumptions and Test Procedure

The following assumptions are used throughout the paper in the framework of SS-PALT.

(1):n identical and independent items are put on a life test. The lifetime of each unit has Chen(λ, α) distribution.

(2): The test is terminated at the time of the m^{th} failure,where m is prefixed $(m \leq n)$.

(3): Each of the n items is run under normal use condition.If it does not fail or remove from the test by a prespecified time τ, it is put under accelerated condition.

(4): At the time of the i^{th} failure, a random number of the surviving items R_i, $i = 1, 2, \ldots, m-1$ are randomly selected and removed from the test. Finally, at the time of the m^{th} failure, the remaining surviving items $R_m = n - m - \sum_{i=1}^{m-1} R_i$ are removed from the test and the test is terminated.

(5): Let n_1 be the number of failures before time τ at normal condition, and let $m - n_1$ be the number of failures after time τ at stress condition, then, the observed progressive censored data are

$$y^R_{1;m,n} < \ldots < y^R_{n_1;m,n} < \tau < y^R_{n_1+1;m,n} \cdots < y^R_{m;m,n},$$

where $R = (R_1, R_2, \ldots, R_m)$ and $\sum_{i=1}^{m-1} R_i = n - m$.

(6):Assume that the random variable Y represents the lifetime of a product which are identical and independent and has Chen distribution with the parameters λ and α, respectively. The probability density function (pdf) of Y is,

$$f(y) = \lambda \alpha y^{\lambda-1} \exp\left\{y^{\lambda} + \alpha\left(1 - e^{y^{\lambda}}\right)\right\}, \quad y > 0, \lambda, \alpha > 0 \tag{2.4}$$

In order to account for the effect of stress change on the remaining lifetime of the testing units of the step stress accelerated model, we have applied the tampered random variable model (TRVM) which was introduced by DeGroot and Goel (1979)[18]. According to tampered random variable model the lifetime of a unit under SSPALT can be written as,

$$Y = \begin{cases} T, & \text{if } T \leq \tau \\ \tau + \beta^{-1}(T - \tau), & \text{if } T > \tau \end{cases} \tag{2.5}$$

where T is the lifetime of the unit under use condition,τ is the stress change time and β and is the acceleration factor which is the ratio of mean life at use condition to that at accelerated condition, usually $\beta > 1$.

The pdf of Y (Chen distribution) under SS-PALT is shown as below: The corrected mathematical form of the expression is as follows:

$$f(y) = \begin{cases} f_1(y), & \text{if } y \leq \tau, \\ f_2(\tau + \beta(y - \tau)), & \text{if } y > \tau, \end{cases}$$

where

$$f(y) = \begin{cases} \lambda \alpha y^{\lambda-1} \exp\left(y^{\lambda} + \alpha(1 - \exp y^{\lambda})\right), & \text{if } y < \tau, \\ \lambda \alpha \beta(\tau + \beta(y - \tau))^{\lambda-1} \exp\left((\tau + \beta(y - \tau))^{\lambda} + \alpha(1 - \exp(\tau + \beta(y - \tau))^{\lambda})\right), & \text{if } y > \tau. \end{cases} \tag{2.6}$$

where $f_1(y)$, is given by (2.4) and $f_2(y)$ is obtained by the transformation variable technique using equations (2.1) and (2.5).

the CDF given in 2.2, the CDF of a test unit under the SSP-ALT is

$$F(y) = \begin{cases} F_1(y), & \text{if } y \leq \tau, \\ F_2(\tau + \beta(y - \tau)), & \text{if } y > \tau, \end{cases}$$

where

$$F(y) = \begin{cases} 1 - \exp\left\{\alpha(1 - \exp(y^{\lambda}))\right\}, & \text{if } y \le \tau, \\ 1 - \exp\left\{\alpha(1 - \exp((\tau + \beta(y - \tau))^{\lambda}))\right\}, & \text{if } y > \tau. \end{cases} \tag{2.7}$$

also, the reliability function is given by ,

$$S(y) = \begin{cases} \exp\left\{\alpha(1 - \exp(y^{\lambda}))\right\}, & \text{if } y \le \tau, \\ \exp\left\{\alpha(1 - \exp((\tau + \beta(y - \tau))^{\lambda}))\right\}, & \text{if } y > \tau. \end{cases} \tag{2.8}$$

In progressive Type II censoring the test terminates when the number of observations is reached to $m < n$.The observed values of the total lifetime Y are : $y_1 < y_2 < ... < y_{<}\tau < y_{n_1+1} < ... < y_m$ where n_1 are the number of items failed at normal conditions and $m - n_1$ at accelerated conditions. Let us define the two indicator functionsalign

$$\delta_{1i} = \begin{cases} 1, & \text{if } y_i \le \tau, \\ 0, & \text{otherwise}, \end{cases}, \delta_{2i} = \begin{cases} 1, & \text{if } y_i > \tau, \\ 0, & \text{otherwise}. \end{cases}, \forall i = 1, 2, ..., m. \tag{2.9}$$

For the lifetimes $y_1 < y_2 < ... < y_m$ of m items are independent and identically distributed random variables,then the likelihood function is given by,

$$L(\lambda, \alpha, \beta \mid y) = c \prod_{i=1}^{m} [f_1(y_i)[S_1(y_i)]^{R_i}]^{\delta_{1i}} [f_2(y_i)[S_2(y_i)]^{R_i}]^{\delta_{2i}} \tag{2.10}$$

$$y_1 < y_2 < \ldots < y_{n_1} < \tau < y_{n_1+1} < \ldots < y_m < \infty$$

then,

$$\begin{aligned} L(\lambda, \alpha, \beta \mid y) = C \prod_{i=1}^{n_1} &\left[\lambda \alpha y_i^{\lambda-1} \exp\left\{ y_i^{\lambda} + \alpha(1 - \exp(y_i^{\lambda})) \right\} \exp\left\{ \alpha(1 - \exp(y_i^{\lambda})) \right\}^{R_i} \right] \\ &\times \prod_{i=n_1+1}^{m} \left[\lambda \alpha \beta (\tau + \beta(y_i - \tau))^{\lambda-1} \exp\left\{ (\tau + \beta(y_i - \tau))^{\lambda} + \alpha(1 - \exp(\tau + \beta(y_i - \tau))^{\lambda}) \right\} \right. \\ &\left. \times \exp\left\{ \alpha(1 - \exp(\tau + \beta(y_i - \tau))^{\lambda}) \right\}^{R_i} \right] \end{aligned} \tag{2.11}$$

where

$$C = n(n-1-R_1)(n-2-R_1-R_2)\dots(n-m+1\sum_{i=1}^{m-1} R_i) \tag{2.12}$$

2.7 Maximum Likelihood Estimation

2.7.1 MLEs

From(2.11), and considering the random variables $R = (R_1, R_2, ..., R_m)$, the likelihood function $L(\lambda, \alpha, \beta \mid y)$ without the normalized constant can be expressed as,

$$L(\lambda,\alpha,\beta \mid y) \propto A\left\{\prod_{i=1}^{n_1} \lambda\alpha y_i^{\lambda-1}\exp(y_i^{\lambda})\exp\left\{(R_i+1)\left[\alpha(1-\exp(y_i^{\lambda}))\right]\right\}\right\}$$

$$\times\left\{\prod_{i=n_1+1}^{m} \lambda\alpha\beta(\tau+\beta(y_i-\tau))^{\lambda-1}\exp(\tau+\beta(y_i-\tau))^{\lambda-1}\exp\left\{(R_i+1)\left[\alpha(1-\exp(\tau+\beta(y_i-\tau))^{\lambda})\right]\right\}\right\}$$

