Deanship of Graduate Studies

Al-Quds University

Generalized Log-Logistic Proportional Hazard Model with Applications in Survival Analysis

Afnan Mahmoud Mustafa AL-sheikh

M.Sc. Thesis

Jerusalem-Palestine

1444/2022

Generalized Log-Logistic Proportional Hazard Model with Applications in Survival Analysis

Prepared by:

Afnan Mahmoud Mustafa AL-sheikh

B.Sc.: Mathematics from-Al-Quds University/ Palestine

Supervisor: Dr. Khalid Salah

A thesis submitted in partial fulfillment of requirements for the degree of Master of Mathematics, Department of Mathematics / Graduate Studies / Al-Quds University.

1444\2022

Al-Quds University

Deanship of Graduate Studies

Mathematics Program

Thesis Approval

Generalized Log-Logistic Proportional Hazard Model with Applications in Survival Analysis

Prepared By: Afnan Mahmoud Mustafa AL-sheikh

Registration No: 21720067

Supervisor: Dr. Khalid Salah

Master thesis submitted and accepted, Date: 10/8/2022

1- Head of Committee: Dr. Khalid Salah Signature:............

2-Internal Examiner : Dr. Jamil Jamal Abu Ismail Signature:............

3- External Examiner: Dr.Mohammad Dawabsha Signature:............

Jerusalem – Palestine

1444/ 2022

Dedication

To my family, my husband and my children, to my mother and father, my brothers and sisters, my instructor, my supervisor, my respected doctors, my colleagues at Al-Quds University, and all math lovers.

Afnan Mahmoud

Declaration

I certify that this submitted for the degree of a master is the result of my own research, except where otherwise acknowledged. And that this (or any part of the same) has not been submitted to a higher degree to any other university or institution.

Signature: Afnan M…………………..

Student's name: Afnan Mahmoud Mustafa AL-sheikh

Date: 10/8/2022

Acknowledgement

First of all, I thank God for reaching this stage. Peace and blessing upon our first teacher and educator, Prophet Mohammed, who has taught the whole world.

Big thanks to my supervisor Dr. Khalid Salah for all his support, guidance, and encouragement to finish this thesis despite of the difficult circumstances we live in. I also extend my sincere thanks to the examiners who contributed in improving the thesis through their constructive comments and advice.

I thank the President of the University, Dr. Imad Abu Kishek, for giving me all the support.

I thank my colleagues at Al-Quds University, especially Ramzi Jaffar and Layla Lahaleeh for their continued support always.

Abstract

Proportional hazard (PH) models can be formulated with or without assuming a probability distribution for survival times. The former assumption leads to parametric models, whereas the latter leads to the semi-parametric Cox model which is by far the most popular in survival analysis. However, a parametric model may lead to more efficient estimates than the Cox model under certain conditions. Only a few parametric models are closed under the PH assumption, the most common of which is the Weibull that accommodates only monotone hazard functions. We study and investigate a generalization of the log-logistic distribution that belongs to the PH family. It has properties similar to those of log-logistic, and approaches the Weibull in the limit. These features enable it to handle both monotone and nonmonotone hazard functions. Application to four data sets and a simulation study revealed that the model could potentially be very useful in adequately describing different types of time-to-event data.

نموذج الخطر النسبي اللوجستي واللوجستي المعمم و تطبيقاته في تحليل البقاء

اعداد: افنان محمود مصطفى الشيخ

إشراف: د. خالد صلاح

الملخص

يمكن صياغة نماذج المخاطر النسبيةProportional Hazard مع أو بدون افتراض توزيع احتمالي لأوقات البقاء Survival Time. يؤدي الافتراض الأول إلى نماذج بارامترية Parametric، في حين أن الأخير يؤدي إلى نموذج كوكس Cox شبه البارامتي الذي يعد الأكثر شيوعًا في تحليل البقاء على قيد الحياة. ومع ذلك، قد يؤدي النموذج المعياري إلى تقديرات أكثر كفاءة من نموذج كوكس في ظل ظروف معينة. عدد قليل فقط من النماذج البارامترية في ظل افتراض PH تكون مغلقة، وأكثرها شيوعًا هو Weibull الذي يستوعب اقترانات الخطر أحادية الاتجاه فقط في هذا البحث سوف ندرس ونتوسع في تعميم التوزيع اللوجيستي الذي ينتمي إلى عائلة PH. لها خصائص مشابهة لتلك الخاصة باللوجستيات ، وتقترب من Weibull في النهاية. هذه الميزات تمكنه من التعامل مع اقترانات الخطر monotone and nonmonotone hazard function. تم تطبيق هذا النموذج على مجموعات مختلفة من البيانات وكذلك دراسة محاكاة، تبين أن النموذج يمكن أن يكون مفيداً جداً في وصف الأنواع المختلفة من بيانات الوقت إلى الحدث time-to-event بشكل مناسب.

Table of Contents

List of Tables

List of Figures

Figure	Page
Figure 5.6. The above figures are the time series convergence diagrams of the estimated parameters κ, γ, and ρ. The figures below is the kernel density estimation of the same parameters for the **Good** data set	57
Figure 5.7. The above figures are the time series convergence diagrams of the estimated parameters κ, γ, and ρ. The figures below is the kernel density estimation of the same parameters for the **Medium** data set.	57
Figure 5.8. The above figures are the time series convergence diagrams of the estimated parameters κ, γ, and ρ. The figures below is the kernel density estimation of the same parameters for the **Poor** data set.	57

Chapter One

Survival Analysis

1.1 Introduction

Survival analysis is a statically method for data analysis where the outcome variable of interest is the time to the occurrence of an event, it deals with the analysis of lifetime data.In this analysis, our objective is to model the survival time, i.e. the time to the occurrence of a given event. The event could be medical field ,common examples are the time to development of a disease, response to treatment, and death. The time to event in this case is defined as the time till death.

When we study survival analysis some quantities of interest should be known, the first one, the survival function, which provides the probability of survival at a given time, and the second quantity is the rate at which a person who is an event at a given point in time will instantaneously experience the event which called hazard function.

Survival data where a set of individuals are observed and the failure time or lifetime of that individual is recorded such as the time until a patient is cured.

Survival analysis estimates duration by computing the survival function. There are several different ways to estimate a survival function such that Bayesian inference.

Recent research that has sought to identify malignant treatments such as cancer has shown that it is important to monitor cured patients, with particular attention to the survival rate.

Relative risk models are a category of survival models in this model where the effect on the common variable is multiplied by the risk rate, for example taking a drug that reduces the risk of a stroke by half. To estimate the healing rate and survival curve of treated patients, some authors provided border, semi-parametric and non-parameter methods, including logarithmic, exponential, and Weibull distributions.

We aim in this thesis to offer a simple extension of the log-logistic model which is closed under the Proportional Hazards relationship. The proposed generalized log-logistic model is a three parameter distribution and has characteristics similar to those of the log-logistic model. Moreover, it approaches the Weibull is the limit. These features enable it to satisfactorily handle both monotone and non-monotone hazard functions.

Survival data where a set of individuals are observed and the failure time or lifetime of that individual is recorded such as the time until a patient is cured.

Survival analysis estimates duration by computing the survival function. There are several different ways to estimate a survival function such that Bayesian inference

1. Bayesian Inference: In Bayesian Inference, the parameter of interest is always considered to be a random variable with a prior distribution. To make inferences about the population parameters, Adaptive Rejection Metropolis Sampling (ARMS) and Gibbs sampling techniques are used, see Salah (2019). These methods are kind of Markov chain Monte Carlo (MCMC) technique used to draw dependent samples from complex high-dimensional distributions. However, the joint posterior distribution of the parameters in the proposed model is very complicated. Using Open BUGS software can greatly simplify the process of simulating these samples, and we only need to specify the data distribution and prior distributions for the model parameters. Prafulla, et. al. (2016).

2. Non- Bayesian Inference (Maximum Likelihood Approach): It provides a consistent approach to parameter estimation problems. This means that maximum likelihood

estimates can be developed for a large variety of estimation situations. Also, it has desirable mathematical and optimality properties. The disadvantages of this method are: The likelihood equations need to be specifically worked out for a given distribution and estimation problem, the numerical estimation is usually non-trivial, and it can be heavily biased for small samples. The optimality properties may not apply to small samples, and it is sensitive to the choice of starting values.

1.2 Basic Concepts

T is a continuous random variable with probability density function (Pdf) $f(t)$ and cumulative distribution function (cdf)

$$F(t) = pr\{T < t\} \tag{1.1}$$

giving the probability that the event has occurred by duration t. Where

$$f(t) = \frac{dF(t)}{dt} \tag{1.2}$$

Survival function which gives the probability that the event of interest has not occurred by

duration t.

$$\begin{aligned} S(t) &= \Pr\{T \geq t\} = 1 - F(t) \\ &= \int_t^{\infty} f(x)dx \end{aligned} \tag{1.3}$$

The survival function gives the probability that a subject will survive past time *t*, where $t \geq 0$.The survival function has properties
It is non – increasing, $S(0) = 1$ and as $t \longrightarrow \infty$, $S(t) \longrightarrow 0$.

All the above properties are shown in Figure 1.1

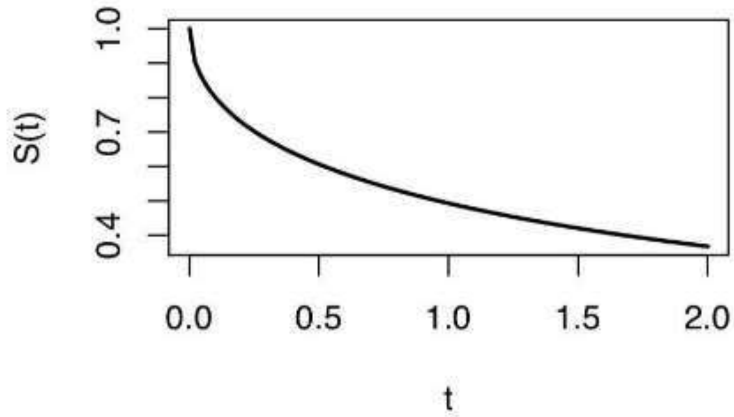


Figure 1.1. The Survival Function

Different kinds of proportional hazard models may be obtained by making different assumptions about the baseline survival function, or equivalently, the baseline hazard function.

Example 1.1.1 The exponential distribution, with density function

$$f(t) = \frac{1}{\theta} e^{(-x/\theta)} \tag{1.4}$$

has a survival function of

$$S(t) = e^{(-x/\theta)} \tag{1.5}$$

Example 1.1.2 The Weibull distribution which has survival function

$$S(t) = e^{(-x/\theta)\alpha} \tag{1.6}$$

Example 1.1.3 The Log logistic distribution with density function

$$f(x) = \frac{\beta^{\kappa}\kappa\, x^{\kappa-1}}{(\beta^{\kappa} + x^{\kappa})^2}, \quad x \in (0, \infty) \tag{1.7}$$

The Hazard function:

An alternative characterization of the distribution of T is given by the hazard function, or instantaneous rate of occurrence of the event (Collett,D 2003) defined as

$$\lambda(t) = \lim_{dt \to 0} \frac{P(t \le T < t + dt \mid T \ge t)}{dt}$$
$$= -\frac{d}{dt} \log S(t) \qquad (1.8)$$

The cumulative Hazard function is defined by

$$H(t) = \int_{-\infty}^{t} \lambda(x)\, dx$$
$$= -\log S(t) \qquad (1.8)$$

1.3 Literature Review

The Log logistic distribution is among the class of survival time parametric models where the hazard rate initially increases and then decreases and at times can be hump-shaped. Brain (1974) modeled the Log Logistic distribution using a transformation of a well-Known logistic variate, according to the literature on the field. Ragab and Green (1984) who also worked on the order statistics for the given distribution discussed the properties of the Log logistic distribution. The sample plane proposed by Kantam et al (2001) was based on the Log logistic distribution. The modified maximum likelihood estimation (MLE) of this distribution was developed by kantam and Rao (2002). The latest research aims to model survival data using the LL distribution and drive MLE using the Bayes estimates related probability intervals. Under the assumption of independent uniform priors for parameters, the Bayesian estimates may not be obtained straightforwardly.

The authors will proceed with the assumption that the LL model's shape and scale are unknown.

Using the open Bugs program, the authors will design a method to generate Markov chain Monte Carto(MCMC) samples based on posterior samples generated from the samples present research discussed the Log logistic model with two parameters, MLEs, and Bayesian estimates are obtained from a real life sample using the Markov chain Monte Carlo(MCMC) technique using open BUGS software, Bayesian analysis under a different set of priors has been carried out and convergence pattern was studied using different diagnostics procedures. visual reviews are witnessed that the Log Logistic model whether used with MLES or with Bayesian Estimates fits the data well. A key assumption in the PH models is that the hazard ratio comparing any two speciating of covariates is constant over time, (Klein Baum and Klein 2012) put the proportional hazard assumption may be. Handled using time-dependent covariates. The cox model is the most popular in survival analysis. (Kalbfleisch and prentice 2002, Lawless2002). They saw fit distributional assumption is required for a fully parametric model proportional hazard.

Efron (1977) and Oakes (1977), see that parametric models lead to more efficient estimates than Cox's model under certain conditions. The Cox proportional hazard in joint modeling of time to the event and longitudinal data (Wulfsohn and Tsiatis 1997) leads to underestimation of the standard errors of the parameter estimates (Shieh et al.2006; Rizopoulos 2012). The parametric time-to-event models are the Weibull and the Log logistic is useful in modeling non-monotone hazard rates (Lawless 2002) the parametric like Weibull that Monotone hazard functions (Kalbfleisch and prentice 2002).

Madhukar et al (1996) proposed a generalization of the Weibull distribution.
Based on the results in the (international journal of Tomography and statistics 2021) research, Gamma and LOG-normal generalized distributions, as well as other models, are better suited to the breast cancer data set because these distributions are attracted to heavy peripheral data. In population-based cancer studies, treatment occurs when the mortality rate (risk) in the group of infected individuals returns to the same level of the population as expected, so treatment must study the survival of the cancer patient.

1.4. Objective

In this thesis, we study, investigate, and examine the extent to which the log-logistic model closes under the proportional hazards relationship, approaching the Weibull limit. These advantages make both monotonic and non-monotonic hazard functions acceptable. We investigate the generalized log-logistic model and tested the parameters estimation using Bayesian and maximum likelihood approaches. The aim of this work is to provide a simple extension and application of the closed log-logistic model under the proportional hazards relationship. More specifically, the model we study is a generalized log-logistic model with a three-parameter distribution with properties similar to the log-logistic model. Weibull is also approaching his limits.

Chapter Two

Survival Analysis and Regression Models

2.1 Introduction

Survival analysis is a statically method for data analysis where the outcome variable of interest is the time to the occurrence of an event, often referred to as a failure time. It is used in different fields such that: medicine and engineering, and in engineering time till failure in mechanical systems is studied, for example, in medical studies the effect of a drug on the lifetime of a patient with a certain disease can be studied. The time to event in this case is defined as time till death.

