INFERENTIAL PROCEDURES FOR THE GENERALIZED EXPONENTIAL MODEL HAVING COVARIATE, WITH RIGHT AND INTERVAL CENSORED DATA

By

ALHARBI NADA MOHAMMEDSAEED M

Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia, in Fulfillment of the Requirements for the Degree of Doctor of Philosophy

May 2023

IPM 2023 2

© COPYRIGHT UPM

COPYRIGHT

All material contained within the thesis, including without limitation text, logos, icons, photographs and all other artwork, is copyright material of Universiti Putra Malaysia unless otherwise stated. Use may be made of any material contained within the thesis for non-commercial purposes from the copyright holder. Commercial use of material may only be made with the express, prior, written permission of Universiti Putra Malaysia.

Copyright © Universiti Putra Malaysia

© COPYRIGHT UPM

DEDICATIONS

To my loving parents for their unconditional love.
To my siblings for their understanding, affection, prayers and wholeheartedness.
To my husband and my kids.
To all of my love

© COPYRIGHT UPM

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfillment of the requirement for the degree of Doctor of Philosophy

INFERENTIAL PROCEDURES FOR THE GENERALIZED EXPONENTIAL MODEL HAVING COVARIATE, WITH RIGHT AND INTERVAL CENSORED DATA

By

ALHARBI NADA MOHAMMEDSAEED M

May 2023

Chairman: Associate Professor Dr. Jayanthi a/p Arasan
Institute: Institute for Mathematical Research

In literature, there are various studies that incorporate censoring mechanisms to the generalized exponential model (GEM). This research aims to analyse generalized exponential models in the presence of right and interval-censored data with fixed covariates. The analysis starts with the extension of the GEM to incorporate fixed covariates in the presence of right and interval censored data. Parameters of the models under both censoring were estimated using the maximum likelihood estimation (MLE) method. The performance of these estimates were assessed at various sample sizes (n) and censoring proportion (cp) via the bias, standard error (SE) and root mean square error (RMSE). Next the model was extended to incorporate interval censored data with covariate. The performance of the MLE using the midpoint, right, left, random imputations were compared at various sample sizes and censoring proportions via a simulation study.

In addition, three asymptotic confidence interval procedures which included Wald, likelihood ratio, and score confidence intervals procedures were investigated through a coverage probability study when the data were both right and interval censored at various n and cp. Then, five alternative confidence intervals procedures, which included the jackknife, bootstrap-normal, bootstrap-t, bootstrap-p, bias correction acceleration bootstrap procedures were studied via a coverage probability study. This simulation study showed that overall, the Wald asymptotic and bootstrap normal alternative confidence intervals methods are recommended as a suitable inferential to estimate the parameters of the model using different sample sizes, interval length and censoring proportions.

© COPYRIGHT UPM

In summary, the simulation studies for each category indicate that the bias, standard error, and root mean square error are large when the cp is high, which indicates that the estimators perform better when the sample size is large, and the cp is low. Furthermore, the performance of the asymptotic confidence interval estimate indicates that the Wald confidence interval for the parameter β_1 in the generalized exponential model, under both right and interval censoring, represent the most effective approach. In comparison to alternative confidence intervals, the bootstrap normal (b-n) method yields results significantly closer to the nominal error probability for parameters β_0 and β_1.

Finally, to further support the findings of the simulation studies, we employ two real datasets with right and interval-censored data from lung and breast cancer datasets, respectively. The first dataset is an interval censored data from a breast cancer study with age as the covariate. The second dataset consists of right censored lung cancer data with age as the covariate. The results indicated that the GEM was a better fit for both datasets compared to the exponential distribution. The confidence interval estimation techniques were obtained for the covariate parameter of both models. Additionally, the findings of the real data indicate that the Wald method for the covariate β_1 is significant within the context of the lung cancer data. For the breast censer data with age as the covariate, the bootstrap normal, bootstrap-t, BCa and the jackknife have a similar confidence interval for μ, α, β_0 and β_1. The results indicate that the generalized exponential model outperforms the submodel based on the exponential distribution.

© COPYRIGHT UPM

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk ijazah Doktor Falsafah

PROSEDUR INFERENSI BAGI MODEL EKSPONEN TERITLAK MEMPUNYAI KOVARIAT, DENGAN DATA TERTAPIS KE KANAN DAN SELANG TERTAPIS

Oleh

ALHARBI NADA MOHAMMEDSAEED M

Mai 2023

Pengerusi: Prof Madya Dr. Jayanthi a/p Arasan
Institut: Fakulti Sains

Dalam tinjauan literatur, terdapat pelbagai kajian yang menggabungkan mekanisma tapisan kepada taburan eksponen umum. Penyelidikan ini bertujuan untuk menganalisis model eksponen teritlak terhadap kehadiran data tertapis kanan dan selang dengan kovariat tetap. Analisis dimulakan dengan melanjutkan model eksponen teritlak bagi menggabungkan kovariat tetap dengan kehadiran data tertapis kanan. Parameter bagi model bagi tertapis kanan dianggar dengan menggunakan kaedah anggaran kebolehjadian maksimum (MLE). Prestasi anggaran ini dinilai pada pelbagai saiz sampel (n) dan kadaran data tertapis (cp) melalui pincang, ralat piawai (SE) dan punca min ralat kuasa dua (RMSE). Seterusnya, model tersebut dilanjutkan bagi mengabungkan data tertapis selang dengan kehadiran kovariat. Prestasi MLE yang menggunakan imputasi titik tengah, kanan, kiri, rawak dibandingkan pada pelbagai saiz sampel dan kadaran data tertapis menerusi kajian simulasi.

Di samping itu, tiga prosedur selang keyakinan asimptotik termasuk Wald, nisbah kebolehjadian dan prosedur selang keyakinan skor disiasat menerusi kajian kebarangkalian liputan apabila data tertapis secara kanan dan selang pada pelbagai n dan cp. Seterusnya, lima prosedur selang keyakinan alternatif, iaitu prosedur selang keyakinan $jackknife$, $bootstrap-normal$, $bootstrap-t$, $bootstrap-p$, dan $bootstrap$ pembetulan kepincangan telah dikaji menerusi kajian kebarangkalian liputan. Hasil kajian simulasi ini menunjukkan bahawa secara keseluruhan, selang keyakinan Wald dan selang keyakinan bootstrap-normal disyorkan sebagai inferensi yang sesuai untuk menganggar parameter model menggunakan sampel saiz, panjang selang dan kadaran tertapis yang berbeza.

© COPYRIGHT UPM

Secara ringkasnya, kajian simulasi untuk setiap kategori menunjukkan bahawa pincang, ralat piawai, dan punca ralat kuasa dua adalah besar apabila cp adalah tinggi, yang menunjukkan bahawa penganggar berfungsi lebih baik apabila saiz sampel besar, dan cp adalah rendah. Selain itu, prestasi anggaran selang keyakinan asimptotik menunjukkan bahawa selang keyakinan Wald untuk parameter β_1 dalam model eksponen teritlak, di bawah kedua-dua tapisan kanan dan selang, mewakili pendekatan yang paling berkesan. Berbanding dengan selang keyakinan alternatif, kaedah bootstrap normal (b-n) memberikan hasil yang lebih hampir secara signifikan dengan kebarangkalian ralat nominal untuk parameter β_0 dan β_1.

Akhirnya, untuk menyokong penemuan kajian simulasi ini, kami menggunakan dua set data tertapis kanan dan selang sebenar kanser paru-paru dan kanser payudara, masing-masing. Set data pertama adalah data yang tertapis selang dari kajian kanser payudara dengan usia sebagai pembolehubah kovariat. Set data kedua terdiri daripada data kanser paru-paru yang tertapis kanan dengan usia sebagai pembolehubah kovariat. Keputusan menunjukkan bahawa GEM adalah lebih padan untuk kedua-dua set data berbanding dengan taburan eksponen. Teknik penentuan selang keyakinan diperoleh untuk parameter kovariat kedua-dua model. Selain itu, penemuan data sebenar menunjukkan bahawa kaedah Wald untuk kovariat β_1 adalah signifikan dalam konteks data kanser paru-paru. Bagi data kanser payudara dengan usia sebagai kovariat, penganggaran selang keyakinan yang diperolehi adalah hampir sama untuk μ, α, β_0 dan β_1 menggunakan kaedah bootstrap normal, $bootstrap-t$, BCa, dan $jackknife$. Hasil menunjukkan bahawa model eksponen teritlak adalah lebih baik berbanding submodel berdasarkan taburan eksponen.