Then, the log-likelihood of the function $L(\lambda, \alpha, \beta \mid y)$ is given by,

$$\begin{aligned} \ell(\lambda,\alpha,\beta \mid y) &= m\log\lambda + m\log\alpha + (m-n_1)\log\beta + (\lambda-1)\sum_{i=1}^{n_1}\log(y_i) \\ &+\sum_{i=1}^{n_1} {y_i}^{\lambda} + \alpha\sum_{i=1}^{n_1}(R_i+1)(1-\exp(y_i^{\lambda})) + (\lambda-1) \\ &\times\sum_{i=n_1+1}^{m}\log(\tau+\beta(y_i-\tau)) + \sum_{i=n_1+1}^{m}(\tau+\beta(y_i-\tau))^{\lambda} \\ &+\alpha\sum_{i=n_1+1}^{m}(R_i+1)(1-\exp(\tau+\beta(y_i-\tau))^{\lambda}) \end{aligned} \tag{2.13}$$

Calculating the first partial derivatives of (2.13) with respect to λ, α and β , we obtain,

$$\begin{aligned} \frac{\partial\ell}{\partial\lambda} &= \frac{m}{\lambda} + \sum_{i=1}^{n_1}\log(y_i) + \sum_{i=1}^{n_1}(y_i)^{\lambda}\ln(y_i) - \alpha\sum_{i=1}^{n_1}(1+r_i)\exp(y_i)^{\lambda}({y_i}^{\lambda}\ln(y_i)) \\ &+\sum_{i=n_1+1}^{m}\log(\tau+\beta(y_i-\tau)) + \sum_{i=n_1+1}^{m}(\tau+\beta(y_i-\tau))^{\lambda}\ln(\tau+\beta(y_i-\tau)) \\ &-\alpha\sum_{i=n_1+1}^{m}(1+r_i)\exp(\tau+\beta(y_i-\tau))^{\lambda}((\tau+\beta(y_i-\tau))^{\lambda}\ln((\tau+\beta(y_i-\tau))) = 0 \end{aligned} \tag{2.14}$$

$$\frac{\partial \ell}{\partial \alpha} = \frac{m}{\alpha} + \sum_{i=1}^{n_1} (R_i + 1)(1 - \exp(y_i)^{\lambda}) + \sum_{i=n_1+1}^{m} (R_i + 1)(1 - \exp(\tau + \beta(y_i - \tau))^{\lambda}) = 0 \tag{2.15}$$

and

$$\begin{aligned} \frac{\partial \ell}{\partial \beta} &= \frac{m - n_1}{\beta} + (\lambda - 1) \sum_{i=n_1+1}^{m} \frac{(y_i - \tau)}{(\tau + \beta(y_i - \tau))} \\ &+ \lambda \sum_{i=n_1+1}^{m} (y_i - \tau)(\tau + \beta(y_i - \tau))^{\lambda - 1} \\ &- \alpha\lambda \sum_{i=n_1+1}^{m} (R_i + 1)(y_i - \tau)(\tau + \beta(y_i - \tau))^{\lambda - 1} \exp(\tau + \beta(y_i - \tau))^{\lambda} = 0 \end{aligned} \tag{2.16}$$

From (2.15), we can write

$$\alpha(\lambda, \beta) = \frac{m}{D}. \tag{2.17}$$

where

$$D = \sum_{i=1}^{n_1} (r_i + 1)(\exp(y_i{}^{\lambda}) - 1) + \sum_{i=n_1+1}^{m} (r_i + 1)(\exp(\tau + \beta(y_i - \tau))^{\lambda} - 1) \tag{2.18}$$

By substituting (2.17) in (2.14) and (2.16) we have,

$$\begin{aligned} &\frac{m}{\lambda} + \sum_{i=1}^{n_1} \log(y_i) + \sum_{i=1}^{n_1} (y_i)^{\lambda} \ln(y_i) - \frac{m}{D} \sum_{i=1}^{n_1} (1 + r_i) \exp(y_i)^{\lambda} (y_i{}^{\lambda} \ln(y_i)) \\ &+ \sum_{i=n_1+1}^{m} \log(\tau + \beta(y_i - \tau)) + \sum_{i=n_1+1}^{m} (\tau + \beta(y_i - \tau))^{\lambda} \ln(\tau + \beta(y_i - \tau)) \\ &- \frac{m}{D} \sum_{i=n_1+1}^{m} (R_i + 1) \exp(\tau + \beta(y_i - \tau))^{\lambda} ((\tau + \beta(y_i - \tau))^{\lambda} \ln((\tau + \beta(y_i - \tau))) = 0 \end{aligned} \tag{2.19}$$

and

$$\begin{aligned}&\frac{m-n_1}{\beta}+(\lambda-1)\sum_{i=n_1+1}^{m}\frac{(y_i-\tau)}{(\tau+\beta(y_i-\tau))}\\&-\frac{m\lambda}{D}\sum_{i=n_1+1}^{m}(1+r_i)\exp(\tau+\beta(y_i-\tau))^{\lambda}(y_i-\tau)(\tau+\beta(y_i-\tau))^{\lambda-1}\\&+\lambda\sum_{i=n_1+1}^{m}(y_i-\tau)(\tau+\beta(y_i-\tau))^{\lambda-1}=0\end{aligned}\tag{2.20}$$

By simultaneously solving equations (2.14,2.15,2.16), the MLEs $(\widehat{\lambda},\widehat{\alpha},\widehat{\beta})$ of the unknown parameters (λ,α,β) of the model given here can be found. Unfortunately, because the system of equations (2.14,2.15,2.16) is n nonlinear, there is no analytical way to arrive at a closed-form solution. Due to this issue, several iterative techniques must be utilized to acquire estimates of the model's unknown parameters (λ,α,β) . There are various iterative strategies for resolving nonlinear equations.

Thus, likelihoods equations are reduced to a two nonlinear equations (2.19) and (2.20) which could be solved numerically with respect to λ and β using any iteration procedure such as quasi-Newton Raphson, to get the MLE, $\widehat{\lambda}$ and $\widehat{\beta}$ and hence $\widehat{\alpha}$ by using (2.17).In this instance, we used the R statistical language and the software's optim() function to solve our nonlinear equations.

2.7.2 Confidence Intervals of Estimates

In this subsection, using the asymptotic properties of MLE, we determine the CI of model parameters. Given specific regularity requirements asymptotic features indicate that MLE are approximately distributed according to the normal distribution with mean zero and variance I^{-1}, which can be represented mathematically as follows

$$(\widehat{\alpha}-\alpha,\widehat{\lambda}-\lambda,\widehat{\beta}-\beta)\sim N\left(0,(I)^{-1}\right)$$

where $(I)^{-1}$is the inverse of the observed Fisher information matrix and is commonly referred to as a variance-covariance matrix for MLEs. It is possible to derive it as follows:

$$(I)^{-1}=\begin{bmatrix}\frac{\partial^2\log\ell}{\partial\lambda^2}&\frac{\partial^2\log\ell}{\partial\lambda\partial\alpha}&\frac{\partial^2\log\ell}{\partial\lambda\partial\beta}\\\frac{\partial^2\log l}{\partial\alpha\partial\lambda}&\frac{\partial^2\log l}{\partial\alpha^2}&\frac{\partial^2\log l}{\partial\beta\partial\alpha}\\\frac{\partial^2\log l}{\partial\beta\partial\alpha}&\frac{\partial^2\log l}{\partial\alpha\partial\beta}&\frac{\partial^2\log l}{\partial\beta^2}\end{bmatrix}^{-1}_{\widehat{\alpha},\widehat{\lambda},\widehat{\beta}}=\begin{bmatrix}var(\widehat{\lambda})&covar(\widehat{\lambda\alpha})&covar(\widehat{\lambda\beta})\\covar(\widehat{\alpha}\widehat{\lambda})&var(\widehat{\lambda})&covar(\widehat{\beta}\widehat{\alpha})\\covar(\widehat{\beta}\widehat{\alpha})&covar(\widehat{\lambda\beta})&var(\widehat{\beta})\end{bmatrix}$$