In the recent year's survival has been introduced into credit scoring it is the area of statistics that deals with the analysis of lifetime data. The variable of interest is the time to event. In the case of credit risk, the event of interest is default. The major advantage of survival analysis compared to other credit scoring models, is that the model is capable of including censored and truncated data in the development sample. In the current logistic regression approaches these observations are removed from the dataset. Right censoring is the most common type of censoring and states that the event is not observed within the study period. In the case of credit risk, a customer who doesn't default, because most of the customers do not default a lot of the data is right censored.

2.2 Censoring

In survival data there is important part, which is worrying constraint an inconvenience in the analysis if not adequately controlled. Also, the presence of some hidden observation in the survival data cannot be ignored or neglected. In the other word, the presence of individual who do not know when the event occurred and cannot be tracked over a period of time. There are different types of censoring, but the most common types are:

Type Ⅰ censoring. In this type, the duration of the study is constant, while the number of events (i; e) the number of individuals for whom the event occurred) it is a random variable, and this type of censoring is called constant censoring in which the time of the study cessation is determined after a specific period of time.

It is one the following types

Right Censoring. the most common case in survival data, and this case is related to vocabulary that does not some vocabulary remains in life at the end of the study, i;e the time of survival of this singular above the end point of the study, and this vocabulary denotes it censoring right, the reasons for the occurrence of 111the right censoring: Some individuals did not have an event., When the researcher decides to end the study before the event occurs .

Left Censoring To illustrate this concept, we assume that we have individuals who entered the study, but the time of exposure to the risk is unknown, while the time of the occurrence of the event is known only, for example cancer patients and AIDS patients, the time of the onset of the disease is unknown, but the time of death due to that disease is known and this word is called the left control. Survival methods are applicable when the measure of interest is time to an event such as mortality or occurrence of disease. The concept of censoring makes survival methods unique.

Interval censoring. In this case the exact time of occurrence of the event is unknown to some individual, the period of time in which the event occurred, and it is said about this vocabulary that it has censored.

Type Ⅱ censoring:

In this type, the number of individuals for which the event is known (constant)in advanced while period, the study is a random variable that cannot be known in advance. And in it, the end time of study is determined, after a certain number of cases the event occurs.

2.3 Basic Survival Function

2.3.1 The Survival Function

Let T be a non-negative random variable with probability density function $(p.d.f)\ s(t)$ and cumulative distribution function $(c.d.f)\ S(t) = p\{T < t\}$ representing the waiting time until the occurrence of an event, we consider survival analysis referring to the event of interest as death and to the waiting time, but the techniques to be studied have much wider applicability. They can be used for example to see the duration of stay in a city (or in a job) and the length of life. Survival function defined by:

$$R(t) = P\{T \geq t\} = 1 - S(t) \tag{2.1}$$

$$= \int_t^\infty s(x)dx \qquad , t \geq 0 \tag{2.2}$$

While we note that about R(t)

1- $R(t) = 1 \quad if \ \ t < 0$

2- $R(\infty) = \lim_{t \to \infty} R(t) = 0$

3- $R(t)$ is non-increasing in t.

If there are no censored observations(t) is estimated by the proportion of patients surviving longer than t . time

$$\hat{R}(t) = \frac{\#\, patients\ surviving\ longer\ than\ t}{total\ \#\ of\ patients} \tag{2.3}$$

Now let T be a discrete random variable that takes the values $\{t_i\}$ where $0 \leq t_1 < t_2 < \cdots$ with probabilities

$$s(t_j) = p\{T = t_j\} \tag{2.4}$$

And we define the survivor function at time t_j as the probability that the survival time T is at least t_j

$$\mathrm{R}(\mathrm{t_j}) = \mathrm{R_j} = \mathrm{P}\{\mathrm{T} \geq \mathrm{t_j}\} = \sum_{\mathrm{k=j}}^{\infty} \mathrm{s_j} \tag{2.5}$$

2. The Hazard Function:

An alternative characterization of the distribution of T is given by the hazard function, or instantaneous rate of occurrence of the event, defined as

$$Z(t) = \lim_{dt \to 0} \frac{P(t \leq T < t + dt \mid T \geq t)}{dt} \tag{2.6}$$

This is the limit of the conditional probability that the event will occur in the interval $[t, t + dt]$ given that it has not occurred before per the width of the interval. In words, the rate of occurrence of the event at duration t equals the density of events at t divided by the probability of surviving to that duration without experiencing the event

3. The Cumulative Hazard Function

The cumulative risk function is defined as the sum of risk that have occurred up to time t.

$$H(x) = \int_{-\infty}^{x} Z(t)dt \tag{2.7}$$

For discrete data, the Hazard Function is

$$Z_i = p(t_{i-1} < T \leq t_i/T > t_{i-1}) = 1 - \frac{s_i}{s_{i-1}} \tag{2.8}$$

the survival in terms of Hazard Function

$$\mathrm{s_i} = \prod_{j=2}^{i} \frac{\mathrm{s_j}}{\mathrm{s_{j-1}}} = \prod_{j=2}^{i}(1 - \mathrm{Z_j}) = \prod_{j=1}^{i}(1 - \mathrm{Z_j}) \tag{2.9}$$

The cumulative Hazard function is

$$H_i = \sum_{j=1}^{i} h_i \tag{2.10}$$

2.3.4. **Relationships Between Survival Functions**

The former (2.5) may be written as $s(t)dt$ for small dt, while the latter is $R(t)$ by definition.

Dividing by dt and passing to the limit gives the useful result

$$Z(t) = \frac{f(t)}{R(t)} \tag{2.11}$$

since the density function is defined as the derivative of the cumulative distribution function, we get

$$s(t) = \frac{d}{dt}[1 - R(t)] = -R'(t) \tag{2.12}$$

Inserting (2.11) into (2.12), we got

$$Z(t) = \frac{R'(t)}{R(t)} = \frac{d}{dt} ln\, R(t) \qquad (2.13)$$

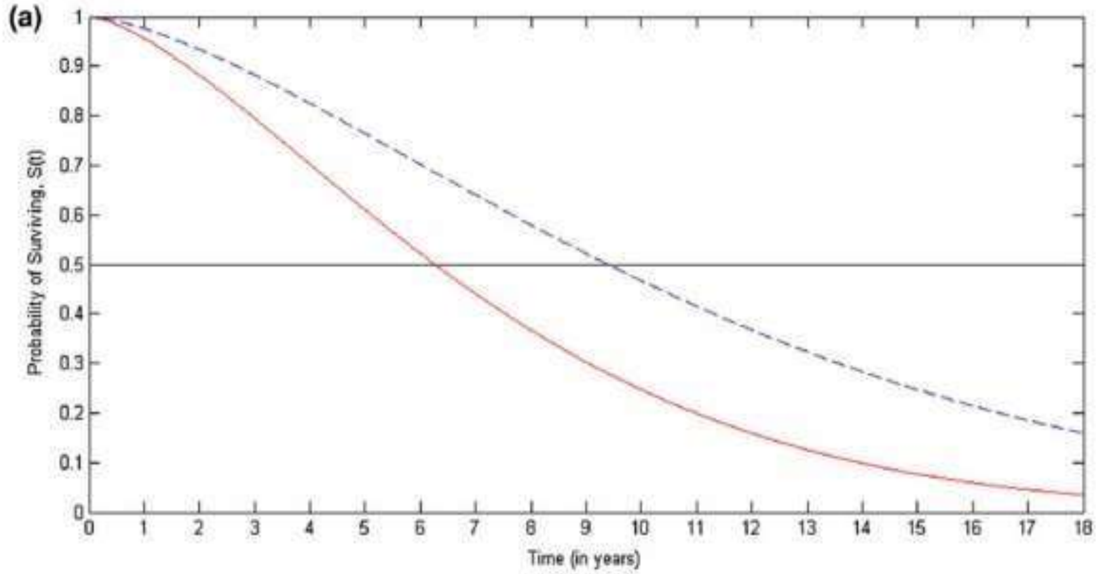
Referring to (2.7) and use (2.11) we get

$$\begin{aligned} H(t) &= \int_0^t \frac{s(x)}{R(x)} dx \\ &= \int_0^t \frac{1}{R(x)} \left(\frac{d}{dx} R(x) \right) dx \qquad (2.14) \\ &= -\ln(R(t) \end{aligned}$$

we can solve the above expression (2.8) to obtain a formula for the probability of surviving to duration *t* as a function of the hazard at all durations up to *t*:

$$R(t) = Exp\left(- \int_0^t Z(x)\, dx \right) \qquad (2.15)$$

Suppose that death is the event of interest, and time is measured in years. For example, in Figure 2.1.a shows two functions for two individuals(group), the first group represented by the red line, which has a higher risk of death than the second group, which is represented by the blue dashed line because the survival curve decreases faster than the other group.



Figuer 2.1.a. Plots of Survival curves.

Figure 2.2.b shows that the risk function relates to how quickly the survival function over time

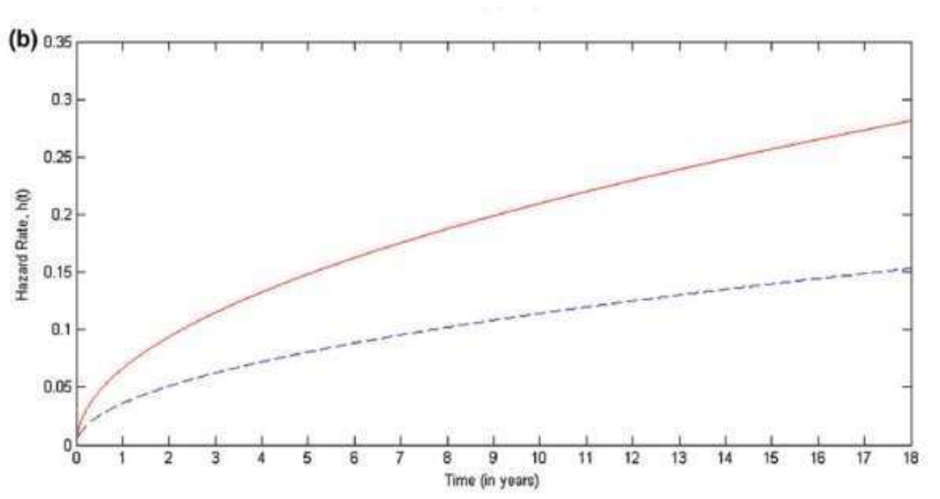


Figure 2.2.b. Plots of Hazard curves corresponding to survival function

2.4 Types of Survival

The choice of which approach to use should be driven by the research question of interest. Often, more than one approach can be appropriately utilized in the same analysis. In survival analysis, we have three options for modeling the survival function: Non-parametric like Kaplan-Meir, Semi parametric like cox regression, and parametric (such as the Weibull distribution).

2.4.1. Non-parametric

Non-parametric approaches do not rely on assumptions about the shape or form of parameters in the underlying population, where are used to describe the data by estimating the survival function R(t). These descriptive statistics cannot be calculated directly from the data due to censoring, which underestimates the true survival time in censored subjects, leading to skewed estimates of the mean, median and other descriptive. Nonparametric approaches are often used as the first step in an analysis to generate unbiased descriptive statistics, and are often used in conjunction with semi-parametric or parametric approaches.

The Non-parametric Kaplan-Meier Estimate

In case of censored data, raw empirical estimators will not produce good results. In order to determine distribution function of these data, one of basic technique can be applied Kaplan Meier (KM), The Kaplan-Meier estimator works as a tool to estimate the number of patients who may have survived treatment. It is a non-parametric estimation which is commonly used to describe a community's survival and to compare two sample population groups and is the best statistics to make predictions in a sample to measure the survival chances of patients living after treatment for a certain period of time

The Kaplan-Meier estimator works by dividing the estimation of S(t) into a series of steps (intervals) based on the timing of the events observed. Observations contribute to the estimation of S(t) until such time as the event occurs or is censored. For each period, the probability of survival is calculated at the end of the period, given that people are at risk at the beginning of the period.

Let n_j be the number of individuals at risk (uncensored and alive) just before t_i. And let d_j be the number of observed deaths at t_i

This is commonly referred to as $p_j = (\frac{n_{j-d_j}}{n_j})$.The Kaplan-Meier estimator is obtained by calculating the probability of an event occurring at a given time, and then the successive probabilities are multiplied by any previous calculated probabilities to determine the final estimate. The main assumptions of this method are that censorship occurs after failure and that there is no collective effect on survival, so that people are equally likely to survive regardless of when they study.

Suppose r individuals experience events in a group of individuals. Let the observed event times be given by $0 \leq t_{(1)} \leq t_{(2)} \leq t_{(3)} \leq \cdots \leq t_{(r)} < \infty$.

The KM estimator of the survival function $s(t)$ is defined (Efron, B.(1977) by

$$KM(t) = \hat{S}\,(t) = \prod_{s<t} (1 - \frac{d_j}{n_j}) \tag{2.16}$$

Where d_j is the number of events at time x, generally either zero or one, but in case of tied survival times $d_j \geq 1$,n_j is the number of items at risk at time x. In the figure 2.3 a KM survival estimator is illustrated. The ‘x’ on the time axes determine an event and the ‘o’ determines a censored data.

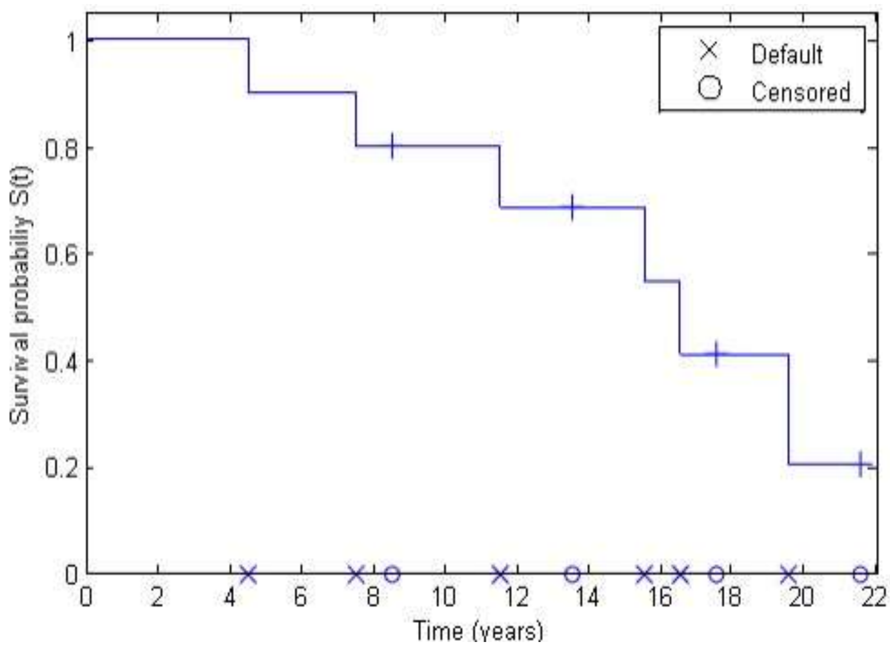


Figure 2.3. Kaplan-Meier survival estimator.

