SEMIPARAMETRIC INFERENCE PROCEDURE FOR THE ACCELERATED FAILURE TIME MODEL WITH INTERVAL-CENSORED DATA

By

MOSTAFA KARIMI

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

May 2019

IPM 2019 27

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

SEMIPARAMETRIC INFERENCE PROCEDURE FOR THE ACCELERATED FAILURE TIME MODEL WITH INTERVAL-CENSORED DATA

By

MOSTAFA KARIMI

May 2019

Chairman : Professor Noor Akma Ibrahim, PhD
Institute : Mathematical Research

In this thesis new inference procedures are proposed for estimating the parameters of the accelerated failure time (AFT) model in the presence of interval-censored data. In the literature, a variety of semiparametric inference procedures are suggested by previous research for estimating the parameters of AFT models with censored data, and rank-based methods are popular among all.

The main difficulty with the existing rank-based methods is that they involve nonparametric estimation of the probability distribution of the model's error terms. Another problem is estimating the covariance matrix of the parameter estimators, since the existing methods involve derivative of the hazard function of the model's error terms.

Considering the stated problems, the major objectives of this thesis include developing new rank-based estimating procedures for AFT models with interval-censored data based on the actual rank and the expected rank estimating functions, for both univariate and multivariate models. Other research objectives include developing new resampling methods for estimating the covariance matrix of the estimators based on random sampling within censoring intervals and based on the perturbed estimating function.

The findings of this research provide two new iterative algorithms for estimating the parameters of the AFT model with interval-censored data, and also two new resampling techniques for estimating the covariance matrix of estimators. The rank-based methods, estimating algorithms, and resampling techniques that are developed do not involve the difficulties of the existing estimating procedures.

A computationally simple two-step iterative algorithm, called estimation-approximation algorithm, is introduced for estimating the parameters of the model on the basis of the rank estimators. Also, a one-step iterative algorithm, called expected rank algorithm, is introduced which is more complicated than the estimation-approximation algorithm, but more accurate. For estimating the covariance matrix of the proposed estimators two new resampling techniques are proposed, one based on random sampling within censoring interval and another based on perturbed estimating function.

Inference procedures are developed for modelling multiple events interval-censored data through AFT models. Computational properties of the proposed parameter estimating methods and the proposed resampling techniques are comprehensively discussed. The proposed inference procedures are assessed through simulation studies and their performance in applications is demonstrated through analysing real data sets in health science and transportation.

The significance of the study from the results of the numerical analysis shows that the proposed estimators and their corresponding resampling methods are accurate and computationally simpler than the existing methods. The results also imply that influential factors such as the length of censoring intervals and the distribution of the error terms do not significantly affect their efficiency and accuracy. The main contribution of this research is developing statistical approaches, and introducing new algorithms and resampling methods for analysing interval-censored data through AFT models.

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

INFERENS BERPARAMETER SEPARA BAGI MODEL MASA KEGAGALAN TERPECUT DENGAN KEHADIRAN DATA TERTAPIS SELANG

Oleh

MOSTAFA KARIMI

Mei 2019

Pengerusi : Profesor Noor Akma Ibrahim, PhD
Institut : Penyelidikan Matematik

Dalam tesis ini, prosedur inferens baharu dicadangkan untuk menganggar parameter bagi model masa kegagalan terpecut (MKT) dengan kehadiran data tertapis selang. Dalam kesusasteraan literatur, pelbagai prosedur berparameter separa dicadangkan oleh penyelidikan terdahulu untuk menganggarkan parameter model MKT dengan data tertapis selang dan kaedah berasaskan pangkat sering digunakan di kalangan semua.

Kesukaran utama kaedah berasaskan pangkat yang sedia ada adalah ianya melibatkan anggaran tak berparameter bagi sebutan ralat model taburan kebarangkalian. Satu lagi masalah ialah untuk menganggar matriks kovarians parameter penganggar, kerana kaedah yang sedia ada melibatkan terbitan fungsi bahaya sebutan ralat bagi model.

Dengan mempertimbangkan masalah yang dinyatakan, objektif utama tesis ini adalah untuk membangunkan prosedur penganggaran berasaskan pangkat baharu bagi model MKT dengan data tertapis selang berdasarkan pangkat sebenar dan fungsi anggaran pangkat dijangka bagi kedua-dua model univariat dan multivariat. Objektif penyelidikan lain termasuk membangunkan kaedah pensampelan semula baharu untuk menganggar matriks kovarians bagi penganggar berdasarkan pensampelan rawak dalam tertapis selang dan berdasarkan fungsi anggaran yang diganggu.

Penemuan kajian ini memberikan dua algoritma lelaran baharu untuk menganggar parameter model MKT dengan data tertapis selang, dan juga dua teknik pensampelan semula baharu untuk menganggar matriks kovarians bagi penganggar. Kaedah berasaskan pangkat, algoritma anggaran, dan teknik

pensampelan semula baharu yang dibangunkan tidak melibatkan kesulitan seperti prosedur anggaran yang sedia ada.

Satu algoritma lelaran dua-langkah yang mudah pengiraannya, dipanggil algoritma penghampiran-penganggaran, diperkenalkan untuk menganggarkan parameter model berdasarkan penganggar pangkat. Juga, algoritma lelaran satu-langkah, dipanggil algoritma pangkat dijangka diperkenalkan yang lebih rumit daripada algoritma anggaran-penganggaran, tetapi lebih tepat. Untuk menganggar matriks kovarians bagi penganggar yang dicadangkan, dua teknik pensampelan semula baharu dicadangkan, satu berdasarkan pensampelan rawak dalam tertapis selang dan yang lagi satu berdasarkan fungsi anggaran yang diganggu.