The elements of I can be expressed by the following equations:

$$\begin{aligned}
\frac{\partial^2 \log \ell}{\partial \lambda^2} &= -\frac{m}{\lambda^2} + \sum_{n_1+1}^{m} (y_i)^{\lambda} \log(y_i)^2 + \sum_{i=n_1+1}^{m} (\tau + \beta(y_i - \tau))^{\lambda} (\log(\tau + \beta(y_i - \tau)))^2 \\
&\quad -\alpha \sum_{i=1}^{n_1} (r_i + 1) {y_i}^{\lambda} (1 + y_i^{\lambda}) (\log y_i)^2 \exp(y_i^{\lambda}) \\
&\quad -\alpha \sum_{i=n_1+1}^{m} (r_i + 1)(\log(\tau + \beta(y_i - \tau)))^2 (\tau + \beta(y_i - \tau))^{\lambda} (1 + (\tau + \beta(y_i - \tau))^{\lambda}) \\
\frac{\partial^2 \log \ell}{\partial \alpha^2} &= -\frac{m}{\alpha^2}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 \log \ell}{\partial \beta^2} &= -\frac{m - n_1}{\beta^2} - (\lambda - 1) \sum_{i=n_1+1}^{m} \frac{(y_i - \tau)^2}{(\tau + \beta(y_i - \tau))^2} \\
&\quad -\alpha\lambda(\lambda - 1) \sum_{i=n_1+1}^{m} (r_i + 1)(y_i - \tau)^2 (\tau + \beta(y_i - \tau))^{\lambda - 2} \exp((\tau + \beta(y_i - \tau))^{\lambda}) \\
&\quad -\alpha\lambda^2 \sum_{i=n_1+1}^{m} (r_i + 1)(y_i - \tau)^2 (\tau + \beta(y_i - \tau))^{2(\lambda - 1)} \exp((\tau + \beta(y_i - \tau))^{\lambda})
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 \log l}{\partial \lambda \partial \alpha} = \frac{\partial^2 \log l}{\partial \alpha \partial \lambda} &= -\sum_{i=1}^{n_1} (r_i + 1) \exp((y_i)^{\lambda}) y_i^{\lambda} \ln(y_i) \\
&\quad - \sum_{i=n_1+1}^{m} (r_i + 1) \exp((\tau + \beta(y_i - \tau))^{\lambda}) (\tau + \beta(y_i - \tau))^{\lambda} \ln(\tau + \beta(y_i - \tau))
\end{aligned}$$

$$
\begin{aligned}
\frac{\partial^2 \log l}{\partial\lambda\partial\beta} = \frac{\partial^2 \log l}{\partial\beta\partial\lambda} = & \sum_{i=n_1+1}^{m} \frac{(y_i-\tau)}{(\tau+\beta(y_i-\tau))} \\
& + \sum_{i=n_1+1}^{m} (y_i-\tau)(\tau+\beta(y_i-\tau))^{\lambda-1} \\
& +\lambda \sum_{i=n_1+1}^{m} (y_i-\tau)(\tau+\beta(y_i-\tau))^{\lambda-1} \log(\tau+\beta(y_i-\tau)) \\
& -\alpha \sum_{i=n_1+1}^{m} (R_i+1)(y_i-\tau)(\tau+\beta(y_i-\tau))^{\lambda-1} \exp(\tau+\beta(y_i\tau))^{\lambda} \\
& -\alpha\lambda \sum_{i=n_1+1}^{m} (R_i+1)(y_i-\tau)(\tau+\beta(y_i-\tau))^{\lambda-1} \log(\tau+\beta(y_i-\tau)) \exp(\tau+\beta(y_i-\tau))^{\lambda} \\
& -\alpha\lambda \sum_{i=n_1+1}^{m} (R_i+1)(y_i-\tau)(\tau+\beta(y_i-\tau))^{2\lambda-1} \log(\tau+\beta(y_i-\tau)) \exp(\tau+\beta(y_i-\tau))^{\lambda}
\end{aligned}
$$

$$
\frac{\partial^2 \log l}{\partial\beta\partial\alpha} = \frac{\partial^2 \log l}{\partial\alpha\partial\beta} = \quad -\lambda \sum_{i=n_1+1}^{m} (r_i+1)(y_i-\tau)(\tau+\beta(y_i-\tau))^{\lambda-1} \exp((\tau+\beta(y_i-\tau)^{\lambda}))
$$

Now, two-sided $100(1-\Delta)\%$ ACIs for the parameter λ, α and β can be obtained as follows:

$$
\widehat{\lambda} \pm z_{\Delta/2}\sqrt{\mathrm{var}\,(\lambda)};
$$

$$
\widehat{\alpha} \pm z_{\Delta/2}\sqrt{\mathrm{var}\,(\alpha)};
$$

$$
\widehat{\beta} \pm z_{\Delta/2}\sqrt{\mathrm{var}\,(\beta)};
$$

where $\pm z_{\Delta/2}$ represents standard normal distribution's upper and lower $\Delta/2^{th}$ percentile. $var(\widehat{\lambda})$, $var(\widehat{\alpha})$, and $var(\widehat{\beta})$ are the diagonal entries of $(I)^{-1}$

2.8 Bootstrap Confidence Intervals

The bootstrap method is a resampling technique commonly used for statistical inference. It is widely applied to estimate confidence intervals, assess the bias and variance of an estimator, and calibrate hypothesis tests. There are various variations of the bootstrap method, including nonparametric and parametric approaches. Some notable references for the nonparametric and parametric bootstrap methods are Davison and Hinkley [45] and Efron and Tibshirani [46].

In this section, we will focus on two parametric bootstrap methods for estimating confidence intervals: the percentile bootstrap method proposed by Efron [46] and the

bootstrap-t method introduced by Hall [47]. Below, we provide algorithms for estimating confidence intervals using both of these methods. the parametric bootstrap methods are proposed:
(i)percentile bootstrap method Efron [46],**(ii)** bootstrapt method Hall, [47].

The algorithms for estimating the confidence intervals of parameters using both methods are illustrated below.

1. Based on the original progressively Type-II sample,$y = (y_1 < y_2 < ... < y_{n_1} < \tau < y_{n_1+1} < ... < y_m)$, obtain $\widehat{\lambda}$ and $\widehat{\beta}$ from (2.19) and (2.20) and hence $\widehat{\alpha}$ from(2.17).

2. Based on $\widehat{\lambda}$ and $\widehat{\alpha}$ and the values of n and m and τ^* with the same values of r, $(i = 1, 2, ..., m_j)$, generate $t^* = (t_1^* < t_2^* < ... < t_m^*)$ using the algorithm described in Balakrishnan and Sandhu [47], and hence $\widehat{\beta}$ in (2.5)the sample obtained $y^* = (y_1^* < y_2^* < ... < y_{n_1}^* < \tau^* < y_{n_1+1}^* < ... < y_m^*)$.

3. As in step 1 based on y^* compute the bootstrap sample estimates of $\widehat{\lambda}$, $\widehat{\alpha}$,$\widehat{\beta}$ and $\widehat{\lambda}^*$ say $\widehat{\alpha}^*$ and $\widehat{\beta}^*$.

4. Repeat the above steps 2 and 3 N times representing N different bootstrap samples. The value of N has been taken to be 1000.

5. Arrange all $\widehat{\lambda}^*$, $\widehat{\alpha}^*$ and $\widehat{\beta}^*$ in an ascending order to obtain the bootstrap sample $(\hat{\varphi}_k^{*[1]}, \hat{\varphi}_k^{*[2]}, ..., \hat{\varphi}_k^{*[N]})$,$k = 1, 2, 3$ where $(\hat{\varphi}_1^* = \lambda^*, \hat{\varphi}_2^* = \alpha^*, \hat{\varphi}_3^* = \beta^*)$.

