Nonparametric Survival Analysis under Shape Restrictions

Shabnam Fani

A thesis submitted in partial fulfilment of the requirements
for the degree of Doctor of Philosophy in Statistics,
The University of Auckland,
2014.

Abstract

The main problem studied in this thesis is to analyse and model time-to-event data, particularly when the survival times of subjects under study are not exactly observed. One of the primary tasks in the analysis of survival data is to study the distribution of the event times of interest. In order to avoid strict assumptions associated with a parametric model, we resort to nonparametric methods for estimating a function. Although other nonparametric approaches, such as Kaplan-Meier, kernel-based, and roughness penalty methods, are popular tools for solving function estimation problems, they suffer from some non-trivial issues like the loss of some important information about the true underlying function, difficulties with bandwidth or tuning parameter selection. In contrast, one can avoid these issues at the cost of enforcing some qualitative shape constraints on the function to be estimated. We confine our survival analysis studies to estimating a hazard function since it may make a lot of practical sense to impose certain shape constraints on it. Specifically, we study the problem of nonparametric estimation of a hazard function subject to convex shape restrictions, which naturally entails monotonicity constraints.

In this thesis, three main objectives are addressed. Firstly, the problem of nonparametric maximum-likelihood estimation of a hazard function under convex shape restrictions is investigated. We introduce a new nonparametric approach to estimating a convex hazard function in the case of exact observations, the case of interval-censored observations, and the mixed case of exact and interval-censored observations. A new idea to handle the problem of choosing the minimum of a convex hazard function estimate is proposed. Based on this, a new fast algorithm for nonparametric hazard function estimation under convexity shape constraints is developed. Theoretical justification for the convergence of the new algorithm is provided. Secondly, nonparametric estimation of a hazard function under smoothness and convex shape assumptions is studied. Particularly, our nonparametric maximum-likelihood approach is generalized for smooth estimation of a function by applying a higher-order smoothness assumption of an estimator. We also evaluate the performance of the estimators using simulation studies and real-world data. Numerical studies suggest that the shape-constrained estimators generally outperform their unconstrained competitors. Moreover, the empirical results indicate that the smooth shape-restricted estimator has more capability to model human mortality data compared to the piecewise linear continuous estimator, specifically in the infant mortality phase. Lastly, our nonparametric estimation of a hazard function approach under convex shape restrictions is extended to the Cox proportional hazards model. A new algorithm is also developed to estimate both convex baseline hazard function and the effects of covariates on survival times. Numerical studies reveal that our new approaches generally dominate the traditional partial likelihood method in the case of right-censored data and the fully semiparametric maximum likelihood estimation method in the case of interval-censored data. Overall, our series of studies show that the shape-restricted approach tends to provide more accurate estimation than its unconstrained competitors, and further investigations in this direction can be highly fruitful.

Acknowledgments

I would like to convey and express my gratitude and thanks via this piece of writing to those who helped me to bring this research journey to an end. First and foremost, I thank the almighty God for giving me strength and wisdom to finish my study. I want to express my most sincere gratitude and deepest appreciation to my academic supervisor, Dr. Yong Wang, not only for his technical knowledge and insight, but also for his encouragement and thoughtful guidance in the preparation of this thesis. Sincere thanks to Professor Chris Wild, my co-supervisor for his advice and help. Special thanks are also extended to Professor Chris Triggs, Head of the Department of Statistics for his exceptional supports and helps during my study in Auckland.

I would like to extend my deepest gratitude to my parents for their endless kindness, compassion and support, which led my studies to be accomplished. Thank you for your understanding, love, and patience. Without you, this would not have been possible.

Table of Contents

Abstract **iii**

Acknowledgments **v**

Table of Contents **vii**

List of Tables **xi**

List of Figures **xiii**

List of Abbreviations **xv**

1 Introduction **1**
- 1.1 Nonparametric Survival Analysis 1
- 1.2 Nonparametric Survival Analysis under Shape Restrictions 3
- 1.3 Contributions 4
- 1.4 Outline of the Thesis 5

2 Literature Review **7**
- 2.1 Basics of Survival Analysis 7
- 2.2 Censoring Types in Survival Analysis 8
  - 2.2.1 Right Censoring 9
  - 2.2.2 Left Censoring 9
  - 2.2.3 Interval Censoring 9
- 2.3 Nonparametric Estimation of a Survival Function 11
  - 2.3.1 Non-smooth Estimation 11
  - 2.3.2 Smooth Estimation 15
- 2.4 Cox Proportional Hazards Model 20

2.5 Shape-restricted Estimation of a Function 22

3 Nonparametric Estimation of a Convex Hazard Function 27

3.1 Introduction 27
3.2 Nonparametric Maximum Likelihood Estimation 30
3.2.1 Characterization 31
3.2.2 Theoretical Properties 33
3.3 Computation 38
3.3.1 Main Idea 38
3.3.2 Updating Masses 38
3.3.3 Derivatives of the Modified Log-likelihood Function 40
3.3.4 QR Factorization of $\mathbf{D}$ 41
3.3.5 Expanding and Contracting Support Sets 42
3.3.6 The Algorithm 43
3.4 Convergence 44
3.5 Real Data Examples 47
3.5.1 Canadian Human Mortality Data 48
3.5.2 Air Conditioner Failure Data 49
3.5.3 Angina Pectoris Survival Data 51
3.6 Simulation Study 55
3.6.1 Setup 55
3.6.2 Results 56
3.7 Summary 57

4 Nonparametric Smooth Estimation of a Convex Hazard Function 59

4.1 Introduction 59
4.2 Smooth Estimation of a Hazard Function 60
4.2.1 Kernel-based Estimation 60
4.2.2 Spline Estimators 62
4.3 Maximum Likelihood Estimation with Smoothness Assumption 63
4.3.1 Computational Algorithm 66
4.3.2 Derivatives of the Modified Log-likelihood Function 69
4.3.3 QR Factorization of $\mathbf{D}$ 71
4.4 Simulation Studies 73
4.4.1 With Exact Observations 73
4.4.2 With Exact and Interval-censored Observations 76
4.4.3 When Hazard May Become Zero 78
4.5 Real Data 80
4.5.1 Kidney Transplant Data 81

4.5.2 Canadian Lifetime Data 82
4.6 Summary 85

5 Nonparametric Estimation of a Convex Baseline Hazard Function in the Cox Proportional Hazards Model 87
5.1 Introduction 87
5.2 Nonparametric Estimation of a Convex Baseline Hazard in the Cox Model 89
5.2.1 Maximum Likelihood Estimation of a Proportional Hazards Model 89
5.2.2 Computation 91
5.2.3 Derivatives of the Modified Log-likelihood Function 93
5.2.4 QR Factorization of $\mathbf{D}$ 96
5.3 Simulation Study 99
5.4 Real Data Examples 102
5.4.1 Kidney Transplant Data 102
5.4.2 Larynx Cancer Data 103
5.5 Summary 108

6 Summary and Future Works 109
6.1 Summary 109
6.2 Future Works 111

References 113

List of Tables

3.1 Results of the algorithms for the Canadian lifetime data in 2008. 50
3.2 Air Conditioner failure data of Boeing airplanes. 51
3.3 Performance of the algorithms on the air conditioner failure data. 52
3.4 Life table of 2418 males with angina pectoris. 53
3.5 Performance of the algorithms on the survival of males with angina pectoris. 55
3.6 Results of simulation studies for $n = 400, 1600$. 56

4.1 Distributions for the simulation study. 74
4.2 Simulation results for two distributions in terms of the MISE and MHD. 77
4.3 Simulation results for a mixture of two Weibull distributions. 78
4.4 Cross-validation results for the three nonparametric density estimations for kidney transplant data. 82
4.5 Results based on three nonparametric density estimations for the Canadian lifetime data. 83

5.1 Proportional hazards model, simulation results for interval-censored data. 100
5.2 Analysis of variance table for race, gender, and age for the kidney transplant patients. 103
5.3 Analysis of variance table for stage of the laryngeal cancer patients. 105
5.4 Estimated survival probabilities for a 60- and 70-year-old patient at different years after diagnosis. 105

List of Figures

2.1 Illustration example of different types of censoring. 10

3.1 Schematic representation of a convex bathtub-shaped hazard function. 43
3.2 Histograms of the entire and a random sample of Canadian life table for year 2008. 49
3.3 NPMLE of the CNMCH algorithm for the Canadian lifetime data in 2008. 50
3.4 Column-wise plots correspond to the hazard function and the gradient curve. 52
3.5 NPMLE for the life table of males with angina pectoris. 54

4.1 Hazard plots from the distributions used in the simulation study. 75
4.2 Comparisons of hazard and density estimates for the simulated data. 79
4.3 Smooth and non-smooth estimates of the hazard function for the kidney transplant data set. 82
4.4 Smoothed hazard and density function estimates for the Canadian mortality data set. 84

5.1 Box plots of the squared errors versus each of the covariates over 100 replications. 101
5.2 Estimated survival functions for each stage of the larynx cancer. 106
5.3 Estimated baseline hazard and cumulative hazard functions for larynx cancer data. 107

List of Abbreviations

BFGS	Broyden-Fletcher-Goldfarb-Shanno
BIC	Baysian Information Criteria
BT	bathtub
CNM	constrained Newton method
CNM-MS	constrained Newton-Broyden-Fletcher-Goldfarb-Shanno method via modifying the support set
CNMCH	constrained Newton method for convex hazard
CNMSCH	consrained Newton method for smooth convex hazard
EM	expectation-maximization
EW	exponentiated Weibull
GR	generalized Rosen
HD	Hellinger distance
ICM	iterative convex minorant
ISE	integrated squared error
KL	Kullback-Leibler
KM	Kaplan-Meier
L-BFGS	limited-memory BFGS
LCH	linear convex hazard
LCHCOX	linear baseline convex hazard within the Cox model
LS	logspline
MHD	mean Hellinger distance
MISE	mean integrated squared error
MLE	maximum likelihood estimator
MSE	mean squared error
MWK	Müller-Wang kernel
NA	Nelson-Aalen
NNLS	non-negative least squares
NPMLE	nonparametric maximum likelihood estimator
PAV	pool adjacent violators
PL	partial likelihood
PS	presmoothed
SCH	smooth convex hazard
SCHCOX	smoothed baseline convex hazard in the Cox model
SNPMLE	smooth nonparametric maximum likelihood estimate
SPMLE	semiparametric maximum likelihood estimate
SR	support reduction
SRB	support reduction and bisection
TTT	total time on test

TW	Tanner-Wong

Chapter 1

Introduction

1.1 Nonparametric Survival Analysis

Survival analysis is a class of statistical methods that deal with situations where one is interested in the study of the time related to a certain event. In actuarial applications, the event of interest is most often the age of death. In biomedical sciences, this could be the time of a disease onset or the time of the first heart attack. In reliability engineering, it can be the time of failure of an electrical component. The time to the occurrence of a particular event is generally referred to as survival or lifetime data. The terminology for the analysis of the occurrence and the timing of events varies in different fields. The terms survival analysis, failure time analysis, and duration analysis are applied in biomedical, engineering, and economic or sociology studies, respectively. Throughout this thesis, we will often adopt the terminology of survival analysis.

A peculiar characteristic that often arises in survival data is known as censoring. By censored data, we mean that we have only partial information about the random variable of interest, instead of knowing it exactly. Various types of censoring schemes include right censoring, left censoring, and interval censoring. Right-censored data implies that not all of the survival times are observed, so for some of them it is only known that the event of interest occurs sometime after the end of the study. Left-censored data indicates that the subject under study has experienced the event of interest prior to the start of the study. Interval-censored data arises when one only knows that the event of interest occurred within a certain time interval; e.g., two consecutive patient monitoring times. We note that right and left censoring are special cases of interval censoring. Another well-known phenomenon that exists in time-to-event data is truncation. Truncation occurs when subjects whose failure times fail to meet certain conditions are excluded from a study. In this thesis, we consider univariate uncensored data, purely interval-censored data, and the case of interval-censored data mixed with exact observations.

The primary concern in survival analysis is to find the distribution of the event times of interest. Many statistical approaches study the direct estimation of the density or the survival function. However, assessing the risk of an individual at certain times has received considerable attention in recent years. This instantaneous rate of occurrence of a time event, given subjects having survived so far, is known as hazard. Parametric survival models, such as exponential, Weibull, and Gompertz distributions, are commonly used to model lifetimes. For reviews, see, e.g., Juckett and Rosenberg (1993), Kalbfleisch and Prentice (2002), Murthy et al. (2004), Rinne (2008), and Lai (2013). Although the advantages of parametric models include the ease in their computation, interpretation and prediction, the main limitation of parametric methods is the necessary model assumptions. This can cause the risk of model misspecification, which may lead to invalid estimates and incorrect inferences (Gómez et al., 2009; Sun, 2006).

On the other hand, nonparametric methods require fewer restrictive assumptions of the models and are more flexible and robust to model misidentification. The most popular nonparametric technique for modelling survival data is the Kaplan-Meier (KM) or product limit estimator for a survival function (Kaplan and Meier, 1958). One can derive an alternative of the KM estimator for the survival function based on the Nelson-Aalen (NA) estimator for the cumulative hazard rate function (Nelson, 1972; Aalen, 1978). The great advantages of the KM and NA estimators are their straightforward implementation and intuitive interpretation. Both estimators are step functions with jumps at the observed event times. Due to discontinuities, these estimators can have low efficiencies, depending on how censoring times are distributed. The low efficiency of the KM estimator compared to parametric survival estimators is extensively discussed by Miller (1983) (see also Efron; 1988). Moreover, the KM estimator is not able to estimate survival probabilities beyond the last observed time. In the case of heavily censored data, the traditional nonparametric estimators tend to have a few jumps with larger sizes and, therefore, we can not rely on the accuracy of the estimation. In order to retrieve the lost information in the traditional nonparametric estimators, smooth and more efficient estimators are highly desirable in such a situation. Nonparametric smoothing approaches to estimating hazard functions include kernel, spline, local likelihood, and penalized likelihood methods. Smooth hazard function estimators are practically preferred to crude estimators since these estimators can be more accurate or efficient, and they also provide flexible estimates that lend themselves to nice graphical presentation. Smooth nonparametric estimation of the hazard or density function was covered in details in many monographs; see, e.g., Watson and Leadbetter (1964), Tanner and Wong (1983), Silverman

(1986), Kooperberg and Stone (1992), Tibshirani and Hastie (1987), and Wand and Jones (1995).

1.2 Nonparametric Survival Analysis under Shape Restrictions

In a nonparametric framework, smoothing approaches, such as kernel-based and penalized likelihood methods, define specific families of estimators for solving density or hazard estimation problems. The density or hazard estimators based on kernels or penalized likelihood are frequently used in the literature; however, they suffer from non-trivial issues like bandwidth or tuning parameter selection. Both the selection difficulty and the restrictions of a parametric method can be avoided at the price of imposing some qualitative shape constraints on the true underling density or hazard function.

In many applications, one often has some prior knowledge about the shape of the underlying density or hazard function. Thus, it is reasonable to impose natural qualitative constraints on these functions. In such situations, this prior knowledge can be turned into restrictions on the shape of the function. The most plausible types of shape restrictions on a density or hazard function are monotonicity (increasing or decreasing), convexity, or bathtub-shapes. Two popular criteria for nonparametric estimation of a function under shape constraints are maximum likelihood and least squares. The milestone work of Grenander (1956) proposed a maximum likelihood estimator of a nondecreasing failure rate or a nonincreasing density. Recently, there has been a great deal of attention paid to the other kinds of shape constraints. We draw the attention of the reader to Hall et al. (2001) and Banerjee (2008) for nonparametric estimation of monotone failure rates, to Reboul (2005) for the U-shaped or unimodal functions and to Groeneboom et al. (2001) and Jankowski and Wellner (2009a) for the situations of convex density and hazard function, respectively. The reader is also referred to Hildreth (1954), Groeneboom et al. (2001), Balabdaoui and Wellner (2007) and Jankowski and Wellner (2009b) for nonparametric estimation of a function under shape restrictions through the least square method.

In particular, the nonparametric maximum likelihood estimator (NPMLE) of a function has no closed-form solution and a numerical algorithm has to be used. Although some theoretical properties of the convexity shape-constrained estimators of a function are established in the literature, there is no fast algorithm to solve this problem. In essence, the main motivation of this thesis is to propose new techniques that address the problems of nonparametric estimation of a function subject to shape constraints and then develop fast and simple computational algorithms to solve these problems. Another interest is to provide a comprehensive algorithm that covers the case of exact observations, interval-censored data, and the situation of interval-censored data mixed with exact observations. In a nutshell, without any distribution assumption, but with only knowledge of the shape of the underlying density or hazard function, one can potentially obtain a more accurate estimate by a shape-constrained method.

Specifically, we focus on nonparametric estimation of a hazard function, since it can be practically more sensible to impose certain shape restrictions on it than a survival function. In this thesis, we consider the bathtub-shaped or convexity shaped constraint on a hazard function, which naturally entails both the increasing and decreasing shape restrictions. We describe a new idea to overcome the issue of choosing the minimum of a convex hazard function. Particularly, a series of research studies on nonparametric convex hazard function estimators are conducted to compare our proposed estimators with the other existing ones. We further study the smoothed nonparametric hazard function under convexity restriction. The resulting density estimators are compared with those based on the kernel-based and logspline approaches. In the presence of some additional features such as age, gender, and blood pressure that may affect the failure times, one may be interested in estimating the baseline hazard function and the effects of covariates simultaneously. Consequently, we consider nonparametric estimation of a convex baseline hazard function within the Cox model. We illustrate our proposed approaches with some simulated and real data sets and assess the performance and efficiency of the new methods compared to unconstrained ones.

1.3 Contributions

The primary goal of this thesis is to propose a novel approach to estimating a hazard function under convexity restriction. The proposed framework is not only applicable to the case of exact observations, but can also deal well with the situation of censored data. Moreover, this framework extends in a straightforward manner for the Cox proportional hazards model in order to estimate the baseline hazard function along with the effects of covariates that may influence the times to failure of a system. Our proposed approaches can also be adjusted to accommodate monotone increasing or decreasing shape constraints.

The main contributions of this thesis are given as follows:

- Study nonparametric estimation of a hazard function subject to convex shape constraint.
- Present a novel idea to overcome the difficulty of choosing the minimum of a convex hazard function estimate during its computation.
- Develop a new algorithm for computing the nonparametric maximum likelihood estimate of a convex hazard function.
- Assess the performance of the nonparametric maximum likelihood estimate of a convex hazard function using both simulation studies and real data sets in different scenarios.

- Study nonparametric likelihood-based estimators under smoothness and convex shape assumptions.
- Propose a new algorithm for nonparametric estimation of a convex hazard function under smoothness assumption.
- Evaluate and compare the performance of the smoothed and non-smoothed density estimators with the other competitors.
- Derive the nonparametric maximum likelihood estimate of a convex baseline hazard within the Cox model.
- Develop a new algorithm to simultaneously estimate both convex baseline hazard function and the effects of covariates on survival times.
- Investigate the performance of our proposed methods through simulation studies and illustrate them by a variety of real-world data sets.

1.4 Outline of the Thesis

The thesis is organized as follows. In Chapter 2, we first provide a description of the fundamental concepts in survival analysis. Then, a brief overview of the existing survival analysis techniques to model survival data is given. Furthermore, we give a review of the nonparametric estimation methods for shape-restricted functions. The strengths and weaknesses of different methods are also discussed.

Chapter 3 introduces a nonparametric maximum likelihood estimator of a convex hazard function in the case of, respectively, exact observations, interval-censored data, and interval-censored data mixed with exact observations. The computational difficulty with estimating the minimum of a convex hazard function and the new idea for solving this problem are also discussed. A general algorithm is presented for estimating a hazard function under convex shape restriction. Moreover, in the case of exact observations, some theoretical properties of the nonparametric convex hazard function estimator are established and a proof of convergence for the proposed algorithm is given.

As an extension of Chapter 3, Chapter 4 is concerned with the nonparametric estimation of a function under both convex shape and smoothness assumptions. We derive the nonparametric maximum likelihood estimator of a convex hazard function based on the smoothness assumption. Moreover, a new algorithm for nonparametric hazard function estimation under these assumptions is outlined. Furthermore, numerical studies to compare the performance of our proposed methods with the presmoothed kernel-based and logspline approaches using simulated and real data sets are performed.

Chapter 5 considers the problem of nonparametric maximum likelihood estimation of a shape-constrained baseline hazard function in the semiparametric Cox model. A general hybrid algorithm is proposed for computing the nonparametric maximum likelihood estimator of a convex baseline hazard along with the estimation of the covariate effects. In addition, we report the results of empirical studies of simulation and real-world data sets that compare the performance of our shape-restricted estimators to the traditional estimator.

Finally, Chapter 6 presents the conclusion of this thesis and depicts some interesting aspects for future works related to the dissertation.

Chapter 2

Literature Review

2.1 Basics of Survival Analysis

This section describes the basic entities used in modeling univariate survival data. We intend to present the essential notations, definitions, and fundamental facts about survival analysis that will be used in subsequent sections and chapters.

Specifically, let T denote a nonnegative continuous random variable representing the failure time of an individual in a homogeneous population, which is the survival variable of interest. Four functions are commonly used to describe the distribution of failure or survival time T, namely the survival function, the probability density function, the hazard rate function, and the cumulative hazard function. For a continuous random variable T, let f be the probability density function and $F = \Pr(T \leq t) = \int_0^t f(u)\,\mathrm{d}u$ the corresponding cumulative distribution function. Then, the first basic quantity used to characterize the distribution of T is the survival function and is defined as

$$S(t) = 1 - F(t) = \Pr(T > t) = \int_t^\infty f(u)\,\mathrm{d}u, \quad 0 < t < \infty,$$

which is the probability of an individual to survive to time t or beyond. Another major function that describes the distribution of failure time T is the hazard rate function, which is given by

$$h(t) = \lim_{\Delta t \to 0} \frac{\Pr(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t}.$$

The hazard function can be interpreted as the instantaneous failure or death rate at time t, given that the individual has not failed before time t. In what follows, we simply refer to $h(t)$ as the "hazard function" instead of the" hazard rate function" for short. Assume that the distribution of T is absolutely continuous. Therefore, the hazard function can also be represented as

$$h(t) = \frac{f(t)}{1 - F(t)} = \frac{f(t)}{S(t)} = -\frac{\mathrm{d} \log S(t)}{\mathrm{d}t}. \tag{2.1}$$

Another feature of the event time distribution is the cumulative hazard function, which is defined as

$$H(t) = \int_0^t h(u) \, \mathrm{d}u.$$

On the other hand, the cumulative hazard function can also be obtained in terms of the cumulative distribution function or the survival function in the following way

$$H(t) = -\log[1 - F(t)] = -\log S(t).$$

Thus, it can be shown that

$$S(t) = \exp[-H(t)] = \exp\left[-\int_0^t h(u) \, \mathrm{d}u\right]. \tag{2.2}$$

Eventually, it follows immediately from (2.1) and (2.2) that

$$f(t) = h(t) \exp[-H(t)]. \tag{2.3}$$

Therefore, there is a one-to-one relationship between functions h, S, H and f.

2.2 Censoring Types in Survival Analysis

A major challenge in survival analysis is how to deal adequately with a peculiar characteristic that often arises in time-to-event data. This special feature of survival data is the presence of censoring, such as right censoring, left censoring, and interval censoring. By censoring, it implies that the information about an observation on a survival time of interest is incompletely known due to the temporal limits of the observation interval or cost. Obviously, survival times can also be observed exactly.

2.2.1 Right Censoring

The most common type of censored data is right-censored data. Right censoring exists when the event of interest has not occurred before the study ends at a predetermined point of time or a subject under study is lost to follow-up before an event happens. Within the right censoring model, let us assume that there exists a lifetime T and a fixed censoring time C_r. Suppose that $T_1, T_2, \cdots, T_n$ be independent identically distributed (i.i.d.) event times from an absolutely continuous cumulative distribution function F_T with density f and the cumulative distribution function F_{C_r} of the i.i.d. censoring times $C_{r1}, C_{r2}, \cdots, C_{rn}$ is absolutely continuous with density g. The event times T_i and the censoring times C_{ri} are usually assumed to be independent. Then, the observed data can be represented by pairs of random variables (X_i, δ_i), where $X_i = \min(T_i, C_{ri})$ and $\delta_i = I(T_i \leq C_{ri})$.

2.2.2 Left Censoring

The other kind of censoring data, which is relatively rare, is left censoring. By left-censored data, we mean that the event of interest is only known to have already occurred before the start of the study. Assume that the lifetime of an individual is denoted by T and C_l is the left censoring time. The observed data in the left censoring scheme consists of the pairs of random variables $(X_i, \delta_i), i = 1, \cdots, n$, where $X_i = \max(T_i, C_{li})$ and $\delta_i = I(T_i \geq C_{li})$.

2.2.3 Interval Censoring

The more challenging censoring mechanism that predominantly takes place in medical, clinical trials and the longitudinal studies that entail periodic follow-ups is interval censoring. Interval-censored data implies that the failure time T can not be directly observed, but is only known to have occurred within a time window or interval, e.g., two adjacent examination times in a sequence of clinical visits.

Let T_i denote a nonnegative random variable representing the survival time of interest for subject i, $i = 1, \cdots, n$. An exact observation T may not be directly observable, but instead may be censored by an interval of time between two values L and R such that $T \in (L, R]$, where $L \leq R$. Therefore, the observed event times are defined by

$$O_i = (L_i, R_i] \subset [0, \infty), \; i = 1, \cdots, n.$$

Interval-censored failure time data can be regarded as a generalization of exact, right-censored, and left-censored observations. That is, if $L = R$, we have an exact observation, whereas $R = \infty$ gives a right-censored observation, and $L = 0$ represents a left-censored observation. For illustration purposes, we consider an example data set that contains interval-censored data. This example consists of four subjects under study. As shown in Figure 2.1, subject A represents a left-censored event time with its interval $O_{\mathrm{A}} = (0, 1]$, whereas subject B is an example of exact observation since $O_{\mathrm{B}} = [2, 2] = \{2\}$. Moreover, subject C illustrate a right-censored event time with $O_{\mathrm{C}} = (4, \infty]$ and subject D presents an example of interval-censored observation as $O_{\mathrm{D}} = (2, 4]$.

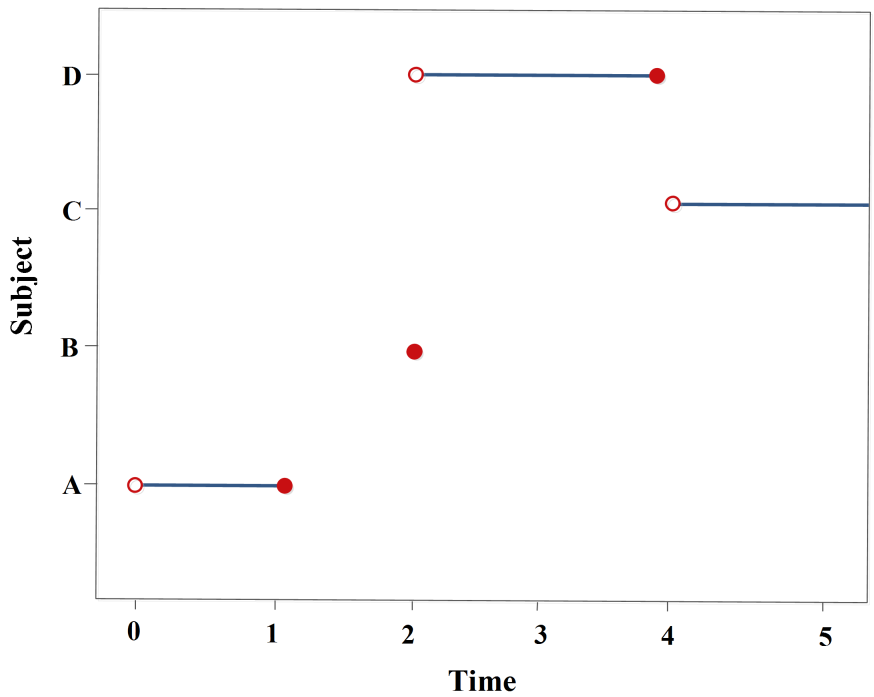

Figure 2.1: Illustration example of different types of censoring.

The two common types of interval censored data is case I interval-censored data and case II interval censored data. In the situation of case I interval-censored data or current status data, each subject under study is observed only once and the only information about the failure time T is that it has been observed to be either smaller or larger than the examination or observation time. In this case, the observed failure time is either left- or right-censored, which means either $L = 0$ or $R = \infty$. Current status data are frequently encountered in animal tumorigenicity experiments, since the time of tumor onset is not observed and only the death or sacrifice time of an animal is known (see Finkelstein and Wolfe, 1985 and Gómez and Van Ryzin, 1992).

On the other hand, case II or general interval-censored data occur when the failure time of interest T is only known to have occurred between two adjacent examination times such that $L < T \leq R$, before the first examination time ($T \leq L$), or after the last examination time

$(T > R)$. The situation of mixed exact and interval-censored observations was generalized by Yu et al. (2000) in order to involve exact observations as well.

As noted by Self and Grossman (1986), the interval censoring independence assumption specifies

$$P(T \leq t \mid L = l, R = r, L \leq T < R) = P(T \leq t \mid l \leq T < r),$$

that is, the interval $(L, R]$ does not provide any extra information about the survival time T except for the fact that T lies between the two realised values l and r (see also Oller et al., 2004 and Sun, 2006).

2.3 Nonparametric Estimation of a Survival Function

2.3.1 Non-smooth Estimation

Parametric distributions, such as the exponential and Weibull distributions, are commonly used for modeling lifetime data. However, the exponential and Weibull models can not adequately describe many functions in practical problems owing to their relative simplicity and restrictiveness. The popularity of parametric models is due to their simplicity and manageable closed-form analytical solutions, while their main drawback is the model validity issue that may result in invalid inferences and conclusions. The pros and cons of different parametric models can be found in the books by Kalbfleisch and Prentice (2002) and Lawless (2003). As a consequence, the model misspecification problem of parametric methods motivated many researchers to develop nonparametric approaches, which demand fewer restrictive assumptions for modeling survival data.

Let $T_i, i = 1, \cdots, n$, be a sample of the failure times. If there are no censored observations in the sample, the estimate of the distribution or survival function can be obtained by the empirical distribution or survival function as follows, respectively,

$$\tilde{F}(t) = \frac{1}{n}\sum_{i=1}^{n} I(T_i \leq t) \quad \text{or} \quad \tilde{S}(t) = \frac{1}{n}\sum_{i=1}^{n} I(T_i > t).$$

In the case of right-censored data, classical nonparametric estimators are well established in the literature. The simplest nonparametric estimator of the survival function was proposed by Kaplan and Meier (1958). Following the notation mentioned in Section 2.2.1 about right censoring observations, let $(X_{(i)}, \delta_{(i)})$, $i = 1, \cdots, n$, be the (X_i, δ_i) ordered with regard to the X_i's. Thus, the KM estimator, also known as the product limit estimator, of the survival function is given by

$$\hat{S}^{\mathrm{KM}}(t) = \begin{cases} 1, & \text{if } 0 \leq t \leq X_{(1)} \\ \prod_{i=1}^{j-1} \left(\frac{n-i}{n-i+1}\right)^{\delta_{(i)}}, & \text{if } X_{(j-1)} < t \leq X_{(j)},\ j = 2, \cdots, n, \\ 0, & \text{if } t > X_{(n)} \end{cases} \tag{2.4}$$

(Wand and Jones, 1995; Bantis et al., 2012). The asymptotic variance of $\hat{S}^{\mathrm{KM}}(t)$ can easily be computed by Greenwood's formula. The straightforward computation and ease of understanding this classical nonparametric estimator have made it very attractive. Nevertheless, Miller (1983) noted that the asymptotic efficiency of this nonparametric estimator can be as low as 50 percent of that of parametric ones (Whittemore and Keller, 1986). In the case of no censoring, the KM estimator coincides with the empirical survival function estimator. Furthermore, if the largest event time corresponds to a censored observation, the KM estimator of the survival function is restricted to the last observed event $t_{\max}$. Several schemes have been suggested to remedy this deficiency. For example, estimating survival probabilities at time points beyond the largest event time $t_{\max}$ by 0 was suggested by Efron (1967), which results in a negatively biased estimator. Also, setting survival probabilities equal to $\hat{S}^{\mathrm{KM}}(t_{\max})$ for $t > t_{\max}$ was proposed by Gill (1980), which leads to a positively biased estimator.

As an alternative estimator of the survival function to the KM, the NA estimator of $H(t)$ was proposed by Nelson (1972), further developed by Aalen (1978), and is known to perform better in the case of small-sized data (Klein and Moeschberger, 2003). The NA estimator for the cumulative hazard function has the form

$$\hat{H}^{\mathrm{NA}}(t) = \sum_{i:X_{(i)} \leq t} \frac{\delta_{(i)}}{n-i+1},$$

(Cao et al., 2005). Apparently, $\hat{S}^{\mathrm{KM}}(t)$ and $\hat{H}^{\mathrm{NA}}(t)$ are step functions with discontinuities or jumps located only at the observed death times t_i's. In the situation with heavily censored data, the KM and NA estimators tend to have only a few number of jumps and thus larger jump sizes (López-de Ullibarri and Jácome, 2013b). Therefore, they might waste some important information about the true underlying function and fail to efficiently use the information that is available in the data.

In the case of current status data, suppose that T_i denotes the failure time of the ith subject with distribution F. Let U_i be an observation or examination time for subject i such that T_i is independent of U_i. Thus, the observed data consists of the random pairs

$(\delta_i, U_i), i = 1, 2, \cdots, n$, where $\delta_i = I(T_i \leq U_i)$. Then the likelihood function is

$$L(F) = \prod_{i=1}^{n} \{1 - F(U_i)\}^{1-\delta_i} \{F(U_i)\}^{\delta_i}.$$

The nonparametric maximum likelihood estimate (NPMLE) of a distribution function based on current status data has received considerable attention starting with the forefront work of Ayer et al. (1955), who introduced the pool adjacent violators (PAV) algorithm (see also Robertson et al., 1988). The asymptotic properties of the NPMLE of a distribution function can be found in Groeneboom and Wellner (1992) and Huang and Wellner (1995a).

For case II or general interval censored data, we consider a study with n independent subjects and let T_i denote the survival time of interest for subject i with distribution F. Each subject has a sequence of inspection times where the inspection times and survival times are independent. This implies that the censoring is noninformative. Recall that the observed data have the form $\{O_i\}_{i=1}^n$, where $O_i = (L_i, R_i]$ is the interval known to contain the unobserved survival time T_i. Therefore, the likelihood function can be written as

$$L(F) = \prod_{i=1}^{n} \{F(R_i) - F(L_i)\}.$$

Let $\{s_j\}_{j=0}^m$ define the unique ordered elements of $\{0, \{L_i\}_{i=1}^n, \{R_i\}_{i=1}^n\}$. Peto (1973) and Turnbull (1976) noted that the likelihood does not depend on how F changes between s_j's, but it only relies on F through the values $\{F(s_j)\}_{j=1}^m$. Also, let $\boldsymbol{p} = (p_1, \cdots, p_m)^\top$ be a point in the $(m-1)$-dimensional probability simplex

$$\mathcal{P} \equiv \left\{\boldsymbol{p} : \boldsymbol{p}^\top \mathbf{1} = 1, \boldsymbol{p} \geq \mathbf{0}\right\},$$

where $\mathbf{0} = (0, \cdots, 0)^\top$ and $\mathbf{1} = (1, \cdots, 1)^\top$. Define $p_j = F(s_j) - F(s_{j-1})$ and $\delta_{ij} = 1$ if $s_j \in (L_i, R_i]$, and $\delta_{ij} = 0$ otherwise. Given $\boldsymbol{p}$, the probability for the survival time to be in O_i is

$$P_i \equiv P_i(\boldsymbol{p}) = \sum_{j=1}^{m} \delta_{ij} p_j.$$

Thus, the log-likelihood function of $\boldsymbol{p}$ can be written as

$$\ell(\boldsymbol{p}) = \sum_{i=1}^{n} \log \left(\sum_{j=1}^{m} \delta_{ij} p_j \right).$$

The NPMLE $\hat{\boldsymbol{p}}$ maximises $\ell(\boldsymbol{p})$ among all $\boldsymbol{p} \in \mathcal{P}$ (Gentleman and Geyer, 1994). For each interval O_i, the different interpretation of each left and right endpoint can be open, semi-open or closed. The continuous failure time T has no effect on how the likelihood function is expressed, whereas the discrete failure time T has an influence. Therefore, in the case of a discrete lifetime, one should pay attention to how the NPMLE assigns the probability masses on $\mathbb{R}_+$. Consequently, failing properly to deal with them might lead to strange results (Ng, 2002).

The nonparametric inference in the case of interval-censored data is much more difficult to handle than for right-censored data from both practical and theoretical perspectives. Specifically, iterative algorithms are frequently used to find the NPMLE of a function because there is no explicit solution for it. For computing the NPMLE of F, a number of algorithms have been developed in the literature; e.g., the constrained Newton-Raphson method (Peto, 1973), the self-consistency algorithm (Turnbull, 1976) being an earlier application of the expectation-maximization (EM) algorithm (Dempster et al., 1977), the iterative convex minorant (ICM) algorithm (Groeneboom and Wellner, 1992; Jongbloed, 1998), the hybrid ICM-EM algorithm (Wellner and Zhan, 1997), the generalized Rosen (GR) and hybrid GR-EM methods (Zhang and Jamshidian, 2004), the subspace-based Newton (Dümbgen et al., 2006), the constrained Newton method (CNM) (Wang, 2007), the support reduction (SR) algorithm (Groeneboom et al., 2008), and the hierarchical CNM (Wang and Taylor, 2013). In addition, a general dimension reduction technique was suggested by Wang (2008) for efficient computation of the NPMLE of a survival function. Numerical studies by Wang (2008) showed that the dimension-reducing technique performs reasonably well, especially for purely interval-censored data. The advantages and disadvantages of various algorithms have also been discussed. For a review of methods for analysing interval-censored data, we direct the reader to Gómez et al. (2009) and Zhang and Sun (2010).

Due to the difficulty of computing the NPMLE for interval-censored data, a common approach is to utilize imputation technique in order to reduce the problem of analysing interval-censored data to that of analysing right-censored data (see Sun, 2006, Chapter 2). By substituting an interval-censored observation with one or more points from that interval, the classical nonparamteric approaches can be used. However, the traditional nonparametric techniques throw away some of the information available in the data, and, therefore, produce biased estimates (Wang and Taylor, 2013).

2.3.2 Smooth Estimation

Kernel-based Estimation

In order to recover some parts of lost information in the classical nonparametric estimators of a function, the underlying function can often be assumed safely to be smooth. There is a rich literature in the area of smoothing techniques for nonparametric estimation of a function. Kernel-based methods are one of the most popular smoothing nonparametric approaches. The earliest work on the problem of kernel-based density estimation in the case of no censoring was investigated by Silverman (1986) (see also Simonoff, 1996). Assume that $T_1, \cdots, T_n$ are a random sample from a continuous univariate density f. The kernel-smoothed estimator of $f(t)$ in the case of no censoring is given by

$$\hat{f}(t) = \frac{1}{nb} \sum_{i=1}^{n} K\left(\frac{t - T_{(i)}}{b}\right),$$

where $K(\cdot)$ is the kernel function for which $\int K(t)\,\mathrm{d}t = 1$ and $b > 0$ is called the bandwidth.

The kernel-based density estimation in the presence of censoring has been discussed in Wand and Jones (1995) and Bowman and Azzalini (1997). In the case of right censoring, the kernel density estimator can be obtained based on the KM estimator (2.4). Therefore, the kernel density estimator is given by

$$\hat{f}_T(t; b) = \sum_{i=1}^{n} s_i K_b(t - X_{(i)}),$$

where $K_b(u) = b^{-1}K(u/b)$ and s_i denote the size of jump of the KM estimator at $X_{(i)}$ (see Wand and Jones, 1995 and Bantis et al., 2012). The reader is referred to Wand and Jones (1995) for a review of commonly used bandwidth selection methods. The downside of the kernel smooth estimator in the case of censoring is the phenomenon that the last event time is censored due to the fact that $s_i = 0$ if and only if $X_{(i)}$ corresponds to a censored observation. Essentially, the kernel estimate is established by centering a kernel at each event time. In essence, the density estimator can not be expanded beyond the tail of the kernel located at the last event time, even if the last event tracked by censored data (Bantis et al., 2012).

The kernel hazard function estimator was firstly introduced by Watson and Leadbetter (1964) for the case of uncensored data and then further developed by Ramlau-Hansen (1983) and Tanner and Wong (1983) to right-censored data. The asymptotic properties of the kernel-based hazard function estimator are investigated by many authors with different techniques;

e.g., Rice and Rosenblatt (1976) for exact observations and Tanner and Wong (1983) for censored data.

Let $T_1, \cdots, T_n$ be a random sample of independent failure times with distribution F and density f. The kernel-based estimator of the hazard function in the case of uncensored data is given by

$$\hat{h}(t) = \frac{1}{b} \sum_{i=1}^{n} K\Big(\frac{t - T_{(i)}}{b}\Big) (n - i + 1)^{-1}.$$

where $T_{(i)}$ be the i-th order statistic.