Prosedur inferens dibangunkan untuk memodelkan data tertapis selang bagi peristiwa gandaan melalui model MKT. Ciri pengiraan bagi kaedah anggaran parameter dan teknik pensampelan semula yang dicadangkan dibincangkan secara komprehensif. Prosedur inferens yang dicadangkan dinilai melalui kajian simulasi dan prestasi prosedur dalam aplikasi ditunjukkan melalui analisis set data sebenar dalam sains kesihatan dan pengangkutan.

Kesignifikan kajian melalui keputusan analisis berangka menunjukkan bahawa penganggar yang dicadangkan dan kaedah pensampelan semula adalah sangat tepat dan mudah dikomput berbanding kaedah sedia ada. Keputusan juga menunjukkan bahawa faktor yang berpengaruh seperti panjang selang tertapis dan taburan sebutan ralat tidak menjejaskan kecekapan dan ketepatan penganggar. Sumbangan utama penyelidikan ini adalah membangunkan pendekatan statistik, dan memperkenalkan algoritma baharu dan kaedah pensampelan semula untuk menganalisis data tertapis selang melalui model MKT.

ACKNOWLEDGEMENTS

In the name of Allah, the most gracious, the most merciful.

First and foremost, I would like to express my sincere gratitude to my supervisor Professor Dr. Noor Akma Ibrahim for the continuous support of my PhD study and related research, for her patience, motivation, and immense knowledge. Her excellent guidance and insightful comments helped me in all the time of research and writing of this thesis.

Besides my supervisor, I would like to thank the rest of my thesis committee: Associate Professor Dr. Mohd Rizam Abu Bakar, and Associate Professor Dr. Jayanthi Arasan for their precious comments and encouragement.

I would like to express my gratitude to Dr. Ardalan Shariat for providing the stretching exercise training data and Dr. Mohammad Hossein Moosavi for providing the bus dwell time and the bus headway intervals data for this research.

Last but not the least, I would like to thank my parents for supporting me throughout writing this thesis and my life in general.

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

Noor Akma Ibrahim, PhD
Professor
Faculty of Science
Universiti Putra Malaysia
(Chairman)

Mohd Rizam Abu Bakar, PhD
Associate Professor
Faculty of Science
Universiti Putra Malaysia
(Member)

Jayanthi Arasan, PhD
Associate Professor
Faculty of Science
Universiti Putra Malaysia
(Member)

ROBIAH BINTI YUNUS, PhD
Professor and Dean
School of Graduate Studies
Universiti Putra Malaysia

Date: 29 October 2019

Declaration by Members of 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: Professor Dr. Noor Akma Ibrahim

Signature: ______________________
Name of Member of Supervisory Committee: Associate Professor Dr. Mohd Rizam Abu Bakar

Signature: ______________________
Name of Member of Supervisory Committee: Associate Professor Dr. Jayanthi Arasan

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

AFT	Accelerated Failure Time
EA	Estimation-Approximation
ER	Expected Rank
NPMLE	Nonparametric Maximum Likelihood
PH	Proportional Hazards

CHAPTER 1

INTRODUCTION

Survival data or time-to-event data that usually contain censored observations are common in health sciences and engineering. Regression analysis and modelling time-to-event data is challenging in the presence of censored observations. Ordinary linear regression models are not suitable for analysing survival data, since they do not provide a straightforward inference procedure for handling censored observations. Moreover, linear regression models that are developed for analysing censored observations have essential statistical assumptions that are not easy to satisfy in applications. Therefore, there is a need for regression models that are specifically proposed for analysing time-to-event data.

The most common regression models for analysing survival data are the proportional hazards (PH) model and the accelerated failure time (AFT) model. Estimating the parameters of the model and interpretation of regression coefficients of the AFT model is more direct than PH model. This is mainly because the AFT model associates the explanatory variables linearly to the logarithm of survival times. There have been theoretical advances in inference procedures for estimating the parameters of AFT models, but these models are less common than the Cox PH models. The main reason for such matter is the lack of reliable and efficient inference procedures and estimating approaches in the literature.

In this chapter, first the basic concepts in survival analysis such as survival data and failure times are described. Also, the most well-known parametric and nonparametric methods in survival analysis are summarized. Then the preliminary concepts such as right, left, and interval censoring, survival function, and hazard function are defined. In addition, the relationship between the survival function and the hazard function is explained. The Cox PH model and the AFT model and their basic characteristics are discussed. Some parametric and semiparametric approaches for estimating the parameters of such models are reviewed. The main theoretical and computational issues that motivated proposing new inference procedures for modelling interval-censored data are clarified. The objectives of this research are determined based on the statement of the problems. In the end, the outline of this thesis is provided, as well as an overview of the next chapters of the thesis.

1.1 Background of the Study

Survival analysis is a branch of statistics that is defined as a set of methods and techniques for analysing the duration of time until the occurrence of one or more events of interest. The event of interest in clinical trials could be remission of a subject, response of a subject to a specific treatment, or death of a subject. In fact, the exact definition depends on the objectives and hypotheses of the experiment. In engineering research this topic is called reliability analysis or reliability theory. In such studies, the event of interest could be the failure of a mechanical or electronical system, or the failure of a component or a device. However, in both biological survival and industrial reliability studies the definition of "failure" is rather ambiguous. Therefore, the concentration of survival analysis methods is only on well-defined events that occur at specific times.

