# INFERENCE AND DIAGNOSTICS FOR GENERALIZED EXPONENTIAL DISTRIBUTION WITH FIXED AND TIME-DEPENDENT COVARIATES AND INTERVAL CENSORED DATA

**By**

**AL HAKEEM HUSSEIN ALI GHAFFOORI**

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

**December 2022**

**IPM 2022 14**

© 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

**DEDICATION**

*To my beloved mother with a kind heart*
*Who has supported me all the way since the beginning of my life*

*To the best memory of my father*
*who passed away and who always wished to see my dreams come true*

*To my sister and brother*
*For their great encouragement and support*

*To my love and fiancee, Raghda*
*Who has been a great source of love and motivation*

© COPYRIGHT UPM

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

**INFERENCE AND DIAGNOSTICS FOR GENERALIZED EXPONENTIAL DISTRIBUTION WITH FIXED AND TIME-DEPENDENT COVARIATES AND INTERVAL CENSORED DATA**

By

**AL HAKEEM HUSSEIN ALI GHAFFOORI**

**December 2022**

**Chairman : Associate Professor Jayanthi a/p Arasan, PhD**
**Institute : Mathematical Research**

The aim of this research is to analyse the Generalized Exponential distribution in the presence of interval-censored data with fixed and time-dependent covariates. The analysis starts with a thorough simulation study to compare the performance of the estimation procedure by evaluating the bias, standard error (SE) and root mean square error (RMSE) of the maximum likelihood estimates (MLE) with and without imputation at various censoring proportions and sample sizes. The results clearly indicate that the estimates, based on the random imputation method, work slightly better than the traditional method when dealing with the interval censored data and fixed covariate. Thereafter, we assessed the goodness of fit for this model by comparing the performances of the Cox-Snell and modified Cox-Snell residuals based on the empirical geometric and harmonic means via simulation study at various censoring proportions and sample sizes. The results indicate that the residuals based on the harmonic mean perform slightly better than other residuals, especially when sample sizes in the data are high.

Subsequently, the Generalized Exponential distribution is further extended to incorporate time-dependent covariates with interval-censored data as well as uncensored data. The model is then investigated thoroughly via a comprehensive simulation study at various sample sizes and attendance probabilities when the time-dependent covariate has two levels, before and after update time. Following that, comparison using the values of RMSE is made when a fixed covariate model was fitted wrongly to a data set with time-dependent covariate. The results clearly indicate that the estimates, based on the time dependent covariate, work slightly better than the time dependent covariate when dealing with the interval censored data time dependent covariate. Then we studied two methods of constructing confidence interval estimates namely the Wald and jackknife for the parameters of this model with time-dependent covariate and conclusions were drawn based on the results of the coverage probability



© COPYRIGHT UPM

study. The results indicate that the Wald technique works slightly better than the jackknife technique when dealing with interval censored data and time dependent covariate.

Finally, the methods in the simulation study were applied to real interval-censored data from Diabetic Nephropathy (DN) study with fixed and time-dependent covariates. The results indicate that the Generalized Exponential distribution performs well with interval censored data, fixed, and time-dependent covariates while providing a good fit for dataset. The modified Cox-Snell residual using the harmonic mean was also very useful at assessing the model adequacy using fixed covariates. The Wald confidence interval outperformed the jackknife confidence interval estimation technique was applied to the parameters of model and was useful at indicating the significance of both the fixed and time-dependent covariate parameters.

© COPYRIGHT UPM

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

**INFERENS DAN DIAGNOSTIK UNTUK TABURAN EKSPONEN TERITLAK DENGAN KOVARIAT TETAP DAN BERSANDAR MASA SERTA DATA DITAPIS SELANG**

Oleh

**AL HAKEEM HUSSEIN ALI GHAFFOORI**

**Disember 2022**

**Pengerusi : Profesor Madya Jayanthi a/p Arasan, PhD**
**Institut : Penyelidikan Matematik**

Matlamat penyelidikan ini adalah untuk menganalisis taburan Eksponen Teritlak dengan kehadiran data yang ditapis selang dengan kovariat tetap dan bersandar pada masa. Analisis bermula dengan kajian simulasi menyeluruh untuk membandingkan prestasi prosedur anggaran dengan menilai bias, ralat piawai (SE) dan ralat purata kuasa dua (RMSE) bagi penganggar kebolehjadian maksimum (MLE) dengan dan tanpa imputasi pada pelbagai perkadaran penapisan dan saiz sampel. Keputusan jelas menunjukkan bahawa anggaran, berdasarkan kaedah imputasi rawak, berfungsi sedikit lebihbaik daripada kaedah tradisional apabila berurusan dengan data ditapis selang dan kovariat tetap. Selepas itu, kami menilai kebangusan penyuaian untuk model ini dengan membandingkan prestasi Cox-Snell dan reja Cox-Snell yang diubah suai berdasarkan min empirikal geometri dan harmonik melalui kajian simu- lasi pada pelbagai perkadaran penapisan dan saiz sampel. Keputusan menunjukkan bahawa baki berdasarkan min harmonik berprestasi sedikit lebihbaik daripada reja lain, terutamanya apabila saiz sampel dalam data adalah tinggi.

Berikutan itu, taburan Eksponen Teritlak diperluaskan lagi untuk menggabungkan kovariat bersandar masa dengan data yang ditapis selang serta data yang tidak ditapis. Model ini kemudiannya disiasat secara menyeluruh melalui kajian simulasi komprehensif pada pelbagai saiz sampel dan kebarangkalian kehadiran apabila kovariat bersandar masa mempunyai dua tahap, sebelum dan selepas kemaskini masa. Berikutan itu, perbandingan menggunakan nilai RMSE dibuat apabila model kovariat tetap dipadan secara salah ke pada set data dengan kovariat bersandar masa. Keputusan jelas menunjukkan bahawa anggaran, berdasarkan kovariat bersan- dar masa, berfungsi sedikit lebihbaik daripada kovariat bersandar masa apabila berurusan dengan kovariat bersandar masa data ditapis selang. Kemudian kami mengkaji dua kaedah membina anggaran selang keyakinan iaitu Wald dan jackknife bagi parameter model ini dengan kovariat bersandar masa dan kesimpulan dibuat berdasarkan keputusan kajian



© COPYRIGHT UPM

kebarangkalian liputan. Keputusan menunjukkan bahawa teknik Wald berfungsi sedikit lebihbaik daripada teknik jackknife apabila berurusan dengan data ditapis selang dan kovariat bersandar masa.

Akhir sekali, kaedah dalam kajian simulasi diapplikasikan kepada data sebenar yang ditapis selang sebenar daripada Nefropati Diabetik (DN) dengan kajian kovariat tetap dan bersandar pada masa. Keputusan menunjukkan bahawa taburan Eksponen Teritlak beraksi baik dengan data yang ditapis selang dan kovariat yang bersandar pada masa sambil menyediakan kesesuaian yang baik untuk set data. Reja Cox-Snell yang diubah suai menggunakan min harmonik juga sangat berguna untuk menilai kecukupan model menggunakan kovariat tetap. Selang keyakinan Wald yang mengatasi teknik penganggaran selang keyakinan jackknife telah digunakan pada parameter kedua-dua model dan berguna dalam menunjukkan signifikasi kedua-dua parameter kovariat tetap dan bersandar pada masa.

© COPYRIGHT UPM

**ACKNOWLEDGEMENTS**

In the name of Allah, the most Gracious, most Merciful

First and foremost, I am grateful to Allah s.w.t, for providing me his strength and grace throughout my doctoral pursue and giving me the opportunity for me to continue my studies in Malaysia.

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.

Special thanks and appreciation goes to Dr. Mohd Shafie Mustafa and Dr. Lim Fong Peng being as a members of supervisory committee for their cooperation.

I would like to thank all members of Institute for Mathematical Research (INSPEM) and Department of Mathematics and Statistics for all their friendly support and for their great effort to provide a unique academic environment for the graduate students. I am highly grateful to the Universiti Putra Malaysia (UPM) for all the fruitful years of my study that left an enduring positive impression on my life and professional development.

My special thanks goes to my beloved mother (Mrs. Najeeba Hilo), my beloved father (late Mr. Ali Ghaffoori), my sister and her husband (Mrs. Duaa Ali and Mr. Ali Abdulhassan), my brother and his wife (Mr. Abass Ali and Mrs. Shahad Hameed), for their prayers, love, encouragement, support and all their favours.

Finally, I would like to express my deepest gratitude to my beloved fiancee, Raghda Hameed for her continuous encouragement and kind patience. Without her i never finished my studies. Raghda I Love You.