The fixed-bandwidth kernel-smoothed hazard function estimator for right-censored data is described by

$$\hat{h}(t) = \frac{1}{b} \sum_{i=1}^{n} K\Big(\frac{t - X_{(i)}}{b}\Big) \frac{\delta_{(i)}}{n - i + 1},$$

where $X_{(i)}$ is the i-th ordered time, K is a fix kernel function and b denotes the bandwidth that determines the degree of smoothing (Ramlau-Hansen, 1983). The most well-known kernels are the Epanechnikov, the Biweight, the Triweight, the Gaussian, the Triangular and the Uniform. The choice of kernel function does not much effect on the results of estimates as noted by Silverman (1986) (see also Wand and Jones, 1995 and Simonoff, 1996). In contrast, the choice of smoothing or bandwidth parameter is a crucial factor in the performance of the kernel-based estimator (Gefeller and Michels, 1992). Also, the idea of the presmoothed procedure based on considering a smoother estimator of $p(X_{(i)})$ instead of $\delta_{(i)}$ for estimating a density or a hazard function can be found in work of Cao and Jácome (2004) and Cao and López-de Ullibarri (2007). A more detail description of this approach is investigated in Chapter 4.

The popularity of the kernel smooth estimator is due to its simplicity and intuitive nature, whereas its main drawback is the difficulty of determining the optimal bandwidth. The main shortcoming of the fixed-bandwidth kernel estimators is caused by the constant bandwidth which results in unpleasant effects whenever the data are not equally distributed throughout the range of interest (Gefeller and Michels, 1992). In such situation, the fixed-bandwidth kernel estimator tends to over-smooth in dense regions, while under-smooth in a sparse region with many misleading peaks (Hess et al., 1999). To achieve a uniform degree of smoothing over the range of data, one should choose a large bandwidth when the data are sparse, whereas the bandwidth should be small when the data are dense (Gefeller and Michels, 1992). In order to overcome this problem, one can apply the idea of the nearest neighbor into the definition of the bandwidth (Gefeller and Dette, 1992) .

Furthermore, the boundary effects problem arises when the kernel estimation near the endpoints of the data is biased. These effects are aggregated when the hazard function has high derivatives near the extremes of the data; e.g., bathtub-shaped hazard function (Müller and Wang, 1994). Due to the incident of boundary effects, some authors have used different kernels near the endpoints of the data; see e.g., Gray (1990) and Müller and Wang (1994). A solution to removing boundary effects was developed by Hougaard (1988). Moreover, Müller and Wang (1994) introduced a new class of the hazard function estimator with varying kernels and varying bandwidths as follows

$$\hat{h}(t) = \frac{1}{b(t)} \sum_{i=1}^{n} K_t\Big(\frac{t - X_{(i)}}{b(t)}\Big) \frac{\delta_{(i)}}{n - i + 1},$$

where the bandwidth $b(t)$ and K_t depend on the point t where the estimate is to be assessed (Müller and Wang, 1994). Some new boundary kernels and local and global bandwidth choices can be found in R package **muhaz** (Hess and Gentleman, 2010).

Following notations in Section 2.2.3, the kernel hazard function estimator in the case of interval-censored data is given by

$$\hat{h}(t) = \sum_{j=1}^{m} w_j(t)\hat{h}_j,$$

where

$$w_j(t) = \frac{w_j^{\star}(t, b)}{\sum_{u=1}^{m} w_u^{\star}(t, b)}$$

and

$$w_j^{\star} = b^{-1} K \left\{(t - s_j)/b\right\},$$

for $j = 1, \cdots, m$. The reader is referred to Eubank (1999) and Sun (2006) for a more detailed description. The kernel-based method is not likelihood-based and, therefore, straightforward inferences and conclusions can not be made, as noted by Sun (2006).

Spline Estimators

Another type of approaches use splines functions to approximate hazard functions. The hazard function is considered as a piecewise polynomial function of some spline functions joined at knots. A quadratic spline estimator with knots placed at each decile of the sample's distribution was first introduced by Bloxom (1985) in the case of exact observations. A linear spline hazard function estimator where knots are placed at each unique failure time was also developed by Whittemore and Keller (1986). The deficiency of the former model is due to imposing constraints on parameters in order to determine the shape of the hazard function and the limitation of the latter is that a large number of parameters are required to be estimated (Herndon and Harrell, 1990).

An alternative class of methods for smooth hazard function estimation is the logspline model. Logspline models have been studied by many authors; see e.g., Stone (1987, 1990) and Kooperberg and Stone (1991). Kooperberg and Stone (1992) extended the idea of the logspline density estimation for censored data (see also Koo et al., 1999). Basically, the logspline approach models the logarithm of a density function as a spline function. A cubic spline often is used, which is a twice-continuously differentiable function. Suppose that L and U are some numbers such that $-\infty \leq L < U \leq \infty$ and the time line is partitioned into K intervals $(0, \tau_1], \cdots, [\tau_k, \infty)$ where τ_i are knots. Then, the logspline density model is written as

$$f(x; \boldsymbol{\theta}) = \exp\left(\sum_{i=1}^{p} \theta_i B_i(x) - C(\boldsymbol{\theta})\right), \quad L < x < U,$$

where

$$C(\theta) = \log\left\{\int_L^U \exp(\sum_{i=1}^{p} \theta_i B_i(x))\, \mathrm{d}x\right\} < \infty$$

makes $f(x; \boldsymbol{\theta})$ a proper density function and $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_p)^\top$ is the parameter vector. Also, the basis functions $B_1, \cdots, B_p$ can be selected such that B_1 and B_p are linear with negative and positive slope on $(L, \tau_1]$ and $[\tau_k, U)$, respectively, $B_2, \cdots, B_p$ and $B_1, \cdots, B_{p-1}$ are constant on $(L, \tau_1]$ and $[\tau_k, U)$, respectively, and we have a cubic polynomial in each of the intervals $[\tau_1, \tau_2], \cdots, [\tau_{k-1}, \tau_k]$. For specifying the number of knots, a stepwise knot addition and deletion procedure based on the Wald statistic was applied by Kooperberg and Stone (1992). For model selection, they suggested the Baysian Information Criteria (BIC). The implementation of the knot addition and deletion algorithm can be found in R package **logspline** (Kooperberg, 2007). It is noteworthy that in the case of heavily censored data, the logspline estimation algorithm may fail to converge with an error message of "no convergence" (Pan, 2000b).

Likelihood-based Methods

Smooth estimation of hazard functions can also be obtained by using some likelihood-based approaches such as penalized likelihood methods and local likelihood methods. Recall that T denotes the survival time of interest in a survival study and an interval-censored observation is represented by $O_i = (L_i, R_i]$, $i = 1, \cdots, n$. An exact observation is indicated by $T_i = L_i = R_i$. For simplicity, we assume that the first n_1 observations are exact and the rests are interval-censored. Then the log-likelihood function has the form

$$L(h) = \prod_{i=1}^{n_1} \left\{ h(T_i) e^{-H(T_i)} \right\} \prod_{i=n_1+1}^{n} \left\{ e^{-H(L_i)} - e^{-H(R_i)} \right\}.$$

Without further restrictions, the resulting estimate is not smooth, whereas the true hazard function is often smooth. Assume that $h(\cdot)$ is a member of the class of continuous, twice differentiable functions, and its second derivative is square integrable. For a smooth estimate of $h(t)$, the penalized likelihood approach which maximizes the log-likelihood function adjusted by a penalty function is defined as follows

$$\ell_p(h) = \ell(h) - kJ(h), \; k > 0,$$

where $\ell(h) = \log L(h), \; J(h) = \int \left\{ h^{''}(u) \right\}^2 \mathrm{d}u$ is a known penalty function that measures the roughness of the hazard function and k is the smoothing parameter which controls the balance between the fit to the observed data and the smoothness of the hazard function. A combination of the penalized likelihood approach and the modeling of the hazard function was also suggested by Rosenberg (1995). For a review of the penalized likelihood approach, we direct the reader to Senthilselvan (1987) and Joly et al. (1998).

One can also use local likelihood methods for smooth estimation of a hazard function. In this method, the hazard rate is estimated at each time point by some parametric functions, e.g., linear function of times. Tibshirani and Hastie (1987) introduced the concept of using local fitting for likelihood-based regression models to the class of generalized linear models, which was further developed by Fan et al. (1998). A smooth hazard estimation using the local likelihood method was also proposed by Betensky et al. (1999) for interval-censored data. In the region of sparse data such as the right tail of the hazard function, numerical instability is a disadvantage of this technique (Cai and Betensky, 2003; Sun, 2006).

2.4 Cox Proportional Hazards Model

The proportional hazards or Cox model is one of the most popular methods for anlayzing right-censored survival data in the presence of covariates. Basically, the Cox model is semi-parametric since it explores the relationship between the survival time and the covariates through an arbitrary baseline hazard function and the exponential of a regression function of the covariates. Within the Cox model, the hazard function of the continuously distributed survival time T with covariate vector $\boldsymbol{Z} \in \mathbb{R}^p$ has the form

$$h(t; \boldsymbol{Z}) = h_0(t) \exp(\boldsymbol{\beta}^\top \boldsymbol{Z}),$$

where $h_0(t)$ is an unspecified baseline hazard function and $\boldsymbol{\beta} \in \mathbb{R}^p$ is the vector of the unknown regression parameters. The Cox's partial likelihood (PL) approach was developed by Cox (1972, 1975) to estimate the regression coefficients of proportional hazard models, where the baseline hazard function is not involved. The main advantage of the PL approach is that it does not require estimation of the baseline hazard function h_0 when the main interest is to estimate the regression coefficients $\boldsymbol{\beta}$. Moreover, the asymptotic properties of the maximum partial likelihood $\boldsymbol{\beta}$ were well established by Tsiatis (1981) (see also Andersen and Gill, 1982).

Let X_i denote the event time with the corresponding censoring time C_i. Then, the observed data consists of i.i.d. samples of the triple $(T_i, \delta_i, \boldsymbol{Z}_i)$ for $i = 1, \cdots, n$, where $T_i = \min(X_i, C_i)$ is the follow-up time, $\delta_i = I(X_i \leq C_i)$, and $\boldsymbol{Z}_i \in \mathbb{R}^p$ denotes the covariate vector. Also, the event time X and the censoring time C are assumed to be conditionally independent given $\boldsymbol{Z} = \boldsymbol{z}$. The covariate vector $\boldsymbol{Z}$ is assumed to be time fixed. We also assume that the survival time X, conditionally on $\boldsymbol{Z} = \boldsymbol{z}$, is continuous with density $f(x \mid \boldsymbol{z})$. The simplified likelihood function based on the baseline hazard h_0 and the effect parameters $\boldsymbol{\beta}$ is given by

$$\mathcal{L}(h_0, \boldsymbol{\beta}) = \prod_{i=1}^{n} \left\{ h_0(T_i) e^{(\boldsymbol{\beta}^\top \boldsymbol{Z}_i)} \right\}^{\delta_i} \exp\left\{ -e^{(\boldsymbol{\beta}^\top \boldsymbol{Z}_i)} H_0(T_i) \right\},$$

where $H_0(t) = \int_0^t h_0(u)\, du$ denotes the baseline cumulative hazard function. Therefore, the log-likelihood function can be written as

$$\ell(h_0, \boldsymbol{\beta}) = \sum_{i=1}^{n} \left\{ \delta_i \log h_0(T_i) + \delta_i \boldsymbol{\beta}^\top \boldsymbol{Z}_i - e^{(\boldsymbol{\beta}^\top \boldsymbol{Z}_i)} H_0(t) \right\}.$$

Let $X_{(1)} < X_{(2)} < \cdots < X_{(D)}$ denote the ordered distinct event times and $\boldsymbol{Z}_{(j)}$ the covariate vector of the subject whose failure time is equal to $X_{(j)}$. Also let $R(X_{(j)})$ define the risk set of individuals at time $X_{(j)}$, where all individuals who are still under study at a time just prior to $X_{(j)}$. Assume that there are no ties between the event times. Cox (1975) showed that the $\hat{\boldsymbol{\beta}}$ is the maximizer of the partial likelihood function

$$\mathcal{L}_p(\boldsymbol{\beta}) = \prod_{j=1}^{D} \frac{\exp(\boldsymbol{\beta}^\top \boldsymbol{Z}_{(j)})}{\sum_{i \in R(X_{(j)})} \exp(\boldsymbol{\beta}^\top \boldsymbol{Z}_{(i)})}.$$

In the case of tied failure times, the partial likelihood is given by

$$\mathcal{L}_p(\boldsymbol{\beta}) = \prod_{j=1}^{D} \frac{\exp(\boldsymbol{\beta}^\top \boldsymbol{Z}_{(j)})}{\left\{\sum_{i \in R(X_{(j)})} \exp(\boldsymbol{\beta}^\top \boldsymbol{Z}_{(i)})\right\}^{d_j}},$$

where d_j is the number of individuals whose failure times are equal to $X_{(j)}$ (Breslow, 1974).

In practice, one may also be interested in estimating the baseline function. The estimated baseline hazard function along with the estimation of regression coefficients can be obtained by the Breslow estimator (Breslow, 1972). For a review of smooth estimation of a baseline hazard function in the Cox model; see, e.g., Gray (1990) and Anderson and Senthilselvan (1980) based the kernel and penalized likelihood smoothing methods, respectively. Others who discussed the proportional hazards model in the case of right-censored data include Gray (1994), Joly et al. (1998) and Cai and Betensky (2003).

In the Cox model with interval-censored data, the partial likelihood approach is not available in a closed form as noted by Sun (2006), since h_0 can not be easily removed. Let the observed data be an i.i.d. sample $\{(L_i, R_i], \boldsymbol{Z}_i; i = 1, \cdots, n\}$, where $(L_i, R_i]$ denotes the observation interval for the event time T_i and $\boldsymbol{Z_i} \in \mathbb{R}^p$ is the vector of covariates for subject i. Then the likelihood function is

$$\mathcal{L}(S_0, \boldsymbol{\beta}) = \prod_{i=1}^{n} \left\{ S_0(L_i)^{\exp(\boldsymbol{\beta}^\top \boldsymbol{Z}_i)} - S_0(R_i)^{\exp(\boldsymbol{\beta}^\top \boldsymbol{Z}_i)} \right\},$$

where $S_0(t)$ denotes the baseline survival function.

Finkelstein (1986) is one of the pioneers who investigated the fitting of the proportional hazards model to interval-censored data by using a modified Newton-Raphson algorithm. Huang (1996) proposed the fully semiparametric maximum likelihood analysis for current status data, where he applied a monotone step function whose number of jumps increases with sample size in order to estimate the baseline cumulative hazard function. Betensky et al. (2002) investigated the use of the local likelihood method to simultaneously estimate the regression coefficients and the baseline hazard function. Cai and Betensky (2003) derived a smoothed estimate of the hazard function by maximizing the penalized likelihood through a mixed model-based approach. They modeled the log-baseline hazard as a linear spline.

In order to avoid the complexity of analysing interval-censored data, one can employ the (multiple) imputation approach to replace an interval-censored observation with one or more points from that interval. Thus, the standard PL analysis could be used for imputed interval-censored observations as for right-censored data. Some literature has already discussed the multiple imputation procedures in order to change an interval-censored data problem to a right-censored data one; see e.g., Satten (1996), Goggins et al. (1998) and Pan (2000a). In spite of its simplicity, this approach can lead to biased results and invalid inferences (Lindsey and Ryan, 1998).

We draw the attention of the reader to Lesaffre et al. (2005) for a review of different approaches for fitting a proportional hazards model in the case of interval-censored data. A general framework for semiparametric regression analysis of different censoring schemes was also proposed by Zhang and Davidian (2008). Recently, Zhang et al. (2010) introduced a spline-based semiparametric maximum likelihood approach within the Cox model with interval-censored data by fitting the baseline cumulative hazard function via a monotone B-spline function. As noted by Gómez et al. (2009), the available methods in the software for the proportional hazards in the case of interval-censored data are rare. The implementation of a modification of the ICM algorithm to the Cox model (Pan, 1999) for interval-censored data is available in the R package **intcox** developed by Henschel and Mansmann (2009). This package is not able to yield the standard errors for the estimated regression coefficients, and they recommended computing bootstrap intervals.

2.5 Shape-restricted Estimation of a Function

Several general parametric models have been suggested to model the bathtub-shaped failure rates; see e.g., Hjorth (1980), Haupt (1992), Mudholkar and Srivastava (1993). However, one can totally remove the issue of model misspecification of parametric models and also the problem of choosing the tuning parameters and penalty terms of smooth nonparametric approaches by virtue of the shape constraint. Two popular approaches to nonparametric estimation under shape constraints are maximum likelihood and least squares. The earliest work on shape-constrained maximum likelihood estimation can be traced back to Grenander (1956), who derived a maximum likelihood estimator (MLE) for a distribution function with a nondecreasing hazard rate for uncensored data. Bray et al. (1967) extended it to the case of a U-shaped hazard function. Some properties of distributions with monotone failure rates were studied by Barlow et al. (1963). The MLE of a nondecreasing hazard function was demonstrated by Marshall and Proschan (1965) to be a right-continuous step function which is 0 before the first observation and jumps to infinity at the largest observation. Prakasa Rao

(1969) also showed that these estimators are piecewise constant functions and converge at a rate of $n^{1/3}$.

For a sample of ordered observations $T_{(1)}, \cdots, T_{(n)}$, the log-likelihood function as a function of the hazard function h is given by

$$\ell(h) = \sum_{i=1}^{n} \left\{ \log h(T_{(i)}) - \int_0^{T_{(i)}} h(u)\,\mathrm{d}u. \right\}. \tag{2.5}$$

Since $h(T_{(n)})$ can be made arbitrarily large, Grenander (1956) considered that $\ell(h)$ can be maximised over a nondecreasing h which is bounded by some $M > 0$. Therefore, (2.5) reduces to maximizing the following function

$$\tilde{\ell}(h) = \sum_{i=1}^{n-1} \left\{ \log h(T_{(i)}) - (n-i)(T_{(i+1)} - T_{(i)})h(T_{(i)}) \right\},$$

subject to $0 \leq h_{(T_{(1)})} \leq \cdots \leq h_{(T_{(n-1)})} \leq M$ (see Robertson et al., 1988, Chapter 7). The MLE is then a right-continuous step function with values at the observations given by

$$\hat{h}(T_{(i)}) = \min_{i \leq x \leq n-1} \max_{1 \leq s \leq i} \frac{x - s + 1}{\sum_{j=s}^{x} (n-j)(T_{(j+1)} - T_{(j)})}, \quad \text{for } i = 1, \cdots, n-1,$$

(see Robertson et al., 1988, Chapter 1). The full MLE is then found by additionally letting $M \to \infty$ such that $\hat{h}(t) = \infty$ for $t \geq T_{(n)}$.

The MLE of an increasing hazard function with right-censored data was first studied by Padgett and Wei (1980). In addition, Mykytyn and Santner (1981) derived the NPMLE of a hazard rate function based on monotonicity assumptions with different censoring schemes. Huang and Zhang (1994) and Huang and Wellner (1995b) also provided the asymptotic properties of the MLE of a monotone density and a monotone hazard respectively for right-censored data. Nonparametric estimation for U-shaped or unimodal hazard function with right censoring data was also investigated by Reboul (2005). In the case of right censoring with no covariates, the limiting distribution of the likelihood ratio test has been derived by Banerjee (2008). More recently, the nonparametric maximum likelihood estimator and the least squares estimator of a convex hazard function was studied by Jankowski and Wellner (2009b). An estimator of a convex hazard function was proposed by Jankowski and Wellner (2009a) using the profile likelihood method. They showed that both estimators are consistent and converge at rate $n^{2/5}$. The MLE of a convex hazard function was proved to be a piecewise-linear function as shown in Jankowski and Wellner (2009b). Therefore, it can be expressed as

$$h(t) = \alpha + \sum_{j=1}^{k} \nu_j(\tau_j - t)_+ + \sum_{j=1}^{m} \mu_j(t - \eta_j)_+,$$

where $t_+ \equiv \max(0, t)$, $\alpha, \nu_j, \mu_j \geq 0$ denote the masses and the τ_j's and η_j's correspond to the knots, or the support points of the corresponding nonnegative measures.

Following the log-likelihood function (2.5), one could find the NPMLE $\hat{h}$ for exact observations by maximising the modified log-likelihood function, since $\ell(h)$ can be made arbitrarily large by increasing the value of h at the largest observed value, as noted by Grenander (1956). That is, one needs to maximise the modified log-likelihood

$$\tilde{\ell}(h) = \sum_{i=1}^{n-1} \log h(T_{(i)}) - \sum_{i=1}^{n} H(T_{(i)}).$$

Jankowski and Wellner (2009a) proposed an iterative two-step optimization method. They introduced a hybrid algorithm that iterates between the SR (Groeneboom et al., 2008) and bisection algorithms, which we call SRB. In the first step, the SR algorithm is applied to find the MLE over all convex hazard functions with a fixed minimum (anti-mode). In the second step, the bisection method is employed in order to find the minimum of the convex hazard function. Their double looping method makes the convergence of the algorithm very slow. The implementation of the SRB algorithm can be found in the R package **convexHaz** (Jankowski et al., 2009). Also, this package can not deal with the situation of tied event times. It fails to handle the issue of duplicated largest observation, since it only removes one largest observation, instead of all of them from the modified log-likelihood function. More recently, the literature on shape-constrained estimation techniques has been upgraded by the contributions of Meyer and Habtzghi (2011) who studied the NPMLE of a decreasing density and also increasing, convex, and increasing and convex hazard functions using regression splines. For recent work on smooth monotone hazard function estimation in the case of exact observations, we refer the interested reader to Groeneboom and Jongbloed (2013).

Within the Cox model, Chung and Chang (1994) proposed a maximum likelihood estimator for a baseline hazard function under nondecreasing shape constraint. Recently, Lopuhaä and Nane (2013) derived a nonparametric maximum likelihood estimator for estimating a monotone baseline hazard function and a decreasing baseline density in the Cox model with right-censored observations. The asymptotic properties of their estimators were also studied. The computation of different shaped-constrained NPMLEs of a baseline hazard function along with the estimation of regression coefficients within the Cox model for the case of right-censored data is available in the R package **CPHshape** (Hui and Jankowski, 2011).

This package can not provide the standard errors for the estimated regression coefficients. In addition, their estimated baseline hazard function is piecewise linear and hence is not smooth.

Chapter 3

Nonparametric Estimation of a Convex Hazard Function

3.1 Introduction

The study of hazard functions is one of the major topics of interest in biomedical studies and reliability engineering. Many parametric models have received considerable attention in this regard, mainly for reasons of straightforward implementation and ease of analysis. To avoid the strict assumptions associated with a parametric model that may lead to biased inference and invalid conclusions, one could resort to nonparametric approaches for estimating a hazard function. As a nonparametric maximum likelihood estimator, the renowned KM estimator for a survival function has been widely used in the case of right-censored data.

In practice, some prior knowledge may be available regarding the shape of the underlying hazard function; therefore, it may be natural to make use of this knowledge in estimation. With respect to such information, a reliability engineer and a medical practitioner may assess the necessity of applying some prevention or intervention action to save the mechanism or patient from failing. In reliability contexts, the test statistic based on the total time on test (TTT) plot can be used for testing if a random sample is generated from a life distriution with constant versus bathtub. If the scale TTT curve changes from convex to concave in $(0,1)$, then the hazard function has a bathtub hazard rate shape; see e.g., Bergman and Klefsjo (1984) and Aarset (1987). For nonparametric estimation of increasing, decreasing, monotone, U-shaped failure rates, we refer the reader to Grenander (1956), Proschan (1963), Hall et al. (2001) and Reboul (2005) respectively. Two popular approaches to nonparametric estimation under shape constraints are maximum likelihood and least squares. Obtaining a more acceptable, fruitful and rational nonparametric estimate of a hazard function subject to shape constraints based on maximum likelihood methods is the main motivation of this chapter.

The most important types of shape restrictions are monotonicity (increasing or decreasing) and bathtub shapes or convex. In this thesis, a convex or bathtub-shaped hazard function is particularly investigated. A bathtub-shaped hazard function consists of three periods, including burn-in, useful, and wear-out periods. The configuration of the bathtub-shaped hazard function starts at optimum burn-in time, followed by a gradual stabilization at a level for a certain time in the useful period, and finally, a hazard rate rise during the wear-out period. This kind of hazard is predominantly applied to the human population followed from birth. An increasing hazard rate often arises when there is natural aging or wear. In many practical applications, the producers frequently use a burn-in process of the products. For preventing an early failure of defective items, the operation of products is investigated before being shipped to the customers. However, after being sold, the intact released items are subject to gradual aging. Decreasing hazards appear occasionally when there is an early failure, such as modelling survival times after a successful medical treatment.

The most noticeable feature of time-to-event data is incompleteness such as censoring. By censored data, we mean that the event of interest is never observed exactly, but only known to occur to the right or left of a time boundary, or within a time interval. The interval-censored data is a generalization of the exact, left-censored or right-censored data; in such cases an interval can reduce to a single point, has a lower bound of zero, or is unbounded on the right, respectively. A typical example of interval-censored data arises often in medical, longitudinal studies, and clinical trials that entail periodic monitoring of the progression or changes in a disease status. In such situations, many patients may miss one or more predetermined observation times and then come back to clinical centres with a changed status, or more likely have irregular visits. Accordingly, we only know that the true event time is greater than the prior clinic visit time at which there was no symptom of disease progression, and is less than or equal to the clinic visit time where the change was observed to occur. Hence, it results in an interval which includes the real but unobservable time of occurrence of the change. The readers are referred to Sun (2006) for different kinds of examples in the case of interval-censored data. Despite the high incidence of the interval-censored data in practice, literature on nonparametric estimation of a hazard function under shape constraints is relatively sparse. Therefore, we also extend our study of nonparametric hazard function estimation under convex shape constraint to the case of interval-censored data.

Nonparametric estimation of a bathtub hazard function in the uncensored situation was pioneered by Bray et al. (1967) and was extended to the case of right-censored data by Mykytyn and Santner (1981). We also urge the interested reader to see Tsai (1988), and Huang and Wellner (1995b). Banerjee (2008) presented pointwise confidence intervals for nonparametric estimators of monotone, unimodal and U-shaped hazard functions for right-censored data using nonparametric likelihood ratios. Recently, Jankowski and Wellner (2009a) studied the estimator of a convex hazard function by the profile likelihood method. The estimate is piecewise linear and has a local rate of convergence $n^{2/5}$. More recently, a monograph on NPMLE of a hazard function with increasing, convex, and increasing and convex via regression splines for both uncensored and right-censored observations was provided by Meyer and Habtzghi (2011). They showed that the estimator converges at rate $r = (p+1)/(2p+3)$ where p is the degree of the polynomial spline.

Iterative algorithms have to be used to find the NPMLE of a bathtub-shaped hazard function since it is not explicitly available. An iterative two-step optimization method was introduced by Jankowski and Wellner (2009a). Their method iterates between the SR (Groeneboom et al., 2008) and the bisection algorithm. The former is used to find the MLE over all convex hazard functions with a fixed minimum (anti-mode), and the latter is applied to optimize over other possible minimum values of a convex hazard function. This approach has drawbacks such as singling-out the minimum of a convex hazard function estimate and having a very slow convergence speed as the result of a double-looping method. Solving these problems is the motivation of this chapter.

Our new algorithm is an appropriate extension of the constrained Newton method (CNM) (Wang, 2007), which was proposed for fitting a nonparametric mixture. The main idea behind the new algorithm to overcome the above-mentioned problems is to always maintain a constant hazard segment, even if its length is zero. Therefore, we consider two situations, namely positive and zero length constant part, which are computationally interchangeable during the computation of the algorithm. In each iteration, the extended algorithm finds and adds all the global maxima of two gradient functions between every two neighbouring support points to the support sets corresponding to the decreasing and increasing parts of a convex hazard function. The procedure continues by finding a new support point that has the maximum gradients value between the last support point of the decreasing and the first support point of the increasing part of a convex hazard function. Then, we update all masses under the convexity restriction. Finally, points with zero masses are removed straightaway after each updating.

Essentially, two main aims in this chapter are to: (a) Study the nonparametric estimation problem involving convex hazard function in the case of exact observations and develop a fast computational algorithm to solve this problem, (b) Extend our algorithm to the situation with interval-censored data, which naturally includes the cases with right-censored or left-censored data, and to the mixed situation with both exact observations and interval-censored data. The procedure can also be narrowed down to a monotone (increasing or decreasing) shape.

3.2 Nonparametric Maximum Likelihood Estimation

In this section, we derive a general form for shape-restricted estimation of a hazard rate function based on maximum likelihood method from exact observations and interval censored data. In order to obtain the likelihood for types of uncensored or censored data, it is necessary to define the following terms.

Let T be a nonnegative random variable representing the time until some specified event occurs. One of the informative functions which can characterize the distribution of T is the hazard function. For a sample of times to an event with a continuous distribution function F and density f, the hazard function of T is defined as

$$h(t) = \frac{f(t)}{S(t)} = \frac{f(t)}{\exp\{-H(t)\}},$$

where $S(t) = 1 - F(t)$ is the survival function and $H(t) = \int_0^t h(u)\,\mathrm{d}u$ the cumulative hazard function. Given h, other functions such as H, F, S and f can all be easily derived. For example, one can express the density as

$$f(t) = h(t)\exp\{-H(t)\} = h(t)\exp\left\{-\int_0^t h(u)\,\mathrm{d}u\right\}.$$

We are interested in finding the NPMLE of h (or equivalently f), under the restriction of a convex h.

In our general formulation, an observation T for each subject can be either exact or interval-censored. By interval-censored, we mean that a random variable of interest is known only to lie within an interval instead of being observed exactly. For an event time T_i ($i = 1, \cdots, n$) that is interval-censored, we denote it by its censoring interval $O_i = (L_i, R_i]$, while an exact observation is indicated by $T_i = L_i = R_i$. To make representation simpler, let us assume that the first n_1 observations are exact. Hence, the log-likelihood function is given by

$$\ell(h) = \sum_{i=1}^{n_1}\{\log h(T_i) - H(T_i)\} + \sum_{i=n_1+1}^{n}\log\{S(L_i) - S(R_i)\}. \tag{3.1}$$

As suggested by Grenander (1956) in the case when all the observations are exact, $\ell(h)$ can be made arbitrarily large by increasing the value of h at the largest observed value and a modified log-likelihood should be maximized. In the mixed case with both exact and interval-censored observations, following the same argument the modified log-likelihood should take the form of

$$\tilde{\ell}(h) = \sum_{\substack{i=1 \\ T_i \neq T_{\max}}}^{n_1} \log h(T_i) - \sum_{i=1}^{n_1} H(T_i) + \sum_{i=n_1+1}^{n} \log \{S(L_i) - S(R_i)\}, \tag{3.2}$$

where $T_{(n_1)} = \max\{T_1, \cdots, T_{n_1}\}$, $I_{\max}$ is the maximum of all the finite values of L_i and R_i and $T_{\max} = \max\left\{T_{(n_1)}, I_{\max}\right\}$. Note that in the case of interval-censored data mixed with exact observations, we make a comparison between $T_{(n_1)}$ and $T_{\max}$ to take a decision on whether to choose the full or the modified log-likelihood function. If $T_{(n_1)} < T_{\max}$, then the full log-likelihood (3.1) is used as a special case of the modified log-likelihood (3.2); otherwise, we always work with the modified log-likelihood (3.2) and may even omit the adjective "modified". Therefore, the full NPMLE is obtained by additionally setting $\hat{h}(T_{(n_1)}) = \infty$ when $T_{(n_1)} = T_{\max}$.

Jankowski and Wellner (2009b) showed that the NPMLE of a convex hazard function is piecewise linear. It can hence be expressed as a piecewise linear function with three parts as follows,

$$h(t) = \alpha + \sum_{j=1}^{k} \nu_j (\tau_j - t)_+ + \sum_{j=1}^{m} \mu_j (t - \eta_j)_+, \tag{3.3}$$

where $\alpha, \nu_j, \mu_j \geq 0$. Let us denote $\boldsymbol{\pi} = (\alpha, \boldsymbol{\nu}, \boldsymbol{\mu})^\top$, the vector of positive masses, and $\boldsymbol{\theta} = (\boldsymbol{\tau}, \boldsymbol{\eta})^\top$, the vector of support points, where $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_k)$, $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_m)$, $\boldsymbol{\tau} = (\tau_1, \cdots, \tau_k)$, and $\boldsymbol{\eta} = (\eta_1, \cdots, \eta_m)$. Note that a τ_j indicates a point of slope changing of h where h is decreasing and an η_j a point of slope changing where h is increasing. Always, we let $0 < \tau_1 < \cdots < \tau_k \leq \eta_1 < \cdots < \eta_m$. Apparently, h is constant, being α, on $[\tau_k, \eta_1]$. Since h is fully defined by its $\boldsymbol{\pi}$ and $\boldsymbol{\theta}$, with k and m always implicitly assumed known, we treat h and $(\boldsymbol{\pi}, \boldsymbol{\theta})$ interchangeably below and hence may write the modified log-likelihood as $\tilde{\ell}(\boldsymbol{\pi}, \boldsymbol{\theta})$.

3.2.1 Characterization

For both exact and censored observations, the NPMLE $\hat{h}$ maximizes $\tilde{\ell}(h)$ among all h in the space of non-negative convex functions. Its characterization involves gradient functions of the log-likelihood. Let $e_{1,\tau} = (\tau - t)_+$, and $e_{2,\eta} = (t - \eta)_+$. The two gradient functions are defined as, respectively,

$$\begin{aligned}
d_1(\tau; h) &= \lim_{\epsilon \to 0} \frac{\tilde{\ell}(h + \epsilon e_{1,\tau}) - \tilde{\ell}(h)}{\epsilon} \\
&= \sum_{\substack{i=1 \\ T_i \neq T_{\max}}}^{n_1} \frac{(\tau - T_i)_+}{h(T_i)} - \sum_{i=1}^{n_1} \frac{\tau^2 - (\tau - T_i)_+^2}{2} \quad + \\
&\quad \sum_{i=n_1+1}^{n} \left\{ -\frac{\tau^2 - (\tau - L_i)_+^2}{2} - \frac{(\tau - R_i)_+^2 - (\tau - L_i)_+^2}{2} \, \Delta_i(H) \right\}, \\
&\quad \text{for } 0 < \tau \leq \ \eta_1,
\end{aligned} \tag{3.4}$$

and

$$\begin{aligned}
d_2(\eta; h) &= \lim_{\epsilon \to 0} \frac{\tilde{\ell}(h + \epsilon e_{2,\eta}) - \tilde{\ell}(h)}{\epsilon} \\
&= \sum_{\substack{i=1 \\ T_i \neq T_{\max}}}^{n_1} \frac{(T_i - \eta)_+}{h(T_i)} - \sum_{i=1}^{n_1} \frac{(T_i - \eta)_+^2}{2} \quad + \\
&\quad \sum_{i=n_1+1}^{n} \left\{ -\frac{(L_i - \eta)_+^2}{2} - \frac{(L_i - \eta)_+^2 - (R_i - \eta)_+^2}{2} \, \Delta_i(H) \right\}, \\
&\quad \text{for } \tau_k \ \leq \ \eta < T_{\max}.
\end{aligned} \tag{3.5}$$

Let us also set $e_{0,\alpha} = 1$. For the sake of complement, we can define

$$\begin{aligned}
d_0(\alpha; h) &= \lim_{\epsilon \to 0} \frac{\tilde{\ell}(h + \epsilon e_{0,\alpha}) - \tilde{\ell}(h)}{\epsilon} \\
&= \sum_{\substack{i=1 \\ T_i \neq T_{\max}}}^{n_1} \frac{1}{h(T_i)} - \sum_{i=1}^{n_1} T_i \quad + \\
&\quad \sum_{i=n_1+1}^{n} \left\{ -L_i - (L_i - R_i) \, \Delta_i(H) \right\}, \\
&\quad \text{for } \alpha \geq 0,
\end{aligned}$$

where $\Delta_i(H) = \exp(H(L_i) - H(R_i))/(1 - \exp(H(L_i) - H(R_i)))$. Note that d_1 and d_2 are piecewise quadratic functions of τ and η, respectively.

3.2.2 Theoretical Properties

Some of the characteristics of gradient functions and support point properties of the NPMLE of a convex hazard function are established for exact observations in this section. For interval-censored data or the case mixed with exact and interval-censored observations, these theoretical properties can be similarly derived, only by changing to more complicated notations and employing the same idea as for exact observations.

Given a hazard function h and an i.i.d. sample of exact observations $T_1, \ldots, T_n \in [0, \infty)$, the log-likelihood function is given by

$$\ell(h) = \sum_{i=1}^{n} \{\log h(T_i) - H(T_i)\}.$$

As mentioned in Section 2.5, since $h(T_{(n)})$ can be made arbitrarily large, Grenander (1956) considered that $\ell(h)$ can be maximised over a nondecreasing h which is bounded by some $M > 0$. Instead, one should maximize the modified log-likelihood $\tilde{\ell}(h)$

$$\tilde{\ell}(h) = \sum_{i \in \mathcal{I}} \log h(T_i) - \sum_{i=1}^{n} H(T_i),$$

where $\mathcal{I} = \{i : T_i \neq T_{(n)}\}$.

Let us denote by $\mathcal{K}$ the space of all convex hazard functions defined on $[0, \infty)$. Jankowski and Wellner (2009b) have shown that the MLE of a convex hazard function must be a piecewise linear function. Hence, we only need to consider the following hazard function

$$h(t) = \alpha + \sum_{j=1}^{k} \nu_j (\tau_j - t)_+ + \sum_{j=1}^{m} \mu_j (t - \eta_j)_+,$$

for $\alpha, \nu_j, \mu_j \geq 0$. The cumulative hazard function is then

$$H(t) = \alpha t + \frac{1}{2} \sum_{j=1}^{k} \nu_j \{\tau_j^2 - (\tau_j - t)_+^2\} + \frac{1}{2} \sum_{j=1}^{m} \mu_j (t - \eta_j)_+^2.$$

Here any piecewise linear $h \in \mathcal{K}$ is determined by three parameters (h_0, h_1, h_2): a nonnegative scalar $h_0 \equiv \alpha$ and two nonnegative measures $h_1 \equiv (\boldsymbol{\nu}, \boldsymbol{\tau})$ and $h_2 \equiv (\boldsymbol{\mu}, \boldsymbol{\eta})$, whose dimensions k and m need to be estimated as well.

Theorem 1. *$\tilde{\ell}$ is concave on $\mathcal{K}$ and any non-empty level set $\mathcal{K}(c) \equiv \{h \in \mathcal{K} : \tilde{\ell}(h) \geq c \geq -\infty\}$ is convex.*

Proof. For $h, g \in \mathcal{K}$ and $\epsilon \in (0, 1)$,

$$\frac{\partial \tilde{\ell}\{(1-\epsilon)h + \epsilon g\}}{\partial \epsilon} = \sum_{i \in \mathcal{I}} \frac{g(T_i) - h(T_i)}{(1-\epsilon)h(T_i) + \epsilon g(T_i)} + \sum_{i=1}^{n} \{G(T_i) - H(T_i)\},$$

where $G(t) = \int_0^t g(u)\,\mathrm{d}u$. Further,

$$\frac{\partial^2 \tilde{\ell}\{(1-\epsilon)h + \epsilon g\}}{\partial \epsilon^2} = -\sum_{i \in \mathcal{I}} \frac{\{g(T_i) - h(T_i)\}^2}{\{(1-\epsilon)h(T_i) + \epsilon g(T_i)\}^2},$$

which is negative, unless $h(T_i) = g(T_i)$ for all i. If $h, g \in \mathcal{K}(c)$, then

$$\tilde{\ell}\{(1-\epsilon)h + \epsilon g\} \geq \min\{\tilde{\ell}(h), \tilde{\ell}(g)\} \geq c,$$

meaning that $(1-\epsilon)h + \epsilon g \in \mathcal{K}(c)$, or $\mathcal{K}(c)$ is convex. ∎

Theorem 2. *For all $h \in \mathcal{K}(c)$, $c > -\infty$, $\max_{i \in \mathcal{I}}\{h(T_i)\}$ is bounded above.*

Proof. Denote by $T_{(i)}$ the i-th smallest value in the set of the unique values of T_i's, $i \in \mathcal{I}$. Assuming that h is increasing at $T_{(i)}$ and that $h(T_{(i)}) \to \infty$, we have

$$\lim_{h(T_{(i)}) \to \infty} \frac{h(T_{(i)})}{e^{H(T_{(i+1)})}} \leq \lim_{h(T_{(i)}) \to \infty} \frac{h(T_{(i)})}{e^{h(T_{(i)})(T_{(i+1)} - T_{(i)})}} = 0.$$

Similarly, we can obtain that $\frac{h(T_{(i)})}{e^{H(T_{(i)})}} \to 0$, if h is decreasing at $T_{(i)}$ and $h(T_{(i)}) \to \infty$. This means that if any $h(T_i) \to \infty$, $i \in \mathcal{I}$, then $\tilde{\ell}(h) \to -\infty < c$ and the limiting $h \notin \mathcal{K}(c)$. ∎

Corollary 1. *$\tilde{\ell}$ has a finite maximum on $\mathcal{K}$.*

Proof. This follows immediately from Theorems 1 and 2. ∎

Corollary 2. *For any $h \in \mathcal{K}(c)$, $c > -\infty$,*

$$|h| \equiv h_0 + \int \mathrm{d}h_1(\tau) + \int \mathrm{d}h_2(\eta)$$

is bounded above.