Survival data is any type of data that contains survival time, failure time, lifetime, time-to-event, time-to-respond. It also includes the characteristics of the experiment subject that are related to the response variable. Some examples of survival data in biological studies are provided by Lee and Wang (2003), including the demographical data, treatment received, and survival time of 30 melanoma patients. They also provided datasets that contain the number of years after diagnosis of rectum cancer of 749 male patients, and datasets that contain the time to remission of 42 patients that were diagnosed with acute leukaemia. Examples of lifetime data in industrial studies are provided by Lawless (2011), including the time to breakdown of 7 groups of a specific electrical fluid specimens that were exposed to voltage stress. They also provided datasets that contain the failure times of 10 pieces of equipment in a system that were installed at different times.

Survival analysis classic methods describe the survival times of subjects through survival functions, hazard functions, and Kaplan-Meier curves (Kaplan and Meier, 1958). They compare the survival times of subjects by using log-rank test (Peto and Peto, 1972), Generalized Wilcoxon test (Gehan, 1965), and Cox-Mantel test (Cox 1959, 1972; Mantel, 1966). The effect of variables on survival times of subjects are described through survival regression models (Cox, 1972; Kalbfleisch and Prentice,1980). The main focus of survival analysis is on specifying models that accurately represent the distribution of survival times and the inference procedure with regard to these models. More precisely, modelling survival times includes estimating survival distributions, comparing survival distributions, and prediction based on explanatory variables. The type of model that is being used in survival analysis can be parametric, nonparametric, or semiparametric. Semiparametric models are the type that have the features and characteristics of both parametric and nonparametric models.

Survival times can be modelled by being described as a function of explanatory variables or covariates. However, the ordinary linear regression may not be an efficient choice for modelling survival times. In fact, survival times are positive numbers regarding to the nature of time variables. Therefore, they are required to be transformed first in a way that satisfies the statistical assumptions of the ordinary linear regression. Moreover, censoring is an important property of survival times and must be carefully considered while modelling survival data. The ordinary linear regression is not able to handle censored observations effectively. For analysing and modelling survival times in the presence of censored observations a common regression model is the AFT model.

1.2 Preliminary Concepts

The event of interest in analysing time-to-event data is generically referred to as "failure". This is even in studies that the interest of the analysis is not necessarily modelling the time to an actual failure in the experiment. In another word, failure time could represent time to death of a subject in one analysis while in another analysis could represent time to recovery of the subject. "Time" refers to the time period from the beginning of an experiment to the occurrence of the event. It also could be the end of experiment, or loss of contact with the subject before the end of experiment. With regard to the interest of the experiment and the nature of the event, failure times can be measured and collected in different time scale units. The common time scale units are seconds, minutes, hours, days, weeks, months, and years. Failure times of subjects of the experiment are assumed to be independent.

Censoring is a special feature of survival data and an important issue in survival analysis. Censored data are categorized as incomplete data. Time-to-event is described as censored observation if the failure time of the subject is not observed during the study period. This means that the subject does not experience failure at the end of the study. In case of a censored observation, subject of the study may or may not experience the event of interest after the end of experiment. Therefore, the actual failure time of the subject is unknown after its censoring time. Examples of censored observations in clinical trials are situations that a patient is withdrawn from the experiment, or a patient is lost to follow-up in the experiment.

In an experiment with a specific number of subjects sometimes the study ends at a predetermined time. Then those subjects that did not experience the event of interest at the end of the study are facing type-I censoring. In such cases, the censoring times would be equal to the time of the end of study, if there is no accidental loss. For example, in many clinical trials researchers cannot wait until all subjects experience the event of interest. This is due to time constraints or cost limitations. Therefore, they need to end the experiment at a predetermined time, which may result type-I censoring.

In some experiments with a specific number of subjects sometimes the study ends when a predetermined failure times are observed. Then those subjects that did not experience the event of interest at the end of the study are facing type-II censoring. In such cases, the censoring times would be equal to the largest observed failure time, if there is no accidental loss. There are examples clinical trials that researchers cannot wait until all subjects experience the event of interest. Therefore, they need to end the experiment when a predetermined failure times are observed, which may result type-II censoring.

Right censoring occurs when for a subject the event of interest happens after a certain time point. This means that the actual failure time of the subject is unknown. It is not difficult to see that both type-I and type-II censoring are right censoring. Left censoring occurs when for a subject the event of interest happens before a certain time point. Thus, the actual failure time of the subject is unknown. Interval censoring occurs when for a subject the event of interest happens between two certain time points that are called examination times. In such case, the actual failure time of the subject is unknown. Figure 1.1 displays an example of right, left, and interval censoring. Note that right censoring can be considered as a special case of interval censoring when the end of the interval is infinity. Left censoring can also be considered as a special case of interval censoring when the beginning of the interval is zero.

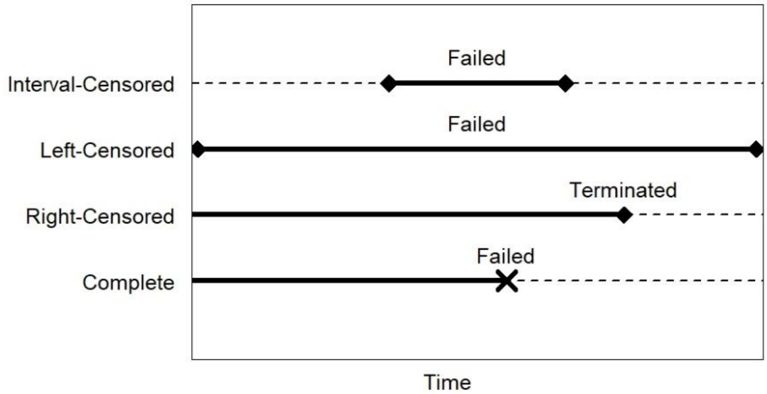

Figure 1.1 : Example of right, left, and interval censoring

Similar to any random variables, since failure times are random variables, they form a probability distribution. The probability distribution of failure times can be described through the density function, the survival function, and the hazard function. The survival function gives the probability that the failure time of a subject, denoted by T is greater than a specific time t. The survival function is given by $S(t) = P(T > t)$. It is clear that $S(t) = 1 - P(T \leq t) = 1 - F(t)$, thus $f(t) = -\frac{d}{dx}S(t)$. The hazard function can be defined in terms of $F(t)$, $f(t)$, and $S(t)$. The hazard function takes the form $h(t) = f(t)/1 - F(t) = f(t)/S(t)$, therefore $h(t) = -\frac{d}{dx}\log S(t)$. Note that the equivalence relationship among $f(t)$, $S(t)$, and $h(t)$ implies that having one of these three functions is enough to drive the other two.

