Analysis of Dependent Competing Risks using Downton's Bivariate Weibull Distribution

by

Nayanthi M. Karunanayake

A Thesis submitted to the Faculty of Graduate Studies of
The University of Manitoba
in partial fulfilment of the requirements of the degree of

Master of Science

Department of Statistics
University of Manitoba
Winnipeg

Abstract

In this thesis, the focus is developing an inference for the Downton's bivariate Weibull (DBW) distribution, particularly in the context of dependent competing risks. The study begins by investigating the statistical properties of the DBW distribution. Next, the thesis delves into likelihood inference for the DBW distribution using complete bivariate data, specifically in the context of dependent competing risks. Additionally, moment-based estimates based on complete bivariate data are developed and examined. To assess the performance of the point and interval estimates, extensive Monte Carlo simulations are conducted. Furthermore, two real datasets are presented as illustrative examples of the application of the DBW distribution with the inferential method. These illustrations serve to demonstrate the practical relevance and usefulness of the proposed model and methods. The thesis concludes with a comprehensive evaluation of the estimation techniques and suggestions for further research in this area.

Keywords: Downton's bivariate Weibull, lifetimes, competing risk model, dependent competing risks, likelihood inference

Acknowledgment Page

I am deeply grateful to my supervisor, Dr. Katherine Davies, for her exceptional guidance, unwavering support, and endless patience throughout my Master's study. Her invaluable advice and mentorship have played a pivotal role in shaping the direction of my research and enhancing my academic growth. Working under her supervision has been a privilege and a truly enriching experience.

I extend my heartfelt thanks to my advisory committee members, Dr. Saumen Mandal and Dr. Po Yang, for their valuable time, insightful feedback, and thorough evaluation of my thesis work.

I would like to express my appreciation to the staff, supporting staff, and my colleagues in the Department of Statistics at the University of Manitoba for their assistance and camaraderie.

I am grateful to the Department of Statistics for the financial support provided during the past two years, which has enabled me to pursue my studies and research.

I would like to convey my deepest gratitude to my family for their unwavering belief in me and their constant encouragement to pursue my dreams. A special thanks to my dear husband, whose relentless motivation and support have been a constant source of strength during the past two years.

To all those who have supported me in various ways, I am sincerely thankful. This accomplishment would not have been possible without each and every one of you.

Dedication Page

I dedicate this work to my father, mother, brother, and loving husband. Their unconditional love and unwavering support have been the driving force behind my research journey. Their presence in my life has enriched my soul and inspired me to pursue and successfully complete this endeavour.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Analysis of time-to-event data is the examination of data measured over a period from a specified point of origin to an endpoint of interest Collett (1994). For example, in medical studies, patients are followed from an initial date of study until the date of death or study closing date. If an individual dies before the end of the medical study, that individual is considered to had an ‘event’. Failure of a machine or a product also can be considered as the termination point of a lifetime or failure time study. In time-to-event analysis, some individuals may experience the event, and some individuals may have not determined their experience of the event at the end of the study period. As an example, patients who experience death after the study closing date of a medical study will have an unknown survival time. These individuals with unknown survival times would be considered as ‘censored data’. However, in many lifetime experiments, it is possible that we can observe more than one cause of failure. They have often been termed competing risks. For example, an individual in a medical study where death is the event of interest might be recorded as dying from cancer, cardiovascular disease, or other causes.

Researchers like Lawless (1982) and Kalbfleish (1980) have done studies on different competing risks models. It has generally been assumed in previous studies that competing causes of failure are independent, which are referred to as latent failure time models (see for example Wang et al. (2022) and Ren and Gui (2021)). In practice, the independence between several failure causes is a rare case. It is more common to observe dependency between the failure modes. This is due to the fact that the occurrence of one cause of failure can impact the occurrence of the other cause. Lin et al. (1999) describes the modes of failure (cancer recurrence or death) as dependent in a colon cancer study. Univariate distributions for the lifetimes can be used for the models with independent competing risks. However, bivariate or multivariate distributions are more appropriate when the causes of failure are considered dependent.

The primary goal of this thesis is to use Downton's Bivariate Weibull distribution in a competing risks scenario and develop appropriate inferential methods. Since the exponential distribution has a constant failure rate, it is well known that it has serious limitations in modelling lifetime data. In this thesis, we consider the latent failure times model formulation and it is assumed that the latent failure times are independent Weibull random variables with the same shape parameter within a particular set of simulations but different scale parameters. Since the Weibull distribution has both increasing and decreasing failure rates, it is well known that the Weibull distribution can be used more effectively than the exponential distribution to analyze lifetime data. We compute the maximum likelihood estimators (MLEs) of the different parameters of Downton's Bivariate Weibull distribution by proposing a simple iterative scheme.

1.1 Thesis organization

Chapter 2 of the thesis provides an overview of key concepts related to survival analysis. The chapter begins by introducing lifetime distributions and their significance in survival analysis. Next, the concept of multivariate lifetime models is discussed, emphasizing the analysis of multiple correlated lifetimes and their dependence structures. This section lays the foundation for the subsequent exploration of competing risks. The chapter then delves into different types of censoring mechanisms and their implications for survival analysis. Lastly, various approaches to maximum likelihood estimation, a commonly used method for estimating model parameters in survival analysis, are presented.

Chapter 3 of the thesis focuses on presenting the Downton's Bivariate Weibull (DBW) model and data description. Additionally, the chapter highlights key properties and characteristics of the DBW distribution, which is the primary distribution under investigation. In Chapter 4, the construction of the likelihood function is detailed. The process of formulating the likelihood function based on the DBW distribution is discussed.

Chapter 5 presents a comprehensive numerical assessment of inference for the proposed model through a simulation study. Chapter 6 provides illustrative examples to demonstrate the practical application of the developed methodology.

Finally, the thesis concludes with a summary of the findings, implications, and potential areas for future research in Chapter 7.

Chapter 2

Preliminaries

2.1 Survival Analysis

The time-to-event analysis involves the analysis of the length of time to the occurrence of a predefined outcome of interest. Simply, it is the study of the time between a fixed starting point and a terminating event. In many different cases, the time-to-event data is referred to as “survival data”, “lifetime data”, “failure times”, “reliability data,” and “event history data,” depending on the focus and context where it is used. The time-to-event analysis is mainly introduced as the survival or failure time analysis. Time to the event could vary from time to fatal event i.e., death, or time to occurrence of a clinical endpoint such as disease, or failure of a product or machine. In medical studies, the lifespan of a certain population is measured to estimate the duration between the birth and death of each individual. Death is considered as the outcome of interest for survival/ lifetime data. Again, the failure of a machine or the durability of a product also can be considered an endpoint of interest for the failure time/survival time analysis. Lawless (2003) mentions that the time scale may not always be real or chronological time to consider in lifetime measures. Miles driven might be used as a time scale with motor vehicles, and the number of pages of output for a computer printer or photocopier might be considered as the time measurement.

Lawless (2003) explains that the lifetime variable of the lifetime distribution models can consist of continuous, discrete or mixed discrete-continuous random variables. The lifetime variable is considered discrete when the lifetimes are grouped or measured as several cycles. Most lifetimes are nonnegative continuous random variables. We are more concerned about the continuous lifetime variables.

Let T be a positive or nonnegative random variable representing the lifetimes of individuals in some population. From here onwards, the probability density function (PDF) of T is denoted by *f(t)* and the cumulative distribution function (CDF) is denoted by *F(t)*. Then the CDF of T is:

$$F(t) = Pr(T \leq t) = \int_0^t f(x)\ dx.$$

The survivor/survival function *S(t)* gives the probability that an individual of interest will be surviving until the time t :

$$S(t) = Pr(T \geq t) = \int_t^\infty f(x)\ dx. \tag{2.1.1}$$

The CDF and survival function can be interconnected as follows:

$$S(t) = 1 - Pr(T \leq t) = 1 - F(t).$$

The survival function is $0 \leq S(t) \leq 1$ for any t and it is a monotone decreasing function. If $S(t)$ is a proper survival function, then $\lim_{t\to 0}\ S(t) = 1$ and $\lim_{t\to\infty}\ S(t) = 0$.

The hazard function or the hazard rate function is the probability of an individual's failure during a very small time interval, assuming that the individual has survived to the beginning of the interval. It represents the instantaneous failure in the time interval $(t + \Delta t)$ given that the individual has survived up to t. The hazard function of T is denoted by *h(t)* and is defined as:

$$h(t) = \lim_{\Delta t \to 0} \frac{Pr(t \leq T < t + \Delta t | T \geq t)}{\Delta t} = \frac{f(t)}{S(t)} \tag{2.1.2}$$

The cumulative hazard function *H(t)* is defined as:

$$H(t) = \int_0^t h(x)\, dx.$$

Further, it is easy to derive expressions for *S(t)* and *f(t)* in terms of *h(t)*.

2.1.1 Lifetime Distributions

Lifetime distributions such as exponential, Weibull, and gamma distributions are most frequently used in studies involving lifetime data. The exponential distribution is one of the probability distributions that come under the exponential family of probability distributions, which includes other distributions such as normal, binomial and gamma. Exponential distributions are probabilistic representations of the intervals between events in processes when events occur continuously and independently at a constant average rate. The PDF of an exponential distribution for lifetimes can be defined as:

$$f(t) = \lambda e^{-\lambda t} \quad , t > 0$$

where $\lambda > 0$ is the rate parameter and the survival function can be derived from Eq: 2.1.1 as $S(t) = e^{-\lambda t}$. The hazard function derived from Eq: 2.1.2 is $h(t) = \lambda$, which is a constant and independent of time. One of the key properties of exponential distributions is that they are memoryless. Usually, the lack of memory refers to situations in which a “waiting time” distribution isn’t affected by the amount of time that's already passed. Simply it indicates that the age of a subject does not affect future survival. That is not true in general and many survival data cannot fit into the exponential distribution.

The Weibull distribution is a generalization of exponential distribution proposed by Weibull (1939). The Weibull distribution is a popular choice for modeling survival times in a wide range of applications. It is particularly useful for modeling data with varying hazard rates and can be parameterized to reflect different shapes of hazard functions. The PDF of the Weibull distribution is defined as follows:

$$f(t) = \frac{\beta}{\alpha}\left(\frac{t}{\alpha}\right)^{\beta-1} e^{-\left(\frac{t}{\alpha}\right)^{\beta}} \qquad t > 0,$$

where $\alpha > 0$ and $\beta > 0$ are scale and shape parameters respectively.

If t is a "time-to-event", the Weibull distribution gives a distribution for which the event rate is proportional to a power of time. When $\beta < 1$, the event rate decreases over time. This situation may interpret the significant infant mortality in a survival study or the high early failings rate of products that gradually decreases over time. The case of $\beta > 1$ indicates that the event rate increase with time. This happens when subjects of interest in a time-to-event study experience the outcome due to wear or age, or if there is a process of "aging". A value of $\beta = 1$ indicates that the event rate is constant over time, and the Weibull distribution reduces to an exponential distribution. Figure 2.1 shows density values of Weibull distribution for different shape parameters with the same scale parameter.

The survival function for T can be derived as $S(t) = e^{-\left(\frac{t}{\alpha}\right)^{\beta}}$. Similarly, the hazard function can be obtained as hazard function is $h(t) = \frac{\beta}{\alpha}\left(\frac{t}{\alpha}\right)^{\beta-1}$, where hazard function

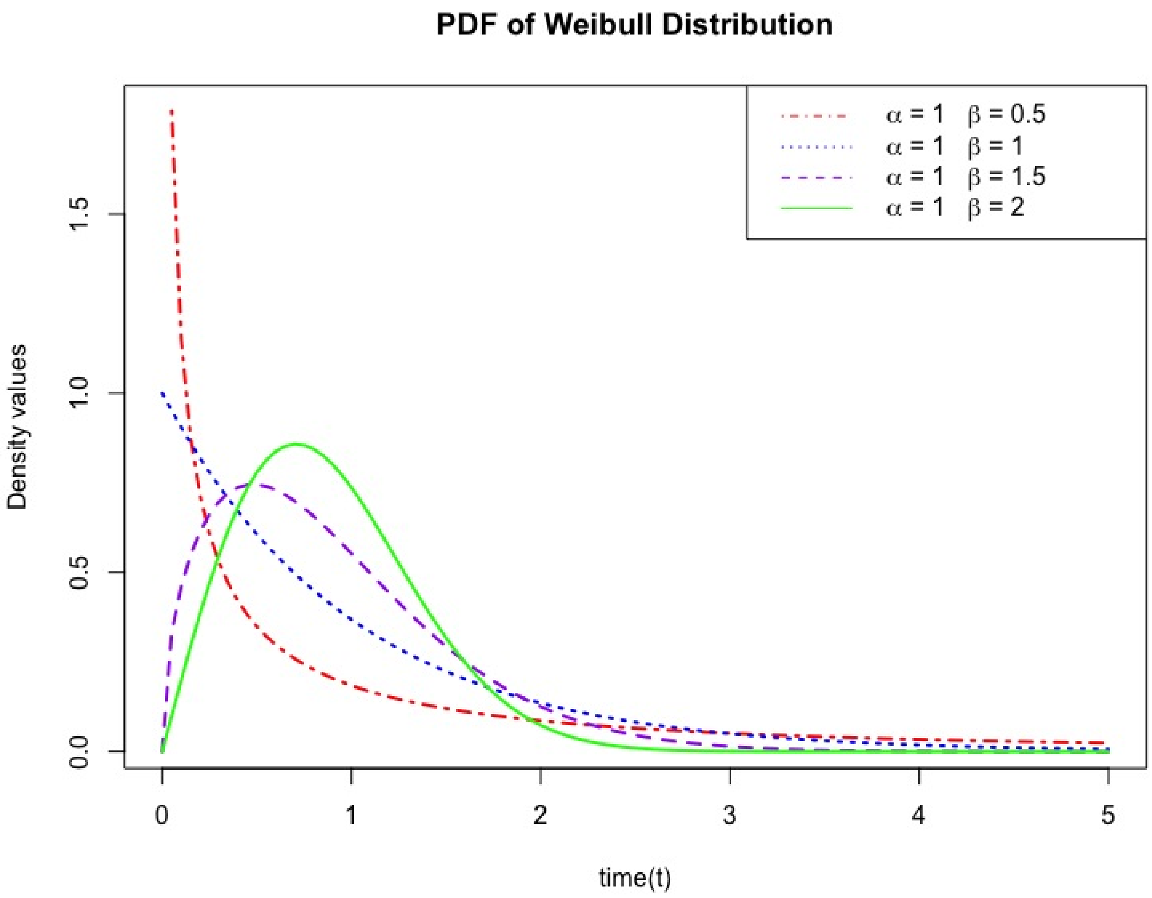


Figure 2.1: PDFs of Weibull distributions with $\alpha = 1$ and $\beta = 0.5, 1.0, 1.5, 2.0$

increases monotonically with respect to time. Conversely, based on the initial term *S(t)*, *f(t)* and *h(t)* can be obtained using the above-derived equation expressions.

The gamma distribution is also a continuous probability distribution that is widely used in statistics and probability theory. The gamma distribution has a shape parameter (α) and a rate parameter (β) and is often denoted as $Gamma(\alpha, \beta)$. The gamma distribution is particularly useful when the hazard rate, or the probability of an event occurring at a given time, is not constant over time. The gamma distribution can be used to model both increasing and decreasing hazard rates and can also be used to model hazard rates that initially increase and then decrease over time. The PDF of the gamma distribution is defined as follows:

$$f(t) = \frac{t^{\alpha-1}e^{-\beta t}\beta^{\alpha}}{\Gamma(\alpha)} \qquad for \quad t > 0 \quad and \quad \alpha, \beta > 0$$

Overall, the gamma distribution is a flexible and widely used probability distribution in survival analysis that can be used to model a wide range of hazard rates and duration data.

2.1.2 Multivariate lifetime models

Univariate survival analysis focuses on the analysis of the survival concerning the single variable under the investigation of study but necessarily ignores the influence of other variables. It is more common to have a situation where several types of events lead to processes with multiple responses, duration, or lifetimes for units in the study.

From Hougaard (2014), multivariate data can be a family of data(lifetimes) or multivariate data(lifetimes) for a single subject. When subjects in a study fall into a “group” or “cluster” of individuals whose lifetimes have some degree of association, those lifetime data have termed a family of data. Crowder (2012) discusses an example of studying the effects of heredity on the life length of identical twins. Pairs of twins or animals in the same litter would expect some association in lifetimes as they have similar genetics.

Lawless (2003) describes the consideration of multivariate lifetime distributions, which can be specified in terms of a joint PDF or a joint survival function,

$$S(t_1, ..., t_k) = Pr(T_1 \geq t_1, ..., T_k \geq t_k).$$

Multivariate data for a single subject is the most common situation that occurs in lifetime data analysis. The situation basically arises where there is more than one separate lifetime for one unit in the study. Eruption of teeth can be considered one of the most common examples of multivariate lifetimes for the same subject. All teeth are checked at once at visits to the dentist. When considering the failure time of paired organs, like eyes, lungs, and kidneys, we are dealing with multivariate data for a single individual. This is because the failure time of the paired organs is taken from the same subject. Similarly, this type of multivariate data can be the failure times of paired components, such as left and right brake pads in a car or left and right aircraft engines. Lawless (2003) explains that the lifetime T_j of this circumstances can be observed only if $T_1, ..., T_{j-1}$ have already been observed.

In Hougaard (2014)'s article, he explains that analyzing multivariate data is important for studying dependence, which is particularly relevant in situations such as dental applications (teeth for one person is considered as multivariate data) and paired organ or component failure times. This dependence allows for the use of available information at a one-time point to predict future events, such as predicting the likelihood of a twin developing an inherited disease if the other twin has already been diagnosed. Additionally, the data can be used in survival analysis to model failure times with dependent causes of failure.

2.2 Censoring

The first and most crucial difficulty faced in the time-to-event analysis is censoring. Censoring is said to be present when certain subjects experience the event but information about the time to outcome event is unavailable for some fraction of subjects before the end of the study. Practically speaking, it may not be possible to examine every subject's entire lifetime during the study period due to various reasons. Simply, when survival times are unknown for a subset of the study group, the rest of the data are considered "censored data".

Censoring may occur due to reasons like loss of follow-up, the non-occurrence of outcome events before the end of the study or the subject experiencing a different event than the one under study.

2.2.1 Point and interval censoring

Censoring mechanisms can be broadly classified into two types, i.e., point and interval censoring. Leung et al. (1997) extensively portray each of these distinguish censoring mechanisms. When the subjects in the study have exactly known censoring time or survival time, then those subjects are introduced as point-censored data. Continuously monitored or properly documented data have the exact censoring times, which leads to point-censored data. A more descriptive illustration of different types of point-censored observations examined between T_0 to T_1 is shown by Leung et al. (1997) depicted in Figure 2.2. The line ending with an asterisk (*) at the end of the line is to indicate an occurrence of the event of interest (true survival time), and the open point (o) is the occurrence of any other event besides the event of interest (censoring time).

Left and right censored data are considered in the point censoring mechanism. Tang et al. (2021) explains a subject is said to be left-censored if that subject has experienced the event sometime prior to entering the study, and the exact time is unknown. Case D in Figure 2.2 represents the left censoring. However, left censoring is usually not considered a crucial issue in most studies. For example, consider a study investigating the duration of pregnancy. Where pregnant women were examined at 7 months but some of them already had their babies. Those who had a baby before 7

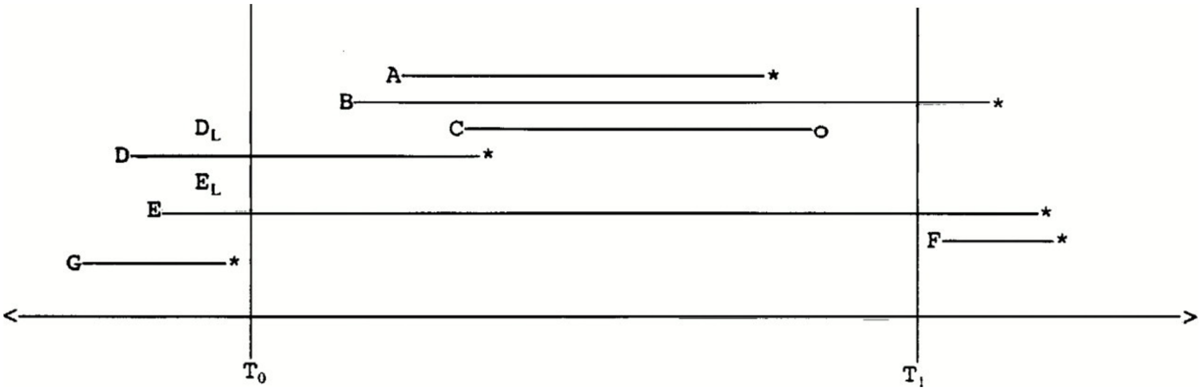


Figure 2.2: Types of point-censored data

months would have a survival time that was left censored because the actual births had happened before the 7 months, and the birth times are not exactly known.