© 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**
Assosiate Professor
Faculty of Science
Universiti Putra Malaysia
(Chairman)

**Mohd Shafie bin Mustafa, PhD**
Senior Lecturer
Faculty of Science
Universiti Putra Malaysia
(Member)

**Lim Fong Peng, 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: 9 March 2023

© COPYRIGHT UPM

**Declaration by the Graduate Student**

I hereby confirm that:

- this thesis is my original work;
- quotations, illustrations and citations have been duly referenced;
- this thesis has not been submitted previously or concurrently for any other degree at any institutions;
- intellectual property from the thesis and copyright of thesis are fully-owned by Universiti Putra Malaysia, as according to the Universiti Putra Malaysia (Research) Rules 2012;
- written permission must be obtained from supervisor and the office of Deputy Vice-Chancellor (Research and innovation) before thesis is published (in the form of written, printed or in electronic form) including books, journals, modules, proceedings, popular writings, seminar papers, manuscripts, posters, reports, lecture notes, learning modules or any other materials as stated in the Universiti Putra Malaysia (Research) Rules 2012;
- there is no plagiarism or data falsification/fabrication in the thesis, and scholarly integrity is upheld as according to the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) and the Universiti Putra Malaysia (Research) Rules 2012. The thesis has undergone plagiarism detection software

Signature: _____ Date: _____

Name and Matric No: Al Hakeem Hussein Ali Ghaffoori

© COPYRIGHT UPM

**Declaration by Members of the Supervisory Committee**

This is to confirm that:

- the research conducted and the writing of this thesis was under our supervision;
- supervision responsibilities as stated in the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) are adhered to.

Signature: _____
Name of Chairman of Supervisory Committee: Associate Professor Dr. Jayanthi a/p Arasan

Signature: _____
Name of Member of Supervisory Committee: Dr. Mohd Shafie bin Mustafa

Signature: _____
Name of Member of Supervisory Committee: Dr. Lim Fong Peng

© COPYRIGHT UPM

**TABLE OF CONTENTS**





© COPYRIGHT UPM

© COPYRIGHT UPM

© COPYRIGHT UPM

## LIST OF TABLES





© COPYRIGHT UPM

© COPYRIGHT UPM

## LIST OF FIGURES





© COPYRIGHT UPM

© COPYRIGHT UPM

## LIST OF ABBREVIATIONS

| | |
|---|---|
| AC | Anticonservative |
| AP | Attendance Probabilities |
| ASY | Asymmetrical |
| C | Conservative |
| CDF | Cumulative Distribution Function |
| CI | Confidence Interval |
| CP | Censoring Proportion |
| $CS_i$ | Cox Snell Residuals |
| DN | Diabetic Nephropathy |
| FC | Fixed Covariate |
| GED | Generalized Exponential Distribution |
| GM | Geometric Mean |
| HM | Harmonic Mean |
| IC | Interval Censoring |
| KM | Kaplan-Meier |
| LC | Left Censoring |
| MLE | Maximum Likelihood Estimate |
| $MCS_i$ | Modified Cox Snell Residuals |
| OE | Observed Exactly |
| PDF | Probability Density Function |
| RC | Right Censoring |
| RMSE | Root Mean Square Error |
| SE | Standard Error |
| TD | Time Dependent |
| TID | Time Independent |



© COPYRIGHT UPM

# CHAPTER 1

# INTRODUCTION

## 1.1 Survival Analysis

Survival analysis is defined as a group of statistical approaches determined by a well-defined time variable $T$ until the appearance of some particular event or end-point for the individual in a homogeneous population. The response variable $T$ is a positive continuous random variable representing the survival time. Other terms such as lifetime and failure time are frequently used to refer to variable $T$. Time is typically measured in days, weeks, months, years or even fractions of a second from the time an individual enters the study until the event of interest. This includes, for example, relapse, disease occurrence, recovery, death, the collapse of a political system in a country or other events of interest Kleinbaum and Klein (2012).

Survival analysis can also be applied to other fields such as engineering and biomedical sciences. In a study by Olivi (2016), survival analysis is applied to mechanical components installed in gas turbines. This is to estimate survival functions depending on the different environmental attributes of the sites where the gas turbines operate and obtain optimal time points for preventive maintenance. Moreover, Kalbfleisch and Lawless (1992) widened the application of survival analysis to approximate brake pad life for a specific car line. Furthermore, Kachman (1999) investigated the effect of survival on animal production and determined the factors that affect production and the challenges facing the breeder by using several programs such as Survival Kit. Nevertheless, clinical studies remain always dominant in survival analysis studies.

In most medical and clinical trials, the main target is determining and estimating the survival distribution of the lifetime $T$ by estimating the probability that a person survives from a well-specified time $t$ or beyond. For example, patients with fatal diseases such as leukemia, lung cancer and other dangerous diseases survive. Additionally, the statistical analysis and modeling of survival time data are usually done by applying various kinds of non-parametric, semi-parametric or parametric models. Suppose the researcher has all the information related to the appropriate distribution for the survival time $T$. In that case, the parametric method can estimate the survival and risk (hazard) function at a well-defined time $t$. On the contrary, if complete information is unavailable about the appropriate distribution or the study was distribution-free, the non-parametric methods will ensure the estimation of the survival function regarding the study's observations. The Kaplan-Meier and Nelson Aalen estimators are the most often used non-parametric approaches. As for semi-parametric models, there is no detailed definition provided. However, we will refer to a semi-parametric model if it is not fully parametric but has a finite- dimensional parameter of interest Kiani (2012). A brief discussion of the non-parametric, semi-parametric and parametric survival models is given in Section 1.4.



© COPYRIGHT UPM

## 1.2 General Definitions in Survival Analysis

In summarising survival data, there are three functions of central interest, namely the survival function, the hazard function, and the cumulative hazard function. These functions are therefore defined as follows:

### 1.2.1 The Survival Function

If $T$ is a positive $(T \geq 0)$ continuous random variable representing survival, lifetime or failure times the survival function, $S(t)$, defined as follows, then

$$\begin{aligned} S(t) &= \Pr(T \geq t) \\ &= 1 - F(t) \\ &= \int_t^{\infty} f(x)d(x), \end{aligned} \tag{1.1}$$

where $S(t)$ is a survival function that can be described as the probability that an individuals survives $T$ longer than or equal to the specified time $t$, while $F(t)$ is the cumulative distribution function (CDF). 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 and Klein (2012). As $t$ ranges from 0 to ∞, the survival function has the following properties:

1. for $t = 0$

   $$S(0) = \int_0^{\infty} f(x)d(x) = 1.$$

2. for $t = \infty$

   $$S(\infty) = \lim_{t\to\infty} \int_t^{\infty} f(x)d(x) = 0.$$

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



© COPYRIGHT UPM

### 1.2.2 The Hazard Function

The hazard function $h(t)$ is the probability that an individual dies at time $t$, conditional n he or she has survived to that time. The hazard function, therefore, represents the instantaneous death rate for an individual surviving to time $t$, which is mathematically defined as follows:

$$h(t) = \lim_{\Delta t \to 0} \frac{\Pr(\, t \le T < t + \Delta t \mid T \ge t\,)}{\Delta t}$$

$$= \lim_{\Delta t \to 0} \frac{\Pr(t \le T < t + \Delta t, T \ge t)}{\Pr(T \ge t)\, \Delta t}$$

$$= \lim_{\Delta t \to 0} \frac{\Pr(t \le T < t + \Delta t)}{\Pr(T \ge t)\, \Delta t}$$

$$h(t) = \frac{f(t)}{S(t)} \tag{1.2}$$

where $f(t)$ is a non negative probability density function (pdf) that might be considered as an approximation to the probability that the event of interest will occur by the time *t*, represented as follows:

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

$$= -\frac{dS(t)}{dt} \tag{1.3}$$

where $F(t) = 1 - S(t)$ represents the probability that a randomly selected subject from the population will fail before time $t$. This function is also called the cumulative incidence function since it summarises the cumulative probability of death occurring before time *t*, Collett (2015).

Additionally, the relationship between the survival $S(t)$ and the hazard function $h(t)$ can be derived using equations (1.2) and (1.3) as follows:

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

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



© COPYRIGHT UPM

$$= \frac{-dlogS(t)}{dt}, t \geq 0. \quad (1.4)$$

Therefore,

$$S(t) = \exp(-\int_0^t h(x)d(x)), t \geq 0. \quad (1.5)$$

### 1.2.3 The Cumulative Hazard Function

The cumulative hazard function, $H(t)$, can be defined as follows:

$$H(t) = \int_0^t h(x)d(x), t \geq 0. \quad (1.6)$$

The relationship between the survival function and cumulative hazard function can be written as follows:

$$H(t) = \int_0^t \frac{f(x)}{S(x)}\, dx = \int_0^t \frac{-S'(x)}{S(x)}\, dx$$

$$= [-\log S(x)]|_0^t$$

$$H(t) = -\log S(t). \quad (1.7)$$

Therefore,

$$S(t) = exp\big(-H(t)\big). \quad (1.8)$$

## 1.3 Some Censoring Schemes

In long-term medical studies, exact survival time is only known for those individuals who show the event of interest during the follow-up period. Therefore, a researcher must determine the start and end point of the study. Consequently, the observer may notice the emergence of an interesting event for some individuals during the follow-up period. Thus, the survival times $t_i$ for these individuals are known or observed exactly (OE). On the contrary, there will be individuals whose exact survival times are never observed for reasons such as exit from a place where they are being monitored. Other examples include transferring the patient to another part of the country that can no longer be traced as well as terminating the study before all recruited subjects have



© COPYRIGHT UPM

shown the event of interest or other reasons. This leads to the researcher having partial (uncomplete) information about the subjects' survival times but not privy to the exact survival times. The times for these individuals are categorized as censored survival times $c_i$. In the clinical studies, there are three main types of censoring comprising left censoring, interval censoring and right censoring. These censoring mechanisms are discussed in detail in the following section.