The most common survival model for analysing failure times is PH model (Cox, 1972). Cox PH model, also known as Cox regression model, indicates that covariates of the model are associated to the hazard rate multiplicatively. This means that when a covariate increases one unit the effect of this increment is multiplicative on the hazard rate. Interpretation of the inference procedure of Cox model is not always straightforward in applications. As an alternative for commonly used Cox PH model, the AFT model relates the covariates linearly to the logarithm of failure times. The assumption of Cox PH model is that a unit change in a covariate multiplies the hazard. The AFT model assumes that a unit change in a covariate accelerate or decelerate the failure time. This assumption makes the AFT model appealing to researchers, since the interpretation of the coefficients and the inference procedures are quite straightforward.

Estimating the regression parameters of the model in the presence of censored observations can be carried out through parametric methods or nonparametric methods. If the probability distribution of the failure times is one of the known statistical distributions, then the model and its corresponding parameter estimation method is called parametric. Otherwise, the model would be nonparametric or semiparametric. The most common statistical distributions that are suitable for AFT models are log-logistic, Weibull, log-normal, and gamma distribution. In the category of nonparametric and semiparametric methods for AFT models one of the most well-known techniques is rank-based estimating procedure, also known as rank regression.

1.3 Statement of the Problem

In many research fields, such as medical sciences and engineering, the AFT model is common for regression analysis and statistical modelling of censored observations. For estimating the unknown regression parameters of the AFT model with censored data a variety of parametric and semiparametric methods have been proposed since the AFT model was introduced. The main privilege that makes semiparametric methods more appealing than parametric methods is that they do not involve any assumptions about the probability distribution of the data or the model's error terms.

A well-known semiparametric approach for estimating the parameters of the AFT model with censored data is rank-based inference procedure. The main focus of the previous research on rank estimators was developing estimating methods and iterative algorithms that are efficient and reliable in practice. To provide a great performance, rank regression requires a parameter estimating procedure that is theoretically and computationally simple. For analysing interval-censored data through AFT models, the proposed rank estimators need to be developed specifically based on interval-censored data. In application, rank estimating methods are more common that suggest iterative algorithms based on monotone estimating functions and do not involve complicated computations.

When the probability distribution of the model's error terms is unspecified the model is called semiparametric. There have been significant theoretical advances in semiparametric methods for estimating the parameters of the model in the presence of censored data. But such methods are not common in applications because most of the existing methods are computationally complicated or not completely reliable. There are four main difficulties with the existing semiparametric methods. These problems include the lack of reliable methods for modelling interval-censored data, computational challenges that the estimating procedure involves, non-monotonicity of the estimating functions, and the problems with estimating the covariance matrix of parameter estimators.

For the AFT model with interval-censored data the rank-based estimating procedure that is proposed by Sun (2007) is more common among other existing methods. Such rank estimators were developed based on the approach that was previously suggested by Betensky et al. (2001). The suggested rank estimators are not computationally simple, since they involve likelihood functions based the estimated distribution of the examination times. Their proposed estimating function and score function require nonparametric maximum likelihood estimation (NPMLE) of the distribution of the examination times. Therefore, such estimating procedure is not computationally simple in application. Moreover, the method that they suggested for estimating the covariance matrix of the rank estimators is complicated, due to involving estimation of conditional expected values of the model's covariates.

In general, there are four major problems with the existing rank estimators based on interval-censored data. The first problem is that only a few rank estimators are specifically developed for estimating the parameters of AFT models with interval-censored observation. In fact, most of the proposed methods in the literature concentrated on estimating procedures based on right-censored data. Therefore, making inference in the presence of interval-censored data requires modifications of methods that are proposed for right-censored data. The second problem with the existing rank-based methods is that they require nonparametric estimation of the distribution of the model's error terms. Such estimation involves Kaplan-Meier method and makes the inference procedure computationally intensive.

The third problem is that the existing rank estimating functions are in general step functions of the model's parameters. This means that such estimating functions are not necessarily monotone. It is clear that such issue may cause multiple solutions for the estimating equation, and the possibility of multiple roots for the estimating function. The fourth problem is that for making inference about the distribution of the parameter estimators the existing rank-based methods propose estimating procedures that involve nonparametric estimation of the density function and the hazard function of the model's error terms. Clearly, such methods are challenging for estimating the covariance matrix of the parameter estimators. Thus, conducting hypothesis testing or constructing confidence intervals for the parameter estimators would be problematic.

In this thesis, the major problems with the existing rank estimators for AFT models with interval-censored data are considered and new inference procedures are proposed to overcome such difficulties. New rank-based methods are developed for estimating the regression parameters of the AFT model and approximating the covariance matrix of the estimators. Iterative algorithms are developed that do not involve the computational difficulties of the existing methods, such as NPMLE. The suggested rank estimating functions are based on the examinations times and do not require estimating the distribution of the model's error terms. Two new resampling techniques are developed for estimating the covariance matrix of estimators, one based on random sampling within censoring intervals and another based on perturbed estimating function.