Right censoring occurs when a subject is lost to follow-up, or the event does not take place during the study duration. Subjects B and C in Figure 2.2 portray the right censoring. More specifically, subject C can occur due to the loss of follow-up before the end of the study while subject B occurs due to the termination of the study before experiencing the event. Consider the previous example of a pregnancy duration study. Some women in the study may still be pregnant at the completion of the study, and it is unknown how long their pregnancies will last. These women will be considered as the right-censored subjects of the study. Moreover, there might be situations like case E in Figure 2.2 which shows a subject censored from both cases, i.e., left and right censoring.

Interval censoring arises when a specific subject is followed only for an interval of time in the study. The study subjects may be lost to follow-up for some period of time but return for the examinations during the study period. The same study of the pregnancy period can be used to explain interval censoring. If the exact time of the birth of a baby is not known, but it is known to be within some interval of time, then that subject is said to be interval censored.

2.2.2 Type I, Type II, Hybrid and Random censoring

Right censored data can be either Type I censored or Type II censored data. When the completion of a study is set at a fixed time, subjects that did not experience the event are said to be Type I right censored. In Type I censoring, the number of censored subjects (and the uncensored subjects) is considered a random variable, even with the fixed total sample size.

Type II right censoring occurs when the study is intended to end at a pre-specified number of events of interest among the subjects has occurred. As an example, the study is terminated when 100^{th} failure occurs among the 200 subjects in the study group. As well as the sample size, the number of censored subjects is also fixed in advance. In type II censoring, the completion times of the study are random outcomes.

Epstein (1954) originally introduced the hybrid censoring scheme as a combination of Type I and Type II censoring. Again, in hybrid censoring, two types of censoring schemes are introduced, i.e., Type-I hybrid censoring scheme (Type-I HCS) and Type-II hybrid censoring scheme (Type-II HCS). Balakrishnan and Kundu (2013) broadly discusses and analyses the hybrid censoring scheme.

When both the number of subjects in the study and the termination times of the study turns out to be random, then random censoring arises. According to Tang et al. (2021), the censored subjects do not all have the same censoring time. Some subjects may experience the event, while some subjects are alive or do not experience the studying event.

Progressive censoring is a type of censoring mechanism in survival analysis where the event times of some individuals are observed only at specific intervals, leading to progressively increasing levels of censoring over time. This censoring also has two types of censoring, named Type I progressive censoring and Type II progressive censoring. For more details on progressive censoring see, Balakrishnan and Cramer (2014).

More details on different censoring mechanisms, examples and analyses can be found on Lagakos (1979), Hosmer et al. (1999), Radke (2003), Suttner (2019) and Tang et al. (2021).

2.3 Competing risks

For a single lifetime of a unit in a study, the occurrence of the failure or the endpoint of an event may have one cause, or sometimes it may have more than one distinct cause. In many lifetime studies, it is more common to examine different causes of failures instead of the specific cause of failure which is focused on investigating in the study. One of the common examples described in the literature is an individual who is in a study of prostate cancer may die due to a fatal accident or any other disease (Gross and Clark, 1975). The existence of different competing risk factors for the failure of subjects (including the cause of failure of interest) of the study is referred to as competing risks. Each risk competes with the other to deliver the event of interest. Then the order of the occurrence of each cause of failure has effects on each other. For example, the warranty period of a product may end due to the expiration of time or the exceeding of the usage limit; a marriage may end due to death or divorce. If the study is unaware of competing causes of failures other than the event of interest, then there will be subjects with unobserved events due to the occurrence of the unknown event (regardless of the duration of the study or follow-up). So, it is important to identify and include the competing risks in lifetime models. In such cases, the lifetime of the subject will be decided based on the first occurrence of the failure type. The unique characteristic of competing risks data is that the data consist of two indications about lifetime component; i.e. the time when the failure occurred and an indicator of the cause of failure. Further, Crowder (2012) provides detailed examples involving different multiple failure mode situations, which characterize the competing risks.

The competing risks model is referred to in the literature when different methods of analysis are used to analyze the competing risks. The analysis of competing risk data can be performed using many methods. Some researchers may ignore the presence of competing risks, other than the cause of failure of interest in the study. Some may consider analyzing the lifetime data based on different risk factors separately. Many researchers carried out different studies considering the presence of competing risks in the lifetime data analysis in both parametric and non-parametric designs. Researchers like Cox (1959), David and Moeschberger (1978), Berkson and Elveback (1960) considered parametric setup, assuming different lifetime distributions follow a parametric distribution such as exponential, gamma or Weibull.

Bhattacharya et al. (2014) mentions two parametric competing risk model adaptations mainly used in survival studies when all the information in competing risk data is considered in the analysis process; i.e. the latent failure times model proposed by Cox (1959) and the cause-specific hazard function model introduced by Prentice et al. (1978). If the lifetime of the subjects is T and the cause of failure is C, then the pair (T, C) is defined for each subject. The model for the pair (T, C) can be approached by specifying the model for $Pr(T \geq t, C = j)$ or by specifying cause-specific hazard functions (Lawless, 2003). When the lifetime is taken as the minimum lifetime of the multiple competing risks for the model, then it refers to the latent lifetimes model. Consider an individual with k different causes of failures (or risks) and let T_j be the time of the failure for each of the risk j $(j = 1, ..., k)$. The subject of interest will fail when the first cause of failure arises. Then the lifetime for the latent model would be

$T = min(T_1, ..., T_k)$.

Miyakawa (1982) specifically mentioned the fact that specifying the cause of failure may be challenging in practical scenarios. There might be situations where the observed lifetimes are without any observable cause of failure. Miyakawa (1984) first addressed this situation and obtained the maximum likelihood estimators (MLEs) and uniformly minimum variance unbiased estimators (UMVUEs) in the context of the exponential lifetime for incomplete competing risks data. Guess et al. (1991) also consider the same issue, i.e. partial information on the cause of failure. Kundu and Basu (2000) further investigated model analyzed Miyakawa (1984), with the purpose of addressing different distributional properties missed in earlier studies and additionally considering the Weibull distribution for the lifetimes. In Park and Kulasekera (2004), the authors developed the parametric inference of missing causes of failures among several groups assuming lifetimes are exponentially distributed, based on the original work by Nelson (1970). Later Kundu and Sarhan (2006) examined the same study with several groups, considering the Weibull case.

Different researchers have done impressive work on the analysis of lifetime data in the presence of competing risks, especially in the case of censored data. Even Kundu and Basu (2000) considered complete lifetime data (incomplete in terms of causes of failures), and worked with several authors in case of incomplete lifetimes under competing risks models. Kundu et al. (2017) considered the left truncated and right censored data in the analysis of competing risk data. Bhattacharya et al. (2014) examined hybrid censored lifetime data in the existence of competing risks, while Pareek et al. (2009) did the work on progressively censored competing risks data. Additionally, analysis of competing risks data progressively censored data have been further extended with the Bayesian inference in the study by Kundu and Pradhan

(2011). All of these mentioned analyses of censored data were carried out considering the latent failure times model with lifetime distribution under the assumption of the Weibull type.

In the case of the non-parametric settings, Kaplan and Meier (1958), Efron (1967) and Peterson Jr (1977) have done some impressive work in analyzing the competing risks data.

2.3.1 Independent and dependent competing risks

Two competing causes of failures are said to be independent if the possibility that one kind of failure will occur does not affect the possibility that the other type of failure will occur. Lifetimes are assumed to be independent in Cox's formulated latent failure time model. Most of the previous studies, including the parametric cases mentioned above with the latent lifetime model, have assumed independent lifetimes under the causes of failures. Even though it is common to assume independence in competing risk problems, the more reasonable assumption would be the dependency between lifetime distributions for competing causes. The cause-specific hazard function model introduced by Prentice can accommodate situations where lifetime distributions of the competing causes may not be independent. However, in literature, the identifiability of the dependent competing risks is a well-known argument. Kalbfleish (1980), Crowder (1991, 1994) etc. implied that based on the observed data alone, it is impossible to test the assumption of failure time distributions in the presence of competing risks without the information of covariates.

In the case of competing causes of failures, the occurrence of one failure mode will affect the occurrence of other failure modes. Prentice et al. (1978) specially mentions that the dependence between the causes of failures is something we cannot ignore in lifetime models with competing risks. The dependent competing risks problem was addressed by Babu et al. (1992) by their initiative work using a mixture model formulation in a non-parametric setting. Kundu et al. (1992) formulated the corresponding parametric version under the assumption of observable competing risks. Kundu and Basu (2003) assumed a partially incomplete dependent competing risks model by expanding the analysis of Miyakawa (1982) and Kundu and Basu (2000). In early stages, Moeschberger and Klein (1996) also examined the bivariate lifetime distributions in presence of competing modes of failures.

More recently many works such as Feizjavadian and Hashemi (2015), Wu et al. (2017) have focused on the extending statistical analysis of dependent competing risks with censoring. In Samanta and Kundu (2021) authors addressed the Bayesian perspective of the dependent competing risks. It seems more and more research works can be seen focusing on the analysis of parametric lifetime models with dependent competing risks.

2.4 Maximum Likelihood Estimation

Following the formulation of a statistical model for observed data, interpreting how it works in a larger population is an essential aspect of any statistical study. That is where inferential statistics comes into the picture. The statistical inference process involves analyzing data to determine the properties of an underlying probability distribution. More commonly model-based statistical inference can usually be reached by evaluating the likelihood of the observed data. Boos et al. (2013) describes that likelihood methods are basically asymptotically optimal under the assumption that the statistical model is correct. The likelihood of data is not necessarily a probability, but it defines as proportional to a probability. Because the likelihood function is defined up to a multiplicative constant. The constant is often disregarded in practice since it does not affect the relative likelihoods of different parameter values.

Maximum likelihood estimation (MLE) is a statistical technique for estimating the parameters of a probability distribution using observed data. The goal of MLE is to find the values of the parameters that maximize the likelihood function, which is the probability of observing the data given the model and its parameters

The likelihood function is defined as the joint probability density function or mass function of the observed data, treated as a function of the unknown parameters. If a set of observations $X = \{x_1, x_2, ..., x_n\}$ are assumed to be independently and identically distributed (IID) according to a probability distribution with an unknown set of parameters θ, then the likelihood function is generally expressed as:

$$L(\theta|x) = f(x|\theta) = f(x_1, x_2, ..., x_n|\theta),$$

where $L(\theta|x)$ is the likelihood function, θ is the vector of unknown parameters, x is the observed data, and f is the probability density function or mass function of the distribution.

The maximum likelihood estimate is the value of the parameter vector that maximizes the likelihood function. This is typically found by taking the derivative of the likelihood function with respect to the parameter vector and setting it equal to zero:

$$\frac{\partial L(\theta|x)}{\partial \theta} = 0.$$

Solving this equation for θ gives the maximum likelihood estimate for the parameter vector, and taking the second derivative would verify it is indeed a global maximum.

2.4.1 Likelihood Methods

There are various methods to determine the likelihood function for maximum likelihood estimation (MLE) depending on the nature of the data and the problem being solved. It is important to carefully choose the appropriate likelihood method and to be aware of any assumptions underlying the method. Some of the commonly used likelihood procedures are mainly focusing discrete, continuous, censored or dependent data types.

In the discrete case, the likelihood function for discrete data is calculated by taking the product of the probability mass function of the observed data. For example, if we have a sample of counts following a Poisson distribution, then the likelihood function is calculated as the product of the Poisson probability functions for each count.

For continuous data, the likelihood function is calculated by taking the product of the probability density function of the observed data. For example, if we have a random sample of data following a normal distribution, then the likelihood function is calculated as the product of the normal density functions for each observation.

In survival analysis, where the time-to-event data may be censored, the likelihood function is calculated using the Kaplan-Meier estimator or the Nelson-Aalen estimator, which provides an estimate of the survival function based on the censored data.

When the data is not independent, the likelihood function can be calculated using the conditional probability of the observed data given the value of the parameter. This is known as the conditional likelihood function and is often used in clustered or longitudinal data analysis.

The marginal likelihood function, also known as the integrated likelihood, is obtained by integrating the likelihood function over the parameter space, weighted by the prior distribution of the parameters. The marginal likelihood is commonly used for model selection, where different models are compared based on their respective marginal likelihood values.

The profile likelihood function is a variation of the likelihood function in which the likelihood is maximized with respect to one or more parameters while holding the other parameters constant. This is used to estimate the standard error of the maximum likelihood estimate, construct confidence intervals and perform hypothesis testing for the parameter of interest.

2.4.2 Methods for finding Maximum Likelihood Estimators

Depending on the nature of the data and likelihood function, Maximum Likelihood Estimators (MLEs) can be determined in different ways.

For some probability distributions, the likelihood function can be expressed in a closed-form analytical expression. In such cases, the MLEs can be sometimes obtained by solving the equation for the first derivative of the likelihood function with respect to the parameter(s) of interest and equating it to zero. The resulting equations can then be solved analytically to obtain the MLEs.

When the likelihood function is not available in a closed-form expression or when analytical solutions are not possible, numerical optimization methods can be used to find the MLEs. Commonly used optimization methods include gradient descent, Newton-Raphson, and quasi-Newton methods. These methods involve starting with an initial guess for the parameters and iteratively updating the parameter values until convergence to the maximum likelihood estimate.

The Expectation-Maximization (EM) algorithm, introduced by Dempster et al. (1977), is a widely used iterative algorithm for finding MLEs in cases where the likelihood function involves latent variables in an incomplete data setup. It involves two steps in each iteration i.e., the E-step, where the expectation of the latent variables is computed given the current estimate of the parameters, and the M-step, where the parameters are updated to maximize the expected complete log-likelihood function.

Monte Carlo methods: In some cases, such as in Bayesian inference, the MLEs may not be the focus of interest, but rather the posterior distribution of the parameters. In such cases, Monte Carlo methods such as Markov Chain Monte Carlo (MCMC) can be used to sample from the posterior distribution and estimate the parameters of interest.

It is important to carefully choose the appropriate method for finding MLEs depending on the nature of the problem, the likelihood function, and the computational resources available.

Chapter 3

Model and data description

In competing risk models, the assumption of independence between causes of failures is not always valid, especially when the causes share common factors or dependencies. In such cases, it is more appropriate to use bivariate or multivariate distributions for the lifetimes under the causes.

Various researchers have proposed different generalizations of exponential distributions to bivariate or multivariate models. Gumbel (1960) proposed the use of the bivariate Gumbel distribution to model the dependence between the two lifetimes. This model assumes that the two lifetimes follow independent exponential distributions, and the dependence structure between them is captured by a copula function.

Henrich and Jensen (1995) proposed a bivariate model based on the Marshall-Olkin distribution, which is a generalization of the exponential distribution. This model assumes that the two lifetimes follow Marshall-Olkin distributions with different parameters and a shared dependence parameter.

Block and Basu (1974) introduced the Farlie-Gumbel-Morgenstern (FGM) bivariate exponential distribution, which is a copula-based model that allows for both positive and negative dependence between the two lifetimes.

Downton (1970) proposed the use of the bivariate Weibull distribution to model the dependence between two lifetimes. This model assumes that the two lifetimes follow Weibull distributions with different parameters and a shared dependence parameter.

Basu (1988) further discussed multivariate exponential distributions and their applications in reliability. These models allow for the modeling of dependence between more than two lifetimes.

Overall, the choice of which bivariate or multivariate distribution to use depends on the specific nature of the dependence between the lifetimes and the assumptions made about the underlying distribution. These models have important applications in reliability, survival analysis, and risk management, where understanding the dependence between multiple factors affecting a system is crucial.

3.1 Downton's bivariate Weibull distribution

Downton (1970) introduced a bivariate exponential distribution, which is commonly referred to as Downton's bivariate exponential distribution. It is a popular choice for modelling the lifetime distribution of systems that are subject to competing risks. The distribution assumes that the two failure times of the system due to each cause are exponentially distributed and that the failure times due to different causes are independent. The bivariate exponential distribution has two parameters that represent the failure rates for the two causes.

Downton's bivariate exponential distribution has been extensively used in reliability and survival analysis and has many applications in various fields such as engineering, medical sciences, and economics. One of the advantages of using this distribution is that it allows for the estimation of the probability of each cause of failure, as well as the probability of failure due to any cause. Additionally, it provides a useful tool for studying the relationship between the causes of failure and the system performance.

Downton's bivariate exponential (DBE) distribution is given as follows. If a random vector (Y_1, Y_2) follows the DBE distribution, the corresponding joint probability density function (PDF) is given by:

$$f_{Y_1,Y_2}(y_1,y_2) = \frac{1}{\theta_1\theta_2(1-\rho)} exp\left\{\frac{1}{1-\rho}\left(\frac{y_1}{\theta_1}+\frac{y_2}{\theta_2}\right)\right\} I_0\left(\frac{2(\rho y_1 y_2)^{\frac{1}{2}}}{(1-\rho)\sqrt{\theta_1\theta_2}}\right), y_1, y_2 > 0,$$

where $\theta_1, \theta_2 > 0$, $0 \leq \rho < 1$ and $I_0(\cdot)$ is the modified Bessel function of the first kind of zero.

Downton's bivariate exponential distribution is denoted by $(Y_1, Y_2) \sim DBE(\theta_1, \theta_2, \rho)$. Also Y_1 and Y_2 have exponential marginal distributions. DBE distribution can be used in situations where the hazard rates of the marginal distributions can be assumed (or verified) to be constants. However, for the bivariate data which does not have unimodal marginal probability density functions or if the hazard functions are not constant, it does not seem to be suitable to fit the data. In this case, univariate Weibull marginals are more appropriate to fit models with non-constant marginal hazard rates.

For the $(Y_1, Y_2) \sim DBE(\theta_1, \theta_2, \rho)$, consider the transformations $X_1 = Y_1^{\alpha_1}$ and $X_2 = Y_2^{\alpha_2}$, where $\alpha_1, \alpha_2 > 0$. The joint density of the transformed random variables X_1 and X_2 can be expressed as follows:

$$f_{X_1,X_2}(x_1,x_2) = \frac{\eta_1\eta_2 x_1^{\eta_1-1} x_2^{\eta_2-1}}{\theta_1\theta_2(1-\rho)} exp\left\{-\frac{1}{1-\rho}\left(\frac{x_1^{\eta_1}}{\theta_1}+\frac{x_2^{\eta_2}}{\theta_1}\right)\right\} I_0\left(\frac{2(\rho x_1^{\eta_1} x_2^{\eta_2})^{1/2}}{(1-\rho)\sqrt{\theta_1\theta_2}}\right),$$

$$x_1, x_2 > 0. \quad (3.1.1)$$

where $\theta_1, \theta_2, \eta_1, \eta_2 \; > \; 0$ while, $\eta_1 = 1/\alpha_1$ and $\eta_1 = 1/\alpha_1$. Hence by applying the power transformation to the variables Y_1 and Y_2, we can achieve a transformed bivariate distribution known as Downton's bivariate Weibull (DBW) distribution. Downton's bivariate Weibull distribution allows for flexibility in accommodating non-constant marginal hazard rates, making it a more suitable choice for analyzing the joint behaviour of variables with varying hazard rates.

The Downton's bivariate Weibull (DBW) distribution can be denoted by $(X_1, X_2) \sim DBW(\eta_1, \theta_1, \eta_2, \theta_2, \rho)$ with five parameters. The marginal distributions of X_1 and X_2 are Weibull distributions with parameters (η_1, θ_1) and (η_2, θ_2), respectively. With this, η_1 and η_2 are shape parameters and θ_1 and θ_2 are the scale parameters of the two univariate marginal Weibull distributions.

The PDF of the univariate Weibull distribution is given by:

$$f_{X_i}(x_i) = \frac{\eta_i}{\theta_i} x_i^{\eta_i - 1} exp\left(-\frac{x_i^{\eta_i}}{\theta_i}\right), x_i > 0, \tag{3.1.2}$$

The CDF of the univariate Weibull distribution is as follows:

$$F_{X_i}(x_i) = 1 - exp\left(-\frac{x_i^{\eta_i}}{\theta_i}\right), x_i > 0. \tag{3.1.3}$$

where, i= 1,2 for each random variable.