Proof. Follows easily from Theorem 2 and the convexity of h. ■

Letting $e_{1,\tau} = (\tau - t)_+$ and $e_{2,\eta} = (t - \eta)_+$, the two gradient functions are defined as, respectively,

$$\begin{aligned} d_1(\tau; h) &\equiv \left. \frac{\partial \tilde{\ell}(h + \epsilon e_{1,\tau})}{\partial \epsilon} \right|_{\epsilon=0} \\ &= \sum_{i \in \mathcal{I}} \frac{(\tau - T_i)_+}{h(T_i)} - \frac{1}{2} \sum_{i=1}^{n} \{\tau^2 - (\tau - T_i)_+^2\} \end{aligned}$$

and

$$\begin{aligned} d_2(\eta; h) &\equiv \left. \frac{\partial \tilde{\ell}(h + \epsilon e_{2,\eta})}{\partial \epsilon} \right|_{\epsilon=0} \\ &= \sum_{i \in \mathcal{I}} \frac{(T_i - \eta)_+}{h(T_i)} - \frac{1}{2} \sum_{i=1}^{n} (T_i - \eta)_+^2. \end{aligned}$$

For purposes of completeness, let us define

$$d_0(h) \equiv \left. \frac{\partial \tilde{\ell}(h + \epsilon e_0)}{\partial \epsilon} \right|_{\epsilon=0} = \frac{\partial \tilde{\ell}(h)}{\partial \alpha},$$

where $e_0 = 1$. The directional derivative from h to g is given by

$$\begin{aligned} d(g;h) &\equiv \left.\frac{\partial\tilde{\ell}\{(1-\epsilon)h+\epsilon g\}}{\partial\epsilon}\right|_{\epsilon=0} \\ &= \sum_{i\in\mathcal{I}} \frac{g(T_i)-h(T_i)}{h(T_i)} + \sum_{i=1}^{n}\{G(T_i)-H(T_i)\} \\ &= d_0(h)(g_0-h_0) + \int d_1(\tau;h)\,\mathrm{d}(g_1-h_1)(\tau) \\ &\quad + \int d_2(\eta;h)\,\mathrm{d}(g_2-h_2)(\eta). \end{aligned}$$

Lemma 1. *For any $g, h \in \mathcal{K}$,*

$$\tilde{\ell}(g) \le \tilde{\ell}(h) + d(g;h).$$

Proof. This follows easily from the concavity of $\tilde{\ell}$. ■

Theorem 3 (Characterization)**.** *$\widehat{h}$ maximizes $\tilde{\ell}(h)$ if and only if the following conditions are satisfied:*

(i) $d_0(\widehat{h}) \le 0$, *if* $\hat{\alpha} = 0$;

(ii) $d_0(\widehat{h}) = 0$, *if* $\hat{\alpha} > 0$;

(iii) $d_1(\tau;\widehat{h}) \le 0$, *for* $\tau \in [0, \hat{\eta}_1]$;

(iv) $d_1(\tau;\widehat{h}) = 0$, *for* $\tau \in \{\hat{\tau}_1, \ldots, \hat{\tau}_{\hat{k}}\}$;

(v) $d_2(\eta;\widehat{h}) \le 0$, *for* $\eta \in [\hat{\tau}_{\hat{k}}, T_{(n)}]$;

(vi) $d_2(\eta;\widehat{h}) = 0$, *for* $\eta \in \{\hat{\eta}_1, \ldots, \hat{\eta}_{\hat{m}}\}$.

Proof: Necessity can be established by contradiction. For any h that fails to satisfy any of the six conditions, $\tilde{\ell}$ must have an ascent direction at h and a sufficiently small movement in the direction will increase $\tilde{\ell}$. Hence this h does not maximize $\tilde{\ell}$.

If the six conditions are satisfied, then $d(g,\widehat{h}) \le 0$ for every $g \in \mathcal{K}$. Sufficiency follows from Lemma 1. ■

The first derivatives of the modified log-likelihood are

$$\frac{\partial\tilde{\ell}}{\partial\alpha} = d_0(h) = \sum_{i\in\mathcal{I}} \frac{1}{h(T_i)} - \sum_{i=1}^{n} T_i,$$
$$\frac{\partial\tilde{\ell}}{\partial\nu_j} = d_1(\tau_j;h) = \sum_{i\in\mathcal{I}} \frac{(\tau_j - T_i)_+}{h(T_i)} - \frac{1}{2}\sum_{i=1}^{n}\{\tau_j^2 - (\tau_j - T_i)_+^2\},$$
$$\frac{\partial\tilde{\ell}}{\partial\mu_j} = d_2(\eta_j;h) = \sum_{i\in\mathcal{I}} \frac{(T_i - \eta_j)_+}{h(T_i)} - \frac{1}{2}\sum_{i=1}^{n}(T_i - \eta_j)_+^2.$$

The second derivatives are

$$\frac{\partial^2\tilde{\ell}}{\partial\alpha^2} = -\sum_{i\in\mathcal{I}} \frac{1}{h(T_i)^2},$$
$$\frac{\partial^2\tilde{\ell}}{\partial\alpha\partial\nu_j} = -\sum_{i\in\mathcal{I}} \frac{(\tau_j - T_i)_+}{h(T_i)^2},$$
$$\frac{\partial^2\tilde{\ell}}{\partial\alpha\partial\eta_j} = -\sum_{i\in\mathcal{I}} \frac{(T_i - \eta_j)_+}{h(T_i)^2},$$
$$\frac{\partial^2\tilde{\ell}}{\partial\nu_j\partial\nu_k} = -\sum_{i\in\mathcal{I}} \frac{(\tau_j - T_i)_+(\tau_k - T_i)_+}{h(T_i)^2},$$
$$\frac{\partial^2\tilde{\ell}}{\partial\nu_j\partial\mu_k} = -\sum_{i\in\mathcal{I}} \frac{(\tau_j - T_i)_+(T_i - \eta_k)_+}{h(T_i)^2},$$
$$\frac{\partial^2\tilde{\ell}}{\partial\mu_j\partial\mu_k} = -\sum_{i\in\mathcal{I}} \frac{(T_i - \eta_j)_+(T_i - \eta_k)_+}{h(T_i)^2}.$$

Hence one can write

$$\frac{\partial\tilde{\ell}}{\partial\boldsymbol{\pi}} = \mathbf{S}^\top\mathbf{1} + \boldsymbol{\beta},$$
$$\frac{\partial^2\tilde{\ell}}{\partial\boldsymbol{\pi}\partial\boldsymbol{\pi}^\top} = -\mathbf{S}^\top\mathbf{S},$$

where $\boldsymbol{\pi} = (\alpha, \boldsymbol{\nu}^\top, \boldsymbol{\mu}^\top)^\top$, $\boldsymbol{\beta} = -\sum_{i=1}^{n}(T_i, \frac{1}{2}\{\tau_1^2 - (\tau_1 - T_i)_+^2\}, \ldots, \{\tau_k^2 - (\tau_k - T_i)_+^2\}, (T_i - \eta_1)_+^2, \ldots, (T_i - \eta_m)_+^2)^\top$, $\mathbf{s}_{(i)} = (1, (\tau_1 - T_i)_+, \ldots, (\tau_k - T_i)_+, (T_i - \eta_1)_+, \ldots, (T_i - \eta_m)_+)^\top / h(T_i)$, for $i \in \mathcal{I}$, and $\mathbf{S}^\top$ is the matrix with columns being $\mathbf{s}_{(i)}$.

3.3 Computation

In this section, we discuss some computational aspects of a convex hazard function estimation and then present an algorithm for computing the NPMLE. In Section 3.3.1, we will first present the computational complication for estimating a convex hazard function and then propose a novel idea for solving this problem. In the next two subsections, we describe how to update $\boldsymbol{\pi}'$ with $\boldsymbol{\theta}$ fixed and then how to expand and contract $\boldsymbol{\theta}$ appropriately. The algorithm is summarized in Section 3.3.6.

3.3.1 Main Idea

From a computational perspective, a main difficulty is how to tackle the problem of finding the minimum of a convex hazard function. A relevant work to our study is that of Jankowski and Wellner (2009a). The authors considered a two-loops profile likelihood method. In the inner loop, the support reduction algorithm is used to find the MLE over all convex hazard functions with a fixed minimum. In the outer loop, the bisection algorithm is utilized to optimize over all other possible minimum values of a convex hazard function. This approach contains some deficiencies such as including single support point in each iteration of the inner loop and using a time-consuming algorithm for finding the minimum of a convex hazard function estimate in the outer loop.

The key idea of our solution is to sustain a constant hazard part, even if its length is zero. Specifically, two cases of the constant hazard segment of, respectively, a positive and a zero length are treated in such a way that they are computationally convertible into each other. The resulting algorithm is remarkably faster and simpler than the profile likelihood method because of its reduction from the double loop to a single loop and the inclusion of multiple new support point to the support sets corresponding to the decreasing and increasing parts of a convex hazard function rather than the inclusion of only one new support point in each iteration.

3.3.2 Updating Masses

In order to update $\boldsymbol{\pi}$ to $\boldsymbol{\pi}'$ with $\boldsymbol{\theta}$ fixed, the second-order Taylor series expansion of the modified log-likelihood function in the neighbourhood of $\boldsymbol{\pi}$ is applied. Denote the gradient vector and Hessian matrix of the modified log-likelihood (3.2) (see Section 3.3.3) with respect to $\boldsymbol{\pi}$ by

$$\mathbf{g} \equiv \mathbf{g}(\boldsymbol{\pi}, \boldsymbol{\theta}) = \frac{\partial \tilde{\ell}}{\partial \boldsymbol{\pi}},$$
$$\mathbf{H} \equiv \mathbf{H}(\boldsymbol{\pi}, \boldsymbol{\theta}) = \frac{\partial^2 \tilde{\ell}}{\partial \boldsymbol{\pi} \partial \boldsymbol{\pi}^\top}.$$

The following quadratic approximation can be obtained by using the Taylor series expansion for $\tilde{\ell}(\boldsymbol{\pi}', \boldsymbol{\theta})$ around $\boldsymbol{\pi}$

$$\tilde{\ell}(\boldsymbol{\pi}', \boldsymbol{\theta}) - \tilde{\ell}(\boldsymbol{\pi}, \boldsymbol{\theta}) \approx \mathbf{g}^\top(\boldsymbol{\pi}' - \boldsymbol{\pi}) + \frac{1}{2}(\boldsymbol{\pi}' - \boldsymbol{\pi})^\top \mathbf{H}(\boldsymbol{\pi}' - \boldsymbol{\pi}).$$

Therefore, in a small neighbourhood of $\boldsymbol{\pi}$, the problem of maximizing $\tilde{\ell}(\boldsymbol{\pi}', \boldsymbol{\theta})$ can be approximated by the following least squares linear regression problem with non-negativity constraints:

$$\text{minimize} \quad \|\mathbf{R}\boldsymbol{\pi}' - (\mathbf{R}\boldsymbol{\pi} + \mathbf{b}_1)\|^2, \quad \text{subject to} \quad \boldsymbol{\pi}' \geq \mathbf{0}, \tag{3.6}$$

where $\mathbf{R} \equiv \mathbf{R}(\boldsymbol{\pi}, \boldsymbol{\theta})$ satisfies $\mathbf{H} = -\mathbf{R}^\top\mathbf{R}$ and $\mathbf{b}_1$ is the solution of $\mathbf{R}^\top\mathbf{b}_1 = \mathbf{g}$. In our implementation, $\mathbf{R}$ is the upper triangular matrix of the Cholesky decomposition of $-\mathbf{H}$ (see Section 3.3.4). Numerically, $\mathbf{R}$ may turn out to be singular when there exist very similar support points. To overcome this numerical difficulty, we add a very small positive value to each diagonal element of $\mathbf{R}$, e.g., $R_{ii} \times 10^{-10}$ to the ith diagonal element so the Cholesky factorization can be confidently utilized. To solve problem (3.6), the non-negativity least squares (NNLS) algorithm of Lawson and Hanson (1974) is applied.

A backtracking line search is conducted to ensure monotone increase of the log-likelihood. Let $\boldsymbol{\pi}'$ be the solution of problem (3.6). The updated vector $\boldsymbol{\pi} + \sigma^k(\boldsymbol{\pi}' - \boldsymbol{\pi})$ is chosen by using the smallest $k \in \{0, 1, 2, \cdots\}$ that satisfies the following inequality

$$\tilde{\ell}(\boldsymbol{\pi} + \sigma^k(\boldsymbol{\pi}' - \boldsymbol{\pi}), \boldsymbol{\theta}) \geq \tilde{\ell}(\boldsymbol{\pi}, \boldsymbol{\theta}) + \alpha\sigma^k\mathbf{g}^\top(\boldsymbol{\pi}' - \boldsymbol{\pi}), \quad 0 < \alpha < \frac{1}{2}. \tag{3.7}$$

We use $\sigma = \frac{1}{2}$, as for the popular step halving, and $\alpha = \frac{1}{3}$ in our implementation. For numerical reasons, similar support points are then collapsed, if after collapsing, the log-likelihood value either increases or does not decrease by more than a small threshold value. In the implementation, the threshold value used is 10^{-6}.

3.3.3 Derivatives of the Modified Log-likelihood Function

The first partial derivatives of the modified log-likelihood function (3.2) are derived as follows:

$$
\begin{aligned}
\frac{\partial \tilde{\ell}}{\partial \alpha} &= \sum_{\substack{i=1 \\ T_i \neq T_{\max}}}^{n_1} \frac{1}{h(T_i)} - \sum_{i=1}^{n_1} T_i + \sum_{i=n_1+1}^{n} \left\{ -L_i - (L_i - R_i)\Delta_i(H) \right\}, \\
\frac{\partial \tilde{\ell}}{\partial \nu_j} &= \sum_{\substack{i=1 \\ T_i \neq T_{\max}}}^{n_1} \frac{(\tau_j - T_i)_+}{h(T_i)} - \sum_{i=1}^{n_1} \frac{\tau_j^2 - (\tau_j - T_i)_+^2}{2} + \\
&\quad \sum_{i=n_1+1}^{n} \left\{ -\frac{\tau_j^2 - (\tau_j - L_i)_+^2}{2} - \frac{(\tau_j - R_i)_+^2 - (\tau_j - L_i)_+^2}{2} \Delta_i(H) \right\}, \\
\frac{\partial \tilde{\ell}}{\partial \mu_j} &= \sum_{\substack{i=1 \\ T_i \neq T_{\max}}}^{n_1} \frac{(T_i - \eta_j)_+}{h(T_i)} - \sum_{i=1}^{n_1} \frac{(T_i - \eta_j)_+^2}{2} + \\
&\quad \sum_{i=n_1+1}^{n} \left\{ -\frac{(L_i - \eta_j)_+^2}{2} - \frac{(L_i - \eta_j)_+^2 - (R_i - \eta_j)_+^2}{2} \Delta_i(H) \right\},
\end{aligned}
$$

where $\Delta_i(H) = \exp(H(L_i) - H(R_i))/(1 - \exp(H(L_i) - H(R_i)))$.

The second derivatives of the modified log-likelihood function that form the Hessian matrix $\mathbf{H}$ are given by

$$
\begin{aligned}
\frac{\partial^2 \tilde{\ell}}{\partial \alpha^2} &= \sum_{\substack{i=1 \\ T_i \neq T_{\max}}}^{n_1} \left\{ -\frac{1}{h(T_i)^2} \right\} + \sum_{i=n_1+1}^{n} \left\{ -(L_i - R_i)^2 \Delta_i(H) \right\}, \\
\frac{\partial^2 \tilde{\ell}}{\partial \mu_j^2} &= \sum_{\substack{i=1 \\ T_i \neq T_{\max}}}^{n_1} \left\{ -\frac{(\tau_j - T_i)_+^2}{h(T_i)^2} \right\} + \sum_{i=n_1+1}^{n} \left\{ -\left(\frac{(\tau_j - R_i)_+^2 - (\tau_j - L_i)_+^2}{2} \right)^2 \Delta_i(H) \right\}, \\
\frac{\partial^2 \tilde{\ell}}{\partial \nu_j^2} &= \sum_{\substack{i=1 \\ T_i \neq T_{\max}}}^{n_1} \left\{ -\frac{(T_i - \eta_j)_+^2}{h(T_i)^2} \right\} + \sum_{i=n_1+1}^{n} \left\{ -\left(\frac{(L_i - \eta_j)_+^2 - (R_i - \eta_j)_+^2}{2} \right)^2 \Delta_i(H) \right\}, \\
\frac{\partial^2 \tilde{\ell}}{\partial \nu_j \partial \alpha} &= \sum_{\substack{i=1 \\ T_i \neq T_{\max}}}^{n_1} \left\{ -\frac{(\tau_j - T_i)_+}{h(T_i)^2} \right\} + \\
&\quad \sum_{i=n_1+1}^{n} \left\{ -(L_i - R_i) \left(\frac{(\tau_j - R_i)_+^2 - (\tau_j - L_i)_+^2}{2} \right) \Delta_i(H) \right\},
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 \tilde{\ell}}{\partial \mu_j \partial \alpha} = & \sum_{\substack{i=1 \\ T_i \neq T_{\max}}}^{n_1} \left\{ -\frac{(T_i - \eta_j)_+}{h(T_i)^2} \right\} + \\
& \sum_{i=n_1+1}^{n} \left\{ -(L_i - R_i) \left(\frac{(L_i - \eta_j)_+^2 - (R_i - \eta_j)_+^2}{2} \right) \Delta_i(H) \right\}, \\
\frac{\partial^2 \tilde{\ell}}{\partial \mu_j \partial \nu_j} = & \sum_{\substack{i=1 \\ T_i \neq T_{\max}}}^{n_1} \left\{ -\frac{(T_i - \eta_j)_+ (\tau_j - T_i)_+}{h(T_i)^2} \right\} + \\
& \sum_{i=n_1+1}^{n} \left\{ -\left(\frac{(\tau_j - R_i)_+^2 - (\tau_j - L_i)_+^2}{2} \right) \left(\frac{(L_i - \eta_j)_+^2 - (R_i - \eta_j)_+^2}{2} \right) \Delta_i(H) \right\},
\end{aligned}
$$

where $\Delta_i(H) = \exp(H(L_i) - H(R_i))/(1 - \exp(H(L_i) - H(R_i))^2$.

3.3.4 QR Factorization of $\mathbf{D}$

The Hessian matrix $\mathbf{H}$ can be written as follows,

$$
\mathbf{H} = -\mathbf{D}^\top \mathbf{D}
$$

where

$$
\mathbf{D} = \begin{pmatrix}
\frac{1}{h(T_1)} & \frac{(\tau_1 - T_1)_+}{h(T_1)} & \cdots & \frac{(\tau_k - T_1)_+}{h(T_1)} & \frac{(T_1 - \eta_1)_+}{h(T_1)} & \cdots & \frac{(T_1 - \eta_m)_+}{h(T_1)} \\
\frac{1}{h(T_2)} & \frac{(\tau_1 - T_2)_+}{h(T_2)} & \cdots & \frac{(\tau_k - T_2)_+}{h(T_2)} & \frac{(T_2 - \eta_1)_+}{h(T_2)} & \cdots & \frac{(T_2 - \eta_m)_+}{h(T_2)} \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
\frac{1}{h(T_{n_1})} & \frac{(\tau_1 - T_{n_1})_+}{h(T_{n_1})} & \cdots & \frac{(\tau_k - T_{n_1})_+}{h(T_{n_1})} & \frac{(T_{n_1} - \eta_1)_+}{h(T_{n_1})} & \cdots & \frac{(T_{n_1} - \eta_m)_+}{h(T_{n_1})} \\
\alpha_{n_1+1} & \beta_{n_1+1} & \cdots & \gamma_{n_1+1} & \lambda_{n_1+1} & \cdots & \delta_{n_1+1} \\
\alpha_{n_1+2} & \beta_{n_1+2} & \cdots & \gamma_{n_1+2} & \lambda_{n_1+2} & \cdots & \delta_{n_1+2} \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
\alpha_n & \beta_n & \cdots & \gamma_n & \lambda_n & \cdots & \delta_n
\end{pmatrix}
$$

and

$$
\begin{aligned}
\alpha_i &= (L_i - R_i)\ \Delta_i(H), \\
\beta_i &= \frac{(\tau_1 - L_i)_+^2 - (\tau_1 - R_i)_+^2}{2}\ \Delta_i(H), \\
\gamma_i &= \frac{(\tau_k - L_i)_+^2 - (\tau_k - R_i)_+^2}{2}\ \Delta_i(H),
\end{aligned}
$$

$$\lambda_i = \frac{(L_i - \eta_1)_+^2 - (R_i - \eta_1)_+^2}{2} \Delta_i(H),$$
$$\delta_i = \frac{(L_i - \eta_k)_+^2 - (R_i - \eta_k)_+^2}{2} \Delta_i(H),$$
$$\Delta_i(H) = \frac{\sqrt{\exp(H(L_i) - H(R_i))}}{1 - \exp(H(L_i) - H(R_i))},$$

for $i = n_1 + 1, n_1 + 2, \cdots, n$.

Therefore, we have the following Taylor series expansion about $\boldsymbol{\pi}$ with $\boldsymbol{\theta}$ fixed:

$$\begin{aligned}\tilde{\ell}(\boldsymbol{\pi}') - \tilde{\ell}(\boldsymbol{\pi}) &\approx \mathbf{g}^\top(\boldsymbol{\pi}' - \boldsymbol{\pi}) + \frac{1}{2}(\boldsymbol{\pi}' - \boldsymbol{\pi})^\top \mathbf{H}(\boldsymbol{\pi}' - \boldsymbol{\pi}) \\ &= -\frac{1}{2}\|\mathbf{D}(\boldsymbol{\pi}' - \boldsymbol{\pi}) - \mathbf{d}\|^2 + c \\ &= -\frac{1}{2}\|\mathbf{R}(\boldsymbol{\pi}' - \boldsymbol{\pi}) - \mathbf{b}_1\|^2 - \frac{1}{2}\|\mathbf{b}_2\|^2 + c, \end{aligned} \tag{3.8}$$

where $\mathbf{D} = \mathbf{QR}$ by a QR decomposition and $\mathbf{d} = (\mathbf{R}^\top)^{-1}\mathbf{g}^\top$. In order to yield problem (3.6), we partition $\mathbf{Q} = \begin{pmatrix} \mathbf{Q_1} \\ \mathbf{Q_2} \end{pmatrix}$, where $\mathbf{Q_1} \in \mathbb{R}^{n\times(1+k+m)}$. Hence, $\mathbf{Q}_1^\top \mathbf{D} = \begin{pmatrix} \mathbf{R} \\ \mathbf{0} \end{pmatrix}$, where $\mathbf{R} \in \mathbb{R}^{(1+k+m)\times(1+k+m)}$, and $\mathbf{Q}^\top \mathbf{b} = \begin{pmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{Q}_1{}^\top \mathbf{d} \\ \mathbf{Q}_2{}^\top \mathbf{d} \end{pmatrix}$. The upper triangular matrix $\mathbf{R}$ obtained this way is mathematically the same as by the Cholesky decomposition of $-\mathbf{H}$, but numerically more stable.

3.3.5 Expanding and Contracting Support Sets

One of the critical components of our new algorithm is enlarging and shrinking the two support sets $\boldsymbol{\tau}$ and $\boldsymbol{\eta}$. The gradient functions $d_1(\tau; h)$ and $d_2(\eta; h)$ are applied as a great tool to locate new elements for the support sets. As $d_1(\tau; h)$ and $d_2(\eta; h)$ are piecewise quadratic functions of τ and η, it is simple to locate all global maxima between every two neighbouring support points of these functions. A detailed description of the support sets expansion technique is presented below.

The procedure of enlarging the two support sets corresponding to the decreasing and increasing parts of a convex bathtub-shaped hazard function needs to be carried out on three intervals: (i) the decreasing part on $[0, \tau_k)$, (ii) the constant part on $[\tau_k, \eta_1]$; and (iii) the increasing part on $(\eta_1, T_{\max})$.

Let us define an ordered set $J = \{0, \tau_1, \cdots, \tau_k, \eta_1, \cdots, \eta_m, T_{\max}\}$. In the first step of the new algorithm, we expand the two support sets corresponding to the decreasing and increasing parts of a convex hazard function by finding and adding one new support point between every two neighbouring support points in set J. For the decreasing part $[0, \tau_k)$, we

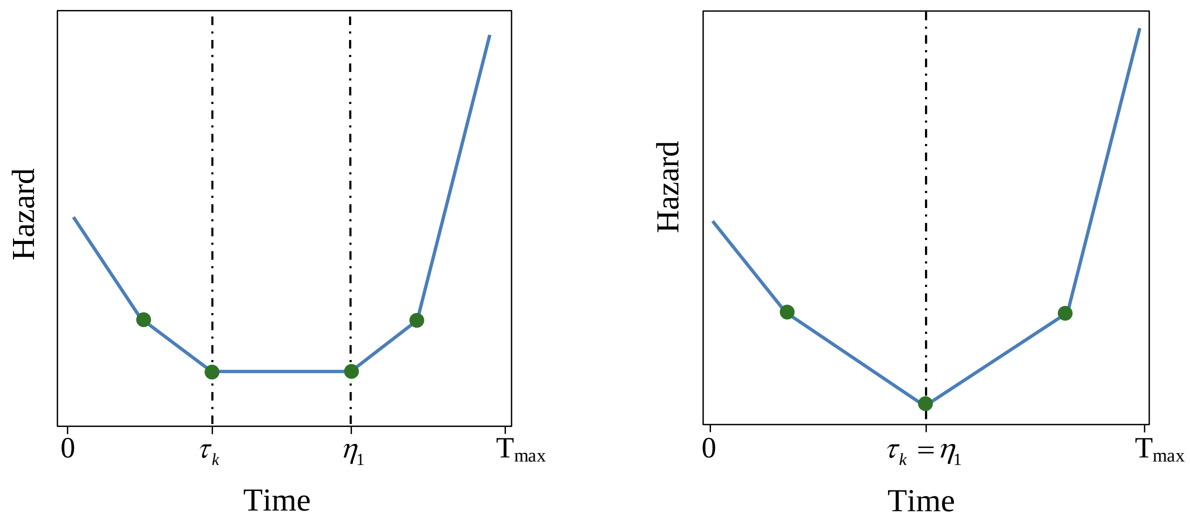

Figure 3.1: Schematic representation of a convex bathtub-shaped hazard function for the constant part.

use the gradient function $d_1(\tau; h)$ (3.4). For the increasing segment $[\eta_1, T_{\max})$, the gradient function $d_2(\eta; h)$ (3.5) is applied. For the constant hazard interval $[\tau_k, \eta_1]$, we employ the two gradient functions $d_1(\tau; h)$ and $d_2(\eta; h)$ when $\tau_k \neq \eta_1$. This means that the greatest value of d_1 and d_2 on $[\tau_k, \eta_1]$ corresponds to the new support point added, to either set $\boldsymbol{\tau}$ or set $\boldsymbol{\eta}$, depending on either d_1 or d_2 is greater at the point.

In the second step of the new algorithm, we update the mass vectors $\boldsymbol{\nu}$ and $\boldsymbol{\mu}$ by solving problem (3.6). Then, we contract the support sets $\boldsymbol{\tau}$ and $\boldsymbol{\eta}$, by removing the elements of $\boldsymbol{\tau}$ and $\boldsymbol{\eta}$ that have zero masses. The length of the constant part can thus increase or decrease after each iteration. This includes the situation when a positive length becomes zero, and the opposite situation. This is because, e.g., if the updated mass of η_1 becomes zero, the length of the constant part increases, while if there is a new support point between $[\tau_k, \eta_1]$ which recieves a positive mass, then the length of constant segment decreases. With following these two steps, this process is repeated until the final solution is found, which must be the global maximum $\hat{h}$.

By implementing this new idea, the double loop needed by Jankowski and Wellner (2009a) which finds the minimum of the convex hazard function by the bisection algorithm method while computing the MLE of the hazard function with the minimum held fixed by the SR algorithm, is reduced to a single loop. Thus, the convergence of our algorithm is much faster and resulting in a remarkable reduction in computation time.

3.3.6 The Algorithm

In this section, we present our new algorithm for computing the nonparametric estimate of a convex hazard function. It is a generalization of the CNM algorithm of Wang (2007, 2008) for computing the NPMLE of a mixing distribution. However, the new one does not need to comply with the restriction that the sum of masses must be one. Moreover, two gradient functions $d_1(\tau; h)$ and $d_2(\eta; h)$ instead of one are employed through computational steps of the new algorithm.

Let us denote $\boldsymbol{\theta}^+$ the extended support point vector by finding and including some new support points, and $\boldsymbol{\pi}^+$ the corresponding mass vector enlarged on $\boldsymbol{\pi}$ by adding 0s for new support points. Let $\boldsymbol{\pi}'$ and $\boldsymbol{\theta}'$ be the updated mass vector and support point vector by discarding the redundant points from $\boldsymbol{\theta}^+$ with zero masses in $\boldsymbol{\pi}^+$. The new algorithm, which we call CNMCH as an extension of CNM for computing a convex hazard function, is summarized as follows.

Algorithm 3.1 (CNMCH)**.** Choose a small $\gamma > 0$ and set $s = 0$. From an initial estimate h_0 with finite support and $\ell(h_0) > -\infty$, repeat the following steps.

Step 1: compute all global maxima of gradients between every two adjacent points in set J, as described above, hence giving $\tau_{s1}^*, \cdots, \tau_{sk}^*$ for the decreasing part, $\eta_{s1}^*, \cdots, \eta_{sm}^*$ for the increasing part, and ρ_s^* for the constant part.

Step 2: set $\boldsymbol{\tau}_s^* = (\tau_{s1}^*, \cdots, \tau_{sk}^*)^\top$, $\boldsymbol{\eta}_s^* = (\eta_{s1}^*, \cdots, \eta_{sm}^*)^\top$, $\boldsymbol{\theta}_{s+} = (\boldsymbol{\tau}_s^\top, \boldsymbol{\tau}_s^{*\top}, \rho_s^*, \boldsymbol{\eta}_s^\top, \boldsymbol{\eta}_s^{*\top})^\top$ and $\boldsymbol{\pi}_s^+ = (\alpha_s, \boldsymbol{\nu}_s^\top, \mathbf{0}^\top, 0, \boldsymbol{\mu}_s^\top, \mathbf{0}^\top)^\top$. Find $\boldsymbol{\pi}_{s+1}^-$ by solving problem (3.6), with $\mathbf{R}$ replaced by $\mathbf{R}_s^+ = \mathbf{R}(\boldsymbol{\pi}_s^+, \boldsymbol{\theta}_s^+)$ and $\mathbf{b}_{1s}^+ = (\boldsymbol{\pi}_s^{+\top}, \boldsymbol{\theta}_s^{+\top})^\top$, followed by a backtracking line search.

Step 3: discard all support points with zero masses in $\boldsymbol{\pi}_{s+1}^-$, which gives h_{s+1} with $\boldsymbol{\theta}_{s+1}$ and $\boldsymbol{\pi}_{s+1}$. If $|\tilde{\ell}(h_{s+1}) - \tilde{\ell}(h_s)| \le \gamma$ stop; otherwise, set $s = s + 1$.

3.4 Convergence

The theoretical justification for the convergence of the CNMCH algorithm for computing the NPMLE of a convex hazard function in the case of exact observations is established in this section. In fact, the proof below also holds true for the other similar situations, e.g., the case of interval-censored data and the case mixed with exact and interval-censored observations. Let $\mathcal{K}_0 \equiv \mathcal{K}(c_0)$ and $\mathbf{b} = \mathbf{R}\boldsymbol{\pi} + \mathbf{b}_1$, refer to (3.8).

Lemma 2. *Let $\boldsymbol{\pi}'$ be obtained by minimizing $\|\mathbf{S}\boldsymbol{\pi}' - \mathbf{b}\|$ in the algorithm. Then always*

$$\|\mathbf{S}\boldsymbol{\delta}\| \leq u,$$

for some $u < \infty$ independent of s.

Proof. Since $\min_{\boldsymbol{\pi}' \geq 0} \|\mathbf{S}\boldsymbol{\pi}' - \mathbf{b}\|$, it holds that $\|\mathbf{S}\boldsymbol{\pi}' - \mathbf{b}\| \leq \|\mathbf{b}\|$ and hence $\|\mathbf{S}\boldsymbol{\pi}'\| \leq \|\mathbf{S}\boldsymbol{\pi}' - \mathbf{b}\| + \|\mathbf{b}\| \leq 2\|\mathbf{b}\|$. Therefore, $\|\mathbf{S}\boldsymbol{\delta}\| \leq 2\|\mathbf{b}\| + \sqrt{n}$. Because $\|\mathbf{b}\|$ only depends on $h(T_i)$, for $i \in \mathcal{I}$, which is bounded away from zero for all $h \in \mathcal{K}_0$, the proof is completed. ■

Lemma 3. *The Armijo search used in the algorithm always succeeds in a finite number of steps independent of s.*

Proof. Since $\boldsymbol{\delta} \equiv \boldsymbol{\pi}' - \boldsymbol{\pi}$ maximizes

$$(\mathbf{S}^{+\top}\mathbf{1} + \boldsymbol{\beta})^\top \boldsymbol{\delta} - \frac{1}{2}\boldsymbol{\delta}^\top \mathbf{S}^{+\top}\mathbf{S}\boldsymbol{\delta}$$

under restriction $\boldsymbol{\pi}' \geq 0$, we have

$$\boldsymbol{\delta}^\top \mathbf{S}^{+\top}\mathbf{S}^{+}\boldsymbol{\delta} \leq (\mathbf{S}^{+\top}\mathbf{1} + \boldsymbol{\beta})^\top \boldsymbol{\delta}.$$

Noting the Taylor series expansion

$$\begin{aligned}
&\tilde{\ell}(\boldsymbol{\pi} + \boldsymbol{\delta}, \boldsymbol{\theta}) - \tilde{\ell}(\boldsymbol{\pi}, \boldsymbol{\theta}) \\
&= (\mathbf{S}^{+\top}\mathbf{1} + \boldsymbol{\beta})^\top \boldsymbol{\delta} - \frac{1}{2}\boldsymbol{\delta}^\top \mathbf{S}^{+\top}\mathbf{S}\boldsymbol{\delta} + o(\|\mathbf{S}\boldsymbol{\delta}\|^2),
\end{aligned}$$

for any $0 < \alpha < \frac{1}{2}$, there is a $\lambda > 0$ such that if $\|\mathbf{S}\boldsymbol{\delta}\| \leq \lambda$, then

$$\tilde{\ell}(\boldsymbol{\pi} + \boldsymbol{\delta}, \boldsymbol{\theta}) - \tilde{\ell}(\boldsymbol{\pi}, \boldsymbol{\theta}) \geq \alpha(\mathbf{S}^{+\top}\mathbf{1} + \boldsymbol{\beta})^\top \boldsymbol{\delta},$$

thus satisfying the Armijo rule.

If $\|\mathbf{S}\boldsymbol{\delta}\| > \lambda$, then $\|\sigma^k \mathbf{S}\boldsymbol{\delta}\| \le \lambda$ holds for some $k > 0$. Because $\|\mathbf{S}\boldsymbol{\delta}\| \le u$ from Lemma 2, we need at most

$$\bar{k} \equiv \max\left\{\left\lceil \log_\sigma \left(\frac{\lambda}{u}\right)\right\rceil, 0\right\}$$

steps for the Armijo rule to be satisfied in all situations. ∎

Theorem 4. *Let $\{h_s\}$ be any sequence created by the algorithm. Then $\tilde{\ell}(h_s) \to \tilde{\ell}(\widehat{h})$ monotonically as $s \to \infty$.*

Proof. Owing to its monotone increase, $\tilde{\ell}(h_s)$ will converge to a finite value no greater than $\tilde{\ell}(\widehat{h})$. Further,

$$\begin{aligned}
\tilde{\ell}(h_{s+1}) - \tilde{\ell}(h_s) &\ge \alpha\sigma^{\bar{k}}(\mathbf{S}_s^{+\top}\mathbf{1} + \boldsymbol{\beta}_s)^\top \boldsymbol{\delta}_s \\
&\ge \alpha\sigma^{\bar{k}}\left\{(\mathbf{S}_s^{+\top}\mathbf{1} + \boldsymbol{\beta}_s)^\top \boldsymbol{\delta}_s - \frac{1}{2}\boldsymbol{\delta}_s^\top \mathbf{S}_s^{+\top}\mathbf{S}_s^{+}\boldsymbol{\delta}_s\right\},
\end{aligned}$$

because of Armijo's rule and the non-negative definiteness of $\mathbf{S}_s^{+\top}\mathbf{S}_s^{+}$.

Consider all point-mass directions $e \in \{\pm e_0, \pm e_{1,\tau}, \pm e_{2,\eta}\}$ from h_s, that are valid in the meaning that there exists an $\epsilon > 0$ such that $h_s + \epsilon e \in \mathcal{K}$. Denote the steepest ascent direction by

$$e_s^* = \arg\max_e d(h_s + e; h_s)$$

and $\boldsymbol{\delta}_s^*$ the direction resulting from h_s to $h_s + e_s^*$. Hence, from any $\epsilon \in \mathbb{R}$ such that $h_s + \epsilon e_s^* \in \mathcal{K}$, we have

$$\tilde{\ell}(h_{s+1}) - \tilde{\ell}(h_s) \ge \alpha\sigma^{\bar{k}}\left\{\epsilon(\mathbf{S}_s^{+\top}\mathbf{1} + \boldsymbol{\beta}_s)^\top \boldsymbol{\delta}_s^* - \frac{\epsilon^2}{2}\boldsymbol{\delta}_s^{*\top}\mathbf{S}_s^{+\top}\mathbf{S}_s^{+}\boldsymbol{\delta}_s^*\right\},$$

because of the optimality of $\boldsymbol{\delta}_s$.

Now, let us assume that $d(h_s + e_s^*; h_s)$ does not approach 0 as $s \to \infty$. There are, hence, infinitely many s such that $d(h_s + e_s^*; h_s) \geq \tau$, for some $\tau > 0$. For such an s and noting that

$$d(h_s + e_s^*; h_s) = (\mathbf{S}_s^{+\top}\mathbf{1} + \boldsymbol{\beta}_s)^\top \boldsymbol{\delta}_s^*,$$

we have, with Lemma 2,

$$\tilde{\ell}(h_{s+1}) - \tilde{\ell}(h_s) \geq \alpha\sigma^{\bar{k}} \left\{ \epsilon\tau - \frac{\epsilon^2 u^2}{2} \right\}.$$

Without loss of generality, assume $\tau \leq u^2$ and let $\epsilon = \tau/u^2$. As a result,

$$\tilde{\ell}(h_{s+1}) - \tilde{\ell}(h_s) \geq \frac{\alpha\sigma^{\bar{k}}\tau^2}{2u^2},$$

a positive value that is independent of s. Since this violates the Cauchy property of a convergent sequence, we must have $d(h_s + e_s^*; h_s) \to 0$ as $s \to \infty$. Therefore, $d(\widehat{h}; h_s) \leq d(h_s + e_s^*; h_s)(|h_s|+|\widehat{h}|) \to 0$ from Corollary 2, and $\tilde{\ell}(h_s) \to \tilde{\ell}(\widehat{h})$ from Lemma 1. ■

3.5 Real Data Examples

This section provides the results of our numerical studies of three real-world data sets. In the case of all exact observations, we investigated and compared the performance of our proposed CNMCH algorithm with the SRB algorithm to the same problem in the first and second real data examples. In particular, we focused on the computational aspects of these algorithms. In the situation of interval-censored data in the third example, we are interested in comparing the estimates of the hazard function provided by the CNMCH algorithm for computing a nonparametric estimation of a hazard function under convexity shape restriction with the constrained Newton method (CNM) (Wang, 2008) without using any shape constraint.

In survival analysis, it is quite common for real data sets to consist of tied event times, where a tie means that two or more subjects in the data set share the same time. For computational efficiency, tied observations should better be grouped. Hence, we implemented two different versions of our algorithm CNMCH, namely ties grouped and ungrouped. The SRB algorithm is available in the R package **convexHaz** (Jankowski et al., 2009) and unfortunately does not handle this issue. As a result, it also fails to deal with duplicated largest observations. Instead of removing all the greatest duplicated observations for maximising the modified log-likelihood, it only takes out one, which leads to an incorrect likelihood maximization. Two different implementations of the SRB algorithm, gridded and gridless, for finding the location of the minimum of the directional derivatives were used. In their implementation, $[0, T_{(n)}]$ was split into M intervals. We compare the performance of the algorithms based on the number of iterations and running times. The number of iterations in the SRB algorithm is attributed to the outer loop of an algorithm.

In our studies, the CNMCH algorithm was terminated when the following condition was satisfied:

$$|\tilde{\ell}(h_s) - \tilde{\ell}(h_{s-1})| \leq \gamma,$$

for $\gamma = 10^{-6}$. All algorithms are performed in R (R Development Core Team, 2013), except for the NNLS algorithm which is in Fortran and embedded in CNM and CNMCH for solving a least squares problem with non-negativity constraint.