2.4.2 Parametric.

In parametric approaches, both the hazard function and the effect of the covariates are specified. The hazard function is estimated based on an assumed distribution in the underlying population. In this model the researchers assume completely the form of the model and its assumptions, the most commonly used distributions are: the exponential distribution and the Weibull distribution, In order to estimate R(t), maximum likelihood estimation is used. parametric form is the most difficult part of parametric survival analysis. The specification of the parametric form should be driven by the study hypothesis, along with prior knowledge and biologic plausibility of the shape of the baseline hazard. For example, if it is known that the risk of death increases dramatically right after surgery and then decreases and flattens out, it would be inappropriate to specify the exponential distribution, which assumes a constant hazard over time. The data can be used to assess whether the specified form appears to fit the data, but these data-driven methods should complement, not replace, hypothesis-driven selections.

The exponential and Weibull Models:

Different kinds of proportional hazard models may be obtained by making different assumptions about the base line survival function, or equivalently, the baseline hazard function. For example, if the baseline risk is constant, over time, so $\lambda_0(t) = \lambda_0$, say we obtain the exponential regression model, where

$$\lambda_i(t, x_i) = \lambda_0 \exp\{x_i' \beta\} \qquad (2.17)$$

For example, consider the one-parameter exponential distribution with density function

$$s(t) = \begin{cases} \lambda e^{-\lambda t} & t \geq 0, \lambda > 0 \\ 0 & t < 0 \end{cases} \qquad (2.18)$$

and the survivor function

$$R(t) = e^{-\lambda t} \qquad t \geq 0 \qquad (2.19)$$

also the hazard function

$$Z(t) = \lambda, \qquad t \geq 0 \qquad (2.20)$$

The Weibull distribution is similar to the exponential distribution. While the exponential distribution assumes a constant hazard, the Weibull distribution assumes a monotonic hazard that can either be increasing or decreasing but not both. Which has survival function

For parameter $\theta > 0$ and $\alpha > 0$.

$$R(t) = \exp\left(-\left(x/\theta\right)\right)^{\alpha} \tag{2.21}$$

And hazard function

$$Z(t) = (\alpha/\theta)\left(\left(x/\theta\right)^{\alpha-1}\right. \tag{2.22}$$

In Figure 2.4 two models are fitted to the survival data: KM estimator, exponential and Weibull.

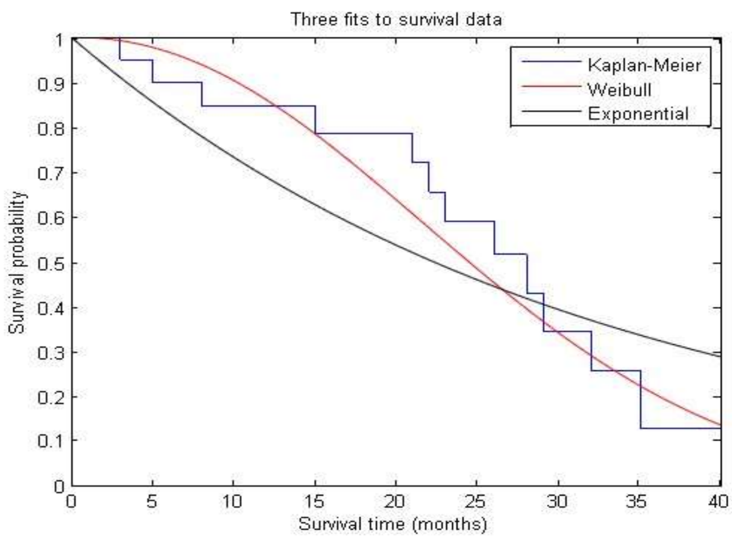


Figure 2.4 plot of two types of survival models.

2.4.3. Semi parametric

Cox Regression Model

The Cox Regression Model is a technique used in Survival Analysis, which deals with time in the analysis. This method has several advantages, the most important of which is that it is considered one of the modern methods in addition to the ease of dealing with the disappearance data that appears when taking time into account. This model is used in cases where the time variable that precedes the occurrence of a particular event is of importance in the analysis of the phenomenon concerned with the study, and it is also called the Proportional Hazards Model.

Terms of Use of the Cox Form

1-The dependent variable is made up (a binary descriptive variable with a value and a time variable that precedes Event occurrence).

2-Independent variables, regardless of their nature, structural, descriptive, or mixed, and it is expected to have an impact on the phenomenon concerned with the study.

Characteristics of the Cox Model

The characteristics of the cox model has the following properties

1- It does not require that you choose some specific probability model to represent the number of survivals martials.
2- Parametric model.
3- It is easy to combine the variables that depend on time, which are the variables whose value changes on.
4- Over the course of the observation period.
5- The Cox model means the effect of variables on the risk rate, but leaves the basic risk rate undefined.

The archetype of Cox's regression was proposed by (Cox 1972), so T A continuous random variable. The mathematical form of the model is written as

$$Z\left(t/x_p\right) = Z_0(t)e^{\beta_1 x_1+\cdots+\beta_p x_p} \qquad (2.23)$$

where $h\,(t/x)$ is the conditional hazard time t for a subject with a set of predictors $x_1, \ldots, x_p$, $Z_0(t)$is the baseline hazard function, and $\beta_1, \ldots, \beta_p$ are the model parameters describing the effect of the predictors on the overall hazard. It is considered a semi-parametric approach because the model contains a non-parametric component and a parametric component. The nonparametric component is the baseline hazard, Z_0(t). This is the value of the hazard when all covariates are equal to 0, which highlights the importance of centering the covariates in the model for interpretability. Do not confuse the baseline hazard to be the hazard at time 0. The baseline hazard function is estimated non-parametrically, and so unlike most other statistical models, the survival times are not assumed to follow a particular statistical distribution and the shape of the baseline hazard is arbitrary. The baseline hazard function doesn't need to be estimated in order to make inferences about the relative hazard or the hazard ratio, so the effect of any covariate is the same at any time during follow-up, and this is the basis for the proportional hazard's assumption.

- β_j represent the increase in the log hazard ratio for one-unit increase in x_j
- e^{β_j} represent the hazard ratio for one unit increase in x_j
- $\beta_j < 0$ means increasing x_j associated with lower risk and longer survival times.
- $\beta_j > 0$ means increasing x_j associated with increased risk and shorter survival times.

Now we have used different approaches for parameter estimation.

2.5. Maximum Likelihood Estimation (MLE)

The most general of estimation is known as maximum likelihood Estimators (MLE), suppose that we have n units with lifetimes governed by a survivor function $R(t)$ with associated density

$s(t)$ and hazard $Z(t)$. Suppose unit i is observed for a time t_i. If the unit died at t_i, its contribution to the likelihood function is the density at that duration, which can be written as the product of the survivor and hazard functions

$$l_i = s(t_i) = R(t_i)Z(t_i) \tag{2.24}$$

If the unit is still a live at t_i, all we know under non-informative censoring is that the lifetime exceeds t_i , the probability of this event is

$$l_i = R(t_i) \tag{2.25}$$

Let d_i be a death indicator, taking the value one, if unit i died and the value 0 otherwise. Then

$$l = \prod_{i=1}^{n} l_i = \prod_{j=1}^{i} Z(t_i)^{d_i} R(t_i) \tag{2.24}$$

Where $d_i = 1$ when the event occurred and take the value 0 otherwise (censoring).

2.6. Bayesian Method of Estimation.

Bayesian statistics is a theory of statistics based on the Bayesian interpretation of probabilities in which probability expresses a certain degree of belief in an event, which can change as new information is gathered, rather than a fixed value dependent on frequency or slope. The degree of belief may be based on prior knowledge of the event, such as results from past experiences, or on personal beliefs about the event. This differs from a number of other explanations for probability, such as the repeated interpretation that sees probability as a certain limit to the relative frequency of an event after a large number of experiences.

Bayesian statistics use Bayes' Theorem to calculate and update probabilities after new data is obtained. Bayes' theorem describes the conditional probability of an event based on data as well as on prior information or beliefs about the event or conditions related to the event. For example, Bayes 'theorem can be used to estimate parameters through a probability distribution or a statistical model in Bayesian inference, and Bayes' theorem can also be a specific probability distribution that defines a belief as a parameter or a set of parameters.in probability theory and statistics, Bayes' theorem (also known as Bayes' law or Bayes' rule) describes the probability of an event occurring, based on prior knowledge of the conditions that may be related to the event. For example, if cancer is related to aging, then when using Bayes' Theorem, a person's age can be used to make a more accurate assessment of their cancer probability than can be done without knowing the person's age.

Consider we have a sample random variable of the continuous type, the joint marginal p.d.f of $x_1, x_2, \dots, x_n$ is given by

$$f_1(x_1, x_2, \dots, x_n) = \int_{-\infty}^{\infty} f\,(x_1, x_2, \dots, x_n, \theta) d\theta \tag{2.27}$$

If the random variable of the discrete type, integration would be replaced by summation. In either case the conditional p.d.f of sample, given $X_1 = x_1, \dots, X_n = x_n$ is

$$\begin{aligned} k(\theta | x_1, x_2, \dots, x_n) &= \frac{f\,(x_1, x_2, \dots, x_n, \theta)}{f_1(x_1, x_2, \dots, x_n)} \\ &= \frac{f\,(x_1/\theta) f\,(x_2/\theta) \dots f\,(x_n/\theta) h(\theta)}{f_1(x_1, x_2, \dots, x_n)} \end{aligned} \tag{2.28}$$

This relationship is another form of Bayes formula.

Chapter Three

Logistic Distributions

3.1 Introduction

The logistic distribution (LD) has been used in applications in modeling life data, the shape of the logistic distribution and the normal distribution are very similar. This distribution has no shape parameter; this means that the logistic pdf has only one shape, the bell shape.

The log logistic distribution (LLD) has the same relationship to the logistic distribution that the log normal distribution has to the normal distribution. The log logistic distribution has certain similarities to the logistic distribution; A random variable is log logistic distribution if the logarithm of the random variable is logistic distribution, because of this there are many mathematical similarities between the two distribution.

3.2 Logistic Distribution

The log logistic distribution is a commonly used lifetime distribution data analysis since the logarithm of the lifetime variables are logistically distributed because of the well- known properties of the logistic distribution and because it belongs to the location- Scale family. The random variable X has the logistic distribution if it has the following cumulative distribution function (CDF) (Gupta 2010)

$$F\ (x;\ \mu,\sigma) = \frac{1}{1 + e^{(-(x-\ \mu/\sigma)}}\ , \qquad -\infty < x < \infty \tag{3.1}$$

for any arbitrary location parameter μ and for the scale parameter $\sigma > 0$. The probability density function (pdf)(Gupta 2010) corresponding to the CDF (3.1) is

$$f(x) = \frac{exp(-\ (x-\ \mu/\sigma))}{\sigma\ (1 + exp(-(x-\ \mu/\sigma))^2} \tag{3.2}$$

Clearly, the pdf given in (2) is symmetric about the location parameter μ. However, as μ decreases, the pdf is shifted to the left, and when μ increases, the pdf is shifted to the right. Moreover, as σ decreases, the pdf gets pushed toward the mean, or it becomes narrow and taller and as σ increases, the pdf spread out away from the mean, or it becomes broader and shallower. Figure 3.1 shows, pdf in (3.2) with different values of μ and σ=s. Clear, when the location parameter μ is 0 and the scale parameter σ is 1, then the probability density function of the logistic distribution is given by

$$f(x) = \frac{e^{-x}}{(1 + e^{-x})^2} \tag{3.3}$$

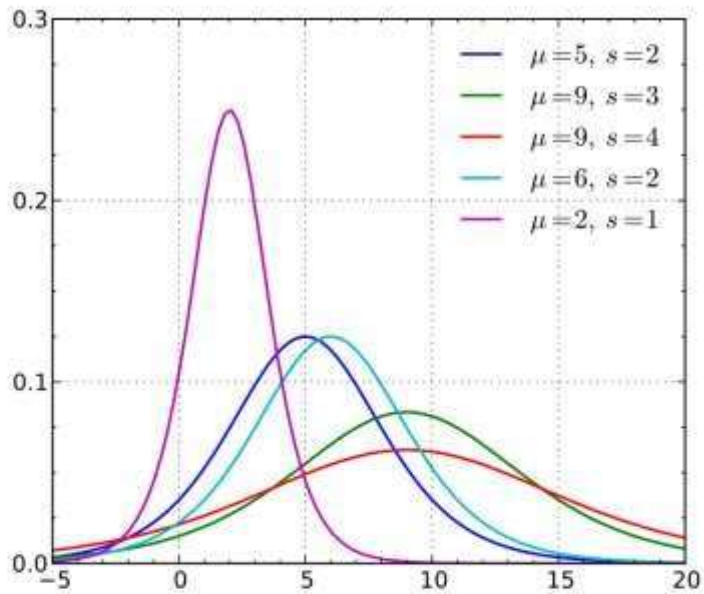


Figure 3.1. pdf of logistic distribution

Note, that the main difference between the normal distribution and logistic distribution lies in the tails and in the behavior of the failure rate function. The logistic distribution has slightly longer tails compared to the normal distribution. Also, in the upper tail of the logistic distribution, the failure rate function levels out for large *x* approaching 1/ δ.

One of the disadvantages of using the logistic distribution for reliability calculations is the fact that the logistic distribution starts at negative infinity. This can result in negative values for some of the results. Negative values for time are not accepted in most of the components of Weibull, nor are they implemented. Certain components of the application reserve negative values for suspensions, or will not return negative results. For example, the Quick Calculation Pad will return a null value (zero) if the result is negative. Only the Free-Form (Probit) data sheet can accept negative values for the random variable (x-axis values).

3.3 Log- Logistic Distribution

In probability and statistics, the log-logistic distribution (known as the Fisk distribution in economics) is a continuous probability distribution for a non-negative random variable. It is used in survival analysis as a parametric model for events whose rate increases initially and decreases later, as, for example, mortality rate from cancer following diagnosis or treatment. It has also been used in hydrology to model stream flow and precipitation, in economics as a simple model of the distribution of wealth or income, and in networking to model the transmission times of data considering both the network and the software.

The log-logistic distribution is the probability distribution of a random variable whose logarithm has a logistic distribution. It is similar in shape to the log-normal distribution but has heavier tails. Unlike the log-normal, its cumulative distribution function can be written in closed form.

3.3.1 Basic Log-Logistic Distribution

The basic log-logistic distribution with shape parameter $\kappa \in (0, \infty)$ is a continuous distribution on $[0, \infty)$ with distribution function given by

$$G(z) = \frac{z^{\kappa}}{1 + z^{\kappa}}, \quad z \in [0, \infty) \tag{3.4}$$

In the special case that $\kappa = 1$, the distribution is the standard log-logistic distribution. The probability density function $g(z)$ is given by

$$g(z) = \frac{\kappa z^{\kappa-1}}{(1 + z^{\kappa})^2}, \quad z \in (0, \infty) \tag{3.5}$$

$g(z)$ in (3.5) has a rich variety of shapes, and is unimodal if $\kappa > 1$. When $\kappa \geq 1$, is defined at 0 as well.

Definition 3.3.1

The quantile function G^{-1} is given by

$$G^{-1}(p) = \left(\frac{p}{1-p}\right)^{1/\kappa}, \quad p \in [0, 1) \tag{3.6}$$

Therefore, the quartiles are given by

1. The first quartile $Q_1 = P_{25} = (1/3)^{1/\kappa}$
2. The second quartile (Median) $Q_2 = P_{50} = 1$
3. The third quartile $Q_3 = P_{75} = 3^{1/\kappa}$

Recall that $p/(1-p)$ is the odds ratio associated with probability $p \in (0, 1)$. Thus, the quantile function of the basic log-logistic distribution with shape parameter κ is the κ^{th} root of the odds ratio function. In particular, the quantile function of the standard log-logistic distribution is the odds ratio function itself. Also of interest is that the median is 1 for every value of the shape parameter.