### 1.3.1 Left Censoring

Left censoring (LC) is one of the common censoring mechanisms that take place in clinical and medical cases. This type of censoring indicates that the event of interest has already occurred or has been experienced by an individual even before they are registered in the study. In other words, the actual survival time of an individual will be less than that observed. An example of this type of censoring is when some of the patient's cancer cells may have progressed even before they were recorded in the study. Therefore, the exact time of metastasis is unknown, although it occurred before the start of the study. Mathematically, in the event of left censoring, the exact survival time is available when $t_i \geq c_i$. As for the survival times, the $i^{th}$ individual will be a left-censored when $t_i < c_i$ and $t_i \in (0, l_i]$, where $l_i$ is the left-censored survival times.

### 1.3.2 Right Censoring

The observation is said to be right-censored if the exact survival time of observation is unknown, but it is larger than or equal to censoring time $t_i \geq c_i$. In real-life settings, right censoring is more widespread. The right censoring occurs due to the following factors: The subject is lost to follow- up within the study period. Here, the subject withdraws the treatment variable on purpose; the subject is forced to withdraw from the treatment owing to circumstances beyond their control, and the subject withdraws from the study for another reason (i.e., death, if death is not the event of interest) Turkson et al. (2021). With regard to the right censoring, it is divided into three categories: Type I censoring, Type II censoring and Type III censoring (random censoring) Lawless (2003).

In most cases, Type I censoring is associated with a predetermined observation period defined in keeping with the research design. Only a few observations will experience a particular event of interest during the study time, and some will make it to the end. Therefore, it is necessary to determine the start and end point of the study according to the required research. The sole information available to the researcher for individuals who survive the entire observation period is that the actual survival time is located to the right of the study period's end-point along the time axis, mathematically denoted by $t_i > c_i$, where $t_i$ is the event time while $c_i$ is a fixed censored time.

The circumstance in which a specific number of occurrences is targeted for a particular study is called Type II right censoring. In this type of censoring, research would automatically terminate when the desired number of events is observed. Here, all those individuals whose survival times are longer than the time of termination will be right-



© COPYRIGHT UPM

censored. Type II right censoring does not have a set end time; instead, it is related to a time that is decided by the date when a certain number of occurrences are observed. Because of this limitation, Type II right censoring is far less common than other types of right censoring in surveys and clinical trials.

Random censoring occurs randomly at any time during a study period. This could be due to individuals dropping out of the study because they wish to discontinue treatment or lose to the study at a random point due to migration or death related to other possibilities regardless of the event being observed. This type of censoring differs from Type I because the censored time is not fixed but behaves like a random variable. Some participants may join the study after predetermined start date and then be right-censored at the end of the study time. The censored survival time for random censoring is calculated as the time interval between the entry into the study to the time when random censoring occurs, Liu (2012).

### 1.3.3 Interval Censoring

Interval-censoring (IC) occurs in survival analysis when the time until an event of interest is unknown precisely (and instead, only is known to fall into a particular interval). In medical settings, death is not the only occurrence that requires observation. Here, disease progression has piqued medical practitioners' attention or concern. One common example occurs in medical or health studies that entail periodic follow-up. For example, a person may miss one or more scheduled observation times to clinically examine possible changes in illness state and then return with a different status. Alternatively, individuals may visit clinical centers at times convenient to them rather than at predetermined observation times. In both cases, the data changes in the state are classified as interval censoring data Zhang and Sun (2010).

Another application of interval-censored data is in the field of acquired immune deficiency syndrome (AIDS), in terms of the time it takes for human immunodeficiency virus (HIV) infected people. In these circumstances, determining the onset of AIDS is normally based on blood testing, which can be done periodically and not continually. As a result, only interval- censored data is useful for AIDS diagnosis periods. A comparable situation is for studies on HIV infection times. If a patient is HIV positive at the start of a study, the HIV infection period is usually specified by reviewing the analysis of his/her medical history. Therefore, we can only obtain an interval for the HIV infection time based on the last HIV negative test date and the first HIV positive test date Chen et al. (2012). In other words, interval-censored data arises when a survival time (exact event time) $T$ cannot be observed but can be specified to lie within the interval $(t_{L_i}, t_{R_i}]$ obtained from a sequence of examination times, where $t_{L_i} \leq T \leq t_{R_i}$. Here, $t_{L_i}$ and $t_{R_i}$ are defined as the left and right end-points, respectively.

In the clinical and medical studies, the most common types of interval censoring mechanisms are case-I, case-II, case-k, mixed case and doubly IC data as well as panel count data following Schick and Yu (2000) and Sun (2006).



© COPYRIGHT UPM

Case I interval-censored data or current status data, arises when each individual is subjected to only inspected once say at actual inspection time ($ac_{i1}$), producing either a left- or a right censored observation. Thus, if $t_i$ is the event time, then $t_i$ could be less than or larger than $ac_{i1}$, Gómez et al. (2009). This type of IC data could be represented by:

$$\{ac_{i1}, \delta_{I_{i1}} = (t_i \leq ac_{i1}), \delta_{I_{i2}} = (t_i > ac_{i1}), i = 1,2, \dots, n\}$$

then

$$(t_{L_i}\, t_{R_i}] = \begin{cases} (0, ac_{i1}\,]; & \delta_{I_{i1}} = 1 \\ (ac_{i1}, \infty); & \delta_{I_{i2}} = 1. \end{cases} \tag{1.9}$$

where $\delta_{I_i}$ is an indicator variable.

Case II interval-censored data, also called general interval-censored data, arises when the $i^{th}$ individual is inspected twice say at $ac_{i1}$ and $ac_{i2}$ with $ac_{i1} < ac_{i2}$. As a result, the event time $t_i$ may have occurred before the initial inspection time $(0, ac_{i1}]$, in between the two specified inspection times ($ac_{i1}$ & $ac_{i2}$] or after the last inspection time ($ac_{i2}$,∞], Groeneboom and Wellner (1992); Huang and Wellner (1997); Schick and Yu (2000); Sun et al. (2005); and Gómez et al. (2009). This type of IC data could be represented by:

$$\{ac_{i1}, ac_{i2}, \delta_{I_{i1}} = (t_i \leq ac_{i1}), \delta_{I_{i2}} = (ac_{i1} < t_i \leq ac_{i2}), \delta_{I_{i2}} = (t_i > ac_{i2}), i = 1,2,, n\}$$

where $ac_{i1}$ and $ac_{i2}$ are actual inspection times and $ac_{i1} \leq ac_{i2}$, then

$$(t_{L_i}\, t_{R_i}] = \begin{cases} (0, ac_{i1}\,]; & \delta_{I_{i1}} = 1 \\ (ac_{i1}, ac_{i2}]; & \delta_{I_{i2}} = 1 \\ (ac_{i2}, \infty); & \delta_{I_{i3}} = 1. \end{cases} \tag{1.10}$$

Case-k IC data commonly arise in longitudinal studies when there are $k$ actual inspection times for $i^{th}$ individual, $ac_{i1} \leq ac_{i2} \leq \dots \leq ac_{ik}$, where $k$ is a constant number, Schick and Yu (2000) and Gómez et al. (2009). In this case, the event times $t_i$ could have occurred before the first inspection time, in between any two following inspection times or after the last inspection times. This type of IC data could be represented by:

$$\left\{ac_{ij}, k, \delta_{I_{ij}} = \left(ac_{i(j-1)} < t_i \leq ac_{ij}\right), \delta_{I_{i(k+1)}} = (t_i > ac_{ik}), i = 1,2, \dots, n, j = 1,2, \dots, k\right\}$$

where $ac_{i0} = 0$, then



© COPYRIGHT UPM

$$\left(t_{L_i}\, t_{R_i}\right] = \begin{cases} (0, ac_{i1}\,]; & \delta_{I_{i1}} = 1 \\ \left(ac_{i(j-1)}, ac_{ij}\right]; & \delta_{I_{ij}} = 1 \;\; and \;\; 1 < j \leq k, \\ (ac_{ik}, \infty); & \delta_{I_{i(k+1)}} = 1 \;\; and \;\; \mathrm{j} > \mathrm{k}. \end{cases} \tag{1.11}$$

Case I and case II IC are special cases of the case-k IC Wellner (1995).