3.5.1 Canadian Human Mortality Data

We considered the Canadian human mortality table for the year 2008 (`http://www.mortality.org/`). A random sample of size 1000 was extracted from the entire life table with sample size 238612, so that a comparison with the observed hazard would be possible. The histogram for the entire life table and a random sample of size 1000 are shown in Figure 3.2. The observed hazard is computed by the number of people who died at time t over the number of people who were still alive at time t. The result of the NPMLE for lifetimes is shown with a dashed line and the observed hazard with a solid line in the left panel of Figure 3.3. The Canadian mortality data reveals the hazard rate is high in the infant mortality phase, but decreases rapidly. Subsequently, the hazard rate stabilizes before the age of 70, followed by an increasing hazard rate due to the natural aging process. As a desirable attribute, the estimated turning point of the onset of late life observation remarkably corresponds to the observed hazard point. Our proposed approach has a good performance after age 80. However, the subplot in the left panel of Figure 3.3 demonstrates that the estimated hazard function at the early infant mortality phase deviates from the observed hazard function. To overcome the raised problem, one can employ the higher order smoothness of an estimator, which will be studied in Chapter 4. For the complete mortality table, the resulting fitted hazard functions for the total, females and males are also presented in the right panel of

Figure 3.3. Obviously, the death rates of males were higher than those of females in 2008 at any specified age.

Table 3.1 provides a brief summary of the estimation results for a random sample and the entire life table. As mentioned above, we are unable to execute the SRB algorithm for the entire life table since it does not tackle the problem of duplicated observations. In terms of the modified log-likelihood, the CNMCH algorithm has an overall better performance than the SRB algorithm for the random sample drawn from the entire life table. By looking at the computation time, one can easily see that the SRB algorithm is far more expensive than the CNMCH algorithm.

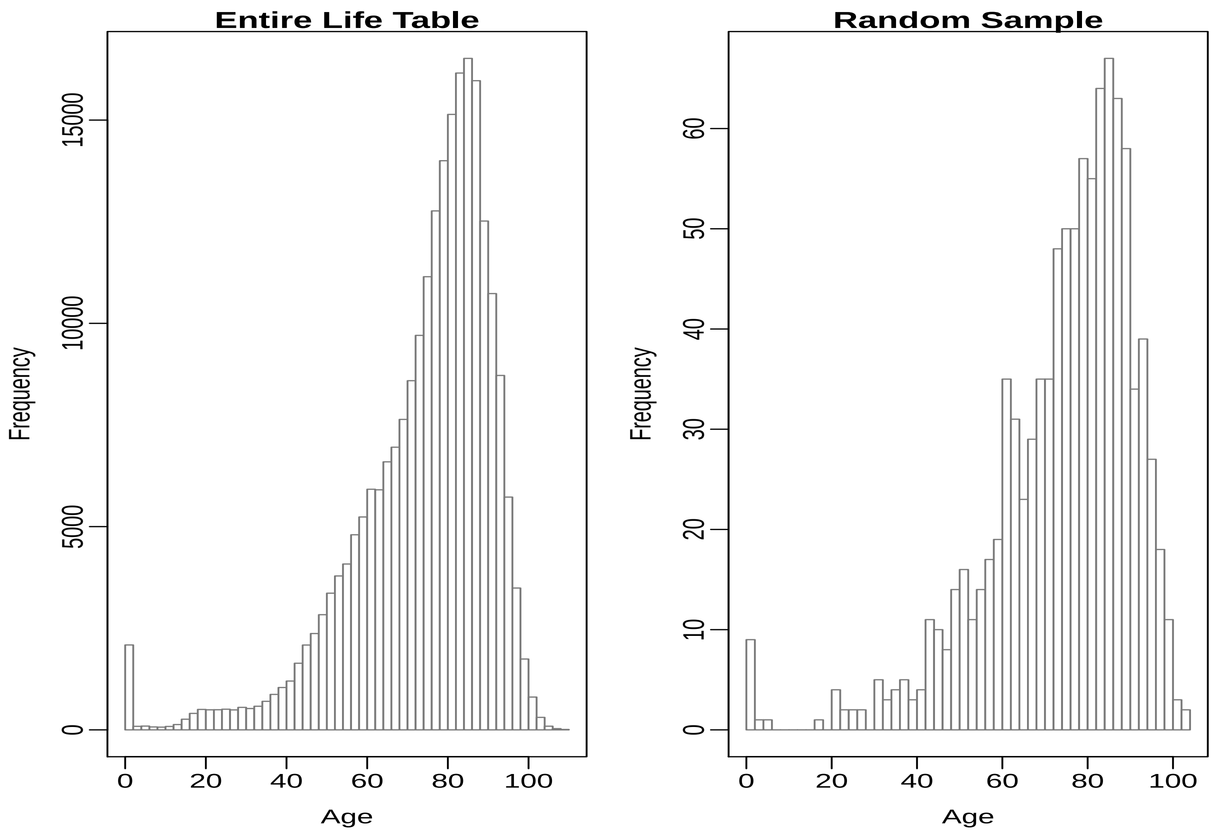

Figure 3.2: Histograms of the entire and a random sample of Canadian life table for year 2008.

3.5.2 Air Conditioner Failure Data

The second real data set, studied by Proschan (1963) and Jankowski and Wellner (2009a), concerns the successive failures of the air conditioning system from Boeing airplanes and the interest is in the estimation of hazard rates of the intervals between successive failures. Table 3.2 gives the 213 observations which are the number of operating hours between successive failures of air conditioning equipment in 13 Boeing 720 jet aircrafts. Figure 3.4 provides a couple of plots depicting the hazard function and the gradient curve at the initial, fourth, and tenth iteration, respectively, and finally at convergence. Computationally, the conversion of

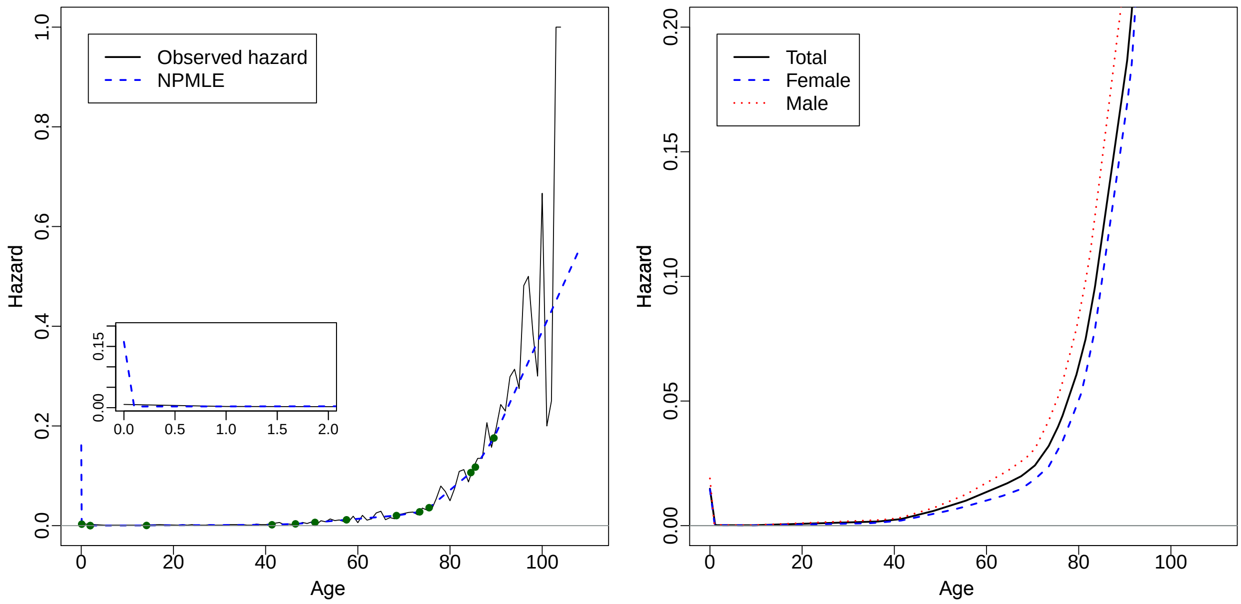

Figure 3.3: NPMLE of the CNMCH algorithm for the Canadian lifetime data in 2008: the observed hazard rates (solid line) compared with the NPMLE of the hazard function (dashed line) based on the generated sample size of 1000 from the entire life table (left); the fitted hazard rates of the total (solid line), males (dotted line) and females (dashed line) for the complete life table (right).

Table 3.1: Results of the algorithms for the Canadian lifetime data in 2008.

Method	#Iter.	Time (s)	Modified log-likelihood
Random sample			
CNMCH - ties grouped	18	0.17	-4043.84
CNMCH - ties ungrouped	18	0.78	-4043.84
SRB (M = 100, GRIDLESS = 0)	5	233.02	-4068.83
SRB (M = 1000, GRIDLESS = 0)	9	1552.02	-4044.49
SRB (M = 100, GRIDLESS = 1)	8	391.14	-4096.98
SRB (M = 1000, GRIDLESS = 1)	11	1218.45	-4096.97
Entire life table			
CNMCH - ties grouped	21	0.28	-963700.10
CNMCH - ties ungrouped	22	307.02	-963700.10

a positive-lengthed constant part of the convex hazard function to a zero-lengthed one can be clearly observed in Figure 3.4. The intersection points between the last gradient function and x-axis are the support points of the NPMLE.

The performance of the CNMCH and SRB algorithms is given in Table 3.3. We executed the SRB algorithm for both the grid and gridless methods with different numbers of grid points. It can be seen from the results that the CNMCH gave a slightly larger likelihood value and was also significantly faster than the SRB algorithm, especially in the case of treating data set by grouping tied observations.

For this data set, the piecewise linear hazard function is given by

$$\hat{h}(t) = 0.00657 + 0.00003\,(136.2989 - t)_+ + 0.00001\,(376.5684 - t)_+ + \\ 0.00003\,(t - 376.5735)_+ + 0.00001\,(t - 396.1332)_+.$$

Table 3.2: Air conditioner failure data of Boeing airplanes.

Plane	Intervals between failures													
7907	194	15	41	29	33	181								
7908	413	14	58	37	100	65	9	169	447	184	36	201	118	34	31
	18	18	67	57	62	7	22	34							
7909	90	10	60	186	61	49	14	24	56	20	79	84	44	59	29
	118	25	156	310	76	26	44	23	62	130	208	70	101	208	
7910	74	57	48	29	502	12	70	21	29	386	59	27	153	26	326
7911	55	320	56	104	220	239	47	246	176	182	33	15	104	35	
7912	23	261	87	7	120	14	62	47	225	71	246	21	42	20	5
	12	120	11	3	14	71	11	14	11	16	90	1	16	52	95
7913	97	51	11	4	141	18	142	68	77	80	1	16	106	206	82
	54	31	216	7	46	111	39	63	18	191	18	163	24		
7914	50	44	102	72	22	39	3	15	197	188	79	88	46	5	5
	36	22	139	210	97	30	23	13	14						
7915	359	9	12	270	603	3	104	2	438						
7916	50	254	5	283	35	12									
7917	130	493													
8044	487	18	100	7	98	5	85	91	43	230	3	130			
8045	102	209	14	57	54	32	67	59	134	152	27	14	230	66	61
	34														

3.5.3 Angina Pectoris Survival Data

The third real-world data is the life table data set that has been studied by Lee and Wang (2003). It is used here to particularly illustrate our motivation in the case of purely interval-censored data. The survival data for 2418 males patients with angina pectoris is tabulated in Table 3.5. Survival times are recorded as years from the time of diagnosis and collected in each one year time interval for 16 intervals. Two cases were investigated: (i) exact data, where the CNMCH algorithm applied to the midpoints of the time intervals for only those who died during the study, (ii) purely interval-censored data, where we also included those

Table 3.3: Performance of the algorithms on the air conditioner failure data.

Method	#Iter.	Time (s)	Modified log-likelihood
CNMCH - ties grouped	15	0.13	-1169.983165
CNMCH - ties ungrouped	15	0.15	-1169.983165
SRB (M = 100, GRIDLESS = 0)	7	34.38	-1169.989235
SRB (M = 1000, GRIDLESS = 0)	7	122.73	-1169.983180
SRB (M = 100, GRIDLESS = 1)	7	222.50	-1169.983472
SRB (M = 1000, GRIDLESS = 1)	7	440.96	-1169.983177

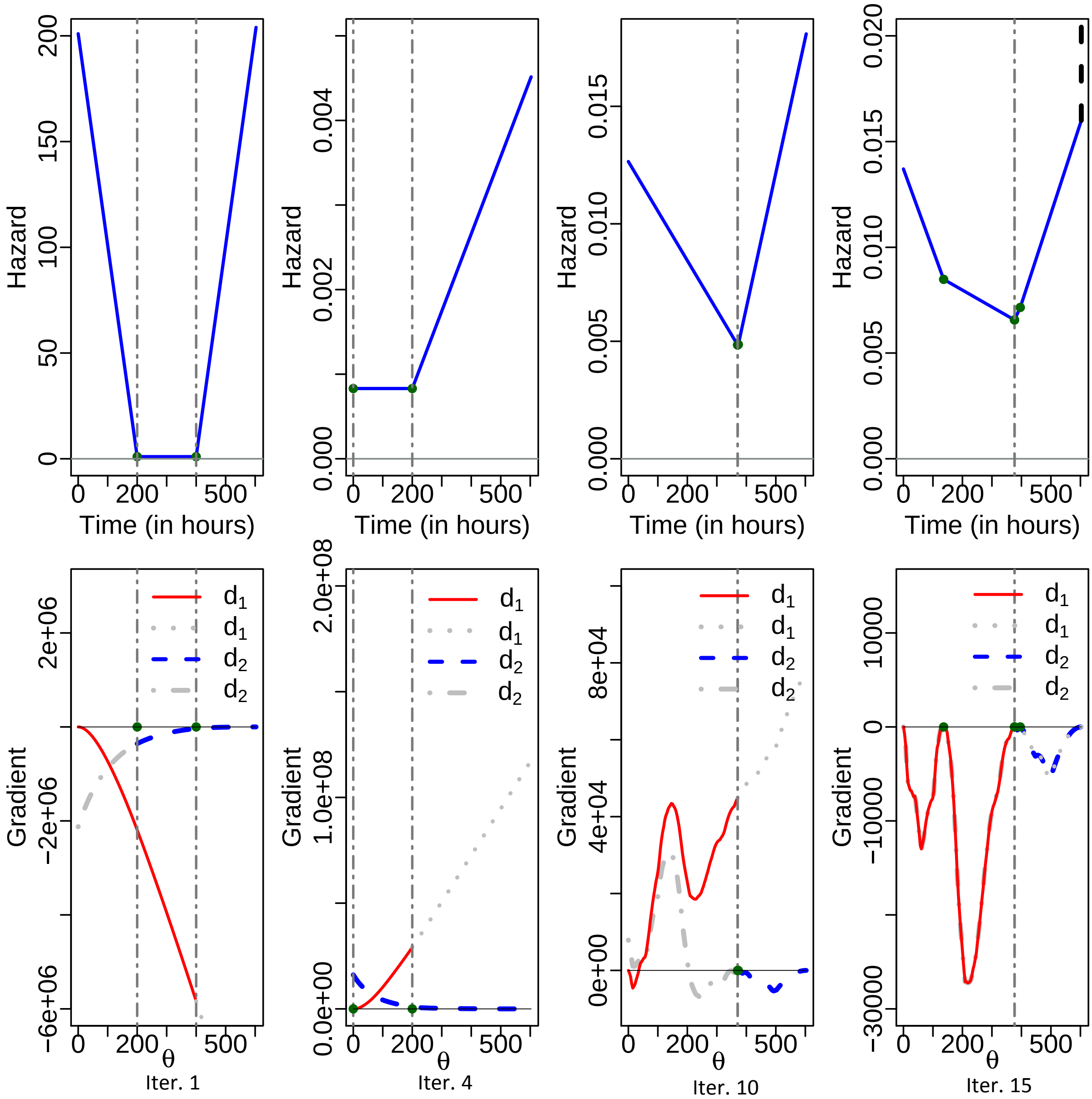

Figure 3.4: Column-wise plots correspond to the initial, fourth, and tenth iteration, and the NPMLE of the hazard function and the gradient curve.

who are lost to follow-up during the study. We are interested in comparing the estimates of these two cases.

Table 3.4: Life table of 2418 males with angina pectoris.

Years after Diagnosis	Number Died	Number Lost to Follow-up
0	456	0
1	226	39
2	152	22
3	171	23
4	135	24
5	125	107
6	83	133
7	74	102
8	51	68
9	42	64
10	43	45
11	34	53
12	18	33
13	9	27
14	6	23
15	0	30

In the case of all exact observations, we just carried out the CNMCH algorithm since the R package **convexHaz** can not cope with the situation of repeated largest observations. The top row of Figure 3.5 shows the NPMLE of the hazard function and the gradient curve at the NPMLE, with solid points as the support points on the x-axis.

The dimension-reduced constrained Newton method (Wang, 2008) and the proposed CNMCH algorithm were used in the case of purely interval-censored data. The hazard function at two NPMLE's by the CNMCH and CNM algorithms are plotted in the bottom left panel of Figure 3.5. The plots of the estimated hazard functions illustrate that the death rate is the highest immediately after diagnosis. The hazard rate stays approximately constant between the end of the first year and the beginning of the tenth year. After the tenth year, the hazard rate gradually increases. It is intuitively obvious that the hazard function estimate based on the convex shape restriction captures the gradual change of hazard over time far better than its discrete competitor without shape constraint. The bottom right panel of Figure 3.5 shows the gradient curve at the NPMLE by the CNMCH algorithm.

The performance of the algorithms for males with angina pectoris for both the case of exact data (in the upper panel) and the case of purely interval-censored data (in the lower panel) is displayed in Table 3.5. In both cases, the CNMCH algorithm with ties grouped took a far shorter computation time than the version without grouping ties.

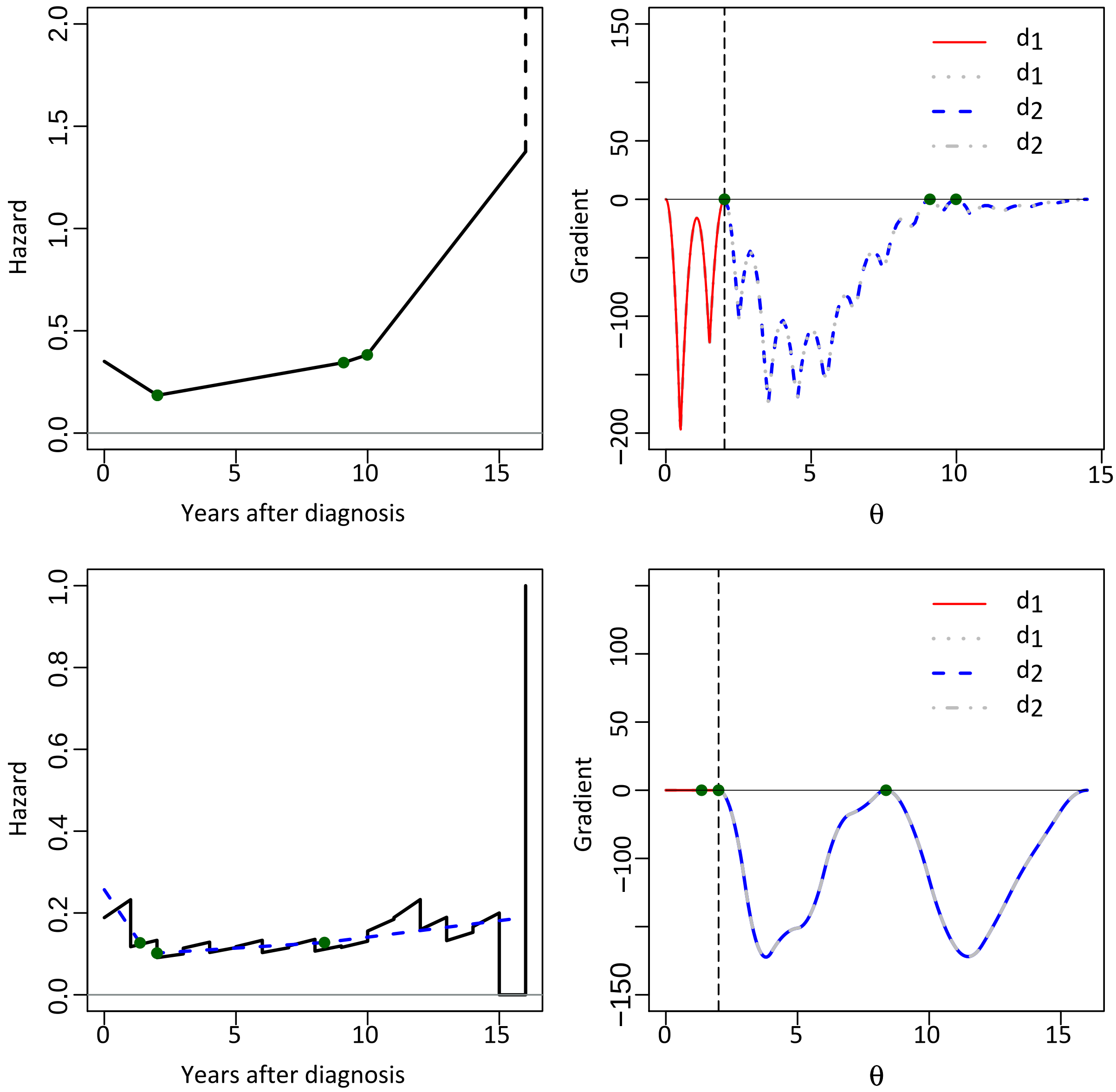

Figure 3.5: NPMLE for the life table of males with angina pectoris: the fitted hazard rates for the exact observations by the CNMCH algorithm (top left); the gradient at the NPMLE for the exact observations (top right); the fitted hazard function for the purely-interval-censored data by the CNM algorithm (solid line) and the CNMCH algorithm (dashed line) (bottom left); the gradient at the NPMLE for purely interval-censored data of the CNMCH algorithm (bottom right). The solid dots represents the support points at the NPMLE.

The estimated hazard function in the case of exact and purely interval-censored data are given respectively as follows:

$$\begin{aligned}\hat{h}(t) = {} & 0.185 + 0.082\,(2.017 - t)_+ + 0.023\,(t - 2.017)_+ + 0.020\,(t - 9.087)_+ + \\ & 0.123\,(t - 9.985)_+,\end{aligned}$$

and

$$\hat{h}(t) = 0.102 + 0.069\,(1.615 - t)_+ + 0.02\,(2 - t)_+ + 0.004\,(t - 2)_+$$
$$0.004\,(t - 8.36)_+.$$

Table 3.5: Performance of the algorithms on the survival of males with angina pectoris.

Method	#Iter.	Time (s)	Modified log-likelihood
	Exact data		
CNMCH - ties grouped	10	0.08	-3699.095
CNMCH - ties ungrouped	10	0.32	-3699.095
	Purely interval-censored data		
CNMCH - ties grouped	8	0.11	-4823.722
CNMCH - ties ungrouped	8	0.80	-4823.722
CNM	7	0.06	-4812.727 (Log-likelihood)

3.6 Simulation Study

3.6.1 Setup

A simulation study was carried out to investigate the performance of the CNMCH algorithm in different scenarios. To generate random samples from a bathtub-shaped hazard function in the simulation study, we used the exponentiated Weibull distribution, which was introduced by Mudholkar and Srivastava (1993). The exponentiated Weibull distribution has cumulative distribution function

$$F(x; \alpha, \theta) = \{1 - \exp(-x^{\alpha})\}^{\theta}, \quad x > 0,$$

and hence density

$$f(x; \alpha, \theta) = \alpha\theta x^{\alpha-1} e^{-x^{\alpha}} \{1 - \exp(-x^{\alpha})\}^{\theta-1}, \quad x > 0,$$

where $\alpha > 0$ and $\theta > 0$ are the shape and scale parameters, respectively. As noted by Nassar and Eissa (2003), the exponentiated Weibull distribution has a bathtub-shaped hazard function when $\alpha > 1$ and $\alpha\theta < 1$.

We performed a simulation study, which covers the situations of exact data, purely interval censored data and interval-censored data mixed with exact observations. A scheme for data generation similar to the one described in Wang (2008) was used. To create a data set, n exact event times were first drawn from the exponentiated Weibull distribution. For this purpose, we used R package **reliaR** (Kumar and Ligges, 2011) with shape parameters $\alpha = 4$ and $\theta = 0.2$. Afterwards, only $r \times n$ $(0 \leq r \leq 1)$ observations chosen randomly remain uncensored. A random sample of size 10 from the exponential distribution with mean 1 was taken for each censored observation, which divides the domain $[0, \infty)$ into 11 disjoint subintervals. The subinterval that contains the exact event time replaces it. Obviously, when $r = 1$, all observations are exact; when $0 < r < 1$, the data are made up of both exact and interval-censored observations; and when $r = 0$, all observations are purely interval-censored. We also included the simulation results produced by the SRB algorithm in the case of exact data. We had to give up the execution of this algorithm for a sample size of 1600 due to extremely high computational costs.

3.6.2 Results

The performance of the CNMCH algorithm is investigated in these situations: $(r, n) \in \{0\%, 50\%, 100\%\} \times \{400, 1600\}$. The experimental results based on 10 replications in each situation are summarized in Table 3.6, including the mean and standard deviation (in parentheses) of both the number of iterations and computation time (in seconds). From Table 3.6, we can conclude that the computational cost reduces as the proportion of exact observations increases. Apparently, from the simulation study, the SRB algorithm needed a much longer computation time when it is applicable.

Table 3.6: Results of simulation studies for $n = 400, 1600$.

Method	#Iter.	Time (s)	#Iter.	Time (s)	#Iter.	Time (s)
	$n = 400$					
	$r = 0\%$		$r = 50\%$		$r = 100\%$	
CNMCH	30.1 (2.52)	0.83 (0.12)	22.6 (3.83)	0.66 (0.14)	13.1 (1.52)	0.25 (0.06)
SRB					10 (1.33)	303.84 (25.57)
	$n = 1600$					
	$r = 0\%$		$r = 50\%$		$r = 100\%$	
CNMCH	31.1 (2.07)	5.12 (1.17)	26.8 (2.59)	2.83 (0.61)	13.8 (2.54)	2.20 (1.03)

3.7 Summary

In this chapter, we apply nonparametric approaches for estimating a hazard function under convexity restriction. A new algorithm, CNMCH, is proposed for computing the NPMLE of a convex hazard function, which can well deal with the situations of exact data, purely interval-censored data, and interval-censored data mixed with exact observations. We suggested a novel technique for expanding the support sets which avoids the problem of estimating the minimum of a convex hazard function. Therefore, it helps reduce the double loop needed by the SRB algorithm to a single loop and save a remarkable computation cost.

Through a simulation study and three real-world data sets, we have investigated and compared the performance of our proposed approach with that of the state-of-the-art one to the same problem. In comparison with the SRB algorithm, the CNMCH algorithm is computationally much faster. Furthermore, the CNMCH algorithm can also deal well with interval-censored data and the case of mixed exact and interval-censored observations, whereas the SRB algorithm can only apply to the case of exact observations. Also, the SRB algorithm does not handle the problem of duplicated observations. Numerical studies in the case of interval-censored data suggest that the shape-constrained method yields a more flexible and adaptable estimator than the unconstrained one.

In order to obtain a smooth estimation of a hazard function, one can generalize the NPMLE approach by applying a higher-order smoothness of an estimator. In addition, one could also be interested in investigating the effects of covariates on the distribution of the failure time data. We undertake detailed studies of these extensions in the subsequent chapters.

Chapter 4

Nonparametric Smooth Estimation of a Convex Hazard Function

4.1 Introduction

One of the primary goals of analyzing time to event data is to estimate the functions that characterize the distribution of a lifetime variable. The hazard rate function is a very useful tool to illustrate the instantaneous risk of observing the event of interest over time as one of the functions. Estimating a hazard function has many feasible applications in various fields such as clinical research, longitudinal and medical studies. For example, one may need to estimate hazard functions to graphically compare several different treatments or processes, or to predict hazard rates to clarify the way in which the process or treatment leaves the stationary status. Parametric models, in particular the exponential and Weibull distributions, have been widely used where the estimation of a hazard function is of interest. However, in the case of interval-censored data, parametric approaches may lead to an inconsistent estimator under model misspecification in which the corresponding asymptotic bias may result in invalid statistical inference (Gómez et al., 2009). In order to avoid parametric assumptions, one could employ nonparametric approaches for estimating a hazard function. The merit of nonparametric models is that the bias, due to model misspecification, can be reduced through nonparametric methods.

Traditional nonparametric estimators in the presence of right censoring were appropriately described in the literature. The two most popular of this type of estimators are the KM or product limit estimator of the survival function (Kaplan and Meier, 1958) and the NA estimator of the cumulative hazard function (Nelson, 1972; Aalen, 1978). The KM and NA estimators are basically step functions with jumps at the observed event times. Apart from their simplicity and ease of comprehension, they might lose some important information about the true underlying hazard function. In the case of heavily censored data, the accuracy of the estimation could not be reliable or even acceptable, since these estimators have only a few jumps with increasing sizes and overestimate the survival probabilities (Murray, 2001; López-de Ullibarri and Jácome, 2013b). With nonparametric shape constrained techniques and using the concept of smoothing, one can facilitate generating estimators that better adapt to a data set. Simulation studies by Pan (2000b) showed that smooth estimators, e.g., the kernel and logspline of a survival function, are more efficient than the classical NPMLE approaches.

The gist of this chapter is to present an outline for nonparametric estimation of a hazard function under convex shape restrictions and smoothness assumption. Our motivation comes from the desire to produce a smooth estimation of a hazard, or equivalently, a density function. We show that if the smoothness is introduced properly, it can help to increase estimation efficiency and also provide a more pleasant graphical representation.

4.2 Smooth Estimation of a Hazard Function

There exist a number of nonparametric approaches, as briefly summarized below, that provide smooth estimation of a hazard function. We first provide a broad overview of the existing methods and then present our new approach for solving the problem of nonparametric smooth estimation of a hazard function subject to the convex shape constraint.

4.2.1 Kernel-based Estimation

The simplest and well-known nonparametric approach is perhaps the kernel-based method. Kernel-based techniques for nonparametric function estimation are widely used in the literature, particularly for density function. Silverman (1986) addressed the problem of kernel-based density estimation in the case of no censoring. For issues of censoring, we refer the interested reader to Wand and Jones (1995) and Bowman and Azzalini (1997). In the case of interval-censored data, Braun et al. (2005) proposed the imputation method based on local likelihood density estimation, which applied an extension of kernel smoothing where the kernel weight is specified by the conditional expectation of the kernel over the interval.

Kernel estimation of the hazard function was first suggested by Watson and Leadbetter (1964) in the uncensored situation. Ramlau-Hansen (1983) and Tanner and Wong (1983) extended the kernel hazard function estimator to the case of censored data and investigated the properties of estimators.

Let $T_1, \cdots, T_n$ denote i.i.d. lifetimes from the continuous distribution function F, and $C_1, \cdots, C_n$ be the i.i.d. corresponding censoring times from continuous distribution G. The times T_i and C_i are usually assumed to be independent. It is not possible to observe both T_i and C_i. Instead, the observed variables are $X_i = \min(T_i, C_i)$ and $\delta_i = I(T_i \leq C_i)$. Let $(X_{(i)}, \delta_{(i)})$ be the ordered paired observations. The general form of a fixed bandwidth kernel-smoothed Tanner-Wong (TW) estimator of the hazard function in the case of right-censored observations can be written as

$$\hat{h}(t) = \frac{1}{b}\sum_{i=1}^{n} K\Big(\frac{t - X_{(i)}}{b}\Big)\frac{\delta_{(i)}}{n-i+1},$$

where $K(\cdot)$ is a fixed kernel function, $b > 0$ is a specified bandwidth or window parameter that determines the degree of smoothing and $\delta_{(i)}$ denotes the censoring indicator corresponding to the i-th order statistic $X_{(i)}$. Kernel functions are generally chosen to be symmetric probability density functions. Many authors have demonstrated that the resulting estimates are not greatly affected by the choice of the kernel function; see, e.g., Silverman (1986), Wand and Jones (1995) and Simonoff (1996). A general choice for the kernel function is the so-called "Epanechnikov" kernel ($K(x) = 0.75(1 - x^2)$ for $-1 \leq x \leq 1$) which is used in many numerical studies.

The idea of the presmoothed approach based on considering a smoother estimator of $p(X_{(i)})$ instead of $\delta_{(i)}$ for estimating a density or a hazard function can be traced back to the work of Cao and Jácome (2004) and Cao and López-de Ullibarri (2007). The smoother estimator is grounded on using the Nadaraya-Watson kernel estimator (Nadaraya, 1964; Watson, 1964)

$$\widehat{p}_{b_1}(t) = \frac{\sum_{i=1}^{n} K_{b_1}(t - X_{(i)})\delta_{(i)}}{\sum_{i=1}^{n} K_{b_1}(t - X_{(i)})}.$$

Hence, the presmoothed TW estimator has the form

$$\widehat{h}^{P}_{b_1,b_2}(t) = \frac{1}{n}\sum_{i=1}^{n}\frac{K_{b_2}(t - X_{(i)})\widehat{p}_{b_1}(X_{(i)})}{1 - H_n(X_{(i)}) + 1/n}.$$

This estimator relies on two parameters, the presmoothing bandwidth b_1 and the smoothing bandwidth b_2. The presmoothed estimator can also be perceived as the convolution of the TW estimator when b_1 tends to zero. López-de Ullibarri and Jácome (2013b) proposed two bandwidth selection methods, namely plug-in and bootstrap bandwidth selectors, which are available in R package **survPresmooth** (López-de Ullibarri and Jácome, 2013a).

The major advantage of the kernel-based estimation approach is its flexibility and interpretative appeal. Nevertheless, the main difficulty with the kernel method is the problem of determining the appropriate smoothing parameter or bandwidth parameter. As noted by Gefeller and Michels (1992), a deficiency of the fixed bandwidth kernel estimators is that the constant bandwidth results in unpleasant effects whenever the data are not equally distributed throughout the range of interest. For uneven distributed data, the fixed bandwidth kernel estimator tends to over-smooth dense regions and under-smooth sparse regions. Thus, the bandwidth should be large when the data are sparse, whereas in regions with many observations, the bandwidth should become small. This problem can be remedied by incorporating the concept of nearest neighbour, which originally introduced by Fix and Hodges (1989), into the definition of the bandwidth (see Gefeller and Dette, 1992). In addition, the fixed bandwidth kernel-based hazard function estimation is rather formidable due to the boundary effects that occur near the endpoints of the support of the hazard function. These effects are aggravated when the hazard rates are changing quickly near the endpoints of the data such as with a bathtub-shaped hazard function. Müller and Wang (1994) have proposed the solutions for boundary effects and balancing variance and bias problems by using a new class of boundary kernels and data-adaptive variable bandwidths, respectively. The hazard function estimates based on the new boundary kernels, and local and global bandwidth choices are available in R package **muhaz** (Hess and Gentleman, 2010). Nielsen (2003) considered the numerical studies of several variable bandwidth kernel estimators (see also Bagkavos and Patil, 2009). Another limitation of this approach is that it is not likelihood-based; therefore, it is not easy to make inferences and draw conclusions.

4.2.2 Spline Estimators

An alternative class of methods applies spline functions to estimate hazard functions. In this vein, hazard functions are piecewise polynomial functions of some spline functions jointed in a distinct sequence of cut points or "knots". For uncensored samples, Bloxom (1985) introduced a quadratic spline with knots placed at each decile of the sample's distribution. Then the quadratic spline function which approximates the hazard function for the case of uncensored data can be written as

$$\hat{h}(t) = \beta_0 + \beta_1 t + \beta_2 t^2 + \sum_{i=1}^{k} b_i (t - a_i)_+^2,$$

where $0 < a_1 < \cdots < a_k$ are k knots. The necessity of this model is imposing constraints on parameters to determine the shape of the hazard function. A cubic spline hazard model with tails which are linearly constrained was proposed by Herndon and Harrell (1990). Recently, Meyer and Habtzghi (2011) suggested cubic regression spline with the knots placed at equal data quantiles for approximating a convex hazard function for both uncensored and right-censored data.

In order to retrieve the lost information in the discrete NPMLE, one can also apply the logspline model for smooth estimation of a hazard function which is a smooth estimate directly from the data. In essence, the logspline density estimation models the logarithm of a density function by using a polynomial spline. A number of papers have dealt with the logspline density estimation; see e.g., Stone (1990) and Kooperberg and Stone (1991). The logspline density estimation was developed for the case of censored data by Kooperberg and Stone (1992).

Let the integer $K \geq 3$ and $t_1, \cdots, t_K$ be a knot sequence such that $-\infty \leq L < t_1 < \cdots < t_K < U \leq \infty$ where L and U are some numbers. The K-dimensional natural cubic spline space has a basis of the form $B_1, \cdots, B_p, p = K - 1$. Then, the logspline density model based on these basis functions has the form

$$f(x; \boldsymbol{\theta}) = \exp\left(\sum_{i=1}^{p} \theta_i B_i(x) - C(\boldsymbol{\theta})\right), \quad L < x < U,$$

where $C(\boldsymbol{\theta}) = \log\left\{\int \exp(\sum_{i=1}^{p} \theta_i B_i(u))\, \mathrm{d}u\right\} < \infty$ is the normalizing constant and $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_p)^\top$ is the parameter vector.

The number of knots is selected based on a stepwise process of addition and deletion of knots depending on Wald statistics, as suggested by Kooperberg and Stone (1992). The final model is chosen that minimises the BIC. The knot addition and deletion algorithm is available in R package **logspline** (Kooperberg, 2007).

4.3 Maximum Likelihood Estimation with Smoothness Assumption

Consider a sample of n subjects from a homogeneous population with survival function $S(t)$. In practice, the observed data for each individual under study is either an exact survival time or, in the case of right censoring, a censoring time. In many clinical trials and longitudinal studies that entail periodic follow-ups, instead of having the survival times T_i observed directly, it is only known that the occurrence of the event of interest for each individual lies in the interval between visits; e.g., the interval $O_i = (L_i, R_i]$, where $L_i < T_i \leq R_i$. By convention, the exact failure time T_i is obtained when $L_i = R_i$. For the sake of simplicity, we assume that the first n_1 observations are exact, and the rest are interval-censored data. Then the log-likelihood function for both the exact and interval-censored observations can be written as

$$\ell(h) = \sum_{i=1}^{n_1} \{\log h(T_i) - H(T_i)\} + \sum_{i=n_1+1}^{n} \log \{S(L_i) - S(R_i)\}. \tag{4.1}$$

As mentioned in Chapter 3, the log-likelihood function can be made arbitrarily large by increasing value of h at the largest observed value. We therefore maximise the modified log-likelihood

$$\tilde{\ell}(h) = \sum_{\substack{i=1 \\ T_i \neq T_{\max}}}^{n_1} \log h(T_i) - \sum_{i=1}^{n_1} H(T_i) + \sum_{i=n_1+1}^{n} \log \{S(L_i) - S(R_i)\}, \tag{4.2}$$

where $T_{(n_1)} = \max\{T_1, \cdots, T_{n_1}\}$, $I_{\max}$ is the maximum of all the finite values of L_i and R_i, and $T_{\max} = \max\left\{T_{(n_1)}, I_{\max}\right\}$.

To obtain a smooth estimator of a hazard function, we describe it as a continuous piecewise quadratic function

$$h(t) = \alpha_0 + \alpha_1\, t + \alpha_2\, t^2 + \sum_{j=1}^{k} \nu_j (\tau_j - t)_+^2 + \sum_{j=1}^{m} \mu_j (t - \eta_j)_+^2, \tag{4.3}$$

where, for example,

$$(\tau_j - t)_+^2 = \begin{cases} (\tau_j - t)^2, & \text{if } t < \tau_j, \\ 0, & \text{if } t \geq \tau_j. \end{cases}$$

and $\boldsymbol{\theta} = (\boldsymbol{\tau}, \boldsymbol{\eta})^\top$ is a support point vector of h with its corresponding mass vector $\boldsymbol{\pi} = (\alpha_0, \alpha_1, \alpha_2, \boldsymbol{\nu}, \boldsymbol{\mu})^\top$, all are positive except α_0 and α_1 which are free parameters, and $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_k)$, $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_m)$, $\boldsymbol{\tau} = (\tau_1, \cdots, \tau_k)$, and $\boldsymbol{\eta} = (\eta_1, \cdots, \eta_m)$. Notationally, h is exchangeable with its mass vector $\boldsymbol{\pi}$ and support point vector $\boldsymbol{\theta}$, namely $\tilde{\ell}(h) \equiv \tilde{\ell}(\boldsymbol{\pi}, \boldsymbol{\theta})$. One of the properties of this model is that its first derivative is continuous. The main advantage of the smoothed approach rather than the non-smoothed one is that the smoothness, if introduced correctly, should help increase estimation efficiency, and it also gives an intuitively pleasant graphical presentation. The smooth nonparametric maximum likelihood problem of estimating the unknown hazard function h, owing to the imposition of convex shape restriction, is an optimization problem defined as

$$\begin{aligned} &\text{maximize} \quad \tilde{\ell}(h) \\ &\text{subject to} \quad h \in \mathcal{K}, \end{aligned} \tag{4.4}$$

where $\mathcal{K}$ is a convex set of hazard functions with a decreasing first derivative and the modified log-likelihood $\tilde{\ell}(h)$ is a concave function on $\mathcal{K}$. By maximizing the modified log-likelihood function (4.2) over all hazard functions h in $\mathcal{K}$, we obtain the smooth nonparametric maximum likelihood estimate (SNPMLE) of $(\boldsymbol{\pi}, \boldsymbol{\theta})$, denoted by $(\widehat{\boldsymbol{\pi}}, \widehat{\boldsymbol{\theta}})$. This problem has no closed-form solution, and hence iterative methods must be used.