2.8.1 Percentile bootstrap confidence interval

Let $G(z) = P(\hat{\varphi}_k^{*[j]} \leq z)$ be cumulative distribution function of $\hat{\varphi}_k^*$. Define $\hat{\varphi}_{kboot}^* = G^{-1}(z)$ for given z. The approximate bootstrap $100(1-\gamma)\%$ confidence interval of $\hat{\varphi}_k^*$ given by

$$\left[\hat{\varphi}_{kboot}^*(\frac{\gamma}{2}), \hat{\varphi}_{kboot}^*(1-\frac{\gamma}{2})\right]$$

2.8.2 Bootstrap-t confidence interval

First, find the order statistics $\delta_k^{*[1]} < \delta_k^{*[2]} ... < \delta_k^{[N]}$, where

$$\delta_k^{*[j]} = \frac{\hat{\varphi}_k^{*[j]} - \hat{\varphi}_k}{\sqrt{\operatorname{var}\left(\hat{\varphi}_k^{*[j]}\right)}}, j = 1, 2, \ldots, N, k = 1, 2, 3$$

where $(\hat{\varphi}_1 = \widehat{\lambda}, \hat{\varphi}_2 = \widehat{\alpha}, \hat{\varphi}_3 = \widehat{\beta})$
Let $H(z) = P(\delta_k^{*[j]} \geq z)$ be the cumulative distribution function of $\delta_k^{*[j]}$ For a given z, define

$$\hat{\varphi}_{kboot-t} = \hat{\varphi}_k + \sqrt{\operatorname{Var}(\hat{\varphi}_k)} H^{-1}(z);$$

The approximate $100(1-\gamma)\%$ confidence interval of $\hat{\varphi}_k$ is given by

$$\left(\hat{\varphi}_{kboot-t}(\frac{\gamma}{2}), \hat{\varphi}_{kboot-t}(1-\frac{\gamma}{2})\right)$$

CHAPTER 3

SIMULATIONS AND REAL DATA APPLICATIONS

3.1 R packages for fitting distributions

Fitting distributions to data is a very common task in statistics and consists in choosing a probability distribution modeling the random variable, as well as finding parameter estimates for that distribution. This requires judgment and expertise and generally needs an iterative process of distribution choice, parameter estimation, and quality of fit assessment. In the R (R Core Team 2014[48]):

- **Package MASS:** (Venables and Ripley 2010[49]), maximum likelihood estimation is available via the fitdistr function; other steps of the fitting process can be done using other R functions (Ricci 2005[50]). The fitdistr function in the MASS package estimates distribution parameters by maximizing the likelihood function using the optim function.

- **Package fitdistrplus:** (DelignetteMuller, Pouillot, Denis, and Dutang 2015[51]) could be preferred, such as maximum goodness-of-fit estimation (also called minimum distance estimation), as proposed in the R package actuar with three different goodness-of-fit distances (Dutang, Goulet, and Pigeon 2008[52]). While developing the fitdistrplus package, a second objective was to consider various estimation methods in addition to maximum likelihood estimation (MLE). Functions were developed to enable moment matching estimation (MME), quantile matching estimation (QME), and maximum goodness-of-fit estimation (MGE) using eight different distances. Moreover, the fitdistrplus package offers the possibility to specify a user-supplied function for optimization, useful in cases where classical optimization techniques, not included in optim, are more adequate.

- **Package distrMod:** The distrMod package (Kohl and Ruckdeschel 2010[53]) provides an object-oriented (S4) implementation of probability models and includes distribution fitting procedures for a given minimization criterion. This criterion is a user-supplied function which is sufficiently flexible to handle censored data, yet not in a trivial way. The fitting functions MLEstimator and MDEstimator return an S4 class for which a coercion method to class 'mle' is provided so that the respective functionalities (e.g., confint and logLik) from package stats4 are available, too. In fitdistrplus, we chose to use the standard S3 class system because of its ease of understanding by most R users. When designing the fitdistrplus package, we did not forget to implement method functions for the S3 classes.

- **Package moments:** This package focuses on calculating various statistical moments and related properties of probability distributions. It provides functions for estimating the parameters of several distributions based on moments.

```
fit <- fitdistmoments(data, "distribution_name")
```

- **Package actuar:** This package is specifically designed for actuarial science and insurance applications. It offers functions for fitting and working with various probability distributions commonly used in actuarial modelling.

- **Package maxLik** It contains Functions for Maximum Likelihood (ML) estimation, non-linear optimization, and related tools. It includes a unified way to call different optimizers, and classes and methods to handle the results from the Maximum Likelihood viewpoint. It also includes a number of convenience tools for testing and developing your own models.

In applied statistics, it is frequent to have to fit distributions to censored data (Klein and Moeschberger 2003[54]; Busschaert, Geeraerd, Uyttendaele, and VanImpe 2010[55]; Leha, Beissbarth, and Jung 2011[56]; Commeau, Parent, Delignette-Muller, and Cornu 2012[57]). The fitdistr function in the MASS package does not enable maximum likelihood estimation with this type of data. Some packages can be used to work with censored data, especially survival data (Therneau 2014[58]; Hirano, Clayton, and Upper 1994[59]), but those packages generally focus on specific models, enabling the fit of a restricted set of distributions. A third objective is thus to provide R users with a function to estimate univariate distribution parameters from right-, left- and interval-censored data. Few packages on the Comprehensive R Archive Network (CRAN, http://CRAN.R-project. org) provide estimation procedures for any user-supplied parametric distribution and support different types of data.

3.2 Simulation Study

In this section, Monte Carlo simulation techniques were used to determine the unknown parameters of the Chen distribution. MLE and RMSE, Bais, CI in type-II PC simulation were run for prefixed values of n, m, τ. The parameters are estimated using type-II PC data obtained through simulation under SSPALT, estimation process is carried out in accordance with the following steps using numerical simulation:

Step 1: Specify the values of n, m, τ.

Step 2: Specify the values of the parameters λ, α, and β.

Step 3: Generate n independent random observations from a uniform distribution $n \sim Uniforme(0,1)$ and sort the data.

Step 4: Generate Chen random variable using transformation as follows:
If U represents the uniform random variable from $[0,1]$. Then U_1 and U_2 are selected such that U_1 consists of the uniform random numbers that are $<= F_1(\tau)$ and U_2 consists of the uniform random numbers that are $> F_1(\tau)$. Then using inverse transformation ,

$$y_1 = \left(\log\left(1 - \frac{1}{\alpha}\ln(1-u_i)\right)\right)^{\frac{1}{\lambda}}$$

has CD with cdf given in Equation [2.7] if $y < \tau$. But if $y > \tau$ then using inverse transformation

$$y_2 = \frac{1}{\beta}\left\{\log\left(\frac{1}{\alpha}\ln(1-u_i)\right)^{\frac{1}{\lambda}} - \tau\right\} + \tau,$$

Step 5: Generate a censored data set can be considered as below
Set: $y_1 < \; < y_{n_1} \leq \tau < y_{n_1+1} < \; < y_m,$

Step 6: Create function in R for the log-likelihood functions 2.13.

Step 7: Use the MaxLiK function; which is found in a package in R to obtain the maximum likelihood estimators for unknown parameters $(\widehat{\lambda})$, $(\widehat{\alpha})$, and $(\widehat{\beta})$.

Step 8: Repeat steps $1-7$ up to 50 times, Obtain the average MLEs with their RMSE and Bias.

Step 9: Obtain CI.

Taking into account the above-mentioned algorithm, we set the initial values for $\tau = (0.3)$ and the combinations of sample sizes $(n,m) = (100,40),(100,30),(100,25),$ $(80,40),(80,30),(80,25)$. Assuming that the true values of parameters $(\lambda,\alpha,\beta) = (1.5,1.2,0.05)$, the MLEs of the parameters with their respective Bias are obtained and given in Tables 3.1 under types II of censored data. RMSE and CIs of corresponding 95% are provided in Tables 3.2.