Mixed case IC data is another type of interval-censored data similar to case-k interval censored data, with the exception that the inspection times, $k$, are random ($k_i$) rather than fixed, where $ac_{i1} \leq ac_{i2} \leq .... \leq ac_{ik_i}$ , Schick and Yu (2000); Lawless and Babineau (2006); and Kiani and Arasan (2013). This type of IC data could be represented by:

$$\left\{ac_{ij}, k_i, \delta_{I_{ij}} = \left(ac_{i(j-1)} < t_i \leq ac_{ij}\right), \delta_{I_{i(k_i+1)}} = \left(t_i > ac_{ik_i}\right), i = 1,2, \ldots, n, j = 1,2, \ldots, k_i\right\}$$

where $ac_{i0} = 0$, then

$$\left(t_{L_i}\, t_{R_i}\right] = \begin{cases} (0, ac_{i1}\,]; & \delta_{I_{i1}} = 1 \\ \left(ac_{i(j-1)}, ac_{ij}\right]; & \delta_{I_{ij}} = 1 \;\; and \;\; 1 < j \leq k_i, \\ \left(ac_{ik_i}, \infty\right); & \delta_{I_{i(k_i+1)}} = 1 \;\; and \;\; \mathrm{j} > k_i. \end{cases} \tag{1.12}$$

In this study, equation (1.12) is the general form of equations (1.9), (1.10) and (1.11) in expressing interval censoring data. Mixed case IC data is quite prevalent in clinical and medical trials because the number of actual inspection times is different from one patient to other.

Doubly interval-censored data (DIC) refers to the survival time of interest, which is the elapsed time between two related events, the initial and the end events. The observations on the occurrences of both events could be interval-censoring. This kind of data often arises in fields such as biometry studies and reliability research. The articles that addressed the doubly interval-censored data arising from survival studies include, Gruttola and Lagakos (1989); Sun (2003); and Kiani and Arasan (2018).

Partly interval-censored data arise when the exact failure times are observed with respect to some subjects. Still, 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). An example of this type of partly interval-censored data is presented by the Framingham Heart Disease study. In this study, 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) and Odell et al. (1992).



© COPYRIGHT UPM

### 1.3.4 Independent Random Censoring

A key assumption in the analysis of censored survival data is that an individual's actual survival time, $t_i$, is independent of any process that causes that individual's survival time to be censored at time ci, where $c_i \leq t_i$. This implies the censoring times $c_i$ are random and independent of the survival or failure times $t_i$ Lawless (2011). In other words, the random censoring mechanism in a survival study is either non-informative or includes no information on all observations' survival/failure times. Considering that the individuals in this study have the same probability of being censored, some individuals may have a less or higher failure risk than others.

With regard to independent censoring mechanism, it is equivalent to non-informative censoring, but the reverse may not always be true, Betensky (2000); Oller et al. (2004); and Sun (2006). Here, dependent censoring and informative censoring are always the same.

Subsequently, for right censoring (RC) data, let the survival/failure time $T$ be independent continuous random variable with prevalent probability density function (pdf) $f(t)$ and survival function $S(t)$. Also, let the censoring time $C$ be continuous random variable with prevalent pdf $g(t)$ and survival function $G(t)$. We all know that the censoring times are $c_i$, non-informative of the survival/failure times $t_i$. Thus, the distribution of $G(t)$ does not depend upon any parameters in $G(t)$.

Furthermore, if censoring process produces censoring time ci for the $i^{th}$ individual and survival time $t_i$ of this individual, then the observed survival/failure time would be $ot_i$ = min ($t_i$, $c_i$). Here, if $t_i \leq c_i$, the indicator variable is $\delta_{Ri}$ = 1 and $ot_i = c_i$. However, if $t_i > c_i$, the indicator variable is $\delta_{Ri}$ = 0 and $ot_i = t_i$. The probability density function (pdf) of independent (non-informative) random censoring for RC data scheme can be represented as follows:

$$\Pr(ot_i = t \mid \delta_{Ri} = 1) = \Pr(T_i = t, t_i \leq c_i) = \Pr(T_i = t)pr(t_i \leq c_i) = f(t)G(t),$$

$$Pr(ot_i = t \mid \delta_{Ri} = 0) = Pr(C_i = t, t_i > c_i) = Pr(C_i = t)Pr(t_i > c_i)\, g(t)S(t), \quad (1.13)$$

By using the result in equation (1.13), the joint likelihood of $(t_i, \delta_i)$ for $n$ of random samples with $i = 1,2,\ldots,n$, can be represented as follows:

$$L(t_i) = \prod_{i=1}^{n} \{f(t_i)G(t_i)\}^{\delta_i} \{g(t_i)\, S(t_i)\}^{1-\delta_i}$$

$$L(t_i) = \prod_{i=1}^{n} f(t_i)^{\delta_i} S(t_i)^{1-\delta_i} \prod_{i=1}^{n} g(t_i)^{1-\delta_i} G(t_i)^{\delta_i} \quad (1.14)$$



© COPYRIGHT UPM

Since $g(t_i)$ and $G(t_i)$ do not involve any of the parameters in $f(t_i)$, they can be neglected and the likelihood function will be shown as follows:

$$L(t_i) = \prod_{i=1}^{n} f(t_i)^{\delta_i} S(t_i)^{1-\delta_i} \tag{1.15}$$

For interval censoring data IC, let $f(t)$ and $S(t)$ represent probability density function (pdf). Moreover, survival function of survival/failure time $T$ and $(t_L, t_R)$ represents joint survivor function of left and right censoring time, where $T \in (t_L, t_R)$. If censoring process produces $t_{L_i}$ and $t_{R_i}$ as a left and right censoring times and $t_i$ be the survival time of the individuals, where $t_i \in (t_{L_i}, t_{R_i})$, Kiani (2012), then independent (non-informative) random censoring for interval censoring data scheme could be represented by:

$$\Pr\left(t_{L_i} = L_i, t_{R_i} = R_i \mid t_i = \mathrm{t}\right) = \Pr(L_i < t_i \leq R_i). \tag{1.16}$$

## 1.4 Survival Models

In survival study, estimating the survival or hazard function is critical to provide an overview of an individual's failure/hazard rate in accordance with the event of interest being studied. In addition, it is usual in clinical research to have situations like covariates or risk factors, such as age or blood pressure level, which may affect patient diagnosis and are recorded to explore the effect of these variables on an individual's survival times Clark et al. (2003) and Bradburn et al. (2003). This can be achieved using three options for modeling the survival function: parametric such as (Generalized Exponential distribution), semi-parametric such as (Cox regression), and non-parametric such as (Kaplan-Meier).

### 1.4.1 Parametric Survival Models

Parametric survival models determine the lifetime distribution up to a parameter of a limited (and usually small) dimension. Even though Cox's semi-parametric model is the most commonly used regression tool for survival data, fully parametric models have some advantages Nardi and Schemper (2003). The main advantage is the availability of straightforward methods of estimation and inference based on the likelihood function, constructing the confidence intervals and performing hypothesis testing to assess the significance of parameter estimates. When the sample size decreases, relative efficiencies may change in favor of parametric models. In addition, the fully parametric models involve stronger assumptions compared to semi- or nonparametric models Lawless (2014). Although they may have advantages over Cox's model, parametric models are rarely employed in the analysis of clinical studies of survival Nardi and Schemper (2003).

The most common parametric distributions of survival time $T$ are exponential, Weibull, gamma, generalized gamma, generalized Weibull, log-logistic, log-normal, Gompertz and Generalized Exponential distribution (GED). Moreover, different types



© COPYRIGHT UPM

of censoring mechanisms and covariates have motivated researchers to create novel parametric models or extend these models to accommodate these different components. Since our research focuses on modeling clinical survival data, we choose to deal with three GED parameters. Initially, GED was first discussed by Mudholkar and Srivastava (1993) as an alternative to the commonly used gamma and Weibull distributions. The three parameters of GED are introduced by the Gupta and Kundu (1999), where they compared the theoretical characteristics of this family to the well-studied characteristics of the Weibull and the Gamma distributions.

The GED has many different theoretical characteristics and can be used effectively to analyze several skewed lifetime data. It also easily handles inference based on the censored data instead of studying Weibull and the Gamma distributions. Recently, many authors have studied different characteristics and statistical methodologies of GED with interval-censored data, for instance, Chen and Lio (2010) and Alharbi et al. (2022).

### 1.4.2 Semi-Parametric Survival Models

In medical and clinical trials, it is frequently required to explore how the risk factors or covariates interact to determine patient survival. To achieve this objective, Bradburn et al. (2003) and Selingerova et al. (2021) used statistical models (semi-parametric and fully parametric models) to measure survival while taking into account multiple factors at the same time. This also includes measuring the degree of influence of the respective factors. In this section, we will deal with the semi-parametric model. There are several approaches for semi-parametric models, for example, Cox proportional hazard (CPH) model. It is the most commonly used multivariate approach for analyzing survival time data, Cox (1972). It is also describes the relationship between the event incidence in terms of the association of covariates with failure time and for studying the effect of these covariates on the hazard function while adjusting for other variables. The hazard function of the CPH model is given as follows:

$$h(t,x) = h_0(t)e^{(\beta' X)}, \tag{1.17}$$

where $X$ is the vector of the $p$ covariates, $X = (x_1, x_2, \dots, x_p)$, $\beta$ is the vector of $p$ parameters, $\beta = (\beta_1, \beta_2, \dots, \beta_p)$ and $h_0(t)$ is the baseline hazard function. Additionally, the survival function for this model can be expressed as follows:

$$S(t,x) = S_0(t)e^{(\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p)}, \tag{1.18}$$

where $S_0(t)$ is the baseline survival function. The vector of parameters $\beta$ can be estimated by using maximum likelihood technique (MLE). Additionally, the baseline hazard function $h_0(t)$ also need to be estimated using several methods, for example the marginal likelihood method as suggested by Kalbfleisch and Prentice (1973), partial likelihood proposed by Cox (1972) and profile likelihood proposed by Breslow (1974, 1975).