Let us define the following two gradient functions in aid of the basis functions $e_{1,\tau} = (\tau - t)_+^2$, and $e_{2,\eta} = (t-\eta)_+^2$:

$$\begin{aligned} d_1(\tau; h) &= \lim_{\epsilon \to 0} \frac{\tilde{\ell}(h + \epsilon e_{1,\tau}) - \tilde{\ell}(h)}{\epsilon} \\ &= \sum_{\substack{i=1 \\ T_i \neq T_{\max}}}^{n_1} \frac{(\tau - T_i)_+^2}{h(T_i)} - \sum_{i=1}^{n_1} \frac{\tau^3 - (\tau - T_i)_+^3}{3} \quad + \\ &\quad \sum_{i=n_1+1}^{n} \left\{ -\frac{\tau^3 - (\tau - L_i)_+^3}{3} - \frac{(\tau - R_i)_+^3 - (\tau - L_i)_+^3}{3} \Delta_i(H) \right\}, \\ &\quad \text{for } 0 < \tau \leq \eta_1, \end{aligned} \tag{4.5}$$

and

$$\begin{aligned} d_2(\eta; h) &= \lim_{\epsilon \to 0} \frac{\tilde{\ell}(h + \epsilon e_{2,\eta}) - \tilde{\ell}(h)}{\epsilon} \\ &= \sum_{\substack{i=1 \\ T_i \neq T_{\max}}}^{n_1} \frac{(T_i - \eta)_+^2}{h(T_i)} - \sum_{i=1}^{n_1} \frac{(T_i - \eta)_+^3}{3} \quad + \\ &\quad \sum_{i=n_1+1}^{n} \left\{ -\frac{(L_i - \eta)_+^3}{3} - \frac{(L_i - \eta)_+^3 - (R_i - \eta)_+^3}{3} \Delta_i(H) \right\}, \\ &\quad \text{for } \tau_k \leq \eta < T_{\max}. \end{aligned} \tag{4.6}$$

where $\Delta_i(H) = \exp(H(L_i) - H(R_i))/(1 - \exp(H(L_i) - H(R_i)))$. Note that d_1 and d_2 are piecewise cubic functions of τ and η, respectively. For the purpose of completeness, let us also define

$$
\begin{aligned}
d_0(\alpha_0; h) &= \lim_{\epsilon \to 0} \frac{\tilde{\ell}(h + \epsilon e_0) - \tilde{\ell}(h)}{\epsilon} \\
&= \sum_{\substack{i=1 \\ T_i \neq T_{\max}}}^{n_1} \frac{1}{h(T_i)} - \sum_{i=1}^{n_1} T_i + \sum_{i=n_1+1}^{n} \left\{ -L_i - (L_i - R_i)\, \Delta_i(H) \right\}, \\
&\quad \text{for } -\infty < \alpha_0 < \infty,
\end{aligned}
$$

$$
\begin{aligned}
d_1(\alpha_1; h) &= \lim_{\epsilon \to 0} \frac{\tilde{\ell}(h + \epsilon e_1) - \tilde{\ell}(h)}{\epsilon} \\
&= \sum_{\substack{i=1 \\ T_i \neq T_{\max}}}^{n_1} \frac{T_i}{h(T_i)} - \sum_{i=1}^{n_1} \frac{T_i^2}{2} + \sum_{i=n_1+1}^{n} \left\{ \frac{-L_i^2 - (L_i - R_i)^2}{2}\, \Delta_i(H) \right\}, \\
&\quad \text{for } -\infty < \alpha_1 < \infty,
\end{aligned}
$$

and

$$
\begin{aligned}
d_2(\alpha_2; h) &= \lim_{\epsilon \to 0} \frac{\tilde{\ell}(h + \epsilon e_2) - \tilde{\ell}(h)}{\epsilon} \\
&= \sum_{\substack{i=1 \\ T_i \neq T_{\max}}}^{n_1} \frac{T_i^2}{h(T_i)} - \sum_{i=1}^{n_1} \frac{T_i^3}{3} + \sum_{i=n_1+1}^{n} \left\{ \frac{-L_i^3 - (L_i - R_i)^3}{3}\, \Delta_i(H) \right\}, \\
&\quad \text{for } \alpha_2 \geq 0,
\end{aligned}
$$

where $e_0 = 1$, $e_1 = t$, and $e_2 = t^2$.

4.3.1 Computational Algorithm

In this section, we discuss some computational aspects of the smooth nonparametric convex hazard function estimator and then present an algorithm for computing the SNPMLE of a convex hazard function. Similar to the CNMCH algorithm, this algorithm involves two main phases. The first phase entails expanding and contracting $\boldsymbol{\theta}$, and the second consists of updating $\boldsymbol{\pi}$. With $\boldsymbol{\theta}$ fixed, $\boldsymbol{\pi}$ can be efficiently updated to $\boldsymbol{\pi}'$ as follows. Denoting the gradient vector and Hessian matrix of the modified log-likelihood (4.2) (see Section 4.3.2) as follows

$$\mathbf{g} \equiv \mathbf{g}(\boldsymbol{\pi}, \boldsymbol{\theta}) = \frac{\partial \tilde{\ell}}{\partial \boldsymbol{\pi}},$$
$$\mathbf{H} \equiv \mathbf{H}(\boldsymbol{\pi}, \boldsymbol{\theta}) = \frac{\partial^2 \tilde{\ell}}{\partial \boldsymbol{\pi} \partial \boldsymbol{\pi}^\top}.$$

By using the Taylor series expansion to second order, one obtains the following quadratic approximation to $\tilde{\ell}(\boldsymbol{\pi}', \boldsymbol{\theta})$ around $\boldsymbol{\pi}$:

$$\tilde{\ell}(\boldsymbol{\pi}', \boldsymbol{\theta}) - \tilde{\ell}(\boldsymbol{\pi}, \boldsymbol{\theta}) \approx \mathbf{g}^\top(\boldsymbol{\pi}' - \boldsymbol{\pi}) + \frac{1}{2}(\boldsymbol{\pi}' - \boldsymbol{\pi})^\top \mathbf{H}(\boldsymbol{\pi}' - \boldsymbol{\pi}). \tag{4.7}$$

With known $\boldsymbol{\theta}$, the problem of maximizing (4.7) can be turned into a least squares linear regression problem as follows. Let $\mathbf{H} = -\mathbf{R}^\top\mathbf{R}$, where the upper triangular matrix $\mathbf{R} \equiv \mathbf{R}(\boldsymbol{\pi}, \boldsymbol{\theta})$ can be attained by the Cholesky decomposition of $-\mathbf{H}$ (see Section 4.3.3). Then maximising $\tilde{\ell}(\boldsymbol{\pi}', \boldsymbol{\theta})$ can be approximated by

$$\text{minimize} \quad \left\| \mathbf{R}\begin{pmatrix} \boldsymbol{\pi}'_1 \\ \boldsymbol{\pi}'_2 \end{pmatrix} - \mathbf{R}\begin{pmatrix} \boldsymbol{\pi}_1 \\ \boldsymbol{\pi}_2 \end{pmatrix} - \mathbf{b}_1 \right\|^2, \tag{4.8}$$

where $\boldsymbol{\pi}'_1 \in \mathbb{R}^2, \boldsymbol{\pi}'_2 \in \mathbb{R}_+^{1+k+m}$, and $\mathbf{b}_1$ the solution of the lower triangular system $\mathbf{R}^\top\mathbf{b}_1 = \mathbf{g}$. As problem (4.8) has only non-negativity constraints on $\boldsymbol{\pi}'_2$, this problem can be solved using the NNLS algorithm of Lawson and Hanson (1974) by leaving $\boldsymbol{\pi}'_1$ unconstrained. Also, the popular step-halving backtracking is conducted to ensure monotone increase of the log-likelihood function.

To find the smoothed NPMLE of a convex hazard function, one needs to locate all its new support points by incorporating the gradient functions. At each iteration, the new algorithm expands the two support sets $\boldsymbol{\tau}$ and $\boldsymbol{\eta}$ corresponding to the decreasing and increasing parts of a convex hazard function by including all local maxima of the gradient functions $d_1(\tau; h)$ (4.5) and $d_2(\eta; h)$ (4.6). In order to locate all these local maxima, we replace the combined Newton-bisection method, basically applied in Wang (2007), with the combined secant-bisection method which has the merit of not computing the second derivative of the gradient functions. For the constant hazard part, the global maximum of the gradient functions $d_1(\tau; h)$ and $d_2(\eta; h)$ is chosen and then added to the support set to which it corresponds. The redundant support points with zero masses are discarded after updating $\boldsymbol{\pi}$ by optimising the quadratic approximation (4.8).

This modified algorithm, which we call CNMSCH for computing a smooth convex hazard function, is outlined as follows.

Algorithm 4.1 (CNMSCH)**.** Set $s = 0$ and choose an initial estimate h_0 with finite support such that $\ell(h_0) > -\infty$. Repeat the following steps:

Step 1: compute all local maxima of gradients, hence giving $\tau^*_{s1}, \cdots, \tau^*_{sk}$ for the decreasing part, $\eta^*_{s1}, \cdots, \eta^*_{sm}$ for the increasing part, and ρ^*_s for the constant part of the convex hazard function.

Step 2: set $\boldsymbol{\tau}^*_s = (\tau^*_{s1}, \cdots, \tau^*_{sk})^\top$, $\boldsymbol{\eta}^*_s = (\eta^*_{s1}, \cdots, \eta^*_{sm})^\top$, $\boldsymbol{\theta}_{s^+} = (\boldsymbol{\tau}_s^\top, \boldsymbol{\tau}_s^{*\top}, \rho^*_s, \boldsymbol{\eta}_s^\top, \boldsymbol{\eta}_s^{*\top})^\top$ and $\boldsymbol{\pi}^+_s = (\alpha_{0s}, \alpha_{1s}, \alpha_{2s}, \boldsymbol{\nu}_s^\top, \mathbf{0}^\top, 0, \boldsymbol{\mu}_s^\top, \mathbf{0}^\top)^\top$. Find $\boldsymbol{\pi}^-_{s+1}$ by solving problem (4.8), with $\mathbf{R}$ replaced by $\mathbf{R}^+_s = \mathbf{R}(\boldsymbol{\pi}^+_s, \boldsymbol{\theta}^+_s)$ and $\mathbf{b}^+_{1s} = (\boldsymbol{\pi}_s^{+\top}, \boldsymbol{\theta}_s^{+\top})^\top$, respectively, followed by conducting a backtracking line search.

Step 3: discard all support points with zero masses in $\boldsymbol{\pi}^-_{s+1}$, which gives h_{s+1} with $\boldsymbol{\theta}_{s+1}$ and $\boldsymbol{\pi}_{s+1}$. Stop if converged; otherwise, set $s = s + 1$.

For testing convergence in step 3, the criterion

$$|\tilde{\ell}(h_{s+1}) - \tilde{\ell}(h_s)| \leq \gamma$$

can be used, for some small $\gamma > 0$, e.g., $\gamma = 10^{-6}$.

4.3.2 Derivatives of the Modified Log-likelihood Function

The first partial derivatives of the modified log-likelihood function is given as follows:

$$
\begin{aligned}
\frac{\partial\tilde{\ell}}{\partial\alpha_0} &= \sum_{\substack{i=1\\T_i\neq T_{\max}}}^{n_1}\left\{\frac{1}{h(T_i)}\right\} - \sum_{i=1}^{n_1} T_i + \sum_{i=n_1+1}^{n}\left\{-L_i - (L_i - R_i)\Delta_i(H)\right\},\\
\frac{\partial\tilde{\ell}}{\partial\alpha_1} &= \sum_{\substack{i=1\\T_i\neq T_{\max}}}^{n_1}\left\{\frac{T_i}{h(T_i)}\right\} - \sum_{i=1}^{n_1}\frac{T_i^2}{2} + \sum_{i=n_1+1}^{n}\left\{-\frac{L_i^2}{2} - \frac{(L_i^2 - R_i^2)}{2}\Delta_i(H)\right\},\\
\frac{\partial\tilde{\ell}}{\partial\alpha_2} &= \sum_{\substack{i=1\\T_i\neq T_{\max}}}^{n_1}\left\{\frac{T_i^2}{h(T_i)}\right\} - \sum_{i=1}^{n_1}\frac{T_i^3}{3} + \sum_{i=n_1+1}^{n}\left\{-\frac{L_i^3}{3} - \frac{(L_i^3 - R_i^3)}{3}\Delta_i(H)\right\},\\
\frac{\partial\tilde{\ell}}{\partial\nu_j} &= \sum_{\substack{i=1\\T_i\neq T_{\max}}}^{n_1}\left\{\frac{(\tau_j - T_i)_+^2}{h(T_i)}\right\} - \sum_{i=1}^{n_1}\frac{\tau_j^3 - (\tau_j - T_i)_+^3}{3} + \\
&\sum_{i=n_1+1}^{n}\left\{-\frac{\tau_j^3 - (\tau_j - L_i)_+^3}{3} - \frac{(\tau_j - R_i)_+^3 - (\tau_j - L_i)_+^3}{3}\Delta_i(H)\right\},\\
\frac{\partial\tilde{\ell}}{\partial\mu_j} &= \sum_{\substack{i=1\\T_i\neq T_{\max}}}^{n_1}\left\{\frac{(T_i - \eta_j)_+^2}{h(T_i)}\right\} - \sum_{i=1}^{n_1}\frac{(T_i - \eta_j)_+^3}{3} + \\
&\sum_{i=n_1+1}^{n}\left\{-\frac{(L_i - \eta_j)_+^3}{3} - \frac{(L_i - \eta_j)_+^3 - (R_i - \eta_j)_+^3}{3}\Delta_i(H)\right\},
\end{aligned}
$$

where $\Delta_i(H) = \exp(H(L_i) - H(R_i))/(1 - \exp(H(L_i) - H(R_i)))$.

The second derivatives of the modified log-likelihood function that form the Hessian matrix $\mathbf{H}$ are given by

$$\frac{\partial^2 \tilde{\ell}}{\partial \alpha_0^2} = \sum_{\substack{i=1 \\ T_i \neq T_{\max}}}^{n_1} \left\{ -\frac{1}{h(T_i)^2} \right\} + \sum_{i=n_1+1}^{n} \left\{ -(L_i - R_i)^2 \Delta_i(H) \right\},$$

$$\frac{\partial^2 \tilde{\ell}}{\partial \alpha_1^2} = \sum_{\substack{i=1 \\ T_i \neq T_{\max}}}^{n_1} \left\{ -\frac{T_i^2}{h2(T_i)^2} \right\} + \sum_{i=n_1+1}^{n} \left\{ -\left(\frac{L_i^2 - R_i^2}{2} \right)^2 \Delta_i(H) \right\},$$

$$\frac{\partial^2 \tilde{\ell}}{\partial \alpha_2^2} = \sum_{\substack{i=1 \\ T_i \neq T_{\max}}}^{n_1} \left\{ -\frac{T_i^4}{h(T_i)^2} \right\} + \sum_{i=n_1+1}^{n} \left\{ -\left(\frac{L_i^3 - R_i^3}{3} \right)^2 \Delta_i(H) \right\},$$

$$\frac{\partial^2 \tilde{\ell}}{\partial \mu_j^2} = \sum_{\substack{i=1 \\ T_i \neq T_{\max}}}^{n_1} \left\{ -\frac{(\tau_j - T_i)_+^4}{h(T_i)^2} \right\} + \sum_{i=n_1+1}^{n} \left\{ -\left(\frac{(\tau_j - R_i)_+^3 - (\tau_j - L_i)_+^3}{3} \right)^2 \Delta_i(H) \right\},$$

$$\frac{\partial^2 \tilde{\ell}}{\partial \nu_j^2} = \sum_{\substack{i=1 \\ T_i \neq T_{\max}}}^{n_1} \left\{ -\frac{(T_i - \eta_j)_+^4}{h(T_i)^2} \right\} + \sum_{i=n_1+1}^{n} \left\{ -\left(\frac{(L_i - \eta_j)_+^3 - (R_i - \eta_j)_+^3}{3} \right)^2 \Delta_i(H) \right\},$$

$$\frac{\partial^2 \tilde{\ell}}{\partial \alpha_0 \partial \alpha_1} = \sum_{\substack{i=1 \\ T_i \neq T_{\max}}}^{n_1} \left\{ -\frac{T_i}{h(T_i)^2} \right\} + \sum_{i=n_1+1}^{n} \left\{ -(L_i - R_i) \left(\frac{L_i^2 - R_i^2}{2} \right) \Delta_i(H) \right\},$$

$$\frac{\partial^2 \tilde{\ell}}{\partial \alpha_0 \partial \alpha_2} = \sum_{\substack{i=1 \\ T_i \neq T_{\max}}}^{n_1} \left\{ -\frac{T_i^2}{h(T_i)^2} \right\} + \sum_{i=n_1+1}^{n} \left\{ -(L_i - R_i) \left(\frac{L_i^3 - R_i^3}{3} \right) \Delta_i(H) \right\},$$

$$\frac{\partial^2 \tilde{\ell}}{\partial \alpha_1 \partial \alpha_2} = \sum_{\substack{i=1 \\ T_i \neq T_{\max}}}^{n_1} \left\{ -\frac{T_i^3}{h(T_i)^2} \right\} + \sum_{i=n_1+1}^{n} \left\{ -\left(\frac{L_i^2 - R_i^2}{2} \right) \left(\frac{L_i^3 - R_i^3}{3} \right) \Delta_i(H) \right\},$$

$$\frac{\partial^2 \tilde{\ell}}{\partial \nu_j \partial \alpha_0} = \sum_{\substack{i=1 \\ T_i \neq T_{\max}}}^{n_1} \left\{ -\frac{(\tau_j - T_i)_+^2}{h(T_i)^2} \right\} +$$

$$\sum_{i=n_1+1}^{n} \left\{ -(L_i - R_i) \left(\frac{(\tau_j - R_i)_+^3 - (\tau_j - L_i)_+^3}{3} \right) \Delta_i(H) \right\},$$

$$\frac{\partial^2 \tilde{\ell}}{\partial \nu_j \partial \alpha_1} = \sum_{\substack{i=1 \\ T_i \neq T_{\max}}}^{n_1} \left\{ -\frac{(\tau_j - T_i)_+^2 \, T_i}{h(T_i)^2} \right\} +$$

$$\sum_{i=n_1+1}^{n} \left\{ -\left(\frac{L_i^2 - R_i^2}{2} \right) \left(\frac{(\tau_j - R_i)_+^3 - (\tau_j - L_i)_+^3}{3} \right) \Delta_i(H) \right\},$$

$$\frac{\partial^2 \tilde{\ell}}{\partial \nu_j \partial \alpha_2} = \sum_{\substack{i=1 \\ T_i \neq T_{\max}}}^{n_1} \left\{ -\frac{(\tau_j - T_i)_+^2 \, T_i^2}{h(T_i)^2} \right\} +$$

$$\sum_{i=n_1+1}^{n} \left\{ -\left(\frac{L_i^3 - R_i^3}{3} \right) \left(\frac{(\tau_j - R_i)_+^3 - (\tau_j - L_i)_+^3}{3} \right) \Delta_i(H) \right\},$$

$$
\begin{aligned}
\frac{\partial^2 \tilde{\ell}}{\partial \mu_j \partial \alpha_0} = & \sum_{\substack{i=1 \\ T_i \neq T_{\max}}}^{n_1} \left\{ -\frac{(T_i - \eta_j)_+^2}{h(T_i)^2} \right\} + \\
& \sum_{i=n_1+1}^{n} \left\{ - (L_i - R_i) \left(\frac{(L_i - \eta_j)_+^3 - (R_i - \eta_j)_+^3}{3} \right) \Delta_i(H) \right\}, \\
\frac{\partial^2 \tilde{\ell}}{\partial \mu_j \partial \alpha_1} = & \sum_{\substack{i=1 \\ T_i \neq T_{\max}}}^{n_1} \left\{ -\frac{(T_i - \eta_j)_+^2 \; T_i}{h(T_i)^2} \right\} + \\
& \sum_{i=n_1+1}^{n} \left\{ - \left(\frac{L_i^2 - R_i^2}{2} \right) \left(\frac{(L_i - \eta_j)_+^3 - (R_i - \eta_j)_+^3}{3} \right) \Delta_i(H) \right\}, \\
\frac{\partial^2 \tilde{\ell}}{\partial \mu_j \partial \alpha_2} = & \sum_{\substack{i=1 \\ T_i \neq T_{\max}}}^{n_1} \left\{ -\frac{(T_i - \eta_j)_+^2 \; T_i^2}{h(T_i)^2} \right\} + \\
& \sum_{i=n_1+1}^{n} \left\{ - \left(\frac{L_i^3 - R_i^3}{3} \right) \left(\frac{(L_i - \eta_j)_+^3 - (R_i - \eta_j)_+^3}{3} \right) \Delta_i(H) \right\}, \\
\frac{\partial^2 \tilde{\ell}}{\partial \mu_j \partial \nu_j} = & \sum_{\substack{i=1 \\ T_i \neq T_{\max}}}^{n_1} \left\{ -\frac{(T_i - \eta_j)_+^2 \; (\tau_j - T_i)_+^2}{h(T_i)^2} \right\} + \\
& \sum_{i=n_1+1}^{n} \left\{ - \left(\frac{(\tau_j - R_i)_+^3 - (\tau_j - L_i)_+^3}{3} \right) \left(\frac{(L_i - \eta_j)_+^3 - (R_i - \eta_j)_+^3}{3} \right) \Delta_i(H) \right\},
\end{aligned}
$$

where $\Delta_i(H) = \exp(H(L_i) - H(R_i))/(1 - \exp(H(L_i) - H(R_i))^2$.

4.3.3 QR Factorization of $\mathbf{D}$

The Hessian matrix $\mathbf{H}$ can be rewritten

$$\mathbf{H} = -\mathbf{D}^\top \mathbf{D}$$

where

$$\mathbf{D} = \begin{pmatrix} \frac{1}{h(T_1)} & \frac{T_1}{h(T_1)} & \frac{T_1^2}{h(T_1)} & \frac{(\tau_1 - T_1)_+^2}{h(T_1)} & \cdots & \frac{(\tau_k - T_1)_+^2}{h(T_1)} & \frac{(T_1 - \eta_1)_+^2}{h(T_1)} & \cdots & \frac{(T_1 - \eta_m)_+^2}{h(T_1)} \\ \frac{1}{h(T_2)} & \frac{T_2}{h(T_2)} & \frac{T_2^2}{h(T_2)} & \frac{(\tau_1 - T_2)_+^2}{h(T_2)} & \cdots & \frac{(\tau_k - T_2)_+^2}{h(T_2)} & \frac{(T_2 - \eta_1)_+^2}{h(T_2)} & \cdots & \frac{(T_2 - \eta_m)_+^2}{h(T_2)} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{1}{h(T_{n_1})} & \frac{T_{n_1}}{h(T_{n_1})} & \frac{T_{n_1}^2}{h(T_{n_1})} & \frac{(\tau_1 - T_{n_1})_+^2}{h(T_{n_1})} & \cdots & \frac{(\tau_k - T_{n_1})_+^2}{h(T_{n_1})} & \frac{(T_{n_1} - \eta_1)_+^2}{h(T_{n_1})} & \cdots & \frac{(T_{n_1} - \eta_m)_+^2}{h(T_{n_1})} \\ \alpha_{0_{n_1+1}} & \alpha_{1_{n_1+1}} & \alpha_{2_{n_1+1}} & \beta_{n_1+1} & \cdots & \gamma_{n_1+1} & \lambda_{n_1+1} & \cdots & \delta_{n_1+1} \\ \alpha_{0_{n_1+2}} & \alpha_{1_{n_1+2}} & \alpha_{2_{n_1+2}} & \beta_{n_1+2} & \cdots & \gamma_{n_1+2} & \lambda_{n_1+2} & \cdots & \delta_{n_1+2} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \alpha_{0_n} & \alpha_{1_n} & \alpha_{2_n} & \beta_n & \cdots & \gamma_n & \lambda_n & \cdots & \delta_n \end{pmatrix}$$

and

$$\begin{aligned} \alpha_{0_i} &= (L_i - R_i)\ \Delta_i(H), \\ \alpha_{1_i} &= \left(\frac{L_i^2 - R_i^2}{2}\right)\ \Delta_i(H), \\ \alpha_{2_i} &= \left(\frac{L_i^3 - R_i^3}{3}\right)\ \Delta_i(H), \\ \beta_i &= \frac{(\tau_1 - L_i)_+^3 - (\tau_1 - R_i)_+^3}{3}\ \Delta_i(H), \\ \gamma_i &= \frac{(\tau_k - L_i)_+^3 - (\tau_k - R_i)_+^3}{3}\ \Delta_i(H), \\ \lambda_i &= \frac{(L_i - \eta_1)_+^3 - (R_i - \eta_1)_+^3}{3}\ \Delta_i(H), \\ \delta_i &= \frac{(L_i - \eta_k)_+^3 - (R_i - \eta_k)_+^3}{3}\ \Delta_i(H), \\ \Delta_i(H) &= \frac{\sqrt{\exp\left(H(L_i) - H(R_i)\right)}}{1 - \exp\left(H(L_i) - H(R_i)\right)}, \end{aligned}$$

for $i = n_1 + 1, n_1 + 2, \cdots, n$.

Therefore, the Taylor series expansion about $\boldsymbol{\pi}$ with $\boldsymbol{\theta}$ fixed gives

$$
\begin{aligned}
\tilde{\ell}(\boldsymbol{\pi}') - \tilde{\ell}(\boldsymbol{\pi}) &\approx \mathbf{g}^\top(\boldsymbol{\pi}' - \boldsymbol{\pi}) + \frac{1}{2}(\boldsymbol{\pi}' - \boldsymbol{\pi})^\top \mathbf{H}(\boldsymbol{\pi}' - \boldsymbol{\pi}) \\
&= -\frac{1}{2}\|\mathbf{D}(\boldsymbol{\pi}' - \boldsymbol{\pi}) - \mathbf{d}\|^2 + c \\
&= -\frac{1}{2}\|\mathbf{R}(\boldsymbol{\pi}' - \boldsymbol{\pi}) - \mathbf{b}_1\|^2 - \frac{1}{2}\|\mathbf{b}_2\|^2 + c \\
&= -\frac{1}{2}\left\|\mathbf{R}\begin{pmatrix}\boldsymbol{\pi}'_1 \\ \boldsymbol{\pi}'_2\end{pmatrix} - \mathbf{R}\begin{pmatrix}\boldsymbol{\pi}_1 \\ \boldsymbol{\pi}_2\end{pmatrix} - \mathbf{b}_1\right\|^2 - \frac{1}{2}\|\mathbf{b}_2\|^2 + c,
\end{aligned}
$$

where $\mathbf{D} = \mathbf{QR}$ by a QR decomposition, $\mathbf{d} = (\mathbf{R}^\top)^{-1}\mathbf{g}^\top$, and $\boldsymbol{\pi}'_1 \in \mathbb{R}^2, \boldsymbol{\pi}'_2 \in \mathbb{R}_+^{1+m+k}$. To obtain the least squares linear regression problem (4.7) with only nonnegativity constraints, we partitioned $\mathbf{Q} = \begin{pmatrix}\mathbf{Q_1} \\ \mathbf{Q_2}\end{pmatrix}$, where $\mathbf{Q_1} \in \mathbb{R}^{n\times(3+k+m)}$, $\mathbf{Q}_1^\top\mathbf{D} = \begin{pmatrix}\mathbf{R} \\ \mathbf{0}\end{pmatrix}$, where $\mathbf{R} \in \mathbb{R}^{(3+k+m)\times(3+k+m)}$, and $\mathbf{Q}^\top\mathbf{b} = \begin{pmatrix}\mathbf{b}_1 \\ \mathbf{b}_2\end{pmatrix} = \begin{pmatrix}\mathbf{Q}_1{}^\top\mathbf{d} \\ \mathbf{Q}_2{}^\top\mathbf{d}\end{pmatrix}$. Note that the $\mathbf{R}$ obtained by the QR decomposition of $\mathbf{D}$ is the same as by the Cholesky decomposition of $-\mathbf{H}$.

4.4 Simulation Studies

4.4.1 With Exact Observations

Setup

For the case of exact data, we carried out a simulation study to compare the performance of four nonparametric density estimators: the logspline (LS) density estimator of Kooperberg and Stone (1992), the presmoothed (PS) density estimator of López-de Ullibarri and Jácome (2013b) and our piecewise linear convex hazard (LCH) and smooth convex hazard (SCH) estimators. To compute the parameter estimates of the LS approach, we used the knot addition and deletion algorithm described in Stone et al. (1997), which is available by the function logspline in R package **logspline** (Kooperberg, 2007). For the PS density estimator, a presmoothed estimate of a density function is computed by using the biweight kernel function, which is available by the function presmooth in R package **survPresmooth** (López-de Ullibarri and Jácome, 2013a). Throughout the study, we implemented a plug-in bandwidth selection according to Cao and López-de Ullibarri (2007) and Jácome et al. (2008) in the case of density and hazard function estimation.

Table 4.1 lists two different distributions with bathtub-shaped hazard functions, namely the exponentiated Weibull (EW) (Nassar and Eissa, 2003) and the bathtub (BT) (Haupt and Schäbe, 1997) distributions, used in this simulation study. It is worth pointing out that the BT distribution has convex hazards for all values of $\beta > -1/2$. These distributions were studied by Wang (2000) and Jankowski and Wellner (2009b). The appealing feature of the former distribution is the capability of describing the bathtub-shaped failure rate lifetime data of a mechanical or electronic component; the latter will be able to model lifetime behaviour appropriately. The upper panel of Figure 4.1 displays the plots of the hazard functions from the EW distribution with $\alpha = 4, 3$ and $\theta = 0.2, 0.3$, respectively; whereas, the lower panel of Figure 4.1 shows the hazard functions from the BT distribution with $\beta = 0, 0.2$ and $\alpha = 1, 3$, respectively.

Table 4.1: Distributions for the simulation study.

Distribution	Notation	Density Function	Parameters
Exponentiated Weibull	EW(α, θ)	$\alpha\theta x^{\alpha-1}e^{-x^\alpha}\left\{1-\exp(-x^\alpha)\right\}^{\theta-1}$	$\alpha > 0$, $\theta > 0$
Bathtub	BT(β, α)	$\frac{1+2\beta}{2\alpha\sqrt{\beta^2+(1+2\beta)x/\alpha}}$	$0 \le x \le \alpha$, $-1/3<\beta<1$

Performance Measures

Given a convex function $f : [0, \infty) \to \mathbb{R}$ such that $f(1) = 0$. The f-divergence between two probabilities densities $p(x)$ and $q(x)$ over a probability space $\mathcal{X}$ is defined by

$$D_f(p, q) = \int_{\mathcal{X}} pf\left(\frac{p(x)}{q(x)}\right) \mathrm{d}x.$$

Two popular divergences such as the Kullback-Leibler (KL) divergence and the Hellinger distance (HD) are special cases of f-divergence by choosing f function $x\log(x)$ and $(\sqrt{x}-1)^2$, respectively. In our simulation study, we occasionally obtain the KL values as infinite when p is not absolutely continuous with respect to q, which means that there exists a measurable set A such that $p(A) \neq 0$ and $q(A) = 0$. Therefore, it is often more convenient to work with the HD which is a symmetric and non-negative distance. Moreover, the HD provides a lower bound for the KL divergence (see Tsybakov, 2004, page 73).

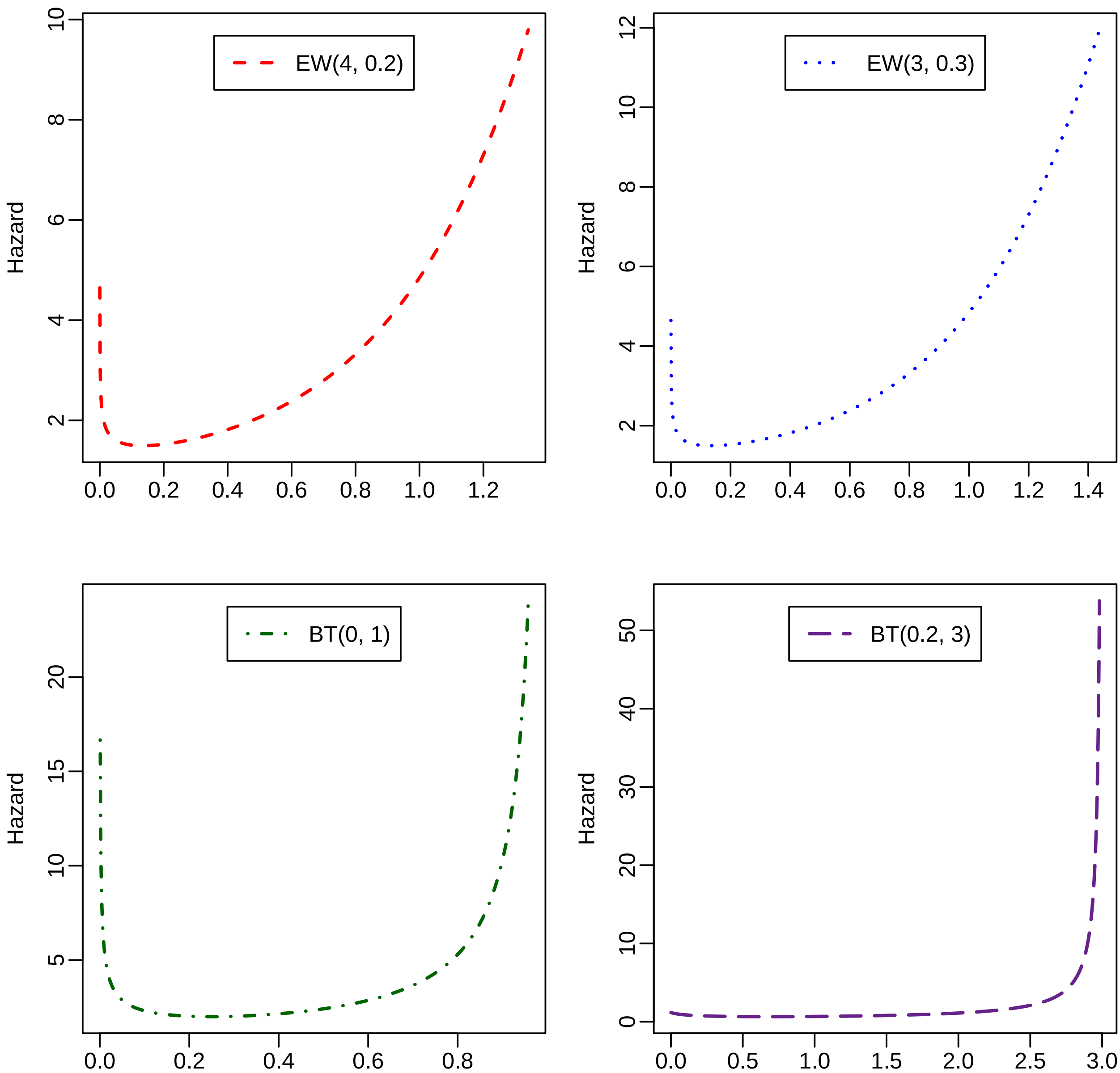

Figure 4.1: Hazard plots from the distributions used in the simulation study: exponentiated Weibull (top) and bathtub (bottom).

In order to assess the performance of a density estimator, two loss functions, namely, the integrated squared error (ISE) and HD are used, which are given by, respectively,

$$\mathrm{ISE}(f, \widehat{f}\,) = \int_{\mathbb{R}} \left\{ f(x) - \widehat{f}(x) \right\}^2 \, \mathrm{d}x,$$
$$\mathrm{HD}(f, \widehat{f}\,) = \int_{\mathbb{R}} \left\{ f(x)^{\frac{1}{2}} - \widehat{f}(x)^{\frac{1}{2}} \right\}^2 \, \mathrm{d}x,$$

where $\widehat{f}$ is an estimator of the true density f. The expectation of a loss function with regard to the true density f is then applied to evaluate the performance of an estimator. For each entry in the table, we compute the ISE and HD values respectively for each combination of estimator and replication. The mean integrated squared error (MISE) and mean Hellinger distance (MHD) are eventually estimated by the average of these ISE and HD values.

Results

For each distribution, the results of the simulation study based on 100 replications with sample sizes 100 and 1000 are summarized in Table 4.2. Each entry in the table is one of the empirical MISE and MHD values with their corresponding standard errors in parentheses. Also, the best expected loss value among density estimators for a given density is highlighted in boldface.

It can be seen from the results that the performance of the SCH estimator dominates the others, except in one case where the LCH is the best. Broadly speaking, the LCH and SCH estimators perform better than their LS and PS density estimators' competitors in terms of the MISE and MHD. For the case of the PS density estimator, this can be relatively attributed to the problem of bandwidth selection methods that involve different performance measures. Furthermore, one should also notice that the shape-constrained estimators obviously outperform their unconstrained counterparts in terms of both loss functions.

4.4.2 With Exact and Interval-censored Observations

Another simulation study is conducted to illustrate the performance of three different nonparametric density estimators for various ways of treating the censoring in the training set. Sometimes, to avoid the difficulty with dealing with interval-censored data, one can employ an imputation technique in order to reduce the problem of analysing interval-censored data to one of analysing exact observations along with right-censored data. For this purpose, we can replace an interval-censored observation with one point from that interval; e.g., its left endpoint or middle point.

In this simulation study, two different ways with which the data can be observed were considered: purely interval-censored data without any imputations and imputation approach for the analysis of interval-censored data where the two common choices of imputation, the left endpoint and midpoint of a censoring interval, were used. Each data set is generated from a two-component Weibull mixture with shape parameters $k = 0.8, 4$ and scale parameters $\lambda = 1, 10$, respectively, where the mixing proportion for the first component is $p = 0.1$. For the simulation study here, 100 random samples, each of size 500, were generated from the mentioned mixture model. In order to construct the model, the training sets can be obtained

Table 4.2: Simulation results for two distributions in terms of the MISE and MHD, with standard error given in parentheses.

Estimator	Density			
	EW$(4, 0.2)$	EW$(3, 0.3)$	BT$(0, 1)$	BT$(0.2, 3)$
$n = 100$				
	MISE			
PS	0.0888 (0.0028)	0.0517 (0.0019)	2.8408 (0.0142)	0.0465 (0.0012)
LS	0.1235 (0.0050)	0.0910 (0.0041)	2.7753 (0.0183)	0.0561 (0.0027)
LCH	0.0894 (0.0097)	0.0497 (0.0045)	1.8743 (0.0813)	0.0432 (0.0042)
SCH	**0.0768** (0.0087)	**0.0439** (0.0043)	**1.8400** (0.0317)	**0.0338** (0.0030)
	MHD			
PS	0.0222 (0.0008)	0.0184 (0.0007)	0.0593 (0.0015)	0.0236 (0.0007)
LS	0.0291 (0.0012)	0.0278 (0.0012)	0.0525 (0.0014)	0.0322 (0.0014)
LCH	0.0162 (0.0011)	0.0146 (0.0011)	0.0250 (0.0009)	0.0185 (0.0011)
SCH	**0.0145** (0.0011)	**0.0140** (0.0013)	**0.0228** (0.0008)	**0.0162** (0.0010)
$n = 1000$				
	MISE			
PS	0.0562 (0.0009)	0.0271 (0.0006)	2.6753 (0.0057)	0.0284 (0.0004)
LS	0.0150 (0.0005)	0.0105 (0.0004)	2.4905 (0.0252)	0.0077 (0.0002)
LCH	0.0103 (0.0007)	0.0052 (0.0003)	1.3491 (0.0211)	0.0056 (0.0004)
SCH	**0.0089** (0.0007)	**0.0046** (0.0003)	**1.3461** (0.0157)	**0.0045** (0.0003)
	MHD			
PS	0.0104 (0.0002)	0.0073 (0.0002)	0.0304 (0.0003)	0.0118 (0.0002)
LS	0.0029 (0.0004)	0.0030 (0.0001)	0.0229 (0.0006)	0.0042 (0.0004)
LCH	0.0019 (0.0004)	0.0015 (0.0001)	**0.0044** (0.0004)	0.0031 (0.0004)
SCH	**0.0017** (0.0004)	**0.0014** (0.0001)	0.0048 (0.0004)	**0.0025** (0.0004)

by rounding generated real values to their nearest integer values as in human mortality data. However, to evaluate the performance of each approach more precisely, we use a test set that directly uses the generated real values without rounding.

The results of the simulation study are given in Table 4.3, which includes the case of left endpoint imputation (in the top panel), midpoint imputation (in the middle panel), and the case of purely interval-censored data without any imputations (in the bottom panel). From the results, we can draw the following conclusions: First, it is clear that the best way of treating this kind of simulated data is the case of purely interval-censored data. The SCH estimator is superior to its competitors in terms of both loss functions in all cases studied.