© COPYRIGHT UPM

ACKNOWLEDGEMENTS

First and foremost, I am grateful to ALLAH (s.w.t.) for giving me the strength and patience to complete this thesis.

I am extremely indebted to my supervisor, Assoc. Prof. Dr. Jayanthi Arasan for her excellent supervision, invaluable guidance, patience and support through the past years. Her charisma and enthusiasm during these years has made these challenging life period a very enjoyable and useful experience. She is the one always leading me to keep faith during some difficult times. She really deserves special recognition because without her help, this work would not be accomplished. I have learned immensely from her not only in survival analysis but also in various other aspects of life. She has helped me so much in so many ways that I do have not enough words to express them.

To my family, I am outmost grateful to my beloved mother for her unlimited love, care, support and encouragement, without my mother and her prayers I would not be at this stage of my career. My special thanks go to my kids (Sadeem, Asseel, Raneem, Alhakm, Jaser, and Misk) their sacrifices and care. Also my thanks go to all my relatives and friends for their encouragement. My final thanks go to Taibah University, Al Madinah Al Munawwarah, Saudi Arabia, for offering me the opportunity to complete my studies in Malaysia.

© COPYRIGHT UPM

This thesis was submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfilment of the requirement for the degree of Doctor of Philosophy. The members of the Supervisory Committee were as follows:

Jayanthi a/p Arasan, PhD
Associate Professor
Faculty of Science
Universiti Putra Malaysia
(Chairman)

Wendy Ling Shin Yie, PhD
Senior Lecturer
Faculty of Science
Universiti Putra Malaysia
(Member)

Nur Haizum Abd.Rahman, PhD
Senior Lecturer
Faculty of Science
Universiti Putra Malaysia
(Member)

ZALILAH MOHD SHARIFF, PhD
Professor and Dean
School of Graduate Studies
Universiti Putra Malaysia

Date: 14 December, 2023

© COPYRIGHT UPM

TABLE OF CONTENTS

© COPYRIGHT UPM

© COPYRIGHT UPM

© COPYRIGHT UPM

LIST OF TABLES

© COPYRIGHT UPM

© COPYRIGHT UPM

© COPYRIGHT UPM

© COPYRIGHT UPM

© COPYRIGHT UPM

© COPYRIGHT UPM

LIST OF FIGURES

© COPYRIGHT UPM

© COPYRIGHT UPM

LIST OF ABBREVIATIONS

AC	Anti-Conservative
AS	Asymmetrical
AFT	Accelerated failure time
C	Conservative
CI	Confidence Interval
Cp	Censoring proportion
CDF	Cumulative distribution function
DIC	Doubly interval-Censored
EO	Exactly Observed
GE	Generalized Exponential
GEMR	Generalized Exponential Model with
RCDFC	Right-Censored Data and Fixed Covariate
GEMI	Generalized Exponential Model with
ICDFC	Interval-Censored Data and Fixed Covariate
GEMDIC	Generalized exponential with
DICD	Doubly Interval-Censored Data
GEM	Generalized Exponential Model
IC	Interval-Censored
ICDT	Interval-censored and doubly truncated
IME	Inverse moment estimators
LC	Left-Censored
MLE	Maximum likelihood estimate
n	Sample size
LR	Likelihood ratio
JK	Jackknife

© COPYRIGHT UPM

n-b	Normal bootstrap
b-t	Bootstrap-t
b-p	Percentile bootstrap
BCa	Bias-corrected and accelerated bootstrap
pdf	Probability distribution function
SE	Standard error
SCIs	Simultaneous Confidence Intervals
RMSE	Root Mean Square Error
RC	Right-Censored
t	Time
TD	Time Dependent
MCMC	Markov chain Monte Carlo
MOIPL	Marshall-Olkin inverse power Lomax
BLUEs	Best linear unbiased estimates

© COPYRIGHT UPM

CHAPTER 1

INTRODUCTION

1.1 Survival Analysis

Survival analysis is described as a collection of statistical methods used for a dataset, which is estimated using a well-defined time variable denoted as T until the occurrence of a specific event for an individual or observation. In survival studies, the time variable T is a non-negative continuous or discrete variable representing failure, duration, or survival time. According to Kleinbaum and Klein (2012) this period is usually measured in days, weeks, months, or years from the time the individual joins the study until the occurrence of significance, such as death, recurrence, or disease incidence, occurs. Lawless (2011) pointed out that the time variable needs a well-defined reference point that refers to as the time origin; $t = 0$; the starting time from which the individual's survival is measured, which means results in the incidence of the event of interest and, eventually, a reason for the observation time to be ended.

Kalbfleisch and Prentice (2011) emphasized that, in any survival study, the investigator should verify the starting point and the study endpoint. In other words, survival analysis can be described as the study of time-dependent data sets i.e., time to failure of a mechanical or physical component from time of installation in a particular machine or time to death for a patient diagnosed with leukemia. So also, survival analysis equally plays an essential role in industrial or reliability studies, although the methodologies involved are most commonly used in the biomedical field (Jennison and Turnbull, 1999). The aim of clinical research is often to identify and estimate the survival distribution of time T, by estimating the probability that an individual survives from a specified time t or beyond (Kleinbaum and Klein, 2012). Mathematically, the survival of these patients can be expressed as a function $S(t) = Pr(T \geq t)$. Due to the instantaneous risk at failure time t, the conditional probability of dying given that patients have survived up to time t can be obtained using the hazard function, $h(t)$. Furthermore, because some individuals may be diagnosed with a disease, for instance lung cancer disease, at random time points after the study begins, a straightforward method is to shift the start time to the time origin, $t = 0$ (Klein and Moeschberger, 2006).

The presence of incomplete observation of the survival times due to censoring and truncation mechanisms is another unique feature of survival data. If censoring is present in a survival study, the exact survival time or time to failure of a randomly selected individuals remains unknown to the researcher. For instance, in the cancer study discussed earlier, some individuals may have survived death even after the end of the study.

© COPYRIGHT UPM

Therefore, the exact death time for these individuals will be unknown, because these times are longer than the end time point of the study. Additionally, some of these individuals maybe sampled from a medical registry records due to cost or time constraints. The observation times of these individuals might be truncated due to a selection mechanism applied to select only individuals who are free from failure into the study. If no assumptions are made on the distribution of T, a non-parametric method can be applied to estimate the survival probabilities of the observations in the study. The non-parametric methods that are most commonly used are the Kaplan-Meier and the Nelson Aalen estimators. Also, a semi-parametric model known as Cox's proportional hazard model or parametric regression models can be applied (Mitra, 2013).

1.2 Basic Formulation in Survival Analysis

Survival time, T, is a non-negative random variable which has a continuous distribution function. The survival, hazard, and cumulative hazard function can be defined as follows:

1.2.1 Survival Function

The essential quantity employed to describe failure time phenomena is the survival function, the probability of an individual survives beyond time t. It is defined as,

$$S(t) = Pr(T \geq t) = 1 - F(t),$$

where $F(t)$ is the cumulative distribution function (CDF). Meanwhile, the survival function is given by,

$$S(t) = Pr(T \geq t) = \int_t^{\infty} f(u)du,$$

thus,

$$f(t) = -\frac{dS(t)}{dt},$$

where $f(t)$ is the probability distribution function (pdf) for the random variable T.