3.3.2 General Log-Logistic Distribution

The basic log-logistic distribution is generalized, like so many distributions on $(0, \infty)$, by adding a scale parameter. Recall thata scale transformation often corresponds to a change of units (gallons into liters, for example), and so such transformations are of basic importance.

If Z has the basic log-logistic distribution with shape parameter $\kappa \in (0, \infty)$ and if $\beta \in (0, \infty)$ then $X = \beta Z$ has the log-logistic (LL) distribution with shape parameter κ and scale parameter β.

X has distribution function F given by

$$F(x) = P(X \le x) = \frac{x^\kappa}{\beta^\kappa + x^\kappa}, \quad x \in [0, \infty) \tag{3.7}$$

X has probability density function f given by

$$f(x) = \frac{\beta^\kappa \kappa\, x^{\kappa-1}}{(\beta^\kappa + x^\kappa)^2}, \quad x \in (0, \infty) \tag{3.8}$$

As shown in Figure 2. When $\kappa \ge 1$, f is defined at 0 also. f satisfies the following properties:

1. If $0 < \kappa < 1$, f is decreases with $f(x) \longrightarrow \infty$ as $x \downarrow 0$
2. If $\kappa = 1$, f is decreases with mode $x = 0$
3. If $\kappa > 1$, f is increases and then decreases with mode $x = \beta\left(\frac{\kappa-1}{\kappa+1}\right)^{1/\kappa}$.

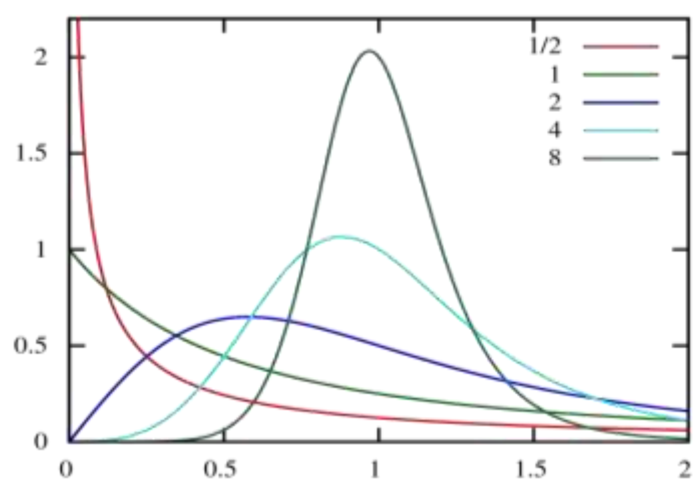


Figure 3.2. pdf of log-logistic distribution with β = 1, and values of κ as shown in legend

Definition 3.3.2

X has quantile function(Alshomrani 2016) F^{-1} given by

$$F^{-1}(p) = \beta\left(\frac{p}{1-p}\right)^{1/\kappa}, \quad p \in [0,1) \tag{3.9}$$

1. The first quartile $Q_1 = P_{25} = \beta(1/3)^{1/\kappa}$
2. The second quartile (Median) $Q_2 = P_{50} = \beta$
3. The third quartile $Q_3 = P_{75} = \beta 3^{1/\kappa}$

3.4 Maximum Likelihood Estimation (MLE)

In this section, we briefly discuss the maximum likelihood estimators (MLE's) of the two parameter log-logistic distribution and discuss their asymptotic properties to obtain approximate confidence intervals based on MLE's.

Let a random sample $X = (x_1, x_2, \dots, x_n)$ of size n can be taken from $LL(\kappa, \beta)$, then the likelihood function $L(\kappa, \beta)$ can be written as;

$$L(x_1, x_2, \dots, x_n; \kappa, \beta) = \prod_{i=1}^{n} f(x_i) = \frac{\prod_{i=1}^{n} \beta^{\kappa} \kappa\, x_i^{\kappa-1}}{\prod_{i=1}^{n} (\beta^{\kappa} + x_i^{\kappa})^2} \tag{3.10}$$

by taking the logarithm of (3.10) then the log-likelihood function $l(\kappa, \beta)$ can be written as;

$$l(\kappa, \beta) = n\kappa \log(\beta) + n\log(\kappa) + (\kappa - 1)\sum_{i=1}^{n} \log(x_i) - 2\sum_{i=1}^{n} \log(\beta^{\kappa} + x_i^{\kappa}) \tag{3.11}$$

Therefore, in order to obtain the MLE's of κ and β, we can maximize (3.11) directly with respect to κ and β or the following two non-linear equations can be solved using iteration method such as Newton-Raphson method (Alshomrani 2016)

$$\frac{\partial l}{\partial \kappa} = n\,log(\beta) + \frac{n}{\kappa} + \sum_{i=1}^{n} log(x_i) - 2\sum_{i=1}^{n} \frac{(\beta^{\kappa}\,log(\beta) + x_i^{\kappa} log(x_i))}{\beta^{\kappa} + x_i^{\kappa}} = 0 \qquad (3.11.a)$$

$$\frac{\partial l}{\partial \beta} = \frac{n\kappa}{\beta} - 2\sum_{i=1}^{n} \frac{\kappa \beta^{\kappa-1}}{\beta^{\kappa} + x_i^{\kappa}} = 0 \qquad (3.11.b)$$

Let us denote the parameter vector by $\boldsymbol{\theta} = (\kappa, \beta)$ and the corresponding MLE of $\boldsymbol{\theta}$ as $\hat{\boldsymbol{\theta}} = (\hat{\kappa}, \hat{\beta})$, then the asymptotic normality results in

$$(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \longrightarrow N_2\left(0, (I(\hat{\boldsymbol{\theta}}))^{-1}\right)$$

where $I(\widehat{\theta})$ is the Fisher's information matrix given by (Alshomrani 2016)

$$I(\hat{\boldsymbol{\theta}}) = \begin{bmatrix} E\left(\frac{\partial^2 l}{\partial \kappa^2}\right) & E\left(\frac{\partial^2 l}{\partial \kappa \partial \beta}\right) \\ E\left(\frac{\partial^2 l}{\partial \beta \partial \kappa}\right) & E\left(\frac{\partial^2 l}{\partial \beta^2}\right) \end{bmatrix} \qquad (3.12)$$

`In practice, as we do not know $\hat{\boldsymbol{\theta}}$, it is useless that the MLE has asymptotic variance $(I(\hat{\boldsymbol{\theta}}))^{-1}$. Hence, the asymptotic variance can be approximated by "plugging in" the estimated value of the parameters. The common procedure is to use the observed Fisher information matrix $O(\underline{\hat{\boldsymbol{\theta}}})$(as an estimate of the information matrix $(\hat{\boldsymbol{\theta}})$) given by

$$I(\underline{\hat{\boldsymbol{\theta}}}) = -\begin{bmatrix} \frac{\partial^2 l}{\partial \kappa^2} & \frac{\partial^2 l}{\partial \kappa \partial \beta} \\ \\ \frac{\partial^2 l}{\partial \beta \partial \kappa} & \frac{\partial^2 l}{\partial \beta^2} \end{bmatrix}_{|(\hat{\kappa},\hat{\beta})} = -H(\boldsymbol{\theta})|_{\bar{\theta}=\underline{\hat{\theta}}} \tag{3.13}$$

where H is the Hessian matrix, $\bar{\boldsymbol{\theta}} = (\kappa, \beta)$ and $\underline{\hat{\boldsymbol{\theta}}} = (\hat{\kappa}, \hat{\beta})$.

To maximize the likelihood, the Newton-Raphson algorithm produces the observed information matrix. Therefore, the variance-covariance matrix is given by

$$\left(-H(\boldsymbol{\theta})|_{\bar{\theta}=\underline{\hat{\theta}}}\right)^{-1} = \begin{bmatrix} var(\hat{\kappa}) & cov(\hat{\kappa},\hat{\beta}) \\ \\ cov(\hat{\beta},\hat{\kappa}) & var(\hat{\beta}) \end{bmatrix} \tag{3.14}$$

Hence, approximate $100(1-\alpha)\%$ confidence intervals for κ and β can be constructed from the asymptotic normality of MLEs as

$$\hat{\kappa} \pm z_{\alpha/2}\sqrt{var(\hat{\kappa})} \quad \text{and} \quad \hat{\beta} \pm z_{\alpha/2}\sqrt{var(\hat{\beta})}$$

Where $0 < \alpha < 1$, $z_{\alpha/2}$ is the upper percentile of standard normal variate.

As shown above, the MLEs can be obtained by setting the two equations (3.11.1) and (3.11.b) to zero. Due to the lack of explicit solutions to these two Equations, we numerically estimate the MLEs using the llogisMLE function from the r STAR package.

3.5. Generalized log-logistic Distribution

Some general aspects of the parametrized log-logistic distribution (LLD) are discussed in the previous section. The three-parameter log-logistic distribution is a generalization of the twoparameter log-logistic distribution. The generalized log-logistic distribution can be obtained from the log-logistic distribution by addition of a shift parameter δ. Thus if X has a log-logistic distribution then $X + \delta$ has a shifted (generalized) log-logistic distribution (GLLD). So Y has a GLL distribution if log $(Y - \gamma)$ has a logistic distribution. The shift parameter adds a location parameter to the scale and shape parameters of the (unshifted) log-logistic.

The properties of this distribution are straightforward to derive from those of the log-logistic distribution. However, an alternative parameterization, similar to that used for the generalized Pareto distribution and the generalized extreme value distribution, gives more interpretable parameters and also aids their estimation.

The generalized log-logistic distribution with location parameter $\mu \in \mathbb{R}$, scale parameter $\sigma \in (0, \infty)$, and shape parameter $\xi \in \mathbb{R}$ is a continuous distribution on $\mathbb{R}$ with distribution function given by

$$F(x;\mu,\sigma,\xi) = \frac{1}{1 + \left(1 + \frac{\xi(x-\mu)}{\sigma}\right)^{-1/\xi}}, \quad x \in \mathbb{R} \tag{3.15}$$

for $1 + \frac{\xi(x-\mu)}{\sigma} \geq 0$.

The probability density function (Pdf)

$$f(x;\mu,\sigma,\xi) = \frac{\left(1 + \frac{\xi(x-\mu)}{\sigma}\right)^{-(1/\xi+1)}}{\sigma\left[1 + \left(1 + \frac{\xi(x-\mu)}{\sigma}\right)^{-1/\xi}\right]^2}, \quad x \in \mathbb{R} \tag{3.16}$$

also for $1 + \frac{\xi(x-\mu)}{\sigma} \geq 0$.

The shape parameter ξ is often restricted to lie in $[-1, 1]$, when the probability density function is bounded. When $|\xi| > 1$, it has an asymptote at $x = \mu - \frac{\sigma}{\xi}$. Reversing the sign of ξ reflects the pdf and the cdf about $x = \mu$. See Figure 3

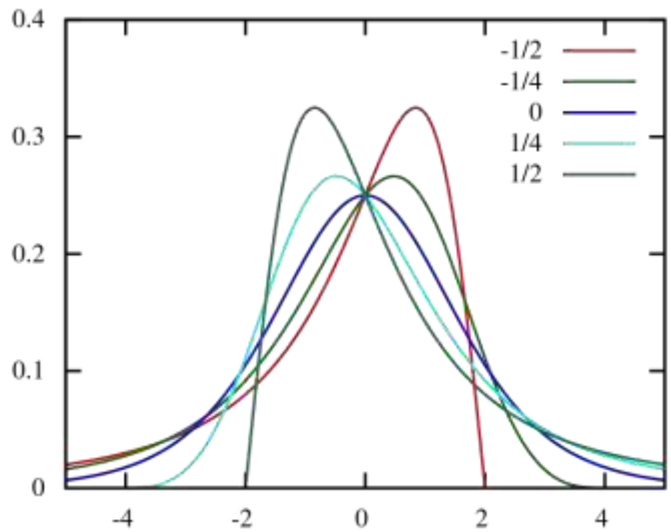


Figure 3.3. pdf of generalized log-logistic distribution with $\mu = 0$, $\sigma = 1$ and values of ξ as shown in legend

An alternate parameterization with simpler expressions for the cdf in (3.15) and the pdf in (3.16) is as follows. For the shape parameter κ, scale parameter ρ, and location parameter μ, the PDF is given by

$$F(x; \mu, \kappa, \rho) = \frac{(x-\mu)^{\kappa}}{(x-\mu)^{\kappa} + \rho^{\kappa}}, \quad x \in \mathbb{R} \tag{3.17}$$

The probability density function (pdf) is then

$$f(x; \mu, \kappa, \rho) = \frac{\kappa}{\rho}\left(\frac{x-\mu}{\rho}\right)^{\kappa-1}\left(1 + \left(\frac{x-\mu}{\rho}\right)^{\kappa}\right)^{-2}, \quad x \in \mathbb{R} \tag{3.18}$$

In survival analysis and for a non-negative random variable T, by introducing a location parameter $\gamma > 0$ the pdf in (3.18) becomes

$$f(t;\gamma,\kappa,\rho) = \frac{\kappa\rho(\rho t)^{\kappa-1}}{[1+(\gamma t)^{\kappa}]^{(\rho^{\kappa}/\gamma^{\kappa})+1}}, \quad t \geq 0 \tag{3.19}$$

hence, the cdf is then given by

$$F(t;\mu,\kappa,\rho) = \frac{[1+(\gamma t)^{\kappa}]^{\rho^{\kappa}/\gamma^{\kappa}} - 1}{[1+(\gamma t)^{\kappa}]^{\rho^{\kappa}/\gamma^{\kappa}}}, \quad t \geq 0 \tag{3.20}$$

for $\rho > 0$, and $\kappa > 0$.

3.6 Bayesian Estimation

Let T be a non-negative random variable follows the generalized log-logistic distribution with unknown parameters $\Omega = \{\kappa, \gamma, \rho\}$, and $D = \{t\}$ be the observed data. Using the pdf in (3.19) the likelihood function of Ω, is then given by

$$L(\Omega) = \prod_{i=1}^{n} f(t_i) = \prod_{i=1}^{n} \frac{\kappa\rho(\rho t_i)^{\kappa-1}}{[1+(\gamma t_i)^{\kappa}]^{(\rho^{\kappa}/\gamma^{\kappa})+1}} \tag{3.21}$$

The joint posterior distribution of the parameters is obtained by combining the joint priors distribution with the likelihood function Ω. In the case of non-informative prior, the posterior distribution is given by

$$P(\Omega|D) = P(\kappa, \gamma, \rho|D) \propto L(\kappa, \gamma, \rho)\ \phi[\delta, \gamma, p] \tag{3.22}$$

where $\phi[\cdot]$ represents the joint prior distribution, and $L(\cdot\ )$ is the likelihood function given in (3.21). In the case of informative priors, Chen et al. (1999) considered a joint posterior distribution from historical data as joint prior distribution by adding an extra parameter $\alpha_0 \in (0, 1)$ that controls the influence of the historical data on the current data, and given by

$$P(\Omega, \alpha_0|D_0) = \phi(\kappa, \gamma, \rho, \alpha_0|D_0) \propto [L(\kappa, \gamma, \rho, \alpha_0)]^{\alpha 0}\ \phi_0[\kappa, \gamma, \rho]\phi_0[\alpha_0] \tag{3.23}$$

where D_0 is the vector of the observed historical data, $L(\cdot\ )$ is the likelihood function (3.21) from these historical data, and $\phi_0[\cdot]$ represents the joint prior distribution considered for $\{\kappa, \gamma, \rho, \alpha_0\}$ from historical data.