Utilizing the infinite series representation of the Bessel function presented in Gradshteyn and Ryzhik (2000),

$$I_0(z) = \sum_{k=0}^{\infty} \frac{1}{(k!)^2} \left(\frac{z}{2}\right)^{2k},$$

Davies, Mitra and Volterman (submitted) obtained a representation of the PDF of the DBW distribution as follows:

$$f_{X_1,X_2}(x_1,x_2) = \eta_1\eta_2 exp\left\{-\frac{1}{1-\rho}\left(\frac{x_1^{\eta_1}}{\theta_1}+\frac{x_2^{\eta_2}}{\theta_2}\right)\right\}$$

$$\times\sum_{k=0}^{\infty}\frac{\rho^k}{(k!)^2(1-\rho)^{2k+1}}\frac{x_1^{-1+(k+1)\eta_1}x_2^{-1+(k+1)\eta_2}}{(\theta_1\theta_2)^{k+1}}, x_1, x_2 > 0, \quad (3.1.4)$$

The corresponding joint cumulative density function(CDF) of X_1 and X_2 is as follows:

$$F_{X_1,X_2}(x_1,x_2) = 1 - exp\left(\frac{-x_1^{\eta_1}}{\theta_1}\right)Q_1\left(\sqrt{\frac{2}{1-\rho}}\frac{x_2^{\eta_2/2}}{\sqrt{\theta_2}}, \sqrt{\frac{2\rho}{1-\rho}}\frac{x_1^{\eta_1/2}}{\sqrt{\theta_1}}\right)$$

$$- exp\left(\frac{-x_2^{\eta_2}}{\theta_2}\right)\left[1-Q_1\left(\sqrt{\frac{2\rho}{1-\rho}}\frac{x_2^{\eta_2/2}}{\sqrt{\theta_2}}, \sqrt{\frac{2}{1-\rho}}\frac{x_1^{\eta_1/2}}{\sqrt{\theta_1}}\right)\right], \quad (3.1.5)$$

where $Q_1(\cdot,\cdot)$ is the first order Marcum's Q-function. Then the bivariate survival function for DBW can be found using:

$$S_{X_1,X_2}(x_1,x_2) = 1 - F_{X_1}(x_1) - F_{X_2}(x_2) + F_{X_1,X_2}(x_1,x_2).$$

Therefore the bivariate survival function is as follows:

$$S_{X_1,X_2}(x_1,x_2) = exp\left(\frac{-x_1^{\eta_1}}{\theta_1}\right)\left[1-Q_1\left(\sqrt{\frac{2}{1-\rho}}\frac{x_2^{\eta_2/2}}{\sqrt{\theta_2}}, \sqrt{\frac{2\rho}{1-\rho}}\frac{x_1^{\eta_1/2}}{\sqrt{\theta_1}}\right)\right]$$

$$+ exp\left(\frac{-x_2^{\eta_2}}{\theta_2}\right)Q_1\left(\sqrt{\frac{2\rho}{1-\rho}}\frac{x_2^{\eta_2/2}}{\sqrt{\theta_2}}, \sqrt{\frac{2}{1-\rho}}\frac{x_1^{\eta_1/2}}{\sqrt{\theta_1}}\right). \quad (3.1.6)$$

The conditional distributions can be obtained by using the joint density and the marginal densities. The PDFs of the conditional distributions of X_1 given $X_2 = x_2$ and X_2 given $X_1 = x_1$, respectively are given by:

$$f_{X_1|X_2=x_2}(x_1) = \frac{\eta_1 x_1^{\eta_1 - 1}}{\theta_1(1-\rho)} exp\left\{-(A_1 + A_2)\right\} I_0\left(\frac{2(\rho x_1^{\eta_1} x_2^{\eta_2})^{1/2}}{(1-\rho)\sqrt{\theta_1\theta_2}}\right) x_1 > 0 \qquad (3.1.7)$$

and,

$$f_{X_2|X_1=x_1}(x_2) = \frac{\eta_2 x_2^{\eta_2 - 1}}{\theta_2(1-\rho)} exp\left\{-(B_1 + B_2)\right\} I_0\left(\frac{2(\rho x_1^{\eta_1} x_2^{\eta_2})^{1/2}}{(1-\rho)\sqrt{\theta_1\theta_2}}\right) x_2 > 0 \qquad (3.1.8)$$

where

$$A_1 = \frac{1}{1-\rho}\frac{x_1^{\eta_1}}{\theta_1}, \quad A_2 = \frac{\rho}{1-\rho}\frac{x_2^{\eta_2}}{\theta_2}, \quad B_1 = \frac{1}{1-\rho}\frac{x_2^{\eta_2}}{\theta_2} \quad \text{and} \quad B_2 = \frac{\rho}{1-\rho}\frac{x_1^{\eta_1}}{\theta_1}.$$

Similarly, if $(X_1, X_2) \sim DBW(\eta_1, \theta_1, \eta_2, \theta_2, \rho)$, expressions for the CDFs of the conditional distributions of X_1 given $X_2 = x_2$, and X_2 given $X_1 = x_1$ are given by:

$$F_{X_1|X_2=x_2}(x_1) = A_1 exp(-A_2)\int_0^1 exp(-A_1 t) I_0(2\sqrt{A_1 A_2 t}))dt, \qquad (3.1.9)$$

and

$$F_{X_2|X_1=x_1}(x_2) = B_1 exp(-B_2)\int_0^1 exp(-B_1 t) I_0(2\sqrt{B_1 B_2 t}))dt, \qquad (3.1.10)$$

repectively.

Finally, the dependence parameter ρ captures the dependence between X_1 and X_2. But ρ does not refer to Pearson's correlation coefficient between X_1 and X_2 for the DBW distribution. For $(X_1, X_2) \sim DBW(\eta_1, \theta_1, \eta_2, \theta_2, \rho)$, X_1 and X_2 are independent if and only if $\rho = 0$. When $0 < \rho < 1$, X_1 and X_2 are correlated. The product-moment correlation coefficient (PMC) of X_1 and X_2 of the DBW distribution is given by:

$$\rho_{PMC} = \frac{\Gamma\left(1+\frac{1}{\eta_1}\right)\Gamma\left(1+\frac{1}{\eta_2}\right)\left[(1-\rho)^{1+\frac{1}{\eta_1}+\frac{1}{\eta_2}} \times \ {}_2F_1\left(1+\frac{1}{\eta_1}, 1+\frac{1}{\eta_2}; 1; \rho\right) - 1\right]}{\sqrt{\Gamma\left(1+\frac{2}{\eta_1}\right)\left\{\Gamma\left(1+\frac{1}{\eta_1}\right)\right\}^2}\sqrt{\Gamma\left(1+\frac{2}{\eta_2}\right)\left\{\Gamma\left(1+\frac{1}{\eta_2}\right)\right\}^2}}, \tag{3.1.11}$$

where ${}_2F_1(\cdot, \cdot)$ is the generalized hypergeometric function.

The generalized hypergeometric function can be obtained as:

$$E\left(X_1^{m_1} X_2^{m_2}\right) = (1-\rho)^{1+\frac{m_1}{\eta_1}+\frac{m_2}{\eta_2}} \theta_1^{\frac{m_1}{\eta_1}} \theta_2^{\frac{m_2}{\eta_2}} \Gamma\left(1+\frac{m_1}{\eta_1}\right)\Gamma\left(1+\frac{m_2}{\eta_2}\right) \times {}_2F_1\left(1+\frac{m_1}{\eta_1}, 1+\frac{m_2}{\eta_2}; 1; \rho\right). \tag{3.1.12}$$

The product-moment correlation is a measure of the linear association between two variables X_1 and X_2. The Eq.3.1.11 for PMC clearly indicates that it depends only on the shape parameters, η_1 and η_2, and the dependence parameter, but not on the scale parameters, θ_1 and θ_2. This means that the scale parameters do not affect the correlation coefficient for the DBW distribution. Therefore, while PMC can still be used to measure the association between X_1 and X_2 for the DBW distribution, it is important to note that it is not the same as the correlation coefficient. Instead, the correlation coefficient should be calculated directly from the joint distribution of X_1 and X_2, which is given by the DBW bivariate distribution.

3.2 Data Description

3.2.1 Bivariate data generation

In this scenario, unit/individuals are considered to experience two dependent competing risks. The time to failure for an individual under these risks, denoted as (X_1, X_2), are assumed to follow a bivariate Weibull distribution known as the Downton's bivariate Weibull (DBW) distribution. The parameters of this distribution are η_1, θ_1, η_2, θ_2, and ρ. To generate bivariate data observations (x_1, x_2), we utilize the $DBW(\eta_1, \theta_1, \eta_2, \theta_2, \rho)$ distribution.

Davies, Mitra and Volterman(submitted) describe the process of generating random variables from the distribution(when $0 < \rho < 1$) as follows:

STEP 1: Generate x_1 from Weibull(η_1, θ_1).

STEP 2: Transform x_1 into z_1, where $z_1 = \frac{\rho}{1-\rho}\frac{x_1^{\eta_1}}{\theta_1}$.

STEP 3: Generate z_2 from the conditional distribution with CDF $F_{Z_2|Z_1=z_1}(z_2) = z_2 exp(-z_1) \int_0^1 exp(-z_2 t) I_0(2\sqrt{z_1 z_2 t}) dt$.

STEP 4: Obtain x_2 using the transformation $x_2 = (\theta_2(1-\rho)z_2)^{1/\eta_2}$.

The above algorithm should only be used when $0 < \rho < 1$. For $\rho = 0$, x_1 and x_2 can be generated directly from their respective marginal distributions.

3.2.2 Data description

In a lifetime experiment where a unit/individual can experience an event with two dependent competing risks. Let us consider n subjects and assume that the units/individuals do not experience both risks simultaneously. In this scenario, the n

subjects will be considered as complete data. Also suppose that the times to failure of a subject under these risks, (X_1, X_2), follows a DBW distribution with parameters $(\theta_2, \theta_1, \eta_1, \eta_2, \rho)$.

For a competing risks model, we define $Y = min(X_1, X_2)$ as the time to the first failure (latent failure model), and the observations will be in the following form:

$$D = \{(y_1, \delta_1^*), \cdots\cdots, (y_n, \delta_n^*)\}$$

where $y_i = min(x_{1i}, x_{2i})$ is the observed minimum time of failure for the i^{th} subject, and δ_i^* is the corresponding event indicator for specifying the causes of failure. The observed failure time y_i is informative for both risks. The competing risks model includes two types of failures: type 1, which occurs when $x_{1i} < x_{2i}$, and type 2, which occurs when $x_{2i} < x_{1i}$. Then the δ_i^* will represent as:

$$\delta_i^* = \begin{cases} 1 \; if \; x_{i1} < x_{i2} \\ 2 \; if \; x_{i1} > x_{i2} \end{cases}$$

for $i = 1, ..., n$.

To estimate the parameters of the bivariate DBW distribution and perform inference, the joint likelihood function of the observed data D needs to be maximized. However, the joint likelihood function (as shown in the next chapter), involves a complex integral that is analytically intractable. Hence, numerical methods can be used to estimate the parameters and here will be discussed.

Chapter 4

Likelihood inference

The likelihood function of the bivariate competing risks model introduced by Lawless (1982) can be utilized for the purpose of dependency between multiple competing risks. In their work, Feizjavadian and Hashemi (2015) present a likelihood inference approach for the bivariate Weibull distribution introduced by Marshall and Olkin (1967). The method involves the use of the likelihood function of the bivariate competing risks model introduced by Lawless (1982), which is applied to the Marshall-Olkin bivariate Weibull distribution. The obtained likelihood function of the bivariate competing risks model with lifetime random variables X and Y are as follows:

$$L \propto \prod_{i=1}^{m} \left[f_{X,Y}\left(t_{xi}, t_{yi}\right)\right]^{\delta_{i0}} \left[-\frac{\partial}{\partial x} S_{X,Y}\left(x,y\right)|_{x=t_{xi}, y=t_{yi}}\right]^{\delta_{i1}} \left[-\frac{\partial}{\partial y} S_{X,Y}\left(x,y\right)|_{x=t_{xi}, y=t_{yi}}\right]^{\delta_{i2}} \times \left[S_{X,Y}\left(t_{xi}, t_{yi}\right)\right]^{1-\delta_{i0}-\delta_{i1}-\delta_{i2}}, \quad (4.0.1)$$

where $f_{X,Y}(\cdot,\cdot)$ is a joint probability density function (PDF) and $S_{X,Y}(\cdot,\cdot)$ is a joint survival function corresponding to $f_{X,Y}(\cdot,\cdot)$; also, δ_{ij} , $j = 0, 1, 2$ are event indicators for specifying the causes of failure and $i = 1, ..., n$ are number of individuals in the sample. Here, $j = 1$ indicates type 1 cause of failure, which occurs when $x < y$, and

$j = 2$ is the type 2, which occurs when $y < x$. $j = 0$ is the tie case when $x = y$. The approach presented by Feizjavadian and Hashemi (2015) provides a useful framework for analyzing lifetime data under the DBW distribution.

4.1 Inference based on complete bivariate data

Based on observed bivariate data, maximum likelihood estimates (MLEs) for the parameters of the DBW distribution will be obtained considering the complete data case. We consider a bivariate competing risks model in which two random variables X_1, X_2 indicate the time to failure under the first and second risks, respectively and (X_1, X_2) follows a $DBW(\eta_1, \theta_1, \eta_2, \theta_2, \rho)$ distribution.

Using the Eq.4.0.1, the joint likelihood function for the dependent competing risks model for the complete dataset can be rewritten as:

$$L \propto \prod_{i=1}^{m} \left[-\frac{\partial}{\partial x_1} S_{X_1,X_2}(x_1, x_2) \,|_{(y_i,y_i)} \right]^{\delta_{i1}} \left[-\frac{\partial}{\partial x_2} S_{X_1,X_2}(x_1, x_2) \,|_{(y_i,y_i)} \right]^{\delta_{i2}} \tag{4.1.1}$$

where the two terms are the probabilities of the failure of a subject by the first and second risks at y_i respectively. Also, δ_{ij}, $j = 1, 2$ are event indicators for specifying the causes of failure.

Since the DBW distribution assumes no ties, the likelihood function expressed in Eq.4.1.1 only considers the non-tied observations. Any tied observations would need to be dealt with separately, using methods such as tied failure time models. In the context of the bivariate competing risks model with DBW distribution, ignoring the tied observations in the likelihood function is a valid approach since it simplifies the computation and the likelihood can still be maximized to obtain the MLEs of the parameters. However, it is important to note that ignoring tied observations can potentially lead to biased estimation if the proportion of tied observations is large. In such cases, it may be more appropriate to use methods that can handle tied data or to ensure that the sample size is sufficiently large so that the proportion of tied observations is negligible.

To simplify the analysis, the joint survival function Eq.3.1.6 is rewritten in a new form as follows:

$$S_{X_1,X_2}(x_1,x_2) = exp\left(\frac{-x_1^{\eta_1}}{\theta_1}\right)\left[1 - Q_1\left(Ax_2^{\eta_2/2}, Bx_1^{\eta_1/2}\right)\right] + exp\left(\frac{-x_2^{\eta_2}}{\theta_2}\right)Q_1\left(Cx_2^{\eta_2/2}, Dx_1^{\eta_1/2}\right) \tag{4.1.2}$$

where

$$A = \sqrt{\frac{2}{(1-\rho)\theta_2}}, \quad B = \sqrt{\frac{2\rho}{(1-\rho)\theta_1}}, \quad C = \sqrt{\frac{2\rho}{(1-\rho)\theta_2}} \quad and \quad D = \sqrt{\frac{2}{(1-\rho)\theta_1}}.$$

The first term of the likelihood function in Eq.4.1.1 can be obtained by differentiating the bivariate survival function for DBW in Eq.4.1.2 by x_1. Then, the partial derivative of $S_{X_1,X_2}\left(x_1,x_2\right)|_{(y_i,y_i)}$ with respect to x_1 is obtained as:

$$\begin{aligned}\frac{\partial}{\partial x_1}S_{X_1,X_2}\left(x_1,x_2\right)|_{(y_i,y_i)} = -\frac{\eta_1}{\theta_1}x_1^{\eta_1-1}exp\left(\frac{-x_1^{\eta_1}}{\theta_1}\right)\left[1 - Q_1\left(Ax_2^{\eta_2/2}, Bx_1^{\eta_1/2}\right)\right] \\ -exp\left(\frac{-x_1^{\eta_1}}{\theta_1}\right)\left[\frac{\partial}{\partial x_1}Q_1\left(Ax_2^{\eta_2/2}, Bx_1^{\eta_1/2}\right)\right] \\ +exp\left(\frac{-x_2^{\eta_2}}{\theta_1}\right)\left[\frac{\partial}{\partial x_1}Q_1\left(Cx_2^{\eta_2/2}, Dx_1^{\eta_1/2}\right)\right].\end{aligned} \tag{4.1.3}$$

Pratt (1968) obtained the partial differentials of Marcum's Q function. As the article describes, the obtained partial derivative of $Q_1(a,b)$ with respect to a can be delivered by:

$$\frac{\partial}{\partial a}Q_1(a,b) = a\left(\frac{b}{a}\right)e^{-(a^2+b^2)/2}I_1(ab),$$

where $I_1(z)$ is the Bessel function of the first kind.

Then the partial derivative of $Q_1(a,b)$ with respect to b is given by:

$$\frac{\partial}{\partial b}Q_1(a,b) = -be^{-(a^2+b^2)/2}I_0(ab)$$

Referencing the above to equations, the partial derivative of $Q_1\left(Ax_2^{\eta_2/2}, Bx_1^{\eta_1/2}\right)$ with respect to x_1 can be obtained as:

$$\frac{\partial}{\partial x_1}Q_1\left(Ax_2^{\eta_2/2}, Bx_1^{\eta_1/2}\right) = -\frac{B^2\eta_1}{2}x_1^{\eta_1-1}exp\left[-\frac{1}{2}\left(A^2x_2^{\eta_2}+B^2x_1^{\eta_1}\right)\right]I_0\left(Ax_2^{\eta_2/2}Bx_1^{\eta_1/2}\right). \tag{4.1.4}$$

Similarly, the partial derivative of $Q_1\left(Cx_2^{\eta_2/2}, Dx_1^{\eta_1/2}\right)$ with respect to x_1 can be generated as:

$$\frac{\partial}{\partial x_1}Q_1\left(Cx_2^{\eta_2/2}, Dx_1^{\eta_1/2}\right) = -\frac{D^2\eta_1}{2}x_1^{\eta_1-1}exp\left[-\frac{1}{2}\left(C^2x_2^{\eta_2}+D^2x_1^{\eta_1}\right)\right]I_0\left(Cx_2^{\eta_2/2}Dx_1^{\eta_1/2}\right). \tag{4.1.5}$$

The partial derivative of $S_{X_1,X_2}\left(x_1,x_2\right)|_{(y_i,y_i)}$ with respect to x_2 is derived to obtain the second term of the likelihood function in Eq.3.3.2.

$$\begin{aligned}\frac{\partial}{\partial x_2} S_{X_1,X_2}\left(x_1,x_2\right)|_{(y_i,y_i)} = -exp\left(\frac{-x_1^{\eta_1}}{\theta_1}\right)\left[\frac{\partial}{\partial x_2} Q_1\left(Ax_2^{\eta_2/2}, Bx_1^{\eta_1/2}\right)\right] \\ -\frac{\eta_2}{\theta_2} x_2^{\eta_2-1} exp\left(\frac{-x_1^{\eta_1}}{\theta_1}\right) Q_1\left(Cx_2^{\eta_2/2}, Dx_1^{\eta_1/2}\right) \\ +exp\left(\frac{-x_2^{\eta_2}}{\theta_1}\right)\left[\frac{\partial}{\partial x_2} Q_1\left(Cx_2^{\eta_2/2}, Dx_1^{\eta_1/2}\right)\right].\end{aligned} \tag{4.1.6}$$

The partial derivative of $Q_1\left(Ax_2^{\eta_2/2}, Bx_1^{\eta_1/2}\right)$ with respect to x_2 is given by:

$$\frac{\partial}{\partial x_2} Q_1\left(Ax_2^{\eta_2/2}, Bx_1^{\eta_1/2}\right) = -\frac{AB\eta_2}{2} x_1^{\eta_1/2} x_2^{\eta_2/2-1} exp\left[-\frac{1}{2}\left(A^2 x_2^{\eta_2} + B^2 x_1^{\eta_1}\right)\right] I_1\left(Ax_2^{\eta_2/2} Bx_1^{\eta_1/2}\right). \tag{4.1.7}$$

Similarly, the partial derivative of $Q_1\left(Cx_2^{\eta_2/2}, Dx_1^{\eta_1/2}\right)$ with respect to x_2 can be generated as:

$$\frac{\partial}{\partial x_2} Q_1\left(Cx_2^{\eta_2/2}, Dx_1^{\eta_1/2}\right) = -\frac{CD\eta_2}{2} x_1^{\eta_1/2} x_2^{\eta_2/2-1} exp\left[-\frac{1}{2}\left(C^2 x_2^{\eta_2} + D^2 x_1^{\eta_1}\right)\right] I_1\left(Cx_2^{\eta_2/2} Dx_1^{\eta_1/2}\right). \tag{4.1.8}$$

Finally, all these expressions can be used in deriving the likelihood function for the DBW distribution. Ideally, upon establishing the likelihood function expression for the maximum likelihood estimators(MLEs) can be produced. Clearly, however, the likelihood function involves complex integrals that are analytically rebellious to deal with. Therefore, as mentioned earlier it is more reasonable to consider numerical optimization methods to compute the maximum likelihood estimates.