The survival function is fundamental to a survival analysis because obtaining survival probabilities for different values of t provides crucial summary information from survival data (Kleinbaum et al., 2012). As t ranges from 0 to ∞, the survival function has the following properties:

- For $t = 0$:
$$S(0) = \int_0^{\infty} f(x)dx = 1$$

- For $t = \infty$:
$$S(\infty) = \lim_{t\to\infty} S(t) = \lim_{t\to\infty} \int_t^{\infty} f(x)dx = 0$$

© COPYRIGHT UPM

- $S(t)$ is a decreasing continuous function.

1.2.2 Hazard Function

The hazard function of T is also fundamental in survival analysis. Hazard function is also known as the conditional failure rate in reliability and can be defined as follows:

$$\begin{aligned} h(t) &= \lim_{\triangle t \to 0} \frac{1}{\triangle t} P(t \leq T < t + \triangle t \mid T \geq t) \\ &= \frac{f(t)}{S(t)}. \end{aligned}$$

Furthermore, Breslow (1975) emphasized that the relationship between the hazard and the survival functions can be derived as follows:

$$h(t) = \frac{f(t)}{S(t)} = \frac{-\frac{d}{dt}S(t)}{S(t)} = -\frac{d}{dt}\log S(t) \text{ for } t \geq 0.$$

Also,

$$S(t) = exp(-\int_0^t h(u)du) \text{ for } t \geq 0. \tag{1.1}$$

1.2.3 Cumulative Hazard Function

The cumulative hazard function, $H(t)$, is define as

$$H(t) = \int_0^t h(u)du \text{ for } t \geq 0. \tag{1.2}$$

By substituting (1.1) in (1.2), the relationship between the survival function and the cumulative hazard function can be written as follows:

$$S(t) = exp(-H(t)).$$

1.3 Censoring

In any medical survival study, a researcher must determine the start and endpoint of the study based on the fact that survival studies do not last indefinitely. Subsequently, this results in the incomplete observation of failure times, and the recorded time for these individuals is categorized as censored survival or failure time. In most clinical studies, three types of censoring, right, left, and interval censoring, are encountered (Bogaerts et al., 2017). These censoring mechanisms are given in more detail in the following sections.

© COPYRIGHT UPM

1.3.1 Right Censoring

In some survival study, all the individuals may have the same fixed and known censoring times. Usually this occurs when individuals enter the study at the same time origin $t = 0$ and the study terminates later at a time. In other words, the study comes to an end at a pre-determined end time and individuals can leave the study only through failure (Lawless et al., 2003). The right censoring are of three forms. They are:

1.3.1.1 Type I Censoring

Fixed type I censoring occurs when a study is designed to end after C years of follow-up. In this case, everyone who does not have an event observed during the course of the study is censored at C years. In a simple generalization of this scheme, each unit has a potential maximum observation time ψ_i for $i = 1, \ldots, n$ which may differ from one case to the next but is nevertheless fixed in advance. The probability that unit i will be alive at the end of her observation time is $S(\psi_i)$, and the total number of deaths is again random (Lawless et al., 2003).

1.3.1.2 Random Censoring

In random type I censoring, the study is designed to end after C years, but censored subjects do not all have the same censoring time. It is the most common type of right-censoring use in several studies. In addition to the grounds that random censoring occurs because the study has come to an end, an observation is equally censored for a reason that individuals drop out of the study because the patient wishes to discontinue treatment or lost to the study at a random point due to migration or death related to other possibilities regardless of the event being observed (Bogaerts et al., 2017).

1.3.1.3 Fixed Type II Censoring

In type II censoring, a study ends when there is a pre-specified number of events. A sample of n units is followed as long as necessary until d units have experienced the event. In this design the number of deaths d, which determines the precision of the study, is fixed in advance and can be used as a design parameter. Unfortunately, the total duration of the study is then random and cannot be known with certainty in advance. Other forms of Type-I and Type-II right censoring includes the generalized Type-I, progressive Type-I and progressive Type-II right censoring, see (Lawless, 2011; Klein and Moeschberger, 2006).

1.3.2 Left Censoring

Left censoring is also a common censoring mechanism that takes place in any medical setting. Under this type of censoring, an event of interest has already occurred

© COPYRIGHT UPM

or has been experienced by an individual even before he/she is enrolled in the study. Thus, the exact survival time for this individual remains unknown. Left censoring is most likely to occur when you begin observing a sample at a time when some of the individuals may have already experienced the event. The survival times for the i^{th} individual is left-censored (LC) when $t_i < c_i$ and $t_i \in (0, l_i]$ with l_i the left-censored survival times.

1.3.3 Interval Censoring

Interval censoring arises when the exact event time is unknown except to fall between pre-specified interval. Since death is not the only event to be observed in medical settings, disease progression has equally gained interest or concern among medical practitioners. There are several different types of interval-censored data. They are as follows:

1. Case I interval-censored data, also called current status data, arises when each individual is subjected to observation only at a single follow-up time, and thus, the event of interest (failure) is only observed either to have or have not occurred before the observation time. That is, the failure time of interest is either left- or right-censored data (Keiding, 1991; Groeneboom and Wellner, 1992; Koul and Yi, 2006).

   Case I interval-censored data usually occur in tumorigenicity tests. In these tests, the tumor start time of animals is always of prime interest but not observable. Instead, tumor status is commonly known at death (either natural death for scientific study). Hence, the tumor start time is expected to be less or greater than the death time. There are many authors who discussed the current status data arising from survival studies such as (Huang and Wellner, 1995; Huang et al., 1996; Rossini and Tsiatis, 1996; Lin et al., 1998; Shen, 2000; Ghosh, 2001; Martinussen and Scheike, 2002; Xue et al., 2004; Sun, 2007).

2. Case II interval-censored data, also known as general interval-censored data, is defined as data which refers to a situation when the event of interest cannot be immediately observed and is only known to have appeared through a random interval of time. Left (right) censoring is a special case of interval censoring in which the left (right) endpoint is $0(\infty)$.

   Case II interval-censored data arises in several medical and health studies. For example, in a study which compares time to cosmetic deterioration of breasts for breast cancer patients treated with radiotherapy and radiotherapy plus chemotherapy, patients were examined at each clinical visit for breast retraction. The breast retraction is only known to occur between two clinical visits or right-censored at the end of the study. The study objective is to compare the patients who received adjuvant chemotherapy to those who did

© COPYRIGHT UPM

not and to decide whether chemotherapy affects the rate of deterioration of the cosmetic state (Finkelstein, 1986; Pan, 2000; Lim and Sun, 2003; Huang et al., 2008).

3. Mixed interval-censored data refers to the survival time of interest, which is observed either to belong to an interval, or to be in right-censoring (Zhao and Sun, 2004).

4. Partly interval-censored data arise when the exact failure times are observed of some subjects but, for the remaining subjects, the failure time of interest is not observable, but is only known to be bracketed between two examination times (Huang, 1999). Example of partly interval-censored data is given by the Framingham Heart Disease study. In this research, times of the first occurrence of the subcategory angina pectoris in coronary heart disease patients are of interest. For some patients, the event time is recorded precisely, but for the remaining patients, time is recorded only between two clinical examinations see (Feinleib et al., 1975; Odell et al., 1992).

1.3.4 Independent Random Censoring

A more realistic and common postulation in a clinical study is to assume that the censoring times are random and independent of the survival or failure times. In other words, the random censoring mechanism is non-informative or contains no information of the survival/failure times of all observations in a survival study. Thus, under independent random censoring, all individuals in the study have the same probability being censored despite the fact that the risk/hazard of failure for some individuals may be higher than the others.

Subsequently for RC data, let the failure time T be independent continuous random variable with common survival function, density function and hazard function given as $S(t)$, $f(t)$ and $h(t)$ respectively. Also, let the censoring time C be continuous random variable with common survival function $G(t)$ and density function $g(t)$. Assuming that the censoring times are non-informative of the failure times, we know that the distribution of $G(t)$ does not rely on any parameters in $S(t)$ (Lawless, 2011).