τ	(n,m)	MLE			Bias		
		$\widehat{\lambda}$	$\widehat{\alpha}$	$\widehat{\beta}$	$\widehat{\lambda}$	$\widehat{\alpha}$	$\widehat{\beta}$
0.3	(100,40)	1.50029	1.18390	0.03143	0.50125	-0.01335	-1.98415
	(100,30)	1.51386	1.17243	0.02421	0.51042	-0.02376	-1.97985
	(100,25)	1.52491	1.16193	0.01697	0.52961	-0.04046	-1.96987
0.3	(80,40)	1.50076	1.20203	0.03951	0.50154	-0.00546	-1.98367
	(80,30)	1.52750	1.16076	0.02671	0.50420	-0.01794	-1.98028
	(80,20)	1.52834	1.15905	0.01547	0.52699	-0.03707	-1.97437

Table 3.1: MLE estimator and Bias with the real valus $\lambda = 1.5, \alpha = 1.2, \beta = 0.05$

τ	(n,m)	RMSE			CI		
		$\widehat{\lambda}$	$\widehat{\alpha}$	$\widehat{\beta}$	$\widehat{\lambda}$	$\widehat{\alpha}$	$\widehat{\beta}$
0.3	(100,40)	0.00260	0.00656	0.01123	(1.50052,1.50196)	(1.18592,1.18737)	(0.01513,0.09257)
	(100,30)	0.00536	0.00532	0.01401	(1.50893,1.51190)	(1.17475,1.17772)	(0.01866,0.08164)
	(100,25)	0.01249	0.00863	0.01576	(1.52614,1.53307)	(1.15607,1.16300)	(0.01066,0.07359)
0.3	(80,40)	0.00336	0.01064	0.00813	(1.50060,1.50247)	(1.20452,1.20639)	(0.01540,0.09726)
	(80,30)	0.00581	0.00972	0.01047	(1.50258,1.50580)	(1.18045,1.18367)	(0.01811,0.07133)
	(80,20)	0.01076	0.00706	0.01573	(1.52400,1.52997)	(1.15994,1.16591)	(0.01265,0.06861)

Table 3.2: RMSE and CI with real values of $\lambda = 1.5, \alpha = 1.2, \beta = 0.05$

From the results in tables 3.1 and 3.2, it can be observed that the results are consistent and the estimates are quite close to the predefined values, especially for λ and α , the MLE increases when of n increases till some value and then it decreases, Bias, RMSE decrease for λ,α,β when the values of n,m increase.

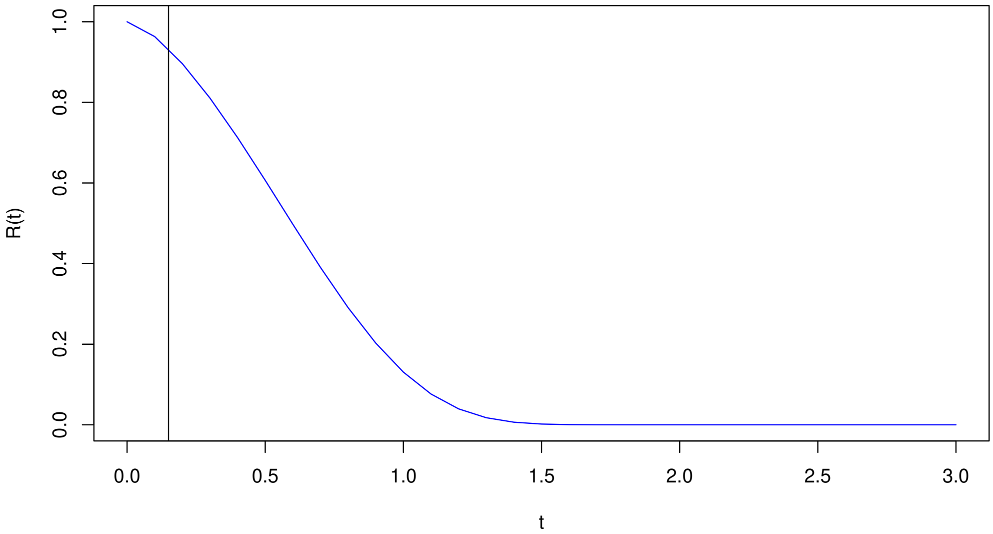

Figure 3.1: Reliability function plot for simulation study

3.3 Real data application 1

In this section, we considered the data of the tensile strength, measured in GPa, of 69 carbon fibers tested under tension at gauge lengths of 20 mm (Bader and Priest, 1982 [60]), which is fitted by power Chen distribution by Ghitany et al., 2013 [61], before. The data set is given as

---

1.312, 1.314, 1.479, 1.552, 1.700, 1.803, 1.861, 1.865, 1.944, 1.958, 1.966, 1.997, 2.006, 2.021, 2.027, 2.055, 2.063, 2.098, 2.140, 2.179, 2.224, 2.240, 2.253, 2.270, 2.272, 2.274, 2.301, 2.301, 2.359, 2.382, 2.382, 2.426, 2.434, 2.435, 2.478, 2.490, 2.511, 2.514, 2.535, 2.554, 2.566, 2.570, 2.586, 2.629, 2.633, 2.642, 2.648, 2.684, 2.697, 2.726, 2.770, 2.773, 2.800, 2.809, 2.818, 2.821, 2.848, 2.880, 2.954, 3.012, 3.067, 3.084, 3.090, 3.096, 3.128, 3.233, 3.433, 3.585, 3.585.

The summary of the sample is presented by the boxplot is represent in fig3.2 and the descriptive parameters of the sample are given in the following table:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	sd
1.312	2.098	2.478	2.451	2.773	3.585	0.495

Table 3.3: Descriptive parameters of carbon fibers sample

To assess the fit of the data on the selected model, we considered Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Kolmogorov-Smirnov

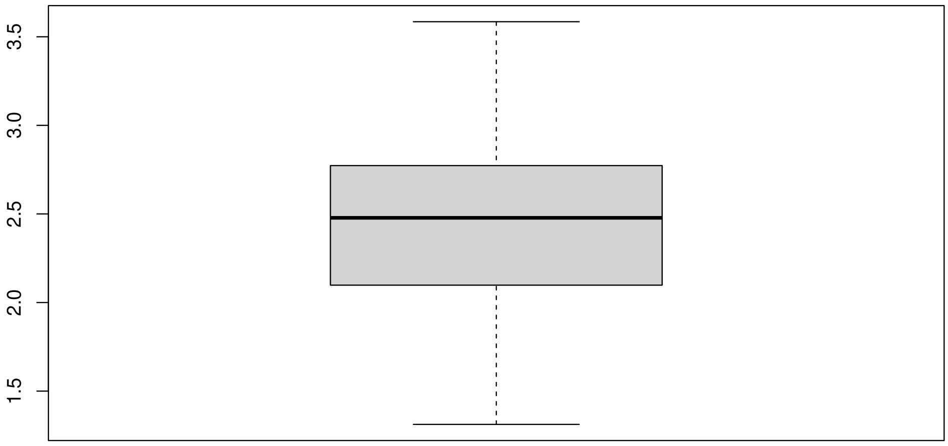

Figure 3.2: Boxplot of the carbon fiber data set

(KS) statistic, Chi-squared Test(chisq) statistic and Anderson-Darling (AD) statistics as the goodness of fit criteria. The Chen distribution is also compared to the Normal, log-normal, Weibull, exponential, Gamma,cauchy, log-logistic, student, uniform, and Chi-Square distributions, and it is noticed that among the mentioned distributions, Chen is a decent contender to fit this data set as the least value of all information criteria is observed in the case of this distribution.

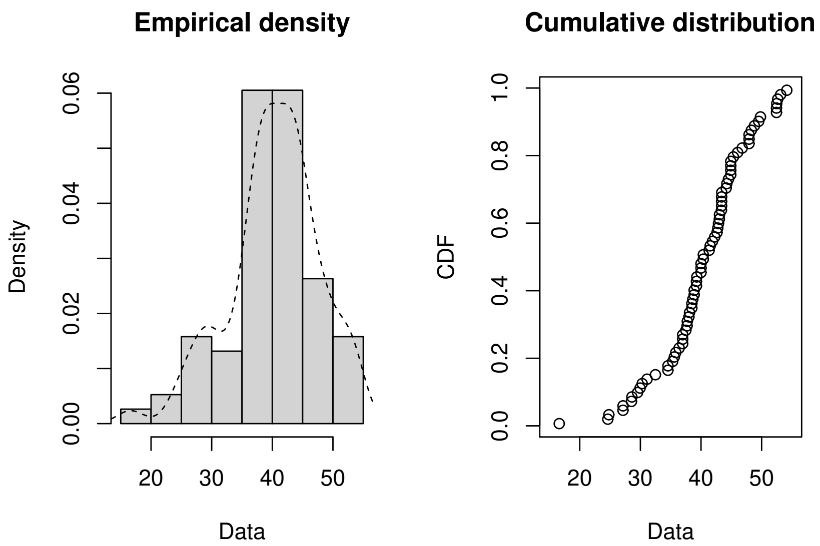

Figure 3.3: Histogram, polygon, and cdf of carbon fiber data

The fitting of the data set with multiple probability distributions is given in the table below:

	logL	AIC	BIC	Chisqv	Chisq(p)	AD(value)	KS(value)	H(KS)
Chen	**-85.28**	**166.58**	**161.81**			9.57	0.0282	
Normal	-259.65	523.31	527.97	12.08	0.06	0.60	0.09	not rejected
Cauchy	-269.2	542.41	547.07	9.69	0.14	0.96	0.10	not rejected
Logistic	-259.27	522.55	527.21	10.36	0.11	0.40	0.06	not rejected
Exponential	-356.89	715.79	718.12	299.91	0.00	23.56	0.45	rejected
Chi-square	-264.93	531.86	534.2	19.81	0.01	1.65	0.13	not rejected
Uniform	NULL	NULL	NULL	24.30	0.00	Inf	0.13	not rejected
Gamma	-263.76	531.52	536.18	16.07	0.01	1.20	0.12	not rejected
Lognormal	-266.79	537.59	542.25	18.88	0.00	1.64	0.14	not rejected
Weibull	-257.93	519.85	524.52	9.89	0.13	0.39	0.07	not rejected
Student	-531.6	1065.2	1067.53	3723.82	0.00	73.59	0.81	rejected

Table 3.4: Results of the fitting of the carbon fiber data

The Maximum Likelihood estimation of the fitted model is given by:

$\hat{\lambda}$	1.3555360
$\hat{\alpha}$	0.0219726

Table 3.5: MLE of the Chen distribution

We have generated a progressively censored sample using a random censoring scheme as follows, see table 3.3.

The statistical inference of the sample of fiber carbon, by the use of the MLE method, is established by the MaxLik r packages, and we obtain the estimator of the three parameters. Also by using the bootstrap technique, we establish the confidence bounds of three parameters, the results are summarized in the table 3.6 and they show that the estimations are close to the real values:

Parameters	Estimate value	Confidence bounds
$\hat{\lambda}$	1.491019	(1.2980075,1.5928101)
$\hat{\alpha}$	0.054053	(0.0201827,0.0630441)
$\hat{\beta}$	1.971302	(1.9722682,3.0294652)

Table 3.6: Paremeters and confidence bounds of the parameters

N	odata	cdata	cindicator	N	odata	cdata	cindicator
1	16.62	16.62	1	39	47.92	9.25	1
2	24.63	16.62	0	40	52.45	39.25	0
3	24.80	24.63	1	41	44.91	39.31	1
4	27.13	24.80	1	42	29.54	40.00	1
5	27.14	24.80	0	43	27.13	40.00	1
6	28.53	27.13	1	44	35.60	40.00	1
7	28.60	27.14	1	45	45.34	40.00	0
8	29.54	28.53	1	46	43.37	41.34	1
9	29.95	28.60	1	47	54.15	41.34	0
10	30.30	29.54	1	48	42.77	41.51	1
11	31.10	29.54	0	49	42.88	41.89	1
12	32.48	29.95	1	50	44.26	41.89	0
13	34.54	29.95	0	51	27.14	42.26	1
14	34.55	30.30	1	52	39.31	42.64	1
15	35.33	31.10	1	53	24.80	42.88	1
16	35.60	31.10	0	54	16.62	43.02	1
17	35.84	32.48	1	55	30.30	43.02	0
18	36.39	34.54	1	56	36.39	43.05	1
19	36.98	34.55	1	57	28.60	43.05	0
20	36.98	35.33	1	58	28.53	43.05	0
21	36.98	35.33	0	59	35.84	43.05	0
22	37.50	35.60	1	60	31.10	43.05	0
23	37.70	35.84	1	61	34.55	43.05	0
24	37.74	36.98	1	62	52.65	43.05	0
25	38.00	36.98	0	63	48.81	43.05	0
26	38.11	36.98	1	64	43.42	43.05	0
27	38.49	36.98	1	65	52.49	43.05	0
28	38.49	37.50	1	66	38.00	43.05	0
29	38.65	37.70	1	67	38.65	43.05	0
30	38.87	37.70	0	68	34.54	43.05	0
31	38.87	37.74	1	69	37.70	43.05	0
32	39.25	38.11	1	70	38.11	43.05	0
33	39.25	38.49	1	71	43.05	43.05	0
34	39.31	38.49	1	72	29.95	43.05	0
35	40.00	38.87	1	73	32.48	43.05	0
36	40.00	38.87	0	74	24.63	43.05	0
37	40.00	38.87	1	75	35.33	43.05	0
38	40.38	39.25	1	76	41.34	43.05	0

Table 3.7: Results of the scheme of progressively censoring of type II of fiber carbon sample

3.4 Real data application 2

In this section, a real data set taken from Lawless [62] is considered to illustrate the proposed methodology. The data comprise 50 observations, which represent the quantity of 1000s of cycles to failure for electrical appliances in a life test.

0.014 0.034 0.059 0.061 0.069 0.080 0.123 0.142 0.165 0.210 0.381 0.464 0.479 0.556 0.574 0.839 0.917 0.969 0.991 1.064 1.088 1.091 1.174 1.270 1.275 1.355 1.397 1.477 1.578 1.649 1.702 1.893 1.932 2.001 2.161 2.292 2.326 2.337 2.628 2.785 2.811 2.886 2.993 3.122 3.248 3.715 3.790 3.857 3.912 4.100.

The descriptive summary of the sample are given in the following table:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	sd
0.0140	0.4983	1.3150	1.5607	2.3342	4.1000	1.2176