The maximum likelihood estimates of the unknown parameters can be computed by maximizing the logarithm of the likelihood function for the complete dataset. In order to maximize the logarithm of the likelihood function numerically in R, initial values of the parameters are required.

4.2 Choice of initial values

To obtain maximum likelihood estimates (MLEs) of the parameters in a statistical model, one typically needs to provide initial values for the algorithm to start searching for the optimal values. However, choosing appropriate initial values can be a challenging task. In this context, a method for selecting initial parameter values is proposed by Davies, Mitra, and Volterman (submitted) based on the observation that the marginal distribution of one of the variables does not depend on certain parameters.

More specifically, if we consider the DBW model where the marginal distribution of X_1 does not depend on parameters θ_2 and η_2, then we can compute closed-form initial values for those parameters. The same applies if the marginal distribution of X_2 does not depend on θ_1 and η_1. These initial values can be obtained by solving the corresponding equations that express the marginal distribution in terms of the parameters of interest.

When X follows a Weibull distribution with the PDF as given in Eq.3.1.2,

$$f_X(x) = \frac{\eta}{\theta_i} x^{\eta-1} exp\left(-\frac{x^\eta}{\theta}\right)$$

where $x > 0$, $\eta > 0$ and $\theta > 0$.

If we let $Z = logX$, then the PDF of Z can be derived as follows. Using the transformation of variables formula for probability density functions, we have:

$$f_Z(z) = \frac{1}{\sigma} exp\left[\left(\frac{z-\mu}{\sigma}\right) - exp\left(\frac{z-\mu}{\sigma}\right)\right] \tag{4.2.1}$$

where z is any real number, μ is the location parameter, and σ is the scale parameter, and

$$\mu = \frac{1}{\eta}\, log\theta \quad and \quad \sigma = \frac{1}{\eta}.$$

Eq.4.2.1 is the PDF of an extreme-value distribution, specifically the Gumbel distribution. The Gumbel distribution is a two-parameter distribution that is commonly used in extreme value analysis. It is often used to model the distribution of the maximum or minimum of a large number of samples from a distribution.

In this parameterization, the expected value of the extreme value distribution is:

$$E(Z) = \mu + \gamma\sigma$$

and the variance is:

$$Var(Z) = \frac{\pi^2}{6}\sigma^2,$$

where $\gamma \approx 0.5772$ is Euler's constant.

The given results and relations can be used to obtain approximate estimates for the population parameters by equating the population moments with sample quantities and transforming them back. The estimates for the parameters, namely $\theta_1,\ \theta_2,\ \eta_1,\ and\ \eta_2$, can be obtained as:

$$\theta_i = exp\left(\bar{z}_i\left(\frac{\sqrt{6}}{\pi}s_i\right)^{-1} + \gamma\right)$$

and

$$\eta_i = \left(\frac{\sqrt{6}}{\pi} s_i \right)^{-1},$$

where $i = 1, 2$. Here $z_i = log(x_i)$ and s_i is the standard deviation of z_i.

For the DBW distribution, the dependence parameter ρ, as noted earlier, is not the same as Pearson's correlation coefficient, although the latter can be used as an initial value for ρ. By using the sample Pearson's correlation coefficient as an initial value, one can incorporate the dependence structure of the data into the optimization process. This approach works efficiently in practice as the sample Pearson's correlation coefficient provides a good approximation of the true dependence structure.

Monte Carlo simulation studies have shown that using the sample Pearson's correlation coefficient as an initial value helps in finding the global maximum of the likelihood function. It is worth noting that this approach is not guaranteed to work in all cases, and in some cases, more sophisticated initialization methods may be required. Nonetheless, it is a simple and effective approach that can be used as a starting point for optimizing the likelihood function for the DBW distribution.

This method can be particularly useful when dealing with complex models with several parameters, as it provides a systematic way of obtaining initial values that can lead to faster convergence and more reliable MLEs. However, it is important to keep in mind that these initial values are only a starting point and that the optimization algorithm may still need to explore a large parameter space before finding the optimal values. Therefore, it is we use different initial values and check the sensitivity of the results to ensure the robustness of the method.

Once initial values are found, we choose to use the function `‘optim’`, which is an inbuilt function in R. We specify the method as `"BFGS"` which was developed by Byrd et al. (1995). The `"BFGS"` method is a type of quasi-Newton optimization algorithm, which is commonly used in optimization problems for its efficiency and ability to handle a wide range of functions.

By providing appropriate initial values and the associated log-likelihood function, `‘optim’`uses a quasi-Newton method to iteratively search for the parameter values that maximize the log-likelihood function. Along with the estimated parameter values, `‘optim’` also provides the value of the log-likelihood function at these parameter values and a convergence code indicating whether the optimization algorithm has successfully converged. Therefore, the estimated parameter values obtained through `‘optim’` are the MLEs that maximize the likelihood function for a given data and are of great interest in our statistical inference approach.

Chapter 5

Numerical results

5.1 Simulation study

Monte Carlo simulation is one of the popular methods for evaluating the performance of statistical methods, such as maximum likelihood estimators (MLEs), under different parameter settings and sample sizes. In this study, we carried out a Monte Carlo simulation study to evaluate the performance of the MLEs of parameters θ_1, θ_2, η_1, η_2 and ρ. Without loss of generality, the scale parameters θ_1 and θ_2 of the DBW distribution are fixed at unity. We investigate the effect of varying the shape parameters η_1 and η_2 and the dependence parameter ρ on the performance of the MLEs. Since the shape parameter is constrained to the positive real line, η_1 and η_2 parameters are set at 0.5,1.0,1.5 and 2.0 for the simulations. The dependence parameter is constrained to be between 0 and 1. Then ρ is taken as 0.25 (small), 0.5 (moderate), and 0.75 (high). The specific shape parameter settings were selected in the same way described in Section 2.1.1, while dependence parameter settings are chosen to represent lower, middle and upper levels between the constraints. Each of these setting combinations are then repeated for four different sample sizes namely n = 20, 50, 100, and 200. Each simulation was performed with 1000 replications using the `R` software. In each replication, we generated a random sample from the DBW distribution with the specified parameter values and sample size. We then used the MLEs to estimate the parameter values from the sample and recorded the estimated values and the associated standard errors.

The random samples of observations from the DBW distribution are generated through the algorithm described in Section 3.2.1. The generated lifetimes of the observations (x_1, x_2) are then categorized under two competing risks. The competing risks in this case are considered dependent. Finally, the observations used in the competing risk model of the simulations are obtained as $y_i = min(x_1, x_2)$, where $(i = 1, ..., n)$. Numerical MLEs were then obtained using the direct optimization of the log-likelihood function. The likelihood function is generated as the description of Section 4.1. In this study, we use the `R` programming language to compute the MLEs using the '`optim()`' function, which is a generic optimization function that can be used to minimize or maximize a function. By default '`optim`' performs minimization, therefore in the purpose of maximization of the log-likelihood function, it is transformed to a negative log-likelihood function. The initial values of the parameters required for the optimization were generated based on the method described in Section 4.2.

One challenge that is often faced in numerically calculating MLEs is in the context of constrained parameter spaces. To overcome this, one may consider a transformation that transforms the parameter space into the entire real line. One such transformation is the log. For example, if one needs to find the MLE of theta where $\theta > 0$, one may write the likelihood (log-likelihood) as a function of $\theta^* = log(\theta)$, and find the MLE of θ^*, and then by the invariance property of MLEs, one can easily find the MLE of θ. In our case we transformed θ_1, θ_2, η_1 and η_2 using log transformation and dependence parameter ρ was transformed by logit tranformation.

The performance of the MLEs is evaluated using several measures, such as the mean bias, root mean squared error (RMSE), and coverage probability (CP) of the estimated parameters. The mean bias measures the difference between the true value of a parameter and its estimated value averaged over the replications. The RMSE measures the average magnitude of the differences between the true value of a parameter and its estimated value. The CP measures the proportion of times the true value of a parameter falls within its 95% confidence interval, which is estimated using the asymptotic normality of the MLEs. After obtaining the MLEs for each dataset, we computed the Hessian matrix to estimate the Fisher Information and standard errors of the estimates. The confidence intervals for the estimates were calculated using the standard errors and then used to calculate CP measures.

The estimates of the parameters and evaluating measures of each parameter setting are reported in Tables 5.1 through 5.12. The results of the simulation study indicated that the performance of the MLEs was generally satisfactory for all parameter settings and sample sizes. Generally, the bias of the estimators was small and decreased as the sample size increased. The RMSE of the estimators was also small and decreased as the sample size increased for all the parameter settings. In the case of coverage probability of the confidence intervals, all the parameters except for ρ were close to the nominal level of 95%, indicating that the intervals were well-calibrated. However, for $\rho = 0.25$ value settings (which represents the lower set of ρ values) showed considerable difference in CP compared to the nominal level. Increasing the sample size indicates an improvement in the CP and increases reasonably closer to the nominal level. Overall, the simulation study demonstrates MLEs perform well in estimating the parameters of the DBW distribution, across a wide range of simulated scenarios.

Table 5.1: Estimates, mean bias, root mean square error (RMSE) and coverage probability (CP) for dependent competing risk model parameters using direct optimization method with $\theta_1 = \theta_2 = 1.0$, $\eta_1 = \eta_2 = 0.5$ and $\rho = 0.25$ for each sample size (n)

n	Parameter	True Value	Estimate	Mean Bias	RMSE	95% CP
20	θ_1	1.00	1.0264	0.0264	0.3044	0.986
	θ_2	1.00	1.0022	0.0022	0.3013	0.983
	η_1	0.50	0.5493	0.0493	0.1548	0.964
	η_2	0.50	0.5449	0.0449	0.1486	0.966
	ρ	0.25	0.2977	0.0477	0.2057	0.635
50	θ_1	1.00	1.0004	0.0004	0.1859	0.987
	θ_2	1.00	1.0284	0.0284	0.1959	0.992
	η_1	0.50	0.5186	0.0186	0.0866	0.969
	η_2	0.50	0.5163	0.0163	0.0885	0.967
	ρ	0.25	0.2697	0.0197	0.1739	0.656
100	θ_1	1.00	1.0052	0.0052	0.1186	0.996
	θ_2	1.00	1.0105	0.0105	0.1202	0.998
	η_1	0.50	0.5059	0.0059	0.0512	0.989
	η_2	0.50	0.5041	0.0041	0.0495	0.987
	ρ	0.25	0.2417	-0.0083	0.1445	0.778
200	θ_1	1.00	1.0150	0.0150	0.0942	0.998
	θ_2	1.00	1.0067	0.0067	0.0917	0.999
	η_1	0.50	0.4998	-0.0002	0.0383	0.992
	η_2	0.50	0.5017	0.0017	0.0409	0.978
	ρ	0.25	0.2356	-0.0144	0.1197	0.829

Table 5.2: Estimates, mean bias, root mean square error (RMSE) and coverage probability (CP) for dependent competing risk model parameters using direct optimization method with $\theta_1 = \theta_2 = 1.0$, $\eta_1 = \eta_2 = 0.5$ and $\rho = 0.5$ for each sample size (n)

n	Parameter	True Value	Estimate	Mean Bias	RMSE	95% CP
20	θ_1	1.00	1.0563	0.0563	0.3254	0.985
	θ_2	1.00	1.0328	0.0328	0.3053	0.978
	η_1	0.50	0.5406	0.0406	0.1369	0.968
	η_2	0.50	0.5343	0.0343	0.1299	0.979
	ρ	0.50	0.4537	-0.0463	0.2197	0.621
50	θ_1	1.00	1.0236	0.0236	0.1845	0.991
	θ_2	1.00	1.0422	0.0422	0.1863	0.994
	η_1	0.50	0.5144	0.0144	0.0816	0.975
	η_2	0.50	0.5156	0.0156	0.0827	0.976
	ρ	0.50	0.4646	-0.0354	0.1804	0.734
100	θ_1	1.00	1.0142	0.0142	0.1201	0.996
	θ_2	1.00	1.0199	0.0199	0.1247	0.994
	η_1	0.50	0.5041	0.0041	0.0509	0.980
	η_2	0.50	0.5019	0.0019	0.0478	0.992
	ρ	0.50	0.4481	-0.0519	0.1613	0.867
200	θ_1	1.00	1.0274	0.0274	0.0990	0.999
	θ_2	1.00	1.0230	0.0230	0.0978	0.996
	η_1	0.50	0.4999	-0.0001	0.0371	0.982
	η_2	0.50	0.4997	-0.0003	0.0397	0.982
	ρ	0.50	0.4536	-0.0464	0.1343	0.899

Table 5.3: Estimates, mean bias, root mean square error (RMSE) and coverage probability (CP) for dependent competing risk model parameters using direct optimization method with $\theta_1 = \theta_2 = 1.0$, $\eta_1 = \eta_2 = 0.5$ and $\rho = 0.75$ for each sample size (n)

n	Parameter	True Value	Estimate	Mean Bias	RMSE	95% CP
20	θ_1	1.00	1.0655	0.0655	0.3152	0.987
	θ_2	1.00	1.0490	0.0490	0.3028	0.987
	η_1	0.50	0.5391	0.0391	0.1313	0.978
	η_2	0.50	0.5362	0.0362	0.1309	0.976
	ρ	0.75	0.6509	-0.0991	0.1880	0.724
50	θ_1	1.00	1.0234	0.0234	0.1730	0.995
	θ_2	1.00	1.0320	0.0320	0.1725	0.998
	η_1	0.50	0.5133	0.0133	0.0764	0.975
	η_2	0.50	0.5148	0.0148	0.0780	0.971
	ρ	0.75	0.6802	-0.0698	0.1467	0.840
100	θ_1	1.00	1.0183	0.0183	0.1238	0.997
	θ_2	1.00	1.0234	0.0234	0.1259	0.996
	η_1	0.50	0.5063	0.0063	0.0496	0.978
	η_2	0.50	0.5051	0.0051	0.0477	0.983
	ρ	0.75	0.6817	-0.0683	0.1299	0.891
200	θ_1	1.00	1.0203	0.0203	0.0899	0.995
	θ_2	1.00	1.0186	0.0186	0.0882	0.996
	η_1	0.50	0.5034	0.0034	0.0344	0.968
	η_2	0.50	0.5018	0.0018	0.0349	0.982
	ρ	0.75	0.6952	-0.0548	0.1046	0.901

Table 5.4: Estimates, mean bias, root mean square error (RMSE) and coverage probability (CP) for dependent competing risk model parameters using direct optimization method with $\theta_1 = \theta_2 = 1.0$, $\eta_1 = \eta_2 = 1.0$ and $\rho = 0.25$ for each sample size (n)

n	Parameter	True Value	Estimate	Mean Bias	RMSE	95% CP
20	θ_1	1.00	1.0229	0.0229	0.3000	0.985
	θ_2	1.00	0.9970	-0.0030	0.2979	0.984
	η_1	1.00	1.0982	0.0982	0.3099	0.961
	η_2	1.00	1.0909	0.0909	0.2968	0.963
	ρ	0.25	0.3089	0.0589	0.1783	0.681
50	θ_1	1.00	0.9877	-0.0123	0.1824	0.983
	θ_2	1.00	1.0117	0.0117	0.1859	0.993
	η_1	1.00	1.0451	0.0451	0.1776	0.964
	η_2	1.00	1.0410	0.0410	0.1788	0.967
	ρ	0.25	0.2891	0.0391	0.1358	0.717
100	θ_1	1.00	0.9980	-0.0020	0.1130	0.996
	θ_2	1.00	1.0027	0.0027	0.1149	0.998
	η_1	1.00	1.0152	0.0152	0.1036	0.985
	η_2	1.00	1.0121	0.0121	0.0987	0.991
	ρ	0.25	0.2710	0.0210	0.1078	0.817
200	θ_1	1.00	1.0025	0.0025	0.0876	0.998
	θ_2	1.00	0.9947	-0.0053	0.0853	1.000
	η_1	1.00	1.0069	0.0069	0.0787	0.984
	η_2	1.00	1.0105	0.0105	0.0837	0.977
	ρ	0.25	0.2658	0.0158	0.0800	0.874

Table 5.5: Estimates, mean bias, root mean square error (RMSE) and coverage probability (CP) for dependent competing risk model parameters using direct optimization method with $\theta_1 = \theta_2 = 1.0$, $\eta_1 = \eta_2 = 1.0$ and $\rho = 0.5$ for each sample size (n)

n	Parameter	True Value	Estimate	Mean Bias	RMSE	95% CP
20	θ_1	1.00	1.0475	0.0475	0.3093	0.982
	θ_2	1.00	1.0226	0.0226	0.2933	0.985
	η_1	1.00	1.0826	0.0826	0.2786	0.967
	η_2	1.00	1.0708	0.0708	0.2624	0.974
	ρ	0.50	0.4812	-0.0188	0.1679	0.694
50	θ_1	1.00	1.0009	0.0009	0.1718	0.994
	θ_2	1.00	1.0192	0.0192	0.1677	0.987
	η_1	1.00	1.0365	0.0365	0.1619	0.963
	η_2	1.00	1.0400	0.0400	0.1682	0.956
	ρ	0.50	0.4963	-0.0037	0.1179	0.822
100	θ_1	1.00	1.0034	0.0034	0.1136	0.993
	θ_2	1.00	1.0088	0.0088	0.1168	0.991
	η_1	1.00	1.0121	0.0121	0.1029	0.971
	η_2	1.00	1.0078	0.0078	0.0956	0.984
	ρ	0.50	0.4930	-0.0070	0.0934	0.867
200	θ_1	1.00	1.0086	0.0086	0.0880	0.997
	θ_2	1.00	1.0032	0.0032	0.0863	0.996
	η_1	1.00	1.0089	0.0089	0.0756	0.973
	η_2	1.00	1.0092	0.0092	0.0805	0.965
	ρ	0.50	0.4974	-0.0026	0.0674	0.931

Table 5.6: Estimates, mean bias, root mean square error (RMSE) and coverage probability (CP) for dependent competing risk model parameters using direct optimization method with $\theta_1 = \theta_2 = 1.0$, $\eta_1 = \eta_2 = 1.0$ and $\rho = 0.75$ for each sample size (n)

n	Parameter	True Value	Estimate	Mean Bias	RMSE	95% CP
20	θ_1	1.00	1.0593	0.0593	0.3117	0.986
	θ_2	1.00	1.0388	0.0388	0.2963	0.983
	η_1	1.00	1.0795	0.0795	0.2650	0.974
	η_2	1.00	1.0758	0.0758	0.2639	0.972
	ρ	0.75	0.6784	-0.0716	0.1308	0.744
50	θ_1	1.00	1.0087	0.0087	0.1616	0.993
	θ_2	1.00	1.0176	0.0176	0.1606	0.993
	η_1	1.00	1.0239	0.0239	0.1463	0.970
	η_2	1.00	1.0265	0.0265	0.1444	0.972
	ρ	0.75	0.7084	-0.0416	0.0864	0.864
100	θ_1	1.00	1.0075	0.0075	0.1175	0.996
	θ_2	1.00	1.0129	0.0129	0.1179	0.994
	η_1	1.00	1.0150	0.0150	0.1006	0.968
	η_2	1.00	1.0120	0.0120	0.0962	0.976
	ρ	0.75	0.7193	-0.0307	0.0654	0.909
200	θ_1	1.00	1.0073	0.0073	0.0803	0.997
	θ_2	1.00	1.0052	0.0052	0.0782	0.998
	η_1	1.00	1.0074	0.0074	0.0657	0.973
	η_2	1.00	1.0056	0.0056	0.0680	0.978
	ρ	0.75	0.7305	-0.0195	0.0448	0.938