© COPYRIGHT UPM

The Cox proportional hazards (CPH) model in (1.19), is extended to include covariates incorporated which changes and time that significantly affect the hazard function as in Pierre and Hans (1995); Collett (2003); and Murray et al. (2016). The hazard function of the CPH with Time Dependent (TD) covariates will be introduced as follows:

$$h(t,x) = h_0(t)e^{\Sigma_{j=1}^{p}\beta' X(t)}, \tag{1.19}$$

where $j = 1,2,\dots p$ is the number of TD covariates in the extended CPH model.

### 1.4.3 Non-Parametric Survival Models

In this section, we discuss survival data analysis without parametric assumptions about the form of the distribution. It is said to be nonparametric or distribution-free since they do not require specific assumptions about the underlying distribution of the survival times. The nonparametric methods are widely used in statistics. It is often attracts researchers as it has more flexibility than parametric and semi-parametric models. The most frequent non-parametric technique is the Kaplan-Meier estimate. The Kaplan-Meier estimator, often known as the "Product- Limit" (PL) estimator, is a popular non-parametric method for estimating the survival function $S(t)$ for right censoring (RC) data. This estimator was proposed first time by Kaplan and Meier (1958), where it uses both uncensored and right-censored data. The Kaplan-Meier (KM) estimator is derived from the following definitions. Let $(n_j)$ be the number of individuals who are alive before time $t_{(j)} = t_1, t_2, \dots, t_r$, including those who are about to die at this time, where $j = 1,2,\dots,r$ , and $(d_j)$ be the number who die at $t_{(j)}$. Then, the estimated survivor function at any time, $t$, in the jth constructed time interval from $t_{(j)}$ to $t_{(j+1)}$ where $t_{(r+1)}$ is defined to be $\infty$, will be the estimated probability of surviving beyond $t_{(j)}$. This is the probability of surviving through the interval from $t_{(j)}$ to $t_{(j+1)}$ and all preceding intervals. In addition to the lifetimes $t_1, t_2, \dots, t_r$, there are also censoring times for individuals whose lifetimes are not observed. Then, the Kaplan-Meier (KM) estimator of the survivor function is given as follows:

$$\hat{S}_{(t)} = \prod_{t_j \le t} \frac{n_j - d_j}{n_j}. \tag{1.20}$$

We can also write equation (1.20) in a simple form as follows:

$$\hat{S}_{(t_j)} = \hat{S}_{\left(t_{(j-1)}\right)}\left(1 - \frac{d_j}{n_j}\right) \tag{1.21}$$

When $\frac{d_j}{n_j}$ is the hazard function $h(t_j)$ at time $t_j$ based on the equations in (1.20) and (1.21), there exists some assumptions in estimating Kaplan Meier given as follows:

- When $t = 0, \hat{S}(0) = 1$.



© COPYRIGHT UPM

- Suppose no censored observations exist after the maximum recorded failure time $t_r$. Then, the number of individuals who are alive (number at risk) $(n_k)$ will be equal to the number who are dead $(d_k)$, in which $n_k = d_k$. Here, the Kaplan-Meier estimator will be zero Lawless (2011).
- Suppose censored observations exist after the maximum recorded failure time $t_r$. In that case, the number of individuals who are alive (number at risk) will be greater than the number who are die, in which $n_k > d_k$. Here, the Kaplan-Meier estimator will never be zero.

Regarding the interval censored data, the nonparametric maximum likelihood estimation (MLE) of a distribution function for case-I interval censored data was first introduced by Ayer et al. (1955) and Van (1956). Then, Peto (1973) and Turnbull (1976) studied the estimations using (MLE) for case-II interval censored data. Furthermore, Sun (2006) studied in detail the nonparametric maximum likelihood estimation (MLE) for case-I and case-II interval censored data. For case-II IC, suppose $t_i \in (t_{L_i}, t_{R_i}]$, where, $i = 1,2,\dots,n$ with survival function $S(t)$. Also let $s_j = 0 < s_{(0)} < s_{(1)} < s_{(2)}, \dots, s_{(m-1)} < s_m = \infty$, $j = 1,2,\dots,m$, called the unique ordered elements of the set $\{0, t_{L_i}, t_{R_i}, \infty, i = 1,2,\dots,n\}$. Then, the likelihood function for case-II IC data is written as follows:

$$L(t_i) = \prod_{i=1}^{n} \{S(t_{L_i}) - S(t_{R_i})\}$$

$$= \prod_{i=1}^{n} \sum_{j=1}^{m} \alpha_{ij} p_j$$

where

$\alpha_{ij} = I(S_j \in (t_{L_i}, t_{R_i}])$ and $p_j = S(s_{j-1}) - S(s_j)$.

As for the rest of the interval censoring cases, Groeneboom and Wellner (1992); Huang and Wellner (1995); and Yu et al. (1996) studied nonparametric MLE of the case I and case II IC data. Wellner (1995) and Huang and Wellner (1997) discussed the nonparametric MLE of the case-k IC data while van der Vaart and Wellner (2000) and Schick and Yu (2000) studied nonparametric MLE of the mixed case IC data.

## 1.5 Types of Covariates

Fixed covariates, stays constant throughout the study period. An example of this type of covariate is the race and gender of individuals in the study. Usually, fixed covariates are measured at the beginning time point of the study, although these covariates can also be measured at the middle or end of the study.



© COPYRIGHT UPM

Contrary to the fixed covariate, the Time-dependent (TD) covariate values change over time. These types of covariates are measured on a regular basis for an individual over the study period. In medical and clinical trials or observational studies, there are many examples of TD covariates, such as the age of subjects, tumour sizes, blood pressure and cholesterol level. The impact of TD covariates will be evaluated on the survival and hazard rate of the individuals in the study. The changes in the level of covariates and constant monitoring for individuals will provide a better indication of future life expectancy than the value at the time of origin, Collett (2015).

In a survival study, TD covariates are categorized into two types as either internal or external covariates. The internal time dependent (ITD) covariates can be measured frequently through a specified period, where the patient is alive during this period. Examples of this covariate include a blood pressure level, white and red blood cell count and serum cholesterol level, Collett (2015) and Kalbfleisch and Prentice (2011). This covariate is observable always as long as the patient is under monitoring.

On the other hand, external time-dependent (ETD) covariates are covariates that do not require the survival of a patient to exist. One type of ETD is the variable that changes so that we can know its value at any time in the future. The clearest example of this type is the age of a patient, where the age of that patient can be known at any future time relying on the time origin without requiring the presence of this patient in the study, Collett (2015). Another type of ETD may affect the patient's survival time exists independently. An example of this type is the respiratory survival studies, where the presence of air pollutants may affect the life span of individuals with lung cancer or heart disease where the change in the air quality is independent of any patient in the study Kalbfleisch and Prentice (2011). Moreover, Arasan and Lunn (2008); Kiani et al. (2012); Kiani and Arasan (2013); and Manoharan et al. (2016) extended their parametric models to include TD covariates with different types of censoring.

### 1.6 Cox Snell Residuals

Residuals are a common method for evaluating the adequacy of a model. In survival study, it is not as easy to determine a residual when modelling survival data compared to a general linear model. As a result, a set of different residuals has been suggested. Cox-Snell ($CS_i$) residuals are widely used in the analysis of survival data as a measure of goodness-of-fit of the models. It was first discussed by Cox and Snell (1968). It is also a type of standardized residuals evaluated both formally and graphically using a test of goodness of fit. The general formula of Cox-Snell residuals for the $i^{th}$ individual, $i = 1,2,\ldots,n$, $CS_i$ is given as follows:

$$CS_i = -log\hat{S}_i(t_i), \tag{1.22}$$

where $\hat{S}_i(t_i)$ is the estimated value of the survival function of the $i^{th}$ individual at $t_i$ By following Collett (2015), we can clearly notice that the $CS_i$ residuals is a special case from the estimated cumulative hazard function, written as follows:



© COPYRIGHT UPM

$$\widehat{H}_i(t_{i)} = -log\hat{S}_i(t_i), \quad (1.23)$$

where $\widehat{H}_i(t_{i)}$ is the estimated value of the cumulative hazard function of the $i^{th}$ individual at $t_i$. If the model fitted to the observed data is satisfactory, then the estimated value of the survival function $\hat{S}_i(t_i)$ for the $i^{th}$ individual at $t_i$ will be close to the survival function's true value $S_i(t_i)$. This implies that if the correct model has been fitted, $\hat{S}_i(t_i)$ values will have properties comparable to $S_i(t_i)$. Then, the negative logarithms of the estimated survivor functions, $-log\hat{S}_i(t_i)$, $i = 1,2,\ldots,n$, will behave as $n$ observations from a unit exponential distribution. The properties of $CS_i$ residuals are quite different from those residuals utilized in linear regression analysis. In particular, Cox-Snell residuals cannot be negative and will not be symmetrically distributed around zero.