In order to make inference about the population parameters, Adaptive Rejection Metropolis Sampling (ARMS) and Gibbs sampling techniques are used, see Salah (2019). These methods are a kind of MCMC technique used to drawing dependent samples from complex highdimensional distributions. However, the joint posterior distribution of the parameters in the proposed model is very complicated. Using OpenBUGS software can greatly simplify the process of simulating these samples, and we only need to specify the data distribution and prior distributions for the model parameters. Prafulla, et. al. (2016).

Chapter Four

Generalized Log Logistic Distribution in Survival Model

4.1 Introduction

The Log Logistic distribution (LL) has attracted wide applicability in Survival and reliability over the last few decades, particularly in probability and statistics, it has a continuous probability distribution for a non-negative random variable, and it is used in Survival analysis as a parametric model for events whose failure rate increases initially and decreases later, for example, mortality from cancer following diagnosis or treatment.

In survival analysis when the mortality rate reaches a peak after a specified period and then slowly back down, it is appropriate to use a model with a non-pulmonary failure rate.

In this thesis, we studied the logistic model as it is similar in shape to a normal logarithmic distribution, but it is more suitable for use in the analysis of survival data, this is because of its greater mathematical tracking ability when dealing with Controlled Observations that occur frequently in such data. The presence of controlled observations causes difficulties when using normal or inverse Gaussian models since the Survival functions in these states are complex.

On the other hand, although the logarithms of small positive numbers are large negative numbers,

log-Normal distribution may give Undue weight to extremely short Survival times.

4.2 The Generalization Log Logistic in Survival Analysis

The generalized Log-Logistic(GLL) model is a three-parameter distribution, and has characteristics similar to those of the log-logistic model. Also it approaches the Weibull in the limit. These advantages enable it to satisfactorily handle both monotone and nonmonotone (unimodal) hazard functions. The nonnegative random variable T has the generalized LogLogistic distribution (GLLD) if it has the following cumulative distribution function (CDF)

$$\begin{aligned} H(t;\gamma,\kappa,\rho) &= -Log\ (R(t\ ;\gamma,k,\rho) \\ &= \frac{\rho^k}{\gamma^k} Log(1+\gamma t)^k \end{aligned} \tag{4.1}$$

for $\rho > 0$, $k > 0$, and $\gamma > 0$ are parameters. The (pdf) corresponding to the CDF (4.1) (Shahedul A. Khan, and Saima K. Khosa 2016) is

$$z(t;\gamma,\kappa,\rho) = \frac{k\rho(\rho t)^{k-1}}{(1+(\gamma t)^k)^{(\rho^k/\gamma^k)+1}} \qquad ,t > 0 \tag{4.2}$$

the survivor function to the (GLLD) is given by (Shahedul A. Khan, and Saima K. Khosa 2016)

$$\begin{aligned} R(t;\gamma,\kappa,\rho) &= 1 - F(t;\gamma,\kappa,\rho) \\ &= \frac{1}{[1+(\gamma t)^{\kappa}]^{-(\rho^{\kappa}/\gamma^{\kappa})}} \end{aligned} \qquad (4.3)$$

Thus, the GLLD for anon negative random variable of the hazard function for $t > 0$ defined as

$$\begin{aligned} Z(t;\gamma,\kappa,\rho) &= \frac{f(\mathrm{t};\gamma,\kappa,\rho)}{F(\mathrm{t};\gamma,\kappa,\rho)}) \\ &= \frac{k\rho(\rho t)^{k-1}}{1+(\gamma t)^{k}} \end{aligned} \qquad (4.4)$$

Let $\alpha = (k, \gamma, \rho)'$, when γ depends on ρ by and $\gamma = \rho\eta^{-1/k}$ with $\eta > 0$ then (4.4) reduces the hazard function of the LLD , otherwise it is closed under Proportional Hazard relationship. Clearly if $k \leq 1$ the hazard function is monotone decreasing , when

$k > 1$ be a unimodal ; which mean Z(t; α) = 0 at t = 0 and when $t = (k - 1)/\gamma^{k}(^{1/k})$ it increases to a maximum) ,also as $t \longrightarrow \infty$ it is approaches zero monotonically). Figure 1 shows the different examples. The GLLD of the hazard function from (4.4) the Weibull hazard function as $\gamma^{k} \rightarrow 0$. All the above gives us that the GLL model is distinguished for being able to satisfactorily deal with monotonous increased risks by k > 1 and small γ.

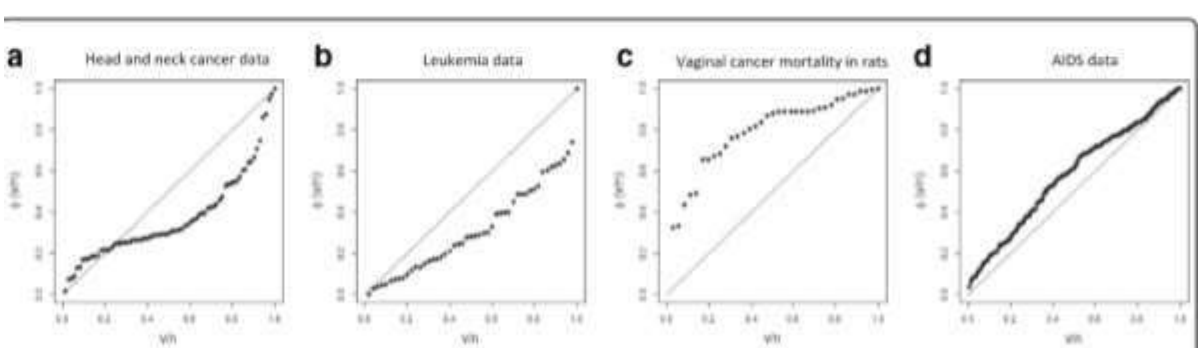


Figure 4.1. A unimodal hazard shape of the survival time

For the family of Proportional hazard models with covariates where $v = (v_1, v_2, v_3, \dots, v_p)'$ and $\hat{\rho} = e^{v\hat{\beta}/k}$, the proportional hazard model can be formulated by

$$Z(t; v) = (Z_0(t; \alpha)e^{v\hat{\beta}}$$

Where the Z_0 is the baseline hazard function , if $Z_0(t; \alpha)$ is definte by the (GLL) (4.5) hazard function from (4.4), then

$$Z(t; v) = \frac{k\hat{\rho}(\hat{\rho}t)^{k-1}}{1 + (\gamma t)^k} \tag{4.6}$$

,$t > 0$

where $\hat{\rho} = e^{v\hat{\beta}/k}$. Hence the generalized log-logistic is closed under proportionality of hazards. The (GLLD) fall under the parametric proportional hazard family is the Weibull when $Z_0 = K\rho(\rho t)^k$, and classified the Cox proportional hazard is semiparametric for the baseline hazard function in (4.6) is denoted by Z_0 .

4.3. Some Basic characteristics of the (GLLD)

The GLLD has r^{th} moment given by

$$E(T)^r = \frac{\rho^k}{\gamma^{K+r}} \frac{\Gamma(\frac{\rho^k}{\gamma^k} - \frac{r}{k})\Gamma(\frac{r}{k} + 1)}{\Gamma(\frac{\rho^k}{\gamma^k} + 1)} \tag{4.7}$$

where $\frac{k\rho^K}{\gamma^k} > 1$

The GLLD has mean given by

$$E(T) = \frac{\rho^k}{\gamma^K} \frac{\Gamma(\frac{\rho^k}{\gamma^k} - \frac{1}{k})\Gamma(\frac{1}{k} + 1)}{\Gamma(\frac{\rho^k}{\gamma^k} + 1)} \tag{4.8}$$

where $\frac{k\rho^\kappa}{\gamma^k} > 1$

4.4 Maximum Likelihood Estimation (MLE)

In this section, we briefly discuss the maximum likelihood estimators (MLE's) of the threeparameter generalized Log-Logistic Distribution and discuss their asymptotic properties to obtain approximate confidence intervals based on MLE's.

Let a censored random sample (t_i, δ_i, y_i) of size n , where t_i is a lifetime and relative to whether $\delta_i = 1 \; or \; 0$ .We assume the vector of Covariates $v_i = (v_{i1}, v_{i2}, \dots, v_{ip})$ for the i^{th} individual can be taken from GLLD$(\acute{\alpha}, \acute{\beta})$, then the likelihood function L$(\acute{\alpha}, \acute{\beta})$ can be written as ;

$$\begin{aligned} L(y_1, y_2, \dots, y_n ; \acute{\alpha}, \acute{\beta}) &= \prod_{i=1}^{n} f(y_i) \\ &= \frac{\prod_{i=1}^{n} \kappa\rho \, (\rho t_i)^{\kappa-1}}{\prod_i^n [1 + (\gamma t_i)^\kappa]^{(\rho^\kappa/\gamma^\kappa)+1}} \end{aligned} \tag{4.9}$$

Let $h = \sum_{i=1}^{n} \delta_i$, $a_i = e^{(\acute{z}_i \beta)}$, $b_i = (\gamma t_i)^k$ .Now by taking the logarithm of (4.9) then the log-likelihood function $\ell\,(\acute{\alpha}, \acute{\beta})$ can be written as;

$$\begin{aligned} \ell(\theta) &= h \log \kappa + h\kappa \log \rho + (\kappa - 1) \sum_{i=1}^{n} \delta_i \log t_i \\ &- \sum_{i=1}^{n} \delta_i \log(1 + b_i) + \sum_{i=1}^{n} \delta_i \log a_i - \left(\frac{\rho}{\kappa}\right) \sum_{i=1}^{n} a_i \log(1 + b_i) \end{aligned} \tag{4.10}$$

where $\theta = (\acute{\alpha}, \acute{\beta})$

The first derivatives of the log-likelihood function are

$$
\begin{aligned}
\frac{\partial \ell(\theta)}{\partial \kappa} &= \frac{h}{\kappa} + h\log\rho + \sum_{i=1}^{n}\delta_i \log t_i - \sum_{i=1}^{n}\delta_i \frac{\partial \log(1+b_i)}{\partial \kappa} \\
&- \left(\frac{\rho}{\gamma}\right)^{\kappa}\sum_{i=1}^{n} a_i \frac{\partial \log(1+\ b_i)}{\partial \kappa} - \left(\frac{\rho}{\gamma}\right)^{\kappa}\sum_{i=1}^{n} a_i \log(1+b_i)\frac{\partial}{\partial \kappa}\left(\frac{\rho}{\gamma}\right)^{k} \\
&= \frac{h}{\kappa} + h\log\rho + \sum_{i=1}^{n}\delta_i \log t_i - \sum_{i=1}^{n}\delta_i \frac{b_i \log(b_i)}{\kappa(1+b_i)} - \left(\frac{\rho}{\gamma}\right)^{\kappa}\left(\frac{1}{\kappa}\right) b_i c_i \\
&- \left(\frac{\rho}{\gamma}\right)^{\kappa}\sum_{i=1}^{n} a_i \log(1+b_i)\, log\left(\frac{\rho}{\gamma}\right) \\
&= \frac{h}{\kappa} + h\, log\,\rho + \sum_{i=1}^{n}\delta_i\, log\, t_i - \frac{1}{k}\sum_{i=1}^{n}\delta_i b_i c_i - \left(\frac{\rho}{\gamma}\right)^{\kappa}\left(\frac{1}{\kappa}\right)\sum_{i=1}^{n} b_i c_i \\
&- \left(\frac{\rho}{\gamma}\right)^{\kappa} log\left(\frac{\rho}{\gamma}\right)\sum_{i=1}^{n} a_i log(1+b_i)
\end{aligned} \tag{4.11}
$$

$$
\begin{aligned}
\frac{\partial \ell(\theta)}{\partial \gamma} &= -\sum_{i=1}^{n}\delta_i \frac{\partial \log(1+b_i)}{\partial \gamma} - \left(\frac{\rho}{\gamma}\right)^{\kappa}\sum_{i=1}^{n} a_i \frac{\partial \log(1+b_i)}{\partial \gamma} \\
&- \sum_{i=1}^{n} a_i\, log(1+b_i)\frac{\partial}{\partial \gamma}\left(\frac{\rho}{\gamma}\right)^{\kappa} \\
&- \sum_{i=1}^{n}\delta_i\left(\frac{k}{\gamma}\right) b_i - \left(\frac{\rho}{\gamma}\right)^{\kappa}\sum_{i=1}^{n} a_i\, log(1-d_i) - \frac{k}{\gamma}\left(\frac{\rho}{\gamma}\right)^{\kappa}\sum_{i=1}^{n} a_i\, d_i \\
&= -\left(\frac{k}{\gamma}\right)\sum_{i=1}^{n}\delta_i b_i - \left(\frac{\rho}{\gamma}\right)^{\kappa}\left(\frac{k}{\gamma}\right)\sum_{i=1}^{n} a_i\, log(1-d_i) - \left(\frac{k}{\gamma}\right)\left(\frac{\rho}{\gamma}\right)^{\kappa}\sum_{i=1}^{n} a_i\, d_i
\end{aligned} \tag{4.12}
$$

(4.13)

$$
\begin{aligned}
\frac{\partial \ell(\theta)}{\partial \beta_j} &= \sum_{i=1}^{n}\delta_i \frac{\partial\, log e^{(\acute{Z}_i\beta)}}{\partial \beta_j} - \left(\frac{\rho}{\gamma}\right)^{\kappa}\frac{\partial e^{(\acute{Z}_i\beta)}\log(1+b_i)}{\partial \beta_j} Z_{ij} \\
&= \sum_{i=1}^{n}\delta_i z_{ij} - \left(\frac{\rho}{\gamma}\right)^{\kappa}\sum_{i=1}^{n} a_i \log(1+b_i) z_{ij}
\end{aligned} \tag{4.14}
$$

where $c_i = \log b_i/(1+b_i)$, and $d_i = b_i/(1+b_i)$.