Table 3.8: Summary of the dataset

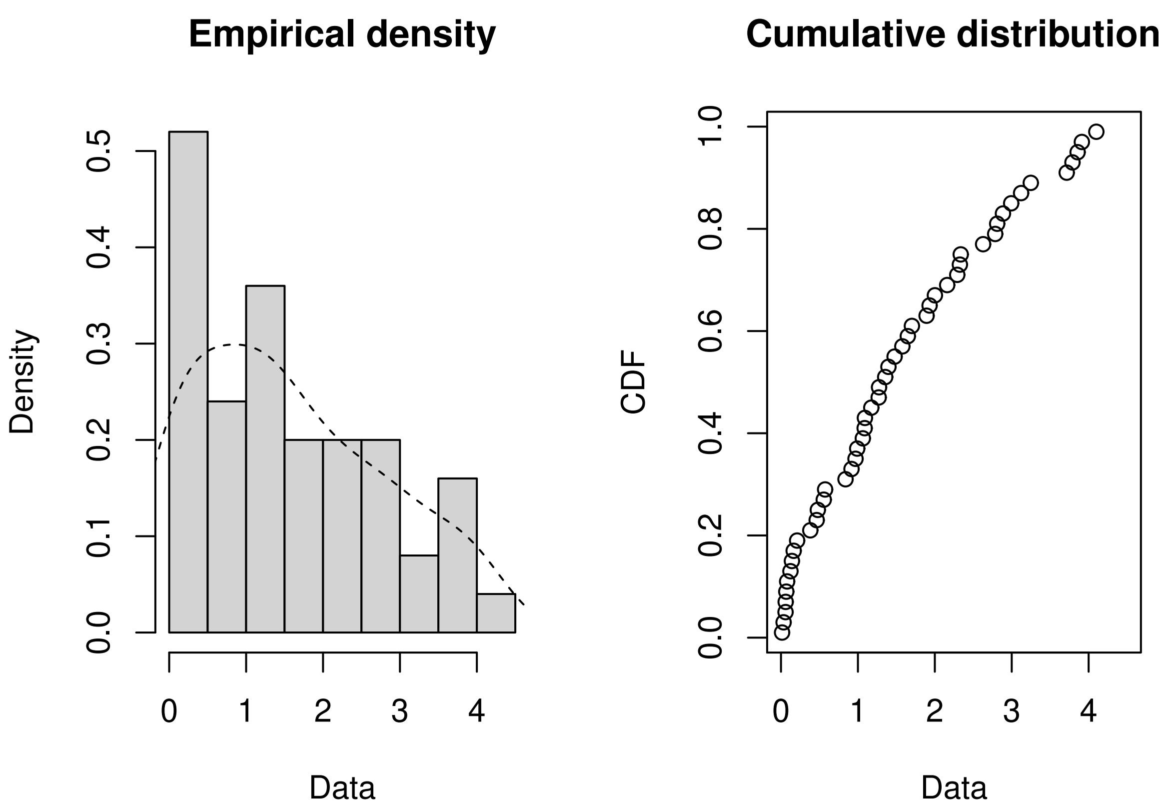

Figure 3.4: Histogram, polygon, and cdf of electrical appliances data

To assess the fit of the data on the selected model, we considered Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Kolmogorov-Smirnov (KS) statistic, Cramer-von Mises (CVM) statistics and Anderson-Darling (AD) statistics as the goodness of fit criteria. The Chen distribution is also compared to the Weibull, gamma, logistic, and exponential distributions, and it is noticed that among the mentioned distributions, Chen is a decent contender to fit this data set as the least value of all information criteria is observed in the case of this distribution (see Table 3.9):

	log-Lik	AIC	BIC	AD(value)	KS(value)	CVM(value)
Chen	**-68.2821**	**140.5643**	**144.3883**	**0.40945**	**0.08522**	**0.04293**
Logistic	-81.8383	167.6766	171.5007	0.88219	0.11436	0.10190
Exponential	-72.2573	146.5147	148.4267	1.07778	0.12431	0.17472
Gamma	-72.2563	148.5127	152.3368	1.07733	0.12564	0.17686
Weibull	-72.1150	148.2301	152.0541	1.08798	0.10539	0.14733

Table 3.9: Results of the fitting of electrical appliances data

The Maximum Likelihood estimator of the fitted model is given by (see Table 3.10)

$\hat{\lambda}$	0.6631230
$\hat{\alpha}$	0.2915201

Table 3.10: MLE of the Chen distribution

The statistical inference of the sample of electrical appliances, by the use of the MLE method, is established by the MaxLik r packages, and we obtain the estimator of the three parameters. Also by using the bootstrap technique, we establish the confidence bounds of three parameters, the results are summarized in the table 3.11, the estimations are approximately the same as real values which again show the efficiency of the proposed approach:

Parameters	Estimate value	Confidence bounds
$\hat{\lambda}$	0.527640	(0.506531,0.881235)
$\hat{\alpha}$	0.259439	(0.198425,0.376297)
$\hat{\beta}$	1.520202	(1.216698,1.791564)

Table 3.11: Paremeters and confidence bounds of the parameters

GENERAL CONCLUSION

In this Master's thesis, the focus was on addressing the challenges associated with step-stress accelerated life testing, particularly in the context of partially accelerated life testing using censoring data. The goal was to develop a robust and efficient model that can accurately estimate the parameters of interest and provide reliable measures of reliability.

The thesis introduced the concept of reliability systems and provided an overview of the fundamental principles of reliability analysis. It discussed various reliability measures, such as the System Mean Time to Failure (MTTF), Mean Time Between Failures (MTBF), failure rate function, maintainability, and availability. Additionally, the thesis explored common distribution functions used in reliability analysis, including the exponential, Weibull, binomial, Poisson, normal, log-normal, gamma, beta, Pareto, Lomax, and Rayleigh distributions.

To estimate the unknown parameters of the SSPALT model with type-II progressively censored data, the thesis employed the Maximum Likelihood Estimation (MLE) method. The MLEs were derived based on the assumption that the lifetimes of the experimental units follow the Chen distribution, which was described in detail. Confidence intervals for the parameter estimates were also calculated to assess their precision and reliability.

To evaluate the performance of the proposed model and estimation methods, a Monte Carlo simulation study was conducted. This simulation study involved generating synthetic data based on the SSPALT model and Chen distribution assumptions, and then applying the estimation techniques to estimate the parameters from the simulated data. The simulation results provided valuable insights into the accuracy and efficiency of the numerical methods employed.

Furthermore, the thesis included the analysis of real data sets to validate the proposed model and estimation methods. Model selection was performed using various criteria, including the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Kolmogorov-Smirnov test, Anderson-Darling test, and information criteria. The results indicated that the Chen distribution performed better than other competitive lifetime distributions, highlighting its suitability for modeling step-stress accelerated life testing with progressive censoring type-II data.

The findings of this thesis contribute to the field of reliability systems by providing a comprehensive framework for analyzing and modeling step-stress accelerated life testing with partially censored data. The developed SSPALT model, along with the estimation methods and model selection criteria, offer practical tools for reliability engineers and researchers to assess and enhance the reliability of systems. Future research endeavors may involve exploring alternative distributions and considering additional factors in the analysis of Partial Step Stress for progressively censored data to further advance the understanding and application of reliability as a quality measure.

BIBLIOGRAPHY

[1] Hoang Pham. *Software reliability*. Springer Science & Business Media, 2000.

[2] DJF Davis. An analysis of some failure data. *Journal of the American Statistical Association*, 47(258):113–150, 1952.

[3] Benjamin Epstein. The exponential distribution and its role in life testing. Technical report, Wayne State Univ Detroit MI, 1958.

[4] Brian A Maguire, Egon S Pearson, and AHA Wynn. The time intervals between industrial accidents. *Biometrika*, 39(1/2):168–180, 1952.

[5] Waloddi Weibull. A statistical distribution function of wide applicability. *Journal of applied mechanics*, 1951.

[6] Kenneth S Lomax. Business failures: Another example of the analysis of failure data. *Journal of the American Statistical Association*, 49(268):847–852, 1954.

[7] Amal S Hassan and Amani S Al-Ghamdi. Optimum step stress accelerated life testing for lomax distribution. *Journal of Applied Sciences Research*, 5(12):2153–2164, 2009.

[8] Gregory Campbell and Makarand V Ratnaparkhi. An application of lomax distributions in receiver operating characteristic (roc) curve analysis. *Communications in Statistics-Theory and Methods*, 22(6):1681–1687, 1993.

[9] N Balakrishnan and M Ahsanullah. Relations for single and product moments of record values from lomax distribution. *Sankhyā: The Indian Journal of Statistics, Series B*, pages 140–146, 1994.

[10] Maurice C Bryson. Heavy-tailed distributions: properties and tests. *Technometrics*, 16(1):61–68, 1974.

[11] Majid Chahkandi and Mojtaba Ganjali. On some lifetime distributions with decreasing failure rate. *Computational Statistics & Data Analysis*, 53(12):4433–4440, 2009.

[12] N Johnson, S Kotz, and N Balakrishnan. Equation 14.65 (from continuous univariate distributions, volume 1,). *Continuous Univariate Distributions*, 1, 1994.

[13] John Aldrich. Ra fisher and the making of maximum likelihood 1912-1922. *Statistical science*, 12(3):162–176, 1997.

[14] Lloyd S Nelson. Comments on significance tests and confidence intervals. *Journal of Quality Technology*, 22(4):328–330, 1990.

[15] Herman Chernoff. Optimal accelerated life designs for estimation. *Technometrics*, 4(3):381–408, 1962.

[16] S Bessler, Herman Chernoff, and Albert W Marshall. An optimal sequential accelerated life test. *Technometrics*, 4(3):367–379, 1962.

[17] PK Goel. Some estimation problems in the study of tampered random variables. pittsburgh (pa): Department of statistics, 1971.

[18] Morris H DeGroot and Prem K Goel. Bayesian estimation and optimal designs in partially accelerated life testing. *Naval research logistics quarterly*, 26(2):223–235, 1979.

[19] Gouri K Bhattacharyya and Zanzawi Soejoeti. A tampered failure rate model for step-stress accelerated life test. *Communications in statistics-Theory and methods*, 18(5):1627–1643, 1989.