Independent interval censoring is the condition whereby the method that generates the censoring is independent of the subject's failure time distribution. For instance, T is failure time of interest and L and R are the two observed values such that $T \in (L,R]$. Then the independent random censoring process for interval-censored data can be expressed by,

$$P(L < T \leq R \mid L = l, R = r) = P(l < T \leq r). \tag{1.3}$$

That is, the joint survival function of the two observed values L and R is free from any

© COPYRIGHT UPM

parameters contributory in the survival function of T. More importantly, it should be noted that the independent interval censoring is non- informative interval censoring while the opposite is not always true (Betensky, 2000; Oller et al., 2004; Sun, 2007).

1.4 Survival Models

In any survival study, estimation of the survival or hazard function is crucial in providing an overview of failure/hazard rate of an individual following the event of interest being studied. Furthermore, it is common in clinical studies to have conditions where several (known) covariates or risk factors such as age and blood pressure level possibly affect patient diagnosis and are recorded to investigate the effect of these variables on the survival times of an individual (Clark et al., 2003). The modeling can be achieved using either the non-parametric, semi-parametric or parametric approaches. In this research, the parametric approach is considered.

1.5 Parametric Approach

Parametric regression often remain a useful tool as it is fitted much faster compared to the semi-parametric models (Venables and Ripley, 2013). Also, under conditions such as the dependency of the survival times on the covariates, either fixed or time-dependent, and when parameter values are far from zero, an asymptotically adequate parametric model offers more efficient estimates as they are based on fewer parameters compared to a semi-parametric model (Cox and Oakes, 1984; Nardi and Schemper, 2003). Some of the advantages of the parametric models are the existence of straightforward methods to obtain the maximum likelihood estimation (MLE) of the parameters, confidence intervals (CI) and hypothesis testing procedures. Also, in a parametric framework, the survival times T is a continuous random variable with a specified distribution function $f(t)$. Sometimes it is vital to extend the existing parametric distributions to incorporate information on the censoring mechanism and covariates. Some of the most commonly used parametric regression models are the extensions of the Weibull, exponential, and the log-normal, which differs from each other with different hazard functions (Bradburn et al., 2003).

1.6 Types of Covariates

Sometimes it is necessary to measure the effect of covariates in survival analysis study. The two types of covariates that are usually encountered in any survival study are the fixed and time-dependent covariates. Fixed covariates are commonly measured at the beginning time point of the study, but these covariates can be equally measured at the middle or end of the study. The covariates are constant throughout the study. Examples of such covariates are the gender and ethnicity of individuals. On the other hand, time-dependent covariates vary over time and are measured on regular basis for an individual in a study. By accessing the record of a time-dependent covariate up to a specified time t, a researcher is able to study the continuous effect

© COPYRIGHT UPM

of these variables on the survival time T. For instance, accounting for the change in the level of covariates such as age, glucose level, blood pressure, or tumor sizes provides an up-to-date effect of these variables on the hazard and survival rate of the individuals in the study. It provides a more reliable prognosis of the future life expectancy comparatively when these covariates are measured only at the time origin (Collett, 2015).

Fixed-time covariates are those whose values are fixed throughout the study, examples are sex, and race, among others. It is typical in many survival studies that individuals are monitored for the duration of the study, and some explanatory variables are recorded whose values may change during the study. For example, the status of neutrophil recovery of patients with leukemia after transplantation (discrete-time) and the number of CD4 T-cells of HIV/AIDS patients measured at irregular intervals (continuous).

Also, some external covariates may influence the survival time of an individual at time t which however exist independently. Such covariate may exist in respiratory survival studies, where the presence of air pollutant may affect the life span of individuals with heart disease or lung cancer where the change in the air quality is independent of any particular individual in the study (Collett, 2015; Kalbfleisch and Prentice, 2011; Kiani, 2012). Another approach is to extend parametric models where the covariates are often modeled through the mean function by incorporating in the hazard function as proposed by (Sparling et al., 2006; Arasan and Lunn, 2008; Kiani, 2012) among others.

1.7 Multiple Imputation

Missing data usually occur in various fields of study due to several reasons. The most popular method used in estimating the missing values is the imputation. For example, in medical and health studies, it is required for the participants to undergo periodic follow-ups for the examination of characteristics related to the condition of interest. In handling the missing observations various imputation methods, such as single imputation, multiple imputation, etc are used (Sterne et al., 2009).

Single imputation is often applied because it is intuitively attractive. In single imputation, we fill in each missing value by a single predicted value. The obvious shortcoming of single imputation is that we replace the missing value with a single value. Hence, single imputation ignores uncertainty and always underestimates the variance. Multiple imputation method on the other hand, rectify the shortcoming of single imputation by taking into account both within imputation uncertainty and between imputation uncertainty. In other words, multiple imputations use multiple values to estimate the unknown missing observation. In survival analysis, the multiple imputation method is used in estimating missing interval-censored failure time

© COPYRIGHT UPM

data. The method reduces the interval-censored data to right-censored data, which can be handled using specified methods for right-censored data (Pan, 2000; Chen and Sun, 2010; Kiani, 2012; Manoharan, 2018). Some of the research in the literature that employed the imputation methods for interval-censored data are the works of (Dorey et al., 1993; Satten et al., 1998; Betensky and Finkelstein, 1999; Bebchuk and Betensky, 2000; Pan, 2001).

Imputation is a common approach to dealing with missing data. A straightforward method is to use right limit for the period in which the infection time is censored to impute the infection time. The procedure is called the right end imputation. Furthermore, another method called the left-end imputation method involves considering the first inspection time for the imputations. Lastly, a random imputation in which a random time is selected within the interval between the first inspection time, and the last inspection time.

1.8 The Generalized Exponential Distribution

The two parameter exponential distribution has been applied in real-life studies in the literature. A random variable T is assumed to follow the two-parameter exponential distribution if T has the cumulative distribution function as,

$$F(t;\lambda,\mu) = (1-e^{-(t-\mu)/\lambda}), \quad t > \mu, \alpha > 0, \lambda > 0, \tag{1.4}$$

with corresponding probability density function,

$$f(t;\lambda,\mu) = \frac{1}{\lambda} e^{-(t-\mu)/\lambda}, \quad t > \mu, \lambda > 0. \tag{1.5}$$

The two parameter exponential distribution was generalized by the introduction of a shape parameter (also called exponent parameter) by Gupta and Kundu (1999) to have a three-parameter generalised exponential (GE) distribution. This shape parameter enables the distribution to have different shapes which makes it more flexible when used in fitting lifetime data. The GE distribution has the following cumulative distribution function (CDF),

$$F(t;\alpha,\lambda,\mu) = (1-e^{-(t-\mu)/\lambda})^{\alpha}, \quad t > \mu, \alpha > 0, \lambda > 0, \tag{1.6}$$

and probability density function (pdf),

$$f(t;\alpha,\lambda,\mu) = \frac{\alpha}{\lambda}(1-e^{-(t-\mu)/\lambda})^{\alpha-1} e^{-(t-\mu)/\lambda}, t > \mu, \alpha > 0, \lambda > 0, \tag{1.7}$$

where α, λ, and μ are respectively the shape, scale, and location parameters of the distribution. Figure 1.1 shows the plots of pdf of GE distribution with different parameter values.

© COPYRIGHT UPM

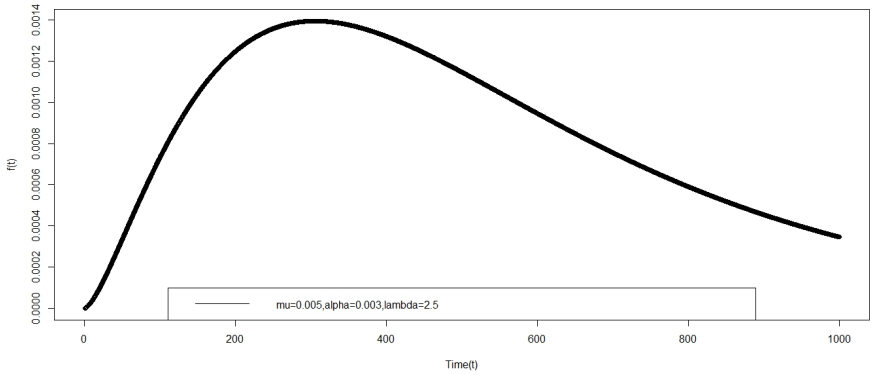

Figure 1.1: Graph of the pdf of GE with various parameter values

Meanwhile, the survival function and hazard function are given respectively by:

$$S(t;\alpha,\lambda,\mu) = 1 - F(t;\alpha,\lambda,\mu) = 1 - (1 - e^{-(t-\mu)/\lambda})^{\alpha}, \quad t > \mu, \lambda, \alpha > 0. \tag{1.8}$$

$$h(t;\alpha,\lambda,\mu) = \frac{f(t;\alpha,\lambda,\mu)}{S(t;\alpha,\lambda,\mu)} = \frac{\alpha}{\lambda} \frac{(1 - e^{-(t-\mu)/\lambda})^{\alpha-1} e^{-(t-\mu)/\lambda}}{1 - (1 - e^{-(t-\mu)/\lambda})^{\alpha}}, \quad (t > \mu). \tag{1.9}$$

Figure 1.2 shows the plots of hazard function of GE distribution with different parameter values.