Table 5.7: Estimates, mean bias, root mean square error (RMSE) and coverage probability (CP) for dependent competing risk model parameters using direct optimization method with $\theta_1 = \theta_2 = 1.0$, $\eta_1 = \eta_2 = 1.5$ and $\rho = 0.25$ for each sample size (n)

n	Parameter	True Value	Estimate	Mean Bias	RMSE	95% CP
20	θ_1	1.00	1.0222	0.0222	0.3002	0.985
	θ_2	1.00	0.9978	-0.0022	0.2983	0.982
	η_1	1.50	1.6474	0.1474	0.4665	0.964
	η_2	1.50	1.6346	0.1346	0.4458	0.963
	ρ	0.25	0.3060	0.0560	0.1660	0.667
50	θ_1	1.00	0.9868	-0.0132	0.1823	0.985
	θ_2	1.00	1.0115	0.0115	0.1860	0.994
	η_1	1.50	1.5673	0.0673	0.2694	0.966
	η_2	1.50	1.5593	0.0593	0.2711	0.966
	ρ	0.25	0.2834	0.0334	0.1235	0.732
100	θ_1	1.00	0.9984	-0.0016	0.1129	0.996
	θ_2	1.00	1.0032	0.0032	0.1150	0.998
	η_1	1.50	1.5225	0.0225	0.1558	0.984
	η_2	1.50	1.5175	0.0175	0.1485	0.991
	ρ	0.25	0.2671	0.0171	0.0948	0.799
200	θ_1	1.00	1.0040	0.0040	0.0875	0.997
	θ_2	1.00	0.9962	-0.0038	0.0849	0.999
	η_1	1.50	1.5082	0.0082	0.1200	0.983
	η_2	1.50	1.5143	0.0143	0.1266	0.975
	ρ	0.25	0.2604	0.0104	0.0697	0.875

Table 5.8: Estimates, mean bias, root mean square error (RMSE) and coverage probability (CP) for dependent competing risk model parameters using direct optimization method with $\theta_1 = \theta_2 = 1.0$, $\eta_1 = \eta_2 = 1.5$ and $\rho = 0.5$ for each sample size (n)

n	Parameter	True Value	Estimate	Mean Bias	RMSE	95% CP
20	θ_1	1.00	1.0466	0.0466	0.3039	0.991
	θ_2	1.00	1.0223	0.0223	0.2927	0.987
	η_1	1.50	1.6207	0.1207	0.4177	0.969
	η_2	1.50	1.6061	0.1061	0.3941	0.968
	ρ	0.50	0.4778	-0.0222	0.1540	0.714
50	θ_1	1.00	0.9993	-0.0007	0.1697	0.988
	θ_2	1.00	1.0184	0.0184	0.1665	0.988
	η_1	1.50	1.5540	0.0540	0.2419	0.965
	η_2	1.50	1.5562	0.0562	0.2503	0.954
	ρ	0.50	0.4887	-0.0113	0.1055	0.829
100	θ_1	1.00	1.0044	0.0044	0.1137	0.997
	θ_2	1.00	1.0097	0.0097	0.1167	0.994
	η_1	1.50	1.5171	0.0171	0.1547	0.974
	η_2	1.50	1.5110	0.0110	0.1436	0.983
	ρ	0.50	0.4872	-0.0128	0.0798	0.881
200	θ_1	1.00	1.0091	0.0091	0.0882	0.997
	θ_2	1.00	1.0050	0.0050	0.0867	0.997
	η_1	1.50	1.5114	0.0114	0.1158	0.968
	η_2	1.50	1.5109	0.0109	0.1214	0.966
	ρ	0.50	0.4893	-0.0107	0.0574	0.931

Table 5.9: Estimates, mean bias, root mean square error (RMSE) and coverage probability (CP) for dependent competing risk model parameters using direct optimization method with $\theta_1 = \theta_2 = 1.0$, $\eta_1 = \eta_2 = 1.5$ and $\rho = 0.75$ for each sample size (n)

n	Parameter	True Value	Estimate	Mean Bias	RMSE	95% CP
20	θ_1	1.00	1.0584	0.0584	0.3101	0.982
	θ_2	1.00	1.0402	0.0402	0.2993	0.983
	η_1	1.50	1.6213	0.1213	0.3990	0.967
	η_2	1.50	1.6090	0.1090	0.3946	0.971
	ρ	0.75	0.6750	-0.0750	0.1238	0.755
50	θ_1	1.00	1.0085	0.0085	0.1605	0.996
	θ_2	1.00	1.0172	0.0172	0.1603	0.994
	η_1	1.50	1.5353	0.0353	0.2174	0.969
	η_2	1.50	1.5382	0.0382	0.2150	0.972
	ρ	0.75	0.7030	-0.0470	0.0804	0.883
100	θ_1	1.00	1.0077	0.0077	0.1175	0.997
	θ_2	1.00	1.0140	0.0140	0.1180	0.994
	η_1	1.50	1.5210	0.0210	0.1519	0.964
	η_2	1.50	1.5167	0.0167	0.1445	0.972
	ρ	0.75	0.7140	-0.0360	0.0599	0.926
200	θ_1	1.00	1.0078	0.0078	0.0800	0.997
	θ_2	1.00	1.0055	0.0055	0.0779	0.998
	η_1	1.50	1.5106	0.0106	0.0986	0.971
	η_2	1.50	1.5081	0.0081	0.1025	0.975
	ρ	0.75	0.7233	-0.0267	0.0425	0.940

Table 5.10: Estimates, mean bias, root mean square error (RMSE) and coverage probability (CP) for dependent competing risk model parameters using direct optimization method with $\theta_1 = \theta_2 = 1.0$, $\eta_1 = \eta_2 = 2.0$ and $\rho = 0.25$ for each sample size (n)

n	Parameter	True Value	Estimate	Mean Bias	RMSE	95% CP
20	θ_1	1.00	1.0225	0.0225	0.3003	0.985
	θ_2	1.00	0.9986	-0.0014	0.2983	0.981
	η_1	2.00	2.1957	0.1957	0.6239	0.961
	η_2	2.00	2.1788	0.1788	0.5948	0.959
	ρ	0.25	0.3013	0.0513	0.1595	0.657
50	θ_1	1.00	0.9889	-0.0111	0.1828	0.989
	θ_2	1.00	1.0129	0.0129	0.1858	0.996
	η_1	2.00	2.0859	0.0859	0.3619	0.964
	η_2	2.00	2.0733	0.0733	0.3645	0.968
	ρ	0.25	0.2754	0.0254	0.1179	0.730
100	θ_1	1.00	0.9997	-0.0003	0.1129	0.993
	θ_2	1.00	1.0045	0.0045	0.1152	0.994
	η_1	2.00	2.0285	0.0285	0.2081	0.985
	η_2	2.00	2.0218	0.0218	0.1981	0.991
	ρ	0.25	0.2593	0.0093	0.0888	0.792
200	θ_1	1.00	1.0069	0.0069	0.0880	0.998
	θ_2	1.00	0.9990	-0.0010	0.0851	1.000
	η_1	2.00	2.0075	0.0075	0.1610	0.984
	η_2	2.00	2.0156	0.0156	0.1697	0.976
	ρ	0.25	0.2517	0.0017	0.0659	0.865

Table 5.11: Estimates, mean bias, root mean square error (RMSE) and coverage probability (CP) for dependent competing risk model parameters using direct optimization method with $\theta_1 = \theta_2 = 1.0$, $\eta_1 = \eta_2 = 2.0$ and $\rho = 0.5$ for each sample size (n)

n	Parameter	True Value	Estimate	Mean Bias	RMSE	95% CP
20	θ_1	1.00	1.0484	0.0484	0.3056	0.993
	θ_2	1.00	1.0239	0.0239	0.2940	0.988
	η_1	2.00	2.1582	0.1582	0.5569	0.967
	η_2	2.00	2.1393	0.1393	0.5291	0.974
	ρ	0.50	0.4696	-0.0304	0.1506	0.712
50	θ_1	1.00	1.0023	0.0023	0.1717	0.988
	θ_2	1.00	1.0197	0.0197	0.1682	0.994
	η_1	2.00	2.0638	0.0638	0.3237	0.967
	η_2	2.00	2.0743	0.0743	0.3358	0.949
	ρ	0.50	0.4765	-0.0235	0.1042	0.827
100	θ_1	1.00	1.0063	0.0063	0.1139	0.996
	θ_2	1.00	1.0119	0.0119	0.1173	0.996
	η_1	2.00	2.0215	0.0215	0.2063	0.972
	η_2	2.00	2.0127	0.0127	0.1918	0.983
	ρ	0.50	0.4752	-0.0248	0.0782	0.882
200	θ_1	1.00	1.0129	0.0129	0.0896	0.997
	θ_2	1.00	1.0093	0.0093	0.0877	0.999
	η_1	2.00	2.0109	0.0109	0.1557	0.969
	η_2	2.00	2.0066	0.0066	0.1630	0.966
	ρ	0.50	0.4759	-0.0241	0.0588	0.927

Table 5.12: Estimates, mean bias, root mean square error (RMSE) and coverage probability (CP) for dependent competing risk model parameters using direct optimization method with $\theta_1 = \theta_2 = 1.0$, $\eta_1 = \eta_2 = 2.0$ and $\rho = 0.75$ for each sample size (n)

n	Parameter	True Value	Estimate	Mean Bias	RMSE	95% CP
20	θ_1	1.00	1.0611	0.0611	0.3117	0.988
	θ_2	1.00	1.0409	0.0409	0.3001	0.982
	η_1	2.00	2.1569	0.1569	0.5342	0.966
	η_2	2.00	2.1458	0.1458	0.5254	0.969
	ρ	0.75	0.6669	-0.0831	0.1272	0.754
50	θ_1	1.00	1.0097	0.0097	0.1611	0.992
	θ_2	1.00	1.0184	0.0184	0.1607	0.987
	η_1	2.00	2.0472	0.0472	0.2888	0.968
	η_2	2.00	2.0507	0.0507	0.2853	0.969
	ρ	0.75	0.6927	-0.0573	0.0853	0.858
100	θ_1	1.00	1.0096	0.0096	0.1183	0.996
	θ_2	1.00	1.0176	0.0176	0.1200	0.994
	η_1	2.00	2.0256	0.0256	0.2017	0.965
	η_2	2.00	2.0164	0.0164	0.1940	0.973
	ρ	0.75	0.7029	-0.0471	0.0656	0.915
200	θ_1	1.00	1.0094	0.0094	0.0806	0.997
	θ_2	1.00	1.0073	0.0073	0.0782	0.999
	η_1	2.00	2.0126	0.0126	0.1325	0.971
	η_2	2.00	2.0097	0.0097	0.1377	0.974
	ρ	0.75	0.7109	-0.0391	0.0505	0.937

In order to evaluate the accuracy of the asymptotic confidence intervals, it is important to assess the asymptotic normality of the MLEs. The Figures 5.1 to 5.3 showcase the histograms of MLEs for all three levels of dependence parameter ρ with $\theta_1 = \theta_2 = 1.0$ and $\eta_1 = \eta_2 = 2.0$ settings. The histograms indicate that the distributions of the MLEs for all parameters are close to normal, indicating that the asymptotic normality assumption is reasonable. These results were consistent across all simulation settings considered. The close-to-normal distributions of the MLEs provide further support for the reliability of the asymptotic confidence intervals obtained.

The figures from 5.4 to 5.6 are the histograms of MLEs of transformed data for the same parameter settings.

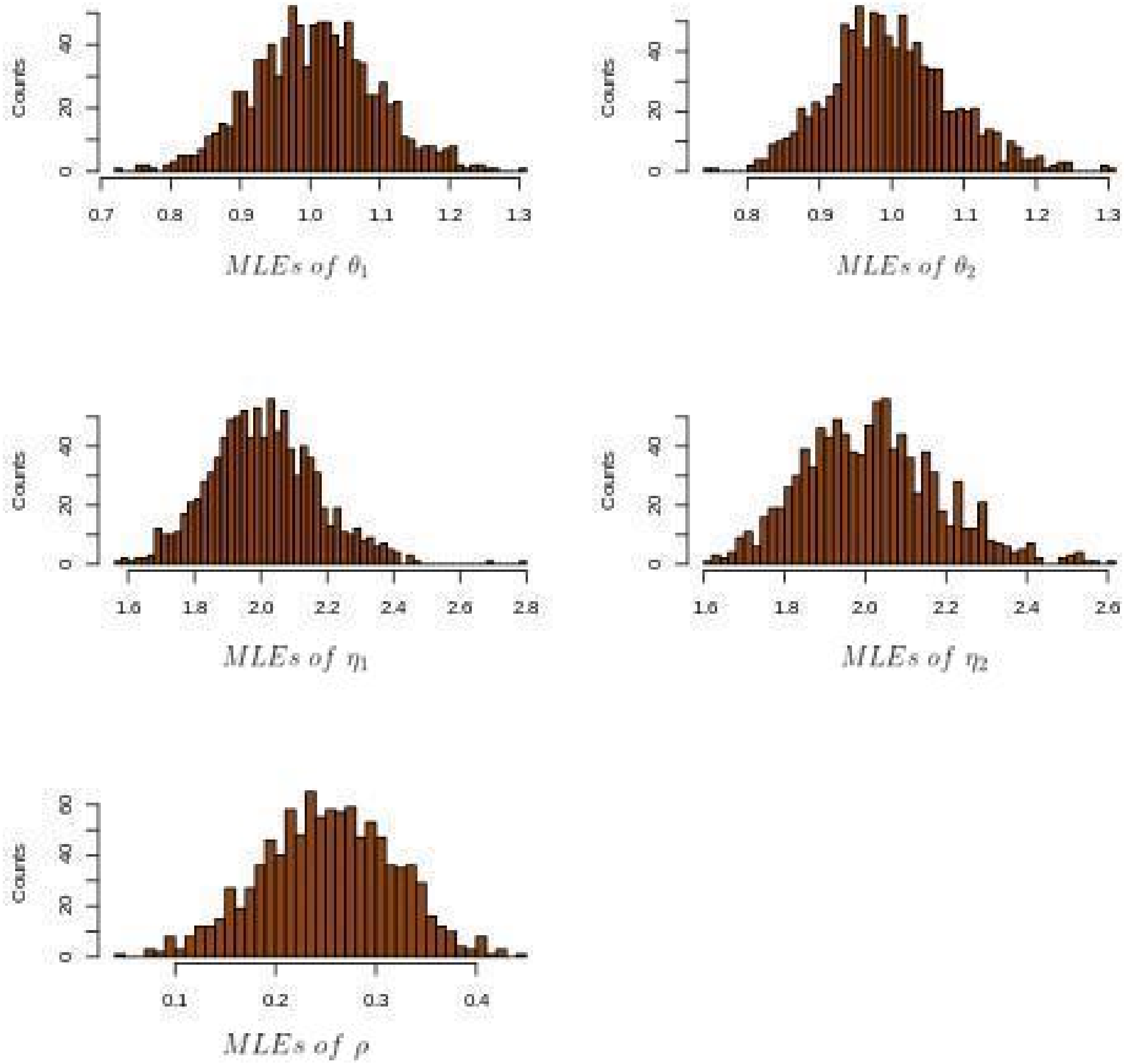


Figure 5.1: Histograms of the MLEs based on complete data when $\theta_1 = \theta_2 = 1.0$, $\eta_1 = \eta_2 = 2.0$ and $\rho = 0.25$

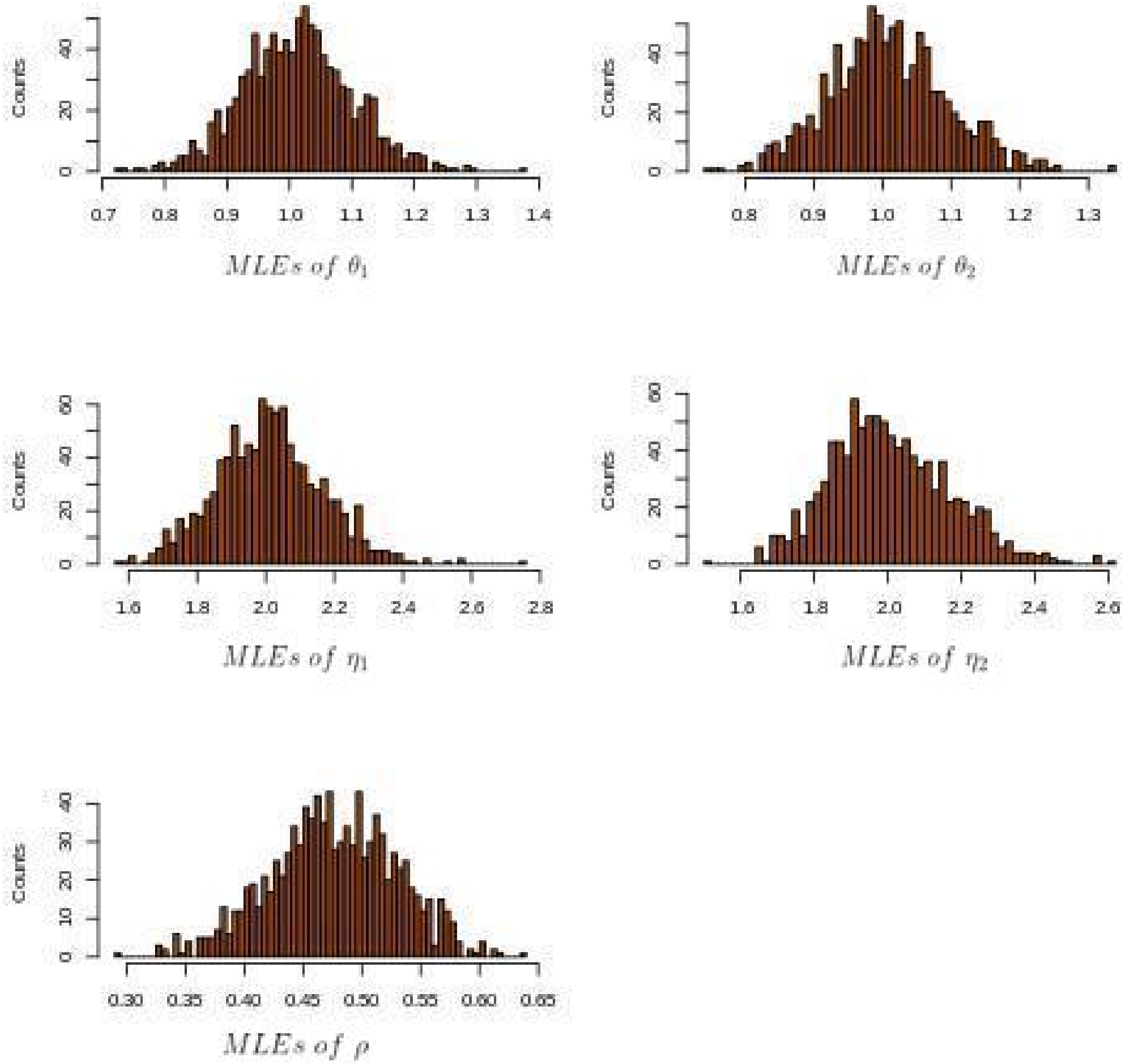


Figure 5.2: Histograms of the MLEs based on complete data when $\theta_1 = \theta_2 = 1.0$, $\eta_1 = \eta_2 = 2.0$ and $\rho = 0.5$