Second, the SCH has an overall better performance than the LCH, in terms of the MISE and MHD, suggesting that the introduction of smoothness is indeed helpful.

Table 4.3: Simulation results for a mixture of two Weibull distributions. Standard errors are included in parentheses.

Method	MISE	MHD
	Left Endpoint	
LS	0.00806 (0.00061)	0.02584 (0.00134)
LCH	0.04660 (0.00199)	0.05148 (0.00112)
SCH	0.03208 (0.00162)	0.04095 (0.00105)
	Midpoint	
LS	0.00833 (0.00055)	0.02149 (0.00110)
LCH	0.00140 (0.00005)	0.00568 (0.00021)
SCH	0.00119 (0.00003)	0.00465 (0.00014)
	Interval-censored	
LS	0.02240 (0.00700)	0.04023 (0.01032)
LCH	0.01005 (0.00142)	0.01524 (0.00139)
SCH	**0.00108** (0.00005)	**0.00461** (0.00019)

4.4.3 When Hazard May Become Zero

Setup

We performed the third simulation study to emphasize that our approach also works when hazard, as well as its estimate, can become zero. The failure times were generated from the density $f(t) = (1-t)^4 \exp\{-(t-1)^5/5 - 1/5\}$ with the convex hazard function $h(t) = (1-t)^4$. We employed the inverse transform method to draw a sample with size 500. Note that the hazard is 0 at $t = 1$.

The left panel of Figure 4.2 displays five nonparametric hazard estimates, whereas the right panel of Figure 4.2 shows nonparametric density estimates. Although, the PS, LS, and the Müller-Wang kernel (MWK) estimates roughly correspond to the true hazard function until failure time $t = 2.5$, they perform poorly in those regions near corners or edges. Instead, the SCH and LCH estimators gave better performance than the other estimators near the edges and also lied closely to the true one.

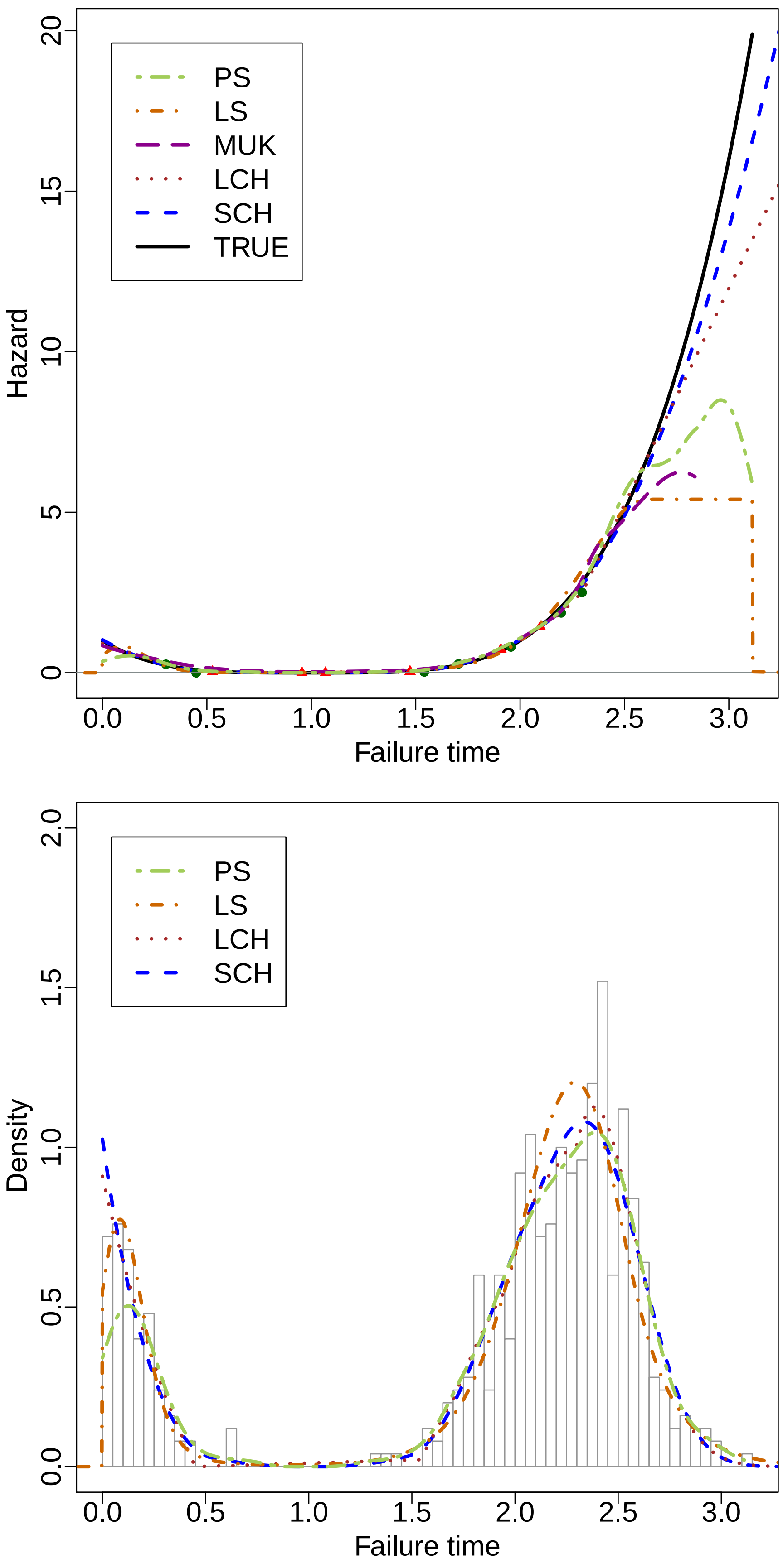

Figure 4.2: Comparisons of hazard and density estimates for the simulated data. The solid dots and triangles represent the support points of the non-smoothed and smoothed NPMLE of a bathtub-shaped hazard function, respectively.

4.5 Real Data

In order to make a further investigation into the performance of different smooth and non-smooth nonparametric estimators of a hazard function, we consider two real examples in this section. In particular, we estimate the hazard and density functions from the kidney transplant data in Section 4.5.1, and the estimation of the hazard and density functions for the Canadian mortality lifetime data set for the year 2008 is given in Section 4.5.2.

The first real data set is a case mixed with exact and right-censored observations. We compare the performance of three nonparametric estimators: the PS, LCH, and SCH. The presmoothed density estimator used the plug-in bandwidth selection method introduced by Cao and López-de Ullibarri (2007) and Jácome et al. (2008). We also intended to include the result produced by the LS approach for density estimation, but had to give up due to the failure of the R package **muhaz** (with an error message "no convergence") for heavily censored samples. Also, the Müller-Wang adaptive kernel estimator (Müller and Wang, 1994) based on the local bandwidth selection algorithm and the boundary kernel formulations is implemented and available in R package **muhaz** (Hess and Gentleman, 2010). However, this implementation only provides hazard values estimated at the prespecified points that are used for computing the estimate, not those at any new testing points, nor any estimated density values. Thus, we can not include it in the performance comparison.

For the second real data example, we treated the entire life table data in three different ways, namely left endpoint imputation, midpoint imputation, and interval-censored data without any imputation. We compared the performance of three nonparametric models: LS, LCH, and SCH. We had to quit the execution of the PS density estimator due to extremely high computational costs.

In the empirical studies based on real-world data, we do not have the knowledge of the true underlying density function, and thus we can not exactly utilize the loss functions given in Section 4.4.1. As a substitute, we replace the true density f with the empirical probability mass function $\widehat{f}_n$ based on a test set of size n. Two loss functions, the ISE and KL, are computed below; these are given for the case of exact data and interval-censored data by, respectively,

$$\mathrm{ISE}(\widehat{f}_n, \widehat{f}\,) = \int_{\mathbb{R}} \left\{\widehat{f}(x)\right\}^2 \mathrm{d}x - \frac{2}{n}\sum_{i=1}^{n} \widehat{f}(x_i),$$

$$\mathrm{KL}(\widehat{f}_n, \widehat{f}\,) = -\frac{1}{n}\sum_{i=1}^{n} \log\left\{\widehat{f}(x_i)\right\},$$

and

$$\begin{aligned}
\mathrm{ISE}(\widehat{f}_n, \widehat{f}\,) &= \int_{\mathbb{R}} \left\{ \widehat{f}(x) \right\}^2 \mathrm{d}x - 2 \int_{\mathbb{R}} \widehat{f}_n(x) \widehat{f}(x)\, \mathrm{d}x \\
&= \int_{\mathbb{R}} \left\{ \widehat{f}(x) \right\}^2 \mathrm{d}x - 2 \int_{\mathbb{R}} \widehat{f}(x)\, \frac{1}{n} \sum_{i=1}^{n} \frac{\mathbf{1}(L_i \,\le\, x < R_i)}{R_i - L_i}\, \mathrm{d}x \\
&= \int_{\mathbb{R}} \left\{ \widehat{f}(x) \right\}^2 \mathrm{d}x - \frac{2}{n} \sum_{i=1}^{n} \frac{\widehat{F}(R_i) - \widehat{F}(L_i)}{R_i - L_i}, \\
\mathrm{KL}(\widehat{f}_n, \widehat{f}\,) &= -\frac{1}{n} \sum_{i=1}^{n} \log \left\{ \widehat{F}(R_i) - \widehat{F}(L_i) \right\},
\end{aligned}$$

where $\widehat{f}_n$ denotes the empirical mass function from a test set of size n and $\widehat{f}$ the density estimate is obtained from a training set. In addition, additive constants are excluded from the above formulas.

4.5.1 Kidney Transplant Data

The first data set described in Klein and Moeschberger (2003) in Section 1.7 contains the times to death of 863 kidney transplant patients treated at the Ohio State University Transplant Center between 1982 and 1992. The observations were considered as right-censored data when the patients moved away from Columbus or were still alive at the end of the study. There were 140 exact and 723 right-censored observations. Klein and Moeschberger (2003) showed the effects of varying bandwidths on the smooth estimation of the hazard rate function.

Four nonparametric hazard function estimates are plotted in Figure 4.3. Visual inspection of Figure 4.3 indicates that there is an early high risk of death in the first year after transplant. After around six years, our estimators show that the estimated hazard rate increases, whereas their unconstrained contenders reveal that the risk of death decreases.

To assess the performance, we ran 10-fold cross-validation, with results produced by averaging over 10 replications. Table 4.4 provides a brief summary of the estimation results for the three nonparametric approaches. Overall, the SCH estimator appears to provide the best fit to this data set by having the smallest values in terms of both criteria: MISE and MKL. It seems the shape-constrained estimators are more accurate and outperform the unconstrained one for estimation of the hazard function. The PS density estimator gives apparently an improper estimate.

Table 4.4: Cross-validation results for the three nonparametric density estimators for kidney transplant data. Standard errors are given in parentheses.

Method	MKL	MISE
PS	4.5045 (0.0302)	−0.0568 (0.0002)
LCH	4.0684 (0.0081)	−0.0920 (0.0007)
SCH	**4.0669** (0.0078)	**-0.0933** (0.0009)

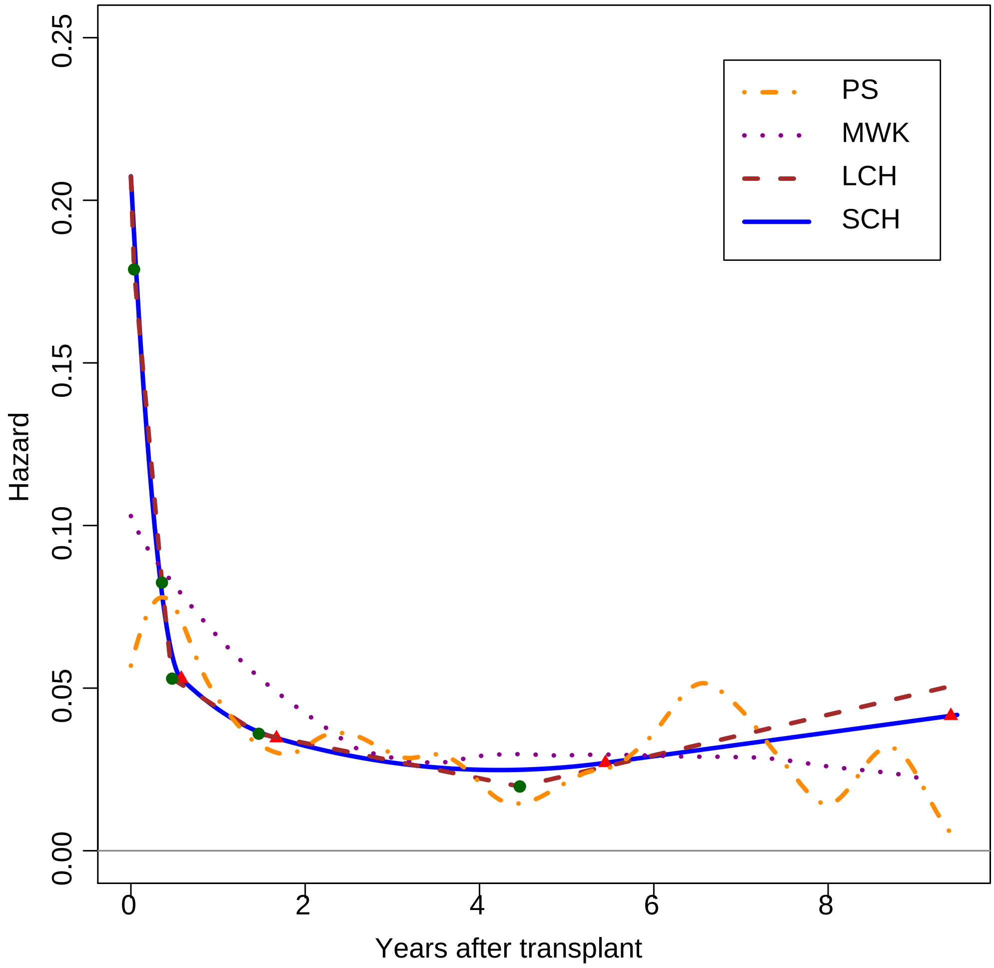

Figure 4.3: Hazard rate function estimates for the kidney transplant data set, with support points shown as solid points and triangles for the non-smoothed and smoothed NPMLE of a hazard function, respectively.

4.5.2 Canadian Lifetime Data

The Canadian mortality table for year 2008 which was utilized in Example 3.5.1 is considered again. Indeed, we used this data set to highlight the comparison of different ways of treating censoring. In particular, three cases were investigated: (i) left endpoint imputation, where the three nonparametric density estimators were applied for the left endpoint of the time intervals, (ii) midpoint imputation, where the midpoint of the time intervals was used, and (iii) integers in years were interpreted as purely interval-censored data as they should be.

We used the holdout method for evaluation of our procedures since we have a large data set ($n = 238612$). Our data was randomly partitioned into two independent sets: a training set (2/3) for model construction and a test set (1/3) for accuracy assessment. In this study, we applied interval-censored data for evaluation. The top part of Table 4.5 reports the results of comparison of density estimates in terms of the KL and ISE for the case of left endpoint data; the results of density estimates for the cases of midpoint imputation and interval-censored data are shown in the middle and bottom parts, respectively. Of the three nonparametric density estimators, the SCH estimator in the case of purely interval-censored data is superior to the others in terms of the KL. It also works better in terms of the ISE than the other two approaches in the case of midpoint imputation, while the differences between the SCH and LCH in both midpoint imputation and no imputation cases are reasonably small.

Four nonparametric hazard function estimates are depicted in the left panel of Figure 4.4, along with bar plots for the observed hazards obtained directly from the data. The right panel of Figure 4.4 shows their density estimates, together with standardized histograms for the data. As shown in Figure 4.4, imputation with left endpoint is somehow inappropriate, especially in the infant mortality phase. In contrast, the midpoint imputation and interval-censored cases work reasonably well in the early mortality phase. As can be seen in the subplots of Figure 4.4, the best way for modelling the infant mortality data is perhaps treating them as purely interval-censored data as they should be.

Table 4.5: Summary of three nonparametric density estimation results for the Canadian mortality table.

Method	KL	ISE
	Left Endpoint	
LS	4.08064	-0.02096
LCH	4.07232	-0.02097
SCH	4.07225	-0.02094
	Midpoint	
LS	4.08037	-0.02098
LCH	4.07172	-0.02099
SCH	4.07208	**-0.02101**
	Interval-censored	
LS	4.08540	-0.02092
LCH	4.07175	-0.02100
SCH	**4.07169**	-0.02099

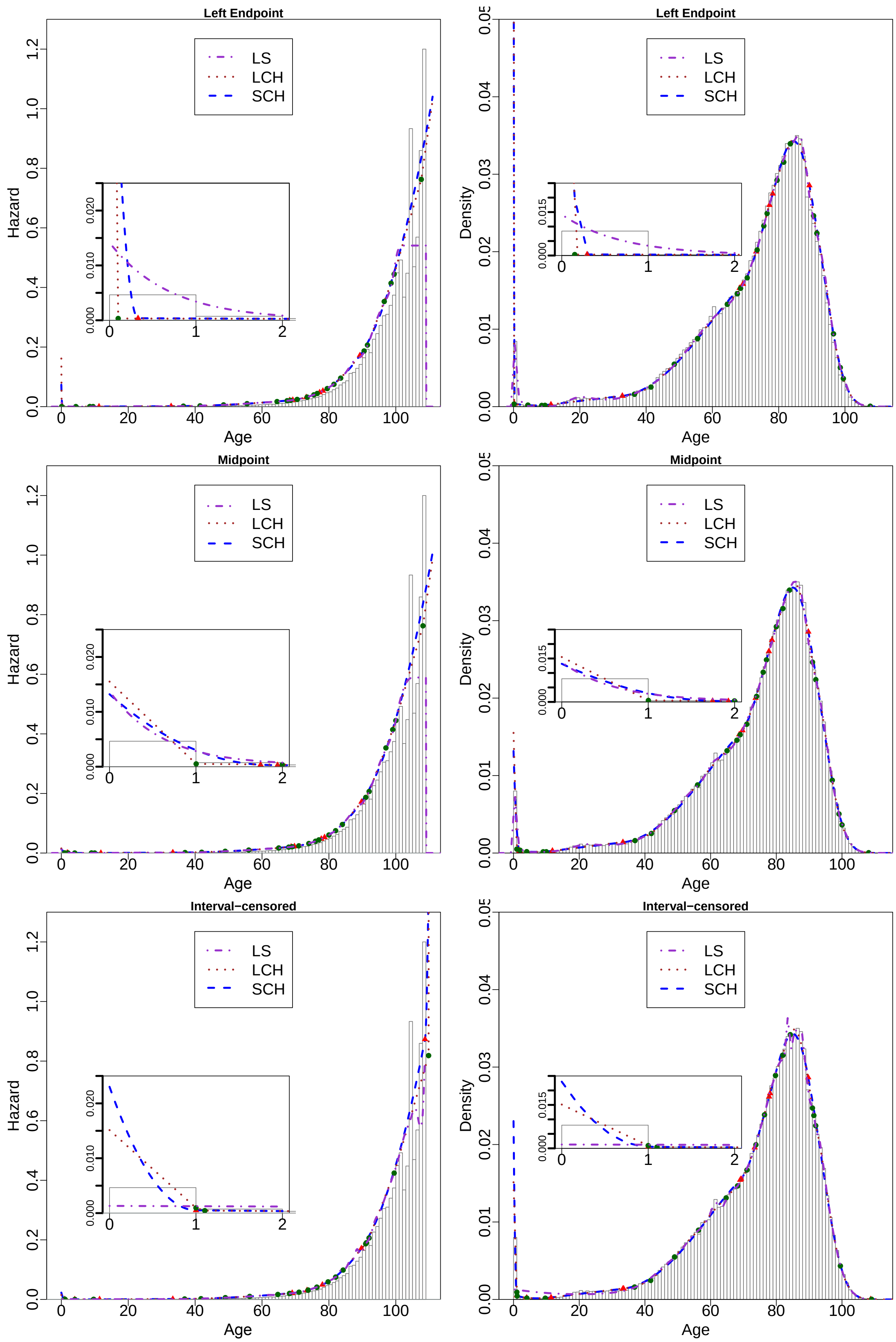

Figure 4.4: Smoothed hazard rate and density function estimates for the Canadian mortality data set for the year 2008. The solid points and triangle indicate the support points of the non-smoothed and smoothed NPMLE of a hazard function, respectively.

4.6 Summary

In this chapter, we study the smooth nonparametric estimation of a hazard function under convexity shape constraint. Through simulation studies and two real-world data sets, we have investigated and compared the performance of our proposed approaches with that of the logspline and presmoothed kernel-based approaches. As seen in this chapter, the empirical results indicate that the smooth shape-restricted estimator provides a better fit to human mortality data than the piecewise linear continuous estimator and the logspline estimator, particularly in the estimation of infant mortality rates. We also generally achieve a better graphical representation of a hazard function estimation through the use of shape restriction and smoothness assumption than the unconstrained nonparametric estimators. Moreover, numerical studies suggest that the shape-restricted estimators generally outperform their unconstrained competitors.

It is worth pointing out that our method also works very well with a small sample size or heavily censored data, whereas the logspline technique may suffer from convergence problems in these situations. In addition, some issues of the presmoothed kernel-based method can be partially associated with their bandwidth selection methods. In contrast, our nonparametric hazard estimation is void of the bandwidth selection problems that are associated with the presmoothed kernel method, by imposing a convex shape restriction on the true underlying hazard function.

Chapter 5

Nonparametric Estimation of a Convex Baseline Hazard Function in the Cox Proportional Hazards Model

5.1 Introduction

In failure time data analysis, the subjects under study might have some additional characteristics that can effect the failure times. Therefore, one is often tempted to analyze how different features of subjects such as age, gender, smoking history, physical activity level, heart rate, or treatment indicator influence the distribution of the event of interest. These features are generally referred to as covariates, which are time-independent and usually recorded at the start of the study. The most renowned method of investigating the effects of covariates on lifetime distribution is the Cox proportional hazards model (Cox, 1972). Its appeal resides mainly in its framework that enables us to efficiently estimate the regression coefficients; whereas, the baseline distribution is completely unknown (Efron, 1977). The proposed estimator was shown by Cox (1975) to be a maximum partial likelihood, and its asymptotic properties were extensively studied by Tsiatis (1981). Breslow (1972) also proposed a different approach that yields the same maximum partial likelihood estimator along with an estimator of the baseline cumulative hazard function.

For the case of right-censored data, the crucial feature of the partial likelihood method is that one can yield the estimates of the regression coefficients without involving the estimation of the underlying baseline function; see e.g., Kalbfleisch and Prentice (2002) and Klein and Moeschberger (2003). In contrast, the estimates of the regression coefficients and the derivation of its asymptotic properties for the case of interval-censored data is a more challenging issue to be investigated due to the fact that we are not able to remove the baseline hazard function. Regression analysis for interval-censored data was first studied by Finkelstein (1986), who considered a parametric method for approximating the baseline hazard distribution together with regression coefficients by maximizing the full likelihood of the observed data. Pan (1999) achieved the NPMLE of covariates along with that of the piecewise constant baseline survival function by extending the ICM algorithm. Others who discussed the proportional hazards model in this case include Huang and Wellner (1997), Pan (2000a), Betensky et al. (2002), and Cai and Betensky (2003).

One of the important problems in survival analysis is to predict the distribution of the time to some event of interest from a set of covariates, although the baseline hazard function is completely unspecified. In practice, there is some additional information about the baseline distribution function which is available from prior studies. Therefore, it is often reasonable to assume a shape for it. Regression analysis of different types of censoring data have been studied in numerous literature, while the nonparametric estimates of the baseline distribution function subject to shape constraints is relatively sparse. Chung and Chang (1994) suggested a maximum likelihood estimator of a nondecreasing baseline hazard function in the Cox proportional hazards model by imposing the assumption that all censoring times were equal to their former observed failure times. Recently, Lopuhaä and Nane (2013) proposed a nonparametric maximum likelihood estimator and a Grenander-type estimator for estimating a monotone baseline hazard function and a decreasing baseline density within the Cox model from right-censored observations.

In this chapter, our aims are two-fold: to derive the nonparametric estimator of a convex baseline hazard function based on the maximum likelihood method for right-censored data, interval-censored data, and the situation with both exact and censored data. Moreover, we propose a hybrid algorithm for simultaneously computing the NPMLE of a convex baseline hazard function and the estimates of the effects of covariates in the Cox model. The shaped-constrained estimators are shown to have similar results compared to the traditional Cox partial likelihood (PL) estimator and the fully semiparametric maximum likelihood estimator (Pan, 1999) using some real and simulated data set in the case of right-censored and interval-censored data, respectively. Our proposed algorithm directly employs the CNMCH or CNMSCH algorithm combined with only an optimization method that can cope with lower bound constraints, e.g., limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS).

5.2 Nonparametric Estimation of a Convex Baseline Hazard in the Cox Model

Consider the Cox proportional hazards model, in which the conditional hazard of the event time T with covariate vector $\boldsymbol{Z} \in \mathbb{R}^p$ is proportional to the baseline hazard. In terms of the hazard or cumulative hazard functions, the proportional hazards model is given by

$$h(t \mid \boldsymbol{Z}) = h_0(t)\exp(\boldsymbol{\beta}^\top \boldsymbol{Z}) \quad \text{or} \quad H(t \mid \boldsymbol{Z}) = H_0(t)\exp(\boldsymbol{\beta}^\top \boldsymbol{Z}),$$

where $h_0(t)$ and $H_0(t)$ represent the unknown baseline hazard and baseline cumulative hazard functions, respectively, which are assumed to be continuous and $\boldsymbol{\beta}$ denotes the vector of unknown regression coefficients.

5.2.1 Maximum Likelihood Estimation of a Proportional Hazards Model

Let the observed data consist of i.i.d. sample $\{(L_i, R_i], \boldsymbol{Z}_i; i = 1, \cdots, n\}$, where $(L_i, R_i]$ denotes the censoring interval within which the event time T for the ith subject is observed to occur and $\boldsymbol{Z}_i$ is a p-dimensional vector of covariates of subject i. As usual we assume that T_i and $(L_i, R_i]$ are conditionally independent given $\boldsymbol{Z} = z$. An exact failure time T_i is obtained, if $L_i = R_i$. Then, the simplified log-likelihood function can be written as

$$\begin{aligned} \ell(h_0, \boldsymbol{\beta}) = &\sum_{i=1}^{n_1} \left\{ -\exp\left(\boldsymbol{\beta}^\top \boldsymbol{Z}_i\right) H_0(T_i) + \log h_0(T_i) + \boldsymbol{\beta}^\top \boldsymbol{Z}_i \right\} \\ &+ \sum_{i=n_1+1}^{n} \log\left\{ [S_0(L_i)]^{\exp(\boldsymbol{\beta}^\top \boldsymbol{Z}_i)} - [S_0(R_i)]^{\exp(\boldsymbol{\beta}^\top \boldsymbol{Z}_i)} \right\}, \end{aligned} \tag{5.1}$$

where $S_0(t)$ correspond to the baseline survival function. Since $h_0(T_{(n)})$ can become arbitrarily large, Grenander (1956) suggested that $\ell(h_0, \boldsymbol{\beta})$ can be maximised over a nondecreasing h_0 which is bounded by some M. Thus, to handle such situations, one can simply use the following modified log-likelihood function

$$\ell(h_0, \boldsymbol{\beta}) = -\sum_{i=1}^{n_1} \exp(\boldsymbol{\beta}^\top \boldsymbol{Z}_i) H_0(T_i) + \sum_{\substack{i=1\\T_i \neq T_{\max}}}^{n_1} \left\{\log h_0(T_i) + \boldsymbol{\beta}^\top \boldsymbol{Z}_i\right\}$$
$$+ \sum_{i=n_1+1}^{n} \log\left\{[S_0(L_i)]^{\exp(\boldsymbol{\beta}^\top \boldsymbol{Z}_i)} - [S_0(R_i)]^{\exp(\boldsymbol{\beta}^\top Z_i)}\right\}, \tag{5.2}$$

where $T_{\max}$ is the maximum of exact observations and finite censoring values. When $T_{(n_1)} = T_{\max}$, the full NPMLE is then achieved by setting $\hat{h}_0(T_{(n_1)}) = \infty$ for all $T \geq T_{(n_1)}$.

The NPMLE of a convex baseline hazard function can be expressed as a piecewise linear (3.3) or quadratic (4.3) function. Therefore, h_0 can be written in the general form as follows:

$$h_0(t;q) = \sum_{i=0}^{q-1} \alpha_i\, t^i + \sum_{j=1}^{k} \nu_j (\tau_j - t)_+^q + \sum_{j=1}^{m} \mu_j (t-\eta_j)_+^q, \quad \text{for } q = 1, 2.$$

One can also add the term $\alpha_2 t^2$ in order to facilitate the computational procedure for nonparametric estimation of a convex baseline hazard function under smoothness assumption. A piecewise linear or quadratic convex hazard function is supported at $\boldsymbol{\theta} = (\boldsymbol{\tau}, \boldsymbol{\eta})^\top$, where $\boldsymbol{\tau} = (\tau_1, \cdots, \tau_k)$ and $\boldsymbol{\eta} = (\eta_1, \cdots, \eta_m)$. The support point vector $\boldsymbol{\theta}$ has a corresponding mass vector $\boldsymbol{\pi}$, where $\boldsymbol{\pi} = (\alpha_0, \boldsymbol{\nu}, \boldsymbol{\mu})^\top$ is a positive mass vector in the case of piecewise linear and $\boldsymbol{\pi} = (\alpha_0, \alpha_1, \boldsymbol{\nu}, \boldsymbol{\mu})^\top$ is the corresponding positive mass vector except the first two elements which can be real numbers in the case of piecewise quadratic.

Of critical importance for computing a nonparametric MLE is the gradient function, in particular, for locating new support points. Let us define $e_{1,\tau} = (\tau - t)_+^q$ and $e_{2,\eta} = (t-\eta)_+^q$, $q = 1, 2$. The two gradient functions are defined as, respectively,

$$\begin{aligned} d_1(\tau; h_0, \boldsymbol{\beta}) &= \lim_{\epsilon \to 0} \frac{\tilde{\ell}(h_0 + \epsilon e_{1,\tau}) - \tilde{\ell}(h_0)}{\epsilon} \\ &= \sum_{\substack{i=1\\T_i \neq T_{\max}}}^{n_1} \frac{(\tau - T_i)_+^q}{h_0(T_i)} - \sum_{i=1}^{n_1} \exp(\boldsymbol{\beta}^\top \boldsymbol{Z}_i) \left(\frac{\tau^{q+1} - (\tau - T_i)_+^{q+1}}{q+1}\right) + \\ &\quad \sum_{i=n_1+1}^{n} - \exp(\boldsymbol{\beta}^\top \boldsymbol{Z}_i) \left(\frac{\tau^{q+1} - (\tau - L_i)_+^{q+1}}{q+1}\right) - \\ &\quad \sum_{i=n_1+1}^{n} \exp(\boldsymbol{\beta}^\top \boldsymbol{Z}_i) \left(\frac{(\tau - R_i)_+^{q+1} - (\tau - L_i)_+^{q+1}}{q+1}\right) \Delta_i(H_0), \\ &\quad \text{for } 0 < \tau \leq \eta_1, \end{aligned} \tag{5.3}$$

and

$$
\begin{aligned}
d_2(\eta; h_0, \boldsymbol{\beta}) &= \lim_{\epsilon \to 0} \frac{\tilde{\ell}(h_0 + \epsilon e_{2,\eta}) - \tilde{\ell}(h_0)}{\epsilon} \\
&= \sum_{\substack{i=1 \\ T_i \neq T_{\max}}}^{n_1} \frac{(T_i - \eta)_+^q}{h_0(T_i)} - \sum_{i=1}^{n_1} \exp\left(\boldsymbol{\beta}^\top \boldsymbol{Z}_i\right) \frac{(T_i - \eta)_+^{q+1}}{q+1} \quad + \\
&\sum_{i=n_1+1}^{n} \left\{ - \exp\left(\boldsymbol{\beta}^\top \boldsymbol{Z}_i\right) \frac{(L_i - \eta)_+^{q+1}}{q+1} - \exp\left(\boldsymbol{\beta}^\top \boldsymbol{Z}_i\right) \left(\frac{(L_i - \eta)_+^{q+1} - (R_i - \eta)_+^{q+1}}{q+1} \right) \Delta_i(H_0) \right\}, \\
&\text{for } \tau_k \ \leq \ \eta < T_{\max},
\end{aligned} \tag{5.4}
$$

where

$$
\Delta_i(H_0) = \frac{\left(\exp(H_0(L_i) - H_0(R_i))\right)^{\exp\left(\boldsymbol{\beta}^\top \boldsymbol{Z}_i\right)}}{\left(1 - \left(\exp(H_0(L_i) - H_0(R_i))\right)^{\exp\left(\boldsymbol{\beta}^\top \boldsymbol{Z}_i\right)}\right)}.
$$

Note that in the case of piecewise linear convex baseline hazard function estimator, d_1 and d_2 are piecewise quadratic functions of τ and η, while they are cubic functions of τ and η in the situation of piecewise quadratic estimator.

5.2.2 Computation

We first discuss some computational characteristics of the convex baseline hazard function and then present a new algorithm for simultaneously computing the nonparametric estimate of a baseline hazard function under convexity constraint and the regression coefficients in proportional hazard models.

Recently, three general algorithms for computing the semiparametric maximum likelihood estimates (SPMLEs) of parameters in semiparametric mixtures were proposed by Wang (2010). All of the algorithms combine the CNM with an unconstrained optimization algorithm; e.g., a quasi-Newton method. According to the results of experimental studies, the algorithm based on modifying the support set of the mixing distribution has superior performance compared to the others. This premier algorithm, called CNM-MS, is a hybrid between the CNM method and the standard BFGS method. Our new algorithm is inherently similar to the CNM-MS, as it comprises two alternative main steps in each iteration: (i) apply one iteration of the CNMCH or CNMSCH to provide an efficient update of the support points and masses $(\boldsymbol{\pi}, \boldsymbol{\theta})$, (ii) update all parameters $(\boldsymbol{\pi}, \boldsymbol{\theta}, \boldsymbol{\beta})$ using the BFGS method with lower bound constraints.

Given $\boldsymbol{\pi}$, we can easily compute a new estimate $\boldsymbol{\pi}'$ by employing the second-order Taylor series expansion of the modified log-likelihood function in the neighbourhood of $\boldsymbol{\pi}$. Let us denote

$$
\begin{aligned}
\mathbf{g} \equiv \mathbf{g}(\boldsymbol{\pi}, \boldsymbol{\theta}) &= \frac{\partial \tilde{\ell}}{\partial \boldsymbol{\pi}}, \\
\mathbf{H} \equiv \mathbf{H}(\boldsymbol{\pi}, \boldsymbol{\theta}) &= \frac{\partial^2 \tilde{\ell}}{\partial \boldsymbol{\pi} \partial \boldsymbol{\pi}^\top},
\end{aligned}
$$

(see Section 5.2.3). Thus, the quadratic approximation is given by

$$
\tilde{\ell}(\boldsymbol{\pi}', \boldsymbol{\theta}) - \tilde{\ell}(\boldsymbol{\pi}, \boldsymbol{\theta}) \approx \mathbf{g}^\top \Delta\boldsymbol{\pi} + \frac{1}{2} \Delta\boldsymbol{\pi}^\top \mathbf{H} \Delta\boldsymbol{\pi}. \tag{5.5}
$$

where $\Delta\boldsymbol{\pi} = \boldsymbol{\pi}' - \boldsymbol{\pi}$ indicates a direction away from $\boldsymbol{\pi}$. Let $\mathbf{R}$ be the upper triangular Cholesky factor of $-\mathbf{H}$ (see Section 5.2.4) and $\mathbf{b}_1$ the solution of the lower triangular system $\mathbf{R}^\top \mathbf{b}_1 = \mathbf{g}$. Let $\|\cdot\|$ denote the L_2-norm. In order to maximise $\tilde{\ell}(\boldsymbol{\pi}, \boldsymbol{\theta})$ over $\boldsymbol{\pi}$ with $\boldsymbol{\theta}$ fixed, one can repeatedly solve the following least squares linear regression problem:

$$
\text{minimize} \quad \|\mathbf{R}\boldsymbol{\pi}' - (\mathbf{R}\boldsymbol{\pi} + \mathbf{b}_1)\|^2, \qquad \text{subject to} \quad \boldsymbol{\pi}' \geq \mathbf{0}, \tag{5.6}
$$

if we assume a piecewise linear baseline hazard function, while maximising $\tilde{\ell}(\boldsymbol{\pi}, \boldsymbol{\theta})$ in the neighbourhood of $\boldsymbol{\pi}$ can be approximated by solving the least squares linear regression problem with non-negativity constraints:

$$
\text{minimize} \quad \left\| \mathbf{R} \begin{pmatrix} \boldsymbol{\pi}'_1 \\ \boldsymbol{\pi}'_2 \end{pmatrix} - \mathbf{R} \begin{pmatrix} \boldsymbol{\pi}_1 \\ \boldsymbol{\pi}_2 \end{pmatrix} - \mathbf{b}_1 \right\|^2, \quad \text{subject to} \quad \boldsymbol{\pi}'_1 \in \mathbb{R}^2, \boldsymbol{\pi}'_2 \in \mathbb{R}_+^{1+m+k} \tag{5.7}
$$

if we assume a piecewise quadratic one, where $\boldsymbol{\pi}$ partitioned into two vectors $\boldsymbol{\pi}_1 = (\alpha_0, \alpha_1)$ and $\boldsymbol{\pi}_2 = (\nu, \mu)$. Both least square problems (5.6) and (5.7) can be solved by the NNLS algorithm of Lawson and Hanson (1974).

The gradient functions $d_1(\tau; h_0, \boldsymbol{\beta})$ (5.3) and $d_2(\eta; h_0, \boldsymbol{\beta})$ (5.4) are utilized to obtain new candidate support points. Specifically, in each iteration the two support point vectors τ and η corresponding to the decreasing and increasing parts of a convex baseline hazard function are expanded by including all local maxima of the gradient functions. For the constant hazard interval $[\tau_k, \eta_1]$, the global maximum of the gradient functions is also found and then added to the support point vector to which it corresponds.

The extensions of CNM-MS for computing a convex baseline hazard function without or with smoothness assumption is called CNMCH-MS and CNMSCH-MS, respectively. The resulting algorithms can be formally described as follows.

Algorithm 5.1 (CNMCH-MS/CNMSCH-MS)**.** Set $s = 0$, and choose $\boldsymbol{\beta}_0 \in \mathbb{R}^p$ and h_0 with finite support such that $\ell(h_0, \boldsymbol{\beta}_0) > -\infty$. Repeat the following steps.

Step 1: update $(\boldsymbol{\pi}_s, \boldsymbol{\theta}_s)$ to $(\boldsymbol{\pi}_s^+, \boldsymbol{\theta}_s^+)$ with $\boldsymbol{\beta} = \boldsymbol{\beta}_s$ fixed, by using one iteration of the CNMCH or CNMSCH method.

Step 2: update $(\boldsymbol{\pi}_s^+, \boldsymbol{\theta}_s^+, \boldsymbol{\beta}_s)$ to a local maximum $(\boldsymbol{\pi}_{s+1}, \boldsymbol{\theta}_{s+1}, \boldsymbol{\beta}_{s+1})$, by using the BFGS method with lower bound constraints, followed by a line search.

Step 3: set $s = s + 1$. If converged, stop.

Note that applying one iteration of the CNMCH or CNMSCH in the first step of the algorithm certifies that: (i) the support point vector $\boldsymbol{\theta}$ is continuously updated with expanding it by including new potential support points and then probably contracting it by discarding those with zero entries in $\boldsymbol{\pi}$, (ii) ${h_0}_s^+$ containing of $(\boldsymbol{\pi}_s^+, \boldsymbol{\theta}_s^+)$ is not far from being the NPMLE of $h_{0\boldsymbol{\beta}_s}$; and (iii) the log-likelihood value is definitely increased from $(\boldsymbol{\pi}_s, \boldsymbol{\theta}_s, \boldsymbol{\beta}_s)$ to the provisional solution $(\boldsymbol{\pi}_s^+, \boldsymbol{\pi}_s^+, \boldsymbol{\beta}_s)$ since it has been followed by a line search.

In order to update all parameters in the second step of the CNMCH-MS or CNMSCH-MS algorithm, we applied the BFGS method with lower bound constraints only on $\boldsymbol{\pi}'$ and $\boldsymbol{\pi}'_2$ in the case of piecewise linear continuous and smooth baseline hazard function estimation, respectively. This method has been described by Nash (1979), which is available using the function Rvmmin in R package **Rvmmin** (Nash, 2011). A line search is then conducted to guarantee monotone increase and global convergence of the log-likelihood function. Subsequently, all zero-massed support points are removed after being found redundant by the bound-constrained optimization method. The above procedure is repeated until the solution has converged to a local maximum $(\boldsymbol{\pi}_{s+1}, \boldsymbol{\theta}_{s+1}, \boldsymbol{\beta}_{s+1})$. Convergence of the algorithm to a global maximum can be assured as the log-likelihood increases continually through computing each iteration of the algorithm.