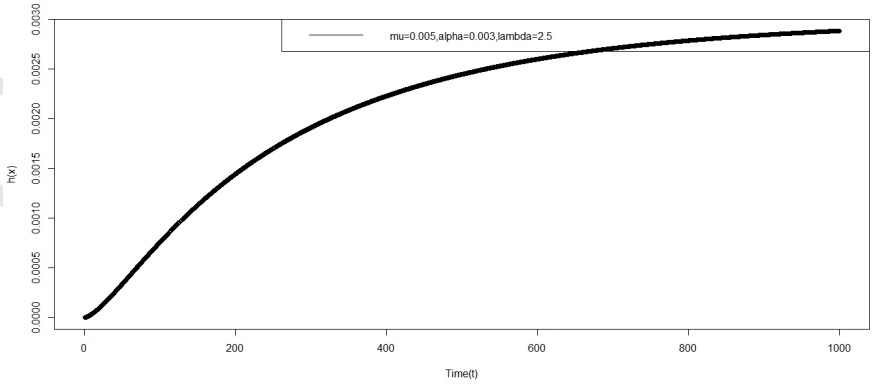

Figure 1.2: Graph of the hazard function of GE with various parameter values

1.9 Problem Statement

Many statistical approaches have been developed to solve the problems that arise in the survival analysis. But these methods still face some problems such as the following :

- The three-parameter GE distribution was shown to perform better than the three-parameter Weibull and Gamma distributions in some cases. However, no work has been conducted to check the performance of the model with covariates.
- The performance of the GE distribution and GE distribution with covariates and mixed case interval-censored data has never been explored as well.
- There is no existing literature on the confidence interval estimation procedures for the GE distribution with covariates and mixed case interval-censored data. Alternative inferential procedures via the bootstrap and jackknife have also not been applied to these distributions and their extensions.

1.10 Research Objectives

- To extend the Generalized exponential distribution to incorporate fixed covariate with right censored data (GEMR model) and fixed covariate with interval censored data (GEMI model).
- To estimate the parameters of the GEMR and GEMI models via simulation studies by evaluating the values of bias, standard error and root mean square error of the parameter estimates at various sample sizes and censoring proportions.
- To assess the performance of the parameter estimates for the GEMI model using midpoint imputation by evaluating the values of bias, standard error and root mean square error of the parameter estimates at various sample sizes, interval lengths and censoring proportions.
- To investigate the performance of confidence interval methods (Wald, likelihood ratio, and score) asymptotic and alternative inferential procedures (jackknife and bootstrap intervals) for the parameters estimates of GEMR and GEMI models with fixed covariate. The evaluation will be conducted through a coverage probability study at various sample sizes, censoring proportions, and nominal levels.
- To apply the proposed models and inferential procedures to real data namely the breast cancer and the lung cancer studies to see how the proposed models work with real-life data.

© COPYRIGHT UPM

1.11 Outline of the Thesis

This thesis is organized into seven chapters. Chapter 1.1 provides a brief introduction to survival data, basic functions in survival analysis, types of censoring, censoring mechanisms, and fixed covariates. This chapter equally discusses parametric survival models that are commonly applied to survival data. The objectives of this research are also discussed in this chapter. Chapter 2 provides a review of related literature to the current work. Special consideration in this literature review is the research conducted on parametric models with RC and IC data with fixed covariates.

Chapter 3 begins with the GE model's extension to incorporate both the right-censored and interval-censored data with fixed covariates. Furthermore, the interval-censored with fixed covariates of the generalized exponential distribution with different forms of imputations were considered. These models' performance is compared at different sample sizes, censoring proportions through a simulation study. Chapter 4 focuses on inferential procedures by evaluating the performance of asymptotic confidence intervals: Wald, likelihood ratio, and score intervals on the parameters of the (GEMR) and (GEMI) models through a coverage probability study at various sample sizes, nominal levels, censoring proportions and interval widths.

Chapter 5 focuses on inferential procedures by evaluating the performance of alternative confidence intervals: jackknife and bootstrap intervals (Normal Bootstrap, Bootstrap Percentile (B-p), Bootstrap-t (B-t), Bias Corrected and Accelerated (BCa) CI estimate methods) on the parameters of the (GEMR) and (GEMI) models through a coverage probability study at various sample sizes, nominal levels, censoring proportions and interval widths.

Chapter 6 discusses the applications of the (GEMR) and (GEMI) to different time-to-event data. The proposed (GEMR) model and confidence interval technique are then applied to modified lung cancer data with age as a fixed covariate. Furthermore, the (GEMI) model was applied to breast cancer data. The necessary inferences were drawn from the applications. Chapter 7 summarizes and concludes the whole thesis.

© COPYRIGHT UPM

REFERENCES

Aho, K., Derryberry, D., and Peterson, T. (2017). A graphical framework for model selection criteria and significance tests: refutation, confirmation and ecology. *Methods in Ecology and Evolution*, 8(1):47–56.

Al-Hakeem, H. A., Arasan, J., Mustafa, M. S. B., and Peng, L. F. (2021). Generalized exponential distribution with interval-censored data and time dependent covariate. *Communications in Statistics-Simulation and Computation*, pages 1–11.

Alizadeh, M., Rezaei, S., Bagheri, S., and Nadarajah, S. (2015). Efficient estimation for the generalized exponential distribution. *Statistical Papers*, 56(4):1015–1031.

Anderson-Bergman, C. (2017). icenreg: regression models for interval censored data in r. *Journal of Statistical Software*, 81:1–23.

Arasan, J. (2009). Interval estimation for parameters of a bivariate time varying covariate model. *Pertanika Journal of Science and Technology*, 17(2):313–323.

Arasan, J. and Adam, M. B. (2014). Double bootstrap confidence interval estimates with censored and truncated data. *Journal of Modern Applied Statistical Methods*, 13(2):22.

Arasan, J. and Lunn, M. (2008). Alternative interval estimation for parameters of bivariate exponential model with time varying covariate. *Computational Statistics*, 23(4):605–622.

Arasan, J. and Midi, H. (2023). Bootstrap based diagnostics for survival regression model with interval and right-censored data. *Austrian Journal of Statistics*, 52(2):66–85.

Awodutire, P., Olapade, A., Kolawole, O., and Ilori, O. (2018). Assessing survival times of breast cancer patients using type i generalized half logistic survival modell. *Journal of Advances in Medicine and Medical Research*, 25(2):1–7.

Azid, N. N. N., Arasan, J., Zulkafli, H. S., Adam, M. B., et al. (2021). Assessing the adequacy of the gompertz regression model in the presence of right censored data. In *Journal of Physics: Conference Series*, volume 1988, page 012114. IOP Publishing.

Balakrishnan, K. (2019). *Exponential distribution: theory, methods and applications*. Routledge.

Balakrishnan, N. and Mitra, D. (2011). Likelihood inference for lognormal data with left truncation and right censoring with an illustration. *Journal of Statistical Planning and Inference*, 141(11):3536–3553.