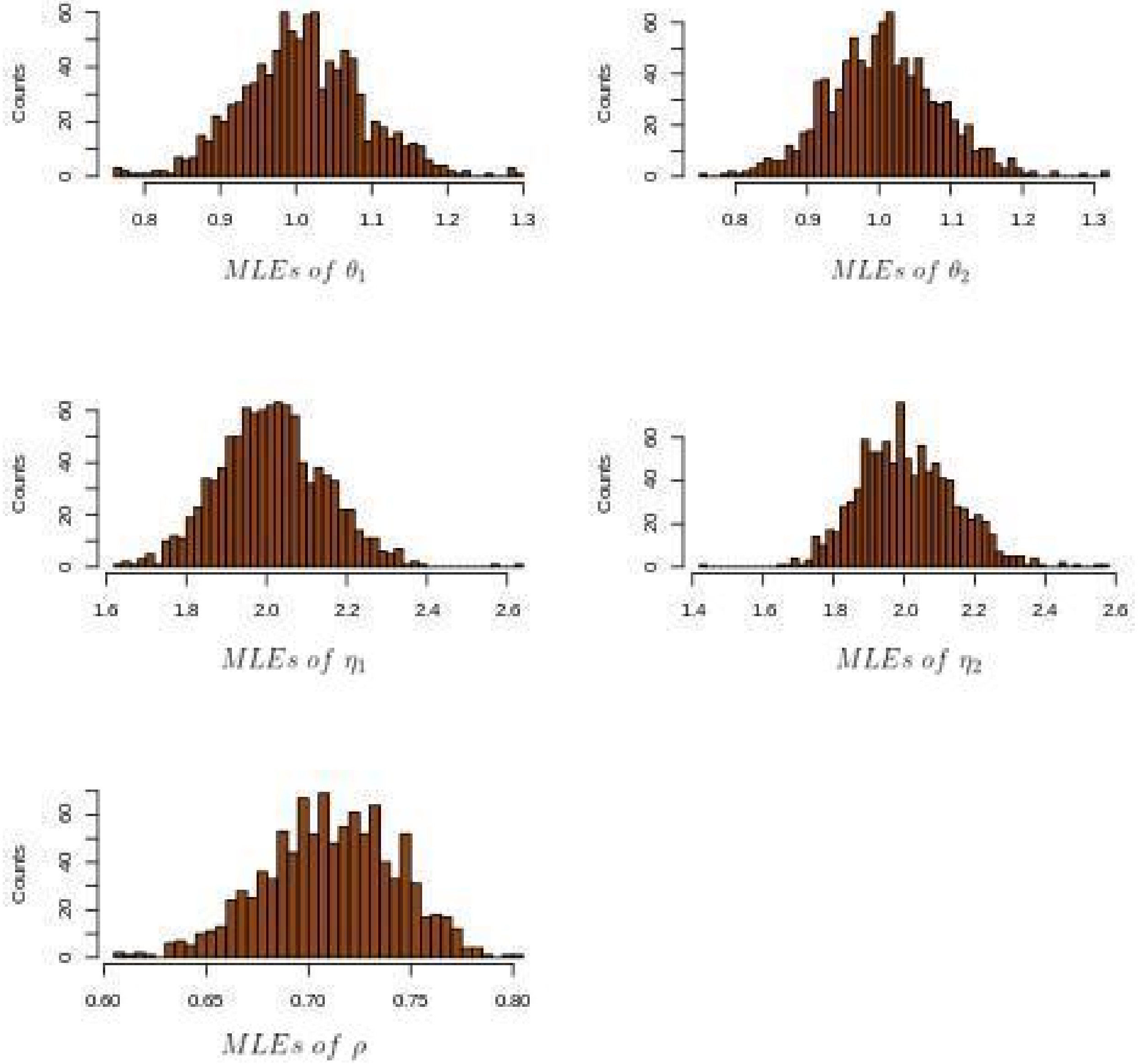


Figure 5.3: Histograms of the MLEs based on complete data when $\theta_1 = \theta_2 = 1.0$, $\eta_1 = \eta_2 = 2.0$ and $\rho = 0.75$

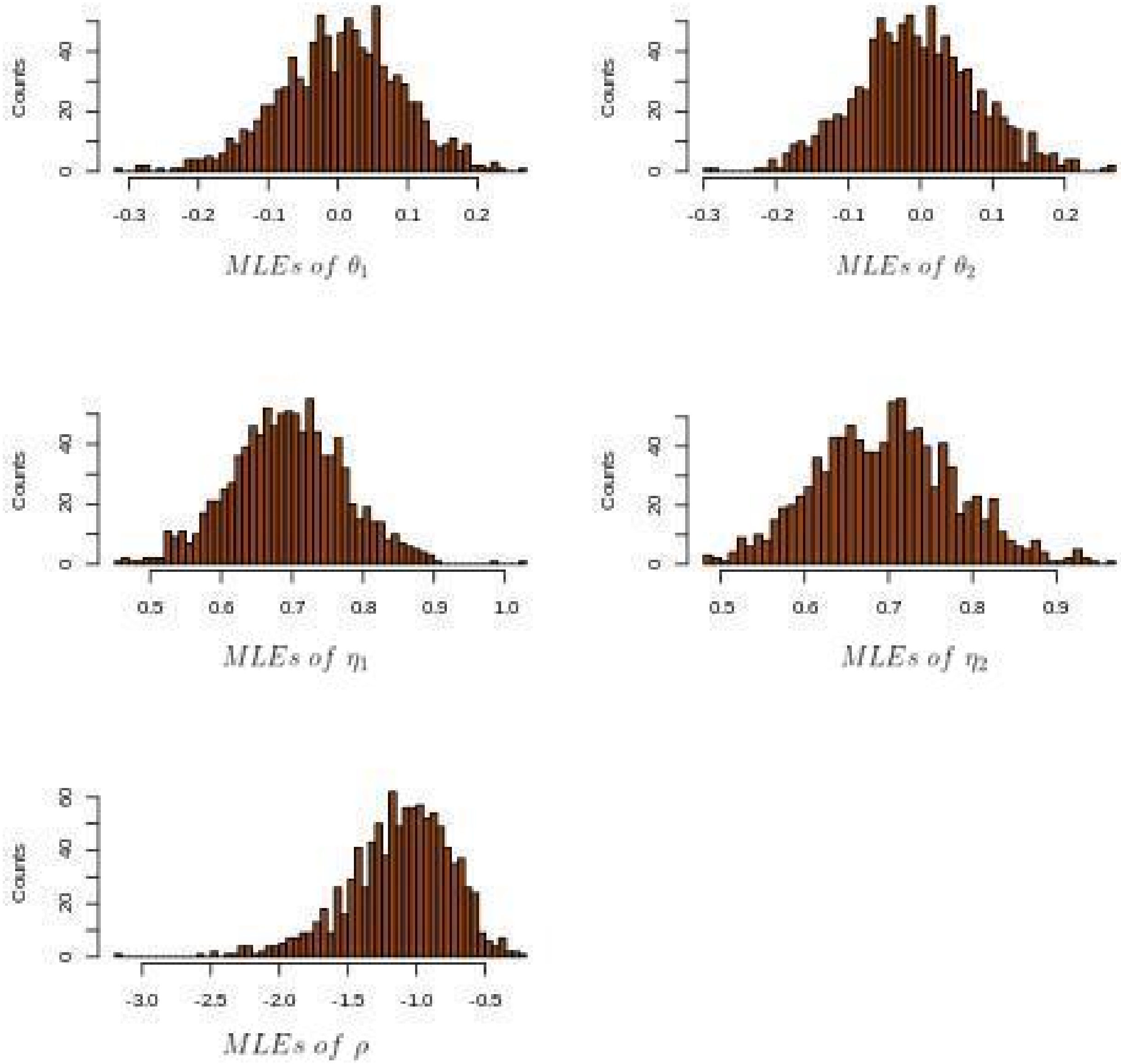


Figure 5.4: Histograms of the MLEs with transformation based on complete data when $\theta_1 = \theta_2 = 1.0$, $\eta_1 = \eta_2 = 2.0$ and $\rho = 0.25$

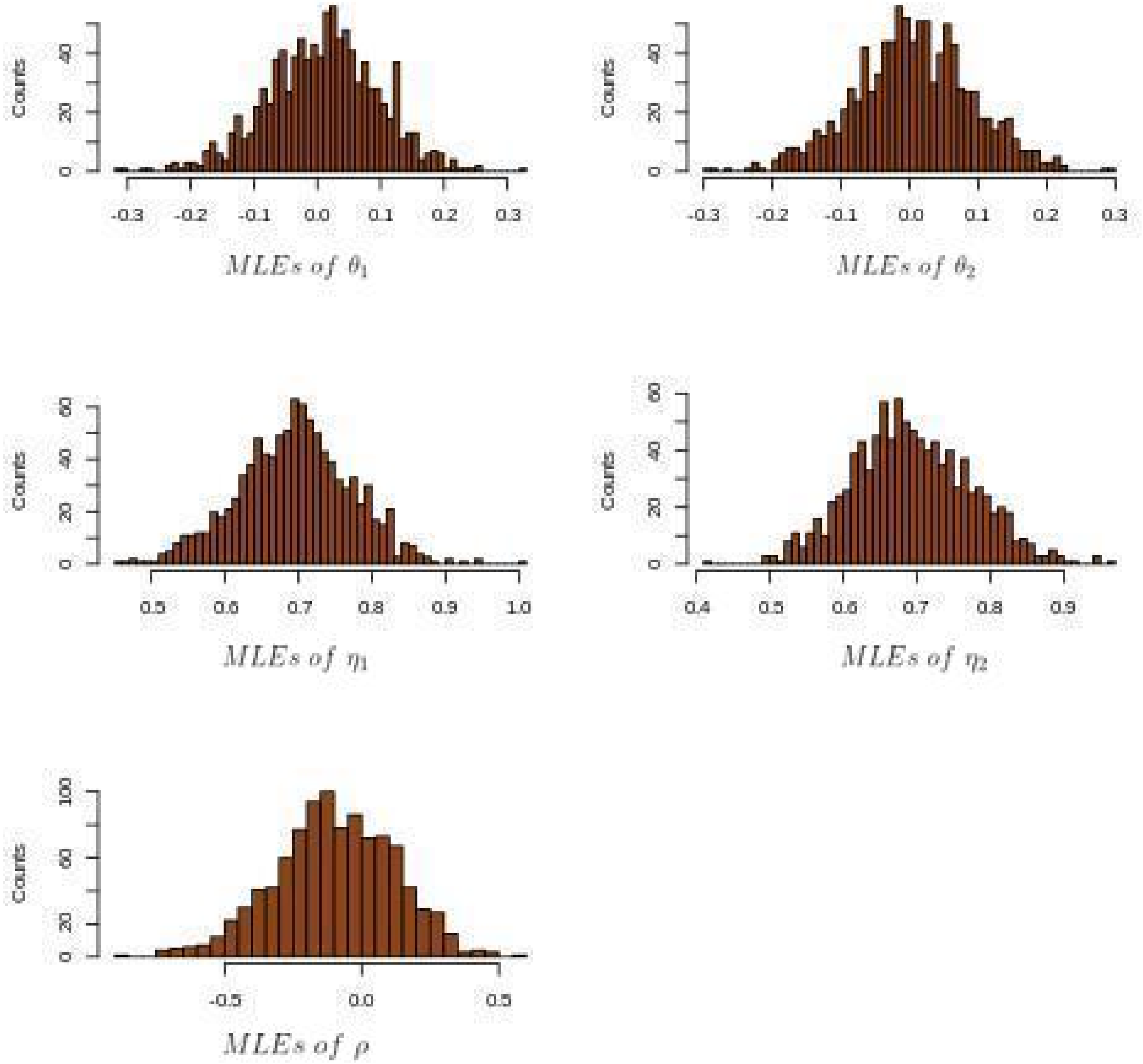


Figure 5.5: Histograms of the MLEs with transformation based on complete data when $\theta_1 = \theta_2 = 1.0$, $\eta_1 = \eta_2 = 2.0$ and $\rho = 0.5$

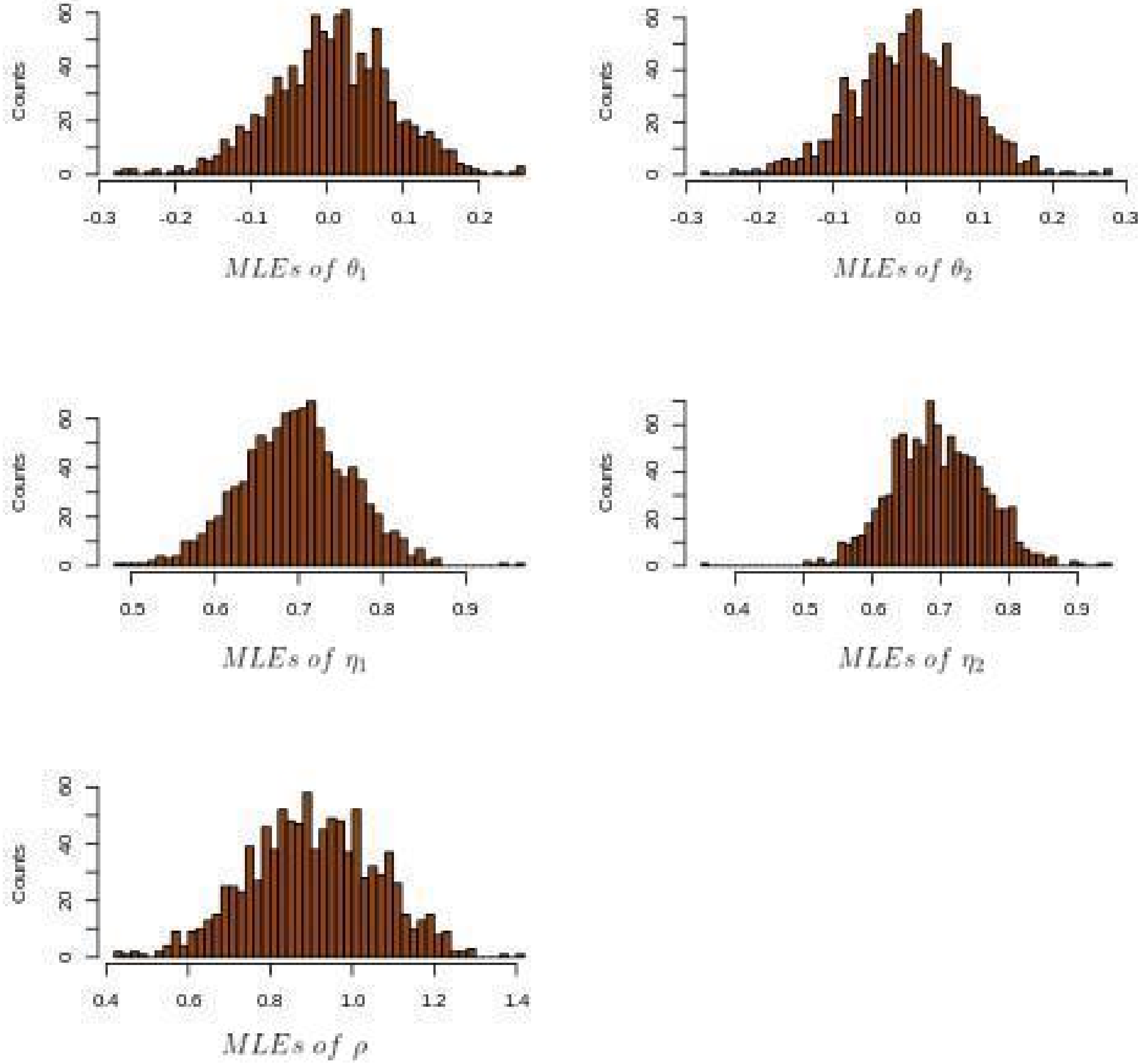


Figure 5.6: Histograms of the MLEs with transformation based on complete data when $\theta_1 = \theta_2 = 1.0$, $\eta_1 = \eta_2 = 2.0$ and $\rho = 0.75$

Chapter 6

Illustrative data analysis

In this section, we apply the DBW distribution and our inferential method to two real datasets to showcase its adaptability in modeling different types of data. The first dataset we analyze is the car warranty data, which is commonly used for reliability analysis. We show how the DBW distribution can be used to model the lifetime of car components and how it can be extended to the two-dimensional warranty model, which allows for modeling the warranty cost as a function of time and mileage. The DBW distribution is well-suited for this type of data because of its flexibility and ability to model both short and long-term failure rates.

The second dataset we analyze is a general bivariate dataset on bone mineral density. Although this dataset is not lifetime data, we demonstrate how the DBW distribution can be used to model different types of data. We show how the DBW distribution can be used to model the joint distribution of two variables, which is useful for analyzing the relationship between two variables in a multivariate setting. We also demonstrate how the DBW distribution can be used to model the marginal distributions of each variable separately, which is useful for understanding the distribution of each variable individually.

6.1 Car Warranty Data

A two-dimensional warranty is often offered with a variety of products, such as heavy-duty machines, industrial equipment, and automobiles, which considers both the age and usage of the product when evaluating warranty claims. To analyze reliability data for such products, a suitable approach would be to use a bivariate distributional model. A two-dimensional warranty model was proposed by Jung and Bai (2007), to analyze the car warranty dataset which is used in this work.

The car warranty dataset contains data on the age and mileage of 40 cars, which are the measurable factors to measure its lifetime. Car age and mileage can be considered as dependent competing risks in warranty data. When a car is used, both age and mileage increase and either factor can lead to failure or the need for maintenance. For example, a car may need a replacement part after being driven a certain number of miles, or it may need a tune-up after a certain number of years. These two risks are competing because the car can fail due to either age or mileage, but not both at the same time. Additionally, age and mileage are dependent because a car that has been driven frequently will typically be older than a car that has been driven less frequently, even if both cars were manufactured at the same time. Therefore, it is important to consider both age and mileage as dependent competing risks when analyzing warranty data for cars or other vehicles.

Figure 6.1 shows the scatter plot between the car mileage and car age data. The observed data indicates that the two variables are highly correlated. The correlation coefficient between the two variables has been calculated to be 0.995 indicating a strong positive relationship.

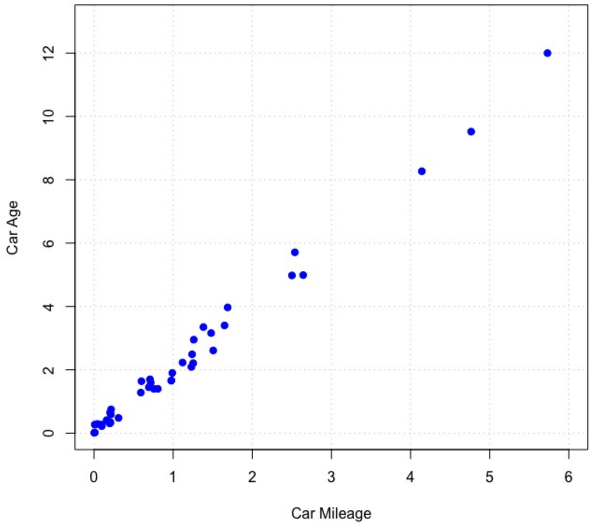


Figure 6.1: Scatterplot of the Age (X_1) and Mileage (X_2) of cars for the car warranty data.

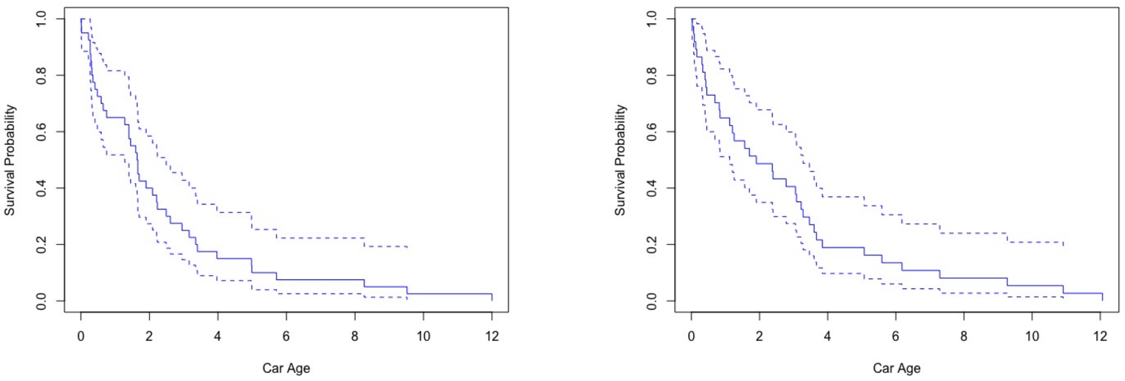
Assuming a DBW distribution for the variables car age (X_1) and car mileage (X_2), we can estimate the maximum likelihood estimators (MLEs) of the model parameters based on the given warranty data. The results are summarized in Table 6.1. Additionally, we calculate asymptotic 95% confidence intervals for the parameters.

Table 6.1: Estimated parameters and 95% confidence intervals of the DBW model based on the warranty data

Parameter	Initial value	Estimate	Confidence Interval
θ_1	2.045	2.2576	(0.513, 9.940)
θ_2	1.0604	0.9218	(0.674, 1.260)
η_1	0.8671	0.8812	(0.342, 2.267)
η_2	0.7605	0.7885	(0.613, 1.015)
ρ	0.9167	0.9191	(0.163, 0.998)

To check if the model fit is reasonable, in Figures 6.2 and 6.3 we plot the marginal survival function as well as the non-parametric estimated survival function, evaluated at the ordered observed and estimated values, separately for both car age and mileage.

To assess the validity of the model, it is common practice to compare the survival functions obtained from the parametric model (which assumes a specific distribution) with the survival function estimated without any distributional assumptions. By comparing the patterns of the two survival functions, we can evaluate if they exhibit similar trends. If the parametric survival function and the non-parametric (empirical) survival function demonstrate consistent patterns and follow a similar shape, it suggests that the chosen parametric model adequately captures the underlying distribution of the data. This comparison allows us to assess the goodness-of-fit of the parametric model and determine if it provides a reasonable approximation of the true survival function. Any significant deviations or discrepancies between the two survival functions may indicate a poor fit of the chosen parametric model to the data, suggesting the need for alternative modeling approaches or further investigation.