One critique of $CS_i$ residuals is that Cox-Snell residuals do not account for censored observations. Therefore, the adjusted $CS_i$ residuals that will deal with this problem were devised by Crowley and Hu (1977). Moreover, Naslina et al. (2020, 2021) and Lai and Arasan (2020) extended their parametric models to include $CS_i$ residuals and modified Cox Snell residuals ($MCS_i$) residuals with different types of censoring and covariates.

## 1.7 Research Objectives

The goal of this research is to analyze the Generalized Exponential model when data is uncensored and interval censored in the presence of fixed and time-dependent covariates. We mainly focused on studying the maximum likelihood estimates with and without imputation. Then we conducted a model adequacy study via the Cox-Snell and several modified Cox-Snell residuals based on geometric and harmonic means. Following that we also thoroughly investigated two confidence interval estimation procedures, the Wald and jackknife for the parameters of this model.

Two main models will be explored,

- Generalized Exponential model with interval censored data and fixed time covariate (FC), for example, race and gender.
- Generalized Exponential model with interval censored data and time-dependent covariate (TD), for example, age, blood pressure, and cholesterol level.

The main objectives of this research are as follows:

- To evaluate the performance of the maximum likelihood estimation with and without imputation for the Generalized Exponential distribution in the presence of interval-censored data and fixed covariates via simulation study at various censoring proportions and sample sizes.



© COPYRIGHT UPM

- Then, conduct a model adequacy study for this model using several modifications of the Cox-Snell residuals based on the geometric mean and harmonic mean at different sample sizes and censoring proportions.
- To extend the Generalized Exponential distribution to incorporate time-dependent covariates in the presence of uncensored and interval-censored data at the various attendance probabilities (AP) and sample sizes.
- To investigate the performance of the Wald and jackknife confidence interval estimates by conducting a coverage probability study at various attendance probabilities (AP), sample sizes and nominal error probabilities for the parameters of the time-dependent covariates model.
- Demonstrate the application of some of the proposed methods to real data from diabetic nephropathy (DN) study using two models. The first model will be fit to uncensored and interval-censored data with fixed covariate and the second model will be fit to a modified version of the data with time-dependent covariate.

## 1.8 Scope of Thesis

This thesis is organized into seven chapters. Chapter 1 provides a brief introduction to survival data, basic functions in survival analysis, types of censoring, fixed covariates and time-dependent covariates. This chapter equally discusses parametric, semi-parametric and nonparametric survival models that is commonly applied with survival data. The objective of this research is also discussed in this chapter.

Chapter 2 provides a literature review highlighting on parametric models in the presence of interval-censored data and different types of covariates. A brief review on parametric, semi-parametric and nonparametric models are discussed to trace the development of these models in the presence of interval-censored data and different types of covariates.

Chapter 3 begins with the extension of Generalized Exponential distribution to incorporate interval-censored data with fixed covariates. Then, the performance of this model is studied by comparing two methods, traditional estimation method and random imputation method at different sample sizes and censoring proportions through a simulation study. Then, the assess the goodness of fit for GED model is studied by comparing the performances of the Cox-Snell and modified Cox-Snell residuals based on the empirical geometric and harmonic means via simulation study at various censoring proportions and sample sizes.

Chapter 4 focuses on extending of Generalized Exponential distribution to incorporate interval-censored data with time-dependent covariates. Then, the performance of this model is studied by comparing the time-dependent covariates with time independent covariates at different sample sizes and attendance probabilities through a simulation study. Following that, the performance of the GED model is studied by fitting data



© COPYRIGHT UPM

from a time-dependent covariates model wrongly to a time-independent covariates model.

Chapter 5 applies the suitability of the Wald and jackknife confidence intervals are compared with the parameters of the GED model with a time-dependent covariate through a coverage probability study.

Chapter 6 applies the methods that were slightly more accurate and efficient in the simulation study to real interval-censored data from Diabetic Nephropathy (DN) study with fixed and time-dependent covariates.

The final chapter, Chapter 7 provides summary and discussion on the overall research. Further, some ideas on future research are equally discussed.

© COPYRIGHT UPM

## REFERENCES


Abdel Hamid, H. (2012). *Flexible parametric survival models with time-dependent covariates for right censored data*. PhD thesis, University of Southampton.

Adeleke, K. A., Abiodun, A. A., and Ipinyomi, R. (2015). Semi-parametric non-proportional hazard model with time varying covariate. *Journal of Modern Ap-plied Statistical Methods*, 14(2):9.

Alharbi, N., Jayanthi, A., Haizum, A., and Ling, W. (2021). Inference for the generalized exponential distributions with covariate and right-censored data. *Malaysian Journal of Mathematical Sciences*, 15(3):411–423.

Alharbi, N., Jayanthi, A., Haizum, A., and Ling, W. (2022). Assessing performance of the generalized exponential model in the presence of the interval censored data with covariate. *Austrian Journal of Statistics*, 51(1):52–69.

Amorim, L. D., Cai, J., Zeng, D., and Barreto, M. L. (2008). Regression splines in the time-dependent coefficient rates model for recurrent event data. *Statistics in medicine*, 27(28):5890–5906.

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

Angelov, A. G. (2018). *Methods for interval-censored data and testing for stochastic dominance*. PhD thesis, Ume$\rho$a*University*.

Arasan, J. (2006). *Lifetime of parallel component systems with dependent failures and multiple covariates*. PhD thesis, University of Oxford.

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 Ehsani, S. (2011). Modeling repairable system failures with interval failure data and time dependent covariate. *Journal of Modern Applied Statistical Methods*, 10(2):19.

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 Lunn, M. (2009). Survival model of a parallel system with dependent failures and time varying covariates. *Journal of Statistical Planning and Inference*, 139(3):944–951.

Arasan, J. and Midi, H. (2021). Jackknife and bootstrap estimates for modified residuals of the log-logistic model. In *AIP Conference Proceedings*, volume 2423, page 070009. AIP Publishing LLC.




© COPYRIGHT UPM

Ateya, S. F. (2014). Maximum likelihood estimation under a finite mixture of generalized exponential distributions based on censored data. *Statistical papers*, 55(2):311–325.

Ayer, M., Brunk, H. D., Ewing, G. M., Reid, W. T., and Silverman, E. (1955). An empirical distribution function for sampling with incomplete information. *The annals of mathematical statistics*, pages 641–647.

Bagdonavicius, V. B. and Nikulin, M. S. (1997). Transfer functions and semiparametric regresion models. *Biometrika*, 84(2):365–378.

Baghestani, A. R., Hajizadeh, E., and Fatemi, S. R. (2010). Parametric model to analyse the survival of gastric cancer in the presence of interval censoring. *Tumori Journal*, 96(3):433–437.

Baklizi, A. S. (2002). Likelihood based estimation in the logistic model with time censored data. *Pertanika Journal of Science and Technology*, 10(1):55–62.

Barabadi, A., Barabady, J., and Markeset, T. (2011). Maintainability analysis considering time-dependent and time-independent covariates. *Reliability Engineering & System Safety*, 96(1):210–217.

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.

Boardman, T. J. (1973). Estimation in compound exponential failure models–when the data are grouped. *Technometrics*, 15(2):271–277.

Borch-Johnsen, K., Andersen, P. K., and Deckert, T. (1985). The effect of proteinuria on relative mortality in type 1 (insulin-dependent) diabetes mellitus. *Diabetologia*, 28(8):590–596.

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. (1974). Covariance analysis of censored survival data. *Biometrics*, pages 89–99.

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



© COPYRIGHT UPM

Cai, Z. and Sun, Y. (2003). Local linear estimation for time-dependent coefficients in cox's regression models. *Scandinavian Journal of Statistics*, 30(1):93–111.

Chen, D. and Lio, Y. (2010). Parameter estimations for generalized exponential distribution under progressive type-i interval censoring. *Computational Statistics & Data Analysis*, 54(6):1581–1591.

Chen, D.-G. D., Sun, J., and Peace, K. E. (2012). *Interval-censored time-to-event data: methods and applications*. CRC Press.

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.

Collett, D. (2003). *Modelling survival data in medical research*. Chap- man and Hall/Crc. ISBN 978-1584883258.

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

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2):187–202.

Cox, D. R. and Hinkley, D. V. (1979). *Theoretical statistics*. CRC Press.

Cox, D. R. and Oakes, D. (2018). *Analysis of survival data*. Chapman and Hall/CRC.

Cox, D. R. and Snell, E. J. (1968). A general definition of residuals. *Journal of the Royal Statistical Society: Series B (Methodological)*, 30(2):248–265.

Crowley, J. and Hu, M. (1977). Covariance analysis of heart transplant survival data. *Journal of the American Statistical Association*, 72(357):27–36.