Balakrishnan, N. and Mitra, D. (2014a). Em-based likelihood inference for some lifetime distributions based on left truncated and right censored data and associated model discrimination. *South African Statistical Journal*, 48:125–171.

© COPYRIGHT UPM

Balakrishnan, N. and Mitra, D. (2014b). Some further issues concerning likelihood inference for left truncated and right censored lognormal data. *Communications in Statistics-Simulation and Computation*, 43(2):400–416.

Barreto-Souza, W., Santos, A. H., and Cordeiro, G. M. (2010). The beta generalized exponential distribution. *Journal of Statistical Computation and Simulation*, 80(2):159–172.

Bebchuk, J. D. and Betensky, R. A. (2000). Multiple imputation for simple estimation of the hazard function based on interval censored data. *Statistics in Medicine*, 19(3):405–419.

Betensky, R. A. (2000). Miscellanea. on nonidentifiability and noninformative censoring for current status data. *Biometrika*, 87(1):218–221.

Betensky, R. A. and Finkelstein, D. M. (1999). An extension of kendall's coefficient of concordance to bivariate interval censored data. *Statistics in Medicine*, 18(22):3101–3109.

Betensky, R. A., Rabinowitz, D. E., and Tsiatis, A. A. (2001). Computationally simple accelerated failure time regression for interval censored data. *Biometrika*, 88:703–711.

Bogaerts, K., Komarek, A., and Lesaffre, E. (2017). *Survival Analysis with Interval-Censored Data: A Practical Approach with Examples in R, SAS, and BUGS*. Chapman and Hall CRC, 2nd edition.

Bradburn, M. J., Clark, T. G., Love, S. B., and Altman, D. G. (2003). Survival analysis part ii: multivariate data analysis–an introduction to concepts and methods. *British Journal of Cancer*, 89(3):431–436.

Breslow, N. E. (1975). Analysis of survival data under the proportional hazards model. *International Statistical Review/Revue Internationale de Statistique*, pages 45–57.

Carpenter, J. and Bithell, J. (2000). Bootstrap confidence intervals: when, which, what? a practical guide for medical statisticians. *Statistics in Medicine*, 19(9):1141–1164.

Chen, L. and Sun, J. (2010). A multiple imputation approach to the analysis of interval-censored failure time data with the additive hazards model. *Computational Statistics & Data Analysis*, 54(4):1109–1116.

Clark, T. G., Bradburn, M. J., Love, S. B., and Altman, D. G. (2003). Survival analysis part i: basic concepts and first analyses. *British Journal of Cancer*, 89(2):232–238.

Cleves, M., Gould, W., Gould, W. W., Gutierrez, R., and Marchenko, Y. (2008). *An introduction to survival analysis using Stata*. Stata press.

Collett, D. (2015). *Modelling survival data in medical research*. CRC press.

© COPYRIGHT UPM

Cook, T. D., Steiner, P. M., and Pohl, S. (2009). How bias reduction is affected by covariate choice, unreliability, and mode of data analysis: Results from two types of within-study comparisons. *Multivariate Behavioral Research*, 44(6):828–847.

Cox, C., Chu, H., Schneider, M. F., and Munoz, A. (2007). Parametric survival analysis and taxonomy of hazard functions for the generalized gamma distribution. *Statistics in Medicine*, 26(23):4352–4374.

Cox, D. R. and Hinkley, D. V. (1979). *Theoretical statistics*. Chapman and Hall/CRC.

Cox, D. R. and Oakes, D. (1984). *Analysis of survival data*, volume 21. CRC press.

Dai, C.-Y., Yu, M., Hou, C., Lu, S.-N., Wang, J.-H., Huang, J.-F., Chen, S.-C., Lin, Z.-Y., Hsieh, M., Tsai, J.-F., Wang, L. Y., Chuang, W.-L., and Chang, W.-Y. (2002). Clinical characteristics and distribution of genotypes of tt virus infection in a hepatitis c virus-hyperendemic township of a hepatitis b virus-endemic country (taiwan). *Journal of Gastroenterology and Hepatology*, 17.

Deng, W., Tian, Y., and Lv, Q. (2012). Parametric estimator of linear model with interval-censored data. *Communications in Statistics - Simulation and Computation*, 41(10):1794–1804.

Dey, S., Alzaatreh, A., Zhang, C., and Kumar, D. (2017). A new extension of generalized exponential distribution with application to ozone data. *Ozone: Science & Engineering*, 39(4):273–285.

DiCiccio, T. J. and Efron, B. (1996). Bootstrap confidence intervals. *Statistical Science*, 11(3):189–228.

Doganaksoy, N. and Schmee, J. (1993). Comparisons of approximate confidence intervals for distributions used in life-data analysis. *Technometrics*, 35(2):175–184.

Dorey, F. J., Little, R. J., and Schenker, N. (1993). Multiple imputation for threshold-crossing data with interval censoring. *Statistics in Medicine*, 12(17):1589–1603.

Dunlap, N., Bailey, W. C., Bass, J. B., Boyd, W. W., Sutton, F. D., and Tally, W. (1994). Occupational exposure to tuberculosis: risk and consequences. *Chest*, 106 3:658–9.

Efron, B. and Stein, C. M. (1981). The jackknife estimate of variance. *Annals of Statistics*, 9:586–596.

Efron, B. and Tibshirani, R. (1985). The bootstrap method for assessing statistical accuracy. *Behaviormetrika*, 12:1–35.

Efron, B. and Tibshirani, R. J. (1994). *An introduction to the bootstrap*. CRC press.

Feinleib, M., Kannel, W. B., Garrison, R. J., McNamara, P. M., and Castelli, W. P. (1975). The framingham offspring study. design and preliminary data. *Preventive Medicine*, 4(4):518–525.

© COPYRIGHT UPM

Finkelstein, D. M. (1986). A proportional hazards model for interval-censored failure time data. *Biometrics*, pages 845–854.

Finkelstein, D. M., Goggins, W. B., and Schoenfeld, D. A. (2002). Analysis of failure time data with dependent interval censoring. *Biometrics*, 58.

Fleming, T. R. and Harrington, D. P. (2013). *Counting processes and survival analysis*, volume 625. John Wiley & Sons.

George, B., Seals, S., and Aban, I. (2014). Survival analysis and regression models. *Journal of Nuclear Cardiology*, 21:686–694.

Ghosh, D. (2001). Efficiency considerations in the additive hazards model with current status data. *Statistica Neerlandica*, 55(3):367–376.

Groeneboom, P. and Wellner, J. A. (1992). *Information bounds and nonparametric maximum likelihood estimation*, volume 19. Springer Science & Business Media.

Grover, G. and Sabharwal, A. (2012). A parametric approach to estimate survival time of diabetic nephropathy with left truncated and right censored data. *International Journal of Statistics and Probability*, 1:128.

Gu, B. and Yue, R. (2013). Some notes on parameter estimation for generalized exponential distribution. *Communications in Statistics - Theory and Methods*, 42(10):1787–1805.

Guan, Q., Tang, Y., Fu, J., and Xu, A. (2014). Optimal multiple constant-stress accelerated life tests for generalized exponential distribution. *Communications in Statistics: Simulation and Computation*, 43.

Gupta, R. D. and Kundu, D. (1999). Theory & methods: Generalized exponential distributions. *Australian & New Zealand Journal of Statistics*, 41(2):173–188.

Gupta, R. D. and Kundu, D. (2007). Generalized exponential distribution: Existing results and some recent developments. *Journal of Statistical Planning and Inference*, 137(11):3537–3547.

Hall, P. (1986). On the bootstrap and confidence intervals. *Annals of Statistics*, 14:1431–1452.

Henderson, N. C., Louis, T. A., Rosner, G. L., and Varadhan, R. (2020). Individualized treatment effects with censored data via fully nonparametric bayesian accelerated failure time models. *Biostatistics*, 21(1):50–68.

Hsu, C.-H., Taylor, J. M., Murray, S., and Commenges, D. (2007). Multiple imputation for interval censored data with auxiliary variables. *Statistics in Medicine*, 26.