1.4 Objectives of the Research

With respect to the problems that were stated in the previous section, there are five main objectives for this research that are listed as follows:

1. Develop new inference procedures for estimating the regression parameters of semiparametric accelerated failure time models with interval-censored data based on the examination times of censored observations through the actual rank estimating function and the expected rank estimating function
2. Develop new methods for estimating the covariance matrix of the parameter estimators through a resampling technique based on random sampling within censoring intervals and a resampling technique based on the perturbed estimating function
3. Develop new inference procedures for estimating the regression parameters of semiparametric accelerated failure time models with multiple events interval-censored data based on the examination times of censored observations through the actual rank estimating function and the expected rank estimating function and estimate the covariance matrix of the parameter estimators
4. Evaluate the performance and operating characteristics of the parameter estimating methods and the resampling techniques in practical settings through conducting extensive simulation studies
5. Illustrate the parameter estimating methods and the resampling techniques in applications through analysing and modelling real data sets

The parameter estimating procedure will be carried out through two iterative algorithms. A two-step iterative algorithm is developed based on the actual rank estimating function which is computationally and theoretically simple. A one-step iterative algorithm is developed based on the expected rank estimating function which is accurate for estimating the parameters of the model in the presence of interval-censored data. Both resampling techniques based on random sampling within censoring intervals and the perturbed estimating function for estimating the covariance matrix of the parameter estimators will be associated to the iterative algorithms.

1.5 Outline of the Thesis

The first chapter of this thesis is dedicated to describing the introductory topics such as the background of the research and defining the fundamental concepts, as well as stating the research problem and specifying the objectives of the research. In chapter 2 the literature on the theory and application of regression models and specifically the AFT model in the presence of censored data is reviewed. Moreover, the advances of semiparametric AFT models in the literature are discussed. Some references that are fundamental for understanding and using rank-based inference procedure for modelling censored data are also provided in chapter 2.

Chapter 3 describes the structure of interval-censored data, the log-rank and Gehan estimating functions, and general weighted estimating function. With respect to the general weighted loss function, a two-step iterative algorithm is developed for approximating the censored failure times and estimating the unknown parameters of the model. A new resampling technique is developed for estimating the covariance matrix of the rank estimators.

In chapter 4 a new inference procedure is introduced for estimating the parameters of the model which is specifically developed for modelling interval-censored data. This new estimating procedure is described based on the idea of replacing the index function with its corresponding expected value in the estimating equation. Such replacement provides an accurate evaluation of censored failure times. A one-step iterative algorithm is developed for estimating the parameter of the model. Also, a resampling technique is developed for approximating the distribution of the parameter estimators.

In chapter 5 the multivariate analysis of failure time data is discussed, as well as the rank-based parameter estimating approach for modelling multivariate interval-censored data. The parameter estimating methods that are proposed in chapter 3 and 4 for univariate analysis are modified and extended to develop a rank-based estimating procedure for AFT models in the presence of multivariate censored observations. Chapter 6 is dedicated to illustrating the efficiency and performance of the proposed methods for modelling univariate and multivariate interval-censored data. The proposed methods are applied on real data sets to be evaluated and assessed.

Chapter 7 of the thesis provides summary of the proposed methods and their characteristics with regard to their technical and computational aspects. This final chapter also provides a comprehensive conclusion about the performance of the proposed methods and techniques in application. The advantages and disadvantages of using the proposed methods are discussed in the final chapter. A comparison between the proposed inference procedures is also provided. Ideas and opinions for future research works on rank-based estimating procedure for AFT models are presented in chapter 7 as well.

REFERENCES

Andersen, P. K., Borgan, O., Gill, R. D., & Keiding, N. (2012). *Statistical models based on counting processes*. Springer Science & Business Media.

Anderson, K. M. (1991). A nonproportional hazards Weibull accelerated failure time regression model. *Biometrics*, 281-288.

Bagdonavicius, V., & Nikulin, M. (2001). *Accelerated life models: modelling and statistical analysis*. Chapman and Hall/CRC.

Betensky, R. A., Rabinowitz, D., & Tsiatis, A. A. (2001). Computationally simple accelerated failure time regression for interval censored data. *Biometrika*, *88*(3), 703-711.

Bouadoumou, M., Zhao, Y., & Lu, Y. (2015). Jackknife empirical likelihood for the accelerated failure time model with censored data. *Communications in Statistics-Simulation and Computation*, *44*(7), 1818-1832.

Brennan, R. L. (2001). *Generalizability theory*. New York: Springer-Verlag.

Brennan, R. L. (2010). Generalizability theory and classical test theory. *Applied Measurement in Education*, *24*(1), 1-21.

Brown, B. M., & Wang, Y. G. (2007). Induced smoothing for rank regression with censored survival times. *Statistics in medicine*, *26*(4), 828-836.

Buckley, J., & James, I. (1979). Linear regression with censored data. *Biometrika*, *66*(3), 429-436.

Chen, P., Zhang, J., & Zhang, R. (2013). Estimation of the accelerated failure time frailty model under generalized gamma frailty. *Computational Statistics & Data Analysis*, *62*, 171-180.

Chiou, S. H., & Xu, G. (2017). Rank-based estimation for semiparametric accelerated failure time model under length-biased sampling. *Statistics and Computing*, *27*(2), 483-500.

Chiou, S. H., Kang, S., Kim, J., & Yan, J. (2014). Marginal semiparametric multivariate accelerated failure time model with generalized estimating equations. *Lifetime data analysis*, *20*(4), 599-618.