[20] Do Sun Bai, MS Kim, and SH Lee. Optimum simple step-stress accelerated life tests with censoring. *IEEE transactions on reliability*, 38(5):528–532, 1989.

[21] Vilijandas Bagdonavičius. Testing hypothesis of the linear accumulation of damages. *Teoriya Veroyatnostei i ee Primeneniya*, 23(2):403–408, 1978.

[22] Reinhard Viertl. *Statistical methods in accelerated life testing*. Number 32. Vandenhoeck & Ruprecht, 1988.

[23] Wayne Nelson. Accelerated life testing-step-stress models and data analyses. *IEEE transactions on reliability*, 29(2):103–108, 1980.

[24] William Q Meeker, Luis A Escobar, and C Joseph Lu. Accelerated degradation tests: modeling and analysis. *Technometrics*, 40(2):89–99, 1998.

[25] CM Kim and Do-Sun Bai. Analyses of accelerated life test data under two failure modes. *International Journal of Reliability, Quality and Safety Engineering*, 9(02):111–125, 2002.

[26] Vilijandas Bagdonavicius and Mikhail Nikulin. *Accelerated life models: modeling and statistical analysis*. CRC press, 2001.

[27] Essam K Al-Hussaini and Alaa H Abdel-Hamid. Bayesian estimation of the parameters, reliability and hazard rate functions of mixtures under accelerated life tests. *Communications in Statistics-Simulation and Computation*, 33(4):963–982, 2004.

[28] Essam K Al-Hussaini and Alaa H Abdel-Hamid. Accelerated life tests under finite mixture models. *Journal of Statistical Computation and Simulation*, 76(8):673–690, 2006.

[29] WANG Rong Hua and He Liang FEI. Statistical inference of weibull distribution for tampered failure rate model in progressive stress accelerated life testing. *Journal of Systems Science and Complexity*, 14(2):237, 2004.

[30] A Clifford Cohen. Progressively censored samples in life testing. *Technometrics*, 5(3):327–339, 1963.

[31] Nancy R Mann. Best linear invariant estimation for weibull parameters under progressive censoring. *Technometrics*, 13(3):521–533, 1971.

[32] N Balakrishnan and J Varadan. Approximate mles for the location and scale parameters of the extreme value distribution with censoring. *IEEE Transactions on Reliability*, 40(2):146–151, 1991.

[33] Narayanaswamy Balakrishnan, Narayanaswamy Balakrishnan, and Rita Aggarwala. *Progressive censoring: theory, methods, and applications*. Springer Science & Business Media, 2000.

[34] Uditha Balasooriya and C-K Low. Competing causes of failure and reliability tests for weibull lifetimes under type i progressive censoring. *IEEE Transactions on Reliability*, 53(1):29–36, 2004.

[35] N Balakrishnan and Erhard Cramer. The art of progressive censoring. *Statistics for industry and technology*, 2014.

[36] Siyi Chen and Wenhao Gui. Statistical analysis of a lifetime distribution with a bathtub-shaped failure rate function under adaptive progressive type-ii censoring. *Mathematics*, 8(5):670, 2020.

[37] N Balakrishnan and RA Sandhu. Best linear unbiased and maximum likelihood estimation for exponential distributions under general progressive type-ii censored samples. *Sankhyā: The Indian Journal of Statistics, Series B*, pages 1–9, 1996.

[38] DR Barot and MN Patel. Bayesian estimation of reliability indexes for cold standby system under general progressive type ii censored data. *International Journal of Quality & Reliability Management*, 2014.

[39] Yi Min Shi and Xiao Lin Shi. Reliability estimation for cold standby series system based on general progressive type ii censored samples. In *Applied Mechanics and Materials*, volume 321, pages 2460–2463. Trans Tech Publ, 2013.

[40] N Balakrishnan and Chien-Tai Lin. Exact linear inference and prediction for exponential distributions based on general progressively type-ii censored samples. *Journal of Statistical Computation and Simulation*, 72(8):677–686, 2002.

[41] N Balakrishnan and Jie Mi. Existence and uniqueness of the mles for normal distribution based on general progressively type-ii censored samples. *Statistics & Probability Letters*, 64(4):407–414, 2003.

[42] Yanbin Ma and Wenhao Gui. Point estimation and two new goodness of fit tests for the scale family based on general progressively type-ii censored samples. *Communications in Statistics-Simulation and Computation*, 50(12):4187–4214, 2021.

[43] Zhenmin Chen. A new two-parameter lifetime distribution with bathtub shape or increasing failure rate function. *Statistics & Probability Letters*, 49(2):155–161, 2000.

[44] Essam A Ahmed, Ziyad Ali Alhussain, Mukhtar M Salah, Hanan Haj Ahmed, and MS Eliwa. Inference of progressively type-ii censored competing risks data from chen distribution with an application. *Journal of Applied Statistics*, 47(13-15):2492–2524, 2020.

[45] Anthony Christopher Davison and David Victor Hinkley. *Bootstrap methods and their application*. Number 1. Cambridge university press, 1997.

[46] Bradley Efron. *The jackknife, the bootstrap and other resampling plans*. SIAM, 1982.

[47] N Balakrishnan and RA Sandhu. A simple simulational algorithm for generating progressive type-ii censored samples. *The American Statistician*, 49(2):229–230, 1995.

[48] R Core Team. R: A language and environment for statistical computing. vienna, austria. *R Foundation for Statistical Computing*, 2014.

[49] William N Venables, Brian D Ripley, K Hornik, and A Gebhardt. Package mass. *Online at: http://cran. r-project. org/web/packages/MASS/MASS. pdf*, 2010.

[50] F Ricci and G Palleschi. Sensor and biosensor preparation, optimisation and applications of prussian blue modified electrodes. *Biosensors and Bioelectronics*, 21(3):389–407, 2005.

[51] Marie Laure Delignette-Muller, Christophe Dutang, Regis Pouillot, Jean-Baptiste Denis, and Maintainer Marie Laure Delignette-Muller. Package 'fitdistrplus'. *Journal of Statistical Software*, 64(4):1–34, 2015.

[52] Christophe Dutang, Vincent Goulet, and Mathieu Pigeon. actuar: An r package for actuarial science. *Journal of Statistical software*, 25:1–37, 2008.

[53] Matthias Kohl, Peter Ruckdeschel, and Helmut Rieder. Infinitesimally robust estimation in general smoothly parametrized models. *Statistical Methods & Applications*, 19:333–354, 2010.

[54] John P Klein and Melvin L Moeschberger. *Survival analysis: techniques for censored and truncated data*, volume 1230. Springer, 2003.

[55] Pieter Busschaert, AH Geeraerd, Mieke Uyttendaele, and JF Van Impe. Estimating distributions out of qualitative and (semi) quantitative microbiological contamination data for use in risk assessment. *International journal of food microbiology*, 138(3):260–269, 2010.

[56] Andreas Leha, Tim Beißbarth, and Klaus Jung. Sequential interim analyses of survival data in dna microarray experiments. *BMC bioinformatics*, 12(1):1–14, 2011.

[57] Natalie Commeau, Eric Parent, Marie-Laure Delignette-Muller, and Marie Cornu. Fitting a lognormal distribution to enumeration and absence/presence data. *International journal of food microbiology*, 155(3):146–152, 2012.

[58] Terry Therneau et al. A package for survival analysis in s. *R package version*, 2(7), 2015.

[59] Susan S Hirano, Murray K Clayton, and Christen D Upper. Estimation of and temporal changes in means and variances of populations of pseudomonas syringae on snap bean leaflets. *Phytopathology*, 84(9):934–939, 1994.

[60] MG Bader and AM Priest. Statistical aspects of fibre and bundle strength in hybrid composites. *Progress in science and engineering of composites*, pages 1129–1136, 1982.

[61] ME Ghitany, Dhaifalla K Al-Mutairi, N Balakrishnan, and LJ Al-Enezi. Power lindley distribution and associated inference. *Computational Statistics & Data Analysis*, 64:20–33, 2013.

[62] Jerald F Lawless. *Statistical models and methods for lifetime data*. John Wiley & Sons, 2011.