To improve the convergence of iterative procedures for maximum likelihood estimation and the accuracy of large-sample methods, we remove range restrictions on parameterizations to put

$\kappa^* = \log \kappa$, $\gamma^* = \log \gamma$, and $\rho^* = \log \rho$, then$\alpha^* = (u^*, \gamma^*, \rho^*)'$. The Maximum likelihood estimate of $\theta^* = (\alpha^{*\prime},\ \beta')'$ can be obtained by solving the equations

$$\frac{\partial \ell(\theta^*)}{\partial \kappa^*} = 0, \quad \frac{\partial \ell(\theta^*)}{\partial \gamma^*} = 0, \quad \frac{\partial \ell(\theta^*)}{\partial \rho^*} = 0, \quad \frac{\partial \ell(\theta^*)}{\partial {\beta_j}^*} = 0$$

where

$$\frac{\partial \ell(\theta^*)}{\partial \kappa^*} = \left[\kappa\left(\frac{\partial \ell(\theta)}{\partial \kappa}\right)\right]_{\alpha=e^{\alpha^*}}, \quad \frac{\partial \ell(\theta^*)}{\partial \gamma^*} = \left[\gamma\left(\frac{\partial \ell(\theta)}{\partial \gamma}\right)\right]_{\alpha=e^{\alpha^*}}$$

$$\frac{\partial \ell(\theta^*)}{\partial \rho^*} = \left[\rho\left(\frac{\partial \ell(\theta)}{\partial \rho}\right)\right]_{\alpha=e^{\alpha^*}} \quad \frac{\partial \ell(\theta^*)}{\partial \beta_j} = \left[\frac{\partial \ell(\theta)}{\partial \beta_j}\right]_{\alpha=e^{\alpha^*}}$$

Many software packages have reliable optimization procedures to maximize loglikelihood functions. We used the r and ObeBugs softwares, (see the Appendex).

4.4 Tests and confidence intervals

Tests and interval estimates for the model parameters are based on the approximate normality of the maximum likelihood estimators. The asymptotic distribution of $\hat{\theta}^*$ is approximately a (p + 3)-variate normal distribution with mean θ^* and covariance matrix $D \sum D$ where

$\Sigma = I\left(\hat{\theta}^*\right)^{-1}$ where

$$I(\hat{\theta}^*) = \begin{bmatrix} \frac{\partial^2 \ell(\theta^*)}{\partial \kappa^{*2}} & \frac{\partial^2 \ell(\theta^*)}{\partial \kappa^* \partial \gamma^*} & \cdots & \frac{\partial^2 \ell(\theta^*)}{\partial \kappa^{*2} \partial \beta_p} \\ \frac{\partial^2 \ell(\theta^*)}{\partial \gamma^* \partial \kappa^*} & \frac{\partial^2 \ell(\theta^*)}{\partial \gamma^{*2}} & \cdots & \frac{\partial^2 \ell(\theta^*)}{\partial \gamma^* \partial \beta_p} \\ \vdots & & \vdots & \vdots \\ \frac{\partial^2 \ell(\theta^*)}{\partial \beta_p \partial \kappa^*} & \frac{\partial^2 \ell(\theta^*)}{\partial \beta_p \partial \gamma^*} & \cdots & \frac{\partial^2 \ell(\theta^*)}{\partial {\beta_p}^2} \end{bmatrix}_{\theta^* = \hat{\theta}^*}$$

is the $(p + 3) \times (p + 3)$ observed information matrix (second derivatives of $\ell(\theta^*)$ are given below). By the multivariate delta method, the asymptotic distribution of $\hat{\theta}$ is also approximately normal with mean θ and covariance matrix ΣD, where D is the $(p + 3) \times (p + 3)$ diagonal matrix $diag(\hat{\alpha}, 1, 1, \ldots, 1)$ and $\hat{\alpha} = e^{\hat{\alpha}^*}$.

Second derivatives of $\ell(\theta^*)$:

$$\begin{aligned} \frac{\partial^2 \ell(\theta)}{\partial \kappa^2} &= \frac{\partial}{\partial \kappa}\frac{\partial \ell(\theta)}{\partial \kappa} \\ &= -\frac{m}{\kappa^2}\sum_{i=1}^{n} \delta_i b_i c_i - \frac{1}{\kappa^2}\sum_{i=1}^{n} \delta_i b_i c_i (1 + c_i) - \left(\frac{\rho}{\gamma}\right)^{\kappa} \log\left(\frac{\rho}{\gamma}\right)\left(\frac{1}{\kappa}\right)\sum_{i=1}^{n} a_i b_i c_i \\ &+ \left(\frac{\rho}{\gamma}\right)^{\kappa}\left(\frac{1}{\kappa^2}\right)\sum_{i=1}^{n} a_i b_i c_i - \left(\frac{\rho}{\gamma}\right)^{\kappa}\left(\frac{1}{\kappa^2}\right)\sum_{i=1}^{n} a_i\, b_i c_i (1 + c_i) \\ &- \left(\frac{\rho}{\gamma}\right)^{\kappa} \log\left(\frac{\rho}{\gamma}\right)\left(\frac{1}{\kappa}\right)\sum_{i=1}^{n} a_i b_i c_i - \left(\frac{\rho}{\gamma}\right)^{\kappa}\left\{\log\left(\frac{\rho}{\gamma}\right)\right\}^2 \sum_{i=1}^{n} a_i \log(1 + b_i) \end{aligned} \tag{4.15}$$

$$\begin{aligned} \frac{\partial^2 \ell(\theta^*)}{\partial \kappa^* \partial \gamma^*} &= \kappa\gamma\left(\frac{\partial^2 \ell(\theta)}{\partial \kappa \partial \gamma}\right) = \frac{\partial}{\partial \kappa}\frac{\partial \ell(\theta)}{\partial \gamma} \\ &= -\left(\frac{1}{\gamma}\right)\sum_{i=1}^{n} \delta_i d_i - \left(\frac{\kappa}{\gamma}\right)\left(\frac{1}{\kappa}\right)\sum_{i=1}^{n} \delta_i c_i d_i - \left(\frac{\kappa}{\gamma}\right)\left(\frac{\rho}{\gamma}\right)^{\kappa}\left(\frac{1}{\kappa}\right)\sum_{i=1}^{n} a_i c_i d_i \end{aligned}$$

n n n

$$-\left(\frac{1}{\gamma}\right)\left(\frac{\rho}{\gamma}\right)^{\kappa}\sum_{i=1} a_i d_i - \left(\frac{\kappa}{\gamma}\right)\left(\frac{\rho}{\gamma}\right)^{\kappa}\log\left(\frac{\rho}{\gamma}\right)\sum_{i=1} a_i d_i + \left(\frac{\kappa}{\gamma}\right)\left(\frac{\rho}{\gamma}\right)^{\kappa}\left(\frac{1}{\kappa}\right)\sum_{i=1} a_i b_i c_i$$

$$-\left(\frac{1}{\gamma}\right)\left(\frac{\rho}{\gamma}\right)^{\kappa}\sum_{i=1}^{n} a_i \log(1-d_i) - \left(\frac{\kappa}{\gamma}\right)\left(\frac{\rho}{\gamma}\right)^{\kappa}\log\left(\frac{\rho}{\gamma}\right)\sum_{i=1}^{n} a_i \log(1-d_i) \tag{4.16}$$

$$\frac{\partial^2 \ell(\theta^*)}{\partial\kappa^*\partial\beta_j} = \kappa\left(\frac{\partial^2 \ell(\theta)}{\partial\kappa\partial\beta_j}\right) = \frac{\partial}{\partial\beta_j}\frac{\partial\ell(\theta)}{\partial\kappa}$$

$$= -\left(\frac{1}{\kappa}\right)\left(\frac{\rho}{\gamma}\right)^{\kappa}\sum_{i=1}^{n} a_i b_i c_i z_{ij} - \left(\frac{\rho}{\gamma}\right)^{\kappa}\log\left(\frac{\rho}{\gamma}\right)\sum_{i=1}^{n} a_i \log(1+b_i) z_{ij} \tag{4.17}$$

$$\frac{\partial^2 \ell(\theta)}{\partial\kappa\partial\gamma} = \frac{\partial}{\partial\kappa}\frac{\partial\ell(\theta)}{\partial\gamma} = -\left(\frac{1}{\gamma}\right)\sum_{i=1}^{n}\delta_i d_i - \left(\frac{\kappa}{\gamma}\right)\left(\frac{1}{\kappa}\right)\sum_{i=1}^{n}\delta_i c_i d_i - \left(\frac{\kappa}{\gamma}\right)\left(\frac{\rho}{\gamma}\right)^{\kappa}\left(\frac{1}{\kappa}\right)\sum_{i=1}^{n} a_i c_i d_i$$

$$-\left(\frac{1}{\gamma}\right)\left(\frac{\rho}{\gamma}\right)^{\kappa}\sum_{i=1}^{n} a_i d_i - \left(\frac{\kappa}{\gamma}\right)\left(\frac{\rho}{\gamma}\right)^{\kappa}\log\left(\frac{\rho}{\gamma}\right)\sum_{i=1}^{n} a_i d_i + \left(\frac{\kappa}{\gamma}\right)\left(\frac{\rho}{\gamma}\right)^{\kappa}\left(\frac{1}{\kappa}\right)\sum_{i=1}^{n} a_i b_i c_i$$

$$-\left(\frac{1}{\gamma}\right)\left(\frac{\rho}{\gamma}\right)^{\kappa}\sum_{i=1}^{n} a_i \log(1-d_i) - \left(\frac{\kappa}{\gamma}\right)\left(\frac{\rho}{\gamma}\right)^{\kappa}\log\left(\frac{\rho}{\gamma}\right)\sum_{i=1}^{n} a_i \log(1-d_i)$$

$$= -\left(\frac{1}{\gamma}\right)\sum_{i=1}^{n}\delta_i d_i(1+c_i) - \left(\frac{1}{\gamma}\right)\left(\frac{\rho}{\gamma}\right)^{\kappa}\sum_{i=1}^{n} a_i(d_i + \log(1+d_i) + c_i(d_i - b_i))$$

$$-\left(\frac{1}{\gamma}\right)\left(\frac{\rho}{\gamma}\right)^{\kappa}\log\left(\frac{\rho}{\gamma}\right)\sum_{i=1}^{n}(a_i d_i + \log(1-d_i) \tag{4.18}$$

$$
\begin{aligned}
\frac{\partial^2 \ell(\theta)}{\partial \gamma^2} &= \frac{\partial}{\partial \gamma}\frac{\partial \ell(\theta)}{\partial \gamma} \\
&= \left(\frac{\kappa}{\gamma^2}\right)\sum_{i=1}^{n} \delta_i d_i - \left(\frac{\kappa}{\gamma}\right)^2 \sum_{i=1}^{n} \delta_i d_i\,(1-d_i) + \kappa\rho^\kappa \left(\frac{\kappa+1}{\gamma^{\kappa+2}}\right)\sum_{i=1}^{n} a_i d_i \\
&- \left(\frac{\kappa}{\gamma}\right)^2 \left(\frac{\rho}{\gamma}\right)^\kappa \sum_{i=1}^{n} a_i d_i (1-d_i) + \kappa\rho^\kappa \left(\frac{\kappa+1}{\gamma^{\kappa+2}}\right)\sum_{i=1}^{n} a_i \log(1-d_i)
\end{aligned} \tag{4.19}
$$

$$
\frac{\partial^2 \ell(\theta^*)}{\partial \gamma^* \partial \beta_j} = \gamma\left(\frac{\partial^2 \ell(\theta)}{\partial \gamma \partial \beta_j}\right) = \frac{\partial}{\partial \beta_j}\frac{\partial \ell(\theta)}{\partial \gamma} \tag{4.20}
$$

$$
\begin{aligned}
&= -\left(\frac{\kappa}{\gamma}\right)\left(\frac{\rho}{\gamma}\right)^\kappa \sum_{i=1}^{n} a_i d_i z_{ij} - \left(\frac{\kappa}{\gamma}\right)\left(\frac{\rho}{\gamma}\right)^\kappa \sum_{i=1}^{n} a_i \log(1-d_i)\, z_{ij} \\
&= -\left(\frac{\kappa}{\gamma}\right)\left(\frac{\rho}{\gamma}\right)^\kappa \sum_{i=1}^{n} a_i [d_i + \log(1-d_i)]\, z_{ij}
\end{aligned} \tag{4.21}
$$

for $j, j' = 1,2, \dots, p$

The asymptotic $(1-\alpha)100\%$ confidence intervals for $\gamma, \kappa,$ and ρ, k respectively can be constructed as

$$
\hat{\gamma} \pm z_{\alpha/2}\sqrt{var(\hat{\gamma})} \quad , \quad \hat{k} \pm z_{\alpha/2}\sqrt{var(\hat{\kappa})}\,, \quad \text{and} \quad \hat{\rho} \pm z_{\alpha/2}\sqrt{var(\hat{\rho})}
$$

where $z_{\alpha/2}$ is the upper percentile of standard normal variate

4.5 Generalized LL Distribution in joint Modeling

Co-models are used to determine the association between an internal time –dependent covariate and time until an event of interest occurs. It involves two separate models: a model that takes into account measurement error in the time dependent covariate to estimate its true values (longitudinal model), and another model that uses these estimated values to quantify the association between this covariate and the time to the occurrence of the event .The relationship between the longitudinal covariate, and the failure time process can be evaluated using the Cox relative risk regression model, the problem in estimating the parameters in the Cox model when the linear variable is measured infrequently and with a measurement error.

Parameter estimates are obtained by maximizing the common probability of the covariate process and the failure time process, it uses the covariate data and the survival data simultaneously. Many longitudinal studies collect information on outcomes such as infection or death as well as covariates that vary over time. In order to study the relationship of covariate to survival, we can use the covariate as a time dependent covariate in the relative risk regression model.

This modeling approach has been advocated on the basis that is reduces the bias of the parameter estimate in the cox mode. The covariate and the survival are processed simultaneously because we estimate the parameters that describe the risk of failure as a function of the covariate process and those that describe the risk of failure as a function of the covariate process at the same time, our method uses not only the observed covariate data but also survival information to get estimates of the true covariate value at any time. If we assume that the random effects that determine the process of the covariate are constant for the individual over time and therefore are identical at all times of events when the individual is at risk, this is logical more than the assumption of common variables for those in the risk group will follow the normal distribution at all times if individuals are excluded from Surveillance due to death or censorship, and another model that uses these estimated values to quantify the association between this covariate and the time to the occurrence of the event (time to event model) .

The idea behind the joint modeling technique is to couple the time-to-event model with the longitudinal model , an advantageous characteristic of mixed models is that it is not only possible to estimate parameters that describe how the mean response change overtime . This is one of the main reasons for the use of these models in the joint modeling framework for longitudinal and time- to-event data.

Estimation:

The estimation of the parameters of linear mixed effects models is often based on maximum likelihood principles. Maximization of the log-likelihood function for joint modeling is computationally difficult, as it involves evaluating multiple integrals that have no analytical solution, except in very special cases, in particular, the marginal density of the observed response data for the ith subject is given by the expression

$$p(y_i) = \int p(y_i \mid b_i) p(b_i) db_i \qquad (4.22)$$

4.6 Goodness of fit

The nonparametric estimates are useful for assessing the quality of fit of a particular parametric time-to-event model (Lawless 2002). For a model without covariate, we use the approach to simultaneously examine plots of parametric and nonparametric estimates of the survival function, superimposed on the same graph. Let $S(t, \theta)$ and $S(t)$ be the estimates of the survivor functions based on the parametric model of interest and the Kaplan-Meier method (Kaplan and Meier 1958), respectively. The estimates $S(t, \theta)$ as a function of t should be close to $S(t)$ if the parametric model is adequate. For a model with covariates, we consider residual diagnostic plots, where the residuals are defined based on the cumulative hazard function $H(t, \theta)$. If $S(H(t, \theta))$ is the Kaplan-Meier estimate of $H(t, \theta)$, then a plot of log $S(H(t, \theta))$ versus $H(t, \theta)$ should be roughly a straight line with unit slope when the model is adequate (Lawless 2002). We also use the Akaike's information criterion (AIC) (Akaike 1974) to compare the fits of different models.