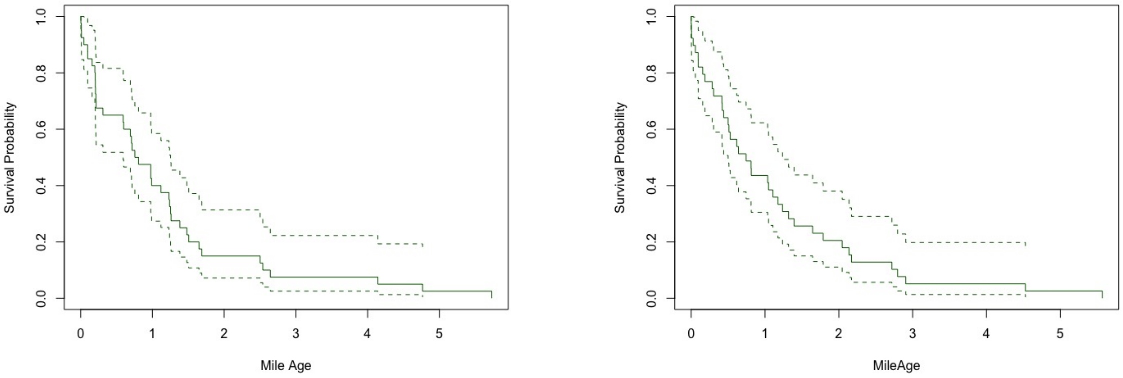


(a) Marginal survival distribution

(b) Estimated survival distribution

Figure 6.2: Marginal and estimated survival distributions for car age in warranty data

The comparison between the marginal and estimated survival functions reveals similar patterns, indicating that the estimated values of the survival function align well. This observation suggests that the model fit is reasonable, as the estimated survival probabilities closely resemble the actual survival probabilities.



(a) Marginal survival distribution

(b) Estimated survival distribution

Figure 6.3: Marginal and estimated survival distributions for car mileage in warranty data

6.2 Bone mineral density data

The bone mineral density (BMD) dataset from Johnson et al. (2002) consists of measurements from 24 individuals. Each observation in the dataset includes two values: X_1, representing the BMD of the bone dominant radius before the study, and X_2, representing the BMD after the study.

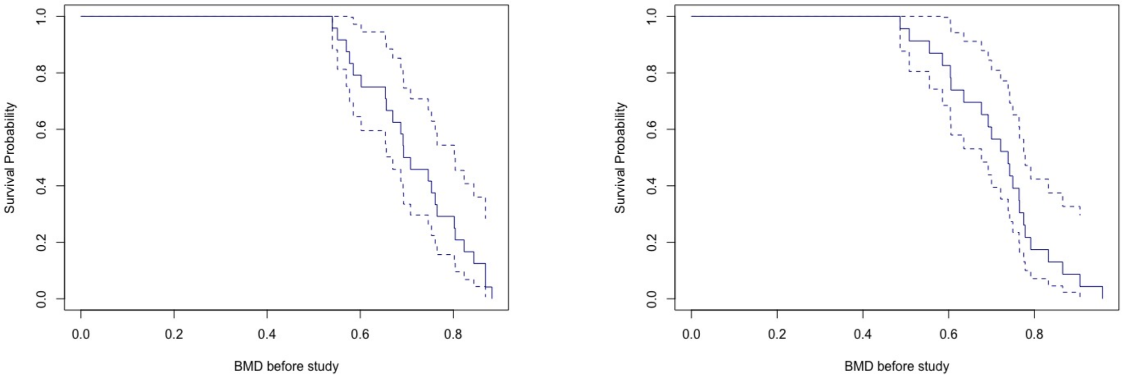
In the context of competing risks, the BMD before the study and after the study are considered as the dependent competing risks. These risks are dependent because the BMD after the study is influenced by the initial BMD before the study. Analyzing these dependent competing risks allows us to understand the changes in BMD over time and evaluate the factors affecting BMD progression or regression.

Assuming a DBW distribution for the variables BMD before the study (X_1) and BMD after the study (X_2), the maximum likelihood estimates of the model parameters are generated. Table 6.2 presents the estimated values of the parameters, including their asymptotic 95% confidence intervals for the estimated parameters.

Table 6.2: Estimated parameters and 95% confidence intervals of the DBW model based on the Bone Mineral Density data

Parameter	Initial value	Estimate	Confidence Interval
θ_1	0.0953	0.0954	(0.028, 0.320)
θ_2	0.0896	0.0899	(0.028, 0.285)
η_1	8.3695	8.2446	(5.414, 12.555)
η_2	7.6876	7.8709	(5.440, 11.388)
ρ	0.6017	0.602	(0.042, 0.981)

In order to evaluate the validity of the model, we analyze the marginal Weibull survival function and the estimated empirical survival function for the bone mineral density (BMD) data. We create plots, similar to the ones used for warranty data, to visualize these functions. These plots, shown in Figures 6.4 and 6.5, display the evaluated values of the survival functions at the ordered observed and estimated values of both X_1 (BMD before the study) and X_2 (BMD after the study).

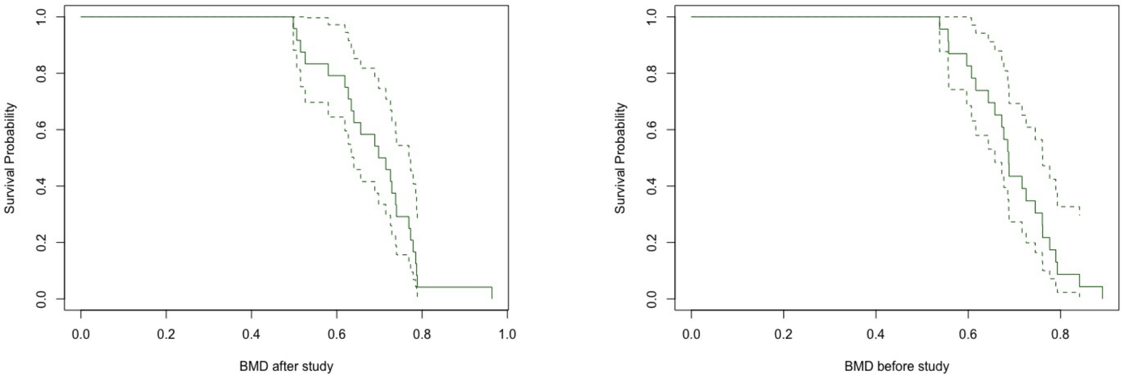


(a) Marginal survival distribution

(b) Estimated survival distribution

Figure 6.4: Marginal and estimated survival distributions for BMD before the study

By comparing the shapes and patterns of the marginal Weibull survival function and the estimated empirical survival function, we can assess the reasonableness of the model. We can visually analyze the agreement between the observed and estimated survival probabilities, providing that the model fit is reasonable. Similar trends and consistency between these two functions indicate a good fit of the model to the BMD data.



(a) Marginal survival distribution (b) Estimated survival distribution

Figure 6.5: Marginal and estimated survival distributions for BMD after the study

Chapter 7

Conclusions

This thesis discusses a bivariate Weibull distribution which is widely known as Downton's bivariate Weibull (DBW) distribution as a novel approach to modelling dependent competing risk scenarios. The DBW distribution has been investigated to explore its significant statistical properties in relation to the dependent competing risk model. Then, this thesis explores inferences that can be made based on complete bivariate data for this distribution. We employ a direct optimization technique to maximize the likelihood function using numerical methods.

In this thesis, we have conducted a simulation study and presented two illustrative examples to demonstrate the performance of the estimation method. Through these evaluations, we were able to observe and comment on the effectiveness of these methods.

During the simulation study, we investigated the behaviour of the estimation methods under various settings. We found that the method provided reasonable estimates of the parameter values. However, it became evident that the model fitting process yielded better results for larger sample sizes compared to smaller ones, which aligns with our expectations. These findings highlight the importance of considering the sample size when applying these estimation methods. It is crucial to have an adequate amount of data to achieve more accurate and reliable parameter estimates. Additionally, the simulation study allowed us to gain insights into the performance characteristics and limitations of the estimation methods in different scenarios.

To illustrate the practical application of the proposed methods, we present two case studies based on real datasets. These case studies serve to provide concrete examples and demonstrate the effectiveness of the estimation methods in real-world scenarios. Specifically, our method takes into account different failure statuses of the units sold under two-dimensional warranty. This generalized methodology provides a more comprehensive analysis of the two-dimensional warranty data, and allows for more accurate modeling and estimation of the underlying distribution.

It is worth noting that this method only considers the complete data and does not take into account any censoring or truncation in the data. Therefore, it may not be suitable for datasets with right-censored or interval-censored observations.

In the context of the bivariate competing risks model with DBW distribution, we ignore the tied observations in the likelihood function. However, it is important to note that ignoring tied observations can potentially lead to biased estimation if the proportion of tied observations is large in real data. In such cases, it may be more appropriate to ensure that the sample size is sufficiently large so that the proportion of tied observations is negligible.

Finally, in this study, our focus was primarily on analyzing complete data within the framework of the proposed model. However, there are opportunities for future research to extend our inference procedures to incorporate different forms of censoring, such as right censoring and progressive censoring. Also, the work in this thesis could be extended to look at alternative distributions to model dependent competing risks data.

APPENDIX

pcondbivweib - Conditional CDF of bivariate weibull (scaled rv's)

Input(s) - A_1, A_2 : scaled versions of the desired RV's $A_2 = x_2^{n_2}/\theta_2 * \rho/(1-\rho), A_1 = x_1^{n_1}/\theta_1/(1-\rho)$

Output(s) - $F_{A_2}|A_1 = a_1(a_2)$

```
# first function won't work too well for large values so commented
# out pcondbivweib <- function(A2, A1){ if(A2 <= 0){return(0)}
# phi3alt.fun <- function(t, w, z){exp(w*(1-t))*besselI(2*sqrt(z*t),
# nu = 0, expon.scaled = FALSE)} A2 * exp(-(A1+A2)) *
# integrate(phi3alt.fun, 0, 1, w = A2, z = A1 * A2)$value }

pcondbivweib <- function(A2, A1) {
    if (A2 <= 0) {
        return(0)
    }
    phi3alt.fun <- function(t, w, z) {
        exp(2 * sqrt(z * t) - w * t - z/w) * besselI(2 * sqrt(z * t), nu = 0,
            expon.scaled = TRUE)
    }
    A2 * integrate(phi3alt.fun, 0, 1, w = A2, z = A1 * A2)$value
}
```

rcondbivweib - Generates conditional bivariate weibull (scaled rv's)

Input(s) - $A_1 A_1$: vector containing A_1, and inititial estimate of A_2 (in that order)

Output(s) - $F_{A_2}|A_1 = a_1(a_2)$

```
rcondbivweib <- function(A1) {
    u <- runif(1)
    loc.fun <- function(A2, A1, u) {
        pcondbivweib(A2, A1 = A1) - u
    }
    uniroot(f = loc.fun, lower = 0, upper = 100, f.lower = -u, f.upper = 1 -
        u, A1 = A1, u = u, tol = 1e-08)$root
}
```

rbivweib - Generates bivariate weibull

Input(s) - n : sample size - theta1, theta2, nu1, nu2, rho ($\theta_1,\ \theta_2,\ \eta_1,\ \eta_2,\ \rho$, respectively) : parameters for the bivariate weibull

Output(s) - n by 2 matrix whose columns are the 2 correlated samples

```
rbivweib <- function(n, theta1, theta2, nu1, nu2, rho) {

    data.out <- matrix(0, ncol = 2, nrow = n, dimnames = list(NULL, c("x1",
        "x2")))

```

```
    data.out[, 1] <- rweibull(n, shape = nu1, scale = theta1^(1/nu1))

    if (rho == 0) {
        data.out[, 2] <- rweibull(n, shape = nu2, scale = theta2^(1/nu2))
        return(data.out)
    }

    A1 <- data.out[, 1]^nu1/theta1 * rho/(1 - rho)
    A2 <- sapply(A1, rcondbivweib)
    data.out[, 2] <- (theta2 * A2 * (1 - rho))^(1/nu2)

    data.out

}
```

downton.jointpdf - Generates joint probability density function for Downton's bivariate Weibull

Input(s) - n : sample size - theta1, theta2, nu1, nu2, rho : parameters for the bivariate weibull

Output(s) - numerical output of the bivariate pdf

```
downton.jointpdf <- function(u, the1, the2, n1, n2, rh) {
    x1 <- u[1]
    x2 <- u[2]
    return(((n1 * n2 * x1^(n1 - 1) * x2^(n2 - 1))/the1 * the2 * (1 - rh)) *
        exp(-(1/1 - rh) * (x1^n1/the1 + x2^n2/the2)) * besselI(2 * sqrt(rh *
        x1^n1 * x2^n2)/((1 - rh) * (sqrt(the1 * the2))), nu = 0))
}
```

marcumq - Compute the generalized Marcum Q function

Input(s) - a, b input arguments, specified as non-negative real numbers. - m order, specified as a positive integer

```
library(lmomco)
marcumq <- function(a, b, m = 1) {
    stats::pchisq(b^2, df = 2 * m, ncp = a^2, lower.tail = FALSE)
}
```

Define bivariate survival function

downton.bisurv - Generates bivariate survival function for Downton's bivariate Weibull

Input(s) - n : sample size - theta1, theta2, nu1, nu2, rho : parameters for the bivariate weibull

Output(s) - numerical output of the bivariate survival function

```
# first function was not defined well so commented out downton.bisurv
# <- function(u,the1,the2,n1,n2,rh) { x1 <- u[1]; x2 <- u[2]
# return((exp(-x1^n1/the1)* (1-
# marcumq(sqrt(2/(1-rh))*(x2^(n2/2)/sqrt(the2)),
# sqrt(2*rh/(1-rh))*(x1^(n1/2)/sqrt(the1)), m = 1))) +
# (exp(-x2^n2/the2)*marcumq(sqrt(2*rh/(1-rh))*(x2^(n2/2)/sqrt(the2)),
```

```
# sqrt(2/(1-rh))*(x1^(n1/2)/sqrt(the1)), m = 1))) }

downton.bisurv <- function(u, the1, the2, n1, n2, rh) {
    x1 <- u[1]
    x2 <- u[2]
    A <- sqrt(2/((1 - rh) * the2))
    B <- sqrt((2 * rh)/((1 - rh) * the1))
    C <- sqrt((2 * rh)/((1 - rh) * the2))
    D <- sqrt(2/((1 - rh) * the1))
    return((exp(-(x1^n1)/the1) * (1 - marcumq(A * (x2^(n2/2)), B * (x1^(n1/2)),
        m = 1))) + (exp(-(x2^n2)/the2) * (marcumq(C * (x2^(n2/2)), D *
        (x1^(n1/2)), m = 1))))
}


#-------- Bivariate survival function deferentiated by X1 ----------

# Marcumq equation 1-deferentiated by X1
marcumq.eq1.derx1 <- function(x1, x2, the1, the2, n1, n2, rh) {
    A <- sqrt(2/((1 - rh) * the2))
    B <- sqrt((2 * rh)/((1 - rh) * the1))
    return(-B * (x1^(n1/2)) * exp(-0.5 * (A^2 * x2^(n2) + B^2 * x1^(n1))) *
        besselI(B * x1^(n1/2) * A * x2^(n2/2), nu = 0) * (B/2) * n1 * x1^((n1/2) -
        1))
}

# Marcumq equation2 -deferentiated by X1
marcumq.eq2.derx1 <- function(x1, x2, the1, the2, n1, n2, rh) {
    A <- sqrt((2 * rh)/((1 - rh) * the2))
    B <- sqrt(2/((1 - rh) * the1))
    return(-B * (x1^(n1/2)) * exp(-0.5 * (A^2 * x2^(n2) + B^2 * x1^(n1))) *
        besselI(B * x1^(n1/2) * A * x2^(n2/2), nu = 0) * (B/2) * n1 * x1^((n1/2) -
        1))
}

# Bivariate survival function deferentiated by X1
derx1.bi.sur <- function(u, the1, the2, n1, n2, rh) {
    x1 <- u[1]
    x2 <- u[2]
    A <- sqrt(2/((1 - rh) * the2))
    B <- sqrt((2 * rh)/((1 - rh) * the1))
    return((((-n1 * x1^(n1 - 1))/the1) * (exp(-(x1^n1)/the1) * (1 - marcumq(A *
        (x2^(n2/2)), B * (x1^(n1/2)), m = 1)))) - (exp(-(x1^n1)/the1) *
        marcumq.eq1.derx1(x1, x2, the1, the2, n1, n2, rh)) + (exp(-(x2^n2)/the2) *
        marcumq.eq2.derx1(x1, x2, the1, the2, n1, n2, rh)))
}
#-------------------------------------------------------------------

#-------- Bivariate survival function deferentiated by X2 ----------

# Marcumq equation 1-deferentiated by X2
marcumq.eq1.derx2 <- function(x1, x2, the1, the2, n1, n2, rh) {
    A <- sqrt(2/((1 - rh) * the2))
```

```
    B <- sqrt((2 * rh)/((1 - rh) * the1))
    return(B * (x1^(n1/2)) * exp(-0.5 * (A^2 * x2^(n2) + B^2 * x1^(n1))) *
        besselI(B * x1^(n1/2) * A * x2^(n2/2), nu = 1) * (A/2) * n2 * x2^((n2/2) -
        1))
}

# Marcumq equation 2-deferentiated by X2
marcumq.eq2.derx2 <- function(x1, x2, the1, the2, n1, n2, rh) {
    A <- sqrt((2 * rh)/((1 - rh) * the2))
    B <- sqrt(2/((1 - rh) * the1))
    return(B * (x1^(n1/2)) * exp(-0.5 * (A^2 * x2^(n2) + B^2 * x1^(n1))) *
        besselI(B * x1^(n1/2) * A * x2^(n2/2), nu = 1) * (A/2) * n2 * x2^((n2/2) -
        1))
}

# Bivariate survival function deferentiated by X2
derx2.bi.sur <- function(u, the1, the2, n1, n2, rh) {
    x1 <- u[1]
    x2 <- u[2]
    A <- sqrt((2 * rh)/((1 - rh) * the2))
    B <- sqrt(2/((1 - rh) * the1))
    return((((-n2 * x2^(n2 - 1))/the2) * (exp(-(x2^n2)/the2) * (marcumq(A *
        (x2^(n2/2)), B * (x1^(n1/2)), m = 1)))) - (exp(-(x1^n1)/the1) *
        marcumq.eq1.derx2(x1, x2, the1, the2, n1, n2, rh)) + (exp(-(x2^n2)/the2) *
        marcumq.eq2.derx2(x1, x2, the1, the2, n1, n2, rh)))
}
#----------------------------------------------------------------------
```

Define log and logit transformations with inverse functions

```
invfun1 <- exp
invfun2 <- function(x) {
    exp(x)/(1 + exp(x))
}
fun1 <- log
fun2 <- function(x) {
    log(x/(1 - x))
}
```

The log likelihood function

downton.logLikelihood - Generates log likelihood function for Downton's bivariate Weibull

Input(s) - parms : vector of parameters for bivariate weibull(order-theta1, theta2, nu1, nu2, rho) - bidata : bivariate data for the bivariate weibull