Dauda, K., Yahya, W. B., and Banjoko, A. (2015). Survival analysis with multivariate adaptive regression splines using cox-snell residual. *Ann Comput Sci Ser*, 13(2):25–41.

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.

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

Frydman, H. (1994). A note on nonparametric estimation of the distribution function from interval-censored and truncated observations. *Journal of the Royal Statistical Society: Series B (Methodological)*, 56(1):71–74.

Goggins, W. B., Finkelstein, D. M., and Zaslavsky, A. M. (1999). Applying the cox proportional hazards model when the change time of a binary time-varying covariate is interval censored. *Biometrics*, 55(2):445–451.

Go´mez, G., Calle, M. L., Oller, R., and Langohr, K. (2009). Tutorial on methods for interval-censored data and their implementation in r. *Statistical Modelling*, 9(4):259–297.

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

Gruttola, V. and Lagakos, S. W. (1989). Analysis of doubly-censored survival data, with application to aids. *Biometrics*, pages 1–11.

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. (2001). Generalized exponential distribution: different method of estimations. *Journal of Statistical Computation and Simulation*, 69(4):315–337.

Gupta, R. D. and Kundu, D. (2006). On the comparison of fisher information of the weibull and ge distributions. *Journal of Statistical Planning and Inference*, 136(9):3130–3144.

Hashimoto, E. M., Ortega, E. M., Cancho, V. G., and Cordeiro, G. M. (2010). The log-exponentiated weibull regression model for interval-censored data. *Computa- tional Statistics & Data Analysis*, 54(4):1017–1035.

Hastie, T. and Tibshirani, R. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society: Series B (Methodological)*, 55(4):757–779.

Henry, B. H., Kvaløy, J. T., and Aaserud, S. (2015). Residual plots to reveal the functional form for covariates in parametric accelerated failure time models. *Lifetime data analysis*, 21(3):353–378.

Hess, K. R. (1994). Assessing time-by-covariate interactions in proportional hazards regression models using cubic spline functions. *Statistics in medicine*, 13(10):1045–1062.

Hinkley, D. (1977). Jackknife confidence limits using student t approximations. *Biometrika*, 64(1):21–28.

Hinkley, D. and Wet, B.-C. (1984). Improvements of jackknife confidence limit methods. *Biometrika*, 71(2):331–339.



© COPYRIGHT UPM

Huang, J. (1996). Efficient estimation for the proportional hazards model with inter- val censoring. *The Annals of Statistics*, 24(2):540–568.

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

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.

Huang, J. and Wellner, J. A. (1997). Interval censored survival data: a review of recent progress. In *Proceedings of the first seattle symposium in biostatistics*, pages 123–169. Springer.

Hudgens, M. G., Satten, G. A., and Longini Jr, I. M. (2001). Nonparametric maximum likelihood estimation for competing risks survival data subject to interval censoring and truncation. *Biometrics*, 57(1):74–80.

Jamil, S. A. M., Abdullah, M. A. A., Kek, S. L., Olaniran, O. R., and Amran, S. E. (2017). Simulation of parametric model towards the fixed covariate of right cen- sored lung cancer data. In *Journal of Physics: Conference Series*, volume 890, page 012172. IOP Publishing.

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

Johansen, M. N., Lundbye-Christensen, S., Larsen, J. M., and Parner, E. T. (2021). Regression models for interval censored data using parametric pseudo- observations. *BMC medical research methodology*, 21(1):1–10.

Kachman, S. D. (1999). Applicaitons in survival analysis. *Journal of animal science*, 77(suppl 2):147–153.

Kalbfleisch, J. and Lawless, J. F. (1992). Some useful statistical methods for truncated data. *Journal of Quality Technology*, 24(3):145–152.

Kalbfleisch, J. D. and Prentice, R. L. (1973). Marginal likelihoods based on cox's regression and life model. *Biometrika*, 60(2):267–278.

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

Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, 53(282):457–481.

Khanal, S. P., Sreenivas, V., and Acharya, S. (2018). Cox proportional hazards model for identification of the prognostic factors in the survival of acute liver failure patients in india. *Nepalese Journal of Statistics*, 2:53–74.

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



© COPYRIGHT UPM

Kiani, K. and Arasan, J. (2012). Simulation of interval censored data in medical and biological studies. In *International Journal of Modern Physics: Conference Series*, volume 9, pages 112–118. World Scientific.

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.

Kiani, K. and Arasan, J. (2018). "survival model with doubly interval –censored data and time-dependent covariates. In *Proceedings of the World Congress on Engineering, I. doi: https://www. semanticscholar. org/paper/Survival-Model-With-Doubly-Interval-Censored-Data-Kiani-Arasan/16509499d65015e3cc2188d1dec25c5b5288f167*.

Kiani, K., Arasan, J., and Midi, H. (2012). Interval estimations for parameters of gompertz model with time-dependent covariate and right censored data. *Sains Malaysiana*, 41(4):471–480.

Kim, D. K. (1997). Regression analysis of interval-censored survival data with co-variates using log-linear models. *Biometrics*, pages 1274–1283.

Kim, J. S. (2003). Maximum likelihood estimation for the proportional hazards model with partly interval-censored data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(2):489–502.

Klein, J. P. and Moeschberger, M. L. (2003). *Survival analysis: techniques for censored and truncated data*, volume 1230. Springer.

Kleinbaum, D. G. and Klein, M. (2012). *Survival analysis: a self-learning text*. Statistics for Biology and Health. Springer-Verlag New York, 3rd edition.

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):1569–1578.

Lawless, G. (2003). *Statistical Models and Methods for Lifetime Data*. New York: John Wiley Sons, 2nd edition.

Lawless, J. (2014). Parametric models in survival analysis. *Wiley StatsRef: Statistics Reference Online*.

Lawless, J. F. (2011). *Statistical models and methods for lifetime data*. volume 362 of Wiley series in probability and statistics. 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.

© COPYRIGHT UPM

Lee, T. C., Zeng, L., Thompson, D. J., and Dean, C. (2011). Comparison of imputation methods for interval censored time-to-event data in joint modelling of tree growth and mortality. *Canadian Journal of Statistics*, 39(3):438–457.

Li, L. and Pu, Z. (2003). Rank estimation of log-linear regression with interval-censored data. *Lifetime data analysis*, 9(1):57–70.

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

Liu, H. and Shen, Y. (2009). A semiparametric regression cure model for interval-censored data. *Journal of the American Statistical Association*, 104(487):1168– 1178.

Liu, X. (2012). *Survival analysis: models and applications*. John Wiley & Sons.

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).

Ma, S. (2010). Mixed case interval censored data with a cured subgroup. *Statistica Sinica*, pages 1165–1181.

Manoharan, T. (2017). Inferential procedures on lognormal model for left-truncated and case-k interval censored data with covariates. PhD thesis, University Putra Malaysia.

Manoharan, T., Arasan, J., Midi, H., and Adam, M. B. (2016). Simulation of left-truncated and case-k interval censored survival data with time-varying covariates. *Malaysian Journal of Mathematical Sciences*, 10(spec. Aug.):159–173.

Manoharan, T., Arasan, J., Midi, H., and Adam, M. B. (2020). Influential measures on log-normal model for left-truncated and case-k interval censored data with time-dependent covariate. *Communications in Statistics-Simulation and Compu- tation*, 49(6):1445–1466.

Martinussen, T., Scheike, T. H., and Skovgaard, I. M. (2002). Efficient estimation of fixed and time-varying covariate effects in multiplicative intensity models. *Scan- dinavian Journal of Statistics*, 29(1):57–74.

Marzec, L. and Marzec, P. (1997). On fitting cox's regression model with time-dependent coefficients. *Biometrika*, 84(4):901–908.

Miller, R. G. (1964). A trustworthy jackknife. *The Annals of Mathematical Statistics*, 35(4):1594–1605.

Miller, R. G. (1974). The jackknife-a review. *Biometrika*, 61(1):1–15.

Mudholkar, G. S. and Srivastava, D. K. (1993). Exponentiated weibull family for analyzing bathtub failure-rate data. *IEEE transactions on reliability*, 42(2):299– 302.



© COPYRIGHT UPM

Murphy, S. A. and Sen, P. K. (1991). Time-dependent coefficients in a cox-type regression model. *Stochastic Processes and their Applications*, 39(1):153–180.

Murray, T. A., Hobbs, B. P., Sargent, D. J., and Carlin, B. P. (2016). Flexible bayesian survival modeling with semiparametric time-dependent and shape-restricted co- variate effects. *Bayesian analysis (Online)*, 11(2):381.

Nan, B., Lin, X., Lisabeth, L. D., and Harlow, S. D. (2005). A varying-coefficient cox model for the effect of age at a marker event on age at menopause. *Biometrics*, 61(2):576–583.

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

Naslina, A. M. N. N., Jayanthi, A., Syahida, Z. H., and Bakri, A. M. (2020). Assessing the goodness of fit of the gompertz model in the presence of right and interval censored data with covariate. *Austrian Journal of Statistics*, 49(3):57–71.