5.2.3 Derivatives of the Modified Log-likelihood Function

The first derivatives of the modified log-likelihood function (5.2) are given by

$$\frac{\partial \tilde{\ell}}{\partial \alpha_0} = \sum_{\substack{i=1 \\ T_i \neq T_{\max}}}^{n_1} \left\{ \frac{1}{h(T_i)} \right\} - \sum_{i=1}^{n_1} \exp{(\boldsymbol{\beta}^\top \boldsymbol{Z}_i)} T_i +$$
$$\sum_{i=n_1+1}^{n} \left\{ -\exp{(\boldsymbol{\beta}^\top \boldsymbol{Z}_i)} L_i - \exp{(\boldsymbol{\beta}^\top \boldsymbol{Z}_i)} \left(L_i - R_i \right) \Delta_i(H_0) \right\},$$

$$
\begin{aligned}
\frac{\partial \tilde{\ell}}{\partial \alpha_1} = & \sum_{\substack{i=1\\ T_i \neq T_{\max}}}^{n_1} \left\{ \frac{T_i}{h_0(T_i)} \right\} - \sum_{i=1}^{n_1} \left\{ \exp\left(\boldsymbol{\beta}^\top \boldsymbol{Z}_i\right) \frac{T_i^2}{2} \right\} + \\
& \sum_{i=n_1+1}^{n} \left\{ - \exp\left(\boldsymbol{\beta}^\top \boldsymbol{Z}_i\right) \frac{L_i^2}{2} - \exp\left(\boldsymbol{\beta}^\top \boldsymbol{Z}_i\right) \left(\frac{(L_i^2 - R_i^2)}{2} \right) \Delta_i(H_0) \right\}, \\
\frac{\partial \tilde{\ell}}{\partial \nu_j} = & \sum_{\substack{i=1\\ T_i \neq T_{\max}}}^{n_1} \left\{ \frac{(\tau_j - T_i)_+^q}{h_0(T_i)} \right\} - \sum_{i=1}^{n_1} \left\{ \frac{\tau_j^{q+1} - (\tau_j - T_i)_+^{q+1}}{q+1} \right\} + \\
& \sum_{i=n_1+1}^{n} \left\{ - \exp\left(\boldsymbol{\beta}^\top \boldsymbol{Z}_i\right) \frac{\tau_j^{q+1} - (\tau_j - L_i)_+^{q+1}}{q+1} - \exp\left(\boldsymbol{\beta}^\top \boldsymbol{Z}_i\right) \left(\frac{(\tau_j - R_i)_+^{q+1} - (\tau_j - L_i)_+^{q+1}}{q+1} \right) \Delta_i(H_0) \right\}, \\
\frac{\partial \tilde{\ell}}{\partial \mu_j} = & \sum_{\substack{i=1\\ T_i \neq T_{\max}}}^{n_1} \left\{ \frac{(T_i - \eta_j)_+^q}{h(T_i)} \right\} - \sum_{i=1}^{n} \exp\left(\boldsymbol{\beta}^\top \boldsymbol{Z}_i\right) \left\{ \frac{(T_i - \eta_j)_+^{q+1}}{q+1} \right\} + \\
& \sum_{i=n_1+1}^{n} \left\{ - \exp\left(\boldsymbol{\beta}^\top \boldsymbol{Z}_i\right) \frac{(L_i - \eta_j)_+^{q+1}}{q+1} - \exp\left(\boldsymbol{\beta}^\top \boldsymbol{Z}_i\right) \left(\frac{(L_i - \eta_j)_+^{q+1} - (R_i - \eta_j)_+^{q+1}}{q+1} \right) \Delta_i(H_0) \right\},
\end{aligned}
$$

where $q = 1, 2$ in the case of piecewise linear and piecewise quadratic convex hazard function estimators, repectively, and

$$
\Delta_i(H_0) = \frac{\left(\exp(H_0(L_i) - H_0(R_i))\right)^{\exp\left(\boldsymbol{\beta}^\top \boldsymbol{Z}_i\right)}}{\left(1 - \left(\exp(H_0(L_i) - H_0(R_i))\right)^{\exp\left(\boldsymbol{\beta}^\top \boldsymbol{Z}_i\right)}\right)}.
$$

The second derivatives of the modified log-likelihood function that form the Hessian matrix $\mathbf{H}$ are given by

$$
\begin{aligned}
\frac{\partial^2 \tilde{\ell}}{\partial \alpha_0^2} &= \sum_{\substack{i=1\\ T_i \neq T_{\max}}}^{n_1} \left\{ -\frac{1}{h_0(T_i)^2} \right\} + \sum_{i=n_1+1}^{n} \left\{ -\exp\left(\boldsymbol{\beta}^\top \boldsymbol{Z}_i\right)(L_i - R_i)^2 \Delta_i(H_0) \right\}, \\
\frac{\partial^2 \tilde{\ell}}{\partial \alpha_1^2} &= \sum_{\substack{i=1\\ T_i \neq T_{\max}}}^{n_1} \left\{ -\frac{T_i^2}{h_0(T_i)^2} \right\} + \sum_{i=n_1+1}^{n} \left\{ -\exp\left(\boldsymbol{\beta}^\top \boldsymbol{Z}_i\right) \left(\frac{L_i^2 - R_i^2}{2} \right)^2 \Delta_i(H_0) \right\},
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 \tilde{\ell}}{\partial \mu_j^2} &= \sum_{\substack{i=1\\ T_i \neq T_{\max}}}^{n_1} \left\{ -\frac{(\tau_j - T_i)_+^q}{h_0(T_i)^2} \right\} + \\
&\quad \sum_{i=n_1+1}^{n} \left\{ -\exp\left(\boldsymbol{\beta}^\top \boldsymbol{Z}_i\right) \left(\frac{(\tau_j - R_i)_+^w - (\tau_j - L_i)_+^w}{w} \right)^2 \Delta_i(H_0) \right\}, \\
\frac{\partial^2 \tilde{\ell}}{\partial \nu_j^2} &= \sum_{\substack{i=1\\ T_i \neq T_{\max}}}^{n_1} \left\{ -\frac{(T_i - \eta_j)_+^q}{h_0(T_i)^2} \right\} + \\
&\quad \sum_{i=n_1+1}^{n} \left\{ -\exp\left(\boldsymbol{\beta}^\top \boldsymbol{Z}_i\right) \left(\frac{(L_i - \eta_j)_+^w - (R_i - \eta_j)_+^w}{w} \right)^2 \Delta_i(H_0) \right\}, \\
\frac{\partial^2 \tilde{\ell}}{\partial \alpha_0 \partial \alpha_1} &= \sum_{\substack{i=1\\ T_i \neq T_{\max}}}^{n_1} \left\{ -\frac{T_i}{h_0(T_i)^2} \right\} + \sum_{i=n_1+1}^{n} \left\{ -\exp\left(\boldsymbol{\beta}^\top \boldsymbol{Z}_i\right) \left((L_i - R_i) \left(\frac{L_i^2 - R_i^2}{2} \right) \right) \Delta_i(H_0) \right\}, \\
\frac{\partial^2 \tilde{\ell}}{\partial \nu_j \partial \alpha_0} &= \sum_{\substack{i=1\\ T_i \neq T_{\max}}}^{n_1} \left\{ -\frac{(\tau_j - T_i)_+^p}{h_0(T_i)^2} \right\} + \\
&\quad \sum_{i=n_1+1}^{n} \left\{ -\exp\left(\boldsymbol{\beta}^\top \boldsymbol{Z}_i\right) \left((L_i - R_i) \left(\frac{(\tau_j - R_i)_+^w - (\tau_j - L_i)_+^w}{w} \right) \right) \Delta_i(H_0) \right\}, \\
\frac{\partial^2 \tilde{\ell}}{\partial \nu_j \partial \alpha_1} &= \sum_{\substack{i=1\\ T_i \neq T_{\max}}}^{n_1} \left\{ -\frac{(\tau_j - T_i)_+^2 \; T_i}{h_0(T_i)^2} \right\} + \\
&\quad \sum_{i=n_1+1}^{n} \left\{ -\exp\left(\boldsymbol{\beta}^\top \boldsymbol{Z}_i\right) \left(\left(\frac{L_i^2 - R_i^2}{2} \right) \left(\frac{(\tau_j - R_i)_+^3 - (\tau_j - L_i)_+^3}{3} \right) \right) \Delta_i(H_0) \right\}, \\
\frac{\partial^2 \tilde{\ell}}{\partial \mu_j \partial \alpha_0} &= \sum_{\substack{i=1\\ T_i \neq T_{\max}}}^{n_1} \left\{ -\frac{(T_i - \eta_j)_+^p}{h_0(T_i)^2} \right\} + \\
&\quad \sum_{i=n_1+1}^{n} \left\{ -\exp\left(\boldsymbol{\beta}^\top \boldsymbol{Z}_i\right) \left((L_i - R_i) \left(\frac{(L_i - \eta_j)_+^w - (R_i - \eta_j)_+^w}{w} \right) \right) \Delta_i(H_0) \right\}, \\
\frac{\partial^2 \tilde{\ell}}{\partial \mu_j \partial \alpha_1} &= \sum_{\substack{i=1\\ T_i \neq T_{\max}}}^{n_1} \left\{ -\frac{(T_i - \eta_j)_+^2 \; T_i}{h_0(T_i)^2} \right\} + \\
&\quad \sum_{i=n_1+1}^{n} \left\{ -\exp\left(\boldsymbol{\beta}^\top \boldsymbol{Z}_i\right) \left(\left(\frac{L_i^2 - R_i^2}{2} \right) \left(\frac{(L_i - \eta_j)_+^3 - (R_i - \eta_j)_+^3}{3} \right) \right) \Delta_i(H_0) \right\}, \\
\frac{\partial^2 \tilde{\ell}}{\partial \mu_j \partial \nu_j} &= \sum_{\substack{i=1\\ T_i \neq T_{\max}}}^{n_1} \left\{ -\frac{(T_i - \eta_j)_+^p \; (\tau_j - T_i)_+^p}{h_0(T_i)^2} \right\} + \\
&\quad \sum_{i=n_1+1}^{n} \left\{ -\exp\left(\boldsymbol{\beta}^\top \boldsymbol{Z}_i\right) \left(\left(\frac{(\tau_j - R_i)_+^w - (\tau_j - L_i)_+^w}{w} \right) \left(\frac{(L_i - \eta_j)_+^w - (R_i - \eta_j)_+^w}{w} \right) \right) \Delta_i(H_0) \right\},
\end{aligned}
$$

where $q = 2, 4$, $w = 2, 3$, and $p = 1, 2$ in the case of piecewise linear and piecewise quadratic hazard function estimators, respectively, and

$$\Delta_i(H_0) = \frac{(\exp(H_0(L_i) - H_0(R_i)))^{\exp(\boldsymbol{\beta}^\top \boldsymbol{Z}_i)}}{\left(1 - (\exp(H_0(L_i) - H_0(R_i)))^{\exp(\boldsymbol{\beta}^\top \boldsymbol{Z}_i)}\right)^2}.$$

Also, the first and second derivative of the modified log-likelihood function with respect to $\boldsymbol{\beta}$ are given by

$$\begin{aligned}
\frac{\partial \tilde{\ell}}{\partial \boldsymbol{\beta}_j} &= \sum_{\substack{i=1 \\ T_i \neq T_{\max}}}^{n_1} \boldsymbol{z}_{ij} - \sum_{i=1}^{n} \left\{ \boldsymbol{z}_{ij} e^{(\boldsymbol{\beta}^\top \boldsymbol{Z}_i)} H_0(T_i) \right\} + \\
&\sum_{i=n_1+1}^{n} \left\{ -\boldsymbol{z}_{ij} e^{(\boldsymbol{\beta}^\top \boldsymbol{Z}_i)} H_0(L_i) - \frac{\boldsymbol{z}_{ij} e^{(\boldsymbol{\beta}^\top \boldsymbol{Z}_i)} (H_0(L_i) - H_0(R_i)) (e^{(H_0(L_i) - H_0(R_i))})^{e^{(\boldsymbol{\beta}^\top \boldsymbol{Z}_i)}}}{1 - \left(e^{(H_0(L_i) - H_0(R_i))}\right)^{e^{(\boldsymbol{\beta}^\top \boldsymbol{Z}_i)}}} \right\}, \\
\frac{\partial^2 \tilde{\ell}}{\partial \boldsymbol{\beta}_j^2} &= -\sum_{i=1}^{n} \left\{ \boldsymbol{z}_{ij} \boldsymbol{z}_{ij}^\top e^{(\boldsymbol{\beta}^\top \boldsymbol{Z}_i)} H_0(T_i) \right\} + \sum_{i=n_1+1}^{n} \left\{ -\boldsymbol{z}_{ij} \boldsymbol{z}_{ij}^\top e^{(\boldsymbol{\beta}^\top \boldsymbol{Z}_i)} H_0(L_i) - A \right\},
\end{aligned}$$

for $j = 1, \cdots, p$ and

$$\begin{aligned}
A = &\Big(\boldsymbol{z}_{ij} \boldsymbol{z}_{ij}^\top e^{(\boldsymbol{\beta}^\top \boldsymbol{Z}_i)} (H_0(L_i) - H_0(R_i)) \left(e^{(H_0(L_i) - H_0(R_i))}\right)^{e^{(\boldsymbol{\beta}^\top \boldsymbol{Z}_i)}} + \boldsymbol{z}_{ij} \boldsymbol{z}_{ij}^\top \left(e^{(\boldsymbol{\beta}^\top \boldsymbol{Z}_i)}\right)^2 (H_0(L_i) - H_0(R_i))^2 \\
&\left(e^{(H_0(L_i) - H_0(R_i))}\right)^{e^{(\boldsymbol{\beta}^\top \boldsymbol{Z}_i)}} \Big) B + \left(\boldsymbol{z}_{ij} \boldsymbol{z}_{ij}^\top \left(e^{(\boldsymbol{\beta}^\top \boldsymbol{Z}_i)}\right)^2 (H_0(L_i) - H_0(R_i))^2 \left(e^{(H_0(L_i) - H_0(R_i))}\right)^{2e^{(\boldsymbol{\beta}^\top \boldsymbol{Z}_i)}} \right) B,
\end{aligned}$$

where $B = 1 / \left(1 - \left(e^{(H_0(L_i) - H_0(R_i))}\right)^{e^{(\boldsymbol{\beta}^\top \boldsymbol{Z}_i)}}\right)^2$.

The information matrix is obtained based on the second derivative of the modified log-likelihood function with respect to $\boldsymbol{\beta}$. The standard errors of the estimator, $\hat{\boldsymbol{\beta}}$, are just the square roots of the diagonal elements in the variance-covariance matrix, i.e., the inverse of the information matrix.

5.2.4 QR Factorization of $\mathbf{D}$

In the case of piecewise linear hazard function, the Hessian matrix $\mathbf{H}$ can be written as follows,

$$\mathbf{H} = -\mathbf{D}^\top \mathbf{D},$$

where

$$\mathbf{D} = \begin{pmatrix} \frac{1}{h_0(T_1)} & \frac{(\tau_1-T_1)_+}{h_0(T_1)} & \cdots & \frac{(\tau_k-T_1)_+}{h_0(T_1)} & \frac{(T_1-\eta_1)_+}{h_0(T_1)} & \cdots & \frac{(T_1-\eta_m)_+}{h_0(T_1)} \\ \frac{1}{h_0(T_2)} & \frac{(\tau_1-T_2)_+}{h_0(T_2)} & \cdots & \frac{(\tau_k-T_2)_+}{h_0(T_2)} & \frac{(T_2-\eta_1)_+}{h_0(T_2)} & \cdots & \frac{(T_2-\eta_m)_+}{h_0(T_2)} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{1}{h_0(T_{n_1})} & \frac{(\tau_1-T_{n_1})_+}{h_0(T_{n_1})} & \cdots & \frac{(\tau_k-T_{n_1})_+}{h_0(T_{n_1})} & \frac{(T_{n_1}-\eta_1)_+}{h_0(T_{n_1})} & \cdots & \frac{(T_{n_1}-\eta_m)_+}{h_0(T_{n_1})} \\ \alpha_{n_1+1} & \beta_{n_1+1} & \cdots & \gamma_{n_1+1} & \lambda_{n_1+1} & \cdots & \delta_{n_1+1} \\ \alpha_{n_1+2} & \beta_{n_1+2} & \cdots & \gamma_{n_1+2} & \lambda_{n_1+2} & \cdots & \delta_{n_1+2} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \alpha_n & \beta_n & \cdots & \gamma_n & \lambda_n & \cdots & \delta_n \end{pmatrix}$$

and

$$\begin{aligned} \alpha_i &= \sqrt{\exp(\boldsymbol{\beta}^\top \boldsymbol{Z}_i)}(L_i - R_i)\ \Delta_i(H_0), \\ \beta_i &= \sqrt{\exp(\boldsymbol{\beta}^\top \boldsymbol{Z}_i)}\frac{(\tau_1 - L_i)_+^2 - (\tau_1 - R_i)_+^2}{2}\ \Delta_i(H_0), \\ \gamma_i &= \sqrt{\exp(\boldsymbol{\beta}^\top \boldsymbol{Z}_i)}\frac{(\tau_k - L_i)_+^2 - (\tau_k - R_i)_+^2}{2}\ \Delta_i(H_0), \\ \lambda_i &= \sqrt{\exp(\boldsymbol{\beta}^\top \boldsymbol{Z}_i)}\frac{(L_i - \eta_1)_+^2 - (R_i - \eta_1)_+^2}{2}\ \Delta_i(H_0), \\ \delta_i &= \sqrt{\exp(\boldsymbol{\beta}^\top \boldsymbol{Z}_i)}\frac{(L_i - \eta_k)_+^2 - (R_i - \eta_k)_+^2}{2}\ \Delta_i(H_0), \end{aligned}$$

where

$$\Delta_i(H_0) = \frac{\sqrt{(\exp(H_0(L_i) - H_0(R_i)))^{\exp(\boldsymbol{\beta}^\top \boldsymbol{Z}_i)}}}{\left(1 - (\exp(H_0(L_i) - H_0(R_i)))^{\exp(\boldsymbol{\beta}^\top \boldsymbol{Z}_i)}\right)},$$

for $i = n_1 + 1, n_1 + 2, \cdots, n$.

Therefore, we have the following Taylor series expansion about $\boldsymbol{\pi}$ with $\boldsymbol{\theta}$ fixed:

$$\begin{aligned} \tilde{\ell}(\boldsymbol{\pi}') - \tilde{\ell}(\boldsymbol{\pi}) &\approx \mathbf{g}^\top(\boldsymbol{\pi}' - \boldsymbol{\pi}) + \frac{1}{2}(\boldsymbol{\pi}' - \boldsymbol{\pi})^\top \mathbf{H}(\boldsymbol{\pi}' - \boldsymbol{\pi}) \\ &= -\frac{1}{2}\|\mathbf{D}(\boldsymbol{\pi}' - \boldsymbol{\pi}) - \mathbf{d}\|^2 + c \\ &= -\frac{1}{2}\|\mathbf{R}(\boldsymbol{\pi}' - \boldsymbol{\pi}) - \mathbf{b}_1\|^2 - \frac{1}{2}\|\mathbf{b}_2\|^2 + c \end{aligned}$$

where $\mathbf{D} = \mathbf{QR}$ by a QR decomposition and $\mathbf{d} = (\mathbf{R}^\top)^{-1}\mathbf{g}^\top$. In order to yield problem (5.6), we partition $\mathbf{Q} = \begin{pmatrix} \mathbf{Q_1} \\ \mathbf{Q_2} \end{pmatrix}$, where $\mathbf{Q_1} \in \mathbb{R}^{n\times(1+k+m)}$. Thus, $\mathbf{Q}_1^\top \mathbf{D} = \begin{pmatrix} \mathbf{R} \\ \mathbf{0} \end{pmatrix}$, where $\mathbf{R} \in$

$\mathbb{R}^{(1+k+m)\times(1+k+m)}$, and $\mathbf{Q}^\top\mathbf{b} = \begin{pmatrix}\mathbf{b}_1\\ \mathbf{b}_2\end{pmatrix} = \begin{pmatrix}\mathbf{Q}_1{}^\top\mathbf{d}\\ \mathbf{Q}_2{}^\top\mathbf{d}\end{pmatrix}$. The upper triangular matrix $\mathbf{R}$ obtained this way is mathematically the same as by the Cholesky decomposition of $-\mathbf{H}$, while it is numerically more stable.

In the case of piecewise quadratic hazard function, the Hessian matrix $\mathbf{H}$ is given by

$$\mathbf{H} = -\mathbf{D}^\top\mathbf{D},$$

where

$$\mathbf{D} = \begin{pmatrix}
\frac{1}{h_0(T_1)} & \frac{T_1}{h_0(T_1)} & \frac{(\tau_1-T_1)_+^2}{h_0(T_1)} & \cdots & \frac{(\tau_k-T_1)_+^2}{h_0(T_1)} & \frac{(T_1-\eta_1)_+^2}{h_0(T_1)} & \cdots & \frac{(T_1-\eta_m)_+^2}{h_0(T_1)} \\
\frac{1}{h_0(T_2)} & \frac{T_2}{h_0(T_2)} & \frac{(\tau_1-T_2)_+^2}{h_0(T_2)} & \cdots & \frac{(\tau_k-T_2)_+^2}{h_0(T_2)} & \frac{(T_2-\eta_1)_+^2}{h_0(T_2)} & \cdots & \frac{(T_2-\eta_m)_+^2}{h_0(T_2)} \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
\frac{1}{h_0(T_{n_1})} & \frac{T_{n_1}}{h_0(T_{n_1})} & \frac{(\tau_1-T_{n_1})_+^2}{h_0(T_{n_1})} & \cdots & \frac{(\tau_k-T_{n_1})_+^2}{h_0(T_{n_1})} & \frac{(T_{n_1}-\eta_1)_+^2}{h(T_{n_1})} & \cdots & \frac{(T_{n_1}-\eta_m)_+^2}{h(T_{n_1})} \\
\alpha_{0_{n_1+1}} & \alpha_{1_{n_1+1}} & \beta_{n_1+1} & \cdots & \gamma_{n_1+1} & \lambda_{n_1+1} & \cdots & \delta_{n_1+1} \\
\alpha_{0_{n_1+2}} & \alpha_{1_{n_1+2}} & \beta_{n_1+2} & \cdots & \gamma_{n_1+2} & \lambda_{n_1+2} & \cdots & \delta_{n_1+2} \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
\alpha_{0_n} & \alpha_{1_n} & \beta_n & \cdots & \gamma_n & \lambda_n & \cdots & \delta_n
\end{pmatrix}$$

and

$$\begin{aligned}
\alpha_{0_i} &= \sqrt{\exp(\boldsymbol{\beta}^\top \boldsymbol{Z}_i)}(L_i - R_i)\ \Delta_i(H_0), \\
\alpha_{1_i} &= \sqrt{\exp(\boldsymbol{\beta}^\top \boldsymbol{Z}_i)}\left(\frac{L_i^2 - R_i^2}{2}\right)\ \Delta_i(H_0), \\
\beta_i &= \sqrt{\exp(\boldsymbol{\beta}^\top \boldsymbol{Z}_i)}\frac{(\tau_1 - L_i)_+^3 - (\tau_1 - R_i)_+^3}{3}\ \Delta_i(H_0), \\
\gamma_i &= \sqrt{\exp(\boldsymbol{\beta}^\top \boldsymbol{Z}_i)}\frac{(\tau_k - L_i)_+^3 - (\tau_k - R_i)_+^3}{3}\ \Delta_i(H_0), \\
\lambda_i &= \sqrt{\exp(\boldsymbol{\beta}^\top \boldsymbol{Z}_i)}\frac{(L_i - \eta_1)_+^3 - (R_i - \eta_1)_+^3}{3}\ \Delta_i(H_0), \\
\delta_i &= \sqrt{\exp(\boldsymbol{\beta}^\top \boldsymbol{Z}_i)}\frac{(L_i - \eta_k)_+^3 - (R_i - \eta_k)_+^3}{3}\ \Delta_i(H_0),
\end{aligned}$$

where

$$\Delta_i(H_0) = \frac{\sqrt{(\exp(H_0(L_i) - H_0(R_i)))^{\exp(\boldsymbol{\beta}^\top \boldsymbol{Z}_i)}}}{\left(1 - (\exp(H_0(L_i) - H_0(R_i)))^{\exp(\boldsymbol{\beta}^\top \boldsymbol{Z}_i)}\right)}.$$

for $i = n_1 + 1, n_1 + 2, \cdots, n$.

Therefore, the Taylor series expansion about $\boldsymbol{\pi}$ with $\boldsymbol{\theta}$ fixed gives

$$\begin{aligned}
\tilde{\ell}(\boldsymbol{\pi}') - \tilde{\ell}(\boldsymbol{\pi}) &\approx \mathbf{g}^\top(\boldsymbol{\pi}' - \boldsymbol{\pi}) + \frac{1}{2}(\boldsymbol{\pi}' - \boldsymbol{\pi})^\top \mathbf{H}(\boldsymbol{\pi}' - \boldsymbol{\pi}) \\
&= -\frac{1}{2}\|\mathbf{D}(\boldsymbol{\pi}' - \boldsymbol{\pi}) - \mathbf{d}\|^2 + c \\
&= -\frac{1}{2}\|\mathbf{R}(\boldsymbol{\pi}' - \boldsymbol{\pi}) - \mathbf{b}_1\|^2 - \frac{1}{2}\|\mathbf{b}_2\|^2 + c \\
&= -\frac{1}{2}\left\|\mathbf{R}\begin{pmatrix}\boldsymbol{\pi}'_1 \\ \boldsymbol{\pi}'_2\end{pmatrix} - \mathbf{R}\begin{pmatrix}\boldsymbol{\pi}_1 \\ \boldsymbol{\pi}_2\end{pmatrix} - \mathbf{b}_1\right\|^2 - \frac{1}{2}\|\mathbf{b}_2\|^2 + c,
\end{aligned}$$

where $\mathbf{D} = \mathbf{QR}$ by a QR decomposition, $\mathbf{d} = (\mathbf{R}^\top)^{-1}\mathbf{g}^\top$, and $\boldsymbol{\pi}'_1 \in \mathbb{R}^2, \boldsymbol{\pi}'_2 \in \mathbb{R}_+^{m+k}$. To obtain the least squares linear regression problem (5.7) with only nonnegativity constraints, we partitioned $\mathbf{Q} = \begin{pmatrix}\mathbf{Q_1} \\ \mathbf{Q_2}\end{pmatrix}$, where $\mathbf{Q_1} \in \mathbb{R}^{n\times(2+k+m)}$. Hence, $\mathbf{Q}_1^\top\mathbf{D} = \begin{pmatrix}\mathbf{R} \\ \mathbf{0}\end{pmatrix}$, where $\mathbf{R} \in \mathbb{R}^{(2+k+m)\times(2+k+m)}$, and $\mathbf{Q}^\top\mathbf{b} = \begin{pmatrix}\mathbf{b}_1 \\ \mathbf{b}_2\end{pmatrix} = \begin{pmatrix}\mathbf{Q}_1{}^\top\mathbf{d} \\ \mathbf{Q}_2{}^\top\mathbf{d}\end{pmatrix}$. Note that the upper triangular matrix $\mathbf{R}$ obtained by the QR decomposition of $\mathbf{D}$ is mathematically the same as by the Cholesky decomposition of $-\mathbf{H}$ and also numerically more stable than that.

5.3 Simulation Study

A simulation study was conducted to evaluate the efficiency of the proposed piecewise linear baseline convex hazard within the Cox model (LCHCOX) and the smoothed baseline convex hazard in the Cox model (SCHCOX) estimators and compare them to its conventional fully semiparametric maximum likelihood estimation method in the case of interval-censored data. For each subject, all n exact event times T are first generated independently based on the Cox model with a BT distribution (Haupt and Schäbe, 1997) baseline hazard function, where $\alpha = 5$ and $\beta = 0.1$. For each censored observation, a random sample of size 10 is drawn from the exponential distribution with mean 0.5, which divides the domain $[0, \infty)$ into 11 disjoint subintervals. Then, the subinterval that includes the exact event time replaces it. In this study, there exist three covariates $\boldsymbol{Z} = (Z_1, Z_2, Z_3)$, following $Z_1 \sim \text{uniform}(-1, 1)$,

$Z_2 \sim$ normal$(0,1)$ and $Z_3 \sim$ Bernoulli(0.5), whose true coefficients are $\boldsymbol{\beta} = (-0.5, 0.5, 1)^\top$. Pan (1999) extended the ICM algorithm proposed by Groeneboom and Wellner (1992) and Joly et al. (1998) for computing the nonparametric maximum likelihood estimates to the Cox model with interval-censored data. This approach was implemented in the R package **intcox** developed by Henschel and Mansmann (2009).

The results of the simulation study of over 100 replications with sample size 200 and 500 are summarized in Table 5.1. We assessed the performance of different $\boldsymbol{\beta}$ estimates by examining their empirical mean of the $\boldsymbol{\beta}$ estimates, the bias of $\hat{\boldsymbol{\beta}}$, and the standard deviations and mean squared errors (MSEs) of $\hat{\boldsymbol{\beta}}$. Also, Figure 5.1 shows the box plots for the squared errors over 100 replications using the LCHCOX, SCHCOX, and fully semiparametric maximum likelihood (INTCOX) estimators. According to Figure 5.1, our shape-constrained estimators generally outperform the unconstrained competitor.

Based on the results of Table 5.1 and Figure 5.1, we have the following conclusions. First, the proposed shape-constrained approaches generally perform better than the INTCOX in terms of the MSE for both sample sizes. Second, the biases of the SCHCOX approach is generally smaller than the others that were considered. Third, the bias and variance decrease as the sample size increases in all cases studied.

Table 5.1: Proportional hazard model, simulation results for interval-censored data.

Method	Coefficient	Estimate	Bias	SD	MSE
$n = 200$					
INTCOX	$\beta_1 = -0.5$	-0.4480	0.0500	0.1394	0.0217
	$\beta_2 = 0.5$	0.4680	-0.0320	0.0822	0.0077
	$\beta_3 = 1$	1.1853	0.2061	0.1515	0.0571
LCHCOX	$\beta_1 = -0.5$	-0.4706	0.0294	0.1367	0.0194
	$\beta_=0.5$	0.4953	-0.0047	0.0898	0.0080
	$\beta_3 = 1$	1.0162	0.0162	0.1152	0.0233
SCHCOX	$\beta_1 = -0.5$	-0.4699	0.0301	0.1370	0.0195
	$\beta_2 = 0.5$	0.4983	-0.0017	0.0912	0.0082
	$\beta_3 = 1$	1.0141	0.0141	0.1532	0.0234
$n = 500$					
INTCOX	$\beta_1 = -0.5$	-0.4675	0.0314	0.0954	0.0051
	$\beta_2 = 0.5$	0.4657	-0.0291	0.0482	0.0019
	$\beta_3 = 1$	1.1729	0.1780	0.1063	0.0299
LCHCOX	$\beta_1 = -0.5$	-0.4850	0.0048	0.0985	0.0055
	$\beta_2 = 0.5$	0.4972	0.0012	0.0456	0.0008
	$\beta_3 = 1$	0.9957	-0.0008	0.1082	0.0043
SCHCOX	$\beta_1 = -0.5$	-0.4840	0.0052	0.0983	0.0054
	$\beta_2 = 0.5$	0.4973	0.0005	0.0448	0.0008
	$\beta_3 = 1$	0.9921	-0.0003	0.1077	0.0046

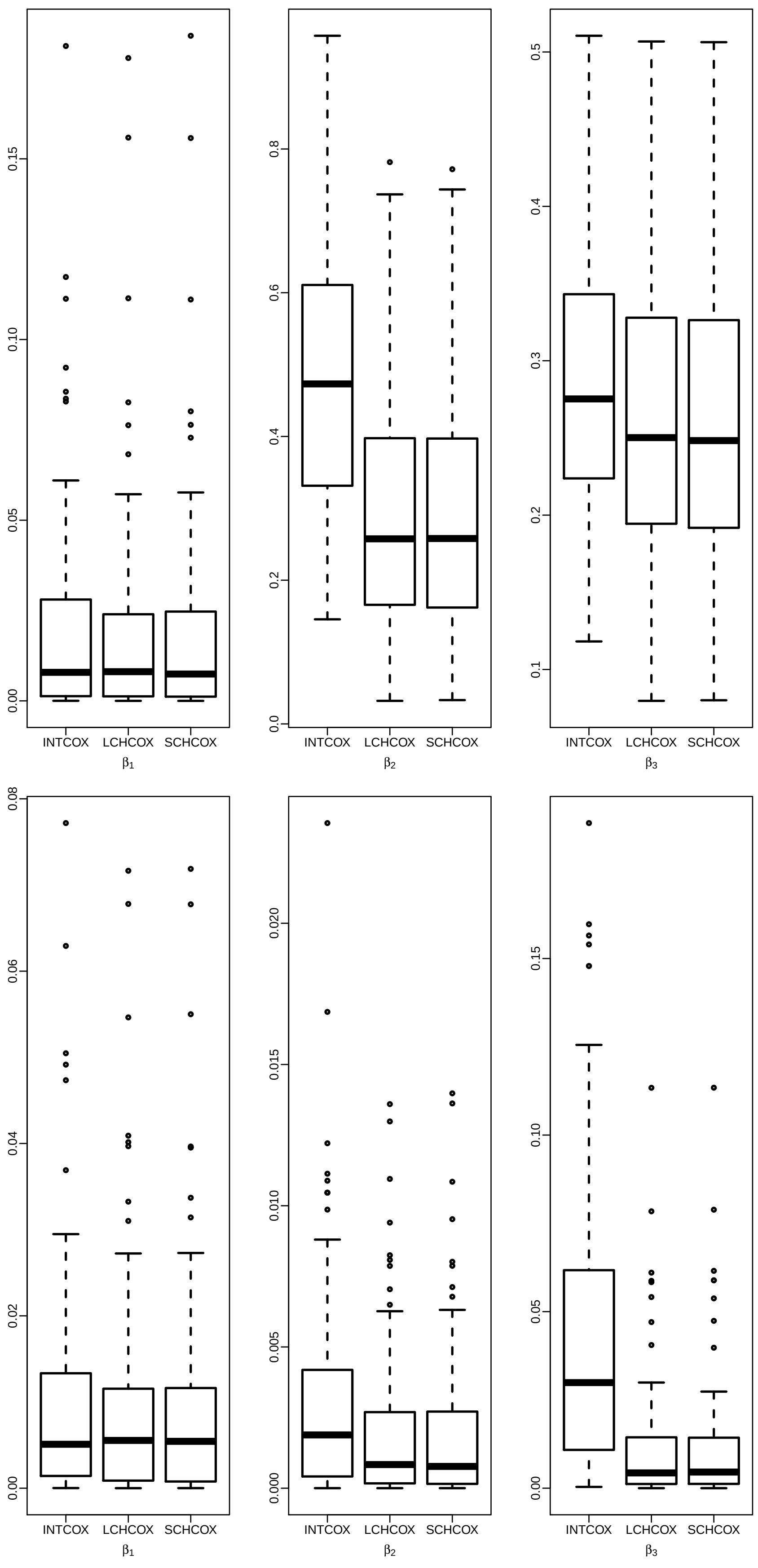

Figure 5.1: Box plots of the squared errors versus each of the covariates over 100 replications for $n = 200$ (upper panel) and $n = 500$ (lower panel) simulated data.

5.4 Real Data Examples

In a further investigation on the performance of different proposed estimators and the traditional partial likelihood estimator, we consider two real data sets named, respectively in our study, "Kidney Transplant" and "Larynx Cancer". The first data set can be obtained from R package **OIsurv** (Diez, 2012) that includes two categorical covariates, gender and race, and one continuous covariate, age at the time of a transplant. The second real data is available in R package **KMsurv** (Yan, 2010) including four variables, namely stage of cancer at the time of the first treatment, death time on study in terms of months, age at diagnosis and death indicator. For the traditional partial likelihood estimator, the Efron's partial likelihood estimator is available by function `coxph` in R package **survival** (Therneau and Lumely, 2009).

5.4.1 Kidney Transplant Data

For illustration and comparison, we apply the proposed methods for the time to death of 863 patients after kidney transplant. In this study, we have one continuous variable age at time of transplant and two categorical variables, namely race and gender as follows:

$$Z1 = 1 \text{ if the patient is female}, 0 \text{ otherwise}$$
$$Z2 = 1 \text{ if the patient is black}, 0 \text{ otherwise}$$

Table 5.2 reports the results of comparison between our proposed constrained approaches and the conventional partial likelihood method in terms of the standard error. As can be seen from Table 5.2, our proposed LCHCOX and SCHCOX approaches and the PL model yield very similar regression coefficients which implies that all methods are successful on this data set. Furthermore, the LCHCOX and SCHCOX approaches provide a piecewise linear and a smooth baseline hazard estimate, respectively. Moreover, the results indicate that the age of the patient at time of transplant has a highly significant effect on the time of death of a patient after kidney transplant surgery, leading to a p-value of around 1.5×10^{-12}. The test of the proportional hazards assumption for two covariates gender and race was non-significant with p-values around 0.88 and 0.53, respectively.

The LCHCOX and SCHCOX for this data set are, respectively,

$$\begin{aligned}\hat{h}_0(t) = 0.002 + 0.023\,(0.040 - t)_+ + 0.003\,(0.356 - t)_+ + 0.019\,(0.473 - t)_+ + \\ 0.001\,(1.464 - t)_+ + 0.001\,(4.477 - t)_+ + 0.001\,(t - 4.477)_+,\end{aligned}$$

and

$$\hat{h}_0(t) = 0.0003 + 0.0004\,t + 0.0198\,(0.5842 - t)_+^2 + 0.0107\,(0.5846 - t)_+^2 + \\ 0.0004\,(1.6665 - t)_+^2 + 0.0001\,(5.5483 - t)_+^2.$$

Table 5.2: Analysis of variance table for race, gender, and age for the kidney transplant patients.

Method	Variables	Estimates	Standard Error	t-value	p-value
	Z1 : Female	0.0265	0.1749	0.1515	0.8796
PL	Z2 : Black	0.1164	0.2115	0.5504	0.5822
	Z3 : Age	0.0510	0.0071	7.1831	0
	Z1 : Female	0.0202	0.1488	0.1358	0.8920
LCHCOX	Z2 : Black	0.1185	0.1867	0.6349	0.5257
	Z3 : Age	0.0511	0.0050	10.32	0
	Z1 : Female	0.0199	0.1488	0.1337	0.8937
SCHCOX	Z2 : Black	0.1204	0.1866	0.6452	0.5190
	Z3 : Age	0.0512	0.0050	10.24	0

5.4.2 Larynx Cancer Data

The larynx cancer data set, descried fully in Kardaun (1983), consists of 90 male patients with larynx cancer, during the period from 1970 to 1978 at a Dutch hospital. In this study, the time variable is the interval, in years, between the fist treatment of the laryngeal cancer and the time of death of the patient or the end of the study (January 1, 1983). For each patient, the age at the time of diagnosis of the cancer, the year of diagnosis, and the stage of the larynx cancer were reported. There are four stages of cancer disease based on the T.N.M. (primary tumor (T), nodal involvement (N) and distant metastasis (M) grading) classification used by the American Joint Committee for Cancer Staging. The stages also denoted by I, II, III and IV which are ordered from the least to the most serious. The three dummy variables are defined as follows.

$$Z1 = 1 \text{ if the patient is in stage II}, 0 \text{ otherwise}$$
$$Z2 = 1 \text{ if the patient is in stage III}, 0 \text{ otherwise}$$
$$Z3 = 1 \text{ if the patient is in stage IV}, 0 \text{ otherwise}$$

We first fitted a Cox proportional hazard model using the proposed models as well as the conventional PL approach and then compared the estimated baseline survival function for each of the four stages of cancer at the ages of 60 and 70. According to our estimations, the risk of death among stage II, III, and stage IV patients are approximately $1.16, 1.91$, and 5.64 times higher than that of stage I patients who are similar with respect to age at time of diagnosis. By looking at the p-value of the covariate Stage IV, one can easily find there is a significant difference between stage I and IV patients. In addition, a one-year increase in age at time of diagnosis is related to around 2% greater risk of death among patients who are similar with respect to disease staging. Also, the estimated survival probabilities of the smooth approach for a patient at age 60 and 70 at one, three, five, and ten years after diagnosis of larynx cancer is summarized in Table 5.4. Obviously, the survival probability of a patient at different stages of the larynx cancer reduces as the age of the patient advances. In addition, the survival of a patient at different stages of disease decreases as the years after diagnosis of the cancer increase.

The estimated survival curves of mentioned strategies for all cancer stages at the patient age 60 and 70 are shown in Figure 5.3. As expected from Figure 5.3, it is clear that the survival probabilities decrease as the stage increases. Particularly, the estimated survival curve of stage IV of the cancer disease is significantly lower than all the other stages, whilst the estimated survival probabilities of the stage I and stage II have smaller differences. Moreover, our proposed approaches can also provide a piecewise linear continuous and a smooth baseline hazard function estimate, respectively, while the traditional method only yields a step function estimate.

For this data set, the estimated piecewise linear continuous and smooth baseline hazard functions are given respectively as follows:

$$\hat{h}_0(t) = 0.019 + 0.002\,(2.616 - t)_+ + 0.004\,(t - 2.616)_+,$$

and

$$\hat{h}_0(t) = 0.012 + 0.003\,t + 0.001\,(3.092 - t)_+^2.$$

Table 5.3: Analysis of variance table for stage of the laryngeal cancer patients.