Chiou, S. H., Kang, S., & Yan, J. (2015). Rank-based estimating equations with general weight for accelerated failure time models: an induced smoothing approach. *Statistics in medicine*, *34*(9), 1495-1510.

Christensen, R., & Johnson, W. (1988). Modelling accelerated failure time with a Dirichlet process. *Biometrika*, *75*(4), 693-704.

Chung, M., Long, Q., & Johnson, B. A. (2013). A tutorial on rank-based coefficient estimation for censored data in small-and large-scale problems. *Statistics and computing*, *23*(5), 601-614.

Cole, B. F., Gelber, R. D., & Anderson, K. M. (1994). Parametric approaches to quality-adjusted survival analysis. *Biometrics*, 621-631.

Cox, D. R. (1959). The analysis of exponentially distributed life-times with two types of failure. *Journal of the Royal Statistical Society. Series B (Methodological)*, 411-421.

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., & Oakes, D. (1984). Analysis of survival data. 1984. *Chapman&Hall, London*.

Crowther, M. J., Look, M. P., & Riley, R. D. (2014). Multilevel mixed effects parametric survival models using adaptive Gauss–Hermite quadrature with application to recurrent events and individual participant data meta-analysis. *Statistics in medicine*, *33*(22), 3844-3858.

Deng, W., Ouyang, F., & Zhang, J. (2017). Semiparametric estimation method for accelerated failure time model with dependent censoring. *Communications in Statistics-Simulation and Computation*, *46*(9), 6947-6958.

Dong, X., Kong, L., & Wahed, A. S. (2016). Accelerated failure time model for case-cohort design with longitudinal covariates subject to measurement error and detection limits. *Statistics in medicine*, *35*(8), 1327-1339.

Emura, T., & Wang, W. (2016). Semiparametric inference for an accelerated failure time model with dependent truncation. *Annals of the Institute of Statistical Mathematics*, *68*(5), 1073-1094.

Fleming, T. R., & Harrington, D. P. (2011). *Counting processes and survival analysis* (Vol. 169). John Wiley & Sons.

Fygenson, M., & Ritov, Y. A. (1994). Monotone estimating equations for censored data. *The Annals of Statistics*, 732-746.

Gao, F., Zeng, D., & Lin, D. Y. (2017). Semiparametric estimation of the accelerated failure time model with partly interval-censored data. *Biometrics*, *73*(4), 1161-1168.

Gehan, E. A. (1965). A generalized Wilcoxon test for comparing arbitrarily singly-censored samples. *Biometrika*, *52*(1-2), 203-224.

Gill, R. D. (1980). Censoring and stochastic integrals. *Statistica Neerlandica*, *34*(2), 124-124.

Glaser, R. E. (1995). Weibull accelerated life testing with unreported early failures. *IEEE transactions on reliability*, *44*(1), 31-36.

Gupta, R. C., & Michalek, J. E. (1985). On the relationship between the proportional hazards and accelerated failure time models in survival analysis. *Statistics & probability letters*, *3*(5), 231-234.

Heller, G. (2007). Smoothed rank regression with censored data. *Journal of the American Statistical Association*, *102*(478), 552-559.

Huang, Y. (2013). Fast censored linear regression. *Scandinavian Journal of Statistics*, *40*(4), 789-806.

Hudec, M., & Platz, H. (1983). Fitting regression models to censored survival data. *Statistics in medicine*, *2*(2), 287-293.

Jin, Z., Lin, D. Y., & Ying, Z. (2006). On least-squares regression with censored data. *Biometrika*, *93*(1), 147-161.

Jin, Z., Lin, D. Y., & Ying, Z. (2006). Rank regression analysis of multivariate failure time data based on marginal linear models. *Scandinavian Journal of Statistics*, *33*(1), 1-23.

Jin, Z., Lin, D. Y., Wei, L. J., & Ying, Z. (2003). Rank-based inference for the accelerated failure time model. *Biometrika*, *90*(2), 341-353.

Jin, Z., Shao, Y., & Ying, Z. (2014). A Monte Carlo method for variance estimation for estimators based on induced smoothing. *Biostatistics*, *16*(1), 179-188.

Jin, Z., Ying, Z., & Wei, L. J. (2001). A simple resampling method by perturbing the minimand. *Biometrika*, *88*(2), 381-390.

Johnson, L. M., & Strawderman, R. L. (2009). Induced smoothing for the semiparametric accelerated failure time model: asymptotics and extensions to clustered data. *Biometrika*, *96*(3), 577-590.

Johnson, L. M., & Strawderman, R. L. (2009). Induced smoothing for the semiparametric accelerated failure time model: asymptotics and extensions to clustered data. *Biometrika*, *96*(3), 577-590.

Kalbfleisch, J. D., & Prentice, R. L. (1980). The statistical analysis of time failure data. *The statistical analysis of time failure data*.

Kang, S., Lu, W., & Liu, M. (2017). Efficient estimation for accelerated failure time model under case-cohort and nested case-control sampling. *Biometrics*, *73*(1), 114.

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

Kellerer, A. M., & Chmelevsky, D. (1983). Small-sample properties of censored-data rank tests. *Biometrics*, 675-682.

Kim, J. P., Sit, T., & Ying, Z. (2016). Accelerated failure time model under general biased sampling scheme. *Biostatistics*, *17*(3), 576-588.

Klein, J. P., Pelz, C., & Zhang, M. J. (1999). Modelling random effects for censored data by a multivariate normal regression model. *Biometrics*, *55*(2), 497-506.

Komárek, A., Lesaffre, E., & Hilton, J. F. (2005). Accelerated failure time model for arbitrarily censored data with smoothed error distribution. *Journal of Computational and Graphical Statistics*, *14*(3), 726-745.