Huang, J. (1999). Asymptotic properties of nonparametric estimation based on partly interval-censored data. *Statistica Sinica*, pages 501–519.

Huang, J. et al. (1996). Efficient estimation for the proportional hazards model with interval censoring. *Annals of Statistics*, 24(2):540–568.

© COPYRIGHT UPM

Huang, J., Lee, C., and Yu, Q. (2008). A generalized log-rank test for interval-censored failure time data via multiple imputation. *Statistics in Medicine*, 27(17):3217–3226.

Huang, J. and Wellner, J. A. (1995). Asymptotic normality of the npmle of linear functionals for interval censored data, case 1. *Statistica Neerlandica*, 49(2):153–163.

Jaheen, Z. F. (2004). Empirical bayes inference for generalized exponential distribution based on records. *Communications in Statistics-Theory and Methods*, 33(8):1851–1861.

Jeng, S.-L. and Meeker, W. Q. (2000). Comparisons of approximate confidence interval procedures for type i censored data. *Technometrics*, 42(2):135–148.

Jennison, C. and Turnbull, B. W. (1999). *Group sequential methods with applications to clinical trials*. CRC Press.

Jia, X. and Guo, B. (2017). Exact inference for exponential distribution with multiply type-i censored data. *Communications in Statistics - Simulation and Computation*, 46:7210 – 7220.

Johnson, V. E. (2013). Uniformly most powerful bayesian tests. *Annals of Statistics*, 41(4):1716.

Kalbfleisch, J. D. and Prentice, R. L. (2011). *The statistical analysis of failure time data*, volume 360. John Wiley & Sons.

Keiding, N. (1991). Age-specific incidence and prevalence: a statistical perspective. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 154(3):371–396.

Khan, M. S., King, R., and Hudson, I. L. (2017). Transmuted generalized exponential distribution: A generalization of the exponential distribution with applications to survival data. *Communications in Statistics-Simulation and Computation*, 46(6):4377–4398.

Kiani, K. (2012). *Parametric survival models with time-dependent covariate for mixed case interval censored data*. PhD thesis, University Putra Malaysia.

Kiani, K. and Arasan, J. (2013). Gompertz model with time-dependent covariate in the presence of interval-, right-and left-censored data. *Journal of Statistical Computation and Simulation*, 83(8):1472–1490.

Kim, C. and Han, K. (2015). Bayesian estimation of generalized exponential distribution under progressive first failure censored sample. *Applied Mathematical Sciences*, 9:2037–2047.

Klein, J. P. and Moeschberger, M. L. (2006). *Survival analysis: techniques for censored and truncated data*. Springer Science & Business Media.

Kleinbaum, D. G. and Klein, M. (2012). Competing risks survival analysis. In *Survival Analysis*, pages 425–495. Springer.

© COPYRIGHT UPM

Kleinbaum, D. G., Klein, M., Kleinbaum, D. G., and Klein, M. (2012). Kaplan-meier survival curves and the log-rank test. *Survival analysis: a self-learning text*, pages 55–96.

Koul, H. L. and Yi, T. (2006). Goodness-of-fit testing in interval censoring case 1. *Statistics & Probability Letters*, 76(7):709–718.

Kundu, D. and Gupta, R. D. (2009). Bivariate generalized exponential distribution. *Journal of Multivariate Analysis*, 100(4):581–593.

Kundu, D. and Gupta, R. D. (2011). An extension of the generalized exponential distribution. *Statistical Methodology*, 8(6):485–496.

Kızılaslan, F. and Nadar, M. (2015). Estimation with the generalized exponential distribution based on record values and inter-record times. *Journal of Statistical Computation and Simulation*, 85(5):978–999.

Lai, M. C. and Arasan, J. (2020). Single covariate log-logistic model adequacy with right and interval censored data. *Journal of Quality Measurement and Analysis*, 16(2):131–140.

Law, C. G. and Brookmeyer, R. (1992). Effects of mid-point imputation on the analysis of doubly censored data. *Statistics in Medicine*, 11 12:1569–78.

Lawless, J. et al. (2003). *Statistical models and methods for lifetime data*. Wiley-Interscience.

Lawless, J. F. (2011). *Statistical models and methods for lifetime data*, volume 362. John Wiley & Sons.

Lawless, J. F. and Babineau, D. (2006). Models for interval censoring and simulation-based inference for lifetime distributions. *Biometrika*, 93(3):671–686.

Lee, C. (1999). An urn model in the simulation of interval censored failure time data. *Statistics & Probability Letters*, 45(2):131–139.

Lee, E. T. and Go, O. T. (1997). Survival analysis in public health research. *Annual Review of Public Health*, 18(1):105–134.

Li, J., Song, W., and Shi, J. (2015). Parametric bootstrap simultaneous confidence intervals for differences of means from several two-parameter exponential distributions. *Statistics & Probability Letters*, 106:39–45.

Lim, H.-J. and Sun, J. (2003). Nonparametric tests for interval-censored failure time data. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, 45(3):263–276.

Lin, D., Oakes, D., and Ying, Z. (1998). Additive hazards regression with current status data. *Biometrika*, 85(2):289–298.

Lindsey, J. C. and Ryan, L. M. (1998). Tutorial in biostatistics methods for interval-censored data. *Statistics in Medicine*, 17 2:219–38.

© COPYRIGHT UPM

Lindsey, J. K. (1998). A study of interval censoring in parametric regression models. *Lifetime Data Analysis*, 4:329–354.

Loh, Y. F., Arasan, J., Midi, H., and Bakar, M. R. A. (2015). Inferential procedures based on the double bootstrap for log logistic regression model with censored data. *Malaysian Journal of Science*, 34(2):199–207.

Loh, Y. F., Arasan, J., Midi, H., and Bakar, M. R. A. (2017). A parametric model for doubly interval censored lifetime data. *Journal of Science and Technology*, 9(2).

Lone, S. A., Alam, I., and Rahman, A. (2022). Statistical analysis under geometric process in accelerated life testing plans for generalized exponential distribution. *Annals of Data Science*, pages 1–13.

Loprinzi, C. L., Laurie, J. A., Wieand, H. S., Krook, J. E., Novotny, P. J., Kugler, J. W., Bartel, J., Law, M., Bateman, M., and Klatt, N. E. (1994). Prospective evaluation of prognostic variables from patient-completed questionnaires. north central cancer treatment group. *Journal of Clinical Oncology*, 12(3):601–607.

Mahdavi, A. and Kundu, D. (2017). A new method for generating distributions with an application to exponential distribution. *Communications in Statistics-Theory and Methods*, 46(13):6543–6557.

Mahmoud, M. A. and Alam, F. M. A. (2010). The generalized linear exponential distribution. *Statistics & Probability Letters*, 80(11-12):1005–1014.

Manoharan, T. (2018). *Inferential procedures on log-normal model for left-truncated and case-k interval censored data with covariates*. PhD thesis, Universiti Putra Malaysia.

Manoharan, T., Arasan, J., Midi, H., and Adam, M. B. (2015). A coverage probability on the parameters of the log-normal distribution in the presence of left-truncated and right-censored survival data. *Malaysian Journal of Mathematical Sciences*, 9(1):127–144.

Martinussen, T. and Scheike, T. H. (2002). Efficient estimation in additive hazards regression with current status data. *Biometrika*, 89(3):649–658.

Mirhosseini, S. M., Amini, M., Kundu, D., and Dolati, A. (2015). On a new absolutely continuous bivariate generalized exponential distribution. *Statistical Methods & Applications*, 24:61–83.

Mitra, D. (2013). *Likelihood inference for left truncated and right censored lifetime data*. PhD thesis, McMaster University.

Muhamad Jamil, S., Abdullah, M. A. A., Kek, S. L., Olaniran, O., and Amran, S. (2017). Simulation of parametric model towards the fixed covariate of right censored lung cancer data. *Journal of Physics: Conference Series*, 890:012172.

Mweleli, R. M., Orawo, L. A., Tamba, C. L., Okenye, J. O., et al. (2020). Interval estimation in a two parameter weibull distribution based on type-2 censored data. *Open Journal of Statistics*, 10(06):1039.