Method	Variables	Estimates	Standard Errors	t-Value	p-Value
	Z1: Stage II	0.1400	0.4625	0.3027	0.7628
PL	Z2 : Stage III	0.6420	0.3561	1.804	0.0747
	Z3 : Stage IV	1.7060	0.4219	4.0436	0.0001
	Z4 : Age	0.019	0.0143	1.3287	0.1874
	Z1: Stage II	0.1786	0.4456	0.3281	0.7436
LCHCOX	Z2 : Stage III	0.6553	0.3427	1.9158	0.0586
	Z3 : Stage IV	1.7661	0.3978	4.0321	0.0001
	Z4 : Age	0.0200	0.0110	1.8181	0.0724
	Z1 : Stage II	0.1795	0.4453	0.4031	0.6879
SCHCOX	Z2 : Stage III	0.6531	0.3431	1.9035	0.0602
	Z3 : Stage IV	1.7671	0.3970	4.4510	2.4×10^{-5}
	Z4 : Age	0.0202	0.0112	1.8035	0.0747

Table 5.4: Estimated survival probabilities for a 60- and 70-year-old patient at different years after diagnosis.

		Year			
Age	Stage	1	3	5	10
60	I	0.9349	0.8219	0.7025	0.4015
	II	0.9226	0.7908	0.6553	0.3356
	III	0.8786	0.6859	0.5073	0.1732
	IV	0.6742	0.3171	0.1265	0.0048
70	I	0.9209	0.7866	0.6491	0.3274
	II	0.9061	0.7504	0.5963	0.2629
	III	0.8536	0.6305	0.4359	0.1170
	IV	0.6173	0.2454	0.0797	0.0015

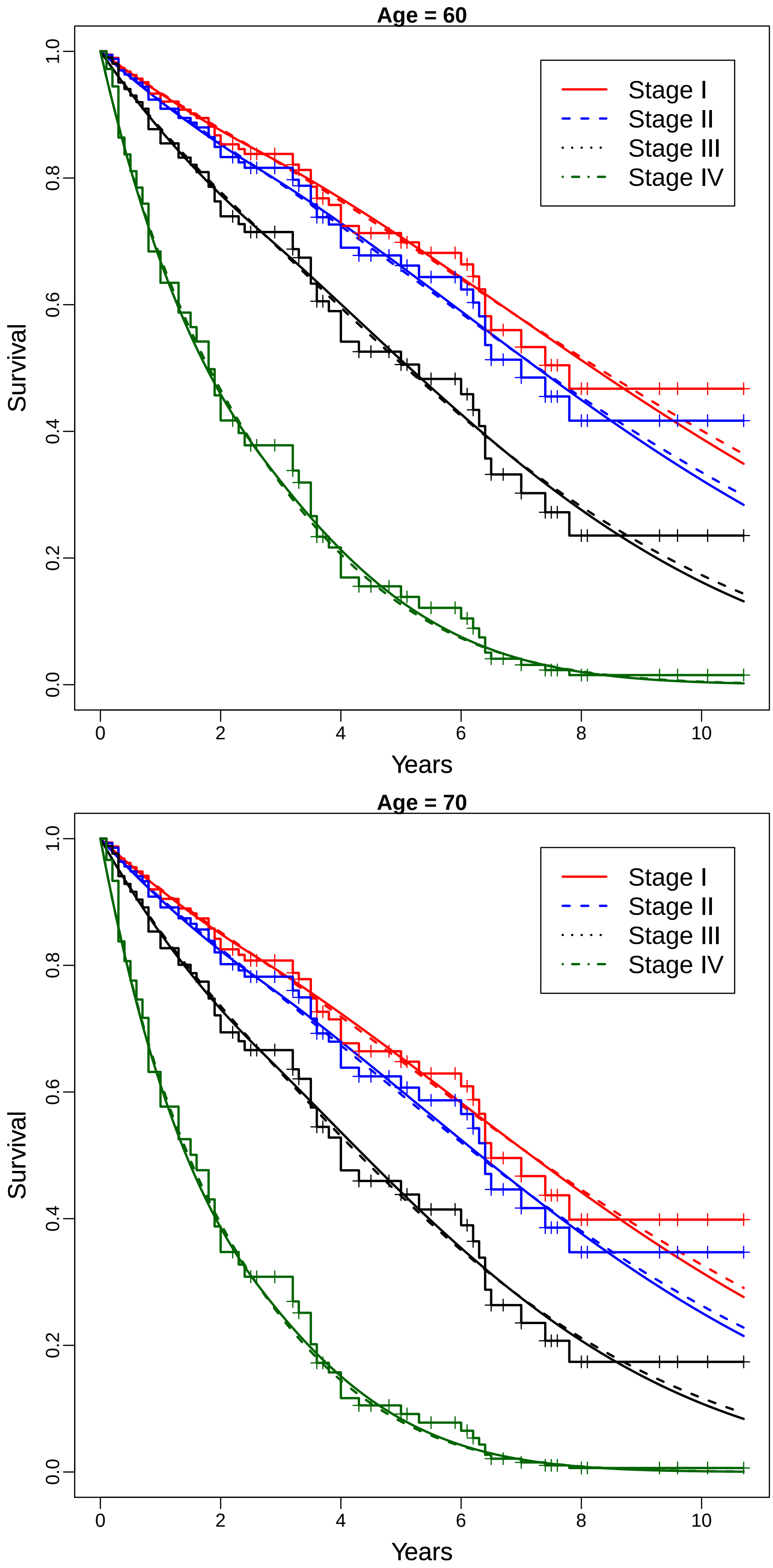

Figure 5.2: Survival estimates of the LCHCOX approach (solid curves), the SCHCOX approach (dashed curves), and the conventional KM estimator (step functions) for each stage of larynx cancer at the ages of 60 and 70.

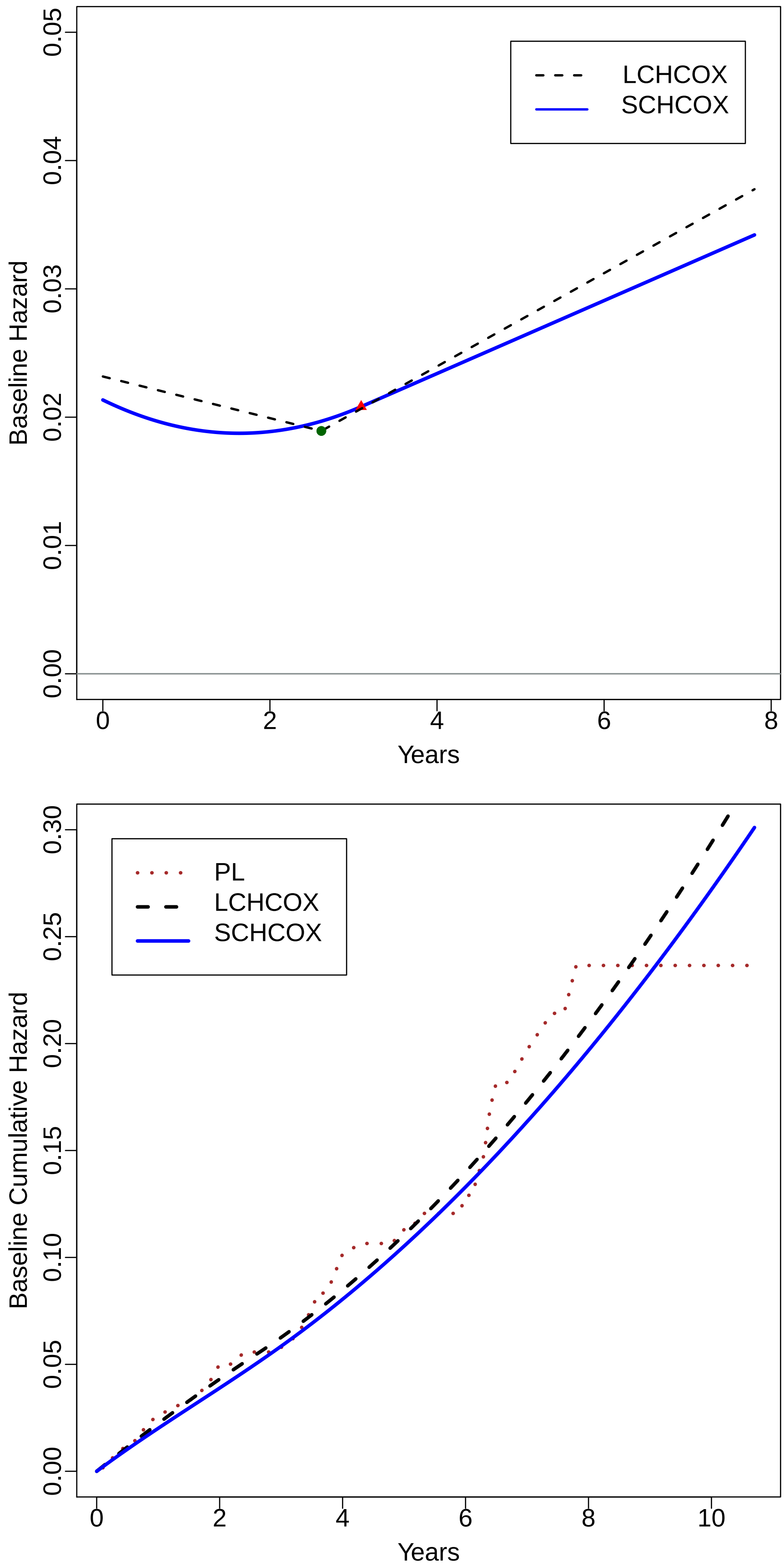

Figure 5.3: Estimated baseline hazard and cumulative hazard functions for larynx cancer data.

5.5 Summary

In this chapter, we study nonparametric estimation of a baseline hazard function under shape constraints as well as the estimation of the effects of covariates on failure times with right censoring, interval censoring and the mixed case with exact observations. Based on maximizing the likelihood function with or without incorporating the smoothness assumption, two different estimators are proposed for a convex baseline hazard function. In addition, a new algorithm was developed for computing the nonparametric estimation of a convex baseline hazard function along with the estimation of the regression coefficients in the Cox proportional hazard models. The new algorithm applies alternately the CNMCH or CNMSCH method and an optimization method that deals with lower bound constraints like L-BFGS.

Through a simulation study and two real-world data sets, we have investigated and compared the performance of our proposed approaches with that of the PL based method for the case of right censoring data and the fully semiparametric maximum likelihood estimation method for the case of interval-censored data. Empirical studies using simulated and real data sets indicate that our shape-constrained approaches generally dominate the unconstrained counterparts. In addition, our shape-constrained techniques provide either a piecewise linear continuous or a smooth baseline hazard estimate subject to convex constraint. The results of simulation study also reveal that our shape-constrained approaches generally perform better than its competitor in terms of the MSE. Furthermore, the shape-constrained baseline hazard function estimators commonly give a slightly smaller bias. Also, the bias and variance decrease as the sample size increases in all cases of simulation study.

Chapter 6

Summary and Future Works

6.1 Summary

Estimation of the distribution of the event times of interest is one of the primary tasks in survival analysis. The major characteristic of the event of interest is the existence of censoring. In survival analysis and reliability applications, one often has a former knowledge about the shape of the underlying function. To eliminate some issues such as bandwidth or tuning parameter selection of the other popular nonparametric approaches, it may be reasonable to make use of this knowledge for estimating a function by imposing natural qualitative constraints on it. Our main research problem is to find the nonparametric estimation of a function under shape restrictions. In order to solve this problem, our methodologies is allocated to nonparametric maximum likelihood approach. Our survival analysis studies is dedicated to the problem of hazard function estimation since, practically, it can be more sensible to enforce certain shape restrictions on it than on, for example, the survival function. Due to various forms of censored data that may arise in survival analysis, we confine our studies to cases with exact observations, general interval-censored data, which naturally includes the cases with right-censored or left-censored data, and interval-censored data mixed with exact observations. Moreover, depending on whether a smoothness assumption is enforced, two different convex shape-constrained hazard function estimators are proposed. In general, there is no explicit solution available for the problem of finding the NPMLE of a function and the most popular approach is to use numerical optimisation through an iterative algorithm. Therefore, two different computational algorithms are also developed to solve these problems.

In Chapter 2, a review of the most popular non-smooth and smooth nonparametric approaches is presented. Although these methods can be utilized to handle the problem of function estimation, the shape-constrained techniques have a potential to provide more accurate estimates than their unconstrained counterparts.

In Chapter 3, the problem of estimating a hazard function subject to convex shape restrictions is studied. A new idea to overcome the difficulty of choosing the minimum of a convex hazard function estimate is proposed. On a computational side, we saved a remarkable computation cost due to reducing the needed double loop method to a single loop. A new fast algorithm is developed for computing the NPMLE of a convex hazard function that cope well with the situations of both exact and interval-censored observations in any proportions. In addition, theoretical justification for the convergence of the new algorithm in the case of exact observations is established, which can be easily extended to other similar situations. Numerical studies using simulation and real-world data sets reveal that our estimator tends to give a more accurate estimate in terms of the maximum likelihood value compared to the state-of-the-art one. In the case of exact observations, our proposed algorithm is significantly faster than its counterpart. Furthermore, it can also handle both the situation of interval-censored data and the situation mixed with exact and interval-censored observations.

In Chapter 4, the nonparametric maximum likelihood estimator of a hazard function under convexity shape constraint and smoothness assumption is derived. A new algorithm for computing the smooth nonparametric maximum likelihood of a convex hazard function is also developed. Empirical studies indicate that our shape-restricted estimators generally outperforms the presmoothed kernel-based and the logspline density estimators in the sense of minimising the mean integrated squared error and the Kullback-Leibler or the Hellinger risk. The results of our numerical studies also show that the smooth shape-restricted estimator has a greater capability to model human mortality data than the piecewise linear continuous estimator does, specifically in the infant mortality phase thanks to the smoothness provided by the former. Furthermore, our proposed estimators can easily deal with the situation of heavily censored data, whereas the logspline technique may suffer from convergence problems in this situation.

In Chapter 5, we investigate the problem of simultaneous estimation of the underlying regression coefficients of the covariates and the convex baseline hazard function using the nonparametric maximum likelihood method. We further propose a new hybrid algorithm for simultaneously computing the nonparametric maximum likelihood estimates of a convex baseline hazard and the effects of covariates on survival times. Although the traditional partial likelihood can address the problem of estimating the regression coefficients, our proposal provides piecewise linear, continuous and/or smooth convex baseline hazard function estimates as well. A simulation study reveals that our proposed estimators generally dominates the fully semiparametric maximum likelihood estimation method in the Cox model with interval-censored data.

6.2 Future Works

The main problem in this thesis is to investigate the effect of applying shape restrictions in survival analysis. There are, of course, interesting issues that can be further investigated in this direction. In the following paragraphs, we list a number of topics for further studies.

More Sophisticated Characteristics of Survival Data

Although, in this thesis, we only focus on exact observations, interval-censored data, and the case mixed with exact and interval-censored observations, it should be a straightforward task to extend the research to more complex situations like truncation and double censoring. Some related work on estimation of a survival function for left-truncated and interval-censored data subject to monotone hazards can be found in Pan and Chappell (1998). In the case of unconstrained nonparametric estimation problems based on the interval-censored and left-truncated, one can also see the discussion in Hudgens (2005). Moreover, some recent works in this area have been done by Shen (2011), and Shen (2012), who studied the NPMLEs of the distribution function of the lifetime with doubly censored and truncated data and then interval-censored and doubly truncated data, respectively.

Multivariate Survival Data

Our present work has been concentrated on univariate survival data analysis. The literature on the analysis of bivariate or multivariate failure time data is relatively sparse. There are still many issues that need to be investigated for the analysis of such failure time data. Our shape-constrained maximum likelihood estimation method seems easily extendable to solve the problem of nonparametric estimation of a joint distribution or survival function, and to regression analysis when several correlated survival times of interest exist. We draw the attention of the reader to Chen et al. (2013) for a discussion of the analysis of multivariate interval-censored failure time data without shape-restricted assumptions.

Other Semiparametric Regression Models

In this thesis, we focused on studying the simultaneous estimation of a shape-constrained baseline hazard function and the regression coefficients in the Cox proportional hazards model. It is also worthwhile to apply the presented methods to other semiparametric regression models such as the proportional odds regression and the additive hazard model. For the unconstrained method, further investigations in this regard can be found in Chen et al.

(2007) and Wang et al. (2010). In addition, the confidence band for an estimated hazard or survival function needs to be further studied in the future.

Informative Censoring and Time-dependent Covariates

The main assumption for establishing our likelihood function is that the censoring mechanism is independent of the event of interest, and furthermore, is noninformative. Also, the covariates are assumed to be time-independent. However, the distribution of censoring time can involve some parameters of interest in the distribution of lifetime. Further study can consider the case of informative censoring and time-dependent covariates. Some related work in these directions are Lu and Zhang (2012) and Sparling et al. (2006).

Test Statistic

One may also be interested in testing the hypothesis that the failure times are a random sample from a population with a parametric hazard rate against a shape-restricted alternative one. In order to conduct such a hypothesis test, one can apply the resampling bootstrap method. We urge the interested reader to see Habtzghi and Datta (2012).

Other Theoretical Justifications

In this thesis, we have established some theoretical properties of our nonparametric maximum likelihood estimator of a convex hazard function. Further theoretical justifications that concern, for example, consistency and efficiency of the shape-restricted hazard function estimators need to be established.

References

Aalen, O. O. (1978). Nonparametric inference for a family of counting processes. *Annals of Statistics 6*(4), 701–726.

Aarset, M. V. (1987). How to identify a bathtub hazard rate. *IEEE Transactions on Reliability R-36*(1), 106–108.

Andersen, P. K. and R. D. Gill (1982). Cox's regression model for counting processes: a large sample study. *Annals of Statistics 10*(4), 1100–1120.

Anderson, J. A. and A. Senthilselvan (1980). Smooth estimates for the hazard function. *Journal of the Royal Statistical Society, Series B 42*(3), 322–327.

Ayer, M., H. D. Brunk, G. M. Ewing, W. T. Reid, and E. Silverman (1955). An empirical distribution function for sampling with incomplete information. *Annals of Mathematical Statistics 26*(4), 641–647.

Bagkavos, D. and P. Patil (2009). Variable bandwidths for nonparametric hazard rate estimation. *Communications in Statistics - Theory and Methods 38*(7), 1055–1078.

Balabdaoui, F. and J. A. Wellner (2007). Estimation of a kmonotone density: Limit distribution theory and the spline connection. *Annals of Statistics 35*(6), 2536–2564.

Banerjee, M. (2008). Estimating monotone, unimodal and U-shaped failure rates using asymptotic pivots. *Statistica Sinica 18*(2), 467–492.

Bantis, L. E., J. V. Tsimikas, and S. D. Georgiou (2012). Survival estimation through the cumulative hazard function with monotone natural cubic splines. *Lifetime Data Analysis 18*(3), 364–396.

Barlow, R. E., A. W. Marshall, and F. Proschan (1963). Properties of probability distributions with monotone hazard rate. *Annals of Mathematical Statistics 34*(2), 375–389.

Bergman, B. and B. Klefsjo (1984). The total time on test concept and its use in reliability theory. *Operations Research 32*(3), 596–606.

Betensky, R. A., J. C. Lindsey, L. M. Ryan, and M. P. Wand (1999). Local EM estimation of the hazard function for interval-censored data. *Biometrics 55*(1), 238–245.

Betensky, R. A., J. C. Lindsey, L. M. Ryan, and M. P. Wand (2002). A local likelihood proportional hazards model for interval censored data. *Statistics in Medicine 21*(2), 263–275.

Bloxom, B. (1985). A constrained spline estimator of a hazard function. *Psychometrika 50*(3), 301–321.

Bowman, A. and A. Azzalini (1997). *Applied Smoothing Techniques for Data Analysis : The Kernel Approach with S-Plus Illustrations*. Oxford University Press, Oxford.

Braun, J., T. Duchesne, and J. E. Stafford (2005). Local likelihood density estimation for interval censored data. *Canadian Journal of Statistics 33*(1), 39–60.

Bray, T. A., G. B. Crawford, and F. Proschan (1967). *Maximum Likelihood Estimation of a U-shaped Failure Rate Function*. Defense Technical Information Center.

Breslow, N. (1974). Covariance analysis of censored survival data. *Biometrics 30*(1), 89–99.

Breslow, N. E. (1972). Contribution to the discussion of the paper by D. R. Cox. *Journal of the Royal Statistical Society, Series B 34*(2), 216–217.

Cai, T. and R. A. Betensky (2003). Hazard regression for interval-censored data with penalized spline. *Biometrics 59*(3), 570–579.

Cao, R. and M. A. Jácome (2004). Presmoothed kernel density estimator for censored data. *Journal of Nonparametric Statistics 16*(1-2), 289–309.

Cao, R. and I. López-de Ullibarri (2007). Product-type and presmoothed hazard rate estimators with censored data. *Test 16*(2), 355–382.

Cao, R., I. López-de Ullibarri, P. Janssen, and N. Veraverbeke (2005). Presmoothed Kaplan-Meier and Nelson-Aalen estimators. *Journal of Nonparametric Statistics 17*(1), 31–56.

Chen, M. H., X. Tong, and J. Sun (2007). The proportional odds model for multivariate interval-censored failure time data. *Statistics in Medicine 26*(28), 5147–5161.

Chen, M. H., X. Tong, and L. Zhu (2013). A linear transformation model for multivariate interval-censored failure time data. *Canadian Journal of Statistics 41*(2), 275–290.

Chung, D. and M. N. Chang (1994). An isotonic estimator of the baseline hazard function in Cox's regression model under order restriction. *Statistics and Probability Letters 21*(3), 223–228.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society. Series B 34*(2), 187–220.

Cox, D. R. (1975). Partial likelihood. *Biometrika 62*(2), 269–276.

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B 39*, 1–38.

Diez, D. M. (2012). R package OIsurv: Survival analysis supplement to OpenIntro guide. http://cran.r-project.org/web/packages/OIsurv/index.html.

Dümbgen, L., S. Freitag-Wolf, and G. Jongbloed (2006). Estimating a unimodal distribution from interval-censored data. *Journal of the American Statistical Association 101*(475), 1094–1106.

Efron, B. (1967). The two sample problem with censored data. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, pp. 831–853. University of California Press.

Efron, B. (1977). The efficiency of Cox's likelihood function for censored data. *Journal of the American Statistical Association 72*(359), 557–565.

Efron, B. (1988). Logistic regression, survival analysis, and the Kaplan-Meier curve. *Journal of the American Statistical Association 83*(402), 414–425.

Eubank, R. L. (1999). *Nonparametric Regression and Spline Smoothing* (2nd ed.). Marcel Dekker, Inc., New York.

Fan, J., M. Farmen, and I. Gijbels (1998). Local maximum likelihood estimation and inference. *Journal of the Royal Statistical Society, Series B: Statistical Methodology 60*(3), 591–608.

Finkelstein, D. M. (1986). A proportional hazards model for interval-censored failure time data. *Biometrics 42*(4), 845–854.

Finkelstein, D. M. and R. A. Wolfe (1985). A semiparametric model for regression analysis of interval-censored failure time data. *Biometrics 41*(4), 933–945.

Fix, E. and J. Hodges, J. L. (1989). Discriminatory analysis, nonparametric discrimination: Consistency properties. *International Statistical Review / Revue Internationale de Statistique 57*(3), 238–247.

Gefeller, O. and H. Dette (1992). Nearest neighbour kernel estimation of the hazard function from censored data. *Journal of Statistical Computation and Simulation 43*(1-2), 93–101.

Gefeller, O. and P. Michels (1992). A review on smoothing methods for the estimation of the hazard rate based on kernel functions. In Y. Dodge and J. Whittaker (Eds.), *Computational Statistics*, pp. 459–464. Physica-Verlag HD.

Gentleman, R. and C. J. Geyer (1994). Maximum likelihood for interval censored data: Consistency and computation. *Biometrika 81*(3), 618–623.

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

Goggins, W. B., D. M. Finkelstein, D. A. Schoenfeld, and A. M. Zaslavsky (1998). A Markov chain Monte Carlo EM algorithm for analyzing interval-censored data under the Cox proportional hazards model. *Biometrics 54*(4), 1498–1507.

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

Gómez, G. and J. Van Ryzin (1992). Estimation of the subsurvival function for time-to-tumor in survival/sacrifice experiments. *Statistics and Probability Letters 13*(1), 5–13.

Gray, R. (1990). Some diagnostic methods for Cox regression models through hazard smoothing. *Biometrics 46*(1), 93–102.

Gray, R. J. (1994). Spline-based tests in survival analysis. *Biometrics 50*(3), 640–652.

Grenander, U. (1956). On the theory of mortality measurement. II. *Skand. Aktuarietidskr. 39*, 125–153.

Groeneboom, P. and G. Jongbloed (2013). Smooth and non-smooth estimates of a monotone hazard. *From Probability to Statistics and Back: High-Dimensional Models and Processes 9*, 174–196.

Groeneboom, P., G. Jongbloed, and J. A. Wellner (2001). Estimation of a convex function: Characterizations and asymptotic theory. *Annals of Statistics 29*(6), 1653–1698.

Groeneboom, P., G. Jongbloed, and J. A. Wellner (2008). The support reduction algorithm for computing non-parametric function estimates in mixture models. *Scandinavian Journal of Statistics 35*(3), 385–399.

Groeneboom, P. and J. Wellner (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation*. DMV Seminar. Birkhäuser Basel.

Habtzghi, D. and S. Datta (2012). One sample goodness of fit tests in presence of shape restrictions on the hazard rate function. *Sankhyā: The Indian Journal of Statistics 74 B*(PART 2), 171–194.

Hall, P., L. S. Huang, J. A. Gifford, and I. Gijbels (2001). Nonparametric estimation of hazard rate under the constraint of monotonicity. *Journal of Computational and Graphical Statistics 10*, 592–614.

Haupt, E. (1992). A new model for a lifetime distribution with bathtub shaped failure rate. *Microelectronics Reliability 32*(5), 633–639.

Haupt, E. and H. Schäbe (1997). The TTT transformation and a new bathtub distribution model. *Journal of Statistical Planning and Inference 60*(2), 229–240.

Henschel, V., H. C. and U. Mansmann (2009). R package intcox: Iterated convex minorant algorithm for interval censored event data. http://cran.r-project.org/web/packages/intcox/index.html.

Herndon, J. E. and F. E. Harrell (1990). The restricted cubic spline hazard model. *Communications in Statistics - Theory and Methods 19*(2), 639–663.

Hess, K. and R. Gentleman (2010). R package muhaz: Hazard function estimation in survival analysis. http://cran.r-project.org/web/packages/muhaz/index.html.

Hess, K. R., D. M. Serachitopol, and B. W. Brown (1999). Hazard function estimators: A simulation study. *Statistics in Medicine 18*(22), 3075–3088.

Hildreth, C. (1954). Point estimates of ordinates of concave functions. *Journal of the American Statistical Association 49*(267), 598–619.

Hjorth, U. (1980). Reliability distribution with increasing, decreasing, constant and bathtub-shaped failure rates. *Technometrics 22*(1), 99–107.

Hougaard, P. (1988). A boundary modification of kernel function smoothing, with application to insulin absorption kinetics. In D. Edwards and N. Raun (Eds.), *Compstat*, pp. 31–36. Physica-Verlag HD.

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

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

Huang, J. and J. A. Wellner (1995b). Estimation of a monotone density or monotone hazard under random censoring. *Scandinavian Journal of Statistics 22*(1), 3–33.

Huang, J. and J. A. Wellner (1997). Interval censored survival data: A review of recent progress. In D. Lin and T. Fleming (Eds.), *Proceedings of the First Seattle Symposium in Biostatistics*, Volume 123 of *Lecture Notes in Statistics*, pp. 123–169. Springer US.

Huang, Y. and C.-H. Zhang (1994). Estimating a monotone density from censored observations. *Annals of Statistics 22*(3), 1256–1274.

Hudgens, M. (2005). On nonparametric maximum likelihood estimation with interval censoring and left truncation. *Journal of the Royal Statistical Society. Series B: Statistical Methodology 67*(4), 573–587.

Hui, R. and H. Jankowski (2011). CPHshape: Find the maximum likelihood estimator of the shape constrained hazard baseline and the effect parameters in the Cox proportional hazards model. http://cran.r-project.org/web/packages/CPHshape/index.html.

Jácome, M. A., I. Gijbels, and R. Cao (2008). Comparison of presmoothing methods in kernel density estimation under censoring. *Computational Statistics 23*(3), 381–406.

Jankowski, H., I. Wang, H. McCague, and J. A. Wellner (2009). R package convexHaz: Nonparametric MLE/LSE of convex hazard. http://cran.r-project.org/web/packages/convexHaz/index.html.

Jankowski, H. K. and J. A. Wellner (2009a). Computation of nonparametric convex hazard estimators via profile methods. *Journal of Nonparametric Statistics 21*(4), 505–518.

Jankowski, H. K. and J. A. Wellner (2009b). Nonparametric estimation of a convex bathtub-shaped hazard function. *Bernoulli 15*(4), 1010–1035.

Joly, P., D. Commenges, and L. Letenneur (1998). A penalized likelihood approach for arbitrarily censored and truncated data: Application to age-specific incidence of dementia. *Biometrics 54*(1), 185–194.

Jongbloed, G. (1998). The iterative convex minorant algorithm for nonparametric estimation. *Journal of Computational and Graphical Statistics 7*, 310–321.

Juckett, D. A. and B. Rosenberg (1993). Comparison of the Gompertz and Weibull functions as descriptors for human mortality distributions and their intersections. *Mechanisms of Ageing and Development 69*(1-2), 1–31.

Kalbfleisch, J. D. and R. L. Prentice (2002). *The statistical analysis of failure time data* (2nd ed.). John Wiley: New York.

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

Kardaun, O. (1983). Statistical survival analysis of male larynx-cancer patients: A case study. *Statistica Neerlandica 37*(3), 103–125.

Klein, J. P. and M. L. Moeschberger (2003). *Survival Analysis: Techniques for Censored and Truncated Data*. Statistics for Biology and Health. Springer.

Koo, J.-Y., C. Kooperberg, and J. Park (1999). Logspline density estimation under censoring and truncation. *Scandinavian Journal of Statistics 26*(1), 87–105.

Kooperberg, C. (2007). R package logspline: Logspline density estimation routines. http://cran.r-project.org/web/packages/logspline/index.html.

Kooperberg, C. and C. J. Stone (1991). A study of logspline density estimation. *Computational Statistics and Data Analysis 12*(3), 327–347.

Kooperberg, C. and C. J. Stone (1992). Logspline density estimation for censored data. *Journal of Computational and Graphical Statistics 1*(4), 301–328.

Kumar, V. and U. Ligges (2011). R package reliaR: Package for some probability distributions. http://cran.r-project.org/web/packages/reliaR/index.html.

Lai, C. D. (2013). *Generalized Weibull Distributions*. Springer.

Lawless, J. (2003). *Statistical Models and Methods for Lifetime Data*. Wiley.

Lawson, C. L. and R. J. Hanson (1974). *Solving Least Squares Problems*. Classics in Applied Mathematics. Society for Industrial and Applied Mathematics.

Lee, E. and J. Wang (2003). *Statistical Methods for Survival Data Analysis*. Wiley Series in Probability and Statistics. Wiley.

Lesaffre, E., A. Komrek, and D. Declerck (2005). An overview of methods for interval-censored data with an emphasis on applications in dentistry. *Statistical Methods in Medical Research 14*(6), 539–552.

Lindsey, J. and L. Ryan (1998). Tutorial in biostatistics. methods for interval-censored data. *Statistics in Medicine 17*(2), 219–238.

López-de Ullibarri, I. and M. A. Jácome (2013a). R package survPresmooth: Presmoothed estimation in survival analysis. http://cran.r-project.org/web/packages/survPresmooth/index.html.

López-de Ullibarri, I. and M. A. Jácome (2013b). SurvPresmooth: An R package for presmoothed estimation in survival analysis. *Journal of Statistical Software 54*(11), 1–26.

Lopuhaä, H. P. and G. F. Nane (2013). Shape constrained non-parametric estimators of the baseline distribution in Cox proportional hazards model. *Scandinavian Journal of Statistics 40*(3), 619–646.

Lu, Z. and W. Zhang (2012). Semiparametric likelihood estimation in survival models with informative censoring. *Journal of Multivariate Analysis 106*, 187–211.

Marshall, A. W. and F. Proschan (1965). Maximum likelihood estimation for distributions with monotone failure rate. *Annals of Mathematical Statistics 36*(1), 69–77.

Meyer, M. C. and D. Habtzghi (2011). Nonparametric estimation of density and hazard rate functions with shape restrictions. *Journal of Nonparametric Statistics 23*(2), 455–470.

Miller, G., J. (1983). What price Kaplan-Meier? *Biometrics 39*(4), 1077–1081.

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

Müller, H.-G. and J.-L. Wang (1994). Hazard rate estimation under random censoring with varying kernels and bandwidths. *Biometrics 50*(1), 61–76.

Murray, S. (2001). Using weighted Kaplan-Meier statistics in nonparametric comparisons of paired censored survival outcomes. *Biometrics 57*(2), 361–368.

Murthy, D. N. P., M. Xie, and R. Jiang (2004). *Weibull Models*. Wiley Series in Probability and Statistics. Wiley.

Mykytyn, S. W. and T. J. Santner (1981). Maximum likelihood estimation of the survival function based on censored data under hazard rate assumptions. *Communications in Statistics - Theory and Methods 10*(14), 1369–1387.

Nadaraya, E. (1964). On estimating regression. *Theory of Probability and Its Applications 9*(1), 141–142.

Nash, J. C. (1979). *Compact numerical methods for computers: Linear algebra and function minimisation*. Wiley.

Nash, J. C. (2011). R package Rvmmin: Variable metric nonlinear function minimization with bounds constraints. http://cran.r-project.org/web/packages/Rvmmin/index.html.

Nassar, M. M. and F. H. Eissa (2003). On the exponentiated Weibull distribution. *Communications in Statistics - Theory and Methods 32*(7), 1317–1336.

Nelson, W. (1972). Theory and applications of hazard plotting for censored failure data. *Technometrics 14*, 945–965.

Ng, M. P. (2002). A modification of Peto's nonparametric estimation of survival curves for interval-censored data. *Biometrics 58*(2), 439–442.

Nielsen, J. P. (2003). Variable bandwidth kernel hazard estimators. *Journal of Nonparametric Statistics 15*(3), 355–376.

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

Padgett, W. J. and L. J. Wei (1980). Maximum likelihood estimation of a distribution function with increasing failure rate based on censored observations. *Biometrika 67*(2), 470–474.

Pan, W. (1999). Extending the iterative convex minorant algorithm to the Cox model for interval-censored data. *Journal of Computational and Graphical Statistics 8*(1), 109–120.

Pan, W. (2000a). A multiple imputation approach to Cox regression with interval-censored data. *Biometrics 56*(1), 199–203.

Pan, W. (2000b). Smooth estimation of the survival function for interval censored data. *Statistics in Medicine 19*(19), 2611–2624.

Pan, W. and R. Chappell (1998). Estimating survival curves with left-truncated and interval-censored data under monotone hazards. *Biometrics 54*(3), 1053–1060.

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

Prakasa Rao, B. L. S. (1969). Estimation of a unimodal density. *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002) 31*(1), 23–36.

Proschan, F. (1963). Theoretical explanation of observed decreasing failure rate. *Technometrics 5*, 375–383.

Ramlau-Hansen, H. (1983). Smoothing counting process intensities by means of kernel functions. *Annals of Statistics 11*(2), 453–466.

Reboul, L. (2005). Estimation of a function under shape restrictions: Applications to reliability. *Annals of Statistics 33*(3), 1330–1356.

Rice, J. and M. Rosenblatt (1976). Estimation of the log survivor function and hazard function. *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002) 38*(1), 60–78.

Rinne, H. (2008). *The Weibull Distribution: A Handbook*. Taylor & Francis.

Robertson, T., F. Wright, and R. Dykstra (1988). *Order Restricted Statistical Inference*. Probability and Statistics Series. Wiley.

Rosenberg, P. S. (1995). Hazard function estimation using B-splines. *Biometrics 51*(3), 874–887.

Satten, G. A. (1996). Rank-based inference in the proportional hazards model for interval censored data. *Biometrika 83*(2), 355–370.

Self, S. G. and E. A. Grossman (1986). Linear rank tests for interval-censored data with application to PCB levels in adipose tissue of transformer repair workers. *Biometrics 42*(3), 521–530.

Senthilselvan, A. (1987). Penalized likelihood estimation of hazard and intensity functions. *Journal of the Royal Statistical Society, Series B 49*(2), 170–174.

Shen, P. (2011). Nonparametric estimation with doubly censored and truncated data. *Computational Statistics 26*(1), 145–157.

Shen, P. (2012). Nonparametric analysis of interval censored and doubly truncated data. *Journal of Statistical Computation and Simulation 82*(12), 1845–1854.

Silverman, B. (1986). *Density Estimation for Statistics and Data Analysis.* Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis.

Simonoff, J. (1996). *Smoothing Methods in Statistics.* Springer Series in Statistics. Springer.

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

Stone, C. J. (1987). Uniform error bounds involving logspline models. In *Probability, statistics, and mathematics.* Department of Statistics, University of California.

Stone, C. J. (1990). Large-sample inference for log-spline models. *Annals of Statistics 18*(2), 717–741.

Stone, C. J., M. H. Hansen, C. Kooperberg, and Y. K. Truong (1997). Polynomial splines and their tensor products in extended linear modeling. *Annals of Statistics 25*(4), 1371–1425.

Sun, J. (2006). *The Statistical Analysis of Interval-censored Failure Time Data.* Statistics for Biology and Health Series. Springer.

Tanner, M. A. and W. H. Wong (1983). The estimation of the hazard function from randomly censored data by the kernel method. *Annals of Statistics 11*(3), 989–993.

Therneau, T. and T. Lumely (2009). R package survival: Survival analysis, including penalised likelihood. http://cran.r-project.org/web/packages/survival/index.html.

Tibshirani, R. and T. Hastie (1987). Local likelihood estimation. *Journal of the American Statistical Association 82*(398), 559–567.

Tsai, W.-Y. (1988). Estimation of the survival function with increasing failure rate based on left truncated and right censored data. *Biometrika 75*(2), 319–324.

Tsiatis, A. A. (1981). A large sample study of Cox's regression model. *Annals of Statistics 9*(1), 93–108.

Tsybakov, A. B. (2004). *Introduction à l'estimation non-paramétrique.* Springer-Verlag.

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

Wand, P. and C. Jones (1995). *Kernel Smoothing.* Monographs on statistics and applied probability. Chapman and Hall, London.

Wang, F. K. (2000). A new model with bathtub-shaped failure rate using an additive BurrXII distribution. *Reliability Engineering and System Safety 70*(3), 305–312.

Wang, L., J. Sun, and X. Tong (2010). Regression analysis of case ii interval-censored failure time data with the additive hazards model. *Statistica Sinica 20*(4), 1709–1723.

Wang, Y. (2007). On fast computation of the non-parametric maximum likelihood estimate of a mixing distribution. *Journal of the Royal Statistical Society, Series B 69*(2), 185–198.

Wang, Y. (2008). Dimension-reduced nonparametric maximum likelihood computation for interval-censored data. *Computational Statistics and Data Analysis 52*(5), 2388–2402.

Wang, Y. (2010). Maximum likelihood computation for fitting semiparametric mixture models. *Statistics and Computing 20*(1), 75–86.

Wang, Y. and S. M. Taylor (2013). Efficient computation of nonparametric survival functions via a hierarchical mixture formulation. *Statistics and Computing 23*(6), 713–725.

Watson, G. (1964). Smooth regression analysis. *Shankyā 9*(26), 359–372.

Watson, G. S. and M. R. Leadbetter (1964). Hazard analysis II. *Sankhyā: The Indian Journal of Statistics, Series A 26*(1), 101–116.

Wellner, J. A. and Y. Zhan (1997). A hybrid algorithm for computation of the nonparametric maximum likelihood estimator from censored data. *Journal of the American Statistical Association 92*(439), 945–959.

Whittemore, A. S. and J. B. Keller (1986). Survival estimation using splines. *Biometrics 42*(3), 495–506.

Yan, J. (2010). R package KMsurv: Data sets from Klein and Moeschberger (1997), survival analysis. http://cran.r-project.org/web/packages/KMsurv/index.html.

Yu, Q., L. Li, and G. Y. C. Wong (2000). On consistency of the self-consistent estimator of survival functions with interval-censored data. *Scandinavian Journal of Statistics 27*(1), 35–42.

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

Zhang, Y., L. Hua, and J. Huang (2010). A spline-based semiparametric maximum likelihood estimation method for the Cox model with interval-censored data. *Scandinavian Journal of Statistics 37*(2), 338–354.

Zhang, Y. and M. Jamshidian (2004). On algorithms for the nonparametric maximum likelihood estimator of the failure function with censored data. *Journal of Computational and Graphical Statistics 13*(1), 123–140.

Zhang, Z. and J. Sun (2010). Interval censoring. *Statistical Methods in Medical Research 19*(1), 53–70.