Kong, S., & Nan, B. (2016). Semiparametric approach to regression with a covariate subject to a detection limit. *Biometrika*, *103*(1), 161-174.

Lai, T. L., & Ying, Z. (1991). Rank regression methods for left-truncated and right-censored data. *The Annals of Statistics*, *19*(2), 531-556.

Lambert, P., Collett, D., Kimber, A., & Johnson, R. (2004). Parametric accelerated failure time models with random effects and an application to kidney transplant survival. *Statistics in medicine*, *23*(20), 3177-3192.

Latta, R. B. (1981). A Monte Carlo study of some two-sample rank tests with censored data. *Journal of the American Statistical Association*, *76*(375), 713-719.

Lawless, J. F. (1986). A note on lifetime regression models. *Biometrika*, 509-512.

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

Lawless, J. F., & Singhal, K. (1987). Regression methods and the exploration of large medical data bases. In *Biostatistics* (pp. 1-22). Springer, Dordrecht.

Lee, E. T., & Wang, J. (2003). *Statistical methods for survival data analysis* (Vol. 476). John Wiley & Sons.

Lee, E. T., Desu, M. M., & Gehan, E. A. (1975). A Monte Carlo study of the power of some two-sample tests. *Biometrika*, *62*(2), 425-432.

Li, H., & Yin, G. (2009). Generalized method of moments estimation for linear regression with clustered failure time data. *Biometrika*, *96*(2), 293-306.

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

Li, Z., Xu, X., & Shen, J. (2017). Semiparametric Bayesian analysis of accelerated failure time models with cluster structures. *Statistics in medicine*, *36*(25), 3976-3989.

Lin, D. Y., & Geyer, C. J. (1992). Computational methods for semiparametric linear regression with censored data. *Journal of Computational and Graphical Statistics*, *1*(1), 77-90.

Lin, D. Y., & Ying, Z. (1995). Semiparametric inference for the accelerated life model with time-dependent covariates. *Journal of statistical planning and inference*, *44*(1), 47-63.

Lin, Y., & Chen, K. (2012). Efficient estimation of the censored linear regression model. *Biometrika*, *100*(2), 525-530.

Lininger, L., Gail, M. H., Green, S. B., & Byar, D. P. (1979). Comparison of four tests for equality of survival curves in the presence of stratification and censoring. *Biometrika*, *66*(3), 419-428.

Louis, T. A. (1981). Nonparametric analysis of an accelerated failure time model. *Biometrika*, *68*(2), 381-390.

Louzada-Neto, F. (1997). Extended hazard regression model for reliability and survival analysis. *Lifetime Data Analysis*, *3*(4), 367-381.

Lu, W. (2010). Efficient estimation for an accelerated failure time model with a cure fraction. *Statistica Sinica*, *20*, 661.

Lu, X., & Cheng, T. L. (2007). Randomly censored partially linear single-index models. *Journal of Multivariate Analysis*, *98*(10), 1895-1922.

Luo, J., Li, H., & Zhang, J. (2016). Robust Smoothed Rank Estimation Methods for Accelerated Failure Time Model Allowing Clusters. *Communications in Statistics-Simulation and Computation*, *45*(6), 1865-1884.

Luo, S. (2014). A Bayesian approach to joint analysis of multivariate longitudinal data and parametric accelerated failure time. *Statistics in Medicine*, *33*(4), 580-594.

Mandel, M., & Ritov, Y. A. (2010). The accelerated failure time model under biased sampling. *Biometrics*, *66*(4), 1306-1308.

Mantel, N. (1966). Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemother Rep*, *50*, 163-170.

Newby, M. (1988). Accelerated failure time models for reliability data analysis. *Reliability Engineering & System Safety*, *20*(3), 187-197.

Ng, E. S. W., Klungel, O. H., Groenwold, R. H., & van Staa, T. P. (2015). Risk patterns in drug safety study using relative times by accelerated failure time models when proportional hazards assumption is questionable: an illustrative case study of cancer risk of patients on glucose-lowering therapies. *Pharmaceutical statistics*, *14*(5), 382-394.

Oakes, D. (1983). Comparison of models for survival data. *Statistics in medicine*, *2*(2), 305-311.

Odell, P. M., Anderson, K. M., & D'Agostino, R. B. (1992). Maximum likelihood estimation for interval-censored data using a Weibull-based accelerated failure time model. *Biometrics*, 951-959.

Pang, L., Lu, W., & Wang, H. J. (2015). Local Buckley-James estimation for heteroscedastic accelerated failure time model. *Statistica Sinica*, *25*, 863.

Parzen, M. I., Wei, L. J., & Ying, Z. (1994). A resampling method based on pivotal estimating functions. *Biometrika*, *81*(2), 341-350.

Peace, K. E., & Flora, R. E. (1978). Size and power assessments of tests of hypotheses on survival parameters. *Journal of the American Statistical Association*, *73*(361), 129-132.

Peto, R., & Peto, J. (1972). Asymptotically efficient rank invariant test procedures. *Journal of the Royal Statistical Society. Series A (General)*, 185-207.

Peto, R., Pike, M. C., Armitage, P., Breslow, N. E., Cox, D. R., Howard, S. V., ... & Smith, P. G. (1977). Design and analysis of randomized clinical trials requiring prolonged observation of each patient. II. analysis and examples. *British journal of cancer*, *35*(1), 1.

Picciotto, S., Ljungman, P. L., & Eisen, E. A. (2016). Straight Metalworking Fluids and All-Cause and Cardiovascular Mortality Analyzed by Using G-Estimation of an Accelerated Failure Time Model With Quantitative Exposure: Methods and Interpretations. *American journal of epidemiology*, *183*(7), 680-688.