Output(s) - numerical output of negative log likelihood function

```
library(numDeriv)
library(stringr)

downton.logLikelihood <- function(parms, x1, x2) {
```

```
    bidata <- data.frame(x1, x2)

    # deriving y by taking min value of x2 and x2
    bidata <- transform(bidata, y = pmin(x1, x2))

    # Defining the cause and delta
    bidata$cause <- ifelse(bidata$x1 == bidata$x2, "X0", names(bidata[,
        c("x1", "x2")])[apply(bidata[, c("x1", "x2")], MARGIN = 1, FUN = which.min)])
    bidata$delta <- as.numeric(str_extract(bidata$cause, "[0-9]+"))

    thetax1 <- invfun1(parms[1])  #exp brings back to original scale
    thetax2 <- invfun1(parms[2])  #exp brings back to original scale
    nux1 <- invfun1(parms[3])  #exp brings back to original scale
    nux2 <- invfun1(parms[4])  #exp brings back to original scale
    rho <- invfun2(parms[5])  #invlogit brings back to original scale

    for (i in 1:nrow(bidata)) {
        x0 <- c(bidata$y[i], bidata$y[i])
        # Partial differentiation by X1-Survival function
        bidata$x1_sur[i] <- grad(downton.bisurv, x0, the1 = parms[1], the2 = parms[2],
            n1 = parms[3], n2 = parms[4], rh = parms[5])[1]
        bidata$x1_der[i] <- derx1.bi.sur(x0, the1 = thetax1, the2 = thetax2,
            n1 = nux1, n2 = nux2, rh = rho)
        # Partial differentiation by X2-Survival function
        bidata$x2_sur[i] <- grad(downton.bisurv, x0, the1 = parms[1], the2 = parms[2],
            n1 = parms[3], n2 = parms[4], rh = parms[5])[2]
        bidata$x2_der[i] <- derx2.bi.sur(x0, the1 = thetax1, the2 = thetax2,
            n1 = nux1, n2 = nux2, rh = rho)
    }
    # Survival function
    bidata$x1x2_sur[i] <- downton.bisurv(x0, the1 = thetax1, the2 = thetax2,
        n1 = nux1, n2 = nux2, rh = rho)
    # Density function
    bidata$x1x2_dens[i] <- downton.jointpdf(x0, the1 = thetax1, the2 = thetax2,
        n1 = nux1, n2 = nux2, rh = rho)

    likeliHood <- prod(bidata$x1_der[bidata$delta == 1], bidata$x2_der[bidata$delta ==
        2])
    neg_loglik <- (-log((likeliHood)))

    return(neg_loglik)

}

# If NAN values occurs in the way, gives a penalty
wrapper_fun <- function(par, x1, x2, n) {
    out <- tryCatch(downton.logLikelihood(par, x1, x2), error = function(e) NaN)
    if (is.nan(out)) {
        out <- (1e+20)  # return a large penalty value
    }
    return(out)

}
```

Initial values for parameters

bweib.init.est - generates initial values for the maximization

Input(s) - parms : bivariate data (x1, x2) - w : constant between 0 and 1

Output(s) - initial values of $\theta_1,\ \theta_2,\ \eta_1,\ \eta_2,\ \rho$, respectively for the maximization

```
bweib.init.est <- function(x1, x2, w) {
    sx1 <- sd(log(x1)) * sqrt(6)/pi
    sx2 <- sd(log(x2)) * sqrt(6)/pi
    return(list(thetax1 = (exp(mean(log(x1)) - digamma(1) * sx1))^(1/sx1),
        thetax2 = (exp(mean(log(x2)) - digamma(1) * sx2))^(1/sx2), nux1 = 1/sx1,
        nux2 = 1/sx2, rho = w/2 + (1 - w) * max(0, cor(x1, x2))))
}
```

Maximum likelihood estimates

bweib.mle.fun - generates the estimations of the parameters $\theta_1,\ \theta_2,\ \eta_1,\ \eta_2,\ \rho$

```
bweib.mle.fun <- function(x1, x2, n, w = 1/sqrt(n), alpha = 0.05, invfun1 = exp,
    invfun2 = function(x) {
        exp(x)/(1 + exp(x))
    }, fun1 = log, fun2 = function(x) {
        log(x/(1 - x))
    }) {

    # Check

    if (length(x1) != n | length(x2) != n) {
        stop("x1 and x2 must be vectors of length n")
    }
    if (w <= 0 | w >= 1) {
        stop("w must be between 0 and 1 exclusive")
    }

    # Create output items
    out.est.table <- matrix(0, ncol = 5, nrow = 5, dimnames = list(c("init",
        "mle", "ci.lower", "ci.upper", "ci"), c("theta1", "theta2", "nu1",
        "nu2", "rho")))
    out.vcovmat.trans <- matrix(0, ncol = 5, nrow = 5, dimnames = list(c("theta1",
        "theta2", "nu1", "nu2", "rho"), c("theta1", "theta2", "nu1", "nu2",
        "rho")))

    # Obtain initial estimates and convert to alternate scale
    init.est <- unlist(bweib.init.est(x1, x2, 1/sqrt(n)))

    init.est[1:4] <- fun1(init.est[1:4])  #transforming initial values, log
    init.est[5] <- fun2(init.est[5])  #transforming initial values, logit

    # Compute MLE
    optOut <- optim(par = init.est, fn = wrapper_fun, x1 = x1, x2 = x2,
        method = "BFGS", hessian = TRUE, control = list(maxit = 20000,
            abstol = 1e+16, reltol = 1e+16))
```

```
    # outputs putting back to original scale
    out.est.table[1, 1:4] <- round(invfun1(init.est[1:4]), 4)
    out.est.table[1, 5] <- round(invfun2(init.est[5]), 4)
    out.est.table[2, 1:4] <- round(invfun1(optOut$par[1:4]), 4)
    out.est.table[2, 5] <- round(invfun2(optOut$par[5]), 4)

    out.vcovmat.trans[1:5, 1:5] <- solve(optOut$hessian)

    out.est.table[3, ] <- c(invfun1(optOut$par[1:4] - sqrt(diag(out.vcovmat.trans)[1:4]) *
        qnorm(1 - alpha/2)), invfun2(optOut$par[5] - sqrt(diag(out.vcovmat.trans)[5]) *
        qnorm(1 - alpha/2)))
    out.est.table[4, ] <- c(invfun1(optOut$par[1:4] + sqrt(diag(out.vcovmat.trans)[1:4]) *
        qnorm(1 - alpha/2)), invfun2(optOut$par[5] + sqrt(diag(out.vcovmat.trans)[5]) *
        qnorm(1 - alpha/2)))

    # Concatenate the lower and upper confidence intervals into one
    # row
    out.est.table["ci", ] <- paste(round(out.est.table["ci.lower", ], 3),
        round(out.est.table["ci.upper", ], 3), sep = ", ")

    out.est.table <- cbind(c("$\\theta_1$", "$\\theta_2$", "$\\eta_1$",
        "$\\eta_2$", "$\\rho$"), t(out.est.table))
    colnames(out.est.table)[1] = "Parameter"
    # print(my_table)
    return(out.est.table)
}
```

Marginal survival function distribution

```
library(survival)

# Define the data
car.age <- c(1.66, 0.35, 2.49, 1.9, 0.27, 0.41, 0.59, 0.75, 2.23, 9.52,
    3.35, 1.64, 1.45, 1.7, 1.4, 4.98, 5.71, 4.99, 3.4, 1.6, 1.28, 0.31,
    0.65, 2.21, 3.16, 0.22, 2.61, 0.32, 3.97, 0.48, 0.01, 0.27, 2.95, 1.4,
    8.27, 0.02, 2.09, 0.29, 1.66, 12)

# Fit the Weibull survival model
fit <- survreg(Surv(car.age) ~ 1, dist = "weibull")

# Extract the estimated parameters
shape <- coef(fit)["(Intercept)"]
scale <- exp(coef(fit)["(Intercept)"])

# Define the time points for survival probability calculation
times <- seq(0, 12, by = 0.1)

# Calculate the survival probabilities from the parametric model
parametric_surv_probs <- 1 - pweibull(times, shape, scale)

# Plot the survival curves
plot(times, parametric_surv_probs, type = "l", col = "blue", lwd = 2, xlab = "Time",
```

```
    ylab = "Marginal Survival Probability")
```

```
# lines(times, empirical_surv_probs, type = 'l', col = 'red', lwd =
# 2) legend('topright', legend = c('Marginal Survival', 'Empirical'),
# col = c('blue', 'red'), lwd = 2)
```

Emperical Survival function distribution

```
# Define the parameters for the bivariate Weibull distribution
the1 <- 2.2576
the2 <- 0.9218
n1 <- 0.8812
n2 <- 0.7885
rh <- 0.9191
x_rb <- rbivweib(40, the1, the2, n1, n2, rh)

# Calculate the empirical survival probabilities
empirical_surv_probs <- 1 - ecdf(x_rb[, 1])(times)

plot(times, empirical_surv_probs, type = "l", col = "red", lwd = 2, xlab = "Time",
    ylab = "Emperical Survival Probability")
```

Data sets

Warranty Data

car.age	car.mileage
1.66	0.9766
0.35	0.2041
2.49	1.2392
1.90	0.9889
0.27	0.0974
0.41	0.1594
0.59	0.2128
0.75	0.2158
2.23	1.1187
9.52	4.7660
3.35	1.3827
1.64	0.5992
1.45	0.6925
1.70	0.7078
1.40	0.7553
4.98	2.5014
5.71	2.5380
4.99	2.6433
3.40	1.6494
1.60	0.7162
1.28	0.5922
0.31	0.1974
0.65	0.2030
2.21	1.2532
3.16	1.4796
0.22	0.0979
2.61	1.5062
0.32	0.2062
3.97	1.6888
0.48	0.3099
0.01	0.0028
0.27	0.0095
2.95	1.2600
1.40	0.8067
8.27	4.1425
0.02	0.0105
2.09	1.2302
0.29	0.0447
1.66	0.9766
12.00	5.7304

Bone density data

before.study	after.study
1.103	1.027
0.842	0.857
0.925	0.875
0.857	0.873
0.795	0.811
0.787	0.640
0.933	0.947
0.799	0.886
0.945	0.991
0.921	0.977
0.792	0.825
0.815	0.851
0.755	0.770
0.880	0.912
0.900	0.905
0.764	0.756
0.733	0.765
0.932	0.932
0.856	0.843
0.890	0.879
0.688	0.673
0.940	0.949
0.493	0.463
0.835	0.776

Weibull distribution plots

```r
curve(dweibull(x, shape=0.5, scale = 1), from=0, to=5,
      main = 'PDF of Weibull Distribution', # add title
      ylab = 'Density values', # change y-axis label
      xlab = 'time(t)',
      lwd = 2, # increase line width to 2
      lty = 4,
      col = "red"
      )

curve(dweibull(x, shape=1.0, scale=1.0),
      from=0, to=5, add=TRUE,
      lwd = 2, # increase line width to 2
      lty = 3,
      col = "blue")

curve(dweibull(x, shape=1.5, scale=1.0),
      from=0, to=5, add=TRUE,
      lwd = 2, # increase line width to 2
      lty = 2,
      col = "purple")

curve(dweibull(x, shape=2.0, scale=1.0),
```

```
      from=0, to=5, add=TRUE,
      lwd = 2, # increase line width to 2
      lty = 1,
      col = "green")

legend("topright", legend = c(expression(paste(alpha, " = ", 1,"   " , beta, " = ", 0.5,)),
                              expression(paste(alpha, " = ", 1,"   " , beta, " = ", 1.0,)),
                              expression(paste(alpha, " = ", 1,"   " , beta, " = ", 1.5,)),
                              expression(paste(alpha, " = ", 1,"   " , beta, " = ", 2.0,))),
lty = c(4,3,2,1),
col = c("red","blue","purple","green"))
```

Bibliography

Babu, G. J., C. R. Rao, and M. B. Rao (1992). Non-parametric estimation of specific exposure rate in risk and survival analysis. *Journal of American Statistical Association 87*. (Cited on page 19.)

Balakrishnan, N. and E. Cramer (2014). The art of progressive censoring. *Statistics for industry and technology*. (Cited on page 15.)

Balakrishnan, N. and D. Kundu (2013). Hybrid censoring: Models, inferential results and applications. *Computational Statistics and Data Analysis 57*(1). (Cited on page 14.)

Basu, A. P. (1988). 23 multivariate exponential distributions and their applications in reliability. In *Quality Control and Reliability*, Volume 7 of *Handbook of Statistics*. Elsevier. (Cited on page 26.)

Berkson, J. and L. Elveback (1960). Competing exponential risks, with particular reference to the study of smoking and lung cancer. *Journal of the American Statistical Association 55*(291). (Cited on page 16.)

Bhattacharya, S., B. Pradhan, and D. Kundu (2014). Analysis of hybrid censored competing risks data. *Statistics 48*(5). (Cited on pages 16 and 17.)

Block, H. W. and A. Basu (1974). A continuous, bivariate exponential extension. *Journal of the American Statistical Association 69*(348). (Cited on page 25.)

Boos, D. D., L. A. Stefanski, et al. (2013). *Essential statistical inference*. Springer. (Cited on page 19.)

Byrd, R. H., P. Lu, J. Nocedal, and C. Zhu (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal on scientific computing 16*(5). (Cited on page 42.)

Collett, D. (1994). Modelling survival data. In *Modelling survival data in medical research*. Springer. (Cited on page 1.)

Cox, D. R. (1959). The analysis of exponentially distributed life-times with two types of failure. *Journal of the Royal Statistical Society: Series B (Methodological) 21*(2). (Cited on page 16.)

Crowder, M. (1991). On the identifiability crisis in competing risks analysis. *Scandinavian Journal of Statistics*. (Cited on page 18.)

Crowder, M. (1994). Identifiability crises in competing risks. *International Statistical Review/Revue Internationale de Statistique*. (Cited on page 18.)

Crowder, M. J. (2012). *Multivariate survival analysis and competing risks*. CRC Press. (Cited on pages 10 and 16.)

David, H. A. and M. L. Moeschberger (1978). *The Theory of Competing Risks: HA David, ML Moeschberger*. C. Griffin. (Cited on page 16.)

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society: series B (methodological) 39*(1). (Cited on page 22.)

Downton, F. (1970). Bivariate exponential distributions in reliability theory. *Journal of the Royal Statistical Society: Series B (Methodological) 32*(3). (Cited on pages 25 and 26.)

Efron, B. (1967). The two sample problem with censored data. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, Volume 4. (Cited on page 18.)

Epstein, B. (1954). Truncated life tests in the exponential case. *The Annals of Mathematical Statistics 25*(3). (Cited on page 14.)

Feizjavadian, S. and R. Hashemi (2015). Analysis of dependent competing risks in the presence of progressive hybrid censoring using marshall–olkin bivariate weibull distribution. *Computational Statistics and Data Analysis 82*. (Cited on pages 19, 35 and 36.)

Gradshteyn, I. and I. Ryzhik (2000). Table of integrals, series, and products 6th edn (new york: Academic). (Cited on page 28.)

Gross, A. J. and V. A. Clark (1975). *Survival distributions: reliability applications in the biomedical sciences*, Volume 11. Wiley New York. (Cited on page 15.)

Guess, F. M., J. S. Usher, and T. J. Hodgson (1991). Estimating system and component reliabilities under partial information on cause of failure. *Journal of Statistical Planning and Inference 29*. (Cited on page 17.)

Gumbel, E. J. (1960). Bivariate exponential distributions. *Journal of the American Statistical Association 55*(292). (Cited on page 25.)

Henrich, G. and U. Jensen (1995). Parameter estimation for a bivariate lifetime distribution in reliability with multivariate extension. *Metrika 42*(1). (Cited on page 25.)

Hosmer, D. W., S. Lemeshow, and S. May (1999). Regression modeling of time to event data. *Lifetime Data Analysis*. (Cited on page 15.)

Hougaard, P. (2014). Multivariate interval-censored survival data: Parametric, semi-parametric and non-parametric models. In *Statistical Modelling in Biostatistics and Bioinformatics*. Springer. (Cited on pages 10 and 11.)

Johnson, R. A., D. W. Wichern, et al. (2002). Applied multivariate statistical analysis. (Cited on page 71.)

Jung, M. and D. S. Bai (2007). Analysis of field data under two-dimensional warranty. *Reliability Engineering & System Safety 92*(2). (Cited on page 68.)

Kalbfleish, J. (1980). The statistical analysis of failure time data. (Cited on pages 1 and 18.)

Kaplan, E. L. and P. Meier (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association 53*(282). (Cited on page 18.)

Kundu, D. and S. Basu (2000). Analysis of incomplete data in presence of competing risks. *Journal of Statistical Planning and Inference 87*(2). (Cited on pages 17 and 19.)

Kundu, D. and S. Basu (2003). *Analysis of incomplete data in presence of dependent competing risks*. Alpha Science Int'l Ltd. (Cited on page 19.)

Kundu, D., N. Kannan, and M. Mazumdar (1992). Inference on risk rates based on mortality data under censoring and competing risks using parametric models. *Biometrical Journal 34*(3). (Cited on page 19.)

Kundu, D., D. Mitra, and A. Ganguly (2017). Analysis of left truncated and right censored competing risks data. *Computational Statistics & Data Analysis 108*. (Cited on page 17.)

Kundu, D. and B. Pradhan (2011). Bayesian analysis of progressively censored competing risks data. *Sankhya B 73*. (Cited on page 17.)

Kundu, D. and A. M. Sarhan (2006). Analysis of incomplete data in presence of competing risks among several groups. *IEEE Transactions on Reliability 55*(2). (Cited on page 17.)

Lagakos, S. W. (1979). General right censoring and its impact on the analysis of survival data. *Biometrics 35*(1). (Cited on page 15.)

Lawless, J. F. (1982). *Statistical models and methods for lifetime data*. John Wiley & Sons, New York. (Cited on pages 1 and 35.)

Lawless, J. F. (2003). *Statistical Models and Methods for Lifetime Data* (Second ed.). University of Waterloo: John Wiley Publications. (Cited on pages 5, 6, 10, 11 and 16.)

Leung, K. M., R. M. Elashoff, and A. A. Afifi (1997). Censoring issues in survival analysis. *Annual Review of Public Health 18*(1). PMID: 9143713. (Cited on page 12.)

Lin, D., W. Sun, and Z. Ying (1999). Nonparametric estimation of the gap time distribution for serial events with censored data. *Biometrika 86*(1). (Cited on page 2.)

Marshall, A. W. and I. Olkin (1967). A multivariate exponential distribution. *Journal of the American Statistical Association 62*(317). (Cited on page 35.)

Miyakawa, M. (1982). Analysis of incomplete data in competing risks model. *Journal of the Japanese Society of Quality Control 12*. (Cited on pages 17 and 19.)

Miyakawa, M. (1984). Analysis of incomplete data in competing risks model. *IEEE Transactions on Reliability 33*(4). (Cited on page 17.)

Moeschberger, M. and J. P. Klein (1996). Statistical methods for dependent competing risks. *Lifetime Data: Models in Reliability and Survival Analysis*. (Cited on page 19.)

Nelson, W. (1970). Hazard plotting methods for analysis of life data with different failure modes. *Journal of Quality Technology 2*. (Cited on page 17.)

Pareek, B., D. Kundu, and S. Kumar (2009). On progressively censored competing risks data for weibull distributions. *Computational Statistics & Data Analysis 53*(12). (Cited on page 17.)

Park, C. and K. Kulasekera (2004). Parametric inference of incomplete data with competing risks among several groups. *IEEE Transactions on Reliability 53*(1). (Cited on page 17.)

Peterson Jr, A. V. (1977). Expressing the kaplan-meier estimator as a function of empirical subsurvival functions. *Journal of the American Statistical Association 72*(360a). (Cited on page 18.)

Pratt, W. (1968). Partial differentials of marcum's q function. *Proceedings of the IEEE 56*(7). (Cited on page 38.)

Prentice, R. L., J. D. Kalbfleisch, A. V. Peterson Jr, N. Flournoy, V. T. Farewell, and N. E. Breslow (1978). The analysis of failure times in the presence of competing risks. *Biometrics*. (Cited on pages 16 and 18.)

Radke, B. R. (2003). A demonstration of interval-censored survival analysis. *Preventive Veterinary Medicine 59*(4). (Cited on page 15.)

Ren, J. and W. Gui (2021). Statistical analysis of adaptive type-ii progressively censored competing risks for weibull models. *Applied Mathematical Modelling 98*. (Cited on page 2.)

Samanta, D. and D. Kundu (2021). Bayesian inference of a dependent competing risk data. *Journal of Statistical Computation and Simulation 91*(15). (Cited on page 19.)

Suttner, L. H. (2019). *Statistical Methods for Censored and Missing Data in Survival and Longitudinal Analysis*. Ph. D. thesis, University of Pennsylvania. (Cited on page 15.)

Tang, N., A. J. Turkson, F. Ayiah-Mensah, and V. Nimoh (2021). Handling censoring and censored data in survival analysis: A standalone systematic literature review. *International Journal of Mathematics and Mathematical Sciences 2021*. (Cited on pages 12, 14 and 15.)

Wang, L., Y. M. Tripathi, C. Lodhi, and X. Zuo (2022). Inference for constant-stress weibull competing risks model under generalized progressive hybrid censoring. *Mathematics and Computers in Simulation 192*. (Cited on page 2.)

Weibull, W. (1939). Ingeniörsvetenskapsakademiens handlingar. *A statistical theory of the strength of materials (Stockholm: Generalstabens litografiska anstalts förlag) 151*. (Cited on page 8.)

Wu, M., Y. Shi, and C. Zhang (2017). Statistical analysis of dependent competing risks model in accelerated life testing under progressively hybrid censoring using copula function. *Communications in Statistics-Simulation and Computation 46*(5). (Cited on page 19.)