The AIC is defined by

$$AIC = -2 \log(Max\ \ Likelih\,ood) + 2(p + k)$$

where p is the number of covariates and k is the number of parameters of the assumed probability distribution (k = 3 for the generalized log-logistic model). In general, when comparing two or more models, we prefer the one with the lowest AIC value. A rule of thumb is that if $\Delta_M = AIC_M - AIC_{Min} > 2$, then there is considerably less support for

Model M compared to the model with minimum *AIC* (Burnham and Anderson 2002).

Chapter Five
Applications

In this chapter, a simulation study is conducted to demonstrate the importance and utility of the GLL distribution in survival data, and then three datasets from the literature are applied to demonstrate the power of the GLL distribution in modeling time_to_event data. Details of the MCMC algorithm we use are in the appendix, including details on how the GLL distribution is implemented in the OpenBUGS development interface.

5.1. Simulation Study.

In all simulations, two covariates were considered in the proportional hazards regression framework: a Continuous covariate (C_1) generated from standard normal distribution; and binary covariate (C_2) generated from Bernoulli (0.5) distribution. α = (0.5, 0.75) corresponding to the covariate vector $\boldsymbol{C} = (C_1,\ C_2)$ was chosen as the regression parameter value. To evaluate the performance of the GLL model, we considered three simulation cases based on the shape of the hazard function. For each case, survival data were generated from a generalized Weibull distribution with a probability density function

$$f(t;\, \gamma, \kappa, \rho, \boldsymbol{\alpha}) = \kappa\rho(\rho t)^{\kappa-1} e^{(C^T\alpha)}[1-\gamma(\rho t)^{\kappa}]^{\frac{e^{(C^T\alpha)}}{\gamma}-1} \tag{5.1}$$

Where $\rho > 0, \kappa > 0,$ and $-\infty < \gamma < \infty$ are parameters. Note that the hazard function of the generalized Weibull distribution is

i. Monotone increasing for $\kappa \geq 0$ and $\gamma \geq 1$

ii. Monotone decreasing for $0 < \kappa \leq 1$ and $\gamma \leq 1$

iii. Unimodal for $\kappa > 1$ and $\gamma < 0$.

The simulations are then specified as follows:

- **Case 1**: The hazard function decreases. Survival times were generated from generalized Weibull with $\kappa = 0.5, \gamma = 0.2,$ and $\rho = 0.2$, and censoring times were generated from the exponential distribution with rate parameter $\lambda = 0.05$.
- **Case 2**: The hazard function increases. Survival times were generated from generalized Weibull with $\kappa = 2, \gamma = 0.2,$ and $\rho = 0.2,$ and censoring times were generated from the exponential distribution with rate parameter $\lambda = 0.07$.
- **Case 3**: The hazard function has a unimodal. Survival times were generated from generalized Weibull with $\kappa = 2,\ \gamma = -0.2,$ and $\rho = 0.2$, and censoring times were generated from the exponential distribution with rate parameter $\lambda = 0.07$.

The choice of model parameter values led to, on average, 40% to 45% censored observations for the above 3 cases. Given the covariates and censoring indicator, we also fit the generalized log-logistic, Weibull and Cox PH models to the simulated survival data. The simulation was performed on finite samples of size n = 100. This simulation was performed to calculate the average biases (AVB), root mean square error (RMSE), and the coverage probability for 95% highest probability density HPD (CP) of all model parameters. Results are presented in Tables 5.1 – 5.3.

As shown in Tables 5.1 - 5.3, the parameter estimates do not show much difference in the first case. However, for continuous covariate (C_1), all three models provided estimates with similar RMSEs and good coverage probabilities; while for binary covariate (C_2), GLL showed the smallest RMSE. In terms of bias, GLL and Cox proportional hazards are almost the same and both outperform Weibull.

In the second case, we assume that the hazard function increases. The results from the GLL and Weibull supporting this assumption in terms of smallest bias and similar RMSE. and both outperform Cox.

According to AVB values, GLL and Cox proportional hazards provided comparable estimates of regression coefficients. However, GLL provided the most accurate estimates in terms of RMSE.

Tables 5.1. GLL Proportional Hazard Model performance.

Case	Parameter	Mean	AVB	RMSE	CP
1	α_1	0.513	0.018	0.174	94%
True values:	α_2	0.744	– 0.023	0.087	96%
α = (0.5, 0.75)	κ	0.517	0.012	0.133	95%
$\kappa = 0.5,\ \gamma = 0.2,$	γ	0.212	0.028	0.218	94%
$\rho = 0.2$	ρ	0.193	– 0.016	0.197	94%
2	α_1	0.516	0.019	0.175	94%
True values:	α_2	0.741	– 0.024	0.088	96%
α = (0.5, 0.75)	κ	2.021	0.024	0.145	95%
$\kappa = 2,\ \gamma = 0.2,$	γ	0.207	0.019	0.212	94%
$\rho = 0.2$	ρ	0.198	– 0.013	0.108	95%
3	α_1	0.517	0.027	0.181	94%
True values:	α_2	0.758	0.017	0.063	97%
α = (0.5, 0.75)	κ	2.039	0.031	0.166	94%
$\kappa = 2,\ \gamma = -0.2,$	γ	– 0.187	0.028	0.263	94%
$\rho = 0.2$	ρ	0.208	0.021	0.137	94%

AVB: Average Biases. **RMSE:** Root Mean Square Error . **CP:** coverage probability for 95% highest probability density HPD.

Tables 5.2. Cox Proportional Hazard Model performance.

Case	Parameter	Mean	AVB	RMSE	CP
1	α_1	0.511	0.017	0.173	94%
	α_2	0.745	– 0.024	0.088	96%
2	α_1	0.516	0.018	0.174	94%
	α_2	0.742	– 0.023	0.087	96%
3	α_1	0.518	0.027	0.180	94%
	α_2	0.757	0.017	0.062	97%

True values: $\boldsymbol{\alpha}$ = (0.5, 0.75). AVB: Average Biases. **RMSE:** Root Mean Square Error . **CP:** coverage probability for 95% highest probability density HPD.

Tables 5.3. Weibull Proportional Hazard Model performance.

Case	Parameter	Mean	AVB	RMSE	CP
1	α_1	0.519	0.021	0.176	94%
True values:	α_2	0.762	0.027	0.094	95%
$\boldsymbol{\alpha}$ = (0.5, 0.75)	κ	0.509	0.011	0.130	95%
$\kappa = 0.5, \quad \rho = 0.2$	ρ	0.202	0.013	0.168	95%
2	α_1	0.515	0.019	0.174	94%
True values:	α_2	0.743	– 0.023	0.086	96%
$\boldsymbol{\alpha}$ = (0.5, 0.75)	κ	2.019	0.024	0.144	95%
$\kappa = 2, \quad \rho = 0.2$	ρ	0.203	0.010	0.107	95%
3	α_1	0.529	0.046	0.196	93%
True values:	α_2	0.786	0.097	0.188	94%
$\boldsymbol{\alpha}$ = (0.5, 0.75)	κ	2.031	0.028	0.162	94%

$\kappa = 2, \quad \rho = 0.2$	ρ	0.202	0.018	0.136	94%

AVB: Average Biases. **RMSE:** Root Mean Square Error . **CP:** coverage probability for 95% highest probability density HPD.

5.2. Breast Cancer Data Analysis.

In this section, we will provide real data for the application, which illustrates the performance and practicality of the Bayesian approach. This is done by applying the GLL to a real data of 686 lymph node positive breast cancer patients. The data is provided as bc.dat in the r software package.

The result of interest is relapse-free survival time; that is, the duration from the start of the study to the beginning of death or recurrence of the disease (usually the time of diagnosis of primary breast cancer) (whichever occurs first). There were 299 events leading to this result, with a median follow-up time of approximately 5 years. In addition, the data were divided into three equal-sized prognostic groups. The three groups are labeled as "Good", "Medium" and "Poor".

We compared the Kaplan–Meier (KM) survival curves with the GLL curves, for all groups. In all estimates the 95% C.I. of GLL curves were approximately covering the Kaplan-Meier curves. Results are presented in Figures 5.1 – 5.4. “r” codes were presented in Appendix A.

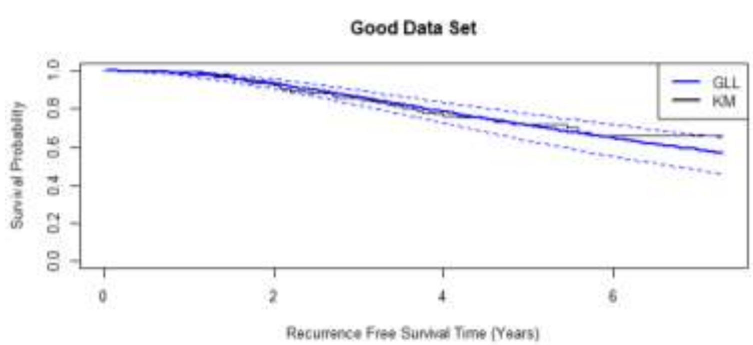


Figure 5.1. Fitted survival curves for prognostic group in the good data.

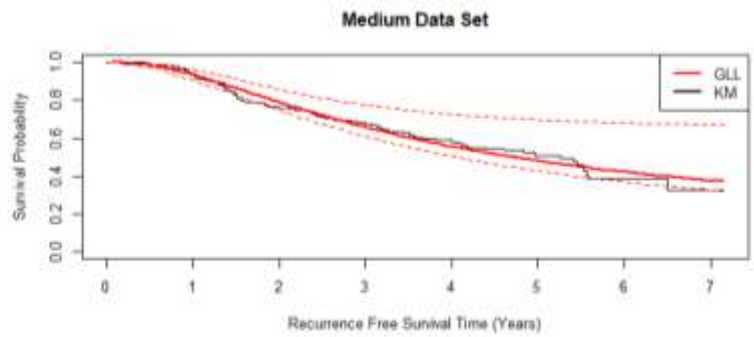


Figure 5.2. Fitted survival curves for prognostic group in the medium data.

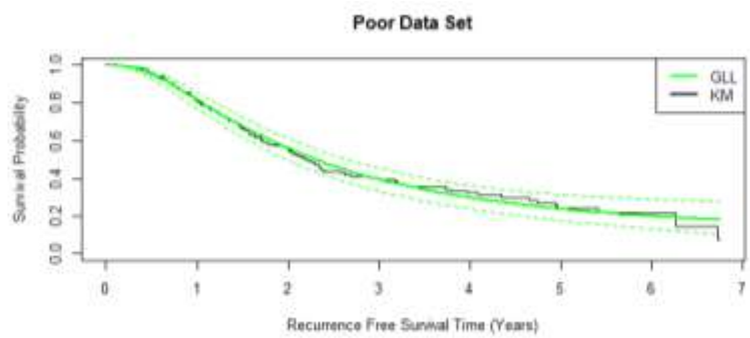


Figure 5.3. Fitted survival curves for prognostic group in the poor data.

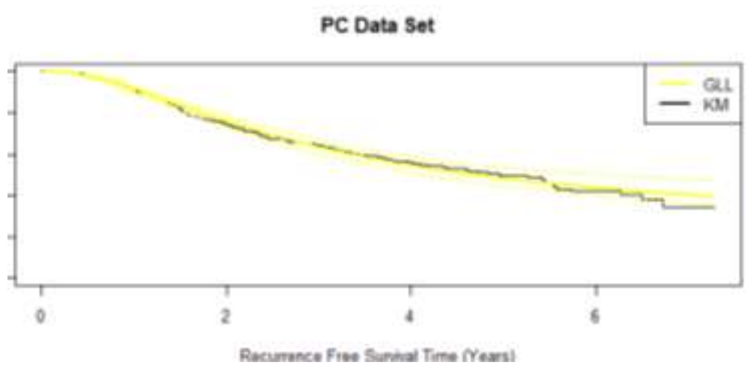


Figure 5.4. Fitted survival curves for prognostic group in the poor data.

As, shown in the previous figures, when the censoring rate is moderate or low, the model will provide a better fit. See Figures 5.2 and 5.3. Moreover, the presence of cured patients (cencored) in all data groups is usually suggested by a KM plots of the survival functions, which show a long and stable plateau with different heavy censoring at the extreme right of the plots, leading to three scenarios of censoring rate (high, moderate, and low) that cover a wide range of tail behaviors.

For the Bayesian analysis of the GLL it was assumed a non-informative gamma G(0.5, 1) prior distributions for κ and ρ, and Normal N(0.5, 1) for γ. The convergence of the MCMC algorithm was checked using the software OpenBUGS choosing different values for these hyperparameters, the same results were obtained when using a larger burn in period for the algorithm.

Since the GLL model is not the default probabilistic model in OpenBUGS, it guarantees the integration of modules for parameter estimation of the GLL model. The Bayesian analysis of the probability model can only be performed on the default probability model in OpenBUGS. For the OpenBUGS code, please refer to the Appendix B.

Using the OpenBUGS software, the model was specified in the same form, and 10,000 MCMC iterations were run. After discarding the first 5,000 burn-in iterations, to eliminate the initial value effects, there are a total of 5,000 samples for summarization and convergence checking. The following figure shows the time series plot and kernel density plot of each simulation parameter used for convergence diagnosis. From the upper graphs of Figure 5.5 to 5.8 we can safely conclude that since the graph does not show an expanded growth or decline trend, but looks like a horizontal band, the chain has converged.

Table 5.4 shows posterior summaries based on MCMC samples from posterior characteristics of GLL model. From the DIC values for each data set, we can conclude that the model has a reasonable fit. However, when we compare the obtained DIC values for each data set, we note that lowest censoring rate laving smallest DIC, as it was expected, since GLL distribution can be considered as “Modiarate Heavy Tailed Distribution”.

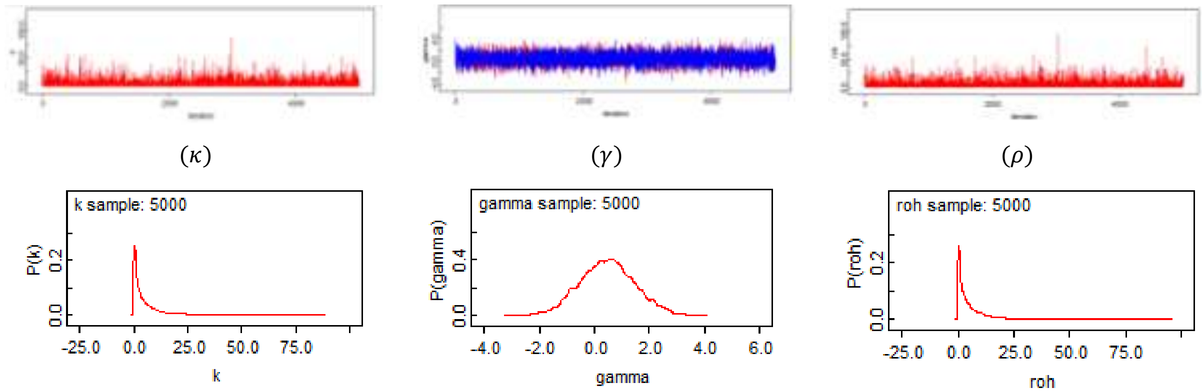


Figure 5.5: The above figures are the time series convergence diagrams of the estimated parameters κ, γ, and ρ. The figures below is the kernel density estimation of the same parameters for the whole set of **bc** data.