Naslina, A. M. N. N., Jayanthi, A., Syahida, Z. H., and Bakri, A. M. (2021). Assessing the adequacy of the gompertz regression model in the presence of right censored data. In *Journal of Physics: Conference Series*, volume 1988. IOP Pub- lishing.

Ng, H. K. T. and Wang, Z. (2009). Statistical estimation for the parameters of weibull distribution based on progressively type-i interval censored sample. *Journal of Statistical Computation and Simulation*, 79(2):145–159.

Odell, P. M., Anderson, K. M., and D'Agostino, R. B. (1992). Maximum likelihood estimation for interval-censored data using a weibull-based accelerated fail- ure time model. *Biometrics*, pages 951–959.

Olivi, A. (2016). Survival analysis of gas turbine components. Master thesis in Statistics and Data Mining, Linko¨ping University.

Oller, R., Go´mez, 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 multiple imputation approach to cox regression with interval-censored data. *Biometrics*, 56(1):199–203.

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

Petersen, T. (1986). Fitting parametric survival models with time-dependent covariates. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 35(3):281–288.

Peto, R. (1973). Experimental survival curves for interval-censored data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 22(1):86–91.



© COPYRIGHT UPM

Pierre and Hans (1995). Time-dependent effects of fixed covariates in cox regression. *Biometrics*, pages 1550–1556.

Pu, Z. and Li, L. (1999). Regression models with arbitrarily interval-censored observations. *Communications in Statistics-Theory and Methods*, 28(7):1547–1563.

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.

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. (2002). Inferences for generalized exponential distribution based on record statistics. *Journal of statistical planning and inference*, 104(2):339–350.

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.

Raqab, M. Z., Madi, M. T., and Kundu, D. (2008). Estimation of p (y¡ x) for the three-parameter generalized exponential distribution. *Communications in Statis- tics—Theory and Methods*, 37(18):2854–2864.

Reeds, J. A. (1978). Jackknifing maximum likelihood estimates. *The Annals of Statistics*, pages 727–739.

Robins, J. and Tsiatis, A. A. (1992). Semiparametric estimation of an accelerated failure time model with time-dependent covariates. *Biometrika*, 79(2):311–319.

Sabathe, C., Andersen, P. K., Helmer, C., Gerds, T. A., Jacqmin-Gadda, H., and Joly, P. (2020). Regression analysis in an illness-death model with interval- censored data: A pseudo-value approach. *Statistical methods in medical research*, 29(3):752–764.

Samuelsen, S. O. and Kongerud, J. (1994). Interval censoring in longitudinal data of respiratory symptoms in aluminium potroom workers: a comparison of methods. *Statistics in Medicine*, 13(17):1771–1780.

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. *Jour- nal of the American Statistical Association*, 93(441):318–327.

Scallan, A. (1999). Regression modelling of interval-censored failure time data using the weibull distribution. *Journal of Applied statistics*, 26(5):613–618.

Scheike, T. H. (2003). Time-varying effects in survival analysis. *Handbook of Statistics*, 23:61–85.



© COPYRIGHT UPM

Schick, A. and Yu, Q. (2000). Consistency of the gmle with mixed case interval-censored data. *Scandinavian Journal of Statistics*, 27(1):45–55.

Schmid, M. and Hothorn, T. (2008). Flexible boosting of accelerated failure time models. *BMC bioinformatics*, 9(1):1–13.

Seaman, S. and Bird, S. (2001). Proportional hazards model for interval-censored failure times and time-dependent covariates: application to hazard of hiv infection of injecting drug users in prison. *Statistics in medicine*, 20(12):1855–1870.

Selingerova, I., Katina, S., and Horova, I. (2021). Comparison of parametric and semiparametric survival regression models with kernel estimation. *Journal of Sta- tistical Computation and Simulation*, 91(13):2717–2739.

Sen, B. and Banerjee, M. (2007). A pseudolikelihood method for analyzing interval censored data. *Biometrika*, 94(1):71–86.

Sleeper, L. A. and Harrington, D. P. (1990). Regression splines in the cox model with application to covariate effects in liver disease. *Journal of the American Statistical Association*, 85(412):941–949.

Song, S. (2004). Estimation with univariate" mixed case" interval censored data. *Statistica Sinica*, 14(1):269–282.

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

Sun, J. (2003). Statistical analysis of doubly interval-censored failure time data. *Handbook of Statistics*, 23:105–122.

Sun, J. (2006). *The statistical analysis of interval-censored failure time data*, volume 3. New York: Springer-Verlag, first edition.

Sun, J., Zhao, Q., and Zhao, X. (2005). Generalized log-rank tests for interval-censored failure time data. *Scandinavian journal of statistics*, 32(1):49–57.

Sun, Y., Sundaram, R., and Zhao, Y. (2009). Empirical likelihood inference for the cox model with time-dependent coefficients via local partial likelihood. *Scandi- navian Journal of Statistics*, 36(3):444–462.

Tukey, J. (1958). Bias and confidence in not quite large samples. *Ann. Math. Statist.*, 29:614.

Turkson, A. J., Ayiah-Mensah, F., and Nimoh, V. (2021). Handling censoring and censored data in survival analysis: A standalone systematic literature review. *In- ternational Journal of Mathematics and Mathematical Sciences*, 2021.

Turnbull, B. W. (1976). The empirical distribution function with arbitrarily grouped, censored and truncated data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 38(3):290–295.



© COPYRIGHT UPM

Van, C. (1956). Maximum likelihood estimation of ordered probabilities. *Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen, lndagtiones Mathematicae*, 18:444–455.

van der Vaart, A. and Wellner, J. A. (2000). Preservation theorems for glivenko-cantelli and uniform glivenko-cantelli classes. In *High dimensional probability II*, pages 115–133. Springer.

Wang, C., Sun, J., Wang, D., and Shi, N. (2017). Nonparametric estimation of interval-censored failure time data in the presence of informative censoring. *Acta Mathematicae Applicatae Sinica, English Series*, 33(1):107–114.

Wang, P., Zhao, H., and Sun, J. (2016). Regression analysis of case k interval-censored failure time data in the presence of informative censoring. *Biometrics*, 72(4):1103–1112.

Wang, S., Wang, C., Wang, P., and Sun, J. (2020). Estimation of the additive hazards model with case k interval-censored failure time data in the presence of informa- tive censoring. *Computational Statistics & Data Analysis*, 144:106891.

Wellner, J. A. (1995). Interval censoring, case 2: alternative hypotheses. *Lecture Notes-Monograph Series*, pages 271–291.

Wen, C.-C. (2012). Cox regression for mixed case interval-censored data with covariate errors. *Lifetime data analysis*, 18(3):321–338.

Wen, C.-C., Huang, S. Y., and Chen, Y.-H. (2011). Cox regression for current status data with mismeasured covariates. *Canadian Journal of Statistics*, 39(1):73–88.

Xiuyun, P. and Zaizai, Y. (2011). Parameter estimations with gamma distribution based on progressive type-i interval censoring. In *2011 IEEE international confer- ence on computer science and automation engineering*, volume 1, pages 449–453. IEEE.

Yu, Q., Schick, A., Li, L., and Wong, G. Y. (1996). Asymptotic properties of the gmle with case 2 interval-censored data. *Statistics & probability letters*, 37(3):223–228.

Yuan, A., Wei, Y., Makambi, K., and Wang, H. (2014). Approximate nonparamet- ric maximum likelihood estimation for interval censoring model case ii (running head: Npmle for interval censoring case ii). *International Journal of Statistics and Probability*, 3(3):44.

Zeng, D., Cai, J., and Shen, Y. (2006). Semiparametric additive risks model for interval-censored data. *Statistica Sinica*, pages 287–302.

Zeng, D., Mao, L., and Lin, D. (2016). Maximum likelihood estimation for semiparametric transformation models with interval-censored data. *Biometrika*, 103(2):253–271.



© COPYRIGHT UPM

Zhang, H. and Huang, C. (2006). Nonparametric survival analysis on time-dependent covariate effects in case-cohort sampling design. *Statistica Sinica*, pages 267–285.

Zhang, Y., Liu, W., and Zhan, Y. (2001). A nonparametric two-sample test of the failure function with interval censoring case 2. *Biometrika*, 88(3):677–686.

Zhang, Z. and Sun, J. (2010). Interval censoring. *Statistical methods in medical research*, 19(1):53–70.

Zhang, Z., Sun, J., and Sun, L. (2005). Statistical analysis of current status data with informative observation times. *Statistics in Medicine*, 24(9):1399–1407.

Zhang, Z., Sun, L., Sun, J., and Finkelstein, D. M. (2007). Regression analysis of failure time data with informative interval censoring. *Statistics in medicine*, 26(12):2533–2546.

Zhao, X., Zhao, Q., Sun, J., and Kim, J. S. (2008). Generalized log-rank tests for partly interval-censored failure time data. *Biometrical Journal: Journal of Math- ematical Methods in Biosciences*, 50(3):375–385.

Zyoud, A., Elfaki, F. M., and Hrairi, M. (2017). Parametric model based on imputations techniques for partly interval censored data. In *Journal of Physics: Confer-ence Series*, volume 949, page 012002. IOP Publishing.

© COPYRIGHT UPM