© COPYRIGHT UPM

Nadarajah, S. and Kotz, S. (2006). The beta exponential distribution. *Reliab. Eng. Syst. Saf.*, 91:689–697.

Nardi, A. and Schemper, M. (2003). Comparing cox and parametric models in clinical studies. *Statistics in Medicine*, 22(23):3597–3610.

Nayak, B. K. and Hazra, A. (2011). How to choose the right statistical test? *Indian Journal of Ophthalmology*, 59(2):85.

Odell, P., Anderson, K. M., and D'Agostino, R. B. (1992). Maximum likelihood estimation for interval-censored data using a weibull-based accelerated failure time model. *Biometrics*, 48 3:951–9.

Oller, R., Gómez, G., and Calle, M. L. (2004). Interval censoring: model characterizations for the validity of the simplified likelihood. *Canadian Journal of Statistics*, 32(3):315–326.

Pan, W. (2000). A two-sample test with interval censored data via multiple imputation. *Statistics in Medicine*, 19(1):1–11.

Pan, W. (2001). A multiple imputation approach to regression analysis for doubly censored data with application to aids studies. *Biometrics*, 57(4):1245–1250.

Patil, B. M., Joshi, R. C., Toshniwal, D., and Biradar, S. (2011). A new approach: role of data mining in prediction of survival of burn patients. *Journal of Medical Systems*, 35:1531–1542.

Peto, R. (1973). Experimental survival curves for interval-censored data. *Journal of The Royal Statistical Society Series C-applied Statistics*, 22:86–91.

Quenouille, M. H. (1949). Approximate tests of correlation in time-series 3. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 45, pages 483–484. Cambridge University Press.

Ramsey, C. B. (2009). Bayesian analysis of radiocarbon dates. *Radiocarbon*, 51(1):337–360.

Rao, C. R. (1948). Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 44, pages 50–57. Cambridge University Press.

Raqab, M. M. and Ahsanullah, M. (2001). Estimation of the location and scale parameters of generalized exponential distribution based on order statistics. *Journal of Statistical Computation and Simulation*, 69(2):109–123.

Raqab, M. Z. and Madi, M. T. (2005). Bayesian inference for the generalized exponential distribution. *Journal of Statistical Computation and Simulation*, 75(10):841–852.

Ristić, M. M. and Kundu, D. (2015). Marshall-olkin generalized exponential distribution. *Metron*, 73(3):317–333.

© COPYRIGHT UPM

Robinson, J. A. (1983). Bootstrap confidence intervals in location-scale models with progressive censoring. *Technometrics*, 25:179–187.

Rossini, A. and Tsiatis, A. (1996). A semiparametric proportional odds regression model for the analysis of current status data. *Journal of the American Statistical Association*, 91(434):713–721.

Saha, M., Dey, S., and Maiti, S. S. (2018). Parametric and non-parametric bootstrap confidence intervals of cnpk for exponential power distribution. *Journal of Industrial and Production Engineering*, 35(3):160–169.

Sarhan, A. M. (2007). Analysis of incomplete, censored data in competing risks models with generalized exponential distributions. *IEEE Transactions on Reliability*, 56(1):132–138.

Sattar, A., Sinha, S. K., and Morris, N. J. (2012). A parametric survival model when a covariate is subject to left-censoring. *Journal of Biometrics & Biostatistics*, (2).

Satten, G. A., Datta, S., and Williamson, J. M. (1998). Inference based on imputed failure times for the proportional hazards model with interval-censored data. *Journal of the American Statistical Association*, 93(441):318–327.

Schenker, N. (1985). Qualms about bootstrap confidence intervals. *Journal of the American Statistical Association*, 80:360–361.

Schober, P. and Vetter, T. R. (2018). Survival analysis and interpretation of time-to-event data: the tortoise and the hare. *Anesthesia and Analgesia*, 127(3):792.

Seo, S.-K. and Yum, B.-J. (1993). Estimation methods for the mean of the exponential distribution based on grouped and censored data. *IEEE Transactions on Reliability*, 42(1):87–96.

Shen, X. (2000). Linear regression with current status data. *Journal of the American Statistical Association*, 95(451):842–852.

Singh, R. S. and Totawattage, D. P. (2013). The statistical analysis of interval-censored failure time data with applications. *Open Journal of Statistics*, 3(2):155–166.

Sparling, Y. H., Younes, N., Lachin, J. M., and Bautista, O. M. (2006). Parametric survival models for interval-censored data with time-dependent covariates. 7(4):599–614.

Sterne, J. A., White, I. R., Carlin, J. B., Spratt, M., Royston, P., Kenward, M. G., Wood, A. M., and Carpenter, J. R. (2009). Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *Bmj*, 338.

Stewart, A. (2022). *Basic statistics and epidemiology: a practical guide*. CRC Press.

Sun, J. (2007). *The statistical analysis of interval-censored failure time data*. Springer.

© COPYRIGHT UPM

Therneau, T. M., Grambsch, P. M., Therneau, T. M., and Grambsch, P. M. (2000). *The cox model*. Springer.

Tibshirani, R. J. and Efron, B. (1993). An introduction to the bootstrap. *Monographs on Statistics and Applied Probability*, 57(1).

Tse, S.-K., Yuen, H.-K., and Yang, C. (2002). Statistical analysis of exponential lifetimes under an integrated type-ii interval censoring scheme. *Journal of Statistical Computation and Simulation*, 72(6):461–471.

Tukey, J. W. (1958). Abstracts of Papers. *The Annals of Mathematical Statistics*, 29(2):614 – 623.

Turnbull, B. W. (1976). The empirical distribution function with arbitrarily grouped, censored and truncated data. *Journal of The Royal Statistical Society Series B*, 38:290–295.

Vandormael, A., Dobra, A., Bärnighausen, T., de Oliveira, T., and Tanser, F. (2017). Incidence rate estimation, periodic testing and the limitations of the mid-point imputation approach. *International Journal of Epidemiology*, 47(1):236–245.

Vannieuwenhuyze, J. T., Loosveldt, G., and Molenberghs, G. (2014). Evaluating mode effects in mixed-mode survey data using covariate adjustment models. *Journal of Official Statistics*, 30(1):1–21.

Venables, W. N. and Ripley, B. D. (2013). *Modern applied statistics with S-PLUS*. Springer Science & Business Media.

Venzon, D. and Moolgavkar, S. (1988). A method for computing profile-likelihood-based confidence intervals. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 37(1):87–94.

Verdu, S. (1996). The exponential distribution in information theory. *Problemy Peredachi Informatsii*, 32(1):100–111.

Wang, B. X., Geng, Y., and Zhou, J. X. (2018). Inference for the generalized exponential stress-strength model. *Applied Mathematical Modelling*, 53:267–275.

Wu, S.-F. and Lin, Y.-P. (2016). Computational testing algorithmic procedure of assessment for lifetime performance index of products with one-parameter exponential distribution under progressive type i interval censoring. *Mathematics and Computers in Simulation*, 120:79–90.

Xiang, B., Yan, F., Wu, T., Xia, W., Hu, J., and Shen, L. (2020). An improved multiple imputation method based on chained equations for distributed photovoltaic systems. *2020 IEEE 6th International Conference on Computer and Communications (ICCC)*, pages 2001–2005.

Xue, H., Lam, K., and Li, G. (2004). Sieve maximum likelihood estimator for semiparametric regression models with current status data. *Journal of the American Statistical Association*, 99(466):346–356.

© COPYRIGHT UPM

Younes, N. and Lachin, J. M. (1997). Link-based models for survival data with interval and continuous time censoring. *Biometrics*, 53:1199–1211.

Zhao, Q. and Sun, J. (2004). Generalized log-rank test for mixed interval-censored failure time data. *Statistics in Medicine*, 23(10):1621–1629.

Zhou, J., Zhang, J., McLain, A., and Cai, B. (2016). A multiple imputation approach for semiparametric cure model with interval censored data. *Computational Statistics & Data Analysis*, 99:105–114.

© COPYRIGHT UPM