Prentice, R. L. (1978). Linear rank tests with right censored data. *Biometrika*, *65*(1), 167-179.

Qiu, Z., Qin, J., & Zhou, Y. (2016). Composite Estimating Equation Method for the Accelerated Failure Time Model with Length-biased Sampling Data. *Scandinavian Journal of Statistics*, *43*(2), 396-415.

Rabinowitz, D., Tsiatis, A., & Aragon, J. (1995). Regression with interval-censored data. *Biometrika*, *82*(3), 501-513.

Rao, C. R., & Zhao, L. C. (1992). Approximation to the distribution of M-estimates in linear models by randomly weighted bootstrap. *Sankhyā: The Indian Journal of Statistics, Series A*, 323-331.

Ritov, Y. (1990). Estimation in a linear regression model with censored data. *The Annals of Statistics*, 303-328.

Robins, J. (1992). Estimation of the time-dependent accelerated failure time model in the presence of confounding factors. *Biometrika*, *79*(2), 321-334.

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

Savage, I. R. (1956). Contributions to the theory of rank order statistics-the two-sample case. *The Annals of Mathematical Statistics*, *27*(3), 590-615.

Shen, J., Li, Z., Yu, H., & Fang, X. (2017). Semiparametric Bayesian inference for accelerated failure time models with errors-in-covariates and doubly censored data. *Journal of Systems Science and Complexity*, *30*(5), 1189-1205.

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

Sun, L., & Su, B. (2008). A class of accelerated means regression models for recurrent event data. *Lifetime data analysis*, *14*(3), 357-375.

Tian, L., & Cai, T. (2006). On the accelerated failure time model for current status and interval censored data. *Biometrika*, *93*(2), 329-342.

Tsiatis, A. A. (1990). Estimating regression parameters using linear rank tests for censored data. *The Annals of Statistics*, 354-372.

Walker, S., & Mallick, B. K. (1999). A Bayesian semiparametric accelerated failure time model. *Biometrics*, *55*(2), 477-483.

Wang, H., Dai, H., & Fu, B. (2012). Accelerated failure time models for censored survival data under referral bias. *Biostatistics*, *14*(2), 313-326.

Wang, S., Nan, B., Zhu, J., & Beer, D. G. (2008). Doubly penalized Buckley–James method for survival data with high-dimensional covariates. *Biometrics*, *64*(1), 132-140.

Wang, W., Ma, Y., Huang, Y., & Chen, H. (2017). Generalizability analysis for clinical trials: a simulation study. *Statistics in medicine*, *36*(10), 1523-1531.

Wang, Y. G., & Fu, L. (2011). Rank regression for accelerated failure time model with clustered and censored data. *Computational Statistics & Data Analysis*, *55*(7), 2334-2343.

Wang, Z., & Wang, C. Y. (2010). Buckley-James boosting for survival analysis with high-dimensional biomarker data. *Statistical Applications in Genetics and Molecular Biology*, *9*(1).

Wei, L. J., Ying, Z., & Lin, D. Y. (1990). Linear regression analysis of censored survival data based on rank tests. *Biometrika*, *77*(4), 845-851.

Xue, H., Lam, K. F., Cowling, B. J., & de Wolf, F. (2006). Semi-parametric accelerated failure time regression analysis with application to interval-censored HIV/AIDS data. *Statistics in medicine*, *25*(22), 3850-3863.

Yang, M., Chen, L., & Dong, G. (2015). Semiparametric Bayesian accelerated failure time model with interval-censored data. *Journal of Statistical Computation and Simulation*, *85*(10), 2049-2058.

Ying, Z. (1993). A large sample study of rank estimation for censored regression data. *The Annals of Statistics*, 76-99.

Ying, Z., Jung, S. H., & Wei, L. J. (1995). Survival analysis with median regression models. *Journal of the American Statistical Association*, *90*(429), 178-184.

Yu, L., Yu, R., Liu, L., & Chen, D. G. (2012). Extended quasi-likelihood with fractional polynomials in the frame of the accelerated failure time model. *Statistics in medicine*, *31*(13), 1369-1379.

Zeng, D., & Lin, D. Y. (2007). Efficient estimation for the accelerated failure time model. *Journal of the American Statistical Association*, *102*(480), 1387-1396.

Zhang, J., & Peng, Y. (2007). An alternative estimation method for the accelerated failure time frailty model. *Computational statistics & data analysis*, *51*(9), 4413-4423.

Zhang, M., & Davidian, M. (2008). “Smooth” semiparametric regression analysis for arbitrarily censored time-to-event data. *Biometrics*, *64*(2), 567-576.

Zhao, Y., Brown, B. M., & Wang, Y. G. (2014). Smoothed rank-based procedure for censored data. *Electronic Journal of Statistics*, *8*(2), 2953-2974.

Zheng, M., & Yu, W. (2011). Empirical likelihood method for the multivariate accelerated failure time models. *Journal of Statistical Planning and Inference*, *141*(2), 972-983.

Zheng, M., Lin, R., & Yu, W. (2016). Competing risks data analysis under the accelerated failure time model with missing cause of failure. *Annals of the Institute of Statistical Mathematics*, *68*(4), 855-876.

Zhou, M. (2005). Empirical likelihood analysis of the rank estimator for the censored accelerated failure time model. *Biometrika*, *92*(2), 492-498.

Zhou, X., Fang, M., Li, J., Prows, D. R., & Yang, R. (2012). Characterization of genomic imprinting effects and patterns with parametric accelerated failure time model. *Molecular genetics and genomics*, *287*(1), 67-75.