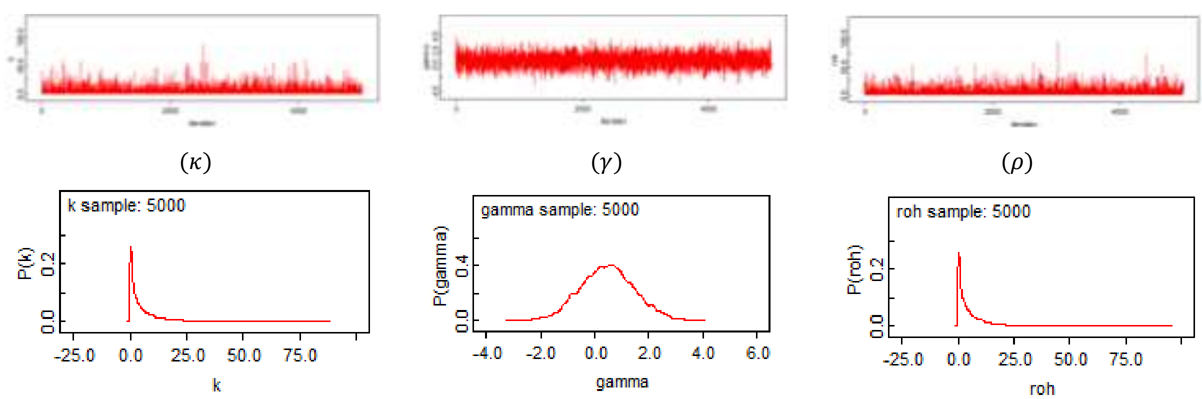
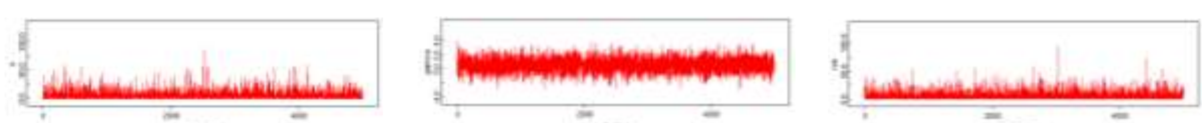


Figure 5.6: The above figures are the time series convergence diagrams of the estimated parameters κ, γ, and ρ. The figures below is the kernel density estimation of the same parameters for the **Good** data set.



(κ) (γ) (ρ)

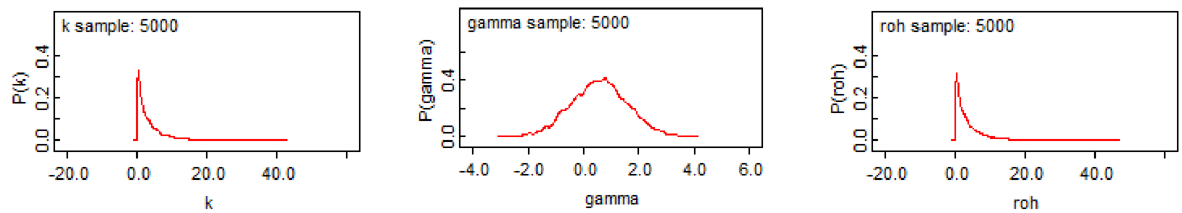


Figure 5.7: The above figures are the time series convergence diagrams of the estimated parameters $\kappa, \gamma,$ and ρ. The figures below is the kernel density estimation of the same parameters for the **Medium** data set.

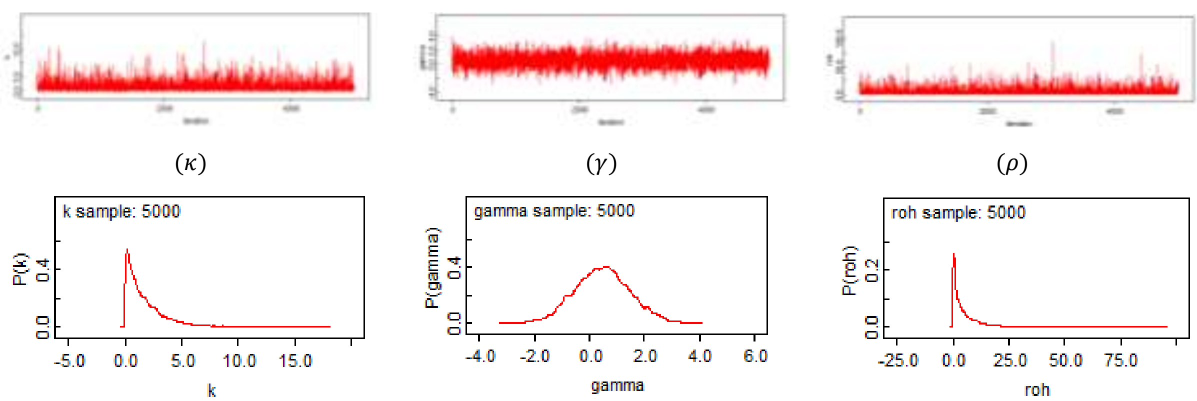


Figure 5.8: The above figures are the time series convergence diagrams of the estimated parameters $\kappa, \gamma,$ and ρ. The figures below is the kernel density estimation of the same parameters for the **Poor** data set.

In addition, the results in the same table show the posterior summary obtained from the simulated samples. Obviously, since the standard error and MC error of the estimated value are small, and all the credible Intervals do not include zero, all the posterior estimates are applicable. Likewise, as shown in the lower graphs of Figures 5.5 to 5.8, all the posterior have approximately unimodal densities.

Table 5.4: Bayesian Estimates of Model Parameters and Deviance Information Criterion (DIC).

Data	Parameter	Estimated			95% Credible Interval		DIC
		Parameter	Std. Error	MC. Error	Lower	Upper	
bc	κ	5.076	0.2745	0.11220	4.5389	5.6130	5343.0
	γ	0.473	0.0384	0.01457	0.3977	0.5482	
	ρ	4.988	0.2655	0.1035	4.4676	5.5083	
Good	κ	5.076	0.2745	0.11220	4.5389	5.6130	1878.7
	γ	0.473	0.0384	0.01457	0.3977	0.5482	
	ρ	4.988	0.2655	0.1035	4.4676	5.5083	
Medium	κ	3.473	0.2788	0.05778	2.9265	4.0194	1257.0
	γ	0.579	0.0660	0.01481	0.4496	0.7083	
	ρ	3.475	0.2749	0.06519	2.9361	4.0138	
Poor	κ	1.801	0.1286	0.02878	1.5489	2.0530	1078.0
	γ	0.4962	0.0662	0.01404	0.3664	0.6259	
	ρ	2.913	0.3470	0.07329	2.2328	3.5931	

5.3 Conclusion

The current study discusses three-parameter GLL models; Bayesian estimates are obtained from simulations and real-world data by using "r" and the Markov Chain Monte Carlo (MCMC) technique of OpenBUGS software. Perform Bayesian analysis under different prior sets and examine convergence patterns using different diagnostic methods. Numerical summaries of MCMC samples based on the late distribution of the GLL model are elaborated on the basis of non-informative priors. As shown in the simulation studies, the GLL model takes into account naturally decreasing and unimodal hazard functions. As demonstrated in bc-data, it turns out that GLL may provide better tuning when describing moderate and low censored data. Furthermore, our simulation studies show that GLL can provide more accurate results in describing monotonically decreasing and unimodal hazard functions than the Weibull and Cox proportional hazards models. In conclusion, the flexibility provided by the GLL model is very useful for adequately describing different types of event-time data.

Appendix (A)

r codes for fitted survival data

```
library(flexsurvcure)
library("survival")
library("splines")
library("lattice") library(ggplot2)
library(ggfortify)
########################data sets ######################### good_data <-
subset(bc, group == "Good" ,select=censrec:recyrs) medium_data <- subset(bc, group
== "Medium" ,select=censrec:recyrs) poor_data <- subset(bc, group == "Poor"
,select=censrec:recyrs) par(mfrow=c(2,2))    # set the plotting area into a 1*2 array
######################## Fit Good #########################
```

```
cure_good <- flexsurvreg(Surv(recyrs, censrec)~1, data = good_data, link="identity", dist="glogis",
mixture=T) print(cure_good) plot(cure_good, ci=TRUE, conf.int=FALSE, main="Good Data Set",
ylab="Survival Probability", xlab="Recurrence Free Survival Time (Years)", lwd=2, col="blue")
legend("topright", lty=c(1,1), lwd=c(2,2), col=c("blue","black"), c("GLL","KM"))
######################## Fit Medium ######################### cure_medium <-
flexsurvcure(Surv(recyrs, censrec)~1, data = medium_data, link="logistic", dist="glogis", mixture=T)
print(cure_medium) plot(cure_medium, ci=TRUE, conf.int=FALSE, main="Medium Data Set", ylab="Survival
Probability", xlab="Recurrence Free Survival Time (Years)", lwd=2, col="red") legend("topright", lty=c(1,1),
lwd=c(2,2), col=c("red","black"), c("GLL","KM"))
######################## Fit Poor ########################
cure_poor <- flexsurvcure(Surv(recyrs, censrec)~1, data = poor_data, link="identity", dist="glogis",
mixture=T) print(cure_poor) plot(cure_poor, ci=TRUE, conf.int=FALSE, main="Poor Data Set",
ylab="Survival Probability", xlab="Recurrence Free Survival Time (Years)", lwd=2, col="green")
legend("topright", lty=c(1,1), lwd=c(2,2), col=c("green","black"), c("GLL","KM"))
######################## Fit bc ########################
cure_all <- flexsurvcure(Surv(recyrs, censrec)~1, data = bc, link="probit", dist="lnorm", mixture=T)
print(cure_all) plot(cure_all, ci=TRUE, conf.int=FALSE, main="PC Data Set", ylab="Survival
Probability", xlab="Recurrence Free Survival Time (Years)", lwd=2, col="yellow") legend("topright",
lty=c(1,1), lwd=c(2,2), col=c("yellow","black"), c("GLL","KM"))
```

Appendix (B)

OpenBugs codes for fitted survival data model

```
{
          for( i in 1 : N ) # observed failure times
        {
         a[i] <-  pow(roh,k)/pow(gamma,k)

         f[i] <- (k*roh* pow(roh*t[i],(k-1)))/ pow((1+pow(gamma*t[i],k)),(a[i]+1))
  S[i]  <-   pow((1+pow(gamma*t[i],k)),(-1*(pow(roh,k)/pow(gamma,k))))     h[i]  <-  (k*roh*
pow(roh*t[i],(k-1)))  /  (1+pow(gamma*t[i],k))      L[i]  <-  pow(f[i],  d[i])*  pow(S[i],  1-d[i])
logL[i] <- log(L[i])   zeros[i] <- 0
        zeros[i]~dloglik(logL[i])
    }
```

```
# Prior distributions of the model parameters     k~dgamma(0.5,0.1)
            roh~dgamma(0.5,0.1)        gamma~dnorm(0.5,1)
 }

# Initial model parameter values  list(k =
2.5,  roh = 2.5, gamma = 0.5)
```

References

[1] Akaike, H. (1974). A new look at the statistical model identification. IEEE Trans. Autom. Control. 19, 716–723.

[2] Burnham, KP, Anderson, DR (2002) Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach. Springer, New York.

[3] Collett, D. (2003). Modelling Survival Data in Medical Research. Chapman and Hall/CRC, Florida.

[4] Cox, DR (1972). Regression models and life-tables. J. R. Stat. Soc. Ser. B. 34, 187–220.

[5] Efron, B. (1977). The efficiency of Cox's likelihood function for censored data. J. Am. Stat. Assoc. 72, 557–565.

[6] Efron, B. (1988). Logistic regression, survival analysis, and the kaplan-meier curve. J. Am. Stat. Assoc. 83, 414–425.

[7] Al-Shomrani, A. A., Shawky, A. I., Arif, O. H., & Aslam, M. (2016). Log-logistic distribution for survival data analysis using MCMC. SpringerPlus, *5*(1), 1-16.

[8] Hsieh, F, Tseng, YK, Wang, JL. (2006). Joint modeling of survival and longitudinal data: Likelihood approach revisited. Biometrics. 62, 1037–1043.

[9] Gupta, R. D., & Kundu, D. (2010). Generalized logistic distributions. Journal of Applied Statistical Science, 18(1), 51.

[10] Shahedul A. Khan, and Saima K. Khosa (2016). Generalized log-logistic proportional hazard model with applications in survival analysis, Journal of Statistical Distributions and Applications (2016)

[11] Hwang, BS, Pennell, ML (2014). Semiparametric bayesian joint modeling of a binary and continuous outcome with applications in toxicological risk assessment. Stat. Med. 33, 1162–1175.

[12] Kalbfleisch, JD, Prentice, RL. (2002). The Statistical Analysis of Failure Time Data. John Wiley & Sons, New Jersey.

[13] Kaplan, EL, Meier, P. (1958). Nonparametric estimation from incomplete observations. J. Am.
Stat. Assoc. 53, 457–481.

[14] Klein, JP, Moeschberger, ML. (2003). Survival Analysis: Techniques for Censored and Truncated Data. Springer, New York.

[15] Kleinbaum, DG, Klein, M. (2012). Survival Analysis: A Self-Learning Text. Springer, New York.

[16] Lawless, JF (2002). Statistical Models and Methods for Lifetime Data. John Wiley & Sons, New Jersey.

[17] Mudholkar, GS, Srivastava, DK, Kollia, GD (1996). A generalization of the weibull distribution with application to the analysis of survival data. J. Am. Stat. Assoc. 91, 1575–1583.

[18]Nardi, A, Schemper, M. (2003). Comparing Cox and parametric models in clinical studies. Stat.
Med. 22, 3597–3610.

[19] Oakes, D. (1977). The asymptotic information in censored survival data. Biometrika. 64, 441– 448.

[20] Pike, MC. (1966). A method of analysis of certain classes of experiments in carcinogenesis. Biometrics. 22, 142–161.

[23] R Core Team: R (2016). A Language and Environment for Statistical Computing. Foundation for Statistical Computing, R. R Core Team, Vienna.

[24] Rizopoulos, D. (2010). Jm: An r package for the joint modelling of longitudinal and time-toevent data. J. Stat. Softw. 35, 1–33.

[25] Rizopoulos, D. (2012). Joint Models for Longitudinal and Time-to-Event Data With Applications in R. Chapman and Hall/CRC, Florida.

[26] Wang, Y, Hossain, AM, Zimmer, WJ. (2008). (Useful properties of the three-parameter Burr XII distribution. In: Ahsanullah, M (ed.) Applied Statistics Research Progress, pp. 11–20. Nova Science Publishers, New York.

[27]Wulfsohn, MS, Tsiatis, AA. (1997). A joint model for survival and longitudinal data measured with errror. Biometrics. 53, 330–339.